Characterization of A Ribose Metabolism Pathway in *Bacteroides thetaiotaomicron* and New Insights into the Nutrients Degraded by this Bacterium

by

Robert W.P. Glowacki

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
at the University of Michigan
2020

Doctoral Committee:

Associate Professor Eric C. Martens, Chair
Professor Matthew R. Chapman
Assistant Professor Nicole K. Koropatkin
Professor Thomas M. Schmidt

Robert W.P. Glowacki

rwglowac@umich.edu

ORCID iD: 0000-0001-7718-5056

Dedication

There is one person who stands out in my mind that, without her guidance, I would not be where I am today. Therefore, I am dedicating my dissertation to the memory of Dr. Jody Modarelli. Jody was my undergraduate mentor and head of the Biochemistry department at Hiram College. When I began my college journey, I was undecided in major field of study, but leaning towards political science. Well, the way my liberal arts education was structure allowed for the crossing of our paths via a semester long colloquium in my first year. I chose to take "The molecular basis of infectious disease", having given few thoughts to a science major, Jody made learning about the science of diseases fun and engaging. So, at my first year advising meeting, Jody convinced me that I needed to be a biochemistry major and that she would support me in any way possible throughout my undergraduate education and beyond. She truly exemplified the joy that comes with discoveries in science and had a passion for teaching that I don't know I've experienced since taking her classes. Jody was the first person to make me truly believe that I had what it takes to be a scientist and her countless letters of recommendation allowed me to get research experience while an undergraduate. Her reference letter undoubtedly aided greatly, my entrance into the University of Michigan PIBS program. Halfway through the first semester of biochemistry Jody was diagnosed with pancreatic cancer. Of course, Jody being Jody, she would not let this get her down, she continued to lecture and mentor and live the happiest and fullest life possible. Sadly, Jody passed away not long after I had decided on my thesis lab. Prior to her death and in failing health, she emailed me just weeks before her passing to ask how graduate school was treating me and to offer advice on how to survive. Her memory, and in some ways her life, lives on in the trainees that she advised and inspired. Jody continues to inspire me and I strive to be as wonderful of a mentor and researcher as she was. I truly would not have been here today if not for the guidance and support of Dr. Jody Modarelli.

Acknowledgements

I would first like to thank my thesis advisor Dr. Eric Martens for his support and guidance in developing my research skills and for career development during the extent of my doctoral training. He has taught me the knowledge required to become an independent scientist and how to formulate testable and interesting hypothesis. He has always been a proponent of going where the data leads to build a better, more complete picture of underlying biological mechanisms. Most importantly, he has always been understanding when science has dealt setbacks. Eric, thank you for providing me the research and writing skills necessary to complete this body of work and for allowing me to pursue my teaching and outreach interests.

I would also like to thank my thesis committee members, Dr. Nicole Koropatkin, Dr. Matthew Chapman, and Dr. Thomas Schmidt for helpful guidance throughout the years and for their commitment to my training and career goals.

I also appreciate the Department of Microbiology and Immunology as a whole for being an excellent resource for reagents, tools, and scientific discussion. I want to call special attention to Dr. David Friedman for providing plasmids and strains as well as being an excellent conversationalist for both science, and general discussions. Further, I want to recognize the unwavering support of the department administrators who ensure everything in the background runs smoothly. Beyond the department, the staff of the Germ Free Facility have been invaluable in completion of the work within this dissertation. Additionally, I would like to thank the collaborators in other departments at the University of Michigan and Dr. Anton Terekov and Dr. Bruce Hamaker of the Department of Food Science at Purdue University for running samples and training me in analytical techniques.

This work would not have been possible if not for the generous support of the Rackham Merit Fellowship, the Molecular Mechanisms of Microbial Pathogenesis training grant and the support of NIH grants.

One of the most important groups of people I need to thank is my lab family. Getting through this journey would not have been possible without their support and encouragement and helpful troubleshooting of failed experiments. Although many individuals have come through the lab, each helping me along the way, Nick Pudlo stands out as the person who I owe the most gratitude. Nick is our lab manager, our science Sherpa, the labs motivational speaker, and guidance counselor, all while being a fantastic and easy going person who is generally fun to hang out with, and for these qualities I must say thanks for everything you do Nick for me, the lab, and the trainees within the Martens lab.

Outside of lab, I have had the pleasure of having many friends that I met during my first year at Michigan, all from a wide range of departments under the PIBS umbrella. These friends have gotten me through a lot of tough times, provided so many laughs and memories, and were even on board when I formed a softball team from a motley crew of graduate students. My friends have made being in Michigan a much more enjoyable experience and I know, our forged friendships are lifelong.

Finally, I would not be here without the love and support of my family. As a first generation college graduate and now a soon-to-be first generation PhD, my mother, father, and step-mother have always encouraged me to do the best that I could, work hard, and stay strong. To my fiancée Jessica Cory, if it were not for your love and the support of your family, the last few years would have been so much harder and way less fun. Thank you for putting up with my late nights, my science discussions over dinner, your willingness to proofread my jargon-laden writing, and especially for being my rock through the last few hectic months.

Table of Contents

List of Tables

List of Figures

Abstract

Bacteria of the *Bacteroidetes* phylum, dominant members within the gut microbiota, devote large genomic capacity towards nutrient acquisition via gene clusters termed polysaccharide utilization loci (PULs). The model organism, *Bacteroides thetaiotaomicron* (*Bt*) contains 88 PULs that target complex polysaccharides of host- microbial- or dietary origin. However, many PULs remain uncharacterized in terms of cognate substrate, enzyme functionality, and regulation. I have expanded the known substrates targeted through characterizing the ribose utilization system (*rus*) PUL in *Bt.* I created gene deletions based on predicted functionality within *rus.* Using these strains allowed for *in vitro* characterization of the substrates (*e.g.* ribose, nucleosides and RNA) that are catabolized through this PUL. The ability to access these nutrients confers a competitive advantage *in vivo* on a fiber-rich diet containing nucleosides. Additionally, through biochemical and *in vivo* studies I have connected the actions of a genomically unlinked nucleoside phosphorylase (BT4554) and the *rus* ribokinases (RusK1/K2). Determining that these two enzymes work together by BT4554 cleaving nucleosides which produces ribose-1-phosphate (R1P), which is subsequently phosphorylated by RusK1/K2 yielding ribose-1,5-bisphosphate (PRibP). Further, RusK2 accepts ribose-5-phosphate (R5P) as a substrate and synthesizes PRibP by phosphorylating the 1'C position. The functions displayed by RusK1 and RusK2 are the first described in eubacteria generating PRibP from R1P or R5P, and represents new metabolism in *Bt*. Further, the ability of *Bt* to sense ribose transcriptionally alter genes located within other PULs and loci.

Contrastingly, to the *rus* PUL, mucin-*O*-glycan (MOG) PULs are strongly upregulated *in vivo* on a fiber-free diet (FF diet); a condition where *Bt* relies on host-derived glycans for growth. This FF diet resembles Westernized human diets that have been implicated in inflammatory bowel disorders (IBD) leading to colitis by bacteria eroding the host mucosa. By deleting MOG-responsive, sulfatase-encoding PULs in *Bt* as single PUL deletions and sequentially, (up to a strain lacking 10 PULs), I abrogated growth of *Bt* on MOG. This approach has assisted in narrowing the gene-encoded functions responsible for disease. Additionally, using a transposon mutagenesis screen, I was able to discover a *Bt*-specific T-cell epitope

recognized *in vivo* during disease. The expression of this epitope is affected both by glucose and salt concentrations, demonstrating even more the interesting and largely unknown regulatory strategies employed by *Bt.*

The regulatory network in *Bt* is complicated, with each PUL encoding its own regulatory protein (ECF-σ/anti-σ proteins, hybrid two-component systems, etc.). Additionally, *Bt* encodes 22 ECF-σ proteins as well as 4 LacI-type regulators not associated with known metabolic loci, making them orphan regulatory proteins. I have deleted most of these genes, resulting in discovery of a single ECF-σ gene, *BT2492*, which when deleted, reduces growth on 12 of the polysaccharides *Bt* degrades. Further, two LacI deletion strains result in drastically improved growth on normally low-priority monosaccharides. Lastly, as suggested by *in vitro* RNAseq data of ribose growth, the presence of ribose affects priority of other nutrients. This phenomenon extends to other simple sugars as arabinose and xylose RNAseq data reveal that they also exert changes in gene expression for loci not associated with their catabolism, including orphan ECF-σ factors. Together these data point to a complex regulatory cascade through a multi-faceted system involving PUL-encoded activators, *trans*-encoded proteins, and sugar-dependent prioritization through these mechanisms.

Chapter I

Introduction

The importance of resident intestinal microbes to human health has been appreciated since the 1880s beginning with Theodor Escherich's investigation of fecal bacteria, including *Escherichia coli,* and his hypothesis that these indigenous micoorganisms play roles in both digestion and intestinal diseases[1]. Since then, our understanding of the bacteria, viruses, archaea and eukaryotes that inhabit the gut has expanded alongside the rest of the field of microbiology and numerous fundamental roles have been established for this community, now termed the microbiome. As speculated by Escherich, these roles definitively include nutrient digestion[2,3] and protection from invading pathogens[4], but also extend to short- and long-term instruction of the immune system[5-7] and production of a wide range of metabolites that could not be produced by human physiology. While the gut microbiome is typically described as being composed of non-pathogenic, "commensal" organisms, it is now appreciated that both individual species[8] or multiple community members acting together[9,10] can exert pathogenic effects, which are often more subtle than those of classical pathogens. Indeed, the presence of common microorganisms with discrete virulence factors (*e.g.*, enterotoxins, genotoxins) that may only manifest in diseases like colorectal cancer or inflammatory bowel disease (IBD) over long time periods or in certain host genetic backgrounds obscures the definition of pathogen.

Accelerated in the 2000's by the "-omics" revolution, along with a recent resurgence of cultivation[11-13], countless studies in the past two decades have implied or established connections between altered gut microbiomes and many diseases. These studies have demonstrated the malleability (or fragility) of the microbiome in the face of environmental and dietary perturbations encompassing antibiotic use[14], geography[15], immigration[16], and dietary changes, including fiber deprivation[17,18]. While Escherich's original ideas were logically predicted with respect to microbiome effects in the gut, less anticipated connections between gut microbes and health have extended to neurobiology[19-21] and systemic immune responses that impact allergy[22]. Emerging studies, often extending from -omics based observations, are providing causal or

mechanistic understanding of the relationships that connect host responses with microbiome-derived metabolic functions. Here I look at recent examples that illustrate how the gut microbiome can augment or perturb host physiology through complementary or novel metabolism often changing the outcome of disease trajectories. The studies I highlight provide details that underscore the importance of gut microbes in human health, which Escherich postulated long ago.

The impact of gut bacterial metabolites on host physiology

The collective diversity of microbial species that compose the gut microbiome harbor at least ~10 million unique, annotated genes[23]—probably many more[24]—that are not present in the human genome. Through our individual microbiomes, each of us has a personalized subset of this gene repertoire that substantially exceeds the genes in our human genome. Our microbiomes are equipped to produce an astonishing array of microbiome-produced products (MPPs), metabolites, and cellular products like polysaccharides and curli fibers that in many cases do not remain confined to the gut. The impacts of specific MPPs, and the presence/absence of individual species/strains that produce them, have been implicated in a wide-range of diseases both in and out of the gastrointestinal tract (Figure 1.1). Effects in the gut include preventing pathogen invasion through bile salt modifications[25] and mucus layer erosion when the host lacks dietary fiber[26]. More surprisingly, studies have drawn connections to neurological conditions such as Parkinson's Disease (PD)[27], depression[28,29], and autism spectrum disorders (ASD)[30-32], showing that certain bacteria and their MPPs (*e.g.*, curli fibers in PD and the metabolites, 4-ethylphenylsulfate, *p*-cresol, taurine, and 5-aminovaleric acid in ASD) can contribute to these states (Figure 1.1).

A recent study using a forward chemical genetics culture-based screen showed that MPPs from several dozen bacteria promote direct interactions with G-Protein Coupled Receptors (GPCRs), a wide class of host receptors important in many aspects of physiology including mood regulation, immune function and the autonomic nervous system, such as peristalsis of the digestive tract[33]. This included a strain of *Morganella morganii* converting L-phenylalanine into phenethylamine, a psychoactive compound that can be fatal in individuals taking monoamine oxidase inhibitor drugs[33]. Studies have also shown that bacteria that

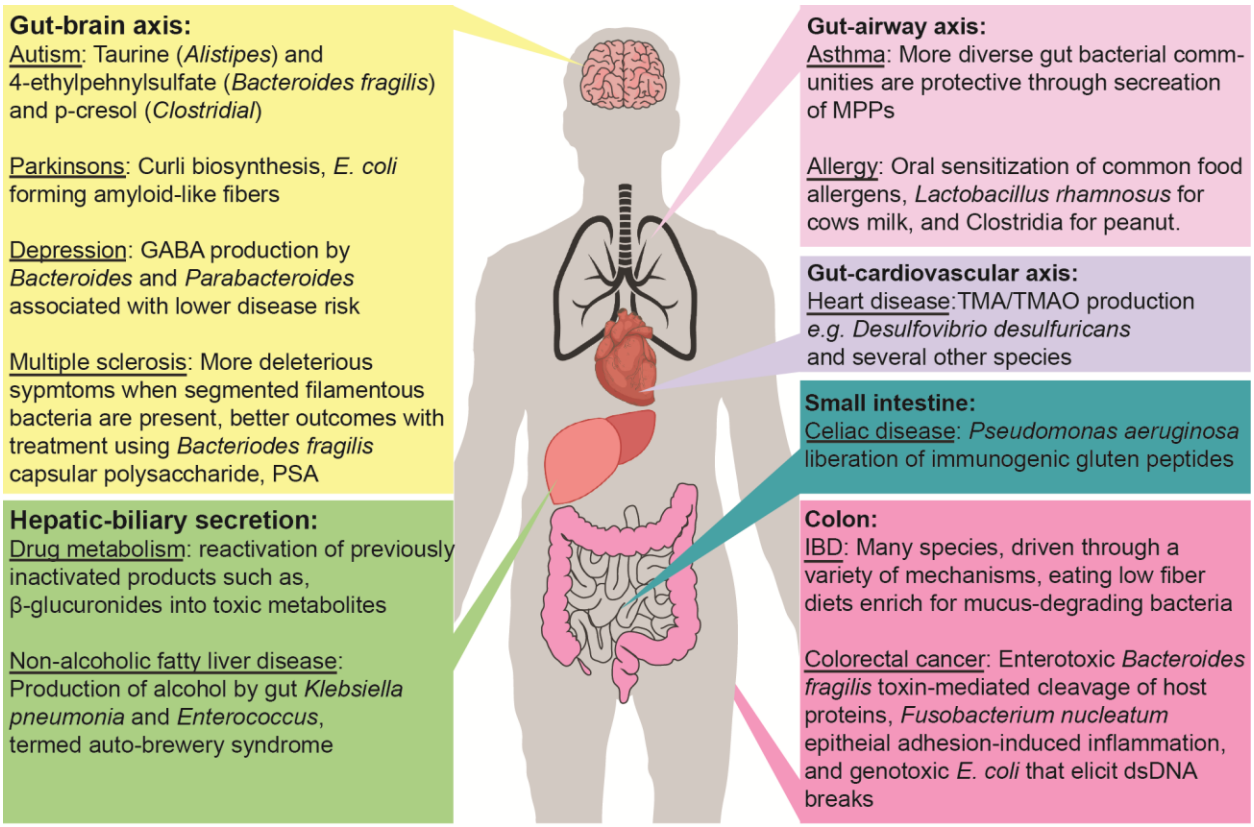


Figure 1.1 Effects of the gut microbiome on host health.

A diagram highlighting many of the known effects of the gut microbiome within various organ systems. Each of the callout boxes highlights a different organ site and within each box several examples of diseases with an emerging connection to the gut microbiota is described. Abbreviations not defined in the figure in order of appearance are: GABA, gamma-aminobutyric acid, TMA/TMAO, Trimethylamine N-oxide, and IBD, inflammatory bowel disease, a collection of several intestinal disorders that includes Crohn's disease and ulcerative colitis. References for each noted association: autism[21,32], Parkinson's[34,35], depression[28], multiple sclerosis[36], drug metabolism[37,38], non-alcoholic fatty liver disease[39,40], asthma[41], allergy[42,43], heart disease[44], Celiac disease[45], IBD[26], colorectal cancer[46-49].

convert tryptophan to tryptamine stimulate the colonic-restricted GPCR, $5HT_4R$, resulting in increased intestinal transit time[50]. Additionally, bacterial production of *N*-acyl amides regulate glucose homeostasis and possibly appetite[51]. MPPs also interact with other receptors, such as the aryl-hydrocarbon receptor (AhR), for example by production of the Ahr ligand indole 3-aldehyde by *Lactobacillus reuteri* leading to increased IL-22 production and increased mucosal immune responses against *Candida albicans*[52]. Studies like the ones noted above often reveal beneficial and detrimental effects from variants of the same species, suggesting strain-level differences in the accessory genome mediate these effects. This is an important consideration when formulating potential probiotics or other bacterial-based treatments or therapies. A recent example of this is implication of *L. reuteri* (strain SP-C2-NAJ0070) as an exacerbator of systemic lupus erythematosus (SLE) symptoms in a TLR7-dependent manner[53].

Another class of molecules, which have previously been well-studied in pathogenic bacteria, the cyclic di- and trinucleotides (CDNs/CTNs), are also emerging as molecules that interact with host sensors. The structural diversity of these compounds has expanded from purine-based to include pyrimidine-based examples[54]. While not definitively linked to aspects of host health, some of these CDNs which can be sensed by the host after bacteriolysis from antibiotics or host response can activate host immune pathways through pattern recognition receptors (PRRs), such as Stimulator of Interferon Genes (STING) and Reductase Controlling NF-κB (RECON) protein. Homologs of CDN synthesis operons are widespread in both commensal and pathogenic bacteria, including the prevalent *Bacteroides* genus. A recent study suggests that bacteria have evolved new ways of evading/enhancing host PRR recognition through synthesis of unique CTNs or modified CDNs not efficiently sensed by host PRRs[54].

A final group of MPPs that is just beginning to be explored is bacterial capsular polysaccharides (CPS), which are enriched and highly diversified in several lineages of gut bacteria[55]. For example, just 14 sequenced strains of the common Gram-negative symbiont *Bacteroides thetiotaomicron* harbor 47 different configurations of gene clusters for producing CPS[56]. A subset of zwitterionic CPS, first discovered in *Bacteroides fragilis* but present in other species, has immunomodulatory properties, as do CPS and extracellular polysaccharides produced by members of different phyla, the Bifidobacteria[57], Proteobacteria[58] and Firmicutes[59,60]. These bacterial surface coatings are likely to be under intense pressure to diversify their glycan structures, perhaps to evade host immune responses, bacteriophages and microbe-mediated killing. In the process, they have fortuitously synthesized chemical structures that interact with the host epithelium and immune system (Figure 1.1), providing additional advantages during colonization and also opportunities for researchers to exploit these molecules for potential drug development[61].

Collectively, the studies highlighted above illustrate how host cells have evolved to sense and interact with a variety of metabolites that are uniquely microbial, which is the basis of much innate immune recognition and of central importance in the tolerance of the dense human gut microbiome[62]. Better understanding these interactions may prove helpful in leveraging these existing chemical relationships to design new drugs that alter immune responses or other aspects of host cellular biology.

Metabolism of drugs and other xenobiotics by gut microbes

Just as members of the microbiome produce novel molecules that interact with human physiology, they also have the capacity to modify exogenous chemicals (xenobiotics), many of which are the drugs used to treat diseases. Two of the most prominent examples are inactivation of the cardiac drug digoxin by *Eggerthella lenta* (*E. lenta*)[63], and related plant-derived cardenolides[64], and conversion of the common dietary compound choline to trimethylamine (TMA), which is subsequently converted by the host to harmful trimethylamine-*N*-oxide (TMAO) (Figure 1.1) that promotes cardiovascular disease[44,65,66]. Another process that has been characterized mechanistically is drug reactivation following β-glucuronic acid conjugation in the liver and biliary secretion back into the gut. This process is catalyzed by gut bacterial β-glucuronidases, which are widely present in gut bacteria[37] and have broad substrate specificities[38,67], which allow them to reactivate toxic drugs like the chemotherapeutic irinotecan. This process may be circumvented by drugs that, in turn block, β-glucuronidases to halt drug re-toxification.

More recent examples highlight how the gut commensal *Bacteroides thetaiotaomicron* (*Bt*) and related Bacteroidetes metabolize a range of xenobiotics using previously undescribed mechanisms. One of these involves degradation of the nucleoside-based antiviral drugs brivudine and sorivudine to the hepatotoxic compound bromovinyluracil (BVU) through the action of a nucleoside phosphorylase[68]. Homologs of this gene are found in many members of the phylum, suggesting that toxic BVU could accumulate at faster rates based on which members of the microbiota are present. Another study expanded the repertoire of drugs that can be metabolized by *Bt*, identifying 18 drugs that are modified by an additional 17 unique enzymes[69]. Further highlighting that multiple bacteria can work synergistically in the gut, a recent study discovered a pathway for enzymatic conversion and inactivation of the Parkinson's drug, levodopa (L-dopa). This step-wise mechanism involves *Enterococcus faecalis*, which first decarboxylates L-dopa to active dopamine, followed by an uncommon enzymatic dehydroxylase from *E. lenta* that inactivates L-dopa and produces *m*-tyramine[70]. These studies and several others like it point to variations in the gut microbiome as an often overlooked reason why therapy fails or patients have intolerable side-effects to treatments. Thus, the microbiota is another factor that needs consideration during treatment of disease, which may eventually require both sequencing and

culture/biochemistry-based approaches.

Beyond commensal bacteria altering the effects of therapeutic drugs, recent studies involving *Clostridium difficile* (*Cd*) have potentially uncovered an indirect link as to why patients taking common calcium supplements, NSAIDs, and proton-pump inhibitors may be predisposed to infection or have a more severe outcome once infected with *Cd.* The germination signal(s) for *Cd* is known to be intestinal bile salts, with co-germinates such as taurocholate and glycine. However, recent studies *in vitro*[71] and *in vivo*[72] have identified a role for both host-derived (endogenous) and more importantly, dietary supplements or vitamins (exogenous) Ca^{2+} from the host. The presence of Ca^{2+} circumvents the requirement for glycine and suggests a plausible mechanism for why individuals with impaired Ca^{2+} absorption (high levels of intestinal Ca^{2+}) are at greater risk of *C. diff infection* (CDI). Beyond predisposition to infection through calcium effects, NSAIDs were recently shown to alter the community structure of the microbiota potentially creating an environment where CDI is more severe[73]. Although the study only examined responses to the NSAID indomethacin, dysregulation of intestinal tight junctions was observed leading to more severe disease through translocation of *Cd* across the epithelium. Although no direct drug-*Cd* interaction was uncovered, the observation that the abundance of other strains is altered implies that they are affected by, or act on, this drug and in turn allow for invasion and infection by *Cd*. Findings such as the ones described above can be leveraged to design tools to guide drug selection and therapeutic interventions. A recent study detailed a new tool developed to model *in silico* interactions between drug classes and bacterial enzymes with activities against these drugs[74]. This approach was used to successfully predict three previously unknown xenobiotic metabolic pathways by gut microbes that were confirmed through *in vitro* studies[74]. It is likely that in the future, personalized medicine approaches will utilize similar predictive tools coupled with *in vitro* and *in vivo* models to guide treatment regimes in a myriad of diseases and states of health.

A way forward in the search for better therapeutics: detailed mechanistic studies

From the studies highlighted here, the picture of commensal and mutualistic bacteria always being “neutral” or “beneficial” to host biology is almost certainly wrong. Rather, commensals, and even mutualists, may also have potential to exhibit pathogenic activities, albeit in more subtle ways. While true pathogens are equipped with toxins and machinery that directly

damages cells, our non-pathogenic symbionts may not be as directly insidious. The means by which these commensal organisms exhibit pathogenic tendencies are highly context dependent on factors such as diet, drug intake and production of MPPs. Further, when considering if the presence of a species is beneficial or detrimental based on approaches like metagenomics or 16S approaches, the functional or accessory genome and not just phylogeny needs to be considered as strain level variations cause different outcomes. These pathogenic, condition-specific activities of commensal bacteria may have both transient (acute) and chronic (long term) health effects that likely influence disease states across organ systems. Moving forward, personalized medicine will need to consider these microbiome variations and incorporate deeper screening methodologies and functional studies. Leveraging the results of these approaches will hopefully generate new interventions that either prevent or cure the deleterious effects of the microbiome on diseases. In order to achieve these personalized medicine goals, engineering of the microbiome will need to be guided by in-depth mechanistic studies of the organisms that compose this community. A logical starting point is to examine the wealth of physiological knowledge available for prominent members such as the human gut *Bacteroides,* which target polysaccharides of diverse structure and origin.

If you eat it, or secrete it, they will grow: The expanding cornucopia of nutrients that gut bacteria, especially *Bacteroides* utilize

Successful bacterial inhabitants within the gut, or any other ecosystem, need to be able to adapt to changing nutrient conditions if they are to persist. This means for the *Bacteroides* and members of other phyla, the requirement to encode many different metabolic loci that equip them to sense and respond to a variety of endogenous host-, dietary-, and bacteria-derived carbohydrate nutrients. Here, I focus mainly on the *Bacteroides*, describing recent advances in understanding polysaccharide utilization loci (PULs), the mechanisms of the multi-protein systems they encode, regulation, and expanding substrate diversity. I also briefly describe important studies involving marine *Bacteroidetes,* as some of the PULs found in these organisms have been naturally transferred to gut-dwelling *Bacteroides*. We highlight that previously under-considered substrates such as monosaccharides and Maillard reaction products can also affect the gut microbiota. Since many invading pathogens preferentially utilize these nutrients, they may represent nutrient niches competed for by commensals and pathogens. Additionally, we mention

recent work on other important gut species and the strategies they employ to access nutrients in the gut. Finally, given constantly expanding examples of the importance of the gut microbiota to human health and disease, we showcase advances in the field of synthetic biology, engineering, and manipulation of key members of the gut microbiota. The tools and strategies reviewed here may one day help to construct synthetic, altered microbiota-based therapeutics for the promise of engineering the microbiome to modulate host health during infection and disease.

The importance of the gut microbiome of humans and animals has been realized in some capacity for several decades. New associations between disease states and microbiome alterations, which are most often characterized by changes in the abundance of certain microbes, are constantly emerging. While microbial abundance changes are not necessarily causal to disease, studies describing functional and mechanistic relationships are becoming increasingly more frequent[26,44,53,61]. A key theme underlying the persistence of many gut bacteria is their ability to utilize carbohydrate-based nutrients, with much of the focus to date on the prominent *Bacteroidetes* phylum and their ability to assimilate carbon from polysaccharides that often contain numerous covalently linked sugars and are complex in structure and linkage. This ability in *Bacteroidetes* is accomplished via the concerted actions of proteins encoded in Polysaccharide Utilization Loci (PULs). Previous studies have highlighted the broad abilities of bacteria in this phylum to catabolize diverse classes of polysaccharides from host mucosal glycans, dietary fibers, other microbes (capsules and exopolysaccharides), fungi and less traditionally consumed sources such as algal polysaccharides obtained from eating seaweed[75-82]. The ability of gut bacteria to utilize complex carbohydrates has been reviewed several times[2,3,83]. However, a number of recent studies have expanded the known repertoire of polysaccharides and other nutrients, such as simpler carbohydrates, that are targeted by gut bacteria and the enzymatic mechanisms responsible for their breakdown. Some of these, such as the effects of uncommon nutrients derived from Maillard reaction products or the ingestion of food preservatives may have unexpected consequences. While the regulation of individual loci by dedicated transcriptional factors has been studied for several systems, additional global regulation strategies controlling their prioritized expression in complex nutrient environments and potential co-regulation with other nutrient utilization systems have been described. One of the motivations for studying these nutrient systems is the possibility of engineering novel functions into bacteria so that they can be replaced or controlled in ecosystems such as the human gut and agricultural

livestock. Here we review recent findings of the mechanism and regulation of gut bacterial nutrient degradation capabilities, with a focus on newly identified substrates that Gram-positive and Gram-negative bacteria have adapted to forage and some emerging aspects of their cellular regulation. We approach this review with the idea that complex carbohydrates are of ubiquitous importance in shaping the ecology and physiology of gut microbes and, therefore, are a convenient lever to intentionally manipulate these communities.

Mechanisms of Polysaccharide Utilization of Human Gut Microbes

Within the human gut microbiota (HGM), members of the Gram-negative *Bacteroidetes* phylum compose a substantial portion of all bacteria present, while members of the *Bacteroides* genus frequently make up the majority of this phylum[84]. As such, major work has concentrated on understanding the metabolic abilities of the *Bacteroides* who devote large portions of the genome to polysaccharide catabolism, a feat that is accomplished through regulated expression of PULs[85]. These gene clusters usually encode all of the functions required to sense, import, and degrade polysaccharide substrates or their products. These mechanisms can often be categorized as "selfish, sharing, or scavenging" depending on how much of the target substrate is primarily degraded by the producing bacterium or released for other bacteria to utilize (Figure 1.2). The first described system, the Starch Utilization System (Sus) has served as an archetype for characterizing and discovering new PULs[86]. The original definition of a PUL required at least one set of homologs of outer membrane (OM) TonB-dependent transporter (TBDT), *susC*, and OM glycan-binding protein, *susD*[87]. PULs also typically encode two—often many more—Carbohydrate Active Enzymes (CAZymes)[88], and additional carbohydrate binding proteins, regulators, and other enzymes (kinases, proteases, sulfatases and transporters)[89,90]. Degenerate or incomplete PULs also clearly exist, challenging the original definition to some degree.
While I do not provide an in-depth description or catalog of specific PULs, which have already been reviewed[2,83,91], recent studies on substrates and the bacterial mechanisms of glycan degradation are primarily considered. As an example, the well-studied *B. thetaiotaomicron* (*Bt*) Sus locus described over 30 years ago[92], continues to provide new insight into the function of PUL-encoded proteins with recent mechanistic characterization of the role of the accessory binding protein SusE, which has distinct function compared to homologous SusF[93]. *In vitro* live-cell imaging has also demonstrated substantial variability of the surface mobility of Sus proteins

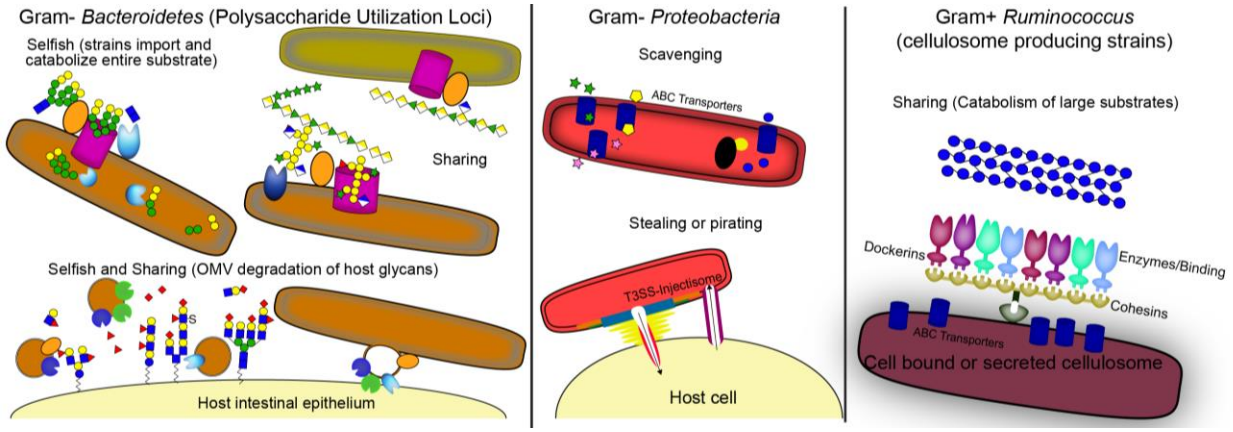


Figure 1.2 Selfish, Sharing, Scavenging: Different means to an end for carbohydrate utilization.

Schematic highlighting several of the major mechanisms of carbohydrate utilization for prominent gut bacteria from the Gram-negative Bacteroidetes and Proteobacteria, and the Gram-positive Ruminococcus. In the left panel the Bacteroides PUL-encoded Sus-like systems are depicted (objects in figure: purple barrel is SusC-like outer membrane (OM) transporter, orange is SusD-like OM binding protein, enzymes are shown as Pacman-like ovals, green ovals are sulfatases, blue ovals are glycoside hydrolases. Bacteroides use several mechanisms to access nutrients including sharing, where one strain partially degrades a polysaccharide substrate and a related species can import and catabolize leftover, liberated oligosaccharides. Further, selfish mechanisms where the bacterium imports the entire substrate and breaks it down in the periplasm and cytoplasm. Lastly, Bacteroides produce outer membrane vesicles that target both host and fiber-derived glycans which can be both selfish and sharing in the mechanism of attack (lower, left panel). Although all of the bacteria shown can use a scavenging mechanism, whereby they import small mono- or disaccharides, the proteobacteria rely mainly on this method of obtaining nutrients and is shown in the middle panel with ABC transporters shown in blue. In the cytoplasm of the cell, common proteins such as kinases and isomerases are shown in black and yellow. In the lower section, a stealing or pirating mechanism is shown that was recently described in E. coli via a type III secretion system/injectisome where the bacterium delivers effector proteins or toxins into host cells, forcing nutrients into an adjacent tube that shuttles carbohydrates or amino acids back to the bacterium. In the right panel, a complex, multi-modular and sometimes extracellular complex produced by Gram-positive Ruminococcus and other related Gram-positive bacteria is shown. The cellulosome is composed of cohesins and dockerins which form the backbone of the complex and then each dockerin can pair with specific enzymes or carbohydrate-binding proteins that work in concert to degrade polysaccharides such as cellulose and hemicelluloses that are largely insoluble fibers found in the human gut, and these cleaved sugars are then imported through ABC transporters. In each panel, the small shapes depict the official glycobiology symbols for invidual monosaccharides (green circle, mannose; yellow, galactose; blue square, N-acetlyglucosamine; green star, xylose; green triangle, rhamnose; yellow and white diamond, Galacturonic acid; blue and white diamond, Glucuronic acid; red triangle,

with SusE being stationary[94], while the hydrolytic enzyme, SusG, moves dynamically and rapidly around the length of the bacterium[95]. These studies provide important understanding into the assembly of PUL-encoded machinery on the cell surface and have further served as a reference for xyloglucan and cereal-derived beta-glucan degradation by related gut *Bacteroides*[96,97]. Previously known substrates that *Bacteroides* PULs target range from dietary plant-polysaccharides[98-103] to host-derived polysaccharides such as the *O*-linked glycans attached to mucus and the *N*-linked glycans and glycosaminoglycans attached to other host proteins and glycoconjugates[77,85,104-106]. Expression of some of these PULs has been shown to cause colitis in a sulfatase-dependent manner[8,26,107]. Still other PULS have been shown to target human milk oligosaccharides, often overlapping with those that target similar structured *O*-glycans[108]. Other

PUL-encoded abilities involve degradation of bacterial exopolysaccharides[80, 109], the α-mannans of yeast[75], and β-glucan of fungal cell walls[76]. This plethora continues to expand, with a recent study describing *Bt* degradation of the most complex dietary polysaccharide rhamnogalacturonan II[79] through the action of three separate PULs encoding 26 different enzymes. Similar studies on the less complex pectin rhamnogalacturonan I and related galacturonic acid-containing pectins and pectic side chains in *Bt* and *B. ovatus* have demonstrated a cross-feeding degradation pathway between the two organisms and mapped the specific PULs and functions required for pectin catabolism[65]. Similar to these studies are those investigating the ability of *Bt* to breakdown complex host *N*-glycans such as high mannose *N*-glycans[110] that decorate host glycoproteins, including those of mucosal immunoglobulin A, through degradation orchestrated by several, non-adjacent PULs[110]. Interestingly, *Bacteroides* hydrolytic enzymes are sometimes preferentially (almost entirely) packaged into outer membrane vesicles (OMVs), which could play important biological roles for these bacteria[111]. For example, the liberated glycans released by the OMVs of some species can serve as nutrients for the bacterium that produced the OMVs and as a communal resource among some but not all *Bacteroides.* (Figure 1.2)[112].

Marine *Bacteroidetes* use similar PUL-encoded mechanisms to utilize polysaccharides and may transfer these abilities to human gut *Bacteroides*

Beyond the degradative abilities directed towards plant- and host-derived polysaccharides, studies into seaweed-derived polysaccharides have shown some human gut *Bacteroides* have the ability to degrade seaweed-derived polysaccharides like porphyran, agarose and alginate[81,82,113-116]. In at least some cases, these abilities were transferred into gut *Bacteroides* from marine *Bacteroidetes* through the action of integrative chromosomal elements, and this ability is transferrable and can modulate gut composition[117]. The PULs found in marine *Bacteroidetes* often encode a larger gene content and an expanded repertoire of enzymatic functions than those present in gut *Bacteroides* and contain additional genes coding for adhesin proteins to help the producing bacterium remain anchored to the nutrient source[118]. This vast enzyme repertoire is likely due to the composition of the available substrates in the form of algal polysaccharides derived from cell walls of red, green, and brown algae as well as those from cyanobacteria and potential EPS structures that they produce[119-121]. The importance of encoding these PULs is exemplified during algae blooms with genes and proteins of these systems being

some of the most highly expressed products in the microbial ecosystem[122,123], likely allowing for competitive advantages for strains able to utilize complex algae polysaccharides[124,125]. These blooms, along with metagenomics and predictive phenotyping by glycan arrays can be used to measure bacterial phyla present and to what extent they degrade diverse polysaccharides[118,126,127]. Beyond *in silico* modeling, detailed degradation capabilities for the algae polysaccharides laminarin[128,129], alginate[130], agarose[131,132], ulvan[133,134], and carrageenan[135,136] have been described. Similarly, to human gut *Bacteroides*, marine *Bacteroidetes* also utilize α- and β-mannans[137]. Like gut Bacteroidetes, the ways in which these marine bacteria utilize polysaccharides can also be characterized as selfish, sharing, or scavenging and may derived from the same substrate sequestration mechanisms (or lack thereof) (Figure 1.2)[75,138,139]. Recent in-depth studies have highlighted the complex nature of PULs in marine bacteria. The genes responsible for utilization of ulvan, a highly-sulfated polysaccharide with a backbone of repeating rhamnose, xylose, glucuronic acid, and iduronic acid with sidechains consisting of rhamnose or glucuronic acid, have been described in *Formosa agariphila*, and requires 39 PUL-encoded and 20 non-PUL encoded genes[140]. Additionally, the genes required for ulvan utilization are contained on a large plasmid in the gammaproteobacterial species, *Alteromonas* sp. 76-1, suggesting this locus may be transferrable[141]. Further, the largest PUL described to date is found in *Paraglaciecola hydrolytica* S66^T, for the degradation of furcellaran, (a mixture of κ- and β-carrageenan's and agarose) and contained in a 167kb genomic region encoding 116 genes[136].

Gram-positive bacteria utilize polysaccharides via alternative multi-protein systems

In contrast to the Gram-negative, TonB-dependent transporter-dominated systems, polysaccharide degradation in the prominent Gram-positive *Firmicutes* phylum is built around different machinery and mechanistic studies are also beginning to emerge[142-147]. Similar to the mechanisms described above for marine *Gammaproteobacteria*, variant examples of multi-protein-encoding genomic loci have been described in Gram-positive species, and termed Gram-positive PULs (gpPULs)[144]. These gpPULs lack homologs of the TonB-dependent transporter *susC* and binding protein *susD*, due to the lack of an OM, and instead rely on ABC-transporter dominated systems coupled with diverse hydrolytic enzymes and binding proteins[133]. Recently, a large gpPUL was characterized in the butyrate producer *Roseburia intestinalis* that assimilates degradation of β-mannans[142]. An additional strategy that Gram-positives use is cellulosomes

(Figure 1.2). These multienzyme, modular complexes often target cellulose or plant-derived structures such as xylans and lignocellulose for degradation[128]. Additionally, recent work has shown that cellulosome-like structures, termed amylosomes, utilize a similar framework of proteins containing cohesin and dockerin domains to assemble the resistant-starch cleaving enzymes of *Ruminococcus bromii*, a keystone species of the rumen and HGM[129-131]. It is likely that strategies employed by Gram-positive members of the HGM will become clearer, and possibly new mechanisms added, as additional studies are undertaken in this very diverse phylum.

Monosaccharides and other overlooked nutrients

While it is clear that gut *Bacteroidetes, Firmicutes* and other phyla are readily equipped to degrade complex host and dietary polysaccharides, the digestive fate and importance of monosaccharides have been less studied. Perhaps this negligence is due to a common thought that monosaccharides do not affect the microbiota due to host absorption in the small intestine, such that the microbiota does not access these in large quantities. However, it is increasingly clear that either from the cleavage of larger polysaccharides or from the diet directly, that monosaccharides can affect the composition and physiology of the gut microbiota. For instance, within the human gut, some of the initially investigated nutrient niches were those delineated by mono- and disaccharides[148]. Many studies have focused on the ability of invading pathogenic bacteria that preferentially utilize monosaccharides during infection. For instance, both enterohemorrhagic *E. coli* (*EHEC*) and *Salmonella* strains upregulate genes for the utilization of ribose and other sugars during infection of cows, chickens, and mice[149-151] and *EHEC* displays preference for this sugar and several other available monosaccharides[152]. However, genes and mechanisms for the assimilation of ribose have been reported in non-pathogenic bacteria such as *Bifidobacterium breve* UCC200 and *Lactobacillus sakei*[153,154]. Further, related compounds such as deoxyribose and DNA can be utilized during pathogenic invasion by *E. coli*[155,156]. Additionally, the monosaccharides fucose and sialic acid are used by pathogens such as *Clostridium difficile* and *Salmonella typhimurium* by profiting off of the activity of gut commensals like *Bt*, which release these sugars with specific enzymes during breakdown of host-derived mucosal polysaccharides[157]. Often these systems include a kinase, sugar import protein such as an ABC-transporter, a regulator, and a few additional genes specific for the sugar such as

aldolases or isomerase, which are generally less complex than PUL-encoded mechanisms of the Gram-negative *Bacteroides* (Figure 1.2). However, in *Bt* there are at least two PULs that encode functions required for the utilization of monosaccharides along with the polysaccharides they are contained in. These include a system for β2,6-linked fructan utilization, which in addition to the fructose-containing polysaccharide levan, is also required for the utilization of fructose through actions of the PUL-encoded fructokinase and fructose permease[99]. Similarly, *Bt* encodes a PUL for the catabolism of ribose, nucleosides, and RNA (Chapter II). Interestingly, this ribose PUL is found throughout the phylum in many different genomic architectures, suggesting that individual species have evolved to access ribose from diverse sources. Additionally, *Bt* also catabolizes the monosaccharides L-fucose, L-rhamnose, and arabinose through loci that look similar to those found in *Proteobacteria*, encoding a sugar-specific kinase and dedicated transport machinery, with the exception that they all contain either novel regulatory protein or unique genetic arrangements[158-160]. This perhaps suggests that the fructan and ribose PUL have evolved from more common sugar utilization clusters, and that future studies may uncover PULs responsible for the catabolism of arabinose or rhamnose for instance, based on the acquisition of *sus-like* genes built around these core sugar assimilation systems. Lastly, emerging studies focusing on amino acids have demonstrated that these nutrients remodel the gut community composition and are important substrates for invading pathogens. For example, enteropathogenic *E. coli* (EPEC) employs a Type 3 secretion system (T3SS) injectisome to steal or "pirate" amino acids from host cells to gain an advantage over commensal bacteria[161] (Figure 1.2). Similarly, dietary L-serine in an inflamed intestine provides *Enterobactericea* a competitive edge[162]. It is clear from these studies that both gut pathogens and commensals have found niches by catabolizing monosaccharides.

To cook or not to cook, let's ask the microbiota: cooking, food preservation, and ultra-refined foods alter the gut microbiota

Produced through cooking and overlooked until recently, advanced glycation end-products (AGEs) and Mailliard reaction products (MRPs) can also perturb the gut microbiota by feeding certain species[163,164]. MRPs are formed through the heat-induced crosslinking of reducing sugars and amino acids yielding novel molecules that can be selectively accessed by

bacteria with the right tools. For instance, the MRP of fructose-asparagine is metabolized almost exclusively by *Salmonella enterica*, which excludes other members of the microbiota by using high affinity systems for import of this nutrient[165]. This ability is not confined to pathogenic bacteria, as a recent study revealed that commensal *Collinsella intestinalis* and *Collinsella aerofaciens* are able to utilize the MRP, fructoselysine *in vivo,* via actions of a phosphotransferase system. Additionally, the AGE, *N*-ε-carboxymethyllysine is degraded by yet uncultured members of the gut microbiota[166]. These three studies along with others examining the effects of MRPs and AGEs on the microbiota indicate that this area requires further study as microbiota composition can be affected by these nutrients and in-turn aspects of host health. Lastly, MRPs and AGEs are not the only substances generated or modified by actions of cooking. Resistant starch (RS) is categorized into types: raw or uncooked, cooked starch that has been cooled and retrograded, chemical modification, or physically inaccessible[167]. Each of these types is accessible to varying degrees by gut microbes such as *Ruminococcus bromii*, while cooking RS yields soluble starch accessible to many gut bacterial species[168]. In light of these studies of MRPs, AGEs, and resistant starch, the need to recognize the importance of these compounds as well as other compounds such as food emulsfiers[169], artificial sweeteners[170], and inadvertently consumed plastic-byproducts[171] is clear due to the capacity to alter the gut microbiota. These substances, although perhaps not commonly considered as nutrients could be targeted for degradation and catabolized as nutrients through actions of gut bacteria.

The central bank, how *Bacteroides* regulate their carbohydrate metabolism

With their plethora of carbohydrate degrading systems, *Bacteroides* require a finely-tuned series of local and global regulatory networks to optimally detect when specific nutrients are available and manage responses from multiple nutrient utilization systems that are activated in parallel so that they can optimize energy expenditure while also staying primed for the next meal. Previous studies and reviews have focused on this topic[172-174]; however, new studies examining the phenomenon of *Bacteroides* nutrient prioritization in complex mixtures (nutrient hierarchies) and global regulation mechanisms warrant discussion. Within the *Bacteroides* there is a clear nutrient utilization hierarchy during growth in a complex mixture of polysaccharides, the PUL-encoded machinery is transcriptionally upregulated in a specific order, similar to carbon catabolite repression in organisms like *E. coli*[173,174]. These hierarchies are present in multiple

species and can be slightly different between *Bacteroides* species[175]. This indicates that the co-existence of many strains could be aided by this differential hierarchy of polysaccharide utilization.

One mechanism that is implicated in the ability of *Bacteroides* to rapidly switch between targeting different polysaccharides is the relatively restricted, but not absolute, transcriptional control such that PULs are highly expressed only when their cognate substrate is detected. For instance, within *Bt* there is a sensory state of low-level transcript production and protein translation of PUL machinery for the starch utilization system (Sus) and other systems. This can be seen through immunofluorescent imaging of the outer-membrane Sus proteins in the presence of glucose, but absence of the cognate signaling molecule, maltose[176]. Upon exposure to starch or maltose, the *sus* PUL is activated and the outer membrane of the cell is flooded with Sus proteins. The regulatory protein within the *sus* PUL is SusR, a transcriptional activator, and is one of three recognized broad classes of regulatory proteins in *Bacteroides.* SusR-type regulators are thought to mainly affect glucose polysaccharides (although approximately 1/3 of these are unknown)[177]. The second type, ECF-σ/anti-σ pairs are seen primarily in systems responsible for the breakdown of host-derived mucin polysaccharides, these regulators function similar to other described sigma factor pairs[77]. Lastly, *Bacteroides* uniquely combine the normal two-component regulatory system domain of sensor and DNA-binding activator into a single protein termed a hybrid two-component system (HTCS), and these are found primarily in PULs involved in degrading fiber polysaccharides, although several are found in PULs associated with host mucin degradation[99] (Figure 1.3).

The local, positive acting feedback loops described above are PUL-encoded regulatory mechanisms and are on the front lines of regulation, interacting directly with cognate substrates. However, this is not the only level of regulation that is required to establish more complex, nutrient hierarchies which require some form of catabolite repression. Recently, antisense small-RNAs (sRNAs) have been described within *B. fragilis* and *Bt* found directly upstream of PULs involved in catabolism of both *N*- and *O*-linked mucus-derived polysaccharides in PULs regulated by ECF-σ/anti-σ pairs, suggesting that this sRNA is substrate restricted primarily to host glycans[178]. These sRNAs repress the transcription of the PULs leading to less breakdown and catabolism of host-derived polysaccharides, likely for more favored metabolism of dietary fiber polysaccharides. An additional, more global regulatory layer of metabolism has been

uncovered in two separate studies trying to determine if *Bacteroides* utilize a mechanism similar to catabolite repressor protein (CRP) in *E. coli*. One such homolog within *Bt* was found to be encoded by *BT4338* which was previously named MalR (maltose regulator) for its requirement

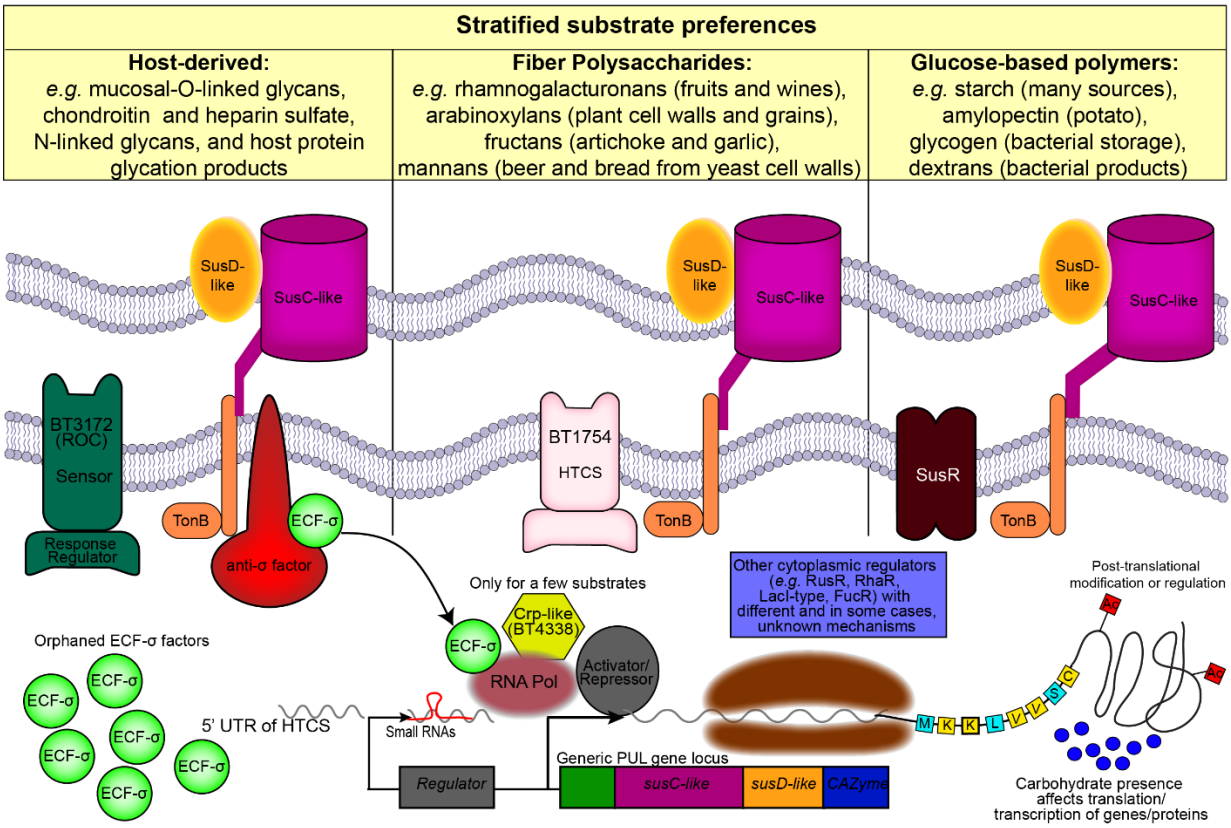


Figure 1.3 Known regulatory control mechanisms in Bacteroides.

Diagram of common regulatory mechanisms based on the types of substrates that the systems sense and degrade. At the bottom of the figure, a generic schematic of a polysaccharide utilization loci (PUL) is shown in the cytoplasm displaying the susC/D-like genes that denote PULs, a generic upstream regulatory gene, a carbohydrate active enzyme (CAZyme), and accessory genes, green boxes. Also shown in the cytoplasm for all three panels: SusC- and SusD-like proteins in the outer membrane as they are required for import and binding/stabilization of the complex regardless of the type or origin of the polysaccharide. Also shown, is the TonB-dependent energization mechanism, thought to be required for transportation through SusC-like proteins. Additional regulators known to operate in Bt are shown in the cytoplasm including, a CRP-like protein and a blue boxed area calling attention to regulators responsible for monosaccharide catabolism that do not fall in the three types shown in the panels. Also, in the cytoplasm, small RNAs and post-translational modifications such as acetylation or elongation termination are shown. The left panel shows ECF-σ/anti-σ factor based regulators which are mainly found in PULs responsible for host-polysaccharide breakdown. These types of systems have the anti-σ spanning the inner membrane (IM) and periplasmic space, while keeping the soluble, activating ECF-σ factor bound until an inducer is sensed in the periplasmic space, when it is then released and can help in recruiting RNA pol for transcript initiation. Additionally, several genomically unlinked, orphaned ECF-σ factors without adjacent encoded anti-σ factors are present within Bt and related Bacteroides genomes, suggesting an unknown role for these regulators. Also, one hybrid two-component system (HTCS) regulator, BT3172 (also known as the regulatory of colonization, Roc) is shown. This is one of the only known HTCS found in a host-responsive PUL and affects persistence in the gut in a diet-dependent manner. Mainly HTCS regulators are associated with fiber-responsive PULs (middle panel). Unlike classical two-component systems composed of two separate response regulator and histidine kinase-encoding genes, these functions have been fused into one, multi-domain containing protein. In the right panel, the SusR family of regulatory proteins is shown. Normally these regulators are in PULs targeting glucose-based polysaccharides for degradation. The mechanism of signal transduction is unknown for this family of regulators, but in the Bt starch system, the inducing molecule is maltose, which when sensed by SusR causes a large transcriptional and translation upregulation leading to catabolism of starch.

in maltose catabolism in the absence of SusR[179]. In a more recent study, the authors found that a deletion strain lacking *BT4338* either lost or had diminished growth on several substrates, many of which were monosaccharides, rather than polysaccharides[180]. It is possible then that *BT4338* acts at the level of the monosaccharide utilization, whereby PULs are still upregulated, but metabolism genes required for growth on these substrates are not functionally expressed. Further, a recent study demonstrated that *BT3172,* a PUL-associated HTCS that is likely responsible for upregulating the adjacent PUL genes *BT3173-3180* in response to mucosal glycans is highly repressed at the post-transcriptional level by glucose and a few other monosaccharides. This is interesting, as it suggests that the dietary monosaccharides fructose and glucose affect the activity of Roc, preventing it from upregulating the adjacent genes, which are clearly important *in vivo* while Roc ("regulator of colonization") is not required for catabolism of these monosaccharides. Further, the suppressive effect of glucose/fructose was localized to an upstream mRNA leader sequence, that when deleted, alleviated Roc repression from glucose or fructose. Taken together, these studies into the regulatory networks underpinning carbohydrate metabolism in *Bt* show that there is still much to discovery in this field (and model organism). These studies also provide insight into the hierarchical or global regulation that has many different layers (Figure 1.3). Lastly, I want to highlight that within *Bt's* genome there are at least 20 separate ECF-σ factors that are "orphans" throughout the genome, meaning that they are not apparently adjacent to known PUL genes[181]. It is possible that these orphan regulators function in the assimilation of non-carbohydrate nutrients such as cofactors, amino acids, vitamins, or in a manner previously unknown for PUL-encoded functions. Uncovering the function(s) of these regulators will likely add important knowledge to the overall understanding of the regulatory mechanisms within *Bacteroides* species.

Synthetic engineering of gut bacteria and generation of synthetic communities

With the above mentioned examples, it should come as no surprise that strides in the area of synthetic biology and engineering of the HGM have been made in the past several years. One impetus for this renewed interest in generating genetic toolkits has been the realization that the majority of species and isolates within the gut microbiota are genetically intractable. Although new genetic tools have been generated for some strains, many of the prominent members of Bacteroidetes and Firmicutes have had few new systems developed until recently. Here we

discuss how systems developed for use elsewhere in non-HGM isolates may be relevant to the HGM. Although *proteobacteria* are often less prominent or more transient members of the gut microbiota, they are still commonly found[182,183] and several important pathogens such as *E. coli, V. cholera, S. enterica* and others have numerous genetic tools. However, even within this phylum, new tools have been developed with broad-host range to target strains inhabiting the bee gut microbiome that can be used in *Alpha-*, *Beta-*, and *Gammaproteobacteria*[184]. It is conceivable that these tools could also be used in proteobacteria of the HGM. In sticking with proteobacteria, an elegant example of synthetic biology that used 12-independent, inducible sensors in a single strain of *E. coli*[185], demonstrated some of the types of tools required for HGM isolates to assay aspects of bacteria-host interactions. These techniques will allow for identification of genes coding for important, direct effectors of disease such as the proteins required for host polysaccharide degradation.

For strains that are not tractable, an interesting and useful technique being used to isolate and transfer DNA is metaparental mating[186]. However, this strategy is limited to strains that are not multi-antibiotic resistant, and is more often used to identify genetically manipulatable strains rather than to transfer or edit DNA for experimental or engineering purposes. Antibiotic resistance is a common reason why strains are hard to genetically manipulate and is likely to only get worse as strains become multi- and pan-drug resistant[187]. However, a new set of plasmids have been created for a large range of *Bacteroidetes* isolates[188]. This new system is based off of a Gain-of-Function phenotype for the ability to utilize inulin, a polysaccharide that only a small percentage of strains can normally utilize[99,188,189]. Presumably this strategy can be adapted to different substrates for recipient strains already able to grow on inulin. Further, the novel counter-selection strategy bypasses the need for antibiotics by using a toxin encoded from a T6SS gene in the *Bacteroides fragilis* genome. Most of the *Bacteroides* plasmids are constructed off of a NBU2 integration plasmid backbone[190], that has previously been adapted for use in up to 8 different channels of *in vivo* fluorescence imaging of *Bacteroides*[191], tunable expression of genes *in vivo*[192], and ability to respond to endogenous signals and record these exposures via a CRISPR-Cas9 system[193]. Although all of these tools represent a nice toolkit, to fully take advantage of engineering strains for health in the HGM, the transfer of PULs, which are often several kb in length, require new approaches to transfer functional, intact PULs into new species. Recently, a method of transferring PULs using yeast as a platform to assemble

piece-wise, the components of larger PULs into a bacterial artificial chromosome has been described[117]. Applications of new techniques such as the transfer of entire PULs, allows for creative, niche engineering approaches to modulate individual species of the HGM and may one day lead to better precision editing of the HGM in disease states. These built niches may also guide the formulation of synthetic microbial communities that can be introduced by consumption of prebiotics or probiotics or in the event of advanced disease states, fecal microbiota transfer.

Prospectus

Given the central importance of complex carbohydrate-based nutrients in the human gut and other environments, more in-depth mechanistic studies are needed to understand the metabolism of these nutrients. One of the most exciting areas that can guide these studies involve the capability of using recombinant or whole cell lysates of *Bacteroidetes* SusD-like proteins in glycan binding arrays in an effort to determine substrate and growth preferences of either uncultivated strains or PULs with unknown substrates[126]. Importantly, within the gut microbiota there are likely additional nutrients such as microbial capsular polysaccharides that can serve as carbohydrate nutrients. This mechanism could essentially be described as predation of other bacteria or their products but it might also be the basis of enforcing species-species interactions. This avenue is one that requires much more in-depth study, as it is likely that capsules represent a massive nutrient pool in the gut. For example, within individual strains there can be several different capsules (*Bt* encodes 8 separate capsular biosynthesis loci for instance)[56] and within the umbrella of *E. coli* there are at least 80 different capsule structures or K antigens[194]. Relatedly, certain exopolysaccharides from *Lactobacillus* and *Bifidobacteria* are degraded and used for growth by *Bt* and *B. fragilis*[80,109], and this is yet another possible nutrient source in the gut. I therefore speculate based on the diversity of origin, structure, and linkage of substrates catabolized by the gut microbiota, that it is highly likely that inhabitants and invaders of the gut ecosystem have developed the collective capacity to target many, tens or hundreds of additional, unknown nutrients. Lastly, I believe that a significant aspect of future work in the gut microbiota and nutrient utilization will be to connect these functions to host health and disease states. In order to do that we both need to continue developing synthetic biology and genetic tools and increase efforts aimed at mapping phenotype to genotype to better understand niche partitioning in the human gut. I expect that this expanded understanding will come to fruition through the use

of interdisciplinary approaches involving enzymology, metabolic modeling, informatics, microbial growth and phenotype assays, and the use of omics-based approaches.

Chapter outline

The carbohydrate nutrients present in the human gastrointestinal tract often define niches based on the degrading or catabolizing abilities of certain species. This phenomenon can often influence human health and disease states. Members of the prominent Bacteroidetes have evolved and developed the ability to degrade and grow on a broad range of dietary-, host-, microbial-, and fungal-derived polysaccharides, monosaccharides, and additional carbohydrate-containing nutrients. Although significant work has been performed examining the genetic loci (polysaccharide utilization loci, PUL) and protein functions (Sus-like systems) that they encode, there are still critical details lacking. The cognate substrates for many of these systems as well as the regulatory mechanisms used to distinguish and respond to available nutrients are unknown. Building on previous knowledge, molecular mechanisms of carbohydrate degradation and recognition in the model system *Bacteroides thetaiotaomicron* (*Bt*) are explored here. The major focus of this dissertation is to elucidate the molecular mechanisms by which a ribose utilization system (*rus*) encoded by a PUL degrades ribose-containing compounds, and the impact this has *in vivo*. Additional work has focused on host-polysaccharide degradation through the identification and study of 12 related PULs that work in concert to break down mucin polysaccharides (host mucosal-derived). Lastly, global regulatory mechanisms were examined.

In Chapter II, I describe the essential *in vitro* and *in vivo* functions encoded in the *rus* locus for the catabolism of ribose and nucleosides. Demonstrating, that *rus*-encoded ribokinases perform a previously unidentified (in eubacteria) mechanism of ribose phosphorylation in cooperation of an upstream, unlinked nucleoside phosphorylase. This mechanism was found to be important *in vivo* on a high fiber diet containing a nucleoside or nucleoside-like substrate. This work also expands the known substrate diversity of PULs to include nucleosides and the monosaccharide ribose, as well as demonstrating that ribose utilization is penetrant across the phylum in at least 70 different genomic configurations.

In Chapter III, I examine a *Bt*-specific, *in vivo* T cell clone by confirming the epitope recognized by these T cells and identify the exact amino acids recognized. The protein epitope, BT4295 is regulated by presence of glucose and salts. Presence of glucose *in vivo* reduces the T

cell stimulation due to reduced transcription of the epitope. Additionally, *BT4295* is found in a PUL responsive to growth in host mucosal polysaccharides, and additional proteins of this locus also served as weaker T cell epitopes. Further, in identifying the epitope, a transposon screen implicated functions of the pentose phosphate pathway as serving as potential epitopes.

In Chapter IV, I highlight regulatory mechanisms in *Bt*. This work uncovered a previously unknown, potentially global regulator of polysaccharide utilization, BT2492. BT2492 is one of many orphan ECF-σ factors without functional knowledge. Additionally, 3 LacI family transcriptional regulators where investigated. They display regulation towards uronic acids or uronic acid-containing polysaccharides. Further, in Chapter II, I observed that ribose altered the expression of other PULs and metabolic loci, this was followed up with experiments in arabinose and xylose, yielding similar results and suggesting this cross-metabolism phenomenon may be an additional layer of regulation in *Bt.* Interestingly, one of the PULs identified during arabinose growth was the one containing the epitope, BT4295 from Chapter III, providing a potential reason why pentose phosphate metabolism genes were found during that study.

Finally, in Chapter V, I expand upon previous work examining host mucosal polysaccharide utilization via PUL-encoded mechanisms. Within this chapter is the characterization of a complex genetic deletion in *Bt*, lacking 11 different PULs associated with host polysaccharide degradation. This study revealed some of the PULs and genes required for mucosal polysaccharide utilization including sulfatases and fucosidases which have proved valuable in guiding additional *in vivo* studies in attempts to decrease colitogenic responses of *Bt* in fiber free dietary conditions.

Notes

Portions of this chapter have been adapted from review articles in preparation with permission from Glowacki, R.W. and Martens, E.C. The first half of this chapter is from an invited, “Pearls” mini-review for *PloS Pathogens*, with a working title of “In sickness and health: effects of gut microbial metabolites on human physiology”. The second half of this chapter is being prepared for an invited review to *Journal of Bacteriology.*

References

1. Escherich, T. and K.S. Bettelheim, The Intestinal Bacteria of the Neonate and Breast-Fed

Infant. *Rev. Infect. Diseases* **10**, 1220-1225 (1988).

2. Koropatkin, N.M., E.A. Cameron, and E.C. Martens, How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323-35 (2012).

3. Porter, N.T. and E.C. Martens, The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu. Rev. Microbiol*. **71**, 349-369 (2017).

4. Britton, R.A. and V.B. Young, Role of the intestinal microbiota in resistance to colonization by Clostridium difficile. *Gastroenterology* **146**, 1547-53 (2014).

5. Nash, M.J., D.N. Frank, and J.E. Friedman, Early Microbes Modify Immune System Development and Metabolic Homeostasis-The "Restaurant" Hypothesis Revisited. *Front. Endocrinol*. **8**, 349 (2017).

6. Gollwitzer, E.S. and B.J. Marsland, Impact of Early-Life Exposures on Immune Maturation and Susceptibility to Disease. *Trends. Immunol*. **36**, 684-696 (2015).

7. Chung, H., et al., Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149**, 1578-93 (2012).

8. 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**, 672-80 (2015).

9. Dejea, C.M., et al., Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592-597 (2018).

10. Tomkovich, S., et al., Human colon mucosal biofilms from healthy or colon cancer hosts are carcinogenic. *J. Clin. Invest*. **130**, 1699-1712 (2019).

11. Lagkouvardos, I., et al., The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. *Nat. Microbiol*. **1**, 16131 (2016).

12. Browne, H.P., et al., Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543-546 (2016).

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

14. Dethlefsen, L., et al., The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **6**, e280 (2008).

15. Smits, S.A., et al., Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* **357**, 802-806 (2017).

16. Vangay, P., et al., US Immigration Westernizes the Human Gut Microbiome. *Cell* **175**, 962-972 (2018).

17. David, L.A., et al., Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559-63 (2014).

18. Sonnenburg, E.D., et al., Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212-5 (2016).

19. Cekanaviciute, E., et al., Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10713-10718 (2017).

20. Hsiao, E.Y., et al., Modeling an autism risk factor in mice leads to permanent immune dysregulation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12776-81 (2012).

21. Sharon, G., et al., Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. *Cell* **177**, 1600-1618 (2019).

22. Feehley, T., et al., Healthy infants harbor intestinal bacteria that protect against food allergy. *Nat. Med*. **25**, 448-453 (2019).

23. Li, J., et al., An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol*. **32**, 834-41 (2014).

24. Pasolli, E., et al., Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649-662 (2019).

25. Thanissery, R., J.A. Winston, and C.M. Theriot, Inhibition of spore germination, growth, and toxin activity of clinically relevant C. difficile strains by gut microbiota derived secondary bile acids. *Anaerobe* **45**, 86-100 (2017).

26. Desai, M.S., et al., A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353 (2016).

27. Sampson, T.R., et al., Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell* **167**, 1469-1480 (2016).

28. Strandwitz, P., et al., GABA-modulating bacteria of the human gut microbiota. *Nat. Microbiol*. **4**, 396-403 (2019).

29. Valles-Colomer, M., et al., The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat. Microbiol*. **4**, 623-632 (2019).

30. Hosie, S., et al., Gastrointestinal dysfunction in patients and mice expressing the autism-associated R451C mutation in neuroligin-3. *Autism Res*. **12**, 1043-1056 (2019).

31. Grimaldi, R., et al., A prebiotic intervention study in children with autism spectrum disorders (ASDs). *Microbiome* **6**, 133 (2018).

32. Hsiao, E.Y., et al., Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155**, 1451-63 (2013).

33. Chen, H., et al., A Forward Chemical Genetic Screen Reveals Gut Microbiota Metabolites That Modulate Host Physiology. *Cell* **177**, 1217-1231 (2019).

34. Lundmark, K., et al., Protein fibrils in nature can enhance amyloid protein A amyloidosis in mice: Cross-seeding as a disease mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6098-6102 (2005).

35. Friedland, R.P. and M.R. Chapman, The role of microbial amyloid in neurodegeneration. *PLoS Pathog*. **13**, e1006654 (2017).

36. Gandy, K.A.O., et al., The role of gut microbiota in shaping the relapse-remitting and chronic-progressive forms of multiple sclerosis in mouse models. Sci Rep, 2019. 9(1): p. 6923.

37. Pollet, R.M., et al., An Atlas of beta-Glucuronidases in the Human Intestinal Microbiome. *Structure*, **25**, 967-977 (2017).

38. Dashnyam, P., et al., beta-Glucuronidases of opportunistic bacteria are the major contributors to xenobiotic-induced toxicity in the gut. *Sci. Rep*. **8**, 16372 (2018).

39. Yuan, J., et al., Fatty Liver Disease Caused by High-Alcohol-Producing Klebsiella pneumoniae. *Cell Metab*. **30**, 675-688 (2019).

40. Llorente, C., et al., Gastric acid suppression promotes alcoholic liver disease by inducing overgrowth of intestinal Enterococcus. *Nat. Commun*. **8**, 837 (2017).

41. Carr, T.F., R. Alkatib, and M. Kraft, Microbiome in Mechanisms of Asthma. *Clin. Chest Med*. **40**, 87-96 (2019).

42. Berni Canani, R., et al., Formula selection for management of children with cow's milk allergy influences the rate of acquisition of tolerance: a prospective multicenter study. *J. Pediatr*. **163**, 771-777 (2013).

43. Stefka, A.T., et al., Commensal bacteria protect against food allergen sensitization. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13145-50 (2014).

44. Wang, Z., et al., Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57-63 (2011).

45. Caminero, A., et al., Duodenal Bacteria From Patients With Celiac Disease and Healthy Subjects Distinctly Affect Gluten Breakdown and Immunogenicity. *Gastroenterology* **151**, 670-683 (2016).

46. Wu, S., et al., The Bacteroides fragilis toxin binds to a specific intestinal epithelial cell receptor. *Infect. Immun*. **74**, 5382-90 (2006).

47. Haghi, F., et al., The association between fecal enterotoxigenic B. fragilis with colorectal cancer. *BMC Cancer* **19**, 879 (2019).

48. Wu, J., Q. Li, and X. Fu, Fusobacterium nucleatum Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity. *Transl. Oncol*. **12**, 846-851 (2019).

49. Arthur, J.C., et al., Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120-3 (2012).

50. Bhattarai, Y., et al., Gut Microbiota-Produced Tryptamine Activates an Epithelial G-Protein-Coupled Receptor to Increase Colonic Secretion. *Cell Host Microbe* **23**, 775-785 (2018).

51. Cohen, L.J., et al., Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* **549**, 48-53 (2017).

52. Zelante, T., et al., Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity* **39**, 372-85 (2013).

53. Zegarra-Ruiz, D.F., et al., A Diet-Sensitive Commensal Lactobacillus Strain Mediates TLR7-Dependent Systemic Autoimmunity. *Cell Host Microbe* **25**, 113-127 (2019).

54. Whiteley, A.T., et al., Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194-199 (2019).

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

56. Porter, N.T., et al., 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).

57. Fanning, S., et al., Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2108-13 (2012).

58. Campos, M.A., et al., Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infect. Immun.* **72**, 7107-14 (2004).

59. Lee, I.C., et al., Strain-Specific Features of Extracellular Polysaccharides and Their Impact on Lactobacillus plantarum-Host Interactions. *Appl. Environ. Microbiol.* **82**, 3959-3970 (2016).

60. Remus, D.M., et al., Impact of 4 Lactobacillus plantarum capsular polysaccharide clusters on surface glycan composition and host cell signaling. *Microb. Cell. Fact.* **11**, 149 (2012).

61. Ramakrishna, C., et al., Bacteroides fragilis polysaccharide A induces IL-10 secreting B and T cells that prevent viral encephalitis. *Nat. Commun.* **10**, 2153 (2019).

62. Vitetta, L., G. Vitetta, and S. Hall, Immunological Tolerance and Function: Associations Between Intestinal Bacteria, Probiotics, Prebiotics, and Phages. *Front. Immunol.* **9**, 2240 (2018).

63. Haiser, H.J., et al., Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. *Science* **341**, 295-8 (2013).

64. Koppel, N., et al., Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *Elife* **7**, e33953 (2018).

65. Koppel, N., V. Maini Rekdal, and E.P. Balskus, Chemical transformation of xenobiotics by the human gut microbiota. *Science* **356**, eaag2770 (2017)

66. Martinez-del Campo, A., et al., Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *MBio* **6**, e00042-15 (2015).

67. Biernat, K.A., et al., Structure, function, and inhibition of drug reactivating human gut microbial beta-glucuronidases. *Sci. Rep*. **9**, 825 (2019).

68. Zimmermann, M., et al., Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science* **363**, eaat9931 (2019).

69. Zimmermann, M., et al., Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **570**, 462-67 (2019).

70. Maini Rekdal, V., et al., Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **364**, eaau6323 (2019).

71. Kochan, T.J., et al., Germinant Synergy Facilitates Clostridium difficile Spore Germination under Physiological Conditions. *mSphere* **3**, e00335-18 (2018).

72. Kochan, T.J., et al., Intestinal calcium and bile salts facilitate germination of Clostridium difficile spores. *PLoS Pathog*. **13**, 1006443 (2017).

73. Maseda, D., et al., Nonsteroidal Anti-inflammatory Drugs Alter the Microbiota and Exacerbate Clostridium difficile Colitis while Dysregulating the Inflammatory Response. *MBio* **10**, e02292-18 (2019).

74. Guthrie, L., S. Wolfson, and L. Kelly, The human gut chemical landscape predicts microbe-mediated biotransformation of foods and drugs. *Elife* **8**, e42966 (2019).

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

76. Temple, M.J., et al., A Bacteroidetes locus dedicated to fungal 1,6-beta-glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6-beta-glucanase. *J. Biol. Chem*. **292**, 10639-10650 (2017).

77. Martens, E.C., H.C. Chiang, and J.I. Gordon, Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447-57 (2008).

78. Luis, A.S., et al., Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat. Microbiol*. **3**, 210-219 (2018).

79. Ndeh, D., et al., Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65-70 (2017).

80. Lammerts van Bueren, A., et al., Differential Metabolism of Exopolysaccharides from Probiotic Lactobacilli by the Human Gut Symbiont Bacteroides thetaiotaomicron. *Appl Environ. Microbiol.* **81**, 3973-83 (2015).

81. Hehemann, J.K., A.G.; Pudlo, P.A.; Martens, E.C.; Boraston, A.B., Bacteria of the human gut microbiome catabolize seaweed glycans with carbohydrate-active enzymes update from extrinsic microbes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19786-19791 (2012).

82. Pluvinage, B., et al., Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. *Nat. Commun.* **9**, 1043 (2018).

83. Grondin, J.M., et al., Polysaccharide Utilization Loci: Fueling Microbial Communities. *J. Bacteriol.* **199**, e00860-16 (2017).

84. Eckburg, P.B., et al., Diversity of the Human Intestinal Microbial Flora. *Science* **308**, 1635-1638 (2005).

85. Bjursell, M.K., E.C. Martens, and J.I. Gordon, 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).

86. Anderson, K.L. and A.A. Salyers, Genetic evidence that outer membrane binding of starch is required for starch utilization by Bacteroides thetaiotaomicron. *J. Bacteriol.* **171**, 3199-204 (1989).

87. Martens, E.C., et al., Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J. Biol. Chem. 2009. **284**(37): p. 24673-7 (2009).

88. 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).

89. Ravcheev, D.A., et al., Polysaccharides Utilization in B. theta; Comparatice Genomics Reconstruction of Metabolic and Regulatory Networks. *BMC Genomics* **14** (2013).

90. Terrapon, N., et al., PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* **46**, D677-D683 (2018).

91. Martens, E.C., et al., The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. *J. Mol. Biol.* **426**, 3851-65 (2014).

92. Anderson, K.L. and A.A. Salyers, Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *J. Bacteriol.* **171**, 3192-8 (1989).

93. Foley, M.H., E.C. Martens, and N.M. Koropatkin, SusE facilitates starch uptake independent of starch binding in B. thetaiotaomicron. Mol. Microbiol. **108**, 551-566 (2018).

94. Tuson, H.H., et al., The Starch Utilization System Assembles around Stationary Starch-Binding Proteins. *Biophys. J.* **115**, 242-250 (2018).

95. Karunatilaka, K.S., et al., Superresolution imaging captures carbohydrate utilization dynamics in human gut symbionts. *MBio* **5**, e02172 (2014).

96. Tamura, K., et al., Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont Bacteroides ovatus. *Cell. Mol. Life Sci*. **76**, 4319-40 (2019).

97. Foley, M.H., et al., A Cell-Surface GH9 Endo-Glucanase Coordinates with Surface Glycan-Binding Proteins to Mediate Xyloglucan Uptake in the Gut Symbiont Bacteroides ovatus. *J. Mol. Biol*. **431**, 981-995 (2019).

98. Mardo, K., et al., A Highly Active Endo-Levanase BT1760 of a Dominant Mammalian Gut Commensal Bacteroides thetaiotaomicron Cleaves Not Only Various Bacterial Levans, but Also Levan of Timothy Grass. *PLoS One* **12**, e0169989 (2017).

99. Sonnenburg, E.D., et al., Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, 1241-52 (2010).

100. 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*. **292**, 229-243 (2017).

101. Tamura, K., et al., Molecular Mechanism by which Prominent Human Gut Bacteroidetes Utilize Mixed-Linkage Beta-Glucans, Major Health-Promoting Cereal Polysaccharides. *Cell Rep*. **21**, 417-430 (2017).

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

103. Martens, E.C., et al., Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol*. **9**, e1001221 (2011).

104. Pudlo, N.A., et al., Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. *MBio* **6**, e01282-15 (2015).

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

106. Ulmer, J.E., et al., Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont Bacteroides thetaiotaomicron reveals the first GAG-specific bacterial endosulfatase. *J. Biol. Chem*. **289**. 24289-303 (2014).

107. Bloom, S.M., et al., Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9**, 390-403 (2011).

108. Marcobal, A., et al., Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10**, 507-14 (2011).

109. Rios-Covian, D., et al., Bacteroides fragilis metabolises exopolysaccharides produced by bifidobacteria. *BMC Microbio*. **16**, 150 (2016).

110. Briliute, J., et al., Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat. Microbiol*. **4**, 1571-81 (2019).

111. Valguarnera, E., et al., Surface Exposure and Packing of Lipoproteins into Outer Membrane Vesicles Are Coupled Processes in Bacteroides. *mSphere* **3**, e00559-18 (2018).

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

113. Hehemann, J.H., et al., Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908-12 (2010).

114. Li, M., et al., Degradation of Marine Algae-Derived Carbohydrates by Bacteroidetes Isolated from Human Gut Microbiota. *Mar. Drugs* **15**, 92 (2017).

115. 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. *Environ. Microbiol*. **14**, 2379-94 (2012).

116. Mathieu, S., et al., Ancient acquisition of "alginate utilization loci" by human gut microbiota. *Sci. Rep*. **8**, 8075 (2018).

117. Shepherd, E.S., et al., An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature* **557**, 434-438 (2018).

118. Fernandez-Gomez, B., et al., Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J*. **7**, 1026-37 (2013).

119. Poli, A., G. Anzelmo, and B. Nicolaus, Bacterial exopolysaccharides from extreme marine habitats: production, characterization and biological activities. *Mar. Drugs* **8**, 1779-802 (2010).

120. Muhlenbruch, M., et al., Mini-review: Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. *Environ. Microbiol*. **20**, 2671-2685 (2018).

121. Decho, A.W. and T. Gutierrez, Microbial Extracellular Polymeric Substances (EPSs) in Ocean Systems. *Front. Microbiol*. **8**, 922 (2017).

122. Mann, A.J., et al., The genome of the alga-associated marine flavobacterium Formosa agariphila KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Appl. Environ. Microbiol*. **79**, 6813-22 (2013).

123. Teeling, H., et al., Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**, 608-11 (2012).

124. Gobet, A., et al., Evolutionary Evidence of Algal Polysaccharide Degradation Acquisition by Pseudoalteromonas carrageenovora 9(T) to Adapt to Macroalgal Niches. *Front. Microbiol*. **9**, 2740 (2018).

125. Unfried, F., et al., Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. *ISME J*. **12**, 2894-2906 (2018).

126. Kappelmann, L., et al., Polysaccharide utilization loci of North Sea Flavobacteriia as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J*. **13**, 76-91 (2019).

127. Tang, K., et al., Characterization of Potential Polysaccharide Utilization Systems in the Marine Bacteroidetes Gramella Flava JLT2011 Using a Multi-Omics Approach. *Front. Microbiol.* **8**, 220 (2017).

128. Xing, P., et al., Niches of two polysaccharide-degrading Polaribacter isolates from the North Sea during a spring diatom bloom. *ISME J*. **9**, 1410-22 (2015).

129. Alderkamp, A.C., M. van Rijssel, and H. Bolhuis, Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. *FEMS Microbiol. Ecol*. **59**, 108-17 (2007).

130. Kabisch, A., et al., Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes 'Gramella forsetii' KT0803. *ISME J*. **8**, 1492-502 (2014).

131. Lee, C.H., et al., A novel agarolytic beta-galactosidase acts on agarooligosaccharides for complete hydrolysis of agarose into monomers. *Appl. Environ. Microbiol*. **80**, 5965-73 (2014).

132. Kim, S.G., et al., Agarivorans aestuarii sp. nov., an agar-degrading bacterium isolated from a tidal flat. *Int. J. Syst. Evol. Microbiol*. **66**, 3119-24 (2016).

133. Konasani, V.R., et al., A novel ulvan lyase family with broad-spectrum activity from the ulvan utilisation loci of Formosa agariphila KMM 3901. *Sci. Rep*. **8**, 14713 (2018).

134. Foran, E., et al., Functional characterization of a novel “ulvan utilization loci” found in Alteromonas sp. LOR genome. *Algal Research* **25**, 39-46 (2017).

135. Ficko-Blean, E., et al., Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat. Commun*. **8**, 1685 (2017).

136. Schultz-Johansen, M., et al., A Novel Enzyme Portfolio for Red Algal Polysaccharide Degradation in the Marine Bacterium Paraglaciecola hydrolytica S66(T) Encoded in a Sizeable Polysaccharide Utilization Locus. *Front. Microbiol.* **9**, 839 (2018).

137. Chen, J., et al., Alpha- and beta-mannan utilization by marine Bacteroidetes. *Environ. Microbiol*. **20**, 4127-4140 (2018).

138. Reintjes, G., et al., Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. *ISME J*. **13**, 1119-1132 (2019).

139. Rakoff-Nahoum, S., K.R. Foster, and L.E. Comstock, The evolution of cooperation within the gut microbiota. *Nature* **533**, 255-9 (2016).

140. Reisky, L., et al., A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. *Nat. Chem. Biol*. **15**, 803-812 (2019).

141. Koch, H., et al., Adaptations of Alteromonas sp. 76-1 to Polysaccharide Degradation: A CAZyme Plasmid for Ulvan Degradation and Two Alginolytic Systems. *Front. Microbiol*. **10**, 504 (2019).

142. La Rosa, S.L., et al., The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary beta-mannans. *Nat. Commun*. **10**, 905 (2019).

143. La Rosa, S.L., et al., Wood-Derived Dietary Fibers Promote Beneficial Human Gut Microbiota. *mSphere*, **4**, e00554-18 (2019).

144. Sheridan, P.O., et al., Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microb. Genom*. **2**, e000043 (2016).

145. 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).

146. Bule, P., et al., Higher order scaffoldin assembly in Ruminococcus flavefaciens cellulosome is coordinated by a discrete cohesin-dockerin interaction. *Sci. Rep*. **8**, 6987 (2018).

147. Tauzin, A.S., et al., Sucrose 6(F)-phosphate phosphorylase: a novel insight in the human gut microbiome. *Microb. Genom*. **5**, e000253 (2019).

148. Freter, R.B., H; Botney, M; Cleven, D; Aranki, A, Mechanisms That Control Bacterial Populations in Continuous-Flow Culture Models of Mouse Large Intestinal Flora. *Infect. Immun*. **39**, 676-685 (1983).

149. Segura, A., et al., Transcriptomic analysis reveals specific metabolic pathways of enterohemorrhagic Escherichia coli O157:H7 in bovine digestive contents. *BMC Genomics* **16**, 766 (2018).

150. Harvey, P.C., et al., Salmonella enterica serovar typhimurium colonizing the lumen of the chicken intestine grows slowly and upregulates a unique set of virulence and metabolism genes. *Infect. Immun*. **79**, 4105-21 (2011).

151. Rollenhagen, C. and D. Bumann, Salmonella enterica highly expressed genes are disease specific. *Infect. Immun*. **74**, 1649-60 (2006).

152. Fabich, A.J., et al., Comparison of Carbon Nutrition for Pathogenic and Commensal Escherichia coli Strains in the Mouse Intestine. *Infect. Immun*. **76**, 1143-1152 (2008).

153. Pokusaeva, K., et al., Ribose utilization by the human commensal Bifidobacterium breve UCC2003. *Microb. Biotechnol*. **3**, 311-23 (2010).

154. McLeod, A., et al., Global transcriptome response in Lactobacillus sakei during growth on ribose. *BMC Microbiol*. **11**, 145 (2011).

155. Bernier-Febreau, C., et al., Use of deoxyribose by intestinal and extraintestinal pathogenic Escherichia coli strains: a metabolic adaptation involved in competitiveness. *Infect. Immun*. **72**, 6151-6 (2004).

156. Palchevskiy, V. and S.E. Finkel, Escherichia coli competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *J. Bacteriol*. **188**, 3902-10 (2006).

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

158. Patel, E.H., et al., Rhamnose catabolism in Bacteroides thetaiotaomicron is controlled by the positive transcriptional regulator RhaR. *Res. Microbiol*. **159**, 678-84 (2008).

159. Hooper, L.V., et al., 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**, 9833-9838 (1999).

160. Chang, C., et al., A novel transcriptional regulator of L-arabinose utilization in human gut bacteria. *Nucleic Acids Res*. **43**, 10546-59 (2015).

161. Pal, R.R., et al., Pathogenic E. coli Extracts Nutrients from Infected Host Cells Utilizing Injectisome Components. *Cell* **177**, 683-696 (2019).

162. Kitamoto, S., et al., Dietary L-serine confers a competitive fitness advantage to Enterobacteriaceae in the inflamed gut. *Nat. Microbiol*. (2019).

163. Snelson, M. and M.T. Coughlan, Dietary Advanced Glycation End Products: Digestion, Metabolism and Modulation of Gut Microbial Ecology. *Nutrients* **11**, e215 (2019).

164. Carmody, R.N., et al., Cooking shapes the structure and function of the gut microbiome. *Nat. Microbiol*. **4**, 2052–2063 (2019).

165. Wu, J., et al., Salmonella-Mediated Inflammation Eliminates Competitors for Fructose-Asparagine in the Gut. *Infect. Immun*. **86**, e00945-17 (2018).

166. Bui, T.P.N., et al., Anaerobic Degradation of N-epsilon-Carboxymethyllysine, a Major Glycation End-Product, by Human Intestinal Bacteria. *J. Agric. Food Chem*. **67**, 6594-6602 (2019).

167. Birt, D.F., et al., Resistant starch: promise for improving human health. *Adv. Nutr*. **4**, 587-601 (2013).

168. Maier, T.V., et al., Impact of Dietary Resistant Starch on the Human Gut Microbiome, Metaproteome, and Metabolome. *MBio* **8**, e01343-17 (2017).

169. Chassaing, B., et al., Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. *Nature* **519**, 92-6 (2015).

170. Wang, Q.P., et al., Non-nutritive sweeteners possess a bacteriostatic effect and alter gut microbiota in mice. *PLoS One* **13**, e0199080 (2018).

171. Reddivari, L., et al., Perinatal Bisphenol A Exposure Induces Chronic Inflammation in Rabbit Offspring via Modulation of Gut Bacteria and Their Metabolites. *mSystems* **2**, e00093-17 (2017).

172. Schwalm, N.D., 3rd and E.A. Groisman, Navigating the Gut Buffet: Control of Polysaccharide Utilization in Bacteroides spp. *Trends Microbiol*. **25**, 1005-1015 (2017).

173. Rogers, T.E., et al., Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol. Microbiol*. **88**, 876-90 (2013).

174. Schwalm, N.D., 3rd, G.E. Townsend, 2nd, and E.A. Groisman, Prioritization of polysaccharide utilization and control of regulator activation in Bacteroides thetaiotaomicron. *Mol. Microbiol.* **104**, 32-45 (2017).

175. Tuncil, Y.E., et al., Reciprocal Prioritization to Dietary Glycans by Gut Bacteria in a Competitive Environment Promotes Stable Coexistence. *MBio* **8**, e01068-17 (2017).

176. 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).

177. D'ella, J.N. and A.A. Salyers, Effect of Regulatory Protein Levels on Utilization of Starch by Bacteroides thetaiotaomicron. *J. Bacteriol*. **178**, 7180-7186 (1996).

178. Cao, Y., et al., cis-Encoded Small RNAs, a Conserved Mechanism for Repression of Polysaccharide Utilization in Bacteroides. *J Bacteriol*. **198**, 2410-8 (2016).

179. Cho, K.H., et al., New regulatory gene that contributes to control of Bacteroides thetaiotaomicron starch utilization genes. *J. Bacteriol*. **183**, 7198-205 (2001).

180. Schwalm, N.D., 3rd, G.E. Townsend, 2nd, and E.A. Groisman, Multiple Signals Govern Utilization of a Polysaccharide in the Gut Bacterium Bacteroides thetaiotaomicron. *MBio* **7**, e01342-16 (2016).

181. Xu, J., et al., A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis. *Science* **299**, 2074-2076 (2003).

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

183. Backhed, F., et al., Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-20 (2005).

184. Leonard, S.P., et al., Genetic Engineering of Bee Gut Microbiome Bacteria with a Toolkit for Modular Assembly of Broad-Host-Range Plasmids. *ACS Synth. Biol*. **7**, 1279-1290 (2018).

185. Meyer, A.J., et al., Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol*. **15**, 196-204 (2019).

186. Cuiv, P.O., et al., Isolation of Genetically Tractable Most-Wanted Bacteria by Metaparental Mating. *Sci. Rep*. **5**, 13282 (2015).

187. Snydman, D.R., et al., Trends in antimicrobial resistance among Bacteroides species and Parabacteroides species in the United States from 2010-2012 with comparison to 2008-2009. *Anaerobe* **43**, 21-26 (2017).

188. Garcia-Bayona, L. and L.E. Comstock, Streamlined Genetic Manipulation of Diverse Bacteroides and Parabacteroides Isolates from the Human Gut Microbiota. *MBio* **10**, e01762-19 (2019).

189. Joglekar, P., et al., Genetic Variation of the SusC/SusD Homologs from a Polysaccharide Utilization Locus Underlies Divergent Fructan Specificities and Functional Adaptation in Bacteroides thetaiotaomicron Strains. *mSphere* **3**, e00185-18 (2018).

190. Wang, J., et al., Characterization of a Bacteroides mobilizable transposon, NBU2, which carries a functional lincomycin resistance gene. *J. Bacteriol*. **182**, 3559-71 (2000).

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

192. Lim, B., et al., Engineered Regulatory Systems Modulate Gene Expression of Human Commensals in the Gut. *Cell* **169**, 547-558 (2017).

193. Mimee, M., et al., Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell Syst*. **1**, 62-71 (2015).

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

Chapter II

A Genetically Adaptable Strategy for Ribose and Nucleoside Scavenging in a Human Gut Symbiont Plays A Diet-Dependent Role in Colon Colonization

Abstract

Efficient nutrient acquisition in the competitive human gut is essential for microbial persistence. While polysaccharides have been well-studied nutrients for the gut microbiome, other resources such as nucleic acids and nucleosides are less studied. We describe a series of ribose utilization systems (RUSs) that are broadly represented in Bacteroidetes and appear to have diversified to allow access to ribose from a variety of substrates. One *Bacteroides thetaiotaomicron* RUS variant is critical for competitive gut colonization in a diet-specific fashion. We used molecular genetics to probe the required functions and nature of the nutrient source(s) underlying this phenotype. Two RUS-encoded ribokinases were the only components required for this effect, presumably because they generate ribose-phosphate derivatives from products of an unlinked, but essential nucleoside phosphorylase. Our results underscore the extensive mechanisms that gut symbionts have evolved to access nutrients and the potential for unexpected dependencies between systems that mediate colonization and persistence.

Introduction

Symbiotic microorganisms that inhabit the human intestine complement digestive capacity in numerous ways, with the most mechanistically understood examples involving degradation of diverse dietary polysaccharides[1]. In contrast, the digestive fates of nucleic acids (from diet, host or microbial origin) and their component ribo- and deoxyribonucleosides are less understood, as are their contributions to gut microbiota community structure and physiology. Mutualistic *Lactobacillus*[2] and *Bifidobacterium*[3] as well as pathogenic and non-pathogenic *Escherichia coli*[4] and *Salmonella enterica*[5] have characterized ribose degrading systems. Additional systems containing nucleoside-cleaving enzymes have been defined in *E. coli* and fecal isolates of *Corynebacterium*[6,7]. In *E. coli,* DNA can serve as a sole-carbon source through

the action of competence genes and exonucleases[8,9]. Mechanisms for assimilating exogenous RNA have not been explored.

Members of the phylum *Bacteroidetes* constitute a major portion of bacteria in the human gut, with individual species devoting large portions of their genomes towards carbohydrate utilization via coordinately regulated polysaccharide utilization loci (PULs). A number of these PULs targeting dietary polysaccharides from plant cell walls or fermented foods have been thoroughly characterized[10-14]. Other characterized PULs degrade infrequent dietary substrates such as agarose and porphyran in edible seaweed[15,16] or host-derived glycans such as those in mucus[17,18]. Despite variations in the substrates they target, the cellular "Sus-like systems" encoded by *Bacteroidetes* PULs are similarly patterned—each containing one or more TonB-dependent receptors (SusC homologs) and corresponding substrate binding lipoproteins (SusD homologs). These two proteins form a complex[19] and work in concert with a variable repertoire of carbohydrate-degrading enzymes, substrate binding proteins and regulators to bind, degrade and import substrates. Despite these studies, many identified PULs within genomes of gut and environmental *Bacteroidetes* lack existing knowledge of their target substrates[20], suggesting that they have evolved to target a broader range of nutrients beyond the common plant and host polysaccharides that have been evaluated[21,22].

Here we describe a ribose-responsive PUL in the human gut symbiont *Bacteroides thetaiotaomicron* (*Bt*). Variants of this PUL exist in a diverse range of human gut and environmental *Bacteroidetes*, but based on enzymatic diversity have likely evolved to target a variety of different ribose-containing nutrients. Using *Bt* as a model, we investigated the functions of this PUL *in vivo* in multiple diet conditions and *in vitro* in defined media. We show that this PUL is essential for utilization of ribose through the activity of two ribokinases, enzymes that catalyze formation of ribose-5-phosphate from ribose or ribose-1,5-bisphosphate from the product of a genomically unlinked nucleoside phosphorylase that is required for growth on nucleosides. The ability to catabolize ribose through PUL-encoded functions and the unlinked nucleoside phosphorylase confers a strong, diet-specific competitive advantage to *Bt in vivo*. This suggests a model in which a diet-specific nucleoside-scavenging pathway has become dependent on cellular ribokinases, which are critical for creating phosphorylated ribose intermediates and are persistently activated in the gut by an unknown signal. Our results reveal that a variety of host-associated and terrestrial bacteria have evolved mechanisms to scavenge

ribose and nucleosides that are important for colonization. The common regulation of a family of highly diversified PULs by ribose, which occurs in nucleic acids, co-factors, modifications (ADP- and poly-ADP-ribose), bacteriocins, and bacterial capsules, suggests that these systems have adapted at the level of encoded enzymes to release ribose from varied sources, diversifying the nutrient niches available to these bacteria. However, the results of our *in vivo* studies highlight that underlying mechanisms for observed colonization advantages are context specific and not always directly attributable to the most obvious function performed or predicted by a particular system.

Results

A ribose-inducible gene cluster is highly active in vivo and required for fitness in a diet-dependent fashion

Members of the human gut *Bacteroidetes* typically encode coordinated degradative functions within discrete polysaccharide utilization loci (PULs), facilitating identification of components that work together to access particular nutrients[23]. Previous work using gnotobiotic mice colonized with only *B. thetaiotaomicron* (*Bt*) identified one such locus (*BT2803-2809*) for which all individual genes are upregulated between 10- and 139-fold in mice fed high or low fiber diets (Figure 2.1A). During low fiber, *Bt's* physiology shifts to expression of genes involved in host glycan foraging[17,24,25]. Thus, expression of *BT2803-09* in the absence of dietary fiber suggested that it may also target endogenous nutrients.

Typically, PULs involved in host glycan foraging encode enzymes required for liberating sugars from mucins and other glycoconjugates (fucosidases, sulfatases, etc.), but the content of the *BT2803-09* PUL was different in several ways (Figure 2.1B). Three predicted enzymes (one nucleoside hydrolase, two ribokinases) suggested a role in assimilating ribose from substrate(s) such as nucleosides. A previous study determined that *Bt* grows on ribose[21], but the genes involved, relevant source(s) of ribose, and whether enzymatic liberation is required from complex substrates were not explored. The immediate upstream gene (*BT2802*) is predicted to have DNA-binding motifs and may act as a regulator, but shares no homology to regulators previously associated with PULs. In addition to the enzymes noted above, other PUL genes encode homologs of the *Bacteroides* SusC and SusD outer-membrane proteins (*BT2805, BT2806*), a glycoside hydrolase of unassigned family and function (*BT2807*), a predicted

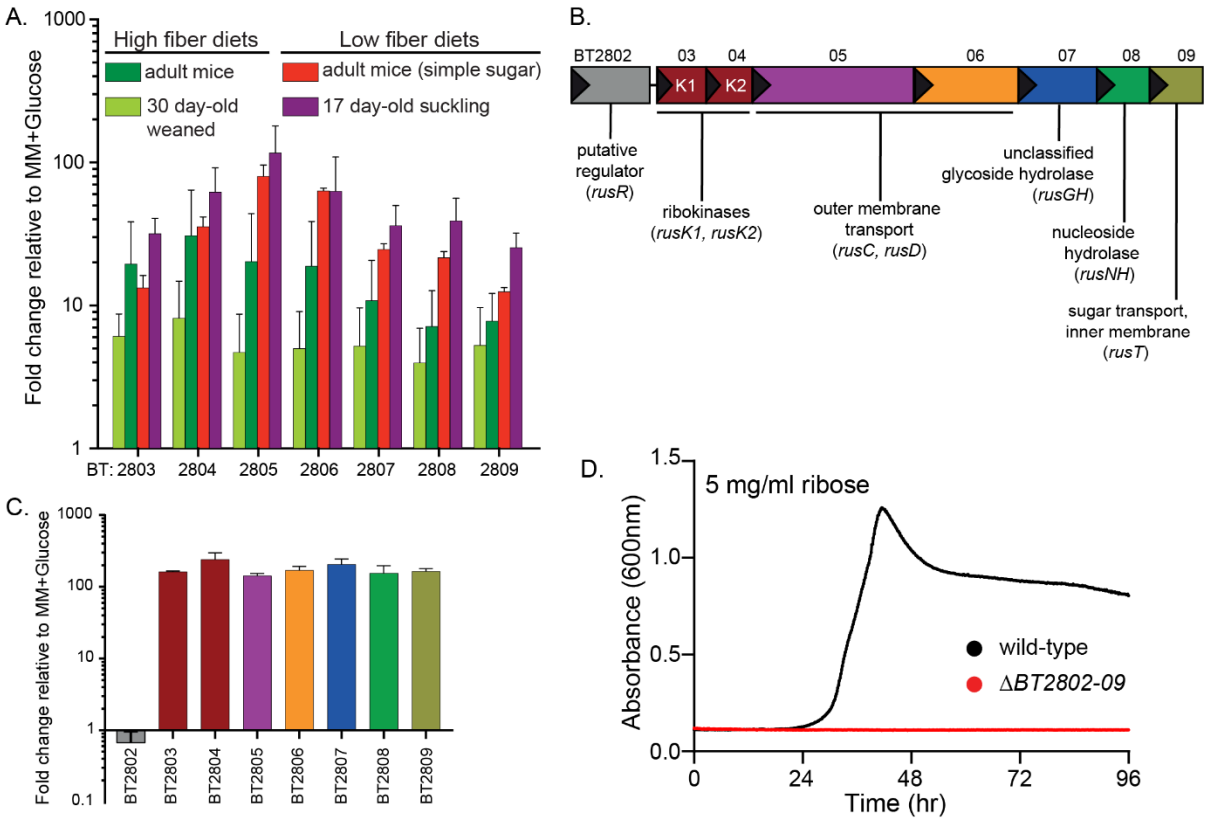


Figure 2.1 Bt upregulates a PUL for ribose metabolism in vivo and in vitro in response to ribose.

(A) In vivo Genechip data showing fold-change relative to in vitro growth MM, plus glucose for BT2803-09 in mice fed high fiber (dark and light green bars) or low fiber diets, including pre-weaned, suckling mice (red and purple bars, respectively). (B) Organization of the rus locus with locus tag numbers, names and predicted functions. (C) In vitro transcriptional response of Bt rus genes in MM-ribose compared to MM-glucose reference (n=3, error bars are SD of the mean). (D) Growth in MM-ribose (5 mg/ml) for wild-type Bt (black) or a strain lacking rus (red) (minimum of n=5 separate replicates).

nucleoside hydrolase (*BT2808*), and a sugar permease (*BT2809*).

The enzymes encoded in this PUL suggested the hypothesis that it is responsible for *Bt's* ability to catabolize ribose and possibly liberate it from more complex sources such as nucleosides. To test if this gene cluster is transcriptionally responsive to growth on ribose, we performed *in vitro* growth in minimal-medium (MM) containing ribose as the sole carbon source and measured expression of *BT2803-09*. All genes were activated 142-240 fold by ribose compared to a MM-glucose reference (Figure 2.1C). Other mono- and disaccharides did not activate this PUL as sole carbon sources (Figure 2.2A). We next examined the requirement for this locus by deleting *BT2802-09*. Loss of the PUL eliminated growth on free ribose (Figure 2.1D) but did not affect growth on non-ribose substrates (Table 2.1, shown in Methods). Based on these findings, we classified this PUL as the *Bt* ribose utilization system, *rus*, with gene annotations listed in (Figure 2.1B).

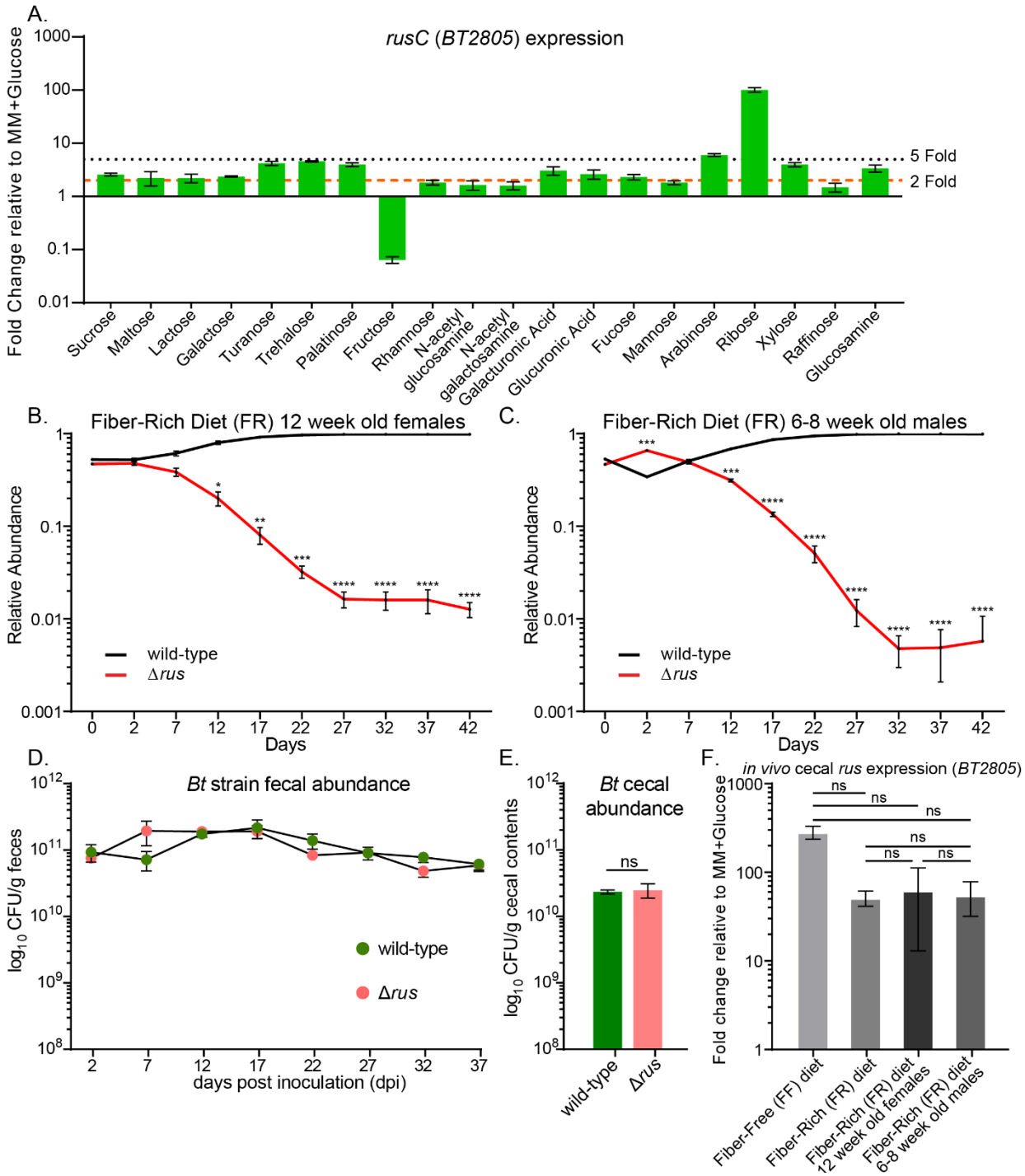


Figure 2.2 In vitro rus activation specificity and supplemental in vivo competitions.

Data is related to Figures 2.1 and 2.3 demonstrating the in vitro activation specificity of the rus locus and further in vivo competition or monocolonization experiments. (A) rusC (BT2805) expression during mid-log phase growth in MM containing a sole carbon source from either mono-, di-, or trisaccharide's compared to growth on MM-glucose, the dashed orange line represents a 2-fold upregulation, while the dotted black line is a 5-fold upregulation, error bars represent the SD of n=3 replicates. (B-F), in vivo competitions all performed in germfree Swiss Webster mice and fed a fiber rich (FR) diet. Relative abundance was enumerated by qRT-PCR of unique chromosomally encoded barcodes for wild type (black line) vs. Δrus strain (red line) in FR diet in (B) 12 week old female mice or (C) 6-8 week old male mice. (D) 6-8 week old female mice mono-associated with either wild-type Bt (pink circles) or Δrus strain (green circles) where absolute abundance was assayed by dilution plating from fresh fecal

samples to obtain CFU/g feces. Error bars show SEM of the biological average of n=3 mice. (E) As in (D), but enumerating absolute cecal abundance in mono-associated mice. (F) Wild type rus expression from cecal contents from experiments in Fig. 2A-B and Fig. S2A-B, with SEM of each sample. P-values were calculated for B-F by Student's t test, specifically in B-E, the relative abundance of each strain is tested with an (*) indicates a statistically significant difference of (P<0.05), while (**) represents (P<0.01), and (***) indicates (P<0.001) and (****) is used to express (P<0.0001) or no significant (ns). In (F) the values used for t test compare all other samples to rus expression in 6-8 week old females on FR diet shown as lines above the bars.

Based on these findings, we classified this PUL as the *Bt* ribose utilization system, rus, with gene annotations listed in (Figure 2.1B).

Because *rus* exhibits high transcriptional activity in the gnotobiotic mouse gut and is elevated in fiber starved mice, we next hypothesized that the ability to utilize endogenous sources of ribose is advantageous *in vivo* during fiber-deficient diets. To test this, we inoculated 6-8 week old germfree (GF) female Swiss-Webster mice with an equal mixture of wild-type and Δ*rus Bt* strains and maintained mice on either a fiber-rich (FR) diet containing several unprocessed plant-derived fiber polysaccharides or an accessible fiber-free (FF) diet consisting mainly of glucose, protein, lipids, and cellulose[26]. We measured the relative abundance by qPCR of each strain for 42 days in DNA extracted from feces. Opposite to our initial hypothesis, the Δ*rus* strain was strongly outcompeted (~100-fold) in mice fed the FR diet (Figure 2.3A). In contrast, in mice fed the FF diet, Δ*rus* exhibited similar abundance to wild-type *Bt* (Figure 2.3B). A similar competitive defect of the Δ*rus* strain in mice fed the FR diet was observed in separate experiments with 12-week-old female and 6-8 week old male mice (Figure 2.2B-C), suggesting the effect is not influenced by sex or age within the range tested. The FR diet-associated defect was not due to lack of colonization or persistence, as the levels of each strain were similar over time in mice colonized with either strain alone (Figure 2.2D-E). Additionally, the defect in the FR diet could not be attributed to the wild-type strain exhibiting different expression of the *rus* PUL, as wild-type *Bt* exhibited similarly high levels of *rus* expression in mice fed either diet when present alone or in competition with the Δ*rus* mutant (Figure 2.2F).

Gas chromatography-mass spectrometry (GC-MS) analysis of the diets revealed ribose present only in the FR diet, in levels similar to other common monosaccharides, in an acid-hydrolyzable (*i.e.*, covalently linked) but not free form. This suggested the presence of a ribose-containing molecule(s), such as RNA, nucleosides or cofactors (Figure 2.4A). In cecal contents of FR diet-fed mice mono-colonized with wild-type *Bt* or Δ*rus* strains, ribose was not detectable above our limit of detection (LOD) (Figure 2.4B). However, the LOD for ribose in the cecal contents was near the amount observed in the uneaten FR diet, raising the specter that substantial

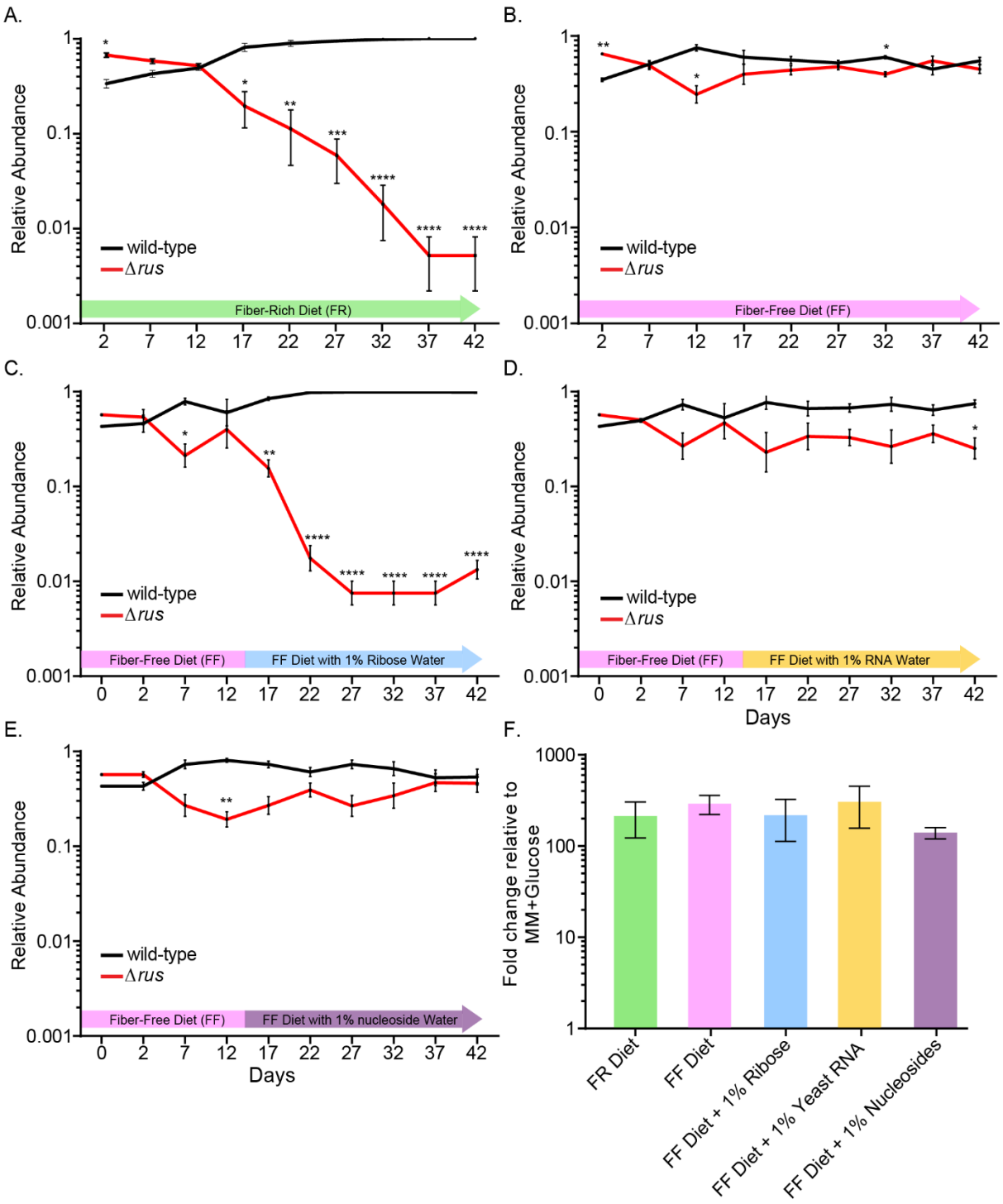


Figure 2.3 The Bt rus locus confers a competitive advantage in vivo in a diet-dependent context.

(A-E) Log-scale relative abundance of wild type (black line) and Δrus (red line) strains enumerated by qRT-PCR from feces of 6-8 week old germfree Swiss-Webster mice. (A) Mice fed a high fiber diet (green arrow; n = 4 mice). (B) Mice pre-fed a fiber-free (FF) diet for one week prior to colonization and maintained for 42d (pink arrow). (C-E) Same diet and strain competition as in (B), but mice were given water containing 1% w/v ribose (C), 1% w/v RNA from type IV Torula yeast tRNAs (D), or a 1% w/v mixture of nucleosides (0.25% each of uridine, cytidine, thymidine, and 5-methyl uridine) (E). The period of water supplementation is shaded either blue, orange, or purple. (F) rusC transcript levels measured by qRT-PCR from cecal contents of mice in panels A-

*E. For all panels, the mean ± SEM is shown at each time point. In panels A-E, asterisks indicate significant differences (*P<0.05, **P<0.01, ***P<0.001, and ****P<0.0001) calculated by Student's t test between strains at the same day.*

amounts reach the cecum but are obscured. This ambiguity about the amount of diet-derived ribose *in vivo* prompted us to test if different sources of dietary ribose impact *Bt* in the gut. We colonized three separate groups of GF mice with a mixture of wild-type and Δ*rus* strains and maintained them on the FF diet. After 14 days of stable competition between strains, water was supplemented with either 1% ribose, 1% RNA, or 1% pyrimidine nucleosides (purines were not tested due to insolubility). The results clearly show that free ribose in the water exerts an effect against the Δ*rus* strain similar in magnitude to the defect in mice fed the FR diet (Figure 2.3C). Little or no defect was observed in mice provided water containing RNA or nucleosides (Figure 2.3D-E) despite increased acid-hydrolyzable ribose being detectable in the cecum (Figure 2.4C). There was comparable expression of the *rus* locus in all conditions, suggesting *rus* expression differences did not account for different fitness outcomes (Figure 2.3F).

A subset of ribose-utilization functions is required for competitive colonization in mice

The experiments described so far used a mutant lacking all 8 *rus* genes, but only a subset of the functions may be important for competition. We therefore took a molecular genetic approach to more precisely probe the required functions and get a clearer idea of the nature of the important nutrient(s) in the FR diet. We constructed single and double gene deletions based on predicted functionality (Figure 2.1B) and performed additional competitive colonization experiments in FR diet-fed mice. Each individual mouse group was inoculated with wild-type *Bt* and one of the following competing strains: Δ*rusK1/2*, Δ*rusC/D*, Δ*rusGH/NH*, Δ*rusT*, or Δ*rusR*, to test the predicted contributions of phosphorylation, outer membrane transport, hydrolase activity, inner membrane transport and regulation, respectively. Surprisingly, only the Δ*rusK1/K2* strain which lacks both predicted ribokinases, exhibited a competitive fitness defect similar to the full Δ*rus* mutant (Figure 2.5A). In contrast, the other deletion strains exhibited equal or better competition compared to wild-type (Figure 2.5B and Figure 2.4D-F). These results show that the required functions underlying the competitive defect in the Δ*rus* strain are encoded by the *rusK1* or *rusK2* genes, while other functions provide no advantage and perhaps a fitness disadvantage on the FR diet. We speculate that the advantages exhibited by the other mutants are due to not incurring the cost of expressing these proteins in a condition where they

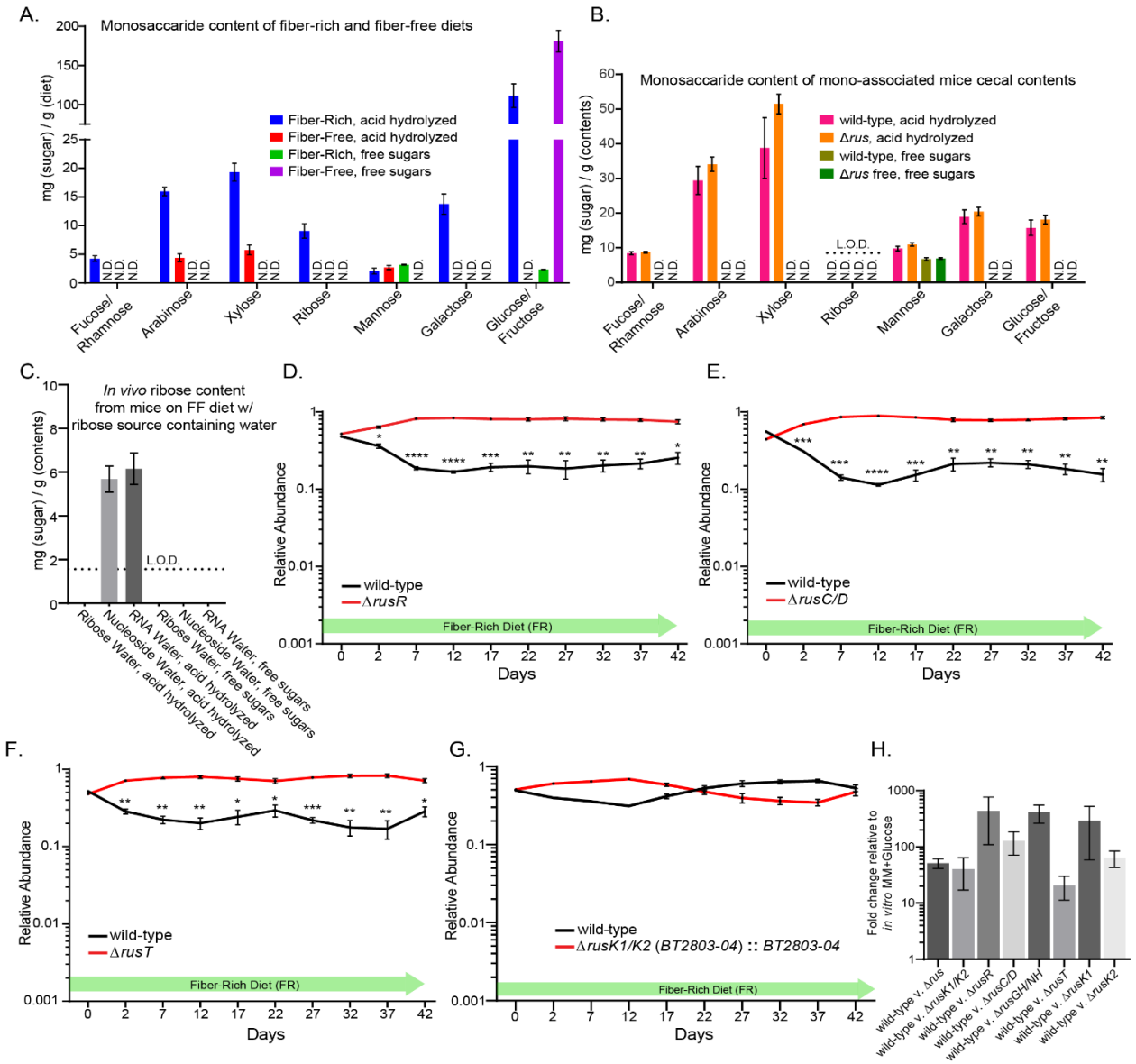


Figure 2.4 Monosaccharide content of diets and cecal contents, additional in vivo experiments and in vivo complementation.

Data is related to Figures 2.3 and 2.5. (A-C) GC/MS analysis of free and linked (acid hydrolyzed) monosaccharides from fiber-rich (FR) or fiber-free (FF) diets (B) cecal contents of wild type or Δrus mono-associated 6-8 week old, female Swiss Webster mice on the FR diet or (C) ribose content only analysis from mice maintained on the FF diet provided water containing ribose, nucleosides, or RNA. Data is presented as mg of sugar per gram of diet or cecal contents. (A) Blue and green bars represent the FR diet linked or free respectively, while the red or purple bars represent the FF diet linked or free respectively. (B) The pink and olive green bars represent wild type linked and free respectively, while the orange and green represent linked or free for the Δrus colonized mice. In figures A-C error bars show the SEM of n=3 biological replicates. N.D. indicates that the sugar was not detectable above our limit of detection (L.O.D.) which is shown as a dotted line above the ribose bars only in (B), or as a dotted line in (C). (D-G) In vivo competition with 6-8 week old Swiss Webster mice inoculated with both wild-type Bt (black line) against one other mutant or complemented strain (red line) with fecal relative abundance shown as enumerated by qRT-PCR with error bars showing the SEM. (D) Wild-type Bt and ΔrusR strain, (E) wild-type Bt and ΔrusC/D strain, (F) wild-type Bt and ΔrusT strain; all deletion strains in D-F displayed slight, but significant competitive advantages over wild-type Bt with n=4 biological replicates. (G) Competition of wild-type Bt against the complementation of the ΔrusK1/K2 strain (ΔBT2803-04::BT2803-04) n=5 biological replicates. (H) in vivo rus transcript measured by qRT-PCR from cecal contents where the gene deleted in the mutant strain was probed to assay rus transcript levels, thus transcript is representative of the wild-type strain only, in the case of Δrus and ΔrusR, the rusC gene was probed. In panels D-G, P-values were calculated by Student's t test. An (*) indicates a statistically significant

difference of (P<0.05), while (**) represents (P<0.01), and (***) indicates (P<0.001) and (****) is used to express (P<0.0001) in the relative abundance of each strain compared to the other at each separate time point.

do not participate in acquiring nutrients, a phenomenon observed with *Bt* fungal mannan utilization[11]. To test if one ribokinase is most important *in vivo,* we repeated the above competition with single Δ*rusK1* and Δ*rusK2* deletion strains. Each of these single kinase mutants also competed better than wild-type, suggesting functional redundancy in this context (Figure 2.5C-D). Genetic complementation of the Δ*rusK1/K2* strain restored the competitive ability of the defective mutant strain, allowing equal competition against wild-type (Figure 2.4G). Finally, variations in competitive behavior were not attributable to significant differences in *rus* expression in wild-type *Bt* for any of the *in vivo* competitions (Figure 2.4H).

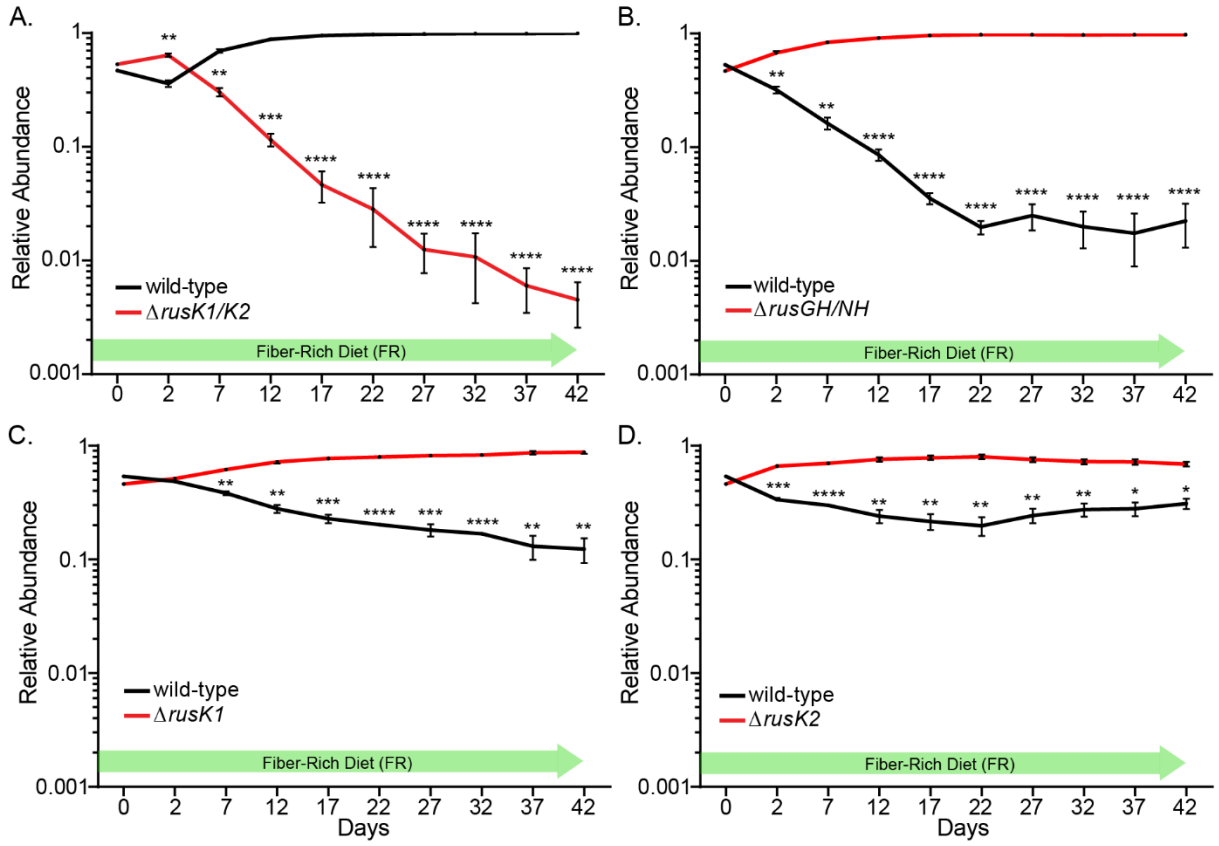


Figure 2.5 Ribokinases are required for competitive advantage in vivo.

*(A-D) In vivo competition between wild-type Bt (black line) and individual mutant strains indicated (red line) in 6-8 week old, germfree Swiss-Webster mice fed the FR diet. Relative abundance is displayed as in Fig. 2. In all panels, the mean of n=4 biological replicates ± SEM is shown. Asterisks indicate significant differences (*P<0.05, **P<0.01, ***P<0.001, and ****P<0.0001) calculated by Student's t test between strains at the same day.*

Rus functions are required for sensing and utilization of RNA, nucleosides and other nutrients in vitro

The results described above indicate a diet-specific advantage for *Bt* strains containing *rus*-encoded ribokinases. To further define this system's function, we tested our panel of deletion

mutants in a variety of growth conditions, including free ribose, nucleosides, RNA, and other sources of ribose. Consistent with *in vivo* data, a mutant lacking both *rusK1* and *rusK*2 could not grow on free ribose (Figure 2.6A). Arguing against purely redundant functions, the mutant lacking just *rusK2* displayed a complete loss of growth phenotype, while a mutant lacking only *rusK1* reproducibly displayed a substantial growth lag, but eventually grew with slightly slower rate than wild-type (Figure 2.6B-C). The delayed growth of this mutant may be due to a genetic suppressor mutation since cells that eventually grew were able to grow quickly on ribose after being isolated and passaged in rich media (Figure 2.7A). Deletion of the flanking gene *rusR*, a candidate transcriptional regulator, also failed to grow on ribose. Suggesting that although it is not transcriptionally activated in response to ribose, it plays an essential role in ribose catabolism (Figure 2.6D). The Δ*rusT* strain exhibited increased lag, slower growth rate, and lower overall growth compared to wild-type (Figure 2.6E).

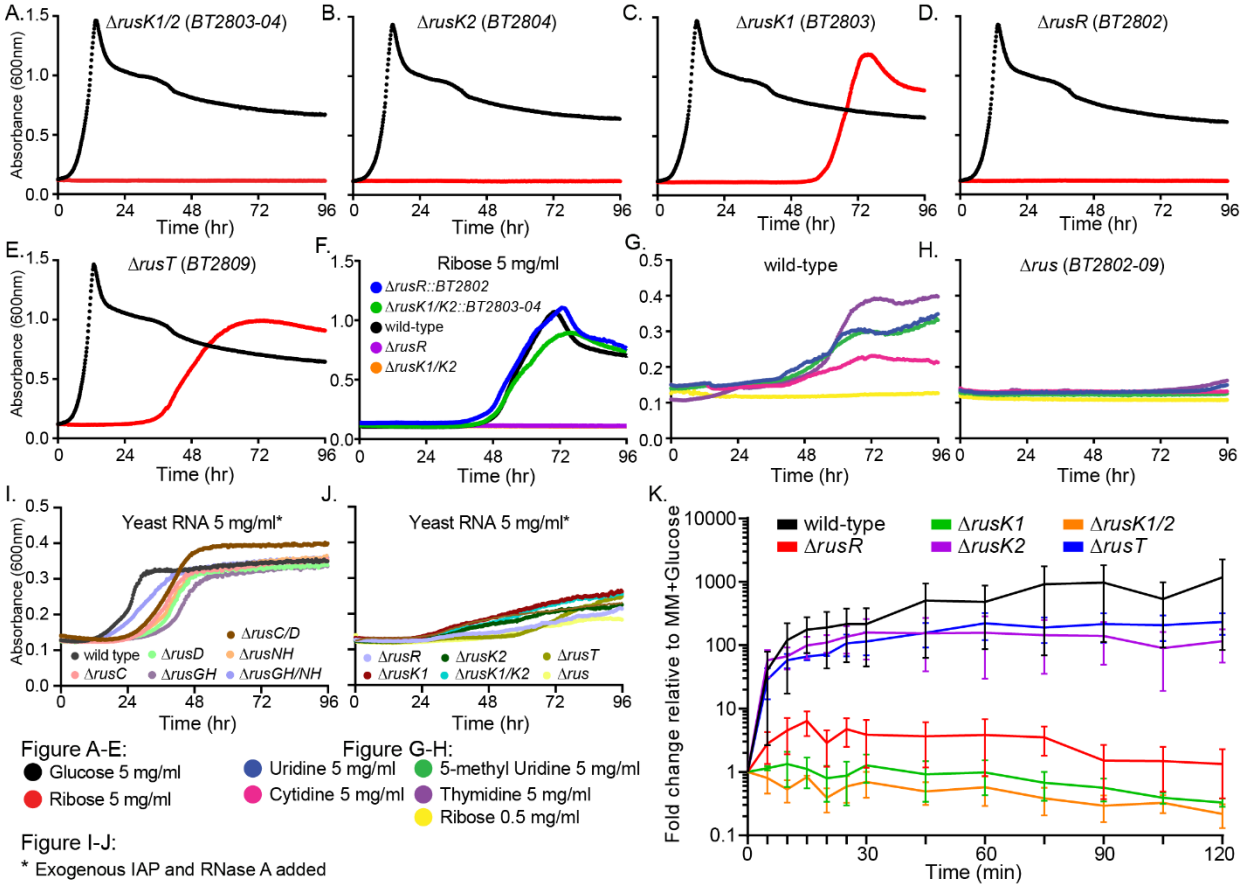


Figure 2.6 The Bt rus PUL encodes functions required for growth and transcript activation on ribose containing nutrients.

(A-E) Growth curves of the individual rus deletion strains indicated (red lines) with growth on glucose (black line) as a control. (F) Growth of genetically complemented ΔrusR (blue line) and ΔrusK1/K2 (green line) on ribose, showing restored growth compared to wild-type (black line) and corresponding deletions (purple and orange lines). (G,H) Wild-type or Δrus growth on nucleosides in the presence of 0.5 mg/ml ribose (yellow line is medium with only 0.5 mg/ml ribose). Legend in bottom right shows

substrates tested. (I,J) Wild-type Bt and rus deletion strains grown in MM, plus 5 mg/ml yeast RNA with RNase A and IAP. Mutants with similar growth phenotypes as wild-type (I), are compared to mutants with reductions in growth (J). (K) Bt rusC transcript activation measured by qRT-PCR after mid-log phase cells grown in MM-glucose were washed in carbon-free medium and transferred to MM-ribose. For all strains, samples were taken every 5 minutes for 30 minutes, then every 15 minutes until 120 minutes. Strains are tinted according to the key provided. Data shown are the mean of n=3 separate experiments ± SEM.

Unlike the Δ*rusK1* mutant this mutant did not exhibit increased growth after passage (not shown), suggesting that suppressor mutations are not involved, but perhaps a lower-affinity sugar permease imports ribose less efficiently. All of the other single or double deletion mutants (Δ*rusC*, Δ*rusD*, Δ*rusC/D*, Δ*rusGH*, Δ*rusNH*, Δ*rusGH/NH*), exhibited no measurable differences in growth on ribose compared to wild-type *Bt* (Figure 2.7B-G and Table 2.1). The growth defects associated with Δ*rus*K1/K2 and Δ*rusR* were fully repaired by a single, complementing copy of each gene *in trans* (Figure 2.6F, *rusT* was not attempted).

Owing to their larger and more complex structure, we hypothesized that utilization of covalently linked ribose sources would require the additional *rus*-encoded outer membrane transport and hydrolase functions. To test this, we assayed growth of our *rus* mutants and wild-type *Bt* on nucleosides and RNA. Wild-type *Bt* displayed no or poor growth on all nucleosides tested as well as on RNA (Figure 2.7H-I and Table 2.1). We hypothesized that free ribose may be required to activate transcription of the *rus* locus, generating proteins necessary for catabolism of these substrates. We determined a concentration (0.5 mg/ml) at which ribose elicited strong *rus* expression but little measurable growth based on absorption measurement (Figure 2.7J-K). We then re-evaluated the ability of wild-type *Bt* to grow on nucleosides, observing considerably higher levels of growth on pyrimidine nucleosides (Figure 2.6G). Growth was comparatively poor relative to growth on ribose, increased growth was not observed by doubling nucleoside concentrations, suggesting that something else related to nucleoside catabolism limits growth (Figure 2.7L).

Importantly, growth on nucleosides was eliminated in mutants lacking the full *rus* locus (Figure 2.6H), either or both ribokinases, the candidate regulator (*rusR*) or the putative transporter (*rusT*) (Figure 2.7M-Q). Growth on RNA was not observed after addition of ribose, suggesting that *Bt* does not produce sufficient extracellular RNAse and phosphatase enzymes required to liberate nucleosides. Therefore, we tested if exogenously added RNase A and intestinal alkaline phosphatase (IAP), both present in the gut from pancreatic secretions (RNAse) or the enterocyte brush boarder (IAP), could enhance growth on RNA at physiologically relevant concentrations[27,28]. These enzymes supported appreciably more growth on RNA (Figure 2.6I),

not attributable to *Bt* growing on the exogenous enzymes themselves (Figure 2.7R). As with individual nucleosides, reductions or eliminations in growth on enzyme-degraded RNA were observed in mutants lacking the entire *rus* locus, *rusK1*, *rusK2*, *rusK1/K2*, *rusT*, and *rusR*

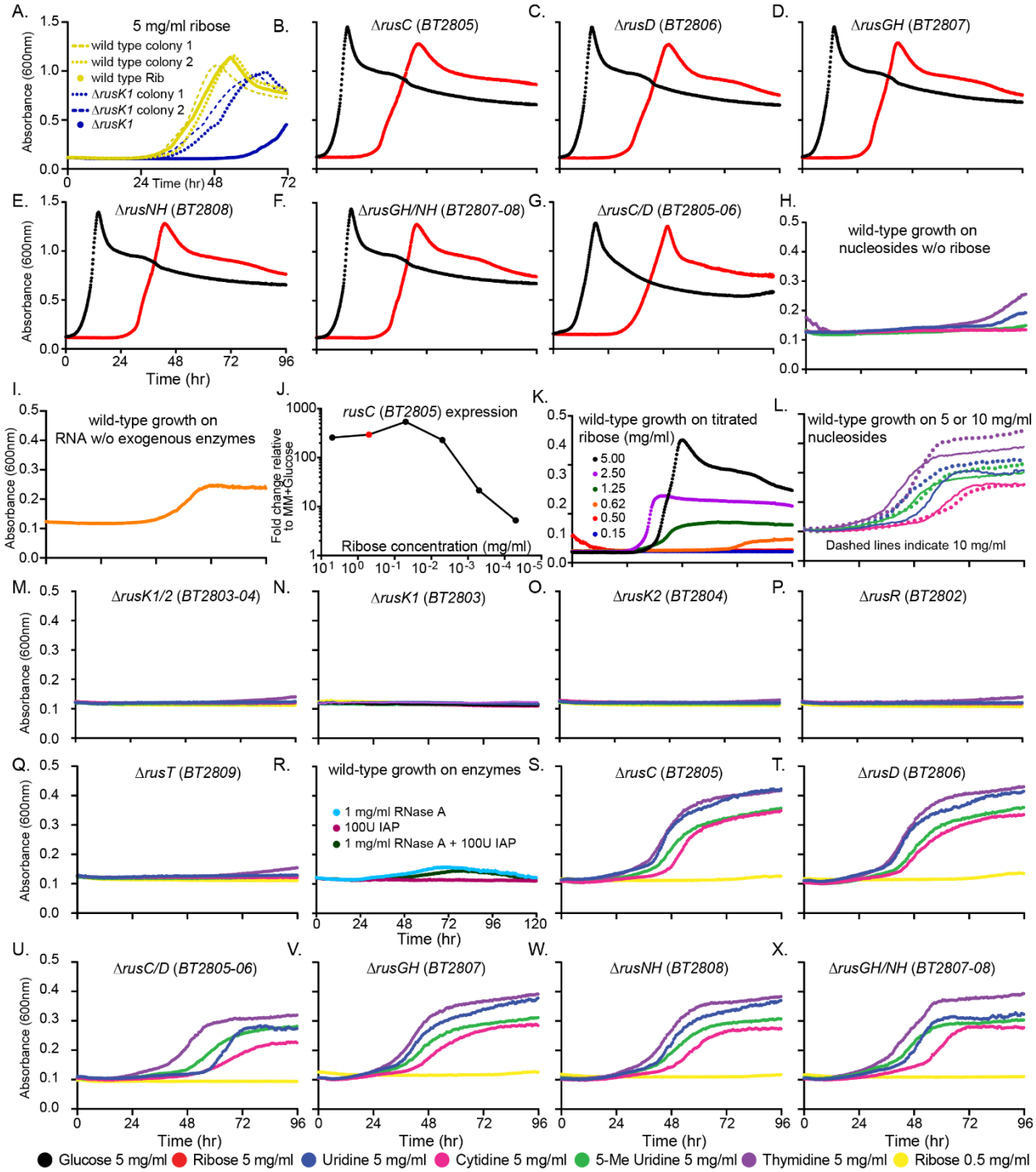


Figure 2.7 Detailed growth of rus mutants on ribose, nucleosides, and RNA.

Data is related to Fig. 2.6 (A) Wild-type (maize, solid line) or ΔrusK1 (blue, solid line) strains grown in MM-ribose or wild-type and ΔrusK1 strains that had been previously grown on MM-ribose, struck on BHI-blood plates, two separate colonies picked into rich media, (TYG) and then grown in MM-ribose again shown as maize dashed lines (wild type) or blue dashed line (ΔrusK1) to

check if the delayed growth phenotype associated with the ΔrusK1 strain was the product of a genetic suppressor mutation or similar epigenetic/reprogramming for MM-ribose growth, all strains were grown for 72 hours. (B-G) Growth of rus deletion strains exhibiting similar growth as wild-type on MM-ribose (red line) with no obvious growth defects, growth on MM-glucose (black line) is also shown for comparison: ΔrusC (B), ΔrusD (C), ΔrusGH (D), ΔrusNH (E), ΔrusGH/NH (F), ΔrusC/D (G). For panel (H) wild-type Bt was grown in MM containing 5 mg/ml of one of the following nucleosides (uridine, blue line; cytidine, pink line; 5-methyl uridine, green line; or thymidine, purple line) without any ribose added. (I)Wild-type Bt growth on MM containing 5 mg/ml of yeast RNA without any exogenous enzymes. (J) rusC transcriptional response when wild-type Bt was exposed to titrated amounts of ribose (mg/ml) where each data point represents a different 10-fold dilution of ribose. The red dot represents 0.5 mg/ml ribose which induces rus activation to comparable levels as 5 mg/ml ribose compared to growth in MM-glucose. (K) Wild-type growth on different concentrations of MM containing ribose at the following concentrations (mg/ml): 5, black line; 2.5, purple; 1.25, green; 0.625, orange; 0.5, red; or 0.15, blue. Growth was not detectable at levels ≤0.5 mg/ml. (L) Wild-type growth on MM containing nucleosides at concentrations of 5 mg/ml (solid lines) or 10 mg/ml (dashed lines) in the presence of 0.5 mg/ml ribose. Individual nucleoside growths are colored same as (G). rus deletion strains exhibiting a complete lack of growth phenotype on all nucleosides tested are shown in (M-Q) as follows with lines colored same as (H): ΔrusK1/2 (M), ΔrusK1 (N), ΔrusK2 (O), ΔrusR (P), and ΔrusT (Q). (R) Wild-type growth curves on enzyme only controls for media supplemented with 1 mg/ml RNase A (teal line), 100 U calf Intestinal Alkaline Phosphatase (IAP) (plum line), or 1 mg/ml RNase A and 100 U IAP (green line), demonstrating that without a carbon source, these enzymes do not support growth of Bt. (S-X) Growth of deletion strains on nucleosides with 0.5 mg/ml ribose added where growth was similar to wild type levels: ΔrusC (S), ΔrusD (T), ΔrusC/D (U), ΔrusGH (V), ΔrusNH (W), or ΔrusGH/NH (X). For (B-I) and (K-X), growth is plotted over 96 hours with dashes denoting 24 hour increments.

(Figure 2.6J). Further, mutants lacking predicted transport and hydrolytic functions grew similarly to wild-type on both nucleosides and degraded RNA (Figure 2.6I and Figure 2.7S-X). In addition, we determined that *Bt* utilizes deoxyribose and lyxose, as well as ADP-ribose, UDP-galactose and UDP-α-glucose. All of these required the presence of a low amount of ribose and the *rus* locus, while 21 other substrates did not support *Bt* growth under any conditions tested (Table 2.1).

Based on our mutant growth phenotypes, we sought to determine if the genes required for ribose growth were also required for activating expression of *rus.* We examined the kinetics of *rus* transcriptional responses when *Bt* was exposed to ribose, an assay that allows us to measure response independent of ability to grow on ribose. Interestingly, the Δ*rusK2* strain, which cannot grow on ribose, generated transcript at a similar rate/level to wild-type up to 2h (Figure 2.6K). In contrast, the Δ*rusK1* mutant, which exhibits an extensive lag before growth on ribose, was unable to quickly generate transcript within 2h, but eventually achieves near wild-type *rus* expression once actively growing on ribose due to its suspected suppressor (Figure 2.6K and Figure 2.8D). As expected, the Δ*rusK1/K2* double mutant did not generate transcript. The Δ*rusR* mutant achieved partial (~10%) activation, supporting the hypothesis that RusR is a positive-acting regulator. The Δ*rusT* strain only has a slight defect, suggesting that another, non-specific permease can transport ribose. We also measured *rus* expression dynamics in our Δ*rusC* and Δ*rusD* strains, but failed to detect any differences compared to wild type, consistent with the lack of their requirement for ribose growth (Figure 2.8E). Finally, the nucleosides uridine and inosine did not serve as *rus*-inducing molecules in wild-type *Bt* (Figure 2.8F).

Non Rus-encoded functions are required for nucleoside utilization

The lack of a requirement for the *rus* hydrolase functions in nucleoside catabolism is noteworthy as we confirmed through biochemical experiments with recombinant enzyme that

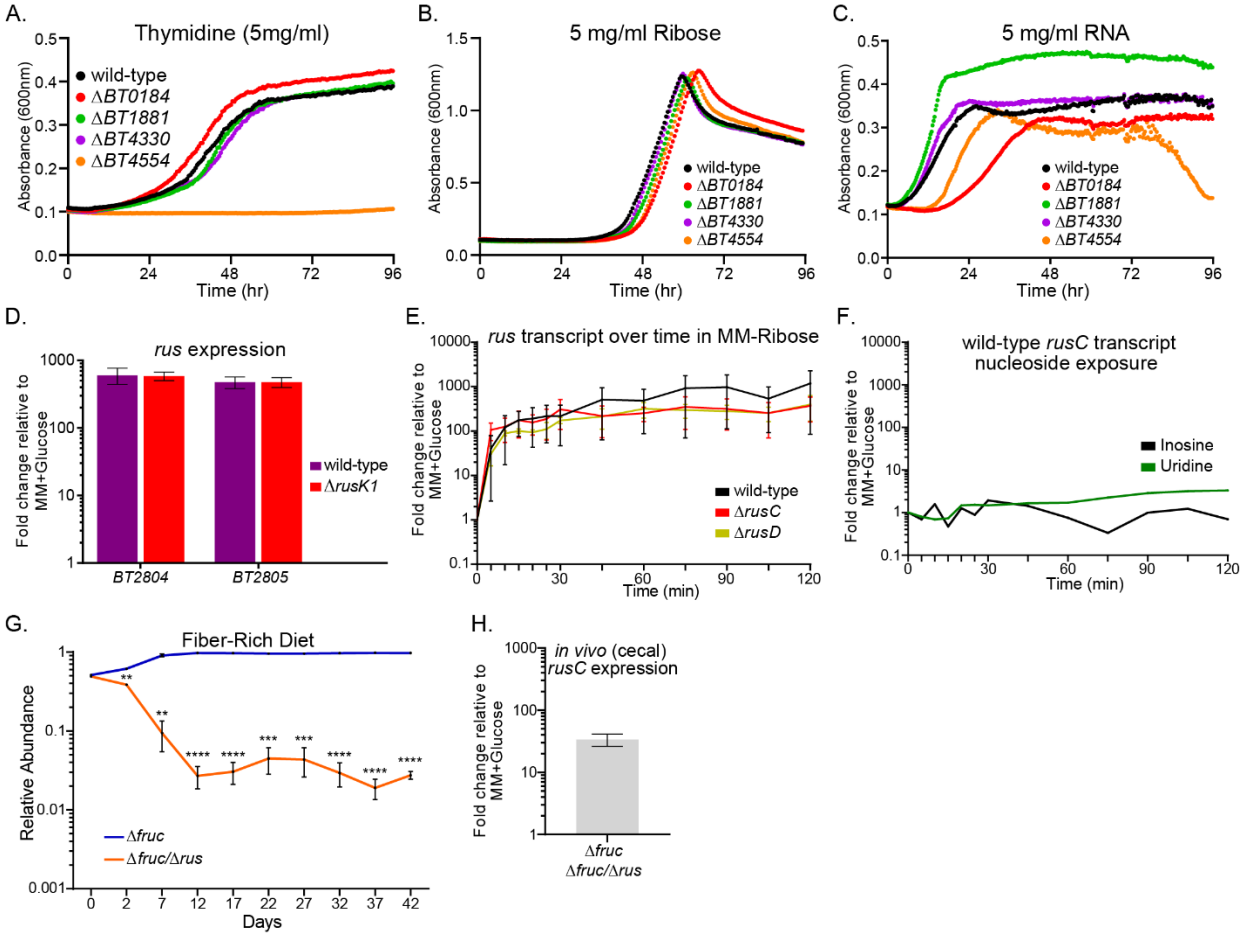


Figure 2.8 Potential crosstalk between ribose and other metabolism genes and PULs

Data is related to Fig. 2.9 (A-C) Growth of NSSs deletion strains or wild-type Bt on thymidine (A), ribose (B), or RNA (C) (with added enzymes) out to 96 hours. Strains are color coded (wild-type, black; ΔBT0184, red; ΔBT1881, green; ΔBT4330, purple; or ΔBT4554, orange). (D) rusK2 (left bars) or rusC (right bars) transcript expression of wild-type Bt (purple) or ΔrusK1 strains when grown to mid-log phase in MM-ribose compared to growth in MM-glucose, error bars represent the SD of n=3 replicates. (E) rus transcript during a time course experiment where cells were shifted from growth on MM-glucose to MM-ribose, and transcript probed over time with points every 5 minutes for the first 30 minutes and every 15 minutes after out to the conclusion at 120 minutes post-ribose exposure. For the wild-type (black line) and ΔrusD (dark yellow line) strains, the rusC gene was probed, while for the ΔrusC strain (red line), the rusD gene was probed (similar kinetics were seen in the wild-type strain when the rusD gene was used to assay rus activation, data not shown). Error bars represent the SEM of n=3 replicates performed on separate days. (F) Similar to the experiment in (E) but using nucleosides (inosine, black line or uridine, green line) and probing rusC expression to address if nucleosides could stimulate rus activation, with no response detected compared to growth in MM-glucose. (G) in vivo competition of a strain lacking the entire fructan PUL, BT1754-1765 (Δfruc, blue line) against a strain lacking both the fructan PUL and the rus PUL (Δfruc/Δrus, orange line) in 6-8 week old Swiss-Webster female mice on the FR diet. The relative fecal abundance is shown on a log scale as assayed by qRT-PCR over the course of the experiment, error bars show the SEM of n=4 mice. (H) in vivo rus expression from cecal contents of the mice from (G) probing the rusC gene, error bar shows the SEM of n=4 mice. P-values for (G) were calculated using Student's t test for the relative abundance of each strain (*) indicates a statistically significant difference of (P<0.05), while (**) represents (P<0.01), and (***) indicates (P<0.001) and (****) is used to express (P<0.0001)

RusNH is a genuine, albeit weak, nucleoside hydrolase (Table 2.2A) and that RusGH can cleave *p*-nitrophenyl-β-D-riboside (Table 2.2B-C). The lack of a phenotype associated with loss of RusNH, suggested that other functions in *Bt* are responsible for cleavage of free pyrimidine nucleosides or those liberated from RNA. To identify alternative enzymes, we searched the *Bt* genome for functions from known nucleoside scavenging systems (NSSs) and identified several candidates. We made deletions of 4 genes predicted to encode nucleoside phosphorylase (*BT1881*, *BT4554*), uridine kinase (*BT0184*) and nucleoside permease (*BT4330*) activities and tested growth of these mutants on pyrimidine nucleosides (Figure 2.9A-C and Figure 2.8A).

Table 2.2A Detailed cleavage activities of the RusNH (BT2808)

UV-based assays of nucleosides	This study		Parkin et al. 1991 Nucleoside Hydrolase from C. fasciculata	
Enzyme: BT2808 (RusNH)	**K_{cat}/k_m (s^{-1}/M^{-1})**		**K_{cat}/k_m (s^{-1}/M^{-1})**	
Substrate	Average	Standard deviation	Average	Standard deviation
Adenosine	15.06	10.75	9800	NR
Cytidine	16.75	8.96	4500	NR
Thymidine	9.63	1.10	NT	NT
5-methyl Uridine	9.91	0.88	NT	NT
Uridine	5.64	0.11	120000	NR
Inosine	9.42	1.91	76000	NR
Xanthosine	10.05	0.89	NT	NT
Guanosine	5.54	0.98	3400	NR
DeoxyUridine	8.80	2.26	NT	NT
DeoxyInosine	9.73	0.78	NT	NT
DeoxyAdenosine	15.51	8.03	NT	NT
DeoxyCytidine	5.01	0.60	NT	NT

Footnotes: NT = Not Tested, NR = Not Reported

One strain (Δ*BT4554*) displayed loss of growth on all nucleosides tested, suggesting that it encodes an essential enzyme for cleaving nucleosides and might work upstream of the *rus* functions, which are also required. The Δ*BT4330* mutant exhibited reductions in growth on uridine, cytidine, and 5-methyl uridine (Figure 2.9A-C), with only a slight defect on thymidine (Figure 2.8A). The Δ*BT0184* mutant displayed enhanced growth that began quicker than wild type and reached a higher total growth level on all nucleosides, except thymidine. This phenotype could be due to its role in 5'- phosphorylating scavenged nucleosides and shunting them towards anabolic pathways, such that its loss favors catabolism. Δ*BT1881* did not display any detectable growth defects compared to wild-type, suggesting that the product of this gene is not essential for pyrimidine catabolism.

To understand how these NSS functions may impact gut colonization, we tested the Δ*BT4554* mutant in our *in vivo* competition assay. In mice fed the FR diet, this mutant exhibited a similar 2-3 order of magnitude defect that closely resembles the Δ*rus* and Δ*rus*K1/K2 mutant

strains (Figure 2.9D). This finding helps connect the role of *rus* functions, which in all of the conditions assayed have been ubiquitously expressed *in vivo*, and the FR diet-specific fitness advantage experienced by wild-type *Bt in vivo*. We cannot definitively determine that

Table 2.2B pNP-assays for RusGH (BT2807)

Enzyme: BT2807 (RusGH), Substrate	Active (Y/N)
p-nitrophenyl β-D-ribofuranoside*	Y
4-nitrophenyl N-acetyl-α-D-galactosamide	N
4-nitrophenyl N-acetyl-β-D-glucosaminide	N
4-nitrophenyl α-D-galactopyranoside	N
4-nitrophenyl α-D-glucopyranoside	N
4-nitrophenyl α-D-mannopyranoside	N
4-nitrophenyl α-D-xylopyranoside	N
4-nitrophenyl α-L-arabinofuranoside	N
4-nitrophenyl α-L-arabinopyranoside	N
4-nitrophenyl α-L-fucopyranoside	N
4-nitrophenyl α-L-rhamnopyranoside	N
4-nitrophenyl β-D-galactopyranoside	N
4-nitrophenyl β-D-glucopyranoside	N
4-nitrophenyl β-D-glucuronide	N
4-nitrophenyl β-D-mannopyranoside	N
4-nitrophenyl β-D-xylopyranoside	N
4-nitrophenyl β-L-fucopyranoside	N

*Footnotes: *This activity was calcium dependent*

Table 2.2C TLC reactions for RusGH (BT2807)

Enzyme: BT2807 (RusGH), Substrates:	Active (Y/N)	Substrates Cont.	Active (Y/N)
Deoxycytidine	N	ADP-ribose	N
Thymidine	N	UDP	N
Cytidine	N	UDP-*N*-acetyl-D-glucosamine	N
5-Methyl uridine	N	AICAR	N
Deoxyuridine	N	D-lyxose	N
Uridine	N	D-psicose	N
Inosine	N	Ribostimycin	N
Deoxyinosine	N	Myo-inositol	N
Deoxyadenosine	N	Neomycin	N
Adenosine	N	Rebaudioside-A	N
Xanthosine	N	Ribitol	N
Deoxyguanosine	N	Amygdalin	N
Guanosine	N		
UDP-α-D-Glucose	N		
UDP-β-D-Glucose	N		
UDP-α-D-Galactose	N		
UMP	N		

nucleosides are the nutrients scavenged *in vivo* that drive this competitive advantage. However, *i*) similarity of the Δ*rus*, Δ*rusK1/K2* and Δ*BT4554* phenotypes and *ii*) the dependence on both a small amount of ribose (*i.e.*, to induce *rus*) and a functional *rus* system for *in vitro* growth on nucleosides via BT4554, supports a model in which ribose-induced Rus kinases are essential for the *in vivo* scavenging of nucleosides processed by BT4554. Although growth on nucleosides in some NSS mutants was reduced or eliminated this phenotype did not extend to growth on RNA or ribose, as the mutant strains exhibited similar levels of growth as wild-type (Figure 2.8B-C). This suggests that, while Rus functions are required to use RNA, the NSS functions interrogated here are not individually essential for catabolism of RNA-derived nucleosides or oligos.

Rus kinases are active towards ribose and nucleoside-derived ribose-1-phosphate

To scrutinize the activities of the *rus*-encoded kinases in detail, we produced recombinant forms and performed *in vitro* phosphorylation assays against pentose sugars and other monosaccharides (*E. coli* RbsK was a positive control). RusK2 has a preferred specificity towards ribose and deoxyribose, while exhibiting weaker activity on arabinose and xylose (Table

2.2D). RusK1 displayed nearly 10-fold weaker activity on ribose and deoxyribose compared to RusK2 and weak activity towards other sugars tested (Table 2.2D). The initial assay used to measure activity from RusK1 and K2 did not determine positional phosphorylation specificity. We hypothesized that an important difference in these kinases might be their positional

Table 2.2D Specific Activities of Bt RusK1, RusK2, and E. coli RbsK ribokinases

Phosphotase-coupled phosphorylation assay		Enzyme: BT2803 (RusK1)		Enzyme: BT2804 (RusK2)		Enzyme: RbsK *E. coli*	
Substrate	**Concentration**	**Average**	**Standard deviation**	**Average**	**Standard deviation**	**Average**	**Standard deviation**
D-Ribose	10 mM	13.797	1.920	336.263	32.437	2273.630	504.526
Deoxyribose	10 mM	13.956	1.156	44.245	6.795	141.407	59.810
D-Arabinose	200 mM	0.670	0.023	0.585	0.098	4.471	0.888
D-Xylose	200 mM	0.677	0.053	0.597	0.199	6.382	0.304
D-Fructose	200 mM	0.653	0.067	ND	NA	ND	NA
D-Glucose	200 mM	+	+	ND	NA	ND	NA
D-Galactose	200 mM	+	+	ND	NA	ND	NA
L-Fucose	200 mM	+	+	ND	NA	ND	NA
D-Mannose	200 mM	+	+	ND	NA	ND	NA
L-Rhamnose	200 mM	+	+	ND	NA	ND	NA
D-Sucrose	200 mM	+	+	ND	NA	ND	NA

phosphorylation at either the 1 or 5 carbon of ribose. When RusK1 and RusK2 enzymes were incubated with ribose and analyzed by LC/MS/MS, both generated ribose-5-phosphate (R5P) as the major detectable product (Figure 2.9E). We did not detect formation of ribose-1-phosphate (R1P) from ribose despite being able to reliably distinguish this compound as a standard (Table 2.2E). We next performed reactions using R1P as the substrate to test if this product, which we expect to be generated by BT4554 phosphorolysis of nucleosides, could be a substrate for the Rus kinases. Interestingly, our results show that ribose-1,5-bisphosphate (PRibP) is generated from R1P by both RusK1 and RusK2 (Figure 2.9F). In addition, RusK2 could generate a product with the same predicted mass as PRibP when given R5P as a substrate, despite RusK2 not forming R1P from ribose (Figure 2.9F). These are both novel findings as either mechanism involved in generating PRibP, or ribokinases capable of phosphorylation in the 1 position, have yet been identified in eubacteria. Rather, generation of PRibP by a different family of kinases has been described in archaea and plants as part of the RuBISCO pathway[29]. Our results help connect the function of Rus-encoded kinases with BT4554-mediated nucleoside scavenging via generation of intermediate PRibP. The route that PRibP takes after it is produced is still uncertain, since *Bt* lacks a clear homolog of the *E. coli* ribose 1,5-bisphosphokinase (*phnN*) that consumes PRibP to generate phosphoribosyl pyrophosphate (PRPP), which can be used both in nucleic acid synthesis or catabolically via the pentose phosphate pathway. Nevertheless, our data

suggest that the similar *in vivo* defects associated with loss of either RusK1/K2 or BT4554 are due to the requirement of both systems for utilization of exogenous nucleosides.

Global responses to ribose catabolism

We hypothesized that growth on ribose may affect expression of a global regulon. To test this, we performed RNAseq-based whole-genome transcriptional profiling on wild-type *Bt*

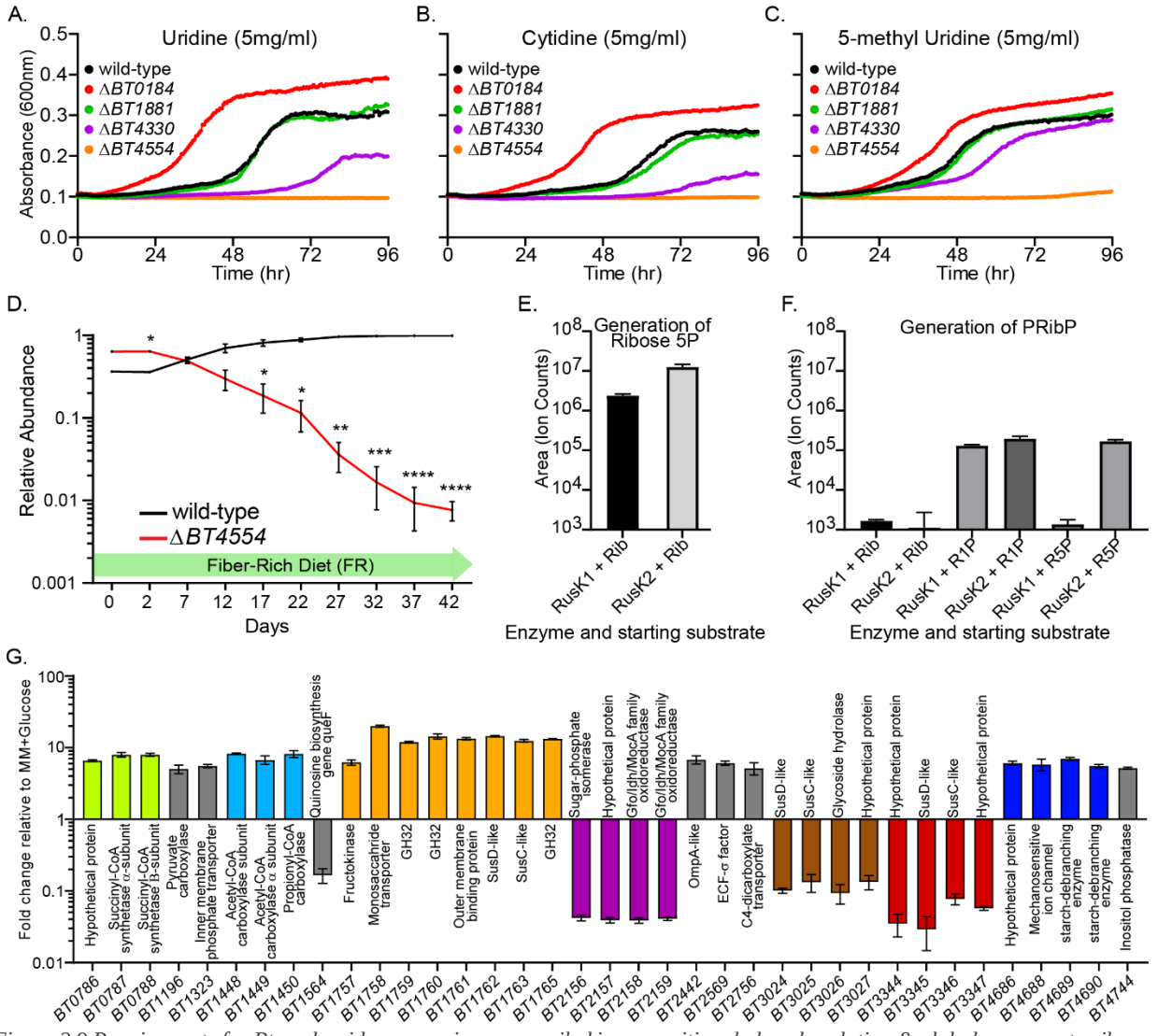


Figure 2.9 Requirements for Bt nucleoside scavenging genes, ribokinase positional phosphorylation & global response to ribose.

(A-C) Growth curves of nucleoside scavenging gene deletion strains (colored according to key) versus wild type Bt (black) on uridine (A), cytidine (B), or 5-methyl uridine (C). (D) In vivo competition between wild-type Bt (black) and ΔBT4554 strains (red) showing the relative abundance on the FR diet with mean ± SEM of n=4 biological replicates. (E-F) Positional ribose phosphorylation by RusK1 or RusK2 measured by LC/MS/MS for ribose 5-phosphate (E) or ribose-1,5-bisphosphate (F), for each bar the mean of n=3 biological replicates ± SD is shown. The y-axis minimum of 10^3 was determined from negative control reactions that included control enzyme or buffer only (Table 2.2E). Detection of PRibP was based off of a PRPP standard fragmenting into the major species of PRibP of the exact expected mass. (G) RNAseq-based global transcriptomic responses in Bt

*grown on MM-ribose compared to MM-glucose. Same bar color indicates genes in the same locus. Genes with gray bars are not physically linked in the genome. For each bar, the mean of n=3 replicates are shown ± SD. In (E), asterisks indicate significant differences (*P<0.05, **P<0.01, ***P<0.001, and ****P<0.0001) calculated by Student's t test.*

grown on ribose or glucose. Indeed, the data revealed a global response in which 81 genes were differentially expressed based on the parameters and thresholds used. Many of the genes (46%) belong to other PULs or metabolic pathways. (Table 2.3, shown below in Methods). Notable changes included upregulation of a previously defined PUL for fructose and β2,6-linked fructan metabolism (*BT1757-1765;* average upregulation of 15-fold), which interestingly liberates fructose that initiates the PPP[30] and suppresses *rus* expression (Figure 2.2A). At the same time, two other PULs of unknown specificity were repressed (*BT3024-3027*, *BT3344-3347*). Further, several genes encoding TCA cycle enzymes leading to generation of succinate and propionate, of which *Bt* has a partial pathway[31], were upregulated. In contrast, genes predicted to participate in sugar-phosphate isomerization and metabolism were strongly repressed (*BT2156-2159;* average of 24-fold) (Figure 2.9G). An experiment to test if cross-regulation between ribose metabolism and the fructan PUL contributes to the FR-diet specific competitive defect failed to support this model (Figure 2.8G-H).

Table 2.2E LC/MS/MS results for ribokinase reactions and controls showing raw data for area under the curve measured as ion counts

Starting Compound	**Detection**	**Ribose**	**Ribose-1P**	**Ribose-5P**	**Ribose-1,5-bisphosphate**
	Transition	149.0->89.0	229.0 ->210.9	229.0 ->97.0	309.0 -> 211.0
1mM Ribose Standard	Area	268,210	2,270	3,726	24
1mM Ribose-1P Standard	Area	910	24,518,553	3,777	108
1mM Ribose-5P Standard	Area	916	6,353	17,002,705	44
BT 2803 (RusK1) Enzyme Only	Area	173	524	420	205
BT 2804 (RusK2) Enzyme Only	Area	175	249	66	108
BT2803+Ribose n=1	Area	759,589	936	2,178,751	1,735
BT2803+Ribose n=2	Area	805,800	627	2,620,909	1,528
BT2803+Ribose n=3	Area	897,783	1,048	2,407,413	1,735
BT2803+Ribose-1P n=1	Area	104	7,706,247	3,078	135,812
BT2803+Ribose-1P n=2	Area	101	8,684,800	5,001	122,344
BT2803+Ribose-1P n=3	Area	78	8,117,289	1,438	137,551
BT2803+Ribose-5P n=1	Area	332	720	32,223,656	1,828
BT2803+Ribose-5P n=2	Area	278	323	28,552,913	1,036
BT2803+Ribose-5P n=3	Area	297	531	24,072,099	1,230
BT2804+Ribose n=1	Area	139,132	734	10,376,139	172
BT2804+Ribose n=2	Area	452,072	858	12,367,333	2,986
BT2804+Ribose n=3	Area	364,124	936	14,799,917	227
BT2804+Ribose-1P n=1	Area	133	7,654,273	8,161	161,399
BT2804+Ribose-1P n=2	Area	109	8,195,257	5,517	202,726
BT2804+Ribose-1P n=3	Area	145	8,085,939	5,825	223,148
BT2804+Ribose-5P n=1	Area	270	595	27,915,015	184,873
BT2804+Ribose-5P n=2	Area	250	201	21,565,887	156,384
BT2804+Ribose-5P n=3	Area	254	366	26,182,891	165,367
ATP	Area	175	423	221	74
Buffer Blank1	Area	1,565	17	5	12
Buffer Blank2	Area	801	23	114	40

Buffer Blank3	Area	526	21	372	46
Buffer Blank4	Area	507	5	149	50
Buffer Blank5	Area	363	2,141	3,327	20
Buffer Blank6	Area	306	1,733	3,015	39
Buffer Blank7	Area	454	1,415	2,578	37
Internal QC Standard 1	Area	4,667	11,244	587	15,350
Internal QC Standard 2	Area	3,637	11,303	694	15,071
Internal QC Standard 3	Area	5,026	11,303	724	14,705
Internal QC Standard 4	Area	4,851	11,310	881	14,970
Internal QC Standard 5	Area	4,791	11,296	888	14,531
Internal QC Standard 6	Area	5,337	11,384	1,012	14,702
Internal QC Standard 7	Area	4,523	11,855	1,107	14,806
Internal QC Standard 8	Area	3,245	13,122	8,026	10,401
Internal QC Standard 9	Area	3,353	14,722	6,436	11,239
Internal QC Standard 10	Area	4,618	10,454	6,329	10,682

An enzyme-diversified family of Rus systems exists throughout the Bacteroidetes

The data described above support the idea that the *Bt rus* PUL is necessary, but not always sufficient, for metabolizing ribose and nucleosides. Since it is strongly activated in response to the simple sugar ribose and not an oligosaccharide cue, *rus* is relatively unique and only the second PUL after the *Bt* fructan PUL shown to be activated in response to a monosaccharide[30]. The architecture of this system suggests that it is equipped to liberate ribose from additional unknown sources via its hydrolases. Therefore, we hypothesized that *rus*-like systems may be found in other gut isolates and perhaps more broadly across the *Bacteroidetes*. To test this, we measured the growth ability of 354 different human and animal gut *Bacteroidetes* in MM-ribose, revealing that ribose utilization is widely but variably present in different species (Figure 2.10A, Figure 2.11A, and shown in the Methods section Table 2.4). To determine if sequenced representatives of the species/strains that grow on ribose contain a homolog of the experimentally validated *Bt rus*, we used comparative genomics to search for homologs of this PUL within these gut isolates. This revealed that all of the sequenced strains that grow on ribose possess a candidate *rus*-like PUL, while none of the strains unable to grow on ribose had a homologous gene cluster. Interestingly, our analysis revealed very similar homologs of some *rus* genes in sequenced gut isolates (*e.g., Prevotella*) beyond those present in our initial survey. When we expanded the search to include *Bacteroidetes* isolates found in other body sites and in the environment, we detected *rus*-like systems across the phylum, with systems found in the genus *Bacteroides* being most similar to the prototype from *Bt*. Remarkably, we identified a total of 70 different *rus* configurations, ranging from simple two gene units (permease and kinase), to *rus* PULs containing as many as 36 genes (Figure 2.10B and Figure 2.11B). For almost all *rus*-

like systems, the following genes were present: *rusC* and *rusD*, an upstream *rusR* (or to a lesser extent different regulator types) either one or two *rusK* genes, and a *rusT* homolog.

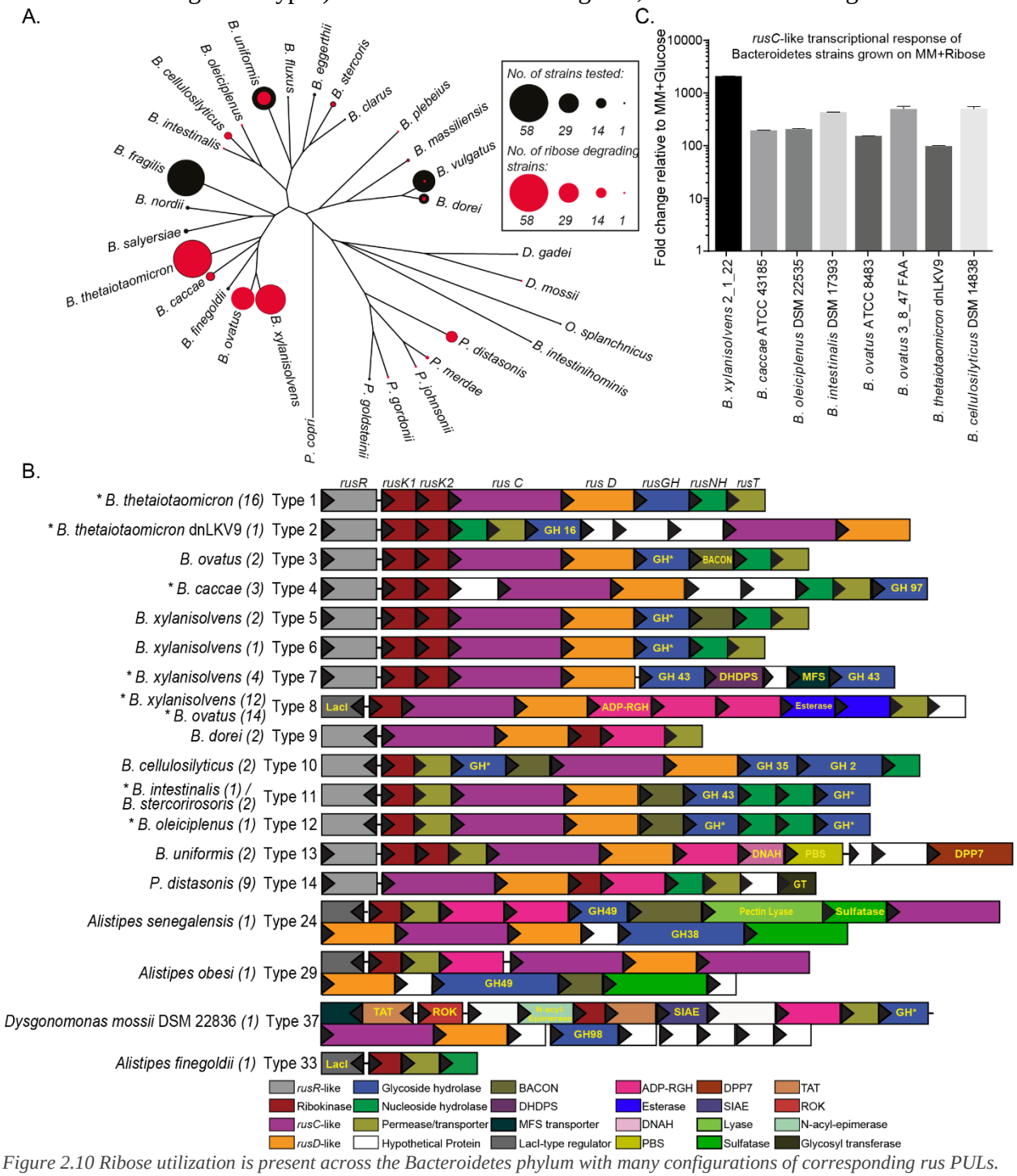


Figure 2.10 Ribose utilization is present across the Bacteroidetes phylum with many configurations of corresponding rus PULs.

(A) Genus-level phylogeny constructed from sequenced isolates showing the presence of ribose utilization. Outer black circles are sized to represent the number of strains tested for each species. Inner red circles indicate the number of tested strains that grow on ribose. (B) Comparisons of several variants of rus PULs throughout the Bacteroidetes. Identical background color indicates the same predicted function(s), which are defined according to the key. The number of sequenced isolates that harbor each PUL type is listed adjacent to each schematic, with each variant assigned an arbitrary type number. Asterisks next to the organism

name indicate that the PUL type shown is upregulated by ribose as the sole carbon source in at least one strain tested. Genes are sized to scale and all species represented here are human gut isolates. A broader representation of rus diversity is shown in Figure 2.11 and includes PULs from environmental and oral Bacteroidetes. Abbreviations not previously defined in the text are: (GH, Glycoside hydrolase of unknown family/function; BACON, Bacteroidetes-Associated Carbohydrate-binding Often N-terminal domain; DHDPS, dihydrodipicolinate synthase; LacI, predicted lacI-type transcriptional regulator; MFS, Major-facilitator superfamily of transporters; ADP-RGH, ADP-ribosyl glycoside hydrolase; DNAH, DNA helicase; PBS, Polysaccharide Biosynthesis and export of O-antigen and techoic acids; DPP7, Dipeptidyl-Peptidase 7 (serine peptidase); GT, Glycosyl Transferase). (C) Fold change of rusC-like transcript from the indicated species/strain showing that several additional rus PULs are activated during growth on ribose compared to glucose. Error bars show the SEM of n=3 biological replicates.*

Perhaps most intriguingly, the predicted enzymes found in different *rus*-like systems are exceptionally variable, with at least 22 different predicted glycoside hydrolase families, ADP-ribosylglycohydrolases, carbohydrate esterases, and nucleoside hydrolases among others. This plethora of enzymatic potential encoded in *rus* homologs across the *Bacteroidetes* suggests individual species or strains target different ribose-containing nutrients. To further connect these predicted *rus*-like systems with ribose utilization, we probed the transcriptional response of 8 different systems during growth on MM-ribose, finding that all strains tested exhibited ~100-1000 fold upregulation relative to a MM-glucose reference (Figure 2.10C).

Discussion

Diet impacts the gut microbiota in many ways and members of the prominent *Bacteroidetes* phylum have developed sophisticated strategies to liberate sugars from very complex dietary fiber polysaccharides such as pectins[13,14]. Such abilities equip these bacteria to compete for dietary and endogenous nutrients to sustain their populations. Diet-, microbiome- and host-derived RNA, nucleosides, cofactors and other sources of ribose have been largely unexplored as potential nutrients scavenged by members of the gut microbiota. Our findings demonstrate that *Bt* utilizes free- and covalently-linked sources of ribose and this metabolic capability contributes to competitive fitness *in vivo* in a diet-dependent fashion—likely through a more complicated metabolic mechanism that interconnects ribose sensing and nucleoside scavenging (Figure 2.12). It is also clear from comparative genomics that the ability to access ribose from diverse sources, extends across the *Bacteroidetes* phylum and is present in many animal gut, oral, and environmental isolates.

Although we have not yet uncovered a more complex ribose containing polymer requiring Rus transport and hydrolase functions, a key aspect of our ribose utilization model is that Rus-encoded kinases are required for growth on free and RNA-derived nucleosides, the latter only after RNAse and IAP degradation (Figure 2.12). In light of this pathway for

nucleoside assimilation, the roles of periplasmic RusNH and cell surface RusGH remain enigmatic. Given the weak activities of these enzymes towards the substrates tested, it is

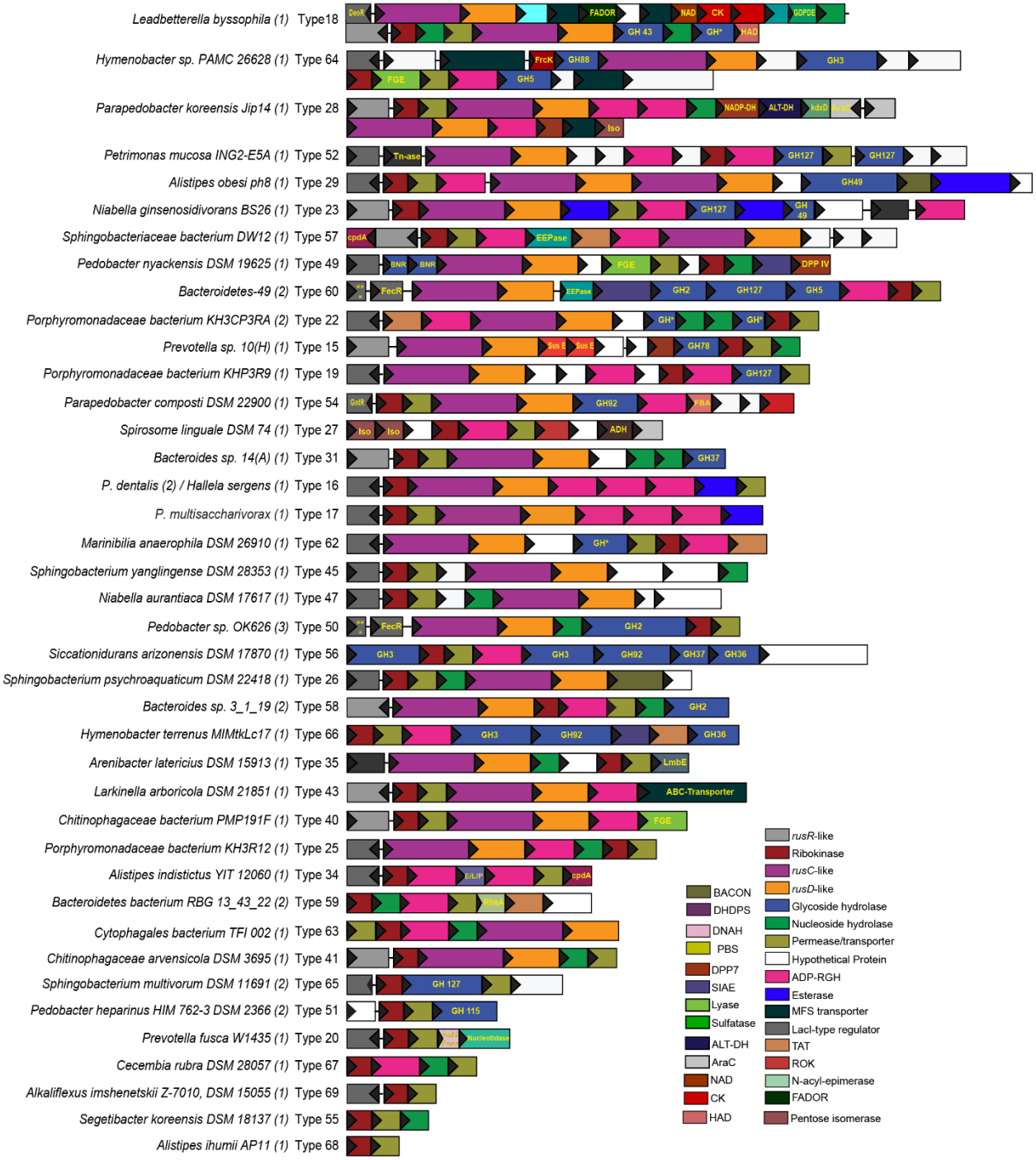


Figure 2.11 An expanded repertoire of rus architectures across the Bacteroidetes phylum.

A comparative genomics approach across the Bacteroidetes phylum revealed many different types and subtypes of the rus locus. This figure displays almost all of the additional types found in both the human gut isolates as an expansion from Figure 2.10B, and those found in aquatic, soil, and human oral cavity isolates. Not shown are subtypes, where the same genes are present, but arranged differently, as well as types 21, 30, 33, 36, 39, 44, 48, 46, and 61, all of these types only had one isolate and were of lesser complexity than the majority shown in this figure. As in Figure 2.10, the gene size represents the amino acid length, and the

*background color is kept constant for genes predicted to encode the same or very similar functions. Gene abbreviations are as follows in order of appearance: (DeoR, DeoR-like family of transcriptional regulator; MFS, Major Facilitator Superfamily of transporters; FADOR, Flavin (FAD) Oxidoreductase; NAD, NAD Binding Protein; CK, Carbohydrate Kinase, unknown family; GDPDE, Glycerophosphoryl Diester Phosphodiesterase; NH, Nucleoside Hydrolase; GH, Glycoside hydrolase; HAD, Haloacid Dehydrogenase; LacI, LacI-type transcriptional regulator; FrcK, fructokinase; FGE, Formylglycine-Generating Enzyme, required for sulfatase activity; NADP-DH, NADP-Dependent aldehyde Dehydrogenase; ALT-DH, Altronate Dehydrogenase; kdxD, 2-dehydro-3-deoxy-D-arabinonate dehydratase; AraC, AraC-like transcriptional regulator; Rib Iso, Ribose-5-Phosphate Isomerase; Tn-ase, transposase; BACON, Bacteroidetes-Associated Carbohydrate-binding Often N-terminal domain; cpdA, 3'.5'-cyclic AMP phosphodiesterase; EEPase, Endo-Exo Nucleoside-Phosphatase; TAT, Twin-Arginine Translocase; BNR, BNR repeat-like domain; SIAE, Sialate O-acetylesterase; DPP IV, Dipeptidyl-peptidase IV; ***, RNA polymerase sigma factor ECF subfamily; FecR, FecR-like transcriptional regulator; SusE, Bacteroides SusE-like outer membrane binding protein; GntR, GntR-like transcriptional regulator; FBA, Fructose Bisphosphate Aldolase; Xyl Iso, Xylose Isomerase; ROK, Repressor/ORF/Kinase domain containing protein; ADH, Alcohol Dehydrogenase; LmbE, N-acetylglucosaminyl deacetylase LmbE-like family; E/L/P, Esterase/Lipase/Peptidase-like domain containing protein; RhaA, Regulator of RNaseE activity; Acid Pase, Acid Phosphotase-like protein). We have included a color-coded legend on the figure with the abbreviations listed next to them to help clarify the relationship between the colored genes.*

probable that they are optimized to cleave substrates that we have not yet been able to test and which are the *bona fide* nutrient targets of the *Bt* Rus system. At least for the pyrimidine nucleosides tested *in vitro*, the BT4554 phosphorylase, which generates a cleaved base plus R1P, is the primary component required. A novel aspect of the model we have determined for *Bt*, is that Rus-encoded ribokinases are required for conversion of R1P to PRibP, and this conversion requires ribose induction of Rus to activate production of the ribokinases. This interconnection may stem from the dual function of the ribokinases, phosphorylating both ribose to R5P and R1P to PRibP. Based on our growth and positional phosphorylation data, it is unlikely that R5P is being shunted directly into catabolism as canonically represented in KEGG maps of the PPP. This is largely based on the observation that either Rus kinase can generate R5P, so if direct assimilatory pathways exist through D-ribulose-5-P or D-sedoheptulose-7-P, both single kinase mutants should grow normally as they would be redundant in this function. Rather the generation of PRibP from R1P or R5P, whose ultimate path(s) are uncertain is more likely the relevant molecule being generated for use in catabolism. The lack of a detectable phosphopentomutase in *Bt* that isomerizes R5P to R1P, (having only a phosphoglucomutase, BT1548) may have driven the evolution of co-dependence on the ribokinases and nucleoside phosphorylase (BT4554) to generate PRibP. Although the full catabolic pathway for PRibP in *Bt* is still unclear, our findings hold important implications for how predicted metabolic maps may be incomplete in some instances and should be interpreted with caution. By investigating this pathway more deeply, our results demonstrate the first known bacterial ATP-dependent ribokinases able to generate PRibP from R1P.

Similar to only one other previously characterized *Bacteroides* PUL for fructan utilization[30], the *rus* PUL is activated in response to a monosaccharide, (fructose and ribose respectively) (Figure 2.1C) and also contains a dedicated permease and kinase revealing that

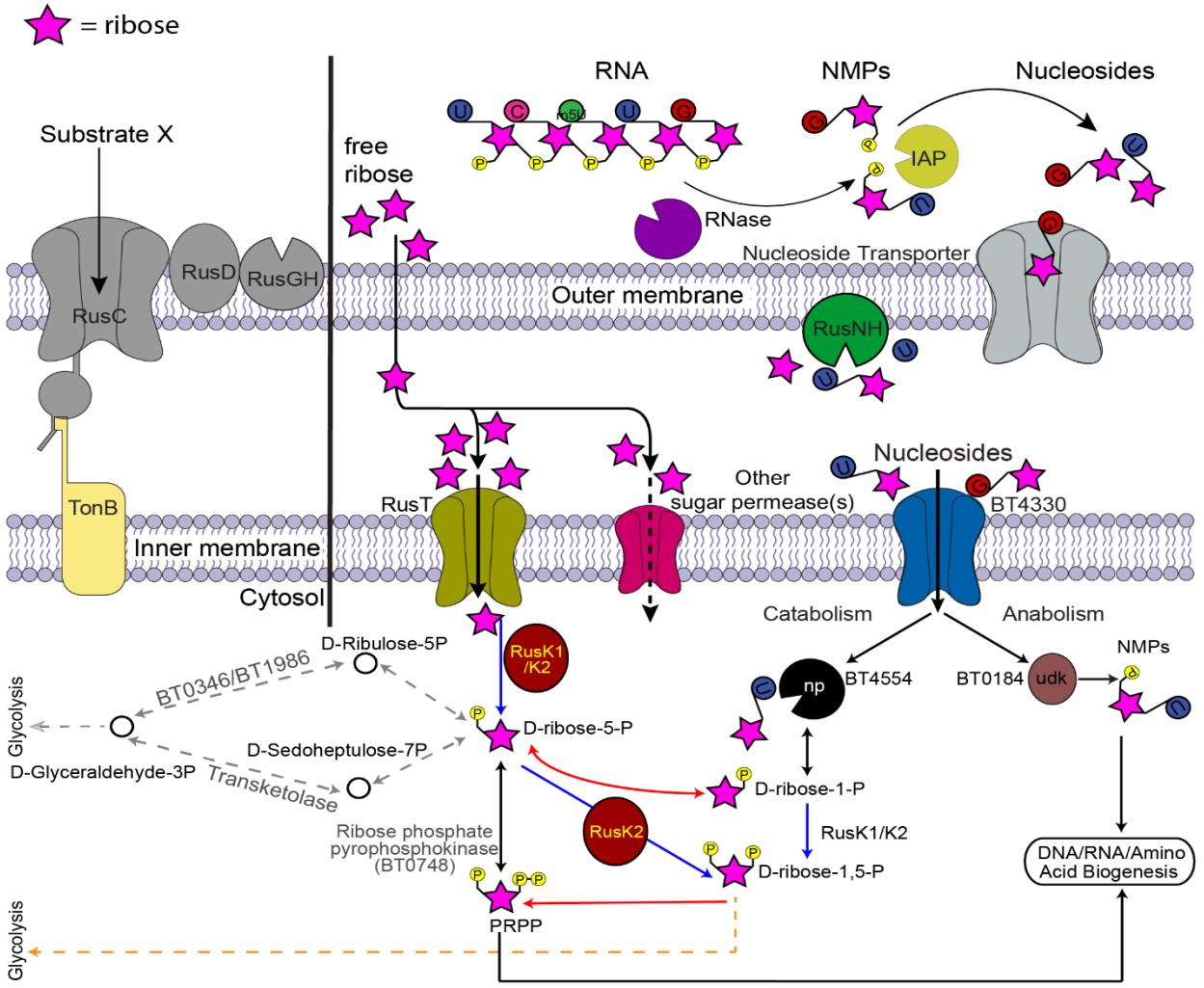


Figure 2.12 Model of ribose utilization by Bt and connection to nucleoside scavenging.

The model shown represents a proposed mechanism of ribose and nucleoside metabolism via rus-encoded functions and nucleoside scavenging systems based on the data in this study and predicted KEGG metabolic maps. Ribose is depicted as a pink star, phosphate groups are represented as yellow circles, while nucleoside bases are shown as colored circles (blue: uridine, green: 5-methyl uridine, dark pink: cytidine, red: guanosine). Some cellular locations of protein products were experimentally determined (e.g., (Figure 2.13) for localization data for RusGH and RusNH). Proteins for which no functional requirement could be assigned are shaded in gray (RusC, RusD, RusGH). Abbreviations: intestinal alkaline phosphatase, IAP, rus-encoded nucleoside hydrolase, RusNH, ribokinases, RusK1/K2, nucleoside phosphorylase, np, uridine kinase, udk. Our cytosolic metabolic model depicts our interpretation of the PPP of Bt based on our results. Dashed arrows with red "x" indicate that homologs of the enzymes normally catalyzing these steps are not detectable by homology searching in Bt. Black arrows are steps of the PPP that likely occur in Bt and we have results indicating the importance of BT0184, BT4330, and BT4554 in nucleoside catabolism.

these two systems are similarly patterned around a core monosaccharide utilization pathway. Although the activation signal for *rus* is derived from extracellular ribose, we are unable to conclude the exact phosphorylation status of the ribose that activates expression. Based on our

results, the kinases and the putative regulator RusR are required for generation of transcript. Often, the enzyme content encoded in *Bacteroidetes* PULs provides a window into the nutrient linkages that any given system has evolved to target[10,11,32]. Ribose is present in many diverse sources with different linkages, including RNA and nucleosides, bacterial capsules, cofactors such as NAD, cellular modification like (poly) ADP-ribose, and more exotic molecules such as microcins[33]. The breadth of enzymatic diversity emphasized by the presence of at least 22 different glycoside hydrolase families plus others, and 70 different configurations of ribose utilization systems across the phylum supports a hypothesis whereby species have adapted to liberate ribose from different and diverse sources. We initially hypothesized that the nutrient mediating the competitive advantage *in vivo* for *Bt rus* would be endogenous nucleosides or RNA from bacteria or host cells in a fiber-free diet (FF-diet). However, our results suggest that in the FR-diet, nucleosides are the nutrients targeted by the combined actions of BT4554 and the *rus* kinases. While pyrimidine nucleoside addition to water in mice fed the FF diet did not reveal

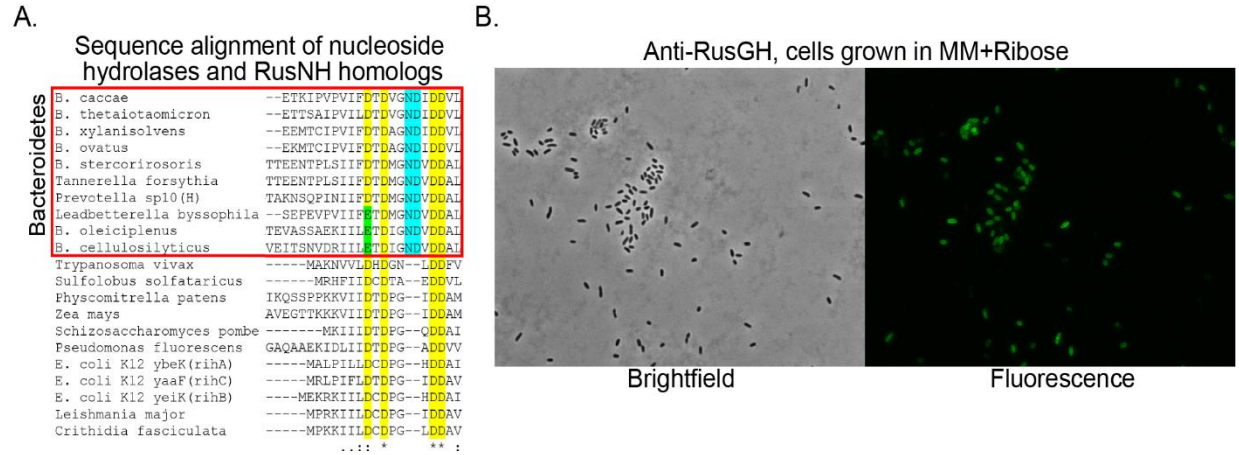


Figure 2.13 Localization of RusGH and sequence alignment of conserved residues of RusNH.

Data is in support of Fig. 2.12 (A) Immunofluorescent microscopy of Bt grown in MM-ribose media staining with anti-BT2807 (RusGH) antibody, indicating that the protein is localized to the outer membrane as the secondary antibody has clearly labeled nearly all of the cells seen in the brightfield image (left), a green color in the fluorescent image at (right). (B) Multiple sequence alignment of BT2808 (RusNH) and other RusNH-like proteins from Bacteroidetes (red boxed region) compared to previously validated nucleoside hydrolases isolated from bacteria (E. coli, RihA,B,C and P. fluorescens), archaea (S. solfataricus), parasitic eukaryotes (T. vivax, L. major, and C. fasciculata), moss (P. patens), maize (Zea mays), and yeast (S. pombe), indicating that the predicted nucleoside hydrolase of BT2808 shares the universally conserved N-terminal DXDXXXDD motif responsible for Ca^{2+} coordination (2^{nd} and 4^{th} yellow-highlighted aspartic acid residues) and ribose binding (3^{rd} yellow-highlighted aspartic acid), as well as the nearly conserved canonical 1^{st} aspartic acid residue denoting the motif (yellow or green highlighted position). Specific to the Bacteroidetes nucleoside hydrolases, there are two residues, an asparagine and an additional aspartic acid within the motif (highlighted in teal) not found IUNH family nucleoside hydrolases outside of the Bacteroidetes.

a competitive defect, it is possible that purine nucleosides, which were not tested due to low solubility, could be present in the FR diet and exert this effect.

The results described here highlight how survival of bacteria in the human gut and other ecosystems has driven adaptations to sense and scavenge the ubiquitous sugar ribose. Since enterohemorrhagic *E. coli* (EHEC) and other pathogenic *E. coli* preferentially utilize ribose *in vivo*[4,34], or upregulate genes for the catabolism of this nutrient in the environment[35] these substrates may represent unexplored nutrient niches competed for by commensal and pathogenic microorganisms and may therefore help mediate colonization resistance against pathogens. The evolution of diverse enzyme functions throughout the *Bacteroidetes* may be analogous to a molecular "Swiss-army knife", in which the core function is utilization of ribose but the various blades and other implements represent the enzymes equipping the system to sense, import or harvest ribose from diverse sources. This molecular adaptability is particularly important in the context of the nutrient niche hypothesis of gut bacterial survival. While some nutrients may be scarce compared to abundant dietary fiber polysaccharides, competition for these lower abundance nutrients may be less intense and organisms capable of accessing them could thereby occupy a stable niche. While a number of gut bacteria, including pathogens, are capable of utilizing free ribose, the *Bacteroides* may have developed a more sophisticated ability to scavenge multiple sources by cleaving it from covalently linked forms. From this perspective, understanding the struggle to access this "simple" sugar may reveal additional layers underpinning the interplay between native gut mutualists and invading pathogens.

Methods

Gnotobiotic mouse experiments

All experiments involving animals, including euthanasia via carbon dioxide asphyxiation, were approved by the University Committee on Use and Care of Animals at the University of Michigan (NIH Office of Laboratory Animal Welfare number A3114-01) and overseen by a veterinarian. Six to eight-week-old, germfree female Swiss-Webster mice were initially maintained on the standard, fiber-rich lab diet (LabDiet 5010, LabDiet, St. Louis, MO), where appropriate, mice were switched to a fiber-free diet (Envigo-Teklad TD 130343) and maintained for one week prior to colonization with *Bt* strains. After stable colonization had been observed, at day 14 some groups of mice were provided water ab libitum containing one of the following: 1% ribose, 1% Nucleoside mixture (0.25% thymidine, 0.25% uridine, 0.25% 5-methyl uridine, and

0.25% cytidine) or Type VI torula yeast RNA. DNA was extracted from fecal pellets throughout the experiment and strain abundance was quantified as described previously[26]. Relative abundance of each strain was normalized to the original abundance on day of gavage (day 0). Post-sacrifice, cecal contents were collected, flash frozen and stored at -80°C. RNA was extracted as described previously[1], briefly, RNA was phenol-chloroform treated and ethanol precipitated, DNA removed by treatment with TURBO^{TM} DNaseI (Ambion), followed by purification using RNeasy mini kit (Qiagen) according to manufactures instructions.

Bacterial strains, culturing conditions, and molecular genetics.

B. thetaiotaomicron ATCC 29148 (VPI-5482) and its genetic variants, as well as other *Bacteroides* strains used in this study, were routinely grown in tryptone-yeast extract-glucose (TYG) broth medium[36], in minimal medium (MM), plus a defined carbon source[17], or on brain heart infusion agar with 10% defibrinated horse blood (Colorado Serum Co.). Unless otherwise noted, carbon sources used in MM were added to a final concentration of 5 mg/ml. Cultures were grown at 37°C in an anaerobic chamber (10% H_2, 5% CO_2, and 85% N_2; Coy Manufacturing, Grass Lake, MI). Genetic deletions and mutations were performed by counter-selectable allelic exchange as previously described[37]. Complementation of deletion strains was performed using pNBU2 vectors as described previously[17], containing 314 bp upstream of *BT2802*, predicted to contain the promoter sequence for the Δ*rusR* strain or 186 bp upstream of *BT2803-04* containing the entire intergenic region for the Δ*rusK1/K2* strain. Primers used in this study are listed in (Table 2.5, located at the end of Methods). To quantify growth on carbon sources and examine mutant phenotypes, increase in culture absorbance (600 nm) in 200µl cultures in 96-well plates was measured at 10 minute intervals for at least 96 hours on an automated plate reader as previously described[21]. To achieve consistent and robust growth on nucleosides and other covalently linked sources of ribose, free ribose was added at a final concentration of 0.5 mg/ml to MM containing 5 mg/ml of carbon source. Growth on 5mg/ml of MM containing Type IV Torula yeast RNA (Sigma) was obtained by adding 100 units of calf-intestinal alkaline phosphatase (CIP) (New England Biolabs) and 2mg/ml RNase A (Sigma). Growth parameters and conditions for all substrates are summarized in (Table 2.1).

Table 2.1 Growth characteristics of B. thetaiotaomicron strains on ribose and other ribose-containing molecules

Substrate	Strain	Rate	Lag (hours)	Corrected Lag (hours)	Total Growth (600nm)

Glucose	Δ*tdk* (isogenic parent strain)	0.1655	8.068	NA	1.301
Glucose	ΔBT2802-2809 (*rus*)	0.1565	7.065	NA	1.269
Glucose	ΔBT2802 (*rusR*)	0.1504	7.650	NA	1.258
Glucose	ΔBT2803 (*rusK1*)	0.1452	7.608	NA	1.292
Glucose	ΔBT2804 (*rusK2*)	0.1392	7.984	NA	1.263
Glucose	ΔBT2803-2804 (*rusK1/K2*)	0.1476	7.023	NA	1.283
Glucose	ΔBT2805 (*rusC*)	0.1461	6.981	NA	1.270
Glucose	ΔBT2806 *(rusD*)	0.1411	7.316	NA	1.280
Glucose	ΔBT2805-2806 (*rusC/D*)	0.0911	6.424	NA	0.946
Glucose	ΔBT2807 (*rusGH*)	0.1379	6.856	NA	1.267
Glucose	ΔBT2808 (*rusNH*)	0.1344	7.441	NA	1.224
Glucose	ΔBT2807-2808 (*rusGH/NH*)	0.1345	8.402	NA	1.263
Glucose	ΔBT2809 (*rusT*)	0.1596	7.024	NA	1.288
Glucose	ΔBT0184	0.1109	3.951	NA	0.877
Glucose	ΔBT1881	0.1053	4.939	NA	0.937
Glucose	ΔBT4330	0.1161	4.447	NA	0.976
Glucose	ΔBT4554	0.0986	8.401	NA	1.025
Ribose	wild-type BT (Δ*tdk*)	0.0802	31.437	23.367	1.101
Ribose	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
Ribose	ΔBT2802 (*rusR*)	NG	NA	NA	NG
Ribose	ΔBT2803 (*rusK1*)	0.0768	62.582	54.972	1.123
Ribose	ΔBT2804 (*rusK2*)	NG	NA	NA	NG
Ribose	ΔBT2803-2804 (*rusK1/K2*)	NG	NA	NA	NG
Ribose	ΔBT2805 (*rusC*)	0.0579	28.385	21.405	1.123
Ribose	ΔBT2806 *(rusD*)	0.0577	32.189	24.869	1.128
Ribose	ΔBT2805-2806 (*rusC/D*)	0.0449	51.992	45.568	1.071
Ribose	ΔBT2807 (*rusGH*)	0.0741	30.015	23.155	1.126
Ribose	ΔBT2808 (*rusNH*)	0.0762	31.771	24.331	1.129
Ribose	ΔBT2807-2808 (*rusGH/NH*)	0.0786	32.774	24.374	1.107
Ribose	ΔBT2809 (*rusT*)	0.0225	40.550	33.530	0.829
Ribose	ΔBT0184	0.0539	46.531	42.580	1.125
Ribose	ΔBT1881	0.0544	44.053	39.114	1.079
Ribose	ΔBT4330	0.0560	43.554	39.107	1.111
Ribose	ΔBT4554	0.0538	46.037	37.636	1.122
Thymidine	wild-type BT (Δ*tdk*)	NC	66.977	49.434	0.224
Thymidine	wild-type BT (Δ*tdk*)	NC	42.127	38.623	0.318
Thymidine	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
Thymidine	ΔBT2802 (*rusR*)	NG	NA	NA	NG
Thymidine	ΔBT2803 (*rusK1*)	NG	NA	NA	NG
Thymidine	ΔBT2804 (*rusK2*)	NG	NA	NA	NG
Thymidine	ΔBT2803-2804 (*rusK1/K2*)	NG	NA	NA	NG
Thymidine	ΔBT2805 (*rusC*)	NC	66.393	52.427	0.248
Thymidine	ΔBT2806 *(rusD*)	NC	74.866	60.977	0.177
Thymidine	ΔBT2805-2806 (*rusC/D*)	NC	46.531	42.084	0.182
Thymidine	ΔBT2807 (*rusGH*)	NC	63.984	55.847	0.167
Thymidine	ΔBT2808 (*rusNH*)	NC	50.850	42.568	0.260
Thymidine	ΔBT2807-2808 (*rusGH/NH*)	NC	67.270	59.932	0.177
Thymidine	ΔBT2809 (*rusT*)	NC	116.377	105.406	0.036
Thymidine	ΔBT0184	NC	33.148	29.198	0.284
Thymidine	ΔBT1881	NC	40.580	35.641	0.271
Thymidine	ΔBT4330	NC	41.078	36.631	0.255
Thymidine	ΔBT4554	NG	NA	NA	NG
Uridine	wild-type BT (Δ*tdk*)	NC	70.321	52.778	0.200
Uridine (10 mg/ml)	wild-type BT (Δ*tdk*)	NC	42.069	38.613	0.198
Uridine	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
Uridine	ΔBT2802 (*rusR*)	NG	NA	NA	NG
Uridine	ΔBT2803 (*rusK1*)	NG	NA	NA	NG
Uridine	ΔBT2804 (*rusK2*)	NG	NA	NA	NG

Uridine	ΔBT2803-2804 (*rusK1/K2*)	NG	NA	NA	NG
Uridine	ΔBT2805 (*rusC*)	NC	73.697	59.730	0.252
Uridine	ΔBT2806 *(rusD*)	NC	76.327	62.438	0.207
Uridine	ΔBT2805-2806 (*rusC/D*)	NC	64.405	59.958	0.134
Uridine	ΔBT2807 (*rusGH*)	NC	59.897	51.760	0.227
Uridine	ΔBT2808 (*rusNH*)	NC	60.991	52.709	0.242
Uridine	ΔBT2807-2808 (*rusGH/NH*)	NC	62.743	55.405	0.220
Uridine	ΔBT2809 (*rusT*)	NG	NA	NA	NG
Uridine	ΔBT0184	NC	31.659	27.709	0.261
Uridine	ΔBT1881	NC	54.472	49.533	0.199
Uridine	ΔBT4330	NC	NC	NC	0.056
Uridine	ΔBT4554	NG	NA	NA	NG
Cytidine	wild-type BT (Δ*tdk*)	NC	101.248	83.705	0.048
Cytidine (10 mg/ml)	wild-type BT (Δ*tdk*)	NC	66.393	62.937	0.107
Cytidine	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
Cytidine	ΔBT2802 (*rusR*)	NG	NA	NA	NG
Cytidine	ΔBT2803 (*rusK1*)	NG	NA	NA	NG
Cytidine	ΔBT2804 (*rusK2*)	NG	NA	NA	NG
Cytidine	ΔBT2803-2804 (*rusK1/K2*)	NG	NA	NA	NG
Cytidine	ΔBT2805 (*rusC*)	NC	108.557	94.591	0.034
Cytidine	ΔBT2806 *(rusD*)	NC	NC	NC	0.011
Cytidine	ΔBT2805-2806 (*rusC/D*)	NC	74.844	70.397	0.081
Cytidine	ΔBT2807 (*rusGH*)	NC	83.345	75.208	0.052
Cytidine	ΔBT2808 (*rusNH*)	NC	86.526	78.243	0.088
Cytidine	ΔBT2807-2808 (*rusGH/NH*)	NC	90.912	83.574	0.072
Cytidine	ΔBT2809 (*rusT*)	NG	NA	NA	NG
Cytidine	ΔBT0184	NC	39.59	35.64	0.20
Cytidine	ΔBT1881	NC	66.89	61.95	0.10
Cytidine	ΔBT4330	NC	NC	NC	0.02
Cytidine	ΔBT4554	NG	NA	NA	NG
5-methyl uridine	wild-type BT (Δ*tdk*)	NC	82.356	64.812	0.151
5-methyl uridine (10 mg/ml)	wild-type BT (Δ*tdk*)	NC	49.014	45.558	0.203
5-methyl uridine	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
5-methyl uridine	ΔBT2802 (*rusR*)	NG	NA	NA	NG
5-methyl uridine	ΔBT2803 (*rusK1*)	NG	NA	NA	NG
5-methyl uridine	ΔBT2804 (*rusK2*)	NG	NA	NA	NG
5-methyl uridine	ΔBT2803-2804 (*rusK1/K2*)	NG	NA	NA	NG
5-methyl uridine	ΔBT2805 (*rusC*)	NC	77.423	63.456	0.187
5-methyl uridine	ΔBT2806 *(rusD*)	NC	64.451	50.562	0.258
5-methyl uridine	ΔBT2805-2806 (*rusC/D*)	NC	58.943	54.496	0.131
5-methyl uridine	ΔBT2807 (*rusGH*)	NC	50.004	41.867	0.255
5-methyl uridine	ΔBT2808 (*rusNH*)	NC	65.765	57.483	0.192
5-methyl uridine	ΔBT2807-2808 (*rusGH/NH*)	NC	51.929	44.591	0.262
5-methyl uridine	ΔBT2809 (*rusT*)	NG	NA	NA	NG
5-methyl uridine	ΔBT0184	NC	38.104	34.153	0.223
5-methyl uridine	ΔBT1881	NC	48.516	43.577	0.197
5-methyl uridine	ΔBT4330	NC	58.943	54.496	0.159
5-methyl uridine	ΔBT4554	NG	NA	NA	NG
RNA w/enzyme	wild-type BT (Δ*tdk*)	NC	28.468	20.398	0.199
RNA w/enzyme	ΔBT2802-2809 (*rus*)	NC	NC	NC	0.067
RNA w/enzyme	ΔBT2802 (*rusR*)	NC	NC	NC	0.059
RNA w/enzyme	ΔBT2803 (*rusK1*)	NC	NC	NC	0.090
RNA w/enzyme	ΔBT2804 (*rusK2*)	NC	NC	NC	0.057
RNA w/enzyme	ΔBT2803-2804 (*rusK1/K2*)	NC	NC	NC	0.082
RNA w/enzyme	ΔBT2805 (*rusC*)	NC	43.693	36.713	0.237
RNA w/enzyme	ΔBT2806 (*rusD*)	NC	46.313	38.993	0.221
RNA w/enzyme	ΔBT2805-2806 (*rusC/D*)	NC	40.997	32.927	0.207
RNA w/enzyme	ΔBT2807 (*rusGH*)	NC	52.141	45.281	0.223

RNA w/enzyme	ΔBT2808 (*rusNH*)	NC	45.828	38.388	0.247
RNA w/enzyme	ΔBT2807-2808 (*rusGH/NH*)	NC	38.357	29.957	0.260
RNA w/enzyme	ΔBT2809 (*rusT*)	NC	NC	NC	0.087
RNA w/enzyme	ΔBT0184	NC	39.278	31.208	0.169
RNA w/enzyme	ΔBT1881	NC	20.630	12.564	0.292
RNA w/enzyme	ΔBT4330	NC	23.634	15.490	0.200
RNA w/enzyme	ΔBT4554	NC	34.020	25.950	0.173
Deoxyribose	wild-type BT (Δ*tdk*)	NC	72.979	64.909	0.142
Deoxyribose	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
AMP	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
AMP	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
Tagatose	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Inosine	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Adenosine	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Xanthosine	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Rebauside A	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Amygdalin	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Ribostymycin	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
N-acetylmuramic acid	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Neomycin	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Myo-Inositol	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
UDP-Galactose	wild-type BT (Δ*tdk*)	NC	88.198	80.128	0.391
UDP-β-Glucose	wild-type BT (Δ*tdk*)	NC	NC	NC	0.300
UDP-α-Glucose	wild-type BT (Δ*tdk*)	NC	98.458	90.390	0.23
UDP	wild-type BT (Δ*tdk*)	NC	NA	NA	NG
Salmon Sperm DNA	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
ADP-Ribose	wild-type BT (Δ*tdk*)	NC	82.048	73.978	0.076
Lyxose	wild-type BT (Δ*tdk*)	NC	66.256	58.186	0.116
Psicose	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Melezitose	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
UDP-N-acetylglucosamine	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
UDP-β-Glucose	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
UDP-α-Glucose	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
UDP	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
ADP-Ribose	ΔBT2802-2809 (*rus*)	NG	NA	NA	NG
UDP-glucuronic acid	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
UDP-glucosamine	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
NADH	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Ribitol (adonitol)	wild-type BT (Δ*tdk*)	NG	NA	NA	NG
Palatinose	Δ*tdk* (isogenic parent strain)	0.0812	48.495	39.632	0.937
Palatinose	ΔBT2802-2809 (*rus*)	0.0621	46.823	37.291	0.872
Turanose	Δ*tdk* (isogenic parent strain)	0.0878	24.080	15.217	1.013
Turanose	ΔBT2802-2809 (*rus*)	0.0876	22.241	12.709	0.982
Trehalose	Δ*tdk* (isogenic parent strain)	0.0930	42.141	33.277	0.995
Trehalose	ΔBT2802-2809 (*rus*)	0.0641	43.813	34.281	0.879
Maltose	Δ*tdk* (isogenic parent strain)	0.0979	11.539	2.676	1.014
Maltose	ΔBT2802-2809 (*rus*)	0.0851	11.371	1.839	1.039
Sucrose	Δ*tdk* (isogenic parent strain)	0.0969	8.027	-0.836	1.053
Sucrose	ΔBT2802-2809 (*rus*)	0.1086	8.361	-1.171	1.035
Lactose	Δ*tdk* (isogenic parent strain)	0.1106	8.696	-0.168	1.036
Lactose	ΔBT2802-2809 (*rus*)	0.1033	9.365	-0.168	0.968
Galactose	Δ*tdk* (isogenic parent strain)	0.1133	11.539	2.676	0.928
Galactose	ΔBT2802-2809 (*rus*)	0.1109	11.371	1.839	0.871
Glucose	Δ*tdk* (isogenic parent strain)	0.1075	8.863	0.000	1.042
Glucose	ΔBT2802-2809 (*rus*)	0.1129	9.532	0.000	1.019
Fructose	Δ*tdk* (isogenic parent strain)	0.1070	8.863	0.000	0.895
Fructose	ΔBT2802-2809 (*rus*)	0.1109	9.030	-0.502	0.853
L-Fucose	Δ*tdk* (isogenic parent strain)	0.0248	42.809	33.946	0.440

L-Fucose	ΔBT2802-2809 (*rus*)	0.0194	40.970	31.438	0.341
Rhamnose	Δ*tdk* (isogenic parent strain)	0.0503	38.796	29.933	0.538
Rhamnose	ΔBT2802-2809 (*rus*)	0.0511	37.626	28.094	0.496
Galacturonic Acid	Δ*tdk* (isogenic parent strain)	0.0255	50.000	41.137	0.511
Galacturonic Acid	ΔBT2802-2809 (*rus*)	0.0155	49.164	39.632	0.486
Glucuronic Acid	Δ*tdk* (isogenic parent strain)	0.0342	46.488	37.625	0.561
Glucuronic Acid	ΔBT2802-2809 (*rus*)	0.0358	48.328	38.796	0.539
N-acetylgalactosamine	Δ*tdk* (isogenic parent strain)	0.0552	20.067	11.204	0.850
N-acetylgalactosamine	ΔBT2802-2809 (*rus*)	0.0536	21.572	12.040	0.816
N-acetylglucosamine	Δ*tdk* (isogenic parent strain)	0.0827	10.368	1.505	0.926
N-acetylglucosamine	ΔBT2802-2809 (*rus*)	0.0788	11.873	2.341	0.856
Xylose	Δ*tdk* (isogenic parent strain)	0.1074	20.569	11.706	1.131
Xylose	ΔBT2802-2809 (*rus*)	0.1074	19.566	10.033	1.096
Mannose	Δ*tdk* (isogenic parent strain)	0.1278	7.526	-1.338	1.047
Mannose	ΔBT2802-2809 (*rus*)	0.1310	8.194	-1.338	1.029
L-Arabinose	Δ*tdk* (isogenic parent strain)	0.1171	16.054	7.190	0.999
L-Arabinose	ΔBT2802-2809 (*rus*)	0.1017	16.388	6.856	0.919
D-Arabinose	Δ*tdk* (isogenic parent strain)	0.1034	20.234	11.371	0.934
D-Arabinose	ΔBT2802-2809 (*rus*)	0.0949	19.900	10.368	0.920
Glucosamine	Δ*tdk* (isogenic parent strain)	0.0189	31.271	22.408	0.590
Glucosamine	ΔBT2802-2809 (*rus*)	0.0202	28.930	19.398	0.589
Raffinose	Δ*tdk* (isogenic parent strain)	0.0245	17.559	8.696	0.732
Raffinose	ΔBT2802-2809 (*rus*)	0.0175	30.603	21.070	0.660
Erlose	Δ*tdk* (isogenic parent strain)	0.0474	16.054	7.190	0.904
Erlose	ΔBT2802-2809 (*rus*)	0.0505	16.221	6.689	0.879
Chondroitin Sulfate	Δ*tdk* (isogenic parent strain)	0.1042	10.536	1.672	0.592
Chondroitin Sulfate	ΔBT2802-2809 (*rus*)	0.1190	10.368	0.836	0.557
Heparin Sulfate	Δ*tdk* (isogenic parent strain)	0.0028	144.520	135.657	0.144
Heparin Sulfate	ΔBT2802-2809 (*rus*)	0.0051	67.391	57.859	0.141
α-mannan	Δ*tdk* (isogenic parent strain)	0.0674	8.529	-0.334	0.541
α-mannan	ΔBT2802-2809 (*rus*)	0.0595	8.529	-1.003	0.497
Mucin-*O*-glycans	Δ*tdk* (isogenic parent strain)	0.0074	8.863	0.000	0.848
Mucin-O-glycans	ΔBT2802-2809 (*rus*)	0.0074	7.693	-1.840	0.813
Dextran	Δ*tdk* (isogenic parent strain)	0.1019	18.897	10.033	0.920
Dextran	ΔBT2802-2809 (*rus*)	0.1086	18.060	8.528	0.926
Arabinan	Δ*tdk* (isogenic parent strain)	0.0366	10.870	2.007	0.754
Arabinan	ΔBT2802-2809 (*rus*)	0.0391	9.030	-0.502	0.725
Arabinogalactan	Δ*tdk* (isogenic parent strain)	0.0642	23.244	14.381	0.817
Arabinogalactan	ΔBT2802-2809 (*rus*)	0.0581	16.890	7.358	0.748
Pectic galactan (potato)	Δ*tdk* (isogenic parent strain)	0.1273	6.689	-2.174	0.809
Pectic galactan (potato)	ΔBT2802-2809 (*rus*)	0.1265	6.355	-3.178	0.740
Pectic galactan (lupin)	Δ*tdk* (isogenic parent strain)		5.853	-3.010	0.961
Pectic galactan (lupin)	ΔBT2802-2809 (*rus*)	0.0954	7.024	-2.509	0.798
Polygalacturonate	Δ*tdk* (isogenic parent strain)	0.0858	8.863	0.000	0.660
Polygalacturonate	ΔBT2802-2809 (*rus*)	0.1078	10.870	1.338	0.595
Rhamnogalacturonan I	Δ*tdk* (isogenic parent strain)	0.0419	6.020	-2.843	0.490
Rhamnogalacturonan I	ΔBT2802-2809 (*rus*)	0.0464	6.020	-3.512	0.497
Inulin	Δ*tdk* (isogenic parent strain)	0.0074	27.593	18.729	0.695
Inulin	ΔBT2802-2809 (*rus*)	0.0059	27.091	17.558	0.637
Levan	Δ*tdk* (isogenic parent strain)	0.1191	8.529	-0.334	0.896
Levan	ΔBT2802-2809 (*rus*)	0.1265	8.696	-0.836	0.868
Pullulan	Δ*tdk* (isogenic parent strain)	0.1040	18.897	10.033	0.974
Pullulan	ΔBT2802-2809 (*rus*)	0.1144	18.897	9.364	0.976
Glycogen	Δ*tdk* (isogenic parent strain)	0.1038	15.218	6.354	0.972
Glycogen	ΔBT2802-2809 (*rus*)	0.1081	13.378	3.846	0.994
AP	Δ*tdk* (isogenic parent strain)				0.334
AP	ΔBT2802-2809 (*rus*)				0.209

Footnote: Abbreviations (NA : Not applicable, NG : No Growth, NT : Not Tested, NC: Not calculated due to poor/weak growth). When no growth was seen for wild-type (Δtdk), which is the isogenic parent strain background that all deletion strains were made in, the indicated substrate was not tested for rus deletion strains. Unless otherwise noted, substrates were a final concentration of 5 mg/ml and 0.5 mg/ml were added to all substrates except glucose

Genetic manipulation and recombinant protein purification in E. coli

To create a nucleoside hydrolase-free expression background, *E. coli* BL21-AI™ One Shot® cells (Invitrogen) were manipulated using lambda red recombineering to introduce genetic deletions of the ribose-inducible hydrolase genes (*rih*) to avoid contaminating activity in downstream applications of purified proteins[38]. The *E. coli* gene deletion procedure developed by Datsenko and Wanner[39] was followed with few modifications. Briefly, BL21-AI cells were transformed with the pKD46 plasmid. Transformed cells were grown overnight in LB + Amp100 and sub-cultured, when the culture absorbance (600 nm) reached 0.1, L-arabinose was added to 10 mM final concentration to induce the P_{BAD} promoter of pKD46, cells were allowed to grow to an OD between 0.6-0.8 and made competent for electroporation by cold water washes and stored in 10% glycerol aliquots. For recombineering, 400ng of gel-purified PCR product was added to freshly made cells and incubated for 10 minutes on ice, electroporated in a 2mm gap cuvette at 2500 V, recovered in 1 ml LB at 30°C for 5 hours. All knockouts were made sequentially in this manner via introduction of the following antibiotic cassettes (spectinomycin from K11497 for Δ*rihA*; hygromycin from K11521 for Δ*rihB*; gentamicin from K11590 for Δ*rihC*), and the following concentrations of antibiotic were used for selection: Spec80, Hygro200, Gent10. Following construction of the last deletion, the pKD46 plasmid was heat-cured by passaging twice at 42°C in LB. To better control background expression of the T7 promoter, the T7 lysozyme containing plasmid, pLysS from BL21 (DE3) (Lucigen) was introduced into the strain via Ca^{2+} chemical competence/heat shock. Protein purification was accomplished using the pETite N-His vector (Lucigen). PCR primers were designed to amplify products for BT2803, BT2804, BT2807 and BT2808 containing all amino acids for BT2804 residues 1-311, or all amino acids downstream of the predicted signal peptide sequences, residues 22-539 for BT2807 and residues 22-338 for BT2808, for BT2803 two constructs were made containing either all amino acids 1-321 or a construct based on an alternative start site containing residues 15-321 (only this construct produced robust expression, while the full length failed to provide active product or good expression), amplified and transformed into Hi-Control 10G cells according to manufactures protocol (Lucigen, *Expresso*™ T7 cloning and expression system). pETite

plasmids containing BT2803, BT2804, or BT2807 were transformed into *E. coli* strains TUNER or for BT2808 into BL21-AI Δ*rihABC* + pLysS. A single colony was grown in 5 mL of LB+Kan[50] for 16h. This pre-inoculum was added to to 1L of Terrific-Broth with 50ng/ul of Kanamycin and 10 ng/ul of Chloramphenicol (BT2808) or 50ng/ul of Kanamycin (BT2807) and culture was grown with shaking at 37 °C until absorbance 0.4 at 600nm. BT2807 and BT2808 cells were induced with a final concentration of 0.2mM or 1 mM IPTG and 0.2% 20mM L-arabinose, respectively, and temperature was reduced to 16°C and outgrown overnight. The recombinant proteins were purified by immobilized metal ion affinity chromatography using cobalt (BT2807) or nickel-affinity (BT2808) columns was accomplished as described previously[40].

Measurements of transcriptional responses by qPCR

Bt and other *Bacteroides* strains were grown to mid-exponential phase 0.6-0.8 (absorbance at 600nm) in MM-ribose, MM-arabinose, MM-xylose, or MM-glucose, two volumes of RNA protect added, followed by centrifugation and storage of cell pellets at -80°C. Total RNA was extracted using the RNeasy mini kit buffers (Qiagen) and purified on RNA-binding spin columns (Epoch), treated with TURBO DNaseI (Ambion) or DNase I (NEB) after elution and purified again using a second RNeasy mini kit isolation column. Reverse transcription was performed using SuperScript III reverse transcriptase and random primers (Invitrogen). The abundance of each target transcript in the resulting cDNA was quantified using either KAPA SYBR® FAST qPCR mix (KAPA Biosystems) or a homemade qPCR mix as described previously[41]. Each 20 uL reaction contained 1X Thermopol Reaction Buffer (NEB), 125uM dNTPs, 2.5mM MgSO4, 1X SYBR Green I (Lonza), 500nM gene specific or 65nM 16S rRNA primer and 0.5 units Hot Start *Taq* Polymerase (NEB), and 10ng of template cDNA. For the KAPA mix, 400 nM of primers specific for genes in the *rus* locus of *Bt* or the *rusC*-like gene of other *Bacteroides* species or 62.5 nM of 16S rRNA primers and 10ng of template cDNA as described previously[42]. Using the ddCT method, raw values were normalized to 16S rRNA values and then MM+ribose values were referenced to the values obtained in MM-glucose to obtain a fold-change. Measurements of transcriptional response over time in MM-ribose or nucleosides was performed similarly to previously described[40]. Briefly, strains were grown in TYG, subcultured 1:50 into MM-glucose, at mid-exponential phase, cells were washed twice in

MM-no carbon and resuspended in MM-ribose with time points being taken every 5 min for the first 30 min and every 15 min for a total of 120 min. Measurements of transcriptional responses to varying amounts of ribose were performed similarly as above, but only one time point was taken after 30 min of exposure to varying concentration of MM-ribose ranging from 0.0005 mg/ml to 5mg/ml.

Antibody production, western blotting and immunofluorescent microscopy

Purified recombinant BT2807 and BT2808 proteins were used as antigens to raise rabbit polyclonal antibodies (Cocalico Biologicals, Inc, Stevens PA). Antibody specificity and cellular localization for BT2807 and BT2808 were determined by western blotting of wild-type and relevant mutant strains and by immunofluorescent microscopy of *Bt* VPI-5482 grown in MM+glucose or MM+ribose. Growth conditions are described above, cells for WB were grown to mid-log optical absorbance (600 nm) 0.6-0.7 or 0.4-0.5 for IF. Western blots of *Bt* whole cell lysates were performed using the primary, polyclonal antibodies mentioned above and secondary antibody conjugated to goat anti-Rabbit IgG conjugated alkaline phosphatase (Sigma) and detected with NBT/BCIP (Roche). Surface expression of BT2807 or BT2808 was examined by staining with a BT2807- or BT2808-specific primary antibody in non-permeabilized formaldehyde-fixed *Bt* cells and detected with Alexa-Flour® 488 conjugated goat anti-Rabbit IgG secondary (Molecular Probes), as described previously[40]. Cells were imaged on an IX-70 inverted microscope (Olympus) with images captured at 100x magnification. A minimum of five fields of view per slide was observed with n=2 biological replicates.

RNAseq analysis

To determine the global transcriptional response to growth in ribose as the sole carbon source, *Bt* cells were grown overnight in rich TYG media then transferred to fresh MM containing either 5 mg/ml glucose or 5 mg/ml ribose. Cells were then grown until mid-log phase (absorbance between 0.6-0.8) and two volumes of RNA Protect (Qiagen) were added to cells. RNA was isolated as described above and purified whole RNA was then rRNA depleted using the Ribo-Zero Bacterial rRNA Removal Kit (Illumina Inc.) and concentrated with the RNA Clean and Concentrator-5 kit (Zymo Research Corp, Irvine, CA). Samples were multiplexed for sequencing on the Illumina HiSeq platform at the University of Michigan Sequencing Core. Data

was analyzed using Arraystar software (DNASTAR, Inc.) using RPKM normalization with default parameters. Gene expression in ribose was compared to gene expression in a glucose reference. Genes with significant up- or down-regulation were determined by the following criteria: genes with an average fold-change ≥5-fold and with at least 2/3 biological replicates with a normalized expression level ≥1% of the overall average RPKM expression level in either glucose or ribose, and a p-value < 0.05 (t test with Benjamini-Hochberg correction) (Table 2.3).

Table 2.3 RNAseq results of wild-type Bt grown on MM+ribose compared to MM+glucose

Gene name	Fold change R/G	P value R/G	Annotation	CAZY	PFAM
BT0094	5.842	4.72E-04	conserved protein found in conjugate transposon	NA	DUF4133
BT0107	6.766	6.22E-03	hypothetical protein	NA	Helix-turn-helix motif, HTH_17
BT0437	5.038	1.05E-02	N-acylglucosamine 2-epimerase	NA	GlcNAc_2-epim
BT0565	8.201	7.66E-04	putative heat shock protein	NA	HSP20
BT0656	5.324	4.72E-04	hypothetical protein	NA	None
BT0786	6.639	4.72E-04	putative integral membrane protein	NA	DUF4396
BT0787	8.058	4.72E-04	succinyl-CoA synthetase alpha chain	NA	None
BT0788 (sucC)	8.087	5.07E-04	succinyl-CoA synthetase (ADP-forming) beta subunit	NA	Ligase_CoA, ATP-grasp_2
BT0805	5.023	4.72E-04	hypothetical protein	NA	DUF2776
BT0854	6.653	3.68E-03	hypothetical protein	NA	None
BT0970	0.121	2.52E-04	haloacid dehalogenase-like hydrolase	NA	HAD_2
BT1009	6.775	1.65E-03	dihydroorotate dehydrogenase	NA	None
BT1096	5.378	4.72E-04	transposase	NA	DDE_Tnp_4, DDE superfamily endonuclease
BT1097	5.207	7.28E-04	hypothetical protein	NA	None
BT1196	5.011	4.72E-04	pyruvate carboxylase subunit B)	NA	HMGL-like, PYC_OADA
BT1259	5.335	4.72E-04	choloylglycine hydrolase	NA	CBAH, Linear amide C-N hydrolases, choloylglycine hydrolase family
BT1323	5.560	4.72E-04	putative ABC transporter permease protein	NA	Binding-protein-dependent transport system inner membrane component
BT1435	6.042	4.72E-04	hypothetical protein	NA	None
BT1448	8.568	1.20E-03	biotin carboxyl carrier protein	NA	PF00364, Biotin_lipoyl
BT1449 (accC1)	6.820	7.16E-04	biotin carboxylase	NA	None
BT1450	8.254	5.00E-04	propionyl-CoA carboxylase beta chain	NA	None
BT1564 (queF)	0.190	1.65E-02	7-cyano-7-deazaguanine reductase (Queuosine biosynthesis)		QueF
BT1757	6.243	4.72E-04	fructokinase	NA	None
BT1758	21.519	1.04E-03	glucose/galactose transporter	NA	None
BT1759	12.264	5.44E-04	levanase precursor (2,6-beta-D-fructofuranosidase)	GH32	None
BT1760	16.220	2.87E-03	glycosylhydrolase	GH32	None
BT1761	15.049	2.87E-03	hypothetical protein	NA	None
BT1762	17.388	4.84E-03	putative outer membrane protein, probably involved in nutrient binding		None
BT1763	13.577	2.00E-03	putative outer membrane protein, probably involved in nutrient binding		None

BT1765	13.231	2.19E-04	levanase precursor (2,6-beta-D-fructofuranosidase)	GH32	None
BT1914	7.540	5.35E-04	thioredoxin-like protein, putative thioredoxin	NA	None
BT1960	5.196	2.28E-02	integrase	NA	Phage integrase SAM-like domain
BT2082	5.906	4.72E-04	hypothetical protein	NA	Outer membrane protein beta-barrel domain
BT2083	6.627	4.72E-04	hypothetical protein	NA	Calycin-like beta-barrel domain
BT2156	0.044	5.59E-04	putative sugar phosphate isomerase/epimerase	NA	AP_endonuc_2 (Xylose isomerase-like TIM barrel)
BT2157	0.040	4.91E-04	hypothetical protein	NA	DUF1080
BT2158	0.042	9.12E-04	putative dehydrogenases and related proteins	NA	GFO_IDH_MocA (Oxidoreductase family, NAD-binding Rossmann fold)
BT2159	0.041	4.72E-04	putative oxidoreductase	NA	GFO_IDH_MocA (Oxidoreductase family, NAD-binding Rossmann fold)
BT2167	8.370	5.77E-03	elongation factor G	NA	None
BT2178	0.056	4.72E-04	hypothetical protein	NA	None
BT2297	5.869	4.91E-04	putative reverse transcriptase	NA	None
BT2298	6.578	1.76E-02	conserved protein found in conjugate transposon	NA	DUF3875
BT2300	5.439	9.78E-03	conserved protein found in conjugate transposon	NA	DUF4134
BT2304	6.707	8.52E-03	hypothetical protein	NA	None
BT2323	5.654	1.31E-03	hypothetical protein	NA	DUF3945, DUF4099
BT2334	6.376	9.85E-03	hypothetical protein	NA	Helix-turn-helix motif, HTH_17
BT2442	6.746	1.75E-03	major outer membrane protein OmpA	NA	None
BT2569	6.166	4.91E-04	RNA polymerase ECF-type sigma factor	NA	Sigma 70
BT2756	5.109	7.30E-04	anaerobic C4-dicarboxylate transporter dcuB	NA	DcuA_DcuB
BT2803	195.309	4.72E-04	ribokinase	NA	None
BT2804	178.415	5.50E-04	ribokinase	NA	None
BT2805	110.793	7.16E-04	SusC-like	NA	None
BT2806	110.268	6.69E-04	SusD-like	NA	None
BT2807	119.753	5.79E-04	GH*	NA	None
BT2808	130.147	4.91E-04	putative inosine-uridine preferring nucleoside hydrolase	NA	None
BT2809	146.585	4.72E-04	Permease	NA	None
BT2872	5.560	2.35E-03	putative capsular polysaccharide synthesis protein	GT32	Capsular polysaccharide synthesis protein
BT3024	0.126	1.92E-02	putative outer membrane protein, probably involved in nutrient binding		TonB_dep_Rec
BT3025	0.156	2.52E-02	putative outer membrane protein, probably involved in nutrient binding		SusD, SusD-like_3
BT3026	0.118	2.76E-02	glycosylhydrolase, putative xylanase	GH30_6	Glyco_hydro_30
BT3027	0.165	3.20E-02	hypothetical protein	NA	None
BT3100	5.076	7.16E-04	lipase, putative esterase	NA	Abhydrolase_3, Peptidase_S9
BT3113	0.186	4.91E-04	putative transmembrane efflux protein	NA	MFS_1
BT3114	0.173	2.97E-03	beta-galactosidase	GH2	Glyco_hydro_2_N, Glyco_hydro_2_N, PA14
BT3167	9.709	1.02E-02	hypothetical protein	NA	HHH-3, Helix-hairpin-helix motif
BT3344	0.044	9.01E-03	hypothetical protein	NA	DUF4361, DUF4973
BT3345	0.034	7.15E-03	conserved hypothetical protein, putative outer membrane protein		SusD, SusD-like_3

BT3346	0.096	1.37E-02	putative outer membrane protein, probably involved in nutrient binding		TonB_dep_Rec
BT3347	0.069	6.69E-03	hypothetical protein	NA	IPT/TIG domain
BT3415	10.103	5.73E-03	hypothetical protein	NA	None
BT3735	6.293	4.72E-04	hypothetical protein	NA	None
BT3823	5.952	4.72E-04	hypothetical protein	NA	Ferritin
BT3916	5.378	8.53E-04	site-specific recombinase IntIA	NA	Phage_integrase
BT4542	5.835	6.40E-04	Type I restriction enzyme EcoR124II specificity protein	NA	Methylase_S
BT4676	5.936	5.55E-04	putative periplasmic protein	NA	Putative beta-lactamase-inhibitor-like, PepSY-like
BT4677	5.813	4.72E-04	hypothetical protein	NA	Putative beta-lactamase-inhibitor-like, PepSY-like
BT4686	8.068	2.07E-02	hypothetical protein	NA	
BT4688	5.808	8.56E-04	hypothetical protein	NA	Mechanosensitive ion channel
BT4689	7.074	4.91E-04	pullulanase precursor	GH13_14	Alpha-amylase, CBM_48
BT4690	5.562	4.72E-04	alpha-amylase precursor	GH13_5	Alpha-amylase
BT4744	5.231	4.72E-04	putative multiple inositol polyphosphate histidine phosphatase 1		Histidine phosphatase superfamily (branch 2)

Footnote: This list was trimmed using the following parameters, gene represented above 1% of total RPKM abundance in at least 2 replicates of one condition either glucose or ribose. The list was further trimmed by only including genes with a p-value < 0.05. Lastly, we only considered genes of greater than a 5 fold up- or downregulation compared to growth in glucose. Abbreviations: NA (not applicable), GH (glycoside hydrolase), GT (glycosyl or glycoside transferase), CAZY (Carbohydrate Active Enzymes).

Functional annotation and comparative genomics of rus PULs across Bacteroidetes genomes

Initial functional annotations of *Bt rus* genes were taken from the Integrated Microbial Genomes (IMG) database using the Pfam, InterPro, COG, or KOG predictions. In cases where multiple annotations, we selected the more inclusive terms (*e.g.* nucleoside phosphorylase instead of purine or pyrimidine-specific nucleoside phosphorylase). A total of 354 different *Bacteroidetes* strains were tested for growth on ribose as a sole carbon source as shown in (Figure 2.10A) and summarized in (Table 2.4)**.** The ability to use ribose is shown in the context of a previously published human gut Bacteroidetes phylogeny that used 14 conserved genes across phylum members[10]. To search for *rus* locus homologs across the Bacteroidetes phylum, we used the amino acid sequences of the *rusK1, rusK2, rusT,* and *rusR* genes from the *Bt* type strain as deletion of these genes yielded growth defects on ribose. We searched the IMG database (current as of May 2018) and performed phylum-level BLAST searches with an E-value cutoff of 1e-50. We chose this stringent cutoff as initial searches using lower values obtained many non-specific hits of genes encoding other kinases and permeases that did not appear to be specific for ribose, including in the *Bt* VPI-5482 genome for which Rusk1 and RusK1 are the only kinases able to promote ribose growth. After we completed our search for *rusK, rusT,* and *rusR* homologs we used the Gene Neighborhood tool in IMG to determine if these hits were located directly next to other genes involved in ribose utilization. The presence of a minimum of

two adjacent *rus* gene homologs was required to count the presence of a candidate utilization locus. Following this first round of searching we observed that many of the *rus* loci contained one or more nucleoside cleaving enzymes such as homologs of *Bt rusNH* or ADP-ribosylglycohydrolases (RGH) and upstream putative regulatory genes. To give our search more power and potentially find additional *rus* homologs we performed additional searches with the same E-value threshold for homologs of *Bt rusNH,* or homologs of the ADP-RGH in *B. xylanisolvens XB1A*. When assembling the comparative genomics data, gene names and glycoside hydrolase family assignments are shown as predicted within IMG by either annotation, Pfam and/or InterPro predictions or confirmed by BLAST of the amino acid sequence of individual genes. Further, in refinement, a handful of genes were found below our E-value, but included in the table as it is clear from gene neighborhood views in IMG that it is likely part of a *rus* locus due to adjacent hits of *rus* homologs. Types of *rus* have been assigned based only on gene content and arrangement as a way to indicate differences, however subtle. In completing our table, we have included the bit score as well as the amino acid % identities compared to *Bt rus* genes or *Bx XB1A* ADP-RGH genes. All of the positive gene hits with locus tag information, isolation location, and other relevant strain information is summarized in the published manuscript of this chapter, but due to concerns of including 75 pages worth of tables has been omitted here.

Table 2.4 Growth of human and animal gut Bacteroidetes on ribose as a sole carbon source

Species[1]	Strain	% identity to species type strain (bp covered)	Host species	Isolation period	Growth (600nm)	Rate
B. caccae	ATCC 43185	100 (type strain)	Human	pre-1980	1.02	0.00156
B. caccae	VPI-3452A	100 (type strain)	Human	pre-1980	0.00	0.00000
B. caccae	VPI-B6-11	99.8 (836/838)	Human	pre-1980	1.13	0.00114
B. caccae	WH110	99.6 (795/798)	Human	1995-99	1.09	0.00146
B. caccae	VPI-C14-3	99.6 (830/833)	Human	pre-1980	0.00	0.00000
B. caccae	VPI-C7-8	99.6 (832/835)	Human	pre-1980	1.28	0.00134
B. caccae	WAL8714	99.6 (832/835)	Human	unknown	0.90	0.00058
B. caccae	WH719	99.5 (770/774)	Human	1995-99	0.72	0.00033
B. caccae	VPI-T1-1	99.5 (850/854)	Human	pre-1980	0.85	0.00040
B. caccae	VPI-8608	99.4 (834/839)	Human	pre-1980	1.20	0.00165
B. caccae	CL03T12C61	99.4 (994/1000)	Human	2000 or later	1.20	0.00153
B. caccae	VPI-C10-2	99.1 (778/785)	Human	pre-1980	0.96	0.00119
B. cellulosilyticus	DSM 14838	100 (type strain)	Human	2000 or later	1.20	0.00192
B. cellulosilyticus	WH1	99.9 (795/796)	Human	1995-99	1.13	0.00178
B. cellulosilyticus	WH403	99.8 (851/853)	Human	1995-99	1.13	0.00211
B. cellulosilyticus	WH206	99.8 (844/846)	Human	1995-99	1.11	0.00227
B. cellulosilyticus	WH2	99.8 (838/840)	Human	1995-99	1.19	0.00185
B. cellulosilyticus	WH401	99.8 (838/840)	Human	1995-99	1.15	0.00195
B. cellulosilyticus	CL02T12C19	99.7 (916/919)	Human	2000 or later	0.00	0.00000

B. cellulosilyticus	WH402	99.6 (844/847)	Human	1995-99	1.11	0.00183
B. cellulosilyticus	WH405	99.5 (827/831)	Human	1995-99	1.19	0.00201
B. cellulosilyticus	WH101	99.4 (802/807)	Human	1995-99	1.22	0.00201
B. clarus	DSM 22519	100 (type strain)	Human	2000 or later	0.00	0.00000
B. dorei	DSM 17855	100 (type strain)	Human	2000 or later	0.00	0.00000
B. dorei	CL02T12C06	100 (1520/1520)	Human	2000 or later	0.31	0.00018
B. dorei	WH106	100 (792/792)	Human	1995-99	0.00	0.00000
B. dorei	WH26	100 (830/830)	Human	1995-99	0.00	0.00000
B. dorei	WH303	100 (870/870)	Human	1995-99	0.92	0.00040
B. dorei	WH512	99.9 (870/871)	Human	1995-99	0.00	0.00000
B. dorei	9_1_42FAA	99.9 (859/860)	Human	2000 or later	0.00	0.00000
B. dorei	3_1_33FAA	99.9 (858/859)	Human	2000 or later	0.00	0.00000
B. dorei	VPI-2277	99.9 (853/854)	Human	pre-1980	0.86	0.00082
B. dorei	WH607	99.9 (852/853)	Human	1995-99	0.00	0.00000
B. dorei	VPI-6598B	99.9 (845/846)	Human	pre-1980	0.00	0.00000
B. dorei	WH104	99.8 (811/813)	Human	1995-99	0.00	0.00000
B. dorei	CL02T00C15	99.7 (882/885)	Human	2000 or later	0.30	0.00016
B. dorei	CL03T12C01	99.6 (956/960)	Human	2000 or later	0.00	0.00000
B. eggerthii	DSM 20697, ATCC 27754	100 (type strain)	Human	pre-1980	0.00	0.00000
B. eggerthii	1_2_48FAA	99.6 (844/847)	Human	2000 or later	0.00	0.00000
B. eggerthii	VPI-S1A-52	99.5 (831/835)	Human	pre-1980	0.00	0.00000
B. finegoldii	DSM 17565	100 (type strain)	Human	2000 or later	0.00	0.00000
B. finegoldii	WH508	99.6 (843/846)	Human	1995-99	0.00	0.00000
B. finegoldii	CL09T03C10	96.9 (991/1023)	Human	2000 or later	0.00	0.00000
B. fluxus	DSM 22534	100 (type strain)	Human	2000 or later	0.00	0.00000
B. fragilis	NCTC 9343, ATCC 25285	100 (type strain)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-4517	100 (836/836)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-4509b	100 (848/848)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF7639	99.9 (850/851)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-6779	99.9 (849/850)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-5383 / 23745	99.9 (843/844)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF7397	99.9 (842/843)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF119	99.9 (841/842)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF7567	99.9 (839/840)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF8371	99.9 (839/840)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-3277	99.9 (830/831)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-2553	99.9 (791/792)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-1582	99.9 (785/786)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF V479	99.8 (871/873)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF-CEST	99.8 (846/848)	Human	pre-1980	0.00	0.00000
B. fragilis	WH709	99.8 (845/847)	Human	1995-99	0.00	0.00000
B. fragilis	VPI-1522	99.8 (845/847)	Human	pre-1980	0.00	0.00000
B. fragilis	WAL8762	99.8 (845/847)	Human	unknown	0.00	0.00000
B. fragilis	VPI-499	99.8 (840/842)	Human	pre-1980	0.00	0.00000
B. fragilis	638R	99.8 (840/842)	Human	pre-1980	0.00	0.00000
B. fragilis	WH706	99.8 (824/826)	Human	1995-99	0.00	0.00000
B. fragilis	WH707	99.8 (820/822)	Human	1995-99	0.00	0.00000
B. fragilis	CL05T12C13	99.7 (985/988)	Human	2000 or later	0.00	0.00000
B. fragilis	CL07T00C01	99.6 (897/901)	Human	2000 or later	0.00	0.00000
B. fragilis	VPI-BF-ERL	99.6 (855/858)	Human	pre-1980	0.00	0.00000
B. fragilis	WAL8916	99.6 (855/858)	Human	unknown	0.00	0.00000
B. fragilis	2_1_56FAA	99.6 (853/856)	Human	2000 or later	0.00	0.00000
B. fragilis	VPI-4361	99.6 (849/852)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF8223	99.6 (846/849)	Human	pre-1980	0.00	0.00000
B. fragilis	WAL8774	99.6 (845/848)	Human	unknown	0.00	0.00000
B. fragilis	WH705	99.6 (836/839)	Human	1995-99	0.00	0.00000
B. fragilis	VPI-2044	99.6 (833/836)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-BF-AK87	99.6 (833/836)	Human	pre-1980	0.00	0.00000

B. fragilis	VPI-12256	99.6 (832/835)	Human	pre-1980	0.00	0.00000
B. fragilis	WH718	99.6 (807/810)	Human	1995-99	0.00	0.00000
B. fragilis	CL03T00C08	99.6 (1026/1030)	Human	2000 or later	0.00	0.00000
B. fragilis	YCH46	99.5 (776/780)	Human	unknown	0.00	0.00000
B. fragilis	VPI-BF8928	99.4 (852/857)	Human	pre-1980	0.00	0.00000
B. fragilis	3_2_5	99.4 (845/850)	Human	2000 or later	0.00	0.00000
B. fragilis	VPI-29765	99.4 (844/849)	Human	pre-1980	0.00	0.00000
B. fragilis	WAL8790	99.4 (835/840)	Human	unknown	0.00	0.00000
B. fragilis	WH605	99.4 (815/820)	Human	1995-99	0.00	0.00000
B. fragilis	CL05T00C42	99.4 (1090/1096)	Human	2000 or later	0.00	0.00000
B. fragilis	VPI-2556I	99.3 (834/840)	Human	pre-1980	0.00	0.00000
B. fragilis	CL07T12C05	99.3 (1051/1058	Human	2000 or later	0.00	0.00000
B. fragilis	CL03T12C07	99.3 (1014/1021)	Human	2000 or later	0.00	0.00000
B. fragilis	VPI-2552	98.7 (820/831)	Human	pre-1980	0.00	0.00000
B. fragilis	3_1_12	98.6 (838/850)	Human	2000 or later	0.00	0.00000
B. fragilis	VPI-2627-J2	98.6 (804/815)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-3392	98.5 (842/855)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-4076	98.5 (835/848)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-4225	98.5 (834/847)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-2393	98.5 (780/792)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-2343	98.5 (779/791)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-A11-24B	98.4 (782/795)	Human	pre-1980	0.00	0.00000
B. fragilis	VPI-4117	98.1 (773/788)	Human	pre-1980	0.00	0.00000
B. intestinalis	DSM 17393	100 (type strain)	Human	2000 or later	0.99	0.00145
B. intestinihominis	DSM 21032	100 (type strain)	Human	2000 or later	0.00	0.00000
B. massiliensis	B84634, DSM 17679	100 (type strain)	Human	2000 or later	0.00	0.00000
B. massiliensis	JCM12982	99.7 (1486/1491)	Human	2000 or later	0.00	0.00000
B. massiliensis	A03	99.5 (865/869)	Human	2000 or later	0.12	0.00002
B. nordii	*CL02T12C05*	100 (type strain)	Human	2000 or later	0.00	0.00000
B. nordii	WAL7936	99 (852/861)	Human	unknown	0.00	0.00000
B. nordii	WH103	99 (806/814)	Human	1995-99	0.00	0.00000
B. nordii	WAL7935	98.8 (848/858)	Human	unknown	0.00	0.00000
B. oleiciplenus	DSM 22535	100 (type strain)	Human	2000 or later	1.04	0.00171
B. ovatus	ATCC 8483	100 (type strain)	Human	pre-1980	0.88	0.00048
B. ovatus	NLAE-zl-C34	99.9 (787/788)	Cow	2000 or later	1.00	0.00057
B. ovatus	NLAE-zl-C11	99.9 (787/788)	Cow	2000 or later	1.00	0.00068
B. ovatus	WH702	99.9 (756/757)	Human	1995-99	1.12	0.00137
B. ovatus	WH711	99.9 (687/688)	Human	1995-99	1.16	0.00083
B. ovatus	WH211	99.8 (850/852)	Human	1995-99	1.14	0.00091
B. ovatus	WH214	99.8 (827/829)	Human	1995-99	1.15	0.00114
B. ovatus	CL02T12C04	99.7 (1032/1035)	Human	2000 or later	1.02	0.00075
B. ovatus	3_8_47FAA	99.6 (849/852)	Human	2000 or later	1.14	0.00148
B. ovatus	VPI-3049	99.6 (845/848)	Human	pre-1980	1.05	0.00085
B. ovatus	VPI-C1-45	99.6 (837/840)	Human	pre-1980	1.13	0.00072
B. ovatus	VPI-4104	99.6 (834/837)	Human	pre-1980	1.05	0.00082
B. ovatus	WH713	99.6 (811/814)	Human	1995-99	1.03	0.00150
B. ovatus	WAL7922	99.5 (859/863)	Human	unknown	1.08	0.00184
B. ovatus	WH208	99.5 (818/822)	Human	1995-99	1.16	0.00137
B. ovatus	VPI-8653	99.5 (786/790)	Human	pre-1980	1.00	0.00090
B. ovatus	VPI-435	99.4 (820/825)	Human	pre-1980	1.16	0.00207
B. ovatus	NLAE-zl-H366	99.4 (1387/1395)	Human	2000 or later	0.93	0.00117
B. ovatus	NLAE-zl-H251	99.4 (1383/1391)	Human	2000 or later	0.81	0.00115
B. ovatus	VPI-38	99.3 (801/807)	Human	pre-1980	0.88	0.00083
B. ovatus	CL03T12C18	99.3 (1023/1030)	Human	2000 or later	0.99	0.00150
B. ovatus	WH514	99.1 (851/859)	Human	1995-99	1.00	0.00062
B. ovatus	VPI-B4-11	99.1 (843/851)	Human	pre-1980	1.11	0.00186
B. ovatus	NLAE-zl-H163	98.7 (1377/1395)	Human	2000 or later	0.93	0.00082
B. ovatus	NLAE-zl-H304	98.6 (1380/1400)	Human	2000 or later	0.94	0.00115

B. ovatus	NLAE-zl-H361	98.5 (1380/1401)	Human	2000 or later	0.96	0.00136
B. ovatus	VPI-C16-22	98.4 (825/838)	Human	pre-1980	1.17	0.00230
B. ovatus	3_1_23	98.3 (834/848)	Human	2000 or later	0.94	0.00045
B. ovatus	WH601	97.9 (834/852)	Human	1995-99	1.04	0.00201
B. ovatus	WH604	97.7 (834/854)	Human	1995-99	1.09	0.00128
B. ovatus	WH606	97.7 (821/840)	Human	1995-99	1.12	0.00109
B. ovatus	D2 (2_1_39)	97.7 (817/836)	Human	2000 or later	1.14	0.00126
B. ovatus	NLAE-zl-H59	97.3 (1249/1284)	Human	2000 or later	0.98	0.00144
B. ovatus	NLAE-zl-H73	95 (1218/1284)	Human	2000 or later	0.98	0.00133
B. plebeius	DSM 17135	100 (type strain)	Human	2000 or later	0.00	0.00000
B. salyersae	DSM 18765, ATCC BAA-997	100 (type strain)	Human	2000 or later	0.00	0.00000
B. salyersae	WAL7960	99.4 (846/851)	Human	unknown	0.00	0.00000
B. salyersae	WAL9166	99.3 (832/838)	Human	unknown	0.00	0.00000
B. salyersae	VPI-2828	99.2 (845/852)	Human	pre-1980	0.00	0.00000
B. salyersae	CL02T12C01	98.9 (842/851)	Human	2000 or later	0.00	0.00000
B. stercoris	ATCC 43183, VPI B5-21	100 (type strain)	Human	pre-1980	0.00	0.00000
B. stercoris	WH102	99.7 (807/809)	Human	1995-99	1.23	0.00163
B. stercoris	WH22	99.6 (833/836)	Human	1995-99	0.80	0.00069
B. stercoris	VPI-B5-21	99.5 (846/850)	Human	pre-1980	0.00	0.00000
B. stercoris	WH24	99.5 (824/828)	Human	1995-99	0.80	0.00075
B. stercoris	VPI-C8-19	99.3 (845/851)	Human	pre-1980	1.02	0.00092
B. stercoris	VPI-C51-6	99 (825/833)	Human	pre-1980	0.00	0.00000
B. thetaiotaomicron	VPI-5482, ATCC 29148	100 (type strain)	Human	pre-1980	1.21	0.00143
B. thetaiotaomicron	NLAE-zl-G288	100 (1223/1233)	Goat	2000 or later	1.39	0.00120
B. thetaiotaomicron	NLAE-zl-H492	100 (1227/1227)	Human	2000 or later	0.98	0.00074
B. thetaiotaomicron	NLAE-zl-H463	100 (1228/1228)	Human	2000 or later	1.05	0.00088
B. thetaiotaomicron	VPI-2808B	100 (751/751)	Human	pre-1980	1.19	0.00184
B. thetaiotaomicron	VPI-BT7853	100 (751/751)	Human	pre-1980	1.31	0.00133
B. thetaiotaomicron	7330	99.9 (840/841)	Human	pre-1980	1.24	0.00182
B. thetaiotaomicron	VPI-3731	99.9 (826/827)	Human	pre-1980	1.13	0.00143
B. thetaiotaomicron	VPI-BT-DOT2	99.8 (854/856)	Human	pre-1980	0.00	0.00000
B. thetaiotaomicron	1_1_6	99.8 (848/850)	Human	2000 or later	0.98	0.00106
B. thetaiotaomicron	23685	99.8 (846/848)	Human	unknown	0.97	0.00103
B. thetaiotaomicron	VPI-3164	99.8 (846/848)	Human	pre-1980	1.19	0.00217
B. thetaiotaomicron	VPI-0633-1	99.8 (844/846)	Human	pre-1980	1.21	0.00197
B. thetaiotaomicron	WH25	99.8 (822/824)	Human	1995-99	1.26	0.00163
B. thetaiotaomicron	NLAE-zl-P32	99.8 (1417/1420)	Pig	2000 or later	1.00	0.00123
B. thetaiotaomicron	NLAE-zl-P699	99.8 (1416/1419)	Pig	2000 or later	1.06	0.00130
B. thetaiotaomicron	NLAE-zl-C523	99.8 (1239/1241)	Cow	2000 or later	1.33	0.00154
B. thetaiotaomicron	WAL8669	99.7 (864/867)	Human	unknown	1.17	0.00173
B. thetaiotaomicron	WH510	99.7 (783/785)	Human	1995-99	1.19	0.00126
B. thetaiotaomicron	WAL8713	99.6 (854/857)	Human	unknown	1.17	0.00162
B. thetaiotaomicron	WH503	99.6 (833/836)	Human	1995-99	1.29	0.00108
B. thetaiotaomicron	WH507	99.6 (810/813)	Human	1995-99	0.00	0.00000
B. thetaiotaomicron	VPI-J19-343	99.5 (850/854)	Human	pre-1980	1.15	0.00164
B. thetaiotaomicron	23722	99.5 (838/842)	Human	unknown	1.11	0.00185
B. thetaiotaomicron	WH502	99.5 (834/838)	Human	1995-99	1.30	0.00127
B. thetaiotaomicron	WH509	99.5 (823/827)	Human	1995-99	1.19	0.00152
B. thetaiotaomicron	1_1_14	99.5 (730/734)	Human	2000 or later	1.11	0.00144
B. thetaiotaomicron	NLAE-zl-C425	99.5 (636/639)	Cow	2000 or later	1.18	0.00141
B. thetaiotaomicron	VPI-0940-1	99.4 (841/846)	Human	pre-1980	1.17	0.00077
B. thetaiotaomicron	NLAE-zl-P750	99.4 (790/795)	Pig	2000 or later	1.00	0.00072
B. thetaiotaomicron	NLAE-zl-H353	99.4 (785/790)	Human	2000 or later	0.99	0.00042
B. thetaiotaomicron	NLAE-zl-H486	99.4 (498/501)	Human	2000 or later	0.97	0.00062
B. thetaiotaomicron	NLAE-zl-G234	99.4 (497/500)	Goat	2000 or later	1.00	0.00044
B. thetaiotaomicron	NLAE-zl-P696	99.4 (496/499)	Pig	2000 or later	1.10	0.00133
B. thetaiotaomicron	NLAE-zl-P737	99.4 (496/499)	Pig	2000 or later	1.13	0.00130
B. thetaiotaomicron	NLAE-zl-C516	99.3 (724/729)	Cow	2000 or later	1.30	0.00156

B. thetaiotaomicron	NLAE-zl-C504	99.3 (711/716)	Cow	2000 or later	1.27	0.00171
B. thetaiotaomicron	VPI-3443	99.2 (851/858)	Human	pre-1980	1.13	0.00159
B. thetaiotaomicron	VPI-C11-15	99.2 (842/849)	Human	pre-1980	1.13	0.00181
B. thetaiotaomicron	NLAE-zl-H39	99.2 (789/795)	Human	2000 or later	1.07	0.00108
B. thetaiotaomicron	NLAE-zl-C15	99.2 (708/714)	Cow	2000 or later	1.04	0.00105
B. thetaiotaomicron	dnlkv9	99.2 (1371/1382)	Mouse	2000 or later	1.25	0.00158
B. thetaiotaomicron	ATCC 29741	99.1 (783/790)	Human	unknown	1.11	0.00163
B. thetaiotaomicron	WH501	99 (829/837)	Human	1995-99	1.27	0.00121
B. thetaiotaomicron	NLAE-zl-G303	98.7 (790/800)	Goat	2000 or later	1.33	0.00104
B. thetaiotaomicron	NLAE-zl-H23	98.7 (779/789)	Human	2000 or later	1.09	0.00108
B. thetaiotaomicron	VPI-BF6436-5	98.7 (777/787)	Human	pre-1980	1.26	0.00215
B. thetaiotaomicron	NLAE-zl-H207	98.7 (1404/1422)	Human	2000 or later	1.10	0.00054
B. thetaiotaomicron	MAJ 27 (*B. faecis*)	98.5 (1402/1424)	Human	2000 or later	1.19	0.00122
B. thetaiotaomicron	NLAE-zl-G493	98.4 (784/797)	Goat	2000 or later	1.09	0.00101
B. thetaiotaomicron	MAJ 26 (*B. faecis*)	98.2 (853/869)	Human	2000 or later	1.22	0.00126
B. thetaiotaomicron	WAL8736	98.1 (839/855)	Human	unknown	1.14	0.00132
B. thetaiotaomicron	WH21	98.1 (836/852)	Human	1995-99	1.17	0.00148
B. thetaiotaomicron	WH3	97.9 (782/799)	Human	1995-99	1.09	0.00140
B. thetaiotaomicron	VPI-BT8702	97.4 (829/851)	Human	pre-1980	1.18	0.00208
B. thetaiotaomicron	NLAE-zl-G295	97.3 (796/818)	Goat	2000 or later	1.10	0.00082
B. thetaiotaomicron	NLAE-zl-P718	97.2 (486/500)	Pig	2000 or later	1.07	0.00121
B. thetaiotaomicron	VPI-11984	97.1 (802/826)	Human	pre-1980	1.15	0.00197
B. uniformis	ATCC 8492	100 (type strain)	Human	pre-1980	0.34	0.00034
B. uniformis	WH506	100 (822/822)	Human	1995-99	0.38	0.00015
B. uniformis	WH505	100 (835/835)	Human	1995-99	0.69	0.00032
B. uniformis	VPI-60-50	99.9 (867/868)	Human	pre-1980	0.00	0.00000
B. uniformis	WH207	99.9 (857/858)	Human	1995-99	0.33	0.00009
B. uniformis	WH703	99.9 (849/850)	Human	1995-99	0.00	0.00000
B. uniformis	WH504	99.9 (835/836)	Human	1995-99	0.35	0.00013
B. uniformis	WH710	99.9 (835/836)	Human	1995-99	0.23	0.00024
B. uniformis	WH23	99.9 (827/828)	Human	1995-99	0.00	0.00000
B. uniformis	VPI-52	99.9 (825/826)	Human	pre-1980	0.00	0.00000
B. uniformis	WH10	99.9 (822/823)	Human	1995-99	0.56	0.00015
B. uniformis	WH701	99.9 (810/811)	Human	1995-99	0.25	0.00019
B. uniformis	WH12	99.9 (743/744)	Human	1995-99	0.51	0.00013
B. uniformis	R3-39	99.8 (989/991)	Human	unknown	0.42	0.00027
B. uniformis	VPI-S5A-14	99.8 (848/850)	Human	pre-1980	0.00	0.00000
B. uniformis	WH4	99.8 (842/844)	Human	1995-99	0.00	0.00000
B. uniformis	WH20	99.8 (841/843)	Human	1995-99	0.00	0.00000
B. uniformis	WH215	99.8 (835/837)	Human	1995-99	0.00	0.00000
B. uniformis	WH704	99.8 (835/837)	Human	1995-99	0.00	0.00000
B. uniformis	dnlkv2	99.7 (1370/1374)	Mouse	2000 or later	0.52	0.00039
B. uniformis	CL03T00C23	99.6 (984/988)	Human	2000 or later	0.00	0.00000
B. uniformis	VPI-C20-25	99.6 (852/855)	Human	pre-1980	0.37	0.00016
B. uniformis	WH203	99.6 (844/847)	Human	1995-99	0.49	0.00010
B. uniformis	2_2_43B	99.6 (834/837)	Human	2000 or later	0.25	0.00019
B. uniformis	WH205	99.6 (817/820)	Human	1995-99	0.34	0.00025
B. uniformis	WH712	99.4 (846/851)	Human	1995-99	0.00	0.00000
B. uniformis	WH204	99.4 (834/839)	Human	1995-99	0.38	0.00023
B. uniformis	WH11	99.4 (798/803)	Human	1995-99	0.98	0.00075
B. uniformis	WH511	99.4 (786/791)	Human	1995-99	0.00	0.00000
B. uniformis	WH717	99.4 (785/790)	Human	1995-99	0.20	0.00009
B. uniformis	WH714	99 (836/844)	Human	1995-99	0.38	0.00022
B. uniformis	CL03T12C37	98.6 (931/944)	Human	2000 or later	0.00	0.00000
B. uniformis	WH17	96.7 (797/824)	Human	1995-99	0.00	0.00000
B. uniformis	WH15	96.7 (794/821)	Human	1995-99	0.00	0.00000
B. uniformis	WH16	96.6 (792/820)	Human	1995-99	0.00	0.00000
B. vulgatus	ATCC 8482	100 (type strain)	Human	pre-1980	0.00	0.00000

B. vulgatus	WH19	100 (813/813)	Human	1995-99	0.00	0.00000
B. vulgatus	WH14	100 (825/825)	Human	1995-99	0.00	0.00000
B. vulgatus	VPI-4025	100 (840/840)	Human	pre-1980	0.00	0.00000
B. vulgatus	WH13	100 (845/845)	Human	1995-99	0.00	0.00000
B. vulgatus	WH18	100 (850/850)	Human	1995-99	0.00	0.00000
B. vulgatus	RJ2H1	100 (965/965)	Mouse	2000 or later	0.00	0.00000
B. vulgatus	WH119	99.9 (850/851)	Human	1995-99	0.00	0.00000
B. vulgatus	PC510	99.9 (1417/1418)	Human	unknown	0.00	0.00000
B. vulgatus	RJ2L3	99.9 (1416/1418)	Mouse	2000 or later	0.00	0.00000
B. vulgatus	274_1D4	99.9 (1111/1112)	Human	unknown	0.00	0.00000
B. vulgatus	WH108	99.8 (849/851)	Human	1995-99	0.00	0.00000
B. vulgatus	WH6	99.8 (841/843)	Human	1995-99	0.00	0.00000
B. vulgatus	WH8	99.8 (840/842)	Human	1995-99	0.00	0.00000
B. vulgatus	VPI-BV8526	99.8 (834/836)	Human	pre-1980	0.00	0.00000
B. vulgatus	VPI-4496.2	99.8 (825/827)	Human	pre-1980	0.00	0.00000
B. vulgatus	VPI-5710	99.8 (804/806)	Human	pre-1980	0.00	0.00000
B. vulgatus	WH516	99.7 (789/791)	Human	1995-99	0.00	0.00000
B. vulgatus	CL09T03C04	99.6 (898/902)	Human	2000 or later	0.00	0.00000
B. vulgatus	WH9	99.6 (848/851)	Human	1995-99	0.00	0.00000
B. vulgatus	WH202	99.6 (847/850)	Human	1995-99	0.00	0.00000
B. vulgatus	VPI-C1-13	99.6 (836/839)	Human	pre-1980	0.00	0.00000
B. vulgatus	VPI-4245	99.6 (818/821)	Human	pre-1980	0.00	0.00000
B. vulgatus	dnlkv7	99.6 (1373/1379)	Mouse	2000 or later	0.25	0.00012
B. vulgatus	VPI-4506	99.5 (848/852)	Human	pre-1980	0.00	0.00000
B. vulgatus	WH109	99.5 (842/846)	Human	1995-99	0.00	0.00000
B. vulgatus	WH715	99.5 (825/829)	Human	1995-99	0.00	0.00000
B. vulgatus	WH515	99.5 (792/796)	Human	1995-99	0.12	0.00006
B. vulgatus	WH7	99.5 (786/790)	Human	1995-99	0.00	0.00000
B. vulgatus	4_3_47FAA	99.4 (840/845)	Human	2000 or later	0.00	0.00000
B. vulgatus	WH5	99.4 (803/808)	Human	1995-99	0.00	0.00000
B. vulgatus	WH716	99.3 (822/828)	Human	1995-99	0.00	0.00000
B. vulgatus	3_1_40A	98.9 (833/842)	Human	2000 or later	0.00	0.00000
B. xylanisolvens	XB1A, DSM 18836	100 (type strain)	Human	2000 or later	1.01	0.00229
B. xylanisolvens	WH301	99.9 (810/811)	Human	1995-99	0.00	0.00000
B. xylanisolvens	WH210	99.8 (849/851)	Human	1995-99	1.09	0.00128
B. xylanisolvens	WH212	99.8 (847/849)	Human	1995-99	1.18	0.00167
B. xylanisolvens	WH213	99.8 (847/849)	Human	1995-99	0.57	0.00096
B. xylanisolvens	VPI-Bov7991	99.8 (847/849)	Human	pre-1980	1.04	0.00165
B. xylanisolvens	WH209	99.8 (846/848)	Human	1995-99	1.17	0.00158
B. xylanisolvens	NLAE-zl-C29	99.7 (769/771)	Cow	2000 or later	1.06	0.00080
B. xylanisolvens	NLAE-zl-C178	99.7 (684/686)	Cow	2000 or later	0.84	0.00077
B. xylanisolvens	NLAE-zl-G310	99.6 (849/852)	Goat	2000 or later	0.68	0.00035
B. xylanisolvens	NLAE-zl-G421	99.6 (849/852)	Goat	2000 or later	0.57	0.00016
B. xylanisolvens	NLAE-zl-C182	99.6 (849/852)	Cow	2000 or later	0.77	0.00045
B. xylanisolvens	NLAE-zl-C339	99.6 (849/852)	Cow	2000 or later	0.84	0.00059
B. xylanisolvens	2_1_22	99.6 (848/851)	Human	2000 or later	1.02	0.00074
B. xylanisolvens	3_1_13	99.6 (848/851)	Human	2000 or later	1.07	0.00083
B. xylanisolvens	D1 (1_1_22)	99.6 (848/851)	Human	2000 or later	0.99	0.00071
B. xylanisolvens	NLAE-zl-P393	99.6 (846/849)	Pig	2000 or later	0.92	0.00111
B. xylanisolvens	NLAE-zl-P352	99.6 (846/849)	Pig	2000 or later	1.02	0.00130
B. xylanisolvens	NLAE-zl-P727	99.6 (846/849)	Pig	2000 or later	1.00	0.00119
B. xylanisolvens	NLAE-zl-P736	99.6 (846/849)	Pig	2000 or later	1.05	0.00135
B. xylanisolvens	WH305	99.6 (842/845)	Human	1995-99	0.99	0.00053
B. xylanisolvens	NLAE-zl-H465	99.6 (833/836)	Human	2000 or later	0.78	0.00031
B. xylanisolvens	CL03T12C04	99.6 (701/704)	Human	2000 or later	1.14	0.00186
B. xylanisolvens	NLAE-zl-P218	99.6 (553/555)	Pig	2000 or later	1.07	0.00140
B. xylanisolvens	NLAE-zl-P349	99.6 (552/554)	Pig	2000 or later	1.04	0.00132
B. xylanisolvens	NLAE-zl-P225	99.6 (552/554)	Pig	2000 or later	1.06	0.00137

B. xylanisolvens	NLAE-zl-G39	99.5 (848/852)	Goat	2000 or later	0.69	0.00041
B. xylanisolvens	WH302	99.5 (847/851)	Human	1995-99	0.93	0.00057
B. xylanisolvens	D22 (1_2_8)	99.5 (847/851)	Human	2000 or later	1.15	0.00115
B. xylanisolvens	NLAE-zl-H194	99.5 (846/850)	Human	2000 or later	0.89	0.00048
B. xylanisolvens	NLAE-zl-P732	99.5 (845/849)	Pig	2000 or later	1.03	0.00136
B. xylanisolvens	WH307	99.5 (789/793)	Human	1995-99	0.75	0.00023
B. xylanisolvens	WH404	99.4 (795/800)	Human	1995-99	1.08	0.00146
B. xylanisolvens	NLAE-zl-G37	99.3 (842/848)	Goat	2000 or later	0.61	0.00025
B. xylanisolvens	1_1_30	99.2 (763/769)	Human	2000 or later	1.21	0.00180
B. xylanisolvens	NLAE-zl-G44	99.1 (846/854)	Goat	2000 or later	0.76	0.00042
B. xylanisolvens	NLAE-zl-H40	99.1 (841/849)	Human	2000 or later	0.84	0.00031
B. xylanisolvens	WH304	99 (788/796)	Human	1995-99	0.82	0.00027
B. xylanisolvens	NLAE-zl-G346	98.7 (844/855)	Goat	2000 or later	0.31	0.00020
B. xylanisolvens	NLAE-zl-G109	98 (837/854)	Goat	2000 or later	0.95	0.00068
B. xylanisolvens	2_2_4	97.7 (814/833)	Human	2000 or later	1.09	0.00153
B. xylanisolvens	NLAE-zl-C257	99.5 (997/1002)	Cow	2000 or later	0.74	0.00047
B. xylanisolvens	NLAE-zl-C233	99.6 (1282/1287)	Cow	2000 or later	0.59	0.00042
B. xylanisolvens	NLAE-zl-C315	99.6 (1259/1264)	Cow	2000 or later	0.84	0.00054
B. xylanisolvens	NLAE-zl-G406	98.8 (1355/1372)	Goat	2000 or later	0.90	0.00086
D. gadei	ATCC BAA-286	100 (type strain)	Human	2000 or later	0.00	0.00000
D. mossii	DSM 22836	100 (type strain)	Human	2000 or later	0.68	0.00044
O. splanchnicus	DSM 20712, ATCC 29572	100 (type strain)	Human	2000 or later	0.00	0.00000
P. distasonis	ATCC 8503	100 (type strain)	Human	pre-1980	0.82	0.00049
P. distasonis	VPI-4243	99.9 (744/745)	Human	pre-1980	0.83	0.00060
P. distasonis	3_1_19	99.6 (923/927)	Human	2000 or later	1.28	0.00095
P. distasonis	VPI-BD6781	99.5 (862/866)	Human	pre-1980	1.26	0.00130
P. distasonis	VPI-C14-2	99.3 (859/865)	Human	pre-1980	1.32	0.00098
P. distasonis	VPI-BD6803	99.2 (858/865)	Human	pre-1980	1.16	0.00113
P. distasonis	VPI-C18-7	99.2 (841/849)	Human	pre-1980	1.34	0.00079
P. distasonis	VPI-56A-56	99.1 (850/858)	Human	pre-1980	1.28	0.00128
P. distasonis	2_1_33B	99 (873/882)	Human	2000 or later	1.28	0.00098
P. distasonis	VPI-T3-25	98.8 (837/847	Human	pre-1980	1.14	0.00091
P. distasonis	VPI-C19-17	98.7 (830/841)	Human	pre-1980	1.36	0.00097
P. distasonis	VPI-C30-45	98.6 (825/837)	Human	pre-1980	1.31	0.00090
P. distasonis	VPI-B1-20	98.5 (834/847)	Human	pre-1980	0.15	0.00008
P. distasonis	WAL8975	98.2 (834/849)	Human	unknown	0.97	0.00053
P. distasonis	WAL9063	98.1 (833/849)	Human	unknown	0.19	0.00003
P. distasonis	WH517	98.1 (807/823)	Human	1995-99	0.00	0.00000
P. goldsteinii	DSM 19448, WAL 12034	100 (type strain)	Human	2000 or later	0.00	0.00000
P. goldsteinii	dnlkv18	99.1 (1374/1382)	Mouse	2000 or later	0.00	0.00000
P. gordonii	DSM 23371	100 (type strain)	Human	2000 or later	0.00	0.00000
P. johnsonii	DSM 18315	100 (type strain)	Human	2000 or later	1.14	0.00073
P. merdae	ATCC 43184	100 (type strain)	Human	2000 or later	1.40	0.00140
P. merdae	T4-1	99.6 (1019/1023)	Human	unknown	1.22	0.00089
P. merdae	VPI-BD6944	96.7 (841/870)	Human	pre-1980	1.32	0.00098

Footnotes: [1]Strains are classified into species based on >97% 16S rDNA sequence identity to the type strain of each species. Type strains for each species are underlined and listed first in each group.

Enzyme assays

Recombinant proteins purified in *E. coli*, were used to determine enzyme kinetics for RusGH, RusNH, RusK1, and RusK2. For RusNH we used a *p*-nitrophenol-ribofuranoside substrate with absorbance readings at 405nm over a 24-hour period as described previously[26], with modifications for using purified protein instead of crude extract, using 0.5mM of enzyme in

a buffer containing 20mM Hepes and 100mM NaCl, at pH 6.7 at 37°C and continuous absorbance readings. For RusGH, a panel of other 4-nitrophenol based substrates in addition to *p*-NP-ribofuranoside were tested at pH 9.0 in 100 mM Tris at 37°C for 16h with 1.5-15 μM of enzyme and using endpoint absorbance measurements. Ion requirements of the RusGH were assayed in *p*-NP-ribofuranoside by addition of divalent cations in the form of $CaCl_2$, $ZnCl_2$, or $MgCl_2$, at 2, 5, or 10 mM concentrations, or in the presence of 10 mM EDTA. Specificity and kinetic parameters for RusNH on natural nucleoside substrates were determined as described previously using a UV-based assay[43]. Briefly, a 96-well, UV-compatible microplate (Santa Cruz Biotechnologies) was used with substrate concentrations ranging from 0.025mM-2.5mM, and enzyme concentrations of 0.25-1uM. Assays were immediately read after addition of enzyme by continuous reading of absorbance at 262nm or 280nm with time points taken every 2.5 minutes over 12-24 hours at 37°C. Volume was 250uL in all assays and carried out in buffer containing 20mM Hepes and 100mM NaCl, at pH 6.7, adjusted with acetic acid. As a measure of catalytic efficiency, (K_{cat}/K_M) was unable to be determined by classical Michaels-Menton kinetics as Vmax was never reached and therefore Km values were not accurate, so we used a previously established method of estimating this value[13]. Briefly, we used a single substrate concentration to calculate (k_{cat}/K_M) and checked to be $<<K_M$ by halving and doubling the substrate concentration and observing a proportionate increase or decrease in rate. Therefore the equation, $V_0 = (k_{cat}/K_M)[S][E]$ was used to calculate k_{cat}/K_M in our case. For, RusGH a panel of other 4-nitrophenol based substrates in addition to *p*-NP-ribofuranoside were tested at pH 9.0 in 100 mM Tris at 37°C for 16h with 1.5-15 μM of enzyme with endpoint absorbance measurements. Ion requirements of the RusGH were assayed in *p*-NP-ribofuranoside by addition of divalent cations in the form of $CaCl_2$, $ZnCl_2$, or $MgCl_2$, at 2, 5, or 10 mM concentrations, or in the presence of 1 mM EDTA. The RusGH was tested against a panel of oligosaccharides, nucleosides and nucleotides. Briefly, the reactions were performed with 10 μM of RusGH, 8mg/ml substrate or 5mM monosaccharide in 50 mM TRIS pH 9.0 at 37 °C for 16h. A control reaction was performed in the same conditions without enzyme. The activity was qualitative determined by thin layer chromatography. 6 μl of the reaction was spotted on foil backed silica plate (Silicagel 60, 20 x 20, Merck) and develop in butanol:acetic acid:water 2:1:1 (mobile phase). The products of the reaction were detected by immersing the TLC plate in developer (sulphuric

acid/ethanol/water 3:70:20 v/v, orcinol 1 %) for 30 seconds and heating to 100 °C for 2 minutes. A standard of ribose was run in all TLC plates. For RusK1/K2 a phosphatase-coupled, universal kinase assay was used according to manufacturer instructions to determine a specific activity of the kinases on pentose sugars. (R&D Systems, Minneapolis, MN)[44]. Specifically, all reactions were carried out in buffer containing 70 mM Tris, 100 mM KCl, and 5 mM $MgCl_2$ at pH 7.5, this buffer is based on previous studies examining ribokinase activities and showing this buffer provided maximal enzymatic activity[45]. Reactions were carried out in 50 µL at 37°C for 30 minutes. All reactions contained 1 mM ATP, 100 ng of coupling phosphatase, and a range of enzyme concentrations ranging from 0.1-10 µM of RusK1, RusK2, or *E. coli* RbsK (MyBioSource, San Diego, CA), as a positive control, and for the acceptor substrate either 10 mM of ribose or deoxyribose or 200 mM of all other sugars tested including: arabinose, xylose, glucose or fructose etc. Determination of specific activity was based off of a coupling rate of 0.399 and a rate constant of 97.78 nmol/min/µg/mM (empirically determined by the kit manufacture). In brief, our specific activity is based on an endpoint observation across a minimum of 5 enzyme concentrations, the resulting absorbance is fit to a known phosphate standard curve equation and the resulting rate is nM (product formed)/min (held standard at 30 min)/uM enzyme/mM of substrate. This is the specific activity at a defined endpoint and so should not be confused with a rate taken at several enzyme concentrations over different time points, but rather used a crude measurement for which to compare different enzymes.

Determination of free and acid hydrolysable monosaccharide content in diets and cecal contents using GC/MS

Prior to analysis, diets were ground to a fine powder using a blender followed by mortar and pestle, while cecal contents were dried by lyophilization. Samples were analyzed for free and linked monosaccharides using the following method described[46]. In brief, all reactions began with 1-3mg of sample and samples were hydrolyzed in 100ul of 2.5 M TFA for 90 min at 121 °C. Samples were allowed to cool to room temperature (RT) and myo-inositol was added as an internal standard (20ul of 2.5mg/ml) and dried under nitrogen. 150ul of methanol was added, dried and repeated once more. Dried samples were then reduced by dissolving in 50ul of 2M NH_4OH followed by addition of 50 ul of freshly made 1M $NaDB_4$ in 2M NaOH. This mixture was sonicated in a water bath for 1 min, followed by incubation at room temperature for 2.5

hours. 23ul of glacial acetic acid was added and samples dried and evaporated 2x with 250ul of 5% (v/v) acetic acid in methanol, followed by 2x evaporation with 250ul of methanol, drying after each step. Acetylation was done by addition of 250ul acetic anhydrate and sonicated 5 min followed by incubation at 100 °C for 2.5 hours. 2ml of ddH_2O was added and sample vortexed to dissolve residue, followed by room temperature incubation for 10 min. 1ml of dichloromethane (DCM) was added and vortexed followed by centrifugation at 2000 rpm for 2.5 min. The aqueous phase was discarded and the DCM phase washed 2x with 2 ml of ddH_2O. DCM phase was dried and residue dissolved in 250 ul acetone. For free monosaccharide analysis the initial hydrolysis step with TFA was not performed. To establish a limit of detection in cecal contents, varying amounts of ribose (0.00002-0.2 mg, in 10-fold increments) were added at the same time as the myo-inositol standard to establish percent recovery throughout the methods used. Acetylated samples were analyzed on a gas chromatography (Agilent Technologies model 7890A) coupled mass spectrometer (Agilent Technologies model 5975C) using a fused silica capillary column (60m x 0.25 mm x 0.2µm SP-2330, Supelco Analytical).

LC/MS/MS Determination of positional ribose phosphorylation by rus ribokinases

Samples were prepared as follows with reactions containing the following: 1µM of either enzyme (RusK1 or RusK2), 10mM of a starting substrate (ribose, ribose 1-phosphate, or ribose 5-phosphate), 1mM ATP, with all components dissolved in a buffer containing 70 mM Tris, 100 mM KCl, and 5 mM $MgCl_2$ at pH 7.5and incubated at 37°C for 30 minutes. Reactions were then flash frozen and stored at -80°C until processing. For analysis, 100% MeOH was added to thawed samples in buffer at a 4:1 ratio to extract metabolites. Samples were then dried down and reconstituted in 45 µL of 1:1 MeOH/H_2O. Samples were run on a 6470 Series Agilent Technologies Triple Quadrupole Mass Spectrometer with Ion-Pairing chromatography. The acquisition method was programed to detect for dynamic multiple reaction monitoring (dMRM) of four compounds of interest: ribose, ribose 1-phosphate, ribose 5-phosphate, and ribose 1,5-bisphosphate. The dMRM scan is used with a 0.07 min peak width and acquisition time of 24 min. The detected fragments displayed the following dMRM transitions: ribose 149->89 at 1.31 min with collision energy (CE) of 5 eV; ribose-1-phosphate 229->210 at 9.4 min with CE of 9eV; ribose-5-phosphate 229-> 97 at 7.9 min with CE of 13eV; ribose-1,5-bisphosphate 309->211 at 14.6 min with CE of 13 eV. The following parameters were incorporated into the

method: delta retention time of plus and minus 1 min, fragmentor of 40 eV and cell accelerator of 5 eV. Agilent Qualitative Analysis version 7.00 was used for post-acquisition analysis. Our empirically determined range of detection was established above a noise baseline determined by running enzyme, buffer, sample, and internal controls for each species of interest where we did not anticipate these species being detected. This was determined to be 10^3 which was our highest background reading (Table 2.2E). Detailed instrumentation running parameters are here described. The following solvents were used during processing, Solvent A: 97% H_2O and 3% MeOH, Solvent B: 15 mM acetic acid and 10 mM tributylamine at pH 5. Solvent C: 15 mM acetic acid and 10 mM tributylamine in MeOH. Washing Solvent D is 100% acetonitrile. LC system seal washing solvent is 90% water and 10% isopropanol, while the needle washing solvent is 75% methanol and 25% water. The Agilent Technologies Triple Quad 6470 LC/MS system used here consists of 1290 Infinity II LC Flexible Pump (Quaternary Pump), 1290 Infinity II Multisampler, 1290 Infinity II Multicolumn Thermostat with 6 port valve and 6470 triple quadrupole mass spectrometer. Agilent Masshunter Workstation Software LC/MS Data Acquisition for 6400 Series Triple Quadrupole MS with Version B.08.02 is used for compound optimization and data acquisition. The following column was used for separation: Agilent ZORBAX RRHD Extend-C18, 2.1 × 150 mm, 1.8 um and ZORBAX Extend Fast Guards for UHPLC are used in the separation. LC gradient profile is: at 0.25 ml/min, 0-2.5 min, 100% A; 7.5 min, 80% A and 20% C; 13 min 55% A and 45% C; 20 min, 1% A and 99% C; 24 min, 1% A and 99% C; 24.05 min, 1% A and 99% D; 27 min, 1% A and 99% D; at 0.8 ml/min, 27.5-31.35 min, 1% A and 99% D; at 0.6 ml/min, 31.50 min, 1% A and 99% D; at 0.4 ml/min, 32.25-39.9 min, 100% A; at 0.25 ml/min, 40 min, 100% A. Column temperature is kept at 35℃, samples at 4℃, and injection volume is 2 µl. The 6470 Triple Quad MS was calibrated with ESI-L Low concentration Tuning mix. Source parameters: Gas temp 150℃, Gas flow 10 l/min, Nebulizer 45 psi, Sheath gas temp 325℃, Sheath gas flow 12 l/min, Capillary -2000 V, Delta EMV -200 V.

Quantification and Statistical Analysis

Student's t-tests for *in vivo* data were performed for each time point in GraphPad Prism version 8.1 with a paired, two-tailed distribution. Detailed statistical information is included in the figure legends where appropriate.

Table 2.5 Strains, vectors, and primers used in this study

Strain	Genotype	Features	Reference
Bacteroides thetaiotaomicron (*B. theta*) *tdk*$^-$	ATCC 29148 *tdk*$^-$	Parent strain of all deletion strains, and refered to in text as "wild-type"	Koropatkin et al. 2008
B. theta Δ*rus*	ATCC 29148 *tdk*$^-$, *BT2802-2809*$^-$	*rus* PUL deletion strain, unable to grow on any ribose containing substrate and unable to activate rus transcript	This study
B. theta Δ*rusR*	ATCC 29148 *tdk*$^-$, *BT2802*$^-$	Putative upstream regulator of *rus* PUL deletion strain, unable to grow on any ribose containing substrate and unable to upregulate rus transcript	This study
B. theta Δ*rusK1*	ATCC 29148 *tdk*$^-$, *BT2803*$^-$	First ribokinase deletion of the *rus* PUL, exhibits delayed growth and transcript activation on ribose, with absent growth on nucleosides/RNA	This study
B. theta Δ*rusK2*	ATCC 29148 *tdk*$^-$, *BT2804*$^-$	Second ribokinase deletion of the *rus* PUL, no growth on any ribose substrates. Normal *rus* transcript activation dynamics	This study
B. theta Δ*rusK1/K2*	ATCC 29148 *tdk*$^-$, *BT2803-2804*$^-$	Both ribokinases deleted from the *rus* PUL, no growth on any ribose substrates, no *rus* transcript activation	This study
B. theta Δ*rusC*	ATCC 29148 *tdk*$^-$, *BT2805*$^-$	Deletion of the *susC* homolog within the *rus* PUL. Wild-type levels of growth on all substrates and normal transcript activation.	This study
B. theta Δ*rusD*	ATCC 29148 *tdk*$^-$, *BT2806*$^-$	Deletion of the *susD* homolog within the *rus* PUL. Wild-type levels of growth on all substrates and normal transcript activation.	This study
B. theta Δ*rusC/D*	ATCC 29148 *tdk*$^-$, *BT2805-2806*$^-$	Deletion of both *susC* and *susD* homologs within the *rus* PUL. Wild-type levels of growth on all substrates.	This study
B. theta Δ*rusGH*	ATCC 29148 *tdk*$^-$, *BT2807*$^-$	Deletion of a predicted glycoside hydrolase of unknown function or family within the *rus* PUL. Wild-type levels of growth on all substrates.	This study
B. theta Δ*rusNH*	ATCC 29148 *tdk*$^-$, *BT2808*$^-$	Deletion of a confirmed nucleoside hydrolase of IUNH family within the *rus* PUL. Wild-type levels of growth on all substrates.	This study
B. theta Δ*rusGH/NH*	ATCC 29148 *tdk*$^-$, *BT2807-2808*$^-$	Deletion of predicted glycoside hydrolase and nucleoside hydrolase of *rus* PUL. Wild-type levels of growth on all substrates.	This study

B. theta Δ*rusT*	ATCC 29148 *tdk*$^-$, *BT2809*$^-$	Deletion of putative *rus*-encoded ribose-specific permease. Exhibits both delayed growth and lower overall growth on all ribose substrates, as well as a decrease in *rus* transcript activation levels	This study
B. theta Δ*BT0184*	ATCC 29148 *tdk*$^-$, *BT0184*$^-$	Deletion of predicted uridine or nucleoside kinase. No growth defects on ribose/nucleosides/RNA observed	This study
B. theta Δ*BT1881*	ATCC 29148 *tdk*$^-$, *BT1881*$^-$	Deletion of predicted purine nucleoside phosphorylase. Growth defects on some nucleosides, but not on ribose or RNA	This study
B. theta Δ*BT4330*	ATCC 29148 *tdk*$^-$, *BT4330*$^-$	Deletion of predicted nucleoside permease. Growth defects on nucleosides, normal growth on ribose and RNA	This study
B. theta Δ*BT4554*	ATCC 29148 *tdk*$^-$, *BT4554*$^-$	Deletion of predicted purine nucleoside phosphorylase. Growth defects on some nucleosides, but not on ribose or RNA	This study
B. theta Δ*rusR :: BT2802*	ATCC 29148 *tdk*$^-$, *BT2802*$^-$ :: pNBU2-bla-ermG *BT2802*	Trans-complementation of the Δ*rusR* (*BT2802*) deletion strain with native promoter, restores growth on ribose	This study
B. theta Δ*rusK1/K2 :: BT2803-04*	ATCC 29148 *tdk*$^-$, *BT2803-04*$^-$ :: pNBU2-bla-ermG *BT2803-04*	Trans-complementation of the Δ*rusK1/K2* (*BT2803-04*) deletion strain with native promoter restoring growth on ribose, also used for *in vivo* studies	This study
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 tdk$^-$ Δ*rus* +Tag14	ATCC 29148 *tdk*$^-$, *BT2802-2809*$^-$ ::pNBU2-bla-tetQb Tag14	Δ*rus* strain with unique barcode inserted used for in vivo studies	This study
B. theta tdk$^-$ Δ*rusR* +Tag14	ATCC 29148 *tdk*$^-$, *BT2802*$^-$ ::pNBU2-bla-tetQb Tag14	Δ*rusR* strain with unique barcode inserted used for *in vivo* studies	This study
B. theta tdk$^-$ Δ*rusK1* +Tag14	ATCC 29148 *tdk*$^-$, *BT2803*$^-$ ::pNBU2-bla-tetQb Tag14	Δ*rusK1* strain with unique barcode inserted used for *in vivo* studies	This study
B. theta tdk$^-$ Δ*rusK2* +Tag14	ATCC 29148 *tdk*$^-$, *BT2804*$^-$ ::pNBU2-bla-tetQb Tag14	Δ*rusK2* strain with unique barcode inserted used for *in vivo* studies	This study
B. theta tdk$^-$ Δ*rusK1/K2* +Tag14	ATCC 29148 *tdk*$^-$, *BT2803-2804*$^-$ ::pNBU2-bla-tetQb Tag14	Δ*rusK1/K2* strain with unique barcode inserted used for *in vivo* studies	This study
B. theta tdk$^-$ Δ*rusC/D* +Tag14	ATCC 29148 *tdk*$^-$, *BT2805-2806*$^-$ ::pNBU2-bla-tetQb Tag3	Δ*rusC/D* strain with unique barcode inserted used for *in vivo* studies	This study
B. theta tdk$^-$ Δ*rusGH/NH* +Tag14	ATCC 29148 *tdk*$^-$, *BT2807-2808*$^-$ ::pNBU2-bla-tetQb Tag9	Δ*rusGH/NH* strain with unique barcode inserted used for *in vivo* studies	This study

B. theta tdk^{-} Δ*rusT* +Tag14	ATCC 29148 *tdk*$^{-}$, *BT2809*$^{-}$ ::pNBU2-bla-tetQb Tag14	Δ*rusT* strain with unique barcode inserted used for *in vivo* studies	This study
B. theta tdk^{-} Δ*BT4554* +Tag14	ATCC 29148 *tdk*$^{-}$, *BT4554*$^{-}$ ::pNBU2-bla-tetQb Tag14	Δ*BT4554* strain with unique barcode inserted used for *in vivo* studies	This study
E. coli BL21-AI™	*B. coli* B strain, F^{-}, *ompT, hsdSB* (r_B'm_B'), *gal, dcm, araB::T7RNAP*$^{-}$, *tetA*	Parent *E. coli* expression strain used to construct deletions in for downstream protein purification of *B. theta* proteins.	Invitrogen
E. coli BL21-AI™ Δ*rihABC*	*B. coli* B strain, F^{-}, *ompT, hsdSB* (r_B'm_B'), *gal, dcm, araB::T7RNAP*$^{-}$, *tetA, rihA*$^{-}$, *rihB*$^{-}$, *rihC*$^{-}$,	*E. coli* strain lacking chromosomal ribose-inducible hydrolases (*rih*) genes ised in the production of the *B. theta* nucleoside hydrolase, BT2808 (RusNH), to avoid contaminating nucleoside hydrolase activity from *E. coli*	This study
Primer	**Sequence (5' to 3')**	**Use**	
Genetic Manipulation Primers			
Ribose PUL Knockout *B. thetaiotaomicron* VPI5482			
Restriction Sites are Underlined			
BT2802-2809 5'Up Xbal	GCGTCTAGACGGCTCCATAAAGGTTATC	BT2802-2809 Ribose PUL Knockout	
BT2802-2809 3'Out	GTTTTCTGTAGCTCTTTGTTGCG	BT2802-2809 Ribose PUL Knockout	
BT2802-2809 5'Out	CGCAACAAAGAGCTACAGAAAACGGGGTGA AATTCAATTCTATGATT	BT2802-2809 Ribose PUL Knockout	
BT2802-2809 3'Down Sal1	GCGGTCGACGCTGTTGTGTTCAATGATCTG	BT2802-2809 Ribose PUL Knockout	
BT2802 5'Up Xbal	GCGTCTAGACGGCTCCATAAAGGTTATC	BT2802 Gene Knockout	
BT2802 3'Out	AGGGAACTCTTTGCATTAGTA	BT2802 Gene Knockout	
BT2802 5'Out	TACTAATGCAAAGAGTTCCCT AGAGTAAGGTGTTTGATTCGT	BT2802 Gene Knockout	
BT2802 3'Down Sal1	GCGGTCGACCTGGACGCGGGAGCCGGATTG	BT2802 Gene Knockout	
BT2803 5'Up Sal1	GCGGTCGACCTGGATAAAGGAGATTTTTCG	BT2803 Gene Knockout	
BT2803 3'Out	GTTTAACCTAGGTCTTATCTCGTGCATAGTTT TCTACATTAATAATTGG	BT2803 Gene Knockout	
BT2803 5'Out	ACGAGATAAGACCTAGGTTAAAC	BT2803 Gene Knockout	
BT2803 3'Down Xbal	GCGTCTAGAGTATCAATCGCATTTACTTTGTA TCCG	BT2803 Gene Knockout	
BT2804 5'Up Sal1	GCGGTCGACCGGCCTGCAAGCATTGGA	BT2804 Gene Knockout	
BT2804 3'Out	GTGTCGATTAGTTTTCATAATCCATGAAATTT GAACAGATTTATGTGTTTAAC	BT2804 Gene Knockout	
BT2804 5'Out	ATTTCATGGATTATGAAAACTAATCGACAC	BT2804 Gene Knockout	
BT2804 3'Down Xbal	GCGTCTAGATACTAAGATTACTCCGAATGGC	BT2804 Gene Knockout	
BT2803-2804 5'Up Sal1	GCGGTCGACCTGGATAAAGGAGATTTTTCG	BT2803-2804 Genes Knockout	

BT2803-2804 3'Out	GTGTCGATTAGTTTTCATAATCCATGAAATGC ATAGTTTTCTACATTAATAATTGG	BT2803-2804 Genes Knockout	
BT2803-2804 5'Out	ATTTCATGGATTATGAAAACTAATCGACAC	BT2803-2804 Genes Knockout	
BT2803-2804 3'Down Xbal	GCGTCTAGATACTAAGATTACTCCGAATGGC	BT2803-2804 Genes Knockout	
BT2805 5'Up Sal1	GCGGTCGACGAAGGTATCACTACTGAT	BT2805 Gene Knockout	
BT2805 3'Out	GTGTCGATTAGTTTTCATAATCCATGAAATGC ATAGTTTTCTACATTAATAATTGG	BT2805 Gene Knockout	
BT2805 5'Out	ATTTCATGGATTATGAAAACTAATCGACAC	BT2805 Gene Knockout	
BT2805 3'Down Xbal	GCGTCTAGATACTAAGATTACTCCGAATGGC	BT2805 Gene Knockout	
BT2806 5'Up Sal1	GCGGTCGACCGTGGGTGAAATATGGG	BT2806 Gene Knockout	
BT2806 3'Out	TAGATTATTGGATTATTGCGGGGTGGTTGAT ATCGTAAGTCTTTTCATAATATCCTG	BT2806 Gene Knockout	
BT2806 5'Out	ACCCCGCAATAATCCAATAATCTA	BT2806 Gene Knockout	
BT2806 3'Down Xbal	GCGTCTAGAGCTTAGAGGGGGCCCATCCCG	BT2806 Gene Knockout	
BT2807 5'Up Sal1	GCGGTCGAC GCCTCTTCTGTTACAGAAAAGTAC	BT2807 Gene Knockout	
BT2807 3'Out	TGAACATATCTTGGGTTTTTATAAGTTAATTG GTCTTTGTTAAATTAGTAG	BT2807 Gene Knockout	
BT2807 5'Out	AACTTATAAAAACCCAAGATATGTTCA	BT2807 Gene Knockout	
BT2807 3'Down Xbal	GCGTCTAGAATATGATACGCACAGAGGATG	BT2807 Gene Knockout	
BT2808 5'Up Sal1	GCGGTCGACGCTCGCCCCAAAGATGTATTC	BT2808 Gene Knockout	
BT2808 3'Out	AAGATTTCAATTCTCTTTTCCTGTGATTGCTA TATATTTGAACATATC	BT2808 Gene Knockout	
BT2808 5'Out	ACAGGAAAAGAGAATTGAAATCTT	BT2808 Gene Knockout	
BT2808 3'Down Xbal	GCGTCTAGAGTGTGTACTTGCTTTTCCGGC	BT2808 Gene Knockout	
BT2809 5'Up Sal1	GCGGTCGACGGGATGAATACAGACGACGGA	BT2809 Gene Knockout	
BT2809 3'Out	ATTTTAGCTTCCTCCTGAAGCTAAGGTGTAAC TATTGACTATATACAT	BT2809 Gene Knockout	
BT2809 5'Out	TCAGGAGGAAGCTAAAAT	BT2809 Gene Knockout	
BT2809 3'Down Xbal	GCGTCTAGACGGTGACTGTGTAGACGACCC	BT2809 Gene Knockout	
BT0184 5'Up Xbal	GCGTCTAGACAGCTTCACTTTCTGCACGAATT T	BT0184 Gene Knockout	
BT0184 3'Out	TGGCACTCATTCTGAACTCTTTATCTTCTCTTT ATAGTGTTATTTGAACCAAATAGA	BT0184 Gene Knockout	
BT0184 5'Out	AAGATAAAGAGTTCAGAATGAGTGCCA	BT0184 Gene Knockout	
BT0184 3'Down Sal1	GCGGTCGACTTTCGTACAGCCCATGAAGAAC GC	BT0184 Gene Knockout	
BT1881 5'Up Sal1	GCGGTCGACCTATCCATCAAAGAACCGTCC	BT1881 Gene Knockout	
BT1881 3'Out	GTTCGCATAAATAGAGTTTAGTAGTGCCAAG GATGATGGCTGTT	BT1881 Gene Knockout	
BT1881 5'Out	TACTAAACTCTATTTATGCGAAC	BT1881 Gene Knockout	
BT1881 3'Down Xbal	GCGTCTAGACAGATATGCATTAATTCGGAC	BT1881 Gene Knockout	

BT4330 5'Up Sal1	GCGGTCGACGCGATGCGCGTCACGAAGCTTTCG	BT4330 Gene Knockout	
BT4330 3'Out	TACTTTATTCTCTCGTTGGTGGACTTTTCTAATTTGTGTTTCTCT	BT4330 Gene Knockout	
BT4330 5'Out	CACCAACGAGAGAATAAAGTA	BT4330 Gene Knockout	
BT4330 3'Down Xbal	GCGTCTAGATAGAAATGTCCCTTACCGTCCGTA	BT4330 Gene Knockout	
BT4554 5'Up Sal1	GCGGTCGACTCCTCCTGTTACTGCAGTAGC	BT4554 Gene Knockout	
BT4554 3'Out	CAAGCCATTTTGTTATGATGGATATATTATCCTATATTCTGTCGGTTATTGAACGGTTTTAGTTTTACAATAAG	BT4554 Gene Knockout	
BT4554 5'Out	CGACAGAATATAGGATAATATATCCATCATAACAAAATGGCTTG	BT4554 Gene Knockout	
BT4554 3'Down Xbal	GCGTCTAGAGATAGTCCTCATCGGAGCTAG	BT4554 Gene Knockout	
cBT2802 5'Up Xba1	GCGTCTAGATTCATCTTTATTATTAAAAAGTACATCGAAGG	BT2802 Gene Complement	
cBT2802 3'Down Sal1	GCGGTCGACACGAATCAAACACCTTACTCT	BT2802 Gene Complement	
cBT2803-04 5'Up Sal1	GCGGTCGACAGAGTAAGGTGTTTGATTCGT	BT2803-04 Gene Complement	
cBT2803-04 3'Down Xbal	GCGTCTAGAGTGTCGATTAGTTTTCATAATCCATGAAAT	BT2803-04 Gene Complement	
NBU2 att1 F	CCTTTGCACCGCTTTCAACG	pNBU2 insertion site determinaton	
NBU2 att1 R	TCAACTAAACATGAGATACTAGC	pNBU2 insertion site determinaton	
NBU2 att2 F	TATCCTATTCTTTAGAGCGCAC	pNBU2 insertion site determinaton	
NBU2 att2 R	GGTGTACCTGGCATTGAAGG	pNBU2 insertion site determinaton	
***Eschericia coli* BL21-AI™ *rihABC* Knockout Genes**			
rihA-Spectinomycin Resistance Forward	cacccaggtaacgcgtggtgatcttaatcaatgacgtgtgagggcttattatgcacgctt	Amplification of SpecR Gene from *E. coli* K11497 and Lamda Red Recombineering into BL21-AI Cells	
rihA-Spectinomycin Resistance Reverse	gcgttacccttgtcgcaaagaagcacacaacaaggagcaacaccgtggaaacggatgaaggc	Amplification of SpecR Gene from *E. coli* K11497 and Lamda Red Recombineering into BL21-AI Cells	
rihB-Hygromycin Resistance Forward	gctgtagaaaaaataacgcaactggaaacagaggaaataaacacaaatatgtatccgctcatg	Amplification of HygroR Gene from *E. coli* K11521 and Lamda Red Recombineering into BL21-AI Cells	
rihB-Hygromycin Resistance Reverse	gcttggatcttgcgatactgacccggcttgttatcaacacgggattttggtcatgagatt	Amplification of HygroR Gene from *E. coli* K11521 and Lamda Red Recombineering into BL21-AI Cells	
rihC-Gentamycin Resistance Forward	gcgaaaatgccggtcttgttaccggcatttttatggagaaaacttgtgtaggctggagctgct	Amplification of GentR Gene from *E. coli* K11590 and Lamda Red Recombineering into BL21-AI Cells	
rihC-Gentamycin Resistance Reverse	catgagtcgatgaatgactgcatgccgataacatgtgaataacaggatatgaatatcctccttagttc	Amplification of SpecR Gene from *E. coli* K11590 and Lamda Red Recombineering into BL21-AI Cells	

qPCR Primers	**Primer**	**Use**	
BT2801 F	taaaccgacggctctccatctg	*B. thetaiotaomicron* Ribose PUL Flanking Gene Expression	
BT2801 R	gccgccgaataatcccactt	*B. thetaiotaomicron* Ribose PUL Flanking Gene Expression	
BT2802 F	tacagagctgccttaaattcatacaaa	*B. thetaiotaomicron* Ribose PUL Expression	
BT2802 R	gctcacagaccgcaggctacc	*B. thetaiotaomicron* Ribose PUL Expression	
BT2803 F	caggtgccggagatgtattttg	*B. thetaiotaomicron* Ribose PUL Expression	
BT2803 R	accgattcgcgtgactgctat	*B. thetaiotaomicron* Ribose PUL Expression	
BT2804 F	ttctgtggtgcattggctgtaa	*B. thetaiotaomicron* Ribose PUL Expression	
BT2804 R	tcgagagtaggaatagacggttgg	*B. thetaiotaomicron* Ribose PUL Expression	
BT2805 F	tccacgccccgatataatgtagg	*B. thetaiotaomicron* Ribose PUL Expression	
BT2805 R	accgtttgcaccccagaagtagtaa	*B. thetaiotaomicron* Ribose PUL Expression	
BT2806 F	taaagcggcacaaatcatagcaga	*B. thetaiotaomicron* Ribose PUL Expression	
BT2806 R	tgtgttgtagcgcctccataaaag	*B. thetaiotaomicron* Ribose PUL Expression	
BT2807 F	tatgcgctggttgccgagaa	*B. thetaiotaomicron* Ribose PUL Expression	
BT2807 R	tgccgccaagccttttatgag	*B. thetaiotaomicron* Ribose PUL Expression	
BT2808 F	caggaaatgccatatgacagagaaa	*B. thetaiotaomicron* Ribose PUL Expression	
BT2808 R	taagtgcgttgcgggttgc	*B. thetaiotaomicron* Ribose PUL Expression	
BT2809 F	tattctcctttccgcctcagtatcc	*B. thetaiotaomicron* Ribose PUL Expression	
BT2809 R	tgttattggttcccgcctttg	*B. thetaiotaomicron* Ribose PUL Expression	
BT2810 F	tgcagcccggtcaaaagtattatta	*B. thetaiotaomicron* Ribose PUL Expression	
BT2810 R	cacaaagcccggaaggtatgg	*B. thetaiotaomicron* Ribose PUL Expression	
BT0348 F	tgcggcaaccaaattcaacaaa	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0348 R	accaagtgccccattcgtcaag	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0349 F	tccgccaagccagtgaagaa	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0349 R	ttagcccggcggaaatgac	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0350 F	attgcggtggtctcctctcctac	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0350 R	tcctccgtgacctgtgattctgt	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0351 F	gattgctgtctggatgcgtgttt	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0351 R	cccatgcgttgttctgcttctac	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0352 F	ataaaagtttggagttcgctgttcg	*B. thetaiotaomicron* arabinose gene cluster expression	

BT0352 R	caatactctgtttcgttcgcttcct	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0353 F	cgaaatcggcggagtggtg	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0353 R	ccgctgtgcaggggattg	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0354 F	cgatccaggcgaaagggtagtt	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0354 R	agtgcggggagttcgttgatg	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0355 F	acgccgttacaatcctcagtcac	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0355 R	ggcggcaatccagaagaagtc	*B. thetaiotaomicron* arabinose gene cluster expression	
BT0790 F	cattgccggttccgattgttc	*B. thetaiotaomicron* xylose gene cluster expression	
BT0790 R	ggcgcgtacccctagagtgtttt	*B. thetaiotaomicron* xylose gene cluster expression	
BT0791 F	gaccgggcactggacaagaatc	*B. thetaiotaomicron* xylose gene cluster expression	
BT0791 R	cctcaactgggcagcggtaaat	*B. thetaiotaomicron* xylose gene cluster expression	
BT0792 F	ccggatgggcagaacaaga	*B. thetaiotaomicron* xylose gene cluster expression	
BT0792 R	caccgcacgagaatcacaccag	*B. thetaiotaomicron* xylose gene cluster expression	
BT0793 F	ggtttcggaaggtgccagtgt	*B. thetaiotaomicron* xylose gene cluster expression	
BT0793 R	ctctccgccaagttcaatcgtt	*B. thetaiotaomicron* xylose gene cluster expression	
BT0794 F	ggcggtttgctcttcggttatga	*B. thetaiotaomicron* xylose gene cluster expression	
BT0794 R	cccccagcacgcagcctatc	*B. thetaiotaomicron* xylose gene cluster expression	
BT0795 F	cgagatatcatgcccgactggttgta	*B. thetaiotaomicron* xylose gene cluster expression	
BT0795 R	atattttcggcacggatttctttg	*B. thetaiotaomicron* xylose gene cluster expression	
BXY19480 F	cccgccaggtggatgagttta	*B. xylanisolvens* XB1A Ribose PUL Flanking Gene Expression	
BXY19480 R	tcgtggcgtcctatggtcctatt	*B. xylanisolvens* XB1A Ribose PUL Flanking Gene Expression	
BXY19510 F	caggataacgcaatgataagagga	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19510 R	aatacgagtaataggagggttcaaatagt	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19520 F	gagctctcgcctatgaaaacaataa	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19520 R	gcacaatccgccgcagaa	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19530 F	atcggctacggctctaccacaaag	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19530 R	gcacctgccaccccatcaat	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19540 F	tgttcccgcttgtggtggttc	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19540 R	tttcggctataatcttctcttttccttca	*B. xylanisolvens* XB1A Ribose PUL Expression	

BXY19550 F	tatgcgggattttggctatgttgt	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19550 R	ggagatggtggctgcctgattat	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19560 F	gtatctacacgccacatcggttcc	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19560 R	ggctggttctactttgcggtctg	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19570 F	aatgcggcttactgcgtgatg	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19570 R	tcgagtgccggtttccagtatt	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19590 F	cttattcggctctgtcgttgcgttat	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19590 R	tgcggagggtggaaagaatgtg	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19600 F	tgccgctatttggggagtatt	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19600 R	cccaaaccaaccaggaagaag	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19610 F	ttggctggttgaaactggtctctt	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19610 R	tccgtcatagtttttgctttcctctt	*B. xylanisolvens* XB1A Ribose PUL Expression	
BXY19620 F	caacgggtagccaatgtgataaataat	*B. xylanisolvens* XB1A Ribose PUL Flanking Gene Expression	
BXY19620 R	gtgcgggcctcttctctacca	*B. xylanisolvens* XB1A Ribose PUL Flanking Gene Expression	
Baccac_00077 F	TTGACCTTACCGCGTTCTTTACTTTT	B. caccae ATCC 43185 rusC homolog Expression	
Baccac_00077 R	TCGGAGCCGGATATTTGTTGTT	B. caccae ATCC 43185 rusC homolog Expression	
Bacint_01219 F	TTCATTGGGCTCGCGTATTAGTG	B. intestinalis DSM 17393 rusC homolog Expression	
Bacint_01219 R	CGACGTTAGCGGTTGTTTTCTTTT	B. intestinalis DSM 17393 rusC homolog Expression	
HMPREF9447_02347 F	CACTCTGGAGCCCGAAAACAAACT	B. oleiciplenus DSM 22535 rusC homolog Expression	
HMPREF9447_02347 R	TCAGACCGCGCATAGATTCCACT	B. oleiciplenus DSM 22535 rusC homolog Expression	
HMPREF9447_02354 F	AACTCCCGAGGCATACGCTTCA	B. oleiciplenus DSM 22535 GH76 in rus PUL homolog	
HMPREF9447_02354 R	CAATAAATGCCGCCTCCCAGAT	B. oleiciplenus DSM 22535 GH76 in rus PUL homolog	
Bacova_02051 F	ATCGGCTACGGCTCTACCACAAAG	B. ovatus ATCC 8483 rusC homolog Expresion	
Bacova_02051 R	GCACCTGCCACCCCATCAAT	B. ovatus ATCC 8483 rusC homolog Expresion	
C799_03204 F	TGCCGACTTAATTGTTCCCTCTTC	B. theta dnLKV9 rusC homolog Expression	
C799_03204 R	GTTTGACCCCGGTTCTCCACTT	B. theta dnLKV9 rusC homolog Expression	
BACCELL_01836 F	TTAGCGCCCAATTATCCAGAGC	B. celluosilyticus DSM 14838 rusC homolog Expression	
BACCELL_01836 R	GTTTTACGACCGCCATTTTTCATT	B. celluosilyticus DSM 14838 rusC homolog Expression	
HMPREF1017_00084 F	TCAGGATTATTTAACGGCGAACAAG	B. ovatus 3_8_47FAA rusC homolog Expression	

HMPREF1017_00084 R	AGCGGAAGCGGGAAGAACC	B. ovatus 3_8_47FAA rusC homolog Expression	
BT2803-04 Scar_3 F	GTAAGGTGTTTGATTCGTATTTCAGACG	*In vivo* Quantification of ΔBT2803-04 :: BT2803-04 Complemented strain	
BT2803-04 Scar_3 R	TTATAAGTTATCAGGTGGACAGCTTTCTTTA	*In vivo* Quantification of ΔBT2803-04 :: BT2803-04 Complemented strain	
Tag1	ATGTCGCCAATTGTCACTTTCTCA	*In vivo* Quantification of barcoded wild-type strain	
Tag3	TTATGACCAGCCGCAAATGAAAAG	*In vivo* Quantification of barcoded Δ*BT2805-06*	
Tag9	TCAAATCCGGGGACTGGGCTTAGA	*In vivo* Quantification of barcoded Δ*BT2807-08*	
Tag14	GGCACGCCATTCTTCATCTAACTG	*In vivo* Quantification of barcoded strains: Δ*BT2802*, Δ*BT2803*, Δ*BT2804*, Δ*BT2803-04*, Δ*BT2809*, Δ*BT4554*,	
Universal Tag R	CACAATATGAGCAACAAGGAATCC	Reverse primer for all barcode qPCR primers	
Protein Cloning Primers			
His Tag/Sticky Ends are Bolded and Protease Cleavage Site Underlined			
***B. thetaiotaomicron* VPI5482**			
BT2803 Forward	**GAAGGAGATATACATATG**GAAACTATATCAATTCATAGACCC	Amplification and Cloning of BT2803, C-terminal His Tag, No protease cleavge sequence	
BT2803 Reverse	**GTGATGGTGGTGATGATG**TAGCACAACTTCTCTTTTATA	Amplification and Cloning of BT2803, C-terminal His Tag, No protease cleavage sequence	
BT2804 Forward	**CATCATCACCACCATCAC**GAGAACCTGTACTTCCAGGGCAAAGTAGTAGTTATTGGAAGT	Amplification and Cloning of BT2804, N-terminal His Tag, No protease cleavage sequence	
BT2804 Reverse	**GTGGCGGCCGCTCTATTA**AAAGTTGAATGATAAGGCTAA	Amplification and Cloning of BT2804, N-terminal His Tag, No protease cleavage sequence	
BT2807 Forward	**CATCATCACCACCATCAC**GAGAACCTGTACTTCCAGGGCTCTGACAATAATAATGAAACATACGTC	Amplification and Cloning of BT2807, N-terminal His Tag, with TEV protease cleavage sequence	
BT2807 Reverse	**GTGGCGGCCGCTCTATTA**TAAGTTATTTCCAATATTCTCAATTGT	Amplification and Cloning of BT2807, N-terminal His Tag, with TEV protease cleavage sequence	
BT2808 Forward	**CATCATCACCACCATCAC**GAGAACCTGTACTTCCAGGGCGCACAACAGAAAGAGACTACTTC	Amplification and Cloning of BT2808, N-terminal His Tag, with TEV protease cleavage sequence	
BT2808 Reverse	**GTGGCGGCCGCTCTATTA**ATTCTCTTTTCCTGTCACTTGTC	Amplification and Cloning of BT2808, N-terminal His Tag, with TEV protease cleavage sequence	

Notes

This work has been reprinted and modified with permission from the following authors: Glowacki, R.W.P., Pudlo, N.A., Tuncil, Y., Luis, A.S., Sajjakulnukit, P., Terekhov, A.I., Lyssiotis, C.A., Hamaker, B.R., and Martens, E.C. A ribose scavenging system confers fitness on the human gut symbiont Bacteroides thetaiotaomicron in a diet-specific manner. *Cell Host Microbe.* **In Press** (2019).

References

1. Porter, N.T. and E.C. Martens, The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu. Rev. Microbiol.* **71**, 349-369 (2017).

2. McLeod, A., et al., Global transcriptome response in Lactobacillus sakei during growth on ribose. *BMC Microbiol.* **11**, 145 (2011).

3. Pokusaeva, K., et al., Ribose utilization by the human commensal Bifidobacterium breve UCC2003. *Microb. Biotechnol.* **3**, 311-23 (2010).

4. Fabich, A.J., et al., Comparison of Carbon Nutrition for Pathogenic and Commensal Escherichia coli Strains in the Mouse Intestine. *Infect. Immun.* **76**, 1143-1152 (2008).

5. Harvey, P.C., et al., Salmonella enterica serovar typhimurium colonizing the lumen of the chicken intestine grows slowly and upregulates a unique set of virulence and metabolism genes. *Infect. Immun.* **79**, 4105-21 (2011).

6. Kim, H.S., et al., Genes encoding ribonucleoside hydrolase 1 and 2 from Corynebacterium ammoniagenes. *Microbiology* **152**, 1169-77 (2006).

7. Hammer-Jespersen, K., et al., Induction of Enzymes Involved in the Catabolism of Deoxyribonucleosides and Ribonucleosides in E.coli K12. *Eur. J. Biochem.* **19**, 533-538 (1971).

8. Finkel, S.E. and R. Kolter, DNA as a nutrient: novel role for bacterial competence gene homologs. *J. Bacteriol.* **183**, 6288-93 (2001).

9. Palchevskiy, V. and S.E. Finkel, A role for single-stranded exonucleases in the use of DNA as a nutrient. *J. Bacteriol.* **191**, 3712-6 (2009).

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

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

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

13. Ndeh, D., et al., Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65-70 (2017).

14. Luis, A.S., et al., Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat. Microbiol.* **3**, 210-219 (2018).

15. Hehemann, J.K., A.G.; Pudlo, P.A.; Martens, E.C.; Boraston, A.B., Bacteria of the human gut microbiome catabolize seaweed glycans with carbohydrate-active enzymes update from extrinsic microbes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19786-19791 (2012).

16. Pluvinage, B., et al., Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. *Nat. Commun.* **9**, 1043 (2018).

17. Martens, E.C., H.C. Chiang, and J.I. Gordon, Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447-57 (2008).

18. Briliute, J., et al., Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat. Microbiol.* **4**, 1571-1581 (2019).

19. Glenwright, A.J., et al., Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407-411 (2017).

20. Terrapon, N., et al., PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* **46**, D677-D683 (2018).

21. Martens, E.C., et al., Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).

22. 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).

23. Martens, E.C., et al., Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673-7 (2009).

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

25. Bjursell, M.K., E.C. Martens, and J.I. Gordon, 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).

26. Desai, M.S., et al., A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353 (2016).

27. Weickmann, J.L., E.M. Olson, and D.G. Glitz, Immunological Assay of Pancreatic Ribonucleases in Serum as an Indicator of Pancreatic Cancer. *Cancer Res.* **44**, 1682-87 (1984).

28. McConnell, R.E., et al., The enterocyte microvillus is a vesicle-generating organelle. *J. Cell Biol.* **185**, 1285-98 (2009).

29. Hove-Jensen, B., D.E. Brodersen, and M.C. Manav, The Prodigal Compound: Return of Ribosyl 1,5-Bisphosphate as an Important Player in Metabolism. *Microbiol. Mol. Biol. Rev.* **83**, e00040-18 (2019).

30. Sonnenburg, E.D., et al., Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, 1241-52 (2010).

31. Pan, N. and J.A. Imlay, How does oxygen inhibit central metabolism in the obligate anaerobe Bacteroides thetaiotaomicron. *Mol. Microbiol.* **39**, 1562-1571 (2001).

32. Temple, M.J., et al., A Bacteroidetes locus dedicated to fungal 1,6-beta-glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6-beta-glucanase. *J. Biol. Chem.* **292**, 10639-10650 (2017).

33. Duquesne, S., et al., Structural and functional diversity of microcins, gene-encoded antibacterial peptides from enterobacteria. *J. Mol. Microbiol. Biotechnol.* **13**, 200-9 (2007).

34. Martinez-Jehanne, V., et al., Role of deoxyribose catabolism in colonization of the murine intestine by pathogenic Escherichia coli strains. *Infect. Immun.* **77**, 1442-50 (2009).

35. Bufe, T., et al., Differential transcriptome analysis of enterohemorrhagic Escherichia coli strains reveals differences in response to plant-derived compounds. *BMC Microbiol.* **19**, 212 (2019).

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

37. Koropatkin, N.M., et al., Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105-15 (2008).

38. Petersen, C. and L.B. Moller, The RihA, RihB, and RihC ribonucleoside hydrolases of Escherichia coli. Substrate specificity, gene expression, and regulation. *J. Biol. Chem.* **276**, 884-94 (2001).

39. Datsenko, K.A. and B.L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640-5 (2000).

40. 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).

41. Speer, M.A. *Development of a genetically modified silage inoculant for the biological pretreatment of lignocellulosic biomass*. (Unpublished doctoral dissertation in Agricultural and Biological Engineering, 2013) University Park, PA

42. Pudlo, N.A., et al., Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. *MBio* **6**, e01282-15 (2015).

43. Parkin, D.W., et al., Nucleoside Hydrolase from Crithidia fasciculata. Metabolic role, purification, specificity, and kinetic mechanism. *J. Biol. Chem.* **266**, 20658-20665 (1991).

44. Wu, Z.L., Phosphatase-coupled universal kinase assay and kinetics for first-order-rate coupling reaction. *PLoS One* **6**, e23172 (2011).

45. Chuvikovsky, D.V., et al., Ribokinase from E. coli: expression, purification, and substrate specificity. *Bioorg. Med. Chem.* **14**, 6327-32 (2006).

46. Pettolino, F.A., et al., Determining the polysaccharide composition of plant cell walls. *Nat. Protoc.* **7**, 1590-607 (2012).

Chapter III

Diet Modulates Colonic T Cell Responses by Regulating the Expression of a *Bacteroides thetaiotaomicron* Antigen

Abstract

T cell responses to symbionts in the intestine drive tolerance or inflammation depending on the genetic background of the host. These symbionts in the gut sense the available nutrients and adapt their metabolic programs to use these nutrients efficiently. Here, we ask whether diet can alter the expression of a bacterial antigen to modulate adaptive immune responses. We generated a $CD4^+$ T cell hybridoma, BθOM, specific for *Bacteroides thetaiotaomicron* (*Bt*). Adoptively transferred transgenic T cells expressing the BθOM TCR proliferated in the colon, colon-draining lymph node, and spleen in *Bt*–colonized healthy mice and differentiated into regulatory T cells (T_{regs}) and effector T cells (T_{effs}). Depletion of *Bt*–specific T_{regs} resulted in colitis, showing that a single protein expressed by *Bt* can drive differentiation of T_{regs} that self-regulate T_{effs} to prevent disease. We found that BθOM T cells recognized a peptide derived from a single *Bt* protein, BT4295, whose expression is regulated by nutrients, with glucose being a strong catabolite repressor. Mice fed a high-glucose diet had a greatly reduced activation of BθOM T cells in the colon. These studies establish that the immune response to specific bacterial antigens can be modified by changes in the diet by altering antigen expression in the microbe.

Introduction

Dietary components and metabolites produced by host and microbial enzymes modulate the function of a variety of host immune cells including T cells[1-3]. These products can have local effects on the intestinal immune system and in more distant organs[4]. For instance, host enzymes break down starch and various disaccharides in the diet to produce glucose, which is required systemically for maximal effector T cell (T_{eff}) stimulation[5,6]. Microbial metabolites derived from dietary fiber, flavonoids, and amino acids such as tryptophan have immunomodulatory

activities[3,7-10]. As examples, short-chain fatty acids from fiber fermentation promote the development of intestinal regulatory T cells (T_{regs})[3], modulate macrophage polarization[11], and suppress innate lymphoid cell development[12]. Further, tryptophan catabolites act via the aryl hydrocarbon receptor to induce T cell cytokine production[13]; taurine-conjugated bile acids formed from milk-derived dietary fat induce a proinflammatory T helper type 1 (T_H1) immune response[14], and, last, the microbial metabolite desaminotyrosine derived from flavonoids stimulates type I interferons (IFNs) and modulates macrophage activation and cytokine production[15]. Recently, ascorbate, a microbial metabolite altered in Crohn's disease, has been shown to modulate T cell activity[16]. Other dietary components such as excess salt can change the composition of the microbiome and favor pathogenic T helper 17 (T_H17) responses[17]. Conversely, an iron-deficient diet can dampen intestinal inflammation[18]. Collectively, these studies reveal the dominant effects of dietary components and their immediate or downstream metabolites on the immune system.

$CD4^+$ T cells play a critical role in the response to specific microbial antigens in the intestine[19-23]. Symbiotic bacteria that do not damage the host produce tolerogenic T_{reg} responses, whereas pathogens that cause damage elicit T_{eff} responses. In both cases, microbe-specific antigens drive these responses, and these intestinal bacteria are well known to be modulated by diet. However, the effect of diet on T cells that recognize these different groups of symbionts has not been tested. This latter question is of importance due to the effects of diet on the composition and physiology of the microbiome, which has a multitude of effects on the host. It is unclear whether specific dietary components have effects at the level of specific bacterial antigens and the T cells that recognize them.

We hypothesized that the $CD4^+$ immune response to specific bacterial antigens can be modified by changes in the diet through effects on antigen expression of the microbe. Progress in this area has been hampered by the lack of a model system in which a $CD4^+$T cell response against a specific gut symbiont can be examined. To this end, we developed a $CD4^+$ T cell model, termed BθOM, specific for an outer membrane (OM) antigen from *Bacteroides thetaiotaomicron* (*Bt*, Bθ). *Bt* is a prototypic gut symbiont that degrades a wide variety of dietary, host, and microbial glycans and is a representative of a prominent genus found in most human microbiomes[24]. In healthy mice gavaged with *Bt*, we found that TCR (T cell receptor) transgenic BθOM T cells responded *in vivo* by differentiating into T_{regs} and T_{effs}. Deletion of the

BθOM T_{regs} induced colitis by activated BθOM T cells, revealing that the symbiont-specific $CD4^+$ T cells were no longer able to self-regulate to prevent T cell–mediated disease.
The *Bt* antigen recognized by BθOM T cells was identified to be BT4295, an OM protein contained in one of *Bt*'s many polysaccharide utilization loci (PULs). We found that we can modify the response of BθOM T cells to their cognate antigen by altering the salts and glycans available to *Bt*. Glucose was identified as a catabolite repressor of BT4295 expression. Mice fed a high-glucose diet had greatly reduced activation of BθOM T cells, establishing a direct link between dietary regulation of a microbial antigen and $CD4^+$ T cell activation. These results show that specific dietary components can alter the T cell–driven immune response to dominant symbiotic antigens.

Results

The Bt–specific $CD4^+$ T cell response is sensitive to changes in Bt growth media

To determine how dietary components and metabolites can affect the interactions between a symbiont and the host immune system, we developed a bacteria-specific $CD4^+$ T cell model. We chose to focus our study on *Bt*, a model gut symbiont that is known to adapt to changes in the available nutrients, especially by changing expression of carbohydrate utilization gene loci. We immunized C57BL/6J mice with the human *Bt* strain VPI-5482 (herein referred to as *Bt*) and produced T cell hybridoma cell lines that responded to *Bt*. We screened the T cell hybridomas for reactivity against *Bt* outer membrane vesicles (OMVs), which have been shown to be a source of antigen to the immune system[25]. To identify a T cell sensitive to changes in available nutrients, we took advantage of a fortuitous observation that *Bt* grown in two different formulations of tryptone-yeast-glucose (TYG) media—classic TYG (TYG) and modified TYG (mTYG) (Table 3.1)—stimulated T cells differently. We chose one T cell hybridoma clone (herein denoted as *Bt* outer membrane or "BθOM") that showed a robust response to both *Bt* and OMVs in T cell stimulation assays (Figure 3.1, A and B). When we cultured BθOM T cell hybridomas with bone marrow–derived macrophages (BMDMs) along with *Bt* grown in the different media, BθOM T cell activation was highest with *Bt* grown in TYG media (Figure 3.1C); no stimulation of these T cells was observed when *Bt* was grown in mTYG media (Figure 3.1C). Thus, BθOM T cells were sensitive to changes in the nutrients in the media used to grow *Bt*.

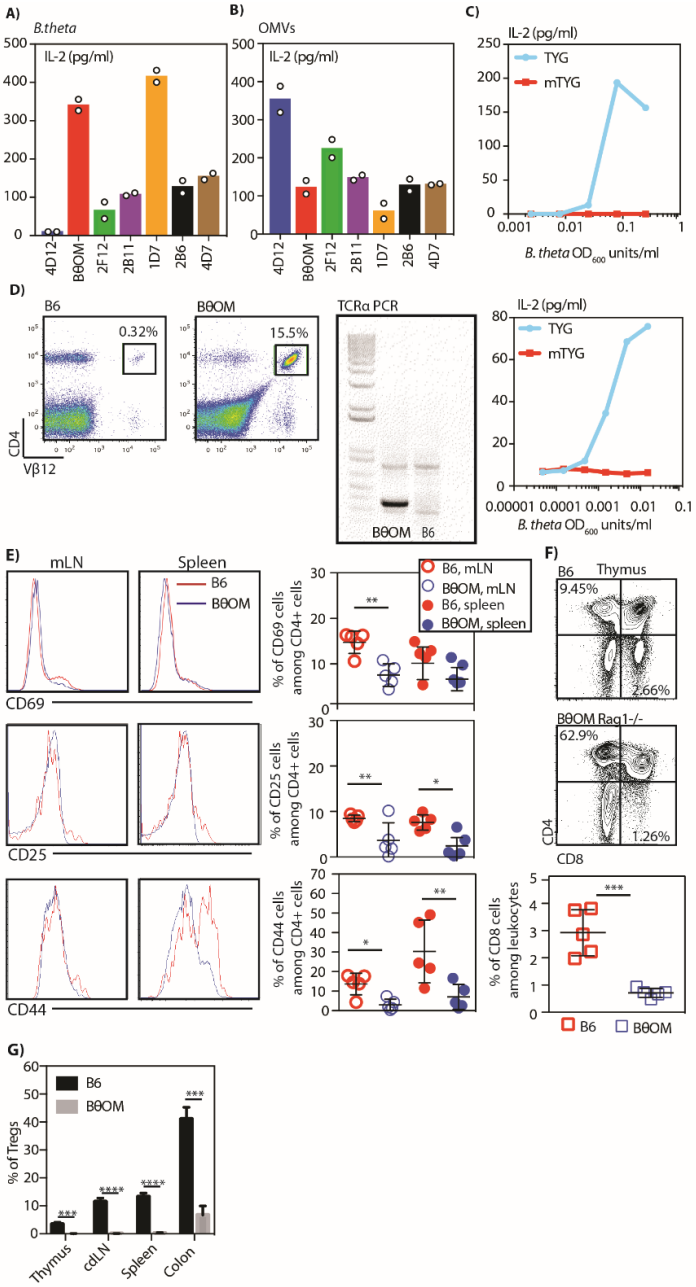


Figure 3.1 Generation and characterization of the BθOM TCR transgenic mouse.

A and B) IL-2 levels in picogram per milliliter after generated T cell hybrid clones were cultured with BMDMs loaded with A) Bt (n = 2, one experiment) or B) OMVs (n = 2, one experiment). C) IL-2 levels in picogram per milliliter after the BθOM T cell hybrid was cultured with BMDMs loaded with Bt grown in TYG or mTYG (n = 2; both replicates are shown). D) Representative flow cytometry plot with Vβ12 staining on blood leukocytes of C57BL/6J mice (left) or BθOM transgenic mice (middle) (n = 3, three experiments). Representative TCRα1 PCR on DNA isolated from tails of C57BL/6J mice and BθOM transgenic mice (right) (x = 3, three experiments). E) Representative histograms of CD69, CD25, and CD44 expression (left) and quantification of the percentage of CD69, CD25, and CD44 cells among all CD4 cells (right) isolated from the mLNs and spleen of C57BL/6J mice (red) or BθOM transgenic mice (blue) (x = 5, three experiments). F) Representative flow cytometry plots of CD4 and CD8 staining of thymic cells isolated from C57BL/6J mice or BθOM transgenic mice (x = 5, three experiments) and quantification of the percentage of CD8 T cells among the thymic leukocyte population. G) Percentage of Tregs in the thymus (n ≥ 6, three experiments), cdLN (n ≥ 10, six experiments), spleen (n ≥ 10, six experiments), and colon (n = 4, four experiments) of C57BL/6J mice (black) or

*BθOM transgenic mice (gray). Student's t test: (E) *P < 0.1 and **P < 0.01; (F) ***P = 0.0004; (G) ****P < 0.0001 and ***P = 0.0001*

We next created a transgenic mouse line expressing the BθOM TCR genes on a C57BL/6J-*Rag1*$^{-/-}$-CD45.1 genetic background (BθOM *Rag1*$^{-/-}$ mouse strain). The TCR transgenic T cells from this line were I-A^{b} restricted, expressed Vα1 and Vβ12 (Figure 3.1D), and were specific for *Bt* (human or mouse isolates) (Figure 3.2A). The peripheral T cells from BθOM *Rag1*$^{-/-}$ mice were essentially all naive, expressing low levels of CD69, CD25, and CD44 proteins (Figure 3.1E); the thymus was also devoid of CD8^{+} T cells (Figure 3.1F). We found that BθOM transgenic mice develop few, if any, thymic or peripheral T_{regs} compared with nontransgenic C57BL/6J mice (Figure 3.1G). Isolated naive T cells from BθOM *Rag1*$^{-/-}$ mice could be activated when stimulated *in vitro* with BMDM incubated with either *Bt* or OMVs (Figure 3.2, A and B). Stimulation of the BθOM TCRtg T cells by *Bt* was confirmed to be sensitive to nutrients in TYG media (Figure 3.2C), enabling the use of BθOM T cells to study the effect of diet on symbiont-host interactions.

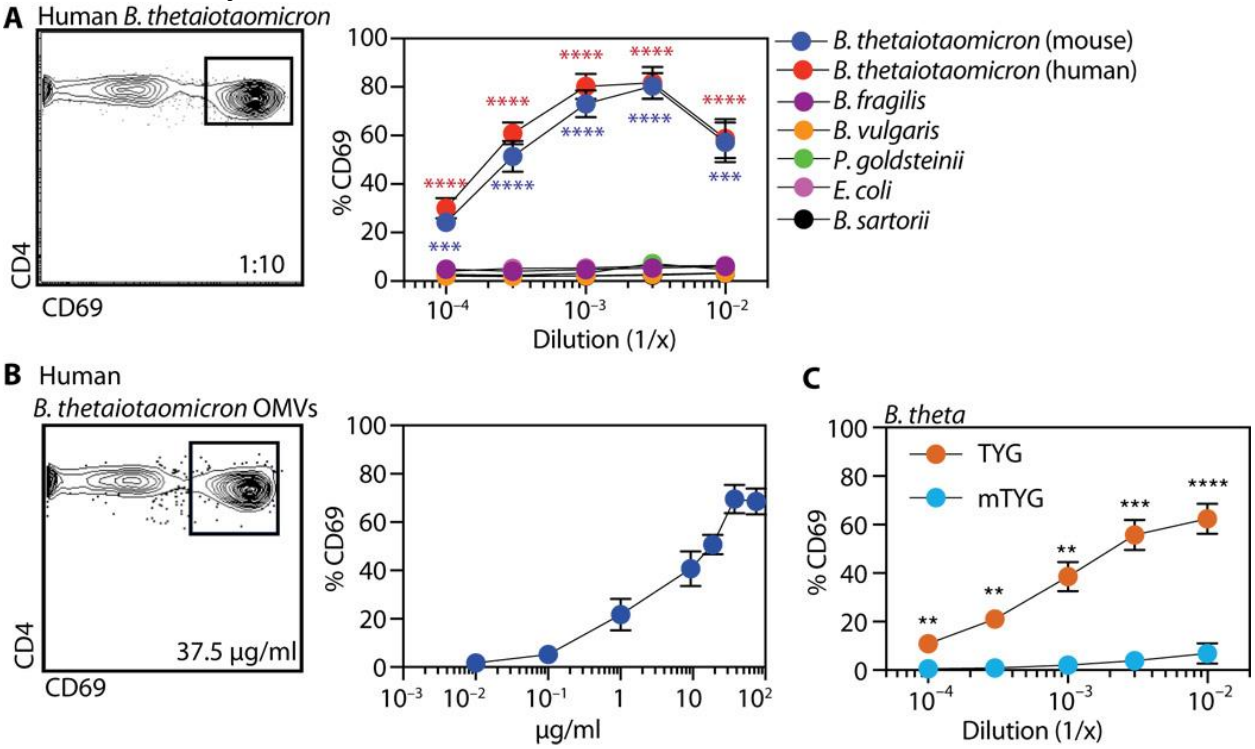


Figure 3.2 Bt activates BθOM T cells in a nutrient-dependent manner.

*A and B) Percentage of CD69 expressing BθOM T cells after a 24-hour culture with BMDM loaded with A) Bacteroidaceae family [human: B. thetaiotaomicron (n = 4, four experiments); mouse: B. fragilis, B. vulgaris, Parabacteroides goldsteinii, E. coli, B. sartorii (n = 3, three experiments)] or B) human Bt OMVs (75 µg/ml: n = 7, six experiments; 37.5 µg/ml: n = 6, six experiments; 18.75 µg/ml: n = 5, four experiments; 10 µg/ml: n= 8, six experiments; 1 µg/ml: n = 3, three experiments; 0.1 µg/ml: n = 4, four experiments; 0.01 µg/ml: n = 3, three experiments). Flow cytometry plots are gated on CD4^{+} CD45.1^{+}leukocytes. C) Percentage of CD69 expressing BθOM hybridoma T cells after a 24-hour culture with BMDM loaded with human Bt grown in TYG (n = 13, five experiments) or mTYG medium (n = 5, five experiments). One-way ANOVA analysis: (A) ***P < 0.001 and ****P < 0.0001.*

*Means with asterisks are significantly different by Tukey's multiple comparisons test. Student's t test: (C) ****P < 0.0001, ***P = 0.0001, and **P < 0.01.*

We then evaluated the function of BθOM T cells *in vivo* by transferring them into antibiotic pretreated *Rag1*$^{-/-}$ mice. Mice were pretreated with antibiotics for 3 weeks to allow colonization with the subsequently gavaged human isolate of *Bt*, which we previously showed colonize mice under these conditions[26]. Sorted naive ($CD44^{lo}CD62L^{hi}$) $CD25^{-}CD4^{+}CD45.1^{+}$ BθOM T cells (Figure 3.3A) were transferred into *Rag1*$^{-/-}$ mice that had

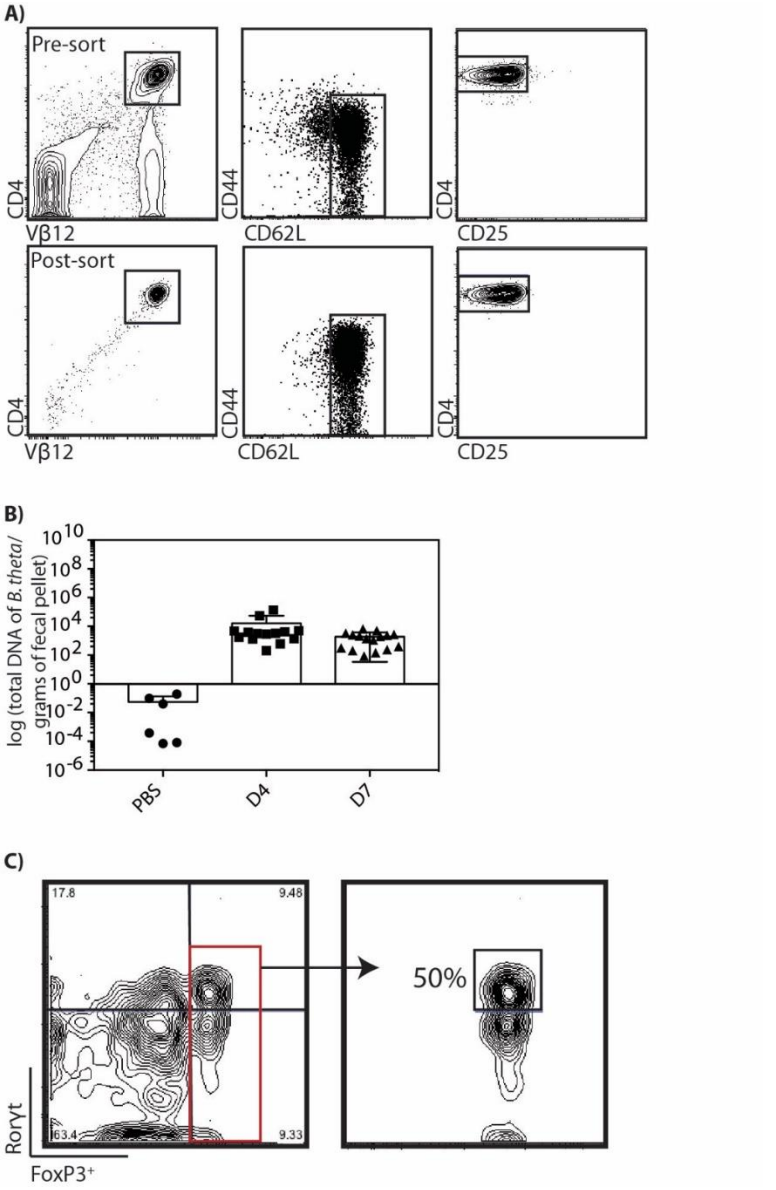


Figure 3.3 Sorting strategy and Bt colonization for in vivo BθOM T cell transfer experiments.

A) Example flow cytometry plots of the gating strategy for sorting $CD4^{+}V\beta12^{+}CD44^{-}CD62L^{+}CD25^{-}$ cells[9] and post-sort flow cytometry plots showing no $CD4^{+}V\beta12^{+}CD44^{-}CD62L^{+}CD25^{+}$ cells in the sorted population (bottom). B) Colonization levels of Bt (total DNA of Bt/gram of fecal matter) on days 4 PBS (n=6, 5 experiments) and day 4 and day 7 Bt (n≥14, 9 experiments) gavaged Rag1$^{-/-}$ mice transferred with BθOM T cells. C) Representative flow cytometry plots of Rorγt and $Foxp3^{+}$ staining of colonic cells isolated from Bt gavaged Rag1$^{-/-}$ mice transferred with BθOM T cells (n=2, 2 experiments).

been previously colonized by *Bt* for 4 days (Figure 3.3B). We identified $CD4^+CD45.1^+$ T cells in the lamina propria, colon-draining lymph node (cdLN), which refers to the lymph node within the mesenteric lymph node (mLN) that drains the colon, and spleen 7 days after T cell transfer (Figure 3.4, A and B). In these mice, BθOM T cell localization in the colon lamina propria and cdLNs was dependent on *Bt* colonization (Figure 3.4C). We also found BθOM T cells in the spleen of *Bt*–colonized *Rag1*$^{-/-}$ mice (Figure 3.4C). The BθOM T cells proliferated in the lamina propria, cdLN, and spleen, revealing that they were exposed to their cognate antigen (Figure 3.4, D and E). *Bt*–gavaged BθOM *Rag1*$^{-/-}$ mice did not have obvious signs of disease such as weight loss (Figure 3.5A).

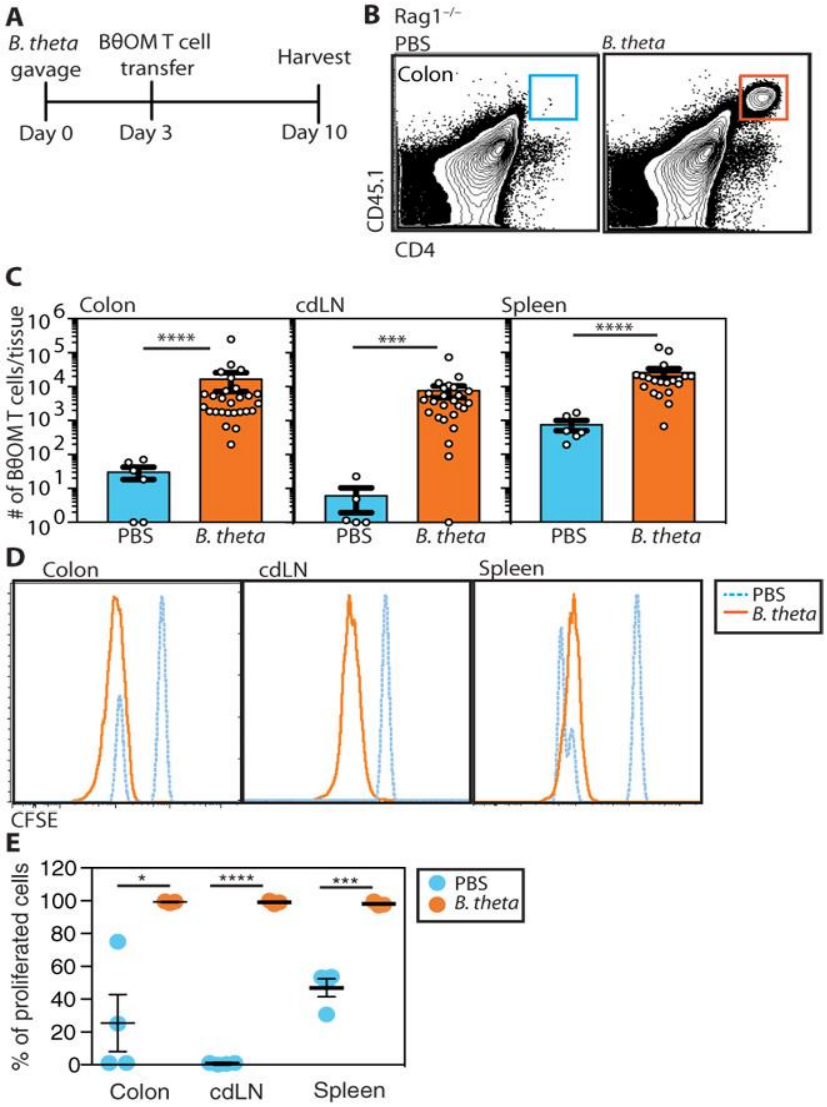


Figure 3.4 BθOM T cells proliferate in the colon in Bt–colonized mice.

A) Schematic of adoptive transfer of BθOM T cells into Rag1$^{-/-}$ *mice gavaged with PBS or Bt. B) Representative flow cytometry plots of CD45.1*$^+$*CD4*$^+$ *BθOM T cells in the colon of Bt–gavaged mice compared with PBS-gavaged mice. C) Number of BθOM T*

*cells among live leukocytes that are CD45.2^{-}CD45.1^{+}CD4^{+} in PBS or Bt–gavaged mice in the colon (n ≥ 6, ≥ five experiments), cdLN (n ≥ 5, ≥ three experiments), and spleen (n ≥ 6, ≥ four experiments). D) Representative histograms of adoptively transferred carboxyfluorescein diacetate succinimidyl ester (CFSE)–labeled BθOM T cells in the colon (n ≥ 3, ≥ three experiments), cdLN (n ≥ 3, three experiments), and spleen (n ≥ 3, ≥ three experiments) of Bt–gavaged mice compared with PBS-gavaged mice. E) Quantification of the percentage of proliferated CFSE low CD45.2^{-}CD45.1^{+}CD4^{+} T cells in the colon (n ≥ 3, ≥ three experiments), cdLN (n ≥ 3, three experiments), and spleen (n ≥ 3, ≥ three experiments). Mann-Whitney test for non-normally distributed data: (C) ****P < 0.0001 and ***P = 0.0006. Student's t test: (E) ****P < 0.0001, ***P = 0.0005, and *P = 0.0160.*

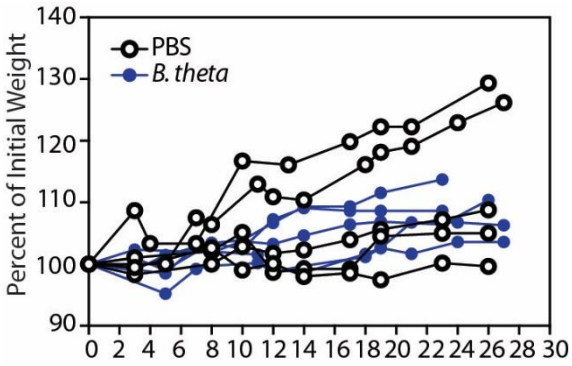


Figure 3.5 BθOM T cells do not cause weight loss in Bt–colonized mice.

Weight changes calculated as the percentage of the initial weight in Rag1-/- mice transferred with BθOM T cells and gavaged with PBS (n=5, 2 experiments) or Bt (n=5, 2 experiments).

BθOM T cells differentiate into T_{effs} and T_{regs} that self-regulate to prevent colitis

Because *Bacteroides* have been previously shown to be strong drivers of T_{reg} induction[27], we reasoned that the BθOM T_{regs} would mediate tolerance to *Bt*. We transferred BθOM T cells into *Rag1*$^{-/-}$ mice; the transferred cells were presorted for CD4^{+}CD44^{lo}CD62L^{hi}CD25^{-} to ensure that there was no transfer of preexisting T_{regs} into recipients (Figure 3.3A). Characterization of the BθOM T cells in multiple locations showed a mixture of T_{eff} and FoxP3^{+} T_{regs} in the lamina propria and cdLN with a lower percentage of T_{regs} found in the spleen (Figure 3.6, A-C). T_{reg} development in the peripheral lymphatics and the colonic tissue was dependent on *Bt* colonization because few to no T_{regs} were found in phosphate-buffered saline (PBS)–gavaged mice (Figure 3.6B and Figure 3.3B). Despite the presence of T_{regs} in both the cdLN and colonic lamina propria, the cdLNs had many more T_{regs} expressing CD25 than the colon, where most of the T_{regs} expressing FoxP3 lacked CD25 expression (Figure 3.6D). Consistent with previous reports with polyclonal T_{regs} exposed to *Bacteroides* in the lamina propria[28,29], 50% of BθOM FoxP3^{+} T_{regs} express RORγt (Figure 3.3C). This finding is also consistent with a report showing that, in healthy wild-type mice, pathobiont-specific T cells differentiate into RORγt-

expressing specific induced T_{regs} (iT_{regs}) in the large intestine[30]. Together, these data reveal that the same TCR can differentiate into both T_{effs} and T_{regs}[19].

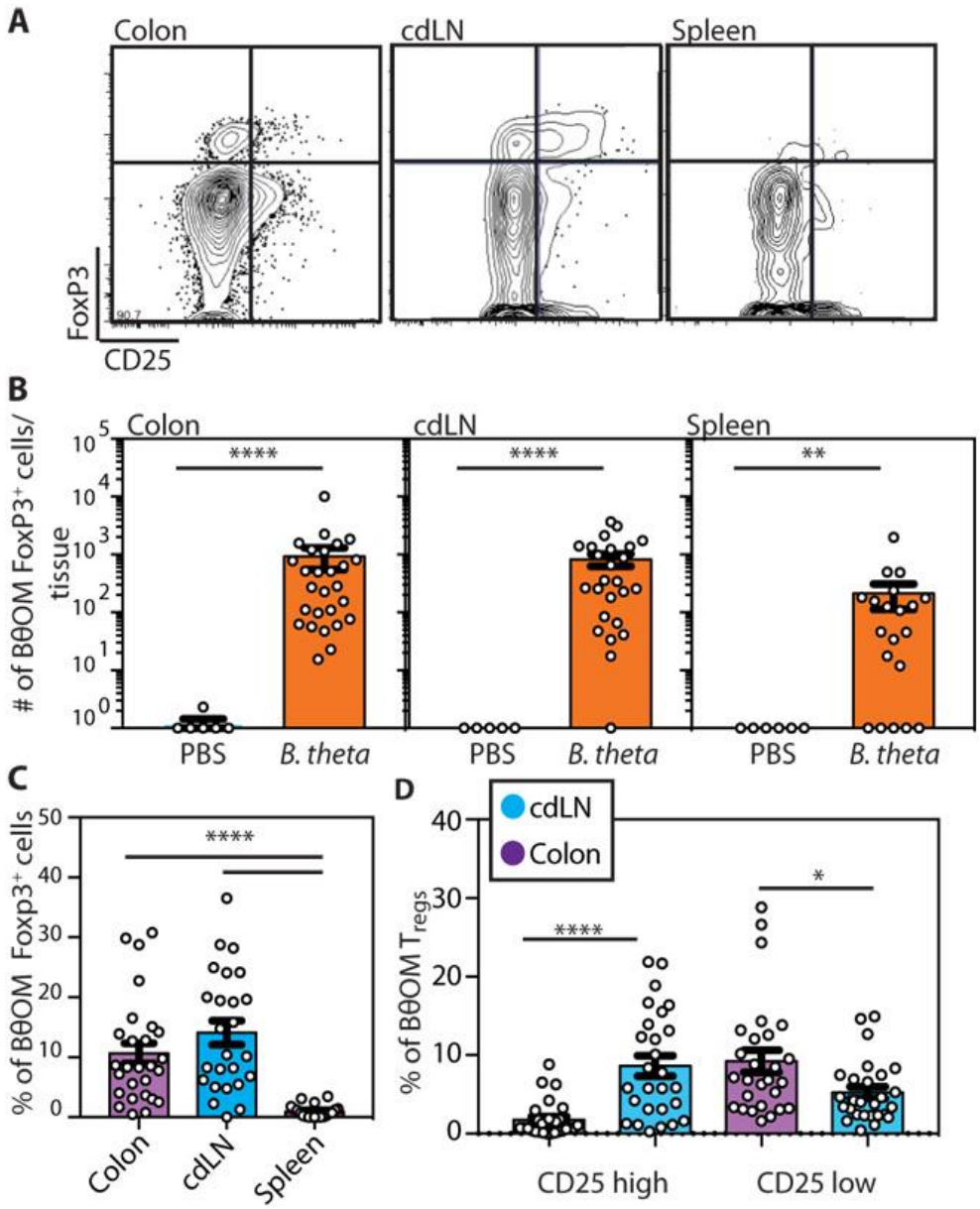


Figure 3.6 BθOM T cells in the colon differentiate into T_{regs}.

*A) Flow cytometry plots of CD45.1$^+$CD4$^+$ BθOM T cells in the colon, cdLN, and spleen of PBS or Bt–gavaged Rag1$^{-/-}$ mice transferred with naive CD25$^-$ BθOM T cells. B) The number of CD4$^+$CD45.1$^+$FoxP3$^+$ BθOM T_{regs} cells in the colon (n ≥ 6, ≥ five experiments), cdLN (n ≥ 5, ≥ three experiments), and spleen (n ≥ 6, ≥ four experiments) of PBS or Bt–gavaged Rag1$^{-/-}$ mice after CD25$^-$ BθOM T cell transfer. C) Percentage of FoxP3$^+$ T_{regs} in the colon (n = 27, nine experiments), cdLNs (n = 25, seven experiments), and spleen (n = 20, seven experiments) of Rag1$^{-/-}$ mice that received naive CD25$^-$ BθOM T cells and were gavaged with Bt. D) Percentage of CD25^{high} versus CD25^{low} CD4$^+$FoxP3$^+$ T_{regs} in the colon (n = 27, nine experiments) and cdLNs (n = 25, seven experiments) of Rag1$^{-/-}$ mice gavaged with Bt and injected with naive BθOM T cells. Mann-Whitney test for non-normally distributed data: (B) ****P < 0.0001 and **P = 0.004. Kruskal-Wallis with Dunn's posttest for non-normally distributed data: (C) ****P < 0.0001. Two-way ANOVA analysis: (D) ****P < 0.0001 and *P = 0.0161.*

We hypothesized that *Bt*–specific T_{regs} produced sufficient regulation in the colonic mucosa to prevent *Bt*–specific $CD4^+$ T cells from inducing colitis upon exposure to *Bt* To test this hypothesis, we crossed the BθOM transgenic mouse to FoxP3-DTR-GFP mice, which permits the *in vivo* depletion of T_{regs} upon diphtheria toxin (DT)[31] treatment and includes a green fluorescent protein (GFP) marker for T_{reg} identification[30]. We transferred naive,

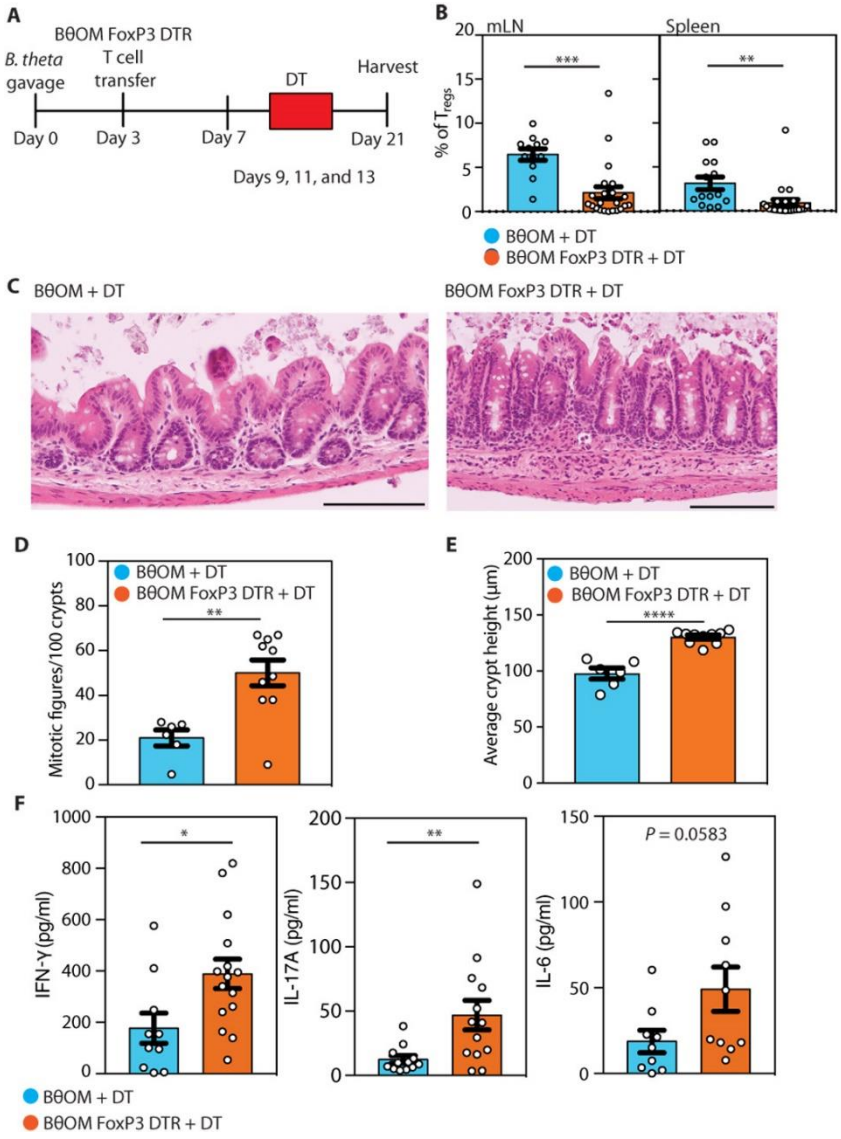


Figure 3.7 Depletion of BθOM T_{regs} drives BθOM $CD4^+$ T_{eff} to cause colitis.

A) Schematic of adoptive transfer of BθOM or BθOM-FoxP3-DTR T cells into $Rag1^{-/-}$ mice gavaged with PBS or Bt and treated with DT (31) to deplete BθOM T_{regs}. B) Percentage of BθOM T_{regs} after depletion in the mLN ($n \geq 12$, five experiments) or spleen ($n \geq 14$, five experiments). (C to E) Histology (C), quantification of the number of mitotic figures/10 crypts (D), and average crypt height (E) in cecal sections from $Rag1^{-/-}$ mice given BθOM T cells and DT ($n = 6$, three experiments) compared with those given BθOM-FoxP3-DTR T cells and DT ($n = 10$, three experiments). Scale bars, 120 μm. (F) Cytometric bead array used to quantify IFN-γ ($n \geq 10$, three experiments), IL-17A ($n \geq 10$, three experiments), and IL-6 ($n \geq 10$, three experiments) after cells isolated from

*the mLN were stimulated with PMA for 5 hours. Student's t test: (B) ***P = 0.0002 and **P = 0.0055; (D) **P = 0.0029; (E) ****P < 0.0001; (F) *P = 0.0205 and **P = 0.098.*

GFP^{lo} BθOM T cells into *Rag1*$^{-/-}$ mice colonized with *Bt* that were treated with DT on days 9, 11, and 13 (Figure 3.7A). We confirmed depletion of T_{regs} in the cdLNs and spleen (Figure 3.7B). We found that *Rag1*$^{-/-}$mice that received BθOM-FoxP3-DTR cells and DT developed colitis, with an increase in hyperproliferative crypts, epithelial proliferation, lymphocyte infiltrate, mitotic figures, and crypt height compared with control mice that received BθOM T cells and DT (Figure 3.7C-E). Cells isolated from the mLN of *Rag1*$^{-/-}$ mice transferred with BθOM-FoxP3-DTR T cells and treated with DT to deplete T_{regs} showed an increase in proinflammatory cytokines [interleukin-17A (IL-17A), IFN-γ, and IL-6] compared with cells isolated from *Rag1*$^{-/-}$ mice receiving wild-type BθOM T cells and treated with DT (Figure 3.7F and Figure 3.8A-B). Both BθOM-FoxP3-DTR T cells and wild-type BθOM T cells isolated from

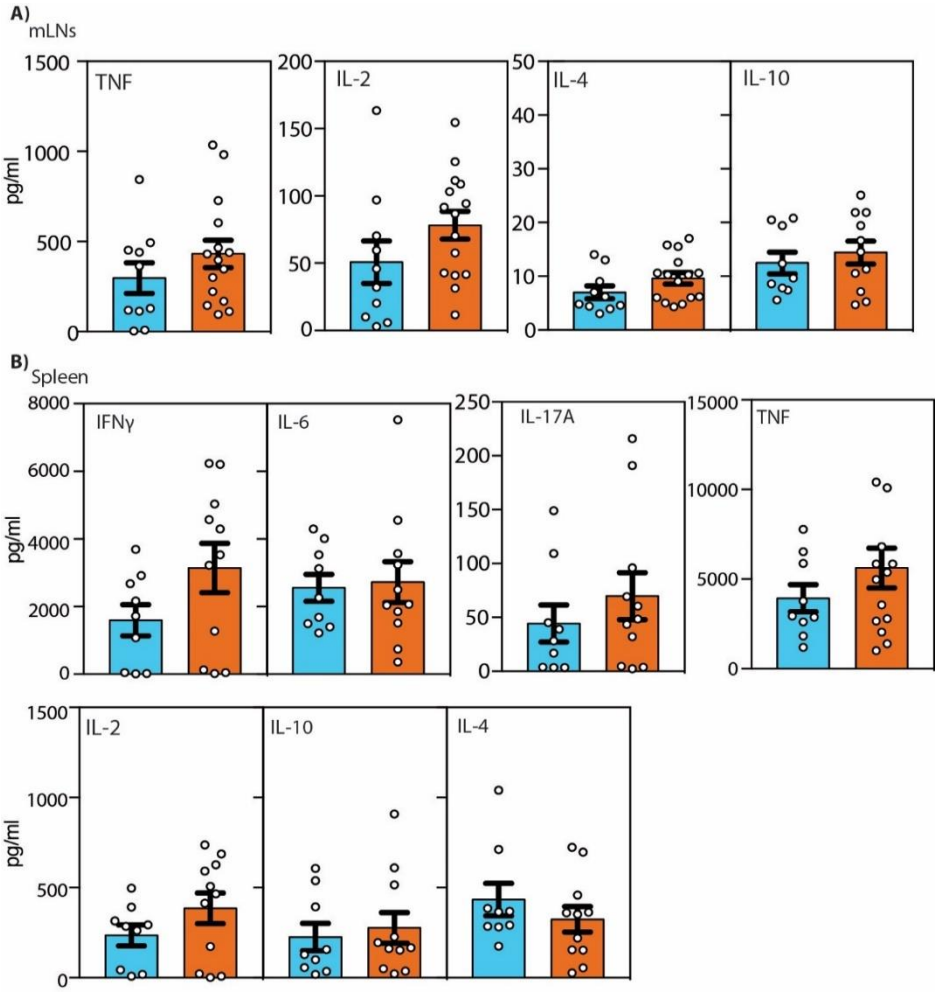


Figure 3.8 Cytokines not altered by BθOM T_{reg} depletion.

(A-B) Cytometric bead array used to quantify (A) TNF-α (n≥10, 3 experiments), IL-2 (n≥10, 3 experiments), IL-4 (n≥10, 3 experiments), and IL-10 (n≥9, 3 experiments) and (B) IFNγ (n≥9, 3 experiments), IL-6 (n≥9, 3 experiments), IL-17A (n≥9, 3

experiments), TNFα (n≥9, 3 experiments), IL-2 (n≥9, 3 experiments), IL-10 (n≥9, 3 experiments), and IL-4 (n≥9, 3 experiments) after cells were isolated from the (A) mesenteric lymph node and (B) spleen and were stimulated with PMA for 5 hours. Student's t test: not significant.

the colon lamina propria and mLN differentiated into T_H1 cells (Figure 3.9A). BθOM-FoxP3-DTR T cells can also differentiate into T_H17 cells; however, variable levels of T_H17 induction were observed between experiments (Figure 3.9B-C). These findings are a direct demonstration that symbiont-specific $CD4^+$ T cells can develop into both T_{effs} and T_{regs} and that these T_{regs} can self-regulate.

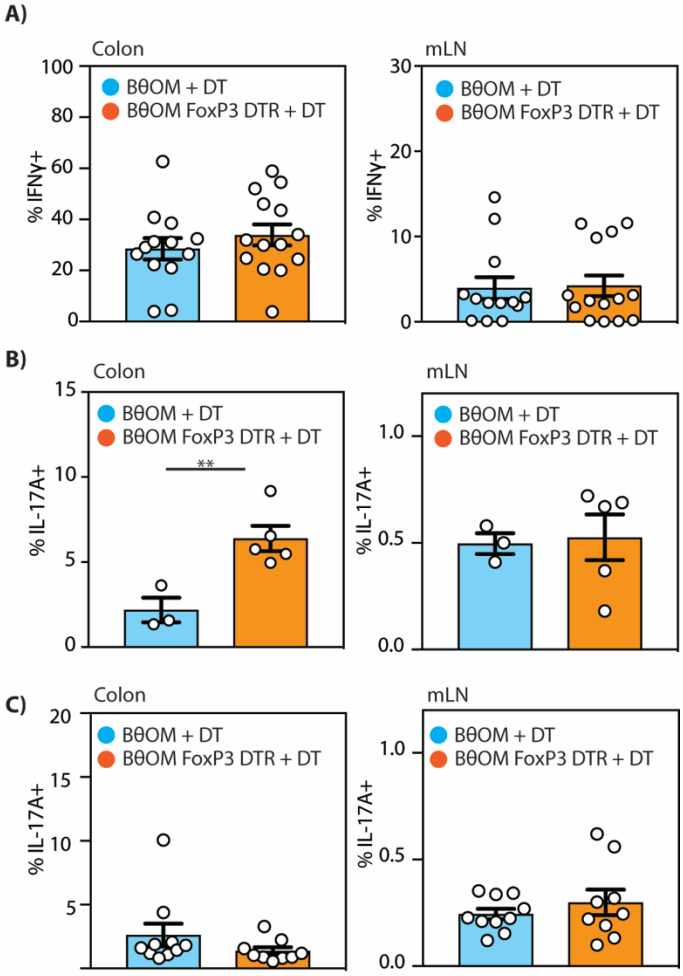


Figure 3.9 BθOM T cells primarily differentiate into T_H1 cells in vivo in the colon lamina propria and mLN.

*The percentage of cells isolated from the colon and mLN and stimulated with PMA and ionomycin for 5 hours that were (A) IFNγ+ (n=27 mice, 2 experiments) or (B-C) IL-17A+. Both IL-17A+ replicates are shown with 1 replicate in (B) n=8, 1 experiment and 1 replicate in (C) n=19, 1 experiment). Student's t test: (B) **P=0.0097*

The antigen recognized by BθOM T cells, BT4295, is expressed in a PUL

To elucidate how diet could affect a bacterial antigen expression, we needed to identify the antigen recognized by BθOM T cells. To identify this *Bt* antigen, we used positive functional fractionation, mass spectrometry, and a loss-of-function screen. Using *Bt* OMVs as the starting

material, we performed a T cell activation assay from 20 fractions of isolated proteins separated on the basis of molecular weight (Figure 3.10A and Figure 3.11A). We found

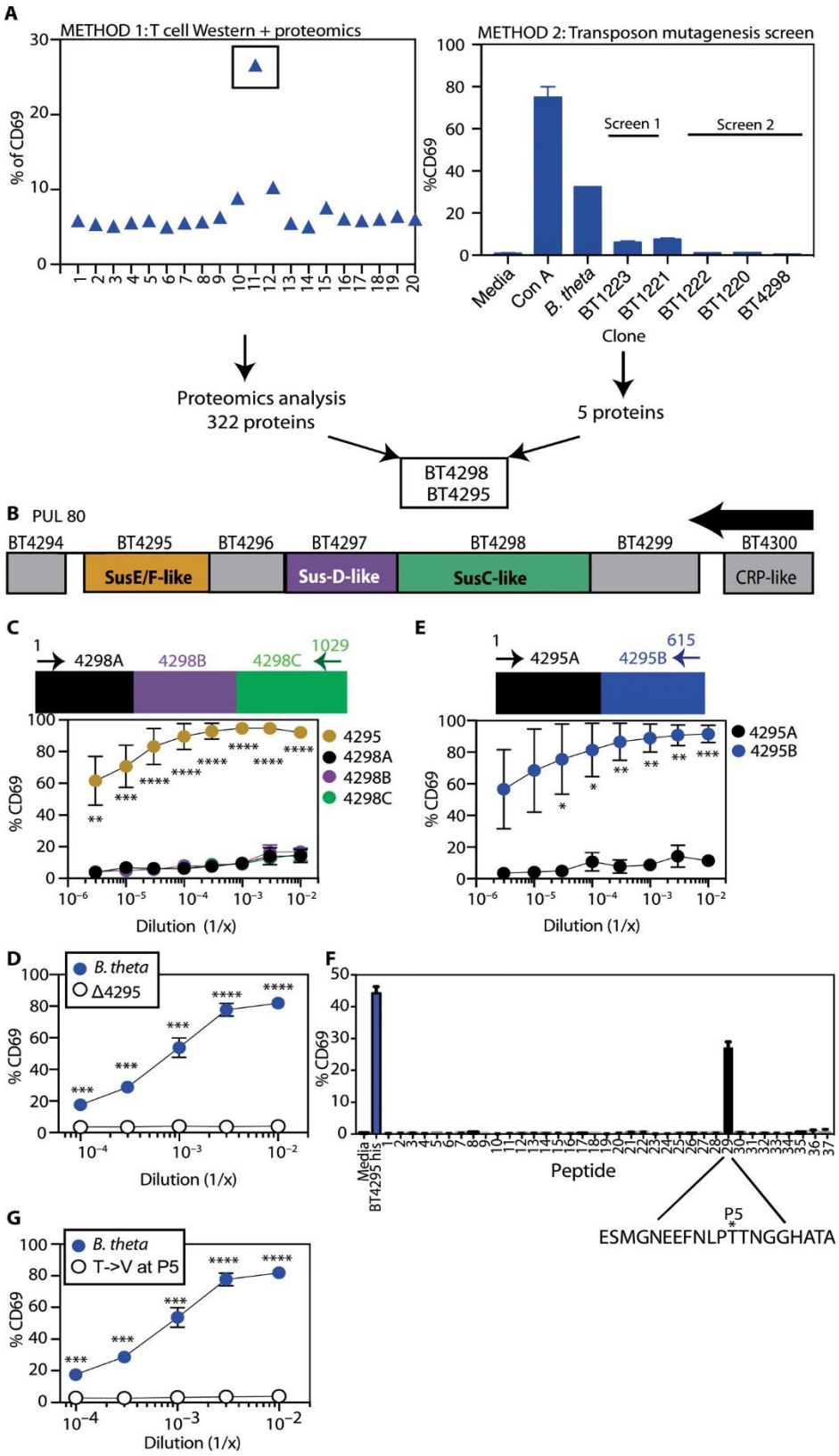
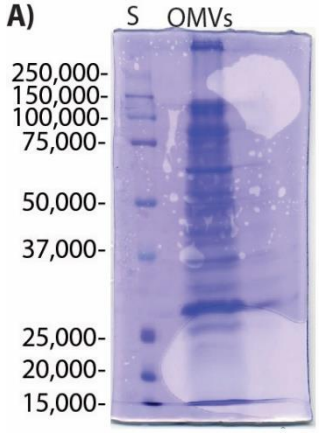


Figure 3.10 BθOM T cells specifically recognize the BT4295$_{(541–554)}$ epitope.

*(A) Two parallel methods, T cell Western with proteomics (left) and transposon mutagenesis (TM) screen[20] (right), used to identify the antigen that stimulates BθOM T cells. (B) Schematic of the PUL80 affected by BT4298 disruption by TM. The arrow represents the direction of transcription. (C to G) Percentage of CD69 expressing BθOM T cells after culture with BMDM loaded with (C) E. coliexpressing the full-length BT4295 (n = 3, three experiments for each dilution) or three consecutive segments of BT4298 (BT4298A, BT4298B, and BT4298C) (n = 3, three experiments for each dilution), (D) Bt (n = 4, four experiments) or Δ4295 (n = 4, four experiments), or (E) E. coli expressing two consecutive segments of BT4295 (BT4295A and BT4295B) (n = 3, three experiments for each dilution). (F) Synthetic 20-amino acid peptides overlapping by 12 amino acids. The asterisks represent the P5 position. (G) Bt (n = 4, four experiments, same data as Figure 3.2E or Δ4295 (n = 3, three experiments). One-way ANOVA analysis: (C) **P < 0.01, ***P < 0.001, and ****P < 0.0001. Means with asterisks are significantly different by Tukey's multiple comparisons test. Student's t test: (D) ***P < 0.001 and ****P < 0.0001; (E) *P < 0.1, **P < 0.01, and ***P < 0.001; (G) ***P < 0.001 and ****P < 0.0001.*

a single fraction of *Bt* OMV proteins that stimulated BθOM T cells (Figure 3.10A). Mass spectrometry analysis of this fraction identified 322 distinct proteins (Figure 3.10A). To refine the list of potential antigens, we generated a *Bt* transposon insertion library[32] and screened individual clones using the *in vitro* T cell activation assay for BθOM T cells (Figure 3.10A). In a screen of 2300 clones, we identified five genes that, when knocked out, no longer stimulated BθOM T cells (Figure 3.10A and Figure 3.11B). One of the five *Bt* gene candidates (BT4298) was identified in the mass spectrometry analysis (Figure 3.10A). The other four hits were all in one additional unlinked locus (BT1220-23) containing genes encoding enzymes in the pentose phosphate pathway.



B)

Gene ID	Annotation	Pathway
BT1220	6-phosphogluconolactonase	PPP
BT1221	Glucose-6-phosphate dehydrogenase	PPP
BT1222	6-phosphogluconate dehydrogenase	PPP
BT1223	Dicarboxylate symporter	Transporter
BT4298	SusC homolog	Sus like system

PPP= pentose phosphate pathway

Figure 3.11 Identification of the epitope recognized by BθOM T cells.

(A) Bt OMVs were separated by molecular weight on a 10% SDS PAGE gel. Lane S is the molecular weight ladder. (B) The five genes identified when screening 2,300 individual clones (n=2 experiments) of a Bt transposon insertion library.

Expression in *Escherichia coli* of the BT4298 protein identified in both the mass spectrometry and transposon library, unexpectedly, did not stimulate BθOM T cells (Figure 3.10C). However, many bacterial genes are organized into cotranscribed operons, and this is likely to be true for *Bt*. For example, the BT4294-4300 PUL was previously shown to be coordinately activated in response to mucus *O*-linked glycans[26,33]. We therefore reasoned that the

transposon insertion in the BT4298 gene exerts loss-of-function effects on downstream genes due to polarity (Figure 3.10B), including BT4295, which was also identified in our mass spectroscopy analysis (Figure 3.10A). Expression of BT4295 in *E. coli* resulted in strong stimulation of BθOM T cells (Figure 3.10C), demonstrating that the BT4295 was the antigen recognized by BθOM T cells. BT4295 is predicted to be a SusE/SusF lipoprotein that is ultimately trafficked to the OM, including OMVs (Figure 3.12A)[26]. We confirmed that BT4295 was the only antigen recognized by BθOM T cells by generating an in-frame deletion mutant of BT4295 that disrupted its expression (BTΔ4295) and abolished its ability to stimulate BθOM T cells (Figure 3.10D).

To identify the epitope in BT4295 recognized by BθOM T cells, we expressed amino and carboxyl halves of the protein in *E. coli* (Figure 3.10E). We found that the carboxyl half of the protein activated BθOM T cells (Figure 3.10E). We then generated overlapping 20-mer peptides for the entire carboxyl half of BT4295 and tested them for their ability to activate BθOM T cells. A single peptide (536 to 555) stimulated BθOM T cells (Figure 3.10F). The antigenic epitope was further defined to be the highly stimulatory 14-mer (541 to 554) (EEFNLPTTNGGHAT), which contains a strong predicted I-A^bbinding motif (P1 = F543) (Figure 3.12B). We identified

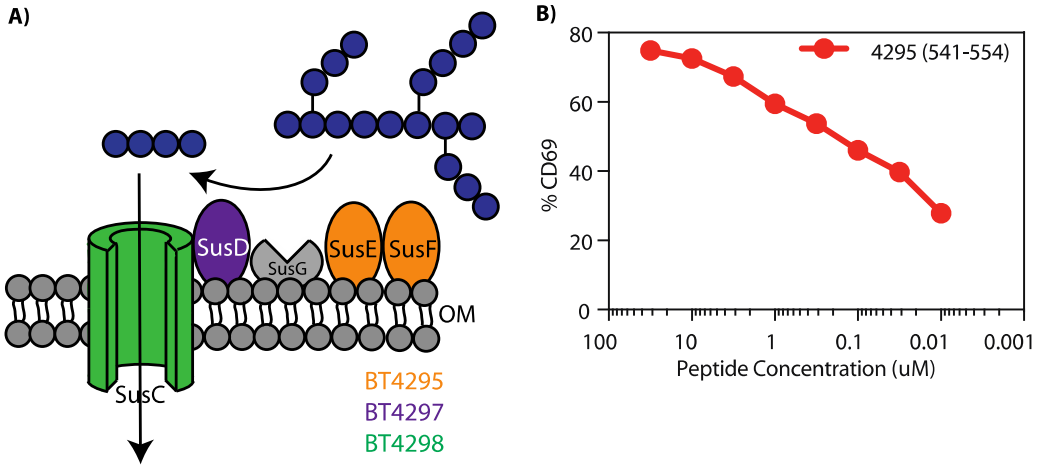


Figure 3.12 BθOM T cells recognize BT4295(541-554) and schematic of the BT4295 PUL.

(A) Schematic of the BT4295 PUL with BT4295 represented as the SusE/SusF proteins. (B) The percentage of CD69⁺ BθOM T cells after culture with BMDMs treated with BT4295(541-554) (n=2, 1 experiment).

the threonine at the P5 position (T547) to be critical for TCR recognition and generated a point mutation at the P5 position (a threonine to a valine substitution, T547V) that resulted in the complete loss of BθOM T cell activation (Figure 3.10G). Together, these findings demonstrate that BθOM T cells strongly and specifically recognize a single peptide epitope ($BT4295_{541–554}$) in

the BT4295 protein, which is expressed in the *Bt* OM in response to mucin-type O-glycan (MOG) cues.

Expression of BT4295 is regulated by available nutrients

Having identified BT4295 as the antigen recognized by BθOM T cells, we determined how specific nutrients altered its expression. On the basis of the differential ability of *Bt* grown in TYG versus in mTYG media to stimulate BθOM T cells (Figure 3.2C), we asked whether removing specific components (Table 3.1) from the TYG media or adding them to the mTYG media would alter the stimulatory ability of *Bt* grown in these modified media.

Table 3.1 Composition of TYG vs. mTYG media

Media Component	TYG	mTYG
Tryptone	10g/L	20g/L
Yeast Extract	5g/L	10g/L
D-glucose	4g/L	5g/L
KH_2PO_4	100mM	0.294mM
K_2HPO_4		0.23mM
$(NH_4)_2SO_4$	8.5mM	
NaCl	15mM	1.4mM
$CaCl_2•2H_2O$	0.0072mM	0.068mM
$FeSO_4•7H_2O$	0.00263mM	
$MgCl_2$	0.1mM	
$MgSO_4•7H_2O$		0.078mM
$NaHCO_3$		0.024mM
Hematin	0.0019mM	0.0079mM
Vitamin K_3	0.01mM	
Vitamin B_{12}	0.00000369mM	
L-histidine	0.2mM	
L-cysteine	0.413mM	8.25mM
Resazurin		0.004mM

Individually removing vitamin B_{12}, vitamin K_3, histidine, cysteine, $FeSO_4$, or $MgCl_2$ from TYG media had no effect on the ability of *Bt* to stimulate BθOM T cells (Figure 3.13A-B). However, when we removed salts [KH_2PO_4, $(NH_2)_4SO_4$, and NaCl] from TYG, *Bt* grown in this altered media no longer stimulated BθOM T cells (Figure 3.13A-B). Because removing salts

A)

Media Component	T stimulation*	
	Removed from TYG	Added to mTYG***
Salts**	-	+
Vitamin B12	+	-
Vitamin K3	+	-
L-histidine	+	-
FeSO4	+	-
MgCl2	+	-

*T cell stimulation was defined as negative (-) when % CD69 was <20% and positive (+) when >65%.
**TYG (KH2PO4, 13.6g/L; (NH2)4SO4, 1.125g/L; NaCL, 0.875g/L);
***mTYG contains a minimal concentration of salts:
(KH2PO4, 0.04mg/L; K2HPO4, 0.04mg/L; NaCL, 0.08mg/L)

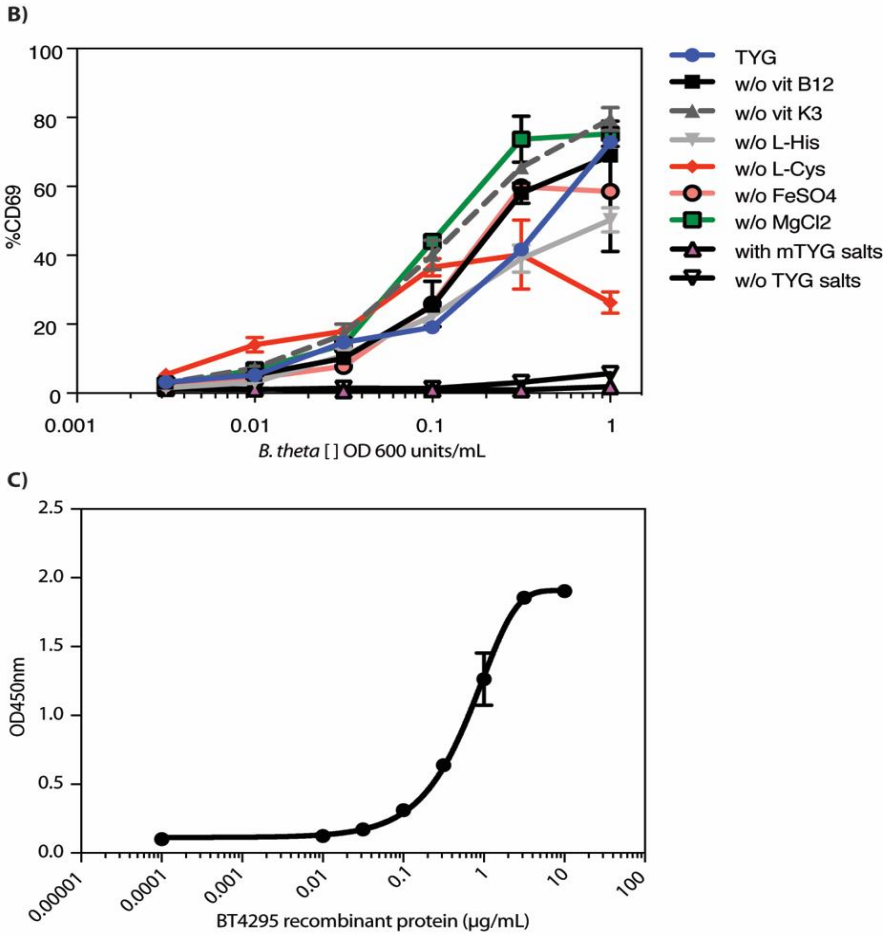


Figure 3.13. The effect of various nutrients on BθOM T cell activation.

(A) T cell stimulation results after various media components were removed from TYG media or added to mTYG media. (B) A representative plot of the percentage of CD69 expressing BθOM T cells after culture with BMDMs treated with the Bt grown in

TYG, TYG lacking various individual media components, TYG without salts, and TYG with mTYG salts (n=6, 3 experiments). (C) A representative plot of the BT4295 protein standard curve used in the BT4295 ELISA.

from the TYG media did reduce *Bt* growth to some extent, we also tested the addition of these salts to mTYG media that contained notably lower concentration of salts (KH_2PO_4, K_2HPO_4, and NaCl) (Figure 3.14A). Adding TYG salts to mTYG media resulted in a significant increase in BθOM T cell activation (Figure 3.14A). The ability of *Bt* grown in TYG, mTYG, and mTYG with TYG salts to stimulate BθOM T cells directly correlated with the level of BT4295 protein expression as determined by a quantitative enzyme-linked immunosorbent assay (ELISA; Figure 3.14B and Figure 3.13C).

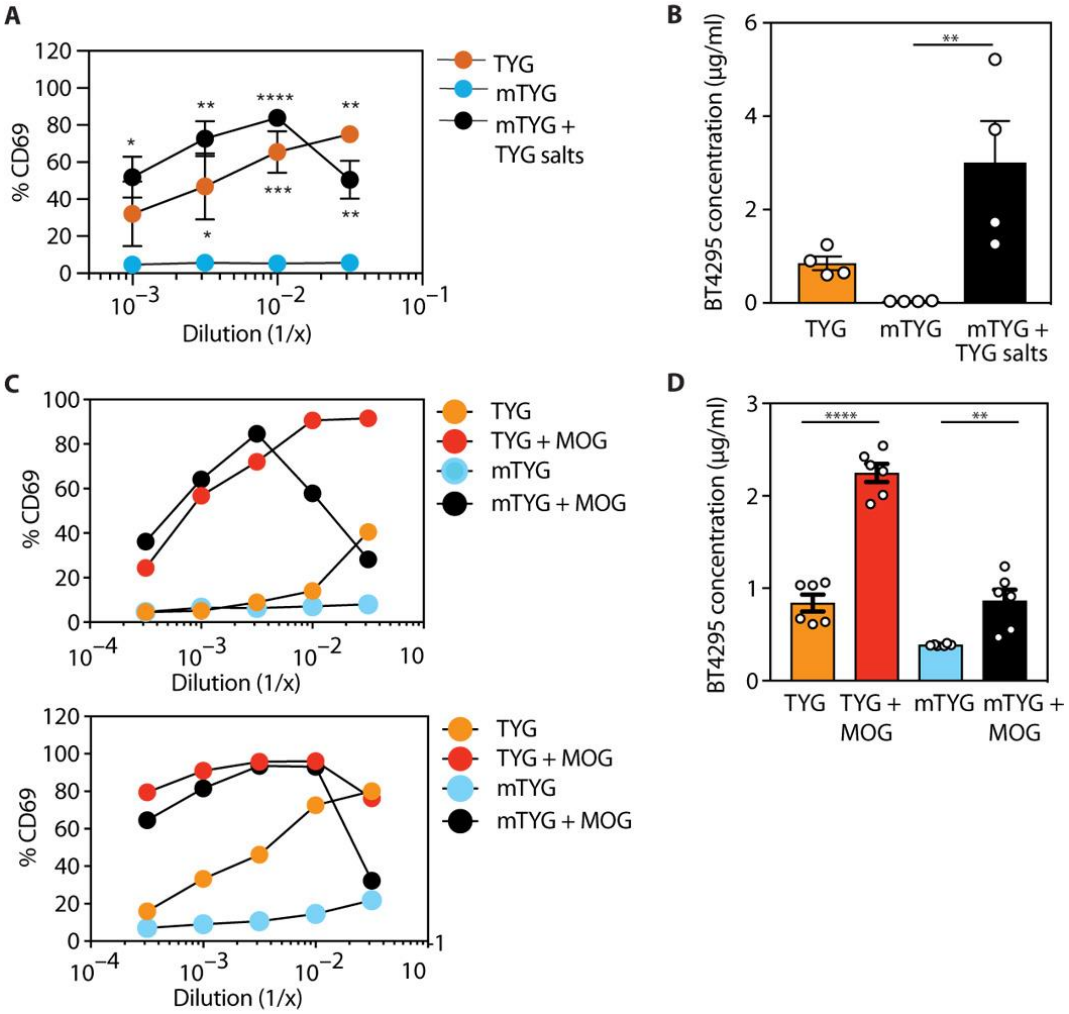


Figure 3.14 Salt and glycan regulate BT4295 expression and alter BθOM T cell activation.

(A) Percentage of CD69 expressing BθOM T cells after a 24-hour culture with BMDM loaded with Bt grown in mTYG (n = 4, four experiments), TYG (n = 2, two experiments), and mTYG supplemented with TYG salts (n = 4, four experiments). (B) The concentration in microgram per milliliter of BT4295 protein expressed in Bt grown in TYG, mTYG, and mTYG supplemented with TYG salts (n = 4, four experiments) as determined by a quantitative ELISA. (C) Percentage of CD69 expressing BθOM T cells after a 24-hour culture with BMDM loaded with Bt grown in mTYG, TYG, mTYG supplemented with MOG and TYG supplemented with MOG (n = 2, two experiments). (D) The concentration in microgram per milliliter of BT4295 protein expressed in Bt grown in mTYG, TYG, mTYG supplemented with MOG and TYG supplemented with MOG (n = 3, three experiments) as determined by a

*quantitative ELISA. One-way ANOVA analysis: (A) *P < 0.1, **P < 0.01, ***P < 0.001, and ****P < 0.0001; (B) **P = 0.0093; (D) ****P < 0.0001 and **P = 0.0065. Means with asterisks are significantly different by Tukey's multiple comparisons test.*

Previous transcriptional analysis showed that, in the absence of dietary glycans, *Bt in vivo* increases the expression of the BT4294-4300 PUL likely to break down endogenous mucin glycans, which is supported by *in vitro* expression of this PUL in response to purified mucin glycans[34,35]. Therefore, we tested whether growing *Bt* in mTYG with porcine MOG would increase the expression of BT4295 and drive BθOM T cell activation. We found that *Bt* grown in mTYG supplemented with MOG now strongly activated BθOM T cells (Figure 3.14C) and led to increased BT4295 protein expression (Figure 3.14D). Thus, BT4295 expression can be up-regulated by MOG in mTYG media, which alone did not induce expression. Together, these findings demonstrate that, by changing available nutrients (salts or glycans), the expression of a specific symbiont-derived antigen can be markedly affected.

Glucose catabolically represses BT4295

The four transposon mutant hits in the pentose phosphate pathway that significantly decreased expression of BT4295 (Figure 3.10A) implicated glucose metabolism as another potential regulator of BT4295 expression. To test the involvement of glucose on the regulation of BT4295 expression, we eliminated glucose from the TYG and mTYG media (Table 3.1). *Bt* grew in both media in the absence of glucose, but at slightly reduced rates. We found that BθOM T cells were now stimulated by *Bt* grown in mTYG in the absence of glucose (Figure 3.15A). Similarly, *Bt* grown in TYG without glucose also stimulated BθOM T cells, even stronger than in the presence of glucose (Figure 3.15A). Thus, glucose appeared to be acting as a repressor of BT4295 expression. Catabolite repression is a well-established regulatory process in bacteria, including *Bt*, in which other metabolic pathways are repressed in the presence of glucose or other high-priority nutrients[36,37]. Using a quantitative ELISA for BT4295 protein, we tested whether the increase in stimulatory ability of *Bt* grown in the absence of glucose was due to increased BT4295 protein expression. Removing glucose from the mTYG media resulted in a 14.5-fold increase in the expression of BT4295, and removing it from TYG media resulted in a 4-fold increase (Figure 3.15B). This finding again shows a direct correlation between the level of BT4295 protein expression and the ability to stimulate BθOM T cells, providing proof that glucose is acting as a repressor of BT4295 expression. From these findings, we conclude that, in

the presence of glucose, *Bt* shuts down the expression of the BT4294-4300 PUL, thereby reducing production of the BT4295 antigen.

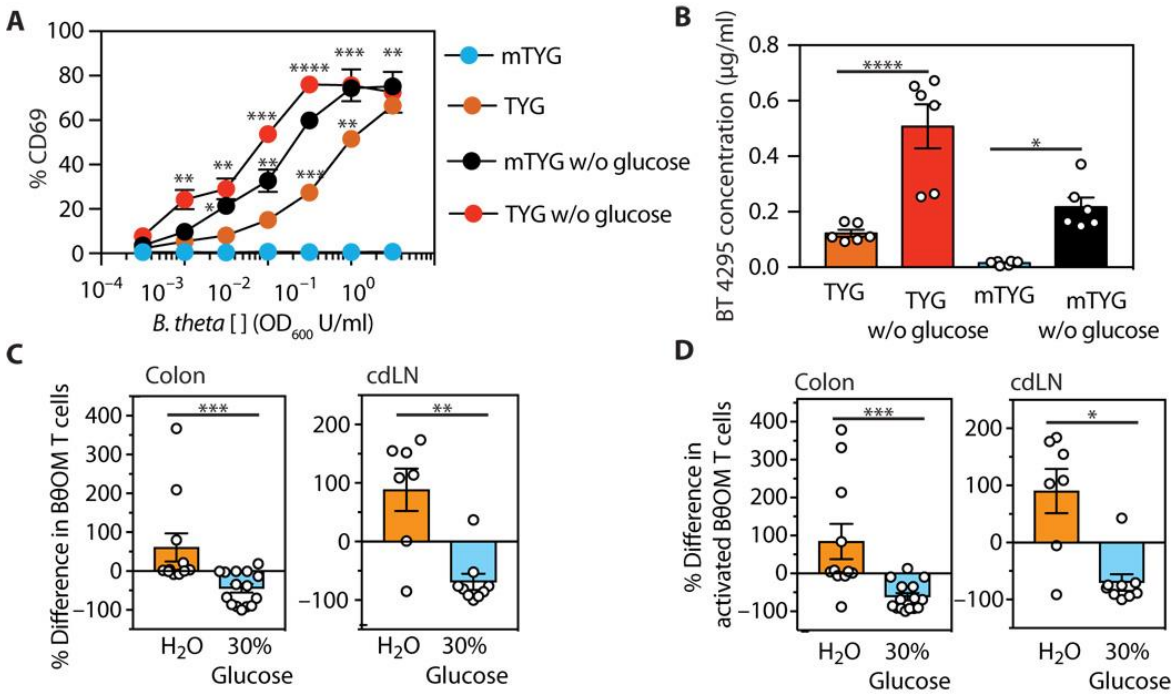


Figure 3.15 Dietary glucose represses BT4295 expression, decreasing the activation of BθOM T cells in vivo.

*(A) Representative plot of the percentage of CD69 expressing BθOM T cells after culture with BMDM loaded with Bt grown in TYG and mTYG media with or without glucose (n = 6, three experiments). (B) The concentration in microgram per milliliter of BT4295 protein expressed in Bt grown in TYG and mTYG media with or without glucose (n = 6, three experiments). The percent difference in the number of (C) CD4⁺CD45.1⁺ BθOM T cells or (D) CD4⁺CD45.1⁺CD44⁺CD62L⁻ activated BθOM T cells in the colon (n = 26, x = 3 experiments) and cdLN (n = 16, two experiments) of Bt–colonized mice given water or 30% glucose water and adoptively transferred with 200,000 CD4-enriched BθOM T cells. (C and D) The percent difference was calculated from the mean of each experiment. ANOVA multiple comparison analysis: (A) *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001; (B) ****P < 0.0001 and *P = 0.0190. Means with asterisks are significantly different by Tukey's multiple comparisons test. Mann-Whitney test for non-normally distributed data: (C) ***P = 0.0002 and **P = 0.0052; (D) ***P = 0.0002 and *P = 0.0115.*

Dietary glucose decreases the stimulation of BθOM T cells in vivo

We next determined whether exogenous glucose affected the ability of BθOM T cells to be stimulated *in vivo* by decreasing BT4295 expression. We added 30% glucose to the drinking water of recipient mice and maintained them on the standard chow throughout the course of the experiment. The addition of 30% glucose to the drinking water had no effect on *Bt* colonization levels (Figure 3.16A). The number of BθOM T cells in the colon and cdLN markedly decreased in the recipient mice fed 30% glucose drinking water (Figure 3.15C). Although there was no difference in T_{regs} (Figure 3.16B), the number of activated BθOM T cells was also decreased (Figure 3.15D). Thus, with a high-glucose diet, BT4295 antigen expression is decreased,

resulting in weaker stimulation of the BθOM T cells. This finding establishes that diet can affect the expression of a specific symbiont antigen and modulate a $CD4^+$ T cell response *in vivo.*

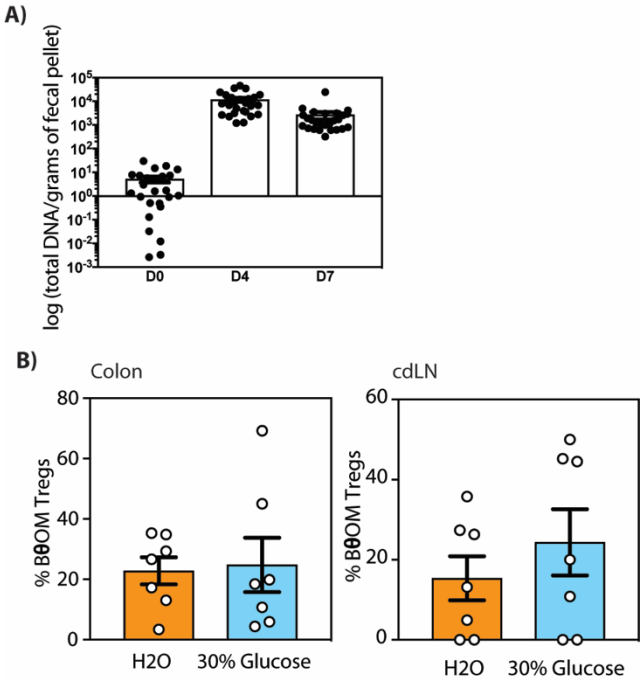


Figure 3.16 The addition of 30% glucose to the drinking water has no effect on Bt colonization or Treg differentiation.

(A) Colonization levels of Bt (total DNA of Bt/gram of fecal matter) on days 0, 4, and 7and the (B) percentage of FoxP3+ T_{regs}in the colon and cdLN (n=14, 2 experiments) in Bt gavaged $Rag1^{-/-}$mice transferred with BθOM T cells and given regular water or water with 30% glucose.

Discussion

We developed a symbiont-specific T cell model to study how diet could affect the interactions between a symbiont and the host immune system. We show that BθOM T cells respond to *Bt* and OMVs but not to other *Bacteroides* family members. Next, we identified BT4295, a SusE/F homolog, as the BθOM T antigen. Transfer of BθOM T cells into *Bt*–colonized $Rag1^{-/-}$ mice showed that antigen-specific T cells differentiate into T_{regs} and T_{effs}. Upon depletion of BθOM T_{regs}, the BθOM T_{effs} cause colitis. We show that the expression of BT4295 can be altered by glycans, salts, and glucose. A high-glucose diet reduced activation of the BθOM T cells, making BT4295 a nutrient-sensitive antigen able to alter T cell responses to microbes. This study definitively shows that diet can play a role in altering antigen expression thereby affecting immune responses.

TCR transgenic models have been previously developed to study antigen-specific responses to gut microbes. T cells specific for segmented filamentous bacteria (SFB) in the small

intestine have revealed how symbiotic microbes contribute to driving organ-specific autoimmunity[23]. The CBir1 TCR transgenic mice are widely used to study antigen-specific microbial interactions[21]; however, CBir1 T cells do not recognize their antigen during homeostasis despite the abundance of microbial antigen in the lumen[38]. More recently, *Helicobacter* species–specific transgenic T cells were shown to respond differently during homeostasis and mucosal injury/inflammation[19,29]. In all of these cases, microbial antigens were not shown to cross the epithelial barrier except in the context of inflammation. Therefore, we developed a symbiont-specific T cell that responds to *Bt* and OMVs, a relevant source of antigen that crosses the colonic epithelium and interacts with the host immune system during homeostasis[25,39].

Although our study focused on a single T cell and its cognate antigen, this approach is likely relevant because of the concept of immunodominance. Despite a theoretically large number of potential microbial epitopes, which can be recognized by $CD4^+$ T cells, the immune system generally focuses on a few immunodominant epitopes. As one example, the $CD4^+$ T cell response in mice to SFB focuses on two dominant antigens of this microbe[23]. We propose that the TCR we identified in this study may be specific for a dominant *Bt* antigen.

Our data directly show the conversion of a naive *Bt*–specific T cell into T_{regs}. Using the diphtheria toxin receptor (DTR) system, we deplete *Bt*–specific T_{regs} and show that, in the absence of these cells, symbiont-specific T cells cause colitis. To determine the mechanism of T_{reg} induction, we identified the antigen driving T cell activation. Previous reports on *B. fragilis*identified capsular polysaccharides on OMVs that induce T_{regs}[40], suggesting that bacterially derived polysaccharides have immunomodulatory effects on the host immune system. Our study extends the types of *Bacteroides* antigens that can participate in T cell development, including induction of T_{regs}.

One potential factor we have not controlled for is a direct effect of glucose on T cells. There is significant literature showing that glucose enhances T cell responses[31,41,42]. To our knowledge, there are no reported studies showing that increased glucose *in vivo* would decrease T cell responses or homeostatic proliferation. Although we cannot definitively rule out that increased glucose *in vivo* was directly inhibiting BθOM T cells, the literature supports our conclusion that increased dietary glucose caused a decrease in T cell proliferation due to a direct effect on BT4295 protein expression.

Inflammatory bowel disease (IBD) involves a potentially definable number of chronically activated T cells and microbial antigen specificities. We now show that specific TCR/cognate antigen pairs can be modulated by altering dietary components to affect gene expression of such a key microbial antigen. Future work developing additional TCR/antigen systems from other symbionts, including those that are enriched in patients with IBD, will be valuable to test whether this paradigm established with *Bt* can be extended to other key microbial antigens. If glucose repression or salt stimulation of dominant microbial antigens is widespread, then such dietary manipulations may become effective for therapy.

Methods

Study design.

The objective of this study was to generate a *Bt*–specific T cell system (BθOM T cells) to identify the interactions between the immune system and an antigen expressed on a highly prevalent colonic symbiont and determine the role that diet plays in altering those interactions. We designed and performed experiments in cellular immunology, protein biochemistry, and mass spectrometry. The number of independent experiments is outlined in the figure legends.

Mice.

All experimental procedures were performed under approval by Washington University's Animal Studies Committee. Mice were housed in an enhanced specific pathogen–free facility. BθOM transgenic mice on the *Rag1*$^{-/-}$ background were maintained by breeding to a nontransgenic *Rag1*$^{-/-}$ mouse. BθOM-FoxP3-DTR mice were generated by breeding BθOM transgenic mice with FoxP3-DTR mice[30].

Generation of the BθOM transgenic mouse.

Bt was grown to confluence and washed with PBS. C57BL/6J mice were immunized subcutaneously in the rear footpads with *Bt* mixed with incomplete Freund's adjuvant (IFA; Difco) in a 1:1 ratio. One week later, draining popliteal lymph nodes were harvested and stimulated *in vitro* with *Bt* for 3 days. Stimulated T cells were fused following a standard protocol. Hybridomas were selected for responsiveness to *Bt* presented by IFN-γ–stimulated

BMDMs. The BθOM clone was selected for further analysis, and its TCR genes were sequenced and cloned into TCR expression vectors (***43***). TCRα and TCRβ constructs were co-injected into C57BL/6J pronuclei in the Washington University Department of Pathology and Immunology's Transgenic Core Facility. Transgenic mice were identified by polymerase chain reaction (PCR) amplification of the Vα1 and Vβ12 transgenes from tail DNA (Vα1 forward primer GTTTCCAAGCAGGTGTGAGGAG and reverse primer CAAAACGTACCAGGGCTTACC; Vβ12 forward primer CTTCTCTTCTAGGTGATGCTG and reverse primer CCCAGCTCACCGAGAACAGTC).

Antibodies and reagents.

The following reagents were purchased: CD62L (MEL-14) and CD45.1 (A20) (BD Biosciences); CD4 (GK1.5), CD69 (H1.2F3), CD45.1 (A20), CD44 (IM7), CD25 (PC61), CD45.2 (104), CD25 (PC61), Vβ12 (MRII-I), and Mouse $T_H1/T_H2/T_H17$ Cytometric Bead Array Kit (BioLegend); CD25 (eBio3C7), CD4 (RM4-5), FoxP3 (FJK-16 s), IFN-γ (XMG1.2), and IL-17A (TC11-18H10.1) (eBiosciences); CellTrace CFSE Cell Proliferation Kit and LIVE/DEAD Fixable Blue Dead Cell Stain Kit (Life Technologies); deoxyribonuclease 1 from bovine pancreas grade II (Roche); and collagenase from *Clostridium histolyticum* (Sigma). Homemade cocktail antibodies for negative selection of $CD4^+$ T cells were purchased: anti-mouse Ter-119, CD11c (clone N418), CD11b (M1/70), CD8α (53-6.7), CD19 (1D3), and CD45R/B220 (RA3-6B2) (Tombo); CD49b (DX5) and CD24 (M1/69) (BioLegend); anti-biotin microbeads (Miltenyi Biotec).

Media recipes.

The following components of TYG medium were purchased: tryptone (10 g/liter) and yeast extract (5 g/liter) (BD Bacto); D-glucose (4 g/liter), 100 mM KH_2PO_4, 8.5 mM $(NH_2)_4SO_4$, 15 mM NaCl, 10 μM vitamin K_3, 2.63 μM $FeSO_4•7H_2O$, 0.1 mM $MgCl_2$, 1.9 μM hematin, 0.2 mM L-histidine, 3.69 nM vitamin B_{12}, and 413 μM L-cysteine (Sigma); 7.2 μM $CaCl_2•2H_2O$ (Mallinckrodt). *mTYG medium:* The following components of mTYG medium were purchased: tryptone (20 g/liter) and yeast extract (10 g/liter) (BD Bacto); D-glucose (5 g/liter), 8.25 mM L-cysteine, 78 μM $MgSO_4•7H_2O$, 294 μM KH_2PO_4, 230 μM K_2HPO_4, 1.4 mM NaCl, 7.9 μM hemin (hematin), 4 μM resazurin, and 24 μM $NaHCO_3$ (Sigma); 68 μM $CaCl_2•2H_2O$ (Mallinckrodt).

Preparation of OMVs.

Bt OMVs were purified with multiple rounds of centrifugation and filtering[25].

Functional in vitro macrophage T cell assay.

BMDM was stimulated with IFN-γ at 2000 U/ml in I-10 medium [Iscove's Modified Dulbecco's Medium (IMDM) 10% fetal bovine serum, glutamine, and gentamicin] and plated on a 96-well plate at 1×10^5 cells per well. The cells were washed with PBS 24 hours later and kept in 100 μl of fresh I-10 medium without IFN-γ for another 24 hours. A total of 5×10^5 splenocytes or 1×10^5 isolated BθOM $CD4^+$ T cells were added per well in 50 μl with 50 μl of half log dilutions of Bacteroidaceae strains and OMV. Bacteroidetes were grown in a 5-ml TYG or mTYG culture at 37°C overnight to mid-log phase. Cultures were washed twice with PBS and resuspended in medium before adding to the assay. Twenty-four hours later, the supernatant containing the T cells was transferred to a fresh 96-well plate and spun down at 1200 rpm. The cells were washed with fluorescence-activated cell sorting buffer and stained for CD69 expression.

In vivo experiments.

Bacterial stocks: Bacteroidetes were grown anaerobically from single isolates in standing culture in TYG at 37°C for 24 hours[33]. Each culture was concentrated by centrifugation, mixed with sterile, prereduced PBS and glycerol to a final concentration of 20% glycerol, and frozen at −80°C in single-use aliquots. *Gavage: Rag1*$^{-/-}$ mice were placed on antibiotics at 3 to 4 weeks of age for 3 to 4 weeks. Antibiotic treatment consisted of ciprofloxacin (0.66 mg/ml), metronidazole (2.5 mg/ml; Sigma), and sugar-sweetened grape Kool-Aid Mix (20 mg/ml; Kraft Foods) in the drinking water[44]. Mice were gavaged with 100 μl of antibiotic water on the first 2 days and the last 2 days of the 3- to 4-week duration. For the bulk of the experiments, mice were taken off antibiotic water and given Kool-Aid. For the *in vivo* glucose experiments, mice were taken off antibiotic water and given water or 30% glucose water. Two days later, mice were gavaged with 100 μl of *Bt* strains at a concentration of 1×10^8 colony-forming units/ml. Fecal pellets were obtained on days 0, 4, and 7 to determine colonization. *BθOM T cell transfer:* Three days after gavage, *Rag1*$^{-/-}$ mice were injected with BθOM T cells isolated from the peripheral lymph nodes (axillary, brachial, and inguinal), mLNs, and spleen. Cells were enriched by

negative selection using a homemade cocktail of antibodies (see reagents) and sorted for $CD4^+CD44^{lo}CD62L^{hi}CD25^-$ T cells. Cells (1×10^5 to 2×10^5) were injected retrorbitally. *Lamina propria dissociation*: Seven days after T cell transfer, mice were euthanized, and leukocytes were isolated from the lamina propria following the Lamina Propria Dissociation Kit protocol published by Miltenyi Biotec. *Peripheral tissue processing*: The cdLN and spleen were removed and processed using frosted microscope slides (Thermo Fisher Scientific). Samples were filtered through a 70-μm filter. *DT depletion of BθOM FoxP3$^+$ $T_{regs.}$ T_{reg} depletion*. Antibiotic treated *Rag1*$^{-/-}$ mice were gavaged with *Bt* and injected with enriched and sorted 1×10^5 BθOM-FoxP3-DTR or BθOM T cells. Intraperitoneal injections of DT (10 μg/kg) were performed on days 9, 11, and 13 after gavage. Depletion was confirmed by staining for T_{regs} on day 21 after gavage in mLNs and spleen. *Cytokines*: On day 21 after gavage, 5×10^4 mLNs and 2×10^6 splenocytes were stimulated with phorbol 12-myristate 13-acetate (PMA; 50 ng/ml) and ionomycin (500 ng/ml) for 5 hours at 37°C. $T_H1/T_H2/T_H17$ cytokines were quantified in the supernatant using the BD Cytometric Bead Array following the manufacturer's instructions. Supernatants from splenocyte samples were diluted 1:2. *T cell differentiation:* On day 24 after gavage, cells isolated from the colon lamina propria and mLN were stimulated with PMA (50 ng/ml) and ionomycin (500 ng/ml) for 1 hour at 37°C, Brefeldin A was added (5 μg/ml), and the cells were stimulated for four additional hours at 37°C. T_H1 and T_H17 cells were identified by intracellular staining with IFN-γ and IL-17A antibodies.

Tissue harvest, fixation, and preparation for histology.

Ceca and colons were fixed in methacarn fixative for 12 to 16 hours at 24°C. Samples were washed two times with 100% methanol for 30 min, followed by 100% ethanol for 20 min (two times), and then stored in 70% ethanol. Five-micrometer sections were stained with hematoxylin and eosin (H&E). Representative images of cecal histology were taken with an Olympus BX51 microscope. Blinded microscopic analysis for mitotic figures using H&E-stained histologic sections was performed at 20× magnification on well-oriented crypts as previously described[44].

Fecal bacterial DNA extraction and quantitative PCR amplification.

Fecal bacterial DNA extraction and quantitative PCR amplification were performed according to a previously published protocol[25,45].

T cell Western assay.

Bt OMV antigens were separated using a T cell Western blot assay as described[46]. Briefly, 500 μg of OMVs was separated on a 10% SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gel on both the left and right sides of the gel with molecular weight standards on both sides. For the left side, the proteins were transferred to nitrocellulose (each lane cut into 20 strips), dissolved in dimethyl sulfoxide, and precipitated with sodium carbonate/sodium bicarbonate. The nitrocellulose particles from each strip were tested for their ability to stimulate BθOM T cells using BMDM as antigen presenting cells (APCs). The corresponding position of the active fraction on the right side of the SDS-PAGE gel was further analyzed by mass spectrometry.

Proteomic analysis of OMVs.

Proteomic analysis of the corresponding T cell stimulatory SDS-PAGE fraction of OMVs from TYG-grown *Bt* was performed using standard procedures at MS Bioworks (Ann Arbor, MI). Briefly, the gel slices were digested with trypsin and analyzed by nano liquid chromatography–tandem mass spectrometry with a Waters NanoAcquity HPLC system interfaced to a Thermo Fisher Q Exactive. The data were searched using Mascot against the UniProt *Bt* reference proteome. Mascot DAT files were parsed into Scaffold for validation, filtering, and creation of a nonredundant list per sample, requiring at least two unique peptides per protein.

B. thetaiotaomicron transposon mutagenesis library and screen.

Transposon mutagenesis of *Bt* was performed as described previously[32]. Briefly, mutagenesis was carried out on an acapsular *Bt* strain (ΔCPS) lacking all capsular polysaccharide loci, which was previously characterized[37]. Here, we used the pSAM_Bt vector containing *mariner* transposon and an *ermG* cassette. S17 *E. coli* was used to deliver the vector through conjugative transfer into *Bt*. DNA isolation from selected mutants was performed using the Qiagen DNeasy Blood and Tissue Kit. Two-round PCR was performed to identify the transposon insertion site with the following conditions: round 1; 1 cycle at 95°C (3 min); 5 cycles at 95°C (30 s), 30°C (30 s), and 72°C (45 s); 32 cycles at 95°C (30 s), 55°C (30 s), and 72°C (45 s). The PCR reactions from step 1 were purified using the Qiagen PCR Purification

Kit, and 100 to 200 ng of product were used as a template for round 2; 1 cycle at 95°C (3 min); 35 cycles at 95°C (30 s), 55°C (30 s), 72°C (45 s). Reactions from round 2 were run on a 2% agarose-Tris-Borate-EDTA gel, and bands were extracted using the Qiagen Gel Extraction Kit. These products were then sequenced using the primers previously described[32].

The library was frozen in 96-well plates. The plates were thawed and spun down, and the medium was removed, washed once in 200 µl of PBS, and then suspended in 100 µl of complete medium. Ten microliters of each was screened using the *in vitro* macrophage T cell assay during the primary screen, and hits were retested in duplicate for conformation before sequencing. *Generation of the BT4295 mutant. BT4295* gene deletion and amino acid substitutions within this gene were done using allelic exchange as described previously[47]. Briefly, all manipulations were done in a Δ*tdk* strain background of *Bt* using the pExchange-tdk vector[48], and primers are listed in Table 3.2. All *Bacteroides* strains and mutants were grown in TYG medium or brain-heart infusion agar with 10% horse blood added. The following antibiotics were used as needed: gentamicin (200 µg/ml), erythromycin (25 µg/ml), and 5-fluoro-2′-deoxyuridine (200 µg/ml).

Table 3.2 BT4295 Primers

Primer	Sequence (5' to 3')	Use
BT4295 5' UpSal1	GCGGTCGACTGCCAAACTGCTTCCCGATGA	Deletion of BT4295
BT4295 3' Out	TTCTTCGTCAGTCTTTTCTTGTTTTACTTGATTTGATTACAAGT TATCTAC	Deletion of BT4296
BT4295 5' Out	GTAAAACAAGAAAAGACTGACGAAGAA	Deletion of BT4297
BT4295 3' DownXbal	GCGTCTAGAGACCAATGAATGCGGTTTCACCCT	Deletion of BT4298
pBT4295_T547V F (ACAàGTA)	TGAAGAATTCAATCTGCCGGTAACAAACGGTGGCCATGCC	T547V Amino Acid conversion of BT4295
pBT4295_T547V R (TGTàCAT)	GGCATGGCCACCGTTTGTTACCGGCAGATTGAATTCTTCA	T547V Amino Acid conversion of BT4296
BT4295_SPdel 5'UpSal1	GCGGTCGACTGCCAAACTGCTTCCCGATGA	Signal peptide deletion of BT4295 (this is the same primer sequence as BT4295 5'UpSal1)
BT4295_SPdel 3'Out	CAGGAACTACGAGCGTTTCTACTTTAAACTCATTTTTTTTCAT ACTCATTTGATTTGATTACAAGTTATCTACTCTTGGGT	Signal peptide deletion of BT4295

BT4295_SPdel 5'Out	GAGTTTAAAGTAGAAACGCTCGTAGTTCCTG	Signal peptide deletion of BT4296
BT4295_SPdel 3'DownXbal	GCGTCTAGACATCACGCAGTGCTCTTGAAGCGG	Signal peptide deletion of BT4297
BT4298 segment A	CATCATCACCACCATCACAGAAATAATTTTCTACTGATTGTA	Forward Primer
BT4298 segment A	GTGGCGGCCGCTCTATTATTGGCACGATAGGTTATTTTT	Reverse Primer
BT4298 segment B	CATCATCACCACCATCACAAAATAACCTATGGTGGCAATATT	Forward Primer
BT4298 segment B	GTGGCGGCCGCTCTATTATCGATACCAAAGTTGAGTT	Reverse Primer
BT4298 segment C	CATCATCACCACCATCACAAGAAACTCAACTTTGGTATCG	Forward Primer
BT4298 segment C	GTGGCGGCCGCTCTATTATATAACTGCAGTTAGAATTTAAG	Reverse Primer
BT4295 Full length	CATCATCACCACCATCACAAAATAACCTATCGTGCCAATATT	Forward Primer
BT4295 Full length	GTGGCGGCCGCTCTATTATATACTGCAGTTAAATGCCTAG	Reverse Primer
BT4295 segment A	CATCATCACCACCATCACAAAATAACCTATCGTGCCAATATT	Forward Primer
BT4295 segment A	GTGGCGGCCGCTCTATTATATTCGTACTCTTGAAGGTTATCT	Reverse Primer
BT4295 segment B	CATCATCACCACCATCACCCTCGTGAAGGAAAGATAACC	Forward Primer
BT4295 segment B	GTGGCGGCCGCTCTATTATATACTGCAGTTAAATGCCTAG	Reverse Primer

Generation of the BT4295 T->V mutant.

Construction of the T547V mutation was done using site-directed mutagenesis via overlapping PCR. Forward and reverse primers were synthesized containing the desired mutation, and outside primers were constructed to contain the entire *BT4295* gene. Once a verified construct was sequenced as containing the mutation, we followed a similar strategy to construct the deletion mutants (e.g., 4295 or SPdeletion). *E. coli* containing the T547V construct was mated with the BT4295 deletion strain, therefore complementing the *BT4295* gene back, but with a T547V mutation so that it no longer stimulated T cells.

Expression of BT4295 and BT4298 in E. coli.

To express BT4295 and BT4298 in *E. coli*, we used the Lucigen Expresso T7 Cloning and Expression System and followed the manufacturer's protocol. Briefly, we expressed BT4295 and BT4298 in the pETite N-His Kan vector and designed oligonucleotides for cloning full-length or partial proteins listed in Table 3.2.
Sequence-confirmed clones of each were transformed into BL21(DE3) *E. coli* and grown overnight at 37°C with shaking. Fresh 2-ml cultures were inoculated and grown to an OD_{600} (optical density at 600 nm) of 0.5, induced with 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG) and grown for 5 hours at 37°C with shaking, harvested by centrifugation, washed once with PBS, and suspended in 1 ml of PBS. Samples were heat-inactivated for 20 min at 95°C and then stored at 4°C until use.

Production of recombinant BT4295.

BT4295 was expressed in Pet-ite expression vector by cloning the sequence distal to the SPII cleavage motif and including a 5′6 His tag using the oligos CATCATCACCACCATCACTCGCCCGATTACGAAACCGAGTT (forward) and GTGGCGGCCGCTCTATTATATACTGCAGTTAAATGCCTAG (reverse)[49]. The construct was verified by sequencing and expressed in the *E. coli* strain BL21(DE3). Bacteria were grown at 37°C until mid-log phase growth was reached. The culture was induced with 1 mM IPTG and grown overnight at 19°C. Cells were collected by centrifugation, lysed [50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, lysozyme (1 mg/ml; HEL), and protease inhibitors at pH value of 8.0] for 30 min on ice, sonicated, and centrifuged to remove insoluble material. Supernatants were passed over a Qiagen NiNTA column, washed, and eluted in 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, and protease inhibitors at pH value of 8.0. Eluted material was buffer-exchanged into PBS with an Amicon Ultra 15 10-kDa concentrator to 1 to 2 ml of the final volume and quantified by absorbance at 280 nm (A_{280}).

Generation of monoclonal antibodies against BT4295.

C57BL/6J mice were immunized subcutaneously with 100 μg of recombinant protein (rBT4295) emulsified in complete Freund's adjuvant and boosted twice with 100 μg of rBT4295

in IFA every 4 weeks, followed by an intravenous (IV) boost of 50 μg rBT4295 3 days before harvest. Splenic B cells were fused with P3Ag8.6.5.3 myeloma cells to create hybridomas. Hybridomas were screened by ELISA against rBT4295, and positives were screened against whole *Bt* or OMV preparations to confirm specificity. Two clones (ERC-11 and 4E9) were selected for further characterization. They were subcloned by limit dilution, and both antibodies isotyped as IgG2b,κ. The antibodies were purified from culture supernatants on a Protein A–Sepharose column. Purified 4E9 was biotinylated using the Pierce Ez-Link Sulfo-NHS-SS-Biotin reagent following the manufacturer's protocol.

Quantitative ELISA for BT4295.

BT4295 protein levels in *Bt* samples were determined using a quantitative ELISA assay. Samples were obtained from equivalent numbers of *Bt* from OD_{600}-measured cultures. Bacteria were lysed in 100 mM CHAPS detergent (Sigma) and incubated with agitation for 1 hour at room temperature (RT). Insoluble material was removed by centrifugation, and samples were stored at 4°C. Purified anti-BT4295 antibody, ERC11, was coated on an Immulon 2 ELISA plate overnight in carbonate coating buffer [5 μg/ml (pH value of 9.6)] at 4°C. Plates were washed and blocked with buffer (PBS with 0.5% bovine serum albumin and 0.1% Tween 20) for 1 hour at RT. Plates were washed and samples were added for 2 hours at RT, washed again, and then, the anti-BT4295 antibody biotin-4E9 (5 μg/ml) was added for 1.5 hours at RT. Plates were washed again, and 1:5000 dilution of streptavidin horseradish peroxidase (SouthernBiotech) was added for 1 hour at RT. Plates were washed and developed with 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) to completion, and A_{405} was determined. Unknown sample concentrations were quantitated by comparison to a standard curve of rBT4295 performed in the same ELISA using GraphPad Prism software.

Statistical analysis.

Differences between two groups were evaluated using Student's *t* test (or Mann-Whitney test, for non-normally distributed data), and those among more than two groups were evaluated using analysis of variance (ANOVA) with Tukey's multiple comparisons test (or Kruskal-Wallis with Dunn's posttest for non-normally distributed data) using GraphPad software. *P* values of less than 0.05 were considered to be significant. Data are summarized as means ± SEM.

Notes

This work has been reprinted and modified with permission from Wegorzewska, M. M., Glowacki, R. W. P., Hsieh, S. A., Donermeyer, D. L., Hickey, C.A., Horvath, S.C., Martens, E. C., Stappenbeck, T. S., and Allen, P. M. Diet modulates colonic T-cell responses by regulating the expression of a *Bacteroides thetaiotaomicron* antigen. *Science Immunology* **4 (32),** eaau9079

References

1. Khalili, H. et al. The role of diet in the aetiopathogenesis of inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 525-535 (2018).

2. Kleinewietfeld, M. *et al.* Sodium chloride drives autoimmune disease by the induction of pathogenic T_H17 cells. *Nature* **496**, 518-522 (2013).

3. Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic T_{reg} cell homeostasis. *Science* **341**, 569-573 (2013).

4. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268-1273 (2012).

5. MacIver, N. J., Michalek, R. D. & Rathmell, J. C. Metabolic regulation of T lymphocytes. *Annu. Rev. Immunol.* **31**, 259-283 (2013).

6. Wei, J., Raynor, J., Nguyen, T. L. & Chi, H. Nutrient and Metabolic Sensing in T Cell Responses. *Front. Immunol.* **8**, 247 (2017).

7. Chang, P. V., Hao, L., Offermanns, S. & Medzhitov, R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc. Natl. Acad. of Sci. U.S.A.* **111**, 2247-2252 (2014).

8. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446-450 (2013).

9. Gao, Z. *et al.* Butyrate improves insulin sensitivity and increases energy expenditure in mice. *Diabetes* **58**, 1509-1517 (2009).

10. Kaiko, G. E. *et al.* The colonic crypt protects stem cells from microbiota-derived metabolites. *Cell* **167**, 1137 (2016).

11. Ji, J. *et al.* Microbial metabolite butyrate facilitates M2 macrophage polarization and function. *Sci. Rep.* **6**, 24838 (2016).

12. Lee, J. S. *et al.* AHR drives the development of gut ILC22 cells and postnatal lymphoid tissues via pathways dependent on and independent of Notch. *Nat. Immunol.* **13**, 144-151 (2011).

13. Cervantes-Barragan, L. *et al.* Lactobacillus reuteri induces gut intraepithelial $CD4^+CD8\alpha\alpha^+$ T cells. *Science* **357**, 806-810 (2017).

14. Devkota, S. *et al.* Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in $Il10^{-/-}$ mice. *Nature* **487**, 104-108 (2012).

15. Steed, A. L. *et al.* The microbial metabolite desaminotyrosine protects from influenza through type I interferon. *Science* **357**, 498-502 (2017).

16. Chang, Y. L. *et al.* A screen of Crohn's disease-associated microbial metabolites identifies ascorbate as a novel metabolic inhibitor of activated human T cells. *Mucosal Immunol.* **12**, 457-467 (2019).

17. Wilck, N. *et al.* Salt-responsive gut commensal modulates T_H17 axis and disease. *Nature* **551**, 585-589 (2017).

18. Kortman, G. A. *et al.* Low dietary iron intake restrains the intestinal inflammatory response and pathology of enteric infection by food-borne bacterial pathogens. *Eur. J. Immunol.* **45**, 2553-2567 (2015).

19. Chai, J. N. *et al.* Helicobacter species are potent drivers of colonic T cell responses in homeostasis and inflammation. *Sci. Immunol.* **2**, (2017).

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

21. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T. R. & Elson, C. O. A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl. Acad. of Sci. U.S.A.* **106**, 19256-19261 (2009).

22. Ivanov, II *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485-498 (2009).

23. Yang, Y. *et al.* Focused specificity of intestinal T_H17 cells towards commensal bacterial antigens. *Nature* **510**, 152-156 (2014).

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

25. 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).

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

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

28. Ohnmacht, C. *et al.* The microbiota regulates type 2 immunity through ROR γ^+ T cells. *Science* **349**, 989-993 (2015).

29. Xu, M. *et al.* c-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature* **554**, 373-377 (2018).

30. Kim, J. M., Rasmussen, J. P. & Rudensky, A. Y. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. *Nat. Immunol.* **8**, 191-197 (2007).

31. Chang, C. H. *et al.* Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell* **162**, 1229-1241 (2015).

32. Goodman, A. L. *et al.* Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279-289 (2009).

33. 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**, 36269-36279 (2006).

34. 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).

35. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).

36. Gorke, B. & Stulke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613-624 (2008).

37. Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol. Microbiol.* **88**, 876-890 (2013).

38. Hand, T. W. *et al.* Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. *Science* **337**, 1553-1556 (2012).

39. Bloom, S. M. *et al.* Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9**, 390-403 (2011).

40. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107-118 (2005).

41. Cham, C. M., Driessens, G., O'Keefe, J. P. & Gajewski, T. F. Glucose deprivation inhibits multiple key gene expression events and effector functions in $CD8^+$ T cells. *Eur. J. Immunol.* **38**, 2438-2450 (2008).

42. Ho, P. C. *et al.* Phosphoenolpyruvate is a metabolic checkpoint of anti-tumor t cell responses. *Cell* **162**, 1217-1228 (2015).

43. Ho, W. Y., Cooke, M. P., Goodnow, C. C. & Davis, M. M. Resting and anergic B cells are defective in CD28-dependent costimulation of naive $CD4^+$ T cells. *J. Exp. Med.* **179**, 1539-1549 (1994).

44. Kang, S. S. *et al.* An antibiotic-responsive mouse model of fulminant ulcerative colitis. *PLoS Med.* **5**, e41 (2008).

45. Nava, G. M. & Stappenbeck, T. S. Diversity of the autochthonous colonic microbiota. *Gut Microbes* **2**, 99-104 (2011).

46. Holsti, M. A. & Allen, P. M. Processing and presentation of an antigen of Mycobacterium avium require access to an acidified compartment with active proteases. *Infect. Immun.* **64**, 4091-4098 (1996).

47. Larsbrink, J. *et al.* A polysaccharide utilization locus from Flavobacterium johnsoniae enables conversion of recalcitrant chitin. *Biotechnol. Biofuels* **9**, 260 (2016).

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

49. Cameron, E. A. *et al.* Multidomain carbohydrate-binding proteins involved in Bacteroides thetaiotaomicron starch metabolism. *J. Biol. Chem.* **287**, 34614-34625 (2012).

Chapter IV

New Regulatory Strategies for *Bacteroides thetaiotaomicron* Polysaccharide Metabolism

Abstract

The ability of commensal gut *Bacteroides* to sense and respond to specific nutrients *in vivo* promotes survival in the competitive and complex milieu of the intestine and may allow individual species to partition to different niches. *Bacteroides* devote large portions of their genomes towards accessing and degrading complex carbohydrates via expression of genetic clusters termed polysaccharide utilization loci (PULs). Generally, each PUL targets a single polysaccharide for depolymerization and subsequent assimilation of the released sugars. PULs are regulated locally through the recognition of one or more sugar cues, which is informative of the larger polysaccharide that is present. This regulatory ability is conferred by each PUL's own regulatory protein(s) that sense and bind to the cognate substrate cue. Although this regulatory strategy is sufficient to activate individual PULs in the presence of particular glycans—usually through the activities of a positive feedback loop that controls only the associated PUL genes—it does not explain the control of previously observed nutrient hierarchies in *Bacteroides*. Here, we report that monosaccharides alter the prioritization hierarchy of *Bacteroides thetaiotaomicron* (*Bt*), and that the sugar ribose affects this hierarchy in the absence of the ability to grow on the sugar, suggesting that other regulatory elements are involved. We observe that members of two regulator families, whose coding genes are not genomically associated with PULs, are transcriptionally altered in the presence of the sugars: arabinose, xylose, and ribose. We identified 22 orphan extracytoplasmic function sigma (ECF-σ) factor regulators as well as 4 orphan LacI family proteins and, through genetic deletions, probed the roles of 9 of these genes. Deletion of several individual genes drastically altered the growth phenotypes of *Bt* on multiple polysaccharides, suggesting that the regulators they encode govern higher-order regulons that may encompass multiple PULs.

Introduction

Many bacteria that successfully compete in the gut by utilizing multiple different nutrients have corresponding regulatory mechanisms to sense available nutrients and respond accordingly, only when cognate substrates are present. For some model organisms like *E. coli*, nutrient regulation mechanisms have been well characterized, revealing a multi-faceted network of interconnected local and global regulons. These networks involve global regulators such as ArcA/B, Crp, Cra, and Mlc [1], which exert effects on many different carbohydrate utilizations pathways at once. While more local (operon-regulating) proteins, such as LacI- and AraC-like regulators, play critical roles in responses to specific nutrients such as lactose and arabinose[2, 3]. Small-regulatory RNAs (sRNAs) also interact directly with the protein Hfq (and others) and alter genes involved in nutrient catabolism such as Crp (which is also a global regulator)[4-6]. Many organisms beyond *E. coli*, including the medically important pathogen *Clostridium difficile* and other Gram-positive *Firmicutes*, employ similar combinations of local and global regulation to express the optimal nutrient acquisition functions[7, 8].

Sigma factors are proteins that direct RNA polymerase to particular promoters[9] and are important in metabolism, stress responses, cell division and, many other bacterial processes. An important set of proteins modulating the activity of some sigma factors are the anti-sigma factors, which directly associate and inhibit transcriptional initiation[10, 11]. Some organisms, like *Bacillus subtilis*, use cascades of sigma factor activation to regulate complex processes like sporulation[12, 13]. Other organisms with amplified repertoires of sigma factors may use these transcription factors to govern multiple, individual regulons instead of cascades. One such organism in which sigma factors have been demonstrated to be important in nutrient metabolism is the important gut bacterial species *Bacteroides thetaiotaomicron (Bt)*, which encodes 54 of these proteins[14]. A subset of *Bt* sigma factors that are associated with PULs and adjacent anti-sigma factors has been shown to be important in the context of foraging of host-mucosal polysaccharides (both *N*- and *O*-linked glycans)[15-18].

Some gut bacteria persist by being able to assimilate many different nutrients in the competitive environment of the mammalian gut, where nutrients are constantly fluctuating due to meal-to-meal fluctuations and other events. The Bacteroidetes devote large portions of their genome to carbohydrate degradation through expression of gene clusters termed polysaccharide utilization loci (PULs)[19]. Perhaps because individual Bacteroidetes may devote up to 20% of

their genome to nutrient foraging, there is a concurrent need to tightly regulate this ability to avoid wasteful expression of the encoded functions[20]. Individual species of *Bacteroides* have nutrient hierarchies that manifest as polyauxic growth in mixtures of polysaccharides[21-24], but are thought to lack cyclic-AMP-based (*i.e*., classical CRP) regulation25 and lack Hfq, so they may rely on other global regulatory mechanisms such as small RNAs (sRNAs)[26, 27], or additional uncharacterized regulators. One layer of PUL regulation is through the activity of local, mostly positive-acting regulatory proteins that are encoded within most individual PULs. The prominent classes of these regulators include a large suite of extracytoplasmic function sigma (ECF-σ) factors that within PULs are always paired with a corresponding anti-s factor, /anti-σ pairs[15, 28], hybrid-two component systems (HTCS)[14], and SusR[29] or RusR-like (described in Chapter II) transcriptional activators. Further, *Bt* is predicted to encode local-acting regulators of AraC-, LacI-, and GntR-type regulators inferred through bioinformatic prediction and preliminary studies on rhamnose and fucose metabolism[30-32].

Recent work has uncovered a new layer of regulation via a protein with functions similar to Crp in *Bt*, BT4338 which has weak amino acid similarity to Crp homologs. When deleted, *Bt* is no longer able to grow on several monosaccharides such as ribose, galactose, uronic acids, and rhamnose to highlight a few, while polysaccharide utilization is relatively unchanged compared to wild-type[33]. Further, this same group of researchers who described the BT4338 Crp-like protein has recently described a separate regulatory protein, BT3172, that when deleted causes a defect in colonization of the mouse gut in a diet specific manner, hence termed the regulator of colonization or Roc[34]. This global effect on carbohydrate utilization via BT4338 and Roc, led us to search for additional regulatory proteins in *Bt* that may predominately act on polysaccharides compared to the BT4338 effects on monosaccharides as well as examine context-dependent (*i.e.* presence of specific sugars affects other metabolic pathways) nutrient utilization previously seen in Chapter II and with Roc.

Here, we investigate how the monosaccharide ribose alters the metabolic priority of other, non-ribose containing nutrients. The presence of the genes involved in ribose catabolism encoded in the ribose utilization system (*rus*) PUL affect the competitive fitness of *Bt* in glucose and rich media similar to the phenotype observed with *Bt* Roc[30]. Growth on pentose sugars xylose and arabinose also elicits changes in expression of other PULs and metabolic functions that do not appear to be directly involved in catabolism of these substrates. Beyond sugar-

induced changes in metabolism we also report the discovery of a single orphan ECF-sigma factor, BT2492, that when deleted in *Bt* causes loss of growth on many of the polysaccharides, but few monosaccharides, that *Bt* normally utilizes for growth. Thus, this sigma factor may act similarly in nature to the global regulator BT4338[33]. In addition, we show that three LacI-type regulators (BT0487, BT1434, and BT3613) are involved in the utilization of uronic-acids and polysaccharides containing these sugars, but not on those that lack these acidic sugars. Taken together, the results here demonstrate that *Bt* employs previously unknown regulatory mechanisms that further enhance its ability to quickly and efficiently respond to changes in available nutrient pools.

Results

Ribose alters polysaccharide prioritization and the ability to use ribose is associated with changes in competitive growth in vitro on non-ribose substrates

Our previous work examining the ribose utilization system (*rus*) PUL showed that, when *Bt* was grown in minimal media (MM) containing ribose as the sole carbon source, expression of genes located in other PULs and metabolic loci were transcriptional altered compared to growth in ribose, despite these genes not being linked to ribose catabolism (Figure 2.9). This suggests that the ability to respond to the presence of ribose may alter the hierarchy of nutrient prioritization. To test if the previously observed effect was dependent on both ribose and a functional *rus* PUL, we performed a transcriptional profiling experiment in which both wild type and Δ*rus Bt* were grown in MM containing equal concentrations of 12 different polysaccharides (PSM12) or the PSM12, plus ribose (PSM12+R). We monitored the transcript of sentinel *susC*-like genes in each of the PULs dedicated to utilization of the 12 individual substrates and compared to pre-PSM exposed (time 0) reference as previously reported[22]. Most of the transcriptional responses to the PSM12 mixture were similar for both wild type and Δ*rus* (Figure 4.1), which was also similar to a previous experiment[22]. However, a PUL involved in arabinan utilization was repressed earlier in the Δ*rus* strain when ribose was both present and absent (Figure 4.1), but this behavior did not extend to the arabinose-containing polysaccharide arabinogalactan. To further explore the basis of this phenomenon, we examined the expression of several non-PUL encoded central metabolism genes and genes found within the arabinose and

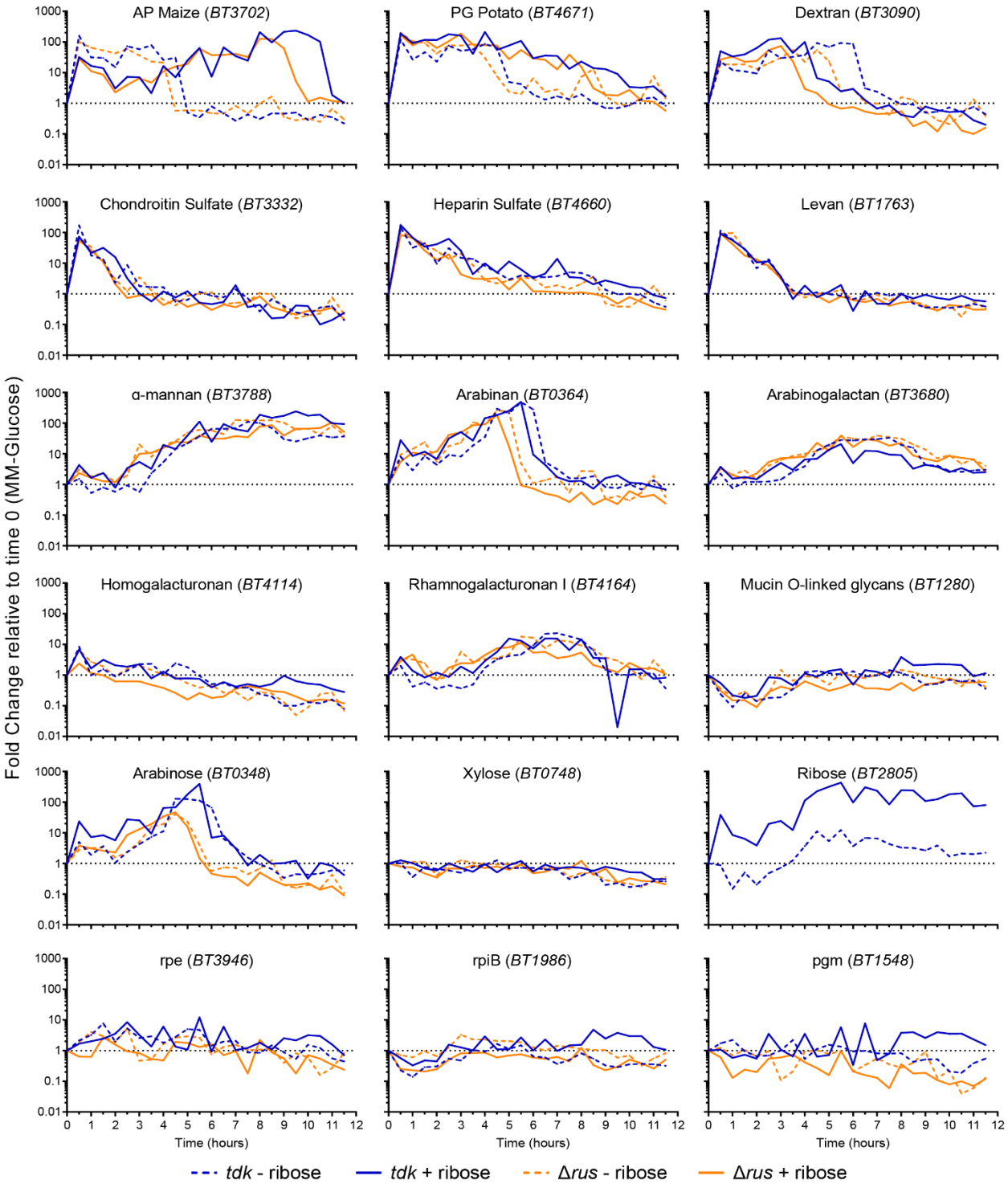


Figure 4.1 Presence of ribose in a polysaccharide mixture alters Bt nutrient hierarchy irrespective of growth ability.

Wild type Bt (blue lines) or a mutant unable to utilize ribose, Δrus (orange lines) assayed for transcript activation in a complex mixture of polysaccharides containing 12, PSM12 (dashed lines) or 13 PSM12+R (solid lines) substrates with ribose being the 13th. In each plot, an individual gene is probed for transcript over the 12 hour time course with samples taken every 30 minutes. For PULs the susC-like homolog was probed, while for arabinose and xylose, the transporter genes were probed. Genes involved in the pentose phosphate pathway were also probed. Levels are compared to the transcript at time 0, when the culture was transferred from MM containing glucose as the sole carbon source and washed in MM with no carbon source.

xylose utilization loci. Similar to the arabinan response, genes in the arabinose utilization locus were repressed more quickly in the Δ*rus* strain (with and without ribose) compared to wild type *Bt* (Figure 4.1). We suspect that the arabinogalactan PUL was not severely altered as it predominantly contains galactose. Interestingly, while changes in responses to arabinan and arabinose were different in the Δ*rus* strain compared to wild type regardless of the presence of ribose, there were also some notable changes in the strain responses in PSM12 compared to PSM12+R. In the PSM12+R experiment, both wild type and Δ*rus* strains displayed early repression in the time course for amylopectin (AP) maize and pectic galactan (PG) potato, but remained active long after expression had ceased in strains not exposed to ribose. In contrast, expression of the dextran PUL remained active for a longer period in the cultures not exposed to ribose versus those of the PSM12+R. Similar responses from the wild type and Δ*rus* strains in the presence and absence of ribose suggests that the ability to utilize ribose is not what mediates the observed change in transcription. Rather, the presence of ribose alone alters the hierarchy by an unknown mechanism, although we cannot rule out that a small amount of ribose gets transported into the cell and phosphorylated by a different set of pentose permeases and kinases. Two of the polymers noted above contain only glucose, albeit AP maize and dextran elicit opposite responses in the presence and absence of ribose. To address if glucose metabolism may be altered in ways not associated with the ability to catabolize ribose as a nutrient, we performed *in vitro* competition assays between wild type *Bt* and the Δ*rus* strain in MM containing glucose as a sole carbon source. For comparison, we used a rich media, tryptone-yeast extract-glucose (TYG) that also contains glucose plus a number of other non-glucose nutrients. We had previously seen *in vivo* that not all genes in the *rus* locus were required for competition, with the presence of some genes involved in outer membrane import and binding (Δ*rusC/D*) and hydrolase functions (Δ*rusGH/NH*) conferring a competitive disadvantage to wild type (Figures 2.4 and 2.5). We used these strains grown in MM-ribose as a control for the phenotype previously seen *in vivo*. In support of our hypothesis that some genes of the *rus* loci are also important for competition in substrates where ribose is not appreciably present (nucleosides, which contain ribose, may be present in TYG), the Δ*rus* mutant was MM-containing glucose and TYG (Figure 4.2G and H). As expected, the Δ*rusC/D* and Δ*rusGH/NH* strains both displayed a competitive advantage of greater than 4 orders of magnitude for Δ*rusGH/NH* compared to wild

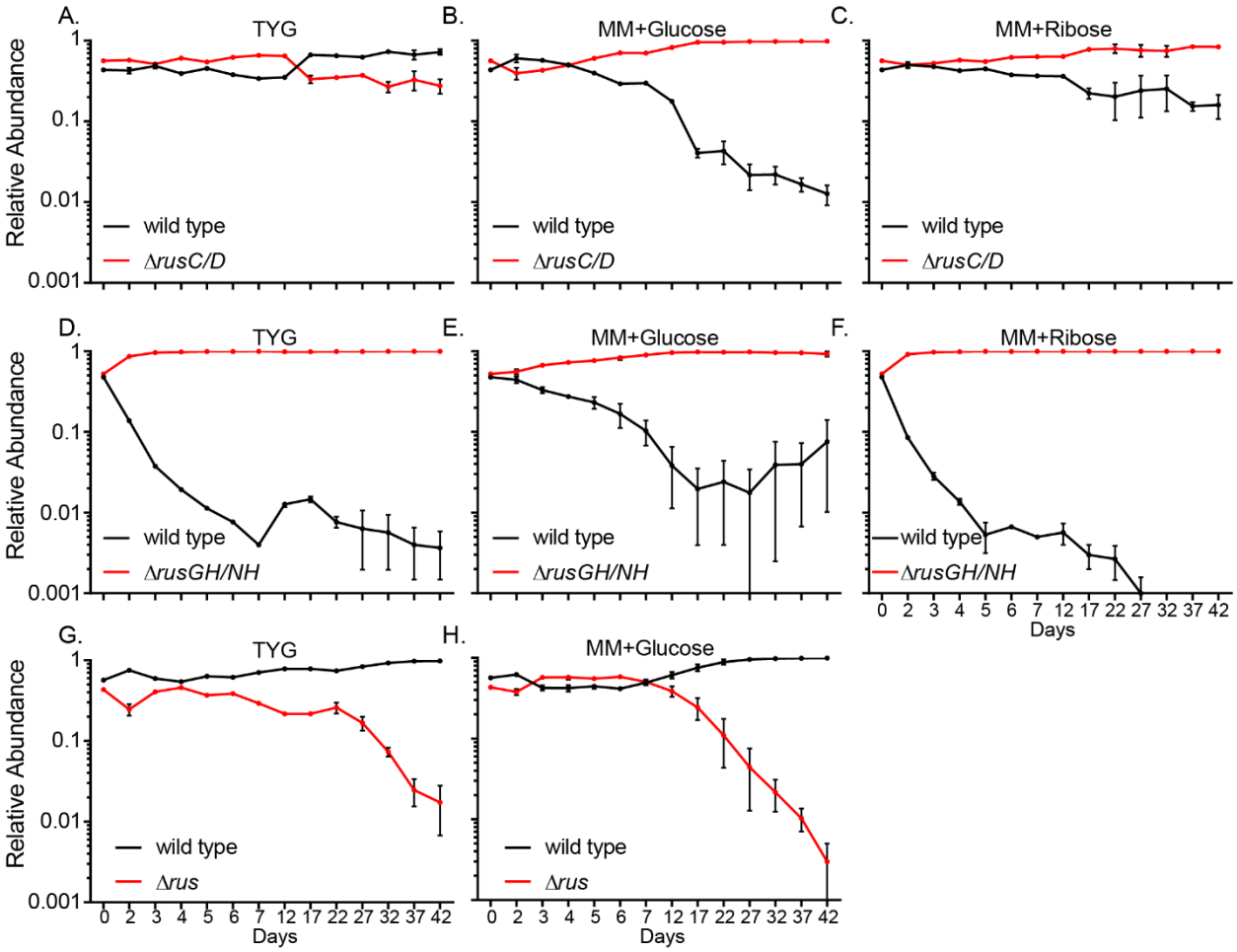


Figure 4.2 Competitive fitness of rus PUL mutant strains are altered in vitro in several media.

Wild type Bt (black line) or deletion mutants within the rus PUL (red lines) were competed in vitro in either tryptone-yeast extract-glucose (TYG) or minimal media (MM) plus glucose or ribose. A-C) ΔrusC/D was competed against wild type in TYG (A), MM+glucose (B) or MM+ribose (C). Similarly, ΔrusGH/NH was competed against wild type Bt (D-F) in TYG (D), MM+glucose (E), or MM+ribose (F). Δrus was competed against wild type in TYG (G) and MM+glucose (H), MM+ribose was not performed as the Δrus strain is unable to grow on ribose and therefore cannot compete for that nutrient. Each time point for each day and strain displays the mean ± the SEM (n=3) with relative abundance shown on the y-axis for all competitions and the number of days is consistent for all competitions.

type, so much so, that after day 27, wild type fell below our detectable limits in MM-ribose (Figure 4.2F). However, the Δ*rusC/D* strain only displayed a slight competitive advantage in MM-ribose, perhaps suggesting a role in ribose scavenging *in vitro* (Figure 4.2C). Contrastingly, both Δ*rusC/D* and Δ*rusGH/NH* strains exhibited a strong advantage over wild type in glucose while only Δ*rusGH/NH* exhibited a strong advantage in TYG with Δ*rusC/D* at nearly equivalent levels as wild type *Bt* (Figure 4.2A-B and D-E). Wild type *Bt* displayed a large competitive advantage over the Δ*rus* strain in both glucose and TYG, suggesting that perhaps nucleosides are present (TYG) or a ribose containing substrate is produced, or a metabolic pathway is affected

during growth on glucose that may be mediated by either the regulator *rusR* or the kinases *rusK1/K2* (Chapter 2).We next wanted to examine further how ribose and the related pentose sugars arabinose and xylose may cause changes in the prioritization of polysaccharides similar to arabinan (Figure 4.1).

Arabinose and xylose growth cause global transcriptional responses beyond direct metabolism of these sugars

Based on our previous result showing ribose causes transcriptional changes in a global regulatory network (Figure 2.9G), and because the absence of *rus* alters the response to arabinose/arabinan, we wanted to test if the related pentoses arabinose and xylose cause similar changes in global metabolism. We hypothesized that growth on these related sugars would cause similar transcriptional changes as growth on ribose and reveal metabolic networks or regulators that mediate these effects. In order to address this, we grew wild type *Bt* on arabinose or xylose as a sole carbon source and performed RNAseq analysis using MM-glucose as a common reference and using a 5-fold cut-off. Our data showed that, like growth on ribose (Figure 2.9G), xylose and arabinose also alter expression of PUL genes (both positively and negatively). Specifically, xylose repressed a single PUL, *BT3344-3347,* responsible for catabolism of an unknown substrate (Figure 4.3A), and this PUL was also repressed during ribose growth (Figure 2.9G). Growth on xylose activated expression of 3 genes in the *rus* PUL (*BT2803, BT2804,* and *BT2809*), although not to nearly as high levels as ribose. Several other glycoside hydrolases were also upregulated and there was repression of the GH13, *susA*. As expected, growth on xylose cause upregulation of the genes involved in xylose catabolism, *BT0791-0794* (Figure 4.3A), (this genetic locus was deleted, confirming involvement in xylose catabolism, data not shown). In contrast to xylose, growth on arabinose almost exclusively elicited transcriptional activation of PUL genes. Specifically, as was the case for ribose growth (Figure 2.9G), arabinose caused upregulation of the fructan PUL, *BT1757-1765*, and the *rus* PUL, *BT2803-09* (Figure 4.3B). Growth on arabinose caused upregulation of an unknown, three gene PUL, *BT4038-4040,* and a PUL, *BT4294-4299* (Figure 4.3B); the latter previously associated with mucin *O*-glycan degradation[15, 35]. The genes involved in fucose metabolism, *BT1272-BT1277* were also upregulated when grown in arabinose (Figure 4.3B), and interestingly fucose is a predominantly host derived sugar and could suggest that arabinose is a metabolite that couple or modifies the

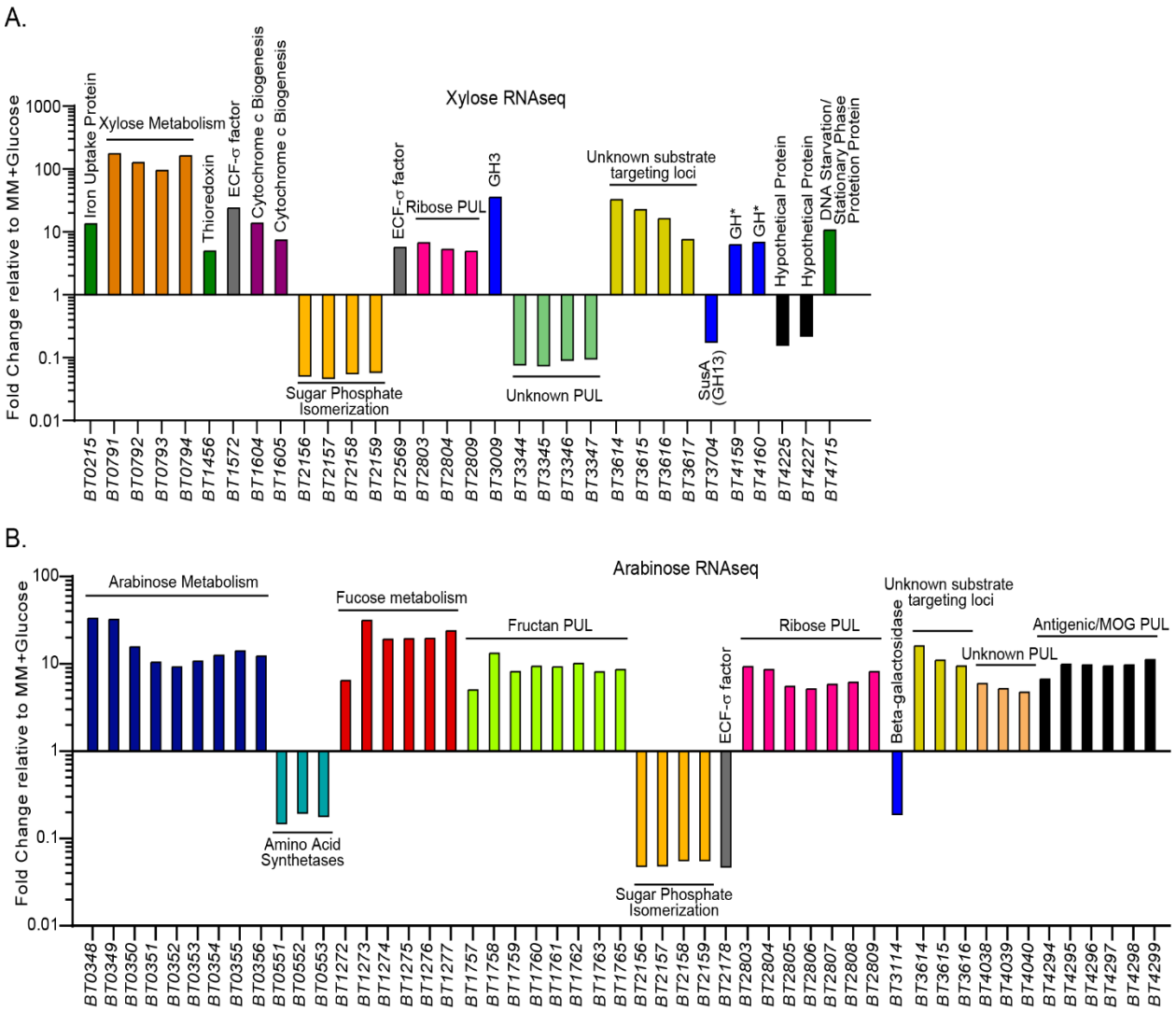


Figure 4.3 RNAseq reveals that the monosaccharides arabinose and xylose alter expression of non-PUL-encoded metabolic loci, including orphan ECF-σ regulatory genes.

RNAseq results of wild type Bt grown in either xylose (A) or arabinose (B) with fold change compared to cells grown in glucose as a sole carbon source. Locus tags are displayed along the x-axis. Genes in the same locus or PUL are colored the same within either A or B. Single genes displaying changes are labeled above or below the bar with the predicted gene product. Bars show the mean of n=2 replicates.

responses of Bt as the host shifts between fiber rich and fiber free diets. As expected, the genes for arabinose metabolism, *BT0348-BT0356,* were upregulated (Figure 4.3B). For the most part, xylose and arabinose appear to affect different global transcriptomes compared to growth on ribose, however, the following loci all behaved similarly in response to growth on any of these pentoses, suggesting either a common regulatory link to metabolic pathways (*i.e.* pentose phosphate) or, more intriguingly, new functions for pentose assimilation. Among these functions were genes for sugar phosphate isomerization, *BT2156-2159,* that were repressed in all three

conditions, and upregulation of a locus, *BT3614-3617*, that may be associated with substrate utilization based on the predicted, encoded enzymes (hydrolase, reductase, permease, dehydrogenase) (Figure 2.9G and Figures 4.3 B and C).

As we hypothesized, growth on each of these monosaccharides also affected the transcript levels of regulatory proteins, more specifically, orphan (non PUL or anti-σ-associated) ECF-σ factors. Within *Bt* there are 20 orphan ECF-σ factors throughout the genome[14]. While most of these did show altered expression, three genes encoding ECF-s factors were changed with *BT2569* upregulated in both ribose and xylose (Figure 2.9G and Figure 4.3A), *BT1572* upregulated in xylose (Figure 4.3A), and *BT2178* repressed in arabinose growth (Figure 4.3B). This result lead us to examine potential roles for additional orphan ECF-σ factors in the metabolism of carbohydrates.

The orphan ECF-σ factor, BT2492 affects growth of Bt on many polysaccharides

In order to test the hypothesis that orphan ECF-σ factors are regulators of carbohydrate utilization, we made strains with single gene deletions for 6 of these regulators and then examined their growth profile on a panel of poly-, mono- and disaccharides that wild type *Bt* normally utilizes for growth as sole carbon sources. Of the 20 known orphan ECF-σ factors, we were able to successfully create deletion strains of *BT0248*, *BT1197*, *BT1572*, *BT1817*, *BT2044*, and *BT2492*. We attempted deletions of the additional ECF-σ factors *BT2184* and *BT2569* that were upregulated in the RNAseq results, as well as *BT0326*, *BT1103*, and *BT1559* but were unable to obtain successful deletions, possibly due to an essential function in metabolism or another biological process as a previous transposon mutagenesis screen identified two of these factors as candidate essential genes[36], and perhaps the frequency of successful allelic recombination is exceedingly rare due to requirements in biological processes. Within the six deletion strains, only Δ*BT2492* displayed reduced growth (both rate and total biomass) on 12 polysaccharides compared to wild type and the other deletion strains (Figure 4.4 A-L). Surprisingly, there does not appear to be a common pattern associated with the substrates that Δ*BT2492* has defects on, meaning they do not readily share a common monosaccharide core or common linkages or source (*i.e.* plant-derived or host-derived). Further, Δ*BT2492* did not display growth defects on monosaccharides and a few additional polysaccharides of inulin, hyaluronan, and RGI (Figure 4.5 A-O) which leads us to hypothesis that the mechanism of

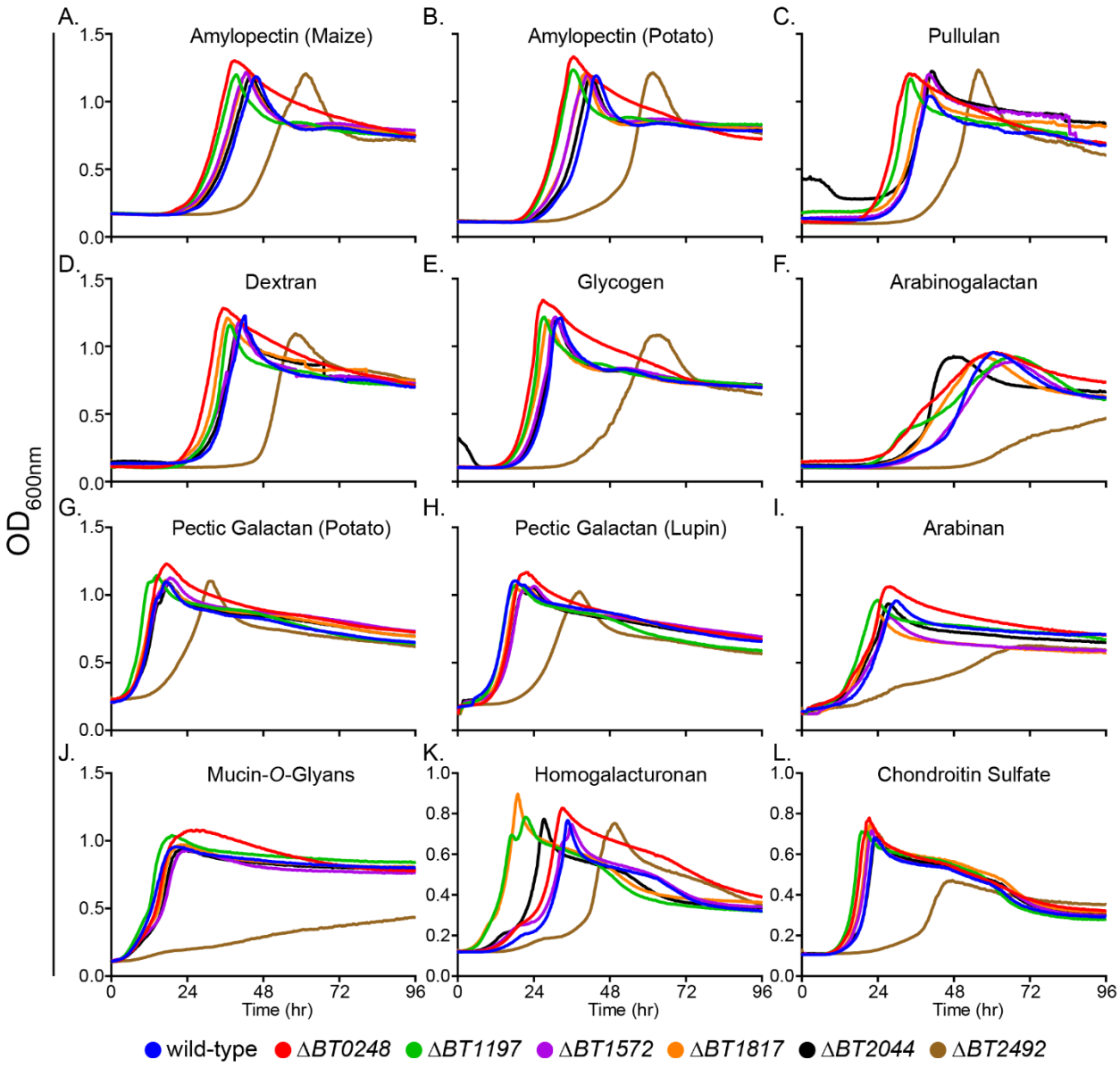


Figure 4.4 The ECF-σ factor, BT2492 controls utilization of Bt metabolized polysaccharides.

(A-L) Growth curves of Bt wild type (blue line), and individual ECF-σ factor deletions: BT0248 (red line), BT1197 (green line), BT1572 (purple line), BT1817 (orange line), BT2044 (black line), and BT2492 (brown line). All strains were grown on the following polysaccharides amylopectin maize (A), amylopectin potato (B), pullulan (C), dextran (D), glycogen (E), arabinogalactan (F), pectic galactan potato (G), pectic galactan lupin (H), arabinan (I), mucin-O-glycans (J), homogalacturonan (K), or chondroitin sulfate (L). Only the ΔBT2492 strain displayed growth defects compared to wild type for substrates shown. Growth was measured by absorbance at 600nm for a minimum of n=3 biological replicates on separate days.

BT2492 promoting growth is by transcribing genes essential for utilization of the indicated polysaccharides. These genes may be any essential component required for growth on these substrates, such as, BT4338 (MalR), PUL-specific regulators, or even machinery required for import (SusC transporters and TonB energizers). Although we do not directly test this, it is interesting that for all of the substrates that Δ*BT2492* displayed defects on, the PUL-encoded regulators are hybrid two-

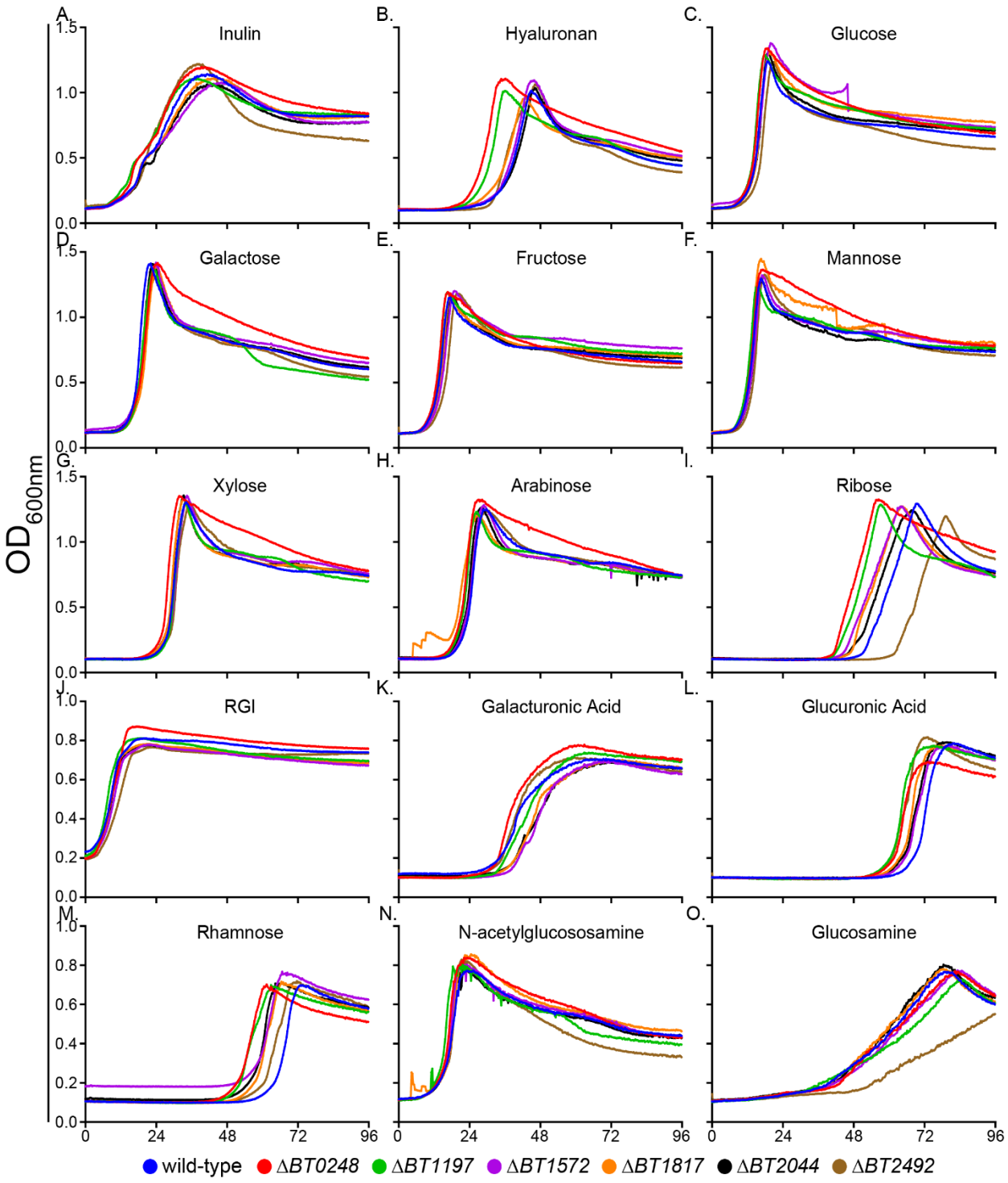


Figure 4.5 BT2492 and other ECF-sigma factors do not affect all substrates that Bt metabolizes.

Data is related to Figure 4.4. Bt wild type (blue line), and individual ECF-σ factor deletions: ΔBT0248 (red line), ΔBT1197 (green line), ΔBT1572 (purple line), ΔBT1817 (orange line), ΔBT2044 (black line), and BT2492 (brown line) were grown on a panel of polysaccharides and monosaccharides. Each panel shows the growth of each strain on a particular substrate as a sole carbon source as measured by absorbance at 600nm. None of the deletion strains exhibited differences in growth compared to wild type except for minor growth defect of ΔBT2492 on ribose and glucosamine. Growth curves displayed are representative of n=3 biological replicates run on separate days.

component systems (HTCS) or SusR. This suggests that the ECF-σ factor *BT2492* acts upstream of PUL-encoded regulation and recruits RNA polymerase to initiate transcription of local regulators of these PULs or another component of their global regulation. A similar mechanism was recently described in the *Bacteroidetes* member *Porphyromonas gingivalis* within type IX secretion systems, whereby expression of two-component system proteins were further regulated by an orphan ECF-σ factor[37]. The promising results that disruption of at least one ECF-σ causes a phenotype that extends to multiple polysaccharides, led us to search within the genome of *Bt* for additional orphan regulators that may provide further evidence of global regulatory mechanisms.

Orphan LacI-type regulators repress the catabolism of uronic acid-containing substrates

In order to identify additional orphan regulators, we used BLAST to search for homologs of LacI- and AraC-type regulators. We focused on four predicted LacI-type regulators, *BT0487, BT0824, BT1434,* and *BT3613* that had been predicted through bioinformatics to be involved in utilization of uronic acid monosaccharides (glucuronate, galacturonate, mannuronante), but not experimentally tested[32]. To experimentally to test the roles of these regulators, we made single gene deletions, of three of these genes, *BT0487, BT1434,* and *BT3613*. We performed growth analysis of these mutants in MM containing poly- and monosaccharides that *Bt* is able to utilize. As predicted by previous bioinformatics, these regulators did show growth differences compared to wild type *Bt* on uronic acids, with better growth (defined as earlier or decreased lag time) than wild type (Figure 4.6). However, our results differ in the predicted substrates that these regulators control, and our analysis also detected altered growth in polysaccharide substrates. For Δ*BT0487*, earlier growth (decreased lag time) was seen in both galacturonic and glucuronic acid with glucuronic acid growth exhibiting a much improved lag compared to wild type (Figure 4.6 A and D). In addition, Δ*BT0487* also displayed better growth on rhamnose, hyaluronan (a polysaccharide of glucuronic acid and *N*-acetyl-D-glucosamine repeating units), homogalacturonan (HG, a polymer of galacturonic acid residues), and arabinogalactan, which in contrast to the other polysaccharides contains few uronic acid residues and consists mainly of galactose and arabinose units (Figure 4.6A-F). Similarly, the Δ*BT1434* strain was only predicted to participate in catabolism of glucuronic acid, which we observed (Figure 4.6A), but we also observed enhanced growth on HG and rhamnose (Figure 4.6B and C). Lastly, the Δ*BT3613*

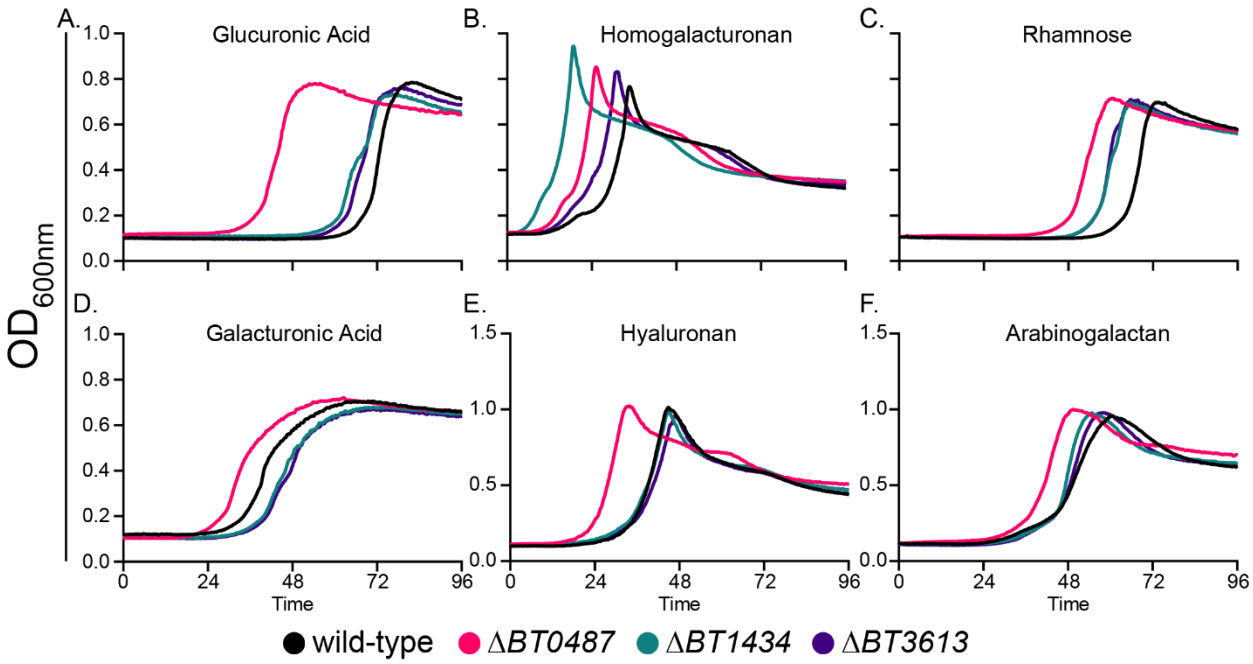


Figure 4.6 Orphan LacI-type regulators in Bt repress metabolism of uronic-acid containing substrates.

(A-E) Growth of wild type Bt (black line) or deletion strains lacking individual orphan LacI-type encoding genes: ΔBT0487 (pink line), ΔBT1434 (teal line), and ΔBT3613 (purple line). Each panel shows growth on an individual substrate as a sole carbon source: glucuronic acid (A), homogalacturonan (B), rhamnose (C), galacturonic acid (D), hyaluronan (E), or arabinogalactan (F). ΔBT0487 displayed increased growth compared to wild type for all substrates shown Growth curves displayed are representative of n=3 biological replicates run on separate days.

strain behaved similarly to the Δ*BT1434* strain, with enhanced growth characteristics on glucuronic acid, HG, and rhamnose (Figure 4.6A-C), despite only being predicted to participate in catabolism of mannuronic acid. We did not test mannuronic acid in our assays, but based on glucuronic acid, HG, and rhamnose (Figure 4.6A-C), despite only being predicted to participate in catabolism of mannuronic acid. We did not test mannuronic acid in our assays, but based on the observed defects, it is likely that all of the mutants would have also displayed enhanced growth. Interestingly, the genes immediately adjacent to the *BT3613* LacI-typeregulator, *BT3614-3617* were upregulated in arabinose, xylose, and ribose (Figure 4.3 A-B and Figure 2.9G). It is possible that this locus participates in catabolism of uronic acids based on the results of the Δ*BT3613* mutant strain growths. We did not observe defects on other non-uronic acid substrates other than rhamnose for any of the three LacI-type regulator deletions (Figure 4.7). The results obtained here help to clarify the function of these orphaned regulators, and for some, confirm what was previously predicted using bioinformatic techniques.

Discussion

We show in this study the presence of additional regulatory mechanisms underpinning nutrient utilization in *Bacteroides thetaiotaomicron* (*Bt*). We demonstrated that the presence of monosaccharide ribose, but necessarily the ability to utilize it with full efficiency if at all, alters previously established nutrient prioritization hierarchy in *Bt*[22]. We determined that ribose alters the competitive fitness of *Bt in vitro* under conditions where ribose containing substrates are not thought to be prominent substrates present, representing only the second example in *Bacteroides* where the ability to catabolize one nutrient may affect the metabolism of or the ability to access other nutrients[34]. We extended these findings by examining the global transcriptional response to the related pentose sugars arabinose and xylose, demonstrating that (like ribose), these individual sugars create an altered global transcriptional response. These responses include activation and repression of genes not directly related to the metabolism of these nutrients was observed compared to growth in glucose. Interestingly, some of the genes for which we observed expression changes were predicted to encoded ECF-σ factors that are orphans, meaning they are not adjacent to known metabolism genes and for the ones we observed, did not have a canonical anti-σ factor located next to it in the genome. This led us to construct genetic deletions of 6 of these orphan ECF-σ factors and 3 orphan LacI-type regulators, leading to the result of 4 of these genes, especially *BT2492*, having substantial effects on the catabolism of 12 carbohydrate nutrients.

There have been previous studies that examined the prioritization of polysaccharides in *Bacteroides* and documented the diauxic or polyauxic growth of *Bacteroides* on polysaccharide mixtures[21-24]. It has been presumed that this transcriptional prioritization hierarchy both matches the utilization profiles of what is being consumed versus retained, and that the hierarchy is relatively stable unless other polysaccharides are introduced. The first assumption has recently been shown inaccurate for some *Bacteroides* as transcriptional activation of PULs for high-priority glycans remains high, even after the substrate is largely[24]. Here, we tested the second assumption, and found that the prioritization changes when additional nutrients are added to the mixture. Interestingly, we were able to drive this change by addition of a monosaccharide (ribose), which in some cases exerted its effect without the presence of the PUL-encoded machinery for its. Nevertheless, our study demonstrated both that the presence of a monosaccharide alters transcriptional profiles of more complex polysaccharides and that the

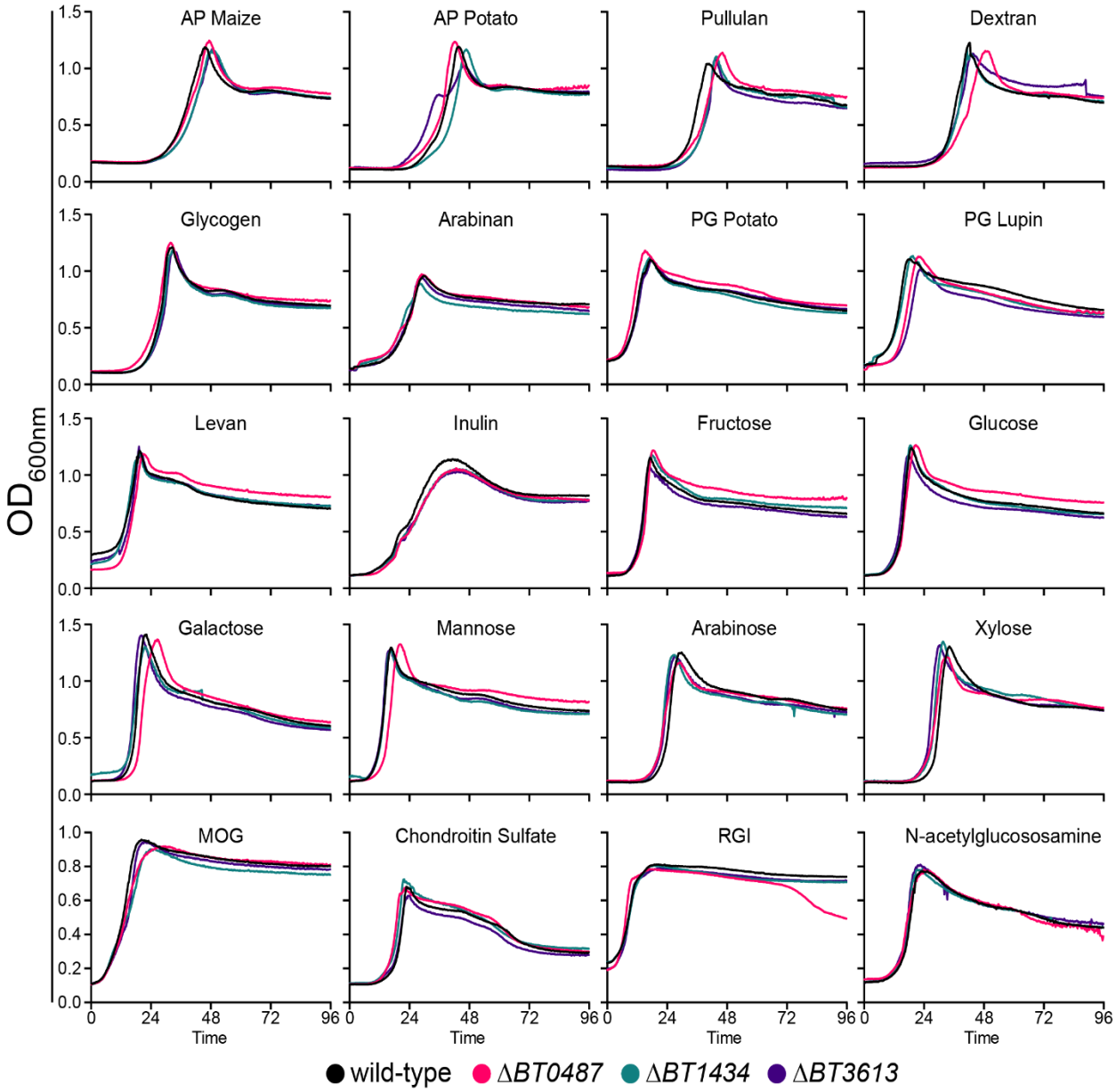


Figure 4.7 Orphan LacI-type regulators do not substantially alter metabolism of non-uronic acid-based polysaccharides.

Data is related to Figure 4.6. Growth curves of wild type or LacI-type deletion strains on polysaccharides or monosaccharides for which there were no detectable differences in growth for deletion strains compared to wild type Bt. Strains are color coded as follows: wild type (black line), ΔBT0487 (pink line), ΔBT1434 (teal line), and ΔBT3613 (purple line). Growth curves represent at least n=3 biological replicates on separate days.

presence of the ribose PUL even in the absence of ribose alters the hierarchy, suggesting a previously undescribed mechanism. It may be that the regulator or ribokinases found in the ribose PUL mediate utilization of other polysaccharides or their constituent monosaccharides, or a mechanism of repression exists, future studies will focus on these possible mechanisms. Additionally, in support of the idea that ribose PUL-encoded functions play a role in utilization

of other nutrients, our *in vitro* competition assay showed that our Δ*rus* PUL mutant was severely outcompeted by wild type *Bt* in both a rich media and a minimal media containing glucose as the sole carbon source. To fully evaluate, complementation studies as well as competitions in non-glucose containing media should be performed in future experiments to rule out an intrinsic defect. We should note that a major component of TYG is glucose and so the defect observed in TYG may be due to presence of glucose, although other nutrients such as yeast extract are also present and could provide additional, albeit less defined, carbon sources that may contain small amounts of ribose-containing nutrients. These results reinforce previous work on the importance of ribose and nucleoside utilization and further suggest that the ribokinases, regulator, or permease encoded in *rus* act on non-ribose substrates, or that small amounts of ribose containing compounds (perhaps nucleosides) are sufficient to drive a competitive defect in glucose. Although previous studies in *Bt* have demonstrated its ability to grow on the monosaccharides: arabinose, xylose, rhamnose, and fucose, we are the first to show that growth on some of these monosaccharides affects transcriptional activation of genes not involved in metabolism of the cognate monosaccharide. Perhaps most intriguingly, arabinose and xylose show differential upregulation or repression of PULs for unknown substrates, stimulation of the ribose PUL, upregulation of a PUL known to be antigenic *in vivo* and the upregulation of the genes for fucose metabolism, a common component of mucosal polysaccharides. This result suggests that perhaps an additional level of regulation in *Bacteroides* is co-regulation of metabolic loci by specific nutrients that prime the cell for additional substrates.

The way in which *Bt* regulates its nutrient degrading capabilities is complex, involving locally-encoded, cis-acting PUL regulators such as HTCS, ECF-σ/anti-σ pairs, and SusR/RusR regulators, as well as small-RNAs. Over-arching these local regulons are global regulators, such as CRP-like proteins. Our observation that at least one orphan ECF-s factor plays a role in utilization of multiple polysaccharides suggest the existence of an undiscovered layer of global regulation mediating carbohydrate metabolism. We suspect that BT2492 acts as a global regulator, possibly aiding in the recruitment of RNA polymerase to transcripts normally under the control of HTCS and SusR/RusR regulatory mechanisms, or turns on expression of the PUL-activating regulators themselves. Although direct evidence of RNA polymerase recruitment to HTCS and similar local regulator transcripts by orphan ECF-σ factors is lacking in *Bt* a mechanism similar to this has recently been described in an oral *Bacteroidetes* member[37].

Although no LacI-type regulators showed altered expression in our RNAseq data, we identified four orphan LacI-type regulators within *Bt.* Deletion of three of LacI-type regulators (one was not attempted, *BT0824*), all yielded increased growth on uronic acid-based monosaccharides and some uronic acid-containing polysaccharides. Although several loci in *Bt* have previously been shown to be regulated by LacI-type transcriptional repressors[31], we are the first to demonstrate how an orphaned repressor affects utilization of monosaccharides using a genetic deletion approach. Take together with the *BT2492* data and the hierarchy reprogramming by monosaccharides, it suggests that there are several additional levels of regulation that *Bt* employs during carbohydrate degradation. The work presented here can be used to guide additional mechanistic studies exploring previously unknown regulatory functions in *Bt* by exploring the genome for orphaned, seemingly unlinked regulatory genes. Future work can be guided by bioinformatic predictions that have identified additional gene encoded regulators with DNA-binding motifs and subsequently deleting these genes individually and sequentially. Further, we believe binding and DNA foot printing studies will be an important aspect of later work in order to identify potential conserved DNA sequences that can be used to search for homologous sequences in related *Bacteroides* and discovery of the molecules mediating these transcriptional regulatory networks.

Methods

Bacterial strains, culturing conditions, and molecular genetics

B. thetaiotaomicron (*Bt*) ATCC 29148 (VPI-5482) and genetic mutants, were grown in tryptone-yeast extract-glucose (TYG) broth medium[38], in minimal medium (MM), plus a single carbon source[15], or on brain heart infusion agar with 10% defibrinated horse blood (Colorado Serum Co.). Carbon sources used in MM were added to a final concentration of 5 mg/ml or 10 mg/ml with RGI and MOG (mucin-*O*-glycans). Cultures were grown at 37°C in an anaerobic chamber (10% H_2, 5% CO_2, and 85% N_2; Coy Manufacturing, Grass Lake, MI). Genetic deletions were made by counter-selectable allelic exchange as described[39]. Primers used in this study are listed in Table 4.1. To quantify growth on carbon sources, increase in culture absorbance (600 nm) in 200µl cultures in 96-well plates was measured an automated plate reader as previously described[19].

In vitro competition assays

Bt wild type and mutant deletion strains (Δ*rus*, Δ*rusC/D*, and Δ*rusGH/NH*) were initially started in TYG media and then washed 2X in MM with no carbon and subcultured together in pairs such with wild type always being present and the varying strain being one of the mutants. These strains were subcultured 1:25 into fresh TYG or MM containing 5 mg/ml glucose or 5 mg/ml ribose. These mixtures of bacteria were continually passaged (subcultured) daily into fresh media of the specific type *i.e.* TYG grown competition was passaged into fresh TYG at roughly the same time every day for 42 days. At each day, DNA was extracted from fecal pellets throughout the experiment and strain abundance was quantified as described previously[40]. Relative abundance was plotted for each strain on a log scale by qPCR enumeration of unique barcoded tags in each strain.

Measuring transcriptional dynamics by qPCR of polysaccharide hierarchy

Measurements of transcriptional response over time in a mixture of 12 polysaccharides (PSM12) or PSM12+ribose was done as described previously[20, 22]. Briefly, strains were grown in TYG, subcultured 1:50 into MM-glucose, at mid-exponential phase defined as 0.6-0.8 (absorbance at 600nm), cells were washed twice in MM-no carbon and resuspended in PSM12 or PSM12+R with time points being taken every 30 min for 12 hours. To stabilize RNA, at each time point, two volumes of RNA protect were added, followed by centrifugation and storage of cell pellets at -80°C. Total RNA was extracted using the RNeasy mini kit buffers (Qiagen) and purified on RNA-binding spin columns (Epoch), treated with TURBO DNaseI (Ambion) or DNase I (NEB) after elution and purified again using a second RNeasy mini kit isolation column. Reverse transcription was performed using SuperScript III reverse transcriptase and random primers (Invitrogen). The abundance of each target transcript in the resulting cDNA was a homemade qPCR mix as described previously[41]. Detailed information regarding composition of each reaction ware included in the Methods of Chapter 2 and are derived from a previously established protocol[21]. The ddCT method was used to normalized to 16S rRNA values and then individual *susC*-like gene values within the PUL of interest were referenced to the values obtained in MM-glucose at time 0 to obtain a fold-change.

RNAseq analysis

To determine the global transcriptional response to growth in arabinose and xylose as sole carbon sources, *Bt* was grown in TYG media then transferred to fresh MM containing either 5 mg/ml glucose or 5 mg/ml arabinose or 5 mg/ml xylose. Cells were then grown until mid-log phase (absorbance between 0.6-0.8) and two volumes of RNA Protect (Qiagen) were added to cells. RNA was isolated as described above and purified whole RNA was then rRNA depleted using the Ribo-Zero Bacterial rRNA Removal Kit (Illumina Inc.) and concentrated with the RNA Clean and Concentrator-5 kit (Zymo Research Corp, Irvine, CA). Samples were multiplexed for sequencing on the Illumina HiSeq platform at the University of Michigan Sequencing Core. Data was analyzed using Arraystar software (DNASTAR, Inc.) using RPKM normalization with default parameters. Gene expression in arabinose and xylose was compared to gene expression in a glucose reference. Genes with significant up- or down-regulation were determined by the following criteria: genes with an average fold-change $\geq$5-fold and a normalized expression level $\geq$1% of the overall average RPKM expression level in either glucose or ribose. Detailed gene information with significant hits for arabinose and xylose are listed in Table 4.2.

Table 4.1 Strains, vectors, and primers used in this study

Strain	**Genotype**	**Features**	**Reference**
Bacteroides thetaiotaomicron (*B. theta*) *tdk*$^-$	ATCC 29148 *tdk*$^-$	Parent strain of all deletion strains, and refered to in text as "wild-type"	Koropatkin et al. 2008
B. theta Δ*BT0248*	ATCC 29148 *tdk*$^-$, *BT0248*$^-$	ECF-σ deletion (Sig1)	This study
B. theta Δ*BT1197*	ATCC 29148 *tdk*$^-$, *BT1197*$^-$	ECF-σ deletion (Sig2)	This study
B. theta Δ*BT1572*	ATCC 29148 *tdk*$^-$, *BT1572*$^-$	ECF-σ deletion (Sig3)	This study
B. theta Δ*BT1817*	ATCC 29148 *tdk*$^-$, *BT1817*$^-$	ECF-σ deletion (Sig4)	This study
B. theta Δ*BT2044*	ATCC 29148 *tdk*$^-$, *BT2044*$^-$	ECF-σ deletion (Sig5)	This study
B. theta Δ*BT2492*	ATCC 29148 *tdk*$^-$, *BT2492*$^-$	ECF-σ deletion (Sig6) that losses growth on ~ half of substrates tested for growth compared to the parent strain	This study
B. theta Δ*BT0487*	ATCC 29148 *tdk*$^-$, *BT0487*$^-$	LacI-type regulator deletion that grows substantially better on all uronic acid-containing substrates tested	This study

		and rhamnose and arabinogalactan	
B. theta Δ*BT1434*	ATCC 29148 *tdk*$^-$, *BT1434*$^-$	LacI-type regulator deletion that grows better on homogalacturonan and rhamnose	This study
B. theta Δ*BT3613*	ATCC 29148 *tdk*$^-$, *BT3613*$^-$	LacI-type regulator deletion that grows better on homogalacturonan and rhamnose	This study
B. theta Δ*rus*	ATCC 29148 *tdk*$^-$, *BT2802-2809*$^-$	Strain lacking the *rus* PUL, unable to grow on or respond to ribose	Glowacki et al. 2019
Primer	**Sequence (5' to 3')**	**Use**	
Genetic Manipulation Primers			
Ribose PUL Knockout *B. thetaiotaomicron* VPI5482			
Restriction Sites are Underlined			
BT2802-2809 5'Up XbaI	GCGTCTAGACGGCTCCATAAAGGTTATC	BT2802-2809 Ribose PUL Knockout	
BT2802-2809 3'Out	GTTTTCTGTAGCTCTTTGTTGCG	BT2802-2809 Ribose PUL Knockout	
BT2802-2809 5'Out	CGCAACAAAGAGCTACAGAAAACGGGGTGAAATTCAATTCTATGATT	BT2802-2809 Ribose PUL Knockout	
BT2802-2809 3'Down Sal1	GCGGTCGACGCTGTTGTGTTCAATGATCTG	BT2802-2809 Ribose PUL Knockout	
BT0248 5' Up Sal1	GCGGTCGACCGGATGCTGGATTTACTTGAC	BT0248 Gene Knockout	
BT0248 3' Out	CAGTTCTTCCAGTTCCATATATGAGAATTAAATGGGTTACTTTTCTGG	BT0248 Gene Knockout	
BT0248 5' Out	TCATATATGGAACTGGAAGAACTG	BT0248 Gene Knockout	
BT0248 3' Down XbaI	GCGTCTAGAGGCAGCCAGTAGAGGATTCTCAGC	BT0248 Gene Knockout	
BT1197 5' Up Sal1	GCG*GTCGAC*GGTACCCAGGTCGAACATGAT	BT1197 Gene Knockout	
BT1197 3' Out	TGACTAATGTGTTATCCCTATACTTTAAGGGTTTTACGGTTAATATTT	BT1197 Gene Knockout	
BT1197 5' Out	ATAGGGATAACACATTAGTCA	BT1197 Gene Knockout	
BT1197 3' Down XbaI	GCGTCTAGAGCGACCGACACCTGCCATAT	BT1197 Gene Knockout	
BT1572 5' Up Sal1	GCGGTCGACCCGTTCGGACCGGCGAAGTCA	BT1572 Gene Knockout	

BT1572 3’ Out	ACATCACTCAATTAGAAAGTTGCATCTCGTATTTTATTTATCTAATTAGTA	BT1572 Gene Knockout	
BT1572 5’ Out	TGCAACTTTCTAATTGAGTGATGT	BT1572 Gene Knockout	
BT1572 3’Down Xbal	GCGTCTAGACACCACTTCCTGCAACGCATAAGT	BT1572 Gene Knockout	
BT1817 5’ Up Sal1	GCGGTCGACGGCTGGTTATCGAAAGAGAAT	BT1817 Gene Knockout	
BT1817 3’ Out	CTTTCCTCTTCCATATCCTCTCTTCTTGACTCTTTAGACGCTGCC	BT1817 Gene Knockout	
BT1817 5’ Out	AAGAGAGGATATGGAAGAGGAAAG	BT1817 Gene Knockout	
BT1817 3’ Down Xbal	GCGTCTAGACCATCGGTTGTGGCAATCGGC	BT1817 Gene Knockout	
BT2044 5’ Up Sal1	GCGGTCGACCTCTCTGGCAAACGCCCAGAA	BT2044 Gene Knockout	
BT2044 3’ Out	AAGAAATCTTTCAGAAGTTTATCATCATCTATTTTCTTCGTTCTT	BT2044 Gene Knockout	
BT2044 5’ Out	TGATAAACTTCTGAAAGATTTCTT	BT2044 Gene Knockout	
BT2044 3’ Down Xbal	GCGTCTAGAGTCGGATATAGGAATCCCTGA	BT2044 Gene Knockout	
BT2492 5’ Up Sal1	GCGGTCGACCGGTGAAATGACTTACGGCGC	BT2492 Gene Knockout	
BT2492 3’ Out	ATATTTTCTAGTTGTTAGACATTTTTTTTAATTCTTCTTTCTTATTTTGATTAATCGTCT	BT2492 Gene Knockout	
BT2492 5’ Out	AAAGAAGAATTAAAAAAAATGTCTAACAACTAGAAAATAT	BT2492 Gene Knockout	
BT2492 3’ Down Xbal	GCGTCTAGACATCGACAAAACGAACAAAAC	BT2492 Gene Knockout	
BT0487 3’ Up Sal1	GCGGTCGACCGGAAATTCAAGTCAACCGAT	BT0487 Gene Knockout	
BT0487 5’ Out	ACGATCACCTTATTTATATAG	BT0487 Gene Knockout	
BT0487 3’ Out	CTATATAAATAAGGTGATCGTCAACAAAATCCAACTAAACAA	BT0487 Gene Knockout	
BT0487 5’ Down Xbal	GCGTCTAGAGGCATATAAATCGCGCAGAATATC	BT0487 Gene Knockout	
BT1434 3’ Up Sal1	GCGGTCGACTGTAGACATGGGAATTTCTGG	BT1434 Gene Knockout	
BT1434 5’ Out	CTTGGCGATGTCTACAATACG	BT1434 Gene Knockout	
BT1434 3’ Out	CGTATTGTAGACATCGCCAAGACCATATAAATAAAAGAACCATGAACG	BT1434 Gene Knockout	
BT1434 5’ Down Xbal	GCGTCTAGAGAATAGTTCCGTAAAGATCTTCCG	BT1434 Gene Knockout	
BT3613 5’Up Sal1	GCGGTCGACCGCACAGACAATAGTGACAGAAGC	BT3616 Gene Knockout	
BT3613 3’Out	TATGTAAAAGGTGCTGTATCTATACTGAAGGTCTCCAACTTAGAT	BT3616 Gene Knockout	
BT3613 5’Out	TATAGATACAGCACCTTTTACATA	BT3616 Gene Knockout	
BT3613 3’Down Xbal	GCGTCTAGAGCGTAATCGCGATACCATGCAGCA	BT3616 Gene Knockout	
BT0824 5’ Up Sal1	GCGGTCGACCGCGTGTCTGGATATGATATTCCA	BT0824 Gene Knockout	
BT0824 3’ Out	TAGTTGTAATTTATAAAGTTGATAACGGGCAATGTCTTTGATGGT	BT0824 Gene Knockout	

BT0824 5’ Out	TATCAACTTTATAAATTACAACTA	BT0824 Gene Knockout	
BT0824 3’ Down Xbal	GCGTCTAGACCCACAGGTGGAAGATTTCAGCCT	BT0824 Gene Knockout	
BT2383 5’ Up Sal1	GCGGTCGACGGAGTATTATCCCAATGTCGA	BT2383 Gene Knockout	
BT2383 3’ Out	CTCTCTGTCGCATTTGAAAAAAAAGTTGATTGTTTTTAAGTTTGA	BT2383 Gene Knockout	
BT2383 5’ Out	TTTTTTTTCAAATGCGACAGAGAG	BT2383 Gene Knockout	
BT2383 3’ Down Xbal	GCGTCTAGAGGCTAAACAGGTCACCCGATA	BT2383 Gene Knockout	
BT2778 5’ Up Sal1	GCGGTCGACCGAAACACACTCCTCCAG	BT2778 Gene Knockout	
BT2778 3’ Out	ACTGCATGGAGAATGAATTTTGTGCGTGTATTATTTTAATATGAT	BT2778 Gene Knockout	
BT2778 5’ Out	AAAATTCATTCTCCATGCAGT	BT2778 Gene Knockout	
BT2778 3’ Down Xbal	GCGTCTAGAATTGACGATTTGATAATCAGT	BT2778 Gene Knockout	
qPCR Primers	**Primer**	**Use**	**Reference**
16S F (Bt/Bo)	ggtagtccacacagtaaacgatgaa	16S rDNA normalization	Rogers et al. 2013
16S R (Bt/Bo)	cccgtcaattcctttgagtttc	16S rDNA normalization	Rogers et al. 2013
BT0348F	tgcggcaaccaaattcaacaaa	Arabinose locus expression	This study
BT0348R	accaagtgccccattcgtcaag	Arabinose locus expression	This study
BT0792F	ccggatgggcagaacaaga	Xylose locus expression	This study
BT0792R	caccgcacgagaatcacaccag	Xylose locus expression	This study
BT0364F	tgaatggcggtaaggtaaaagaaca	Arabinan PUL expression	Rogers et al. 2013
BT0364R	cgggccggaagcgagtag	Arabinan PUL expression	Rogers et al. 2013
BT1280F	tgcgcggtacaaaatccatc	MOG PUL expression	Rogers et al. 2013
BT1280R	ggcggctgcggctgctc	MOG PUL expression	Rogers et al. 2013
BT1548F	tgtctttcgcaatccgtcacctc	*pgm* gene expression	This study
BT1548R	cttgcgcaccatcatcccagtat	*pgm* gene expression	This study
BT1763F	tgccgcatccgcttctatct	Levan PUL expression	Rogers et al. 2013
BT1763R	cgtccgtattgctcagtgttcagt	Levan PUL expression	Rogers et al. 2013
BT1986F	ccccgaccgcaacctgataa	*rpiB* gene expression	This study
BT1986R	gtgggcggggcgtagtgata	*rpiB* gene expression	This study
BT2805F	tccacgccccgatataatgtagg	Ribose PUL expression	Glowacki et al. 2019
BT2805R	accgtttgcaccccagaagtagtaa	Ribose PUL expression	Glowacki et al. 2020
BT3090F	atgctgaatgccgcccaata	Dextran PUL expression	Rogers et al. 2013

BT3090R	cgagaaaaccgccggatacata	Dextran PUL expression	Rogers et al. 2013
BT3332F	tgttcccggagccagtgttc	Chondroitin sulfate PUL expression	Rogers et al. 2013
BT3332R	ttcgtccagcgttttagtatcttcttt	Chondroitin sulfate PUL expression	Rogers et al. 2013
BT3680F	cgggaaattaaatatactgctacgaaact	Arabinogalactan PUL expression	Rogers et al. 2013
BT3680R	ctgccgggtctacattggtga	Arabinogalactan PUL expression	Rogers et al. 2013
BT3702F	gctattggcggggcattgg	Amylopectin PUL expression	Rogers et al. 2013
BT3702R	cagcggattttggggagagttcg	Amylopectin PUL expression	Rogers et al. 2013
BT3788F	aagcgtggggaaaaaggtaagg	α-Mannan PUL expression	Rogers et al. 2013
BT3788R	gctaaacgcgcccaatcataac	α-Mannan PUL expression	Rogers et al. 2013
BT3946F	caatcaatccggcaactcctgt	*rpe* gene expression	This study
BT3946R	atgaacttctgcccgccaaaac	*rpe* gene expression	This study
BT4114F	cgcaacggaagcactaacagg	Homogalacturona n PUL expression	Rogers et al. 2013
BT4114R	gggaagccgtctacaataaataaa	Homogalacturona n PUL expression	Rogers et al. 2013
BT4164F	gaaatgtaatgaatgatgcaaaaggtaga	Rhamnogalacturon an I PUL expression	Rogers et al. 2013
BT4164R	cgaaacgtccgtggaagaaagta	Rhamnogalacturon an I PUL expression	Rogers et al. 2013
BT4660F	agcccgacaaatacttccaacct	Heparin sulfate PUL expression	Rogers et al. 2013
BT4660R	tgtgcggcaaagtgtatcctaaag	Heparin sulfate PUL expression	Rogers et al. 2013
BT4671F	cagcgtggattggaatgtaagatgggtaa	Polygalacturonate PUL expression	Rogers et al. 2013
BT4671R	gtaattcttttgcgggccgtatgtgtagtc	Polygalacturonate PUL expression	Rogers et al. 2013

Table 4.2 RNAseq hits for arabinose and xylose compared to wild type Bt grown in glucose

Name	**Fold change Arabinose/ glucose**	**Fold change Xylose/glucose**	**Annotation**
BT0032	0.196	0.045	hypothetical protein
BT0215	N/A	14.539	iron uptake regulatory protein
BT0348	35.501	N/A	alpha-L-arabinofuranosidase
BT0349	34.634	N/A	hypothetical protein
BT0350	16.794	N/A	xylulose kinase (xylulokinase)
BT0351	11.152	N/A	hypothetical protein
BT0352	9.899	N/A	hypothetical protein
BT0353	11.484	N/A	putative sugar epimerase/aldolase
BT0354	13.432	N/A	hypothetical protein
BT0355	15.076	N/A	Na+/glucose cotransporter
BT0356	13.181	N/A	aldose 1-epimerase precursor

BT0525	0.153	N/A	outer membrane protein, function unknown
BT0551	0.139	N/A	Asparagine synthetase
BT0552	0.182	N/A	glutamate synthase, small subunit
BT0553	0.168	N/A	glutamate synthase, large subunit
BT0791	N/A	191.375	hypothetical protein
BT0792	N/A	138.075	xylulose kinase (xylulokinase)
BT0793	N/A	103.337	hypothetical protein
BT0794	N/A	176.870	D-xylose-proton symporter (D-xylose transporter)
BT0933	0.055	N/A	hypothetical protein
BT0970	0.188	N/A	haloacid dehalogenase-like hydrolase
BT1211	N/A	6.911	hypothetical protein
BT1272	6.882	N/A	FucR
BT1273	33.673	N/A	hypothetical protein
BT1274	20.438	N/A	hypothetical protein
BT1275	20.804	N/A	L-fuculose kinase
BT1276	20.896	N/A	hypothetical protein
BT1277	25.673	N/A	L-fucose permease
BT1419	0.139	N/A	hypothetical protein
BT1420	0.183	N/A	hypothetical protein
BT1456	N/A	5.403	thioredoxin (TRX)
BT1572	N/A	25.976	RNA polymerase ECF-type sigma factor
BT1604	N/A	15.085	cytochrome c biogenesis protein ResB
BT1605	N/A	8.053	cytochrome c biogenesis protein
BT1757	5.371	N/A	fructokinase
BT1758	14.207	N/A	glucose/galactose transporter
BT1759	8.735	N/A	levanase precursor (2,6-beta-D- fructofuranosidase)
BT1760	10.039	N/A	glycosylhydrolase
BT1761	9.887	N/A	hypothetical protein
BT1762	10.844	N/A	putative outer membrane protein, probably involved in nutrient binding
BT1763	8.690	N/A	putative outer membrane protein, probably involved in nutrient binding
BT1765	9.210	N/A	levanase precursor (2,6-beta-D- fructofuranosidase)
BT2156	0.045	0.047	putative sugar phosphate isomerase/epimerase
BT2157	0.045	0.043	hypothetical protein
BT2158	0.052	0.051	putative dehydrogenases and related proteins
BT2159	0.052	0.054	putative oxidoreductase
BT2178	0.044	N/A	hypothetical protein
BT2301	5.044	N/A	conserved protein found in conjugate transposon
BT2490	N/A	0.037	hypothetical protein
BT2555	0.021	N/A	hypothetical protein
BT2569	N/A	6.141	RNA polymerase ECF-type sigma factor
BT2803	9.944	7.306	ribokinase
BT2804	9.213	5.742	ribokinase
BT2805	5.917	N/A	putative outer membrane protein, probably involved in nutrient binding
BT2806	5.517	N/A	hypothetical protein
BT2807	6.233	N/A	hypothetical protein
BT2808	6.582	N/A	putative inosine-uridine preferring nucleoside hydrolase
BT2809	8.714	5.328	putative integral membrane protein
BT2988	N/A	0.191	hypothetical protein
BT3009	N/A	38.513	GH3 (gentobiase)
BT3024	0.118	0.191	putative outer membrane protein, probably involved in nutrient binding
BT3025	0.158	N/A	putative outer membrane protein, probably involved in nutrient binding
BT3026	0.194	N/A	glycosylhydrolase, putative xylanase
BT3027	0.146	N/A	hypothetical protein

BT3114	0.177	N/A	beta-galactosidase
BT3208	0.066	N/A	hypothetical protein
BT3221	0.187	0.176	hypothetical protein
BT3222	0.189	N/A	hypothetical protein
BT3344	0.062	0.070	hypothetical protein
BT3345	0.049	0.069	conserved hypothetical protein, putative outer membrane protein
BT3346	0.071	0.084	putative outer membrane protein, probably involved in nutrient binding
BT3347	0.085	0.088	hypothetical protein
BT3415	N/A	7.670	hypothetical protein
BT3537	0.163	0.196	hypothetical protein
BT3571	4.995	9.968	hypothetical protein
BT3572	5.262	9.262	hypothetical protein
BT3573	N/A	N/A	hypothetical protein
BT3574	N/A	6.489	hypothetical protein
BT3614	17.203	35.446	putative oxidoreductase
BT3615	11.803	24.525	hypothetical protein
BT3616	10.148	17.738	fucose permease
BT3617	N/A	8.187	sorbitol dehydrogenase
BT3669	N/A	5.567	hypothetical protein
BT3704	N/A	0.163	hypothetical protein
BT4038	6.407	N/A	putative outer membrane protein, probably involved in nutrient binding
BT4039	5.552	N/A	putative outer membrane protein, probably involved in nutrient binding
BT4040	5.089	N/A	putative galactose oxidase precursor
BT4159	N/A	6.799	hypothetical protein
BT4160	N/A	7.404	beta-galactosidase precursor
BT4225	N/A	0.145	hypothetical protein
BT4227	0.174	0.203	hypothetical protein
BT4294	7.176	N/A	hypothetical protein
BT4295	10.615	N/A	putative chitobiase
BT4296	10.456	N/A	hypothetical protein
BT4297	10.132	N/A	putative outer membrane protein, probably involved in nutrient binding
BT4298	10.432	N/A	putative outer membrane protein, probably involved in nutrient binding
BT4299	11.937	N/A	hypothetical protein
BT4384	N/A	5.115	hypothetical protein
BT4579	7.366	5.918	hypothetical protein
BT4672	N/A	5.005	hypothetical protein
BT4686	N/A	0.157	hypothetical protein
BT4715	N/A	11.610	non-specific DNA-binding protein Dps

Notes

This work is currently not submitted to a peer-reviewed journal article. However, the plan is to submit this work in the coming months to year after subsequent experiments are performed and analyzed. This work would not have been possible without helpful discussion and assistance from Pudlo, N. A., Porter, N. T., and Martens E. C.

References

1. Perrenoud, A. and U. Sauer, Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli. *J. Bacteriol.* **187**, 3171-3179 (2005).

2. Schleif, R., AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev*. **34**, 779-796 (2010).

3. Matthews, K.S. and J.C. Nichols, Lactose repressor protein: functional properties and structure. *Progress in Nucleic Acid Res. And Mol. Biol.* **58**, 127-150 (1997).

4. De Lay, N. and S. Gottesman, The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J. Bacteriol.* **191**, 461-476 (2009).

5. De Lay, N., D.J. Schu, and S. Gottesman, Bacterial small RNA-based negative regulation: Hfq and its accomplices. *J. Biol. Chem.* **288**, 7996-8003 (2013).

6. Soper, T., et al., Positive regulation by small RNAs and the role of Hfq. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9602-9607 (2010).

7. Fimlaid, K.A. and A. Shen, Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. *Curr. Opin. Microbiol.* **24**, 88-95 (2015).

8. Saujet, L., et al., Genome-wide analysis of cell type-specific gene transcription during spore formation in Clostridium difficile. *PLoS Genet.* **9**, e1003756 (2013).

9. Feklistov, A., et al., Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* **68**, 357-376 (2014).

10. Trevino-Quintanilla, L.G., J.A. Freyre-Gonzalez, and I. Martinez-Flores, Anti-sigma factors in E. coli: common regulatory mechanisms controlling sigma factors availability. *Curr. Genomics* **14**, 378-387 (2013).

11. Paget, M.S., Bacterial sigma factors and anti-sigma factors: structure, function and distribution. *Biomolecules* **5**, 1245-1265 (2015).

12. Fang, F.C., Sigma cascades in prokaryotic regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4933-4934 (2005).

13. Saujet, L., et al., The key sigma factor of transition phase, SigH, controls sporulation, metabolism, and virulence factor expression in Clostridium difficile. *J. Bacteriol.* **193**, 3186-3196 (2011).

14. Xu, J., et al., A genomic view of the Human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**, 2074-2076 (2003).

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

16. Martens, E.C., et al., Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J. Biol. Chem.* **284**, 18445-18457 (2009).

17. Briliute, J., et al., Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat Microbiol*. **4**, 1571-1581 (2019).

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

19. Martens, E.C., et al., Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).

20. 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).

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

22. Rogers, T.E., et al., Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol. Microbiol.* **88**, 876-890 (2013).

23. Schwalm, N.D., 3rd, G.E. Townsend, 2nd, and E.A. Groisman, Prioritization of polysaccharide utilization and control of regulator activation in Bacteroides thetaiotaomicron. *Mol. Microbiol.* **104**, 32-45 (2017).

24. Tuncil, Y.E., et al., Reciprocal prioritization to dietary glycans by gut bacteria in a competitive environment promotes stable coexistence. *MBio* **8**, e01068-17 (2017).

25. Siegel, L.S., et al., Cyclic adenosine 3',5'-monophosphate levels and activities of adenylate cyclase and cyclic adenosine 3',5'-monophosphate phosphodiesterase in Pseudomonas and Bacteroides. *J Bacteriol*. **129**, 87-96 (1977).

26. Cao, Y., et al., cis-Encoded small RNAs, a conserved mechanism for repression of polysaccharide utilization in Bacteroides. *J. Bacteriol.* **198**, 2410-2418 (2016).

27. Comstock, L.E., Small RNAs repress expression of polysaccharide utilization loci of gut Bacteroides species. *J. Bacteriol.* **198**, 2396-2398 (2016).

28. Brooks, B.E. and S.K. Buchanan, Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. *Biochim. Biophys. Acta.* **1778**, 1930-1945 (2008).

29. D'ella, J.N. and A.A. Salyers, Effect of regulatory protein levels on utilization of starch by Bacteroides thetaiotaomicron. *J. Bacteriol.* **178**, 7180-7186 (1996).

30. Patel, E.H., et al., Rhamnose catabolism in Bacteroides thetaiotaomicron is controlled by the positive transcriptional regulator RhaR. Res. Microbiol. **159**, 678-684 (2008).

31. Hooper, L.V., et al., 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**, 9833-9838 (1999).

32. Ravcheev, D.A., et al., Polysaccharides utilization in B. theta; comparatice genomics reconstruction of metabolic and regulatory networks. *BMC Genomics* **14**, (2013).

33. Schwalm, N.D., 3rd, G.E. Townsend, 2nd, and E.A. Groisman, Multiple signals govern utilization of a polysaccharide in the gut bacterium Bacteroides thetaiotaomicron. *MBio* **7**, e01342-16 (2016).

34. Townsend, G.E., 2nd, et al., Dietary sugar silences a colonization factor in a mammalian gut symbiont. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 233-238 (2019).

35. Wegorzewska, M.M., et al., Diet modulates colonic T-cell responses by regulating expression of a Bacteroides thetaiotaomicron antigen. *Sci. Immunol.* **4**, eaau9079 (2019).

36. Goodman, A.L., et al., Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279-298 (2009).

37. Kadowaki, T., et al., A two-component system regulates gene expression of the type IX secretion component proteins via an ECF sigma factor. *Sci. Rep.* **6**, 23288 (2016).

38. Holdeman, L.V.C., E.P. ; Moore, W.E.C., *Anaerobe Laboratory Manual* (Virginia Polytechnic Institute and State University Anaerobe Laboratory, 1977).

39. Koropatkin, N.M., et al., Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105-1115 (2008).

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

41. Speer, M.A., Development of a genetically modified silage inoculant for the biological pretreatment of lignocellulosic biomass, in Agricultural and Biological Engineering. (Penn State University Disseration, University Park, PA, 2013).

Chapter V

Discussion

Introduction

Carbohydrate and nutrient utilization by members of gut microbiota is critical for their survival within the competitive gut environment. The most successful strains could be thought of as belonging to the most abundant bacterial phyla of the Firmicutes and the Bacteroidetes. This is likely due to their vast nutrient degrading abilities, allowing them to switch to a wide range of available nutrients. These phyla, especially members of the Bacteroidetes are also important for host health through the generation of short-chain fatty acids, colonization resistance, gut immune system maturation and barrier function, and several other functions[1-4]. The *Bacteroides* have a particularly large capacity for degrading host, plant (dietary), bacterial, and algal polysaccharides with diverse linkages and compositions through products encoded in genetic loci termed polysaccharide utilization loci (PULs)[5-10]. However, except for a few studies[11-13], little is known about the fate of other nutrients such as monosaccharides and nucleic-acid derived substrates or the detailed mechanistic strategies used by *Bacteroides* to catabolize these substrates. Work presented in this dissertation offers new insights into the variety of carbohydrates that *Bacteroides thetaiotaomicron* (*Bt*) utilizes for growth through the identification of the Ribose Utilization System (Rus) PUL. This is only the second PUL described that targets a monosaccharide and the first to be implicated in processing of a nutrient (ribose-1-phosphate), formed from the actions of an upstream, genomically unlinked pathway for nucleoside scavenging. Importantly, this dissertation describes the actions of the Rus PUL-encoded ribokinases, which is the first described eubacterial ribokinase with 1'-phosphorylating abilities. Further, the work within this dissertation highlights the complexity of host mucosal glycan foraging abilities of *Bt* and the *in vivo* relevance of these systems. Lastly, this work has described new regulatory genes and responses within *Bt* that are responsible for metabolism of many poly- and monosaccharides.

Chapter Summary and Further Results

The underlying current that links the chapters of this dissertation together has been an in-depth examination of the substrates that *Bt* catabolizes, and the mechanisms governing their utilization and regulation. The genes and products required for ribose and nucleoside metabolism were identified and characterized in Chapter II, a PUL-encoded, antigenic protein was identified that was responsive to host glycans (Chapter III), genes that likely code for regulatory proteins and new mechanisms of regulation were uncovered (Chapter IV), and discussed below, PUL-encoded machinery responsible for degrading host-derived nutrients were examined. In sum, this work has advanced the known substrate-degrading abilities and regulatory mechanisms of *Bt* with important implications for the broader Bacteroidetes phylum, important members of the human gut microbiota, oral cavity, and in the environment[14,15].

Of the characterized PULs in *Bacteroides*, most are patterned similar to the starch utilization system (Sus), which was the first PUL described[16]. This system contains a regulatory protein, (SusR), a TonB-dependent outer membrane (OM) transporter (SusC), an OM lipoprotein required for substrate binding (SusD), an OM glycoside hydrolase (SusG), periplasmic amylases (SusA and SusB), and accessory OM binding proteins (SusE/SusF)[17-20]. What typically denotes a PUL is the presence of SusC and SusD homologs, which are often referred to as SusC- or SusD-like proteins[21]. The accessory proteins found in each PUL are what delineates the cognate substrate that is degradation. These often include OM and periplasmic glycoside hydrolases, sulfatases, esterases, sialidases, and regulators[22,23]. Additionally, PULs almost ubiquitously target substrates that are large and complex in both linkage and structure such as polysaccharides, while previously thought to be unimportant or not required for catabolism of smaller nutrients such as monosaccharides[24]. Therefore, the Rus PUL described in Chapter II that targets ribose is unique in many regards.

First and foremost, the Rus is only the second described PUL that encodes functions required for the utilization of a monosaccharide (ribose) (Figure 2.1C and D), with the fructan PUL previously being characterized for fructose utilization[25]. Interestingly, these systems appear to be patterned in similar ways, both containing a dedicated inner membrane permease and a sugar-specific kinase, which suggests these PULs are adapted and built around monosaccharide utilization. Further, both ribose and fructose are connected as sugars of the pentose phosphate pathway (PPP). This supports my observations that fructose suppresses *rus* expression (Figure

2.2A), while ribose upregulates fructan PUL expression (Figure 2.9G). Although these two systems are likely connected, an *in vivo* competition experiment did not clearly show how these systems may work together (Figure 2.8G and H). Components of the fructan system are required for fructose growth while the cognate substrate(s) are fructose polymers such as the polysaccharide levan, inulin, and fructooligosaccharides[25] and the SusC/D-like proteins delineate which fructose polysaccharide is catabolized between different isolates of the same strain and are required for growth on these substrates[26].

In contrast, the RusC and RusD proteins (SusC/D homologs) within the Rus PUL are dispensable for growth on all ribose-containing substrates that I tested for growth (Figure 2.7B,C,G, and S-U). Although this phenomenon has not previously been described, (SusC/D homologs are usually required), this may be due, in part, to not being able to identify the cognate ribose polymer that RusD binds and RusC imports. However, it is also conceivable that this system has evolved a mechanism independent of these PUL-encoded proteins, or perhaps *Bt* has a secondary system containing OM import and binding machinery (Figure 2.9G) that ribose-containing compounds can be shuttled through. A similar mechanism has previously been suggested for the Sus PUL[27]. Although ribose and other monosaccharides can cross the OM of gram negative bacteria through proteins such as OmpF and related Omps[28], nucleosides typically cross the OM through facilitated transporters such Tsx-porins[29]. *Bt* does not have a good homolog of these nucleoside-related transport proteins however, there are numerous uncharacterized OM transporters in *Bt* that could potentially serve this role[30]. Further, this system is unique in the sensing and catabolism of nucleosides as it requires the presence of ribose and the novel regulatory protein RusR (Figure 2.1C and Figure 2.6D).

RusR is a novel regulatory protein, but not just in terms of *Bt* PULs or metabolism. The amino acid sequence coding for RusR (encoded by *BT2802*), does not have homology to previously described families of proteins or regulators. Predictive algorithms have found a helix-turn-helix domain responsible for DNA-binding, this evidence alone is not sufficient to speculate if the sugar bound is ribose or the phosphorylation status of ribose that RusR recognizes or the mechanism of action, other than saying it is a positive acting regulator based on my results. This warrants further characterization through X-ray crystallography and binding assays. Additionally, there are likely many further uncharacterized regulatory proteins in *Bt*[13,31], and the work described in Chapter IV has likely identified a few of these.the most important discovery

from this work was that the ribokinases phosphorylation abilities. RusK1 and RusK2 are the first described ribokinases in eubacteria able to generate ribose 1,5 bisphosphate (PRibP) from ribose-1-phosphate (R1P) (Figure 2.9F). Additionally, RusK2 is the first described ribokinase able to phosphorylate in the 1' position of an already 5'phosphorylated ribose (R5P). This mechanism of phosphorylating R1P derived from the actions of a genomically unlinked nucleoside scavenging, nucleoside phosphorylase (BT4554) (Figure 2.9A-C), is the first time that PUL-encoded functions have been documented to synergize with upstream, non-PUL encoded machinery for the catabolism of a nutrient. The mechanism of ribose and nucleoside utilization had important implications for *Bt in vivo* in mice fed a fiber-rich (FR) diet. Those implications being that the encoded ribokinases and nucleoside phosphorylases are required for successful competition on the FR diet when source(s) of ribose (likely nucleosides) are present.

The original impetus for studying the Rus PUL was that this PUL with unknown function was upregulated in mice fed a fiber free (FF) diet or in neonate mice dependent on mothers' milk for nutrition, which is also a FF condition (Figure 2.1A). This result suggested that this PUL may target an endogenous nutrient source as it behaved similar to PULs responsive to host-glycan degradation[22]. However, when competed against wild type *Bt*, it was clear that *rus* was required for a competitive advantage only in a FR-diet and not the FF-diet (Figure 2.3A-F). This effect was mediated by ribose or a ribose-containing source (nucleosides) from the FR diet (Figure 2.4A-C). Further, this effect was narrowed down to the combined actions of both RusK1 and RusK2, but not their individual actions (Figure 2.4G and Figure 2.5A), as well as the unlinked, nucleoside phosphorylase, BT4554 (Figure 2.9D). Although the exact substrate mediating this effect *in vivo* was not conclusively identified, my results point to a nucleoside-containing nutrient. This study demonstrates the importance of studies on bacterial metabolism, as predicted functionality and metabolic KEGG maps of the PPP were at an incomplete view of how *Bt* processes ribose and nucleosides.

Although the ribose results in Chapter II mostly ruled out this nutrient being important in an FF-diet, it is clear that the FF-diet is an important dietary condition to study as it resembles a Westernized diet rich in fats and simple sugars while depleted of complex dietary fibers. This style of dietary consumption can predispose individuals to colonization by invading pathogens[2], worsen preexisting health conditions such as inflammatory bowel disease (ulcerative colitis)[32] and has been implicated in a myriad of other disorders such as obesity[33,34]. Several studies have

directly linked the effects of a Westernized diet to actions of the gut microbiota, and although *Bt* is often a beneficial gut symbiont that degrades complex dietary fibers, it is also an efficient degrader of the host mucosal lining, rich in *O*- and *N*-linked glycans[7,22]. This ability is associated with the development of colitis in the context of FF diets and predisposed genetic conditions[35,36]. *Bt* degrades mucosal-derived glycans through PUL-encoded machinery. Many of these PULs encode sulfatases which act to remove the sulfate groups from highly sulfated host glycans[36-38]. Work in Chapter III, has built upon foundational knowledge of the sulfatase-dependent mechanism(s) and the generation of outer-membrane vesicles (OMVs) containing sulfatases and other mucin-degrading enzymes[36,39]. The intimate connection between gut bacteria and the host, allows the host immune system to generate a response that is either tolerogenic (T_{regs}) towards resident gut symbionts or more skewed towards $CD4^+$ effector T cells (T_{effs}) when invading pathogens are present, or when resident bacteria perform deleterious functions towards the host[40,41]. Accordingly, the type of T cells generated can be influenced based on the nutrients available for degradation by the gut microbiota, as the antigens that the host senses are different based on what proteins the bacterium has expressed in response to nutrients. In order to test how the host responds to individual bacteria in a diet-dependent manner, a T cell-based assay termed, BθOM was developed specific for an outer membrane (OM) antigen found in OMVs from *Bt* (*Bt*, Bθ) (Figure 3.1 and Figure 3.2B-C).

It was observed that upon screening a T cell hybridoma line for cells active in response to *Bt* antigens, that different *in vitro* growth conditions of *Bt* caused differential T cell activation levels (Figure 3.1). *In vivo* characterization of the T cell response revealed that BθOM T cells differentiate into both T_{regs} and T_{effs} (Figure 3.6A-C and Figure 3.7A-F). An important dynamic that is dependent on T_{regs} being dominant by regulating the T_{effs} to prevent *Bt*-induced colitis (Figure 3.7C) that was observed *in vivo* in *Rag1*$^{-/-}$ mice. Upon deletion of the BθOM T_{regs}, a proinflammatory state was observed (Figure 3.8A-B). The most obvious next question was what is the *Bt*-OM antigen that the BθOM T cells recognize and why is it dependent on host diet.

I positively identified the antigen as an OM SusE/F-like lipoprotein encoded by *BT4295* by using a transposon mutagenesis library of *Bt* (Figure 3.10A). This was refined to a single peptide and was mapped to a single amino acid substitution that abrogated the T cell response to BT4295 (Figure 3.10 F-G and Figure 3.12B). Interestingly, BT4295 is found within one of the many mucosal glycan-responsive PULs in *Bt*. However, unlike the aforementioned PULs that

contain sulfatases, the PUL circumscribed by *BT4294-4300* does not contain sulfatases, or hydrolytic enzymes (Figure 3.10B). Rather, it appears to be involved in the binding and import of a glycan, although hydrolytic or degrading function(s) cannot be ruled out as three of these proteins are hypothetical and contain no previously characterized conserved protein domains. In addition to the BT4295 antigen, the transposon approach located additional potential antigens (Figure 3.11B).

Not surprisingly, due to being in the same PUL as BT4295, BT4298, a SusC-like protein also weakly stimulated T cells (Figure 3.10C). Intriguing are the remaining hits that correspond to the genes, *BT1220-1223*. These genes code for proteins within the PPP and perhaps this represents a similar phenomenon as was observed in Chapter II, with an upstream or downstream metabolic reliance on PUL-encoded and unlinked functions. These findings required follow up study of the initial observation that different *in vitro* growth media affects the expression of the BT4295 T cell antigen. Finding that the salts: K2HPO4 and NaCl and mucin *O*-glycans, strongly increased expression of BT4295 and downstream T cell activation (Figure 3.14A-D). Although the mechanism of how salts stimulate the expression of this glycan are unknown, it could be a response that *Bt* has developed to *in vivo* stresses associated with salt intake. The mucosal *O*-glycan response is more expected with the observation of this system being active in FF dietary conditions where host polysaccharides would be used as a nutrient. Further, expression of this antigen was reduced or repressed *in vivo* when mice were provided water containing 30% w/v glucose (Figure 3.15C-D). This result suggests that dietary glucose can alter the expression of this system, however this was only performed on a standard chow diet similar to the FR diet which contains many plant derived polysaccharides. In the context of a FF diet, where glucose is already a prominent constituent and when *Bt* physiology shifts to catabolizing host polysaccharides, the same suppression may not be observed and should be further tested in follow-up studies.

As Chapters II and III have demonstrated, *Bt* physiology is dynamic. In Chapter IV, I examine some of the ways in which *Bt* regulates and responds to different nutrients, uncovering how monosaccharides alter nutrient hierarchy and competitive fitness, as well as describe the global regulatory proteins. The characterized regulatory mechanisms in *Bt* are diverse, occurring locally within metabolic loci, globally with Crp-like mechanisms, and small RNAs[42-44]. These local regulatory proteins can broadly be classified as hybrid two-component systems (HTCS),

ECF-σ/anti-σ pair, SusR-like activators, LacI-type repressors and GntR-like transcription factors[11,12,22,31,45-47]. These mechanisms allow *Bt* and related *Bacteroides* to degrade and catabolize a diverse set of carbohydrate nutrients in complex mixtures[48-50]. However, previous studies have not given much focus to monosaccharides despite these substrates being monomers of polysaccharides, and as demonstrated in Chapter II, monosaccharides may be important *in vivo* as was the case for ribose and nucleosides.

The results from this study showed that the monosaccharide ribose and presence of the genes response for its catabolism (*rus*) altered the *in vitro* transcriptional response of polysaccharide utilization genes for several glucose-containing polymers as well as a few other arabinose containing nutrients (Figure 4.1). I further confirmed that the Δ*rus* strain competed worse *in vitro* in media containing glucose as the sole carbon source or in a rich media that had glucose as one of its components (Figure 4.2G-H). This result may be due to the presence of nucleosides or ribose in the rich media or may also serve to explain the altered hierarchy of glucose-based polysaccharides. Extending these findings of monosaccharides affecting the transcription of other genes, I examined the global response to arabinose and xylose by RNAseq analysis compared to glucose (Figure 4.3). Interestingly, in arabinose, *Bt* upregulates the PUL (*BT4294-4299*) that contains the T cell antigen BT4295 from Chapter III. This suggests that the hits in the PPP initially observed in the transposon screen (Figure 3.11B), may be related to arabinose being able to upregulate this PUL in addition to host mucin-*O*-glycans (MOG). Additionally, several previously uncharacterized genes predicted to code for orphan ECF-σ regulators were upregulated in both arabinose and xylose (Figure 4.3) as well as in ribose (Figure 2.9G).

These orphan regulators are not located adjacent to clear metabolic genes and are without anti-σ factors. I was unable to make deletions in the two most highly differentially regulated factors, *BT2184* and *BT2569*. However, this result propelled me to make deletions in other orphan ECF-σ factors and orphan LacI-type regulators. Several of the LacI regulators acted as repressors during growth on uronic acid containing mono- and polysaccharides as well as rhamnose (Figure 4.6). Further, only the orphan ECF-σ, BT2492, when deleted caused growth defects on 12 of the polysaccharides *Bt* normally catabolizes (Figure 4.4). The most severe defect was seen for growth in MOG. It is these sulfated, host-derived polysaccharides that are the focus of the final part of this discussion.

Sulfatase containing and host glycan responsive PULs are required for growth on MOG

Within the 88 PULs that *Bt* encodes, there are at least 16 known to be transcriptionally active towards host-derived glycans such as MOG, chondroitin sulfate (CS), heparin sulfate (HS), and keratin sulfate[22,45]. With the degradation of host mucosal glycans being a driver of colitis[36], the importance of elucidating the function of these PULs and the machinery they encode is of critical importance. To this end, I've generated individual mutant strains lacking entire mucin PULs as well as a strain lacking 11 different PULs (11x mutant) containing nearly all of *Bt's* sulfatase genes (Figure 5.1A). The methodology for the order in which PULs were sequentially deleted was based on several factors *i)* best way to delete the most sulfatases, *ii)* deletion of hydrolytic enzymes (fucosidases, sialidases, glycoside hydrolases), and *iii)* deletion of the most highly expressed PULs *in vivo*. I had hypothesized that deletion of successive PULs would lead to a step-wise decrease in the ability to grow on MOG as a sole carbon source. This is largely what I observed, with an initial decrease in growth at the 3x deletion (Figure 5.1C), with slight decreases in growth moving through to the 7x deletion strain (Figure 5.1D-G). A substantially larger defect was seen in the transition to the 8x deletion strain that extend to the 9x and 10x strains (Figure 5.1H-J). I continued the process of deleting strains by completing deletion of *BT4240-50,* yielding the 11x deletion strain (Figure 5.2A).

The large growth defects associated with the 8x and 11x strains suggested that functions encoded in these individual PULs were key to MOG catabolism and led to the hypothesis that an individual deletion strain containing only these genes would also exhibit a strong growth defect. To test, I used single deletion strains and grew these *in vitro* on MOG as a sole carbon source, however none of the individual mutant strains approached the defects seen in the 8x or 11x deletions arguing against a model in which a single PUL is most important. Rather this result supports a model of step-wise degradation (Figure 5.2 B-E). Additionally, a deletion strain of the anaerobic sulfatase maturating enzyme (anSME or ΔChuR) was tested, as this enzyme is required for active sulfatases[37]. Interestingly, this deletion strain did not match the magnitude of the growth defect of the 11x deletion strain (Figure 5.2F). This result suggests that the ability to utilize MOG is not solely dependent on sulfatases but other PUL encoded functions.

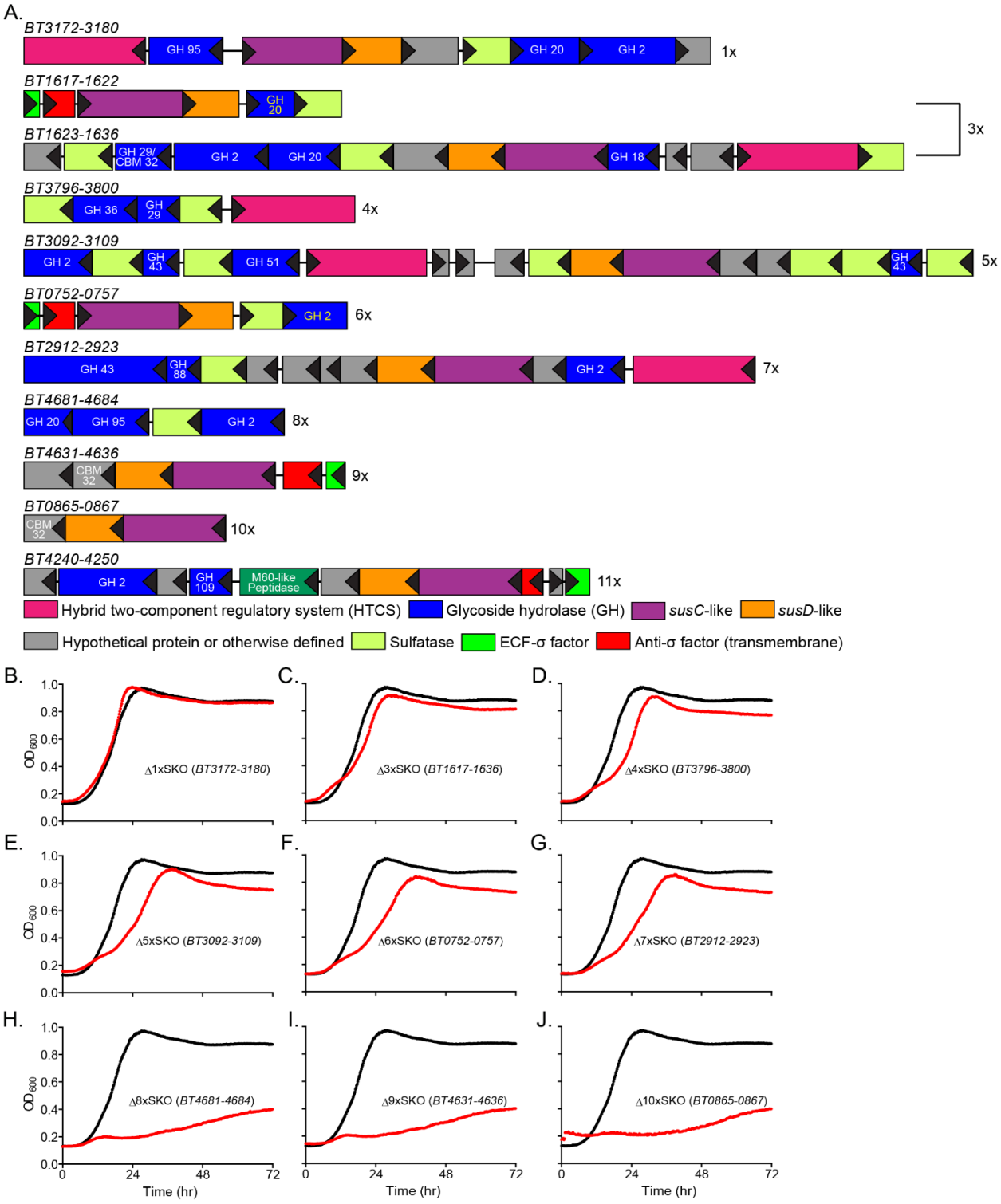


Figure 5.1 Sulfatase PULs code for many additional functions required for degradation and growth in MOG.

A) Schematic of the architecture of PULs deleted individually or in succession leading to a 11x mutant strain lacking all of the PULs shown. Genes are sized relative to the amino acid length and color coded according to predicted function. A key of the predicted functionalities is included as a key at the bottom of this panel. B-J) Growth curves of wild type Bt (black curves) or the indicated mutant strain (red curves on 10 mg/ml MOG). The deleted PUL gene locus tag numbers are shown on each panel. All growth curves were performed on the same time scale and same absorbance scale.

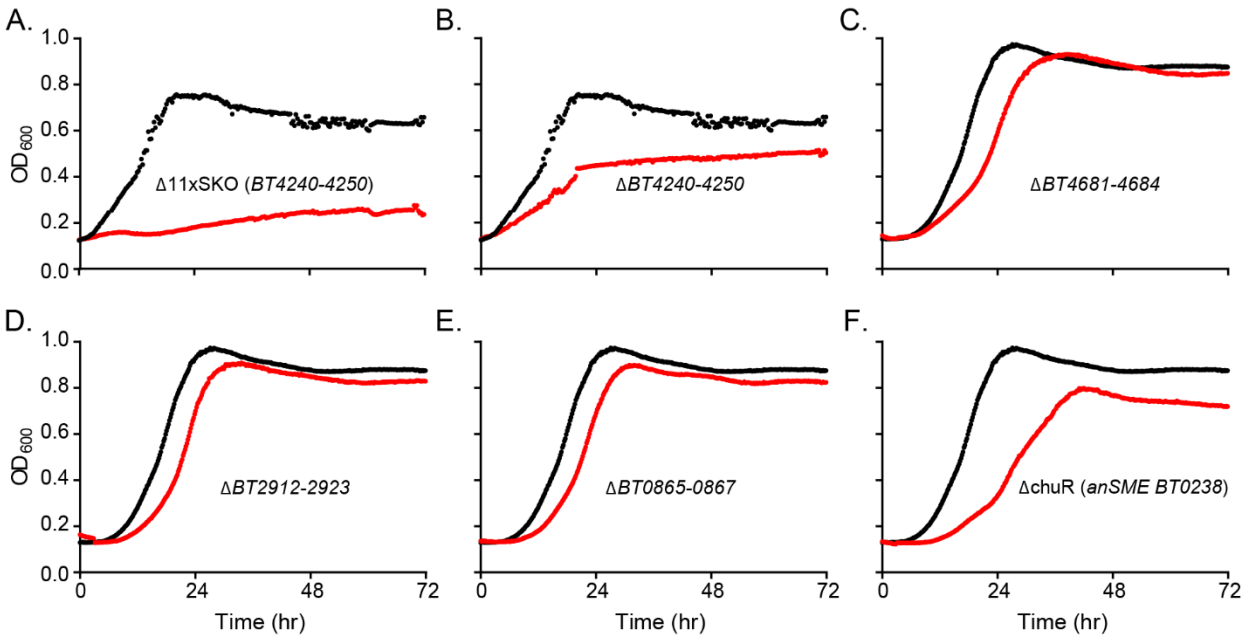


Figure 5.2 Individual MOG-responsive PUL deletions are not sufficient to abrogate growth.

A-F) Growth curves of wild type Bt (black curves) or individual deletion strains lacking a single PUL or a single gene (red curves) on 10 mg/ml MOG. The same time scale and absorbance scale was used for all panels.

I further followed up on this to ensure that orphan sulfatases not found within PULs were not involved in catabolism of MOG. No differences in growth compared to wild type *Bt* in either individual or a 4x deletion strain of orphan sulfatases was observed (Figure 5.3A-B). Further, none of the deletion PULs or orphaned sulfatases exhibited growth defects on non-MOG substrates such as glucose (Figure 5.3C-E).

Lastly, in addition to MOG, host-derived CS and HS can also be used as carbon sources by *Bt*. These polysaccharides are structurally similar to those found in bacterial capsules. Although throughout this thesis the prospect of *Bt* and related *Bacteroides* acting to degrade bacterial capsules has only been mentioned in the introduction (Chapter I) and in Chapter II with certain bacteria putting ribose into their capsules, it is an important underexplored nutrient niche. In order to examine possible capsule degrading abilities *in vivo,* I took advantage of the following *1)* knowledge of *Bt* degrading HS via PUL-encoded functions and *2)* the *E. coli* Nissle 1917 strain that produces a highly sulfated capsular polysaccharide termed heperonsan. This polysaccharide capsule is similar in structure to HS[51,52]. I performed an *in vivo* competition of wild type and a ΔHS PUL mutant in *Bt* in the presence of the *E. coli* Nissle strain and observed that in a FF diet, the capsule from this *E. coli* may be used as a carbon source, until heparan sulfate containing water was supplemented (Figure 5.4B).

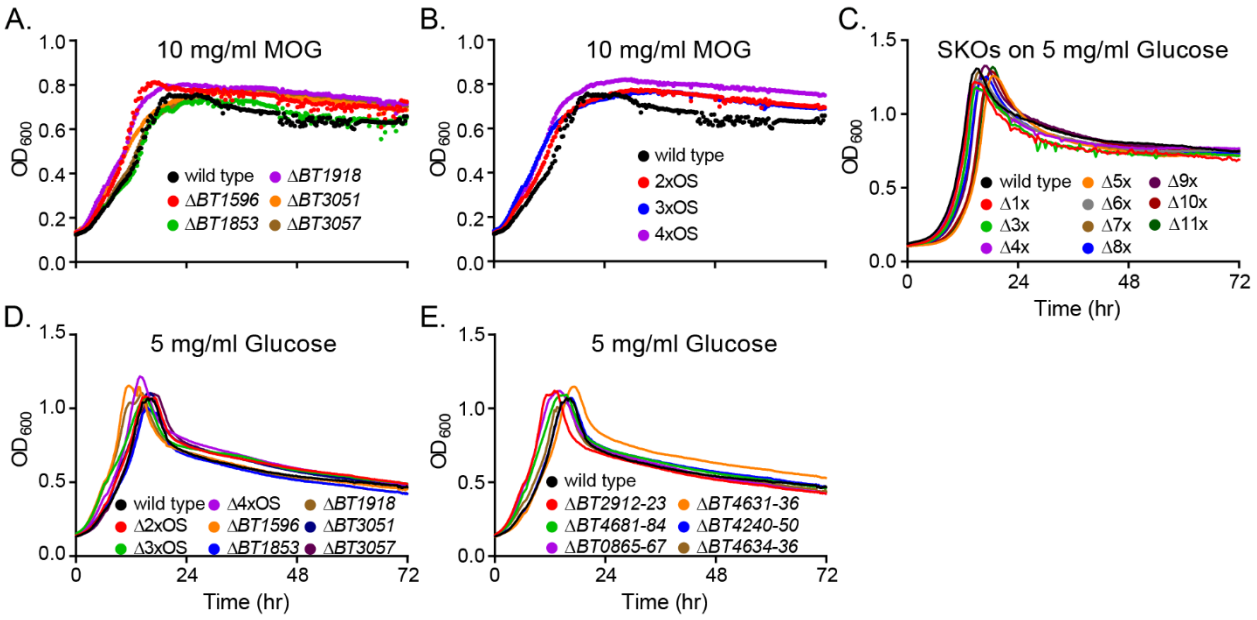


Figure 5.3 Orphan sulfatase deletions do not affect MOG growth and MOG PUL deletions grow normally on glucose.

A-E) Wild type Bt (black line) or individual PUL or single gene deletion strains grown on MOG (A-B) or glucose (C-E). Deletion strains are color coded as listed in the key found within each panel.

The behavior of the ΔHS PUL mutant are interesting. This strain remained stable in both the FR and FF diet (Figure 5.4A-B), until heparan sulfate was introduced in the water. This suggests that although wild type *Bt* can catabolize HS, there are other available nutrients that the ΔHS PUL strain can utilize and establish and maintain a niche.

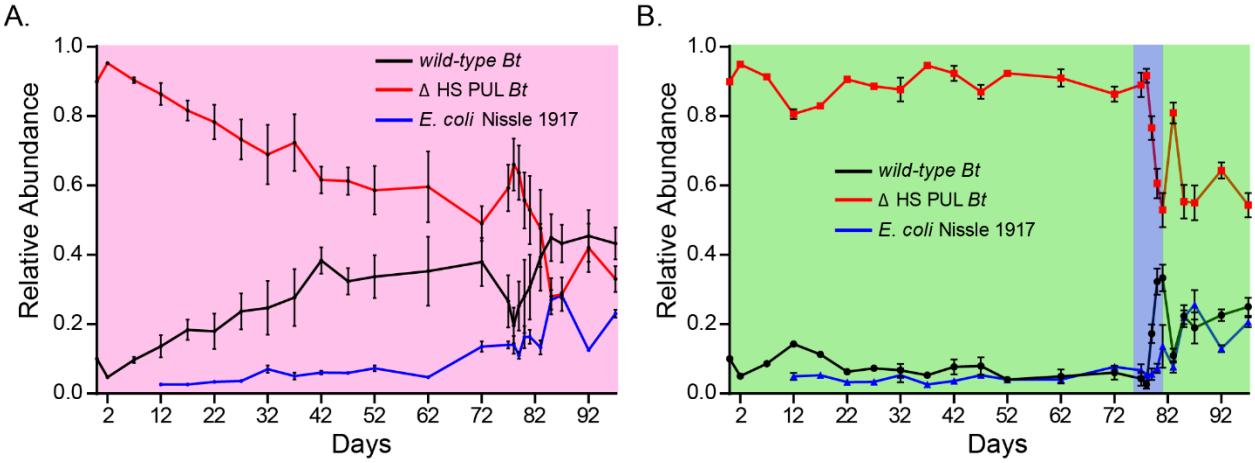


Figure 5.4 Heparin sulfate competition in vivo between Bt and E. coli Nissle 1917.

A-B) Relative fecal abundance enumerated by qPCR of wild type Bt (black line), a mutant lacking the genes for heparin sulfate utilization (red line), and E. coli Nissle 1917 (blue line) on mice fed a fiber rich diet (A) or a fiber free diet (B). The initial time point at day 0 is the inoculum abundance for both Bt strains, the E. coli strain was introduced at day 12. B) The shaded blue boxed region represents where water containing 0.5% w/v heparan sulfate containing water was introduced to the mice ad libitum for 4 days. At each day the average and SEM of n=3 mice is shown.

The combined findings of these results have allowed insight into the complex mechanisms and PULs that *Bt* uses to catabolize highly sulfated host and bacterial capsule-derived glycans. The findings described in this discussion, coupled with the results of the previous chapters have helped to expand the known PUL-encoded functions that are important for successful competition *in vivo*, and broadened the diversity of substrates that *Bt* degrades.

Future Work

Determination of monosaccharide nutrient utilization by Bacteroides

The bulk of this dissertation research has focused on nutrient strategies, mainly PUL-encoded, that allow *Bt* to grow on various carbohydrates. Prior to this work, an in-depth mechanistic study of monosaccharide utilization had not been performed for *Bt*. The finding that ribose and nucleoside utilization occur via PUL-encoded mechanisms while requiring upstream hydrolase functions is an important step towards recognizing additional substrates that *Bt* degrades. In Chapter II, I observed the ribose-induced synergistic metabolism of the sugar's deoxyribose, lyxose, and UDP sugars. Future work should focus on expanding the known repertoire of substrates that can be used as sole carbon sources by testing additional mixtures of carbohydrates to examine if synergism is a common mechanism that *Bacteroides* use. Additional substrates derived from anthropogenic substances such as artificial sweeteners and emulsifiers, and more carbohydrates from environmental derived polysaccharides of fungus and algae should also be investigated. Perhaps the area where the most inroads could be made towards understanding the metabolism and *in vivo* relevance will be the testing of exopolysaccharides and capsular polysaccharides from gut bacteria. Many PULs remain uncharacterized and these substrates have largely been overlooked due to issues in isolation techniques and difficulty in characterization of the monomers present in these polysaccharides. It is likely that the best way to approach this is through growth analysis and transcriptional responses in order to identify genes that respond to these substrates.

Further defining of the sulfatase and fucosidase mechanisms of host glycan catabolism

Work from Chapter III and Chapter V demonstrate the importance of *Bt* PULs involved in degrading host mucosal glycans. Although deletion of 11 PULs in combination led to a large reduction in growth on MOG, it was not a complete elimination of growth. In order to continue

to erode the ability of *Bt* to degrade host polysaccharides, additional deletions in the background of the 11x strain will need to be performed. There are at least 3 additional host-glycan responsive PULs that need to be assayed: *BT4294-99,* (identified in Chapter III), *BT3461-3507* [22], and *BT4355-59* [49]. The observance that additional, non-sulfatase enzymes must be required for growth will need to be explored through genetic deletions of fucosidases for example. The deletion of additional PULs in the 11x background may lead to additional enzymes that are important for catabolism of host glycans. Further, *in vivo* competitions of these strains (*i.e.* the 11x deletion versus wild type) should be performed in both FR and FF diets to examine the impact that lacking these PULs has for different diets. Further, *in vitro* competitions and additional growth and transcriptional experiments need to be performed for the ΔHS PUL versus wild type *Bt* versus *E. coli* Nissle 1917. It is possible that *in vivo,* the *E. coli* strain is not producing the heparan containing capsule and therefore is the reason why it is not outcompeted. Together, these studies into host glycan degradation may one day guide the development of drugs to inhibit the mucosal degrading abilities by *Bt* and related mucosal glycan degrading bacteria through inhibiting enzymes.

Defining the regulation strategies in Bt for nutrient metabolism

In Chapter IV I described several new regulatory proteins as well as the earlier description of RusR in Chapter II. In this dissertation these proteins were not characterized biochemically. Rather, I infer the importance of these through growth-based assays of deletion strains lacking genes coding for these predicted regulators. This approach coupled with transcriptional assays allowed me to comment on the necessity of these proteins for global regulation of carbohydrate nutrients, but did not provide insight into the direct mechanisms of these proteins. Moving forward, research will focus on purification efforts of these regulatory proteins and I will perform binding assays such as isothermal titration calorimetry or gel-shift assays to detect the ligands that these proteins recognize.

Final Conclusions

The work presented in this dissertation has added to the known substrate degrading abilities of *Bt,* and this is the first in-depth mechanistic model for the utilization of a monosaccharide ribose and the degradation nucleosides within *Bt.* Further, the characterization of mucin utilization PULs has advanced knowledge of degrading abilities towards host glycans. I

hypothesize that the observed co-dependence on non-Rus PUL encoded enzymes is likely a common occurrence for other PULs that currently lack knowledge of the cognate substrate they can degrade. Further, the *in vivo* approaches used in the studies here, demonstrate the importance of using multiple diets even when predicted functionality indicates the requirement of these systems is restricted to specific conditions. I also observed that monosaccharides affect the transcription of other PULs and metabolism-related genes, which likely influences the observed nutrient hierarchies of *Bt*. This result together with the deletion strains of Δ*rusR* and Δ*BT2492* suggest that the mechanisms that *Bt* uses to switch between different carbon sources is much more complex and fluid than previously appreciated. This ability to rapidly change nutrient utilization based on the substrates present, no doubt contributes to the survival of *Bt* and related Bacteroidetes. Ultimately, the studies present in this dissertation lay the groundwork for additional inquiry into PULs that currently lack mechanistic studies. In the future, studies like the ones described here will hopefully aid in designing treatments for health disorders caused by the gut microbiota through modulating the niches available to specific bacteria.

References

1. Buffie, C.G. and E.G. Pamer, Microbiota-mediated colonization resistance against intestinal pathogens. *Nat. Rev. Immunol.* **13**, 790-801 (2013).

2. Sun, Y. and M.X. O'Riordan, Regulation of bacterial pathogenesis by intestinal short-chain Fatty acids. *Adv. Appl. Microbiol.* **85**, 93-118 (2013).

3. Mazmanian, S.K., et al., An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107-18 (2005).

4. Nash, M.J., D.N. Frank, and J.E. Friedman, Early Microbes Modify Immune System Development and Metabolic Homeostasis-The "Restaurant" Hypothesis Revisited. Front. Endocrinol. **8**, 349 (2017).

5. Hehemann, J.K., A.G.; Pudlo, P.A.; Martens, E.C.; Boraston, A.B., Bacteria of the human gut microbiome catabolize seaweed glycans with carbohydrate-active enzymes update from extrinsic microbes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 19786-19791 (2012).

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

7. Briliute, J., et al., Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat. Microbiol.* **4**, 1571-1581 (2019).

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

10. Temple, M.J., et al., A Bacteroidetes locus dedicated to fungal 1,6-beta-glucan degradation: Unique substrate conformation drives specificity of the key endo-1,6-beta-glucanase. *J. Biol. Chem.* **292**, 10639-10650 (2017).

11. Hooper, L.V., et al., 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**, 9833-9838 (1999).

12. Patel, E.H., et al., Rhamnose catabolism in Bacteroides thetaiotaomicron is controlled by the positive transcriptional regulator RhaR. *Res. Microbiol.* **159**, 678-84 (2008).

13. Schwalm, N.D., 3rd, G.E. Townsend, 2nd, and E.A. Groisman, Multiple Signals Govern Utilization of a Polysaccharide in the Gut Bacterium Bacteroides thetaiotaomicron. *MBio* **7**, e01342-16 (2016).

14. Thomas, F., et al., Environmental and gut bacteroidetes: the food connection. *Front. Microbiol.* **2**, 93 (2011).

15. Dong, L., et al., Microbial Similarity and Preference for Specific Sites in Healthy Oral Cavity and Esophagus. *Front. Microbiol.* **9**, 1603 (2018).

16. Anderson, K.L. and A.A. Salyers, Genetic evidence that outer membrane binding of starch is required for starch utilization by Bacteroides thetaiotaomicron. *J. Bacteriol.* **171**, 3199-204 (1989).

17. Foley, M.H., E.C. Martens, and N.M. Koropatkin, SusE facilitates starch uptake independent of starch binding in B. thetaiotaomicron. *Mol. Microbiol.* **108**, 551-566 (2018).

18. Tuson, H.H., et al., The Starch Utilization System Assembles around Stationary Starch-Binding Proteins. *Biophys. J.* **115**, 242-250 (2018).

19. Anderson, K.L. and A.A. Salyers, Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *J. Bacteriol.* **171**, 3192-8 (1989).

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

21. Bjursell, M.K., E.C. Martens, and J.I. Gordon, 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).

22. Martens, E.C., H.C. Chiang, and J.I. Gordon, Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447-57 (2008).

23. D'ella, J.N. and A.A. Salyers, Effect of Regulatory Protein Levels on Utilization of Starch by Bacteroides thetaiotaomicron. *J. Bacteriol.* **178**, 7180-7186 (1996).

24. Porter, N.T. and E.C. Martens, The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu. Rev. Microbiol.* **71**, 349-369 (2017).

25. Sonnenburg, E.D., et al., Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, 1241-52 (2010).

26. Joglekar, P., et al., Genetic Variation of the SusC/SusD Homologs from a Polysaccharide Utilization Locus Underlies Divergent Fructan Specificities and Functional Adaptation in Bacteroides thetaiotaomicron Strains. *mSphere* **3**, e00185-18 (2018).

27. Chaudet, M.M. and D.R. Rose, Suggested alternative starch utilization system from the human gut bacterium Bacteroides thetaiotaomicron. *Biochem. Cell. Biol.* **94**, 241-6 (2016).

28. Zgurskaya, H.I., C.A. Lopez, and S. Gnanakaran, Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It. *ACS Infect. Dis.* **1**, 512-522 (2015).

29. Ye, J. and B. van den Berg, Crystal structure of the bacterial nucleoside transporter Tsx. *EMBO J.* **23**, 3187-95 (2004).

30. Zafar, H. and M.H. Saier, Jr., Comparative genomics of transport proteins in seven Bacteroides species. *PLoS One* **13**, e0208151 (2018).

31. Townsend, G.E., 2nd, et al., Dietary sugar silences a colonization factor in a mammalian gut symbiont. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 233-238 (2019).

32. Khalili, H., et al., The role of diet in the aetiopathogenesis of inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 525-535 (2018).

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

34. Zou, J., et al., Fiber-Mediated Nourishment of Gut Microbiota Protects against Diet-Induced Obesity by Restoring IL-22-Mediated Colonic Health. *Cell Host Microbe* **23**, 41-53 (2018).

35. Bloom, S.M., et al., Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9**, 390-403 (2011).

36. 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**, 672-80 (2015).

37. Benjdia, A., et al., 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**, 25973-82 (2011).

38. Cartmell, A., et al., How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7037-7042 (2017).

39. Bryant, W.A., et al., In Silico Analysis of the Small Molecule Content of Outer Membrane Vesicles Produced by Bacteroides thetaiotaomicron Indicates an Extensive Metabolic Link between Microbe and Host. *Front. Microbiol.* **8**, 2440 (2017).

40. Chu, H., et al., Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science* **352**, 1116-20 (2016).

41. Cong, Y., et al., A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19256-61 (2009).

42. Comstock, L.E., Small RNAs Repress Expression of Polysaccharide Utilization Loci of Gut Bacteroides Species. *J. Bacteriol.* **198**, 2396-8 (2016).

43. Cao, Y., et al., cis-Encoded Small RNAs, a Conserved Mechanism for Repression of Polysaccharide Utilization in Bacteroides. *J. Bacteriol.* **198**, 2410-8 (2016).

44. Schwalm, N.D., 3rd and E.A. Groisman, Navigating the Gut Buffet: Control of Polysaccharide Utilization in Bacteroides spp. *Trends Microbiol.* **25**, 1005-1015 (2017).

45. Schwalm, N.D., 3rd, G.E. Townsend, 2nd, and E.A. Groisman, Prioritization of polysaccharide utilization and control of regulator activation in Bacteroides thetaiotaomicron. *Mol. Microbiol.* **104**, 32-45 (2017).

46. Cho, K.H., et al., New regulatory gene that contributes to control of Bacteroides thetaiotaomicron starch utilization genes. *J. Bacteriol.* **183**, 7198-205 (2001).

47. Lynch, J.B. and J.L. Sonnenburg, Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a Bacteroides hybrid two-component system. *Mol. Microbiol.* **85**, 478-91 (2012).

48. Tuncil, Y.E., et al., Reciprocal Prioritization to Dietary Glycans by Gut Bacteria in a Competitive Environment Promotes Stable Coexistence. *MBio* **8**, e01068-17 (2017).

49. Pudlo, N.A., et al., Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. *MBio* **6**, e01282-15 (2015).

50. Rogers, T.E., et al., Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol. Microbiol.* **88**, 876-90 (2013).

51. Ly, M., et al., Analysis of E. coli K5 capsular polysaccharide heparosan. *Anal. Bioanal. Chem.* **399**, 737-45 (2011).

52. Nzakizwanayo, J., et al., Disruption of Escherichia coli Nissle 1917 K5 capsule biosynthesis, through loss of distinct kfi genes, modulates interaction with intestinal epithelial cells and impact on cell health. *PLoS One* **10**, e0120430 (2015).