

Marshall University

Marshall Digital Scholar

Theses, Dissertations and Capstones

2022

A Systems Approach to Dissecting the Role of the Mucosal Microbiome in Disease

Lexie Christine Blalock

Follow this and additional works at: <https://mds.marshall.edu/etd>



Part of the [Medical Microbiology Commons](#)

**A SYSTEMS APPROACH TO DISSECTING THE ROLE OF THE MUCOSAL
MICROBIOME IN DISEASE**

A dissertation submitted to
the Graduate College of
Marshall University
In partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

In

Biomedical Sciences

by

Lexie Christine Blalock

Approved by

Dr. Hongwei Yu, Committee Chairperson

Dr. James Denvir

Dr. Joseph Horzempa

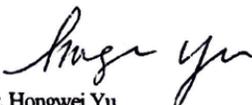
Dr. Jeremy McAleer

Dr. Timothy Long

Marshall University
December 2022

APPROVAL OF THESIS

We, the faculty supervising the work of Lexie Christine Blalock affirm that the dissertation, *A Systems Approach to Dissecting the Role of the Mucosal Microbiome in Disease*, meets the high academic standards for original scholarship and creative work established by the Biomedical Sciences Program and the Joan C. Edwards School of Medicine. This work also conforms to the editorial standards of our discipline and the Graduate College of Marshall University. With our signatures, we approve the manuscript for publication.

 Dr. Hongwei Yu Department of Biomedical Sciences	Committee Chairperson	7/28/2022 Date
 Dr. James Denvir Department of Biomedical Sciences	Committee Member	7/28/2022 Date
 Dr. Timothy Long Department of Pharmaceutical and Biomedical Science	Committee Member	8-1-2022 Date
 Dr. Jeremy McAleer Department of Pharmaceutical and Biomedical Science	Committee Member	7/28/22 Date
 Dr. Joseph Horzempa Department of Natural Science and Mathematics West Liberty University	Committee Member	8/1/2022 Date

© 2022
Lexie Christine Blalock
ALL RIGHTS RESERVED

ACKNOWLEDGMENTS

I am overwhelmed with gratitude for the endless support I received from mentors, colleagues, friends, and family while embarking on this journey. Thank you to my parents, Stephen and Deborah Keding, and husband, Joseph Blalock, for your endless love, support, and belief in me. A very heartfelt thanks to the rest of my family for always being my biggest cheerleaders and an incredible support system. I would also like to acknowledge my past and present four-legged companions, Zar, Abbie, Aussie, and Piper, for enriching my life so much with their fun-loving, kind, and adventurous spirits.

Jaime Friedman, Sarah Stevens, and Morghan Getty Collins: you will forever hold a special place in my heart. My fondest memories of graduate school will always involve the four of us along with our beloved dogs. I would also like to thank my dear friends Steven Sagun and Rebecca Guertal. I am so grateful to have shared so many milestones, including undergraduate and graduate school, and years of friendship with the two of you. I would also like to express my sincere appreciation and love for my cycling buddies who kept me sane over the years with our biking adventures and competitive banter.

A very special thanks to my advisor, Hongwei Yu, for his guidance, encouragement, and patience over the past several years. Your genuine passion and enthusiasm for scientific discovery are infectious, inspiring, and something I hope to impart to others just as you have to me. I would also like to extend a sincere thank you to my committee members, James Denvir, Jeremy McAleer, Joseph Horzempa, and Timothy Long. I have thoroughly enjoyed the opportunity to know and learn from you all as the outstanding scientists and humans you are. I would also like to thank Elaine Hardman for the kindness and hospitality she showed me upon

my arrival to Huntington and start of the Ph.D. program. I will always treasure our discussions and kayaking adventures at Beech Fork Lake. I would also like to thank Todd Green and Richard Eggleton for helping me navigate through the administrative processes of the Ph.D. program and for always having an open door.

Lastly, I would also like to thank my fellow lab members, Ryan Withers, Roy Al Ahmar, Brandon Kirby, and Meagan Valentine, with whom I have had the pleasure to work with while at Marshall University. You were a wonderful team to celebrate the small victories, troubleshoot problems, and laugh with each day in the laboratory, and I am so grateful for all your help and support. To all the other students, staff, and faculty that I engaged with at Marshall University, thank you for your kindness and for making Huntington feel like home. It truly takes a village to support and nurture a Ph.D. student—I appreciate and love each one of you!

TABLE OF CONTENTS

List of Tables	ix
List of Figures	x
ABSTRACT	xii
CHAPTER 1	1
INTRODUCTION TO THE INTESTINAL MICROBIOME AND ITS ROLE IN SHAPING THE MUCOSAL IMMUNITY	1
The gut microbiome	1
Diet-mediated modulation of the intestinal microbiota	2
Intestinal Epithelium-Gut Microbial Crosstalk.....	12
Segmented Filamentous Bacteria.....	16
CHAPTER 2	24
A BIOINFORMATICS APPROACH TO INVESTIGATE THE IMMUNOGENIC PROPERTIES OF SEGMENTED FILAMENOUS BACTERIA	24
Abstract	25
Introduction.....	26
Results.....	28
Conclusion	44
Materials and Methods.....	48
CHAPTER 3	57
INTERACTION OF THE GUT MICROBIOTA, DIET, AND HOST GENETICS IN TALLYHO/JNG MICE: A POLYGENIC MODEL FOR METABOLIC SYNDROME	57
Abstract	58

Introduction.....	59
Results.....	61
Discussion.....	81
Materials and Methods.....	88
Acknowledgements.....	96
CHAPTER 4	97
DEVELOPMENT OF A VECTOR SYSTEM FOR IN VIVO MANIPULATION OF THE GUT MICROBIOTA	97
Abstract.....	97
Introduction.....	98
Results.....	101
Conclusion	117
Materials and Methods.....	119
Acknowledgments.....	126
CHAPTER 5	127
CONCLUSIONS AND FUTURE DIRECTIONS.....	127
REFERENCES	131
Appendix A: Office of Research Integrity Approval Letter	162
Appendix B: Characteristics and peptide sequences of predicted SFB antigens	163
Appendix C: SFB proteins identified in situ in the murine ileal mucosa of c57bL/6J, DBa/2, and C3h mice	177
Appendix D: Hierarchal-all-against-all analysis clusters and associations	193
Appendix E: Identity and importance scores of all ASVs predictive of metabolic clusters	202

Appendix F: Deleterious SNPs and indels in genes essential to antigen processing and immune activation in the TALLYHO/Jng mouse genome 216

Appendix G: Mouse housing information 217

LIST OF TABLES

Table 1. Effect of categorical and continuous variables on alpha-diversity.	63
Table 2. Predictive genera of metS and non-metS clusters	77
Table 3. Summary of diet composition.....	89
Table 4. Summary of metabolic traits.....	90
Table 5. Bacterial plasmids used in this study	119
Table 6. Primers used in this study	121

LIST OF FIGURES

Figure 1. Distribution of SFB in the intestines of male and female mice purchased from Taconic Biosciences and the Jackson Laboratory	30
Figure 2. Diversity and compositional analyses of the Peyer’s Patch-associated microbiota	32
Figure 3. In silico prediction workflow for the identification of SFB antigens.....	34
Figure 4. Functional analysis and network of the SFB candidate antigens identified.....	35
Figure 5. Networks of mouse and rat SFB proteins identified in the murine ileal mucosa via shotgun proteomics	38
Figure 6. Functional characterization of modules identified from mouse SFB proteins	39
Figure 7. Functional characterization of modules identified from rat SFB proteins	40
Figure 8. Operon and protein interaction analyses indicate environmental nutrients are a determinant of flagellar cap protein FliD transcription.	43
Figure 9. Cecal microbiota composition.....	62
Figure 10. Analysis of community diversity and explanatory variables	66
Figure 11. Linear discriminate analysis effect size (LEfSe) analysis of diet-driven microbiota shifts in B6 mice	69
Figure 12. Linear discriminate analysis effect size (LEfSe) analysis of diet-driven microbiota shifts in TH mice.....	71
Figure 13. Linear discriminate analysis effect size (LEfSe) analysis of genetic-driven microbiota shifts in B6 and TH mice	72
Figure 14. Associations between metabolic variables and taxa clusters	74
Figure 15. Microbial features predictive of metabolic phenotypes	76

Figure 16. Dietary factors play a role in the microbial patterns that differentiate healthy from diseased hosts.....	78
Figure 17. Accuracy of Random Forest classifier predictions of non-metS and non-metS features	79
Figure 18. Analysis of community function and explanatory variables	80
Figure 19. Fluorescent and bioluminescent expression by the biome vectors.....	100
Figure 20. Maps of pHERD30T and biome vectors	104
Figure 21. <i>In vitro</i> assay experimental design	105
Figure 22. Aerobic and anaerobic <i>in vitro</i> quantitation of titrated L-arabinose-induced fluorescence and bioluminescence expression.....	107
Figure 23. Glucose supplementation is not necessary to achieve P _{BAD} promoter repression.....	108
Figure 24. Quantitation of titrated L-arabinose-induced fluorescence and bioluminescence protein expression under aerobic and anaerobic conditions	112
Figure 25. Characterization of the biome vectors in an <i>ex vivo</i> gut microbiota model.	113
Figure 26. <i>In vivo</i> colonization of <i>E. coli</i> harboring constitutively expressed bioluminescence vector (pUC18T-mini-Tn7T- <i>lux</i>) and the inducible bioluminescence biome vector.	116
Figure 27. Survival curve of mice post-oral gavage with bacterial vectors or PBS	117

ABSTRACT

The microbial community harbored by the mammalian gastrointestinal tract, collectively called the gut microbiota, plays a critical role in host health, metabolism, and immunity. Quantifiable changes in the gut microbiota have been implicated in a myriad of pathologies, including autoimmune, metabolic, and neurodegenerative diseases. Because of the gut microbiota's plasticity, manipulating this community for therapeutic benefit presents as a novel strategy to combat modern diseases. Yet, only a fraction of the total microbial diversity that exists within the gut has been successfully cultured in a laboratory, leaving the clinical implications and functional attributes of many gut microbes undetermined. To address this issue, studies in this dissertation focused on bioinformatically investigating (i) the role of dietary and genetic factors on the gut microbiota in metabolic syndrome and (ii) the anatomic niche and immunomodulating antigens of the culture-resistant commensal, Segmented Filamentous Bacteria (SFB).

Comprehensive characterization of the cecal microbial community in normal and metabolic syndrome-susceptible mice revealed that dietary factors overshadow the effects of host genetics and are determinist of the microbial patterns that emerge in metabolically healthy or diseased hosts. Moreover, we identified a microbial community proximal to the Peyer's Patches that was phylogenetically distinct from the cecum. Through a multi-omics approach, we show that independent of mouse genetics, vendor, age, or gender, SFB are highly abundant in the Peyer's Patches mucosa. Our *in silico* antigen prediction analyses identified thirty-five potential antigens within the reference proteome of SFB. Shotgun proteomics of the murine ileal mucosa confirmed the *in situ* presence of several potential antigens expressed by SFB, of which included multiple flagellar proteins. To experimentally valid bioinformatically-inferred antigen targets, we developed a collection of broad-host bacterial vectors to enable microbial products of interest to

be precisely expressed *in vitro*, *ex vivo*, and *in vivo*. Collectively, our studies reveal important findings about the mechanisms by which (i) exogenous factors modulate the gut microbiota and (ii) an endogenous microbe modulates the host immune system and provides a novel molecular platform by which microbial products of interest can be causally investigated in polymicrobial communities.

CHAPTER 1

INTRODUCTION TO THE INTESTINAL MICROBIOME AND ITS ROLE IN SHAPING THE MUCOSAL IMMUNITY

The gut microbiome

The human gastrointestinal tract harbors one of the densest microbial communities in existence, collectively called the gut microbiota. Outnumbering human cells by tenfold, the gut microbiota contains approximately 10^{14} bacteria, many of which are critical for their host's well-being and proper physiological function (Ley, Peterson, & Gordon, 2006). The commensal flora provides the host with multiple beneficial functions, including the metabolism of dietary components, defense from pathobionts, and maturation of the intestinal epithelia and associated lymphoid tissue (Ley, Peterson, et al., 2006). The compositional diversity of the mammalian gut microbiota has been characterized by two dominating phyla, Firmicutes and Bacteroidetes, as well as Actinobacteria, Proteobacteria, and Verrucomicrobia phyla to a lesser extent (Backhed, Ley, Sonnenburg, Peterson, & Gordon, 2005).

Establishment of the endogenous flora begins at birth, and many environmental factors such as the method of birth, exposure to antibiotics, dietary regimens, hygiene, socioeconomic status, geographical location, and cultural traditions, can significantly affect the composition and diversity of the gut microbiota (Chan et al., 2016; Human Microbiome Project, 2012).

Interindividual variation in microbial diversity can be further modified by inherited genetic traits; however, multiple cross-sectional studies have concluded that the average heritability of the microbiome is relatively low compared to the variance of the microbiome explained by diet, geographical, and cultural factors (Rothschild et al., 2018; Yatsunencko et al., 2012).

Human genome-wide association studies (GWAS) have correlated over 83 human quantitative trait loci (QTLs) with specific gut microbial phylotypes (Benson, 2015; Blekhman et al., 2015). Many of the associated QTLs overlap with immune-related genes that are often implicated in complex, polygenic diseases such as metabolic syndrome and inflammatory bowel disease (IBD) (Benson, 2015; Blekhman et al., 2015). While there is clear evidence that the host genotype can influence the microbial landscape, we have yet to know the resolution at which genetic traits and their interactions with environmental factors, such as diet, contribute to the host phenotypes and microbial phylotypes related to complex diseases. Animal studies have yielded differing perspectives on this matter, with some investigators associating gut microbial characteristics with the host genotype (Hildebrand et al., 2013), while others have observed changes of the gut microbiota to be diet-dependent, irrespective of the host's genetic background (Carmody et al., 2015).

Diet-mediated modulation of the intestinal microbiota

Through their mutual coevolution, the gut microbiota has developed specialized metabolic processes that provide the host with essential nutrients, comprising an estimated ten percent of the host's total energy intake (Bouter, van Raalte, Groen, & Nieuwdorp, 2017; Holmes, Li, Marchesi, & Nicholson, 2012). Mammalians have become dependent on microbial fermentation of otherwise indigestible plant polysaccharides. The degradation of these complex polysaccharides by intestinal bacteria subsequently renders short-chain fatty acids (SCFAs), such as acetate, propionate, and butyrate, which are substrates in glucose and lipid metabolic pathways within the intestines and peripheral organs (Rios-Covian, Salazar, Gueimonde, & de Los Reyes-Gavilan, 2017; Velagapudi et al., 2010). Considering the unique interplay between host nutrient extraction, metabolism, and the intestinal microbiota, it comes as little surprise that

diet can have significant impacts on the gut microbiota's composition and function. Furthermore, diet-induced microbiota perturbations can affect immune and inflammatory responses, which represent a potential underlying factor of metabolic and inflammatory diseases (Maslowski & Mackay, 2011).

In the absence of endogenous flora, germ-free (GF) mice are resistant to diet-induced obesity and associated metabolic complications (Backhed, Manchester, Semenkovich, & Gordon, 2007; Rabot et al., 2010). Alternatively, the gut microbiota of conventionally colonized mice exhibits extreme plasticity in response to dietary perturbations (Carmody et al., 2015), resulting in altered microbial and host metabolic functions. Consumption of the "Western diet," which is defined by a regimen of high-fat (32.8%) and high-simple carbohydrate (51.8%), and low-protein (15.4%) (Cordain, Eaton, Sebastian, Mann, Lindeberg, Watkins, O'Keefe, et al., 2005), has been shown to result in an augmented abundance of the Firmicutes to Bacteroidetes phyla (F:B ratio) in human and animal models (Carmody et al., 2015; Org et al., 2015), as well as humanized gnotobiotic mice when compared to their low-fat, plant polysaccharide-rich fed-counterparts (Turnbaugh, Ridaura, et al., 2009).

Additionally, studies profiling the compositional differences between lean and obese human populations, including a cohort of monozygotic and dizygotic twins (Turnbaugh, Hamady, et al., 2009), have observed the F:B ratio increasing in parallel with body weight and insulin resistance (Koliada et al., 2017; Ley et al., 2005). Furthermore, reports have shown that transferring the microbiota from obese donors to lean recipients is sufficient to induce an obese phenotype in murine models (Ley et al., 2005; Turnbaugh et al., 2006). Yet, Rabot et al. found that the development of obesity in recipient conventionally-raised and GF mice was independent of the gut microbiome transferred from lean or obese donors; instead, it was determined by a

high-fat diet (Rabot et al., 2016). However, a significant correlation was found between enhanced glucose tolerance and mice that had received a microbiota transplant with a low F:B content, which suggests that the microbiota indirectly affects the host's metabolic health at a very minimum (Rabot et al., 2016). In the subsequent sections of this review, keystone studies and emerging literature on the interplay between dietary macronutrients, host function, and the gut microbiota will be discussed.

Carbohydrates

Investigation of gut microbial diversity over the geographical landscape has revealed culture-specific microbiota compositions, thereby enabling the microbiome to be studied as a function of diet and lifestyle. De Filippo and colleagues conducted a comparative microbiome study between Western diet-consuming European (EU) children and rural African children from the village of Burkina Faso (BF), whose diets are low-fat, and rich in plant-derived polysaccharides, fiber, and protein (De Filippo et al., 2010). The microbiomes of BF children were significantly enriched with Actinobacteria and Bacteroidetes phyla. In contrast, the average EU child's microbiome was dominated by the Firmicutes phyla and harbored an increased abundance of Enterobacteriaceae taxa (De Filippo et al., 2010). Interestingly, augmented abundances of Firmicutes and Enterobacteriaceae taxa are characteristic of the dysbiotic microbiota often observed in IBD patients (Morgan et al., 2012).

Accompanying the diverging microbiota patterns between the EU and BF groups were functional changes, such as significantly attenuated SCFA production in EU children and the exclusive presence of the bacterial genera *Prevotella* and *Xylanibacter* in BF children. These organisms' genomes contain enriched gene sets specific to polysaccharide hydrolysis (De Filippo et al., 2010). Consistent with other studies, De Filippo and colleagues showed that increased

dietary fiber promotes the enrichment of carbohydrate-fermenting and SCFA-producing bacteria, which have been correlated with anti-inflammatory immune responses and healthier metabolic phenotypes (M. S. Kim, Hwang, Park, & Bae, 2013; Martinez et al., 2013).

Digestible carbohydrates, such as starches and sugars, are characterized by a lack of fiber, thereby enabling these foods to be rapidly metabolized by the host. Furthermore, digestible carbohydrates are also high on the glycemic load—a metric representing the blood glucose-raising potential of a food based on the carbohydrate content of a 100-gram portion (Cordain, Eaton, Sebastian, Mann, Lindeberg, Watkins, O’Keefe, et al., 2005). Accordingly, repeated consumption of foods with high glycemic loads leads to acute spikes in blood glucose, which promotes the development of insulin resistance and other metabolic sequelae (Cordain, Eaton, Sebastian, Mann, Lindeberg, Watkins, O’Keefe, et al., 2005). Considering the prevalence of sugar consumption, particularly among Westernized cultures, investigators have begun to study the independent effect of dietary sugar consumption on the gut microbiome, obesity, and even neurological development and function.

Seminal findings by Magnusson et al. revealed that exposing B6 mice to high-sucrose diets significantly impaired short- and long-term cognitive function compared to high-fat and regular chow diets (Magnusson et al., 2015). These behavioral changes were observed within 5 weeks of the prescribed dietary regimens, and were significantly correlated with a gut microbiota enriched in *Lactobacillus* and the Firmicutes order, Clostridiales, with a marked reduction in the Porphyromonadaceae family of the Bacteroidetes phylum (Magnusson et al., 2015). However, because this study was not conducted or repeated in germ-free mice, it’s not clear the degree at which the gut microbiome per se contributed to this observation (Magnusson et al., 2015).

Although sucrose is typically the carbohydrate component of the Western and high-fat diets (HFD) used in animal studies, non-caloric artificial sweeteners (NAS), such as saccharin and aspartame, are becoming popular substitutes for natural sweeteners. Moreover, NAS are added to diet sodas and sugar-free foods and often recommended for weight loss and individuals with T2D, thus understanding the impact of NAS consumption on the gut microbial community and host health is critical, considering many NAS consumers are already afflicted with an obesity-related disorder (Gardner et al., 2012).

Suez et al. determined that supplementation of the NAS, saccharin, sucralose, and aspartame, in water adjusted to the FDA's acceptable daily intake (ADI) (5 mg kg⁻¹ of body weight) exacerbated glucose intolerance in normal B6 mice compared to mice provided glucose and sucrose supplementation or water only (Suez et al., 2014). After 11 weeks of NAS supplementation, saccharin led to the highest blood glucose response curve of all NAS tested (Suez et al., 2014). In a follow-up study, HFD-fed B6 and Swiss Webster mice supplemented with saccharin developed significant glucose intolerance as compared to the glucose supplemented controls, indicating that the glycemic effect of saccharin is independent of other dietary factors (Suez et al., 2014). These metabolic perturbations were proven to be microbiota-mediated as the authors showed (i) four weeks of antibiotic intervention rescued the NAS-induced glucose intolerance in saccharin supplemented mice, and (ii) the saccharin-induced glucose intolerance phenotype could be recapitulated in GF mice through microbiota transfers (Suez et al., 2014).

Work by Thaïss et al. suggests that sugar products advance the pathophysiology of intestinal dysbiosis and metabolic perturbations, such as glucose intolerance, through a unique mechanism separate from those that potentiate obesity (Thaïss et al., 2018). The authors found

that obesity, while often associated with, is not obligatory nor exacerbates intestinal dysbiosis in murine models (Thaiss et al., 2018). Instead, glucose-induced hyperglycemia leads to a dysfunctional intestinal barrier that was accompanied by a systemic spread of intestinal bacteria, causing an infection in streptozotocin-induced type 1 diabetes mellitus mice (Thaiss et al., 2018).

Although this model of hyperglycemia modulated the gut microbiota's composition, the investigators determined the phenotype to be gut microbiota-independent through a series of microbiota transfers (Thaiss et al., 2018). Furthermore, the investigators determined that glucose-induced hyperglycemia affected the function and integrity of the intestinal epithelial cells through the augmented activity of glucose transport between the intestinal epithelium and systemic circulation mediated by the bidirectional glucose transporter, GLUT2 (Thaiss et al., 2018). These changes led to a significant impairment in the tight and adherence junctions of the intestinal epithelial cells, thereby allowing an influx of immune-stimulating microorganisms across the intestinal barrier and into the systemic circulation (Thaiss et al., 2018).

As the understanding of microbiota's role in health and disease becomes more clear, alternative approaches to enhance gut health have emerged. These treatments put emphasis on modulating the gut microbiota through diet, such as the intentional supplementation of fiber and complex carbohydrates, *e.g.*, prebiotics, to improve host health. Prebiotics are composed of non-digestible carbohydrates such as fructosyl-oligosaccharide and galactosyl-oligosaccharide, which promote the growth of favorable gut flora (Boulangue, Neves, Chilloux, Nicholson, & Dumas, 2016). Many animal studies have noted increased glucose tolerance, improved gut barrier function, and the expansion of SCFA-producing bacteria mediated by prebiotics (Boulangue et al., 2016; Liu et al., 2017; Moreno-Indias, Cardona, Tinahones, & Queipo-Ortuno, 2014).

Work by Delzenne et al. demonstrated that prebiotic supplementation in rats could limit the inflammatory effects of metabolic endotoxemia by promoting the function of tight junction proteins and the endogenous production of enteroendocrine peptides (Delzenne, Cani, Daubioul, & Neyrinck, 2005). While the use of prebiotics is growing, research has shown significant inter-individual variation in the response of the gut microbiome to dietary interventions (Smits, Marcobal, Higginbottom, Sonnenburg, & Kashyap, 2016), thereby indicating the need for further understanding and stratifying microbial populations in the context of precision medicine.

Proteins

As evidenced by population-based studies, dietary practices unique to geographical regions or associated cultures lead to distinct microbiota compositions; however, subtle changes of dietary components, such as differing origins of protein and fat, have profound effects on the microbiota and host health (Liisberg, Myrmel, et al., 2016; Yatsunenko et al., 2012). Zhu et al. performed a comprehensive study on the responses of the gut microbiota to red meat (beef and pork), white meat (chicken and fish), and other sources of protein (casein and soy) in Sprague-Dawley rats (Y. Zhu et al., 2015). Protein sources originating from white meats lead to a significant enrichment of *Lactobacillus*, whereas the intake of non-meat proteins increased the abundance of Bacteroides and Prevotella (Y. Zhu et al., 2015). Interestingly, rats maintained on red-meats had the lowest serum level of lipopolysaccharide (LPS) binding protein, an indicator of systemic LPS levels, despite their microbiota's harboring an augmented abundance of the Gram-negative phylum, Proteobacteria (Y. Zhu et al., 2015). However, disagreement still remains on the protective or potentiating effects of different protein sources on body weight and adiposity, and whether the gut microbiota regulates these physiological alterations (Liisberg, Fauske, et al., 2016; Y. Zhu et al., 2015). Overall, evidence from epidemiological studies

demonstrates that individuals who acquire the majority of their protein from animals are at greater risk for obesity, cancer, and premature death compared to those who consume mostly dairy and vegetable-based proteins (Fogelholm, Anderssen, Gunnarsdottir, & Lahti-Koski, 2012; Levine et al., 2014; J. D. Smith et al., 2015).

Increased red meat consumption significantly increases the risk of cardiovascular disease due to an increase in the pro-atherosclerotic microbial metabolite, trimethylamine-N-oxide (TMAO) (Jie et al., 2017). Present at high concentrations in red meats, carnitine, and choline are metabolized to trimethylamine (TMA) by the gut microbiota, which is subsequently transported to the liver (Janeiro, Ramirez, Milagro, Martinez, & Solas, 2018). While in the liver, TMA is rapidly oxidized by hepatic flavin monooxygenases to form the biologically active TMAO metabolite, which at high levels can mediate renin-angiotensin system damage, hypertension, and atherosclerosis (De Filippis et al., 2016; Jie et al., 2017; Zeisel & Warrier, 2017). A cohort study conducted by Jie et al. consisting of 218 individuals with atherosclerotic cardiovascular disease (ACVD) and 187 health controls found that the co-occurrence of *Streptococcus* and *Enterobacteriaceae* in the gut microbiota of ACVD patients accompanied by the depletion of *Bacteroides* and *Prevotella* taxa distinguished diseased from healthy patients. Furthermore, functional analyses of the ACVD microbiome suggested that carbohydrate metabolism and SCFA synthesis were downregulated, while pathways associated with virulence, including O-antigen synthesis of bacterial LPS and TMA-synthesizing enzymes, were significantly enriched compared to healthy controls (Jie et al., 2017).

Lipids

Dietary lipids have also been found to regulate the gut microbiome composition and the host's metabolic health. Diets enriched with saturated or *trans* fats have been shown to promote

obesity and inflammation through the upregulation of cholesterol, insulin resistance, and accumulation of white adipose tissue (WAT) (Buettner, Scholmerich, & Bollheimer, 2007). Recent work has highlighted the differentiation abilities of adipocytes, where dietary factors, such as the composition and amount of consumed lipids, and their interaction with the gut microbiota, can mediate the whitening or browning (BAT) of adipose tissue (Caesar, Tremaroli, Kovatcheva-Datchary, Cani, & Backhed, 2015; Suarez-Zamorano et al., 2015). Neonates are born with a large percentage of the metabolically active BAT, which functions as a heat source to maintain body temperature (Saely, Geiger, & Drexel, 2012). Although BAT typically decreases with age, brown adipocytes have been increasingly implicated in lean and healthy phenotypes (Saely et al., 2012). On the other hand, white adipocytes preferential function to store excess energy in the form of triglycerides. Moreover, adipocytes within WAT can become engorged with lipids, leading to hypertrophy and inflammation-inducing apoptosis (Kennedy, Martinez, Chuang, LaPoint, & McIntosh, 2009).

Caesar and colleagues demonstrated that regardless of isocaloric diets, microbiota and metabolic perturbations could be induced solely by feeding mice different lipid compositions (Caesar et al., 2015). B6 mice fed a diet enriched in saturated fat (lard) had significantly augmented WAT accumulation and associated inflammation. This was accompanied by increased intestinal macrophages and CD45⁺ cells, an indicator of T cell differentiation, when compared to mice that were maintained on a polyunsaturated fat diet (fish oil) (Caesar et al., 2015). Furthermore, the microbiota and toll-like receptor (TLR) signaling was necessary for the activation of WAT, as mice lacking TLR immune signaling pathways were protected from lard-induced adiposity, inflammation, and insulin resistance (Caesar et al., 2015). Additionally, the metabolic and inflammatory effects of the lard diet in WT mice could be minimized by the

transplantation of microbiota originating from fish oil-fed mice (Caesar et al., 2015). In contrast, microbiota transplanted from lard-fed mice further aggravated the obese and inflammatory phenotype in the WT recipient mice.

The work conducted by Caesar and colleagues did not find WAT activation to be entirely microbiota-dependent, as lard-fed GF mice still exhibited mild metabolic and inflammatory responses (Caesar et al., 2015). However, a study by Suárez-Zamorano et al. showed that “being” of adipocytes, that is, the transformation from WAT to BAT, was promoted by the depletion of the gut microbiota in B6, leptin-deficient (*ob/ob*), and HFD-fed mice (Suarez-Zamorano et al., 2015). While this is consistent with other studies that have shown microbiota-depleted mice are resistant to obesity and its associated metabolic consequences (Backhed et al., 2004; Chou, Membrez, & Blancher, 2008; I. Hwang et al., 2015), caution must be used in the interpretation of results originating from unconventionally colonized or GF mice, which have underdeveloped and naïve immune systems (Hooper, Littman, & Macpherson, 2012).

These investigations of the mammalian gut microbiota have begun to demonstrate a profound, albeit extraordinarily complex relationship between the intestinal flora and its host. The remarkable increase in obesity and inflammatory-related disorders in developing countries has been associated with the Western diet and the consequential changes to the gut microbiome, which is substantially different from cultures that maintain plant and fiber-rich diets similar to that of our predecessors (Cordain, Eaton, Sebastian, Mann, Lindeberg, Watkins, O’Keefe, et al., 2005). Strong evidence supports the position that dietary patterns shape the microbiota structure and function, and together play an essential role in host immunity and metabolic health. Although causal mechanisms remain to be elucidated, large-scale association studies have yielded valuable information regarding the interplay between diet, microbial phylotypes, and the

resultant host phenotype. With advancements in multi-omic technologies, e.g., transcriptomics, proteomics, metabolomics, as well as bioinformatics tools, we have entered the era of resolving these complex microbial-immune interactions to a mechanistic level. Given the obvious therapeutic potential that such findings provide, new interventions for long-standing diseases may be on the horizon as our understanding of the human-microbe symbiosis continues to increase.

Intestinal Epithelium-Gut Microbial Crosstalk

Although genetics, as described below, and diet contribute to microbiota diversity and composition, the downstream interactions that occur between the gut microbiota and host intestinal epithelia play a profound role in mammalian health and immunity. At the center of these complex host-microbial interactions is the intestinal barrier, which functions to separate the host immune system from foreign objects within the intestinal lumen, while also promoting tolerance of commensal flora (Turner, 2009). While the etiology of gut-associated inflammatory disorders is multifactorial, dysregulation of the intestinal barrier has been associated with chronic inflammatory diseases such as IBD and metabolic syndrome, which together affect nearly forty percent of the adult American population (Aguilar, Bhuket, Torres, Liu, & Wong, 2015; Dahlhamer, Zammitti, Ward, Wheaton, & Croft, 2016; B. Wang, Yao, Lv, Ling, & Li, 2017).

The intestinal barrier consists of a single layer of epithelial cells lining the intestinal lumen (Turner, 2009). This cellular monolayer creates a physical barrier between the gut microbiota and the immunocytes responsible for orchestrating localized and systemic host immune responses (Turner, 2009). Additionally, the intestinal epithelium is a critical component of the gut-associated lymphoid tissue (GALT), which represents the largest immune organ in the mammalian body (Gao et al., 2012; Turner, 2009). The GALT encompasses approximately

seventy percent of all immune-responding cells, such as immunoglobulin-A (IgA) secreting plasma cells, naïve CD4⁺ and differentiated T cells, and memory B cells (Jung, Hugot, & Barreau, 2010). Thus, a cooperative relationship between the GALT and gut microbiota promotes the tolerance of commensal flora while also affording constant surveillance and selective protection against luminal antigens, such as pathogens or potential food allergens (Hashiguchi et al., 2015; Jung et al., 2010).

To preserve the host-microbe symbiosis, the intestinal immune system has developed tolerogenic and regulatory mechanisms, including soluble mediators that limit the expansion of pathogens and their invasion of the intestinal barrier, which could otherwise lead to localized or systemic infections (Hooper & Macpherson, 2010). Mucin-type glycoproteins produced by the specialized epithelial cells, Goblet cells, aid in the assembly of the protective mucus layer on the apical side of the epithelium, which serves as a spatial barrier between the microbial inhabitants of the lumen and the host's mucosal surfaces (Hooper & Macpherson, 2010).

Intestinal epithelial cells also secrete a variety of antimicrobial peptides (AMPs), with the goal of these secretory molecules being to kill or inactivate microorganisms (Gallo & Hooper, 2012). While eliminating the pathological threat is the main objective of AMPs, the mechanism by which this is accomplished, the breadth of organisms impacted, and the cellular site of expression is specific to each AMP family (Gallo & Hooper, 2012). For example, α - and β -defensins are expressed in the small and large intestines, respectively, and disrupt the membrane of a broad spectrum of bacteria, fungi, viruses, and protozoa, while C-type lectins and phospholipase A2 AMP families have specific bactericidal activity against Gram-positive bacteria (Gallo & Hooper, 2012). The critical importance of and extent to which AMPs can regulate microbiota composition was examined by Salzman et al (Salzman et al., 2010). The

authors showed that mice with an α -defensin deficiency harbored a compositionally distinct gut microbiota compared to wild-type mice (Salzman et al., 2010). These results indicate that AMPs play an instrumental role in regulating the microbial population of the intestines and lends credence to hypothesis that altered AMP production may contribute to IBD—a condition characterized by dysbiosis and bacterial encroachment (Gallo & Hooper, 2012; Salzman et al., 2010).

A delicate balance must be struck between the pro-inflammatory responses necessary to eliminate obtrusive pathogens and anti-inflammatory mechanisms that promote intestinal homeostasis. Indeed, this is dependent on the colonization of commensals, with many studies demonstrating that the intestinal microbiota is required for the proper development of innate immune responses to pathogen-associated molecular patterns (Rakoff-Nahoum, Paglino, Eslami-Varzaneh, Edberg, & Medzhitov, 2004). Constant surveillance of the intestinal lumen's contents occurs through antigen presentation by dendritic cells (DCs), which enables the host immune system to distinguish symbionts versus pathobionts and generate the appropriate immune response in real-time (Hooper & Macpherson, 2010). Antigen-laden DCs migrate to the inductive sites of the GALT, which include the mesenteric lymph nodes (MLNs) and Peyer's Patches (PeyP), where T and B cell responses are initiated (Jung et al., 2010).

Under homeostatic conditions, Forkhead box P3 (Foxp3)⁺ CD4⁺ regulatory T (Treg) cell populations proliferate in response to colonizing commensals, thereby promoting tolerance of the resident microbiota and suppressing inflammatory immune responses (Harrison & Powrie, 2013). Through Foxp3⁺CD4⁺ Treg cell-dependent and independent mechanisms, B cells differentiate into IgA-producing plasma cells (Hapfelmeier et al., 2010; Hooper & Macpherson, 2010). Polymeric IgA is subsequently transcytosed across the intestinal epithelium where it binds

to microorganisms and toxins thereby blocking their attachment to and interaction with the intestinal epithelia (Hapfelmeier et al., 2010; Hooper & Macpherson, 2010). Perhaps the most remarkable aspect of IgA neutralization is the fact this complex process occurs without the induction of a pro-inflammatory response (Corthesy, 2013).

Work by Kawamoto et al revealed a positive feedback loop in which PeyP-derived Foxp3⁺CD4⁺ Treg cells regulate the response of IgA to commensals (Kawamoto et al., 2014). Maintenance of a diverse, homeostatic gut microbiome was found to be dependent on the presence of Foxp3⁺CD4⁺ Treg cells; however, the expansion of these cells is indeed driven by gut microbiota-derived cues (Kawamoto et al., 2014). In the absence of Foxp3⁺CD4⁺ Treg cells, mice exhibited significantly impaired IgA responses and diminished gut microbial diversity that was accompanied by a pro-inflammatory phenotype (Kawamoto et al., 2014). Thus, the diversification and expansion of Foxp3⁺CD4⁺ Treg cells are essential to IgA production as well as the establishment and preservation of a balanced gut microbiome (Kawamoto et al., 2014). Interestingly, mucosal IgA lacks classic immune memory characteristics, indicating the Foxp3⁺CD4⁺ Treg-IgA axis is extremely dynamic, allowing constant immune adaptations in response to microbiota influx and variation over time (Hapfelmeier et al., 2010). Therefore, the exact mechanism of how the gut microbiota induces the isotype switching from IgG- to IgA-producing cells and undergoes clonal expansion is not fully understood.

While physical, chemical, and immune-mediated mechanisms are in place to protect the host's intestinal barrier, a bacterial breach can still occur. Upon pathogenic invasion of the intestinal epithelium, resident macrophages and damaged epithelial cells recruit neutrophils and T cells to the localized tissue through the secretion of pro-inflammatory cytokines and chemokines. These signals promote adaptive immune responses and activation of effector CD4⁺

cells, e.g., T helper cells 1 (Th1), 2 (Th2), or 17 (Th17) (Hooper & Macpherson, 2010). Th1 cells proliferate in response to threats by intracellular bacterial and viral infections, whereas Th2 cells are typically responsive to helminths and extracellular parasitic infections (Maynard & Weaver, 2009). The differentiation of Th17 cells is usually indicative of an extracellular bacterial or fungal infection; however, the cell's population within the intestines is largely driven by the gut microbiota. In particular, mucosal-adherent bacteria, such as *Candidatus arthromitus*, also referred to as Segmented Filamentous Bacteria, are known drivers of the Th17 cell population as described below (Hooper & Macpherson, 2010; Maynard & Weaver, 2009).

Segmented Filamentous Bacteria

Segmented filamentous bacteria (SFB) are gram-positive, spore-forming gut commensals belonging to the Clostridiaceae family (Kuwahara et al., 2011). This bacterium was first noticed nearly six decades ago through microscopy for its unusual, segmented morphology and “holdfast” appendage, which enables the bacterium to tightly adhere to the intestinal epithelial cells (Davis, McAllister, & Savage, 1973; Reimann, 1965; Savage, 1969). SFB exhibit a unique affinity for colonizing the localized regions nearest the aggregated lymphoid follicles of the ileum collectively called Peyer’s Patches (PeyP) (Klaasen, Koopman, Poelma, & Beynen, 1992). Serving as the inductive site of the intestines for localized and systemic innate and adaptive immune responses, PeyP are a critical component of the GALT (Jung et al., 2010). In animal models, colonization of SFB has been shown to be essential for postnatal immune maturation, immunoglobulin A (IgA) production, and regulatory T cell responses (Artis & Grencis, 2008; Macpherson & Uhr, 2004), yet multiple reports implicate SFB in the development of chronic inflammatory diseases, such as inflammatory bowel disease (IBD), autoimmune arthritis, and obesity-related inflammatory sequelae, e.g., nonalcoholic fatty liver disease (NAFLD) (Chappert,

2014; Harley et al., 2014; Kumar et al., 2016). While the spatial relationship that exists between SFB and the PeyP has been well established, recent investigations have focused on defining how SFB affects the host immune system under homeostatic and diseased states.

Microbiological profile of Segmented Filamentous Bacteria

The modern understanding of SFB comes primarily from observations made through electron microscopy imaging as well as molecular methods of detection including 16S ribosomal RNA (rRNA) or whole-genome sequencing (Ericsson et al., 2015; Schnupf et al., 2015). Historically, SFB has been notably difficult to culture *in vitro*, but there were two successful reports published in 2015 by Schnupf et al and Ericsson et al (Ericsson et al., 2015; Schnupf et al., 2015). As a member of the indigenous flora, SFB has been found to inhabit the gastrointestinal tract of multiple vertebrates, including fish, pigs, primates, birds, rodents, cats, and dogs (Klaasen, Koopman, et al., 1993; Yin et al., 2013). Many comparative studies have highlighted the taxonomic divergence of SFB within and between its broad hosts (Pamp, Harrington, Quake, Relman, & Blainey, 2012; Prakash et al., 2011; Yin et al., 2013), suggesting there could be multiple strains of SFB, or the microbe has evolved into host-specific strains for selective advantages.

Phylogenetic analyses align SFB within the Clostridium subphylum, with the provisional name of *Candidatus arthromitus* (Kuwahara et al., 2011; Pamp et al., 2012). However, the bacterium may ultimately be classified as *Candidatus savagella*, in honor of intestinal microbiologist, D.C. Savage, who was first to report the significance of SFB in the gastrointestinal tract (Thompson, Vier, Mikaelyan, Wienemann, & Brune, 2012). At 1.6 MB, the SFB genome is one of the smallest of the Clostridiales class (Kuwahara et al., 2011). While the minimal genome is deficient in the genes required for nutrient and energy metabolism, rendering

the bacterium auxotrophic, the genes dedicated to sporulation and flagellar-mediated motility are intact (Kuwahara et al., 2011; Prakash et al., 2011). Because SFB lack metabolic and aerobic respiration capabilities, cultivating this fastidious, anaerobic microbe remains a great challenge. The partially or completely assembled SFB genomes publicly available originated from murine and turkey hosts, but noticeably none from humans. This may suggest that SFB represents a very small proportion of the human gut microbiota compared to other vertebrate hosts.

T helper 17 cell signaling

The commensal gut flora is a critical regulator of the host's T cell population, albeit through many uncharacterized mechanisms. Conversely, the relationship between SFB, specifically, and its potent induction of Th17 cells has been well defined, and therefore become a microbe of interest in the field of immunology. The role of Th17 signaling in host defense and disease is yet to be completely understood; however, data have shown that these unique cells and their secreted cytokines, such as interleukin-17A (IL-17A) and 17F (IL-17F), interleukin-22 (IL-22), tumor necrosis factor-alpha, and interferon-gamma, are important to the host defenses, particularly in responding to extracellular pathogens, through the localized recruitment of neutrophils and macrophages to infected tissue (Tesmer, Lundy, Sarkar, & Fox, 2008).

Despite its protective effects against extracellular infections, persistent or augmented production of Th17 cells and its associated cytokines have been implicated in the development of autoimmune diseases (Chappert, 2014; Harley et al., 2014; Kumar et al., 2016; Tesmer et al., 2008). Interestingly, the administration of neutralizing IL-17A monoclonal antibody to IBD patients in a clinical trial exacerbated the patient's symptoms, while the overgrowth in *Candida albicans* was also a noted side effect (Hueber et al., 2012). In similar studies with experimental models of autoimmune diseases in mice where the IL-17 receptor (IL-17R) was mutated,

dysbiosis of the commensal flora, including an overgrowth of SFB specifically, and augmented development of autoimmune inflammation occurred (Kumar et al., 2016).

Accordingly, individuals that are deficient in IL-17 signaling pathways are at increased susceptibility for bacterial and fungi infections, further supporting the idea that Th17 cells and IL-17 are necessary for maintaining commensal flora and the mucosal immunity (Gaffen, Jain, Garg, & Cua, 2014; Hirota, Ahlfors, Duarte, & Stockinger, 2012). Together, these data suggest that Th17 cells are necessary for regulating the intestinal microbial community, and generating mucosal immune responses to invading pathogens; however, the persistent generation of Th17 responses may lead to systemic inflammation and ultimately autoimmune disease. Because SFB have the unique ability to induce potent Th17 cell responses via physical contact with the host's immune system, many hypothesize that SFB may function as a driving force behind Th17-mediated autoreactive immune responses. However, Omenetti and colleagues recently made the distinction between the pro-inflammatory Th17 cell response induced by the opportunistic intestinal pathogen, *Citrobacter rodentium*, and the "homeostatic" Th17 cell response induced by SFB (Omenetti et al., 2019). Interestingly, the SFB-induced Th17 cell lineage did not produce interferon-gamma (IFN- γ)—a cytokine with pleiotropic immunological functions, including the orchestration of innate and adaptive immune responses against bacteria, viruses, tumor cells, and other pathogenic threats (Alspach, Lussier, & Schreiber, 2019; Omenetti et al., 2019). In addition to not participating in inflammatory processes, homeostatic Th17 cells exhibited a metabolic phenotype reminiscent of resting memory T cells (Omenetti et al., 2019). Since the major source of IFN- γ is from Th1 and NK cells, together these data suggest that there may be two kinds of inflammatory response, IFN- γ -mediated "reactive" inflammation and Th17-mediated "proactive" inflammation.

Localized and Systemic Immunological Activation by SFB

The specific SFB antigen(s) responsible for the microbe's potent immune response remain broadly elusive; however, the machinery which aid in the bacterium's adherence to the host epithelial cells, e.g., flagella or pili, has been postulated as potential immunogens (Ericsson, Hagan, Davis, & Franklin, 2014). Advances in determining how SFB interacts with the host immune system have been made, including a study by Yang et al, wherein the authors recombinantly expressed a genomic library of SFB DNA fragments in *E. coli* and measured the response of T cell receptor (TCR) hybridomas prepared from Th17 and non-Th17 intestinal CD4⁺ T cells (Yang et al., 2014). Screening the recombinant *E. coli* clones lead to the identification of two potential SFB antigens, which stimulated the Th17 TCR hybridomas in a relatively similar manner than that of fecal material derived from SFB-monocolonized mice (Yang et al., 2014). Unfortunately, the function of these putative SFB antigens is unknown, but they are predicted to exist as extracellular proteins, which is consist with the location at which most extracellular microorganisms present antigens (Chaplin, 2010; Yang et al., 2014). While the generation of an expression genomic library is an unbiased approach for recombinantly expressing and finding antigenic proteins produced by an organism, the comprehensiveness of the library can be greatly affected by cloning efficiency (Festa, Steel, Bian, & Labaer, 2013).

The host cytokine response to SFB has been modeled *in vitro* with mouse and human intestinal epithelial cells lines (Schnupf et al., 2015). Schnupf and colleagues show that after three days of co-culturing mouse and human intestinal epithelial cells with SFB in a transwell system, pleotropic cytokines, including TNF- α and Serum Amyloid A 1 (Saa1) were upregulated by SFB as well as Reg3 γ , a C-type lectin (Schnupf et al., 2015). The finding that SFB augments the expression of Saa1 and Reg3 γ is corroborated by a microarray experiment performed by

Ivanov et al wherein the transcription profile of the terminal ileum of GF mice monocolonized with SFB was investigated (Ivanov et al., 2009). Collectively, these studies suggest that SFB elicits an anti-inflammatory response in the host, as *Saa1* has been shown to suppress the activation and bactericidal activity of neutrophils in a microbiota-dependent manner (Murdoch et al., 2019). While the role of AMPs was discussed earlier in this review, SFB-induced expression of *Reg3 γ* represents a potential mechanism by which the colonization burden of SFB in the intestines is regulated.

Schnupf and colleagues also compared the SFB-induced cytokine profile to that induced by microbe-associated molecular pattern (MAMP) agonists, which are recognized by toll-like receptors (TLRs) (Schnupf et al., 2015). With their expression on many myeloid and lymphoid cells, TLRs are a class of proteins that recognize conserved structural and chemical patterns of microorganisms as a function of the innate immune response (Flo et al., 2001; Schnupf et al., 2015). SFB-induced a response reminiscent of the TLR2 agonists, Pam2Csk4 and Pam3Csk4 (Schnupf et al., 2015). Interestingly, exposing intestinal epithelial cells to flagellin, a TLR5 agonist, did not result in an augmented cytokine profile; however, it should be noted that nuclear factor-kappa B (NF- κ B), a downstream signaling target of TLR5, was not among the cytokines profiled in this study (Schnupf et al., 2015). A previous study by Kuwahara et al found three of the four flagellins genes present in the SFB murine genome to be TLR5 agonists and activators of NF- κ B when modeled in a luciferase reporter system in the human embryonic kidney 293 cell line (Kuwahara et al., 2011). These works, while not fully recapitulating the complex and intimate host-SFB interaction that occur in the intestinal milieu, provide a solid foundation for further investigating how the innate immune system responds to SFB.

One of the most unique properties of SFB is its ability to stimulate the postnatal immune system by priming and inducing T-cell responses and enhancing IgA secretion as efficiently as a naturally developed microbiome (Klaasen, Van der Heijden, et al., 1993; Talham, Jiang, Bos, & Cebra, 1999). Mice monocolonized with SFB have been found to generate 24-63% more intestinal IgA, with less than 1.4% of the IgA exhibiting SFB specificity (Klaasen, Van der Heijden, et al., 1993; Talham et al., 1999). Although a less robust induction was observed compared to controls, mice lacking PeyP can still generate SFB-mediated IgA responses, unlike *Escherichia coli* monocolonized mice, whose IgA responses were abolished (Lecuyer et al., 2014). These data further suggest that SFB is more effective in inducing isotype switching in B cells than *E. coli*. The GALT was found to be indispensable for SFB-induced IgA responses, per contra to Th17 cells, which could be expressed from the MLNs; however, the PeyP are required for IL-17 and Th17 cell induction with specificity to SFB antigens (Lecuyer et al., 2014).

Consistent with these results, a similar study reported that Th17 differentiation in mice lacking GALT was dependent on the expression of Class II Major Histocompatibility Complex (MHC II) on CD11c⁺ cells (intestinal dendritic cells) when presented with SFB or food antigens (Geem et al., 2014). Representing a potential feedback system for controlling Th17 cell production, type 3 innate lymphoid cells (ILC3), which are constitutive residents of the lamina propria (LP) layer of gut mucosa, have been found to inhibit microbiota- or SFB-driven Th17 production in a MHC II-dependent manner (Goto et al., 2014). Furthermore, ILC3-deficient mice exhibited increased Th17 cell production and intestinal inflammation, which was further exacerbated in the presence of SFB (Goto et al., 2014).

ILC3 play an imperative role in the orchestration of immune responses critical to maintaining the intestinal barrier and neutralize inflammatory responses. Altered ILC3 function

or perturbations in the signaling of ILC3's effector molecules, including IL-17A, IL-22, and interleukin-23 (IL-23), are associated with colorectal cancer and chronic inflammatory diseases such as IBD (Withers & Hepworth, 2017). Interestingly, dysbiosis and overgrowth of SFB have been observed in experimental animal models where the IL-22 and/or IL-23 signaling pathway is disrupted, but through the subsequent restoration of these signaling pathways, the SFB population could be controlled (V. F. Shih et al., 2014; Upadhyay et al., 2012).

Recent studies show during mucosal infections, such as with *Staphylococcus aureus*- or *Aspergillus fumigatus*-induced pneumonia, as well as challenges with intestinal pathogens, including *C. rodentium* and Rotavirus, SFB colonization yielded a protective advantage by way of its robust induction of Th17 cells and its associated microbial defenses, thereby highlighting the extreme impact of the gut microbiome on mucosal immunity (Gauguet et al., 2015; Ivanov et al., 2009; McAleer et al., 2016). Based on this evidence, we conclude that SFB is essential to the maturation of the mucosal immune system and provides the host with intrinsic protection against mucosal infections, which likely occurs through its induction of homeostatic Th17 cells.

CHAPTER 2

A BIOINFORMATICS APPROACH TO INVESTIGATE THE IMMUNOGENIC PROPERTIES OF SEGMENTED FILAMENOUS BACTERIA

Lexie C. Blalock¹, Hongwei D. Yu^{1, 2, 3*}

¹ Department of Biomedical Sciences, Joan C. Edwards School of Medicine at Marshall University, Huntington, West Virginia, United States of America

² Department of Pediatrics, Joan C. Edwards School of Medicine at Marshall University, Huntington, West Virginia, United States of America

³ Progenesis Technologies, LLC, One John Marshall Drive, Robert C. Byrd Biotechnology Science Center, Suite 314, Huntington, West Virginia, United States of America

* Corresponding authors

Email: yuh@marshall.edu

Abstract

In concert with its host, the gut microbiota plays an essential role in the development, function, and homeostasis of the immune system. As a member of the gut microbial community, Segmented Filamentous Bacteria (SFB), are distinct from other symbionts as these organisms potentiate post-natal immune maturation, enhance T helper 17 cell responses, and confer protection against mucosal infections of bacterial, viral, and fungal origin. Despite the wide range of vertebrate and invertebrate hosts that SFB colonize, this bacterium remains difficult to detect and culture *in vitro*. As a consequence, much remains unknown surrounding the SFB-specific antigens that stimulate the mucosal immune system. To address these issues, we employed a multi-omics approach to determine i.) the niche within the intestinal microenvironment in which SFB specifically reside, and ii.) bioinformatically investigated the immune-modulating antigens of SFB. Our results show that SFB preferentially colonize the mucosa proximal to the aggregated lymphoid nodules of the ileum, commonly referred to as Peyer's Patches. 16S rRNA analysis of the Peyer's Patches-associated microbiota suggests that this microbial community is phylogenetically distinct from the rest of the gastrointestinal tract. Using a high throughput *in silico* workflow, we identified 35 proteins as putative antigens within the SFB reference proteome, with each protein possessing one or more T cell epitope with binding affinity for Major Histocompatibility Complex Class II molecules. Among the top antigen candidates was flagellar cap protein (FliD), which is involved in bacterial motility and adherence. By implementing shotgun proteomics, we validated the *in situ* expression of FliD by SFB in the murine ileum. Operon-level analyses show that the expression of FliD is controlled by a global carbon storage regulator protein and is only translated in low-nutrient environments; however, experimental studies are necessary to validate this model. Collectively, these data

provide the first account of the proteins expressed by SFB *in situ*, thereby yielding significant insight into the biology of SFB and the mechanism by which this microbe interacts with the host immune system.

Introduction

The intestinal mucosal surfaces are home to a diverse and complex microbial community. Through the coevolution of their mutualistic relationship, humans have become dependent on the community of intestinal microbes, collectively referred to as the gut microbiota, for a variety of immunologic and metabolic functions (Flannigan & Denning, 2018). Emerging evidence suggests that alterations to the gut microbiota, such as those mediated by diet, environmental exposures to toxins and antibiotics, and genetic mutations, have clear implications in immune dysregulation (H. J. Wu & Wu, 2012). For example, Lynn and colleagues showed that antibiotic-induced intestinal dysbiosis in infant mice impaired their ability to generate antibody responses to five vaccines; however, the impaired antibody responses could be rescued through the restoration of commensal gut flora (Lynn et al., 2018)

Although compositionally rich and diverse gut microbial communities are essential to early immune development and subsequent immune homeostasis, Segmented Filamentous Bacteria (SFB), a gram-positive, spore-forming gut commensal has been shown —on its own— to drive maturation of the host immune system to an equal or greater extent than do the comprehensive gut microbiota (Klaasen, Koopman, et al., 1993; Kuwahara et al., 2011).

Colonization of SFB in mammals is believed to occur upon birth via vertical transmission from parent to offspring (Schnupf, Gaboriau-Routhiau, & Cerf-Bensussan, 2013) The intestinal SFB population has been reported to peak during the initial 36 months and 2 weeks of life in

humans and chickens, respectively (Yin et al., 2013). The age-dependent growth patterns of SFB suggest this microbe plays a vital role in post-natal immune development.

In humans, SFB colonization is a correlative of increased levels of secretory immunoglobulin A (IgA), T helper 17 cells (Th17), and B and T cell receptor signaling in the ileum (Chen et al., 2018), which echo the results of immunological studies performed following the monocolonization of SFB in murine models. The immunomodulatory effects elicited by SFB begin when the microbe initiates contact and colonization with the host intestinal epithelial cells (IECs) (Flannigan & Denning, 2018). This adherence triggers serum amyloid A (SAA) expression, an acute-phase response protein that is induced during tissue damage and infections, in IECs (Flannigan & Denning, 2018; Ivanov et al., 2009). In this scenario, SAA acts as a homing signal for CD11c⁺ antigen presenting cells (APCs), such as dendritic cells and macrophages, expressing major histocompatibility class II (MCHII) molecules (Yi, Jung, Han, Surh, & Lee, 2019). APCs subsequently migrate to the Peyer's Patches or mesenteric lymph nodes, where SFB antigen-specific naïve CD4 T cells differentiate to T helper 17 (Th17) cells, which secrete interleukin (IL)-17A, IL-17F, and IL-22 cytokines (Flannigan & Denning, 2018; Ivanov et al., 2009). A recent study revealed that unlike traditional Th17 cells, the SFB-induced "homeostatic" Th17 cell lineage, does not produce interferon-gamma, a cytokine with inflammatory-mediating functions (Omenetti et al., 2019). Moreover, these homeostatic Th17 cells exhibited a metabolic phenotype reminiscent of resting memory T cells, thereby suggesting the SFB-specific Th17 cell may be divergent from the traditional Th17 cell lineage, which is implicated in autoimmunity (Omenetti et al., 2019).

Through the enhanced mucosal immune responses outlined above, SFB confers protection against enteric bacterial and viral pathogens as well as respiratory fungal infections in

animal models (Gauguet et al., 2015; Ivanov et al., 2009; McAleer et al., 2016; Shi et al., 2019). Because SFB acts as a potent adjuvant of the mucosal immune system, this microbe has become a critical, and potentially confounding variable to consider in human and animal experiments (Macpherson & McCoy, 2015). For these reasons, commercially available mice are now screened for SFB, and assigned a SFB carrier status.

Despite the wide-range of vertebrate and invertebrate hosts in which SFB colonize, establishing and maintaining a culture of SFB *in vitro* remains a challenge and significant hindrance to advancing the understanding of SFB, including the mechanism by which SFB stimulates the mucosal immune system (Klaasen, Koopman, et al., 1993; Yin et al., 2013). Therefore, to address these issues, we employed a multi-omics approach to model the niche in which SFB specifically reside and determine the content of proteins expressed by SFB *in situ* within the murine intestinal microenvironment. Furthermore, we bioinformatically investigated immune-modulating antigens produced by SFB through the development of an *in silico* antigen prediction workflow reminiscent of reverse-vaccinology strategies (Rappuoli, 2000).

Results

SFB is present in high abundance in the Peyer's Patches in male and female mice from Jackson Laboratory and Taconic Biosciences

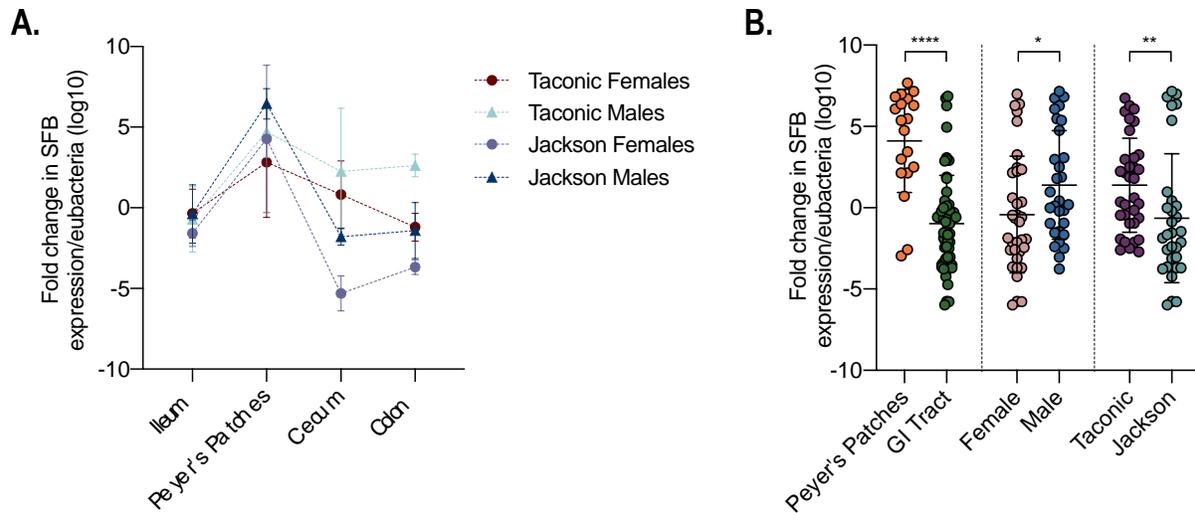
DNA-based detection of SFB has been widely variable among studies, host organisms, and even different mouse vendors (Chen et al., 2018; Ivanov et al., 2009; Sczesnak et al., 2011). For example, previously SFB was reported part of the endogenous gut flora in mice obtained from Taconic Biosciences, but not the Jackson Laboratory (Ivanov et al., 2009). Since this keystone study was published, mice from the Jackson Laboratory have been used as SFB-negative controls in experiments (Ge, Feng, Woods, & Fox, 2015). However, several factors,

including the organism's host-specific genome and age-dependent colonization patterns, may contribute to the disparities in SFB detection (Chen et al., 2018). Another explanation may be the sites at which the samples used for analysis are collected. As evidenced by electron microscopy images (Ivanov et al., 2009), SFB adhere tightly to the host intestinal epithelial cells; yet the majority of studies use the luminal contents or fecal pellets for DNA-based detection of SFB.

Therefore, the first aim of this study was to determine the distribution and relative abundance of SFB at various sites of the murine intestines using male and female mice purchased from Taconic Biosciences and the Jackson Laboratory. We collected the luminal contents from the distal small intestines (ileum), cecum, and large intestines (colon). In addition to these sites, we also collected the individual Peyer's Patches for qPCR analysis using a validated SFB-specific set of primers (Yin et al., 2013). Our results show that SFB was near or below the limit of detection in the sampled luminal sites, but were consistently detected at the site of the Peyer's Patches (Figure 1A). Statistical analysis of the aggregated qPCR data supports the hypothesis that the Peyer's Patches is the anatomic niche of SFB (Figure 1B). Although the relative abundance of SFB was significantly augmented in mice from Taconic Bioscience compared to mice from the Jackson Laboratory, our results illustrate the importance of sample site selection when attempting to detect the presence or absence of SFB. For example, if we had only collected and analyzed the luminal contents for this qPCR analysis, the majority of mice would have been considered SFB-negative.

Our qPCR data also suggest that male mice may naturally harbor more SFB than female mice. Interestingly, SFB colonization has been shown to have sex-dependent effects on disease phenotypes, with SFB-positive male mice exhibiting a decreased incidence of disease compared to their female counterparts (Wolfe, Moskowitz, Franklin, Wiemken, & Ericsson, 2020).

Testosterone and its suppressive effects on Th17 cell differentiation is believed to play a role in the gender dimorphism observed in inflammatory and autoimmune diseases (Arredouani, 2014;



Wolfe et al., 2020); however, the interplay between sex hormones, the gut microbiota, including SFB specifically, and immunity remains an active area of interest.

Figure 1. Distribution of SFB in the intestines of male and female mice purchased from Taconic Biosciences and the Jackson Laboratory

Semi-quantitative PCR (qPCR) was used to determine the relative abundance of SFB at different sites of the intestines (A). Aggregated qPCR data were analyzed to determine differences in the SFB colonization patterns between intestinal locations, genders, and mouse vendors (B). Mean \pm SD, * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$, Mann Whitney test.

The Peyer’s Patches-associated microbiota composition is phylogenetically distinct and the intestinal reservoir of SFB

To validate our qPCR findings of the enriched SFB population within the Peyer’s Patches, we selected five Peyer’s Patches DNA samples to further analyze through next-generation sequencing using primers targeting the V4 region of the bacterial 16S rRNA gene. As a reference point for comparing the composition and diversity of the Peyer’s Patches-associated

microbiota (PPAM), we chose the cecal microbiome because many gut microbiome-centric studies collect and analyze this community, and thus its characteristics are well defined.

We first assessed the community diversity within samples (alpha diversity) using the Shannon diversity index, which is a quantitative metric species richness and evenness (Figure 2A) (Haegeman et al., 2013; C. E. Shannon & Weaver, 1949). Kruskal-Wallis testing showed that Shannon diversity was significantly attenuated in the PPAM compared to the cecum ($p = 0.0022$). Diversity between samples (beta diversity) was evaluated with the Unweight Unifrac distance, a qualitative, phylogenetic metric (Lozupone, Hamady, Kelley, & Knight, 2007). We chose this beta diversity metric because it is more sensitive to low-abundance features, which was what we anticipated observing in the PPAM. Using PERMANOVA, we tested the null hypothesis of no differences in the community structure between the PPAM and cecal microbiota. Notable phylogenetic divergence ($p = 0.001$) was observed between these two communities (Figure 2B). As evidenced by the PcoA, which provides a visual perspective on the phylogenetic dissimilarity between the PPAM and cecal microbiota, the PPAM is a distinct but homogeneous community (Figure 3B). Collectively, these diversity analyses suggest that the microbial community proximal to the Peyer's Patches is comprised of a small number of different microbial species that are extremely specific to the region.

Next, we evaluated the number of sequences found in the PPAM and cecum that mapped to SFB. Amplicon sequence fragments were clustered at the sub-operational taxonomic unit (OTU) level of ASVs (Callahan, McMurdie, & Holmes, 2017). As anticipated, the biomass of the Peyer's Patches was much smaller compared to cecum, with the mean ASV frequency per PPAM and cecum sample being 7,164.8 and 2,246,191.8 sequences, respectively. Of those sequences detected in the PPAM, 2221 ± 2304 mapped to SFB, which was a significant ($p =$

0.0013) increase from the 91 ± 66.71 SFB sequences mapped in the cecal samples (Figure 2D). Taxonomic analysis of the PPAM further confirmed that SFB was the most abundant genus present in this community, representing on average 31% of the total microbial population (Figure 2E). The next most abundant genus was *Acinetobacter* (22%). Although *Acinetobacter baumannii* is an opportunistic pathogen, the majority of *Acinetobacter* species are non-pathogenic, environmental organisms (Peleg, Seifert, & Paterson, 2008). Considering the increased abundance and co-occurrence of SFB and *Acinetobacter* in the PPAM, it is possible that SFB growth is dependent on nutrients or cofactors provided by *Acinetobacter*. In sum, this set of experiments support the notion that the Peyer's Patches are the intestinal reservoir for SFB, and should be the site of sampling for the accurate detection of SFB.

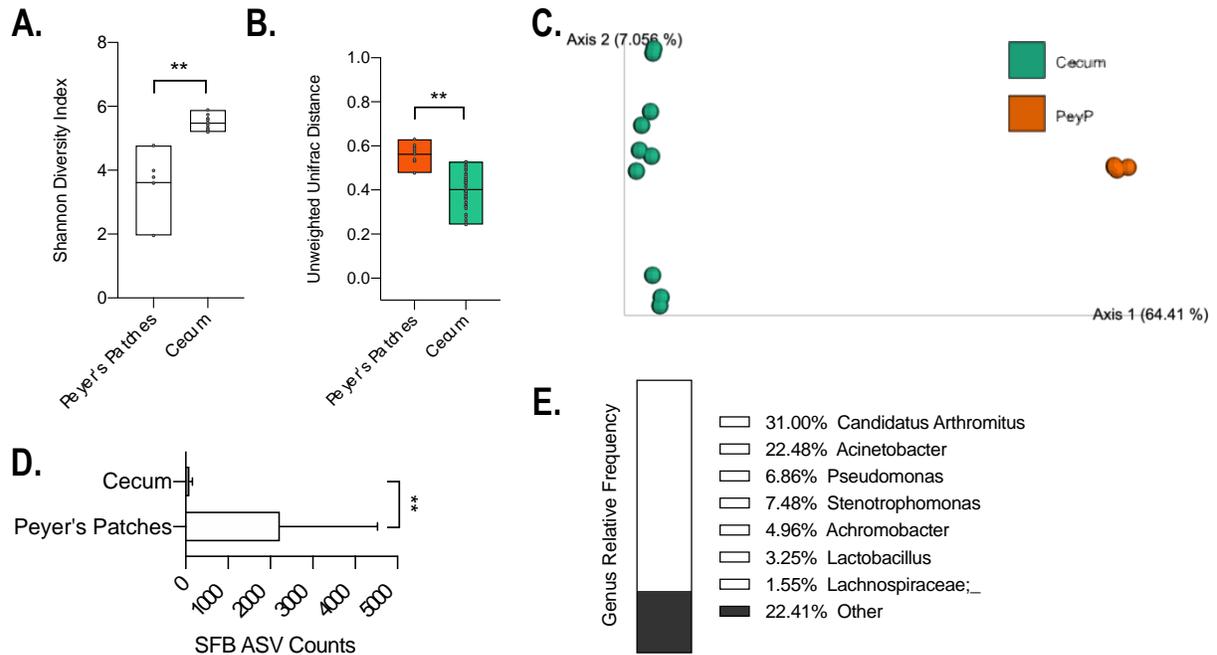


Figure 2. Diversity and compositional analyses of the Peyer's Patch-associated microbiota
 Comparison of species richness and diversity within samples from the Peyer's Patches and cecum (A). Comparison (B) and visualization (C) of the phylogenetic dissimilarity between

samples from the Peyer's Patches and cecum. Bar plot showing the number of sequences detected in the Peyer's Patches and cecal microbial communities that mapped to SFB (D). Mean \pm SD, $**P < 0.01$, Mann Whitney test. Taxonomic profile of the Peyer's Patch-associated microbiota at the genus level (E).

In silico reverse-vaccinology approach identifies thirty-five SFB antigens with CD4 T cell immunogenicity and MHCII recognition

The immunological response to SFB colonization has been well defined, however; the antigen(s) expressed by SFB responsible for this adjuvant-like effect is still unclear and work in this area continues to be impeded by the inability to sustain a mono-culture of this organism *in vitro*. Therefore, we sought to apply a reverse-vaccinology (RV) approach to identify potential SFB antigens. Strategies aimed at predicting bacterial antigens *in silico* have not only expedited the vaccine development process, but also yielded several promising vaccine candidates for long-standing pathogens, including *Campylobacter* (Meunier et al., 2016) and *Streptococci* (Ebrahimi & Mohabatkar, 2018) species.

The pipeline utilized in our *in silico* approach selected SFB antigens based on the criteria of their predicted cellular location, transmembrane helices, adhesion properties, CD4 T cell immunogenicity, and MHCII binding ability (Figure 3). After these criteria were applied in a stepwise fashion, 35 SFB proteins with at least one peptide exhibiting properties of CD4 T cell immunogenicity and natural processing by MHCII molecules remained (see Appendix B for protein and peptide information). We analyzed these proteins as a network in Cytoscape (P. Shannon et al., 2003), and performed a functional enrichment analysis using the StringApp (Doncheva, Morris, Gorodkin, & Jensen, 2019). Results from the functional enrichment show that the majority of candidate antigens are components of bacterial flagella or related to motility

and chemotaxis (Figure 4). These results are consistent with hypotheses proposed by several groups of SFB investigators, who postulate that SFB achieves its immunogenicity through protein(s) which aid in the bacterium's adherence to the host epithelial cells, e.g., pili or flagella (Ericsson et al., 2014; Nkamba et al., 2020; Y. Wang et al., 2019).

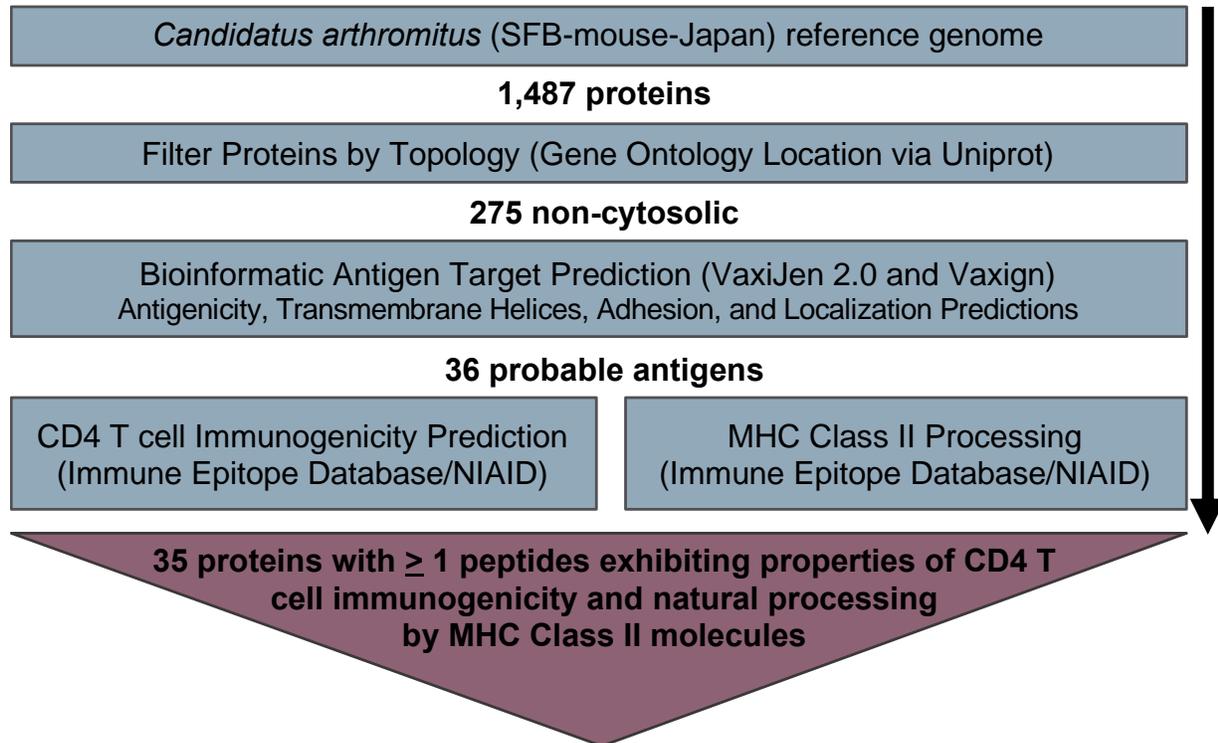
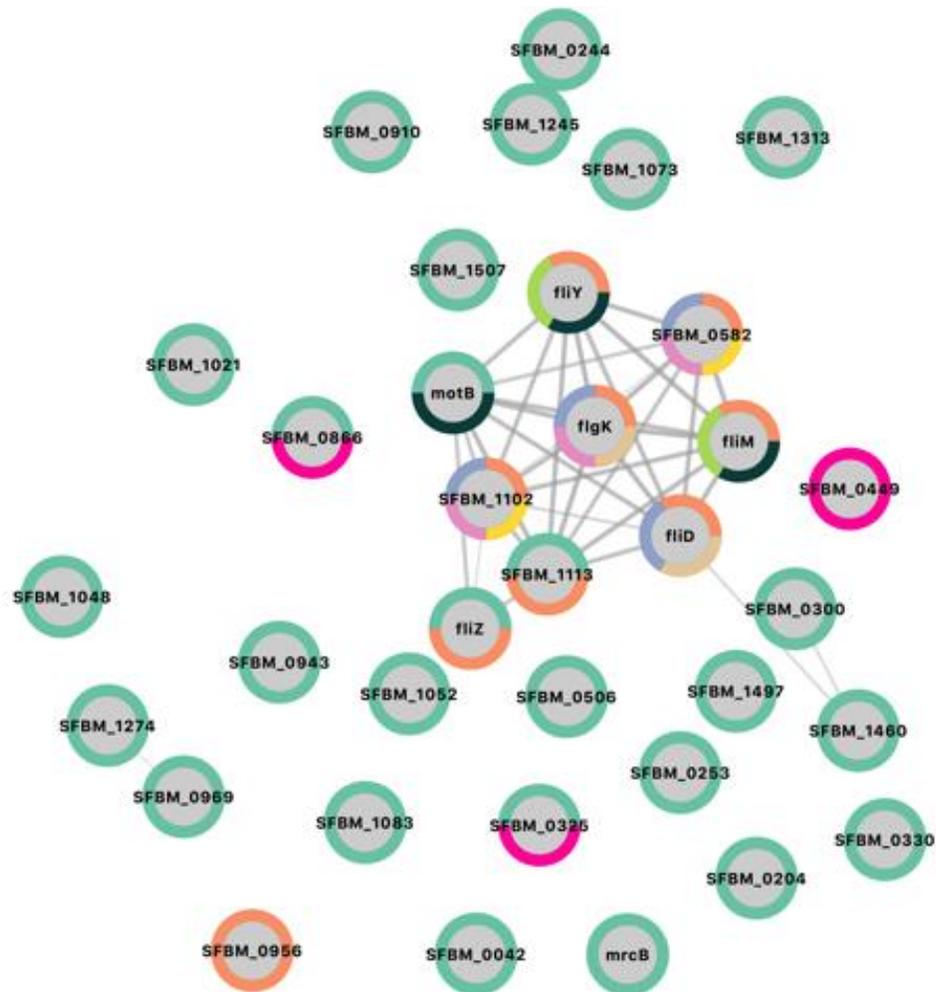


Figure 3. In silico prediction workflow for the identification of SFB antigens
Summary of the workflow and databases used to predict SFB immune-modulation proteins.



category	chart color	term name	description	FDR value
UniProt Keywords	Green	KW-1133	Transmembrane helix	3.3E-11
UniProt Keywords	Orange	KW-0282	Flagellum	1.01E-6
UniProt Keywords	Blue	KW-0975	Bacterial flagellum	0.0018
Pfam	Pink	PF06429	Flagellar basal body rod FlgEFG protein C-terminal	0.0093
Pfam	Light Green	PF01052	Type III flagellar switch regulator (C-ring) FlhC C-term	0.0236
Pfam	Yellow	PF00460	Flagella basal body rod protein	0.0287
UniProt Keywords	Light Orange	KW-0964	Secreted	0.0438
KEGG Pathways	Dark Green	asf02030	Bacterial chemotaxis	0.045
KEGG Pathways	Pink	asf02024	Quorum sensing	0.0454

Figure 4. Functional analysis and network of the SFB candidate antigens identified
 Protein-protein network consisting of the thirty-five SFB proteins selected by our *in silico* antigen prediction pipeline. Each protein is represented by a node, which is color-coded

according to its functional annotation. Nodes connected by a line or “edge” indicate protein-protein interactions.

Shotgun proteomics analysis of the murine ileal mucosa reveals that SFB express flagella *in situ*

Analysis of murine and rat SFB genomes show that while extremely minimal in size, they contain a full set of flagella biosynthetic genes (Kuwahara et al., 2011; Prakash et al., 2011). Therefore, one of the most puzzling elements about SFB has been the failure to detect the expression of SFB flagella *in situ* in a variety of hosts, as reported by numerous authors (Nkamba et al., 2020). Adding to this intrigue has been the historic lack of flagellated SFB in electron microscopy images (Nkamba et al., 2020). Considering the lack of evidence surrounding flagellated SFB, coupled with the fact that the majority of our *in silico* SFB candidate antigens are components on the flagellar machinery, we employed a shotgun proteomics approach to determine the *in situ* expression of SFB proteins, particularly flagellar-related proteins, in the ileal mucosa of mice.

After applying the threshold of at least 2 unique peptides identified for each protein, a total of 200 proteins were annotated as *Candidatus arthromitus*, with 99 proteins mapping to the SFB rat (rSFB) reference genome (SFB-rat-Yit), and 101 mapping the SFB mouse (mSFB) reference genome (SFB-mouse-Japan) (see Appendix C for all protein annotations). Of the SFB mouse and rat proteins identified, 14 proteins were redundant, i.e., seven proteins that mapped to the SFB rat genome also mapped to SFB mouse genome based on a different set of peptides identified. We analyzed the mouse and rat SFB proteins as networks in Cytoscape (P. Shannon et al., 2003), where we applied the Markov Cluster Algorithm (MCL) technique, a bioinformatics method frequently applied to identify, visualize, and delineate functional protein modules from

larger protein networks (Y. K. Shih & Parthasarathy, 2012). Using an inflation value of 4.0, we identified three functional modules from the SFB mouse protein network (Figure 6) and SFB rat protein network (Figure 7). We subsequently performed a functional enrichment analysis on the functional modules using the StringApp (Doncheva et al., 2019). Upon characterization, mSFB modules exhibited significantly enriched functions of metal- and nucleotide-binding as well as antibiotic biosynthesis (Figure 6A), DNA helicase and methylation enzymes (Figure 6B), and two-component systems associated with flagella machinery (Figure 6C). The rSFB modules showed significantly enriched functions related to ATP-binding, pyrimidine biosynthesis, and pyrimidine metabolism (Figure 7A), metal-binding sites as well as the metabolism of pyruvate, propionate, and carbon (Figure 7B), and penicillin- and beta-lactam binding proteins with transpeptidase activity (Figure 7C).

These data provide the first account of the proteins expressed by SFB *in situ*. Consistent with previously characterization of the SFB genome, we did not observe any enriched functions related to amino acid and vitamin/cofactor metabolism, as SFB lack the majority of genes needed for such functions; however, our proteomic analysis provides evidence that SFB do exhibit pyrimidine biosynthetic abilities, which hitherto was undetermined (Kuwahara et al., 2011). Another key finding was the presence of several flagella proteins, including FliK, FliR, FlgK, FlgD, FliD, and FliC. The detection of multiple flagellar proteins expressed *in situ* in the murine ileal mucosa of mice is the first, and perhaps the most essential step in experimentally validating the SFB antigenic targets predicted *in silico*.

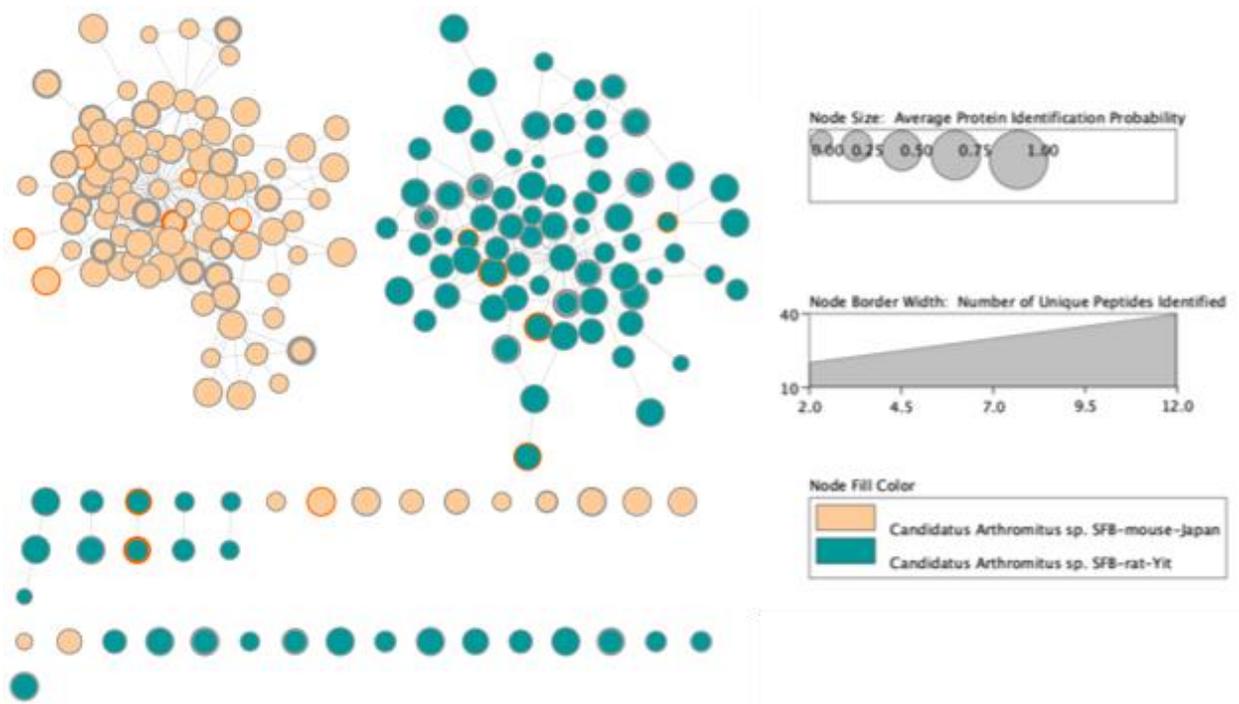


Figure 5. Networks of mouse and rat SFB proteins identified in the murine ileal mucosa via shotgun proteomics

Protein interaction networks showing the SFB proteins annotated as SFB mouse (orange nodes) and SFB rat (green nodes) proteins. Node size is indicative of the average protein identification probability, while the node border width represents the number unique peptides identified for each protein. The node border of redundant proteins is highlighted in red.

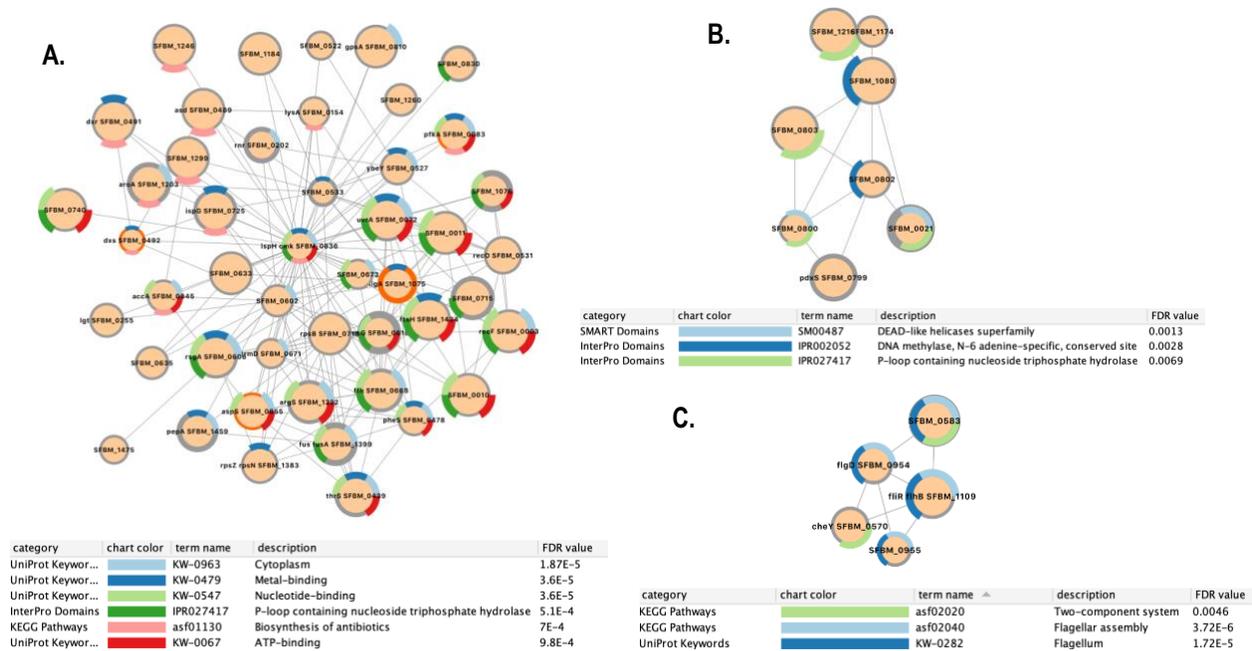


Figure 6. Functional characterization of modules identified from mouse SFB proteins

Functional analysis of the three largest clusters generated by Markov clustering of the mouse

SFB protein network (Figure 5). Using split donut charts, the enriched function(s) of each node

is visible by the color(s) present in the node's border. Within each module, node border colors

correspond to the enriched functional categories, as annotated by the StringApp in Cytoscape.

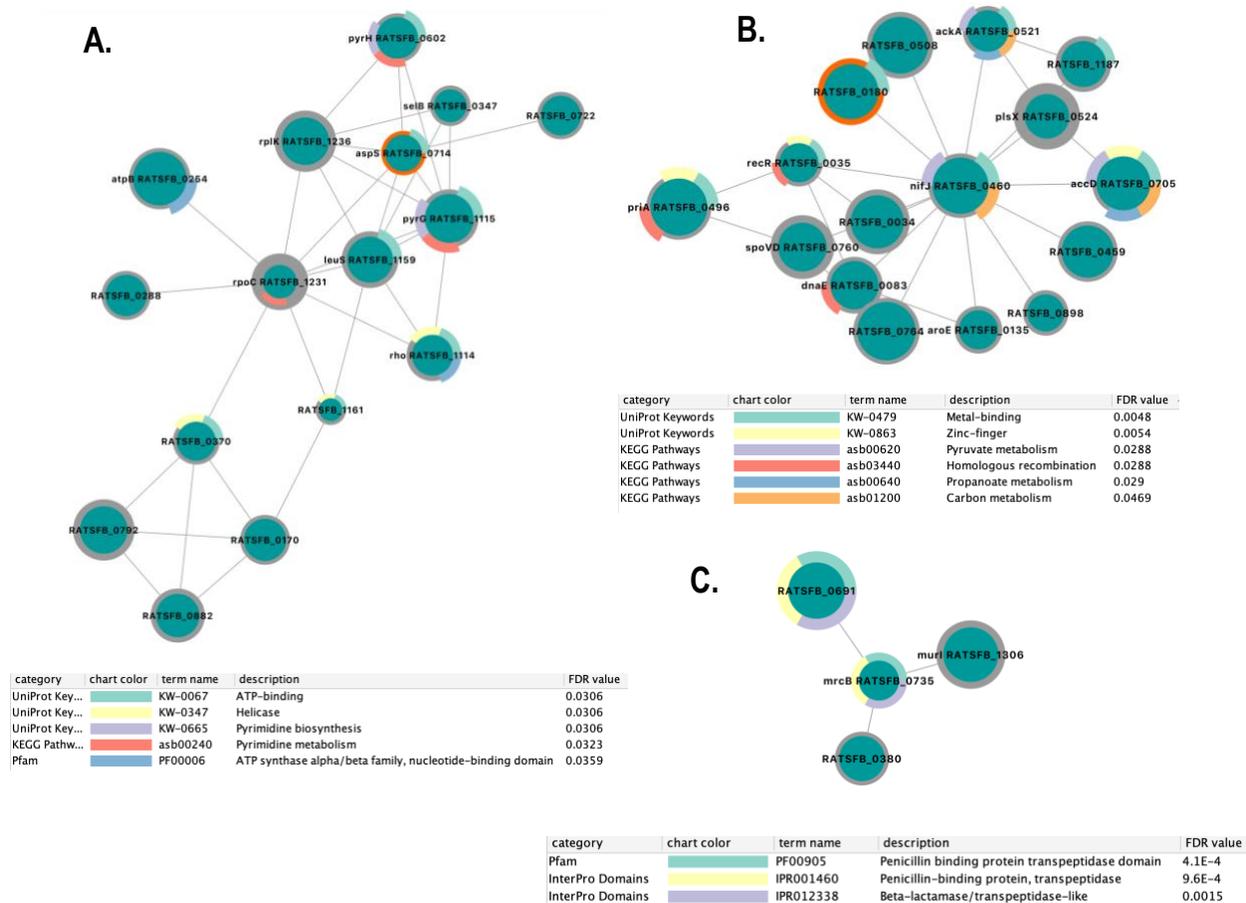


Figure 7. Functional characterization of modules identified from rat SFB proteins

Functional analysis of the three largest clusters generated by Markov clustering of the mouse

SFB protein network (Figure 5). Using split donut charts, the enriched function(s) of each node

is visible by the color(s) present in the node's border. Within each module, node border colors

correspond to the enriched functional categories, as annotated by the StringApp in Cytoscape.

Nutrient signals are postulated to regulate the expression of flagellar cap protein

FliD

Having now validated the *in situ* expression of flagellar proteins by SFB, we choose to further investigate flagellar cap protein, FliD. This protein was among the top predicted antigens that showed strong adhesion and antigenicity score. We also found FliD to encode 10 different

peptide sequences that possessed CD4 T cell immunogenicity and natural recognition by MHCII molecules (see Appendix B for 15mer peptide sequences). FliD is widely conserved among flagellated microbes, and although it functions as the “cap structure” on the distal end of flagella filaments, FliD has been shown to be essential for the adhesion and colonization of *Pseudomonas aeruginosa* and *Clostridium difficile* in respiratory and intestinal tracts, respectively (Arora, Ritchings, Almira, Lory, & Ramphal, 1998; Tasteyre, Barc, Collignon, Boureau, & Karjalainen, 2001). More recently, FliD was shown to mediate the adherence of atypical enteropathogenic *Escherichia coli* and *C. jejuni* to host IECs (Freitag, Strijbis, & van Putten, 2017; Sampaio et al., 2016). Based on this literature, it is likely that FliD plays a vital role in how SFB adhere to and colonize the gut, and quite possibly the immune response that ensues.

To further investigate FliD and what may control its expression, we analyzed the operon on which FliD is located using the Prokaryotic Operon Database available online (Taboada, Ciria, Martinez-Guerrero, & Merino, 2012) coupled with the Genomic Region Viewer available on PATRIC, the online bacterial bioinformatic resource center (Wattam et al., 2017). Using the SFB-mouse-Japan reference genome, our analyses showed that *fliD* is encoded on SFB operon 326, and is located downstream of genes *fliW*, *csrA*, and *fliS* (Figure 8A). Carbon storage regulator A (CsrA) is a widely-conserved small RNA binding protein and pleiotropic regulator of virulence factors, including motility and biofilm formation (Dugar et al., 2016; Muller, Gimpel, Wildenhain, & Brantl, 2019). Based on signals from the environment, CsrA is believed to repress or activate the translation of certain genes by binding to GGA motifs around the Shine-Dalgarno sequence of target mRNAs (Dugar et al., 2016; Muller et al., 2019).

A complex feedback mechanism exists in flagellated bacteria, whereby flagellar machinery transcripts, which often include the *FliD* operon, are governed by a partner switching mechanism involving *FliW* and *CsrA* proteins. In *Bacillus subtilis*, stoichiometry studies have shown the flagellar machinery proteins, *FliW*, and *CsrA* system functions at intracellular concentrations of nearly 1:1:1, which suggests this system is tightly regulated (Oshiro, Rajendren, Hundley, & Kearns, 2019). Although this system and the stoichiometry may be altered in SFB, it is reasonable to believe a similar homeostatic system exists in SFB, as the delicate equilibrium would help prevent states of unnecessary energy expenditure (Oshiro et al., 2019), which is likely essential to the survival of auxotrophs like SFB.

Although we did not detect *CsrA* or *FliW* proteins in our *in situ* proteomic analysis, a predicted protein-protein interaction analysis in STRING (Figure 8B) suggests that SFB proteins *FliW* and *CsrA* have strong binding activity (0.995 combined confidence score). Moreover, in the predicted protein-protein interactions network, *CsrA* appears to be situated between proteins with functions related to a phosphotransferase system (PTS) (red nodes) and bacterial chemotaxis (green nodes) and flagellar assembly (yellow nodes). The primary function of the bacterial PTS is to transport and phosphorylate sugar (Saier, 2015). Based on this protein interaction data and our operon analyses, we have developed a model for the regulation of *FliD* (Figure 8C). In this working model, the translation of *FliD*, although ultimately controlled through the *FliW*-*CsrA* partner switching mechanism, is dependent upon the availability of nutrients in the localized environment. Since SFB is auxotrophic (Kuwahara et al., 2011), and therefore must be in a nutrient-rich environment to survive, it is logical that signals of a nutrient-poor environment initiate a motility-activated response.

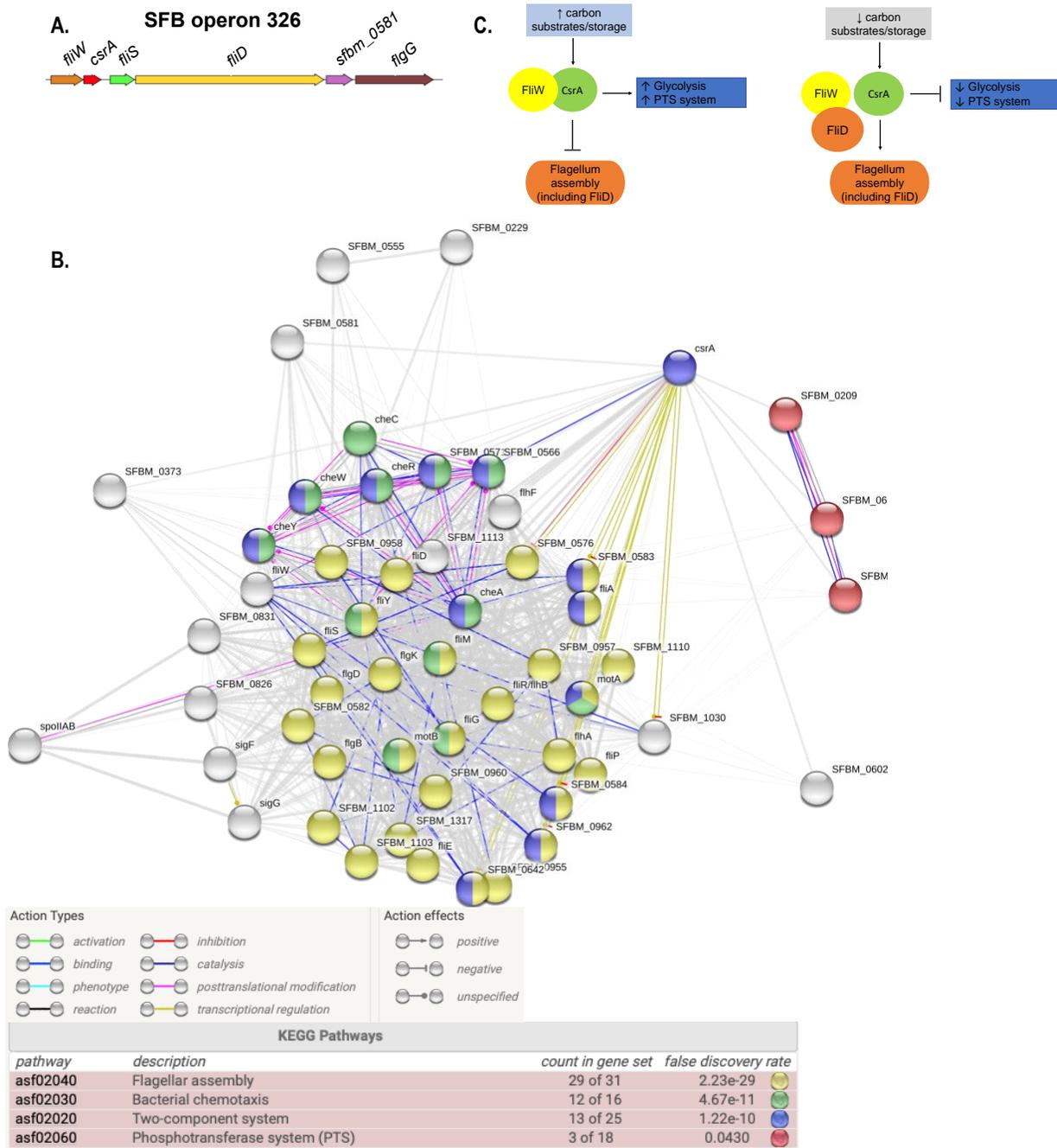


Figure 8. Operon and protein interaction analyses indicate environmental nutrients are a determinant of flagellar cap protein FliD transcription. Schematic diagram SFB-mouse-Japan operon 326, which encodes flagellar cap protein, FliD (A).

To create the predicted protein-protein interaction, FliW, CsrA, FliS, and FliD were queried, and

visualized with 50 additional predicted functional partners on STRING. Functional enrichment was applied to the protein network. Nodes are colored based on their functional annotation by KEGG pathways (B). Our working model of FliD regulation is dependent up on the status of nutrients in the localized environment (C). In the diagram on the left, where carbon substrates are readily available, FliW allosterically antagonizes CsrA RNA-binding activity, thereby preventing FliD transcription. On the other hand, in nutrient-poor environments, signals from the PTS initiate a motility-activated response. Although a sigma factor is likely involved in this initiation step, intracellular FliD binds to FliW, which enables FliD and other flagella machinery transcripts to be translated with the assistance of RNA-binding protein, CsrA.

Conclusion

This work yields critical insight into the biology of SFB and the organisms' localization in the intestines, both of which are likely implicated in the way SFB interacts with the host. Firstly, we determined that SFB specifically colonize the ileal mucosa proximal to the Peyer's Patches in male and female mice obtained from Taconic Biosciences and the Jackson Laboratory. The average relative abundance of SFB in the Peyer's Patches was more than a 5-fold increase compared to the average relative abundance in the ileal, cecal, and colon contents. Considering that mice with immune deficiencies, e.g., in IL-17, IL-22, and IL-23 signaling pathways (Kumar et al., 2016; V. F. Shih et al., 2014; Upadhyay et al., 2012), have a higher burden of SFB, the immune system is likely playing an active role in limiting and/or maintaining the growth of SFB, while also benefiting from the enhanced mucosal immune defenses associated with SFB colonization. Moreover, because the Peyer's Patches represent a site of systemic entry, it is possible SFB spatially or chemically inhibit the invasion of pathogens as a function of an evolved, mutualistic relationship with the host (Lai et al., 2020).

In alignment with this qPCR analysis, next-generation sequencing revealed that SFB was a defining and dominating feature of the PPAM, representing nearly one-third of the total microbial population. Moreover, alpha and beta diversity analyses showed that the PPAM is a phylogenetically distinct, but a surprisingly homogenous microbial community, which hitherto, had not been defined nor characterized. Considering the previously reported discrepancies in the DNA-based detection of SFB, our data suggest that the presence of SFB has likely been underestimated due to the sample locations, e.g., the fecal pellets or cecal contents, commonly used for screening. Furthermore, the specific localization of SFB in the Peyer's Patches suggest that this region is enriched in nutrients conducive to SFB growth. A metabolomics analysis of this region may shed light on cofactors and substrates necessary for SFB growth, particularly *in vitro*.

Secondly, our *in silico* prediction of SFB antigens suggest that one or more flagellar proteins may be involved in inducing the immune response. At the time of this study, no microscopy images existed to prove SFB were indeed flagellated. Since our proteomic analysis of the murine ileal mucosa, during which we identified six flagellar proteins expressed *in situ*, a recent study by Nkamba and colleagues showed through transmission electron microscopy imaging that SFB derived from mono-associated mice and rats exhibit a flagellated phase (Nkamba et al., 2020). Moreover, the authors tested the toll-like receptor five (TLR5), a cell receptor for bacterial flagella, induction potential of recombinantly expressed flagellar filament protein, FliC (Nkamba et al., 2020). In a TLR5-expressing human embryonic kidney cell line, FliC did stimulate TLR5, although these results were not compared to other TLR5 agonists (Nkamba et al., 2020). Other authors have demonstrated that the FliC protein expressed by SFB exhibits Th17 cell induction potential *in vitro* and *in vivo*, although the Th17 cell response and

subsequent cytokines expressed were significantly attenuated in comparison to the orthologous protein produced in *Salmonella* (Y. Wang et al., 2019). While we did identify FliC within the *in situ* proteomics dataset, this protein was not among the thirty-five SFB antigens predicted through the *in silico* reverse-vaccinology pipeline.

Yang et al previously identified two putative SFB antigens by screening an expression library of SFB proteins and measuring their potential to bind and stimulate Th17 T cell receptors hybridomas (Yang et al., 2014). While the function of these two antigenic proteins remains unknown, these proteins are noted as unique to SFB (Yang et al., 2014). Upon comparison, there was no overlap between the core epitope sequences of the putative SFB antigens identified in this study and those reported by Yang et al. Although our SFB antigen prediction results are not directly aligned with these previous works, the antigenic proteins and epitopes previously identified by Nkamba et al and Yang et al will be useful to include during the experimental validation of the SFB antigens predicted in this study, including FliD.

Thirdly, the results of each experiment performed in this study have sequentially led us to the conclusion that environmental nutrients play a significant role in the localization of SFB in the Peyer's Patches and that flagellation is a survival mechanism induced when the concentration of nutritional substrates are low in the environment. In many microbial systems, bacterial motility is triggered by environmental cues, with low nutrient concentrations being the most common to initiate the chemotaxis signaling cascade (Chantranupong, Wolfson, & Sabatini, 2015). Nutrient bioavailability in the Peyer's Patches region could serve as a chemoattractant for SFB. Interestingly, M-cells proximal to the Peyer's Patches secrete a unique profile of glycoproteins (Kimura et al., 2015), including glycoprotein-2, which could potentially be the target of the several proteases and peptidases observed in the SFB genome (Kuwahara et al., 2011).

While motility is likely essential to SFB successfully colonizing in the Peyer's Patches, we propose that the immune response that ensues from physical contact of flagellated SFB is mediated, at least in part, by FliD, which resides at the distal-most end of flagella. In addition to experimentally determining the "homeostatic" Th17 cell induction potential of FliD, future work should focus on validating the working model for FliD's regulation, as the FliW- CsrA partner switching system likely regulates the expression of the other flagellar proteins.

There are several biomedical uses and applications that should incentivize the determination of the antigenic properties associated with SFB. For example, if the full immune response elicited by SFB could be recapitulated by recombinantly-expressing a single SFB protein, it would have extreme potential as a probiotic to enhance mucosal immunity and potentially confer protection against mucosal pathogens. Moreover, because the Peyer's Patches are the portal to the largest immune organ of the body, engineering a microbe that targets the induction site of the mucosal immune system is an attractive target for the precise delivery of therapeutics (Jung et al., 2010) or mucosal vaccines (Lycke, 2012), both of which currently have limited rates of efficacy.

A limitation of this study is the variability in mouse strain and age among experiments. While acknowledging that these extraneous variables could affect the outcome of experiments, such as the native microbiota composition, our justification lies in the principle of reducing the number of animals used in scientific experiments, as the majority of samples used in this study were taken from mice that were sacrificed to satisfy the needs of other experiments. On the other hand, the fact that our results were consistent, and SFB was detected by DNA- and protein-based

methods across mice of different genders, strains, ages, and mouse vendors lends credence to our finding that SFB are widely distributed and function as essential and natural adjuvants to the mucosal immune system.

Materials and Methods

Animal Model

All animal studies were performed under an approved protocol by the Marshall University Institutional Animal Care and Use Committee. For qPCR and 16S rRNA analysis, 8-week old male and female C57BL/6J (B6) mice were purchased from Taconic Farms (Hudson, NY) and The Jackson Laboratory (Bar Harbor, ME). It should be noted that these mice were sacrificed within 48 hours of arriving to the animal resource facilities at Marshall University, as to ensure that their endogenous flora was still representative of the vendor's environments.

The mouse strains used for gut mucosa in-situ proteomic analyses were male B6, DBA/2, and C3H mice purchased from Charles River Laboratories (Wilmington, MA). These mice ranged from 8 to 32 weeks in age at the time of sample collection.

Isolation of intestinal sections and Peyer's Patches

Anatomic-specific sites of the distal small intestine (ileum), cecum, and large intestine (colon) were identified and sectioned off. Intestinal sections were cut longitudinally, and luminal contents were removed with a plastic inoculation loop (Thermo Fisher Scientific, Waltham, MA, USA). Surgical tools were sterilized with ethanol between each intestinal section to avoid cross-contamination of samples. Peyer's Patches were removed from the ileum as previously described (De Jesus, Ahlawat, & Mantis, 2013). Briefly, the terminal ileum was sectioned off, and the Peyer's Patches were physically identified on the anti-mesenteric side of the intestine. The

average mouse yields approximately 5-10 Peyer's Patches. Using either curved surgical scissors or a scalpel, the Peyer's Patches were excised from the surrounding intestinal tissue. The Peyer's Patches were washed with PBS to remove possible contaminants from the lumen. All samples were stored in a 1:1 (v/v) of nitrogen-flushed skim milk and PBS and preserved at -80 °C.

Bacterial DNA Extraction

Genomic DNA was extracted from ileal, cecal, and colon samples via the Easy Fecal DNA Extraction Kit (Zymogen, Irvine, CA) per the manufacturer's instructions. Genomic DNA was extracted from the Peyer's Patches via the Tissue and Blood Miniprep Kit (Qiagen, Hilden, Germany) using the Spin-Column Protocol. Prior to DNA extraction, Peyer's Patches were individually cleaned of additional tissue and washed 3X with PBS to remove residual fecal material. DNA purity and concentration was determined on the SpectraMax i3x Multi-Mode Microplate Reader System (Molecular Devices, San Jose, CA).

Quantitation of Intestinal Microbiota with qPCR

Primers targeting the 16S rRNA gene specific to segmented filamentous bacteria (SFB) and conserved among all bacteria (Eubacteria [Eub]) were purchased from Integrated DNA Technologies (Coralville, IA). SFB primers SFB779F: 5'-TGTGGGTTGTGAATAACAAT-3' and SFB1308R: 5'-GGTTAGCCCACAGGTTTCGG-3' yielded an amplicon of approximately 619 base pairs (bp), while Eubacteria primers Eub27F: 5'-AGAGTTTGATCMTGGCTCAG-3' and Eub1492R: 5'-CGGTTACCTTGTTACGACTT-3' yielded an amplicon of approximately 1500 bp. Each qPCR reaction was performed in technical replicates of two, each containing 10 µL Power SYBR Green PCR Master Mix (Thermo Fisher Scientific, Waltham, MA), 0.4 µM (final concentration) forward primer, 0.4 µM reverse primer, and genomic DNA diluted in RNase/DNase-free water, yielding a total volume of 25 µL. SFB template DNA was set to 200

ng per reaction, while total bacteria (Eub) was set to 1 ng per reaction. 16S rRNA gene amplification was performed on the StepOnePlus RT-PCR System (Applied Biosystems, Waltham, MA) under the following conditions: 25 °C for 1 minute; 95 °C for 3 minutes to denature; 95 °C for 10 seconds, 60 °C for 45 seconds repeated for 40 cycles; 57-95 °C for the disassociation stage. Relative quantification of SFB was calculated using the $\Delta\Delta$ Ct Method (Livak & Schmittgen, 2001).

16S rRNA Library Preparation and Sequencing

Genomic DNA from cecal samples was PCR-amplified using Illumina barcoded forward 341F primer and reverse 518R primer (Bartram, Lynch, Stearns, Moreno-Hagelsieb, & Neufeld, 2011) targeting the V3 region of the bacterial 16S rRNA gene as previously described (Cockburn et al., 2012). Each PCR reaction contained a total of 60 ng genomic DNA, 0.5 μ l of 20 μ M forward primer, 0.5 μ l of 20 μ M reverse primer, 1 μ l AccuPrime PCR Master Mix with Taq Polymerase, 5 μ l 10X AccuPrime Buffer II, and PCR-grade water for a total volume of 50 μ l. PCR reactions were amplified on the MJ Research PTC-200 Peltier Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA) using the following conditions as previous described: 6 minute denaturation step at 95 °C; 30 cycles of 95 °C for 2 minutes, 50 °C for 2 minutes, and 72 °C for 2 minutes; 4 minute extension step at 72 °C. Size and quality of the V3 libraries was assessed by electrophoresis on a 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA, USA) and found to be in 325 to 340 bp range. The pooled libraries were sequenced over two lanes in a 2x180 paired-end fashion in the rapid run mode on an Illumina HiSeq1500 with the addition of 15% PhiX to enhance sequencing quality and diversity.

In a separate sequencing experiment, genomic DNA from Peyer's Patches samples was amplified using Illumina-barcoded 806 R reverse primer and 515 F forward primer targeting the

V4 region of the bacterial 16S rRNA gene (Walters et al., 2016). PCR amplification was performed pursuant to the Earth Microbiome 16S protocol (earthmicrobiome.org/protocols-and-standards/16s/). PCR reactions were set up with 11.375 uL sterile water, 10.625 uL of Ex Taq Master Mix (Takara Bio, Mountain View, CA) with 1 uL of 515F forward primer, 2 uL of extracted DNA, and 1 uL of barcoded 806R reverse primer. The V4 region of the 16S rRNA gene was then amplified using the following thermal cycler conditions: 3 minute denaturation at 94 C; then 35 cycles of 94 C for 45 s, annealing at 50 C for 60 s, and extension at 72 C for 90 s; followed by a 10 minute hold at 72 C for final extension of amplicons. Reactions were then checked on a 2% agarose e-gel (Thermo Fisher Scientific, Waltham, MA) for successful amplicons at approximately 390 bp. Following completion of PCR, each PCR reaction was pooled together in approximately equimolar concentration. The pooled sample was then loaded on a 2% agarose gel for purification to remove primer dimers and other non-specific amplified DNA. Bands at 390 bp were then isolated and cut from the gel and purified with the QIAquick Gel extraction kit (Qiagen, Hilden, Germany) to recover pure DNA. Finally, the purified library was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA) and validated using the Agilent Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA) prior to submission for sequencing. The library was ultimately multiplexed with other pure libraries and weighted to ensure even sequencing between projects and samples. The final sequencing library was checked via Bioanalyzer dsDNA High Sensitivity Assay (Agilent Technologies, Santa Clara, CA, USA) to assess purity. This final multiplexed library was sent for Miseq 2x250 sequencing using v2 chemistry with approximately 20% PhiX spike in.

Microbial Compositional Analyses

Demultiplexed, paired-end sequences were imported into QIIME2 (Caporaso et al., 2010) via the Casava 1.8 paired-end demultiplexed fastq format. Sequences were denoised using the dada2 package (Callahan et al., 2016). Because different sequencing primers were used for Cecum and Peyer's Patches samples, taxonomy had to be assigned independently for each location before merging samples. Taxonomy was assigned to ASVs using the QIIME2-formatted version of SILVA_128_release (Quast et al., 2013; Yilmaz et al., 2014) at the sequence identity cutoff of 99%. Peyer's Patches and cecum representative sequences were merged to perform phylogenetic analyses. A phylogenetic tree was constructed using Fasttree with normalized data that had been rarefied to the sub-sequencing depth of 2,000. Diversity differences within samples (alpha diversity) were analyzed with the nonparametric Kruskal-Wallis test using Shannon's diversity metric (C. E. Shannon & Weaver, 1949). Diversity differences between samples (beta diversity) were analyzed with Permutation Multivariate Analysis of Variance (PERMANOVA) (M. J. Anderson). Testing with PERMANOVA was implemented using the Unweighted Unifrac Distance (Lozupone et al., 2007). Principal-coordinates analysis (PCoA) was visualized with Emperor (Vazquez-Baeza, Pirrung, Gonzalez, & Knight, 2013).

In silico prediction of SFB antigens

The workflow developed for our *in silico* approach to identifying SFB antigens was based on the criteria of their cellular location, transmembrane helices, adhesion properties, CD4 T cell immunogenicity, and MHCII binding ability. Because immune induction by SFB occurs upon the organism's physical contact with host IECs, we reasoned that the immunomodulating antigen is likely a non-cytosolic protein. Therefore, the first step involved topology filtering of all protein in the reference genome (SFB-Mouse-Japan) using the UniProt Knowledgebase

(UniProt, 2019). The number of transmembrane helices detected was also incorporated as a filter, as proteins containing more than one transmembrane helix are difficult to recombinantly express, which is generally an important step in the experimental validation of *in silico* predicted targets (Y. He, Xiang, & Mobley, 2010). Next, we incorporated and cross-referenced the results of two reverse-vaccine applications, Vaxign (Y. He et al., 2010) and VaxiJen (Zaharieva, Dimitrov, Flower, & Doytchinova, 2019). Antigen predictions by VaxiJen are alignment-independent and solely based on the physicochemical properties of each analyzed protein (Zaharieva et al., 2019). In contrast to this approach, Vaxign utilizes protein sequence alignment to predict the subcellular location, transmembrane helix topology, adhesin probability, protein conservation, MHC Class I and II binding, and function annotations of each protein analyzed (Y. He et al., 2010).

Based on immunological studies performed, it has been shown that SFB antigen(s) are (i) recognized in a MHC Class II-dependent manner and (ii) induce the differentiation of CD4 T cells. Thus, we utilized the Immune Epitope Database (IEDB) Analysis Resource provided through the National Institute of Allergy and Infectious Diseases <http://tools.iedb.org/main/> to perform CD4 T cell immunogenicity and MHCII (MHCII-NP) natural processing predictions. Both analyses were performed using the IEDB recommended settings.

Intestinal Mucosa Collection for Proteomic Analysis

The mucosa-associated microbiome was collected from the ileum using Basic Protocol 4 described by Tong et al. 2014 (Tong, Jacobs, McHardy, & Braun, 2014) with some modifications. Briefly, the ileum was identified, sectioned, and the luminal contents were flushed with 30-60 mL of cell culture-grade PBS via a syringe with a 27 gauge x ½ inch hypodermic needle. Next, the ileum was cut longitudinally and flushed 3X with PBS and placed into a 50 mL conical tube containing 16 mL of pre-warmed PBS with 0.1% Tween 80 (Sigma-Aldrich, St.

Louis, MO). The conical tube was then placed horizontally in a 37 °C orbital shaker, set at 180 rpm for 40 minutes. The conical tube was then vortexed for 20 seconds, and its contents were then filtered twice through a 70 µm nylon mesh filter (Thermo Fisher Scientific, Waltham, MA). The filtered supernatant was spun at 4200 rpm for 15 minutes at 4 °C. The pellet was resuspended in 200 µL of PBS and stored at -80 °C.

Total Protein Extraction

Protein was extracted from intestinal mucosa samples with ProteaPrep Non-ionic Cell Lysis Kit (Protea Biosciences, Morgantown, WV) pursuant to the manufacture's protocol. Briefly, mucosa samples were centrifuged at 12,000 x g for 5 minutes at 4 °C, followed by resuspension of the pellets in 1 mL of 1X TBS Buffer (Protea Biosciences, Morgantown, WV). The wash step with TBS Buffer was repeated. Pelleted cells were resuspended in 0.5 mL of ProteaPrep Non-ionic Cell Lysis Reagent with 25 µL Protease Inhibitor Cocktail (Sigma-Aldrich, St. Louis, MO) for a final concentration of 5%. The cell suspension was incubated for 30 minutes on ice. While remaining on ice, cells were lysed during 3 cycles of 1-2 minute sonication with the Sonic Dismembrator D100 (Fisher Scientific, Hampton, NH). Cell lysates were centrifuged at 12,000 x g for 10 minutes at 4 °C, and the supernatant was collected for downstream analyses. Total protein concentrations were determined using the Bradford Protein Assay (Bio-Rad Laboratories, Hercules, CA).

Mass spectrometry

Mass spectrometry was performed by the Proteomics and Mass Spectrometry Facility at the Donald Danforth Plant Science Center in St. Louis, MO. Protein samples (30µg) were reduced with 10 mM TCEP followed by alkylating with 40 mM iodoacetamide. Samples were then digested with trypsin in a 1/50 enzyme/protein ratio at 37 °C overnight. The digested

samples were acidified with 1% TFA then cleaned up with C18 tip. The extracted peptides were dried down and each sample was resuspended in 15 μ L 1% acetonitrile/1% formic acid. For each sample, 3 μ L were analyzed by LC-MS with a Dionex RSLCnano HPLC coupled to a LTQ-Orbitrap Velos Pro (Thermo Fisher Scientific, Waltham, MA) mass spectrometer using a 2h gradient. Peptides were resolved using 75 μ m x 25 cm PepMap C18 column (Thermo Fisher Scientific, Waltham, MA). All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.5.1.0). Mascot was set up to search *Candidatus arthromitus* and *Mus musculus* databases from Uniprot. Digestion enzyme was set as trypsin. Mascot was searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 10 ppm. Oxidation of methionine, carbamidomethylation of cysteine, and acetylation of N-terminal of protein were specified in Mascot as variable modifications.

Criteria for peptide and protein identification

Scaffold (version Scaffold_4.8.9, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, Keller, Kolker, & Aebersold, 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Peptide identifications were accepted if they could be established at greater than 5.0% probability to achieve an FDR less than 5.0% by the Scaffold Local FDR algorithm (Tang, Underwood, Gielbert, Woodward, & Petrovska, 2014). Protein identifications were accepted if they could be established at greater than 95.0% probability (C. Qin et al., 2016) and contained at least 2 identified peptides.

Proteomic Analysis

Mouse and rat SFB protein annotations were further assigned through the UniProt Knowledgebase (UniProt, 2019). Protein networks were imported into Cytoscape (P. Shannon et al., 2003), and clustered using the Markov Cluster Algorithm (inflation value = 4.0) to enable functional modules to be identified and further characterized. Function enrichment analyses of mouse and rat SFB protein modules was performed using the StringApp (Doncheva et al., 2019).

Predicted protein-protein interaction network with SFB Operon 326 proteins was created using STRING database v11 (Szklarczyk et al., 2019). FliW, CsrA, FliS, and FliD were queried, and visualized with 50 additional predicted functional partners.

Statistical analyses

Statistical analyses were performed in GraphPad Prism 8 (GraphPad Software, Inc. La Jolla, CA). Differences between the abundance of SFB were assessed using the Mann Whitney test. Prior to statistical testing, the Shapiro-Wilk normality test was used to determine whether the data were consistent with a Gaussian distribution. Results of statistical testing are denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

CHAPTER 3

INTERACTION OF THE GUT MICROBIOTA, DIET, AND HOST GENETICS IN TALLYHO/JNG MICE: A POLYGENIC MODEL FOR METABOLIC SYNDROME

Lexie C. Blalock¹, Jacaline K. Parkman¹, Jun Fan¹, Donald A. Primerano¹, Jung Han Kim¹,
James Denvir^{1*}, and Hongwei D. Yu^{1, 2, 3*}

¹ Department of Biomedical Sciences, Joan C. Edwards School of Medicine at Marshall
University, Huntington, West Virginia, United States of America

² Department of Pediatrics, Joan C. Edwards School of Medicine at Marshall University,
Huntington, West Virginia, United States of America

³ Progenesis Technologies, LLC, One John Marshall Drive, Robert C. Byrd Biotechnology
Science Center, Suite 314, Huntington, West Virginia, United States of America

* Corresponding authors

Email: yuh@marshall.edu (HYD) denvir@marshall.edu (JD)

Abstract

Metabolic syndrome (metS) afflicts approximately 25% of individuals worldwide. Thus, a clinically relevant model that is characterized by the physiological and genetic complexities of human metS is of critical need for (i) understanding the etiological factors of the disease, and (ii) developing therapeutic strategies—both of which potentially exist within the gut microbiota. Here, we present the first characterization of the cecal microbiota of the TALLYHO/Jng (TH) mouse: an inbred polygenic model for human metS. The metabolic and gut microbial profiles of TH mice maintained on chow, semi-purified high-sucrose high-fat and semi-purified high-sucrose low-fat diets were analyzed in parallel to C57BL/6J (B6) mice. High-resolution analysis of 16S rRNA amplicon sequence variants (ASVs) revealed that variation in gut microbial structure and function was primarily driven by dietary factors. The gut microbiota's response to diet, however, varied by strain, with the TH microbiota exhibiting an increased susceptibility to diet-mediated modulation and the expansion of pathobionts. By integrating the metabolic and 16S rRNA data through computational and machine learning methods, we found several microbial-metabolic interactions and identified 124 ASVs predictive of the absence or presence of metS in mice. By performing a sub-analysis with these ASVs, we determined that diet was the most significant explanatory factor (48%) of these features. Taxa belonging to Enterococcaceae, Leuconostocaceae, Enterobacteriaceae, and Desulfovibrionaceae families were among the most predictive features of metS and strongly associated with augmented triglyceride and blood glucose levels. On the other hand, healthier mice were defined by the presence of Ruminococcaceae, Muribaculaceae, and Eggerthellaceae taxa. Our results suggest that diet is a critical modulator and determinant of the gut microbiota's stability and composition. The

predictive microbial signatures identified in this study may be critical to the development of metS in the TH mouse as well as their human analogs.

Introduction

Over 93.3 million adults and 13.7 million children in the United States are obese (BMI \geq 30 kg/m²) (Hales, Carroll, Fryer, & Ogden, 2017). It is estimated that by 2030, the prevalence of obesity will reach 51% (Finkelstein et al., 2012), which foreshadows an increase in the health and economic burden associated with the obesity epidemic (Hammond & Levine, 2010). Obesity is a complex disease that originates from cumulative interactions between exogenous factors, e.g., diet and exercise (Bouchard, 2010), and biological factors, e.g., inherited predispositions and epigenetic events (Bouchard, 2010; Herrera, Keildson, & Lindgren, 2011; Williams, Mesidor, Winters, Dubbert, & Wyatt, 2015). In addition to its association with mortality, obesity is also a critical risk factor for other chronic conditions, including metabolic syndrome (metS) (Global et al., 2016; Wahba & Mak, 2007).

The National Cholesterol Education Program Adult Treatment Panel III defines metS as the presence of three or more of the following disorders: (i) abdominal obesity, (ii) hypertension, (iii) dyslipidemia, (iv) hypertriglyceridemia, and (v) hyperglycemia (Grundy et al., 2004).

Although the inheritance of obesity-susceptible genes is estimated to explain 50-90% of the variation in intrapersonal bodily fat mass (Parks et al., 2013), the etiological perspective of obesity, as well as other metabolic and inflammatory diseases, has recently shifted to consideration of the gut microbiome as a contributing factor (Hinney, Vogel, & Hebebrand, 2010; Ley, Turnbaugh, Klein, & Gordon, 2006; Org et al., 2015; Yamada et al., 2018).

The host's genetic background (Blekhman et al., 2015) and lifestyle factors, such as nutrition (Carmody et al., 2015; David et al., 2014), antibiotics, and exercise, shape and modify

the gut microbiome. Under homeostatic conditions, the gut microbiome plays an essential role in host nutrient metabolism, immunity, and protection from pathogens (Rios-Covian et al., 2017; Schwartz et al., 2010; Velagapudi et al., 2010). However, an imbalanced microbial community or “dysbiosis” has become a hallmark of immune and metabolic abnormalities in humans and animals (Clemente, Ursell, Parfrey, & Knight, 2012; Cosorich et al., 2017; Ley et al., 2005; Morgan et al., 2012; Turnbaugh et al., 2006). Although the causal mechanisms and temporal relationship between metabolic disorders and gut microbes are yet to be unraveled, the gut microbiome has been implicated in microbiota transfer studies where traits of leanness, obesity, and insulin resistance were transmissible to germ-free mice (Backhed et al., 2007; Carmody et al., 2015; Rabot et al., 2010; Turnbaugh, Backhed, Fulton, & Gordon, 2008).

While these findings have widened the perspective of the profound impact of the gut microbiome on metabolic function, the majority of surrogate animal models of human obesity or type 2 diabetes mellitus (T2D) rarely reflect the genetic mutations underlying human forms of these complex diseases. Genetic variants in the mammalian genome shape the host’s response to diet, the susceptibility to metS-related diseases, and the architecture of the gut microbiome (Org et al., 2015; Parks et al., 2013). A single nucleotide polymorphism (SNP) in the host genome can contribute to quantifiable alterations in the composition and function of the native gut flora (Blekhman et al., 2015; Hall, Tolonen, & Xavier, 2017). For example, a SNP in the lactase-encoding protein LCT has been shown to indirectly regulate the abundance of the lactose-metabolizing genus *Bifidobacterium* by altering the bioavailability of lactose in the gut (Blekhman et al., 2015). Therefore, because the host’s genotype likely impacts the interpretation of microbiome studies, it must be considered when modeling diseases, especially those of polygenic origin (Fuchs et al., 2018).

The TALLYHO/Jng (TH) mouse is an inbred polygenic model can present disease phenotypes similar to human metS including obesity (Parkman et al., 2016), impaired cutaneous wound healing (Nguyen et al., 2013), hypercholesteremia (Parkman et al., 2017), hyperlipidemia (J. H. Kim & Saxton, 2012), and increased β -cell mass and islet insulin secretion in response to glucose (Mao, Dillon, McEntee, Saxton, & Kim, 2014). Additionally, male TH mice exhibit a severe diabetic phenotype of glucose intolerance and hyperglycemia that resembles human T2D (J. H. Kim & Saxton, 2012; Parkman et al., 2016). Parkman et al. (Parkman et al., 2016) showed that the penetrance of the metS phenotype exhibited by male TH mice could be modulated by semi-purified high-sucrose high-fat (HSHF) and or semi-purified high-sucrose low-fat (HSLF) diets. However, the contribution of the gut microbiota to the metabolic dysfunction in the TH mouse is unknown.

This study aimed to characterize the metabolic and cecal microbial profiles of TH mice, and normal C57BL/6J (B6) mice maintained on chow, HSHF, or HSLF diets. We assessed the influence of host genetics and diet on the gut microbiota's diversity, composition, and predicted function. Furthermore, we employed a computational approach to identify genera and amplicon sequence variants (ASVs) associated with metabolic variables as well as taxonomic signatures predictive of metS, which may be critical to the development of metS in the TH mouse as well as their human analogs.

Results

We assessed the interplay of diet and host genetics on the gut microbiota and metabolic profiles of B6 and TH mice fed chow, HSHF, and HSLF diets. At the endpoint of the study, the cecal contents were collected, and the V3 region of the 16S rRNA gene was sequenced via a paired-end strategy. Microbiota data were analyzed with QIIME 2 bioinformatics software

(Caporaso et al., 2010). Amplicon sequence fragments were clustered at the sub-operational taxonomic unit (OTU) level of ASVs (Callahan et al., 2017). A total of 6,708 unique ASVs were identified. The mean frequency (sequence count) per ASV was 22,769.9, and the mean ASV frequency per sample was 2,246,191.8. See Figure 9 for the compositional profile of the cecal microbiota, including an evaluation of the Firmicutes to Bacteroidetes ratio by diet. Consistent with our findings, this ratio has been reported augmented in mice and humans fed high-fat diets (Ley, Turnbaugh, et al., 2006; Turnbaugh et al., 2006).

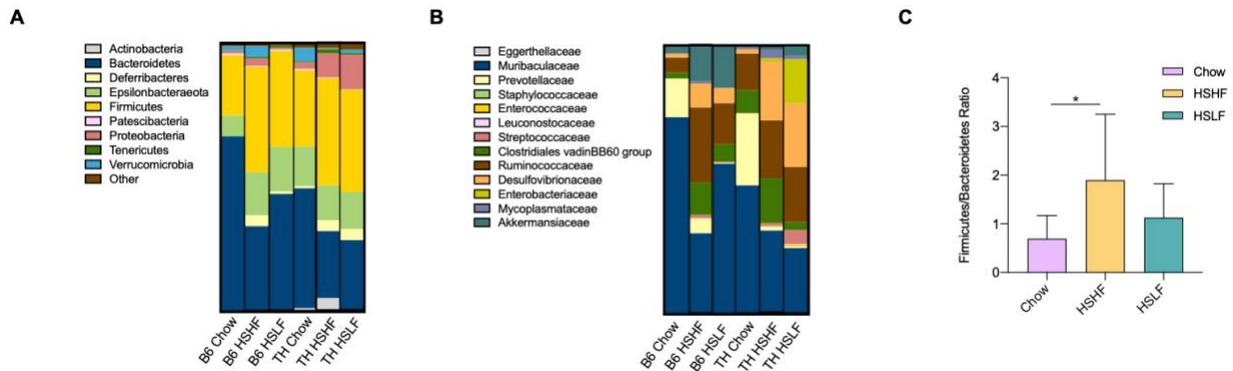


Figure 9. Cecal microbiota composition

Distribution of major microbial phyla (A) and families (B) in B6 and TH mice on chow, HSHF, or HSLF diets. (C) Bar plot showing the Firmicutes to Bacteroidetes ratio across diets. Data are representative of mean \pm SD. Differences among group means were assessed with a one-way ANOVA test followed by Tukey's multiple comparisons test.

Dietary factors primarily drive variation in gut microbial diversity and community structure

Alpha diversity analyses

Patterns in microbial diversity reveal critical information about the stability and functionality of the community (Haegeman et al., 2013). We first evaluated the diversity within

samples (alpha diversity) using the Shannon diversity index (Figure 10A), which is a robust and quantitative estimator of species richness and evenness (Haegeman et al., 2013; C. E. Shannon & Weaver, 1949). Kruskal-Wallis testing at the group level showed that Shannon diversity was similar among the B6 and TH mouse strains ($p = 0.6992$) and metS/non-metS metabolic clusters ($p = 0.5616$), but different between diets ($p = 0.0013$). Pairwise testing among diet groups revealed that Shannon diversity was decreased within the HSLF group compared to the HSHF ($q = 0.0084$) and chow ($q = 0.0014$) groups. Using the QIIME 2 longitudinal plugin (N. Bokulich et al., 2018), we tested the impact of categorical variables (strain, diet, strain x diet interaction, and sequencing lane) on Shannon diversity (Table 1). Interaction between strain and diet variables had the most pronounced effect on Shannon diversity ($p < 0.0001$) followed by diet ($p = 0.00017$), strain ($p = 0.6662$) and sequencing lane ($p = 0.9117$). The impact of continuous covariates (metabolic variables) on Shannon diversity was also examined (Table 1), revealing no linear relationships ($p > 0.1$).

Categorical variable	Statistical Test	F	p-value
Strain x Diet	ANOVA	12.1443	3.83E-05
Diet	ANOVA	10.0707	0.000172674
Strain	ANOVA	0.187897	0.666254
Lane	ANOVA	0.0123838	0.91177
Continuous covariate	Correlation	R	p-value
Body Weight	Spearman	-0.0818	0.549
Insulin	Spearman	0.0472	0.7297
Plasma Triglyceride	Spearman	0.0413	0.7624
Cholesterol	Spearman	0.0181	0.8949
Blood Glucose	Spearman	0.1763	0.1937
Fat Mass	Spearman	0.0364	0.7901
Lean Mass	Spearman	-0.2215	0.1009

Table 1. Effect of categorical and continuous variables on alpha-diversity.

Beta diversity analyses

To understand how the compositional structure of the cecal community varied between samples (beta diversity), community dissimilarity was quantitated with the Weighted Unifrac distance and visualized with PCoA (Figures 10B-D). The Weighted Unifrac distance (Lozupone et al., 2007) is a quantitative metric that accounts for the phylogenetic relationship shared between features. Using PERMANOVA, we tested the null hypothesis of no differences in the community structure between different mouse strains, diet, or metS/non-metS metabolic groups. Phylogenetic divergence in community structure (Figure 10C) was observed between mouse strains ($p = 0.001$), diets ($p = 0.001$), and metabolic clusters ($p = 0.001$). Pairwise PERMANOVA testing showed significant ($q < 0.05$) dissimilarity between all diets, with the greatest phylogenetic distance observed between the chow diet and the HSHF and HSLF diets.

Because PERMANOVA analysis can be affected by intragroup dispersion, i.e., the heterogeneous spread of data within a group (Marti J. Anderson, 2017), we performed PERMDISP to parse out whether the results of PERMANOVA were (i) reflective of true differences between group centroids, or (ii) due to a high degree of intragroup variance. Thus, we tested the null hypothesis of equal variances between strain, diet, and metabolic groups. The results of PERMDISP testing with the Weighted Unifrac distance (Figure 10C) suggest homogeneous dispersions between strain ($p = 0.525$), diet ($p = 0.265$), and metabolic ($p = 0.992$) groups, thereby confirming that the differences observed in PERMANOVA testing are the result of the true difference between group centroids.

Next, we implemented ADONIS analysis to investigate whether variation in the cecal microbiota could be explained by host and environmental factors (Figure 10D). Except for the sequencing lane, which was intentionally added as a negative control, all other variables and

interactions (strain x diet) were significantly ($p < 0.05$) associated with gut microbial variation. Among all the covariates tested, nearly one-third (29.2%) of the variation in the cecal microbiota was explained by diet (Fig 1D). Following diet, the interaction between strain and diet (11.4%), strain (10.2%), and plasma triglyceride (7.3%) were among the top explanatory variables of microbial variance. Collectively, the results of our quantitative diversity analyses show that dietary factors played a significant role in determining community biodiversity, variation, and phylogenetic composition, and overshadowed the effects of host genetics. Metabolic variables did not correlate with biodiversity, suggesting that the number and distribution of microbial features within samples remained consistent when tested over a continuous range of values representing the metabolic traits. In contrast, a significant amount of microbiota variation could be explained by some of the metabolic variables, indicating that metabolic traits had an impact on the phylogenetic composition observed between samples.

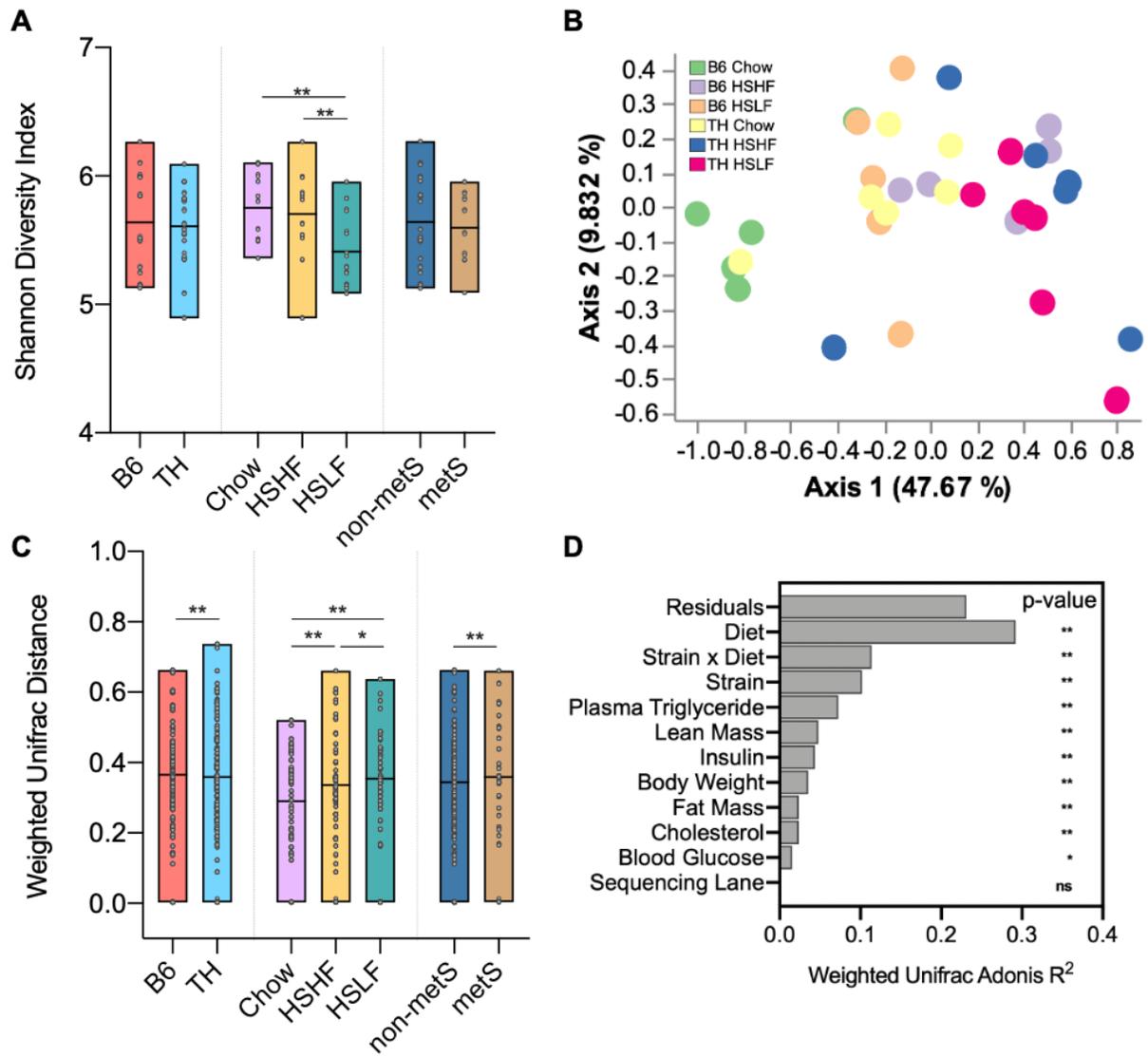


Figure 10. Analysis of community diversity and explanatory variables

(A) Pairwise comparisons of microbial diversity within strain, diet, and metabolic clusters. (B)

Principle coordinate analysis displaying phylogenetic dissimilarity (Weighted Unifrac distance)

between microbiota samples of different mouse strains and diets. (C) Pairwise comparisons of

community composition between strain, diet, and metabolic clusters. (D) Bar plot showing the

percentage of microbiota variation explained by host and environmental factors. Floating bars

are representative of the minimum, maximum, and group mean. Statistical significance is

representative of FDR-adjusted p-values (q-values).

The TALLYHO/Jng microbiota is highly susceptible to diet-mediated modulation

Taxonomy was assigned to ASVs using the SILVA 132 release 16S rRNA reference database, which is reflective of the latest changes in taxonomic classification and nomenclature (Henderson et al., 2019). These classification changes include (i) the recent reassigned of Epsilonproteobacteria, once a class of the Proteobacteria phylum, to the candidate phylum, Epsilonbacteraeota (Waite et al., 2017), (ii) the addition of candidate phylum Patescibacteria (Sanchez-Osuna, Barbe, & Erill, 2017), and (iii) a proposed change in the nomenclature of the Bacteroidales family S24-7 group to Muribaculaceae (Lagkouvardos et al., 2019).

Using LEfSe, a metagenomic biomarker discovery tool that identifies discriminating taxa between defined classes (Segata et al., 2011), we analyzed the impact of diet on the gut microbiota within B6 and TH strains as well as the impact of host genetics by comparing diet-matched strains. Because LEfSe is equipped to analyze hierarchical data, we performed this analysis over the entire phylogeny of features (phylum to genus level) to maximize the discovery of the hierarchical relationships and phylogenetic patterns discovered in the microbiota. LEfSe ranks the results of differential features using effect size, which provides an estimation of the magnitude of the observed phenomenon (Segata et al., 2011). We increased the discriminatory power of this analysis by increasing the default LDA threshold from 2.0 to 3.0.

Effect of diet on the microbiota of B6 mice

A total of 12 discriminating features were found to vary in abundance between the diet groups in B6 mice (Figure 11A). The most abundant features in the B6 chow class were Prevotellaceae and Muribaculaceae families of the Bacteroidetes phylum as well as several taxa from the Clostridiales order, including three uncharacterized genera of the Clostridiales vadinBB60 group and the Butyrate-producing genus, *Butyricoccus*, of the Ruminococcaceae

family (Boesmans et al., 2018). The B6 HSLF class was characterized by three taxa from the Clostridiales order, including unclassified Christensenellaceae and Peptostreptococcaceae taxa and the genus, *Dorea*, of the Lachnospiraceae family. *Lactococcus*, a member of the Streptococcaceae family, was the only discriminating feature of the B6 HSHF class. The phylogenetic distribution of and hierarchical relationships between these features can be visualized as a cladogram (Figure 11B).

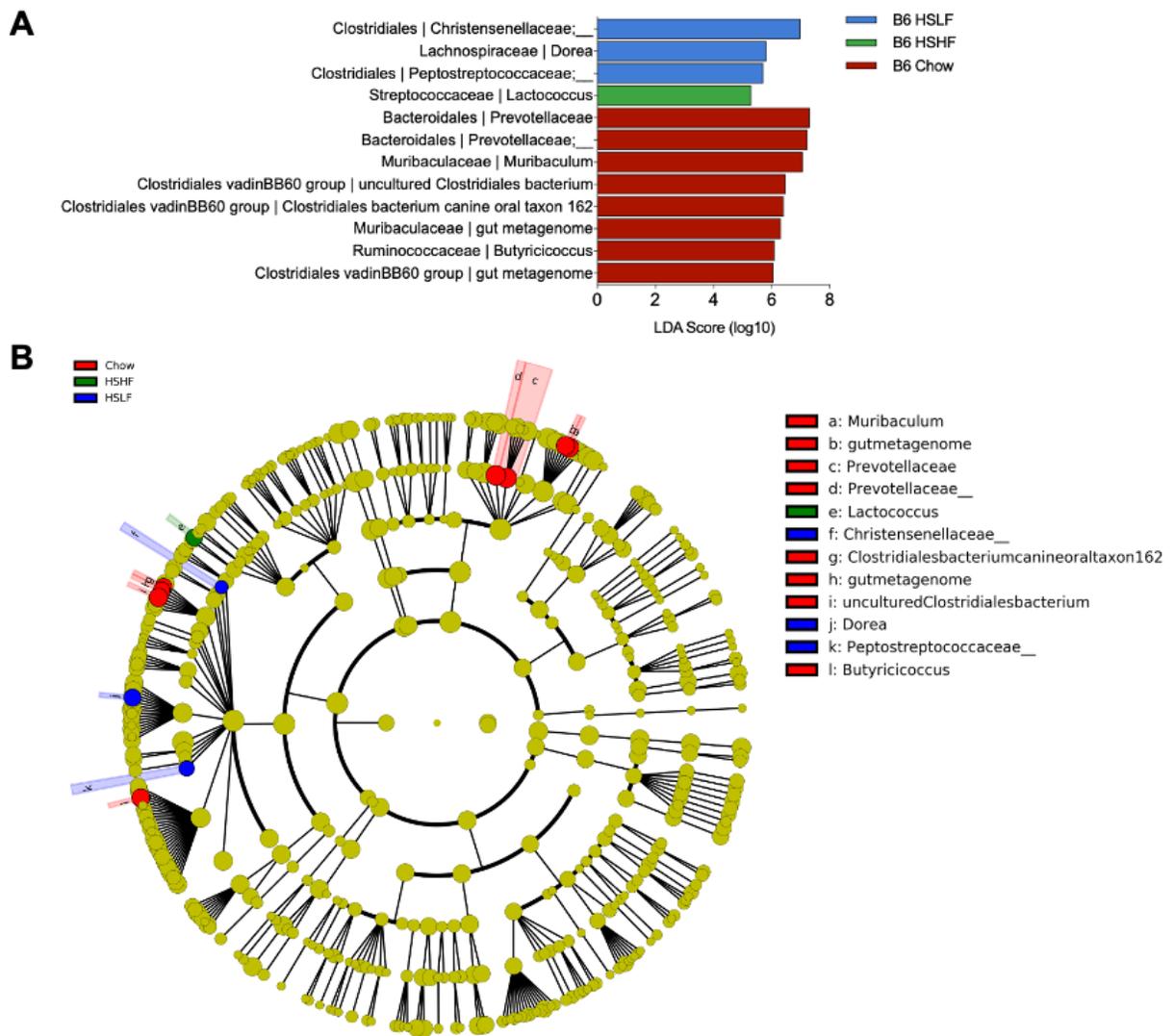


Figure 11. Linear discriminate analysis effect size (LEfSe) analysis of diet-driven microbiota shifts in B6 mice

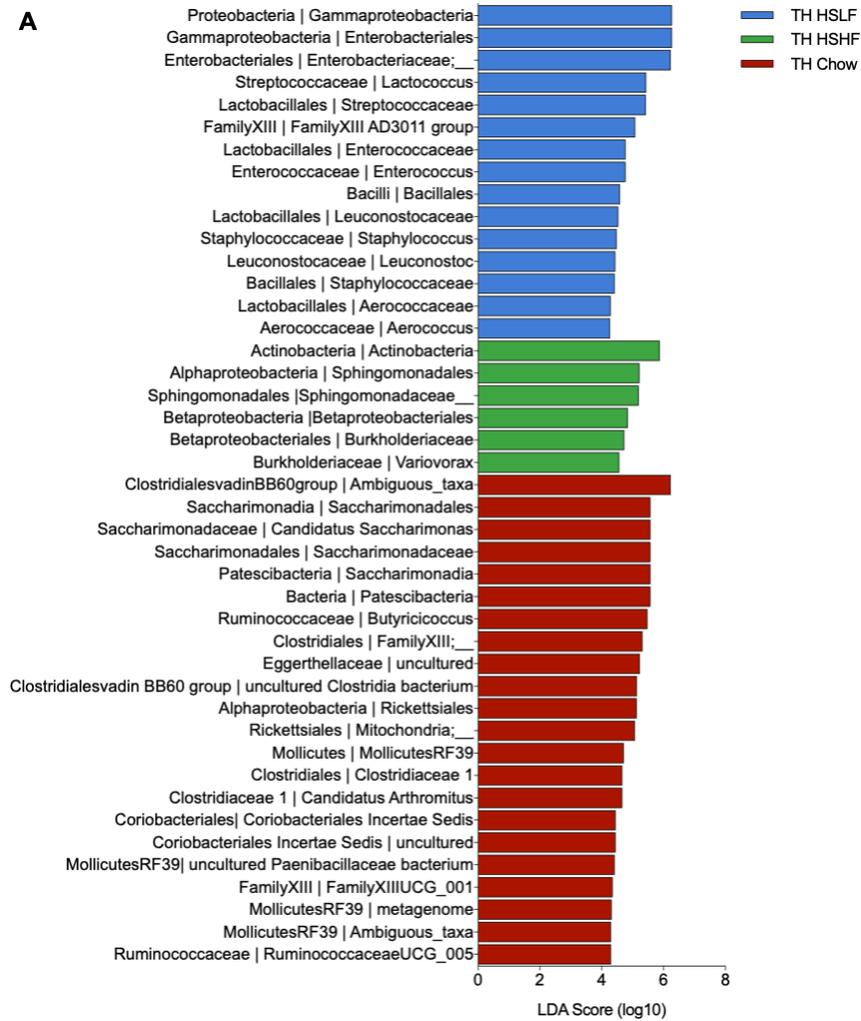
(A) Histogram and (B) cladogram of the linear discriminant analysis (LDA) scores (> 3) for the taxonomic features differentially abundant within B6 mice chow (red), HSHF (green), and HSLF (blue) diets.

Effect of diet on the microbiota of TH mice

A total of 43 discriminating features were found to vary in abundance between the diet groups in TH mice (Figure 12A). Clades of taxa corresponding to the Mollicutes RF39 order (Tenericutes phylum) and the Patescibacteria phylum were among the most enriched features of the TH chow class. Multiple members of the Clostridiales order, including those of the Family XIII group and Ruminococcaceae family, such as *Butyricoccus* and *Ruminococcaceae UCG-005*. While taxa of the Clostridia class were overall highly represented in chow-fed TH mice, multiple clades of taxa within the Bacilli class (Firmicutes phylum) were enriched in HSLF-fed TH mice. More specifically, *Staphylococcus* of the enriched Bacillales order as well as the *Lactococcus*, *Leuconostoc*, *Aerococcus*, and *Enterococcus* genera and their respective families of the Lactobacillales order. The Gammaproteobacteria class (Proteobacteria phylum) was the most overabundant clade within TH HSLF class, with an LDA score greater than 6.0. Within this clade and at finer taxonomic levels, Enterobacteriales and unclassified Enterobacteriaceae taxa were also found enriched. The Actinobacteria phylum, which was most represented mainly by Bifidobacteria in the gut microbiota (Binda et al., 2018), was enriched in the TH HSHF class.

Additionally, unclassified Sphingomonadaceae taxa of the augmented Sphingomonadales order and a clade of taxa corresponding to the Betaproteobacteriales order were also increased in HSHF-fed TH mice. The phylogenetic distribution of and hierarchical relationships between these features can be visualized as a cladogram (Figure 12B). These results suggest that the B6

microbiome is overall more homeostatic in response to dietary factors compared to the TH microbiome, which exhibited a higher degree of plasticity, as evidenced by the number (3.5X more) of differentially abundant taxa at the LDA threshold of 3.0.



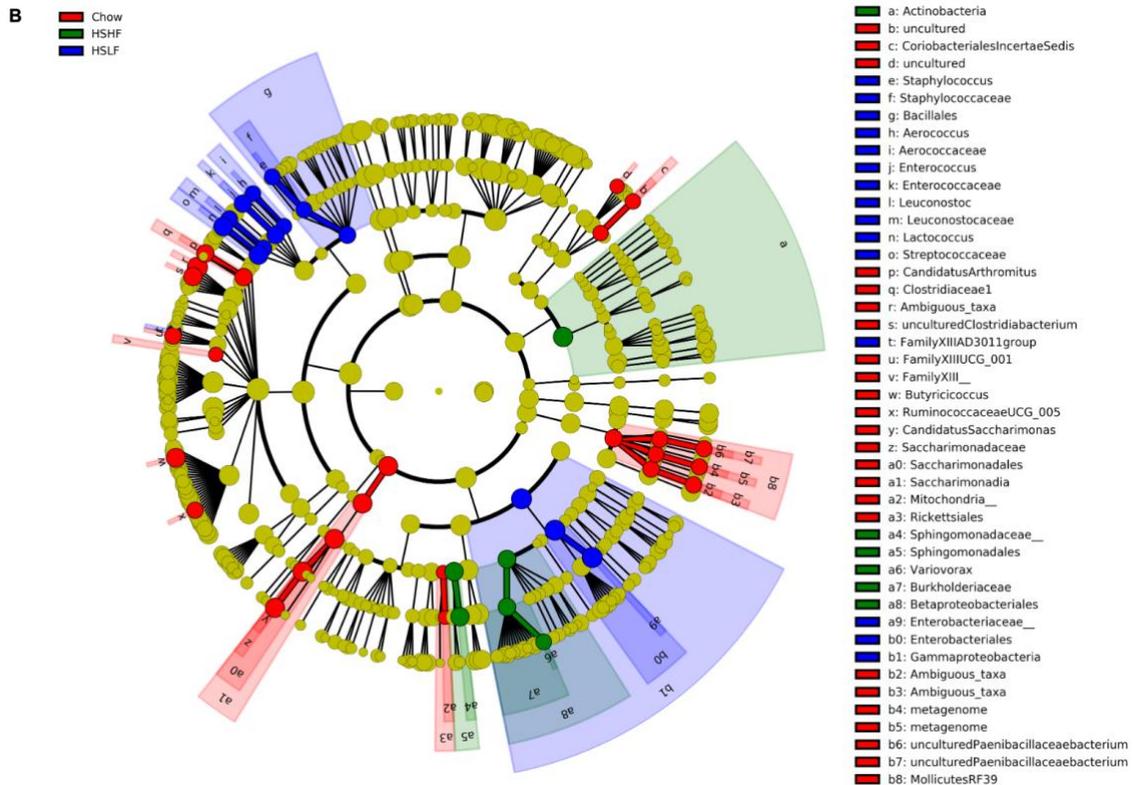


Figure 12. Linear discriminate analysis effect size (LfSe) analysis of diet-driven microbiota shifts in TH mice

(A) Histogram and (B) cladogram of the linear discriminant analysis (LDA) scores (> 3) for the taxonomic features differentially abundant within TH mice chow (red), HSHF (green), and HSLF (blue) diets.

Effect of strain on microbial community composition

A total of 17 discriminating features were found to vary in abundance between the diet-matched mouse strains (Figure 13A). Mycoplasmataceae taxa, including *Mycoplasma* specifically, within the Mycoplasmatales clade, were among the most abundant features in TH mice. *Escherichia-Shigella* of the Enterobacteriaceae family was also enriched in TH mice. In contrast, the gut microbiota of B6 mice was distinguished by the enrichment of Verrucomicrobia phylum and several taxa, including *Akkermansia*, within the clade. Genera belonging to the

Muribaculaceae, Ruminococcaceae, and Eggerthellaceae families were also discriminative of B6 mice. The phylogenetic distribution of and hierarchical relationships between these features can be visualized as a cladogram (Figure 13B). Considering the monophyletic patterns and the low number of abundant differentially taxa between diet-matched mouse strains, these results suggest that the microbiota in B6 and TH mice responded similarly to diet, or their responses did not differ at significant magnitudes.

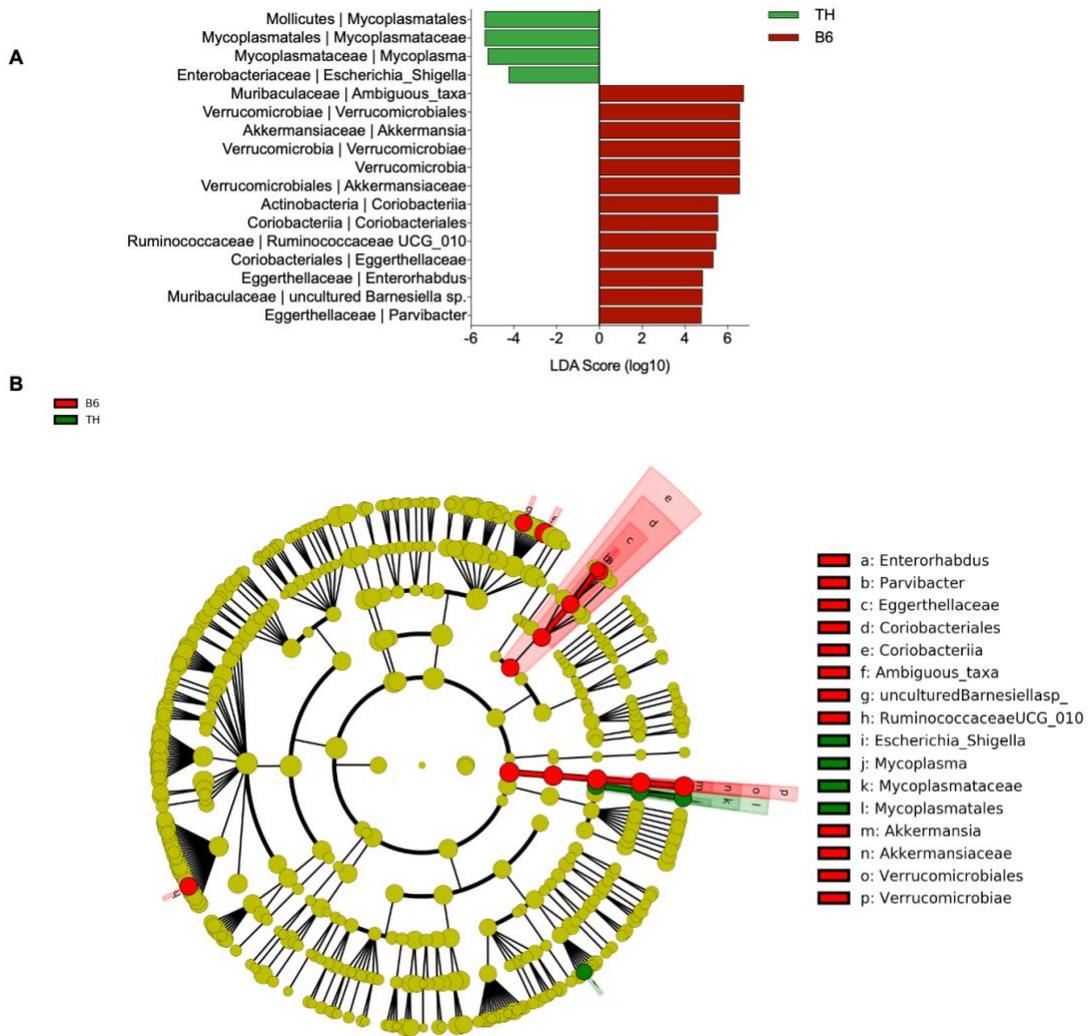


Figure 13. Linear discriminate analysis effect size (LEfSe) analysis of genetic-driven microbiota shifts in B6 and TH mice

(A) Histogram and (B) cladogram of the linear discriminant analysis (LDA) scores (> 3) for the taxonomic features differentially abundant between diet-matched B6 (red) and TH (green) mice.

Metabolic variables strongly associate with taxa clusters

We implemented hierarchical-all-against-all analysis (HALLA) (Rahnavard et al., 2019) to identify associations between the metabolic variables (X) and genus level taxa (Y). HALLA combines hierarchical nonparametric hypothesis testing with false discovery rate correction to enable high-sensitivity pattern discovery among datasets with heterogenous units of measurements (Rahnavard et al., 2019). HALLA testing proceeds by (i) discretizing variables within each dataset, e.g., metabolic variables (X) and genus level taxa (Y), to a uniform representation, (ii) hierarchically clustering each dataset separately to generate two data hierarchies, (iii) coupling clusters of equivalent resolution between the two data hierarchies, and (iv) iteratively testing coupled clusters of increasing resolution for statistically significant association (Rahnavard et al., 2019). In this study, the associations between clusters were measured for significance with the Spearman correlation similarity metric, with similarity scores ranging -1 (strong negative association) to 1 (strong positive association).

HALLA testing resulted in the identification of 66 significant associations between metabolic variables and taxa clusters (Appendix D) We graphed the top ten associations into an interaction network (Figure 14), in which clusters were color-coded, e.g., all genera and metabolic nodes belonging to a cluster were grouped by color. The direction (similarity score) and strength (Benjamini-Hochberg FDR correction-adjusted p-value) of the association between taxa and metabolic variables were illustrated by edge color and thickness, respectively.

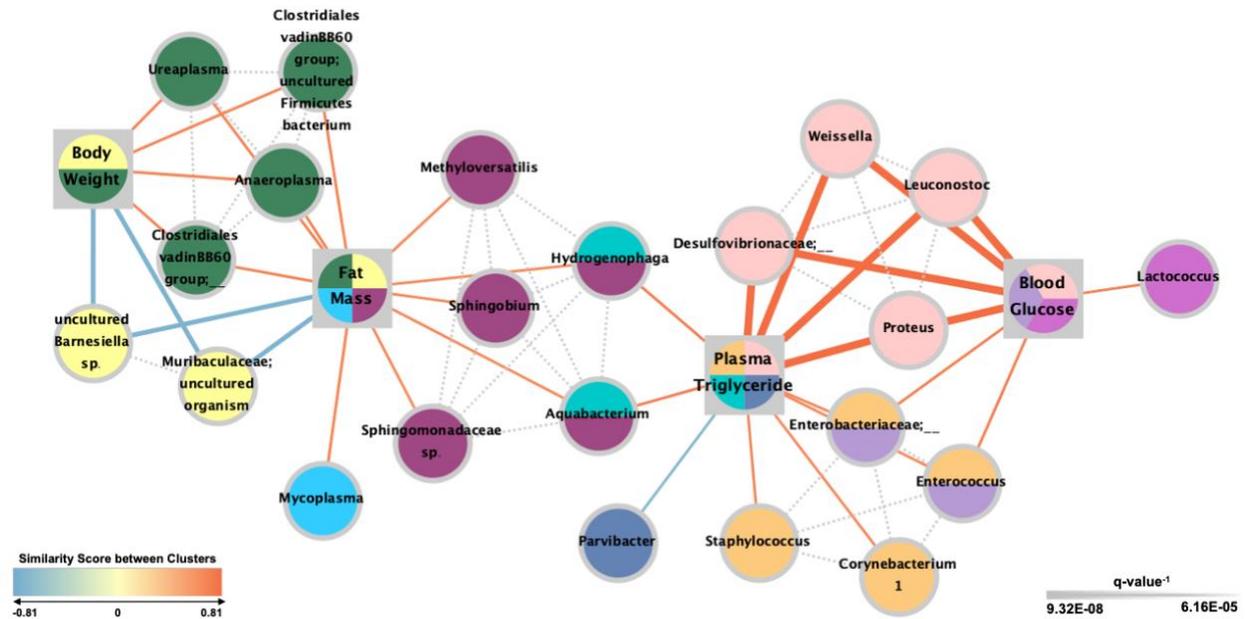


Figure 14. Associations between metabolic variables and taxa clusters

Network showing the top associations between microbial features and metabolic traits identified by hierarchical-all-against-all analysis. All genera and metabolic nodes belonging to a cluster were grouped by color and connected by dotted gray edges. Taxa and metabolic variables are connected by colored edges that illustrate the directionality (similarity score represented with heat gradient) and strength (q-value represented by edge width) of the association.

In general, statistical and network results showed plasma triglyceride and blood glucose variables and associated taxa clustering together, while body weight and fat mass variables and their associated taxa formed a separate cluster. The only exceptions to this were two genera, *Aquabacterium* and *Hydrogenophaga*, which were clustered and associated with plasma triglyceride ($q = 0.0030$; similarity = 0.72) (aqua nodes) as well as fat mass ($q = 0.0047$; similarity = 0.65) as part of a separate taxa cluster that also included *Sphingobium*, *Methyloversatilis*, and unclassified Sphingomonadaceae taxa (maroon nodes).

The strongest association was found between clustered plasma triglyceride and blood glucose variables and *Weissella*, *Leuconostoc*, *Proteus*, and unclassified Desulfovibrionaceae clustered taxa ($q = 0.00014$; similarity = 0.81) (pink nodes). Individually, plasma triglyceride was associated ($q = 0.0062$; similarity = 0.67) with clustered *Staphylococcus*, *Enterococcus*, *Corynebacterium 1*, and unclassified Enterobacteriaceae taxa (orange nodes). In a separate cluster, *Enterococcus* and unclassified Enterobacteriaceae taxa were also associated with blood glucose ($q = 0.0033$; similarity = 0.71) (purple nodes). *Parvibacter* was negatively associated with plasma triglycerides ($q = 0.0031$; similarity = -0.71) (lavender nodes). *Lactococcus* was associated with and blood glucose ($q = 0.0047$; similarity = 0.69) (violet nodes).

Strong negative and positive associations were found between clustered taxa and clustered body weight and fat mass variables. Uncultured *Barnesiella sp.* and Muribaculaceae (uncultured organism) negatively associated with this cluster ($q = 0.00033$; similarity = -0.74) (yellow nodes), while *Anaeroplasma*, *Ureaplasma*, an uncultured Firmicutes bacterium, and unclassified Clostridiales vadinBB60 group taxa were positively associated ($q = 0.0027$; similarity = 0.68) (green nodes). In an individual cluster, fat mass was found associated with *Mycoplasma* ($q = 0.0027$; similarity = 0.68) (blue nodes). Collectively, these associations provide critical insight into the patterns shared between metabolic health factors and their microbial interactors. For example, our results suggest that blood glucose and plasma triglyceride variables are closely connected and share several mutual associations with taxa. At the same time, fat mass is affiliated with body weight, which indicates that significant similarity was observed in the patterns shared between these metabolic variables and their associated microbial features.

Distinct signatures in the gut microbiota are predictive of the metS phenotype

Diversity and HALLA analyses suggested that metabolic traits contributed to variation in the gut microbial community. With this evidence, we clustered each mouse by their comprehensive metabolic profiles and sought to distinguish the microbial features that were predictive of the metS phenotype. To this end, mice were partitioned into two groups based on their comprehensive metabolic profiles, which we designated as non-metS and metS (Figure 15A). The resulting metS cluster was exclusive to HSHF- and HSLF-fed TH mice, while the non-metS cluster encompassed chow-fed TH mice as well as all diet groups of the B6 mice. The inter-cluster distance from the non-metS to the metS cluster was 3.495, and the overall mean silhouette score was 0.399.

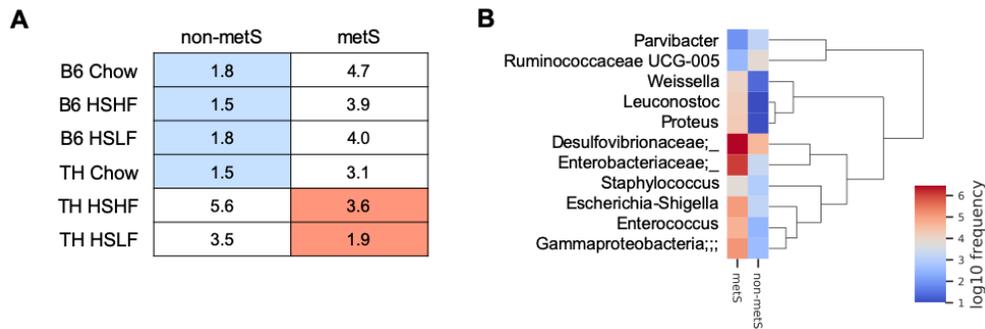


Figure 15. Microbial features predictive of metabolic phenotypes

Using machine learning methods, mice were clustered by their comprehensive metabolic profile, resulting in non-metabolic syndrome (non-metS) and metabolic syndrome (metS) clusters. (A) the distance of samples from non-metS and metS cluster centroids. Samples assigned to the non-metS cluster are colored in blue, while samples assigned to the metS cluster are colored in red. (B) Heatmap of the most predictive genera (feature importance threshold ≥ 0.01) of the non-metS and metS clusters, as determined by Random Forest classification.

A Random Forest classifier was used to identify the microbial features that displayed differential patterns between the non-metS and metS clusters. The most predictive genera (importance score > 0.01) of these clusters were oriented into a heatmap (Figure 15B; see Table 2 for all feature scores).

Feature ID	Importance Score
Enterococcus	0.245660716
Leuconostoc	0.234579092
Enterobacteriaceae;	0.157688912
Escherichia-Shigella	0.088358084
Weissella	0.070722389
Ruminococcaceae UCG-005	0.055046047
Desulfovibrionaceae;	0.048268048
Proteus	0.024102564
Parvibacter	0.020310449
Gammaproteobacteria; ; ;	0.01928438
Staphylococcus	0.012895207
uncultured Paenibacillaceae bacterium	0.009957999
Lactococcus	0.006435218
Eubacterium] nodatum group	0.004654499
Candidatus Saccharimonas	0.002036395

Table 2. Predictive genera of metS and non-metS clusters

While increased abundances of Parvibacter (0.02 importance score) and Ruminococcaceae UCG-005 (0.055) were associated with the non-metS cluster, enrichment of Staphylococcus (0.012), Proteus (0.024), Weissella (0.07), Escherichia-Shigella (0.088), Leuconostoc (0.23), and Enterococcus (0.24) genera and unclassified and Gammaproteobacteria (0.019), Desulfovibrionaceae (0.048) and Enterobacteriaceae (0.15) taxa were most highly associated and predictive of the metS cluster. We also performed this predictive analysis at the

ASV level (Figure 16A), in which multiple ASVs were found to be consistent with the predictive genera (see Appendix E for all ASVs and feature scores).

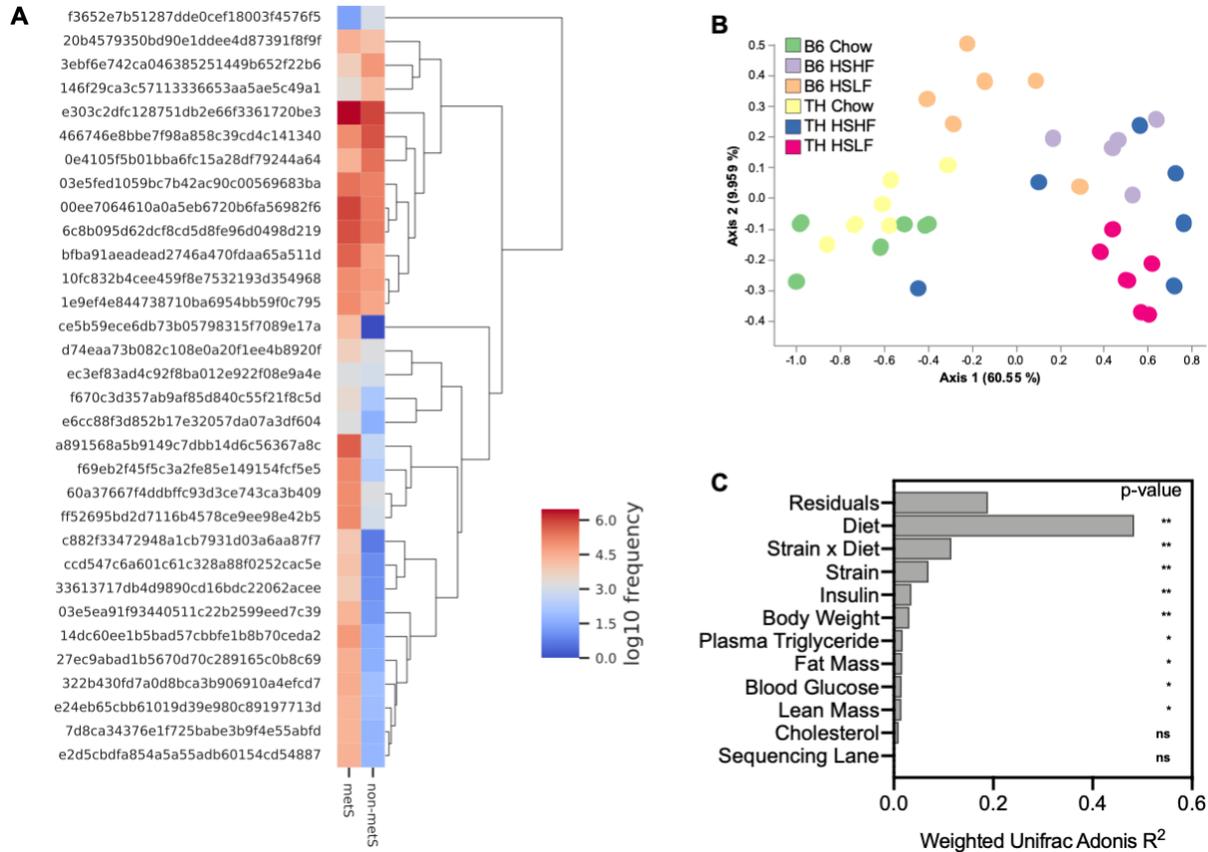


Figure 16. Dietary factors play a role in the microbial patterns that differentiate healthy from diseased hosts

(A) Heatmap of the most predictive (feature importance threshold > 0.01) amplicon sequence variants (ASVs) of the non-metS and metS clusters. (B) Principal coordinate analysis displaying phylogenetic dissimilarity (Weighted Unifrac distance) between microbiota samples based on the microbiota composition of predictive ASVs. (C) Bar plot showing the percentage of microbiota variation explained by host and environmental factors based on the microbiota composition of predictive ASVs.

Using the same methods and measurements of microbial diversity testing as before, we performed a sub-analysis wherein community dissimilarity was visualized, and explanatory variables were tested based on a microbiota composition comprised of only the predictive ASVs (Figures 16B-C). Results from ADONIS testing revealed that diet was the greatest explanatory factor (48.3%) of these select features. The other top explanatory variables included the interaction between strain and diet (11.5%), strain (6.95%), insulin (3.5%), and body weight (3.1%), and plasma triglyceride (1.7%). Illustrative summaries of the predictive models (Figure 17) indicate a high degree of predictive accuracy (N. A. Bokulich et al., 2018).

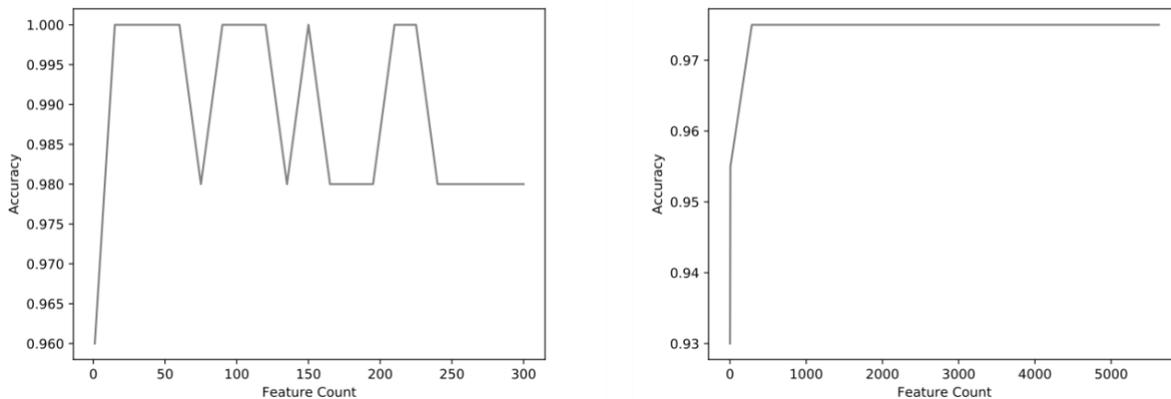


Figure 17. Accuracy of Random Forest classifier predictions of non-metS and non-metS features

Recursive feature elimination plots illustrating changes in model accuracy as a function of (A) genus and (B) ASV count.

Consistent with previous ADONIS results, which showed dietary factors had an overwhelming effect on the gut microbiota’s variation as a whole, these results suggest that diet also played a significant role in driving the population of microbial features that differentiate healthy from diseased hosts.

Microbiota function is influenced by diet

Phylogenetic investigation of communities by reconstruction of unobserved states 2 (PICRUSt2) (Douglas et al., 2019; Langille et al., 2013) was used to infer the approximate function of the microbial community. Unlike its predecessor, PICRUSt2 analyzes each ASV to make metagenomic functional predictions (Douglas et al., 2019). This updated method incorporates the alignment and phylogenetic placement of ASVs against the Integrated Microbial Genomes database and hidden state prediction algorithms to determine the gene (KEGG ortholog) (Kanehisa & Goto, 2000) content of each ASV (Douglas et al., 2019).

We evaluated the functional profile of the microbial community using the beta-diversity techniques as previously performed with the compositional profile. We visualized samples by their functional profile using the Bray-Curtis distance (Figure 18A) and implemented ADONIS to investigate whether variation in the functional profile of the cecal microbiota could be explained by host and environmental factors (Figure 18B).

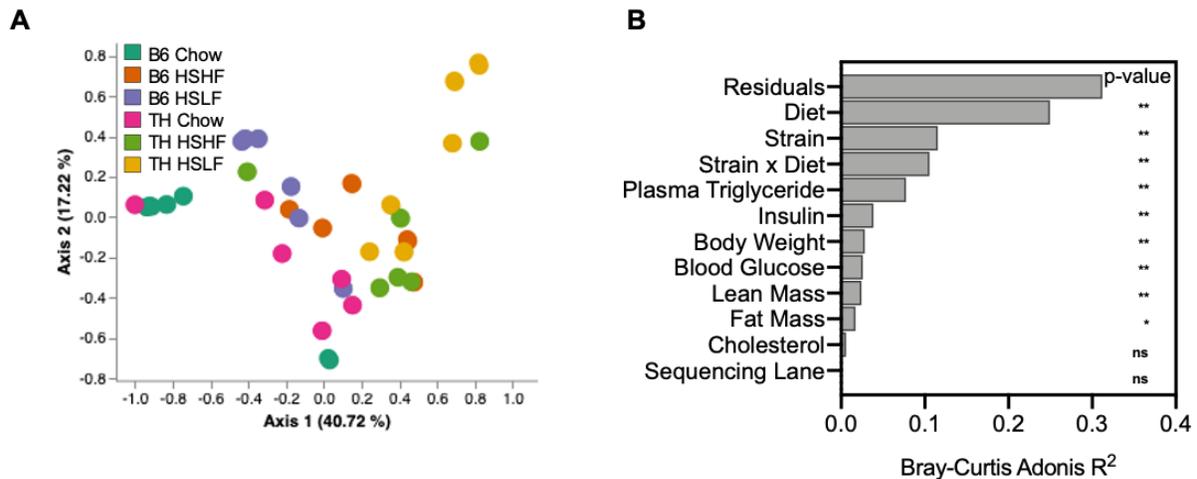


Figure 18. Analysis of community function and explanatory variables

Principle coordinate analysis displaying functional dissimilarity (Bray-Curtis distance) between microbiota samples of different mouse strains and diets. (B) Bar plot showing the percentage of functional variation explained by host and environmental factors.

Except for the sequencing lane and cholesterol, all other variables and interactions (strain x diet) were significantly ($p < 0.05$) associated with gut microbial variation. Among all the covariates tested, nearly one-quarter of the variation in the cecal microbiota was explained by diet (24.9%) Following diet, strain (11.5%), the interaction between strain and diet (10.5%), plasma triglyceride (7.7%) and insulin (3.8%) were among the top explanatory variables of community function. In the future, we would like to assess microbial function through shotgun metagenomics; however, these predictive results support the role of dietary factors as a significant determinant of community function.

Discussion

Gene-diet interactions have been shown to regulate susceptibility to obesity, T2D, and metS in humans and mice (Bouchard et al., 1990; Parkman et al., 2016; Parks et al., 2013; Ussar et al., 2015). These interactions also affect the composition of the gut microbiome and the metabolites it produces, which are contributing factors to metabolic disorders (Fujisaka et al., 2018; Kreznar et al., 2017; Ussar et al., 2015). For these reasons, there is an increasing demand for new animal models that reflect the genetic variance observed in human diseases, particularly when investigating the microbial features that may influence these pathologies (Fuchs et al., 2018). Herein, we establish the TH mouse as a clinically-relevant polygenic model for studying the relationship between the gut microbiome and metS. The results of this study demonstrate that dietary factors (i) play a central role in shaping the ecology of the gut microbiota in TH and B6 mice, (ii) control the microbial patterns that emerge in metabolically healthy or diseased hosts, and (iii) influence the expansion of enteric pathobionts and metS. Collectively, our data suggest host genetics regulate the susceptibility of microbiota perturbations and disease states that are mediated by diet.

Identifying the microbial features implicated in obesity and metS is a critical step in understanding the microbial-host interactions that underlie these pathologies. To this end, we implemented machine learning methods to cluster mice by their comprehensive metabolic profile and predicted the microbial features associated with healthy or compromised metabolic phenotypes. All HSHF- and HSLF-fed TH mice were assigned to the metS cluster, while all B6 mice and chow-fed TH mice were assigned to the non-metS cluster. The latter observation indicates that despite the genetic predispositions harbored by the TH mouse, the manifestation of metS could be controlled by diet in TH mice. Additionally, the accuracy at which microbial features were predictive of the metabolic clusters suggests a strong relationship between the composition of the gut microbiota and host phenotype.

The presence of Proteobacteria in the human gut is a known indicator of microbial perturbations (Morgan et al., 2012) and disease states, including chronic enteropathies (Knights et al., 2014), nonalcoholic steatohepatitis (L. Zhu et al., 2013), obesity (Peters et al., 2018), and T2D (Larsen et al., 2010). In this study, we found that overgrowth of Desulfovibrionaceae and Enterobacteriaceae taxa, including *Proteus* and *Escherichia-Shigella*, to be a defining feature in the microbiota of samples with metS. Prior studies have shown that Enterobacteriaceae exhibit increased fitness under inflammatory conditions compared to resident commensals, thereby rendering these microbes effective inducers and perpetuators of intestinal inflammation, colitis, and metabolic endotoxemia (Forbes, Van Domselaar, & Bernstein, 2016; Hughes et al., 2017; Kitamoto et al., 2020; Knights et al., 2014; Morgan et al., 2012; Shin, Whon, & Bae, 2015; Winter & Baumber, 2014; Zeng, Inohara, & Nunez, 2017).

Lactic acid bacteria (LAB), *Enterococcus*, *Staphylococcus*, *Weissella*, and *Leuconostoc* were also among the strongest predictors of metS. Members of the LAB group are becoming

increasingly implicated in human diseases, including obesity (Peters et al., 2018) and graft-versus-host-disease, particularly in lactose malabsorbers (Stein-Thoeringer et al., 2019). While the byproduct of LAB fermentation is often used by other organisms to produce short-chain fatty acids (SCFAs) (N. Hwang et al., 2017), lactate is also the preferred carbon source of sulfate-reducing bacteria (SRB), such as *Desulfovibrio* which use lactate to reduce sulfate (Vita et al., 2015). *Desulfovibrio* species have been found enriched in humans (Bai et al., 2017) and mice (Kreznar et al., 2017) with metabolic diseases. Furthermore, the hydrogen sulfide gas produced by *Desulfovibrio* as a metabolic byproduct of sulfate reduction is cytotoxic to colonocytes and thought to promote intestinal inflammation and cancer (Linden, 2014; J. Qin et al., 2012). We observed the concurrent enrichment of Desulfovibrionaceae and LAB in the metS cluster, which supports the notion that *Desulfovibrio* is proficient cross-feeders of lactate (Vita et al., 2015). These findings are strengthened by the data indicating that SCFA-producing microbes, e.g., Clostridium cluster XIVa and Bacteroidetes taxa (Parada Venegas et al., 2019; Poeker et al., 2018; B. J. Smith et al., 2019), were noticeably absent in mice with metS, which may have contributed to the accumulation of lactate. Collectively, our findings suggest that maintaining a balanced ratio of lactate-producing to lactate-utilizing bacteria may be critical to gut homeostasis.

Our HAIIA analysis further validated the identity of the metS-associated taxa and provides context for how these organisms may mutually interact to influence metabolic function. For example, we found that blood glucose and plasma triglyceride variables shared several associations, including *Enterococcus*, *Weissella*, *Leuconostoc*, *Proteus*, *Desulfovibrionaceae*, and *Enterobacteriaceae* taxa. The gut microbiome is a well-established determinant of lipid metabolism and circulating triglyceride concentrations (Backhed et al., 2004; Just et al., 2018;

Kuno, Hirayama-Kurogi, Ito, & Ohtsuki, 2018; Matey-Hernandez et al., 2018). Consistent with our findings, Desulfovibrionaceae has been correlated with increased concentrations of circulating and hepatic triglycerides in mice (Just et al., 2018; Kreznar et al., 2017). The involvement of Desulfovibrionaceae in dyslipidemia may occur through the organisms' transformation of bile acids (BAs) into secondary BAs, which are ligands to several glucose and lipid metabolism-regulating nuclear receptors (Hylemon, Harris, & Ridlon, 2018; Kuno et al., 2018). Moreover, under conditions of increased BA secretion, bile-resistant bacteria such as Enterobacteriaceae have a selective advantage (Islam et al., 2011). This may explain the concurrent enrichment of these taxa, as well as the previously reported link between serum concentrations of triglyceride and the endotoxin, lipopolysaccharide (LPS), produced by Gram-negative bacteria (Lassenius et al., 2011). On the other hand, Clostridiales vadin BB60 group and Mollicutes taxa were highly connected and mutually associated with increased fat mass and body weight, while Muribaculaceae taxa were associated with a reduction in these metabolic variables. Our results are in accordance with previous studies, wherein Firmicutes- and Bacteroidetes-rich microbiota have been associated with obesity and leanness, respectively (Koliada et al., 2017; Ley, Turnbaugh, et al., 2006; Turnbaugh et al., 2006).

Akkermansia has been reported as a highly heritable gut microbe, and particularly abundant in lean and metabolically healthy humans and animals (Org et al., 2015; Zhao et al., 2017). While the Verrucomicrobia clade was enriched in B6 mice, we failed to find any relationship between this clade of organisms and metabolic variables or clusters. Instead, Ruminococcaceae and Muribaculaceae taxa, as well as *Parvibacter* (Eggerthellaceae) were predictors of healthier mice, with the latter associated with decreased plasma triglyceride levels. Collectively, these taxa are likely to play a symbiotic role in the gut, where they metabolically

transform dietary nutrients to bioreactive compounds that are critical to host health. For instance, Muribaculaceae and Ruminococcaceae produce short-chain fatty acids (SCFAs) as a byproduct of polysaccharide fermentation (Poeker et al., 2018; B. J. Smith et al., 2019), while Eggerthellaceae convert polyphenolic compounds, e.g., ellagitannins and ellagic acid present in nuts and fruit, to highly bioavailable metabolites such as urolithins, which act as potent anti-inflammatory and anti-oxidant agents (Espin, Larrosa, Garcia-Conesa, & Tomas-Barberan, 2013; Selma et al., 2017). Our results from HALLA further support the beneficial role of these organisms and highlight their functional relationship to metabolic health.

Dietary factors had an overwhelming effect on gut microbial function and variation as a whole, as well as the specific microbial features predictive of the metS and non-metS clusters, explaining 48% of the variation observed in this subset of ASVs. These results highlight a strong relationship between diet and the microbial patterns that differentiate metabolically healthy from diseased hosts. This investigation did not address nor is there a census on the causality of microbiota perturbations and metabolic disorders; however, antibiotic-induced gut microbiota disturbances have been causally linked to the development of obesity in animals, which lends credence to the role of the microbiome in the causation or prevention of metabolic diseases (Leong, Derraik, Hofman, & Cutfield, 2018). In this context, it is reasonable to surmise that dietary changes elicit selective pressures on the gut microbiome in a similar fashion.

Indeed, a large body of evidence has demonstrated that the gut microbiome responds dynamically, reproducibly, and rapidly to dietary shifts, (Carmody et al., 2015; David et al., 2014; Turnbaugh et al., 2008; Turnbaugh, Ridaura, et al., 2009; Zhang et al., 2012), thereby further supporting the hypothesis that environmentally-induced microbiota perturbations may be the triggering event. Components of the HSHF and HSLF diets used in this study are

representative of the Western diet, which is characterized by the increased consumption of saturated and *trans* fats and simple carbohydrates, such as refined grains and sugars (Cordain, Eaton, Sebastian, Mann, Lindeberg, Watkins, O’Keefe, et al., 2005). Although the ratio of Firmicutes to Bacteroidetes was significantly increased in HSHF-fed mice, we conclude that the HSLF diet was most detrimental to the gut microbial community, as evidenced by the significantly decreased biodiversity and expansion of well-defined pathobionts observed in HSLF-fed mice. The disparities in microbiota composition between the HSHF and HSLF diets were somewhat surprising considering both diets are composed of the same ingredients, but with differing ratios of carbohydrate and fat sources. Collectively, these results indicate that the composition of diet, as well as the relative proportion at which these macronutrients are consumed, play a central role in shaping the overall structure of the gut microbiota.

When we accounted for the common effect of diet, we found few differentially abundant taxa between mouse strains, which suggests the microbiota responded similarly to diet. Nevertheless, strain-by-diet analysis with LEfSe showed that the TH microbiota exhibited an enhanced capacity for diet-mediated modulation, thereby intimating that genetic factors played a role. While it is possible that the B6 microbiota exhibited a perturbation of similar magnitude that was not detectable by the endpoint of this study, the specific enrichment of enteric pathobionts in the metS cluster suggests TH mice are less effective at controlling the microbiota population and potentially eradicating pathogens. This observation may be a consequence of impaired immunity, which is a common yet poorly understood complication of metabolic diseases (Casqueiro, Casqueiro, & Alves, 2012). Previously, Denvir et al. (Denvir et al., 2016) identified more than 300 variants in genes orthologous to human genes of obesity, diabetes, and metabolic-related traits in the TH genome. Upon manually reviewing the catalog of variants

specific to the TH genome (Denvir et al., 2016), we identified multiple deleterious (SIFT score < 0.05, or PROVEAN score < -2.5) (Denvir et al., 2016) single nucleotide polymorphisms and indels in genes essential to antigen processing and immune activation, including interleukin (IL)-2 (*Il2*), IL-20 receptor alpha (*Il20ra*), lipopolysaccharide recognition receptor (*Lrba*), and Major Histocompatibility Complex class I and II (*H2-Dma*, *H2-Q1*, *Lrmp*) (see Appendix F for complete list). These results warrant future quantitative genetic studies to further explore the involvement of these variants on immunity, gut microbiome variation, and metS development.

One of the constraints of this and other 16S rRNA-based studies is the limited resolution at lower taxonomic levels, where the symbiotic or pathogenic properties of the microbiota are likely determined at a species or subspecies level (Faith, Colomel, & Gordon, 2015). However, by integrating 16S rRNA and metabolic sample data, we computationally inferred relationships between metabolic and microbial patterns at the molecular level of ASVs. Our intent in performing these analyses with ASVs was to capture the true species diversity of our target community since microbial diversity has been previously linked to disease states (Clemente et al., 2012; Haegeman et al., 2013; Le Chatelier et al., 2013). Moreover, the important ASVs identified in this study can be compared to ASVs inferred independently from different studies and assessed for similar metabolic relationships, including in the human microbiome (Callahan et al., 2017).

In conclusion, our results yield significant insight into the interrelationships among the gut microbiota, dietary macronutrients, and the genetic permissiveness that underlie metS. Our findings show that following the consumption of Western-like diets, the TH microbiome was enriched with enteric pathobionts and that the overgrowth of these pathogens may be the consequence of impaired immune responses in the TH mice. Although the development of metS

and dysbiosis was TH strain-dependent, macronutrients and the relative proportion at which they were consumed ultimately determined the severity of these phenotypes. The metS-associated taxa identified in this study may be critical microbial biomarkers for metabolic dysfunction, particularly dyslipidemia, in humans and future targets for therapeutic countermeasures. Our results suggest the ratio of lactate-producing to lactate-utilizing bacteria, e.g., SCFA producers, is an essential aspect of maintaining a diverse and homeostatic gut ecosystem. With the knowledge imparted here, as well as the publicly available genome and variant catalog, the TH mouse represents a valuable resource for understanding the environmental and genetic underpinnings of human metS, as well as the role of the “hologenome”—the interaction of the host and microbial genome.

Materials and Methods

Animal model

All animal studies were performed under an approved protocol by the Marshall University Institutional Animal Care and Use Committee. TH mice used in this study originated from a breeding colony that has been maintained by J.H.K since 2001 (J. H. Kim & Saxton, 2012). B6 mice were obtained from the Jackson Laboratory (Bar Harbor, ME, USA) and were bred in-house for up to 10 generations. Colonies of B6 and TH mice were established and maintained in the Marshall University Animal Resource Facility (ARF). To optimize the reproducibility of our microbiota analysis, all mice were bred and housed in a common environment within the ARF (Laukens, Brinkman, Raes, De Vos, & Vandenabeele, 2016).

At 3-4 weeks of age, separate litters of male B6 and TH mice were divided over several cages (approximately 2-4 mice per cage) to eliminate the possible bias of a cage effect (see Appendix G for housing details and group sample sizes). Moreover, to control for the effect of

maternal transmission (Laukens et al., 2016), mice originating from different litters were used to achieve the sample size of 5-6 mice per group. Each cage was assigned one of the following diets: standard rodent chow (PMI Nutrition, Purina 5001), semi-purified high-sucrose high-fat diet (HSHF) (Research Diets, D12266B), or semi-purified high-sucrose low-fat diet (HSLF) (Research Diets, D12489B) (see Table 3 for nutritional details of each diet). Mice were provided with food and water *ad libitum* in a temperature and humidity-controlled environment, with a standard 12-hour light/dark cycle.

Diet Components	Chow		High-sugar low-fat (HSLF)		High-sugar high-fat (HSHF)	
kcal/g	3.36		3.90		4.41	
	% of T.W.	kcal%	% of T.W.	kcal%	% of T.W.	kcal%
Protein	23.9	28.5	16.4	16.8	18.5	16.8
Carbohydrate	48.7	58.0	70.8	72.6	56.7	51.4
Fat	5.0	13.5	4.6	10.6	15.6	31.8
Total		100		100		100
Other ingredients						
Sucrose	3.7		24.6		27.8	
Starch	31.9		42.3		20.6	
Crude Fiber	5.1		0.0		0.0	
Cellulose	0.0		2.5		2.8	

Table 3. Summary of diet composition

The table was adapted from Parkman et al. (Parkman et al., 2016). Total weight (T.W.).

Metabolic traits

The metabolic measurements used in this study were derived from a larger dataset from which Parkman et al. previously reported the results, analysis, and interpretation of the metabolic

phenotypes (Parkman et al., 2016). Table 4 summarizes the analysis of the metabolic traits of animals included in this study.

Metabolic Trait	B6 Chow	B6 HSHF	B6 HSLF	TH Chow	TH HSHF	TH HSLF
Body Weight (g)	29.11 ±3.84 ^a	32.09 ±0.86 ^a	34.89 ±5.82 ^{ab}	35.89 ±2.79 ^{ab}	43.12 ±8.01 ^b	35.11 ±4.83 ^{ab}
Plasma Triglyceride (mg/dl)	94.93 ±41.47 ^a	123.0 ±110.8 ^a	46.84 ±28.72 ^a	211.7 ±66.72 ^a	891.5 ±515.1 ^b	826.8 ±374.5 ^b
Blood Glucose (mg/dl)	136.5 ±23.44 ^a	153.6 ±30.37 ^a	105.8 ±26.53 ^a	122.5 ±20.41 ^a	460.7 ±198.8 ^b	496.0 ±193.3 ^b
Insulin (ng/ml)	3.373 ±0.46 ^a	1.394 ±0.74 ^a	1.202 ±1.25 ^a	2.422 ±2.05 ^a	6.823 ±10.72 ^a	0.9540 ±0.41 ^a
Plasma Total Cholesterol (mg/dl)	42.95 ±11.98 ^e	115.1 ±31.28 ^{bd}	81.73 ±51.34 ^{cde}	101.2 ±7.63 ^{bc}	182.3 ±20.64 ^a	147.2 ±10.22 ^{ab}
Fat Mass (g)	1.406 ±0.64 ^a	1.852 ±0.6395 ^a	2.572 ±2.45 ^a	5.740 ±1.40 ^{ac}	10.95 ±3.65 ^b	7.657 ±3.79 ^{bc}
Lean Mass (g)	24.79 ±1.95 ^{ab}	24.20 ±3.024 ^a	27.12 ±1.83 ^{ab}	27.80 ±1.37 ^b	26.27 ±2.27 ^{ab}	27.65 ±0.38 ^{ab}

Table 4. Summary of metabolic traits

The table was adapted from Parkman et al. [31]. Differences among groups were assessed with a one-way ANOVA test followed by Tukey's multiple comparisons test. Data are representative of mean ± SD. Group means labeled with different letters (in superscript) had an adjusted p-value < 0.05.

Methods for collecting and quantifying the metabolic traits were also described by Paet al. et al (Parkman et al., 2016). Briefly, between the ages of 14-20 weeks, the body composition of each mouse was assessed using the quantitative magnetic resonance imager, EchoMRI-100 (Echo Medical Systems, Houston, TX, USA). Median fat mass and lean mass values were representative of five repeated measurements for each animal, as recommended by the manufacturer. Between the ages of 14-28 weeks, blood samples were collected from the mice in the morning via submandibular bleeding. Non-fasting blood glucose measurements were

obtained on the One Touch Ultra2 Blood Glucose Monitoring System (Diagnostics Direct, Cape May Court, NJ, USA). Plasma was subsequently obtained from the blood samples mentioned above samples via centrifugation (1200g) at 4°C. Plasma levels of total cholesterol (Thermo Electron, Louisville, CO, USA) and free and total glycerol (Sigma, St. Louis, MO, USA) were measured using colorimetric assays. True plasma triglyceride concentrations were estimated by the subtraction of free glycerol from total glycerol. Plasma insulin levels were determined with a mouse insulin and ultrasensitive ELISA kit (Crystal Chem, Downers Grove, IL, USA).

Metagenomic analyses

16S rRNA sequencing library preparation

Cecal contents were collected from mice between the ages of 18-32 weeks following sacrifice. Bacterial genomic DNA was extracted from the cecal samples via the DNeasy PowerSoil Kit (Qiagen, 12888-100) using the manufacturer's centrifugation protocol. DNA purity and concentration were determined using Qubit Fluorometry (Thermo Fisher Scientific, Waltham, MA, USA). Polyacrylamide gel electrophoresis purified indexed primers (341F/518R) (Bartram et al., 2011) targeting the variable 3 (V3) region of the bacterial 16S rRNA gene were obtained from Integrated DNA Technologies. AccuPrime Taq DNA Polymerase (Thermo Fisher Scientific, 12346086) was used to amplify the V3 region, yielding an amplicon of approximately 320 base pairs (bp). Each PCR reaction contained a total of 60 ng genomic DNA, 0.5 µl of 20 µM forward primer, 0.5 µl of 20 µM reverse primer, 1 µl AccuPrime PCR Master Mix with Taq Polymerase, 5 µl 10X AccuPrime Buffer II, and PCR-grade water for a total volume of 50 µl. PCR reactions were amplified on the MJ Research PTC-200 Peltier Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA) using the following conditions as previously described by Cockburn et al. (Cockburn et al., 2012): 6-minute denaturation step at 95°C; 30 cycles of 95°C

for 2 minutes, 50°C for 2 minutes, and 72°C for 2 minutes; 4-minute extension step at 72°C. The size and quality of the V3 libraries were assessed by electrophoresis on a 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA, USA) and found to be 325-340 bp. The pooled libraries were sequenced over two lanes in a 2x180 paired-end fashion in the rapid run mode on an Illumina HiSeq1500 with the addition of 15% PhiX to enhance sequencing quality. The average percentage of reads (forward and reverse) with Q30 score or better was 95.06.

Denoising, phylogeny building, and diversity analyses

Demultiplexed, paired-end sequences were imported into QIIME 2 (Bolyen et al., 2019; Caporaso et al., 2010) via the Casava 1.8 paired-end demultiplexed fastq format. Sequences were denoised using the dada2 package (Callahan et al., 2016). A phylogenetic tree of sub-operational taxonomic units was constructed using the SATé-enabled phylogenetic placement (SEPP) fragment-insertion method (Janssen et al., 2018). Prior to diversity analyses, data were normalized by rarifying to a sampling depth of 850,000. Rarifying resulted in the loss of two technical replicates originating from the same sample. Diversity differences within samples (alpha diversity) were analyzed with the nonparametric Kruskal-Wallis test using Shannon's diversity metric (C. E. Shannon & Weaver, 1949). Diversity differences between samples (beta diversity) were analyzed with various iterations of Permutation Multivariate Analysis of Variance (PERMANOVA) (M. J. Anderson), including group and pairwise significance testing, PERMDISP, and ADONIS (Marti J. Anderson, 2017). As standard with pairwise testing in QIIME 2, the Benjamini-Hochberg FDR correction was used to adjust the p-values to account for multiple comparisons. Testing with PERMANOVA was implemented using the following beta diversity metrics: Weighted Unifrac Distance (Lozupone et al., 2007) and Bray-Curtis

Distance (Bray & Curtis, 1957). Principal-coordinates analyses (PCoAs) were visualized with Emperor (Vazquez-Baeza et al., 2013).

Taxonomic analysis

For microbial taxonomic analysis, ASVs were assigned taxonomy using the QIIME 2 feature-classifier plugin (N. A. Bokulich et al., 2018), which was trained on SILVA 132 reference sequences and taxonomy (Quast et al., 2013) at the sequence identity threshold of 99%. ASVs assigned to Cyanobacteria, and Chlorioflexi phyla were removed through filtering. It is common practice to remove these chloroplast-containing bacteria, as their signals have been found to originate from ingested food containing plant polysaccharides, such as the chow diet (Sheik et al., 2018).

Linear discriminant analysis (LDA) effect size (LEfSe) (Segata et al., 2011) (<https://huttenhower.sph.harvard.edu/galaxy/>) was used to identify taxonomic features (phylum through genus level) that explain the difference between strain and diet groups. The following parameters were used for all LEfSe analyses: alpha value for class and subclass testing ≤ 0.05 ; threshold on the logarithmic LDA score for discriminative features ≥ 3.0 ; all-against-all multi-class analysis strategy. Using LEfSe, the effect of diet on the microbiota composition was analyzed within each mouse strain, and the effect of host genetics on the microbiota composition was analyzed by performing pairwise comparisons between mouse strains on the same diet. Thus, the final outputs only contained microbial features that discriminated between diet-matched B6 and TH mice.

Metabolic variable-microbiota association network

Hierarchical All-against-All significance testing (HALLA, v.0.8.17) (<https://huttenhower.sph.harvard.edu/halla>) was employed to identify and quantify relationships

among microbial genera and the continuous metabolic variables using the Spearman correlation similarity metric and default program settings (Rahnavard et al., 2019). The Benjamini-Hochberg FDR correction was used to adjust the p-values to account for multiple comparisons. The most highly associated metabolic variables and taxa clusters were mapped as an interaction network using Cytoscape (P. Shannon et al., 2003).

Clustering of samples with comprehensive metabolic profile

In addition to analyzing each metabolic variable as an associated variable of the microbiota, we aimed to cluster each mouse based on its comprehensive metabolic profile. Because the clinical diagnosis of metS is based on multiple metabolic criteria, this model may provide a global perspective on the microbial features associated with cooccurring metabolic conditions. The following cluster analysis consisted of the 29 mice (4 chow-fed B6 mice; 5 HSHF-fed B6 mice; 6 HSLF-fed B6 mice; 6 chow-fed TH mice; 3 HSHF-fed TH mice; 5 HSLF-fed TH mice) for which metabolic profiles were available. Values for each of the metabolic traits (see Table 4) were standardized by subtracting the mean and dividing by the standard deviation (within each metabolic trait) to remove the multiscalar effect. In order to classify each mouse with respect to metabolic syndrome in an unbiased fashion, the K-means clustering method (MacQueen, 1967) was applied to the standardized metabolic variables using seed randomization over 1000 random runs as implemented by Ginkgo software (Bouxin, 2005; De Caceres, Oliva, Font, & Vives, 2007). The stability of the clustering model was validated by the leave-one-out cross-validation approach (Hastie, Tibshirani, & Friedman, 2001). The number of partitions (k) was chosen *a priori*, with k=2 resulting in two distinct, clustered phenotypes that we designated as non-metabolic syndrome (non-metS) and metabolic syndrome (metS). The choice of k was

evaluated by the non-parametric Silhouette approach (Rousseeuw, 1987), which provides a quantitative measurement of how well each sample clusters within its given partition.

Feature prediction with Random Forest classification

The QIIME 2 sample-classifier plugin was used to identify specific taxa that were associated with the non-metS and metS clusters (as defined above) (N. Bokulich et al., 2018). The supervised learning technique is an unbiased method for predicting microbial patterns that are unique to specific classes of samples (N. Bokulich et al., 2018). ASVs and genera predictive of the metS and non-metS clusters were identified by a Random Forest classifier. To increase the accuracy of the model, feature selection optimization, and parameter tuning were enabled (N. Bokulich et al., 2018).

Functional analysis

Metagenomic functional predictions were assigned to ASVs using the QIIME 2 phylogenetic investigation of communities by reconstruction of unobserved states 2 (PICRUSt2) plugin (<https://github.com/gavinmdouglas/q2-picrust2>) (Douglas et al., 2019; Langille et al., 2013). A table containing KEGG ortholog (KO) abundance values for each ASV was used to analyze and visualize the functional diversity between samples. Prior to diversity analyses, data were normalized by rarifying to the sampling depth of 70,000,000.

Statistical analyses

Statistical analyses on metabolic and taxonomic data were performed in GraphPad Prism 8 (GraphPad Software, Inc. La Jolla, CA). Differences among group means were assessed with a one-way ANOVA test followed by Tukey's multiple comparisons test. Prior to ANOVA testing, the Shapiro-Wilk normality test was used to determine whether the data were consistent with a Gaussian distribution. Results of statistical testing are denoted as follows: * $p < 0.05$, ** $p < 0.01$, ***

$p < 0.001$, **** $p < 0.0001$. Values reported as q-values are representative of the Benjamini-Hochberg FDR corrected p-value.

Acknowledgements

This work was supported by the National Institutes of Health (NIH) grants P20GM103434 and R44GM113545. LCB and JKP are funded by the NASA West Virginia Space Grant Consortium Training Grant NNX15AI01H. JHK is funded by NIH 1 R15 DK113604-01A1 and American Heart Association 18AIREA33960437. We would like to acknowledge the Marshall Genomics and Bioinformatics Core Facility, Huntington, WV for the support provided to help make this publication possible. We would also like to thank Drs. Jeremy McAleer and Joseph Horzempa for their thorough and thoughtful reviews of this manuscript.

CHAPTER 4

DEVELOPMENT OF A VECTOR SYSTEM FOR IN VIVO MANIPULATION OF THE GUT MICROBIOTA

Abstract

During evolution, the human gastrointestinal tract has coevolved with one of the densest and most complex microbial communities in existence: the gut microbiota. While emerging research has demonstrated the critical role of the gut microbiota in human health and disease, the causal relationship between genes expressed by probiotic or pathogenic phylotypes and host phenotypes largely remains unknown and presents as a challenge to study with current molecular techniques. We sought to bridge this gap by developing a collection of broad-host bacterial vectors (biome vectors) encoding fluorescent and bioluminescent markers for controlled, mechanistic studies of the gut microbiota *in vitro* and *in vivo*, respectively. We hypothesize that our biome vectors can be used as a method to establish causality between gut microbes and their impact on host physiological function by tightly controlling the expression of genes of interest *in vivo* and monitoring the phenotypic response in real-time. The biome vectors were constructed by cloning hemagglutinin (HA) epitope-tagged fluorescent or bioluminescent proteins downstream of the promoter region of the backbone vector, pHERD30T. The biome vectors' protein expression is conditionally regulated by the L-arabinose induction of the P_{BAD} promoter. Using the FDA-approved *Escherichia coli* TOP10 strain as a host, fluorescent and bioluminescent protein expression were characterized *in vitro* under aerobic, anaerobic, and gut-simulated conditions via an *ex vivo* gut microbial community. Post-co-cultivation, qPCR analysis of four gut commensals originating from the *ex vivo* microbiota, was performed to determine the impact of the biome vectors on the microbial community. Results of fluorescent and

bioluminescent expression obtained through imaging showed that expression responded in a dose-dependent manner to L-arabinose supplementation, which ranged in concentration from 0.001 - 2.0%. However, results from western blotting suggest bioluminescent and fluorescent proteins reached maximal expression at 0.01% L-arabinose supplementation. Conversely, in the absence of L-arabinose, protein levels were below the limit of detection, indicating P_{BAD} promoter-based systems are tightly regulated and may be an effective method for controlled protein expression in diverse and polymicrobial environments. *In vivo* tracking of *E. coli* harboring constitutively expressed bioluminescence and our bioluminescence biome vector resulted in transient, avirulent colonization of fewer than 18 hours in the intestines of male and female C57BL/6J mice. Future work will entail the engineering of a bacterial delivery vehicle to ensure long-term, rather than transient colonization *in vivo*. In conclusion, this study provides a comprehensive characterization of versatile bacterial vectors that present a novel paradigm for studying microbe-host interactions in a polymicrobial community.

Introduction

The human intestinal tract is colonized with trillions of microbes, collectively defined as the gut microbiota (Hooper & Macpherson, 2010). Over the past decade, the influence of microbes on human development, health, and immunity has been interrogated, primarily through the method of 16s ribosomal RNA sequencing (Yarza et al., 2014). Profiling of the gut microbial community under homeostatic and diseased states has revealed strong correlations between members of the microbiota and their protective or potentiating effects on metabolic- and immune-related (G. D. Wu & Lewis, 2013). However, transitioning research objectives from associations to causal mechanisms remains an ongoing challenge, impart because of the lack of

genetic tools that enable the inference of microbial gene function and host disease phenotype (Braniste et al., 2014; G. D. Wu & Lewis, 2013).

When studying complex ecosystems like the microbiome, the influence of uncontrollable environmental factors during experimentation and interpretation of data must be considered. The variation observed within the gut microbiota can be attributed to the maternal microbiota, diet, exercise, genetics, antibiotic usage, and stress, making the etiology of the observed disease a challenge to identify (Davenport et al., 2014; Gilbert et al., 2016; Livanos et al., 2016). Thus, a method for studying the gut microbial community under controlled conditions is essential to further the understanding of microbe-host interactions. This demand for a controlled, mechanistic approach to microbiota studies inspired us to develop a collection of bacterial vectors, called biome vectors. The purpose of these molecular tools is to enable precise genetic manipulation of the gut microbiota as a method to establish causality between gut microbes and disease phenotypes.

We constructed the biome vectors by cloning hemagglutinin (HA) epitope-tagged fluorescent and bioluminescent proteins downstream of the P_{BAD} promoter region in the backbone vector, pHERD30T—a vector that was previously developed in our lab (Qiu, Damron, Mima, Schweizer, & Yu, 2008) (Figure 19). P_{BAD} -based promoter systems enable genes of interest to be conditionally expressed in a dose-dependent manner through supplementation of the sugar, L-arabinose (Qiu et al., 2008). In the absence of L-arabinose or the presence of D-glucose, the P_{BAD} promoter is actively repressed, and transcriptional activity is inhibited (Qiu et al., 2008). The biome vectors contain replicons that enable replication in Gram-negative organisms, including *Escherichia*, *Pseudomonas*, and *Burkholderia* species (Qiu et al., 2008).

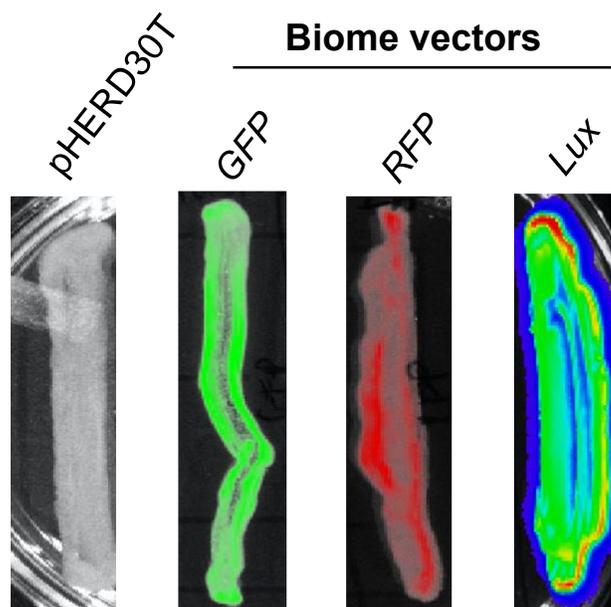


Figure 19. Fluorescent and bioluminescent expression by the biome vectors

The biome vectors were developed by cloning hemagglutinin epitope-tagged green (*gfpmut3*), and red (*dsRed-Express*), and bioluminescence (*luxCDABE*) proteins in the multiple cloning site of pHERD30T, downstream of the P_{BAD} promoter.

Lastly, the biome vectors harbor a Gentamicin-resistance marker, which can be used to maintain selective pressure if necessary. While P_{BAD} promoters are routinely used in bacterial genetics (Szelióva, Krahulec, Safranek, Lisková, & Turna, 2016), we intend to expand the knowledge and applications of these systems in metabolically complex and diverse microbial environments, such as the gut microbiota. In this study, we report on the inducible and repressive dynamics of the biome vectors' P_{BAD} promoter under aerobic, anaerobic, and gut microbiota-simulated conditions. To determine the potential impact of the biome vectors on the endogenous gut flora during *in vivo* use, we developed an *ex vivo* microbiota model, whereby the effect of introducing the exogenous biome vectors into an established polymicrobial community could be

monitored. Lastly, we analyzed the inducible functionality, temporal colonization, and physiological side effects of our bioluminescence-expressing biome vector under *in vivo* gut conditions in C57BL/6J mice. In conclusion, the biome vector system represents a novel paradigm for studying microbe-host interactions in polymicrobial communities.

Results

Construction of biome vectors

To meet the demands of a vector suitable for controlled microbiota studies, we chose pHERD30T as the backbone of our vector (Figure 20A). The pHERD shuttle vectors were created in our laboratory by Qiu et al. in 2008. pHERD30T utilizes the P_{BAD} promoter, which enables cloned genes to be conditionally expressed at titratable levels in the presence of L-arabinose (Qiu et al., 2008). In the absence of L-arabinose or the presence of D-glucose, the P_{BAD} promoter is actively repressed by *araC*, thereby preventing transcription (Qiu et al., 2008). pHERD30T harbors the gentamicin (Gm) resistance marker, *aacC1*, as well as replicons, pBR322, and pRO1600, which support replication in *Escherichia coli* and *Pseudomonas aeruginosa*, respectively. These replicons were recently reported functional in *Burkholderia cepacia*, *Klebsiella pneumoniae*, *Salmonella enterica*, *Shigella flexneri*, as well as *Bordetella bronchiseptica* (Barbier & Damron, 2016). Other genetic features of pHERD30T include the *oriT* region for conjugation-based plasmid transfer, and the *rep* gene encoding the replication-controlling protein (Qiu et al., 2008).

To create the biome vector series, we modified pHERD30T by inserting the sequence of HA-tagged fluorescent or bioluminescent proteins into pHERD30T's multiple cloning site between EcoR1/HindIII restriction enzyme sites. For the generation of the GFP (Figure 20B) and RFP (Figure 20C) biome vectors, fluorescent proteins, GFPmut3 and dsRedExpress, were

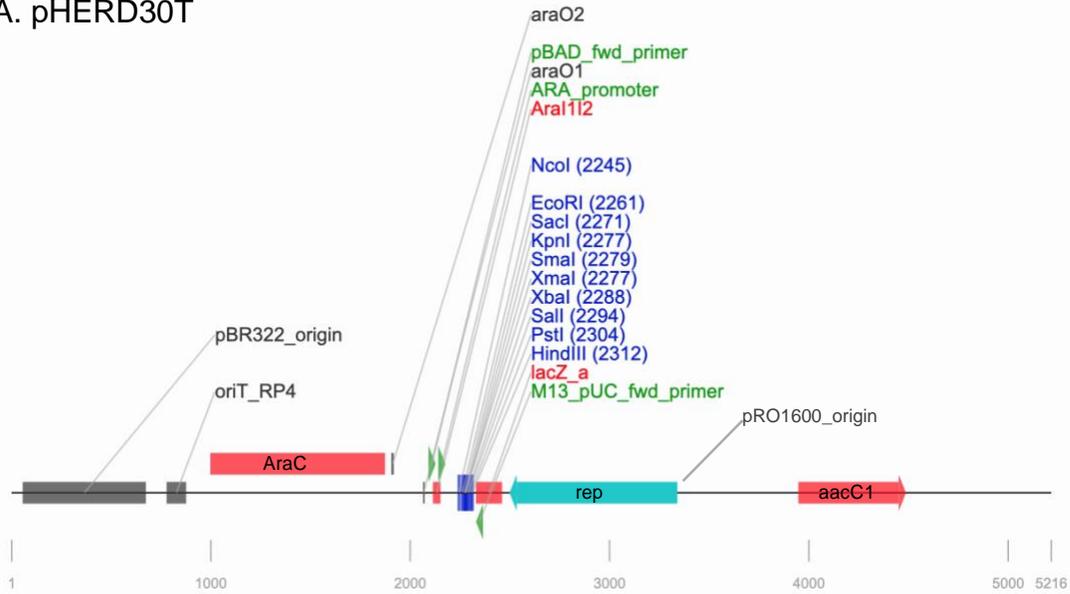
chosen, respectively, because there is no spectral overlap between these proteins, thereby enabling these vectors to be used simultaneously, such as for *in vitro* coinubation competition assays (Speare & Septer, 2019). The excitation and emission for GFPmut3 are 505 and 511 nm, respectively, while the excitation and emission for DsRed-Express are 554 and 586 nm, respectively. These fluorescent vectors are intended for *in vitro* and *ex vivo* studies, as they yield excellent, quantifiable signals that can be directly observed through microscopy and flow cytometry (Troy, Jekic-McMullen, Sambucetti, & Rice, 2004).

During *in vivo* experiments, however, fluorescent signals can become undetectable due to the spectral overlap shared with and autofluorescence of mammalian tissue (Troy et al., 2004). In contrast to fluorescence, which requires a specific wavelength for excitation of its fluorophore, bioluminescence signals are produced endogenously and emitted as the result of a series of chemical reactions. For this reason, bioluminescence is not susceptible to photobleaching like fluorescent proteins. For example, when expressed in *E. coli* strain JM109, GFPmut3 and dsRedExpress lose 50% of their initial brightness within 69 and 357 seconds of excitation, respectively (Barbier & Damron, 2016).

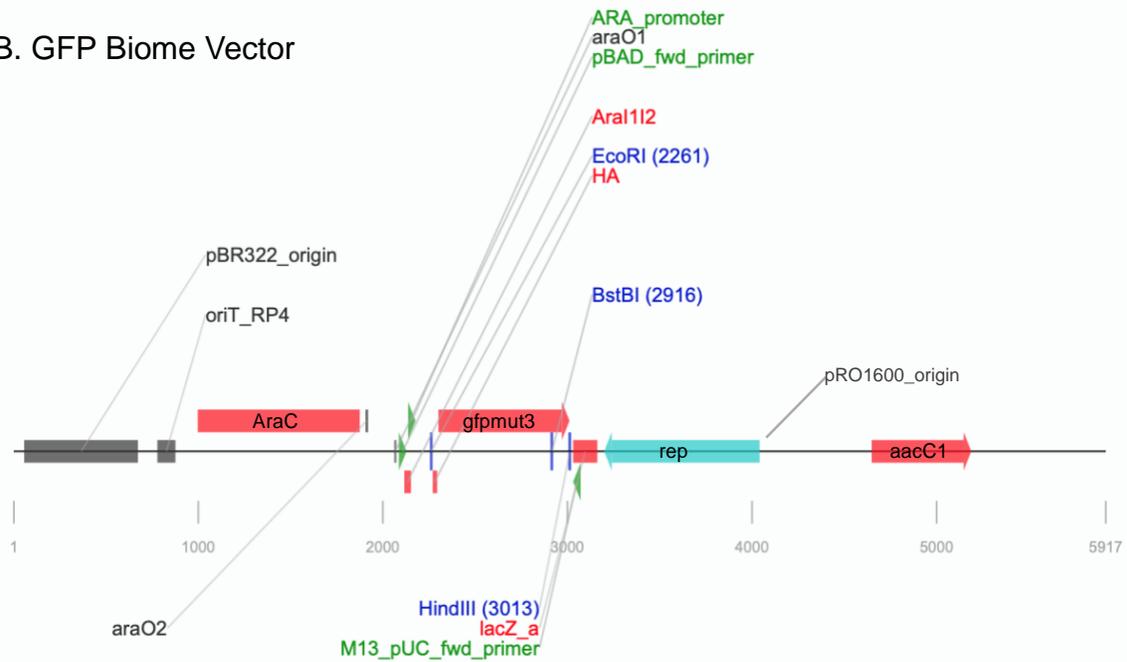
Bioluminescence also exhibits minimal background noise, making it approximately 50 times more sensitive than fluorescence and easily detectable through mammalian tissue (Gregor, Gwosch, Sahl, & Hell, 2018). Thus, we aimed to construct a bioluminescence biome vector for *in vivo* experiment usage. To generate the Lux biome vector (Figure 20D), we chose the *luxCDABE* operon, which originates from *Photorhabdus luminescens*, a symbiotic-nematode bacterium (Rodou, Ankrah, & Stathopoulos, 2010). The production of bioluminescence expression is entirely encoded by genes of the operon itself; however, exogenous factors,

including flavin mononucleotide and molecular oxygen, from metabolically active cells are needed to sustain the chemical reaction resulting in bioluminescence (Gregor et al., 2018).

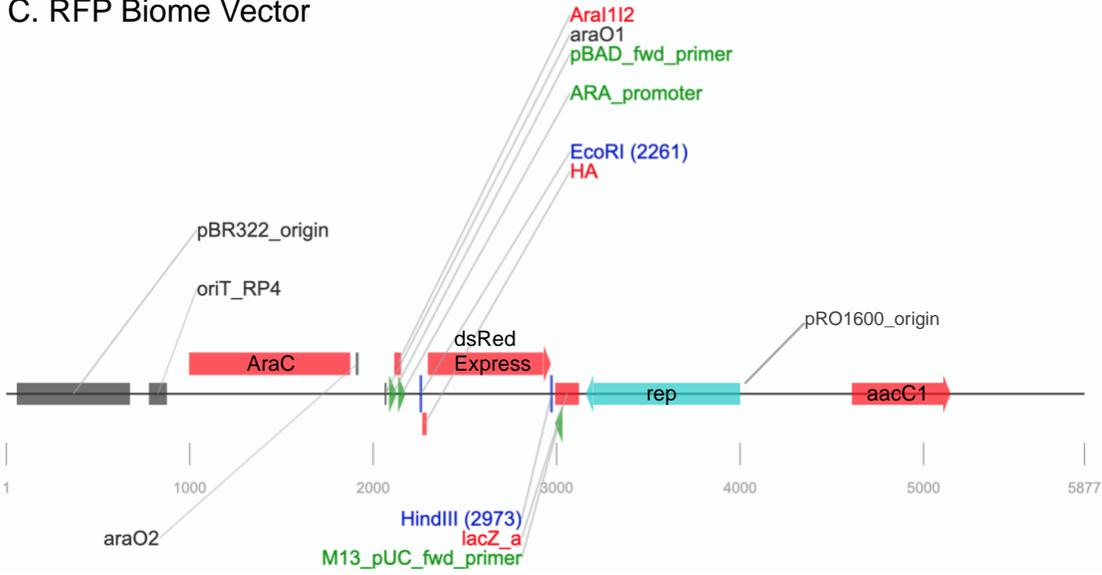
A. pHERD30T



B. GFP Biome Vector



C. RFP Biome Vector



D. Lux Biome Vector

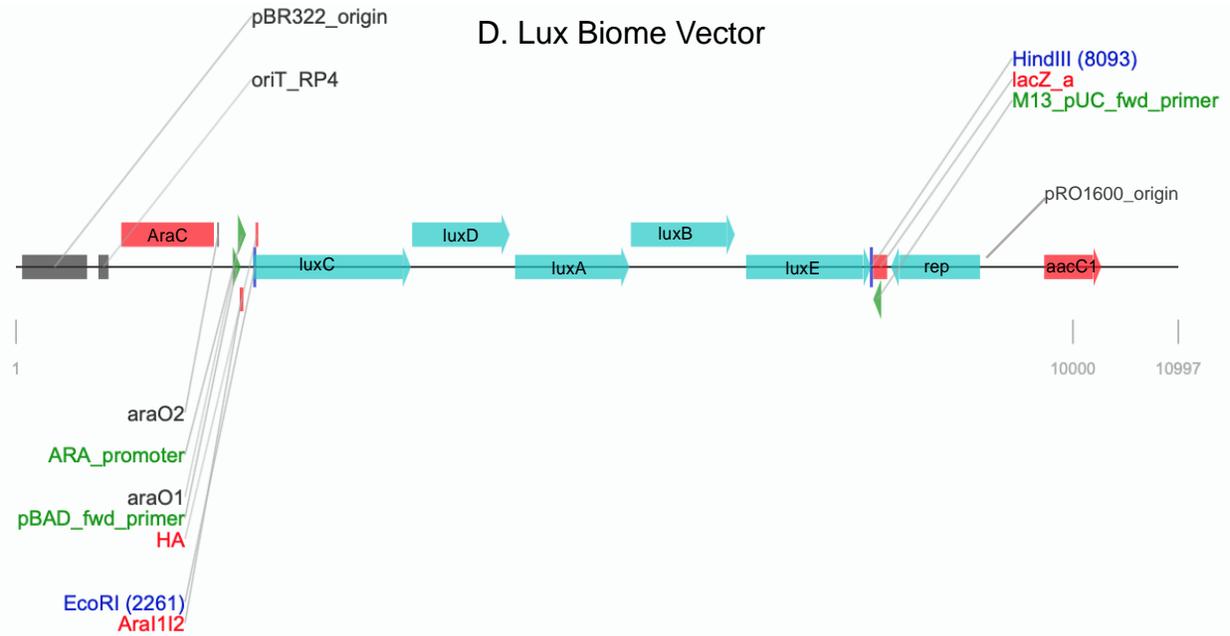


Figure 20. Maps of pHERD30T and biome vectors

The biome vectors were constructed using pHERD30T as the backbone (A). Derived from pHERD30T, the biome vectors harbor an inducible P_{BAD} promoter, gentamicin resistance marker (aacC1), and pBR322 and pRO1600 replicons, which enable replication in *Escherichia* and

Pseudomonads, respectively. GFP (B), RFP (C), and Lux (D) biome vectors were generated by cloning hemagglutinin (HA)-gfpmut3, HA-dsRedExpress, and HA-luxCDABE gene fusions, respectively, into the multiple cloning site of pHERD30T between EcoR1/HindIII restriction enzyme sites.

Quantitative imaging shows the biome vectors' fluorescence and bioluminescence expression increases in a dose-dependent manner to L-arabinose supplementation

Conditional bioluminescence and fluorescence expression of the biome vectors was characterized *in vitro* (Figure 21) to determine the (i) basal P_{BAD} promoter leakage, (ii) minimum concentration of L-arabinose to achieve maximal, reproducible fluorescence or bioluminescence signal, and (iii) necessity of D-glucose for complete promoter repression.

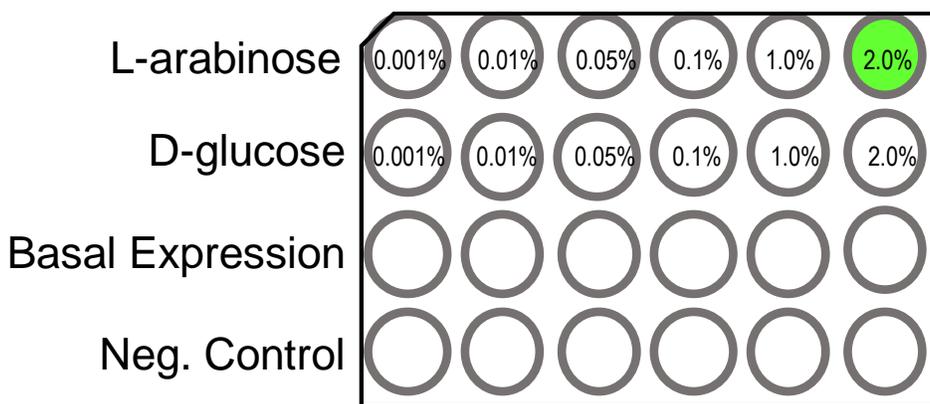


Figure 21. *In vitro* assay experimental design

Overview of the 24-well plate layout for the *in vitro* assay. Induced and repressed fluorescence and bioluminescence expression were quantified over 0.001-2.0% L-arabinose and 0.001-2.0% D-glucose concentrations, respectively. Without the addition of exogenous carbohydrate sources, the observed basal expression of the biome vectors was representative of the baseline leakage of the P_{BAD} promoter, while the backbone of the biome vectors, pHERD30T, served as the negative control.

These experiments were conducted under aerobic and anaerobic conditions to determine the relative fitness of the vectors in anoxic environments to simulate the anaerobic conditions of the gut, which maintains oxygen concentrations of 2-7% (G. He et al., 1999). Under aerobic conditions, GFP, RFP, and Lux biome vectors exhibited fluorescence and bioluminescence expression that increased in a dose-dependent manner in response to augmenting concentrations of L-arabinose supplementation, reaching maximal expression at 2% (Figure 22). At nearly all concentrations of L-arabinose, GFP, and Lux biome vector expression was significantly attenuated under anaerobic conditions compared to their aerobic counterparts. The RFP biome vector was an exception to this trend, albeit the overall expression was generally lower and more variable. This observation is consistent with previous characterizations *DsRedExpress*, which indicates that RFP expression by this protein is weak within the first 24 hours of incubation, with signals reportedly intensifying after 36-48 hours of incubation (Choi & Schweizer, 2006).

Under anaerobic conditions, the bioluminescence expression produced by the Lux biome vector reached its maximal expression at the lowest (0.001%) dose of L-arabinose. Because bioluminescence production is dependent upon the availability of flavin mononucleotide and molecular oxygen within the environment (Gregor et al., 2018), these results may suggest that overexpressing the *luxCDABE* operon under anoxic conditions at increased L-arabinose concentrations may cause localized metabolic resources to be exhausted more quickly than expression at lower concentrations of L-arabinose. Importantly, among all biome vectors, we observed no difference ($p=0.7686$) in the fluorescence and bioluminescence expression between basal and repressed conditions, thereby suggesting that D-glucose supplementation is not necessary to repress the P_{BAD} promoter (Figure 23).

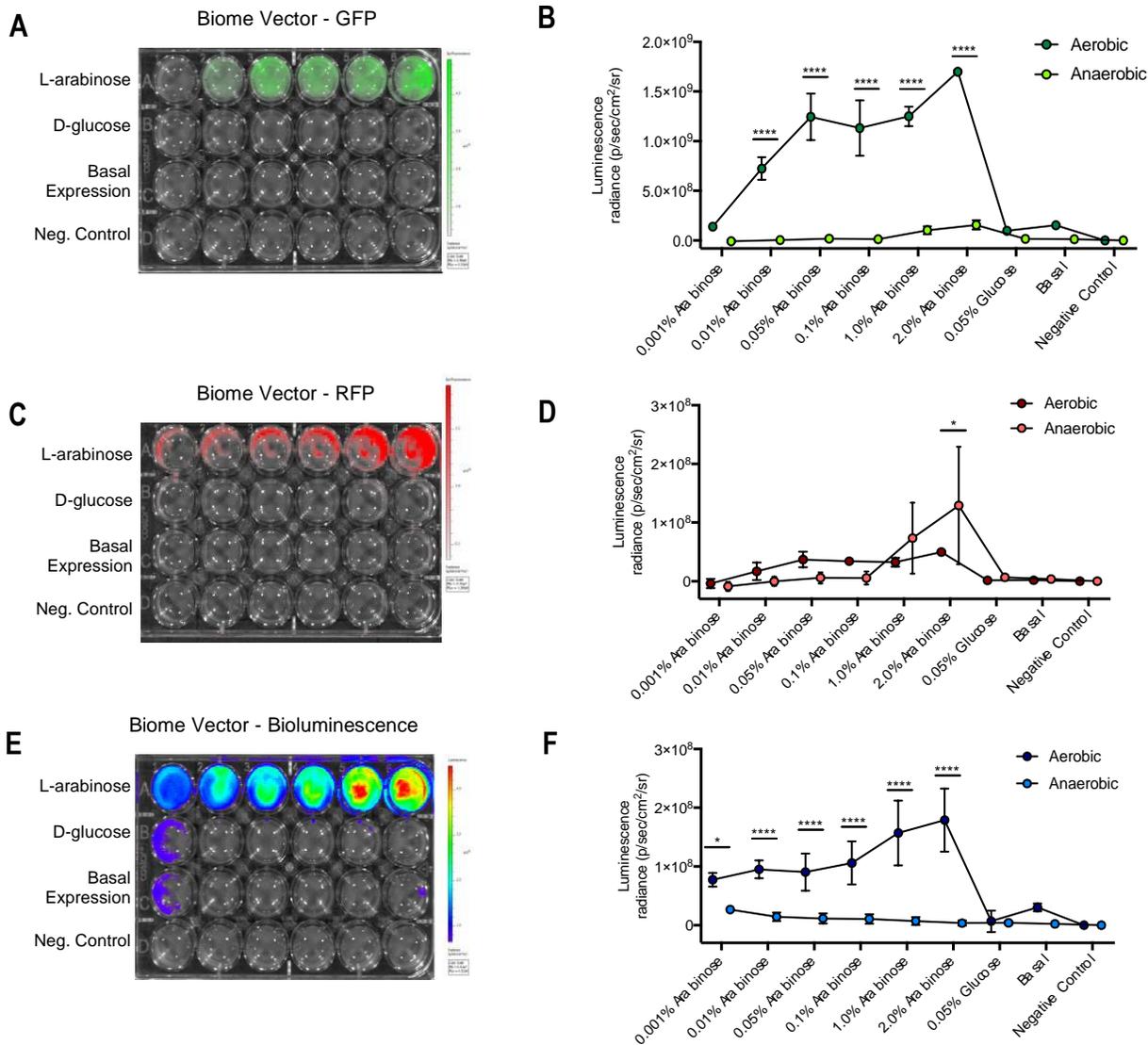


Figure 22. Aerobic and anaerobic *in vitro* quantitation of titrated L-arabinose-induced fluorescence and bioluminescence expression

Images and quantitation of green (A-B) and red (C-D) fluorescence and bioluminescence (E-F)

expression from 24-well plate assays. Images were captured with the IVIS Lumina XRMS III *in*

in vivo imager, and expression quantified using the calibrated units of radiance. The assay images

(A, C, E) are representative of aerobic assays. Anaerobic fitness of the biome vectors was

assessed by comparing the fluorescence or bioluminescence expression to their aerobic counterparts at each supplemented carbohydrate concentration or condition (B, D, F). Each data point is representative of at least two technical replicates. Mean \pm SD, * $P < 0.05$, **** $P < 0.0001$, two-way ANOVA followed by Sidak multiple comparisons test.

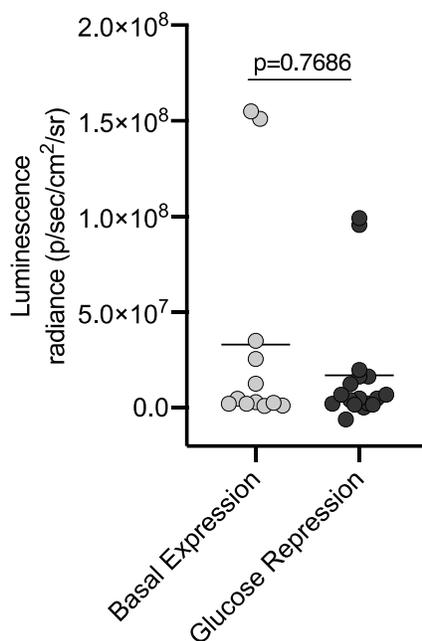


Figure 23. Glucose supplementation is not necessary to achieve P_{BAD} promoter repression
 Fluorescence and bioluminescence expression under basal and 0.5% glucose-supplemented conditions. Data points are aggregated from aerobic and anaerobic *in vitro* assays performed with GFP, RFP, and Lux biome vectors. Data were analyzed using a Mann Whitney test.

Western blot analysis reveals that 0.1% L-arabinose induction leads to optimal protein expression in biome vectors

The induced, repressed, and basal transcriptional activity of the GFP, RFP, and Lux biome vectors was assessed with the Wes, a capillary-based automated immunoassay system.

The total bacterial protein was extracted directly from the aerobic and anaerobic *in vitro* assays described above, and proteins were probed with HA monoclonal antibodies. During the cloning process, HA was fused to GFP, RFP, and the luxCDABE operon, therefore HA is a proxy for measuring the bioluminescence and fluorescence proteins produced by each biome vector at the various experimental conditions.

Consistent with our quantitative imaging analysis, basal fluorescence and bioluminescence expression was below the limit of the immunoassay's detection, thereby further supporting the notion that the P_{BAD} promoter is tightly regulated. In contrast to the quantitative imaging results, which generally showed maximum fluorescence and bioluminescence intensity occurring at 2% L-arabinose supplementation, we observed the biome vectors' protein expression to peak at 0.01% L-arabinose induction, as evidenced by the magnitude of the chemiluminescent signals shown as an electropherograms and virtual western blot images (Figure 24). A likely explanation for the incongruence between these two methods of detection is the difference in the half-life of the fluorescent and bioluminescent proteins detected through imaging and the HA-tag used for protein quantification. For example, the half-life of *gfp* and *dsred* variants is at least 24 and 96 hours, respectively. While the half-life of *P. luminescens* bioluminescence is approximately 4 hours, fluorescent- and bioluminescent-based reporter proteins are continually being modified to have shortened half-lives to enable dynamic gene and protein regulation monitoring (Andersen et al., 1998; Waidmann, Bleichrodt, Laslo, & Riedel, 2011). While the decay kinetics on HA degradation has not been reported, our data suggest that HA degradation may occur relatively quickly, and therefore, provide quantitative data that is representative of the real-time recombinant protein expression, rather than measuring the stable proteins that have accumulated over time.

The anaerobically incubated Lux biome vector was the only condition where a biome vector peaked at 0.001% L-arabinose supplementation (Figure 24F). Under such anoxic conditions, a decreased amount of the endogenous factors needed to produce the bioluminescent signal likely explains the attenuated Lux protein expression observed in the anaerobic immunoassay and quantitative imaging studies. Collectively, these data confirm the P_{BAD} promoter behaves similarly under aerobic and anaerobic conditions and suggests that 0.1% L-arabinose induction of the P_{BAD} promoter supports sustained, high-level recombinant protein expression provided that the endogenous factors and co-factors essential for recombinant protein expression are available.

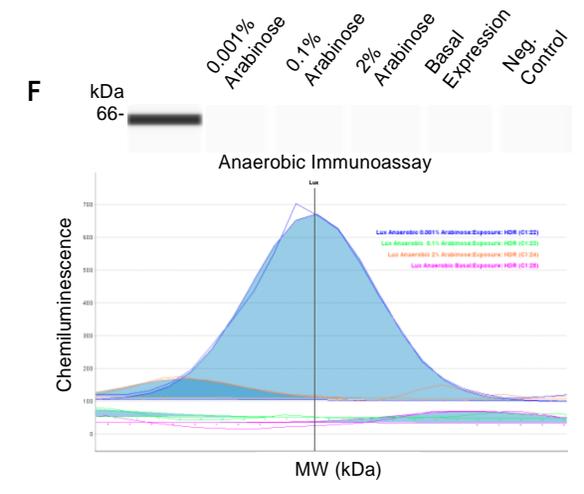
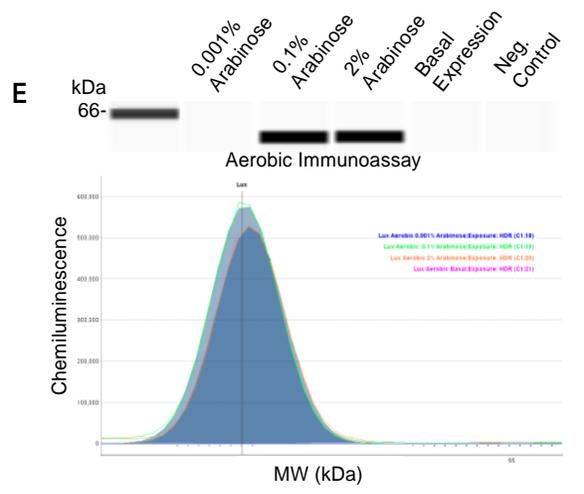
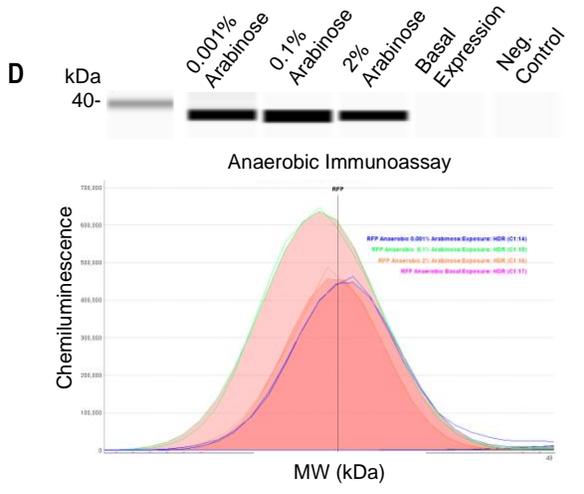
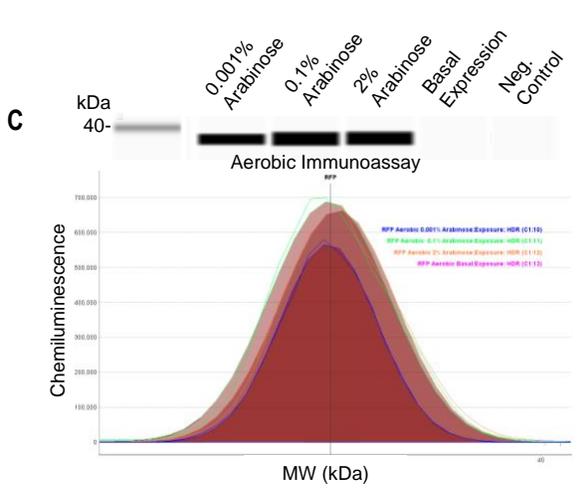
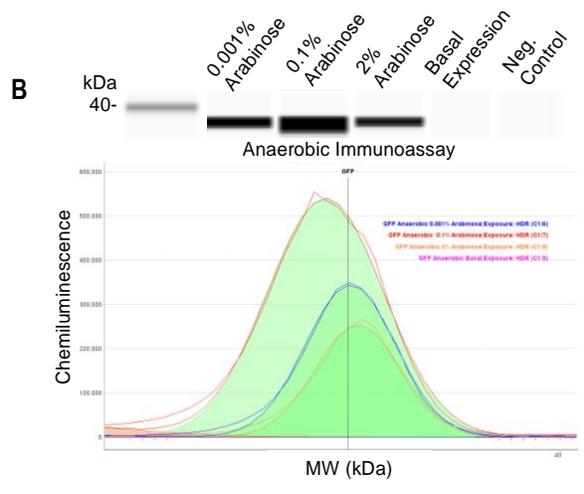
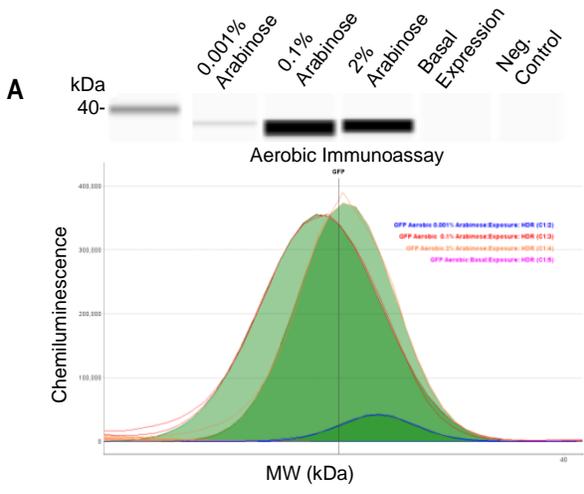


Figure 24. Quantitation of titrated L-arabinose-induced fluorescence and bioluminescence protein expression under aerobic and anaerobic conditions

Bacterial proteins were extracted directly from *in vitro* assays, and the aerobic and anaerobic expression of GFP (A-B) RFP (C-D) and bioluminescence (E-F) at 0.001%, 0.1%, and 2% L-arabinose-induced and basal conditions were quantified using a capillary-based automated immunoassay. Within each aerobic and anaerobic analysis of the GFP, RFP, and Lux biome vectors, the magnitude at which the chemiluminescent HA signal was detected and quantified is displayed as an electropherogram, with a virtual western blot image directly above.

Introduction of biome vectors into *ex vivo* microbiome community results in minimal modifications to the endogenous flora

To determine how the biome vector system may impact the endogenous flora of the gut, we established a co-cultivation model, wherein an *ex vivo* gut microbiota community derived from mice was anaerobically co-cultured with the biome vectors and analyzed under various conditions. Using this co-cultivation strategy, we confirmed that L-arabinose-induced bioluminescence expression was indeed obtainable under gut-simulated conditions (Figure 25A). Additionally, we determined whether Gentamicin and L-arabinose supplementation, as well as the presence of an exogenous *E. coli* strain harboring the pHERD30T vector, resulted in a perturbation to the *ex vivo* microbiota. qPCR analysis revealed Enterobacteriaceae, Bacteroides, and *C. leptum* populations were unaffected by co-cultivation with the exogenous *E. coli* strain and carbohydrate and antibiotic supplementation; however, the *Lactobacillus* population was significantly increased when the *ex vivo* microbiota was co-cultured and media supplemented Gentamicin and L-arabinose (Figure 25B).

In order to discern which element(s) of co-cultivation led to the increased in *Lactobacilli*, we performed a microanalysis, wherein a *Lactobacillus* isolate derived from the *ex vivo*

community was exposed to Gentamicin and/or L-arabinose in the presence and absence of the exogenous *E. coli* strain harboring the pHERD30T vector (Figure 25C). Our results show that *Lactobacillus* exhibited extreme sensitivity to Gentamicin exposure, thereby decreasing the genera's population by eight times compared to the non-treatment group. Conditions involving the supplementation of L-arabinose as well as co-incubation with *E. coli* appeared to have a synergistic effect on the *Lactobacillus* population. Accordingly, previous studies have shown that L-arabinose is a preferred carbon source of certain *Lactobacilli* spp. (Gobbetti, Lavermicocca, Minervini, de Angelis, & Corsetti, 2000). Thus, these results suggest that the biome vectors and other P_{BAD} promoter-based vectors are not ideal for studies of which *Lactobacillus* is the taxa of interest. In sum, from these studies, we conclude that the biome vector system maintains functionality in a metabolically active polymicrobial community while causing minimal change to endogenous gut flora.

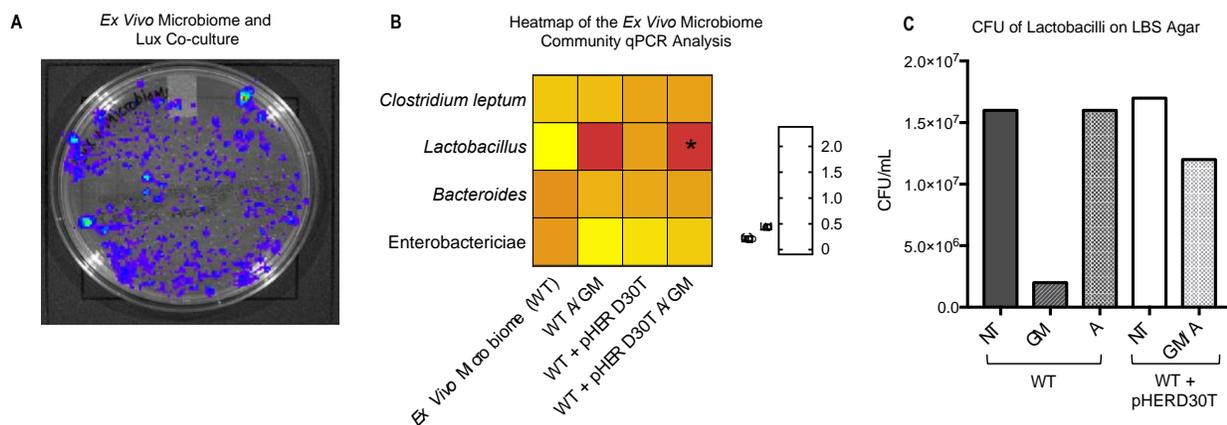


Figure 25. Characterization of the biome vectors in an *ex vivo* gut microbiota model. Induced bioluminescence expression of the Lux biome vector co-cultured with an *ex vivo* microbial community (A). qPCR analysis of four gut commensal groups within an *ex vivo* community under normal (WT) growth and experimental conditions[†] (B). Fold changes were calculated with the $\Delta\Delta C_t$ method. Differential taxa abundance between WT and the experimental

conditions were analyzed for significant changes. * $P < 0.05$, two-way ANOVA followed by the Dunnett multiple comparisons test. Data are representative of two biological replicates.

Lactobacilli CFU/mL on LBS agar after a 5-hour incubation of the *ex vivo* microbiota (WT) alone and with *E. coli* harboring the pHERD30T vector under experimental conditions[†] (C).

[†]GM = Gentamicin (13 $\mu\text{g}/\text{mL}$); A = 0.1% L-arabinose; GM/A = Gentamicin (13 $\mu\text{g}/\text{mL}$) and 0.1% L-arabinose; NT = no treatment.

Biome vectors transiently colonize the murine intestines

To demonstrate the potential applications of the biome vector system for *in vivo* use, we conducted a preliminary study to assess the functionality and safety of the Lux biome vector when introduced into the murine gut microbiota. Male and female B6 mice were orally gavaged with 200 μl of *E. coli* containing the constitutively expressed bioluminescence vector, pUC18T-mini-Tn7T-*lux*, Lux biome vector, or PBS. At 0 hours, the constitutive bioluminescence vector can be observed in the gastrointestinal tract of the orally gavaged hosts (Figure 26). The most notable feature at this time point is that only 50% of mice harbor the bacterial inoculum within their gut. These results highlight an issue with the oral gavage protocol, which originated with the plastic feed tubes used to deliver the bacterial inoculum to the intestines. During the oral gavage process, mice had the tendency of biting the feeding tube, which caused punctures and even fragmentation to the tube. This not only resulted in a different amount of the bacterial inoculum being delivered into the gastrointestinal tract but also led to the one and only observed fatality (Figure 27). In the future, a metal feeding tube will be used to eliminate these issues associated with inoculum delivery.

In the two mice confirmed to receive the bacterial inoculum with pUC18T-mini-Tn7T-*lux* vector, the bacteria were either excreted or killed within 18 hours of oral gavage, as no

bioluminescence was observed at this time point. Accordingly, we did not observe any bioluminescence expression in mice that received the Lux biome vector when imaged at the 18-hour timepoint. Fecal samples were collected at the 18-hour timepoint from all mice that were gavaged with the Lux biome vector. The total bacterial population was isolated through filtration and incubated overnight on LB agar containing Gentamicin and L-arabinose. Plates were subsequently imaged for bioluminescence expression; however, none was detected. These data indicate that the lack of observed bioluminescence was likely due to the rapid elimination of the bacterial inoculum from the gastrointestinal tract extremely, rather than a failure by the vectors to produce a bioluminescence signal. These results also highlight the need for an optimized delivery system that can effectively colonize the gut microbiota without disrupting the existing structure of the endogenous community.

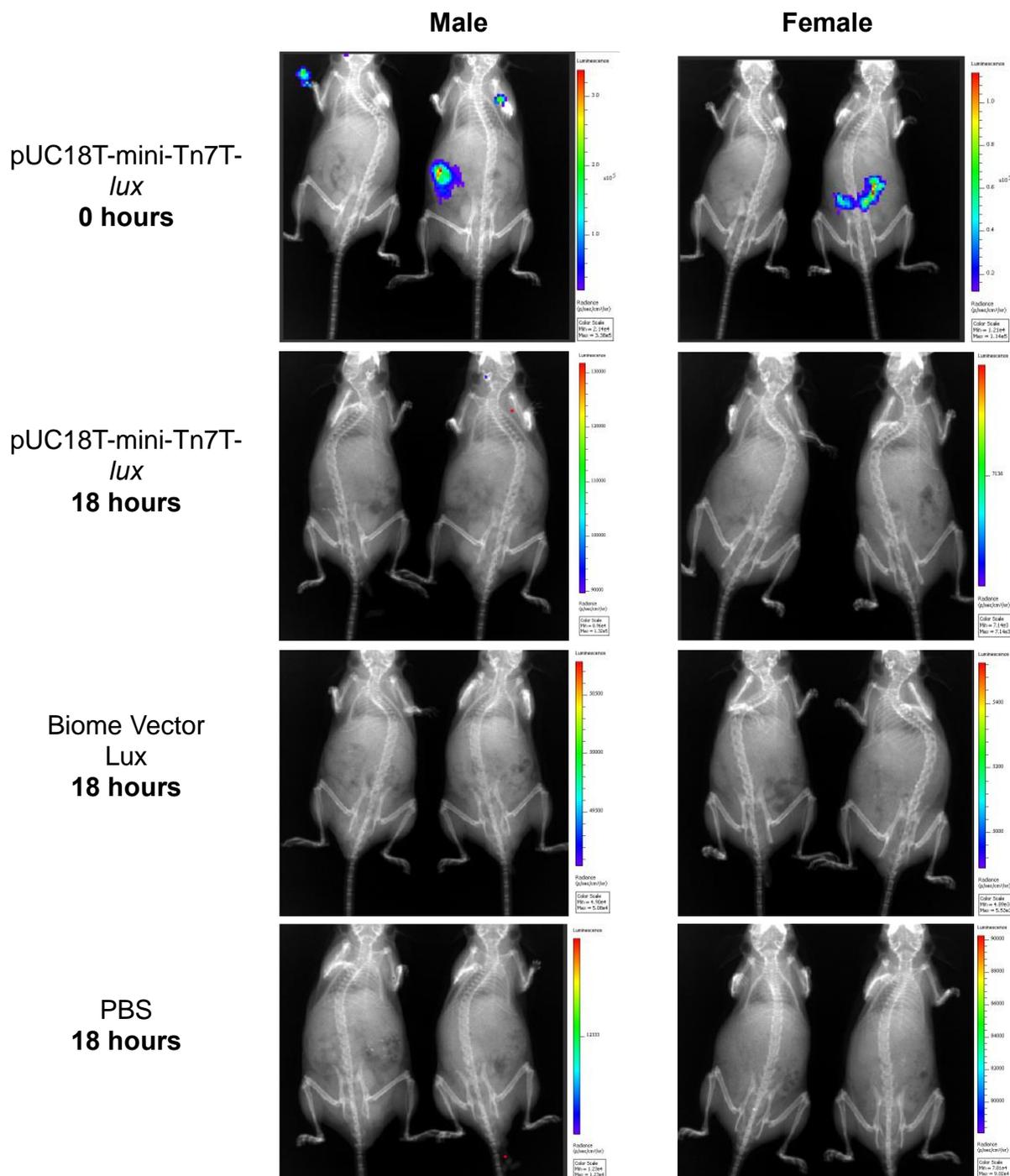


Figure 26. *In vivo* colonization of *E. coli* harboring constitutively expressed bioluminescence vector (pUC18T-mini-Tn7T-*lux*) and the inducible bioluminescence biome vector.

Mice were orally gavaged with 200 μ l of *E. coli* containing the constitutively expressed

bioluminescence vector (pUC18T-mini-Tn7T-*lux*), bioluminescence biome vector, or PBS

(negative control). A subgroup of mice received normal drinking water, while the others received water supplemented with 2% L-arabinose and Gentamicin (13 $\mu\text{g}/\text{mL}$). Each experimental condition consisted of a male and female group, with two mice/group. Mice were imaged at 0 and 18 hours post-gavage and closely monitored for side effects for up to a week.

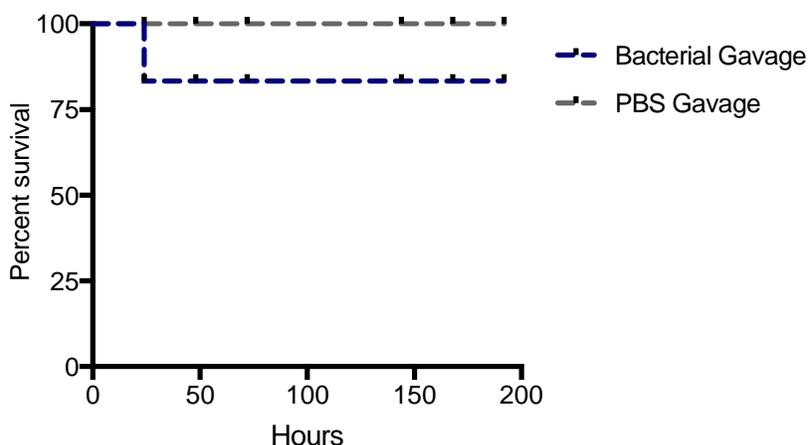


Figure 27. Survival curve of mice post-oral gavage with bacterial vectors or PBS

Survival curve of the mice included in the bacterial oral gavage experiment. Only one fatality occurred, which was the result of a fragmented plastic feeding tube. The mouse was immediately and humanely euthanized to avoid any additional discomfort. No other adverse effects were observed throughout the duration of the study.

Conclusion

The objective of this work was to develop a bacterial vector system for *in vivo* manipulation of the gut microbiota. As demonstrated, the fluorescent and bioluminescent vectors can be utilized for aerobic and anaerobic experiments *in vitro*, *ex vivo*, and *in situ* and readily detected through imaging, flow cytometry, and western blotting. Our results also show that regardless of the gene(s) or operon cloned into the biome vector for recombinant expression,

0.01% L-arabinose supplementation yields sustained, highly-detectable protein expression. Moreover, we showed that glucose supplementation is not necessary to achieve P_{BAD} promoter repression. As proof of concept, we developed an *ex vivo* gut microbiome model to verify the functionality of the biome vectors under gut-simulated conditions. This co-cultivation strategy could also serve as a unique model for manipulating the microbiota's metabolic pathways in a controlled environment and mutant library selection. qPCR analysis showed that the introduction of the biome vectors into the *ex vivo* gut microbiota caused minimal alterations to the microbial community structure. The avirulent nature of the biome vector system was further confirmed through *in vivo* experiments.

Harnessing the ability to modulate the mammalian gut microbiota through the introduction of genetically engineered microbes would lay the foundation for a targeted approach to microbiota manipulation with an array of therapeutic applications (Lee et al., 2018). Aside from this current study, one of the most recent developed platforms for microbiota engineering, Metagenomic Alteration of Gut microbiome by In situ Conjugation (MAGIC), enables mobile genetic elements to be introduced and transferred the native gut microbial community via horizontal gene transfer (Ronda, Chen, Cabral, Yaung, & Wang, 2019). Similar to our results, one of the initial limitations with the MAGIC system was the rapid loss of the *E. coli* strain used as the *in vivo* delivery vehicle (Ronda et al., 2019). The authors hypothesized that using a gut-adapted bacterium would enhance the probability of colonization and stable persistence in the gut, and indeed, showed that genetically engineering gut-derived microbes as delivery vehicles resulted in long-term colonization in the gut (Ronda et al., 2019).

Based on this evidence, future directions include genetically modifying *E. coli* Nissle 1917 (EcN) strain as an *in vivo* delivery system for the biome vectors. Initially isolated from the

human gastrointestinal tract, EcN has since become a commercial probiotic for the treatment of dysbiosis, and ulcerative colitis (Scaldaferri et al., 2016). The biome vector system coupled with a gut-tropic delivery system presents as an excellent model for the precise expression and delivery of proteins and therapeutic peptides as well as a novel method for targeted antigen delivery of mucosal vaccinations, which remain a challenging route of vaccine delivery due to the natural defenses and clearance mechanisms in the gut (Azegami, Yuki, & Kiyono, 2014).

Materials and Methods

Construction of vectors

The *luxCDABE* operon and *gfpmut3* gene were PCR-amplified from the pUC18T-mini-Tn7T-Gm-*lux* and pUC18T-mini-Tn7T-Gm-*gfpmut3* plasmids (Table 5) (Choi & Schweizer, 2006), respectively.

Plasmid and reference	Features
pUC18T-mini-Tn7T-Gm- <i>gfpmut3</i> (Choi & Schweizer, 2006)	Gentamicin resistant; green fluorescence protein for tagging bacteria
pUC18-mini-Tn7T-Gm- <i>lux</i> (Choi & Schweizer, 2006)	Gentamicin resistant; <i>luxCDABE</i> transcriptional fusion vector
pUC18T-mini-Tn7T-Gm dsRedExpress (Choi & Schweizer, 2006)	Gentamicin resistant; red fluorescence protein for tagging bacteria
pHERD30T (Qiu et al., 2008)	Gentamicin resistant; P _{BAD} promoter
Lux biome vector (this study)	Gentamicin resistant; P _{BAD} promoter-controlled bioluminescence expression
GFP biome vector (this study)	Gentamicin resistant; P _{BAD} promoter-controlled GFP expression
RFP biome vector (this study)	Gentamicin resistant; P _{BAD} promoter-controlled RFP expression

Table 5. Bacterial plasmids used in this study

Custom primers (Table 6) were designed so that EcoR1 and HindIII restriction enzyme sites could be added to the 5' and 3' end of the PCR product, respectively. Additionally, the forward primers were designed to add the 5' HA epitope tag. Primers were synthesized by Integrated DNA Technologies (Coralville, IA) using the standard desalting purification procedure. PCR products were ligated into the pCR4-TOPO vector (Invitrogen, Carlsbad, CA), and recombinant clones were selected on Lennox Broth (LB) (Difco, Detroit, MI) agar plates containing Kanamycin (50 µg/mL) (RPI, Mt. Prospect, PA). Plasmid DNA was extracted from recombinant clones grown overnight in LB containing Kanamycin (50 µg/mL) using the QIAprep Spin MiniPrep Kit (Qiagen, Germantown, MD). The PCR products, as well as the recipient vector, pHERD30T, were digested with the restriction enzymes, EcoR1 and HindIII (New England Biolabs, Ipswich, MA), resulting in linearized DNA. The *HA-luxCDABE* and *HA-gfpmut3* gene fusions were ligated into pHERD30T using the Rapid DNA ligation kit (Thermo Fisher Scientific, Waltham, MA), resulting in the Lux and GFP biome vectors. These recombinant plasmids were subsequently transformed into electrocompetent One Shot TOP10 *E. coli* (Invitrogen, Carlsbad, CA). Recombinant clones were selected on LB agar containing Gentamicin (13 µg/mL). The EcoRI/HindIII regions of the GFP and Lux biome vectors were Sanger sequenced at West Virginia University's Genomics Core to verify construct and ensure that no mutations occurred during cloning (see Appendices A and B for partial nucleotide sequences). The RFP biome vector containing *HA-DsRed* gene fusion was commercially synthesized (GenScript Biotech, Piscataway, NJ).

Primer and reference	Sequence (5' - 3')
EcoHA-gfpmut3-F (this study)	AGA ATT CGT ACC CAT ACG ATG TTC CAG ATT ACG CTA TGC GTA AAG GAG AAG AAC TTT TC
Hind-gfpmut3-R (this study)	CCA AGC TTT TAT TTG TAT AGT TCA TCC ATG CCA TGT GTA ATC CCA GCA GCT

EcoHA-luxC-F (this study)	AGA ATT CGT ACC CAT ACG ATG TTC CAG ATT ACG CTA TGA CTA AAA AAA TTT CAT TCA TT
Hind-luxE-R (this study)	CAA GCT TTC AAC TAT CAA ACG CTT CGG TTA AGC TCA AAG CAC ACC CTT TCT
pHERD-SF (Qiu et al., 2008)	ATCGCAACTCTCTACTGTTTCT
pHERD-SR (Qiu et al., 2008)	TGCAAGGCGATTAAGTTGGGT
<i>C. leptum</i> F (Furet et al., 2009)	CCTTCCGTGCCGSAGTTA
<i>C. leptum</i> R (Furet et al., 2009)	GAATTAAACCACATACTCCACTGCTT
<i>Lactobacillus</i> F (LabF362) (Rinttila, Kassinen, Malinen, Krogius, & Palva, 2004)	AGCAGTAGGGAATCTTCCA
<i>Lactobacillus</i> R (LabR677) (Rinttila et al., 2004)	CACCGCTACACATGGAG
<i>Bacteroides</i> F (BactF285) (Dore, Sghir, Hannequart- Gramet, Corthier, & Pochart, 1998)	GGTTCTGAGAGGAGGTCCC
<i>Bacteroides</i> R (UniR338) (Dore et al., 1998)	GCTGCCTCCCGTAGGAGT
Enterobacteriaceae F (Uni515F) (Ludwig, 2007)	GTGCCAGCMGCCGCGGTAA
Enterobacteriaceae (Ent826R) (Ludwig, 2007)	GCCTCAAGGGCACAACCTCCAAG
Eubacteria F (UniF340) (Amann et al., 1990)	ACTCCTACGGGAGGCAGCAGT
Eubacteria R (UniR514) (Amann et al., 1990)	ATTACCGCGGCTGCTGGC

Table 6. Primers used in this study

In vitro assays

In vitro assays were performed to optimize and quantify the biome vectors' fluorescence and bioluminescence expression over titrated concentrations of 0.001 - 2.0% L-arabinose and 0.05% D-glucose. Using 24-well plates (Corning, Corning, NY), 5.0×10^4 CFU/mL *E. coli* inoculum harboring the biome vectors, or pHERD30T was added to 500 μ l LB containing Gentamicin (13 μ g/mL) as well as L-arabinose or D-glucose in each well. Plates were incubated under aerobic conditions at 37 °C for 18-20 hours, and the basal and experimental expression was subsequently measured on the IVIS Lumina XRMS III *in vivo* imager (PerkinElmer, Waltham, MA). To calculate background noise, the luminescence detected from the wells containing pHERD30T was averaged and subtracted from the basal and experimental expression measurements.

To determine if the biome vectors' P_{BAD} promoter retained its inducible functionality in an oxygen-free environment, the *in vitro* assays were repeated and plates incubated at 37 °C for 18-20 hours in an anaerobic chamber (Anaerobe Systems, Morgan Hill, CA). Pre-reduced anaerobically-sterilized (PRAS) LB (Anaerobe Systems, Morgan Hill, CA) containing Gentamicin as well as L-arabinose or D-glucose was implemented in place of the LB media used in the aerobic *in vitro* assays.

Immunoassays

Bacterial proteins were extracted directly from aerobic and anaerobic *in vitro* assay plates described above with the NoviPure Microbial Protein Kit (Qiagen, Germantown, MD). Protein concentrations were measured with the Pierce BCA Protein Assay (Thermo Fisher Scientific, Waltham, MA) microplate procedure, in which the SpectraMax i3x plate reader (Molecular Devices, Downingtown, PA) was used to read the optical density (OD) at 562nm.

Immunoassays were performed on the Wes automated western blotting system (ProteinSimple, San Jose, CA) according to the manufactures' protocol. Briefly, protein samples were diluted to 0.3 mg/mL using 0.1x Sample Buffer and mixed with 5x Fluorescent Master Mix at a 4:1 ratio (Nelson, Guynn, & Chorley, 2017). Protein samples and biotinylated ladder were denatured by heating at 95 °C for 5 minutes. Rabbit anti-HA tag (MAB0601, R&D Systems, Minneapolis, MN) IgG monoclonal antibody was diluted with antibody diluent 2 at a 1:50 ratio (Nelson et al., 2017). Luminal-S and peroxide were mixed at a 1:1 ratio to a create 250 µl final solution. Protein samples, biotinylated ladder, antibody diluent 2, primary antibody, streptavidin-HRP, Luminal-Peroxide mix, wash buffer, and anti-rabbit secondary antibody (DM-001, Protein Simple, San Jose, CA) were loaded into a 12-230 kDa 25 capillary assay plate (Nelson et al., 2017). The loaded plate was centrifuged at ~1000 x g from 5 minutes at ambient temperature. The size immunoassay was run using the following default system settings: 40 minute separation time at 250 volts; 23 minute antibody diluent time; 30 minute primary antibody time; 30 minute secondary antibody time. Results and analysis of immunoassays were performed on Compass for Simple Western software version 3.1.7. (ProteinSimple, San Jose, CA).

Characterization of biome vectors in an *ex vivo* microbiome community

All animal studies were performed under an approved protocol by the Marshall University Institutional Animal Care and Use Committee. *Ex vivo* microbiome communities were derived from the luminal contents of the small intestines of C57BL/6J mice (B6) (Charles River Laboratories, Wilmington, MA). Mice were dissected under anaerobic conditions to maximize the viability of anaerobic microbes. The bacterial fraction of the luminal contents were isolated through a 100 µm vacuum-based filtration system (Millipore Sigma, Burlington, MA) under anaerobic conditions. The *ex vivo* microbiome community was co-cultivated with *E. coli* containing the Lux biome vector (~1:60 CFU/mL *E. coli* to gut flora ratio) on PRAS LB agar

containing 0.1% L-arabinose. The agar plate was incubated anaerobically for 48 hours and subsequently imaged with the IVIS Lumina XRMS III *in vivo* imager to measure bioluminescence expression.

Next, the *ex vivo* bacterial community was cultured alone or co-cultured with *E. coli* harboring the pHERD30T plasmid (~1:60 CFU/mL ratio) on the nutrient-rich, broad-range PRAS medium, YCFA agar with sheep blood (Anaerobe Systems, Morgan Hill, CA). Mono- and co-cultures were incubated anaerobically for 48 hours under normal (no antibiotic or carbohydrate supplementation) and experimental conditions, which consisted of supplementing the agar with Gentamicin (13 µg/mL) and 0.1% L-arabinose. All growth from the agar plates was scraped, and bacterial genomic DNA was subsequently isolated (Zymo Research, Irvine, CA). Using the StepOne Plus Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA), quantitative PCR (qPCR) was performed to measure the abundance of *Clostridium leptum*, *Lactobacillus*, *Bacteroides*, and Enterobacteriaceae relative to the total bacterial content (see Table 2 for primer sequences) in each sample. Fold changes were calculated with the $\Delta\Delta C_t$ method (Livak & Schmittgen, 2001).

In a follow-up study, we evaluated the *Lactobacillus* response to the experimental conditions of the *ex vivo* mono- and co-cultures. Using the same methods as described above, the *ex vivo* microbiome community was mono- and co-cultured with *E. coli* containing the pHERD30T vector (~1:60 CFU/mL *E. coli* to gut flora ratio) in 2 mL of PRAS regular LB and LB containing Gentamicin (13 µg/mL) or 0.1% L-arabinose. Broth cultures were anaerobically incubated for 5 hours, serially diluted, and plated on *Lactobacillus* Selection Agar (Becton Dickinson, Franklin Lake, NJ) to determine the colony-forming units (CFUs) of the surviving *Lactobacillus*.

In vivo mouse model

E. coli harboring the Lux biome vector, pUC18T-mini-Tn7T-Gm-*lux*, and pHERD30T vectors were grown overnight in 500 mL of LB containing Gentamicin (13 µg/mL) until the mid-log growth phase was achieved. Bacterial cultures were chilled on ice and centrifuged until a concentrated stock of 4×10^9 could be achieved in PBS (Thermo Fisher Scientific, Waltham, MA). 16-week old male and female B6 mice were orally gavaged via disposable oral gavage needles (Cadence Science, Cranston, RI) with 200 µl of the previously prepared bacterial stocks or PBS. Mice were divided into the following groups, each of which contained two male and two female mice: A.) subgroup was inoculated with *E. coli* harboring pUC18T-mini-Tn7T-Gm-*lux* and received regular drinking water; B.) subgroup was inoculated with *E. coli* harboring Lux biome vector and received drinking water with Gentamicin (13 µg/mL) and 2% L-arabinose; C.) subgroup was inoculated with *E. coli* harboring pHERD30T and received drinking water with Gentamicin (13 µg/mL) and 2% L-arabinose; D.) subgroup was inoculated with PBS and received drinking water with Gentamicin (13 µg/mL) and 2% L-arabinose. Mice were imaged with the IVIS Lumina XRMS III *in vivo* imager at 0 and 18 hours to track the colonization of the bioluminescence-labeled bacteria. All mice were monitored for adverse reactions for a week post-oral gavage.

Statistical Analyses

Living Image Software (PerkinElmer, Waltham, MA) was used to quantify bacterial bioluminescence and fluorescence expression. Protein quantification from immunoassays was computed on Compass (Protein Simple, San Jose, CA). Mann Whitney, two-way ANOVA, and Sidak or Dunnett multiple comparisons tests were performed using GraphPad Prism version 8.0

(GraphPad Software, La Jolla, CA). Prior to running statistical tests, the Shapiro-Wilk normality test was used to determine whether the data were consistent with a Gaussian distribution.

Acknowledgments

I would like to thank the NASA WVSGC for graciously funding my research and enabling me to pursue my scientific aspirations. I would also like to thank each of the following individuals for their indispensable kindness and help: my research mentor, Dr. Hongwei Yu, for his guidance and infectious enthusiasm for microbiology; Drs. Jeremy McAleer and Emine Koc for their generosity of time, protocols, and lab supplies; my lab mates, Roy Al Ahmar, Brandon Kirby, and Meagan Valentine for their daily help and support.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

This body of work aimed to investigate the complex microbial communities of the gut microbiota through the application of bioinformatics approaches, including machine-learning techniques to identify microbial features predictive of metabolic syndrome in TALLYHO/Jng mice and potentially their human analogs as well as the development of an informatics workflow for the *in silico* prediction of SFB antigens. During this work, we identified a gap between bioinformatics and the molecular tools that enable specific *in silico* targets to be experimentally investigated in a controlled manner, as to allow for causal inference. Therefore, a part of this work was dedicated to developing a series of bacterial vectors to enable the precise expression of microbial genes in *in vitro*, *ex vivo*, and potentially *in vivo* experiments.

Making the connection between specific microbial products, such as metabolites or antigens, and monitoring their effect on the host phenotype is particularly challenging to accomplish *in vivo* due to the complexity of the gut microbiota; however targeted manipulation of the gut microbiota is likely to be one of the pillars of precision-based personalized medicine, which is on the horizon. Thus, the continued development of the biome vector system, particularly the delivery vehicle, is likely a worthy endeavor. Additionally, the utility of the biome vector system in delineating the function of fastidious and unculturable microorganisms became increasingly evident throughout my dissertation research as we, like many others, struggled to establish an *in vitro* culture of SFB. Until the specific nutrient requirements for the *in vitro* growth of SFB are determined, combining computational and recombinant molecular approaches appears to be the best strategy for advancing the knowledge of SFB and other unculturable microorganisms.

At the time of beginning my dissertation research, two reports were published describing the successful cultivation of SFB *in vitro* (Ericsson et al., 2015; Schnupf et al., 2015). Despite many trials of these reported methods, including various modifications to the nutritional and growth conditions, we could not replicate the authors' results. At this point, we decided to employ multi-omics and bioinformatics to further investigate the intestinal niche in which SFB reside as well as the bacterium's antigenic properties.

We developed an *in silico* antigen prediction workflow, by which we identified thirty-five putative antigen candidates within the SFB reference proteome that exhibited CD4 T cell immunogenicity and binding affinity for Major Histocompatibility Complex Class II molecules. Among these predicted targets was flagellar cap protein, FliD, which we computationally predict is regulated by nutrient signals from the environment. Importantly, the real-time expression of FliD by SFB in the murine ileum was validated through shotgun proteomics of the murine ileal mucosa. In addition to and consistent with our proteomic analysis, results from our qPCR, next-generation sequencing consistently show that SFB preferentially colonize the mucosa proximal to the Peyer's Patches.

Based on this evidence, we postulate that a cell model that simulates the environment and expresses the receptors found at the Peyer's Patches may be conducive to SFB growth. One of the defining characteristics of Peyer's Patches is its presence of Microfold cells (M-cells) located on the apical surface of Peyer's Patches facing the intestinal lumen (Jung et al., 2010). M cells are specialized enterocytes that are directly involved in the transcytosis of antigens from the intestinal lumen into the Peyer's Patches (Mabbott, Donaldson, Ohno, Williams, & Mahajan, 2013). M-cell expressing Peyer's Patches cell models are well characterized and documented in the literature (Buda, Sands, & Jepson, 2005; Yeboah et al., 2014). The Peyer's Patches cell

model can be established by culturing the human colonic enterocyte cell line Caco-2, with the human B lymphocyte cell line, Raji-B (Buda et al., 2005; Yeboah et al., 2014). This model, with the addition CD11c⁺ antigen presenting cells could be useful to determine the mechanism by which SFB antigen(s) are recognized, presented, and processed. One consideration for this model is host-specificity SFB exhibits. For example, rat SFB weakly colonize the intestine and stimulate the immune system of mice (Atarashi et al., 2015). Thus human-derived SFB may be the most relevant SFB strain to use in this model and provide the best chance of *in vitro* growth.

Another important observation from this work was our finding that the gut microbiota's structure and function was primarily driven by dietary factors. The effect of diet overshadowed that of host genetics, despite the fact that TALLYHO/Jng mice harbor multiple genetic mutations rendering them susceptible to obesity, hypercholesteremia, hyperlipidemia, glucose intolerance, and type II diabetes (Denvir et al., 2016; Parkman et al., 2017; Parkman et al., 2016). To our knowledge, this is the first study to dissect the interplay of the microbiota, diet, host genetics using a polygenic model for metabolic syndrome. To take this work further, future studies in which the TALLYHO/Jng gut microbiota (i) is monitored during dietary oscillations between chow and Westernized diets, and (ii) transferred to determine whether the chow-fed gut microbiota can rescue the Western-diet induced metabolic phenotype TALLYHO/Jng mice exhibit.

Based on our collective findings, it would be interesting to determine what if any effect diet has on the Peyer's Patches-associated microbiota (PPAM). SFB are likely very sensitive to any metabolic changes that occur in the Peyer's Patches environment, either directly or indirectly from dietary modulations. Thus, additional research is warranted to investigate the interplay between dietary modulations, the PPAM, and mucosal immunity. Although we found the

composition of the PPAM to be extremely homogenous among mice, a larger sample size that includes PPAM samples obtained from mice of different dietary regimens, genetic backgrounds, and environments is needed to validate and extrapolate these findings more broadly to other mammals.

REFERENCES

- Aguilar, M., Bhuket, T., Torres, S., Liu, B., & Wong, R. J. (2015). Prevalence of the metabolic syndrome in the United States, 2003-2012. *JAMA*, *313*(19), 1973-1974. doi:10.1001/jama.2015.4260
- Alspach, E., Lussier, D. M., & Schreiber, R. D. (2019). Interferon gamma and Its Important Roles in Promoting and Inhibiting Spontaneous and Therapeutic Cancer Immunity. *Cold Spring Harb Perspect Biol*, *11*(3). doi:10.1101/cshperspect.a028480
- Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., & Stahl, D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol*, *56*(6), 1919-1925. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/2200342>
- Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P., Givskov, M., & Molin, S. (1998). New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol*, *64*(6), 2240-2246. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9603842>
- Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*, 32-46. doi:10.1111/j.1442-9993.2001.01070.pp.x
- Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). In *Wiley StatsRef: Statistics Reference Online* (pp. 1-15).
- Arora, S. K., Ritchings, B. W., Almira, E. C., Lory, S., & Ramphal, R. (1998). The *Pseudomonas aeruginosa* flagellar cap protein, FlhD, is responsible for mucin adhesion. *Infect Immun*, *66*(3), 1000-1007. doi:10.1128/IAI.66.3.1000-1007.1998
- Arredouani, M. S. (2014). New insights into androgenic immune regulation. *Oncoimmunology*, *3*(9), e954968. doi:10.4161/21624011.2014.954968
- Artis, D., & Grencis, R. K. (2008). The intestinal epithelium: sensors to effectors in nematode infection. *Mucosal Immunol*, *1*(4), 252-264. doi:10.1038/mi.2008.21
- Atarashi, K., Tanoue, T., Ando, M., Kamada, N., Nagano, Y., Narushima, S., . . . Honda, K. (2015). Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells. *Cell*, *163*(2), 367-380. doi:10.1016/j.cell.2015.08.058

- Azegami, T., Yuki, Y., & Kiyono, H. (2014). Challenges in mucosal vaccines for the control of infectious diseases. *Int Immunol*, *26*(9), 517-528. doi:10.1093/intimm/dxu063
- Backhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., . . . Gordon, J. I. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A*, *101*(44), 15718-15723. doi:10.1073/pnas.0407076101
- Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science*, *307*(5717), 1915-1920. doi:10.1126/science.1104816
- Backhed, F., Manchester, J. K., Semenkovich, C. F., & Gordon, J. I. (2007). Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc Natl Acad Sci U S A*, *104*(3), 979-984. doi:10.1073/pnas.0605374104
- Bai, J., Qin, Y., Liu, J., Wang, Y., Sa, R., Zhang, N., & Jia, R. (2017). Proteomic response of oat leaves to long-term salinity stress. *Environ Sci Pollut Res Int*, *24*(4), 3387-3399. doi:10.1007/s11356-016-8092-0
- Barbier, M., & Damron, F. H. (2016). Rainbow Vectors for Broad-Range Bacterial Fluorescence Labeling. *PLoS One*, *11*(3), e0146827. doi:10.1371/journal.pone.0146827
- Bartram, A. K., Lynch, M. D., Stearns, J. C., Moreno-Hagelsieb, G., & Neufeld, J. D. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol*, *77*(11), 3846-3852. doi:10.1128/AEM.02772-10
- Benson, A. K. (2015). Host genetic architecture and the landscape of microbiome composition: humans weigh in. *Genome Biol*, *16*, 203. doi:10.1186/s13059-015-0775-1
- Binda, C., Lopetuso, L. R., Rizzatti, G., Gibiino, G., Cennamo, V., & Gasbarrini, A. (2018). Actinobacteria: A relevant minority for the maintenance of gut homeostasis. *Dig Liver Dis*, *50*(5), 421-428. doi:10.1016/j.dld.2018.02.012
- Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., . . . Clark, A. G. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biol*, *16*, 191. doi:10.1186/s13059-015-0759-1

- Boesmans, L., Valles-Colomer, M., Wang, J., Eeckhaut, V., Falony, G., Ducatelle, R., . . . Verbeke, K. (2018). Butyrate Producers as Potential Next-Generation Probiotics: Safety Assessment of the Administration of *Butyricoccus pullicaecorum* to Healthy Volunteers. *mSystems*, 3(6). doi:10.1128/mSystems.00094-18
- Bokulich, N., Dillon, M., Bolyen, E., Kaehler, B. D., Huttley, G. A., & Caporaso, J. G. (2018). q2-sample-classifier: machine-learning tools for microbiome classification and regression. *bioRxiv*. doi:10.1101/306167
- Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., . . . Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, 6(1), 90. doi:10.1186/s40168-018-0470-z
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., . . . Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*, 37(8), 852-857. doi:10.1038/s41587-019-0209-9
- Bouchard, C. (2010). Defining the genetic architecture of the predisposition to obesity: a challenging but not insurmountable task. *Am J Clin Nutr*, 91(1), 5-6. doi:10.3945/ajcn.2009.28933
- Bouchard, C., Tremblay, A., Despres, J. P., Nadeau, A., Lupien, P. J., Theriault, G., . . . Fournier, G. (1990). The response to long-term overfeeding in identical twins. *N Engl J Med*, 322(21), 1477-1482. doi:10.1056/NEJM199005243222101
- Boulangé, C. L., Neves, A. L., Chilloux, J., Nicholson, J. K., & Dumas, M. E. (2016). Impact of the gut microbiota on inflammation, obesity, and metabolic disease. *Genome Med*, 8(1), 42. doi:10.1186/s13073-016-0303-2
- Buter, K. E., van Raalte, D. H., Groen, A. K., & Nieuwdorp, M. (2017). Role of the Gut Microbiome in the Pathogenesis of Obesity and Obesity-Related Metabolic Dysfunction. *Gastroenterology*, 152(7), 1671-1678. doi:10.1053/j.gastro.2016.12.048
- Bouxin, G. (2005). Ginkgo, a multivariate analysis package. *Journal of Vegetation Science*, 16, 355-359.

- Braniste, V., Al-Asmakh, M., Kowal, C., Anuar, F., Abbaspour, A., Toth, M., . . . Pettersson, S. (2014). The gut microbiota influences blood-brain barrier permeability in mice. *Sci Transl Med*, 6(263), 263ra158. doi:10.1126/scitranslmed.3009759
- Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 326-349. doi:10.2307/1942268
- Buda, A., Sands, C., & Jepson, M. A. (2005). Use of fluorescence imaging to investigate the structure and function of intestinal M cells. *Adv Drug Deliv Rev*, 57(1), 123-134. doi:10.1016/j.addr.2004.07.014
- Buettner, R., Scholmerich, J., & Bollheimer, L. C. (2007). High-fat diets: modeling the metabolic disorders of human obesity in rodents. *Obesity (Silver Spring)*, 15(4), 798-808. doi:10.1038/oby.2007.608
- Caesar, R., Tremaroli, V., Kovatcheva-Datchary, P., Cani, P. D., & Backhed, F. (2015). Crosstalk between Gut Microbiota and Dietary Lipids Aggravates WAT Inflammation through TLR Signaling. *Cell Metab*, 22(4), 658-668. doi:10.1016/j.cmet.2015.07.026
- Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J*, 11(12), 2639-2643. doi:10.1038/ismej.2017.119
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*, 13(7), 581-583. doi:10.1038/nmeth.3869
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 7(5), 335-336. doi:10.1038/nmeth.f.303
- Carmody, R. N., Gerber, G. K., Luevano, J. M., Jr., Gatti, D. M., Somes, L., Svenson, K. L., & Turnbaugh, P. J. (2015). Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe*, 17(1), 72-84. doi:10.1016/j.chom.2014.11.010
- Casqueiro, J., Casqueiro, J., & Alves, C. (2012). Infections in patients with diabetes mellitus: A review of pathogenesis. *Indian J Endocrinol Metab*, 16 Suppl 1, S27-36. doi:10.4103/2230-8210.94253

- Chan, C. W., Wong, R. S., Law, P. T., Wong, C. L., Tsui, S. K., Tang, W. P., & Sit, J. W. (2016). Environmental Factors Associated with Altered Gut Microbiota in Children with Eczema: A Systematic Review. *Int J Mol Sci*, *17*(7). doi:10.3390/ijms17071147
- Chantranupong, L., Wolfson, R. L., & Sabatini, D. M. (2015). Nutrient-sensing mechanisms across evolution. *Cell*, *161*(1), 67-83. doi:10.1016/j.cell.2015.02.041
- Chaplin, D. D. (2010). Overview of the immune response. *J Allergy Clin Immunol*, *125*(2 Suppl 2), S3-23. doi:10.1016/j.jaci.2009.12.980
- Chappert, P. (2014). Role of SFB in autoimmune arthritis: an example of regulation of autoreactive T cell sensitivity in the gut. *Gut Microbes*, *5*(2), 259-264. doi:10.4161/gmic.28134
- Chen, B., Chen, H., Shu, X., Yin, Y., Li, J., Qin, J., . . . Xiang, C. (2018). Presence of Segmented Filamentous Bacteria in Human Children and Its Potential Role in the Modulation of Human Gut Immunity. *Front Microbiol*, *9*, 1403. doi:10.3389/fmicb.2018.01403
- Choi, K. H., & Schweizer, H. P. (2006). mini-Tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*. *Nat Protoc*, *1*(1), 153-161. doi:10.1038/nprot.2006.24
- Chou, C. J., Membrez, M., & Blancher, F. (2008). Gut decontamination with norfloxacin and ampicillin enhances insulin sensitivity in mice. *Nestle Nutr Workshop Ser Pediatr Program*, *62*, 127-137; discussion 137-140. doi:10.1159/000146256
- Clemente, J. C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. *Cell*, *148*(6), 1258-1270. doi:10.1016/j.cell.2012.01.035
- Cockburn, A. F., Dehlin, J. M., Ngan, T., Crout, R., Boskovic, G., Denvir, J., . . . Cuff, C. F. (2012). High throughput DNA sequencing to detect differences in the subgingival plaque microbiome in elderly subjects with and without dementia. *Investig Genet*, *3*(1), 19. doi:10.1186/2041-2223-3-19
- Cordain, L., Eaton, S. B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B. A., . . . Brand-Miller, J. (2005). Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr*, *81*(2), 341-354. doi:10.1093/ajcn.81.2.341

- Cordain, L., Eaton, S. B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B. A., . . . Brand-Miller, J. (2005). Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr*, 81, 341-354.
- Corthesy, B. (2013). Role of secretory IgA in infection and maintenance of homeostasis. *Autoimmun Rev*, 12(6), 661-665. doi:10.1016/j.autrev.2012.10.012
- Cosorich, I., Dalla-Costa, G., Sorini, C., Ferrarese, R., Messina, M. J., Dolpady, J., . . . Falcone, M. (2017). High frequency of intestinal TH17 cells correlates with microbiota alterations and disease activity in multiple sclerosis. *Sci Adv*, 3(7), e1700492. doi:10.1126/sciadv.1700492
- Dahlhamer, J. M., Zammitti, E. P., Ward, B. W., Wheaton, A. G., & Croft, J. B. (2016). Prevalence of Inflammatory Bowel Disease Among Adults Aged ≥ 18 Years - United States, 2015. *MMWR Morb Mortal Wkly Rep*, 65(42), 1166-1169. doi:10.15585/mmwr.mm6542a3
- Davenport, E. R., Mizrahi-Man, O., Michelini, K., Barreiro, L. B., Ober, C., & Gilad, Y. (2014). Seasonal variation in human gut microbiome composition. *PLoS One*, 9(3), e90731. doi:10.1371/journal.pone.0090731
- David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., . . . Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484), 559-563. doi:10.1038/nature12820
- Davis, C. P., McAllister, J. S., & Savage, D. C. (1973). Microbial colonization of the intestinal epithelium in suckling mice. *Infect Immun*, 7(4), 666-672. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/4586864>
- De Caceres, M., Oliva, F., Font, X., & Vives, S. (2007). Ginkgo, a program for non-standard multivariate fuzzy analysis. *Advances in Fuzzy Sets and Systems. Advances in Fuzzy Sets and Systems*, 2(1), 41-46.
- De Filippis, F., Pellegrini, N., Vannini, L., Jeffery, I. B., La Storia, A., Laghi, L., . . . Ercolini, D. (2016). High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut*, 65(11), 1812-1821. doi:10.1136/gutjnl-2015-309957
- De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., . . . Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative

- study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A*, 107(33), 14691-14696. doi:10.1073/pnas.1005963107
- De Jesus, M., Ahlawat, S., & Mantis, N. J. (2013). Isolating and immunostaining lymphocytes and dendritic cells from murine Peyer's patches. *J Vis Exp*(73), e50167. doi:10.3791/50167
- Delzenne, N. M., Cani, P. D., Daubioul, C., & Neyrinck, A. M. (2005). Impact of inulin and oligofructose on gastrointestinal peptides. *Br J Nutr*, 93 Suppl 1, S157-161. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/15877889>
- Denvir, J., Boskovic, G., Fan, J., Primerano, D. A., Parkman, J. K., & Kim, J. H. (2016). Whole genome sequence analysis of the TALLYHO/Jng mouse. *BMC Genomics*, 17(1), 907. doi:10.1186/s12864-016-3245-6
- Doncheva, N. T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res*, 18(2), 623-632. doi:10.1021/acs.jproteome.8b00702
- Dore, J., Sghir, A., Hannequart-Gramet, G., Corthier, G., & Pochart, P. (1998). Design and evaluation of a 16S rRNA-targeted oligonucleotide probe for specific detection and quantitation of human faecal Bacteroides populations. *Syst Appl Microbiol*, 21(1), 65-71. doi:10.1016/S0723-2020(98)80009-X
- Douglas, G. M., Maffei, V. J., Zaneveld, J., Yurgel, S. N., Brown, J. R., Taylor, C. M., . . . Langille, M. G. I. (2019). PICRUSt2: An improved and extensible approach for metagenome inference. *bioRxiv*. doi:10.1101/672295
- Dugar, G., Svensson, S. L., Bischler, T., Waldchen, S., Reinhardt, R., Sauer, M., & Sharma, C. M. (2016). The CsrA-FliW network controls polar localization of the dual-function flagellin mRNA in *Campylobacter jejuni*. *Nat Commun*, 7, 11667. doi:10.1038/ncomms11667
- Ebrahimi, S., & Mohabatkar, H. (2018). Prediction of T-cell epitopes for designing a reverse vaccine against streptococcal bacteria. *Mol Biol Res Commun*, 7(1), 35-41. doi:10.22099/mbrc.2018.28775.1308
- Ericsson, A. C., Hagan, C. E., Davis, D. J., & Franklin, C. L. (2014). Segmented filamentous bacteria: commensal microbes with potential effects on research. *Comp Med*, 64(2), 90-98. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/24674582>

- Ericsson, A. C., Turner, G., Montoya, L., Wolfe, A., Meeker, S., Hsu, C., . . . Franklin, C. L. (2015). Isolation of segmented filamentous bacteria from complex gut microbiota. *Biotechniques*, *59*(2), 94-98. doi:10.2144/000114319
- Espin, J. C., Larrosa, M., Garcia-Conesa, M. T., & Tomas-Barberan, F. (2013). Biological significance of urolithins, the gut microbial ellagic Acid-derived metabolites: the evidence so far. *Evid Based Complement Alternat Med*, *2013*, 270418. doi:10.1155/2013/270418
- Faith, J. J., Colombel, J. F., & Gordon, J. I. (2015). Identifying strains that contribute to complex diseases through the study of microbial inheritance. *Proc Natl Acad Sci U S A*, *112*(3), 633-640. doi:10.1073/pnas.1418781112
- Festa, F., Steel, J., Bian, X., & Labaer, J. (2013). High-throughput cloning and expression library creation for functional proteomics. *Proteomics*, *13*(9), 1381-1399. doi:10.1002/pmic.201200456
- Finkelstein, E. A., Khavjou, O. A., Thompson, H., Trogdon, J. G., Pan, L., Sherry, B., & Dietz, W. (2012). Obesity and severe obesity forecasts through 2030. *Am J Prev Med*, *42*(6), 563-570. doi:10.1016/j.amepre.2011.10.026
- Flannigan, K. L., & Denning, T. L. (2018). Segmented filamentous bacteria-induced immune responses: a balancing act between host protection and autoimmunity. *Immunology*. doi:10.1111/imm.12950
- Flo, T. H., Halaas, O., Torp, S., Ryan, L., Lien, E., Dybdahl, B., . . . Espevik, T. (2001). Differential expression of Toll-like receptor 2 in human cells. *J Leukoc Biol*, *69*(3), 474-481. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11261796>
- Fogelholm, M., Anderssen, S., Gunnarsdottir, I., & Lahti-Koski, M. (2012). Dietary macronutrients and food consumption as determinants of long-term weight change in adult populations: a systematic literature review. *Food Nutr Res*, *56*. doi:10.3402/fnr.v56i0.19103
- Forbes, J. D., Van Domselaar, G., & Bernstein, C. N. (2016). The Gut Microbiota in Immune-Mediated Inflammatory Diseases. *Front Microbiol*, *7*, 1081. doi:10.3389/fmicb.2016.01081
- Freitag, C. M., Strijbis, K., & van Putten, J. P. M. (2017). Host cell binding of the flagellar tip protein of *Campylobacter jejuni*. *Cell Microbiol*, *19*(6). doi:10.1111/cmi.12714

- Fuchs, T., Loureiro, M. P., Macedo, L. E., Nocca, D., Nedelcu, M., & Costa-Casagrande, T. A. (2018). Animal models in metabolic syndrome. *Rev Col Bras Cir*, *45*(5), e1975. doi:10.1590/0100-6991e-20181975
- Fujisaka, S., Avila-Pacheco, J., Soto, M., Kostic, A., Dreyfuss, J. M., Pan, H., . . . Kahn, C. R. (2018). Diet, Genetics, and the Gut Microbiome Drive Dynamic Changes in Plasma Metabolites. *Cell Rep*, *22*(11), 3072-3086. doi:10.1016/j.celrep.2018.02.060
- Furet, J. P., Firmesse, O., Gourmelon, M., Bridonneau, C., Tap, J., Mondot, S., . . . Corthier, G. (2009). Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. *FEMS Microbiol Ecol*, *68*(3), 351-362. doi:10.1111/j.1574-6941.2009.00671.x
- Gaffen, S. L., Jain, R., Garg, A. V., & Cua, D. J. (2014). The IL-23-IL-17 immune axis: from mechanisms to therapeutic testing. *Nat Rev Immunol*, *14*(9), 585-600. doi:10.1038/nri3707
- Gallo, R. L., & Hooper, L. V. (2012). Epithelial antimicrobial defence of the skin and intestine. *Nat Rev Immunol*, *12*(7), 503-516. doi:10.1038/nri3228
- Gao, F., Li, M., Liu, Y., Gao, C., Wen, S., & Tang, L. (2012). Intestinal dysbacteriosis induces changes of T lymphocyte subpopulations in Peyer's patches of mice and orients the immune response towards humoral immunity. *Gut Pathog*, *4*(1), 19. doi:10.1186/1757-4749-4-19
- Gardner, C., Wylie-Rosett, J., Gidding, S. S., Steffen, L. M., Johnson, R. K., Reader, D., . . . American Diabetes, A. (2012). Nonnutritive sweeteners: current use and health perspectives: a scientific statement from the American Heart Association and the American Diabetes Association. *Diabetes Care*, *35*(8), 1798-1808. doi:10.2337/dc12-9002
- Gauguet, S., D'Ortona, S., Ahnger-Pier, K., Duan, B., Surana, N. K., Lu, R., . . . Pier, G. B. (2015). Intestinal Microbiota of Mice Influences Resistance to Staphylococcus aureus Pneumonia. *Infect Immun*, *83*(10), 4003-4014. doi:10.1128/IAI.00037-15
- Ge, Z., Feng, Y., Woods, S. E., & Fox, J. G. (2015). Spatial and temporal colonization dynamics of segmented filamentous bacteria is influenced by gender, age and experimental infection with *Helicobacter hepaticus* in Swiss Webster mice. *Microbes Infect*, *17*(1), 16-22. doi:10.1016/j.micinf.2014.10.005

- Geem, D., Medina-Contreras, O., McBride, M., Newberry, R. D., Koni, P. A., & Denning, T. L. (2014). Specific microbiota-induced intestinal Th17 differentiation requires MHC class II but not GALT and mesenteric lymph nodes. *J Immunol*, *193*(1), 431-438. doi:10.4049/jimmunol.1303167
- Gilbert, J. A., Quinn, R. A., Debelius, J., Xu, Z. Z., Morton, J., Garg, N., . . . Knight, R. (2016). Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature*, *535*(7610), 94-103. doi:10.1038/nature18850
- Global, B. M. I. M. C., Di Angelantonio, E., Bhupathiraju Sh, N., Wormser, D., Gao, P., Kaptoge, S., . . . Hu, F. B. (2016). Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet*, *388*(10046), 776-786. doi:10.1016/S0140-6736(16)30175-1
- Gobbetti, M., Lavermicocca, P., Minervini, F., de Angelis, M., & Corsetti, A. (2000). Arabinose fermentation by *Lactobacillus plantarum* in sourdough with added pentosans and alphaalpha-L-arabinofuranosidase: a tool to increase the production of acetic acid. *J Appl Microbiol*, *88*(2), 317-324. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10736001>
- Goto, Y., Panea, C., Nakato, G., Cebula, A., Lee, C., Diez, M. G., . . . Ivanov, II. (2014). Segmented filamentous bacteria antigens presented by intestinal dendritic cells drive mucosal Th17 cell differentiation. *Immunity*, *40*(4), 594-607. doi:10.1016/j.immuni.2014.03.005
- Gregor, C., Gwosch, K. C., Sahl, S. J., & Hell, S. W. (2018). Strongly enhanced bacterial bioluminescence with the lux operon for single-cell imaging. *Proc Natl Acad Sci U S A*, *115*(5), 962-967. doi:10.1073/pnas.1715946115
- Grundy, S. M., Brewer, H. B., Jr., Cleeman, J. I., Smith, S. C., Jr., Lenfant, C., American Heart, A., . . . Blood, I. (2004). Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation*, *109*(3), 433-438. doi:10.1161/01.CIR.0000111245.75752.C6
- Haegeman, B., Hamelin, J., Moriarty, J., Neal, P., Dushoff, J., & Weitz, J. S. (2013). Robust estimation of microbial diversity in theory and in practice. *ISME J*, *7*(6), 1092-1101. doi:10.1038/ismej.2013.10

- Hales, C. M., Carroll, M. G., Fryer, C. D., & Ogden, C. L. (2017). *Prevalence of Obesity Among Adults and Youth: United States, 2015-2016* (288). Retrieved from Hyattsville (MD): <https://www.cdc.gov/nchs/data/databriefs/db288.pdf>.
- Hall, A. B., Tolonen, A. C., & Xavier, R. J. (2017). Human genetic variation and the gut microbiome in disease. *Nat Rev Genet*, *18*(11), 690-699. doi:10.1038/nrg.2017.63
- Hammond, R. A., & Levine, R. (2010). The economic impact of obesity in the United States. *Diabetes Metab Syndr Obes*, *3*, 285-295. doi:10.2147/DMSOTT.S7384
- Hapfelmeier, S., Lawson, M. A., Slack, E., Kirundi, J. K., Stoel, M., Heikenwalder, M., . . . Macpherson, A. J. (2010). Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science*, *328*(5986), 1705-1709. doi:10.1126/science.1188454
- Harley, I. T., Stankiewicz, T. E., Giles, D. A., Softic, S., Flick, L. M., Cappelletti, M., . . . Divanovic, S. (2014). IL-17 signaling accelerates the progression of nonalcoholic fatty liver disease in mice. *Hepatology*, *59*(5), 1830-1839. doi:10.1002/hep.26746
- Harrison, O. J., & Powrie, F. M. (2013). Regulatory T cells and immune tolerance in the intestine. *Cold Spring Harb Perspect Biol*, *5*(7). doi:10.1101/cshperspect.a018341
- Hashiguchi, M., Kashiwakura, Y., Kojima, H., Kobayashi, A., Kanno, Y., & Kobata, T. (2015). Peyer's patch innate lymphoid cells regulate commensal bacteria expansion. *Immunol Lett*, *165*(1), 1-9. doi:10.1016/j.imlet.2015.03.002
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). The Elements of Statistical Learning. In *The Elements of Statistical Learning* (pp. 219-260). New York (NY): Springer New York Inc.
- He, G., Shankar, R. A., Chzhan, M., Samouilov, A., Kuppusamy, P., & Zweier, J. L. (1999). Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. *Proc Natl Acad Sci U S A*, *96*(8), 4586-4591. doi:10.1073/pnas.96.8.4586
- He, Y., Xiang, Z., & Mobley, H. L. (2010). Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J Biomed Biotechnol*, *2010*, 297505. doi:10.1155/2010/297505

- Henderson, G., Yilmaz, P., Kumar, S., Forster, R. J., Kelly, W. J., Leahy, S. C., . . . Janssen, P. H. (2019). Improved taxonomic assignment of rumen bacterial 16S rRNA sequences using a revised SILVA taxonomic framework. *PeerJ*, 7, e6496. doi:10.7717/peerj.6496
- Herrera, B. M., Keildson, S., & Lindgren, C. M. (2011). Genetics and epigenetics of obesity. *Maturitas*, 69(1), 41-49. doi:10.1016/j.maturitas.2011.02.018
- Hildebrand, F., Nguyen, T. L., Brinkman, B., Yunta, R. G., Cauwe, B., Vandenabeele, P., . . . Raes, J. (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol*, 14(1), R4. doi:10.1186/gb-2013-14-1-r4
- Hinney, A., Vogel, C. I., & Hebebrand, J. (2010). From monogenic to polygenic obesity: recent advances. *Eur Child Adolesc Psychiatry*, 19(3), 297-310. doi:10.1007/s00787-010-0096-6
- Hirota, K., Ahlfors, H., Duarte, J. H., & Stockinger, B. (2012). Regulation and function of innate and adaptive interleukin-17-producing cells. *EMBO Rep*, 13(2), 113-120. doi:10.1038/embor.2011.248
- Holmes, E., Li, J. V., Marchesi, J. R., & Nicholson, J. K. (2012). Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab*, 16(5), 559-564. doi:10.1016/j.cmet.2012.10.007
- Hooper, L. V., Littman, D. R., & Macpherson, A. J. (2012). Interactions between the microbiota and the immune system. *Science*, 336(6086), 1268-1273. doi:10.1126/science.1223490
- Hooper, L. V., & Macpherson, A. J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat Rev Immunol*, 10(3), 159-169. doi:10.1038/nri2710
- Hueber, W., Sands, B. E., Lewitzky, S., Vandemeulebroecke, M., Reinisch, W., Higgins, P. D., . . . Secukinumab in Crohn's Disease Study, G. (2012). Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut*, 61(12), 1693-1700. doi:10.1136/gutjnl-2011-301668
- Hughes, E. R., Winter, M. G., Duerkop, B. A., Spiga, L., Furtado de Carvalho, T., Zhu, W., . . . Winter, S. E. (2017). Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. *Cell Host Microbe*, 21(2), 208-219. doi:10.1016/j.chom.2017.01.005

- Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207-214. doi:10.1038/nature11234
- Hwang, I., Park, Y. J., Kim, Y. R., Kim, Y. N., Ka, S., Lee, H. Y., . . . Kim, J. B. (2015). Alteration of gut microbiota by vancomycin and bacitracin improves insulin resistance via glucagon-like peptide 1 in diet-induced obesity. *FASEB J*, 29(6), 2397-2411. doi:10.1096/fj.14-265983
- Hwang, N., Eom, T., Gupta, S. K., Jeong, S. Y., Jeong, D. Y., Kim, Y. S., . . . Unno, T. (2017). Genes and Gut Bacteria Involved in Luminal Butyrate Reduction Caused by Diet and Loperamide. *Genes (Basel)*, 8(12). doi:10.3390/genes8120350
- Hylemon, P. B., Harris, S. C., & Ridlon, J. M. (2018). Metabolism of hydrogen gases and bile acids in the gut microbiome. *FEBS Lett*, 592(12), 2070-2082. doi:10.1002/1873-3468.13064
- Islam, K. B., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., . . . Yokota, A. (2011). Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology*, 141(5), 1773-1781. doi:10.1053/j.gastro.2011.07.046
- Ivanov, II, Atarashi, K., Manel, N., Brodie, E. L., Shima, T., Karaoz, U., . . . Littman, D. R. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*, 139(3), 485-498. doi:10.1016/j.cell.2009.09.033
- Janeiro, M. H., Ramirez, M. J., Milagro, F. I., Martinez, J. A., & Solas, M. (2018). Implication of Trimethylamine N-Oxide (TMAO) in Disease: Potential Biomarker or New Therapeutic Target. *Nutrients*, 10(10). doi:10.3390/nu10101398
- Janssen, S., McDonald, D., Gonzalez, A., Navas-Molina, J. A., Jiang, L., Xu, Z. Z., . . . Knight, R. (2018). Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems*, 3(3). doi:10.1128/mSystems.00021-18
- Jie, Z., Xia, H., Zhong, S. L., Feng, Q., Li, S., Liang, S., . . . Kristiansen, K. (2017). The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun*, 8(1), 845. doi:10.1038/s41467-017-00900-1
- Jung, C., Hugot, J. P., & Barreau, F. (2010). Peyer's Patches: The Immune Sensors of the Intestine. *Int J Inflam*, 2010, 823710. doi:10.4061/2010/823710

- Just, S., Mondot, S., Ecker, J., Wegner, K., Rath, E., Gau, L., . . . Clavel, T. (2018). The gut microbiota drives the impact of bile acids and fat source in diet on mouse metabolism. *Microbiome*, 6(1), 134. doi:10.1186/s40168-018-0510-8
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1), 27-30. doi:10.1093/nar/28.1.27
- Kawamoto, S., Maruya, M., Kato, L. M., Suda, W., Atarashi, K., Doi, Y., . . . Fagarasan, S. (2014). Foxp3(+) T cells regulate immunoglobulin a selection and facilitate diversification of bacterial species responsible for immune homeostasis. *Immunity*, 41(1), 152-165. doi:10.1016/j.immuni.2014.05.016
- Kennedy, A., Martinez, K., Chuang, C. C., LaPoint, K., & McIntosh, M. (2009). Saturated fatty acid-mediated inflammation and insulin resistance in adipose tissue: mechanisms of action and implications. *J Nutr*, 139(1), 1-4. doi:10.3945/jn.108.098269
- Kim, J. H., & Saxton, A. M. (2012). The TALLYHO mouse as a model of human type 2 diabetes. *Methods Mol Biol*, 933, 75-87. doi:10.1007/978-1-62703-068-7_6
- Kim, M. S., Hwang, S. S., Park, E. J., & Bae, J. W. (2013). Strict vegetarian diet improves the risk factors associated with metabolic diseases by modulating gut microbiota and reducing intestinal inflammation. *Environ Microbiol Rep*, 5(5), 765-775. doi:10.1111/1758-2229.12079
- Kimura, S., Yamakami-Kimura, M., Obata, Y., Hase, K., Kitamura, H., Ohno, H., & Iwanaga, T. (2015). Visualization of the entire differentiation process of murine M cells: suppression of their maturation in cecal patches. *Mucosal Immunol*, 8(3), 650-660. doi:10.1038/mi.2014.99
- Kitamoto, S., Alteri, C. J., Rodrigues, M., Nagao-Kitamoto, H., Sugihara, K., Himpl, S. D., . . . Kamada, N. (2020). Dietary L-serine confers a competitive fitness advantage to Enterobacteriaceae in the inflamed gut. *Nat Microbiol*, 5(1), 116-125. doi:10.1038/s41564-019-0591-6
- Klaasen, H. L., Koopman, J. P., Poelma, F. G., & Beynen, A. C. (1992). Intestinal, segmented, filamentous bacteria. *FEMS Microbiol Rev*, 8(3-4), 165-180. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/1515159>

- Klaasen, H. L., Koopman, J. P., Van den Brink, M. E., Bakker, M. H., Poelma, F. G., & Beynen, A. C. (1993). Intestinal, segmented, filamentous bacteria in a wide range of vertebrate species. *Lab Anim*, *27*(2), 141-150. doi:10.1258/002367793780810441
- Klaasen, H. L., Van der Heijden, P. J., Stok, W., Poelma, F. G., Koopman, J. P., Van den Brink, M. E., . . . Beynen, A. C. (1993). Apathogenic, intestinal, segmented, filamentous bacteria stimulate the mucosal immune system of mice. *Infect Immun*, *61*(1), 303-306. doi:10.1128/IAI.61.1.303-306.1993
- Knights, D., Silverberg, M. S., Weersma, R. K., Gevers, D., Dijkstra, G., Huang, H., . . . Xavier, R. J. (2014). Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med*, *6*(12), 107. doi:10.1186/s13073-014-0107-1
- Koliada, A., Syzenko, G., Moseiko, V., Budovska, L., Puchkov, K., Perederiy, V., . . . Vaiserman, A. (2017). Association between body mass index and Firmicutes/Bacteroidetes ratio in an adult Ukrainian population. *BMC Microbiol*, *17*(1), 120. doi:10.1186/s12866-017-1027-1
- Kreznar, J. H., Keller, M. P., Traeger, L. L., Rabaglia, M. E., Schueler, K. L., Stapleton, D. S., . . . Rey, F. E. (2017). Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes. *Cell Rep*, *18*(7), 1739-1750. doi:10.1016/j.celrep.2017.01.062
- Kumar, P., Monin, L., Castillo, P., Elsegeiny, W., Horne, W., Eddens, T., . . . Kolls, J. K. (2016). Intestinal Interleukin-17 Receptor Signaling Mediates Reciprocal Control of the Gut Microbiota and Autoimmune Inflammation. *Immunity*, *44*(3), 659-671. doi:10.1016/j.immuni.2016.02.007
- Kuno, T., Hirayama-Kurogi, M., Ito, S., & Ohtsuki, S. (2018). Reduction in hepatic secondary bile acids caused by short-term antibiotic-induced dysbiosis decreases mouse serum glucose and triglyceride levels. *Sci Rep*, *8*(1), 1253. doi:10.1038/s41598-018-19545-1
- Kuwahara, T., Ogura, Y., Oshima, K., Kurokawa, K., Ooka, T., Hirakawa, H., . . . Hayashi, T. (2011). The lifestyle of the segmented filamentous bacterium: a non-culturable gut-associated immunostimulating microbe inferred by whole-genome sequencing. *DNA Res*, *18*(4), 291-303. doi:10.1093/dnares/dsr022
- Lagkouvardos, I., Lesker, T. R., Hitch, T. C. A., Galvez, E. J. C., Smit, N., Neuhaus, K., . . . Clavel, T. (2019). Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome*, *7*(1), 28. doi:10.1186/s40168-019-0637-2

- Lai, N. Y., Musser, M. A., Pinho-Ribeiro, F. A., Baral, P., Jacobson, A., Ma, P., . . . Chiu, I. M. (2020). Gut-Innervating Nociceptor Neurons Regulate Peyer's Patch Microfold Cells and SFB Levels to Mediate Salmonella Host Defense. *Cell*, *180*(1), 33-49 e22. doi:10.1016/j.cell.2019.11.014
- Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., . . . Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*, *31*(9), 814-821. doi:10.1038/nbt.2676
- Larsen, N., Vogensen, F. K., van den Berg, F. W., Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., . . . Jakobsen, M. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One*, *5*(2), e9085. doi:10.1371/journal.pone.0009085
- Lassenius, M. I., Pietilainen, K. H., Kaartinen, K., Pussinen, P. J., Syrjanen, J., Forsblom, C., . . . FinnDiane Study, G. (2011). Bacterial endotoxin activity in human serum is associated with dyslipidemia, insulin resistance, obesity, and chronic inflammation. *Diabetes Care*, *34*(8), 1809-1815. doi:10.2337/dc10-2197
- Laukens, D., Brinkman, B. M., Raes, J., De Vos, M., & Vandenabeele, P. (2016). Heterogeneity of the gut microbiome in mice: guidelines for optimizing experimental design. *FEMS Microbiol Rev*, *40*(1), 117-132. doi:10.1093/femsre/fuv036
- Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., . . . Pedersen, O. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature*, *500*(7464), 541-546. doi:10.1038/nature12506
- Lecuyer, E., Rakotobe, S., Lengline-Garnier, H., Lebreton, C., Picard, M., Juste, C., . . . Gaboriau-Routhiau, V. (2014). Segmented filamentous bacterium uses secondary and tertiary lymphoid tissues to induce gut IgA and specific T helper 17 cell responses. *Immunity*, *40*(4), 608-620. doi:10.1016/j.immuni.2014.03.009
- Lee, H. L., Shen, H., Hwang, I. Y., Ling, H., Yew, W. S., Lee, Y. S., & Chang, M. W. (2018). Targeted Approaches for In Situ Gut Microbiome Manipulation. *Genes (Basel)*, *9*(7). doi:10.3390/genes9070351
- Leong, K. S. W., Derraik, J. G. B., Hofman, P. L., & Cutfield, W. S. (2018). Antibiotics, gut microbiome and obesity. *Clin Endocrinol (Oxf)*, *88*(2), 185-200. doi:10.1111/cen.13495
- Levine, M. E., Suarez, J. A., Brandhorst, S., Balasubramanian, P., Cheng, C. W., Madia, F., . . . Longo, V. D. (2014). Low protein intake is associated with a major reduction in IGF-1,

- cancer, and overall mortality in the 65 and younger but not older population. *Cell Metab*, 19(3), 407-417. doi:10.1016/j.cmet.2014.02.006
- Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A*, 102(31), 11070-11075. doi:10.1073/pnas.0504978102
- Ley, R. E., Peterson, D. A., & Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, 124(4), 837-848. doi:10.1016/j.cell.2006.02.017
- Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Microbial ecology: human gut microbes associated with obesity. *Nature*, 444(7122), 1022-1023. doi:10.1038/4441022a
- Liisberg, U., Fauske, K. R., Kuda, O., Fjaere, E., Myrmel, L. S., Norberg, N., . . . Madsen, L. (2016). Intake of a Western diet containing cod instead of pork alters fatty acid composition in tissue phospholipids and attenuates obesity and hepatic lipid accumulation in mice. *J Nutr Biochem*, 33, 119-127. doi:10.1016/j.jnutbio.2016.03.014
- Liisberg, U., Myrmel, L. S., Fjaere, E., Ronnevik, A. K., Bjelland, S., Fauske, K. R., . . . Madsen, L. (2016). The protein source determines the potential of high protein diets to attenuate obesity development in C57BL/6J mice. *Adipocyte*, 5(2), 196-211. doi:10.1080/21623945.2015.1122855
- Linden, D. R. (2014). Hydrogen sulfide signaling in the gastrointestinal tract. *Antioxid Redox Signal*, 20(5), 818-830. doi:10.1089/ars.2013.5312
- Liu, F., Wang, X., Shi, H., Wang, Y., Xue, C., & Tang, Q. J. (2017). Polymannuronic acid ameliorated obesity and inflammation associated with a high-fat and high-sucrose diet by modulating the gut microbiome in a murine model. *Br J Nutr*, 117(9), 1332-1342. doi:10.1017/S0007114517000964
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4), 402-408. doi:10.1006/meth.2001.1262
- Livanos, A. E., Greiner, T. U., Vangay, P., Pathmasiri, W., Stewart, D., McRitchie, S., . . . Blaser, M. J. (2016). Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. *Nat Microbiol*, 1(11), 16140. doi:10.1038/nmicrobiol.2016.140

- Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol*, 73(5), 1576-1585. doi:10.1128/AEM.01996-06
- Ludwig, W. (2007). Nucleic acid techniques in bacterial systematics and identification. *Int J Food Microbiol*, 120(3), 225-236. doi:10.1016/j.ijfoodmicro.2007.06.023
- Lycke, N. (2012). Recent progress in mucosal vaccine development: potential and limitations. *Nat Rev Immunol*, 12(8), 592-605. doi:10.1038/nri3251
- Lynn, M. A., Tumes, D. J., Choo, J. M., Sribnaia, A., Blake, S. J., Leong, L. E. X., . . . Lynn, D. J. (2018). Early-Life Antibiotic-Driven Dysbiosis Leads to Dysregulated Vaccine Immune Responses in Mice. *Cell Host Microbe*, 23(5), 653-660 e655. doi:10.1016/j.chom.2018.04.009
- Mabbott, N. A., Donaldson, D. S., Ohno, H., Williams, I. R., & Mahajan, A. (2013). Microfold (M) cells: important immunosurveillance posts in the intestinal epithelium. *Mucosal Immunol*, 6(4), 666-677. doi:10.1038/mi.2013.30
- Macpherson, A. J., & McCoy, K. D. (2015). Standardised animal models of host microbial mutualism. *Mucosal Immunol*, 8(3), 476-486. doi:10.1038/mi.2014.113
- Macpherson, A. J., & Uhr, T. (2004). Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science*, 303(5664), 1662-1665. doi:10.1126/science.1091334
- MacQueen, J. C. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(Statistics), 281-297.
- Magnusson, K. R., Hauck, L., Jeffrey, B. M., Elias, V., Humphrey, A., Nath, R., . . . Bermudez, L. E. (2015). Relationships between diet-related changes in the gut microbiome and cognitive flexibility. *Neuroscience*, 300, 128-140. doi:10.1016/j.neuroscience.2015.05.016
- Mao, X., Dillon, K. D., McEntee, M. F., Saxton, A. M., & Kim, J. H. (2014). Islet Insulin Secretion, β -Cell Mass, and Energy Balance in a Polygenic Mouse Model of Type 2 Diabetes With Obesity. *Journal of Inborn Errors of Metabolism and Screening*, 2. doi:10.1177/2326409814528153

- Martinez, I., Lattimer, J. M., Hubach, K. L., Case, J. A., Yang, J., Weber, C. G., . . . Walter, J. (2013). Gut microbiome composition is linked to whole grain-induced immunological improvements. *ISME J*, 7(2), 269-280. doi:10.1038/ismej.2012.104
- Maslowski, K. M., & Mackay, C. R. (2011). Diet, gut microbiota and immune responses. *Nat Immunol*, 12(1), 5-9. doi:10.1038/ni0111-5
- Matey-Hernandez, M. L., Williams, F. M. K., Potter, T., Valdes, A. M., Spector, T. D., & Menni, C. (2018). Genetic and microbiome influence on lipid metabolism and dyslipidemia. *Physiol Genomics*, 50(2), 117-126. doi:10.1152/physiolgenomics.00053.2017
- Maynard, C. L., & Weaver, C. T. (2009). Intestinal effector T cells in health and disease. *Immunity*, 31(3), 389-400. doi:10.1016/j.immuni.2009.08.012
- McAleer, J. P., Nguyen, N. L., Chen, K., Kumar, P., Ricks, D. M., Binnie, M., . . . Kolls, J. K. (2016). Pulmonary Th17 Antifungal Immunity Is Regulated by the Gut Microbiome. *J Immunol*, 197(1), 97-107. doi:10.4049/jimmunol.1502566
- Meunier, M., Guyard-Nicodeme, M., Hirschaud, E., Parra, A., Chemaly, M., & Dory, D. (2016). Identification of Novel Vaccine Candidates against *Campylobacter* through Reverse Vaccinology. *J Immunol Res*, 2016, 5715790. doi:10.1155/2016/5715790
- Moreno-Indias, I., Cardona, F., Tinahones, F. J., & Queipo-Ortuno, M. I. (2014). Impact of the gut microbiota on the development of obesity and type 2 diabetes mellitus. *Front Microbiol*, 5, 190. doi:10.3389/fmicb.2014.00190
- Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., . . . Huttenhower, C. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol*, 13(9), R79. doi:10.1186/gb-2012-13-9-r79
- Muller, P., Gimpel, M., Wildenhain, T., & Brantl, S. (2019). A new role for CsrA: promotion of complex formation between an sRNA and its mRNA target in *Bacillus subtilis*. *RNA Biol*, 16(7), 972-987. doi:10.1080/15476286.2019.1605811
- Murdoch, C. C., Espenschied, S. T., Matty, M. A., Mueller, O., Tobin, D. M., & Rawls, J. F. (2019). Intestinal Serum amyloid A suppresses systemic neutrophil activation and bactericidal activity in response to microbiota colonization. *PLoS Pathog*, 15(3), e1007381. doi:10.1371/journal.ppat.1007381

- Nelson, G. M., Guynn, J. M., & Chorley, B. N. (2017). Procedure and Key Optimization Strategies for an Automated Capillary Electrophoretic-based Immunoassay Method. *J Vis Exp*(127). doi:10.3791/55911
- Nesvizhskii, A. I., Keller, A., Kolker, E., & Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*, 75(17), 4646-4658. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/14632076>
- Nguyen, K. T., Seth, A. K., Hong, S. J., Geringer, M. R., Xie, P., Leung, K. P., . . . Galiano, R. D. (2013). Deficient cytokine expression and neutrophil oxidative burst contribute to impaired cutaneous wound healing in diabetic, biofilm-containing chronic wounds. *Wound Repair Regen*, 21(6), 833-841. doi:10.1111/wrr.12109
- Nkamba, I., Mulet, C., Dubey, G. P., Gorgette, O., Couesnon, A., Salles, A., . . . Schnupf, P. (2020). Intracellular offspring released from SFB filaments are flagellated. *Nat Microbiol*, 5(1), 34-39. doi:10.1038/s41564-019-0608-1
- Omenetti, S., Bussi, C., Metidji, A., Iseppon, A., Lee, S., Tolaini, M., . . . Stockinger, B. (2019). The Intestine Harbors Functionally Distinct Homeostatic Tissue-Resident and Inflammatory Th17 Cells. *Immunity*, 51(1), 77-89 e76. doi:10.1016/j.immuni.2019.05.004
- Org, E., Parks, B. W., Joo, J. W., Emert, B., Schwartzman, W., Kang, E. Y., . . . Lusic, A. J. (2015). Genetic and environmental control of host-gut microbiota interactions. *Genome Res*, 25(10), 1558-1569. doi:10.1101/gr.194118.115
- Oshiro, R. T., Rajendren, S., Hundley, H. A., & Kearns, D. B. (2019). Robust Stoichiometry of FliW-CsrA Governs Flagellin Homeostasis and Cytoplasmic Organization in *Bacillus subtilis*. *MBio*, 10(3). doi:10.1128/mBio.00533-19
- Pamp, S. J., Harrington, E. D., Quake, S. R., Relman, D. A., & Blainey, P. C. (2012). Single-cell sequencing provides clues about the host interactions of segmented filamentous bacteria (SFB). *Genome Res*, 22(6), 1107-1119. doi:10.1101/gr.131482.111
- Parada Venegas, D., De la Fuente, M. K., Landskron, G., Gonzalez, M. J., Quera, R., Dijkstra, G., . . . Hermoso, M. A. (2019). Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Front Immunol*, 10, 277. doi:10.3389/fimmu.2019.00277

- Parkman, J. K., Denvir, J., Mao, X., Dillon, K. D., Romero, S., Saxton, A. M., & Kim, J. H. (2017). Congenic mice demonstrate the presence of QTLs conferring obesity and hypercholesterolemia on chromosome 1 in the TALLYHO mouse. *Mamm Genome*, 28(11-12), 487-497. doi:10.1007/s00335-017-9719-2
- Parkman, J. K., Mao, X., Dillon, K., Gudivada, A., Moustaid-Moussa, N., Saxton, A. M., & Kim, J. H. (2016). Genotype-dependent Metabolic Responses to Semi-Purified High-Sucrose High-Fat Diets in the TALLYHO/Jng vs. C57BL/6 Mouse during the Development of Obesity and Type 2 Diabetes. *Exp Clin Endocrinol Diabetes*, 124(10), 622-629. doi:10.1055/s-0042-109605
- Parks, B. W., Nam, E., Org, E., Kostem, E., Norheim, F., Hui, S. T., . . . Lusic, A. J. (2013). Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab*, 17(1), 141-152. doi:10.1016/j.cmet.2012.12.007
- Peleg, A. Y., Seifert, H., & Paterson, D. L. (2008). *Acinetobacter baumannii*: emergence of a successful pathogen. *Clin Microbiol Rev*, 21(3), 538-582. doi:10.1128/CMR.00058-07
- Peters, B. A., Shapiro, J. A., Church, T. R., Miller, G., Trinh-Shevrin, C., Yuen, E., . . . Ahn, J. (2018). A taxonomic signature of obesity in a large study of American adults. *Sci Rep*, 8(1), 9749. doi:10.1038/s41598-018-28126-1
- Poeker, S. A., Geirnaert, A., Berchtold, L., Greppi, A., Krych, L., Steinert, R. E., . . . Lacroix, C. (2018). Understanding the prebiotic potential of different dietary fibers using an in vitro continuous adult fermentation model (PolyFermS). *Sci Rep*, 8(1), 4318. doi:10.1038/s41598-018-22438-y
- Prakash, T., Oshima, K., Morita, H., Fukuda, S., Imaoka, A., Kumar, N., . . . Hattori, M. (2011). Complete genome sequences of rat and mouse segmented filamentous bacteria, a potent inducer of th17 cell differentiation. *Cell Host Microbe*, 10(3), 273-284. doi:10.1016/j.chom.2011.08.007
- Qin, C., Qiu, K., Sun, W., Jiao, N., Zhang, X., Che, L., . . . Yin, J. (2016). A proteomic adaptation of small intestinal mucosa in response to dietary protein limitation. *Sci Rep*, 6, 36888. doi:10.1038/srep36888
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., . . . Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, 490(7418), 55-60. doi:10.1038/nature11450

- Qiu, D., Damron, F. H., Mima, T., Schweizer, H. P., & Yu, H. D. (2008). PBAD-based shuttle vectors for functional analysis of toxic and highly regulated genes in *Pseudomonas* and *Burkholderia* spp. and other bacteria. *Appl Environ Microbiol*, *74*(23), 7422-7426. doi:10.1128/AEM.01369-08
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*, *41*(Database issue), D590-596. doi:10.1093/nar/gks1219
- Rabot, S., Membrez, M., Blancher, F., Berger, B., Moine, D., Krause, L., . . . Chou, C. J. (2016). High fat diet drives obesity regardless the composition of gut microbiota in mice. *Sci Rep*, *6*, 32484. doi:10.1038/srep32484
- Rabot, S., Membrez, M., Bruneau, A., Gerard, P., Harach, T., Moser, M., . . . Chou, C. J. (2010). Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism. *FASEB J*, *24*(12), 4948-4959. doi:10.1096/fj.10-164921
- Rahnavard, G., Franzosa, E. A., McIver, L. J., Schwager, E., Lloyd-Price, J., Weingarten, J., . . . Huttenhower, C. (2019). *High-sensitivity pattern discovery in large multi'omic datasets*. Harvard University. Retrieved from huttenhower.sph.harvard.edu/halla
- Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., & Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell*, *118*(2), 229-241. doi:10.1016/j.cell.2004.07.002
- Rappuoli, R. (2000). Reverse vaccinology. *Curr Opin Microbiol*, *3*(5), 445-450. doi:10.1016/s1369-5274(00)00119-3
- Reimann, H. A. (1965). Microbic Phagocytosis by Enteric Epithelial Cells. *JAMA*, *192*, 1130-1132. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/14301031>
- Rinttila, T., Kassinen, A., Malinen, E., Krogus, L., & Palva, A. (2004). Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol*, *97*(6), 1166-1177. doi:10.1111/j.1365-2672.2004.02409.x
- Rios-Covian, D., Salazar, N., Gueimonde, M., & de Los Reyes-Gavilan, C. G. (2017). Shaping the Metabolism of Intestinal Bacteroides Population through Diet to Improve Human Health. *Front Microbiol*, *8*, 376. doi:10.3389/fmicb.2017.00376

- Rodou, A., Ankrah, D. O., & Stathopoulos, C. (2010). Toxins and secretion systems of *Photobacterium luminescens*. *Toxins (Basel)*, 2(6), 1250-1264. doi:10.3390/toxins2061250
- Ronda, C., Chen, S. P., Cabral, V., Yaung, S. J., & Wang, H. H. (2019). Metagenomic engineering of the mammalian gut microbiome in situ. *Nat Methods*, 16(2), 167-170. doi:10.1038/s41592-018-0301-y
- Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., . . . Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, 555(7695), 210-215. doi:10.1038/nature25973
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.
- Saely, C. H., Geiger, K., & Drexel, H. (2012). Brown versus white adipose tissue: a mini-review. *Gerontology*, 58(1), 15-23. doi:10.1159/000321319
- Saier, M. H., Jr. (2015). The Bacterial Phosphotransferase System: New Frontiers 50 Years after Its Discovery. *J Mol Microbiol Biotechnol*, 25(2-3), 73-78. doi:10.1159/000381215
- Salzman, N. H., Hung, K., Haribhai, D., Chu, H., Karlsson-Sjoberg, J., Amir, E., . . . Bos, N. A. (2010). Enteric defensins are essential regulators of intestinal microbial ecology. *Nat Immunol*, 11(1), 76-83. doi:10.1038/ni.1825
- Sampaio, S. C. F., Luiz, W. B., Vieira, M. A. M., Ferreira, R. C. C., Garcia, B. G., Sinigaglia-Coimbra, R., . . . Gomes, T. A. T. (2016). Flagellar Cap Protein FliD Mediates Adherence of Atypical Enteropathogenic *Escherichia coli* to Enterocyte Microvilli. *Infect Immun*, 84(4), 1112-1122. doi:10.1128/IAI.01001-15
- Sanchez-Osuna, M., Barbe, J., & Erill, I. (2017). Comparative genomics of the DNA damage-inducible network in the *Patescibacteria*. *Environ Microbiol*, 19(9), 3465-3474. doi:10.1111/1462-2920.13826
- Savage, D. C. (1969). Localization of certain indigenous microorganisms on the ileal villi of rats. *J Bacteriol*, 97(3), 1505-1506. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/5776537>
- Scalaferrri, F., Gerardi, V., Mangiola, F., Lopetuso, L. R., Pizzoferrato, M., Petito, V., . . . Gasbarrini, A. (2016). Role and mechanisms of action of *Escherichia coli* Nissle 1917 in

the maintenance of remission in ulcerative colitis patients: An update. *World J Gastroenterol*, 22(24), 5505-5511. doi:10.3748/wjg.v22.i24.5505

- Schnupf, P., Gaboriau-Routhiau, V., & Cerf-Bensussan, N. (2013). Host interactions with Segmented Filamentous Bacteria: an unusual trade-off that drives the post-natal maturation of the gut immune system. *Semin Immunol*, 25(5), 342-351. doi:10.1016/j.smim.2013.09.001
- Schnupf, P., Gaboriau-Routhiau, V., Gros, M., Friedman, R., Moya-Nilges, M., Nigro, G., . . . Sansonetti, P. J. (2015). Growth and host interaction of mouse segmented filamentous bacteria in vitro. *Nature*, 520(7545), 99-103. doi:10.1038/nature14027
- Schwartz, A., Taras, D., Schafer, K., Beijer, S., Bos, N. A., Donus, C., & Hardt, P. D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity (Silver Spring)*, 18(1), 190-195. doi:10.1038/oby.2009.167
- Sczesnak, A., Segata, N., Qin, X., Gevers, D., Petrosino, J. F., Huttenhower, C., . . . Ivanov, II. (2011). The genome of th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe*, 10(3), 260-272. doi:10.1016/j.chom.2011.08.005
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol*, 12(6), R60. doi:10.1186/gb-2011-12-6-r60
- Selma, M. V., Beltran, D., Luna, M. C., Romo-Vaquero, M., Garcia-Villalba, R., Mira, A., . . . Tomas-Barberan, F. A. (2017). Isolation of Human Intestinal Bacteria Capable of Producing the Bioactive Metabolite Isourolithin A from Ellagic Acid. *Front Microbiol*, 8, 1521. doi:10.3389/fmicb.2017.01521
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Champaign, Illinois: University of Illinois Press.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11), 2498-2504. doi:10.1101/gr.1239303
- Sheik, C. S., Reese, B. K., Twing, K. I., Sylvan, J. B., Grim, S. L., Schrenk, M. O., . . . Colwell, F. S. (2018). Identification and Removal of Contaminant Sequences From Ribosomal

Gene Databases: Lessons From the Census of Deep Life. *Front Microbiol*, 9, 840.
doi:10.3389/fmicb.2018.00840

- Shi, Z., Zou, J., Zhang, Z., Zhao, X., Noriega, J., Zhang, B., . . . Gewirtz, A. T. (2019). Segmented Filamentous Bacteria Prevent and Cure Rotavirus Infection. *Cell*, 179(3), 644-658 e613. doi:10.1016/j.cell.2019.09.028
- Shih, V. F., Cox, J., Kljavin, N. M., Dengler, H. S., Reichelt, M., Kumar, P., . . . Ghilardi, N. (2014). Homeostatic IL-23 receptor signaling limits Th17 response through IL-22-mediated containment of commensal microbiota. *Proc Natl Acad Sci U S A*, 111(38), 13942-13947. doi:10.1073/pnas.1323852111
- Shih, Y. K., & Parthasarathy, S. (2012). Identifying functional modules in interaction networks through overlapping Markov clustering. *Bioinformatics*, 28(18), i473-i479. doi:10.1093/bioinformatics/bts370
- Shin, N. R., Whon, T. W., & Bae, J. W. (2015). Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol*, 33(9), 496-503. doi:10.1016/j.tibtech.2015.06.011
- Smith, B. J., Miller, R. A., Ericsson, A. C., Harrison, D. C., Strong, R., & Schmidt, T. M. (2019). Changes in the gut microbiome and fermentation products concurrent with enhanced longevity in acarbose-treated mice. *BMC Microbiol*, 19(1), 130. doi:10.1186/s12866-019-1494-7
- Smith, J. D., Hou, T., Ludwig, D. S., Rimm, E. B., Willett, W., Hu, F. B., & Mozaffarian, D. (2015). Changes in intake of protein foods, carbohydrate amount and quality, and long-term weight change: results from 3 prospective cohorts. *Am J Clin Nutr*, 101(6), 1216-1224. doi:10.3945/ajcn.114.100867
- Smits, S. A., Marcobal, A., Higginbottom, S., Sonnenburg, J. L., & Kashyap, P. C. (2016). Individualized Responses of Gut Microbiota to Dietary Intervention Modeled in Humanized Mice. *mSystems*, 1(5). doi:10.1128/mSystems.00098-16
- Speare, L., & Septer, A. N. (2019). Coincubation Assay for Quantifying Competitive Interactions between *Vibrio fischeri* Isolates. *J Vis Exp*(149). doi:10.3791/59759
- Stein-Thoeringer, C. K., Nichols, K. B., Lazrak, A., Docampo, M. D., Slingerland, A. E., Slingerland, J. B., . . . van den Brink, M. R. M. (2019). Lactose drives *Enterococcus* expansion to promote graft-versus-host disease. *Science*, 366(6469), 1143-1149. doi:10.1126/science.aax3760

- Suarez-Zamorano, N., Fabbiano, S., Chevalier, C., Stojanovic, O., Colin, D. J., Stevanovic, A., . . . Trajkovski, M. (2015). Microbiota depletion promotes browning of white adipose tissue and reduces obesity. *Nat Med*, *21*(12), 1497-1501. doi:10.1038/nm.3994
- Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C. A., Maza, O., . . . Elinav, E. (2014). Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature*, *514*(7521), 181-186. doi:10.1038/nature13793
- Szeliöva, D., Krahulec, J., Safranek, M., Liskova, V., & Turna, J. (2016). Modulation of heterologous expression from PBAD promoter in *Escherichia coli* production strains. *J Biotechnol*, *236*, 1-9. doi:10.1016/j.jbiotec.2016.08.004
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., . . . Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, *47*(D1), D607-D613. doi:10.1093/nar/gky1131
- Taboada, B., Ciria, R., Martinez-Guerrero, C. E., & Merino, E. (2012). ProOpDB: Prokaryotic Operon DataBase. *Nucleic Acids Res*, *40*(Database issue), D627-631. doi:10.1093/nar/gkr1020
- Talham, G. L., Jiang, H. Q., Bos, N. A., & Cebra, J. J. (1999). Segmented filamentous bacteria are potent stimuli of a physiologically normal state of the murine gut mucosal immune system. *Infect Immun*, *67*(4), 1992-2000. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10085047>
- Tang, Y., Underwood, A., Gielbert, A., Woodward, M. J., & Petrovska, L. (2014). Metaproteomics analysis reveals the adaptation process for the chicken gut microbiota. *Appl Environ Microbiol*, *80*(2), 478-485. doi:10.1128/AEM.02472-13
- Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H., & Karjalainen, T. (2001). Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun*, *69*(12), 7937-7940. doi:10.1128/IAI.69.12.7937-7940.2001
- Tesmer, L. A., Lundy, S. K., Sarkar, S., & Fox, D. A. (2008). Th17 cells in human disease. *Immunol Rev*, *223*, 87-113. doi:10.1111/j.1600-065X.2008.00628.x
- Thaiss, C. A., Levy, M., Grosheva, I., Zheng, D., Soffer, E., Blacher, E., . . . Elinav, E. (2018). Hyperglycemia drives intestinal barrier dysfunction and risk for enteric infection. *Science*, *359*(6382), 1376-1383. doi:10.1126/science.aar3318

- Thompson, C. L., Vier, R., Mikaelyan, A., Wienemann, T., & Brune, A. (2012). 'Candidatus Arthromitus' revised: segmented filamentous bacteria in arthropod guts are members of Lachnospiraceae. *Environ Microbiol*, *14*(6), 1454-1465. doi:10.1111/j.1462-2920.2012.02731.x
- Tong, M., Jacobs, J. P., McHardy, I. H., & Braun, J. (2014). Sampling of intestinal microbiota and targeted amplification of bacterial 16S rRNA genes for microbial ecologic analysis. *Curr Protoc Immunol*, *107*, 7 41 41-11. doi:10.1002/0471142735.im0741s107
- Troy, T., Jekic-McMullen, D., Sambucetti, L., & Rice, B. (2004). Quantitative comparison of the sensitivity of detection of fluorescent and bioluminescent reporters in animal models. *Mol Imaging*, *3*(1), 9-23. doi:10.1162/153535004773861688
- Turnbaugh, P. J., Backhed, F., Fulton, L., & Gordon, J. I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe*, *3*(4), 213-223. doi:10.1016/j.chom.2008.02.015
- Turnbaugh, P. J., Hamady, M., Yatsunencko, T., Cantarel, B. L., Duncan, A., Ley, R. E., . . . Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480-484. doi:10.1038/nature07540
- Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027-1031. doi:10.1038/nature05414
- Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., & Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med*, *1*(6), 6ra14. doi:10.1126/scitranslmed.3000322
- Turner, J. R. (2009). Intestinal mucosal barrier function in health and disease. *Nat Rev Immunol*, *9*(11), 799-809. doi:10.1038/nri2653
- UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, *47*(D1), D506-D515. doi:10.1093/nar/gky1049
- Upadhyay, V., Poroyko, V., Kim, T. J., Devkota, S., Fu, S., Liu, D., . . . Fu, Y. X. (2012). Lymphotoxin regulates commensal responses to enable diet-induced obesity. *Nat Immunol*, *13*(10), 947-953. doi:10.1038/ni.2403

- Ussar, S., Griffin, N. W., Bezy, O., Fujisaka, S., Vienberg, S., Softic, S., . . . Kahn, C. R. (2015). Interactions between Gut Microbiota, Host Genetics and Diet Modulate the Predisposition to Obesity and Metabolic Syndrome. *Cell Metab*, 22(3), 516-530. doi:10.1016/j.cmet.2015.07.007
- Vazquez-Baeza, Y., Pirrung, M., Gonzalez, A., & Knight, R. (2013). EMPERor: a tool for visualizing high-throughput microbial community data. *Gigascience*, 2(1), 16. doi:10.1186/2047-217X-2-16
- Velagapudi, V. R., Hezaveh, R., Reigstad, C. S., Gopalacharyulu, P., Yetukuri, L., Islam, S., . . . Backhed, F. (2010). The gut microbiota modulates host energy and lipid metabolism in mice. *J Lipid Res*, 51(5), 1101-1112. doi:10.1194/jlr.M002774
- Vita, N., Valette, O., Brasseur, G., Lignon, S., Denis, Y., Ansaldi, M., . . . Pieulle, L. (2015). The primary pathway for lactate oxidation in *Desulfovibrio vulgaris*. *Front Microbiol*, 6, 606. doi:10.3389/fmicb.2015.00606
- Wahba, I. M., & Mak, R. H. (2007). Obesity and obesity-initiated metabolic syndrome: mechanistic links to chronic kidney disease. *Clin J Am Soc Nephrol*, 2(3), 550-562. doi:10.2215/CJN.04071206
- Waidmann, M. S., Bleichrodt, F. S., Laslo, T., & Riedel, C. U. (2011). Bacterial luciferase reporters: the Swiss army knife of molecular biology. *Bioeng Bugs*, 2(1), 8-16. doi:10.4161/bbug.2.1.13566
- Waite, D. W., Vanwonterghem, I., Rinke, C., Parks, D. H., Zhang, Y., Takai, K., . . . Hugenholtz, P. (2017). Comparative Genomic Analysis of the Class Epsilonproteobacteria and Proposed Reclassification to Epsilonbacteraeota (phyl. nov.). *Front Microbiol*, 8, 682. doi:10.3389/fmicb.2017.00682
- Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., . . . Knight, R. (2016). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems*, 1(1). doi:10.1128/mSystems.00009-15
- Wang, B., Yao, M., Lv, L., Ling, Z., & Li, L. (2017). The Human Microbiota in Health and Disease. *Engineering*, 3(1), 71-82. doi:10.1016/j.Eng.2017.01.008

- Wang, Y., Yin, Y., Chen, X., Zhao, Y., Wu, Y., Li, Y., . . . Xiang, C. (2019). Induction of Intestinal Th17 Cells by Flagellins From Segmented Filamentous Bacteria. *Front Immunol*, *10*, 2750. doi:10.3389/fimmu.2019.02750
- Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., . . . Stevens, R. L. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res*, *45*(D1), D535-D542. doi:10.1093/nar/gkw1017
- Williams, E. P., Mesidor, M., Winters, K., Dubbert, P. M., & Wyatt, S. B. (2015). Overweight and Obesity: Prevalence, Consequences, and Causes of a Growing Public Health Problem. *Curr Obes Rep*, *4*(3), 363-370. doi:10.1007/s13679-015-0169-4
- Winter, S. E., & Baumler, A. J. (2014). Dysbiosis in the inflamed intestine: chance favors the prepared microbe. *Gut Microbes*, *5*(1), 71-73. doi:10.4161/gmic.27129
- Withers, D. R., & Hepworth, M. R. (2017). Group 3 Innate Lymphoid Cells: Communications Hubs of the Intestinal Immune System. *Front Immunol*, *8*, 1298. doi:10.3389/fimmu.2017.01298
- Wolfe, A. E., Moskowitz, J. E., Franklin, C. L., Wiemken, T. L., & Ericsson, A. C. (2020). Interactions of Segmented Filamentous Bacteria (*Candidatus Savagella*) and bacterial drivers in colitis-associated colorectal cancer development. *PLoS One*, *15*(7), e0236595. doi:10.1371/journal.pone.0236595
- Wu, G. D., & Lewis, J. D. (2013). Analysis of the human gut microbiome and association with disease. *Clin Gastroenterol Hepatol*, *11*(7), 774-777. doi:10.1016/j.cgh.2013.03.038
- Wu, H. J., & Wu, E. (2012). The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes*, *3*(1), 4-14. doi:10.4161/gmic.19320
- Yamada, S., Takashina, Y., Watanabe, M., Nagamine, R., Saito, Y., Kamada, N., & Saito, H. (2018). Bile acid metabolism regulated by the gut microbiota promotes non-alcoholic steatohepatitis-associated hepatocellular carcinoma in mice. *Oncotarget*, *9*(11), 9925-9939. doi:10.18632/oncotarget.24066
- Yang, Y., Torchinsky, M. B., Gobert, M., Xiong, H., Xu, M., Linehan, J. L., . . . Littman, D. R. (2014). Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature*, *510*(7503), 152-156. doi:10.1038/nature13279

- Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F. O., Ludwig, W., Schleifer, K. H., . . . Rossello-Mora, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*, *12*(9), 635-645. doi:10.1038/nrmicro3330
- Yatsunenkov, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., . . . Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222-227. doi:10.1038/nature11053
- Yeboah, K. G., Akande, J., Addo, R. T., Siwale, R. C., Aninkorah-Yeboah, K., & Siddig, A. (2014). In vitro and ex vivo characterization of lectin-labeled Mycobacterium tuberculosis antigen-containing microspheres for enhanced oral delivery. *J Drug Target*, *22*(1), 34-47. doi:10.3109/1061186X.2013.833206
- Yi, J., Jung, J., Han, D., Surh, C. D., & Lee, Y. J. (2019). Segmented Filamentous Bacteria Induce Divergent Populations of Antigen-Specific CD4 T Cells in the Small Intestine. *Mol Cells*, *42*(3), 228-236. doi:10.14348/molcells.2018.0424
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., . . . Glockner, F. O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res*, *42*(Database issue), D643-648. doi:10.1093/nar/gkt1209
- Yin, Y., Wang, Y., Zhu, L., Liu, W., Liao, N., Jiang, M., . . . Wang, X. (2013). Comparative analysis of the distribution of segmented filamentous bacteria in humans, mice and chickens. *ISME J*, *7*(3), 615-621. doi:10.1038/ismej.2012.128
- Zaharieva, N., Dimitrov, I., Flower, D. R., & Doytchinova, I. (2019). VaxiJen Dataset of Bacterial Immunogens: An Update. *Curr Comput Aided Drug Des*, *15*(5), 398-400. doi:10.2174/1573409915666190318121838
- Zeisel, S. H., & Warrier, M. (2017). Trimethylamine N-Oxide, the Microbiome, and Heart and Kidney Disease. *Annu Rev Nutr*, *37*, 157-181. doi:10.1146/annurev-nutr-071816-064732
- Zeng, M. Y., Inohara, N., & Nunez, G. (2017). Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol*, *10*(1), 18-26. doi:10.1038/mi.2016.75
- Zhang, C., Zhang, M., Pang, X., Zhao, Y., Wang, L., & Zhao, L. (2012). Structural resilience of the gut microbiota in adult mice under high-fat dietary perturbations. *ISME J*, *6*(10), 1848-1857. doi:10.1038/ismej.2012.27

Zhao, S., Liu, W., Wang, J., Shi, J., Sun, Y., Wang, W., . . . Hong, J. (2017). *Akkermansia muciniphila* improves metabolic profiles by reducing inflammation in chow diet-fed mice. *J Mol Endocrinol*, *58*(1), 1-14. doi:10.1530/JME-16-0054

Zhu, L., Baker, S. S., Gill, C., Liu, W., Alkhouri, R., Baker, R. D., & Gill, S. R. (2013). Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology*, *57*(2), 601-609. doi:10.1002/hep.26093

Zhu, Y., Lin, X., Zhao, F., Shi, X., Li, H., Li, Y., . . . Zhou, G. (2015). Meat, dairy and plant proteins alter bacterial composition of rat gut bacteria. *Sci Rep*, *5*, 15220. doi:10.1038/srep15220

APPENDIX A: OFFICE OF RESEARCH INTEGRITY APPROVAL LETTER



Office of Research Integrity

August 3, 2022

Lexie Blalock
1395 Wakefield Court E
Columbus, OH 43209

Dear Lexie:

This letter is in response to the submitted dissertation abstract entitled "*A Systems Approach to Dissecting the Role of the Mucosal Microbiome in Disease.*" After assessing the abstract, it has been deemed not to be human subject research and therefore exempt from oversight of the Marshall University Institutional Review Board (IRB). The Institutional Animal Care and Use Committee (IACUC) has reviewed and approved the study under protocols #508, #622, and #672. The applicable human and animal federal regulations have set forth the criteria utilized in making this determination. If there are any changes to the abstract you provided then you would need to resubmit that information to the Office of Research Integrity for review and a determination.

I appreciate your willingness to submit the abstract for determination. Please feel free to contact the Office of Research Integrity if you have any questions regarding future protocols that may require IRB review.

Sincerely,

A handwritten signature in blue ink that reads 'Bruce F. Day'.

Bruce F. Day, ThD, CIP
Director
Office of Research Integrity

WE ARE... MARSHALL.

One John Marshall Drive • Huntington, West Virginia 25755 • Tel 304/696-4303
A State University of West Virginia • An Affirmative Action/Equal Opportunity Employer

APPENDIX B: CHARACTERISTICS AND PEPTIDE SEQUENCES OF PREDICTED SFB ANTIGENS

Protein Description	Peptide	Start	End	CD4 Immuno. Score	Peptide Core (processed by MHCII)	Median Percentile Rank (7 allele)	MHCII Cleavage Probability	Localization	Localization Probability	Adhesin Probability	Antigenicity Score
SFBM_0042	NYPHINSINEYLQI	226	240	76.2256	INISINEYL	8.5	0.50764	Unknown	0.515	0.437	0.5061
SFBM_0042	LDIRYASKNNFTHKV	71	85	69.507	LDIRYASKN	27	0.40056	Unknown	0.515	0.437	0.5061
SFBM_0042	YIQKILWDIVPNSNY	126	140	75.7338	IQKILWDIV	11	0.17165	Unknown	0.515	0.437	0.5061
SFBM_0042	AQDKKIMIKNFENT	31	45	76.3623	IMIKNFENT	30	0.14003	Unknown	0.515	0.437	0.5061
SFBM_0042	NISINEYLQIMNNKL	231	245	74.5291	YLQIMNNKL	12	0.12413	Unknown	0.515	0.437	0.5061
SFBM_0042	KIWDAYRPFYQKI	116	130	90.7013	KIWDAYRPF	18	0.0919	Unknown	0.515	0.437	0.5061
SFBM_0042	KNMIKHGFKPIYTE	201	215	75.8072	IMIKHGFKP	25	0.07926	Unknown	0.515	0.437	0.5061
SFBM_0042	NYDVSINGLRLLSNF	51	65	85.6903	INGLRLLSNF	16	0.07906	Unknown	0.515	0.437	0.5061
SFBM_0042	HHEMILDIRYASKNN	66	80	76.3828	MILDIRYAS	19	0.07087	Unknown	0.515	0.437	0.5061
SFBM_0042	NEYLQIMNNKLRGGE	235	249	53.6694	LQIMNNKLR	8.3	0.04908	Unknown	0.515	0.437	0.5061
SFBM_0042	INGLRLLSNFHHEMI	56	70	69.6189	LRLLSNFHH	6.8	0.04297	Unknown	0.515	0.437	0.5061
SFBM_0042	KIMIKNFENTKEYSY	36	50	57.5323	IKNFENTKE	42	0.04227	Unknown	0.515	0.437	0.5061
SFBM_0042	KEYSYNYDVSINGLR	46	60	92.7954	YNYDVSING	20	0.03949	Unknown	0.515	0.437	0.5061
SFBM_0042	KISIIISFGIILIS	6	20	86.1861	IIISFGIIL	9.3	0.03917	Unknown	0.515	0.437	0.5061
SFBM_0042	MNFINKISIIISFG	1	15	91.0139	FINKISIII	8.7	0.03455	Unknown	0.515	0.437	0.5061
SFBM_0042	ILSFGIILISGGLYY	11	25	90.9799	FGIILISGG	15	0.03455	Unknown	0.515	0.437	0.5061
SFBM_0042	NAELLKNMIKHGFK	196	210	57.6864	LLKNMIKH	25	0.03297	Unknown	0.515	0.437	0.5061
SFBM_0204	IDILASNGTVVIPEV	166	180	84.5382	ILASNGTVV	20	0.22785	Extracellular	0.913	0.718	0.6851
SFBM_0204	AKSKLTNLKAYVVRA	196	210	90.3239	KSKLTNLKA	19	0.20659	Extracellular	0.913	0.718	0.6851
SFBM_0204	NNTFGDYVLAHSLGV	1331	1345	97.5318	YVLAHSLGV	12	0.18321	Extracellular	0.913	0.718	0.6851
SFBM_0204	EGYNYWYAHMKPKT	1801	1815	76.7591	YNYWYAHRM	30	0.16403	Extracellular	0.913	0.718	0.6851
SFBM_0204	FEFLVLFNGIKSKEA	21	35	85.5078	LVLNFGIKS	7.6	0.15175	Extracellular	0.913	0.718	0.6851
SFBM_0204	ATQNLNSVKRATIKM	611	625	81.2501	LNSVKRATI	26	0.1448	Extracellular	0.913	0.718	0.6851
SFBM_0204	GVKIKLNKSMRDVFS	346	360	75.6907	VKIKLNKSM	24	0.13376	Extracellular	0.913	0.718	0.6851
SFBM_0204	DLIAALYQIVVNREY	1841	1855	73.9568	YQIVVNREY	21	0.12993	Extracellular	0.913	0.718	0.6851
SFBM_0204	SVKITWSYPSGYIPA	441	455	74.1436	VKITWSYPS	26	0.11794	Extracellular	0.913	0.718	0.6851
SFBM_0204	YELNITLNTNNSVKK	106	120	63.8018	LNTNNSVKK	24	0.11127	Extracellular	0.913	0.718	0.6851
SFBM_0204	MGKTRKFRSLAFL	1	15	83.4155	GKTRKFRS	4	0.0993	Extracellular	0.913	0.718	0.6851
SFBM_0204	DASLYIYYAGSNLGS	266	280	84.181	YIYYAGSNL	22	0.0947	Extracellular	0.913	0.718	0.6851
SFBM_0204	NKFLETGVIIDILA	156	170	98.9751	NKFLETGVI	13	0.08973	Extracellular	0.913	0.718	0.6851
SFBM_0204	GNVDIVSVYYNVAS	251	265	86.4174	VSVYYNVAS	17	0.08755	Extracellular	0.913	0.718	0.6851
SFBM_0204	FGQSFYYPINSSAS	421	435	85.092	FYYPINSS	23	0.08123	Extracellular	0.913	0.718	0.6851
SFBM_0204	TNLKAYVVRARLRES	201	215	79.9677	YVVRARLRE	19	0.07854	Extracellular	0.913	0.718	0.6851

SFBM_0204	KIFEYVYQGNWYSR	731	745	85.615	YVYQGNWNY	22	0.07701	Extracellular	0.913	0.718	0.6851
SFBM_0204	NYHVDLIYTIGGKQI	631	645	87.6875	YHVDLIYTI	18	0.06737	Extracellular	0.913	0.718	0.6851
SFBM_0204	LLYMKTFKASYLSAS	1576	1590	51.9536	LLYMKTFKA	6.2	0.06216	Extracellular	0.913	0.718	0.6851
SFBM_0204	DVSVYYNVASDASLY	256	270	76.4328	YYNVASDAS	24	0.061	Extracellular	0.913	0.718	0.6851
SFBM_0204	VEFIKENTKTSRSD	1031	1045	79.323	VEFIKENT	26	0.05902	Extracellular	0.913	0.718	0.6851
SFBM_0204	LKNLGVNKDYTLKVR	1096	1110	79.4046	VNKDYTLKV	26	0.0576	Extracellular	0.913	0.718	0.6851
SFBM_0204	NKVVIYELNDNLGKT	366	380	88.5649	YELNDNLGK	13	0.0567	Extracellular	0.913	0.718	0.6851
SFBM_0204	YDLGGLLYMKTFKAS	1571	1585	68.4395	LLYMKTFKA	12	0.05272	Extracellular	0.913	0.718	0.6851
SFBM_0204	GDLRTFKSLDLKNLG	1086	1100	90.0235	LRTFKSLDL	9.6	0.04915	Extracellular	0.913	0.718	0.6851
SFBM_0204	LAFFLCFMFIFEFLV	11	25	88.4784	FMFIFEFLV	11	0.04007	Extracellular	0.913	0.718	0.6851
SFBM_0204	IKLEQLKYNEALISW	1481	1495	87.4376	LKYNEALIS	21	0.03834	Extracellular	0.913	0.718	0.6851
SFBM_0204	DLQIVGLNTSNVTKT	996	1010	78.3127	IVGLNTSNV	19	0.03831	Extracellular	0.913	0.718	0.6851
SFBM_0204	VPVVKADGNNLKITG	136	150	84.7918	VVKADGNNL	17	0.03804	Extracellular	0.913	0.718	0.6851
SFBM_0204	GHLVKVHLRKNLSLN	706	720	77.508	VKVHLRKNV	22	0.03633	Extracellular	0.913	0.718	0.6851
SFBM_0204	MVFNPADTLKIFVKP	1291	1305	93.4854	MVFNPADTL	13	0.03503	Extracellular	0.913	0.718	0.6851
SFBM_0204	AGSTAYFNFKTEAFS	781	795	71.2475	YFNFKTEAF	30	0.03285	Extracellular	0.913	0.718	0.6851
SFBM_0204	IEVPRFNTLYNAKVE	861	875	69.4943	FNTLYNAKV	19	0.03142	Extracellular	0.913	0.718	0.6851
SFBM_0204	IDYDFKIVDNKIAFD	46	60	67.2032	FKIVDNKIA	9.7	0.03021	Extracellular	0.913	0.718	0.6851
SFBM_0204	NYEAKIVLYNQRAIT	506	520	85.6653	VLYNQRAIT	18	0.01928	Extracellular	0.913	0.718	0.6851
SFBM_0204	IVLYNQRAITNTFVD	511	525	83.5453	YNQRAITNT	16	0.01376	Extracellular	0.913	0.718	0.6851
SFBM_0204	PPDNYEFVSGDKLLI	1391	1405	91.119	YEFVSGDKL	22	0.00297	Extracellular	0.913	0.718	0.6851
SFBM_0244	RGGNVFTPNKTFDI	351	365	75.805	NYVFTPNKT	7.5	0.29908	Unknown	0.25	0.356	0.5374
SFBM_0244	NVNFTIYNRHFHPVI	241	255	51.945	FTIYNRFHT	31	0.27796	Unknown	0.25	0.356	0.5374
SFBM_0244	SGTNMFIISDFPSSDF	381	395	87.8665	MFISDFPSS	21	0.15786	Unknown	0.25	0.356	0.5374
SFBM_0244	TSVSTDYGFLLNNKLI	536	550	77.784	YGFLLNNKLI	31	0.12942	Unknown	0.25	0.356	0.5374
SFBM_0244	DLDGNFIFYNGHINK	201	215	79.9561	IYNGHINK	17	0.11495	Unknown	0.25	0.356	0.5374
SFBM_0244	DYGFLLNNKLI FDKNK	541	555	66.2723	YGFLLNNKLI	28	0.09866	Unknown	0.25	0.356	0.5374
SFBM_0244	DSFEFRFSYSKDSDF	161	175	79.7321	FEFRFSYSK	19	0.09211	Unknown	0.25	0.356	0.5374
SFBM_0244	DVILNEIIDDKPLG	96	110	89.4328	LNEIIDDK	16	0.09083	Unknown	0.25	0.356	0.5374
SFBM_0244	NIKIPVINSLSAQVD	216	230	92.4579	VINSLSAQV	11	0.09072	Unknown	0.25	0.356	0.5374
SFBM_0244	NRIFDISNRVFMKKN	401	415	76.4595	RIFDISNRV	18	0.08803	Unknown	0.25	0.356	0.5374
SFBM_0244	YENLNILISNSSVLN	81	95	87.5077	LNILISNSS	14	0.08727	Unknown	0.25	0.356	0.5374
SFBM_0244	DGKRIFKIKLKDLLD	121	135	86.651	KRIFKIKLK	22	0.08487	Unknown	0.25	0.356	0.5374
SFBM_0244	FGIVSSVFFITIFLL	11	25	94.1737	IVSSVFFIT	20	0.08342	Unknown	0.25	0.356	0.5374
SFBM_0244	ILTSENRFDISNRV	396	410	88.7501	RIFDISNRV	24	0.08295	Unknown	0.25	0.356	0.5374
SFBM_0244	ENTKNLFLRDNVNFT	231	245	82.6514	FLRDNVNFT	25	0.08234	Unknown	0.25	0.356	0.5374
SFBM_0244	QVYFKYGDTRLVNGR	26	40	85.8826	FKYGDTRLV	12	0.06583	Unknown	0.25	0.356	0.5374
SFBM_0244	TFNLVKNLVTGSVST	526	540	80.3051	YNKLVGTSV	20	0.06466	Unknown	0.25	0.356	0.5374
SFBM_0244	VDIINLIDNYKLYFD	306	320	77.8257	INLIDNYKL	9	0.06263	Unknown	0.25	0.356	0.5374

SFBM_0244	FLSYDSGTNMFILSD	376	390	97.2206	LSYDSGTNM	18	0.06106	Unknown	0.25	0.356	0.5374
SFBM_0244	KLYFDDNLMKTYKAY	316	330	81.2305	LYFDDNLMK	25	0.0607	Unknown	0.25	0.356	0.5374
SFBM_0244	LFLRDNVNFYINRF	236	250	89.5898	VNFYINRF	11	0.06012	Unknown	0.25	0.356	0.5374
SFBM_0244	YKFYDINENKFLSYD	366	380	77.0067	KFYDINENK	29	0.05013	Unknown	0.25	0.356	0.5374
SFBM_0244	SSGDLKISNTYLSLN	261	275	84.7159	LKISNTYSL	13	0.04941	Unknown	0.25	0.356	0.5374
SFBM_0244	DEYIVFRNKELIFLE	661	675	77.5905	YIVFRNKEL	11	0.04794	Unknown	0.25	0.356	0.5374
SFBM_0244	ILISNSSVLNDVILN	86	100	96.6605	ILISNSSVL	12	0.04305	Unknown	0.25	0.356	0.5374
SFBM_0244	SVFFITIFLLQVYFK	16	30	95.2148	FITIFLLQV	11	0.04208	Unknown	0.25	0.356	0.5374
SFBM_0244	NNKLIFDKNNKNNKVR	546	560	74.2214	LIFDKNNKN	20	0.03674	Unknown	0.25	0.356	0.5374
SFBM_0244	GHINKNIKIPVINSL	211	225	72.3411	GHINKNIKI	19	0.03397	Unknown	0.25	0.356	0.5374
SFBM_0244	NFYDKEFKIESNYYS	186	200	76.4692	FKIESNYYS	30	0.02594	Unknown	0.25	0.356	0.5374
SFBM_0244	INSGNYISLSSKGFN	626	640	96.0539	GNYSLSK	13	0.02331	Unknown	0.25	0.356	0.5374
SFBM_0244	FRFNDMSSSILKQVT	436	450	90.5417	RFNDMSSSI	21	0.02057	Unknown	0.25	0.356	0.5374
SFBM_0244	LYYISFGIVSSVFFI	6	20	87.1179	LYYISFGIV	16	0.01541	Unknown	0.25	0.356	0.5374
SFBM_0244	FPSDFILTSENRIFD	391	405	90.3233	FILTSENRI	6.2	0.00861	Unknown	0.25	0.356	0.5374
SFBM_0244	FIYNFGHINKNIKIP	206	220	53.3261	GHINKNIKI	18	0.00462	Unknown	0.25	0.356	0.5374
SFBM_0253	ERLQFYVTVGIGIFF	21	35	92.9301	LQFYVTVGI	19	0.02711	Unknown	0.25	0.000	0.7041
SFBM_0300	LDPEFFYSNSARSRF	166	180	83.3527	FYSNSARSR	9.5	0.51738	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	MRRTNIFSREYERRI	1	15	80.373	TNIFSREYE	26	0.18438	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	KVNINNEMKLLTFSF	256	270	79.5946	NNEMKLLTF	15	0.16059	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	QDIKFMTSIAGENI	231	245	75.7057	IKHFMTSIA	12	0.15607	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	KNTYEGYSWYKGARF	186	200	90.4735	YSWYKGARF	18	0.15049	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	YYSIISKEIVNISEA	51	65	89.6963	YSIISKEIV	19	0.13014	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	TEQGIKVNINNEMKL	251	265	67.8834	IKVNINNEM	29	0.11134	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	ILTILVMVLYMVNF	26	40	91.2421	MVLYMVNF	12	0.09861	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	NLKILIENITQDITY	141	155	73.4893	LKILIENI	22	0.08101	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	VMVLYMVNFSEVNN	31	45	74.6946	MVLYMVNF	21	0.07422	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	TQDITYMIDESFNTFK	151	165	86.8791	YMIDESFNT	19	0.06247	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	FILIIILTILVMVVL	21	35	88.9928	IIIIILTIL	7	0.04405	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	YVRDRYYSIISKEIV	46	60	81.391	YSIISKEIV	6.9	0.03486	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0300	RKKRFFIIIIILTIL	16	30	97.1732	FIIIIILTI	5.5	0.01505	Cytoplasmic Membrane	0.955	0.111	0.5019
SFBM_0325	KSGDRLIYRNYRNI	51	65	72.3756	IYRNYRNI	33	0.18357	Cytoplasmic Membrane	0.999	0.259	0.5303

SFBM_0325	GRAVILFYPIERFKY	161	175	78.2264	VILFYPIER	14	0.10009	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	MINRLKFNRLIYIL	1	15	75.1881	LKFNRLIY	9.2	0.09907	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	IIIPVYLANNLVKS	21	35	86.3989	IPVYLANNL	18	0.09899	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	VFFVTVTSNSMYPTV	36	50	75.9379	VTNSMYPT	23	0.07888	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	LIYRNYRNIKRNDI	56	70	42.2748	YRNIKRNDI	12	0.0618	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	VILFYPIERFKYLF	164	178	79.73	VILFYPIER	17	0.05128	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	NKFVCSNSILGRAVI	151	165	81.0571	NKFVCSNSI	15	0.05052	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	NYRNIKRNDIIVFYS	61	75	59.1164	YRNIKRNDI	7.2	0.03116	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	LWIIIFIIIPVYLAN	16	30	83.6334	IIIFIIIPVYL	6.7	0.03059	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	NLIKRVIAVPGDIID	81	95	88.8318	IKRVIAVPG	17	0.02632	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	VYLANNLVKSVEFT	26	40	86.3699	YLANNLVKS	18	0.01083	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0325	LIYLLWIIIFIIIP	11	25	86.4517	IYLLWIIIF	9	0.00634	Cytoplasmic Membrane	0.999	0.259	0.5303
SFBM_0330	ILPILINIFDNLFNI	16	30	89.4066	NIFDNLFNI	15	0.25955	Unknown	0.25	0.382	0.6307
SFBM_0330	DDIRSLDRTYSYKF	126	140	97.0698	IRSLDRTY	7.9	0.18549	Unknown	0.25	0.382	0.6307
SFBM_0330	MLNFYGFMIKICVLI	1	15	70.5998	MLNFYGFMI	28	0.14481	Unknown	0.25	0.382	0.6307
SFBM_0330	NLFNIFKILNNRISV	26	40	26.65	FNIFKILNN	2.8	0.10634	Unknown	0.25	0.382	0.6307
SFBM_0330	ALEFLMRAYNYRFKI	56	70	77.2755	RAYNYRFKI	5.9	0.0935	Unknown	0.25	0.382	0.6307
SFBM_0330	MRAYNYRFKIKNDSI	61	75	77.6628	YRFKIKNDS	25	0.08541	Unknown	0.25	0.382	0.6307
SFBM_0330	YSYKFKFIHGDDLKL	136	150	87.2418	FKFIHGDDL	22	0.08448	Unknown	0.25	0.382	0.6307
SFBM_0330	LDYSQLVKNFYIYKM	150	164	76.984	VNKFYIYKM	25	0.07804	Unknown	0.25	0.382	0.6307
SFBM_0330	INIFDNLFNIFKILN	21	35	84.5737	DNLFNIFKI	21	0.07038	Unknown	0.25	0.382	0.6307
SFBM_0330	YRFKIKNDSIYLYDD	66	80	82.7954	YRFKIKNDS	28	0.06252	Unknown	0.25	0.382	0.6307
SFBM_0330	FKILNNRISVLDKVL	31	45	62.3525	FKILNNRIS	12	0.04275	Unknown	0.25	0.382	0.6307
SFBM_0330	DDLKLDYSQLVKNFY	146	160	90.6044	LKLDYSQLV	12	0.03969	Unknown	0.25	0.382	0.6307
SFBM_0330	KGCKNFIIFTGIIIN	111	125	80.0842	IFIFTGII	24	0.0038	Unknown	0.25	0.382	0.6307
SFBM_0330	ICVLIIIPILINIFD	11	25	91.3921	ILPILINIF	13	0.00065	Unknown	0.25	0.382	0.6307
SFBM_0449	GEEVTLVANENYYKG	266	280	61.2247	LVANENYYK	36	0.2044	Cellwall	0.921	0.234	0.6001
SFBM_0449	FLGYGYVAINHNEPI	336	350	83.3882	GYGYVAINH	26	0.11504	Cellwall	0.921	0.234	0.6001
SFBM_0449	HNEPIMQDLNVRKAL	346	360	91.7572	IMQDLNVRK	16	0.10047	Cellwall	0.921	0.234	0.6001
SFBM_0449	KDTLIGIISPNGVF	76	90	91.394	IGIISPNGV	12	0.0935	Cellwall	0.921	0.234	0.6001
SFBM_0449	EGFKIIDDKTFVMT	196	210	83.4698	FKIIDDKTF	15	0.08789	Cellwall	0.921	0.234	0.6001
SFBM_0449	KEGKFSYHLAFLAWS	481	495	89.591	FSYHLAFLA	19	0.08284	Cellwall	0.921	0.234	0.6001
SFBM_0449	MKRRKLISILGLVLG	1	15	94.1441	KRKLISILG	13	0.0751	Cellwall	0.921	0.234	0.6001
SFBM_0449	KL.SLKFLASSPNPVN	431	445	73.8086	LKFLASSPN	8.8	0.07367	Cellwall	0.921	0.234	0.6001

SFBM_0449	VMDNYKAIGIELKA	451	465	86.8851	YKAIGIELK	15	0.05793	Cellwall	0.921	0.234	0.6001
SFBM_0449	KTLAYFSKDLKPLAK	216	230	90.1683	LAYFSKDLK	18	0.05258	Cellwall	0.921	0.234	0.6001
SFBM_0449	NLVFRVNETNQMLL	286	300	96.4246	RVVNETNQM	18	0.03405	Cellwall	0.921	0.234	0.6001
SFBM_0449	VRKALAYGLDRSSVV	356	370	90.8811	YGLDRSSVV	12	0.02997	Cellwall	0.921	0.234	0.6001
SFBM_0449	TFKYIMDKTYTGRFE	161	175	78.8818	FKYIMDKTY	13	0.02327	Cellwall	0.921	0.234	0.6001
SFBM_0449	LGFINANIVEFLGYG	326	340	83.3107	LGFINANIV	17	0.02208	Cellwall	0.921	0.234	0.6001
SFBM_0449	DVLVPMIDNYKAIG	446	460	88.2231	VMDNYKAI	19	0.01807	Cellwall	0.921	0.234	0.6001
SFBM_0449	LLLQLGFINANIVE	321	335	88.4247	LGFINANIV	10	0.00848	Cellwall	0.921	0.234	0.6001
SFBM_0506	NIDYEILSSKTFNS	96	110	89.5619	YEILSSKTF	19	0.21391	Unknown	0.333	0.449	0.5166
SFBM_0506	KSTKYKFSILDIDL	151	165	87.5409	YKFSILDI	12	0.20788	Unknown	0.333	0.449	0.5166
SFBM_0506	SKKNNQWYMIVSDKL	166	180	95.5733	QWYMIVSDK	18	0.1935	Unknown	0.333	0.449	0.5166
SFBM_0506	DIAKHIINFFKNDPS	121	135	49.306	INFFKNDPS	34	0.13584	Unknown	0.333	0.449	0.5166
SFBM_0506	IINFFKNDPSFPDDE	126	140	62.9392	INFFKNDPS	20	0.10356	Unknown	0.333	0.449	0.5166
SFBM_0506	NKYEIFLESKSTKY	141	155	92.9941	FLESKSTK	15	0.08863	Unknown	0.333	0.449	0.5166
SFBM_0506	ILFSLISISKSEITF	26	40	97.0471	FSLISIKS	11	0.08651	Unknown	0.333	0.449	0.5166
SFBM_0506	VLGFAILFSLISIKD	21	35	97.6809	FAILFSLI	18	0.07094	Unknown	0.333	0.449	0.5166
SFBM_0506	STLKIRFTYYDIAKH	111	125	86.5438	FTYYDIAKH	25	0.06836	Unknown	0.333	0.449	0.5166
SFBM_0506	QWYMIVSKLLNILT	171	185	80.1402	YMIVSKLL	6.1	0.0656	Unknown	0.333	0.449	0.5166
SFBM_0506	VSDKLLNILTSGIYK	176	190	94.8262	KLLNILTSG	19	0.06289	Unknown	0.333	0.449	0.5166
SFBM_0506	ILSSKTFNSSTLKI	101	115	83.2371	FFNSSTLKI	26	0.06277	Unknown	0.333	0.449	0.5166
SFBM_0506	NKYLLNISIPNINHV	56	70	82.4862	ISIPNINHV	14	0.05382	Unknown	0.333	0.449	0.5166
SFBM_0506	ILISIVLGFMAILFSL	16	30	77.8244	ISIVLGF	3.1	0.03872	Unknown	0.333	0.449	0.5166
SFBM_0506	ESKMEIIEIQNMMF	81	95	93.7953	IIEIQNMM	17	0.02713	Unknown	0.333	0.449	0.5166
SFBM_0506	LLNILTSGIYKNFIT	180	194	78.3768	LNILTSGIY	17	0.02508	Unknown	0.333	0.449	0.5166
SFBM_0506	RFTYYDIAKHINFF	116	130	66.8255	YYDIAKHII	26	0.02294	Unknown	0.333	0.449	0.5166
SFBM_0506	MKITSKNILLKFCVV	1	15	89.0628	MKITSKNIL	21	0.00276	Unknown	0.333	0.449	0.5166
SFBM_0506	KFCCVILISIVLGFA	11	25	93.6723	ILISIVLGF	18	0.0003	Unknown	0.333	0.449	0.5166
fiM	NSPVEIYENRLIFN	291	305	82.9577	PVEIYENR	9.1	0.30574	Cytoplasmic Membrane	0.999	0.171	0.5411
fiM	EFIFMNPNTTIVSF	91	105	84.9512	FIFMNPNT	14	0.13283	Cytoplasmic Membrane	0.999	0.171	0.5411
fiM	RLIFNGVLGLVGENK	301	315	92.6198	FNGVLGLVG	21	0.11619	Cytoplasmic Membrane	0.999	0.171	0.5411
fiM	EEVELIKKSIRKELN	241	255	79.6062	IKKSIRKEL	18	0.09738	Cytoplasmic Membrane	0.999	0.171	0.5411
fiM	YFMYICLPYISIEKI	211	225	87.6906	FMYICLPYI	18	0.09687	Cytoplasmic Membrane	0.999	0.171	0.5411
fiM	MKSLSIHENFTQYL	51	65	88.5023	MKSLSIHE	15	0.08696	Cytoplasmic Membrane	0.999	0.171	0.5411
fiM	SVQDFLSLQTNIII	271	285	86.057	FLSLQTNDI	6.8	0.08423	Cytoplasmic Membrane	0.999	0.171	0.5411
fiM	NRSNILQKNVKKYDF	26	40	67.9278	SNILQKNVK	30	0.08308	Cytoplasmic Membrane	0.999	0.171	0.5411

flM	IYIENRLIFNGVLGL	296	310	84.5103	IENRLIFNG	26	0.0796	Cytoplasmic Membrane	0.999	0.171	0.5411
flM	QITFHEFIFSMNPNT	86	100	94.6783	EFIFSMNP	18	0.05965	Cytoplasmic Membrane	0.999	0.171	0.5411
flM	FGSIFMGLESALSVL	111	125	91.7441	IFMGLESAL	18	0.04555	Cytoplasmic Membrane	0.999	0.171	0.5411
flM	KILTQVSNEILNIR	151	165	87.9161	ILTQVSNEI	17	0.02556	Cytoplasmic Membrane	0.999	0.171	0.5411
flM	SSYFVSVFGIQIKVD	66	80	83.7565	FVSVFGIQI	15	0.01761	Cytoplasmic Membrane	0.999	0.171	0.5411
flM	IKKSIRKELNSVNIE	246	260	76.7843	IKKSIRKEL	29	0.01169	Cytoplasmic Membrane	0.999	0.171	0.5411
flY	HIALKVKYTRGIDGN	76	90	79.8893	HIALKVKYT	22	0.12078	Cytoplasmic Membrane	0.878	0.303	0.5249
flY	KIKFQMNVDLINSN	181	195	80.2875	FQMNVDLI	8.8	0.11728	Cytoplasmic Membrane	0.878	0.303	0.5249
flY	VRTAILSNTERVKN	341	355	86.8982	TAILSNTER	23	0.11357	Cytoplasmic Membrane	0.878	0.303	0.5249
flY	NVLVMSIRDAD	91	105	90.9585	LVMSIRDAI	9.1	0.07156	Cytoplasmic Membrane	0.878	0.303	0.5249
flY	DLILDVPLEISVVLG	276	290	93.3015	LILDVPLEI	15	0.05116	Cytoplasmic Membrane	0.878	0.303	0.5249
flY	EFMQPHIALKVKYTR	71	85	80.8666	MQPHIALKV	16	0.0326	Cytoplasmic Membrane	0.878	0.303	0.5249
flY	KNHINMLNDQNGFK	206	220	74.8913	INIMLNDQN	12	0.02602	Cytoplasmic Membrane	0.878	0.303	0.5249
flgK	MSNLITFQHAYNASA	546	560	86.7382	TFQHAYNAS	11	0.18137	Unknown	0.25	0.459	0.5771
flgK	LVYNLEISRMSVSGV	526	540	78.6305	YNLEISRMS	14	0.12636	Unknown	0.25	0.459	0.5771
flgK	TKSIVFSLNAVHSGK	366	380	87.0598	IVFSLNAVH	17	0.06706	Unknown	0.25	0.459	0.5771
flgK	DYTRQRIRVEASSPL	36	50	93.0629	IRVEASSPL	12	0.05879	Unknown	0.25	0.459	0.5771
flgK	RNIVLQEASNMATEI	136	150	94.1493	IVLQEASNM	18	0.05828	Unknown	0.25	0.459	0.5771
flgK	DSFINQLNQLTKSIV	356	370	88.5819	LNQLTKSIV	18	0.05496	Unknown	0.25	0.459	0.5771
flgK	DPMKLVKVRKNDHMF	416	430	63.2618	LKVRKNDHM	32	0.0012	Unknown	0.25	0.459	0.5771
flID	NSNIFEFASNFTLS	516	530	83.24	IFEFASNFF	18	0.16927	Extracellular	0.972	0.506	0.7302
flID	YYTIKLDKGAYSQDK	216	230	84.7875	IKLDKGAYS	14	0.13692	Extracellular	0.972	0.506	0.7302
flID	SSRAKVNISYNTISR	451	465	79.001	VNISYNTIS	28	0.10787	Extracellular	0.972	0.506	0.7302
flID	YYAQFTRLETALSKL	776	790	81.8408	YAFTRLET	25	0.10541	Extracellular	0.972	0.506	0.7302
flID	ETKLEFMYELRNAI	616	630	80.8954	FMYELRNAI	23	0.05708	Extracellular	0.972	0.506	0.7302
flID	SFMYELRNAIFTVD	621	635	86.4469	FMYELRNAI	22	0.04844	Extracellular	0.972	0.506	0.7302
flID	LTIKQVFSISSRAK	441	455	92.1048	IKQVFSIS	6	0.0481	Extracellular	0.972	0.506	0.7302
flID	KLVFKTNSKEQIVIS	361	375	79.0013	FKTNSKEQI	22	0.02294	Extracellular	0.972	0.506	0.7302
flID	QIVISGNAANSIGIG	371	385	83.1555	VISGNAANS	22	0.01496	Extracellular	0.972	0.506	0.7302
flID	VNISYNTISRTFNIE	456	470	75.1492	ISYNTISRT	23	0.00864	Extracellular	0.972	0.506	0.7302
SFBM_0582	EYKVISYISIGKDG	206	220	84.9707	YKVISYSI	13	0.19913	Extracellular	0.96	0.478	0.5204
SFBM_0582	NGGQDEYKVISYSI	201	215	82.02	YKVISYSI	13	0.13377	Extracellular	0.96	0.478	0.5204
SFBM_0582	FFMVAKGPVAHKDGD	101	115	90.4364	FFMVAKGPV	23	0.06103	Extracellular	0.96	0.478	0.5204

SFBM_0582	QFTDLIVTSRAFQAS	296	310	96.6214	IVTSRAFQA	17	0.0376	Extracellular	0.96	0.478	0.5204
SFBM_0582	IVTSRAFQASSKIIS	301	315	89.5261	RAFQASSKI	22	0.0273	Extracellular	0.96	0.478	0.5204
SFBM_0582	IMYTRDGLSLDSEG	131	145	99.0325	YTRDGLSL	13	0.00977	Extracellular	0.96	0.478	0.5204
SFBM_0866	PDRTRMTFLKNSIGT	71	85	73.4373	MTFLKNSIG	28	0.17723	Unknown	0.25	0.151	0.5129
SFBM_0866	MTFLKNSIGTIVVES	76	90	75.8573	MTFLKNSIG	23	0.11738	Unknown	0.25	0.151	0.5129
SFBM_0866	ILYVIMILPETERRK	16	30	85.1843	VIMILPETR	9	0.09525	Unknown	0.25	0.151	0.5129
SFBM_0866	MSSFVNLLIPIGFLI	1	15	90.0004	VNLLIPIGF	20	0.06733	Unknown	0.25	0.151	0.5129
SFBM_0866	IGFLILLYVIMILPE	11	25	87.1041	LILLYVIMI	16	0.04027	Unknown	0.25	0.151	0.5129
mrcB	SDKFLRSIGSTTKPL	416	430	89.1078	KFLRSIGST	13	0.24982	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	GNPLILASAFSVFAN	541	555	89.1281	LILASAFSV	17	0.23646	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	LDEMPQLRNFALIAI	106	120	82.0209	LQLRNFALIA	20	0.23393	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	RYSLERSINLSAIKI	481	495	90.1837	SINLSAIKI	17	0.16493	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	EDSQFYKNRAITVLN	256	270	71.7909	QFYKNRAIT	12	0.16296	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	FAIKTSPPLTIEAV	61	75	82.6775	FAIKTSPPL	7	0.15221	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	GYTNVRYSLERSINL	476	490	96.6366	GYTNVRYSL	17	0.13803	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	KNNILTNEVSVVRKI	166	180	89.4547	NNILTNEVS	23	0.11821	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	EAGALQYFSKNAKDL	216	230	74.1904	LQYFSKNAK	31	0.11741	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	NAKDLTLAQSAFLAG	226	240	96.0115	LTLAQSAFL	14	0.11712	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	HAQYVLDNATSYNL	356	370	94.8073	YVLDNATS	14	0.10281	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	KIITFLIFLLAIFL	36	50	93.4011	ITFLIFLL	5.1	0.09202	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	YKNRAITVLNKMLEL	261	275	77.8971	NRAITVLNK	21	0.07951	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	VFGFTFAIKTSPPL	56	70	85.048	FAIKTSPPL	9.5	0.07919	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	RSINLSAIKIVDKVG	486	500	81.247	SINLSAIKI	22	0.07412	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	SKKTKKIITFLIFL	31	45	84.2139	KKIITFLI	26	0.07053	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	LQLRNFALIAIEDERF	111	125	86.7189	LQLRNFALIA	19	0.06461	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	PEYLKETRVFLNRKY	756	770	72.2456	TRVFLNRKY	25	0.05864	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	AVVIDFKTGHTKAIH	386	400	85.9587	VVIDFKTGH	14	0.04602	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	IYLNIFVGGNAYGV	201	215	56.8211	NTIFVGGNA	25	0.0453	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	IDIQRIISAVITDIK	131	145	86.1008	IQRIISAVI	12	0.03945	Cytoplasmic Membrane	0.951	0.590	0.5278

mrcB	KNNFIGYTNVRYSL	471	485	85.8565	NNFIGYTNV	20	0.03285	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	FMDEVLKAEKRYVIS	91	105	89.6167	VLKAEKRYV	18	0.03205	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	LAIFLIGAITVFGFT	46	60	85.3144	IFLIGAITV	19	0.02464	Cytoplasmic Membrane	0.951	0.590	0.5278
mrcB	FLIFLLAIFLIGAIT	41	55	86.4523	IFLLAIFLI	7.8	0.01162	Cytoplasmic Membrane	0.951	0.590	0.5278
SFBM_0910	IKIVMVLJLDTKPKI	56	70	71.5943	VMVLJLDTK	8.5	0.16456	Unknown	0.25	0.142	0.8228
SFBM_0910	MKQKLQNLTRPIKI	1	15	96.071	LQNLTRPI	19	0.16067	Unknown	0.25	0.142	0.8228
SFBM_0910	LDITLVLVVGIIYHF	36	50	84.5949	VLMVGIYHF	8	0.10929	Unknown	0.25	0.142	0.8228
SFBM_0910	IYHFIKYLKIKIVM	46	60	69.7574	IKYLLKIKI	5.3	0.0795	Unknown	0.25	0.142	0.8228
SFBM_0910	KYLLKIKIVMVLILD	51	65	86.1039	YLLKIKIVM	19	0.04522	Unknown	0.25	0.142	0.8228
SFBM_0910	RPIKIQTEVSNKIYI	11	25	78.3822	IQTEVSNKI	14	0.03612	Unknown	0.25	0.142	0.8228
SFBM_0910	THHYKLDITLVLVVG	31	45	80.6386	YKLDITLVL	21	0.01136	Unknown	0.25	0.142	0.8228
SFBM_0910	VLMVGIYHFIKYLK	41	55	76.3421	MVGIYHFIK	17	0.00635	Unknown	0.25	0.142	0.8228
SFBM_0943	SYGYDVILRIKGEK	71	85	81.2062	YDVILRIK	23	0.14783	Cytoplasmic Membrane	0.955	0.184	0.7335
SFBM_0943	VKINVVYSSTGIVNS	26	40	86.1212	INVVYSSTG	13	0.10465	Cytoplasmic Membrane	0.955	0.184	0.7335
SFBM_0943	SIYKESYGYDVILRI	66	80	91.6551	YKESYGYDV	16	0.09388	Cytoplasmic Membrane	0.955	0.184	0.7335
SFBM_0943	LFIFVSAGFVKINV	16	30	84.7144	IFVSAGFV	16	0.089	Cytoplasmic Membrane	0.955	0.184	0.7335
SFBM_0943	RNLKYIFGIFLFI	6	20	93.1339	KYIFGIFL	9.3	0.06297	Cytoplasmic Membrane	0.955	0.184	0.7335
SFBM_0943	VILRIKGEKIFSLK	76	90	82.42	VILRIKGEE	28	0.04836	Cytoplasmic Membrane	0.955	0.184	0.7335
SFBM_0943	MFFKRRNLKYIFGIF	1	15	85.0785	MFFKRRNLK	18	0.01521	Cytoplasmic Membrane	0.955	0.184	0.7335
SFBM_0956	EASQKFTDISTRKI	21	35	94.793	FTDISTRKI	20	0.24923	Unknown	0.25	0.494	0.5837
SFBM_0956	ELSLNMYKDLNKED	131	145	68.2475	LSLNMYKDLN	37	0.13573	Unknown	0.25	0.494	0.5837
SFBM_0956	EIIRSNYAYFLENSI	61	75	72.3462	YAYFLENSI	9.8	0.06842	Unknown	0.25	0.494	0.5837
SFBM_0956	YSLNKLNMRLGLEK	6	20	91.1529	LNMRLEK	7.7	0.06715	Unknown	0.25	0.494	0.5837
SFBM_0956	TTRVDEIIRSNYAYF	56	70	95.6251	DEIIRSNYA	14	0.05651	Unknown	0.25	0.494	0.5837
SFBM_0956	NYAYFLENSIRYNEK	66	80	73.0502	YAYFLENSI	15	0.05285	Unknown	0.25	0.494	0.5837
SFBM_0956	MKRFKYSLNKLLNMR	1	15	78.4489	FKYSLNKLL	5.2	0.04284	Unknown	0.25	0.494	0.5837
SFBM_0969	TEPLILKTYNPKIVG	141	155	76.9416	ILKTYNPKI	19	0.25323	Unknown	0.25	0.368	0.6583
SFBM_0969	EDKVKLELKNILSKV	66	80	87.1419	VKLELKNIL	21	0.22833	Unknown	0.25	0.368	0.6583
SFBM_0969	AVSYFNNSTPALLNV	31	45	85.9099	YFNNSTPAL	8	0.13166	Unknown	0.25	0.368	0.6583
SFBM_0969	MEKIKQKIKKSNKF	1	15	74.8997	KIKKSNKF	31	0.10234	Unknown	0.25	0.368	0.6583
SFBM_0969	NKLSYKLTVMVSSLY	166	180	96.4217	YKLTVMVSS	14	0.06306	Unknown	0.25	0.368	0.6583
SFBM_0969	LIIHAGIMIFAVSYF	21	35	77.7549	LIIHAGIMI	11	0.06292	Unknown	0.25	0.368	0.6583
SFBM_0969	KSNKFTYMIILIIIA	11	25	85.6926	FTYMIILII	11	0.06105	Unknown	0.25	0.368	0.6583
SFBM_0969	QKIKKSNKFTYMI	6	20	57.9402	IKKSNKFTY	21	0.02376	Unknown	0.25	0.368	0.6583

SFBM_0969	GIMIFAVSYFNNSTP	26	40	72.0362	IMFVSYF	28	0.01715	Unknown	0.25	0.368	0.6583
SFBM_0969	TYMILIHAGIMIF	16	30	76.0934	ILIHAGIM	6.6	0.00861	Unknown	0.25	0.368	0.6583
SFBM_1021	AYRHLLKFLSRSRA	31	45	88.8333	YRHLLKFLSL	12	0.2631	Extracellular	0.96	0.144	0.5560
SFBM_1021	QIVVVHILICNFDL	6	20	72.8607	IILICNFDL	14	0.10486	Extracellular	0.96	0.144	0.5560
SFBM_1021	MKKYIQIVVVHILI	1	15	84.0062	IQIVVVHII	11	0.09387	Extracellular	0.96	0.144	0.5560
SFBM_1021	IILICNFLLNVMA	11	25	89.1972	IILICNFDL	16	0.09012	Extracellular	0.96	0.144	0.5560
SFBM_1021	LNVMAINGDIAYRHL	21	35	86.183	INGDIAYRH	17	0.08879	Extracellular	0.96	0.144	0.5560
SFBM_1021	VGIPVLYFESTNWN	246	260	90.7507	LYFESTNWN	16	0.05885	Extracellular	0.96	0.144	0.5560
SFBM_1021	LKFLSRSRALVTQS	36	50	90.0934	KFLSRSRA	1.8	0.05544	Extracellular	0.96	0.144	0.5560
SFBM_1021	INGDIAYRHLLKFLSL	26	40	90.379	YRHLLKFLSL	23	0.04297	Extracellular	0.96	0.144	0.5560
SFBM_1021	EKIKYMINLDTLLSG	171	185	90.2774	IKYMINLDT	4.8	0.03654	Extracellular	0.96	0.144	0.5560
SFBM_1021	EHLVGYISLIKEFIL	296	310	82.9924	YISLIKEFI	9.7	0.01883	Extracellular	0.96	0.144	0.5560
SFBM_1048	QTKLIFSRLKTSPET	541	555	83.3784	KLIFSRLKT	20	0.18378	Unknown	0.569	0.855	1.0026
SFBM_1048	SGELRTYERNMATRF	141	155	76.5934	LRTYERNMA	21	0.16932	Unknown	0.569	0.855	1.0026
SFBM_1048	NKRYVQVYVSSGELR	131	145	96.7232	KRYVQVYVS	18	0.10135	Unknown	0.569	0.855	1.0026
SFBM_1048	LYSGSLYVLSVVATM	11	25	94.0934	YVLSVVATM	17	0.05647	Unknown	0.569	0.86	1.0026
SFBM_1048	ISKLVQNFAGQLN	831	845	83.6216	LKLVQNFAG	14	0.04408	Unknown	0.569	0.855	1.0026
SFBM_1048	VTSYIYRSKPKWIIG	176	190	85.558	IYRSKPKWI	7.4	0.03237	Unknown	0.569	0.855	1.0026
SFBM_1048	LYVLSVVATMSLYFS	16	30	98.5013	YVLSVVATM	9.8	0.0224	Unknown	0.569	0.855	1.0026
SFBM_1052	MNFKRKRKRSIGYSGS	1	15	89.9837	FKRKRKRSIGI	17	0.11893	Cytoplasmic Membrane	0.955	0.651	0.6396
SFBM_1052	LCMLSMITAVSLYAN	16	30	86.863	LSMITAVSL	16	0.00063	Cytoplasmic Membrane	0.955	0.651	0.6396
SFBM_1073	GVPIYIVNGGNFLQI	41	55	77.8819	IYIVNGGNF	18	0.61015	Unknown	0.333	0.497	0.9044
SFBM_1073	EEEFMKMVVGMGLKY	1418	1432	79.9466	MKMOVGMGL	14	0.28875	Unknown	0.333	0.497	0.9044
SFBM_1073	NGNTRLRIKSDLVNDM	991	1005	80.6226	IKSDLVNDM	15	0.20584	Unknown	0.333	0.497	0.9044
SFBM_1073	TISNVKFEKLEFKT	861	875	90.1789	FEKLEFKT	22	0.17968	Unknown	0.333	0.497	0.9044
SFBM_1073	AKLVYTISNVKFEKT	856	870	85.1684	YTISNVKFE	7.3	0.12827	Unknown	0.333	0.497	0.9044
SFBM_1073	IEIKIIITGKEITF	1281	1295	88.8384	IKIIITGK	13	0.11472	Unknown	0.333	0.497	0.9044
SFBM_1073	FKFIDNDSIYIKG	251	265	85.9807	FKFIDNSI	13	0.11341	Unknown	0.333	0.497	0.9044
SFBM_1073	LLIQFSFINDVNFKF	336	350	71.3524	IQFSFINDV	7.2	0.10994	Unknown	0.333	0.497	0.9044
SFBM_1073	NVSLVFDKVIKSKDF	211	225	83.2372	FDKVIKSKD	17	0.10843	Unknown	0.333	0.497	0.9044
SFBM_1073	MNNLFKTTMKFMIYF	1	15	81.1279	FKTTMKFMI	14	0.09143	Unknown	0.333	0.497	0.9044
SFBM_1073	DVLLTYKTKANDIAD	956	970	70.4974	YKTKANDIA	34	0.08434	Unknown	0.333	0.497	0.9044
SFBM_1073	INEEFMKMVVGMGL	1416	1430	81.2909	MKMOVGMGL	16	0.08355	Unknown	0.333	0.497	0.9044
SFBM_1073	HTIETDIISVTSFNL	1156	1170	96.4986	IISVTSFNL	17	0.07393	Unknown	0.333	0.497	0.9044
SFBM_1073	FFTLYTDVSVINLKD	321	335	82.2677	FTLYTDVSV	17	0.07181	Unknown	0.333	0.497	0.9044
SFBM_1073	DEVKINKIDKQFFIY	76	90	83.3146	VKINKIDKQ	22	0.07007	Unknown	0.333	0.497	0.9044
SFBM_1073	SFINDVNFKPEDKLN	341	355	86.9402	FINDVNFKF	21	0.06591	Unknown	0.333	0.497	0.9044
SFBM_1073	QYSDKIIASASSYDI	1026	1040	90.8811	DKIIASASS	22	0.06449	Unknown	0.333	0.497	0.9044

SFBM_1073	DSIRLFKCLRNGTSK	596	610	56.6958	IRLFKCLR	19	0.05756	Unknown	0.333	0.497	0.9044
SFBM_1073	GKTKFSVDKIVHTKP	1196	1210	85.3579	FSVDKIVHT	26	0.05116	Unknown	0.333	0.497	0.9044
SFBM_1073	KNDINYSIIPDVKIT	266	280	89.0968	INYSIIPDV	16	0.04953	Unknown	0.333	0.497	0.9044
SFBM_1073	LFMMSCFNLSVKLSY	21	35	71.2751	MSCFNLSVK	11	0.04853	Unknown	0.333	0.497	0.9044
SFBM_1073	LLISYNPHGYFNEVY	1056	1070	90.3682	YNPHGYFNE	11	0.04838	Unknown	0.333	0.497	0.9044
SFBM_1073	KNLIFIDGIEYGR	146	160	98.8036	IDGIEYGR	8.7	0.04745	Unknown	0.333	0.497	0.9044
SFBM_1073	PDEYDVLITYKTKA	951	965	78.3665	YDVLITYKT	21	0.04516	Unknown	0.333	0.497	0.9044
SFBM_1073	DVRYVINLKDNVSLV	201	215	91.9312	VVINLKDNV	9.1	0.04409	Unknown	0.333	0.497	0.9044
SFBM_1073	DISNMSSMNNLI	226	240	75.8046	SMMNNLI	25	0.04067	Unknown	0.333	0.497	0.9044
SFBM_1073	KFVYIDGREYVSYS	506	520	82.549	YIDGREYV	23	0.03949	Unknown	0.333	0.497	0.9044
SFBM_1073	KLLYMFGRRTIERE	301	315	85.3948	LYMFGRRTI	19	0.0356	Unknown	0.333	0.497	0.9044
SFBM_1073	PNFNFRNDDSFKFI	451	465	83.0963	FRNDDSFKI	11	0.03524	Unknown	0.333	0.497	0.9044
SFBM_1073	QISKFKYLENSFGVH	116	130	69.9544	FKYLENSFG	24	0.03429	Unknown	0.333	0.497	0.9044
SFBM_1073	LEFKTISANSSLNIG	871	885	85.143	FKTISANSS	15	0.03241	Unknown	0.333	0.497	0.9044
SFBM_1073	FMIYFGLVFLFMMS	11	25	93.4691	MIYFGLVVF	14	0.03219	Unknown	0.333	0.497	0.9044
SFBM_1073	KLIEILYSIVVDRDP	1366	1380	85.8021	EILYSIVV	21	0.03075	Unknown	0.333	0.497	0.9044
SFBM_1073	GLFVFLFMMSCFNLS	16	30	84.7443	VFLFMMSCF	25	0.02946	Unknown	0.333	0.497	0.9044
SFBM_1073	IVTYMDMDYAKLVY	846	860	85.4257	MDMDYAKL	16	0.0273	Unknown	0.333	0.497	0.9044
SFBM_1073	NNLRNIFKFKDISLH	181	195	82.1566	LRNIFKFKD	25	0.02595	Unknown	0.333	0.497	0.9044
SFBM_1073	NFWINFYNEALNNA	1386	1400	58.7383	FWINFYNE	34	0.02178	Unknown	0.333	0.497	0.9044
SFBM_1073	INLKDNVSLVFDKVI	206	220	96.5242	NVSLVFDKV	16	0.0191	Unknown	0.333	0.497	0.9044
SFBM_1073	FRYDRKNLIFIDGIE	141	155	87.7088	FRYDRKNLI	9.2	0.01861	Unknown	0.333	0.497	0.9044
SFBM_1073	SSMNNNLIIEVKWD	231	245	71.018	SMMNNLI	23	0.01658	Unknown	0.333	0.497	0.9044
SFBM_1073	NFLQIVSGNDRYIED	51	65	81.0705	LQIVSGNDR	28	0.01301	Unknown	0.333	0.497	0.9044
SFBM_1073	NPRQAFLYNVYGRAF	1306	1320	92.4438	FLYNVYGRA	17	0.01003	Unknown	0.333	0.497	0.9044
SFBM_1073	SCDVIFRYDRKNLIF	136	150	79.065	FRYDRKNLI	3.9	0.00073	Unknown	0.333	0.497	0.9044
SFBM_1083	KAILPFVFIILVYF	36	50	89.6009	VFIILVYF	13	0.08048	Cytoplasmic Membrane	0.955	0.378	0.9950
SFBM_1083	FVFIILVYFLFTLF	41	55	83.3792	IILVYFLF	7	0.04738	Cytoplasmic Membrane	0.955	0.378	0.9950
SFBM_1083	VKSAYKAILPFVFI	31	45	91.6056	YKAILPFVF	8.9	0.02882	Cytoplasmic Membrane	0.955	0.378	0.9950
SFBM_1083	ILVYFLFTLTKVI	46	60	82.0194	FTLTKVI	12	0.02754	Cytoplasmic Membrane	0.955	0.378	0.9950
SFBM_1083	VYFLFTLTKVILK	48	62	86.523	FTLTKVI	25	0.02518	Cytoplasmic Membrane	0.955	0.378	0.9950
SFBM_1083	DFWLLVKSAYKAILP	26	40	79.6352	FWLLVKSAY	7.6	0.01473	Cytoplasmic Membrane	0.955	0.378	0.9950
SFBM_1102	MFYILNNSKGMIAQ	1	15	53.5934	FYILNNSKS	14	0.09016	Extracellular	0.972	0.732	0.5091
SFBM_1102	THGYKKIDSSFASLF	31	45	69.4778	YKKIDSSFA	8.9	0.05708	Extracellular	0.972	0.732	0.5091
SFBM_1102	LTASNMIIDRNLVS	146	160	85.9632	NMIIDRNL	16	0.04465	Extracellular	0.972	0.732	0.5091
SFBM_1102	MTEMILAQRAYQMAT	221	235	97.0755	MILAQRAYQ	13	0.04181	Extracellular	0.972	0.732	0.5091

SFBM_1102	QQKLDIISNNMVNVN	16	30	75.8676	IISNNMVNV	8.8	0.02437	Extracellular	0.972	0.732	0.5091
SFBM_1102	FASLFHKDLNKGVP	41	55	65.7767	FHKDLNKG	25	0.00736	Extracellular	0.972	0.732	0.5091
fliz	SKDNYLIIRVIDKY	41	55	82.6528	LIIRVIDKY	20	0.29611	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	FIKVISRVAISKDNY	31	45	68.3286	IKVISRVAI	11	0.16126	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	QSQNLFIKVISRVAI	26	40	77.6273	IKVISRVAI	6.6	0.1284	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	LILILVILMLIYVKL	6	20	84.8063	ILMLIYVKL	14	0.12298	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	VILMLIYVKLNFKGI	11	25	68.3277	ILMLIYVKL	18	0.12099	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	IYVKLNFKGIQSQNL	16	30	86.3706	YVKLNFKGI	8.6	0.09946	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	MNVFSLILILVILML	1	15	96.6537	VFSLILILV	12	0.05888	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	NFKGIQSQNLFIKVI	21	35	95.6135	FKGIQSQNL	6.6	0.03009	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	SRVAISKDNYLIIR	36	50	92.9569	ISKDNYILI	12	0.0273	Cytoplasmic Membrane	0.955	0.398	0.5160
fliz	LIIRVIDKYVLCSS	46	60	73.5797	LIIRVIDKY	15	0.00319	Cytoplasmic Membrane	0.955	0.398	0.5160
SFBM_1113	NKKYVKVSLAMTYDS	66	80	80.0572	YVKVSLAMT	12	0.0938	Unknown	0.25	0.344	0.5951
SFBM_1113	LNIIYKIDAIITFK	91	105	82.5726	IYKIDAI	8	0.06801	Unknown	0.25	0.344	0.5951
SFBM_1113	GFFIISKMNNSDQPK	31	45	80.5391	FIISKMNNS	28	0.06497	Unknown	0.25	0.344	0.5951
SFBM_1113	VVVILLVILGAVG	16	30	87.5791	VVILLVII	17	0.05885	Unknown	0.25	0.344	0.5951
SFBM_1113	KGHIVVILLVII	11	25	92.3456	VVILLVII	5	0.04429	Unknown	0.25	0.344	0.5951
SFBM_1113	VKLEIKNTINSILES	121	135	75.1839	VKLEIKNTI	23	0.03281	Unknown	0.25	0.344	0.5951
motB	ESKHLKDIAGVLNE	136	150	88.7132	ILKDIAGVL	20	0.1015	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	VSVVVVDSLFLQSGR	116	130	97.5896	VVDSLFLQSGR	18	0.08757	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	FASNWELSTARAVV	171	185	92.412	WELSTARAV	18	0.08342	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	KGLKDFISVEIVERG	101	115	95.4089	LKDFISVEI	14	0.07903	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	FILLFSMSTIDAKKE	31	45	79.0712	ILLFSMSTI	12	0.07683	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	INTYMFASNWELSTA	166	180	80.308	TYMFASNWE	17	0.06729	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	HQDSLVLNINELIIE	86	100	87.0925	VNINELII	24	0.04792	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	AKNRRVNILILNKVE	221	235	73.744	VNILILNKV	34	0.03993	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	AQLSMAFNFIENGGG	46	60	76.7342	MAFNFIENG	22	0.03678	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	LLMVFFILLFSMSTI	26	40	81.924	ILLFSMSTI	14	0.02551	Cytoplasmic Membrane	0.878	0.111	0.7307
motB	WMVTFSDAMTLLMVF	16	30	94.1332	MVTFSDAMT	15	0.021	Cytoplasmic Membrane	0.878	0.111	0.7307
SFBM_1245	TINLDFILKTVDIKI	21	35	82.5552	INLDFILKT	12	0.3228	Unknown	0.25	0.160	0.5085

SFBM_1245	SDFVFNIDDGLYQS	141	155	93.4049	FVFNIDDGL	12	0.20875	Unknown	0.25	0.160	0.5085
SFBM_1245	IYYQVLKRYPTKEY	261	275	57.433	YQVLKRYPT	41	0.10145	Unknown	0.25	0.160	0.5085
SFBM_1245	FLKNNKIYMSDIFV	131	145	88.125	FLKNNKIY	20	0.09914	Unknown	0.25	0.160	0.5085
SFBM_1245	KKIIGNLVPGKQYL	211	225	86.4697	IIGNLVPG	16	0.07364	Unknown	0.25	0.160	0.5085
SFBM_1245	GKIFLSSIDTNKLT	46	60	96.0348	SSIDTNKLT	15	0.06862	Unknown	0.25	0.160	0.5085
SFBM_1245	LGKEFVGKIDGKIFL	36	50	94.9164	FVGKIDGKI	18	0.06303	Unknown	0.25	0.160	0.5085
SFBM_1245	CILLVTINLDFILKT	16	30	87.9423	INLDFILKT	8.6	0.05496	Unknown	0.25	0.160	0.5085
SFBM_1245	GLSIDYDKNNKKIII	201	215	74.4463	IDYDKNNKK	31	0.05496	Unknown	0.25	0.160	0.5085
SFBM_1245	SNKMKFIIPKLLMEH	236	250	75.4412	MKFIIPKLL	6	0.04136	Unknown	0.25	0.160	0.5085
SFBM_1245	DKFLNMESSKEYMES	356	370	81.7598	FLNMESSE	28	0.04036	Unknown	0.25	0.160	0.5085
SFBM_1245	RNLGIVAFICLLV	6	20	77.8315	LIGIVAFIC	22	0.03956	Unknown	0.25	0.160	0.5085
SFBM_1245	IPENINISSNSKFTI	176	190	63.9818	INISSNSKF	13	0.03543	Unknown	0.25	0.160	0.5085
SFBM_1245	LYLKIFLKNNEKIYM	126	140	62.3373	LKIFLKNNE	9	0.03093	Unknown	0.25	0.160	0.5085
SFBM_1245	SKFITANFNENKIS	186	200	66.28	FTITANFNE	38	0.02802	Unknown	0.25	0.160	0.5085
SFBM_1245	DDLVKYIMYNLQEI	61	75	86.8266	VKYIMYNLQ	15	0.02475	Unknown	0.25	0.160	0.5085
SFBM_1245	GLYQSYENVITLNDY	151	165	84.4935	YQSYENVIT	27	0.02123	Unknown	0.25	0.160	0.5085
SFBM_1245	EMVDSIYFLSNKKVI	311	325	50.9836	IYFLSNKKV	7.2	0.01929	Unknown	0.25	0.160	0.5085
SFBM_1245	IYFLSNKKVINGRLS	316	330	57.4733	IYFLSNKKV	15	0.01592	Unknown	0.25	0.160	0.5085
SFBM_1245	SYIERFNINIDENFE	96	110	70.0182	FNINIDENF	20	0.01306	Unknown	0.25	0.160	0.5085
SFBM_1245	SYFVKIYYQVLKRYP	256	270	83.1088	VKIYYQVLK	8.4	0.0108	Unknown	0.25	0.160	0.5085
SFBM_1274	GSIYRIRLGNEIIRG	241	255	83.2354	YRIRLGNEI	15	0.17472	Cellwall	0.828	0.294	0.6257
SFBM_1274	NSSDVSFDEPYLKS	176	190	86.9906	FSFDEPYLK	25	0.15349	Cellwall	0.828	0.294	0.6257
SFBM_1274	GEVLYYNDEIVNNAL	151	165	93.112	EVLYYNDEI	19	0.11733	Cellwall	0.828	0.294	0.6257
SFBM_1274	MIKYILKKFSLLLLI	1	15	73.2345	KYILKKFSL	7.1	0.11689	Cellwall	0.828	0.294	0.6257
SFBM_1274	AISKVISGVKIHNVVA	96	110	71.7505	SKVISGVKI	17	0.08487	Cellwall	0.828	0.294	0.6257
SFBM_1274	FSFDEPYLKS VNSLF	181	195	86.5512	PYLKS VNSLF	11	0.07914	Cellwall	0.828	0.294	0.6257
SFBM_1274	VNNALYFAISSGYTE	161	175	89.4735	LYFAISSGY	14	0.05984	Cellwall	0.828	0.294	0.6257
SFBM_1274	MPVKFDIEALKAQAV	76	90	90.2858	IEALKAQAV	20	0.05285	Cellwall	0.828	0.294	0.6257
SFBM_1274	LKHYSYGVKIKKQVSI	310	324	78.3156	HYYSYGVKIK	16	0.04444	Cellwall	0.828	0.294	0.6257
SFBM_1274	PYLKS VNSLFDQDAP	186	200	97.4323	YLKS VNSLF	13	0.0386	Cellwall	0.828	0.294	0.6257
SFBM_1274	LDFRYLFELNSSNID	256	270	94.3001	YLFELNSSN	17	0.03576	Cellwall	0.828	0.294	0.6257
SFBM_1274	ELGYKYDEILKHYS	301	315	89.3014	YKYDEILKH	21	0.02867	Cellwall	0.828	0.294	0.6257
SFBM_1274	LLLLIFMCFVFLPL	11	25	94.2546	IFMCFVFL	20	0.01582	Cellwall	0.828	0.294	0.6257
SFBM_1313	LLILLIIFFLYVYKQ	16	30	88.2239	LILLIIFSF	14	0.15033	Unknown	0.25	0.383	0.5090
SFBM_1313	KIEILKFYIEHIYS	106	120	69.006	ILKFYIEHI	19	0.1364	Unknown	0.25	0.383	0.5090
SFBM_1313	YGSYTLILLIIFSF	11	25	86.7559	LILLIIFSF	25	0.08826	Unknown	0.25	0.383	0.5090
SFBM_1313	IIFFLYVYKQSTYQI	21	35	90.2486	FLYVYKQSTY	14	0.07468	Unknown	0.25	0.383	0.5090
SFBM_1313	STYQINKKLDVIKIS	31	45	67.3868	YQINKKLDV	22	0.01375	Unknown	0.25	0.383	0.5090
SFBM_1460	NFDYSFRNVSSYINY	121	135	89.1935	FRNVSSYIN	11	0.40056	Unknown	0.25	0.534	0.5128

SFBM_1460	FRNVSSYINYLLENL	126	140	93.9255	FRNVSSYIN	20	0.14125	Unknown	0.25	0.534	0.5128
SFBM_1460	KIFSQILIDPIFNY	166	180	89.462	ILIDPIFNY	14	0.12965	Unknown	0.25	0.534	0.5128
SFBM_1460	IINYSNRTKNVILL	61	75	62.4175	IINYSNRTK	36	0.11298	Unknown	0.25	0.534	0.5128
SFBM_1460	YYEKNINYSNRRTK	56	70	66.5293	IINYSNRRT	29	0.1077	Unknown	0.25	0.534	0.5128
SFBM_1460	QFKINNTYITYYEKN	46	60	68.2692	FKINNTYIT	36	0.09864	Unknown	0.25	0.534	0.5128
SFBM_1460	LLICSNFSSIGLNM	146	160	89.4793	LLICSNFSS	15	0.08983	Unknown	0.25	0.534	0.5128
SFBM_1460	SYNYLLENLITDT	131	145	96.0329	NYLLENLNI	15	0.08687	Unknown	0.25	0.534	0.5128
SFBM_1460	NILIYSIILLFFMHI	6	20	93.6835	SIILLFFMH	4.7	0.08621	Unknown	0.25	0.534	0.5128
SFBM_1460	ASKMINNESSPISTLD	231	245	91.165	KMINNESSPI	19	0.08033	Unknown	0.25	0.534	0.5128
SFBM_1460	YIANLSNRFASIIFL	266	280	71.9221	IANLSNRFA	6.4	0.07737	Unknown	0.25	0.534	0.5128
SFBM_1460	IINCMYIFYNSKNL	21	35	52.1123	MYIFYNSKN	22	0.07499	Unknown	0.25	0.534	0.5128
SFBM_1460	SKNNLYYINNFNEK	31	45	70.7675	YYINNFNEK	36	0.07213	Unknown	0.25	0.534	0.5128
SFBM_1460	NSKLFIKNKLSS	186	200	67.3409	SKLFIKNKL	8.9	0.07176	Unknown	0.25	0.534	0.5128
SFBM_1460	YTTYFNKNKSSNKY	216	230	60.6572	TYFNKNKNS	35	0.06956	Unknown	0.25	0.534	0.5128
SFBM_1460	IGLNMISLNDKIFSQ	156	170	74.8509	NMISLNDKI	15	0.06884	Unknown	0.25	0.534	0.5128
SFBM_1460	ISTLDIRSNIPIFAI	241	255	79.714	LDIRSNIPI	12	0.06638	Unknown	0.25	0.534	0.5128
SFBM_1460	IRSNIPIFAINRKS	246	260	90.5248	IFAINRKS	18	0.06495	Unknown	0.25	0.534	0.5128
SFBM_1460	YFNSTYIANLSNRFA	261	275	59.2178	IANLSNRFA	4.2	0.06172	Unknown	0.25	0.534	0.5128
SFBM_1460	FILLNINKSSLNLSN	201	215	57.792	FILLNINKS	15	0.04941	Unknown	0.25	0.534	0.5128
SFBM_1460	IKNKLLSSFIILN	191	205	82.8066	LLSSFIILN	12	0.04778	Unknown	0.25	0.534	0.5128
SFBM_1460	FFMHIIICMYIFYN	16	30	68.1062	FMHIIICM	14	0.04762	Unknown	0.25	0.534	0.5128
SFBM_1460	CNFIIDLDPGHGK	101	115	92.3508	FNIFLIDL	9.9	0.03906	Unknown	0.25	0.534	0.5128
SFBM_1460	NFNEKQFKINNTYIT	41	55	66.4124	FKINNTYIT	26	0.03672	Unknown	0.25	0.534	0.5128
SFBM_1460	YIFYNSKNLNYYIN	26	40	88.5724	IFYNSKNL	18	0.03616	Unknown	0.25	0.534	0.5128
SFBM_1460	LLSSFIILNINKSS	196	210	43.6691	FILNINKS	11	0.01859	Unknown	0.25	0.534	0.5128
SFBM_1460	PIFAINRKSIFYNST	251	265	73.244	PIFAINRK	11	0.01806	Unknown	0.25	0.534	0.5128
SFBM_1460	IILIDPIFYNPWE	171	185	87.2721	IFYNPWE	22	0.01196	Unknown	0.25	0.534	0.5128
SFBM_1460	SIILLFFMHIIICM	11	25	71.4682	LLFFMHII	20	0.00286	Unknown	0.25	0.534	0.5128
SFBM_1497	RDVFNVKRFLFDYY	126	140	84.0691	RDVFNVKK	18	0.1909	Extracellular	0.96	0.599	0.5128
SFBM_1497	KYVGMKVKSTKMLRS	246	260	68.4146	MKKVSTKML	18	0.17536	Extracellular	0.96	0.599	0.5128
SFBM_1497	NGNQIGVTKNKIFN	41	55	74.6136	IGVTKNKI	30	0.16642	Extracellular	0.96	0.599	0.5128
SFBM_1497	NIKSQEIVKMDSLDF	66	80	91.6982	IKSQEIVKM	15	0.15195	Extracellular	0.96	0.599	0.5128
SFBM_1497	FYKAYSFLGDKKI	106	120	85.1597	YSLFLGDKK	17	0.14983	Extracellular	0.96	0.599	0.5128
SFBM_1497	YVILVNGNQIVTKN	36	50	68.1141	YVILVNGNQ	22	0.14741	Extracellular	0.96	0.599	0.5128
SFBM_1497	QNGELFYLNNSPKGN	266	280	73.7767	FYLNNSPKG	16	0.11251	Extracellular	0.96	0.599	0.5128
SFBM_1497	GNMMLNHGKGVETL	326	340	73.0195	YMLNHGKG	25	0.09963	Extracellular	0.96	0.599	0.5128
SFBM_1497	FILILYLISGKTLG	21	35	79.6269	ILYLISGKT	17	0.0735	Extracellular	0.96	0.599	0.5128
SFBM_1497	LYLISGKTLGYVILV	26	40	90.0203	LYLISGKTL	21	0.07243	Extracellular	0.96	0.599	0.5128
SFBM_1497	YSLFLGDKKIANIDN	111	125	83.2373	YSLFLGDKK	18	0.07104	Extracellular	0.96	0.599	0.5128

SFBM_1497	DENLDFYIKAYSLFL	101	115	90.1736	FYIKAYSLF	12	0.06095	Extracellular	0.96	0.599	0.5128
SFBM_1497	GNIFAINIIFILI	11	25	57.9409	GNIFAINI	4.5	0.06002	Extracellular	0.96	0.599	0.5128
SFBM_1497	DSLDFKIGLVNKIEC	76	90	69.1264	FKIGLVNKI	15	0.05939	Extracellular	0.96	0.599	0.5128
SFBM_1497	GKTLGYVILVNGNQI	31	45	68.5977	VILVNGNQI	15	0.05718	Extracellular	0.96	0.599	0.5128
SFBM_1497	EIVKMSLDFKIGLV	71	85	89.6682	KMSLDFKI	14	0.04418	Extracellular	0.96	0.599	0.5128
SFBM_1497	AINIIFILILYLIS	16	30	95.7481	IFILILYL	12	0.03902	Extracellular	0.96	0.599	0.5128
SFBM_1497	MHNFKGKINIGNIIF	1	15	62.9539	GKINIGNI	31	0.03412	Extracellular	0.96	0.599	0.5128
SFBM_1497	AESGFANLVSYNNVY	311	325	61.5145	FANLVSYNN	33	0.03079	Extracellular	0.96	0.599	0.5128
SFBM_1497	ANLVSYNNVYGNVYMK	316	330	67.4112	YNNVYGNVYMK	24	0.0179	Extracellular	0.96	0.599	0.5128
SFBM_1497	GKINIGNIIFAINII	6	20	60.3295	GKINIGNI	14	0.01755	Extracellular	0.96	0.599	0.5128
SFBM_1507	MIKYIKLFFLNIFIL	1	15	87.2215	LLFFLNIFIL	16	0.13825	Unknown	0.25	0.652	0.5042
SFBM_1507	LLNIFIEENKNMKEY	116	130	70.911	LNIFIEENK	27	0.13736	Unknown	0.25	0.652	0.5042
SFBM_1507	RYLILLLNIFIEENK	111	125	65.3549	YLILLLNIF	13	0.09205	Unknown	0.25	0.652	0.5042
SFBM_1507	NQEKFFLRVSEKYDI	31	45	81.2435	FFLRVSEKY	16	0.08003	Unknown	0.25	0.652	0.5042
SFBM_1507	NIFILIGMYNLSAHS	11	25	70.7658	IFILIGMYN	15	0.07753	Unknown	0.25	0.652	0.5042
SFBM_1507	NKYIRNLNQSIRYGN	91	105	80.3069	IRNLNQSIR	17	0.07265	Unknown	0.25	0.652	0.5042
SFBM_1507	EYMEIKKSDSFRLLK	76	90	78.0015	IKKSDSFRLL	15	0.07081	Unknown	0.25	0.652	0.5042
SFBM_1507	TSFEYIKSNILFKSS	136	150	74.9845	FEYIKSNIL	0.83	0.06679	Unknown	0.25	0.652	0.5042
SFBM_1507	FYLYNTSFEYIKSNI	131	145	83.9446	SFEYIKSNI	20	0.06092	Unknown	0.25	0.652	0.5042
SFBM_1507	FEYIKSNILFKSSSTE	138	152	71.8235	FEYIKSNIL	2.4	0.05567	Unknown	0.25	0.652	0.5042
SFBM_1507	SFRLKNKYIRNLNQS	86	100	82.361	LKNKYIRNL	25	0.04297	Unknown	0.25	0.652	0.5042
SFBM_1507	NMKEYFYLYNTSFEY	126	140	83.034	YFYLYNTSFEY	17	0.0403	Unknown	0.25	0.652	0.5042
SFBM_1507	KLFFLNIFILIGMYN	6	20	80.4525	LNIFILIGM	19	0.03265	Unknown	0.25	0.652	0.5042
SFBM_1507	IGMYNLSAHSILELS	16	30	71.7695	YNLSAHSIL	18	0.0285	Unknown	0.25	0.652	0.5042
SFBM_1507	IKKSDSFRLLKNKYIR	81	95	71.4282	FRLLKNKYIR	19	0.02722	Unknown	0.25	0.652	0.5042
SFBM_1507	KKKIKRYLILLLNIF	106	120	80.3266	YLILLLNIF	25	0.01966	Unknown	0.25	0.652	0.5042
SFBM_1507	IKYGNKKKIKRYLIL	101	115	85.3714	NKKKIKRYL	22	0.01892	Unknown	0.25	0.652	0.5042

**APPENDIX C: SFB PROTEINS IDENTIFIED IN SITU IN THE MURINE ILEAL
MUCOSA OF C57BL/6J, DBA/2, AND C3H MICE**

Entry name	Protein names	Length	Redundancy	Gene names (primary)	Gene names (ORF)	Gene ontology IDs	Cross-reference (Pfam)	Average Protein ID Prob.	# of Unique Peptides
F9VJG3_ARTSS	1-deoxy-D-xylulose-5-phosphate synthase (EC 2.2.1.7) (1-deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS)	610	Yes	dxs	SFBM_0492	GO:0000287; GO:0008661; GO:0009228; GO:0016114; GO:0030976; GO:0052865	PF13292;PF02779 ;PF02780;	37%	2
G2IGN5_9CLOT	1-deoxy-D-xylulose-5-phosphate synthase (EC 2.2.1.7) (1-deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS)	609	Yes	dxs	RATSF0401	GO:0000287; GO:0008661; GO:0009228; GO:0016114; GO:0030976; GO:0052865	PF13292;PF02779 ;PF02780;	94%	2
G2IH05_9CLOT	Acetate kinase (EC 2.7.2.1) (Acetokinase)	398	No	ackA	RATSF0521	GO:0000287; GO:0005524; GO:0005737; GO:0006082; GO:0006085; GO:0008776	PF00871;	60%	2
G2IH4_9CLOT	Alanine-tRNA ligase (EC 6.1.1.7) (Alanyl-tRNA synthetase) (AlaRS)	877	No	alaS	RATSF0700	GO:0000049; GO:0004813; GO:0005524; GO:0005737; GO:0006419; GO:0008270	PF02272;PF01411 ;PF07973;	79%	3
F9VLU5_ARTSS	Aldehyde-alcohol dehydrogenase	868	No		SFBM_1344	GO:0004022; GO:0006066; GO:0008774; GO:0015976; GO:0046872	PF00171;PF00465 ;	88%	2
F9VLH5_ARTSS	Arginine-tRNA ligase (EC 6.1.1.19) (Arginyl-tRNA synthetase) (ArgRS)	564	No	argS	SFBM_1222	GO:0004814; GO:0005524; GO:0005737; GO:0006420	PF03485;PF05746 ;PF00750;	100%	4
F9VLF6_ARTSS	3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS)	424	No	aroA	SFBM_1203	GO:0003866; GO:0005737; GO:0009073; GO:0009423	PF00275;	100%	6
F9VKG6_ARTSS	Aspartate-tRNA ligase (EC 6.1.1.12) (Aspartyl-	588	Yes	aspS	SFBM_0855	GO:0003676; GO:0004815; GO:0005524;	PF02938;PF00152 ;PF01336;	77%	2

	tRNA synthetase) (AspRS)					GO:0005737; GO:0006422			
G2IH8_9CLOT	Aspartate--tRNA ligase (EC 6.1.1.12) (Aspartyl- tRNA synthetase) (AspRS)	588	Yes	aspS	RATSF0714	GO:0003676; GO:0004815; GO:0005524; GO:0005737; GO:0006422	PF02938;PF00152 ;PF01336;	56%	2
G2IGW1_9CLOT	Flagellar hook- associated protein 1 (HAP1)	578	No	flgK	RATSF0477	GO:0005198; GO:0005576; GO:0009424; GO:0044780	PF06429;	100%	4
F9VKR2_ARTSS	Basal-body rod modification protein FlgD	343	No	flgD	SFBM_0954	GO:0044781	PF03963;	54%	2
G2IGW6_9CLOT	Flagellar hook- associated protein 2 (HAP2) (Flagellar cap protein)	797	No		RATSF0482	GO:0005576; GO:0007155; GO:0009421; GO:0009424; GO:0071973	PF07195;PF02465 ;	66%	2
F9VJU9_ARTSS	Flagellin	400	No		SFBM_0583	GO:0005198; GO:0005576; GO:0009420; GO:0071973	PF00700;PF00669 ;	92%	9
G2IE06_9CLOT	Flagellin	279	No		RATSF0818	GO:0005198; GO:0005576; GO:0009420; GO:0071973	PF00700;PF00669 ;	48%	2
G2IE86_9CLOT	Aminotransferase (EC 2.6.1.-)	379	No		RATSF0898	GO:0008483; GO:0009058; GO:0030170	PF00155;	49%	2
F9VIJ3_ARTSS	Mannose-1-phosphate guanylyltransferase	358	No		SFBM_0326	GO:0009058; GO:0016779	PF00483;	65%	4
F9VKN1_ARTSS	WecB/TagA/CpsF family glycosyltransferase	245	No		SFBM_0923	GO:0009058; GO:0016757	PF03808;	100%	3
F9VLM2_ARTSS	N-acetylglucosamine-6- phosphate deacetylase	381	Yes	nagA	SFBM_1270	GO:0005975; GO:0006044; GO:0008448; GO:0046872	PF01979;	97%	2
G2IEI7_9CLOT	N-acetylglucosamine-6- phosphate deacetylase	381	Yes	nagA	RATSF01099	GO:0005975; GO:0006044; GO:0008448; GO:0046872	PF01979;	87%	4
F9VJ05_ARTSS	Beta-galactosidase (EC 3.2.1.23) (Lactase)	1187	No		SFBM_0186	GO:0004565; GO:0005975; GO:0009341; GO:0030246	PF02929;PF16353 ;PF00703;PF0283 6;PF02837;	66%	2
F9VLU4_ARTSS	Dipicolinate synthase subunit A	2749	No		SFBM_1343	GO:0003824; GO:0005975;	PF17167;PF06165 ;PF10091;	100%	6

						GO:0016021; GO:0030246			
F9VKZ2_ARTSS	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) (D-fructose-6-phosphate amidotransferase) (GFAT) (Glucosamine-6-phosphate synthase) (Hexosephosphate aminotransferase) (L-glutamine--D-fructose-6-phosphate amidotransferase)	608	No	glmS	SFBM_1036	GO:0004360; GO:0005737; GO:0005975; GO:0097367; GO:1901137	PF01380;	86%	4
G2IENS_9CLOT	Glycosyltransferase 36	781	No		RATSF_1047	GO:0005975; GO:0016740; GO:0030246	PF17167;PF06165 ;	100%	2
G2IES4_9CLOT	Peptidoglycan N-acetylglucosamine deacetylase	254	No	pgdA	RATSF_1086	GO:0005975; GO:0016810	PF01522;	70%	2
F9VJ69_ARTSS	Uncharacterized protein	404	No		SFBM_0398	GO:0003824; GO:0005975; GO:0016021		99%	2
F9VJCS_ARTSS	Uncharacterized protein	473	No		SFBM_0454	GO:0005975; GO:0016773	PF02782;PF00370 ;	93%	2
F9VKC2_ARTSS	Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94) (NAD(P)H-dependent glycerol-3-phosphate dehydrogenase)	332	No	gpsA	SFBM_0810	GO:0004367; GO:0005975; GO:0006650; GO:0008654; GO:0009331; GO:0046167; GO:0046168; GO:0051287	PF07479;PF01210 ;	100%	2
G2IGA4_9CLOT	HPr kinase/phosphorylase (HPrK/P) (EC 2.7.11.-) (EC 2.7.4.-) (HPr(Ser) kinase/phosphorylase)	303	No	hprK	RATSF_0270	GO:0000155; GO:0000287; GO:0004674; GO:0004712; GO:0005524; GO:0005975; GO:0006109	PF07475;PF02603 ;	100%	2
F9VK45_ARTSS	ABC-type sugar transport system	301	No		SFBM_0683	GO:0008643	PF13407;	100%	2
F9VJP8_ARTSS	DNA translocase FtsK/SpoIIIE	758	No		SFBM_0740	GO:0003677; GO:0005524; GO:0005886; GO:0007049; GO:0007059;	PF13491;PF17854 ;PF09397;PF0158 0;	100%	2

						GO:0016021; GO:0051301			
F9VM35_ARTSS	ATP-dependent zinc metalloprotease FtsH (EC 3.4.24.-)	601	No	ftsH	SFBM_1434	GO:0004222; GO:0005524; GO:0005886; GO:0008270; GO:0016021; GO:0016887; GO:0030163; GO:0051301	PF00004;PF17862 ;PF06480;PF01434;	98%	2
F9VM55_ARTSS	Bifunctional protein GlmU [Includes: UDP-N-acetylglucosamine pyrophosphorylase (EC 2.7.7.23) (N-acetylglucosamine-1-phosphate uridylyltransferase); Glucosamine-1-phosphate N-acetyltransferase (EC 2.3.1.157)]	453	No	glmU	SFBM_1454	GO:0000287; GO:0000902; GO:0003977; GO:0005737; GO:0006048; GO:0008360; GO:0009103; GO:0009245; GO:0009252; GO:0019134; GO:0071555	PF00132;PF12804 ;	82%	9
G2IFE4_9CLOT	Glutamate racemase (EC 5.1.1.3)	262	No	murI	RATSFB_1306	GO:0008360; GO:0008881; GO:0009252; GO:0071555	PF01177;	99%	5
F9VKW6_ARTSS	Copper homeostasis protein CutC	249	Yes	cutC	SFBM_1010	GO:0005507; GO:0005623; GO:0005737; GO:0006878	PF03932;	100%	2
G2IE49_9CLOT	Copper homeostasis protein CutC	252	Yes	cutC	RATSFB_0861	GO:0005507; GO:0005623; GO:0005737; GO:0006878	PF03932;	92%	7
G2IH86_9CLOT	Uridylate kinase (UK) (EC 2.7.4.22) (Uridine monophosphate kinase) (UMP kinase) (UMPK)	235	No	pyrH	RATSFB_0602	GO:0005524; GO:0005737; GO:0033862; GO:0044210	PF00696;	69%	2
G2IEV3_9CLOT	CTP synthase (EC 6.3.4.2) (Cytidine 5'-triphosphate synthase) (Cytidine triphosphate synthetase) (CTP synthetase) (CTPS) (UTP--ammonia ligase)	536	No	pyrG	RATSFB_1115	GO:0003883; GO:0005524; GO:0006541; GO:0044210; GO:0046872	PF06418;PF00117 ;	92%	3
F9VJG0_ARTSS	Aspartate-semialdehyde dehydrogenase (ASA dehydrogenase) (ASADH) (EC 1.2.1.11) (Aspartate-beta-	330	No	asd	SFBM_0489	GO:0004073; GO:0009088; GO:0009089; GO:0009097; GO:0019877;	PF01118;PF02774 ;	100%	2

	semialdehyde dehydrogenase)					GO:0046983; GO:0050661; GO:0051287; GO:0071266			
F9VK01_ARTSS	Multifunctional fusion protein [Includes: Cytidylate kinase (CK) (EC 2.7.4.25) (Cytidine monophosphate kinase) (CMP kinase); 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HMBPP reductase) (EC 1.17.7.4)]	850	No	ispH	SFBM_0636	GO:0003676; GO:0004127; GO:0005524; GO:0005737; GO:0006220; GO:0016114; GO:0019288; GO:0046872; GO:0050992; GO:0051539; GO:0051745	PF02224;PF02401 ;PF00575;	54%	2
G2IHC0_9CLOT	Restriction modification system DNA specificity domain-containing protein	387	No		RATSF06_0636	GO:0003677; GO:0006304	PF01420;	100%	8
F9VL36_ARTSS	Type II restriction enzyme, methylase subunit	1155	No		SFBM_1080	GO:0003677; GO:0006304; GO:0008168; GO:0032259	PF07669;	100%	2
F9VJK1_ARTSS	DNA repair protein RecO (Recombination protein O)	246	No	recO	SFBM_0531	GO:0006281; GO:0006310	PF02565;PF11967 ;	80%	2
G2IFL9_9CLOT	Recombination protein RecR	199	No	recR	RATSF06_0035	GO:0003677; GO:0006281; GO:0006310; GO:0046872	PF02132;PF13662 ;	54%	2
F9VL31_ARTSS	DNA ligase (EC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD(+)])	659	Yes	ligA	SFBM_1075	GO:0003677; GO:0003911; GO:0006260; GO:0006281; GO:0046872	PF00533;PF01653 ;PF03120;PF1282 6;	84%	7
G2IEA8_9CLOT	DNA ligase (EC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD(+)])	659	Yes	ligA	RATSF06_0920	GO:0003677; GO:0003911; GO:0006260; GO:0006281; GO:0046872	PF00533;PF01653 ;PF03120;PF1282 6;	100%	3
F9VI35_ARTSS	ATP-dependent helicase/deoxyribonuclease subunit B (EC 3.1.-.-) (EC 3.6.4.12)	1131	No		SFBM_0010	GO:0003677; GO:0003678; GO:0004527; GO:0005524; GO:0006281	PF12705;	100%	2
G2IFJ4_9CLOT	ATP-dependent helicase/deoxyribonuclease subunit B (EC 3.1.-.-) (EC 3.6.4.12)	1126	No	addB	RATSF06_0010	GO:0003677; GO:0003678; GO:0004527; GO:0005524; GO:0006281	PF12705;	77%	2

F9VJZ8_ARTSS	Endonuclease IV	273	No		SFBM_0633	GO:0003677; GO:0004519; GO:0006281; GO:0008270	PF01261;	100%	2
F9VI28_ARTSS	DNA replication and repair protein RecF	366	No	recF	SFBM_0003	GO:0003697; GO:0005524; GO:0005737; GO:0006260; GO:0006281; GO:0009432	PF02463;	100%	2
G2IGY0_9CLOT	Primosomal protein N' (EC 3.6.4.-) (ATP-dependent helicase PriA)	740	No	priA	RATSF0496	GO:0003677; GO:0003678; GO:0005524; GO:0006268; GO:0006269; GO:0008270; GO:1990077	PF00270;PF00271 ;PF17764;PF1807 4;PF18319;	97%	2
G2IFR7_9CLOT	DNA polymerase III subunit alpha (EC 2.7.7.7)	1182	No	dnaE	RATSF0083	GO:0003887; GO:0005737; GO:0006260; GO:0008408	PF07733;PF17657 ;PF14579;PF0281 1;	76%	2
F9VI46_ARTSS	Type I restriction enzyme R Protein (EC 3.1.21.3)	1011	No		SFBM_0021	GO:0003677; GO:0005524; GO:0009035; GO:0009307	PF11867;PF04313 ;PF18766;	100%	11
F9VKB2_ARTSS	Type I restriction enzyme R Protein (EC 3.1.21.3)	1040	No		SFBM_0800	GO:0003677; GO:0005524; GO:0009035; GO:0009307	PF12008;PF04313 ;PF18766;	49%	2
F9VKB4_ARTSS	Type I restriction-modification system, M subunit	524	No		SFBM_0802	GO:0003677; GO:0008170; GO:0009007; GO:0009307	PF12161;PF02384 ;	68%	2
G2IH92_9CLOT	DNA polymerase III PolC-type (PolIII) (EC 2.7.7.7)	1398	No	polC	RATSF0608	GO:0003677; GO:0003887; GO:0005737; GO:0006261; GO:0008408	PF07733;PF17657 ;PF02811;PF0092 9;	85%	3
G2IEV2_9CLOT	Transcription termination factor Rho (EC 3.6.4.-) (ATP-dependent helicase Rho)	467	No	rho	RATSF0114	GO:0003723; GO:0004386; GO:0005524; GO:0006353; GO:0008186	PF00006;PF07498 ;PF07497;	75%	4
F9VI36_ARTSS	ATP-dependent helicase/nuclease subunit A (EC 3.1.-.-) (EC 3.6.4.12)	1193	No		SFBM_0011	GO:0003677; GO:0003678; GO:0004527; GO:0005524; GO:0006302	PF12705;PF00580 ;PF13361;	100%	2

G2IGU4_9CLOT	Pyruvate ferredoxin/flavodoxin oxidoreductase	1177	No	niJ	RATSF0460	GO:0005506; GO:0016903; GO:0022900; GO:0030976; GO:0051539	PF10371;PF17147 ;PF01558;PF0185 5;PF02775;	95%	2
F9VLQ1_ARTSS	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9)	261	No		SFBM_1299	GO:0004318; GO:0006633; GO:0016631		100%	2
F9VKF6_ARTSS	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha (ACCase subunit alpha) (Acetyl-CoA carboxylase carboxyltransferase subunit alpha) (EC 2.1.3.15)	268	No	accA	SFBM_0845	GO:0003989; GO:0005524; GO:0006633; GO:0009317; GO:0016743; GO:2001295	PF03255;	62%	2
G2IH9_9CLOT	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (ACCase subunit beta) (Acetyl-CoA carboxylase carboxyltransferase subunit beta) (EC 2.1.3.15)	289	No	accD	RATSF0705	GO:0003989; GO:0005524; GO:0006633; GO:0008270; GO:0009317; GO:0016743; GO:2001295	PF01039;PF17848 ;	90%	2
G2IH8_9CLOT	Phosphate acyltransferase (EC 2.3.1.274) (Acyl-ACP phosphotransacylase) (Acyl-[acyl-carrier-protein]-phosphate acyltransferase) (Phosphate-acyl-ACP acyltransferase)	330	No	plsX	RATSF0524	GO:0005737; GO:0006633; GO:0008654; GO:0043811	PF02504;	99%	10
F9VIA7_ARTSS	ATP-dependent 6-phosphofructokinase (ATP-PFK) (Phosphofructokinase) (EC 2.7.1.11) (Phosphohexokinase)	319	Yes	pfkA	SFBM_0083	GO:0003872; GO:0005524; GO:0005737; GO:0006002; GO:0046872	PF00365;	69%	2
G2IEQ2_9CLOT	ATP-dependent 6-phosphofructokinase (ATP-PFK) (Phosphofructokinase) (EC 2.7.1.11) (Phosphohexokinase)	364	Yes	pfkA	RATSF01064	GO:0003872; GO:0005524; GO:0005737; GO:0006002; GO:0046872; GO:0047334	PF00365;	55%	2
G2IEV4_9CLOT	UDP-glucuronosyltransferase	373	No		RATSF01116	GO:0009247; GO:0016758	PF04101;PF06925 ;	70%	2

G2IHK1_9CLOT	RelA/SpoT protein, (PppGpp synthetase/pyrophosphohydrolase)	719	No	spoT	RATSF0717	GO:0015969; GO:0016787	PF13291;PF04607 ;PF02824;	100%	3
G2IFA8_9CLOT	Hypoxanthine phosphoribosyltransferase (EC 2.4.2.8)	179	No	hpt_2	RATSF01270	GO:0000166; GO:0004422; GO:0005737; GO:0006166; GO:0032264; GO:0046872; GO:0052657	PF00156;	60%	2
F9VLL7_ARTSS	Protein translocase subunit SecA	833	No	secA	SFBM_1265	GO:0005524; GO:0005737; GO:0005886; GO:0006605; GO:0017038; GO:0046872; GO:0065002	PF02810;PF07517 ;PF01043;PF07516;	72%	2
F9VJL5_ARTSS	Branched-chain-amino-acid aminotransferase (BCAT) (EC 2.6.1.42)	292	No	ilvE	SFBM_0545	GO:0009097; GO:0009098; GO:0009099; GO:0052654; GO:0052655; GO:0052656	PF01063;	59%	2
F9VJG2_ARTSS	1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXP reductoisomerase) (EC 1.1.1.267) (1-deoxyxylulose-5-phosphate reductoisomerase) (2-C-methyl-D-erythritol 4-phosphate synthase)	383	No	dxr	SFBM_0491	GO:0016114; GO:0016853; GO:0019288; GO:0030604; GO:0046872; GO:0070402	PF08436;PF02670 ;PF13288;	100%	2
F9VJN3_ARTSS	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (flavodoxin) (EC 1.17.7.3) (1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase)	349	No	ispG	SFBM_0725	GO:0005506; GO:0016114; GO:0019288; GO:0046429; GO:0051539	PF04551;	100%	4
G2IEZ7_9CLOT	Leucine--tRNA ligase (EC 6.1.1.4) (Leucyl-tRNA synthetase) (LeuRS)	820	No	leuS	RATSF01159	GO:0002161; GO:0004823; GO:0005524; GO:0005737; GO:0006429	PF08264;PF00133 ;PF13603;	88%	4
F9VIC2_ARTSS	Phosphatidylglycerol--prolipoprotein diacylglyceryl	264	No	lgt	SFBM_0255	GO:0005887; GO:0008961; GO:0042158	PF01790;	68%	2

	transferase (EC 2.5.1.145)								
F9VIX4_ARTSS	Diaminopimelate decarboxylase (DAP decarboxylase) (DAPDC) (EC 4.1.1.20)	426	No	lysA	SFBM_0154	GO:0008836; GO:0009089; GO:0030170	PF02784;PF00278	57%	2
G2IFY5_9CLOT	Mannitol-1-phosphate 5-dehydrogenase (EC 1.1.1.17)	376	No	mtlD	RATSF0151	GO:0008926; GO:0019594; GO:0050662	PF01232;PF08125	100%	2
F9VKA2_ARTSS	Alpha-mannosidase	1043	No		SFBM_0790	GO:0004559; GO:0006013; GO:0030246	PF09261;PF17677 ;PF07748;PF01074	60%	2
F9VJ60_ARTSS	Cation transport ATPase	742	Yes		SFBM_0241	GO:0005524; GO:0005886; GO:0016021; GO:0019829; GO:0030001; GO:0046872	PF00403;	60%	2
G2IG14_9CLOT	Copper-translocating P-type ATPase	742	Yes		RATSF0180	GO:0005524; GO:0005886; GO:0016021; GO:0019829; GO:0030001; GO:0046872	PF00403;	100%	4
G2IH36_9CLOT	Hemolysin A	273	No		RATSF0552	GO:0003723; GO:0008168; GO:0032259	PF01728;PF01479	100%	2
G2IHB8_9CLOT	DNA mismatch repair protein MutS	853	No	mutS	RATSF0634	GO:0003684; GO:0005524; GO:0006298; GO:0030983	PF01624;PF05188 ;PF05192;PF05190;PF00488;	36%	2
G2IG95_9CLOT	Redox-sensing transcriptional repressor Rex	208	No	rex	RATSF0261	GO:0003677; GO:0003700; GO:0005737; GO:0045892; GO:0050662; GO:0051775	PF02629;PF06971	43%	2
G2IET4_9CLOT	Putative ComF protein	219	No	comF	RATSF01096	GO:0009116		58%	2
F9VI96_ARTSS	UvrABC system protein A (UvrA protein) (Excinuclease ABC subunit A)	938	No	uvrA	SFBM_0072	GO:0003677; GO:0005524; GO:0005737; GO:0006289; GO:0008270; GO:0009380; GO:0009381; GO:0009432; GO:0016887	PF00005;PF17755 ;PF17760;	100%	3
F9VLJ9_ARTSS	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)	330	No		SFBM_1246	GO:0008460; GO:0009225	PF16363;	100%	2

F9VJE9_ARTSS	Phenylalanine--tRNA ligase alpha subunit (EC 6.1.1.20) (Phenylalanyl-tRNA synthetase alpha subunit) (PheRS)	339	No	pheS	SFBM_0478	GO:0000049; GO:0000287; GO:0004826; GO:0005524; GO:0005737; GO:0006432	PF02912;PF01409	60%	2
G2IF85_9CLOT	Protein-arginine kinase (EC 2.7.14.1)	345	No	mcsB	RATSF01247	GO:0004111; GO:0004672; GO:0005524; GO:0046314	PF00217;	60%	3
G2IE38_9CLOT	Permease IIC component	464	No		RATSF010850	GO:0005886; GO:0008982; GO:0009401; GO:0016021	PF02378;	71%	2
G2IF25_9CLOT	Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9) (Phosphotransferase system, enzyme I)	541	No		RATSF01187	GO:0005737; GO:0008965; GO:0009401; GO:0016301; GO:0046872	PF05524;PF00391 ;PF02896;	74%	3
G2IE39_9CLOT	Phosphotransferase system lactose/cellobiose-specific IIB subunit	100	No		RATSF010851	GO:0008982; GO:0009401	PF02302;	98%	4
G2IH78_9CLOT	PTS system, lactose/cellobiose specific IIB subunit	93	No		RATSF010594	GO:0008982; GO:0009401	PF02302;	87%	5
F9VKA5_ARTSS	AraC family DNA-binding response regulator	509	No		SFBM_0793	GO:0000160; GO:0003700; GO:0043565	PF12833;PF00072	46%	2
F9VJT6_ARTSS	Chemotaxis protein CheY	119	No	cheY	SFBM_0570	GO:0000160	PF00072;	54%	2
G2IHE4_9CLOT	DNA-binding response regulator, AraC family	508	No		RATSF010660	GO:0000160; GO:0003700; GO:0043565	PF12833;PF00072	100%	2
G2IG88_9CLOT	V-type ATP synthase beta chain (V-ATPase subunit B)	460	No	atpB	RATSF010254	GO:0005524; GO:0042777; GO:0046933	PF00006;PF02874	100%	2
F9VL64_ARTSS	Flagellar biosynthetic protein FliR/FliB	611	No	fliR	SFBM_1109	GO:0006605; GO:0009306; GO:0016021	PF01311;PF01312	73%	2
F9VJG5_ARTSS	M16C-associated domain-containing protein	998	No		SFBM_0494	GO:0003824; GO:0006508; GO:0046872	PF08367;PF00675 ;PF05193;	79%	2
G2IGN6_9CLOT	Peptidase M16 domain protein	984	No		RATSF010402	GO:0003824; GO:0006508; GO:0046872	PF08367;PF00675 ;PF05193;	50%	2
F9VKB1_ARTSS	Pyridoxal 5'-phosphate synthase subunit PdxS (PLP synthase subunit	291	No	pdxS	SFBM_0799	GO:0036381; GO:0042823	PF01680;	100%	5

	PdxS) (EC 4.3.3.6) (Pdx1)								
F9VJK3_ARTSS	Pyruvate, phosphate dikinase (EC 2.7.9.1)	875	No		SFBM_0533	GO:0005524; GO:0006090; GO:0016301; GO:0046872; GO:0050242	PF00391;PF02896 ;PF01326;	50%	4
G2IEX5_9CLOT	Iron-sulfur cluster- binding protein	336	No		RATSF_1137	GO:0008033; GO:0008616; GO:0016491; GO:0051536	PF08331;	100%	2
F9VJW8_ARTSS	16S rRNA (cytosine(967)-C(5))- methyltransferase (EC 2.1.1.176)	441	No		SFBM_0602	GO:0003723; GO:0005737; GO:0006355; GO:0008649	PF01189;PF01029 ;	62%	2
F9VJX2_ARTSS	Small ribosomal subunit biogenesis GTPase RsgA (EC 3.6.1.-)	295	No	rsgA	SFBM_0606	GO:0003924; GO:0005525; GO:0005737; GO:0019843; GO:0042274; GO:0046872	PF03193;PF16745 ;	100%	3
G2IEJ9_9CLOT	23S rRNA (Uracil-5)- methyltransferase RumA	446	No	rumA1	RATSF_1011	GO:0006396; GO:0008173	PF05958;	68%	2
G2IGY5_9CLOT	Probable dual-specificity RNA methyltransferase RimN (EC 2.1.1.192) (23S rRNA (adenine(2503)-C(2))- methyltransferase) (23S rRNA m2A2503 methyltransferase) (Ribosomal RNA large subunit methyltransferase N) (tRNA (adenine(37)- C(2))-methyltransferase) (tRNA m2A37 methyltransferase)	336	No	rimN	RATSF_0501	GO:0000049; GO:0002935; GO:0005737; GO:0019843; GO:0046872; GO:0051539; GO:0070040; GO:0070475	PF04055;	100%	2
F9VJJ7_ARTSS	Endoribonuclease YbeY (EC 3.1.-.-)	165	No	ybeY	SFBM_0527	GO:0004222; GO:0004521; GO:0005737; GO:0006364; GO:0008270	PF02130;	67%	2
G2IH10_9CLOT	Ribonuclease J (RNase J) (EC 3.1.-.-)	607	No	mj	RATSF_0696	GO:0003723; GO:0004521; GO:0004534; GO:0005737; GO:0006364; GO:0008270	PF00753;PF07521 ;PF17770;	100%	6

G2IG1_9CLOT	Selenocysteine-specific translation elongation factor	633	No	selB	RATSF0347	GO:0001514; GO:0003746; GO:0003924; GO:0005525; GO:0005737	PF00009;PF09106; ;PF09107;	46%	2
F9VLD7_ARTSS	Glutaconyl-CoA decarboxylase subunit beta (EC 7.2.4.5)	379	No		SFBM_1184	GO:0005886; GO:0006814; GO:0015451; GO:0016021; GO:0018801	PF03977;	100%	2
G2IHF1_9CLOT	Sporulation stage IV	498	No		RATSF0667	GO:0005524; GO:0016887; GO:0043934	PF09547;	99%	2
F9VK33_ARTSS	Signal recognition particle protein (Fifty-four homolog)	454	No	ffh	SFBM_0668	GO:0003924; GO:0005525; GO:0006614; GO:0008312; GO:0048500	PF00448;PF02881; ;PF02978;	100%	4
F9VJB0_ARTSS	Threonine-tRNA ligase (EC 6.1.1.3) (Threonyl-tRNA synthetase) (ThrRS)	644	No	thrS	SFBM_0439	GO:0000049; GO:0004829; GO:0005524; GO:0005737; GO:0006435; GO:0046872	PF03129;PF02824; ;PF00587;PF07973;	97%	5
G2IF69_9CLOT	DNA-directed RNA polymerase subunit beta' (RNAP subunit beta') (EC 2.7.7.6) (RNA polymerase subunit beta') (Transcriptase subunit beta')	1186	No	rpoC	RATSF01231	GO:0000287; GO:0003677; GO:0003899; GO:0006351; GO:0008270	PF04997;PF00623; ;PF04983;PF05000; ;PF04998;	88%	12
F9VLY4_ARTSS	30S ribosomal protein S14 type Z	61	No	rpsZ	SFBM_1383	GO:0003735; GO:0005840; GO:0006412; GO:0008270; GO:0019843	PF00253;	77%	2
F9VJM6_ARTSS	30S ribosomal protein S2	241	No	rpsB	SFBM_0718	GO:0003735; GO:0006412; GO:0015935	PF00318;	100%	2
G2IF74_9CLOT	50S ribosomal protein L11	142	No	rplK	RATSF01236	GO:0003735; GO:0005840; GO:0006412; GO:0070180	PF00298;PF03946; ;	100%	6
G2IEC9_9CLOT	Putative ABC transporter, permease protein	678	No		RATSF0941	GO:0005886; GO:0016021; GO:0055085	PF02687;	100%	6
G2IHA0_9CLOT	tRNA pseudouridine synthase B (EC 5.4.99.25) (tRNA pseudouridine(55)	293	No	truB	RATSF0616	GO:0003723; GO:0031119; GO:0106029	PF16198;PF01509; ;	100%	2

	synthase) (Psi55 synthase) (tRNA pseudouridylylate synthase) (tRNA-uridine isomerase)								
F9VLL3_ARTSS	Glycoprotease family protein	236	No		SFBM_1260	GO:0002949; GO:0008233	PF00814;	63%	2
G2IFH9_9CLOT	tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG (Glucose-inhibited division protein A)	624	No	mnmG	RATSFb_1341	GO:0002098; GO:0005737; GO:0050660	PF13932;	74%	11
F9VLU3_ARTSS	Tyrosine--tRNA ligase (EC 6.1.1.1) (Tyrosyl-tRNA synthetase) (TyrRS)	406	No	tyrS	SFBM_1342	GO:0003723; GO:0004831; GO:0005524; GO:0005737; GO:0006437	PF01479;PF00579	100%	3
G2IE94_9CLOT	Valine--tRNA ligase (EC 6.1.1.9) (Valyl-tRNA synthetase) (ValRS)	885	No	valS	RATSFb_0906	GO:0002161; GO:0004832; GO:0005524; GO:0005737; GO:0006438	PF08264;PF00133 ;PF10458;	74%	2
G2IDV7_9CLOT	ABC transporter, ATP-binding protein	603	No		RATSFb_0769	GO:0005524; GO:0016021; GO:0016887; GO:0042626	PF00664;PF00005	51%	2
F9VIZ3_ARTSS	ABC transporter, permease	360	No		SFBM_0174	GO:0016021		75%	2
G2IDV8_9CLOT	ABC-type multidrug/protein/lipid transport system, ATPase component	576	No		RATSFb_0770	GO:0005524; GO:0016021; GO:0016887; GO:0042626	PF00664;PF00005	53%	2
G2IGC2_9CLOT	Adenylate cyclase	181	No		RATSFb_0288		PF01928;	68%	2
F9VJY3_ARTSS	ATP-dependent DNA helicase RecG	671	No	recG	SFBM_0618	GO:0003676; GO:0004386; GO:0005524	PF00270;PF00271 ;PF17191;	99%	10
F9VKB5_ARTSS	Cation transport ATPase	740	No		SFBM_0803		PF13166;	100%	2
G2IEG9_9CLOT	Cellobiose phosphorylase	1108	No	cbpA	RATSFb_0981	GO:0003824	PF17167;	33%	2
F9VKF4_ARTSS	Cysteine desulfurase NifS	381	No	nifS	SFBM_0843	GO:0003824	PF00266;	50%	2
G2IEW9_9CLOT	DegV family protein	278	No	degV	RATSFb_1131	GO:0008289	PF02645;	100%	2
G2IH99_9CLOT	DHH subfamily 1 protein	319	No		RATSFb_0615	GO:0003676	PF01368;PF02272	94%	2
G2IGJ8_9CLOT	Dipicolinate synthase subunit B	195	No	spoVFB	RATSFb_0364	GO:0003824	PF02441;	67%	2

F9VJL4_ARTSS	DNA helicase (EC 3.6.4.12)	638	No		SFBM_0544	GO:0003677; GO:0003678; GO:0005524	PF00580;PF13361	69%	5
F9VL32_ARTSS	DNA helicase (EC 3.6.4.12)	731	No		SFBM_1076	GO:0003677; GO:0003678; GO:0005524	PF00580;PF13361	93%	8
F9VKY3_ARTSS	DUF4214 domain-containing protein	1084	No		SFBM_1027		PF13946;	58%	2
F9VLJ4_ARTSS	DUF4214 domain-containing protein	1062	No		SFBM_1241		PF13946;	50%	2
G2IE62_9CLOT	DUF4214 domain-containing protein	1070	No		RATSF0874		PF13946;	89%	2
G2IDY0_9CLOT	Dynammin-type G domain-containing protein	745	No		RATSF0792	GO:0003924; GO:0005525; GO:0016021	PF00350;	100%	6
F9VM00_ARTSS	Elongation factor G (EF-G)	686	No	fus	SFBM_1399	GO:0003746; GO:0003924; GO:0005525; GO:0005737	PF00679;PF14492 ;PF03764;PF00009;PF03144;	100%	8
G2IHE0_9CLOT	Extracellular solute-binding protein	491	No		RATSF0656	GO:0016021	PF12010;PF13416	100%	2
F9VLG9_ARTSS	Ferrous iron transport protein B	587	No		SFBM_1216	GO:0005525; GO:0015093; GO:0016021	PF07664;PF02421 ;PF07670;	100%	2
F9VKR3_ARTSS	Flagellar hook-length control protein FliK	490	No		SFBM_0955		PF02120;	38%	2
F9VIF4_ARTSS	Glucokinase	312	No		SFBM_0287	GO:0016301	PF00480;	100%	7
G2IHK6_9CLOT	Heme biosynthesis	462	No		RATSF0722	GO:0003824; GO:0046872; GO:0051539	PF04055;PF13186	64%	2
G2IGC8_9CLOT	KAP NTPase domain-containing protein	786	No		RATSF0294	GO:0016021	PF07693;	76%	2
F9VJD7_ARTSS	Lipoprotein	746	No		SFBM_0466		PF11308;	100%	4
G2IFV9_9CLOT	Lipoprotein, bmp family	358	No		RATSF0125	GO:0005886	PF02608;	74%	2
G2IFN4_9CLOT	LytR family transcriptional regulator	329	No	lytR	RATSF0050	GO:0016021	PF03816;	100%	2
G2IFT0_9CLOT	M18 family aminopeptidase (EC 3.4.11.-)	427	No		RATSF0096	GO:0004177; GO:0008237; GO:0008270	PF02127;	99%	3
G2IGA6_9CLOT	M18 family aminopeptidase (EC 3.4.11.-)	465	No		RATSF0272	GO:0004177; GO:0008237; GO:0008270	PF02127;	73%	3
F9VLB0_ARTSS	Manganese-containing catalase	222	No		SFBM_1155		PF05067;	100%	2
G2IEI7_9CLOT	MATE efflux family protein	443	No	matE	RATSF0999	GO:0015297; GO:0016021; GO:0042910	PF01554;	52%	2
G2IHL9_9CLOT	Membrane carboxypeptidase MrcB	998	No	mrcB	RATSF0735	GO:0004180; GO:0008658;	PF00912;PF00905	53%	2

						GO:0016021; GO:0016740			
G2IGU3_9CLOT	Metallo-beta-lactamase family protein	443	No		RATSF0459		PF10996;PF16661 ;PF07521;	85%	2
F9VJM3_ARTSS	Mg chelatase	509	No		SFBM_0715		PF01078;PF13335 ;	100%	6
G2IH81_9CLOT	Mg chelatase, subunit ChII	508	No	chlI	RATSF0597		PF01078;PF13335 ;	75%	2
G2IFL8_9CLOT	Nucleoid-associated protein RATSF0034	111	No		RATSF0034	GO:0003677; GO:0005737; GO:0043590	PF02575;	96%	4
F9VL73_ARTSS	Oxidoreductase family protein	352	No		SFBM_1118	GO:0016491	PF01408;PF02894 ;	100%	2
F9VL21_ARTSS	Peptidase	413	No		SFBM_1065	GO:0003824; GO:0046872	PF00675;PF05193 ;	57%	2
G2IDV2_9CLOT	Peptidase	410	No		RATSF0764	GO:0003824; GO:0046872	PF05193;	100%	2
G2IGL4_9CLOT	Peptidase, M23/M37 family protein	426	No		RATSF0380		PF01551;	65%	2
F9VM60_ARTSS	Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase) (LAP) (EC 3.4.11.10) (Leucyl aminopeptidase)	479	No	pepA	SFBM_1459	GO:0004177; GO:0005737; GO:0008235; GO:0030145	PF00883;PF02789 ;	92%	7
G2IGZ2_9CLOT	Putative calcium-translocating P-type ATPase, PMCA-type	850	No		RATSF0508	GO:0005388; GO:0005524; GO:0016021	PF00689;PF00690 ;	100%	3
G2IER0_9CLOT	Putative glycosyltransferase	327	No		RATSF01072	GO:0016740	PF00535;	53%	2
G2IEZ9_9CLOT	Putative helicase	1071	No		RATSF01161	GO:0004386; GO:0005524; GO:0008270	PF00271;PF08455 ;PF00176;	20%	2
G2IHH5_9CLOT	Putative peptidoglycan glycosyltransferase	532	No		RATSF0691	GO:0008658; GO:0016021; GO:0016740	PF00905;	99%	4
G2IE70_9CLOT	Putative transposase	162	No		RATSF0882			81%	4
F9VM76_ARTSS	Radical SAM protein	308	No		SFBM_1475	GO:0003824; GO:0051536	PF04055;PF16199 ;	49%	2
F9VJ21_ARTSS	Ribonuclease R (RNase R) (EC 3.1.13.1)	692	No	mr	SFBM_0202	GO:0003723; GO:0005737; GO:0008859	PF17876;PF08206 ;PF00773;PF00575;	74%	6
F9VK38_ARTSS	Ribosome biogenesis GTPase A	289	No		SFBM_0673	GO:0005525; GO:0005737	PF01926;	54%	2
G2IFW9_9CLOT	Shikimate 5-dehydrogenase	278	No	aroE	RATSF0135	GO:0004764	PF08501;	54%	2
F9VIE5_ARTSS	SirA family protein	192	No		SFBM_0278		PF02635;PF01206 ;	100%	2
F9VJG7_ARTSS	Spore coat protein	242	No		SFBM_0496			55%	2

F9VJ72_ARTSS	Spore cortex-lytic enzyme, pre-pro-form	755	No		SFBM_0401		PF01471;	61%	2
G2IDU8_9CLOT	Stage V sporulation protein D	728	No	spoVD	RATSF0760	GO:0008658; GO:0016021	PF03793;PF03717 ;PF00905;	93%	6
G2IGK4_9CLOT	Superfamily I DNA and RNA helicase and helicase subunits-like protein	1306	No		RATSF0370	GO:0004386		62%	2
F9VJX9_ARTSS	TipAS domain-containing protein	113	No		SFBM_0614		PF07739;	37%	2
F9VKA0_ARTSS	Transcriptional regulator	292	No		SFBM_0788		PF00480;	75%	2
F9VK36_ARTSS	tRNA (guanine-N(1)-)-methyltransferase (EC 2.1.1.228) (MIG-methyltransferase) (tRNA [GM37] methyltransferase)	222	No	trnD	SFBM_0671	GO:0005737; GO:0052906	PF01746;	54%	2
F9VKE1_ARTSS	Twitching motility protein PilT	350	No		SFBM_0830		PF00437;	74%	3
F9VL35_ARTSS	Two component regulator	676	No		SFBM_1079		PF07495;	100%	3
G2IE28_9CLOT	Type II secretion system protein E	467	No		RATSF0840		PF00437;	84%	2
F9VJJ2_ARTSS	U32 family peptidase	782	No		SFBM_0522	GO:0008233	PF12392;PF01136 ;	50%	2
F9VL50_ARTSS	Uncharacterized protein	960	No		SFBM_1094	GO:0016021		81%	2
G2IG04_9CLOT	Uncharacterized protein	497	No		RATSF0170	GO:0016021	PF09826;	70%	2
G2IGQ2_9CLOT	Uncharacterized protein	2029	No		RATSF0418	GO:0016021	PF13946;	62%	2
F9VJC9_ARTSS	Uncharacterized protein	1316	No		SFBM_0458			89%	6
F9VK00_ARTSS	Uncharacterized protein	407	No		SFBM_0635			65%	2
F9VK14_ARTSS	Uncharacterized protein	364	No		SFBM_0649	GO:0003824	PF05913;	49%	2
F9VKR0_ARTSS	Uncharacterized protein	372	No		SFBM_0952			81%	3
F9VKT8_ARTSS	Uncharacterized protein	135	No		SFBM_0980			100%	2
F9VLS4_ARTSS	Uncharacterized protein	702	No		SFBM_1322			100%	2
G2IE23_9CLOT	Uncharacterized protein	134	No		RATSF0835			36%	2
G2IEB4_9CLOT	Uncharacterized protein	1059	No		RATSF0926			100%	2
G2IEQ1_9CLOT	Uncharacterized protein	173	No		RATSF1063		PF14270;	100%	2
G2IFX9_9CLOT	Uncharacterized protein	1938	No		RATSF0145		PF13946;PF00041 ;	95%	5
G2IFZ8_9CLOT	Uncharacterized protein	704	No		RATSF0164			100%	4
G2IGF5_9CLOT	Uncharacterized protein	173	No		RATSF0321			99%	3
G2IG23_9CLOT	Xaa-pro aminopeptidase	415	No		RATSF0189	GO:0004177; GO:0030145	PF05195;PF00557 ;	38%	2
F9VLC7_ARTSS	XRE family transcriptional regulator	70	No		SFBM_1174	GO:0003677	PF13443;	49%	2

APPENDIX D: HIERARCHAL-ALL-AGAINST-ALL ANALYSIS CLUSTERS AND ASSOCIATIONS

HALLA association rank	cluster1	cluster2	pvalue	qvalue	similarity score between clusters
1	TG; Glucose	D_0__Bacteria D_1__Proteobacteria D_2__Deltaproteobacteria D_3__Desulfovibrionales D_4__Desulfovibrionaceae D_0__Bacteria D_1__Proteobacteria D_2__Gamma proteobacteria D_3__Enterobacteriales D_4__Enterobacteriaceae D_5__Proteus;D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Leuconostocaceae D_5__Leuconostoc;D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Leuconostocaceae D_5__Weissella	9.32E-08	0.000141	0.811362825
2	Body Weight; Fat	D_0__Bacteria D_1__Bacteroidetes D_2__Bacteroidia D_3__Bacteroidales D_4__Muribaculaceae D_5__uncultured Barnesiella sp.;D_0__Bacteria D_1__Bacteroidetes D_2__Bacteroidia D_3__Bacteroidales D_4__Muribaculaceae D_5__uncultured organism	4.43E-07	0.000335	-0.744640643
3	Body Weight; Fat	D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Anaeroplasmatales D_4__Anaeroplasmataceae D_5__Anaeroplasma;D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Mycoplasmatales D_4__Mycoplasmataceae D_5__Ureaplasma;D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Clostridiales vadinBB60 group D_5__uncultured Firmicutes bacterium;D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Clostridiales vadinBB60 group	6.64E-06	0.00277	0.688922842

4	Fat	D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Mycoplasmatales D_4__Mycoplasmataceae D_5__Mycoplasma	7.33E-06	0.00277	0.686635941
5	TG	D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Betaproteobacteriales D_4__Burkholderiaceae D_5__Aquabacterium;D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Betaproteobacteriales D_4__Burkholderiaceae D_5__Hydrogenophaga	1.02E-05	0.003098	0.72097531
6	TG	D_0__Bacteria D_1__Actinobacteria D_2__Coriobacteriia D_3__Coriobacteriales D_4__Eggerthellaceae D_5__Parvibacter	1.26E-05	0.003183	-0.715906274
7	Glucose	D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Enterococcaceae D_5__Enterococcus;D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Enterobacteriales D_4__Enterobacteriaceae __	1.57E-05	0.003388	0.710552342
8	Glucose	D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Streptococcaceae D_5__Lactococcus	3.00E-05	0.004767	0.693777088
9	Fat	D_0__Bacteria D_1__Proteobacteria D_2__Alphaproteobacteria D_3__Sphingomonadales D_4__Sphingomonadaceae __;D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Betaproteobacteriales D_4__Rhodocyclaceae D_5__Methyloversatilis;D_0__Bacteria D_1__Proteobacteria D_2__Alphaproteobacteria D_3__Sphingomonadales D_4__Sphingomonadaceae D_5__Sphingobium;D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Betaproteobacteriales D_4__Burkholderiaceae D_5__Aquabacterium;D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__	3.15E-05	0.004767	0.65019408

		_Betaproteobacteriales D_4__Burkholderiaceae D_5__Hydrogenophaga			
10	TG	D_0__Bacteria D_1__Actinobacteria D_2__Actinobacteria D_3__Corynebacteriales D_4__Corynebacteriaceae D_5__Corynebacterium 1;D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Bacillales D_4__Staphylococcaceae D_5__Staphylococcus;D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Enterococcaceae D_5__Enterococcus;D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Enterobacteriales D_4__Enterobacteriaceae __	6.16E-05	0.006211	0.673766567
11	Fat	D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Pasteurellales D_4__Pasteurellaceae D_5__Muribacter	6.90E-05	0.00631	0.628449473
12	TG	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Ruminococcaceae D_5__Intestinimonas	7.09E-05	0.00631	-0.669663375
13	TG	D_0__Bacteria D_1__Firmicutes D_2__Erysipelotrichia D_3__Erysipelotrichales D_4__Erysipelotrichaceae D_5__Ileibacterium;D_0__Bacteria D_1__Bacteroidetes D_2__Bacteroidia D_3__Bacteroidales D_4__Tannerellaceae D_5__Parabacteroides;D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Clostridiales vadinBB60 group D_5__uncultured prokaryote	0.000103135	0.008207	-0.65845548
14	TG	D_0__Bacteria D_1__Proteobacteria D_2__Alphaproteobacteria D_3__Sphingomonadales D_4__Sphingomonadaceae __	0.000140949	0.010656	0.648735189
15	Glucose	D_0__Bacteria D_1__Proteobacteria D_2__Alphaproteobacteria D_3__Rhodospirillales D_4__uncultured D_5__unidentified rumen	0.000213448	0.014032	-0.63527676

		proteobacteria D_3__Enterobacteriales D_4__Enterobacteriaceae D_5__Proteus;D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Leuconostocaceae D_5__Leuconostoc;D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Leuconostocaceae D_5__Weissella			
24	TG	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Lachnospiraceae D_5__[Eubacterium] oxidoreducens group	0.000769647	0.031656	-0.589270898
25	Glucose	D_0__Bacteria D_1__Firmicutes D_2__Erysipelotrichia D_3__Erysipelotrichales D_4__Erysipelotrichaceae D_5__Turicibacter;D_0__Bacteria D_1__Proteobacteria D_2__Delta proteobacteria D_3__Desulfovibrionales D_4__Desulfovibrionaceae D_5__Desulfovibrio	0.00077466	0.031656	0.589019236
26	TG	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Ruminococcaceae D_5__Ruminiclostridium 5	0.000810413	0.032246	0.587264446
27	TG	D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Mycoplasmatales D_4__Mycoplasmataceae D_5__Mycoplasma	0.000858412	0.032455	0.585012105
28	Fat	D_0__Bacteria D_1__Verrucomicrobia D_2__Verrucomicrobiae D_3__Verrucomicrobiales D_4__Akkermansiaceae D_5__Akkermansia	0.000858597	0.032455	-0.54503782
29	Glucose	D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Mollicutes RF39 D_4__metagenome D_5__metagenome	0.000951623	0.03417	-0.580935672
30	Cholesterol	D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Streptococcaceae D_5__Lactococcus;D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Ruminococcaceae D_5__GCA-900066225	0.000955155	0.03417	0.580788177

42	Insulin	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Clostridiales vadinBB60 group D_5__uncultured Clostridiales bacterium	0.00240257	0.055041	0.54175007
43	Cholesterol	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Family XIII D_5__[Eubacterium] nodatum group	0.002471187	0.055768	- 0.540480694
44	Lean	D_0__Bacteria D_1__Bacteroidetes D_2__Bacteroidia D_3__Bacteroidales D_4__Muribaculaceae D_5__uncultured Barnesiella sp.	0.002717173	0.059542	- 0.498131713
45	Glucose	D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Betaproteobacteriales D_4__Burkholderiaceae D_5__Hydrogenophaga	0.002769797	0.059828	0.53528758
46	Cholesterol	D_0__Bacteria D_1__Actinobacteria D_2__Actinobacteria D_3__Actinobacteria D_4__Actinobacteria D_5__Actinobacteria	0.002964429	0.06313	0.532156691
47	Fat	D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Pasteurellales D_4__Pasteurellaceae D_5__Rodentibacter	0.003331831	0.06822	0.48909072
48	Cholesterol	D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Mycoplasmatales D_4__Mycoplasmataceae D_5__Ureaplasma	0.003338821	0.06822	0.526600985
49	Insulin	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Ruminococcaceae D_5__UBA1819	0.003563211	0.071834	- 0.523522762
50	Body Weight	D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Betaproteobacteriales D_4__Burkholderiaceae D_5__Aquabacterium	0.003777528	0.075153	0.483401499
51	Lean	D_0__Bacteria D_1__Proteobacteria D_2__Gammaproteobacteria D_3__Pasteurellales D_4__Pasteurellaceae D_5__Muribacter	0.003947066	0.076852	0.481389217
52	Lean	D_0__Bacteria D_1__Actinobacteria D_2__Coriobacteriia D_3__Coriobacteriia D_4__Coriobacteriia D_5__Coriobacteriia	0.004015414	0.076852	0.480599051

		obacterales D_4__Coriobacterales s Incertae Sedis D_5__uncultured			
53	Cholesterol	D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Mollicute s RF39 D_4__uncultured Erysipelotrichaceae bacterium D_5__uncultured Erysipelotrichaceae bacterium	0.0042423 71	0.079191	- 0.51512332 2
54	Fat	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridial es D_4__Ruminococcaceae D_5__ Ruminococcaceae UCG-010	0.0044690 94	0.081413	- 0.47563025 2
55	TG	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridial es D_4__Clostridiaceae 1 D_5__Clostridium sensu stricto 1	0.0046494 43	0.08369	- 0.51062620 4
56	Body Weight	D_0__Bacteria D_1__Bacteroidete s D_2__Bacteroidia D_3__Bactero idales D_4__Muribaculaceae D_5__ gut metagenome	0.0048824 61	0.086051	- 0.47146835 4
57	TG	D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Mycoplas matales D_4__Mycoplasmataceae D_5__Ureaplasma	0.0050552 06	0.087856	0.50646631 7
58	Cholesterol	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridial es D_4__Ruminococcaceae D_5__ Butyricoccus	0.0056560 43	0.09453	- 0.50080108 1
59	Fat	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridial es D_4__Ruminococcaceae D_5__ Intestinimonas	0.0056892 98	0.09453	-0.46415094
60	Glucose	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridial es D_4__Lachnospiraceae D_5__L achnospiraceae NK4A136 group	0.0058506 53	0.096154	0.49907588
61	Fat	D_0__Bacteria D_1__Bacteroidete s D_2__Bacteroidia D_3__Bactero idales D_4__Rikenellaceae D_5__ Mucinivorans	0.0059602 18	0.096902	0.46189365 5
62	Glucose	D_0__Bacteria D_1__Tenericutes D_2__Mollicutes D_3__Mollicute s RF39 D_4__uncultured Paenibacillaceae	0.0062069 87	0.098338	- 0.49603956 2

		bacterium D_5__uncultured Paenibacillaceae bacterium			
63	Glucose	D_0__Bacteria D_1__Proteobacteria D_2__Alphaproteobacteria D_3__Rhodospirillales D_4__uncultured D_5__Azospirillum sp. 47_25	0.006212925	0.098338	- 0.495990224
64	TG	D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Bacillales D_4__Planococcaceae D_5__Kurthia	0.006243664	0.098338	0.495735498
65	Cholesterol	D_0__Bacteria D_1__Firmicutes D_2__Bacilli D_3__Lactobacillales D_4__Streptococcaceae D_5__Streptococcus	0.006324105	0.098578	0.495073892
66	Insulin	D_0__Bacteria D_1__Firmicutes D_2__Clostridia D_3__Clostridiales D_4__Family XIII D_5__Family XIII AD3011 group	0.006506584	0.099592	- 0.493599109

**APPENDIX E: IDENTITY AND IMPORTANCE SCORES OF ALL ASVS PREDICTIVE
OF METABOLIC CLUSTERS**

ASV #	ASV	Taxonomic Assignment	Importance	Cluster Association (as determined by avg. abundance)	Accession
1	14dc60ee1b5bad57cbf1b8b70ceda2	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Enterobacteriales;D_4__Enterobacteriaceae	0.056561064	metS	NR_118333.1
2	ccd547c6a601c61c328a88f0252cac5e	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__Lactobacillales;D_4__Leuconostocaceae;D_5__Leuconostoc	0.053192289	metS	NR_109004.1
3	ce5b59ece6db73b05798315f7089e17a	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Enterobacteriales;D_4__Enterobacteriaceae	0.043489138	metS	NR_114419.1
4	e24eb65cb61019d39e980c89197713d	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__Lactobacillales;D_4__Enterococcaceae;D_5__Enterococcus	0.043442756	metS	NR_113901.1
5	10fc832b4cee459f8e7532193d354968	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Lachnospiraceae	0.03910503	metS	NR_025796.1
6	e2d5cbdfa854a5a55adb60154cd54887	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Enterobacteriales;D_4__Enterobacteriaceae	0.036504094	metS	NR_146667.2
7	33613717db4d9890cd16bdc22062acee	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__Lactobacillales;D_4__Leuconostocaceae;D_5__Weissella	0.035577033	metS	NR_113258.1

8	a891568a5 b9149c7db b14d6c563 67a8c	D_0__Bacteria;D_1__Prot eobacteria;D_2__Gammap roteobacteria;D_3__Entero bacteriales;D_4__Enteroba cteriaceae	0.0295420 94	metS	NR_118568.1
9	bfba91aead ead2746a4 70fdaa65a5 11d	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae	0.0291994 67	metS	NR_044048.
10	e303c2dfc1 28751db2e 66f336172 0be3	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae;D_5__Lachn ospiraceae NK4A136 group;D_6__Trichinella pseudospiralis	0.0219286 14	metS	NR_119035.1
11	03e5fed105 9bc7b42ac 90c005696 83ba	D_0__Bacteria;D_1__Prot eobacteria;D_2__Deltaprot eobacteria;D_3__Desulfovi brionales;D_4__Desulfovi brionaceae;D_5__Desulfov ibrio;D_6__Brachyspira sp. NSH-25	0.0198610 06	metS	NR_104990.1
12	f69eb2f45f 5c3a2fe85e 149154cf5 e5	D_0__Bacteria;D_1__Prot eobacteria;D_2__Gammap roteobacteria	0.0198516 35	metS	NR_117683.1
13	6c8b095d6 2dcf8cd5d8 fe96d0498 d219	D_0__Bacteria;D_1__Bact eroidetes;D_2__Bacteroidi a;D_3__Bacteroidales;D_4 __Rikenellaceae;D_5__Ali stipes;D_6__gut metagenome	0.0188478 29	metS	NR_113152.1
14	27ec9abad 1b5670d70 c289165c0 b8c69	D_0__Bacteria;D_1__Prot eobacteria;D_2__Gammap roteobacteria;D_3__Entero bacteriales;D_4__Enteroba cteriaceae	0.0188244 59	metS	NR_156052.1
15	00ee70646 10a0a5eb6 720b6fa56 982f6	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae;D_5__Lachn ospiraceae NK4A136 group;D_6__Trichinella pseudospiralis	0.0186178 83	metS	NR_116814.1

16	f670c3d357ab9af85d840c55f21f8c5d	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__Bacillales;D_4__Staphylococcaceae;D_5__Staphylococcus	0.018158013	metS	NR_113350.1
17	03e5ea91f93440511c22b2599eed7c39	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Lachnospiraceae	0.01782745	metS	NR_117139.2
18	ff52695bd2d7116b4578ce9ee98e42b5	D_0__Bacteria;D_1__Epsilonbacteraeota;D_2__Campylobacteria;D_3__Campylobacteriales;D_4__Helicobacteraceae;D_5__Helicobacter	0.015415722	metS	NR_041748.2
19	60a37667f4ddbffc93d3ce743ca3b409	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Enterobacteriales;D_4__Enterobacteriaceae;D_5__Escherichia-Shigella	0.015172279	metS	NR_074902.1
20	e6cc88f3d852b17e32057da07a3df604	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__Lactobacillales;D_4__Enterococcaceae;D_5__Enterococcus	0.01463402	metS	NR_104559.2
21	ec3ef83ad4c92f8ba012e922f08e9a4e	D_0__Bacteria	0.014174306	metS	
22	7d8ca34376e1f725babe3b9f4e55abfd	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Lachnospiraceae	0.013991601	metS	NR_117139.2
23	d74eaa73b082c108e0a20f1ee4b8920f	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__Lactobacillales;D_4__Streptococcaceae;D_5__Streptococcus	0.013961096	metS	NR_159228.1
24	20b4579350bd90e1dde4d87391f8f9f	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Rikenellaceae;D_5__Rikenellaceae RC9 gut group	0.013722857	metS	NR_147742.1
25	c882f33472948a1cb7	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__	0.012860797	metS	NR_144587.1

	931d03a6a a87f7	Bacillales;D_4__Planococ aceae;D_5__Kurthia			
26	1e9ef4e844 738710ba6 954bb59f0 c795	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae	0.0109945 91	metS	NR_156081.1
27	322b430fd 7a0d8bca3 b906910a4 efcd7	D_0__Bacteria;D_1__Prot eobacteria;D_2__Gammap roteobacteria;D_3__Betapr oteobacteriales;D_4__Burk holderiaceae;D_5__Hydro genophaga	0.0109651 13	metS	NR_114132.1
28	311e5a9ad 905f9a063 d82edb6ac 07bec	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae	0.0099520 69	metS	NR_025796.1
29	1b3ebb7ecf 71bc99bbf 12b7f738f9 7cc	Ambiguous_taxa;Ambiguo us_taxa;Ambiguous_taxa; Ambiguous_taxa;Ambiguo us_taxa;Ambiguous_taxa; Ambiguous_taxa	0.0099046 86	metS	
30	13494f7a2 8d2db79db 2de2c7d7d 2583b	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae	0.0096404 72	metS	NR_117905.1
31	01ed0ac8d 59046328a 9485ec012 d87f3	D_0__Bacteria;D_1__Firm icutes;D_2__Bacilli;D_3__ Lactobacillales;D_4__Leuc onostocaceae;D_5__Leuco nostoc	0.0091483 53	metS	NR_118557.1
32	333d354df 35affbcc66 cdf1a43dba fb1	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae;D_5__Rum iniclostridium 5	0.0090706 28	metS	NR_144736.1
33	0d2912420 2765d8eae bf345e27a2 3791	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae;D_5__Blauti a	0.0075313 19	metS	NR_042152.1
34	d07c7ec26 91e01baaec e375c59b8 2402	D_0__Bacteria;D_1__Firm icutes;D_2__Bacilli;D_3__ Lactobacillales;D_4__Stre ptococcaceae;D_5__Lactoc occus	0.0063302 77	metS	NR_113925.1

35	38ee90e67 27037cf8e9 4242392e6 c8db	D_0__Bacteria	0.0062207 47	metS	
36	f69c487e31 f5947ce23e ff1aa596ab 96	D_0__Bacteria	0.0060615 76	metS	
37	f85c14977 376e73ec2 e779615f2 3354a	D_0__Bacteria;D_1__Prot eobacteria;D_2__Gammap roteobacteria;D_3__Entero bacteriales;D_4__Enteroba cteriaceae	0.0047870 29	metS	NR_117751.1
38	08b8df0a3c 3104087a6 af89e34426 014	D_0__Bacteria	0.0046099 45	metS	
39	2a51899b2 d61564c2f b8a1b9079 5248a	D_0__Bacteria	0.0045427 3	metS	
40	0af247bade 87e51431f 85ed2ff3e3 9b0	D_0__Bacteria	0.0045343 1	metS	
41	0f0f4217bc bafff0f5e6a 58e244a9a e8	D_0__Bacteria;D_1__Prot eobacteria;D_2__Alphapro teobacteria;D_3__Rhizobia les;D_4__Xanthobacterace ae;D_5__Bradyrhizobium	0.0045191 33	metS	NR_102489.2
42	f9a82fef8fd 4ce66c747 422a0cdd9 a44	D_0__Bacteria	0.0039536 56	metS	
43	086648664 738dcc594 db104b9d4 1182d	D_0__Bacteria	0.0037831 9	metS	
44	16aa96e2d 6029d8ca0 9e2657256 ee987	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae;D_5__Blauti a;D_6__uncultured Clostridiales bacterium	0.0035552 71	metS	NR_042152.1
45	f6f4389068 82dc6adcb	D_0__Bacteria;D_1__Prot eobacteria;D_2__Gammap	0.0034885 68	metS	NR_156986.1

	b8cc79882 b3b3	roteobacteria;D_3__Pseudomonadales;D_4__Pseudomonadaceae;D_5__Pseudomonas			
46	fd073bcd6 d976425ed edbc3859e 8dd9	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Bacteroidaceae;D_5__Bacteroides	0.0024155 79	metS	NR_112931.1
47	18878d46d e68601b15 206f65ff96 b834	D_0__Bacteria	0.0024057 12	metS	
48	0b82eebd9f c63ab8c21 64f2986bf7 db9	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Muribaculaceae;D_5__uncultured Barnesiella sp.;D_6__uncultured Barnesiella sp.	0.0023064 23	metS	NR_133950.1
49	f37b11c58f 63048389c e4dca1199 51b6	D_0__Bacteria	0.0022942 36	metS	
50	0c7c8ec55 56ba94b95 823b56190 4dd64	D_0__Bacteria	0.0021894 89	metS	
51	0452aefc7b 534efc2343 673e30304 1ad	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Bacteroidaceae;D_5__Bacteroides;D_6__Bacteroides caecimuris	0.0020709 56	metS	NR_144606.1
52	117506192 5a32d080d 8c0e42cc1c 872d	D_0__Bacteria	0.0020322 68	metS	
53	04fbf5aa26 b9825d3a7 e2312ebf83 de5	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Ruminococcaceae	0.0019506 99	metS	NR_025670.1
54	1b5e60f2a6 58c644723 03b6996cf 0342	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Enterobacteriales	0.0019258 51	metS	NR_118568.1

		bacteriales;D_4__Enterobacteriaceae			
55	efbc267bad b2cd068c0 5fa01ae6a3 1ba	D_0__Bacteria;D_1__Tenericutes;D_2__Mollicutes;D_3__Anaeroplasmatales;D_4__Anaeroplasmataceae;D_5__Anaeroplasma;D_6__uncultured bacterium	0.0017867 23	metS	NR_029167.1
56	142a8a402 bc6e94ff15 3f337f0e23 63a	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Family XIII;D_5__Family XIII UCG-001	0.0017348 04	metS	NR_041887.1
57	0aa80dc5ae 0a330406f b63062718 b86a	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Ruminococcaceae	0.0016631 88	metS	NR_152059.1
58	0377150a4 bb1b64505 f641daa271 0dc4	D_0__Bacteria	0.0016472 32	metS	
59	05568f8d7 e81baf0800 4b2b7e885 d7d9	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria;D_3__Enterobacteriales;D_4__Enterobacteriaceae	0.0016472 32	metS	
60	0904f784f5 3e6be10ad 4d9ad7e1a 49ac	D_0__Bacteria	0.0016472 32	metS	
61	096c2224d ce880aba1 6a80893e2 a5f69	D_0__Bacteria	0.0016472 32	metS	
62	0beae212b 6c495580f d945a11c8 19b53	D_0__Bacteria	0.0016472 32	metS	
63	0e18ce6efd 7e4d15c80 365b6c789 90b0	D_0__Bacteria	0.0016472 32	metS	
64	f7487ffdf1 b1b852aa6 a95703dbaf 98a	D_0__Bacteria	0.0016472 32	metS	NR_102493.2

65	03bae053e306e6539eff4acc40f073	D_0__Bacteria;D_1__Actinobacteria;D_2__Actinobacteria	0.001485084	metS	NR_159260.1
66	1105982ae2c802e62176c5ee57e850fc	D_0__Bacteria;D_1__Firmicutes;D_2__Erysipelotrichia;D_3__Erysipelotrichales;D_4__Erysipelotrichaceae	0.001460411	metS	NR_044647.2
67	1cf28112fe3855b197a68b7806b91b59	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Lachnospiraceae;D_5__Lachnospiraceae FCS020 group;D_6__uncultured prokaryote	0.001381972	metS	NR_042152.1
68	1c48b4c6a35318d81faae5b507d1ec38	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Clostridiales vadinBB60 group;D_5__uncultured Firmicutes bacterium;D_6__uncultured Firmicutes bacterium	0.00127035	metS	NR_152060.1
69	0090f94a492b23e2a8138b0a4ba75b15	D_0__Bacteria	0.001228643	metS	
70	0a8cf1064cf64956d9a287722506c0f2	D_0__Bacteria	0.001151812	metS	
71	0ab925d01a72a18daaff013acc87783	D_0__Bacteria	0.001151812	metS	
72	126d26ed9b67abcb824b8ca27ad07d77	D_0__Bacteria	0.001151812	metS	
73	036c94fc4506291aa06012041fbdac81	D_0__Bacteria	0.001146284	metS	
74	e51690060afbc0d0294	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Sphingo	0.001124717	metS	NR_152071.1

	bfd244f330011	monadales;D_4__Sphingomonadaceae			
75	066db05a269a9b077fb353acfc8d3932	D_0__Bacteria	0.001048186	metS	
76	10200ce4b503e686c926c638af7ee9ac	D_0__Bacteria;D_1__Firmicutes;D_2__Bacilli;D_3__Bacillales;D_4__Staphylococcaceae;D_5__Staphylococcus	0.001012322	metS	NR_119252.1
77	0c8081a2abccea128499914dbbcdec13	D_0__Bacteria	0.00096598	metS	
78	05e6751a6375d53e2d0a4558dcef6ac7	D_0__Bacteria;D_1__Proteobacteria;D_2__Gammaproteobacteria	0.000943469	metS	NR_133953.1
79	00e29e7f2ea1148a9e115faa37283d45	D_0__Bacteria	0.000821275	metS	
80	f55e20003a34f5e4d6e53d31549e9ee9	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Muribaculaceae;Ambiguous_taxa;Ambiguous_taxa	0.000821275	metS	NR_144616.1
81	ea945e0912117614eab01b879c121dec	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Muribaculaceae	0.000814737	metS	NR_144616.1
82	091758141d8cdbc8feadd792f7146ebe	D_0__Bacteria	0.000697208	metS	
83	128eb6c0c899c1ed5365816cecf151c7	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Lachnospiraceae	0.000646618	metS	NR_025796.1
84	0ccaa036e3d07e116384c53ca40c3c45	D_0__Bacteria	0.000536623	metS	
85	00bba3741cac0cebdd1	D_0__Bacteria	0.000531435	metS	

	1d308429a df73				
86	febd43a7f0 7e35aab5f4 84bf4f46f7 c5	D_0__Bacteria	0.0005185 44	metS	
87	07041512c 92a29a5c0 9333d76a2 272a0	D_0__Bacteria	0.0004718 16	metS	
88	0a64f7f60f a9f1a1af69 629661638 0fa	D_0__Bacteria	0.0004718 16	metS	
89	270f49276 cdc60aac02 a31d57230 8c09	D_0__Bacteria	0.0004621 5	metS	
90	098608154 a070fcdd99 b47022af4 758e	D_0__Bacteria	0.0003151 14	metS	
91	0984b5bc6 23d52fbd6 ead2c9781 71c14	D_0__Bacteria	0.0002999 61	metS	
92	fcb0027c8 252c36d57 5d0c92fbe9 3c8	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae;D_5__Harr yflintia;D_6__uncultured bacterium	0.0002908 47	metS	NR_133956.2
93	06d3b3708 0cd8a3016 8fe96f8423 34b5	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae;D_5__Rum iniclostridium 5	0.0001127 91	metS	NR_146698.1
94	146f29ca3c 571133366 53aa5ae5c4 9a1	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae	0.0193869 18	non-metS	NR_144748.1
95	3ebf6e742c a04638525 1449b652f 22b6	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae;D_5__Ang elakisella	0.0159540 91	non-metS	NR_029355.1

96	466746e8bbe7f98a858c39cd4c141340	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Muribaculaceae;D_5__uncultured bacterium;D_6__uncultured bacterium	0.01544841	non-metS	NR_133950.1
97	f3652e7b51287dde0cef18003f4576f5	D_0__Bacteria;D_1__Actinobacteria;D_2__Coriobacteriia;D_3__Coriobacteriales;D_4__Eggerthellaceae;D_5__uncultured;D_6__uncultured Coriobacteriales bacterium	0.013835868	non-metS	NR_074377.1
98	0e4105f5b01bba6fc15a28df79244a64	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Lachnospiraceae;D_5__Dorea	0.010755806	non-metS	NR_044645.2
99	3fd3f371cbfbf5510936dc29989b632	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Tannerellaceae;D_5__Parabacteroides;D_6__uncultured Parabacteroides sp.	0.008703346	non-metS	NR_113076.1
100	093686784ef286b277452198de3f6fb2	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Bacteroidaceae;D_5__Bacteroides	0.00726423	non-metS	NR_112931.1
101	162fe0768c73672d76126c75c7ba28af	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Rikenellaceae;D_5__Rikenellaceae RC9 gut group	0.007076777	non-metS	NR_148260.1
102	0b7a120b3f9cfb34c9f742eb121fa6e2	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Christensenellaceae;D_5__Catabacter	0.006613662	non-metS	NR_144742.1
103	f86c94b1cccf49dd76ab1b8b3422dfba	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Ruminococcaceae;D_5__Ruminococcaceae UCG-010;D_6__uncultured Clostridia bacterium	0.006425616	non-metS	NR_146698.1

104	039f1c0ab8348f5a4cd119145bb56d39	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Muribaculaceae;D_5__CAG-873;Ambiguous_taxa	0.006087134	non-metS	NR_144616.1
105	266b6f16b2b2eb4801665a0aee246c6f	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Clostridiales vadinBB60 group;Ambiguous_taxa;Ambiguous_taxa	0.005506569	non-metS	NR_147370.1
106	1f92f75a3f0cd4b896552503b57f6624	D_0__Bacteria;D_1__Actinobacteria;D_2__Coriobacteriia;D_3__Coriobacteriales;D_4__Eggerthellaceae;D_5__Enterorhabdus;D_6__mouse gut metagenome	0.004860272	non-metS	NR_148574.1
107	0702fbeb64642bebdb1ecc71692a74d	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Lachnospiraceae;D_5__Roseburia;D_6__Clostridium sp. Clone-44	0.004801091	non-metS	NR_036800.1
108	1a6be0a0d4a99b85527d0cc5990d8744	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Muribaculaceae;Ambiguous_taxa;Ambiguous_taxa	0.004674055	non-metS	NR_144616.1
109	f6456a32cfa4649e30abb8e2697a8dd	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia	0.004628583	non-metS	NR_113271.1
110	0faf39e0c24a07d7bce7a841648b6e24	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Clostridiales vadinBB60 group;D_5__gut metagenome;D_6__gut metagenome	0.004536443	non-metS	NR_134026.1
111	13d11200e7e2ad7a33efd13fa1341a72	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Prevotellaceae;D_5__Prevotellaceae UCG-001;D_6__uncultured bacterium	0.003469965	non-metS	NR_151886.1

112	03a024b6d a22822f29a aef8c32b64 ba0	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae	0.0033885 52	non-metS	NR_156081.1
113	0793a11c3 4c2b2beab e8e4712a2 2616e	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae;D_5__Rum inococcaceae UCG-014	0.0029378 29	non-metS	NG_041947.1
114	110815661 a57172b4b 18a492988 29788	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae;D_5__Osci llibacter	0.0023056 28	non-metS	NR_118156.1
115	fe0aee941d c1bc73615 0ba3cbaf93 e6d	D_0__Bacteria;D_1__Bact eroidetes;D_2__Bacteroidi a;D_3__Bacteroidales;D_4 __Muribaculaceae	0.0022928 81	non-metS	NR_041691.1
116	073e9d76a 75fa6b01be 53a8aa21b 994e	D_0__Bacteria	0.0020377 96	non-metS	
117	073e37fc64 c343acd9a 0286aa409 a098	D_0__Bacteria;D_1__Bact eroidetes;D_2__Bacteroidi a;D_3__Bacteroidales;D_4 __Rikenellaceae;D_5__Ali stipes;D_6__Alistipes sp. N15.MGS-157	0.0015290 73	non-metS	NR_133025.1
118	082fc0f35b fcfe17b1fa 9cdb95be 0d3	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae	0.0013414 25	non-metS	NR_116814.1
119	09df98e69 193ef0f274 e04e878de 7a42	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae;D_5__Rum iniclostridium 5;D_6__uncultured organism	0.0012567 47	non-metS	NR_114789.1
120	04e610814 8c317b415 936a4affc5 9dac	D_0__Bacteria;D_1__Bact eroidetes;D_2__Bacteroidi a;D_3__Bacteroidales;D_4 __Muribaculaceae;D_5__u ncultured bacterium;D_6__unculture d bacterium	0.0012179 29	non-metS	NR_144616.1

121	fc7b26ff8c 122d3a167 d4b09c542 dcad	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Lac hnospiraceae	0.0009219 44	non-metS	NR_042152.1
122	1ea9bcf2fe 1a7547916 3851ec5f93 f18	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae;D_5__Rum iniclostridium 9	0.0009135 29	non-metS	NR_147370.1
123	c622c8334 51f8d311a 20de39e13 25a7c	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Clo stridiales vadinBB60 group	0.0009105 17	non-metS	NR_029034.1
124	09d11d0ed 05e35a543 2a4e529f7f 72d5	D_0__Bacteria;D_1__Firm icutes;D_2__Clostridia;D_ 3__Clostridiales;D_4__Ru minococcaceae	0.0002195 01	non-metS	NR_144736.1

**APPENDIX F: DELETERIOUS SNPs AND INDELS IN GENES ESSENTIAL TO
ANTIGEN PROCESSING AND IMMUNE ACTIVATION IN THE TALLYHO/JNG
MOUSE GENOME**

SNPs							
Gene	Chr	Variant Types	# of private SNPs	Minimum SIFT score for private SNPs	Minimum PROVEAN score for private SNPs	Start	End
Slamf9	1	missense_variant	2	0.09	-3.02	172475 358	17247 8575
Lrba	3	missense_variant	1	0.03	-3.31	862246 90	86782 693
Ly6g6f	17	missense_variant	2	0.06	-2.96	326594 10	32676 687
INDELS							
Gene	Chr	Variant Types	# of private Indels			Start	End
Ii2	3	frameshift_variant	1			371205 23	37125 959
Ii20ra	10	frameshift_variant	1			197125 87	19760 053
H2-Q1	17	frameshift_variant	2			353204 05	35325 099
Lrba	3	splice_region_variant,intron_variant	3			862246 90	86782 693
Vcam1	3	splice_region_variant,frameshift_variant,intron_variant	3			116110 020	11612 9688
Vtcn1	3	splice_region_variant,intron_variant	1			100825 459	10089 5679
Lrmp	6	splice_region_variant,non_coding_transcript_variant,NMD_transcript_variant,intron_variant	7			145115 653	14517 4934
H2-DMa	17	splice_region_variant,non_coding_transcript_variant,intron_variant	2			341351 82	34139 101

APPENDIX G: MOUSE HOUSING INFORMATION

Mouse ID	# of mice per cage
B6-Chow-24	4
B6-Chow-412	4
B6-Chow-413	4
B6-Chow-414	4
B6-Chow-415	2
B6-HSHF-418	3
B6-HSHF-419	3
B6-HSHF-420	3
B6-HSHF-45	5
B6-HSHF-46	5
B6-HSLF-34	3
B6-HSLF-35	3
B6-HSLF-38	2
B6-HSLF-39	2
B6-HSLF-7	2
B6-HSLF-9	2
TH-Chow-1793	3
TH-Chow-1797	3
TH-Chow-1798	2
TH-Chow-1842	2
TH-Chow-1843	4
TH-Chow-1844	3
TH-HSHF-1789	4
TH-HSHF-1800	4
TH-HSHF-1802	4
TH-HSHF-1838	2
TH-HSHF-1885	3
TH-HSHF-1886	3
TH-HSLF-1806	3
TH-HSLF-1808	2
TH-HSLF-1816	2
TH-HSLF-1817	3
TH-HSLF-1851	3
TH-HSLF-1876	3