EXPLORING MICROBIAL METABOLISM OF DIETARY FIBERS WITHIN THE GUT OF AN OMNIVOROUS COCKROACH MODEL

by

RACHEL LOUISE DOCKMAN

(Under the Direction of Elizabeth A. Ottesen)

ABSTRACT

The complex dynamics between a host organism and its accompanying microbiota have long been understood as important to host health. Recent research has focused on uncovering general rules that broadly dictate microbial community composition and host consequences. However, this work is limited by a lack of low-cost models that can replicate the complexity of a mammalian gut microbiome while remaining tractable to manipulation.

The American cockroach is a large and long-lived omnivorous insect that has served as a model for scientific pursuits in neurology and physiology, and more recently as an emerging model for the study of host-microbiome interactions. Despite its history, there is no complete genome assembly of this remarkable insect to guide molecular work. To fill this gap, we have sequenced and assembled the genome of *Periplaneta americana* to chromosomal resolution and identified genes of interest for future work in elucidating host factors involved in microbial community composition.

The cockroach contains a diverse hindgut microbiome that resembles those of omnivorous mammals. Prior research has established the gut community of *P. americana* to be resilient in the face of extreme dietary shifts, so we tested this capacity with synthetic single-

fiber diets. Our results demonstrated that these diets induced fiber-dependent shifts in microbiome composition due to overgrowth of individual taxa in response to specific polysaccharides. The utility of the cockroach as an *in vivo* culture system, where single-source purified components can be studied within the context of a complex community, has exciting prospects for future microbiome research.

Cockroach hindgut microbiota encode functions for systematically degrading substrates. Genomes from culture, metagenomes, and single cells have increased knowledge of cockroach microbial members but how they functionally respond to different fibers is unexplored. To decipher microbial response to fiber, we gave xylan, cellulose, or a mixture as sole carbohydrate source in synthetic diets to adult cockroaches, then sequenced the 16S rRNA gene and metatranscriptomic activity. Our results uncovered different organismal responses to the shifting fiber gradient due to the diverse mechanisms they employed to survive or thrive. These findings showcase our host model for performing in-depth organism-centric microbiome analysis in complex gut communities.

INDEX WORDS: cockroach, Periplaneta americana, gut microbiome, metatranscriptome, genome, immunity, RNAi, ecology

EXPLORING MICROBIAL METABOLISM OF DIETARY FIBERS WITHIN THE GUT OF AN OMNIVOROUS COCKROACH MODEL

by

RACHEL LOUISE DOCKMAN

BSA, University of Georgia, 2019

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2025

EXPLORING MICROBIAL METABOLISM OF DIETARY FIBERS WITHIN THE GUT OF AN OMNIVOROUS COCKROACH MODEL

by

RACHEL LOUISE DOCKMAN

Major Professor:	Elizabeth A. Ottesen
Committee:	Robert J. Maier
	Christine M. Szymanski
	Kevin J. Vogel

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2025

ACKNOWLEDGEMENTS

My love for biology began somewhere between catching my first roly poly and discovering Bill Nye the Science Guy. From that moment, I knew with absolute certainty that I would be either a veterinarian, an entomologist, a neurosurgeon, a pharmacist, a geneticist, a microbiologist, or a high school English teacher (thanks, *Dead Poets Society*). Thankfully, with the incredible support of my family, my mentors, and my friends, I was empowered to explore all these disciplines instead of settling on just one. As this stage of my academic journey comes to an end, I would like to thank some of those who shaped me into the scientist I am today.

First and foremost, I want to thank my thesis advisor, mentor, and scientific enabler, Dr. Elizabeth Ottesen. As a neuroscience undergraduate, I was anxious that you might dismiss me when I approached you about rotations. Instead, you showed genuine interest in my experiences and shared your infectious enthusiasm for research, marking the first of many lively scientific conversations we would have throughout the years. You supported my interdisciplinary explorations, only reining me in when my ideas shifted from ambitious to fantastical, and guided the development of my project management skills. Your mentorship has been invaluable through both academic and personal challenges (including a global pandemic!), and I've enjoyed our conversations about books and terrariums. Thank you for taking a chance on me, Liz.

Thank you to my committee members Rob Maier, Christine Szymanski, and Kevin Vogel, all of whom have cheered me on while providing insightful feedback throughout my project. I would especially like to thank the resident entomologist on my committee, Kevin: I relied extensively on your guidance throughout the eukaryotic aspects of my work, but you

always provided more than just scientific assistance. You believed in me from the beginning and offered infinite kindness and reassurance when I agonized over the little things. I am lucky to have you on my side.

My interest of the moment shifted in lockstep with my textbook chapters, but the love and support of my family remained steady. To my father, Jeff Dockman: you taught me to be inquisitive about the world around me, laying the foundation my achievements rely on today. To my mother, Valerie Corbett: you instilled and nurtured my insatiable love for reading, which became the scaffold that allowed me to build knowledge far beyond my scientific lane. To my grandparents, Walter and Sharon Corbett: our weekly dinners, sparked by the beginning of a pandemic, have provided great conversations and endless encouragement over the years.

My academic journey first became possible thanks to my first advisor, Dr. Nikolay Filipov, and my patient graduate mentor, Jessica Carpenter. With their guidance, I discovered my love for research and was emboldened to become a scientist myself. During graduate school, I've had the honor of working with my amazing lab mates and colleagues Helen, Coralis, Amber, and Kevin, who have become my dear friends. Their technical and emotional support has been a lifeline over the years. I must also thank my lovely undergraduate mentees, Sarah Beth and Melissa, who cheerfully performed much of the benchwork integral to my project while keeping me up to date on undergrad slang.

Last, but certainly not least, I thank my significant other ($p < 0.001$), Kobe Largeman. I could not have made it through this degree without your love, support, and delicious dinners.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction to the gut microbiome

Host-associated microbiota are intimately involved in their host animal's overall well-being. These microorganisms, found on and within their host, influence their host's immunity [1-3], nutrition [4-6], and development [7-10], which are mediated through their interactions with the host itself and how they coexist or compete with fellow microbial community members. The complexity of these interactions scale with the diversity of the resident microbiota and the surrounding chemical ecosystem, creating countless possible configurations that may occur even at the detriment of their host. Achieving and maintaining an optimal gut microbiome, the most diverse of host-associated communities [11], is a major goal of microbiome research, and decades of research has successfully identified outputs that are universally desirable to gut-possessing organisms [6, 12, 13]. However, uncovering possible avenues for obtaining these benefits, or the more numerous paths leading to microbiome-related disease, remains a difficult task [14-19]. Given the many ways a gut microbiome can destabilize, teasing apart complex interactions shaping its structure requires observing the microbial residents and their behaviors *in vivo*, where host factors are considered and consequences detected. Therefore, to further our understanding of the microbial players and activities involved in microbiome homeostasis, it is necessary to develop models that capture intricate microbial interactions as they occur within their natural environment.

The composition of an organism's gut microbiome is dependent on an interconnected web of factors including its habitat [20-23], its genetic makeup [24-26], and its nutritional niche [4, 27-29]. For example, the gut microbiome of blood-feeding insects is shaped by their highly specific diet, which was determined by their evolutionary acquisition of hematophagy [30-32]. The bovine gut microbiome is shaped by their development of a microbial fermentation vat, the rumen, which enables them to support their half-ton bodies on the largest and most indigestible source of biomass on the planet [33, 34]. Finally, the human gut microbiome lacks the extreme specialization inherent in the communities of the two previous examples, but instead is equipped to handle a variety of possible food sources while still supporting shared lineages of human-associated microbes [11, 25, 35, 36]. In general, the microbial community is constrained by what the host evolved to consume, but as omnivores, humans are amenable to a wide range of dietary modifications, making diet a promising avenue to explore for promoting desirable microbiome compositions [37-39]. Towards this end, isolating the effects of specific dietary nutrients on gut microbiota can be accelerated by using a tractable model organism that shares microbial lineages with human microbiota, but is capable of withstanding intense nutritional limitation, such as the cockroach.

Omnivorous cockroaches present an intersection of tractability and complexity for gut microbiome studies. As insects, cockroaches are simple to maintain with fewer barriers to entry than with mammalian models, while their opportunistic feeding habits promote a rich and diverse microbiome that remains stable in the face of extreme dietary shifts [40-43]. Pest cockroaches such as *Periplaneta americana* may be genetically distinct from humans, but their habitat, diet, and microbiome composition overlap substantially [40, 44, 45], positioning these insects as unique candidates for modeling gut microbial dynamics in omnivores.

In this dissertation, I will present the American cockroach (*Periplaneta americana*) as a model host organism that replicates the complexity of the human microbiome while overcoming limitations of mammalian or culture-based investigations. To do this, I will detail traits that uphold *P. americana* as a tractable model organism applicable to humans, especially as related to its gut microbiome. I then describe the microbiota found in this insect and the interactions they share that shape both the host and community activity. Afterwards, I define dietary fiber types and introduce those relevant to my research, as well as what is known about microbial processing of these fibers. Finally, I will outline the research I have performed and how it has enhanced our knowledge of interplay between dietary fiber and gut microbial dynamics in a complex omnivorous insect model.

The American cockroach as a valuable model organism

Cockroaches require little introduction; despite most species peacefully living apart from humans, cockroaches are regarded as nefarious invaders that strike fear into the hearts of those who venture to the kitchen for a midnight snack [46]. However, Crampton [47] has a different perspective that I find quite appealing:

> During the time in which the roach type has been in existence, the great dinosaurs have come and gone, and birds, mammals, and flowering plants have arisen and developed their myriad profusion; but amid there ceaseless comings and goings, the roach type has pursued the even tenor of its way practically unaffected by the passing of the ages. It is therefore of some interest to study the makeup of an organism so perfectly constructed that it has been able to defy the ravages of time and changing conditions (Crampton, 1925, p. 195)

American cockroaches have a long and rich history as research subjects not just in entomology, but in neuroscience, immunity, and physiology [48-50]. Compared to other insects, many features of *P. americana* have established its position as an ideal model system for an assortment of fields. Adults reach approximately 1.5” long, which allows for easier handling and more forgiving dissections than with smaller insects. They are long-lived, with a two-year lifespan that

is comparable to that of mice, and reproduce prolifically once they reach adulthood [51]. Large colonies are simple and inexpensive to maintain on standard laboratory chow, and since they are hemimetabolous insects that grow wings with their final ecdysis (8-12 molts), identifying adult insects is a straightforward task [51].

The cockroach has proven to be compatible with modern genetic tools and techniques, including early genome assemblies [52, 53], which have contributed to knowledge on the genes involved in insect wing development [54], putative allergens shed by cockroaches [53], and highly developed mechanisms underlying *P. americana*'s ability to survive pesticides [55-57]. An increasing number of studies are leveraging this insect's natural susceptibility to RNA interference (RNAi) to further our knowledge of insect gene function [52, 58-60]. RNAi is a conserved immune response that identifies double-stranded RNA (dsRNA) and destroys matching mRNA of suspected viral origin; by introducing synthetic dsRNA matching a gene of interest, the RNA induced silencing complex (RISC) can be manipulated into cleaving host mRNA, knocking down gene expression [61-63]. Unlike common model insects such as *Drosophila*, cockroaches are susceptible to systemic RNAi no matter if the dsRNA is given orally or injected [52, 58, 60, 64-67]. In the context of microbiome research, RNAi will be a powerful tool for removing host contributions to gut microbial community formation, elucidating precisely how some hosts regulate their gut microbiome.

An additional advantage of *P. americana* as an experiment insect is its ability be investigated in the absence of microbiota (germ-free) or with selected individuals (gnotobiotic). Germ-free mouse research has been extremely valuable to deciphering microbial contributions to mammalian host health [10, 68, 69], but maintaining germ-free rodents is both labor-intensive and costly [70, 71]. In comparison, cockroach oothecae can easily be sterilized prior to eclosion

(although they retain their endosymbiotic *Blattabacterium*) and the gnotobiotic nymphs are kept on the bench top, greatly reducing barriers to germ-free research [72]. Research leveraging gnotobiotic cockroaches has explored topics such as importance of gut microbiota to cockroach development [9, 73], the factors involved in initial gut colonization [74, 75], and bacterial metabolic products that exert influence over host behavior and physiology [74, 76].

Finally, American cockroaches are omnivores with voracious appetites that will consume nearly anything edible (**Figure 1.1**), although they do select for their desired macronutrient ratios [51, 77, 78]. This flexibility combined with their long-term survival on low quality foods is especially attractive for experiments testing the gut microbial response to a range of substances including purified nutrients, whole foods, and textiles [40, 79]. Customizable synthetic diets with chemically defined nutritional profiles are a valuable resource for determining the interplay between specific nutrients, host, and microbiome, and are commonly used with media-fed insects such as *Drosophila* [80, 81]. However, fruit flies possess a comparatively small and simple gut microbiome, reducing the inferences that can be adapted to the complex and diverse gut community found in humans. Cockroaches overcome these limitations.

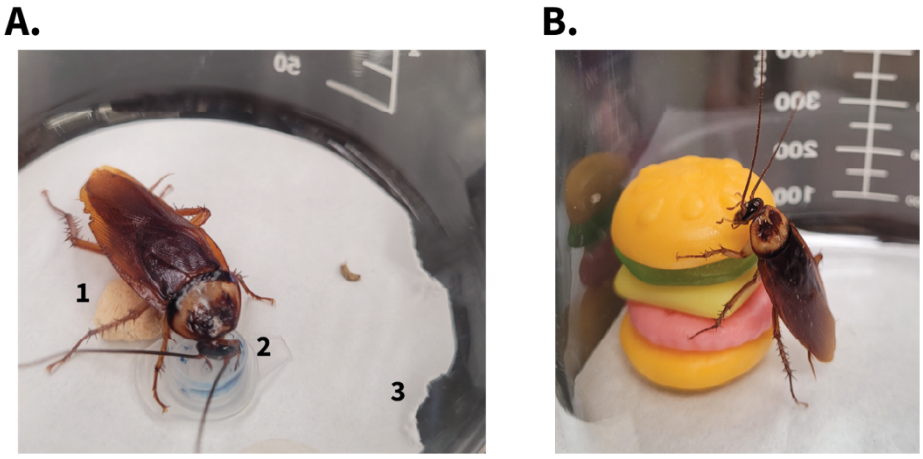


Figure 1.1: Dietary flexibility of laboratory cockroaches. Photographs of (A) an experimental American cockroach maintained on (1) a synthetic diet consuming (2) blue dye, with additional evidence of (3) filter paper consumption and (B) creative depiction of a cockroach eating a humanized western diet (Krabby Patty™).

Cockroach digestive tract physiology

Structurally, the cockroach digestive tract consists of three primary sections that are analogous to the human digestive tract: the foregut, midgut, and hindgut of the cockroach correlate to the stomach, small intestine, and colon with some caveats that I will discuss.

The foregut includes the mouth, esophagus, and the crop, a cuticle-lined compartment that stores food after ingestion. The crop is capable of substantial expansion and is estimated to comprise 40-60% of the total gut volume [82-84], storing food for several days following periods of starvation or dehydration [51]. While stored in the crop, carbohydrate degradation begins through the action of salivary amylases and beta-glucanases [84, 85], while trypsin derived from the midgut enters the crop to initiate protein digestion [84]. Lipase secreted into the midgut lumen also enters the foregut to begin fat hydrolysis, the products of which seem to be able to pass through the cuticle and be incorporated into the host prior to entering the midgut [86]. Microbial activity and density are lower in the foregut with higher abundances of *Pseudomonadota* (Proteobacteria) than in successive regions (**Figure 1.2**); likely, the structure of the foregut microbiome is largely influenced by transient environmental microbes picked up from the diet [83]. The chitinous lining protecting the crop may provide an attachment point for long-term resident microbes that can tolerate influxes of oxygen, but in nymphal stages, this lining along with adherent bacteria are shed with ecdysis.

The foregut and midgut are separated by the proventriculus or gizzard, a grinding organ with chitin spines that shreds food into pieces accessible to enzymatic digestion. While some digestive enzymes are active in the foregut, the midgut is the primary site of host digestion and nutrient absorption like in the mammalian small intestine [27, 84]. There are two distinct sections that make up the midgut: the ventriculus, which visually resembles the small intestine, and the

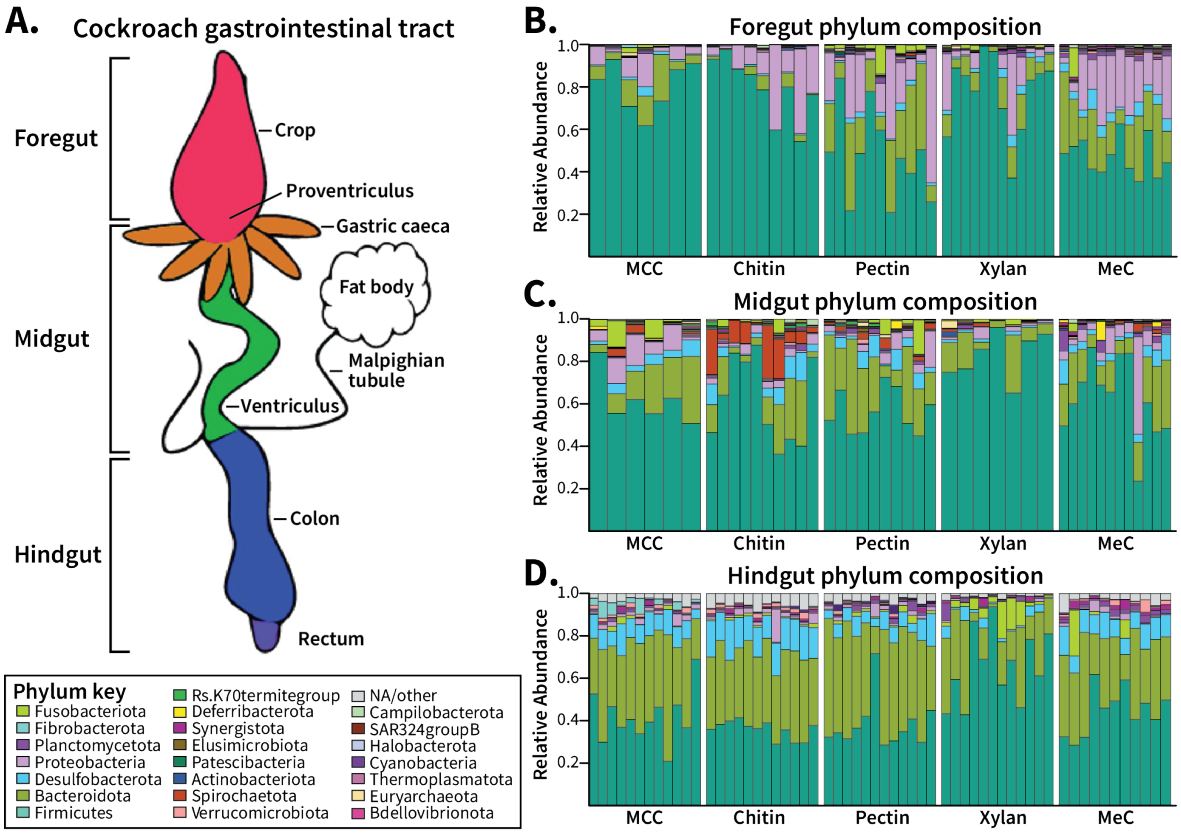


Figure 1.2: Digestive tract anatomy and regional microbiomes of the cockroach. (A) Cartoon depiction of the cockroach gastrointestinal tract with the primary gut regions indicated on the left and specific features on the right. Relative abundances of microbial phyla are displayed in bar plots for the **(B)** foregut, **(C)** midgut, and **(D)** hindgut regions of adult *P. americana* fed different synthetic diets. MCC: microcrystalline cellulose; MeC: methylcellulose

eight gastric caeca, extensions of the midgut which branch off from the ventriculus at the proventriculus [82]. The ventriculus is lined with a sieve-like membrane composed of chitin and glycoproteins called the peritrophic matrix (or membrane) [87]. In cockroaches, this layer continuously delaminates from the midgut epithelial microvilli to form a tight-knit lattice that allows small molecules, such as digestive enzymes from midgut tissue or partially digested foods, to pass through while filtering out microbial cells [87]. The digesta moves in an antiperistaltic motion from the posterior midgut, through the peritrophic matrix, and into the gastric caeca where additional processing and absorption occurs; this flow of digesta is also the

method through which midgut-derived digestive enzymes reach the foregut [84]. In addition to the lipase and trypsin discussed above, enzymes produced by the midgut include trehalase, alpha-glucosidase, beta-glucosidase, beta-galactosidase, chymotrypsin-2, and aminopeptidase [57]. Periodically, the peritrophic matrix sloughs off and proceeds into the hindgut along with indigestible or residual matter to be processed by gut bacteria and ultimately excreted. Due to high levels of host activity within the midgut and the continuous removal of attachment points, the midgut microbial community is typically smaller and more variable than in the hindgut, with fewer sustained community members than are found in the hindgut (**Figure 1.2**).

Before proceeding to the hindgut, it is important to point out the involvement of several insect-specific organs in their digestive function. Adjacent to the gut and filling the body cavity is the fat body, a tissue which stores energy in addition to performing most intermediary metabolism, producing hormones, circulating compounds, and antimicrobial peptides that are released into the surrounding hemolymph, and detoxifying nitrogenous compounds [88]. The waste generated by fat body metabolism is removed through Malpighian tubules (approximately 150 in *P. americana*), hair-like tubes projected throughout the body cavity that empty into the midgut-hindgut junction [89]. In most insects, uric acid is a major waste product transported through these tubules, but *P. americana*, as well as many other cockroaches (excluding the cave-dwelling genus *Nocticola* [90]), handle nitrogen differently; they store uric acid within specialized fat body cells adjacent to an endosymbiont from the bacterial genus *Blattabacterium* [78, 91-96]. This endosymbiont, which is vertically transmitted from mother to offspring, became associated with the cockroach at some point after they diverged from termites and has long lost its ability to survive alone [97, 98]. However, it does encode essential amino acid synthesis pathways that sustain its host during nitrogen starvation using the accumulated uric

acid [99-102]. The trade-off of this symbiosis is that high protein levels can lead to accumulation of uric acid that may become detrimental to the host [103]. While the cockroach certainly benefits from this nutrient provisioning, it is unclear as to if any amino acids or nitrogenous compounds are excreted into the hindgut through the Malpighian tubules during starvation periods as well, or if the level of amino acids produced are regulated based on host need; perhaps, some *Blattabacterium*-derived amino acids are indirectly fed to the gut microbiota through incorporation into the continuously produced peritrophic matrix.

The cockroach hindgut receives waste from Malpighian tubules as well as undigested food and periplasmic matrix material from the midgut. The cockroach hindgut functionally resembles the mammalian colon [51], acting as a site of water resorption, waste consolidation and microbial fermentation of undigested food with little input from host-derived enzymatic metabolism. Like the foregut, the hindgut wall is lined with exoskeleton that is shed with each nymphal molt until adulthood, after which it remains intact. Insect hindgut structures can vary immensely; *Drosophila* have small hindguts in comparison to their elongated midguts, while termites have large hindguts with distinctive paunch(es) that bear similarities to herbivorous hindgut-fermenting mammals with an enlarged cecum [27]. The hindgut of *P. americana* is somewhere in between those two extremes; it is similarly sized as its midgut, and while it lacks the specialized fermentative regions found in herbivores, the hindgut supports an □oolkie microbial community reflecting the microbiome found in humans [84]. Further, cockroach microbiota are capable of fermenting diverse structural polysaccharides, such as hemicellulose (xylan, xyloglucan), cellulose, or pectin, into microbial metabolites such as volatile fatty acids (VFAs; butyrate, propionate, acetate) [41, 51, 104, 105]. A trait uncommon to most insects was discovered in ultra-structure studies of the cockroach hindgut [106]; the *P. americana* hindgut

wall is embedded with intercellular channels that allow VFAs and other microbial products, such as microbially derived amino acids and vitamins, to be absorbed into the hemolymph, confirming the existence of nutritional cross-feeding between *P. americana* and its gut microbiota [41, 101, 107-109]. Altogether, the American cockroach possesses gut morphology suitable for modeling dynamics between an omnivorous host and its complex gut microbiome.

Composition of the cockroach hindgut microbiome

The cockroach hindgut provides an environment conducive to microbial growth. Oxygen levels within the intestinal lumen form a gradient that simultaneously supports facultative and obligate anaerobes while carbon and nitrogen are continuously supplied, if not from the host's diet, then from host glycans and cell material [110-116]. These qualities, in addition to long life span and varied diet, support diverse gut microbes in cockroaches that resemble those found in omnivorous mammals. While they are supported by host nutrient provisioning, gut microbiota must still compete for limited nutrients, devise their own strategies to use uncommon sources for energy, intermediates, or reducing power, or form mutualistic interactions with other community members [117-121]. Understanding the taxonomic structure of the gut community and the general approach these groups take towards competition or coordination is helpful for deciphering their contributions to the gut nutritional landscape and subsequent benefits obtained by their host. In general, the taxonomic structure of the gut microbiome in *P. americana* is dominated by higher proportions of *Bacteroidota* and *Bacillota* with additional representation from *Desulfobacterota*, *Archaea*, and low levels of *Pseudomonadota* [40, 104, 105, 122, 123]. I will cover primary members of the cockroach gut microbiota and their proposed positions in the gut trophic structure shown in **Figure 1.3** as background for understanding the role these microbes play in maintaining the cockroach gut microbiome [104, 124].

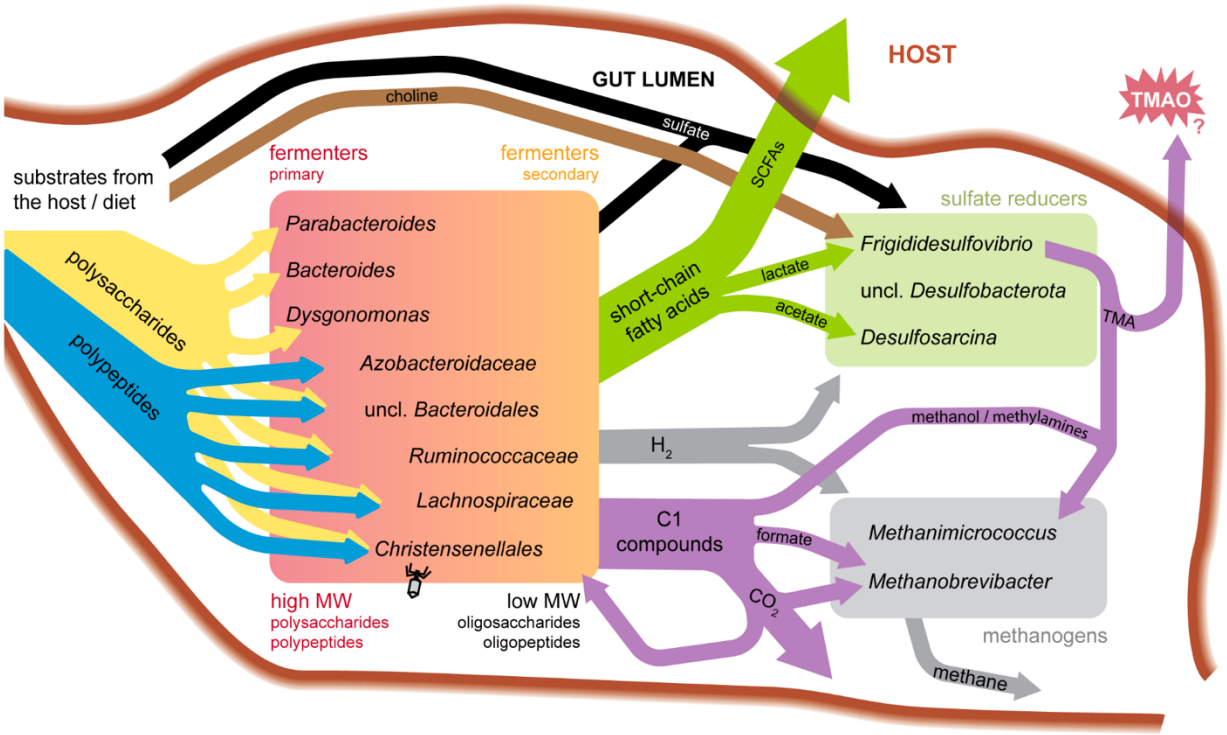


Figure 1.3: Model of substrate utilization in the cockroach hindgut (from Dukes et al., 2023). C1: single carbon; MW: molecular weight; SCFAs: short-chain fatty acids; TMA: trimethylamine; TMAO: trimethylamine N-oxide; uncl: unclassified.

The most abundant macronutrients that reach the gut for microbial processing are polysaccharides and polypeptides, which can be further differentiated into high or low molecular weight [13]. These nutrients are degraded and subsequently fermented primarily by the phyla *Bacteroidota* and *Bacillota* [104, 125]; in humans, these groups are also highly abundant and serve similar functions [11, 126, 127]. The primary degraders largely consist of *Bacteroidota* members, which are commonly associated with polysaccharide degradation and considered important producers of VFAs, especially propionate, that benefit the host [128-130]. During carbohydrate metabolism, they additionally produce both oligosaccharides and intermediate compounds that are picked up by other organisms unable to degrade the intact polysaccharide themselves [131, 132]. *Bacillota* are considered secondary fermenters that use an arsenal of ABC transporters to import sugars, oligosaccharides, and amino acids [133-137].While they are well

known for their ability to ferment peptides and consume lactate released by other organisms [137-140], the phylum is functionally diverse and includes members that specialize in polysaccharide degradation [141-143]. Some members of *Bacillota* are known for butyrate production, the main energy source for intestinal epithelial cells in mammals [144-147].

Together, *Bacteroidota* and *Bacillota* break down complex molecules, converting dietary or host-derived substrates into cell material for themselves while releasing intermediates that are consumed by the host or fed into the subsequent gut trophic levels, the methanogenic *Archaea* and sulfate-reducing *Desulfobacterota* [124, 129]. These groups compete for hydrogen produced by primary and secondary fermentation, with *Desulfobacterota* additionally scavenging sulfate released by *Bacteroidota* during host glycan degradation [148]. The interactions between these trophic levels depend heavily on the functional capacity of present organisms, with different metabolic strategies resulting in distinct profiles of metabolites dumped into the environment; how microorganisms choose their strategy depends on the molecules that reach the gut in the first place: host diet.

Compositional studies on cockroach gut microbiota to dietary composition in general have found conflicting results; some studies have found diet does not change the gut community [149, 150] and others did find effects of diet [151-153]. Work in our lab has established that the composition of the cockroach microbiome remains largely unchanged when confronted with whole-food diets containing various macronutrient forms [40], while synthetic sources of carbohydrates trigger blooms in individual ASVs [79], suggesting that cockroach gut microbiota possess mechanisms to maintain balance when confronted with diverse substrates but can overcome this innate stability when the natural intrinsic structure of food is removed. While the intrinsic fiber found in intact cell matrices is an interesting target for microbiome research, my

dissertation focuses on studying the concerted efforts of gut microbiota in response to individual fibers through synthetic diets.

Microbial fermentation of fiber in the cockroach hindgut

Fiber is a key substrate that influences both microbial composition and functional capacity within a gut system, and sufficient fiber is often correlated with a robust and diverse gut microbiome [154-157]. Dietary fiber encompasses all indigestible structural plant polysaccharides (primarily cellulose, hemicellulose, and pectin), which are further complicated due to factors such as plant species, age at harvesting, and any other environmental factors that shaped the plant's growth [158, 159]. In addition to sugar chain and branch complexity, natural fibers often occur tightly interlaced together, requiring resident microbes to either possess extensive enzymatic machinery to access these fibers or to scavenge manageable intermediates discarded by the primary degraders [131, 160-163]. Not only do different fibers require targeted approaches for microbial degradation, but the metabolites produced during fermentation depend on fiber structure as well [158, 164, 165], highlighting the importance of dissecting out precise microbial responses to both intrinsic and isolated fibers.

While there is a substantial body of work describing microbial degradation of fiber in gut contexts, less is known about the mechanisms employed by omnivorous cockroach gut microbiota. The close relatives of the cockroach, termites, have an extensive literature base describing their gut microbiota and their impressive fiber-degrading abilities, with many shared microbial lineages that are found in cockroach gut communities [40, 104, 122, 123, 166-168]. The host contribution of both termites and *P. Americana* to fiber, particularly cellulose, degradation is similar as well; both *P. americana* [85, 169, 170] and termite relatives [171] secrete endogenous beta-glucanases that liberate cellobiose or glucose from soluble or simpler

cellulose chains, allowing the host to obtain some energy without microbial assistance. However, antibiotic treatment greatly decreases the amount of carbon liberated in the cockroach [43] and termites cannot survive without their gut symbionts [75], underscoring the reliance of these organisms on gut microbiota for energy retrieval from high-fiber diets.

The organisms involved in hindgut fiber degradation differ between cockroaches, higher termites, and lower termites. Higher termites have fiber-degrading prokaryotic gut symbionts within their paunch which consist heavily of spirochetes with some *Fibrobacter* species [123, 166-168]. In our observations, *Spirochaetes* were most abundant in the midgut of chitin-fed *P. americana*, suggesting that cockroach-associated *Spirochaetes* are functionally and phylogenetically distinct from those found in higher termites (**Figure 1.2)**. The American cockroach gut community differs from lower termites as well, which host cellulolytic flagellate protists for fiber degradation; some research has suggested a role for protists in cockroach cellulose degradation [172, 173], but their activity cannot explain all observed cellulose metabolism. As discussed earlier, the cockroach hindgut supports many *Bacteroidota* and *Bacillota* species that are known as key fiber degraders in mammals. I will cover the distinctive mechanisms used in fiber catabolism by these groups as well as in *Fibrobacter*, an uncommon cockroach gut organism which possesses unique cellulolytic machinery.

The fiber-degrading ability of *Bacteroidota* has been well-documented [128, 174, 175]. Many members of this phylum, most notably those within the genus *Bacteroides,* possess an expansive carbohydrate-active enzyme (CAZyme) repertoire with high substrate flexibility [42, 104, 128, 176-181]. Many of these CAZymes are encoded in *Bacteroidota* genomes as clusters of co-regulated genes, regions called polysaccharide-utilization loci (PULs), that orchestrate the degradation of targeted fibers from first detection to final transport [182]. The PULs are

substrate-dependent, and include genes with hydrolase, esterase, binding, and regulatory functions bundled together to be deployed when the appropriate polysaccharide is encountered, bestowing metabolic flexibility that allows *Bacteroidota* to adapt to host dietary decisions [183]. While their modular and targeted PUL structures enable these microbes to make the most use out of diverse polysaccharides, they also participate in cross feeding interactions that support the growth of other organisms [131, 132, 174]. When carbohydrates are not provided via the diet, *Bacteroides* have been found to support themselves with host-derived mucins [114, 115], although this is less explored in those from insect lineages.

Bacillota in omnivore guts are often assumed to be secondary fermenters that scavenge sugars liberated by neighboring *Bacteroidota*. However, in herbivorous systems, *Bacillota* such as *Clostridium thermocellum*, *Clostridium cellulolyticum*, and *Ruminococcus flavefaciens* are important cellulose degrading bacteria through cellulosome activity [184-187]. Cellulosomes are complex multiprotein structures which are secreted by the bacterium and affixed to its cell surface and each other using dockerins and cohesins [188-190]. These structures are highly effective at degrading heterogeneous, insoluble fibers, but may be less flexible than the PUL system in *Bacteroidota*, possibly explaining their lower incidence in omnivorous gut consortia. Despite this, there is recent evidence that humans harbor their own lineages of cellulosome-wielding *Bacillota,* so it is possible the cockroach hosts similar organisms that are currently undiscovered [191]. In the cockroach, *Bacillota* have their own arsenal of CAZymes that are expressed to degrade diverse polysaccharides and increase in abundance on high-fiber diets, although their ability to completely degrade complex substrates alone is unclear [152, 192]. Largely, *Bacillota* benefit from cross-feeding with primary fiber degraders, facilitated by their

abundance of transport mechanisms geared towards importing diverse mono- and oligosaccharides [133-135, 193].

The Gram-negative *Fibrobacterota* are well-characterized members of ruminant and herbivore monogastric gut communities, but they are rarely identified or studied in omnivorous animals [194, 195]. They have been observed in termite guts and in *P. americana* when given a diet high in crystalline cellulose [79, 167]. These bacteria are cellulose specialists with unique mechanisms of fiber metabolism compared to the *Bacteroidota* and *Bacillota* commonly observed in these gut systems, that are highly sensitive to catabolite repression, temperature fluctuations, and pH [196]. Instead of leveraging PULs or cellulosomes, *Fibrobacter* adheres to crystalline cellulose directly via fibro-slime proteins and facilitates fiber degradation by releasing cellulases and hemicellulases prior to transporting cellodextrins into the cell for degradation [197-199]. An especially fascinating aspect of these bacteria is that they encode diverse hemicellulose CAZymes, but cannot transport or use any degradation produce other than glucose and cellobiose, instead providing the surrounding microbiota with cleaved oligo- or monosaccharides [200]. Despite their sensitive disposition, *Fibrobacter* are important members of fiber-degrading communities and represent an unusual specialist in the cockroach microbiome.

Research aims

The work in this dissertation aims to advance *Periplaneta americana* as a model organism for deciphering host-microbe interactions in omnivores with complex and translational gut microbiomes. My second chapter is a manuscript describing the assembly and annotation of a reference genome for the American cockroach, which is currently the highest quality blattid genome available on the NCBI database. This contiguous assembly is organized into 17 putative

chromosomes, which I validated with my own RNA-seq data from host gut tissues, and includes an entire chromosomal scaffold missing from a previous assembly. In my third chapter, I describe work in which I designed synthetic diets that control for the nutritional variability found in whole foods and used them to test the impact of single-source polysaccharides on the 16S rRNA gene composition of the gut microbiome. These results were compared with the whole food diet data previously generated in our lab [40], and revealed the unique influence individual fibers exert on the gut community. Finally, in my fourth chapter I dig deeper into the functional differences in microbiota from insects fed xylan and microcrystalline cellulose diets, and describe how 16S amplicon abundance and transcriptional activity differed in response to diets containing a gradient of xylan to cellulose ratios. The appendices in this dissertation contain supplemental data and figures from these manuscripts. Overall, my work, through molecular biology techniques and bioinformatic analysis, explores the dynamic, complex gut community found in an unexpected omnivorous insect.

References

1. Thaiss, C.A., et al., *The microbiome and innate immunity.* Nature, 2016. **535**(7610): p. 65-74.
2. Mazmanian, S.K., et al., *An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system.* Cell, 2005. **122**(1): p. 107-18.
3. Postler, T.S. and S. Ghosh, *Understanding the Holobiont: How Microbial Metabolites Affect Human Health and Shape the Immune System.* Cell Metab, 2017. **26**(1): p. 110-130.
4. Trevelline, B.K. and K.D. Kohl, *The gut microbiome influences host diet selection behavior.* Proc Natl Acad Sci U S A, 2022. **119**(17): p. e2117537119.
5. Spivak, I., L. Fluhr, and E. Elinav, *Local and systemic effects of microbiome-derived metabolites.* EMBO Rep, 2022. **23**(10): p. e55664.
6. Asnicar, F., et al., *Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals.* Nat Med, 2021. **27**(2): p. 321-332.
7. Donaldson, G.P., S.M. Lee, and S.K. Mazmanian, *Gut biogeography of the bacterial microbiota.* Nat Rev Microbiol, 2016. **14**(1): p. 20-32.
8. Mishra, A., et al., *Microbial exposure during early human development primes fetal immune cells.* Cell, 2021. **184**(13): p. 3394-3409.e20.
9. Vera-Ponce de León, A., et al., *Microbiota Perturbation or Elimination Can Inhibit Normal Development and Elicit a Starvation-Like Response in an Omnivorous Model Invertebrate.* mSystems, 2021. **6**(4): p. e0080221.
10. Kubelkova, K., et al., *Gnotobiotic mouse model's contribution to understanding host-pathogen interactions.* Cell Mol Life Sci, 2016. **73**(20): p. 3961-9.
11. Huttenhower, C., et al., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. **486**(7402): p. 207-214.

12. Ara, K., et al., *From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites.* Cell, 2016.

13. Oliphant, K. and E. Allen-Vercoe, *Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health.* Microbiome, 2019. **7**(1): p. 91.

14. Zaneveld, J.R., R. McMinds, and R. Vega Thurber, *Stress and stability: applying the Anna Karenina principle to animal microbiomes.* Nat Microbiol, 2017. **2**: p. 17121.

15. Dabke, K., G. Hendrick, and S. Devkota, *The gut microbiome and metabolic syndrome.* The Journal of clinical investigation, 2019. **129**(10): p. 4050-4057.

16. Zou, X., et al., *Gut microbiota and its metabolites in Alzheimer's disease: from pathogenesis to treatment.* PeerJ, 2024. **12**: p. e17061.

17. Jia, W., et al., *Gut microbiota alterations are distinct for primary colorectal cancer and hepatocellular carcinoma.* Protein & Cell, 2020. **12**(5): p. 374-393.

18. Øyri, S.F., G. Műzes, and F. Sipos, *Dysbiotic gut microbiome: A key element of Crohn's disease.* Comp Immunol Microbiol Infect Dis, 2015. **43**: p. 36-49.

19. Noble, E.E., T.M. Hsu, and S.E. Kanoski, *Gut to Brain Dysbiosis: Mechanisms Linking Western Diet Consumption, the Microbiome, and Cognitive Impairment.* Front Behav Neurosci, 2017. **11**: p. 9.

20. Rothschild, D., et al., *Environment dominates over host genetics in shaping human gut microbiota.* Nature, 2018. **555**(7695): p. 210-215.

21. Chen, C.Y., et al., *Habitat and indigenous gut microbes contribute to the plasticity of gut microbiome in oriental river prawn during rapid environmental change.* PLoS One, 2017. **12**(7): p. e0181427.

22. Kim, P.S., et al., *Host habitat is the major determinant of the gut microbiome of fish.* Microbiome, 2021. **9**(1): p. 166.

23. Wu, Y., et al., *Habitat environments impacted the gut microbiome of long-distance migratory swan geese but central species conserved.* Scientific Reports, 2018. **8**(1): p. 13314.

24. Goodrich, Julia K., et al., *Human Genetics Shape the Gut Microbiome.* Cell, 2014. **159**(4): p. 789-799.

25. Kurilshikov, A., et al., *Large-scale association analyses identify host factors influencing human gut microbiome composition.* Nat Genet, 2021. **53**(2): p. 156-165.

26. Kurilshikov, A., et al., *Host Genetics and Gut Microbiome: Challenges and Perspectives.* Trends Immunol, 2017. **38**(9): p. 633-647.

27. Engel, P. and N.A. Moran, *The gut microbiota of insects – diversity in structure and function.* FEMS Microbiology Reviews, 2013. **37**(5): p. 699-735.

28. Nishida, A.H. and H. Ochman, *Rates of gut microbiome divergence in mammals.* Mol Ecol, 2018. **27**(8): p. 1884-1897.

29. Zoelzer, F., A.L. Burger, and P.W. Dierkes, *Unraveling differences in fecal microbiota stability in mammals: from high variable carnivores and consistently stable herbivores.* Animal Microbiome, 2021. **3**(1): p. 77.

30. Muturi, E.J., et al., *Blood meal source and mixed blood-feeding influence gut bacterial community composition in Aedes aegypti.* Parasit Vectors, 2021. **14**(1): p. 83.

31. Sonenshine, D.E. and P.E. Stewart *Microbiomes of Blood-Feeding Arthropods: Genes Coding for Essential Nutrients and Relation to Vector Fitness and Pathogenic Infections. A Review*. Microorganisms, 2021. **9**, DOI: 10.3390/microorganisms9122433.

32. Husnik, F., *Host–symbiont–pathogen interactions in blood-feeding parasites: nutrition, immune cross-talk and gene exchange.* Parasitology, 2018. **145**(10): p. 1294-1303.

33. Jami, E., B.A. White, and I. Mizrahi, *Potential Role of the Bovine Rumen Microbiome in Modulating Milk Composition and Feed Efficiency.* PLOS ONE, 2014. **9**(1): p. e85423.

34. Tajima, K., et al., *Diet-dependent shifts in the bacterial population of the rumen revealed with real-time PCR.* Appl Environ Microbiol, 2001. **67**(6): p. 2766-74.

35. *The Integrative Human Microbiome Project.* Nature, 2019. **569**(7758): p. 641-648.

36. Fackelmann, G., et al., *Gut microbiome signatures of vegan, vegetarian and omnivore diets and associated health outcomes across 21,561 individuals.* Nature Microbiology, 2025. **10**(1): p. 41-52.

37. Erica, D.S. and L.S. Justin, *Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates.* Cell Metabolism, 2014.

38. David, L.A., et al., *Diet rapidly and reproducibly alters the human gut microbiome.* Nature, 2014. **505**(7484): p. 559-563.

39. Wolter, M., et al., *Leveraging diet to engineer the gut microbiome.* Nat Rev Gastroenterol Hepatol, 2021. **18**(12): p. 885-902.

40. Tinker, K.A. and E.A. Ottesen, *The Core Gut Microbiome of the American Cockroach, Periplaneta americana, Is Stable and Resilient to Dietary Shifts.* Appl Environ Microbiol, 2016. **82**(22): p. 6603-6610.

41. Cruden, D.L. and A.J. Markovetz, *Microbial ecology of the cockroach gut.* Annu Rev Microbiol, 1987. **41**: p. 617-43.

42. Vera-Ponce de León, A., et al., *Cultivable, Host-Specific Bacteroidetes Symbionts Exhibit Diverse Polysaccharolytic Strategies.* Appl Environ Microbiol, 2020. **86**(8).

43. Bignell, D.E., *An experimental study of cellulose and hemicellulose degradation in the alimentary canal of the American cockroach.* Canadian Journal of Zoology, 1977. **55**: p. 579-589.

44. James, A.G.R., *Man's Uninvited Fellow Traveler--The Cockroach.* The Scientific Monthly, 1945. **61**(4): p. 265-276.

45. Tang, Q., et al., *Solving the 250-year-old mystery of the origin and global spread of the German cockroach, Blattella germanica.* Proceedings of the National Academy of Sciences, 2024. **121**(22): p. e2401185121.

46. Grimaldos, J., et al., *Cockroaches are scarier than snakes and spiders: Validation of an affective standardized set of animal images (ASSAI).* Behav Res Methods, 2021. **53**(6): p. 2338-2350.

47. Crampton, G.C., *The External Anatomy of the Head and Abdomen of the Roach, Periplaneta americana.* Psyche: A Journal of Entomology, 1925. **32**(4-5): p. 195-226.

48. Roeder, K.D., *The effect of potassium and calcium on the nervous system of the cockroach, Periplaneta americana.* J Cell Comp Physiol, 1948. **31**(3): p. 327-38.

49. Bodenstein, D., *Studies on nerve regeneration in Periplaneta americana.* J Exp Zool, 1957. **136**(1): p. 89-115.

50. Troppmann, B., B. Walz, and W. Blenau, *Pharmacology of serotonin-induced salivary secretion in Periplaneta americana.* J Insect Physiol, 2007. **53**(8): p. 774-81.

51. *THE AMERICAN COCKROACH. Edited by William J. Bell and K. G. Adiyodi.* Physiological Entomology, 1982. **7**(1): p. 117-118.

52. Li, S., et al., *The genomic and functional landscapes of developmental plasticity in the American cockroach.* Nature Communications, 2018. **9**(1): p. 1008.

53. Wang, L., et al., *Genome assembly and annotation of Periplaneta americana reveal a comprehensive cockroach allergen profile.* Allergy, 2023. **78**(4): p. 1088-1103.

54. Hrycaj, S., J. Chesebro, and A. Popadić, *Functional analysis of Scr during embryonic and post-embryonic development in the cockroach, Periplaneta americana.* Dev Biol, 2010. **341**(1): p. 324-34.

55. Sun, H., et al., *Elucidating the detoxification efficacy of Periplaneta americana delta glutathione S-transferase 1 (PaGSTd1) against organophosphates.* Pestic Biochem Physiol, 2024. **203**: p. 106013.

56. Zhang, J.H., et al., *New insight into foregut functions of xenobiotic detoxification in the cockroach Periplaneta americana.* Insect Sci, 2018. **25**(6): p. 978-990.

57. Zhang, J., et al., *Midgut Transcriptome of the Cockroach Periplaneta americana and Its Microbiota: Digestion, Detoxification and Oxidative Stress Response.* PloS one, 2016. **11**(5): p. e0155254-e0155254.

58. Li, L., et al., *Applications of RNA Interference in American Cockroach.* J Vis Exp, 2021(178).

59. French, A.S., et al., *Transcriptome analysis and RNA interference of cockroach phototransduction indicate three opsins and suggest a major role for TRPL channels.* Front Physiol, 2015. **6**: p. 207.

60. Hennenfent, A., et al., *RNA interference supports a role for Nanchung–Inactive in mechanotransduction by the cockroach, Periplaneta americana, tactile spine.* Invertebrate Neuroscience, 2020. **20**(1): p. 1.

61. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. **391**(6669): p. 806-11.

62. Vogel, E., et al., *RNA Interference in Insects: Protecting Beneficials and Controlling Pests.* Front Physiol, 2018. **9**: p. 1912.

63. Burand, J.P. and W.B. Hunter, *RNAi: Future in insect management.* Journal of Invertebrate Pathology, 2013. **112**: p. S68-S74.

64. Ciudad, L., X. Bellés, and M.D. Piulachs, *Structural and RNAi characterization of the German cockroach lipophorin receptor, and the evolutionary relationships of lipoprotein receptors.* BMC Mol Biol, 2007. **8**: p. 53.

65. Wang, K., et al., *Variation in RNAi efficacy among insect species is attributable to dsRNA degradation in vivo.* Insect Biochem Mol Biol, 2016. **77**: p. 1-9.

66. Lin, Y.H., et al., *Oral delivery of dsRNA lipoplexes to German cockroach protects dsRNA from degradation and induces RNAi response.* Pest Manag Sci, 2017. **73**(5): p. 960-966.

67. Huang, J.H., et al., *Practical Use of RNA Interference: Oral Delivery of Double-stranded RNA in Liposome Carriers for Cockroaches.* J Vis Exp, 2018(135).

68. Horvath, T.D., et al., *Interrogation of the mammalian gut-brain axis using LC-MS/MS-based targeted metabolomics with in vitro bacterial and organoid cultures and in vivo gnotobiotic mouse models.* Nat Protoc, 2023. **18**(2): p. 490-529.

69. Bayer, F., et al., *Antibiotic Treatment Protocols and Germ-Free Mouse Models in Vascular Research.* Front Immunol, 2019. **10**: p. 2174.

70. Dremova, O., et al., *Sterility testing of germ-free mouse colonies.* Front Immunol, 2023. **14**: p. 1275109.

71. Nicklas, W., L. Keubler, and A. Bleich, *Maintaining and Monitoring the Defined Microbiota Status of Gnotobiotic Rodents.* Ilar j, 2015. **56**(2): p. 241-9.

72. Dukes, H.E., J.E. Dyer, and E.A. Ottesen, *Establishment and Maintenance of Gnotobiotic American Cockroaches (Periplaneta americana).* JoVE, 2021(171): p. e61316.

73. Jahnes, B.C., et al., *Microbial colonization promotes model cockroach gut tissue growth and development.* J Insect Physiol, 2021. **133**: p. 104274.

74. Tegtmeier, D., et al., *Oxygen Affects Gut Bacterial Colonization and Metabolic Activities in a Gnotobiotic Cockroach Model.* Appl Environ Microbiol, 2016. **82**(4): p. 1080-9.

75. Mikaelyan, A., et al., *Deterministic Assembly of Complex Bacterial Communities in Guts of Germ-Free Cockroaches.* Appl Environ Microbiol, 2016. **82**(4): p. 1256-63.

76. Wada-Katsumata, A., et al., *Gut bacteria mediate aggregation in the German cockroach.* Proc Natl Acad Sci U S A, 2015. **112**(51): p. 15678-83.

77. Jones, S.A. and D. Raubenheimer, *Nutritional regulation in nymphs of the German cockroach, Blattella germanica.* J Insect Physiol, 2001. **47**(10): p. 1169-1180.

78. Mullins, D.E. and D.G. Cochran, *Nitrogen metabolism in the American cockroach: an examination of whole body and fat body regulation of cations in response to nitrogen balance.* J Exp Biol, 1974. **61**(3): p. 557-70.

79. Dockman, R.L. and E.A. Ottesen, *Purified fibers in chemically defined synthetic diets destabilize the gut microbiome of an omnivorous insect model.* Frontiers in Microbiomes, 2024. **Volume 3 - 2024**.

80. Piper, M.D., et al., *A holidic medium for Drosophila melanogaster.* Nat Methods, 2014. **11**(1): p. 100-5.

81. Martelli, F., et al., *A defined diet for pre-adult Drosophila melanogaster.* Scientific Reports, 2024. **14**(1): p. 6974.

82. Ma, H., et al., *Morphology and three-dimensional reconstruction of the digestive system of Periplaneta americana.* J Med Entomol, 2009. **46**(1): p. 165-8.

83. Gontang, E.A., et al., *Major changes in microbial diversity and community composition across gut sections of a juvenile Panchlora cockroach.* PLoS One, 2017. **12**(5): p. e0177189.

84. Tamaki, F.K., et al., *Physiology of digestion and the molecular characterization of the major digestive enzymes from Periplaneta americana.* J Insect Physiol, 2014. **70**: p. 22-35.

85. Genta, F.A., W.R. Terra, and C. Ferreira, *Action pattern, specificity, lytic activities, and physiological role of five digestive beta-glucanases isolated from Periplaneta americana.* Insect Biochem Mol Biol, 2003. **33**(11): p. 1085-97.

86. Eisner, T., *The digestion and absorption of fats in the foregut of the cockroach, Periplaneta americana.* J. Exp. Zool, 1955(130): p. 159-181.

87. Hegedus, D., et al., *New insights into peritrophic matrix synthesis, architecture, and function.* Annu Rev Entomol, 2009. **54**: p. 285-302.

88. Arrese, E.L. and J.L. Soulages, *Insect fat body: energy, metabolism, and regulation.* Annu Rev Entomol, 2010. **55**: p. 207-25.

89. Wall, B.J., J.L. Oschman, and B.A. Schmidt, *Morphology and function of Malpighian tubules and associated structures in the cockroach, Periplaneta americana.* J Morphol, 1975. **146**(2): p. 265-306.

90. Lo, N., et al., *Cockroaches that lack Blattabacterium endosymbionts: the phylogenetically divergent genus Nocticola.* Biology letters, 2007. **3**(3): p. 327-330.

91. Park, M.S., P. Park, and M. Takeda, *Roles of fat body trophocytes, mycetocytes and urocytes in the American cockroach, Periplaneta americana under starvation conditions: an ultrastructural study.* Arthropod Struct Dev, 2013. **42**(4): p. 287-95.

92. Mullins, D.E. and D.G. Cochran, *Nitrogen excretion in cockroaches: uric Acid is not a major product.* Science, 1972. **177**(4050): p. 699-701.

93. Mullins, D.E. and D.G. Cochran, *Nitrogen metabolism in the american cockroach—I. An examination of positive nitrogen balance with respect to uric acid stores.* Comparative Biochemistry and Physiology Part A: Physiology, 1975. **50**(3): p. 489-500.

94. Cochran, D.G., D.E. Mullins, and K.J. Mullins, *Cytological Changes in the Fat Body of the American Cockroach, Periplaneta americana, in Relation to Dietary Nitrogen Levels.* Annals of the Entomological Society of America, 1979. **72**(2): p. 197-205.

95. COCHRAN, D.G., *URIC ACID ACCUMULATION IN YOUNG AMERICAN COCKROACH NYMPHS.* 1979. **25**(2): p. 153-157.

96. Latorre, A., et al., *Of Cockroaches and Symbionts: Recent Advances in the Characterization of the Relationship between Blattella germanica and Its Dual Symbiotic System.* Life (Basel), 2022. **12**(2).

97. Vicente, C.S.L., et al., *Genome analysis of new Blattabacterium spp., obligatory endosymbionts of Periplaneta fuliginosa and P. japonica.* PLoS One, 2018. **13**(7): p. e0200512.

98. López-Sánchez, M.J., et al., *Evolutionary convergence and nitrogen metabolism in Blattabacterium strain Bge, primary endosymbiont of the cockroach Blattella germanica.* PLoS Genet, 2009. **5**(11): p. e1000721.

99. Sabree, Z.L., S. Kambhampati, and N.A. Moran, *Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont.* Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19521-6.

100. Patiño-Navarrete, R., et al., *The cockroach Blattella germanica obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont.* Biol Lett, 2014. **10**(7).

101. Ayayee, P.A., T. Larsen, and Z. Sabree, *Symbiotic essential amino acids provisioning in the American cockroach, Periplaneta americana (Linnaeus) under various dietary conditions.* PeerJ, 2016. **4**: p. e2046.

102. Park, M.S. and M. Takeda, *Cloning of PaAtg8 and roles of autophagy in adaptation to starvation with respect to the fat body and midgut of the Americana cockroach, Periplaneta americana.* Cell Tissue Res, 2014. **356**(2): p. 405-16.

103. Haydak, M.H., *Influence of the Protein Level of the Diet on the Longevity of Cockroaches.* Annals of The Entomological Society of America, 1953. **46**: p. 547-560.

104. Dukes, H.E., K.A. Tinker, and E.A. Ottesen, *Disentangling hindgut metabolism in the American cockroach through single-cell genomics and metatranscriptomics.* Front Microbiol, 2023. **14**: p. 1156809.

105. Jahnes, B.C. and Z.L. Sabree, *Nutritional symbiosis and ecology of host-gut microbe systems in the Blattodea.* Current Opinion in Insect Science, 2020. **39**: p. 35-41.

106. Bignell, D.E., *An ultrastructural study and stereological analysis of the colon wall in the cockroach Periplaneta americana.* Tissue and Cell, 1980. **12**(1): p. 153-164.

107. Zurek, L. and B.A. Keddie, *Contribution of the colon and colonie bacterial flora to metabolism and development of the American cockroach Periplaneta americana (L).* Journal of Insect Physiology, 1996. **42**(8): p. 743-748.

108. Bracke, J.W. and A.J. Markovetz, *Transport of bacterial end products from the colon of Periplaneta americana.* Journal of Insect Physiology, 1980. **26**(2): p. 85-89.

109. Hogan, M.E., M. Slaytor, and R.W. O'Brien, *Transport of volatile fatty acids across the hindgut of the cockroach Panesthia cribrata Saussure and the termite, Mastotermes darwiniensis Froggatt.* Journal of Insect Physiology, 1985. **31**(7): p. 587-591.

110. Glover, J.S., T.D. Ticer, and M.A. Engevik, *Characterizing the mucin-degrading capacity of the human gut microbiota.* Scientific Reports, 2022. **12**(1): p. 8456.

111. Konjar, Š., M. Pavšič, and M. Veldhoen, *Regulation of Oxygen Homeostasis at the Intestinal Epithelial Barrier Site.* Int J Mol Sci, 2021. **22**(17).

112. Espey, M.G., *Role of oxygen gradients in shaping redox relationships between the human intestine and its microbiota.* Free Radic Biol Med, 2013. **55**: p. 130-40.

113. Ravcheev, D.A. and I. Thiele, *Systematic genomic analysis reveals the complementary aerobic and anaerobic respiration capacities of the human gut microbiota.* Front Microbiol, 2014. **5**: p. 674.

114. Briliūtė, J., et al., *Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci.* Nat Microbiol, 2019. **4**(9): p. 1571-1581.

115. Martens, E.C., H.C. Chiang, and J.I. Gordon, *Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont.* Cell Host Microbe, 2008. **4**(5): p. 447-57.

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

117. Morris, B.E., et al., *Microbial syntrophy: interaction for the common good.* FEMS Microbiol Rev, 2013. **37**(3): p. 384-406.

118. Ruaud, A., et al., *Syntrophy via Interspecies H₂ Transfer between Christensenella and Methanobrevibacter Underlies Their Global Cooccurrence in the Human Gut.* mBio, 2020. **11**(1): p. 10.1128/mbio.03235-19.

119. Catlett, J.L., et al., *Metabolic Synergy between Human Symbionts Bacteroides and Methanobrevibacter.* Microbiol Spectr, 2022. **10**(3): p. e0106722.

120. Chiang, B.H., et al., *Bacterial interactions on nutrient-rich surfaces in the gut lumen.* Infect Immun, 2024. **92**(9): p. e0048023.

121. Fischbach, M.A. and J.L. Sonnenburg, *Eating for two: how metabolism establishes interspecies interactions in the gut.* Cell Host Microbe, 2011. **10**(4): p. 336-47.

122. Tinker, K.A. and E.A. Ottesen, *Differences in Gut Microbiome Composition Between Sympatric Wild and Allopatric Laboratory Populations of Omnivorous Cockroaches.* Front Microbiol, 2021. **12**: p. 703785.

123. Dietrich, C., T. Köhler, and A. Brune, *The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events.* Appl Environ Microbiol, 2014. **80**(7): p. 2261-9.

124. Wang, T., et al., *Evidence for a multi-level trophic organization of the human gut microbiome.* PLOS Computational Biology, 2019. **15**(12): p. e1007524.

125. DePoy, A.N., et al., *Microbial transcriptional responses to host diet maintain gut microbiome homeostasis in the American cockroach.* bioRxiv, 2024: p. 2024.10.31.621369.

126. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora.* Science, 2005. **308**(5728): p. 1635-8.

127. Li, T.T., et al., *Microbiota metabolism of intestinal amino acids impacts host nutrient homeostasis and physiology.* Cell Host Microbe, 2024. **32**(5): p. 661-675.e10.

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

129. Culp, E.J. and A.L. Goodman, *Cross-feeding in the gut microbiome: Ecology and mechanisms.* Cell Host Microbe, 2023. **31**(4): p. 485-499.

130. Reichardt, N., et al., *Phylogenetic distribution of three pathways for propionate production within the human gut microbiota.* The ISME Journal, 2014. **8**(6): p. 1323-1335.

131. Zeybek, N., R.A. Rastall, and A.O. Buyukkileci, *Utilization of xylan-type polysaccharides in co-culture fermentations of Bifidobacterium and Bacteroides species.* Carbohydrate Polymers, 2020. **236**: p. 116076.

132. Hirmas, B., et al., *Metabolic Modeling and Bidirectional Culturing of Two Gut Microbes Reveal Cross-Feeding Interactions and Protective Effects on Intestinal Cells.* mSystems, 2022. **7**(5): p. e0064622.

133. Cerisy, T., et al., *ABC Transporters Required for Hexose Uptake by Clostridium phytofermentans.* J Bacteriol, 2019. **201**(15).

134. Levy, S., et al., *Convergent evolution of oxidized sugar metabolism in commensal and pathogenic microbes in the inflamed gut.* Nat Commun, 2025. **16**(1): p. 1121.

135. Drousiotis, K., et al., *Characterization of the l-arabinofuranose-specific GafABCD ABC transporter essential for l-arabinose-dependent growth of the lignocellulose-degrading bacterium Shewanella sp. ANA-3.* Microbiology (Reading), 2023. **169**(3).

136. Ejby, M., et al., *An ATP Binding Cassette Transporter Mediates the Uptake of α-(1,6)-Linked Dietary Oligosaccharides in Bifidobacterium and Correlates with Competitive Growth on These Substrates*.* Journal of Biological Chemistry, 2016. **291**(38): p. 20220-20231.

137. Soto-Martin, E.C., et al., *Vitamin Biosynthesis by Human Gut Butyrate-Producing Bacteria and Cross-Feeding in Synthetic Microbial Communities.* mBio, 2020. **11**(4).

138. Smith, E.A. and G.T. Macfarlane, *Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids.* FEMS Microbiology Ecology, 1998. **25**(4): p. 355-368.

139. Van den Abbeele, P., et al., *The Effect of Amino Acids on Production of SCFA and bCFA by Members of the Porcine Colonic Microbiota.* Microorganisms, 2022. **10**(4).

140. Pavao, A., et al., *Reconsidering the in vivo functions of Clostridial Stickland amino acid fermentations.* Anaerobe, 2022. **76**: p. 102600.

141. Rangarajan, A.A., et al., *Ruminococcus bromii enables the growth of proximal Bacteroides thetaiotaomicron by releasing glucose during starch degradation.* Microbiology (Reading), 2022. **168**(4).

142. Ze, X., et al., *Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon.* The ISME Journal, 2012. **6**(8): p. 1535-1543.

143. Walker, A.W., et al., *Dominant and diet-responsive groups of bacteria within the human colonic microbiota.* Isme j, 2011. **5**(2): p. 220-30.

144. Canani, R.B., et al., *Potential beneficial effects of butyrate in intestinal and extraintestinal diseases.* World J Gastroenterol, 2011. **17**(12): p. 1519-28.

145. Liu, B., et al., *Competition Between Butyrate Fermenters and Chain-Elongating Bacteria Limits the Efficiency of Medium-Chain Carboxylate Production.* Frontiers in Microbiology, 2020. **11**.

146. Leth, M.L., et al., *Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut.* Nature Microbiology, 2018. **3**(5): p. 570-580.

147. Nie, K., et al., *Roseburia intestinalis: A Beneficial Gut Organism From the Discoveries in Genus and Species.* Frontiers in Cellular and Infection Microbiology, 2021. **11**.

148. Rey, F.E., et al., *Metabolic niche of a prominent sulfate-reducing human gut bacterium.* Proc Natl Acad Sci U S A, 2013. **110**(33): p. 13582-7.

149. Schauer, C., C. Thompson, and A. Brune, *Pyrotag sequencing of the gut microbiota of the cockroach Shelfordella lateralis reveals a highly dynamic core but only limited effects of diet on community structure.* PLoS One, 2014. **9**(1): p. e85861.

150. Lampert, N., A. Mikaelyan, and A. Brune, *Diet is not the primary driver of bacterial community structure in the gut of litter-feeding cockroaches.* BMC Microbiol, 2019. **19**(1): p. 238.

151. Pérez-Cobas, A.E., et al., *Diet shapes the gut microbiota of the omnivorous cockroach Blattella germanica.* FEMS Microbiol Ecol, 2015. **91**(4).

152. Bertino-Grimaldi, D., et al., *Bacterial community composition shifts in the gut of Periplaneta americana fed on different lignocellulosic materials.* Springerplus, 2013. **2**: p. 609.

153. Zhu, J., et al., *Diet Influences the Gut Microbial Diversity and Olfactory Preference of the German Cockroach Blattella germanica.* Curr Microbiol, 2022. **80**(1): p. 23.

154. Tanes, C., et al., *Role of dietary fiber in the recovery of the human gut microbiome and its metabolome.* Cell Host Microbe, 2021. **29**(3): p. 394-407.e5.

155. Flint, H.J., et al., *Microbial degradation of complex carbohydrates in the gut.* Gut Microbes, 2012. **3**(4): p. 289-306.

156. Jefferson, A. and K. Adolphus, *The Effects of Intact Cereal Grain Fibers, Including Wheat Bran on the Gut Microbiota Composition of Healthy Adults: A Systematic Review.* Frontiers in Nutrition, 2019. **6**.

157. Abreu y Abreu, A.T., et al., *Dietary fiber and the microbiota: A narrative review by a group of experts from the Asociación Mexicana de Gastroenterología.* Revista de Gastroenterología de México (English Edition), 2021. **86**(3): p. 287-304.

158. Hamaker, B.R. and Y.E. Tuncil, *A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota.* J Mol Biol, 2014. **426**(23): p. 3838-50.

159. Guerriero, G., J.F. Hausman, and G. Cai, *No stress! Relax! Mechanisms governing growth and shape in plant cells.* Int J Mol Sci, 2014. **15**(3): p. 5094-114.

160. Puhlmann, M.-L. and W.M. de Vos, *Intrinsic dietary fibers and the gut microbiome: Rediscovering the benefits of the plant cell matrix for human health.* Frontiers in Immunology, 2022. **13**.

161. Rogowski, A., et al., *Glycan complexity dictates microbial resource allocation in the large intestine.* Nature Communications, 2015. **6**(1): p. 7481.

162. Chung, W.S.F., et al., *Impact of carbohydrate substrate complexity on the diversity of the human colonic microbiota.* FEMS Microbiology Ecology, 2019. **95**(1): p. fiy201.

163. Augustin, L.S.A., et al., *Dietary fibre consensus from the International Carbohydrate Quality Consortium (ICQC).* Nutrients, 2020. **12**(9).

164. Tuncil, Y.E., et al., *Subtle Variations in Dietary-Fiber Fine Structure Differentially Influence the Composition and Metabolic Function of Gut Microbiota.* mSphere, 2020. **5**(3).

165. Kane, M.D. and J.A. Breznak, *Effect of host diet on production of organic acids and methane by cockroach gut bacteria.* Applied and Environmental Microbiology, 1991. **57**: p. 2628 - 2634.

166. Schauer, C., C.L. Thompson, and A. Brune, *The bacterial community in the gut of the cockroach Shelfordella lateralis reflects the close evolutionary relatedness of cockroaches and termites.* Appl Environ Microbiol, 2012. **78**(8): p. 2758-67.

167. Tokuda, G., et al., *Fiber-associated spirochetes are major agents of hemicellulose degradation in the hindgut of wood-feeding higher termites.* Proceedings of the National Academy of Sciences, 2018. **115**(51): p. E11996-E12004.

168. Breznak, J.A. and A. Brune, *Role of Microorganisms in the Digestion of Lignocellulose by Termites.* Annual Review of Entomology, 1994. **39**(Volume 39, 1994): p. 453-487.

169. Scrivener, A.M., L. Zhao, and M. Slaytor, *Biochemical and Immunological Relationships Between Endo-β-1,4-Glucanases from Cockroaches.* Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 1997. **118**(4): p. 837-843.

170. Wharton, D.R., M.L. Wharton, and J.E. Lola, *Cellulase in the cockroach, with special reference to Periplaneta americana (L).* J Insect Physiol, 1965. **11**(7): p. 947-59.

171. Watanabe, H., et al., *A cellulase gene of termite origin.* Nature, 1998. **394**(6691): p. 330-331.

172. Gijzen, H.J., et al., *Effect of Host Diet and Hindgut Microbial Composition on Cellulolytic Activity in the Hindgut of the American Cockroach, Periplaneta americana.* Appl Environ Microbiol, 1994. **60**(6): p. 1822-6.

173. Gile, G.H. and C.H. Slamovits, *Phylogenetic position of Lophomonas striata Bütschli (Parabasalia) from the hindgut of the cockroach Periplaneta americana.* Protist, 2012. **163**(2): p. 274-83.

174. Mahowald, M.A., et al., *Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla.* Proc Natl Acad Sci U S A, 2009. **106**(14): p. 5859-64.

175. Koropatkin, N.M., E.A. Cameron, and E.C. Martens, *How glycan metabolism shapes the human gut microbiota.* Nature Reviews Microbiology, 2012. **10**(5): p. 323-335.

176. Déjean, G., et al., *Adaptation of Syntenic Xyloglucan Utilization Loci of Human Gut Bacteroidetes to Polysaccharide Side Chain Diversity.* Appl Environ Microbiol, 2019. **85**(20).

177. Pereira, G.V., et al., *Degradation of complex arabinoxylans by human colonic Bacteroidetes.* Nat Commun, 2021. **12**(1): p. 459.

178. Kmezik, C., et al., *A polysaccharide utilization locus from the gut bacterium Dysgonomonas mossii encodes functionally distinct carbohydrate esterases.* J Biol Chem, 2021. **296**: p. 100500.

179. Luis, A.S., et al., *Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides.* Nat Microbiol, 2018. **3**(2): p. 210-219.

180. Pramono, A.K., et al., *Dysgonomonas termitidis sp. nov., isolated from the gut of the subterranean termite Reticulitermes speratus.* Int J Syst Evol Microbiol, 2015. **65**(Pt 2): p. 681-685.

181. Qu, Z., et al., *Selective utilization of medicinal polysaccharides by human gut Bacteroides and Parabacteroides species.* Nature Communications, 2025. **16**(1): p. 638.

182. Grondin Julie, M., et al., *Polysaccharide Utilization Loci: Fueling Microbial Communities.* Journal of Bacteriology, 2017. **199**(15): p. 10.1128/jb.00860-16.

183. Feng, J., et al., *Polysaccharide utilization loci in Bacteroides determine population fitness and community-level interactions.* Cell Host Microbe, 2022. **30**(2): p. 200-215.e12.

184. Hershko Rimon, A., et al., *Novel clostridial cell-surface hemicellulose-binding CBM3 proteins.* Acta Crystallogr F Struct Biol Commun, 2021. **77**(Pt 4): p. 95-104.

185. Israeli-Ruimy, V., et al., *Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions.* Scientific Reports, 2017. **7**(1): p. 42355.

186. Rincon, M.T., et al., *Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, Ruminococcus flavefaciens FD-1.* PLoS One, 2010. **5**(8): p. e12476.

187. Liu, N., et al., *In vitro and in vivo exploration of the cellobiose and cellodextrin phosphorylases panel in Ruminiclostridium cellulolyticum: implication for cellulose catabolism.* Biotechnology for Biofuels, 2019. **12**(1): p. 208.

188. Galera-Prat, A., et al., *The cohesin module is a major determinant of cellulosome mechanical stability.* J Biol Chem, 2018. **293**(19): p. 7139-7147.

189. Bule, P., et al., *Higher order scaffoldin assembly in Ruminococcus flavefaciens cellulosome is coordinated by a discrete cohesin-dockerin interaction.* Sci Rep, 2018. **8**(1): p. 6987.

190. Bule, P., et al., *Assembly of Ruminococcus flavefaciens cellulosome revealed by structures of two cohesin-dockerin complexes.* Sci Rep, 2017. **7**(1): p. 759.

191. Moraïs, S., et al., *Cryptic diversity of cellulose-degrading gut bacteria in industrialized humans.* Science, 2024. **383**(6688): p. eadj9223.

192. Vera-Ponce de Leon, A., et al., *Genetic drift and host-adaptive features likely underlie cladogenesis of insect-associated Lachnospiraceae.* Genome Biol Evol, 2022. **14**(6).

193. Biddle, A., et al., *Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities.* Diversity, 2013. **5**(3): p. 627-640.

194. Raut, M.P., et al., *Deciphering the unique cellulose degradation mechanism of the ruminal bacterium Fibrobacter succinogenes S85.* Scientific Reports, 2019. **9**(1): p. 16542.

195. Neumann, A.P., C.A. McCormick, and G. Suen, *Fibrobacter communities in the gastrointestinal tracts of diverse hindgut-fermenting herbivores are distinct from those of the rumen.* Environ Microbiol, 2017. **19**(9): p. 3768-3783.

196. Roger, V., et al., *Effects of Physicochemical Factors on the Adhesion to Cellulose Avicel of the Ruminal Bacteria Ruminococcus flavefaciens and Fibrobacter succinogenes subsp. succinogenes.* Appl Environ Microbiol, 1990. **56**(10): p. 3081-7.

197. Suen, G., et al., *The complete genome sequence of Fibrobacter succinogenes S85 reveals a cellulolytic and metabolic specialist.* PLoS One, 2011. **6**(4): p. e18814.

198. Jun, H.S., et al., *Outer membrane proteins of Fibrobacter succinogenes with potential roles in adhesion to cellulose and in cellulose digestion.* J Bacteriol, 2007. **189**(19): p. 6806-15.

199. Burnet, M.C., et al., *Evaluating Models of Cellulose Degradation by Fibrobacter succinogenes S85.* PLoS One, 2015. **10**(12): p. e0143809.

200. Froidurot, A., et al., *Fibrobacter sp. HC4, a newly isolated strain, demonstrates a high cellulolytic activity as revealed by enzymatic measurements and in vitro assay.* Applied and Environmental Microbiology, 2024. **90**(8): p. e00514-24.

CHAPTER 2

GENOME REPORT: IMPROVED CHROMOSOME-LEVEL ASSEMBLY OF THE AMERICAN COCKROACH, *PERIPLANETA AMERICANA*[1]

[1] Dockman, R.L., Simmonds, T.J., Vogel, K.J., Geib, S.M., and E.A. Ottesen. Submitted to *G3: Genes, Genomes, Genetics,* 03/14/25

Abstract

The American cockroach, *Periplaneta americana*, is a cosmopolitan insect notorious for thriving among humans undeterred by attempts to eliminate it. The traits that contribute to its ubiquity as an opportunistic pest, such as long lifespan, expansive neurosensory capacity, and nutritional flexibility, also make *P. americana* an excellent invertebrate model organism with a long history in neuroscience and physiological research. Current genetic resources available for *P. americana* highlight its large, complex genome and richly diverse transcriptional capabilities, but fall short of producing a complete, chromosome-level genome. Here, we present a high-quality de novo genome assembly of a laboratory-raised adult female *P. americana* using a combination of high fidelity PacBio long reads and Hi-C sequencing. The final 3.23 Gb genome was assembled with chromosomal resolution into 17 scaffolds, consistent with previous karyotype analysis, and has a scaffold N50 of 188.1 Mb and genome BUSCO score of 99.7%. This assembly includes a chromosome that was missing from the previous reference genome for this species. Protein prediction and annotation were performed via the NCBI Eukaryotic Genome Annotation Pipeline, which identified 16,780 protein-coding genes and generated an annotation BUSCO score of 97.8%. Ortholog comparisons with available Blattodea assemblies highlight the expanded chemosensory and immune capabilities of *P. americana* compared to termite relatives. This genome assembly is a valuable tool for facilitating future research on the biology and evolution of this remarkable insect.

Introduction

The American cockroach (*Periplaneta americana*) is a notorious pest found across the world, living and thriving alongside humans in widely variable environmental conditions. Despite their unsavory reputation, cockroaches are of considerable interest to researchers across disciplines. Cockroaches combine the large size, complex physiology, and tractable nature of mice with the simpler and cost-effective care requirements characteristic of model insects.

There is a rich body of work compiled throughout the last century describing basic cockroach biology, covering topics ranging from external morphology to internal physiology [1-4]. Early research leveraging *P. americana* as a model system has enhanced understanding of nervous system function, connectivity, and regeneration [5-9] and discovered unique traits that allow *P. americana* to survive extreme environmental stressors, such as endosymbiont-mediated nitrogen cycling [10, 11]. Studies on the cockroach immune system have linked its expansive repertoire of immune-associated proteins to common allergens [12, 13] and mechanisms of pesticide resistance [14-16]. The extensive antimicrobial and regenerative capabilities of *P. americana* have also earned it respect in both ancient and modern Chinese tradition as an important medicinal insect [17, 18]. Modern sequencing technologies and genetic techniques have further supported the use of *P. americana* as a model organism. Its susceptibility and robust response to RNA inhibition (RNAi) via multiple administration methods, *P. americana* is an especially useful organism for deciphering gene involvement in insect physiological development and pesticide resistance, among other investigations [19-21]. *P. americana* also shows potential as an emerging insect model for host-microbiome interactions. The gut microbiome of omnivorous cockroaches reflects that of humans and omnivorous mammalian model systems [22-25], and germ-free nymphs can be easily produced via ootheca sterilization, a

highly desirable trait for defined-community research interests [26-29]. Altogether, these traits support the research potential of this insect and argue towards the necessity of a complete, high-quality American cockroach genome. While there are two previous genome assemblies publicly available, the assembly presented by [30] lacks publicly available protein annotations and is limited by the short-read technology available at the time and the assembly prepared by [13], while a significant improvement, contains more scaffolds than supported by karyotype analysis (male/XO: 33 diploid; female/XX: 34 diploid) [31].

Here, we present the first chromosome-level assembled genome of *P. americana*. We used long-read high fidelity (HiFi) PacBio sequencing in addition to chromatin contact mapping (Hi-C) to produce a genome scaffolded into a 17-chromosome assembly, consistent with previous karyotype findings [31]. This high-quality genome assembly is an important tool for facilitating future genetic research in the American cockroach.

Materials and Methods

2.3.1. Insect origin and selection

An adult female *Periplaneta americana* individual was selected from a stock colony maintained at the University of Georgia by the Ottesen laboratory; this colony has been maintained the laboratory for 10 years and originated in another long-term laboratory colony of unknown origin. The specimen was flash frozen in liquid nitrogen and shipped on dry ice to the United States Department of Agriculture – Agricultural Research Service (USDA-ARS) – Pacific Basin Agricultural Research Center (PBARC) in Hilo, Hawaii.

2.3.2. Sample preparation and sequencing methods

For PacBio sequencing, high molecular weight (HMW) DNA extraction was performed from insect leg tissue using the Qiagen MagAttract HMW DNA Kit (Qiagen, Hilden Germany).

DNA was sheared with the Diagenode Megaruptor 2 (Denville, New Jersey, USA) then prepared for PacBio sequencing using the SMRTBell Express Template kit 2.0 (Pacific Biosciences, Menlo Park, California, USA). The library was size-selected prior to sequencing on a Sequel II System (Pacific Biosciences, Menlo Park, California, USA) using Binding Kit v2.0, Sequencing kit v2.0, and SMRT Cell 8M. To target HiFi reads, the library was sequenced using a 30-hour movie time on three SMRTcells. Raw subreads were converted to HiFi data by processing with CCS to call a single high quality consensus sequence for each molecule, using a 99.5% consensus accuracy cutoff.

For Hi-C sequencing, the Arima Hi-C kit (Arima Genomics, San Diego, California, USA) was used to crosslink leg tissue DNA and perform proximity ligation, following the Arima Hi-C low input protocol. After proximity ligation, DNA was sheared with a Diagenode Bioruptor then size-selected for 200-600bp DNA fragments. The Swift Accel NGS 2S Plus kit (Integrated DNA Technologies, Coralville, Iowa, USA) was used to prepare an Illumina library from the size-selected DNA, which was then sequenced with an Illumina NovaSeq 6000 (Illumina, San Diego, California, USA).

2.3.3. RNA-seq

Transcriptomic data was obtained for the midgut, hindgut, and fat body of 10 individual cockroaches for a dietary experiment investigating the impact of carbohydrate source on the microbial metatranscriptome and host transcriptome (BioProject: PRJNA1105088). Data were obtained as 150bp paired end reads on Illumina NovaSeq from Novogene Corporation in Sacramento, California. The Joint Genome Institute (JGI) programs Bbduk and BBSplit (jgi.doe.gov/data-and-tools/software-tools/bbtools/) and were used to remove sequencing adapters and screen for initial *Blattabacterium* contamination, and SortMeRNA was used to

remove ribosomal RNA reads [32]. Cleaned RNA reads were aligned to the unfiltered genome for contaminant filtering.

2.3.4. Genome assembly and polishing

The first genome assembly was generated using hifiasm v0.19.6-r595 Hi-C integration to obtain primary, alternate, and haplotype-phased assemblies [33]. Three flow cells of PacBio HiFi data were obtained from long-read sequencing and concatenated into a single fastq file for assembly and used with Hi-C data obtained from the same source insect. The primary assembly was selected for polishing and contigs were filtered to retain those with coverage reported from hifiasm as between 6X and 30X.

Contamination filtering was performed as described by Lu and Salzberg [34]. Contigs were separated into individual files, then fragmented with SeqKit v2.5.1 into 100 bp pseudo-reads with 50 bp overlaps [35]. These pseudo-reads were fed through Kraken v2.1.3 with default parameters to align to the default eukaryotic and prokaryotic databases, and contigs identified as *Blattabacterium* or with 70% identity assigned to *Homo sapiens* were discarded [36]. Remaining contigs were masked using RepeatMasker v4.1.5 (www.repeatmasker.org/RepeatMasker/) with Dfam library v3.7 prior to RNA-seq alignment with HISAT2, and contigs with average read depths exceeding 10,000 were removed [37, 38].

The Arima Genomics mapping pipeline (github.com/ArimaGenomics/mapping_pipeline) was used to map Hi-C data to the filtered assembly. The pipeline described utilizes the programs BWA-MEM for separate alignment of the paired Illumina Hi-C reads, the Picard (broadinstitute.github.io/picard/) “MarkDuplicates” command for PCR duplicate removal, and SAMtools for file sorting and handling [39, 40]. The resulting Hi-C alignment files and the hifiasm assembly were fed into Yet Another Hi-C Scaffolding tool (YaHS) without breaking

contigs, and the resulting contact map was visualized with Juicebox for manual curation [41, 42]. The corrected genome was exported and screened for possible telomeres using tidk (github.com/□oolkit/telomeric-identifier) and repeat regions identified *de novo* (AACCTAACCT) were graphed [43]. An additional round of polishing was performed in Juicebox to correct repeat-heavy telomeric loci, and completeness of the final scaffolded genome was assessed with the Insecta set (version odb10) of Benchmarking Universal Single-Copy Orthologs (BUSCOs) [44].

2.3.5. Mitochondrion identification

The program MitoHiFi v2 was used to identify a consensus mitochondrial genome sequence from the original hifiasm assembly [45]. The NCBI reference sequence NC_016956.1 was selected as the reference *P. americana* mitochondrion for identification [46].

2.3.6. Repeat modeling and masking

For in-depth repeat identification, RepeatModeler v2.0.4 was used to create an *ab initio* repeat library specific to *P. americana,* which was then separated into libraries of known and unknown repeat families [47]. Using the script repclassifier.sh (github.com/darencard/GenomeAnnotation/blob/master/repclassifier), unknown repeats were iteratively re-annotated for seven rounds, when the percent of repeats classified as "known" rather than "unknown" plateaued. Four rounds of repeat masking using RepeatMasker v4.1.5 were then performed as described in (darencard.net/blog/2022-07-09-genome-repeat-annotation/), during which simple repeats were identified and masked first, followed by insect-specific Dfam repeats, known *P. americana* repeats, and lastly unknown *P. americana* repeats. The output from the four RepeatMasker rounds were combined to generate an overall masked

genome, annotation files, and table describing the repeats, and repeat landscapes were summarized with the script parseRM.pl (github.com/4ureliek/Parsing-RepeatMasker-Outputs).

2.3.7. Genome annotation

The finished assembly was submitted to NCBI for structural and functional gene annotation via the automated Eukaryotic Genome Annotation Pipeline v10.3. Evidence fed into the GNOMON gene prediction tool included existing RNA-seq data and transcriptome assemblies for *P. americana*, NCBI RefSeq protein sets from *Acromyrmex echinatior*, *Hyalella azteca, Acyrthosiphon pisum, Caenorhabditis elegans, Tribolium castaneum, Drosophila melanogaster*, and *Apis mellifera*, as well as Insecta and *P. americana* GenBank protein sets. Details of the annotation release (GCF_040183065.1-RS_2024_10) are available at www.ncbi.nlm.nih.gov/refseq/annotation_euk/Periplaneta_americana/GCF_040183065.1-RS_2024_10/.

2.3.8. Blattodea comparison

The annotations obtained for this assembly were compared with Blattodea annotations available on NCBI for a previous *P. americana* genome (GCA_025594305.2) [13], *Zootermopsis nevadensis* (GCF_000696155.1) [48], *Blattella germanica* (GCA_003018175.1) and *Cryptotermes secundus* (GCF_002891405.2) [49], *Coptotermes formosanus* (GCA_013340265.1), and *Diploptera punctata* (GCA_030220185.1) [50]. OrthoFinder v2.5.5 was used with GENESPACE v1.3.1 in R Studio to identify orthologous groups and generate riparian plots [51-53]. The distribution of orthologous gene sets between and across genomes were visualized with the R package UpSetR v1.4.0 [54]. GO terms assigned to *P. americana* genes that were sorted into orthologous gene sets were used for GO enrichment analysis in R with the package clusterProfiler v4.6.2 [55, 56].

Results and Discussion

2.4.1. Assembly

American cockroaches have XX/X0 sex chromosome systems [57], so a single female insect was selected for PacBio SMRTBell sequencing, producing over 6.6 million reads covering 90.6 Gb. Paired-end 150 bp sequences generated by Hi-C sequencing produced an additional 217.3 Gb of sequence data. These data were assembled with hifiasm in Hi-C mode to produce a primary genome assembly containing 1818 contigs with a contig N50 over 40 Mb and total length of 3.36 Gb, close to the genome size of 3.338 Gb predicted through previous flow cytometry work [58]. In addition, we identified a complete consensus mitogenome (**Figure S1**).

We evaluated the assembly quality via read coverage and depth (**Figure 2.1**) prior to filtering and found that 96.19% of the genome assembly was contained in contigs with coverage

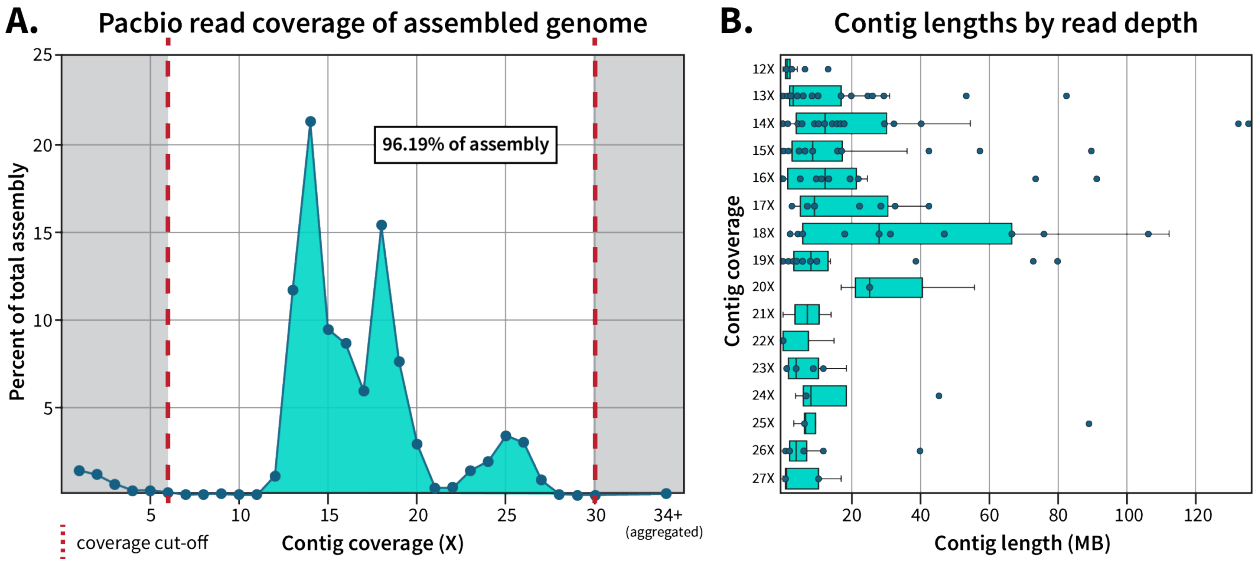


Figure 2.1: Quality profile of initial contig and final scaffold hifiasm assembly. (A) Pacbio reads were mapped to the primary genome assembly to determine overall coverage of individual contigs, and the percent each coverage level contributed to the total genome size was plotted. **(B)** The length of contigs per coverage level were plotted, with individual contigs represented as dots. Retained contigs with coverage between 6X-11X (n=73) or 28X-30X (n=5) were excluded from plotting in **(B)** due to short lengths.

between 6x and 30x. Of these contigs, 19 were removed as they were derived from the cockroach endosymbiont *Blattabacterium* which is a common contaminant in cockroach genomic data. Further filtering based on RNA-seq alignment decreased the number of contigs prior to scaffolding to 243 with a N50 of 42.4 Mb, overall producing a high-quality contig-level genome with no detectable contamination.

Table 2.1: Genome assembly comparison with previous *Periplaneta americana* assemblies available on NCBI.

Assembly name		ASM293952v1	ASM2559430v2	PAMFEO1_priV1 (this assembly)
	Accession	GCA_002939525.1	GCA_025594305.2	GCF_040183065.1
Genome	**Total size (Gb)**	3.4	3.1	3.2
	Ungapped (Gb)	3.2	3.1	3.2
	GC (%)	35.5	35.42	35.5
Scaffolds	**Count (#)**	18,601	48	91
	N50 (Mb)	0.3325	150.7	188.1
	L50 (#)	2951	9	8
Contigs	**Count (#)**	122,589	9217	259
	N50 (Mb)	0.0508	1.9	42.4
	L50 (#)	17,827	416	22
	Repeats (%)	57.80	62.90	65.86
	Genome BUSCO (% complete)	97.60	94.60	99.70
	Protein BUSCO (% complete)	91.20	90.50	97.80
	Protein coding genes (#)	21,336	29,939	16,780

Initial scaffolding with YaHS assigned 95.67% of the genome into 17 scaffolds (**Figure S2**), with manual curation in Juicebox increasing this percentage to 98.93% (**Figure 2.2A**). Telomere analysis identified at least one telomeric region, determined *de novo* as AACCTAACCT, for each of the chromosome-sized scaffolds, with 10 scaffolds flanked on both ends and 6 scaffolds flanked on one end (**Figure 2.2B**). Scaffold #7 contains a long sequence of probable telomeric repeats embedded within a large contig. While this may indicate an assembly error, at this time we do not have evidence to support adjusting its placement. The 74 remaining contigs were unable to be matched to just one scaffold, likely due to a high density of centromeric repeat regions and were therefore left unplaced. BUSCO analysis of the putative

chromosomes found 99.7% of Insecta BUSCOs were present and complete in this genome assembly, of which only 1.5% were duplicated (**Figure S3, Table 2.1**).

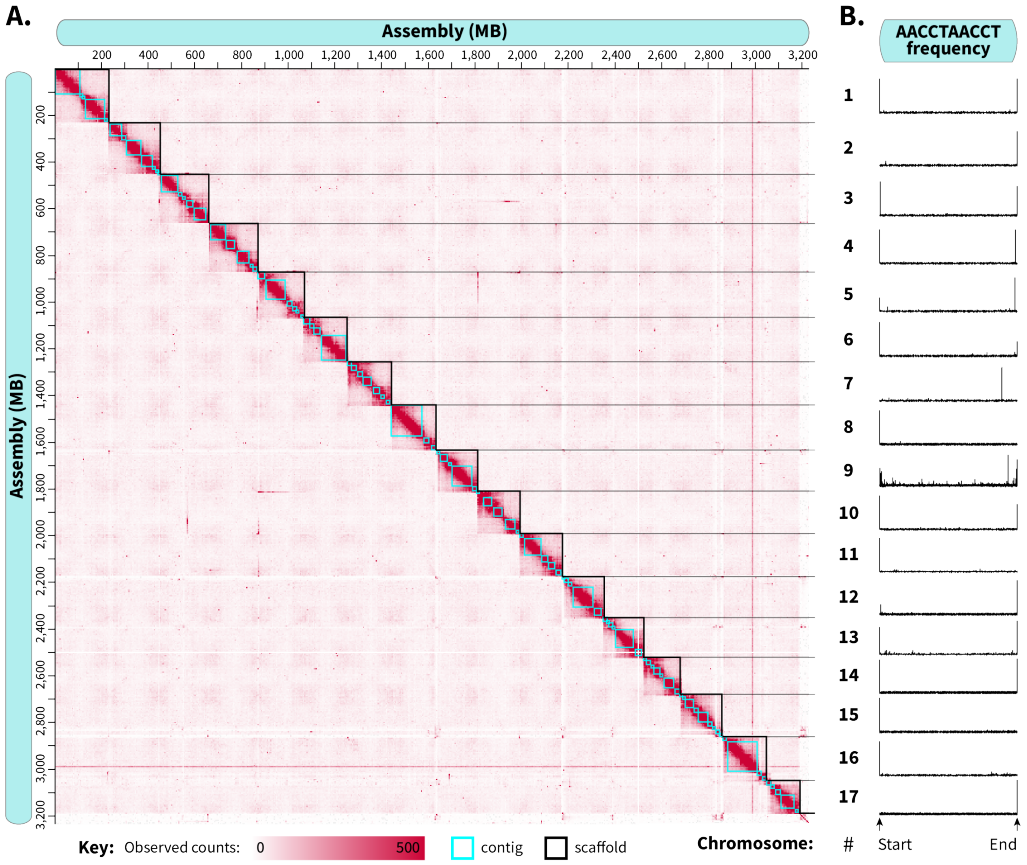


Figure 2.2: Chromatin-contact sequencing produced an assembly with near-chromosomal resolution. Following scaffolding with Yet Another Hi-C Scaffolder (YAHS), scaffold boundaries and contig placement were adjusted in Juicebox to optimize chromatin contacts for the 17 chromosome-level scaffolds. Final chromosomal boundaries are shown in the heatmap in **(A),** and the occurrence of telomeric sequences within each chromosome are displayed in **(B).**

In summary, the assembly presented here is considered high quality across a number of standard metrics. Through a combination of long-read and chromosomal contact sequencing data, we successfully scaffolded 98.9% of this 3.23 Gb assembly into 17 chromosome-scale scaffolds, values which are supported by previous karyotype and flow cytometry findings for *P. americana* of 17 haploid (male/female: 33/34 diploid) chromosomes and a predicted genome size of 3.338 Gb [31, 57, 58]. This represents an improvement on the currently available *P.*

americana genome assemblies in NCBI (**Table 2.1**), which resemble this assembly in length and GC content but have lower contiguity and BUSCO scores [13, 30]. Therefore, we argue that this genome assembly qualifies as both comprehensive and chromosomally resolved.

2.4.2. Repetitive DNA elements

Overall, 50.86% of the genome was identified as repetitive content classified as DNA elements, simple repeat regions, or retroelements, with an additional 15% of the genome determined to be *P. americana*-specific repeats that remain otherwise unclassified **(Figure 2.3A, Table S1**). The DNA transposon and retroelement subgroups contributed similarly to overall repeat content, comprising 24.3% and 20.05% respectively of the genome, but differ in their overall Kimura divergence landscapes (**Figure 2.3B)**. The DNA elements found in this genome primarily belong to the Tc1-Mariner and hobo-Activator-Tam3 subfamilies with higher relative abundances at a Kimura substitution level of 5% (**Figure 2.3C**). In contrast, the primary retrotransposon class of long interspersed nuclear elements (LINEs) contained two peaks in Kimura substitution, with CR1, L1, L2, and CRE subgroups showing elevated substitution levels around 5% while RTE-clade retrotransposons represent a more ancient repeat lineage, with substitution levels peaking between 31-32% **(Figure 2.3D)**. Short interspersed nuclear elements (SINEs) and retroelements containing long-terminal repeats (LTRs) were less commonly identified, only making up 3.47% and 0.94% of the genome respectively. Most SINEs contained internal promoters derived from tRNA and showed more gradual patterns of divergence, with a relatively stable plateau between 5% and 15% divergence before tapering (**Figure 2.3E).** Repeat content varies widely even between insects from the same family, but generally Blattodea species show similar repeat distributions with especially low LTR content and expanded LINEs, with larger genomes correlated to higher repeat content [59].

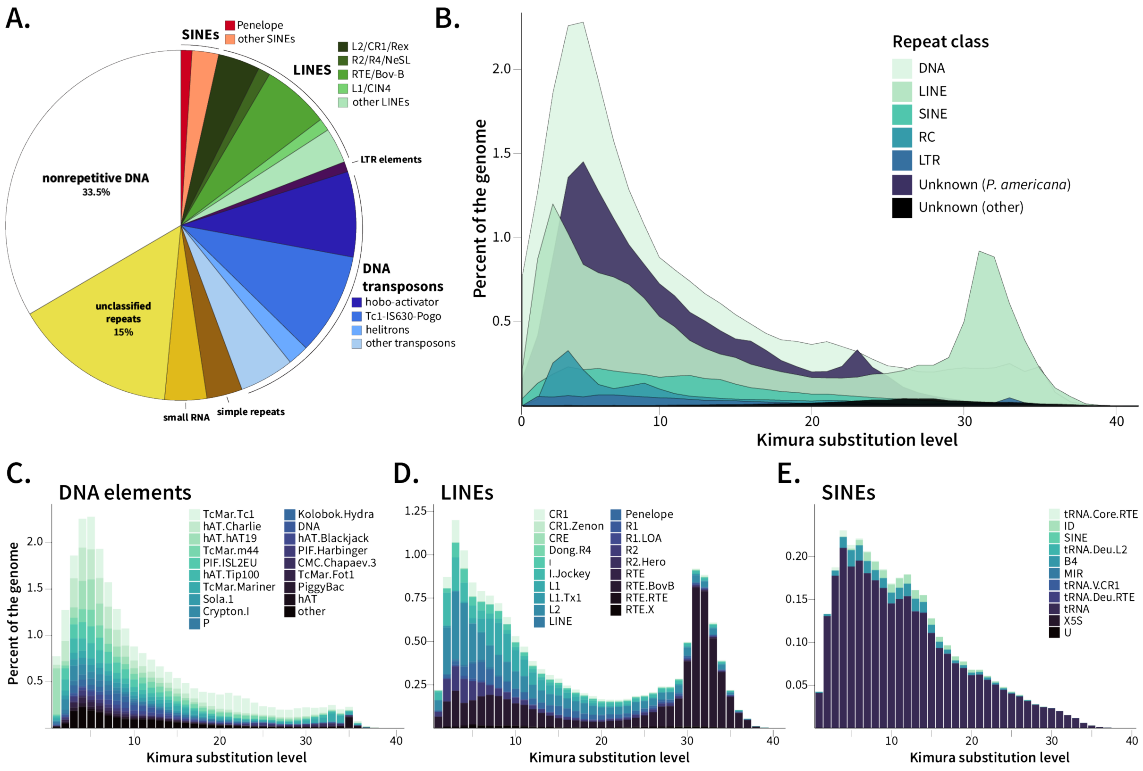


Figure 2.3: Repeat summary of *P. americana*. RepeatMasker and RepeatModeler were used to identify **(A)** the abundance of repetitive element families present in this assembly and **(B)** the relative abundance of each repeat class versus Kimura substitution level. The repeat landscapes for the classes **(C)** DNA elements, **(D)** LINEs, and **(E)** SINEs were further visualized at the repeat subtype level.

2.4.3. Orthology analysis between Blattodea species

The order Blattodea encompasses both termites and cockroaches, with over 4700 species identified in NCBI's taxonomy repository. Despite these many representatives, only 12 species have sequenced genomes [13, 30, 48-50, 60-63], of which half have publicly available annotations uploaded to NCBI (**Figure 2.4A**). The 3.2 Gb genome of *P. americana* is the largest among sequenced Blattodea and more than double the size of available termite genomes (**Figure 2.4B**), consistent with two previous *P. americana* genome assemblies. This assembly has the highest contig N50 among these Blattodea (**Figure 2.4C**) and the second highest scaffold N50 behind *E. pallidus* [60] **(Figure 2.4D).** Compared to those genomes with available annotations, *P. americana* encodes more protein-coding genes than termites, but had less reported than the

other cockroach genomes (**Figure 2.4E**). It is unclear whether this difference between annotations in our assembly and the other cockroach assemblies is biological or a result of annotation technique; this assembly was annotated by NCBI's Gnomon pipeline which produced 16,780 protein-coding genes, while previous *P. americana* genomes reported 21,336 (annotations not available) and 29,939 (annotations available) protein-coding genes [13, 30].

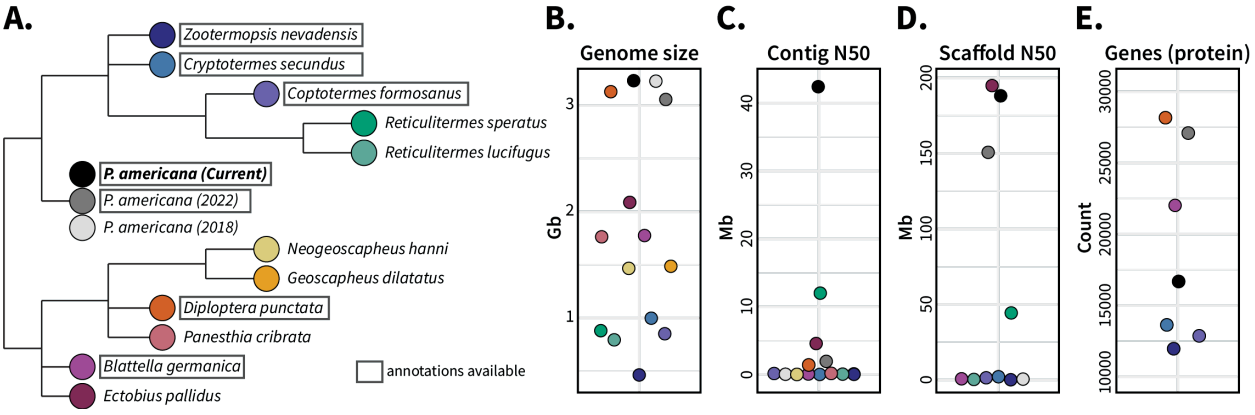


Figure 2.4: Available Blattodea genomes from NCBI. Phylogeny of sequenced Blattodea genomes as they appear in the NCBI taxonomy browser is presented in **(A)**, in addition to their **(B)** genome size, **(C)** contig N50, **(D)** scaffold N50, and **(E)** number of protein-coding genes (if available). Points on the genome statistic plots **(B-E)** correspond to the colors in **(A)** and species in (**A**) with protein annotations available on NCBI are boxed.

We chose to use the genome annotations available on NCBI for ortholog analysis, which included three members of Termitidae (*Zootermopsis nevadensis, Cryptotermes secundus, Coptotermes formosanus*) , two Blaberoidea (*Blattella germanica, Diploptera punctata*) and a previous *P. americana* genome [13]. Initial OrthoFinder results included both *P. americana* assemblies, identifying 16,711 orthologous gene families among these seven annotated genomes (**Table S2.2**). Of these 16,711 gene families, 2,311 were shared across all seven Blattodea with an additional 2,220 gene families shared by all assemblies excluding the previous *P. americana* assembly (**Figure S4**). While both *P. americana* genomes shared 391 gene families absent in the

other species, 198 gene families were only present in our assembly (**Figure S4**). We further evaluated the differences between both *P. americana* assemblies via synteny analysis (**Figure 2.5A**) and found that, while most chromosomes in our assembly were captured in their entirety by 1-2 scaffolds in the previous assembly, our chromosome 10 and half of chromosome 14 were missing from the other *P. americana* annotations. Proteins encoded in chromosome 10 encompass a wide range of functions with immune (Dscam, toll-like receptors, leucine-rich repeat proteins), neurologic (GABA transport, neurotrophic factors), endocrine (vitellogenin synthesis, sterol binding proteins), and nutritional (xanthine dehydrogenase, salivary peptide) importance (**Appendix A: Supplemental File 2.1).** Additional synteny analysis between our *P. americana* genome and other cockroaches (**Figure 2.5B**) and termites (**Figure 2.5C**) supported the existence of these regions. As a result, the previous *P. americana* assembly was excluded from further cross-*Blattodea* orthogroup analysis.

Recalculating OrthoFinder statistics produced 15,491 total gene families, of which 4,531 orthogroups were shared by all species (**Figure 2.5D**). These gene families, which are shared at the order level, comprised large fractions of each insect's total gene count, ranging from 5,674 genes (45.8%) in *Z. nevadensis* to 7,860 genes (27.7%) in *D. punctata* **(Figure 2.5D inset)**. These core gene families encompass a wide range of functions necessary to life and common in insects, with GO terms relating to gene expression and genome maintenance especially prevalent in this subset (**Figure S5**). The full list of GO terms assigned to these shared gene families can be found in **Appendix A: Supplemental File 2.1**.

The percentage of genes unable to be assigned to orthogroups varied between Blaberoidea and Blattoidea (**Figure 2.5D inset**). The termite genomes and our *P. americana* assembly had between 95-97.7% of their genes assigned to orthogroups, while 75.3% and 76.2%

of *D. punctata* and *B. germanica* genes, respectively, were assigned to orthogroups (**Figure 2.5D inset**). Unassigned genes are considered unique both between analyzed genomes and within a single genome, so it is unclear whether these genes stem from expanded single-copy genes within the cockroaches or are misannotated as genes without conferring any biological function. Possibly, the annotations of the termites and our *P. americana* are overly conservative, but gene finding programs rely on homology comparison that requires more sequenced and annotated Blattodea genomes for effective performance. Despite these questions, the difference between cockroaches and termites in their species-specific gene families highlight the gene expansion occurring as these two groups diverged. Our *P. americana* assembly, which encodes over 3,000 more genes than the termite assemblies, has 11.7% of its genes assigned to the 589 *P. americana*-unique gene families, a substantial increase compared to the 1.5-3.2% range in the three termite assemblies **(Figure 2.5D inset**). This expansion of genes within a single species also occurs in *B. germanica* and *D. punctata*, which have 23.2% and 13.5% of their genes respectively assigned to species-specific gene families. Since ortholog analysis is dependent on the data supplied, it is difficult to determine whether this difference stems from biological cockroach-termite delineation or is a consequence of poor representation of closely related cockroach species. Nonetheless, these results, in combination with the large size of this cockroach genome (**Figure 2.4B**) and the late spike in LINE retrotransposon divergence **(Figure 2.3D),** suggests that acquisition and expansion of new genes contributed to cockroach divergence.

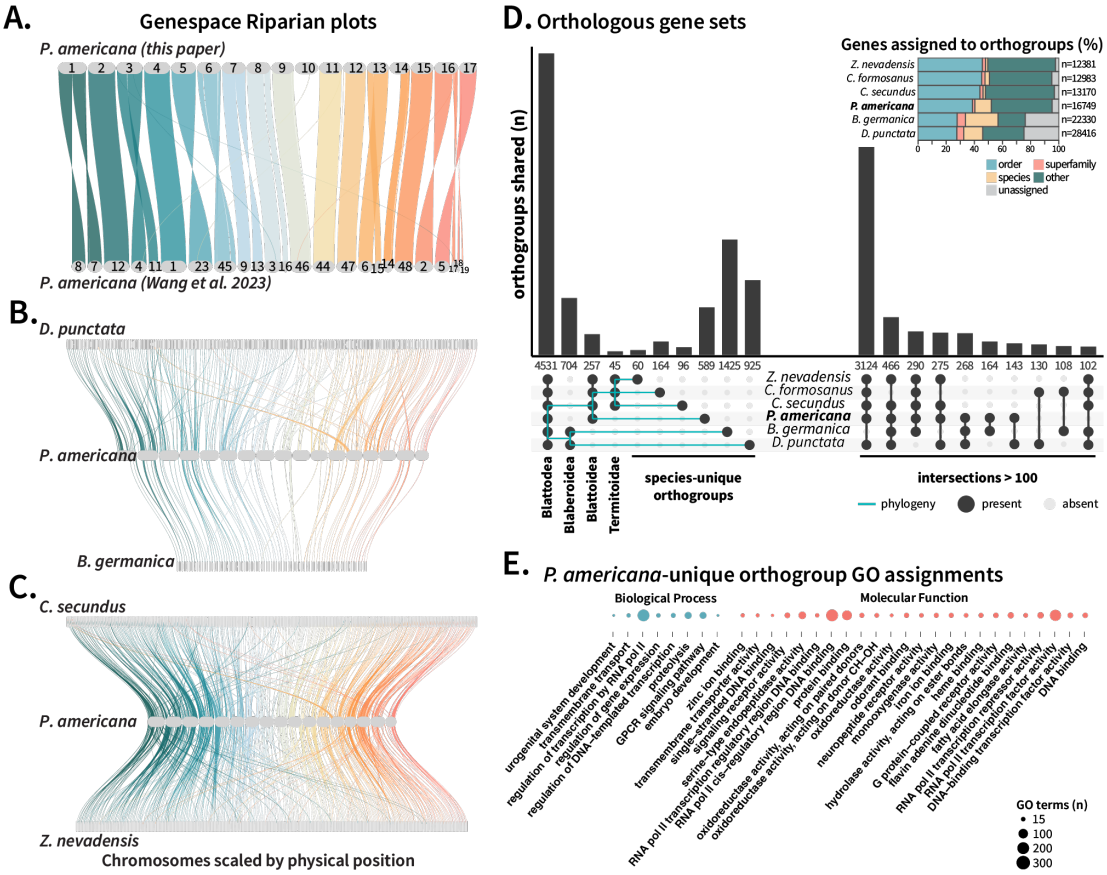


Figure 2.5: Ortholog analysis within Blattodea. Protein GFF3 files available on NCBI for Blattodea species were compared using OrthoFinder and GENESPACE. Riparian plots were generated to assess synteny between the *P. americana* genome presented here and **(A)** a previous *P. americana* assembly (GCA_025594305.2), **(B)** the cockroaches *Diploptera puntata* (GCA_030220185.1) and *Blattella germanica* (GCA_003018175.1), and **(C)** the termites *Cryptotermes secundus* (GCF_002891405.2) and *Zootermopsis nevadensis* (GCF_000696155.1). (**D**) Orthologous gene clusters shared by or unique to the analyzed genomes were visualized via UpSet plot, organized by phylogeny on the left and the next 10 largest sets on the right and summarized by (**inset D)** the percentage of genes assigned to shared or unique orthogroups. Genes belonging to the 589 orthogroups identified as *P. americana*-unique were analyzed by their associated GO terms, and (**E**) biological processes and molecular functions with at least 15 occurrences were visualized.

We performed GO term analysis of the 589 orthologous gene families present in our *P. americana* assembly but absent in the other Blattodea species (**Figure 2.5E**). GO terms related to gene expression wand genomic maintenance were most enriched, reflecting the GO enrichment in shared gene families (**Figure S5**). Further investigation revealed that 157 of these orthogroups

were associated with zinc finger family genes, likely skewing the GO analysis results and masking other expanded functions of interest. As an alternative approach, we evaluated the semantic similarity of named genes and/or GO terms (if name is missing) of these *P. americana*-unique gene families to shed light on the function of these expanded genes (**Figure S6**). These gene families were enriched in immune and digestive functions, such as lipopolysaccharide recognition, protease activity, odorant binding, and lipase activity (**Figure S6**), which may have facilitated cockroach divergence from the protective eusociality found in termite colonies towards a more independent and self-sufficient lifestyle. However, there are a limited number of sequenced cockroach representatives, so further cataloging of diverse Blattodea genomes is necessary to pinpoint exact relationships between these gene families and cockroach-termite evolution. Overall, synteny and ortholog comparison between these Blattodea reveal possible mechanisms of divergence between termites and cockroaches and highlight the potential applications of a chromosomally resolved *P. americana* genome.

Summary

We sequenced the genome of the American cockroach using high fidelity PacBio long reads in conjunction with Hi-C Illumina short reads. This 3.23 Gb assembly is highly contiguous, with a contig N50 of 42 Mb and a scaffold N50 of 188 Mb, and 98.93% of the assembly is contained within 17 putative chromosomes. The quality of this assembly is further exemplified by its genomic and protein BUSCO scores, which are 99.7% and 97.8% complete respectively. This high-quality assembly, generated with cutting edge sequencing technology, is a substantial improvement over existing *P. americana* genomes, and we report an entire chromosome that was missing from a previously published assembly. This genome is expected to facilitate future study of cockroach physiology and Blattodea evolution.

Data Availability

Data associated with this study are available from the NCBI Sequence Read Archive under BioProjects PRJNA1098420 (principal haplotype) and PRJNA1098419 (alternative haplotype). Raw sequencing reads for Hi-C data is available from the SRA with accession SRX24490912, and PacBio HiFi reads may be obtained from accessions SRX24490909, SRX24490910, and SRX24490911. RNA-seq data is deposited under the SRA BioProject PRJNA1105088 (experiments SRX27556002- SRX27556025). Scripts used for assembling and analyzing this genome are available at: https://github.com/rldockman/PAMFEO.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under award number R35GM133789 to EAO. This research was supported by the in-house appropriated USDA-ARS project *Advancing Molecular Pest Management, Diagnostics, and Eradication of Fruit Flies and Invasive Species* (no. 2040-22430-028-000-D) and used resources provided by SCINet, USDA-ARS projects no. 0201-88888-003-000D and 0201-88888-002-000D. USDA is an equal opportunity provider and employer. Mention of trade names does not imply an endorsement from USDA or the Federal Government.

References

1. Crampton, G.C., *The External Anatomy of the Head and Abdomen of the Roach, Periplaneta americana.* Psyche: A Journal of Entomology, 1925. **32**(4-5): p. 127409.
2. Guthrie, D.M. and A.R. Tindall, *The biology of the cockroach*. 1968, London: Edward Arnold. 408.
3. Cornwell, P.B., *The cockroach: a laboratory insect and an industrial pest*. 1968, United Kingdom: Huthinson.
4. *The American Cockroach*. 1982, London: Chapman and Hall.
5. Jacklet, J.W. and M.J. Cohen, *Nerve Regeneration: Correlation of Electrical, Histological, and Behavioral Events.* Science, 1967. **156**(3782): p. 1640-1643.
6. Jankowska, M., et al., *Electromagnetic field and TGF-β enhance the compensatory plasticity after sensory nerve injury in cockroach Periplaneta americana.* Scientific Reports, 2021. **11**(1): p. 6582.
7. Yamasaki, T. and T. Narahashi, *Synaptic Transmission in the Cockroach.* Nature, 1958. **182**(4652): p. 1805-1806.
8. Sattelle, D.B., *Receptors for L-glutamate and GABA in the nervous system of an insect (Periplaneta americana).* Comp Biochem Physiol C Comp Pharmacol Toxicol, 1992. **103**(3): p. 429-38.
9. Roeder, K.D., *The effect of potassium and calcium on the nervous system of the cockroach, Periplaneta americana.* J Cell Comp Physiol, 1948. **31**(3): p. 327-38.
10. Vicente, C.S.L., et al., *Genome analysis of new Blattabacterium spp., obligatory endosymbionts of Periplaneta fuliginosa and P. japonica.* PLoS One, 2018. **13**(7): p. e0200512.
11. Cochran, D.G., D.E. Mullins, and K.J. Mullins, *Cytological Changes in the Fat Body of the American Cockroach, Periplaneta americana, in Relation to Dietary Nitrogen Levels.* Annals of the Entomological Society of America, 1979. **72**(2): p. 197-205.

12. Arumugam, G., et al., *Purification and functional characterization of lectin with phenoloxidase activity from the hemolymph of cockroach, Periplaneta americana.* Arch Insect Biochem Physiol, 2017. **95**(2).

13. Wang, L., et al., *Genome assembly and annotation of Periplaneta americana reveal a comprehensive cockroach allergen profile.* Allergy, 2023. **78**(4): p. 1088-1103.

14. Zhang, J.H., et al., *New insight into foregut functions of xenobiotic detoxification in the cockroach Periplaneta americana.* Insect Sci, 2018. **25**(6): p. 978-990.

15. Zhang, J., et al., *Midgut Transcriptome of the Cockroach Periplaneta americana and Its Microbiota: Digestion, Detoxification and Oxidative Stress Response.* PloS one, 2016. **11**(5): p. e0155254-e0155254.

16. Sun, H., et al., *Elucidating the detoxification efficacy of Periplaneta americana delta glutathione S-transferase 1 (PaGSTd1) against organophosphates.* Pesticide Biochemistry and Physiology, 2024. **203**: p. 106013.

17. Zhang, E., et al., *A minireview of the medicinal and edible insects from the traditional Chinese medicine (TCM).* Front Pharmacol, 2023. **14**: p. 1125600.

18. Tsuneo, N., M.A. Yong-Hua, and I. Kenji, *Insect derived crude drugs in the chinese song dynasty.* Journal of Ethnopharmacology, 1988. **24**(2): p. 247-285.

19. Li, L., et al., *Applications of RNA Interference in American Cockroach.* J Vis Exp, 2021(178).

20. Huang, J.H., et al., *Practical Use of RNA Interference: Oral Delivery of Double-stranded RNA in Liposome Carriers for Cockroaches.* J Vis Exp, 2018(135).

21. French, A.S., et al., *Transcriptome analysis and RNA interference of cockroach phototransduction indicate three opsins and suggest a major role for TRPL channels.* Front Physiol, 2015. **6**: p. 207.

22. Tinker, K.A. and E.A. Ottesen, *The Core Gut Microbiome of the American Cockroach, Periplaneta americana, Is Stable and Resilient to Dietary Shifts.* Appl Environ Microbiol, 2016. **82**(22): p. 6603-6610.

23. Dockman, R.L. and E.A. Ottesen, *Purified fibers in chemically defined synthetic diets destabilize the gut microbiome of an omnivorous insect model.* Frontiers in Microbiomes, 2024. **3**(1477521).

24. Zurek, L. and B.A. Keddie, *Contribution of the colon and colonie bacterial flora to metabolism and development of the American cockroach Periplaneta americana (L).* Journal of Insect Physiology, 1996. **42**(8): p. 743-748.

25. Ayayee, P.A., et al., *The role of gut microbiota in the regulation of standard metabolic rate in female Periplaneta americana.* PeerJ, 2018. **6**: p. e4717.

26. Dukes, H.E., J.E. Dyer, and E.A. Ottesen, *Establishment and Maintenance of Gnotobiotic American Cockroaches (Periplaneta americana).* JoVE, 2021(171): p. e61316.

27. Tegtmeier, D., et al., *Oxygen Affects Gut Bacterial Colonization and Metabolic Activities in a Gnotobiotic Cockroach Model.* Appl Environ Microbiol, 2016. **82**(4): p. 1080-9.

28. Vera-Ponce de León, A., et al., *Microbiota Perturbation or Elimination Can Inhibit Normal Development and Elicit a Starvation-Like Response in an Omnivorous Model Invertebrate.* mSystems, 2021. **6**(4): p. e0080221.

29. Mikaelyan, A., et al., *Deterministic Assembly of Complex Bacterial Communities in Guts of Germ-Free Cockroaches.* Appl Environ Microbiol, 2016. **82**(4): p. 1256-63.

30. Li, S., et al., *The genomic and functional landscapes of developmental plasticity in the American cockroach.* Nature Communications, 2018. **9**(1): p. 1008.

31. John, B. and K.R. Lewis, *Chromosome structure in Periplaneta americana.* Heredity, 1960. **15**(1): p. 47-54.

32. Kopylova, E., L. Noé, and H. Touzet, *SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data.* Bioinformatics, 2012. **28**(24): p. 3211-3217.

33. Cheng, H., et al., *Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm.* Nature Methods, 2021. **18**(2): p. 170-175.

34. Lu, J. and S.L. Salzberg, *Removing contaminants from databases of draft genomes.* PLoS Comput Biol, 2018. **14**(6): p. e1006277.

35. Shen, W., et al., *SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation.* PLOS ONE, 2016. **11**(10): p. e0163962.

36. Wood, D.E., J. Lu, and B. Langmead, *Improved metagenomic analysis with Kraken 2.* Genome Biology, 2019. **20**(1): p. 257.

37. Kim, D., et al., *Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.* Nature Biotechnology, 2019. **37**(8): p. 907-915.

38. Storer, J., et al., *The Dfam community resource of transposable element families, sequence models, and genome annotations.* Mobile DNA, 2021. **12**(1): p. 2.

39. Li, H., *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.* arXiv q-bio.GN, 2013.

40. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2).

41. Durand, N.C., et al., *Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom.* Cell Syst, 2016. **3**(1): p. 99-101.

42. Zhou, C., S.A. McCarthy, and R. Durbin, *YaHS: yet another Hi-C scaffolding tool.* Bioinformatics, 2023. **39**(1).

43. Brown, M., P.M. González De la Rosa, and B. Mark, *A Telomere Identification Toolkit.* Zenodo, 2023.

44. Manni, M., et al., *BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes.* Molecular Biology and Evolution, 2021. **38**(10): p. 4647-4654.

45. Uliano-Silva, M., et al., *MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads.* BMC Bioinformatics, 2023. **24**(1): p. 288.

46. Xiao, B., et al., *Complete mitochondrial genomes of two cockroaches, Blattella germanica and Periplaneta americana, and the phylogenetic position of termites.* Curr Genet, 2012. **58**(2): p. 65-77.

47. Flynn, J.M., et al., *RepeatModeler2 for automated genomic discovery of transposable element families.* Proceedings of the National Academy of Sciences, 2020. **117**(17): p. 9451-9457.

48. Terrapon, N., et al., *Molecular traces of alternative social organization in a termite genome.* Nat Commun, 2014. **5**: p. 3636.

49. Harrison, M.C., et al., *Hemimetabolous genomes reveal molecular basis of termite eusociality.* Nature Ecology & Evolution, 2018. **2**(3): p. 557-566.

50. Fouks, B., et al., *Live-bearing cockroach genome reveals convergent evolutionary mechanisms linked to viviparity in insects and beyond.* iScience, 2023. **26**(10): p. 107832.

51. Emms, D.M. and S. Kelly, *OrthoFinder: phylogenetic orthology inference for comparative genomics.* Genome Biology, 2019. **20**(1): p. 238.

52. Lovell, J.T., et al., *GENESPACE tracks regions of interest and gene copy number variation across multiple genomes.* eLife, 2022. **11**: p. e78526.

53. Emms, D.M. and S. Kelly, *OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy.* Genome Biology, 2015. **16**(1): p. 157.

54. Conway, J.R., A. Lex, and N. Gehlenborg, *UpSetR: an R package for the visualization of intersecting sets and their properties.* Bioinformatics, 2017. **33**(18): p. 2938-2940.

55. Yu, G., et al., *clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.* OMICS: A Journal of Integrative Biology, 2012. **16**(5): p. 284-287.

56. Wu, T., et al., *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* The Innovation, 2021. **2**(3).

57. JANKÁSEK, M., Z. KOTYKOVÁ VARADÍNOVÁ, and F. ŠŤÁHLAVSKÝ, *Blattodea Karyotype Database.* EJE, 2021. **118**(1): p. 192-199.

58. Hanrahan, S.J. and J.S. Johnston, *New genome size estimates of 134 species of arthropods.* Chromosome Res, 2011. **19**(6): p. 809-23.

59. Sproul, J.S., et al., *Analyses of 600+ insect genomes reveal repetitive element dynamics and highlight biodiversity-scale repeat annotation challenges.* Genome Res, 2023. **33**(10): p. 1708-1717.

60. Hunter, T., *The genome sequence of the tawny cockroach, Ectobius (Ectobius) pallidus (Olivier, 1789).* Wellcome Open Res, 2025. **10**: p. 22.

61. Ewart, K., S. Ho, and A.-A. Chowdhury, *Annotated genome assemblies for Geoscapheus dilatatus, Panesthia cribrata and Neogeoscapheus hanni* 2024: Dryad.

62. Shigenobu, S., et al., *Genomic and transcriptomic analyses of the subterranean termite Reticulitermes speratus: Gene duplication facilitates social evolution.* Proceedings of the National Academy of Sciences, 2022. **119**(3): p. e2110361119.

63. Martelossi, J., et al., *Wood feeding and social living: Draft genome of the subterranean termite Reticulitermes lucifugus (Blattodea; Termitoidae).* Insect Mol Biol, 2023. **32**(2): p. 118-131.

CHAPTER 3

PURIFIED FIBERS IN CHEMICALLY DEFINED SYNTHETIC DIETS DESTABILIZE THE GUT MICROBIOME OF AN OMNIVOROUS INSECT MODEL[2]

[2] Dockman, R.L. and E.A. Ottesen. 2024. *Frontiers in Microbiomes*. Volume 3. Reprinted here with permission of the publisher.

Abstract

The macronutrient composition of a host's diet shapes its gut microbial community, with dietary fiber in particular escaping host digestion to serve as a potent carbon source for gut microbiota. Despite widespread recognition of fiber's importance to microbiome health, nutritional research often fails to differentiate hyper-processed fibers from cell-matrix derived intrinsic fibers, limiting our understanding of how individual polysaccharides influence the gut community. We use the American cockroach (*Periplaneta americana*) as a model system to dissect the response of complex gut microbial communities to dietary modifications that are difficult to test in traditional host models. Here, we designed synthetic diets from lab-grade, purified ingredients to identify how the cockroach microbiome responds to six different carbohydrates (chitin, methylcellulose, microcrystalline cellulose, pectin, starch, xylan) in otherwise balanced diets. We show via 16S rRNA gene profiling that these synthetic diets reduce bacterial diversity and alter the phylogenetic composition of cockroach gut microbiota in a fiber-dependent manner, regardless of the vitamin and protein content of the diet. Comparisons with cockroaches fed whole-food diets reveal that synthetic diets induce blooms in common cockroach-associated taxa and subsequently fragment previously stable microbial correlation networks. Our research leverages an unconventional microbiome model system and customizable lab-grade artificial diets to shed light on how purified polysaccharides, as opposed to nutritionally complex intrinsic fibers, exert substantial influence over a normally stable gut community.

Introduction

The gut microbiome is a key player in host metabolism and homeostasis; it extracts energy from recalcitrant dietary components, provisions essential nutrients, and stimulates the host's immune system to protect against pathogens and toxins [1-4]. These benefits to the host are contingent upon the microbiota present, which themselves are selected through external pressure such as host genetics environment, and diet [5-9]. Diet has gained particular attention as the most easily manipulated of these factors, and a clear relationship exists between microbially derived metabolic products from the gut microbiome and overall host health [10, 11].

Shifts in the ratios and sources of metabolizable macronutrients (fats, carbohydrates, protein) are frequently identified as drivers of diet-associated microbiota alterations, but the most important component to resident gut bacteria is what bypasses host digestion relatively untouched: fiber [12, 13]. Dietary fiber consists of plant-derived structural carbohydrates that most animals are unable to process and are thus key to maintaining a diverse, beneficial gut microbial community. However, performing research relating dietary fiber consumption to gut microbiota within a host organism presents several challenges. Whole foods contain "intrinsic fibers", an assortment of carbohydrates characterized by source-specific molecular structures that form close associations with plant proteins and cell matrix components [14-16]. These heterogenous structures can obscure the influence of individual polysaccharides on the gut community, especially considering the high diversity of carbohydrate degrading machinery found across individual lineages of gut microbiota [17-19]. Purified fibers present an alternative that control for these variable compounds, but mammalian models have complex nutritional needs that limit the extent of dietary manipulation possible before introducing host stresses. As a result, *in vivo* dietary research on fiber-microbiome dynamics frequently uses balanced diets

containing host-metabolizable carbohydrates that are supplemented with purified fibers, therefore exposing gut bacteria to a mix of carbon sources. This restricts the conclusions that can be made on microbial response to the fiber itself, since there is no way to prevent gut microbiota from prioritizing an alternative energy source instead of the fiber of interest. Invertebrate models offer more flexibility, but well-known insect models such as *Drosophila* have limited dietary range that poorly reflect the community dynamics found in mammalian host species [20, 21]. To address this challenge, we are developing the omnivorous American cockroach (*Periplaneta americana*) as a model of microbiome dynamics that extends our understanding of human-relevant bacteria while leveraging the benefits of invertebrate research.

The cockroach digestive system is divided into three major regions: the foregut, midgut, and hindgut. The foregut, analogous to the mammalian stomach, consists of a large crop where salivary amylase and midgut-derived trypsin initiate the digestion of carbohydrates and proteins prior to mechanical breakdown via the proventriculus or gizzard [22]. The midgut, analogous to the mammalian small intestine, is lined with a continuously secreted chitinous peritrophic matrix (rather than a mucus layer) and serves as the primary source of host-derived aminopeptidases, cellobiase, and lipase, as well as the primary site of nutrient absorption [22, 23]. The hindgut, analogous to the mammalian large intestine, facilitates microbial fermentation of undigested and/or unabsorbed dietary substrates [24]. Fiber comprises most of this undigested material, but as in mammals, other macronutrients may escape host digestion due to factors such as plant-derived protease inhibitors or the diet's structural complexity [24, 25].

Despite the obvious differences between cockroaches and humans, in the context of host-microbe symbioses, there are both similarities and unique benefits that support the use of omnivorous cockroaches as a promising model for studying diet-gut microbiome dynamics.

American cockroaches are colonized by a complex hindgut microbiome that is taxonomically similar to the human colonic flora, consisting of many shared family and genus-level microbial lineages within the Bacteroidota, Firmicutes (now Bacillota), and Proteobacteria (now Pseudomonadota) [26-28]. These microbiota also play functionally analogous roles to their mammalian counterparts in host nutrition, with hindgut bacteria scavenging escaped nutrients and fermenting otherwise indigestible dietary components into volatile fatty acids that are absorbed by the host for energy [29-32]. Further, cockroaches host in their fat body an endosymbiotic bacterium, *Blattabacterium,* that protects the host from short-term starvation through the conversion of stored uric acid into essential amino acids [33-35]. This unique trait enables the cockroach to survive extreme dietary manipulation for extended periods of time.

Studies of cockroach gut microbiome responses to diet have generated contrasting responses, with multiple large-scale studies finding that the cockroach gut microbiome is highly stable between groups given differing diets [27, 36, 37], while others have demonstrated that diet alterations result in different gut microbiome configurations [38-40]. Currently, there is no consensus on why these studies produced differing results and comparison is difficult due to inconsistent use of synthetic or whole food diets across studies. Structurally complex whole foods may obscure bacterial responses to specific nutritional alterations, but synthetic diets are amenable to precise dietary changes, thus allowing stricter variable control [41-43]. Artificial diets have been successfully developed for insects with far more specialized dietary needs than cockroaches, suggesting that cockroaches are ideal candidates for dietary experimentation with lab-synthesized diets [44-47].

To facilitate precise manipulation of dietary composition in cockroaches, we have developed a series of synthetic cockroach diets based on the work of early entomologists [48-

50]. These artificial diets serve as a nutritionally complete base to isolate the influence of specific dietary components on the *P. americana* hindgut microbiome, a community known to be resistant to dietary manipulation when fed macronutrient-biased whole food diets [27]. Using these synthetic diets as a base, we tested a spectrum of purified polysaccharides as the primary carbon and energy source to identify if the hindgut microbiome responds to specific fibers without obfuscation by intrinsic fiber components. We found that these diets resulted in much stronger impacts on gut microbiome composition than highly divergent whole food diets, with long-chain polysaccharide source exerting the largest effect despite alterations in their protein and micronutrient composition. Our work will facilitate future studies of gut microbiome responses to fine-scale dietary composition in the cockroach and shed light on how hyper-processed synthetic diets, which superficially appear to be nutritionally complete, destabilize a complex gut microbiome.

Materials and Methods

Insects and Experimental Conditions

Our *Periplaneta americana* colony has been maintained in captivity at the University of Georgia for over a decade. Mixed age and sex stock insects are maintained at room temperature in glass aquarium tanks with wood chip bedding and cardboard tubes for shelter in a 12:12 light:dark cycle. Water via cellulose sponge fit to a Tupperware reservoir and dog chow (Purina ONE chicken & rice formula, guaranteed analysis: 26% crude protein (min), 16% crude fat (min), 3% crude fiber (max)) are provided to stock colonies *ad libitum*.

Synthetic Diets

The synthetic diets created for dietary testing were designed to provide balanced nutrition while remaining malleable to component manipulation. Diets contained Vanderzant vitamin mix

Table 3.1. Synthetic diet compositions. CHO: carbohydrate; MeC: methylcellulose; MCC: microcrystalline cellulose; P-: protein deficient; V-: vitamin deficient; *: canned tuna was dried prior to weighing.

Diet Type	CHO %	Casein %	Peptone %	Mineral Mix %	Vitamin Mix %	Cholesterol %	Diet/CHO
Standard: Polysaccharide	70.5	17	8	3	0.5	1	Chitin, MeC, MCC, Pectin, Starch, Xylan
Standard: Simple Sugar	70.5	17	8	3	0.5	1	Cellobiose, Glucose, Xylose
Protein Deficient	95.5	0	0	3	0.5	1	MCC P-, Xylan P-
Vitamin Deficient	72	17	8	3	0	0	MCC V-, Xylan V-
Tuna-Amended	70.5	25% tuna		3	0.5	1	MCC, Xylan

(catalog #: 903244, MP Biomedicals, Irvine, CA, USA), Wesson salt mix (catalog #: 902851, MP Biomedicals), peptone (catalog #: J636, Amresco, VWR International, Radnor, PA, USA), casein (catalog #: C3400, Sigma-Aldrich, St. Louis, MO, USA), and cholesterol (catalog #: 0433, VWR); amounts are listed in **Table 3.1**. The dry ingredients were suspended in sufficient volumes of diH_2O to create a batter or dough, formed into pellets, then dehydrated at 65°C until they were sufficiently dry to maintain shape. Food pellets were stored at -20°C until use.

In most experiments, the only component changed was the carbohydrate source. Polysaccharides used include microcrystalline cellulose (MCC with 51um particle size; catalog #: 435236, Sigma-Aldrich), methylcellulose (catalog #: M0512, Sigma-Aldrich), xylan from corn core (catalog # TCX0078, TCI Chemicals, Portland, OR, USA), pectin from apple (catalog # 93854, Sigma-Aldrich), starch from potato (catalog #: S516, Fisher Chemical), and chitin (catalog #: J61206, Alfa Aesar, Ward Hill, MA, USA). For simple sugar diet variations, cellobiose (catalog #: 22150, Sigma-Aldrich), glucose (catalog #: G7021, Sigma-Aldrich), and

xylose (catalog #: 200001-008, Acros Organics, VWR International) were used as the carbohydrate component.

Experimental Design

Experimental conditions were prepared as described in Tinker and Ottesen [27]. Briefly, mixed-sex healthy adult insects (n=12/diet) were transferred from the stock colony to plastic tanks containing pebbles and bleached polyvinyl chloride tubes for footing and shelter, respectively. Food and water were provided *ad libitum* in rigid plastic or glass dishes following two days of food restriction and habituation. Dietary treatments for the four cohorts **(Table S3.1)** lasted two weeks, during which debris, oothecae, and lethargic insects were removed daily.

Upon completion of dietary treatments, all insects were sacrificed for sample collection. Insects were isolated in a sterile culture plate and placed on ice until torpid, upon which they were decapitated and dissected. Sternites were removed with sterile forceps to expose the intact gut and fat body tissue was cleaned away. The cleaned gut was frozen on a sterile aluminum dish on dry ice and divided into foregut, midgut, and hindgut sections for collection in 500-800µL phosphate-buffered saline (1X PBS). Gut contents and tissue-attached bacteria were disrupted with a sterile pestle, and the samples stored at –20°C until DNA extraction. For this study, only the hindgut community was analyzed due to higher microbial density and activity than in other gut regions.

DNA Extraction

DNA was extracted from 200µL aliquots of all individual samples using the EZNA Bacterial DNA Kit (Omega Biotek, Norcross, GA, USA) with some modifications. Sample aliquots were centrifuged at 5000g for 10min, with the resulting pellet resuspended in 100µL TE buffer plus 10µL lysozyme (50mg/mL) and incubated for 30 min at 37°C. Following incubation,

samples were vortexed with glass beads (25mg, Omega Biotek) for 5 min at 3000rpm, then incubated at 55°C for one hour with 100µL TL buffer, 20µL proteinase K and continuous 600rpm shaking. The kit protocol was followed for additional incubations with BL buffer and DNA isolation using the provided column. DNA was eluted into 50µL of provided Elution Buffer and quantified using either a Nanodrop Lite spectrophotometer (Thermo Scientific) or the Take3 plate for BioTek plate readers (Agilent).

16S rRNA Gene Library Preparation and Sequencing

The V4 region of the 16S rRNA gene was amplified via 2-step polymerase chain reaction (PCR) from individual hindgut lumen samples (n=8-12/diet) as previously described in [27, 51, 52]. Both PCR reactions used 0.02U/L Q5 Hot Start high-fidelity DNA polymerase (New England BioLabs, Ipswich, MA, USA) with 200M dNTPs and 0.5M forward and reverse primers in 1M Q5 reaction buffer. The first 10µL reaction containing 3ng DNA and primers targeting the V4 region (**515F**: GTGCCAGCMGCCGCGGTAA; **806R**: GGACTACHVGGGTWTCTAAT) was performed under the following conditions: activation at 98°C for 30s; 15 cycles of 98°C for 10s, 52°C for 30s, and 72°C for 30s; final extension at 72°C for 2 min. Immediately following amplification, 9µL of the first reaction was added to 21µL of Q5 reaction mix containing barcoded primers with adaptor sequences for Illumina sequencing [53]. Cycling was performed as follows: activation at 98°C for 30s; 4 cycles of 98°C for 10s, 52°C for 10s, and 72°C for 30s; 6 cycles of 98°C for 10s and 72°C for 1 min; final extension at 72°C for 2 min.

After product size verification via gel electrophoresis, samples were cleaned as instructed in Omega Biotek's Cycle Pure kit, quantified, and pooled for equimolar representation of each sample. Prepared libraries were sent to the Georgia Genomics and Bioinformatics Core at the University of Georgia for 250 base pair paired-end Illumina MiSeq sequencing.

Amplicon Sequence Variant Generation

Each dataset collected in this study was processed separately in R (version 4.2.1) by sequencing run using R package DADA2 (version 1.24.0), with the cumulative Amplicon Sequence Variants (ASVs) generated input as a priors table for each successive run [54, 55]. To allow for comparison with this dataset, raw data from previous research in the Ottesen lab were reprocessed to generate ASVs following the same procedures as in this current study [27]. All sequence tables produced by these datasets were combined by ASV sequence prior to taxonomy assignment to ensure continuity in naming. Taxonomy was assigned using DADA2 and the ARB Silva v138 classifier to the species level, uniquely numbered, and filtered to remove sequences matching eukaryotic (chloroplast, mitochondria) or endosymbiotic *Blattabacterium* DNA [54, 56].

Community Analysis

Alpha and beta diversity analyses were performed via the R package vegan (version 2.6-4) [57]. Samples were rarefied prior to diversity analysis to 7924 reads for comparisons between synthetic and/or whole food diets, 9685 reads for analysis of repeat xylan and MCC diet experiments, and 12274 reads for follow-up experiments exploring nutrient deficiencies and simple sugar carbohydrates. Alpha diversity was measured via Shannon index, the count of ASVs observed in rarefied samples, and Pielou's evenness (calculated as Shannon/log(Observed)). Weighted (binary = FALSE) and unweighted (binary = TRUE) Bray-Curtis dissimilarities were calculated, assessed for dispersion, and plotted using the vegan functions vegdist, betadisper, and metaMDS. Unweighted Bray-Curtis dissimilarity, or incidence-based Bray-Curtis dissimilarity, is equivalent to the Sørensen index in that it is based on the number of species shared or unique between groups without accounting for individual

species abundance [58]. Statistics for alpha diversity indices were calculated with the Wilcoxon rank sum test (pairwise comparisons) and Kruskal-Wallis test (multi-group comparisons). The significance of community composition differences observed in beta diversity measures was assessed using PERMANOVA (vegan::adonis2()). Beta dispersion was further examined through the Tukey's HSD test for pairwise comparisons and ANOVA for multi-group comparisons.

Differential abundance analysis was conducted using DESeq2 (version 1.36) [59]. For identification of 'diet-characteristic taxa", raw count data for the synthetic diet set (n=66) were filtered to exclude ASVs present in less than 5 samples and run through the 'DESeq' command with parameters 'fitType = "local"' and 'design = ~ Diet'. Pairwise result tables were obtained for all diet comparisons and filtered for significant data, defined as having an adjusted p-value smaller than 0.05 and a baseMean larger than 10. ASVs significantly upregulated for one diet vs the other five diets were identified as diet-characteristic (n=76) and used to generate the heatmap in **Figure S3.3**. For comparison of "synthetic vs whole", raw count data for both diet sets (n=125) were combined and filtered to exclude ASVs that appeared in fewer than 5 samples. DESeq2 was run with parameters 'fitType = "local"' and 'design = ~ Diet_Type' to identify differentially abundant ASVs between the diet types. The resulting baseMean and log2 fold change were used to generate the MA plots in **Figure 3.4**.

UpSetR (version 1.4) was employed to visualize intersecting sets of taxa, providing insights into the distribution of taxonomic features across samples [60]. For UpSet analysis, samples were rarefied to 7924 reads then collapsed together to obtain total counts per diet. Both a presence/absence table and a proportion table were generated from these data, with the presence/absence table used for UpSet graph generation. Relative abundance of each set was

calculated using the proportion table with ASVs collapsed per set and visualized as pie charts within the UpSet graph.

Co-correlation analysis was conducted to evaluate the impact of synthetic diets on microbial interaction networks using the SparCC procedure [61]. Networks were constructed separately for the synthetic diet group and the whole food diet group using sequence count tables that were filtered to only include ASVs with at least 5 representatives present in 25% of samples (synthetic: 17 samples; whole food: 15 samples), preventing spurious correlations from rare taxa. SparCC was implemented in R with standard parameters, and the resultant networks were characterized and analyzed with the igraph R package (version 1.5.1) [62, 63]. Networks were pruned to contain only edges with a correlation absolute value of at least 0.4 and exported into Cytoscape for visualization using the edge-weighted spring embedded layout method [64].

Results

Impacts of synthetic diets on gut microbiome diversity and community composition

We formulated a series of synthetic diets composed of a fixed base of 25% protein amended with dietary salts, vitamins, and cholesterol while differing only in complex carbohydrate type. Initial experiments utilized five alternative polysaccharide sources: chitin, methylcellulose, microcrystalline cellulose (MCC), pectin, or xylan. Following initial analysis of these results, we tested an additional starch-based diet. The prepared diets were readily consumed by the cockroaches in all cases.

To evaluate the impact of these diets on the gut community, each diet was fed to adult cockroaches (n=12/diet) for a period of 14 consecutive days, after which the insects were sacrificed, and their hindgut dissected out for 16S rRNA gene library sequencing. Following

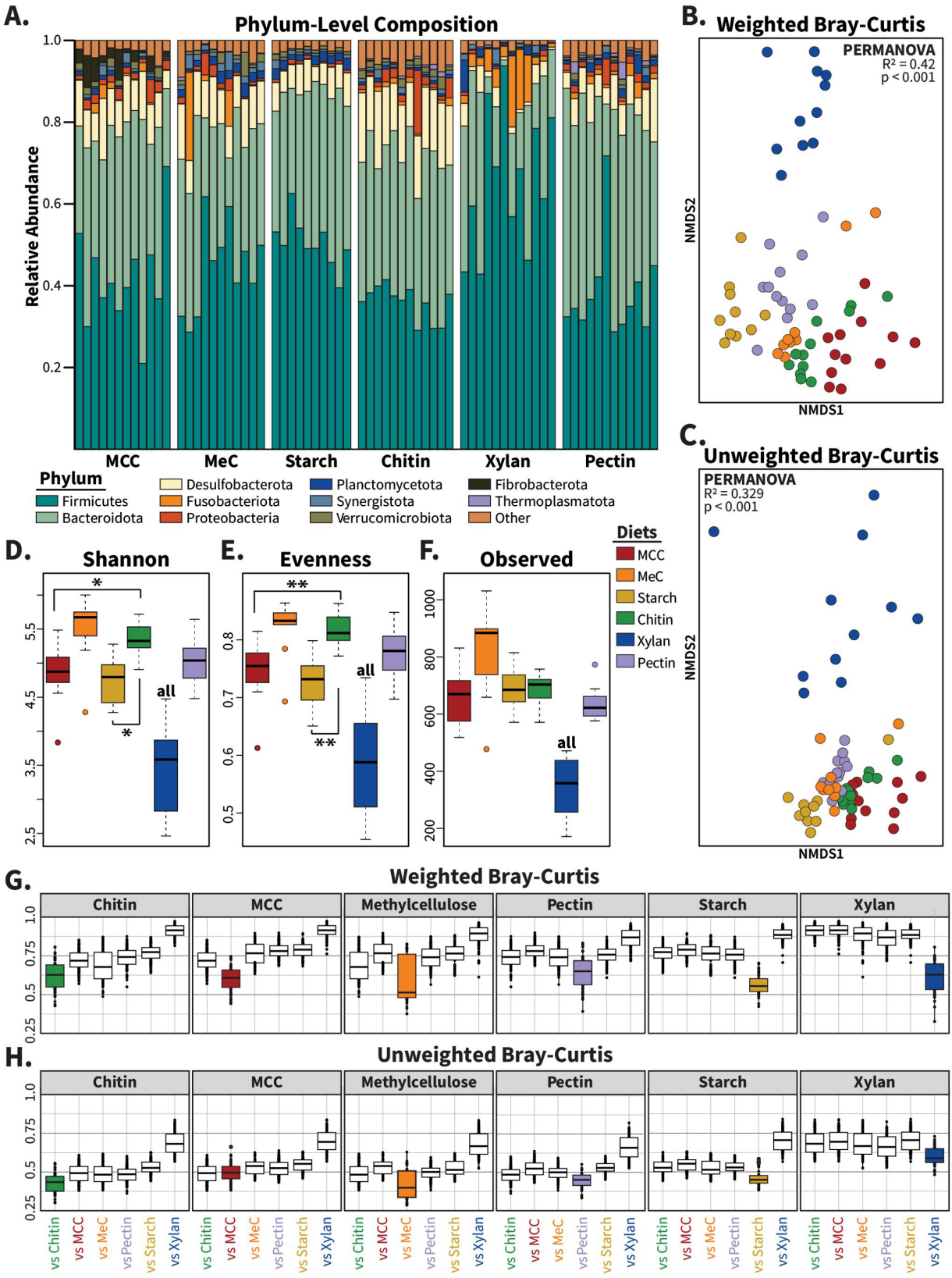

A.
Phylum-Level Composition
Relative Abundance
1.0
0.8
0.6
0.4
0.2
MCC
MeC
Starch
Chitin
Xylan
Pectin
Phylum
Firmicutes
Bacteroidota
Desulfobacterota
Fusobacteriota
Proteobacteria
Planctomycetota
Synergistota
Verrucomicrobiota
Fibrobacterota
Thermoplasmatota
Other
B.
Weighted Bray-Curtis
PERMANOVA
R² = 0.42
p < 0.001
NMDS2
NMDS1
C.
Unweighted Bray-Curtis
PERMANOVA
R² = 0.329
p < 0.001
NMDS2
NMDS1
D.
Shannon
E.
Evenness
F.
Observed
all
Diets
MCC
MeC
Starch
Chitin
Xylan
Pectin
G.
Weighted Bray-Curtis
Chitin
MCC
Methylcellulose
Pectin
Starch
Xylan
H.
Unweighted Bray-Curtis
vs Chitin
vs MCC
vs MeC
vs Pectin
vs Starch
vs Xylan

Figure 3.1: Composition of gut microbiomes from cockroaches fed synthetic diets. (**A**) Barplot showing the relative abundance of phyla across samples for each of the synthetic carbohydrate diets. Bars represent individual hindgut samples, clustered and labeled by diet polysaccharide source. Phyla present at an abundance greater than 1% in at least one sample are plotted. Non-metric multidimensional scaling (NMDS) was used to plot **(B)** weighted and **(C)** unweighted Bray-Curtis dissimilarity, with one point representing the community of one insect. The alpha diversity measures **(D)** Shannon index, **(E)** Pielou's evenness, and **(F)** number of observed taxa were plotted. Boxplots of **(G)** weighted and **(H)** unweighted Bray-Curtis display each diet vs self (colored boxes) and the other five synthetic diets (white boxes). Samples were rarefied to a constant depth of 7924 sequences for alpha and beta diversity calculations. For alpha diversity measures, pairwise statistics were calculated with Wilcoxon rank-sum tests and multivariate analysis was performed using Kruskal-Wallis tests. PERMANOVA was used to generate statistics for ordination analyses. "all" indicates $p<0.05$ vs all other diets; * = $p<0.05$; ** = $p<0.01$

library preparation and sequencing, we used DADA2 to obtain 2,321,848 quality-controlled, assembled sequences assigned to 3308 amplicon sequence variants (ASVs) after removal of endosymbiont (*Blattabacterium* sp.) and mitochondrial sequences [54]. At the phylum level, at least 80% of each sample was dominated by Bacteroidota, Firmicutes, and Desulfobacterota, in agreement with previous studies on the cockroach gut microbiome (**Figure 3.1A**) [27, 51, 65]. The relative abundances of these three phyla were similar across all samples excluding xylan-fed cockroaches; these insects hosted notably more Firmicutes and less Desulfobacterota than cockroaches fed other diets (**Figure 3.1A**).

Alpha diversity, as measured by Shannon index, evenness, and community richness significantly differed across diet treatments (**Figure 3.1D-F;** Kruskal Wallis $p< 0.001$ for each). Pairwise analyses found that chitin-fed insects possessed higher Shannon index values ($p < 0.05$) and community evenness ($p < 0.01$) than that of MCC- and starch-fed insects, while the xylan diet resulted in lower alpha diversity measures than all other diets ($p < 0.05$ for each).

Beta diversity analyses using weighted and unweighted Bray Curtis dissimilarity, which is also known as the Sørensen index, revealed significant impacts of our synthetic diets on gut microbiome composition. On average, between-diet variation was greater than within-diet

variation (**Figure 3.1G and 1H)**, with xylan-fed communities producing distinct communities compared to the other synthetic diets. Ordination analyses using non-metric multidimensional scaling (NMDS) and PERMANOVA analysis showed that samples clustered based on diet composition in both weighted (**Figure 3.1B;** PERMANOVA: R^2=0.42; p<0.001) and unweighted (**Figure 3.1C;** PERMANOVA: R^2=0.329; p<0.001) Bray-Curtis metrics, with especially clear separation of the xylan-based diet from other synthetic diets. Removing xylan-fed samples from diversity calculations did not eliminate diet-based clustering for weighted (**Figure S3.1A;** PERMANOVA: R^2=0.343; p<0.001) or unweighted (**Figure S3.1B;** PERMANOVA: R^2=0.247; p<0.001) measures, suggesting that each carbohydrate source enriched for a unique community composition.

Diet-characteristic taxa enriched by polysaccharide source

We used DESeq2 to identify 76 microbes that exhibited significantly higher abundance in a single synthetic diet across pairwise comparisons against all other treatments, which we termed "diet-characteristic taxa" (**Figure S3.2**) [59]. Diet-characteristic ASVs were primarily assigned to Firmicutes (n=48) and Bacteroidota (n=20); other phyla with diet-responsive taxa include Fusobacteriota, Deferribacterota, Desulfobacterota, Fibrobacterota, and Spirochaetota. We found that the chitin and methylcellulose diets were not associated with any diet-characteristic taxa by this definition, while diets made with xylan, MCC, starch, and pectin enriched for 45, 10, 13, and 8 ASVs respectively.

Cohort effects on diet-driven differences in gut microbiome composition

To confirm that the diet-associated gut community differences we observed are genuine rather than artifacts of natural variation in the insect colony ("cohort effects"), we prepared fresh MCC- and xylan-based diets and repeated the two-week diet experiment with a new cohort of

adult cockroaches. These diets were selected for follow-up experiments due to both their contrasting molecular structure and the dissimilarity they generated in Bray-Curtis analyses (**Figure 3.1B-C**). Data from the first and second experiments exhibited similar alpha diversity measurements (**Figure 3.2A-C**), with the repeated cohorts maintaining the significant shifts in alpha diversity ($p < 0.001$) observed in the initial experiment between these two diets while showing no difference between same-diet cohorts. Beta diversity analysis showed that samples clustered by both cohort and diet (**Figure 3.2D-E**). Diet had large effects on both weighted (PERMANOVA: $R^2=0.34$; $p<0.001$) and unweighted (PERMANOVA: $R^2=0.20$; $p<0.001$) Bray-Curtis dissimilarity, while cohort explained minimal effect sizes of 3.7% and 5.8% for weighted and unweighted measures respectively, with only unweighted reaching significance ($p < 0.01$). Because unweighted Bray-Curtis calculations only consider whether an ASV is present or not in an individual community, we hypothesized that the difference between cohorts was driven by low-abundance microbes that had lesser impact in the calculation of abundance-weighted beta diversity. Using Venn diagrams (**Figure 3.2F**), we confirmed that most ASVs recovered were in fact shared across cohorts. While 616 ASVs (27%) were unique to cohort 1 and 423 ASVs (20.2%) unique to cohort 2 (**Figure 3.2F**, grey pie slices), these ASVs represented only a small fraction of the total sequences (**Figure 3.2F**, black pie slices) obtained from each cohort. Further, 66.7% (cohort 1) and 61% (cohort 2) of these 'cohort-specific' taxa appeared in only one sample (**Figure S3.3**), indicating that most differences in composition due to time between studies stem from transient, rare taxa. Separating the diets for these comparisons confirmed the overall findings, with rare taxa contributing few sequencing reads despite comprising 19.7-33.7% of unique ASVs (**Figure 3.2F**). Altogether, these results show that synthetic diets reproducibly alter the gut microbiome composition in cockroaches.

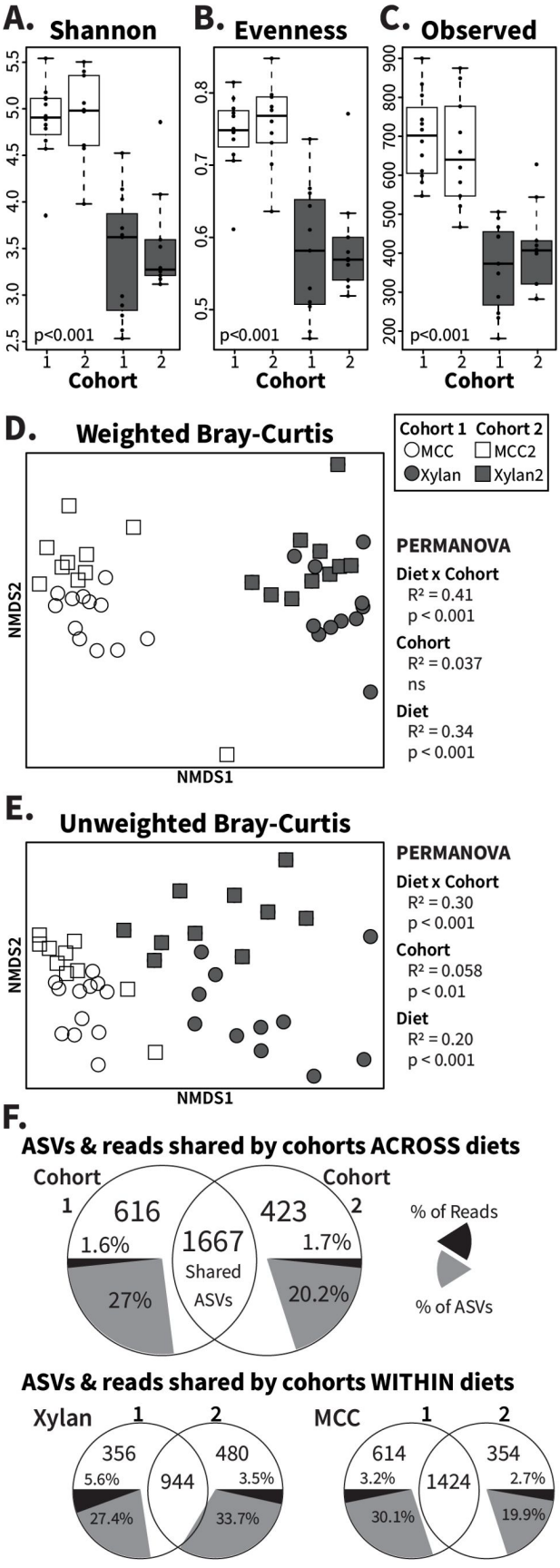


Figure 3.2: Analysis of cohort effects on gut microbiome responses to MCC and xylan diets. Xylan and MCC-fed samples from replicate experiments were rarefied to 9685 ASVs for alpha and beta diversity assessment. Boxplots show **(A)** Shannon index, **(B)** Pielou's evenness**,** and **(C)** number of observed ASVs with Kruskal-Wallis p-values calculated across all individual groups. PERMANOVA was used to calculate R^2 and p-values for diet ("MCC" and "Xylan"), cohort ("Cohort 1" and "Cohort 2"), and diet x cohort for NMDS ordinations of **(D)** weighted and **(E)** unweighted Bray-Curtis dissimilarity, with one point plotted per insect**.** The last panel **(F)** contains Venn diagrams of shared and unique ASVs between cohorts for both diets together as well as separately, constructed using rarefied count tables collapsed by diet and/or cohort. Grey pie slices represent the percent of ASVs observed that are cohort-unique, while black pie slices represent the percentage of sequence reads assigned to the indicated unique ASVs. ns =no significance

gut community, we leveraged our synthetic

<u>Testing the impact of alternative diet formulations</u>

While fiber is the primary component of undigested material that reaches hindgut microbiota, other macro- and micronutrients are known to influence gut community structure in non cockroach host systems [66-68]. To confirm the role of polysaccharides as key modulators of the community, we leveraged our synthetic diets to test the

impacts of xylan and MCC-based diets deficient in protein or micronutrients. For protein-deficient diets (**Table 3.1**), casein and peptone were replaced by mass with either xylan or MCC, while in vitamin-deficient diets, both the vitamin mixture and cholesterol were replaced with additional polysaccharide. We also created simple sugar versions of these two diets to test whether replacing long-chain fibers with their component backbone sugars results in different gut communities: xylose for comparison with xylan, and cellobiose and glucose for comparison with MCC.

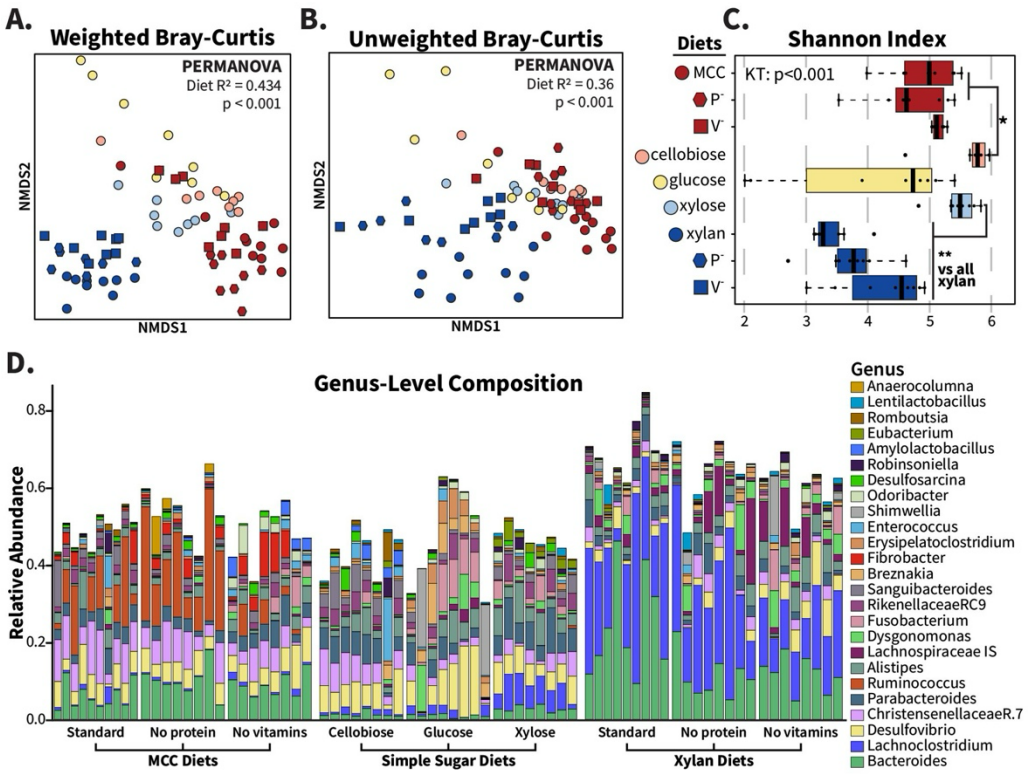


Figure 3.3: Fiber source, not protein, vitamins, or sugar composition, determines community structure from xylan and MCC-based synthetic diets. Deficient and simple sugar variations of MCC and xylan synthetic diets were fed to adult cockroaches for two weeks, and hindgut community compositions were compared with replicated xylan and MCC samples. For these analyses, samples were rarefied to 12274 reads and plotted with each point representing one individual. NMDS ordinations were made for **(A)** weighted and **(B)** unweighted Bray-Curtis dissimilarity, and PERMANOVA used to calculate R^2 and p-values with "diet" as the grouping factor. Alpha diversity is displayed via **(C)** Shannon index with Wilcoxon rank-sum test used for pairwise comparisons. The relative abundance of abundant genera found in the MCC- and xylan-based diets are visualized in **(D).** * =p <0.05; ** =p<0.01.

In unweighted Bray-Curtis analyses, which consider only presence/absence of ASVs (**Figure 3.3B),** communities from all MCC-fed insects overlapped with those of the sugar diets while retaining separation from xylan-fed samples (PERMANOVA: R^2=0.36; p<0.001) suggesting that MCC and the sugar diets supported a shared set of microbiota that are absent in xylan-fed insects. When abundance of ASVs is accounted for via weighted Bray-Curtis ordination (**Figure 3.3A)**, the sugar-fed communities formed their own distinct cluster, while both xylan-fed and MCC-fed samples clustered by polysaccharide regardless of vitamin or protein content (PERMANOVA: R^2=0.434; p<0.001). The alpha diversity profiles of the deficient diets matched the standard MCC or xylan diets as well, as measured by Shannon index (**Figure 3.3C**), evenness (**Figure S3.4A**), and number of observed ASVs (**Figure S3.4B**). The cellobiose communities displayed slightly higher Shannon index values than the standard and protein-deficient MCC communities, while xylose-fed insects possessed noticeably more even and diverse communities than the xylan diets. At the genus level, (**Figure 3.3D**) the community composition of insects fed diets containing the same polysaccharide resembled each other, while sugar-fed microbiota reflected each other more than the polysaccharide they are derived from. Despite xylose-fed samples clustering with the other sugar diets, the pentose did enrich for an abundant *Lachnoclostridium* ASV (**Figure S3.5**) that is heavily associated with xylan, while MCC-associated taxa such as *Fibrobacter* and *Ruminococcus* remained at low levels in the glucose and cellobiose diets. Overall, these results suggest that the communities observed in the fiber diets are driven by the long-chain structures of the polysaccharides rather than their component sugars.

Comparison with whole-food diets

The different microbial communities triggered by our synthetic diets were unexpected given that previous experiments examining the impact of whole food diets with strongly differing macronutrient profiles did not produce substantially different gut microbiome compositions [27]. Therefore, we compared the samples from this current study ("synthetic" diet type) to samples from the previous study ("whole food" diet type) of cockroaches fed butter, tuna, honey, white flour, or whole wheat flour [27]. Both studies fed diet treatments *ad libitum* to groups of adult mixed-sex American cockroaches for two weeks, following the experimental set up described in Methods section 3.3.

We found that gut microbiome samples from cockroaches fed synthetic diets exhibited higher ASV richness ($p<0.01$) but lower evenness ($p<0.001$) and Shannon index ($p<0.01$) than those from insects fed whole foods (**Figure S3.6A-C**). Synthetic and whole food diets produced distinct diet type clusters in NMDS ordinations (**Figure 3.4A-B**) for weighted (PERMANOVA: $R^2=0.105$; $p<0.001$) and unweighted (PERMANOVA: $R^2=0.152$; $p<0.001$) Bray-Curtis dissimilarities. When we analyzed the samples by diet, we found diet explained more variation in NMDS ordination than diet type as interpreted from PERMANOVA R^2 values (weighted: $R^2=0.393$; unweighted: $R^2=0.369$).

Beta dispersion analysis of variation within diet types showed that, together, the gut microbiota of cockroaches fed synthetic diets was more variable than that observed among whole food fed cockroaches (**Figure S3.7A-B**; Tukey's HSD: $p<0.001$). However, when diets were analyzed individually, they were equally dispersed in weighted Bray-Curtis dissimilarity (**Figure S3.7A)** but not unweighted measures **(Figure S3.7B;** ANOVA: $p<0.001$). Xylan-fed cockroaches exhibited significantly greater within-

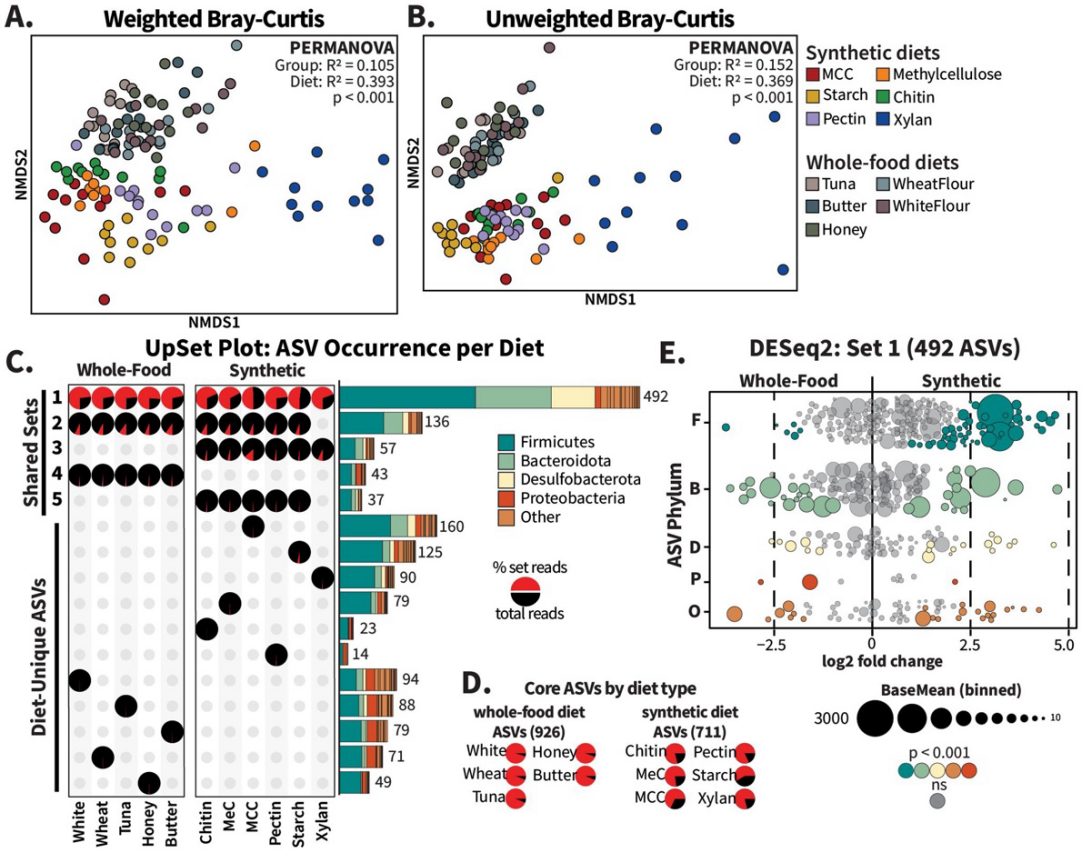


Figure 3.4: Whole-food diets share more ASVs than synthetic diets. Raw sequence data from Tinker and Ottesen (2016) for cockroaches fed tuna, butter, honey, wheat flour, and white flour ("Whole Food Diets") were reprocessed using the methods in this experiment to generate comparable ASVs. For beta diversity comparison, all samples were rarefied to 7924 ASVs and NMDS ordinations, with one point per insect, were generated for **(A)** weighted and **(B)** unweighted Bray-Curtis distances; R^2 and p-values for diet type comparisons were calculated using PERMANOVA. For **(C)** UpSet plot analysis, the five largest intersections (Sets 1-5) and diet-unique sets are displayed. Pie charts represent the percent of reads within a diet that originate from each set (red slices), and the bar charts are colored to display the phylum-level distribution of ASVs assigned to each set. Read abundance of core ASVs, or those present in all synthetic or whole food diets regardless of presence in the other diet type, are visualized in **(D)** pie charts per diet. For the MA plot in panel **(E)**, raw sequence count tables for the ASVs identified as "Set 1" were analyzed using DESeq2 with diet type as the design factor. The ASV circles are scaled according to baseMean size and colored by phylum.

group variability than all ten other diets (Tukey's HSD range: $p = 0.037 – 3.17e^{-06}$), with no significant differences observed in pairwise comparisons of all other diets. When synthetic diets were compared to whole food diets without including xylan-fed samples, we observed no significant differences in unweighted beta dispersion (**Figure S3.7D**) or Shannon index values

(**Figure S3.6B, red boxes**). However, significant differences remained in weighted Bray-Curtis dispersion (**Figure S3.7C**), richness and evenness (**Figure S3.6A, 3.6C**), and in both weighted and unweighted Bray-Curtis ordination analysis **(Figure S3.8A-B)**, highlighting that the altered microbiomes produced by the synthetic diet type were not solely due to biases produced by xylan-fed samples.

To verify that inclusion of whole food dietary components alone was not sufficient to eliminate fiber-dependent gut microbiome configurations, we tested the impact of diets mimicking our synthetic diets but with the purified protein components replaced with canned tuna. These diets induced community compositions similar to those observed in polysaccharide-matched diets containing purified proteins rather than supporting protein-associated communities (**Figure S3.9**). Xylan-containing diets generally produced communities with lower alpha diversity scores (**Figure S3.9D-F**) and clustered away from MCC-containing diets and dog chow-fed insects we included as controls in weighted (**Figure S3.9B;** PERMANOVA: R^2=0.395; p<0.001) and unweighted (**Figure S3.9C;** PERMANOVA: R^2=0.392; p<0.001) analyses. Despite the discordant structural complexities between tuna fish and purified casein/peptone amino acids, the protein portion of the synthetic diets exerted less influence than polysaccharide source.

Core taxa differences between synthetic and whole-food diets

Given these strong differences in community structure, we utilized the R package UpSetR to determine how the ASVs in different diet types overlap [60]. UpSet plots are akin to Venn diagrams, considering only presence/absence of an ASV. Rarefied count tables were aggregated by diet and ASVs were marked as either present or absent per diet. ASVs present in the same subset of diets were grouped into “Sets”, with the phylum-level composition per set

depicted as stacked bar charts labelled with the number of included ASVs (**Figure 3.4C**). We supplemented the UpSet plot with pie charts illustrating the relative abundance (calculated as the fraction of total reads recovered from the collapsed treatment group) of reads assigned to ASVs within each set (**Figure 3.4A**), in addition to pie charts representing "core" ASVs present in all whole-food or all synthetic diets, regardless of presence in the other dietary group (**Figure 3.4D**). For simplicity's sake, **Figure 3.4** and **Supplemental File S1** shows only the five largest intersecting sets as well as all single-diet sets; additional sets are presented in **Figure S3.10**.

A total of 492 ASVs ("Set 1") were shared across all diet treatments (**Figure 3.4C**). These ASVs made up over half of the sequences recovered for all diets except the MCC diet, for which they represented 49% of sequences (**Figure 3.4C,** pie charts). Only 43 ASVs ("Set 4") were exclusive to the whole food diets, contributing between 0.9% and 1.65% of reads in these diet sets. The 57 ASVs ("Set 3") identified as exclusive to synthetic diets made up 1.6-3.4% of the reads recovered from starch-, pectin-, chitin-, and methylcellulose-fed cockroaches, and 7% of xylan- and 13% of MCC- fed cockroaches. Together, these results argue that the synthetic diets did not eliminate core taxa present in the guts of cockroaches fed whole foods, nor did they result in hindgut colonization by a large new set of microbial taxa.

Similarly, individual synthetic diets were not associated with hindgut colonization by large groups of unique microbes. In general, taxa that were unique to individual synthetic or whole food diets represented a very small proportion of sequences recovered (**Figure 3.4C**). ASVs unique to the MCC diet formed the second largest set overall, with 160 diet-unique taxa, yet they only represented 0.4% of total recovered sequences (**Figure 3.4C**). A xylan-based diet, which repeatedly produced the largest community dissimilarities (**Figure 3.4A-B; Figure S3.7**), was associated with 90 diet-specific taxa comprising only 0.63% of reads. Interestingly, our

analysis revealed that a xylan-based diet did result in the loss of 136 taxa that were present in all other diets in abundances ranging from 4.78%-10.14% ("Set 2" in **Figure 3.4C**). However, other sets that excluded individual diets were substantially smaller (**Figure S3.10A**) suggesting that this was not a common mechanism underlying the diet-driven differences in gut microbiome composition. Instead, synthetic diet-driven differences in gut microbiome composition were primarily associated with high relative abundance of individual taxa that were consistently present in the cockroach gut regardless of dietary treatment.

To follow up on these observations, we used DESeq2 to assess enrichment of individual ASVs between synthetic vs whole food fed cockroaches. Of the 492 ASVs found, often at high abundances, in all 11 diets, 95 ASVs were significantly enriched in cockroaches fed synthetic diets while 38 ASVs were enriched in cockroaches fed whole food diets (**Figure 3.4E**; padj < 0.001). The magnitude of these ASV-level differences across diet type were modest (log fold change <5), consistent with the high proportion of reads recovered in all diets that belonged to Set 1 (**Figure 3.4C**). Bacteroidota, Desulfobacterota, and other phyla showed similar enrichment distributions to each other in terms of both number of ASVs and abundance within samples, but the 49 Firmicutes enriched in the synthetic diets included far higher individual abundances than the six enriched in the whole food diets.

We also examined enrichment of ASVs that fell outside of Set 1 but were both somewhat abundant (baseMean > 1) and present in at least 5 samples (**Figure S3.11**). These 1270 ASVs generally had smaller abundances than members of Set 1 (**Figure 3.4E**), but greater log fold changes between diet types. The ASVs *Ruminococcus_NA* and *Fusobacterium ulcerans* were exceptions, being both highly abundant (baseMean = 1020 and 651, respectively) and unique to the synthetic diets. In contrast, the *Christensenellaceae* ASV (R7_NA.68) unique to whole food

diets had high log-fold change, but a baseMean of only 4.29. Apart from a few highly abundant ASVs in the synthetic diets, most abundant, diet-enriched taxa were shared by all diets, supporting that change is driven by common gut bacteria restructuring the community rather than interloping bacteria disrupting the community.

Differential diet-based fluctuations of abundant Firmicutes and Bacteroidota ASVs

To explore the impact of diet on microbial taxa associated with fiber degradation, we evaluated dietary responses of abundant taxa within the Firmicutes and Bacteroidota. For this analysis, we selected the two most abundant representatives of each of these phyla from each diet and examined their abundance across all diets (**Figure 3.5**). We identified 16 Firmicutes and 11 Bacteroidota as most or second-most abundant in at least one diet.

The most abundant taxa from both Firmicutes and Bacteroidota represented a small fraction of reads across the whole foods diets, consistent with higher Shannon diversity and evenness in gut communities of whole food-fed cockroaches (**Figure S3.6B, C**). In contrast, synthetic diets produced strong 'blooms' of individual ASVs, particularly among Firmicutes (**Figure 3.5A**). Several abundant Firmicutes were both present in Set 1 (**Figure 3.4C**) and enriched in individual synthetic diets (**Figure S3.2**), namely *Lachnoclostridium_NA* (xylan), *Lachnoclostridium_NA.1* (xylan), *Enterococcus_NA* (xylan), *Enterococcaceae_NA* (xylan), and *ChristensenellaceaeR7_NA* (pectin). In contrast, Firmicutes enriched in MCC and Starch diets were not typically found across all diets: *Ruminococcus_NA* (MCC), *Lachnospiraceae_NA.5* (starch), and *Ruminococcaceae_NA* (starch). Among Bacteroidota (**Figure 3.5B**), we observed greater overlap in the most abundant taxa present in each diet group, with all but one (*3M1PL1.52termite_NA*) of the abundant Bacteroidota ASVs belonging to Set 1, and only two classified as diet-characteristic: *3M1PL1.52termite_NA* (MCC) and *Bacteroides_NA.12* (pectin).

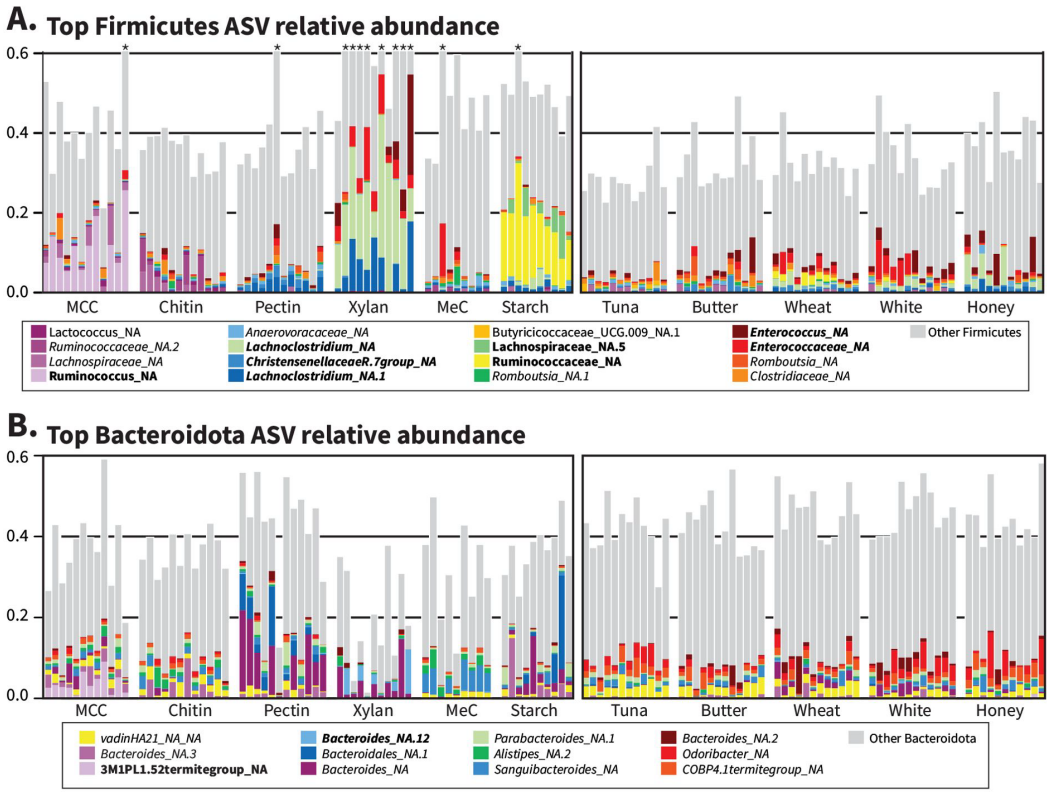


Figure 3.5. Individual ASVs explain large differences between synthetic but not whole-food diets. Variance-stabilized count data from DESeq2 were used to determine the top two ASVs for every diet belonging to **(A)** Firmicutes and **(B)** Bacteroidota, and the relative abundances of ASVs belonging to the combined 'top ASV' set were plotted for all individual samples. Grey bars include all Firmicutes or Bacteroidota not named in the key. * indicates "Other Firmicutes" extends beyond 60% relative abundance; please refer to **Figure 3.1A** for full values. Bolded names indicate diet-characteristic ASVs from **Figure S3.2**, and italics indicate Set 1 ASVs from **Figure 3.4C**.

Analysis of microbial co-correlation networks

We constructed co-correlation networks with SparCC to examine the community structure underlying synthetic and whole food microbiome data sets [61]. To filter out noise and reduce spurious correlations, datasets were filtered to include ASVs present in at least 25% of samples for each diet type, resulting in 976 ASVs for whole food diets and 700 ASVs for synthetic diets. Networks were further pruned to contain edge weights with absolute values larger than 0.4, removing 75 and 168 ASVs from whole food and synthetic networks respectively. Both positive and negative edges were retained for network layout formation

(**Figure S3.12**), but for analysis, only positive edges were considered (**Figure 3.6**). After negative edge removal, the whole food network contained 875 ASVs with 9515 edges forming two connected components (**Figure 3.6A**), while the synthetic diet network contained 497 ASVs with 2536 edges that formed six connected components (**Figure 3.6B**). Networks at SparCC correlation levels of 0.5, 0.6, and 0.7 are included in **Figure S3.13**.

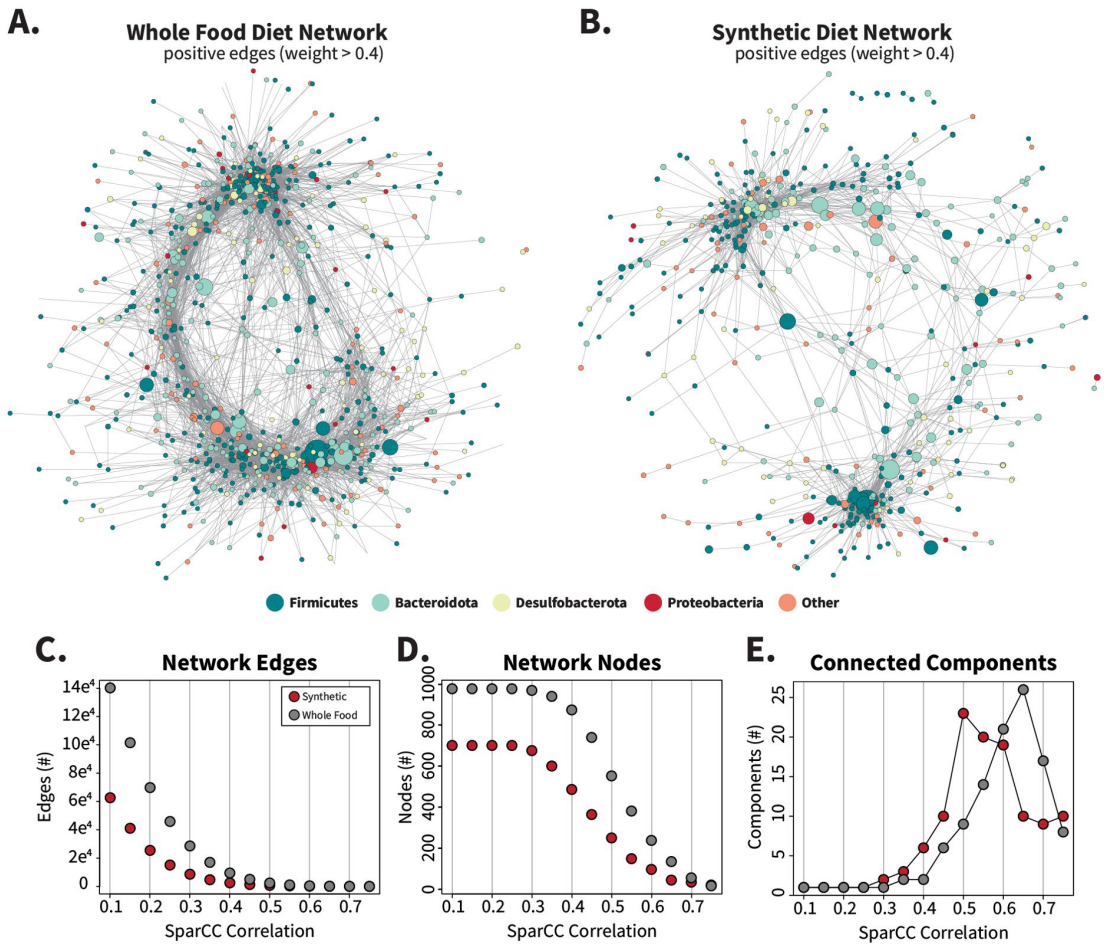


Figure 3.6: Synthetic diet correlation networks are smaller and less interconnected than whole food diet correlations networks. Networks were calculated by SparCC from filtered count tables (ASVs present in > 25% of samples per diet set) for synthetic and whole food diets separately to create two distinct networks containing 976 (whole food) or 700 (synthetic) nodes. Networks were imported into Cytoscape and edges with absolute values < 0.4 were removed to generate panels **(A)** and **(B)**. Negative edges were included during initial layout generation with the edge-weighted spring embedded layout method and are displayed in **Figure S3.12.** The number of nodes, edges, and connected components that remain when the networks are filtered by increasing correlation values are charted in **C**, **D,** and **E** respectively.

Gut microbiota from whole food-fed cockroaches formed an extensive and dense interaction network (**Figure 3.6A**), with higher edge counts (**Figure 3.6C**), node counts (**Figure 3.6D**), node degree (**Figure S3.14A**) network strength (**Figure S3.14B**), and betweenness scores (**Figure S3.14C-D**) than synthetic diets at most levels of filtering based off SparCC correlation values. High-degree nodes, or ASVs with large numbers of neighbors, were present throughout the whole-food network structure, while in the synthetic diets two primary clusters of ASVs appeared with fewer connecting ASVS. When a range of inclusion cutoffs were considered, the synthetic diet network degraded more quickly than the whole food network into separate connected components (**Figure 3.6E, Figure S3.13)** displaying greater fragility and tendency to fragment. These overall network structures suggest that the synthetic diets disrupt the stability of the cockroach microbiome.

Discussion

The core finding of this work is that synthetic diets featuring a single purified complex polysaccharide source induced distinctive and fiber-dependent hindgut microbiome compositions in omnivorous cockroaches. Three of our polysaccharides are abundant structural components in the plant cell wall: cellulose (β-1,4-linked glucose), xylan (β-1,4-linked xylose), and pectin (α-linked galacturonic acid and/or rhamnose). In a typical plant cell, chains of cellulose interface with hemicellulose (xylan or xyloglucan: β-1,4-linked glucose with α-1,6-linked xylose residues) to form a rigid scaffold interspersed by a pectic polysaccharide gel matrix, with these components fortified via hydrogen bonding to create the stable cell wall [69]. In contrast to cell wall polysaccharides, which are expected to require bacterial fermentation for efficient degradation, the polysaccharides starch (amylose: straight chain of α-1,4-linked glucose; amylopectin: branched chain of α-1,4 and α-1,6-linked glucose) and chitin (*N*-acetyl-β-D-

glucosamine chain) are more easily digestible to the cockroach. Starch can be digested by both salivary and midgut-derived α-amylases to provide energy, although the efficiency of these enzymes depends on the amylose/amylopectin makeup of the starch granules. Chitin is a key component of the insect cuticle that is recycled through consumption of the exuviae after molting as well as consumed during cannibalism [70], but the level of chitinase activity in the midgut itself is poorly quantified. Methylcellulose is the only compound tested in this study that is not a naturally occurring polysaccharide; rather, it is a synthetically modified cellulose with an average of 1.8 hydroxyl groups per glucose residue replaced by methoxide. It was selected as a water-soluble cellulose derivative that is also commonly used as an "inert" emulsifier and can act as a laxative.

Key changes observed in this study include alterations in the abundance of multiple organisms, but especially bacteria from the phyla Bacteroidota and Firmicutes (**Figure 3.5**). Members of the Bacteroidota are thought to be key fiber-degrading organisms in the cockroach, supported by the many carbohydrate active enzymes (CAZymes) encoded both within and independently of polysaccharide utilization loci (PULs) in multiple isolates' genomes [65, 71]. The high gene density of CAZymes likely equip Bacteroidota members to efficiently ferment diverse fibers without exclusively relying on a single source [65, 71]. While Firmicutes comprise a large proportion of the bacteria in the cockroach gut, their roles in fiber degradation remain elusive due in part to their extensive genomic diversity [65], and their sparse roles in polysaccharide fermentation in the cockroach's termite relatives [72]. In ruminant research, the Firmicutes species *Ruminococcus flavefaciens* and *Ruminococcus albus* have been extensively studied for their powerful cellulolytic capabilities, [73, 74]. However, in the cockroach and in humans, fiber-degrading Firmicutes are characterized by lower gene densities of CAZymes than

Bacteroidota, suggesting they function as secondary polysaccharide fermenters that scavenge materials released from cell matrices by generalist Bacteroidota [65, 75, 76].

Interestingly, of the six polysaccharides used in this study, only the four plant-derived polysaccharides (MCC, xylan, starch, and pectin) were associated with diet-characteristic ASVs (**Figure S3.2**). The chitin and methylcellulose diets, in contrast, produced communities with the highest alpha diversity values (**Figure 3.1D-E**) and clustered together in beta diversity analyses (**Figures 3.1B-C, S3.1**) closest to the whole food diets (**Figure 3.4A**). Chitin is present in the cockroach hindgut even in the absence of a dietary source due to the continuous shedding of the PM from the midgut, which may explain why no unique organisms were selected for in this diet condition. Methylcellulose, on the other hand, may resemble fiber starvation from the perspective of the gut bacteria, which has previously been shown not to alter the gut microbiota [27]. Methylcellulose has been found to reduce adhesion and inhibit cellulase (but not cellobiase) activity in the rumen bacteria *R. albus*, *R. flavefaciens*, and *Fibrobacter succinogenes* [77, 78]. The extent to which this polymer interacts with other fibrolytic systems, such as those in Bacteroidota, is unclear, but our results suggest that it does not select for a unique set of gut microbes.

Among the polysaccharides tested, the hemicellulose xylan induced the largest shifts in alpha diversity, inter-individual variability, and overall community composition (**Figure 3.1**). Xylans are abundant heteropolysaccharides that vary in branch complexity according to the source they are derived from, ultimately influencing its digestibility for hindgut and rumen microbiota [79-81]. The xylan in this study is derived from corn cob and contains residues of galactose, arabinose, and glucuronic acid with low levels of acetylation [82]. Research performed *in vitro* investigating the xylan degradation ability of gut microbiota mainly focuses

on *Bacteroides* [83-87], although clostridial organisms such as *Roseburia intestinalis* and the rumen bacterium *Butyrivibrio fibrisolvens* have also been identified as key butyrate-producing xylan fermenters [79, 88-90]. Interestingly, we found that in cockroaches, xylan-based diets decreased the relative abundance of Bacteroidota, while increasing Firmicutes (**Figure 3.1A**). Concurrent with this enrichment of specific taxa, xylan decreased the overall diversity of the gut community (**Figure 3.1**); in contrast, feeding the monomer xylose to cockroaches retained high community diversity while still selecting for a *Lachnoclostridium* ASV that was substantially enriched in the xylan diet (**Figure 3.3, S3.5**).

Direct comparison of these results with other *in vivo* dietary studies is difficult, as livestock studies utilizing it as a dietary additive frequently produced harmful effects such as lower ileal digestibility of essential amino acids in pigs and proliferation of pathogens in broilers (reviewed in [91]). However, an abundance of research has been performed analyzing effects of xylan-containing whole foods and its derivatives on gut microbiota [92-95], with notable enrichment of *Lachnospiraceae* species on both xylo-oligosaccharides (XOS) and xylan-containing whole foods [96-99]. Co-culture assays performed using commensal *Bacteroides* and *R. intestinalis* identified different transporter affinities for xylan degradation products based on XOS size [90], while studies using both purified hemicellulose and intact forage found that *R. flavefaciens* effectively converted some xylans to acid-soluble forms but required co-culture with *B. fibrisolvens* to grow [80, 81]. The purified xylan used in this work seems to select for Firmicutes with specialized xylan degradation machinery. This advantage was minimized in the xylose diet, allowing other microbiota to grow concurrent with the enriched *Lachnoclostridium* ASV.

While xylan-based diets induced the largest differences in hindgut community composition, samples clustered by dietary treatment even when the xylan treatment group was excluded (**Figure S3.1)**. These results stand in stark contrast to previous work from multiple investigators, who found minimal to no differences in hindgut microbial community composition in response to diet alterations [27, 36, 37]. A commonality between these experiments is that the investigators utilized whole food diets or animal feeds containing processed complex plant material such as milled bran or soymeal. On the other hand, other investigators have observed substantial influence of diet on the gut microbiome composition [38-40]. These experiments all utilized synthetic diets that contained purified, lab-generated carbohydrate and protein sources without the undefined cell matrix components that are retained in "whole food" or animal feed diets. For example, in experiments using *B. germanica*, Pérez-Cobas, Maiques [39] prepared synthetic diets with a dextrin and micronutrient base amended with either 50% cellulose or 50% casein while Zhu, Wu [40] used diets composed of a cellulose and micronutrient base with supplemented with 40% by mass purified starch, casein, or sesame oil. In *P. americana*, Bertino-Grimaldi, Medeiros [38] utilized purified cellulose to compare with sugarcane bagasse, a complex dietary substrate. Given these conclusions and the findings described in our study, it appears that synthetic diets combined with the selective ability of purified fibers can produce marked differences in the cockroach gut community, although the paucity of cockroach studies with standardized dietary methods limits the strength of these conclusions. Future research is required to conclusively place weight on the purified nature of these diets in the context of cockroach gut microbiota. In addition, we note that the lack of compositional differences does not preclude functional differences resulting from changes in microbial activity, as reported in Schauer et al. (2014) and DePoy, Wall [100].

Looking beyond cockroach models, there appears to be a similar influence of synthetic diets containing highly purified components on gut microbial composition in numerous insect and mammalian studies. Termites, a close relative to cockroaches, responded to single carbohydrate source diets with larger alterations in gut community composition than termites fed mixed-carbohydrate diets [101]. Other insect model systems produced similar results, such as in silkworms [102], ladybugs [103], waxworms [104], and honeybees [105]. Among mammals, dogs provided a purified diet also exhibited reduced alpha diversity compared to those fed a complex diet [106], wild-caught mice transitioned from natural diets to laboratory diets lost large portions of native gut microbes [107], and humans given meal replacement shakes showed loss of biodiversity in their microbiome compositions [108].

Comparison of our dataset with data recovered in previous whole food-based dietary experiments suggest that synthetic diets altered the gut community by inducing overgrowth of microbes already present in the cockroach gut microbiome (**Figure 3.4),** a similar outcome to *in vitro* enrichment one may perform on selective media. Taxa that were unique to individual diets represented <1% of sequence reads in all diets but starch, of which they made up 2.84% (**Figure 3.4**). In contrast, 15 out of 20 highly abundant Firmicutes and Bacteroidota associated with one or more diets were shared across all diet types, while the remaining 5 were found sporadically in other diets (**Figure 3.5)**. The alterations we observed are especially associated with the fibers themselves rather than the other dietary components. Experiments leveraging the xylan and MCC diets without amino acids (**Figure 3.3**) or with tuna (**Figure S3.9**), a complex food, substituted for the casein/peptone mixture of our standard diet configuration largely did not differ from the polysaccharide-associated communities we observed initially. However, small scale changes such as the loss of *Fibrobacter* in MCC when vitamins/cholesterol are removed

provide compelling reasons to study the influence of dietary micronutrients on the gut microbiome in future work.

We hypothesize that highly purified synthetic diets enabled microbial 'specialists' to bloom beyond their former constraints in the whole food diets. The high homogeneity of purified fibers may allow these microbes to grow rapidly without needing to wait for the release of pure polysaccharides from cell matrix degraders, thereby reducing gut microbiome stability. The purified nutrients used in this study differ from "whole foods" in two primary ways: macro/micromolecular composition, and physical accessibility to bacterial degradation. Compositionally, the whole food diets used in Tinker and Ottesen [27] were mostly natural foods that, while highly biased in macronutrients, may have had a more diverse nutritional profile according to the "eye" of a bacterium [15, 84, 109]. For example, the whole and white wheat flour are composed predominantly of endosperm-derived starch but also contain portions of bran and germ, which have structural polysaccharides and bioactive phytochemicals that are targeted by gut bacteria and influence health parameters of the host [41, 110-112]. Honey contains complex mixtures of sugars (glucose, fructose, disaccharides) and fructooligosaccharides (FOS) in addition to organic acids, nitrogenous compounds, vitamins, and bee-derived enzymes [113]. Tuna and butter are similarly high-complexity substrates that offer resident gut microbes diverse metabolizable compounds that are lost in purified dietary components such as those utilized in our study. The purified components used in our synthetic diets are not entirely homogeneous (casein and peptone were used rather than individual amino acids), but as lab-quality reagents their extraction methods remove the bioactive compounds present basally in the source material, while all whole food diets retain some of their source material complexity [15]. Another factor that may contribute to the differences observed between whole and synthetic foods could be level

of processing prior to host feeding. The flours in particular underwent more processing than the other whole foods due to milling, which is known to influence microbial adhesion depending on the resultant particle size [114, 115]. However, even these diets did not shift the gut community composition in the original study [27], while the starch diet used in this study heavily enriched for an unclassified *Ruminococcaceae* (**Figure 3.5A)**. Both the flours and the starch synthetic diet contain approximately 70% starch by weight and are both finely ground substrates, yet the flours did not contain a single ASV with a relative abundance greater than 3% compared to the starch-associated *Ruminococcaceae* relative abundance of 15%. Although the short 16S rRNA gene region used here only starch-enriched cockroach gut ASVs at the family level, human-associated *Ruminococcus bromii* are established as effective degraders of resistant starch that distribute released glucose rather than utilize it themselves [116-119]. The extent to which the starch-associated *Ruminococcaceae* bloomed suggests strong selective enrichment of starch-specialized microbes, although a similar glucose cross-feeding relationship may explain in part why gut community alpha diversity remained high in the starch diet relative to the xylan diet.

Our hypothesis that synthetic diets with purified fibers reducing the need for cooperative metabolism of dietary fiber is supported by our microbial co-occurrence network analysis (**Figure 3.6**), where the whole food network is highly interconnected with numerous significant co-occurrence relationships between ASVs, while the synthetic diet network is easily fragmented into modules of microbes that are weakly or negatively associated with the other network members (**Figure S3.13, Figure S3.14**). Under this hypothesis, when compared to the rich landscape of intrinsic fibers found in whole foods, synthetic diets contain simpler macromolecular structures that may streamline the microbial enzymatic processes of fiber catabolism. This reduction of enzymatic requirements may in turn enable individual fiber

specialists, who possess all or most of the necessary machinery, to metabolize large amounts of these purified fibers without aid from other microbes. While some direct cross-feeding relationships are expected to remain, the loss of metabolite diversity may fragment the more nebulous cross-feeding relationships, therefore pruning the number of significant co-occurrence relationships among gut microbes to produce the network presented in this study.

A key limitation of this study is the fact that the comparison group of "whole food" fed cockroaches were from an earlier experiment and we lack contemporaneous controls fed whole food diets. However, an examination of cohort effects suggest that observed responses to synthetic diets were highly conserved across cohorts in experiments conducted one year apart (**Figure 3.2**). Additional caveats to this work regard the purified components used. The original source of a compound can be difficult to identify and may impact the fine structure of the compound despite it appearing comparable to one from a different source. Our source of starch, for example, was derived from potato, which produces higher resistant starch levels than other starches [120]. The xylan used in this study is highly soluble and may produce a different gut community than if we had used oat or birch xylan. While these caveats limit some of the conclusions that can be formed, it further highlights the utility of the cockroach to perform these studies.

Overall, this study showed that synthetic diets that were highly enriched in a single polysaccharide can produce divergent gut microbiome compositions in the American cockroach, which has previously been shown to be highly resistant to diet-induced differences in gut microbiome composition [27]. The individual polysaccharides featured in the different synthetic diets were associated with diet-specific 'blooms' of native Firmicutes and Bacteroidota rather than the introduction of new microbial specialists into the community. The enrichment of these

ASVs lead to fragmented gut microbiota co-occurrence networks with increased inter-individual variability among insects. Together, these results suggest that overconsumption of a single, purified class of polysaccharides can have destabilizing effects on cockroach gut microbiota. This work highlights the use of omnivorous cockroaches and synthetic diets as an *in vivo* enrichment culture system to pinpoint microbial responses to highly processed dietary ingredients while remaining within the context of a host-microbe system, thus facilitating the isolation and improved characterization of novel gut symbionts that are passed over in traditional benchtop microbiology. Future work will examine the functional and metabolic basis of these alternate microbial community compositions and will further explore the ways in which diet complexity and composition impacts gut microbiome homeostasis.

Data Availability

Data associated with this study are available from the NCBI short-read database under BioProjects PRJNA1096047 and PRJNA1105088. Data associated with the earlier study of the impact of whole food diets on cockroaches is available under BioProject accession PRJNA320546.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under award number R35GM133789.

Acknowledgements

We thank Darian Talamantes for assistance with initial diet formulation, and Dr. Kara Tinker for performing the foundational work serving as the basis of this experiment. We also thank Sarah Beth Griffin for her assistance with cockroach husbandry, dissections, and preparing samples for DNA sequencing. A previous version of this manuscript was submitted as a preprint under the

title “Synthetic diets containing a single polysaccharide disrupt gut microbial community structure and microbial interaction networks in the American cockroach” [121].

References

1. Huttenhower, C., et al., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. **486**(7402): p. 207-214.
2. Belkaid, Y. and O.J. Harrison, *Homeostatic immunity and the microbiota.* Immunity, 2017. **46**(4): p. 562-576.
3. Thaiss, C.A., et al., *The microbiome and innate immunity.* Nature, 2016. **535**(7610): p. 65-74.
4. Bäckhed, F., et al., *Host-bacterial mutualism in the human intestine.* Science, 2005. **307**(5717): p. 1915-20.
5. Sonnenburg, E.D., et al., *Diet-induced extinctions in the gut microbiota compound over generations.* Nature, 2016. **529**(7585): p. 212-5.
6. Kurilshikov, A., et al., *Large-scale association analyses identify host factors influencing human gut microbiome composition.* Nat Genet, 2021. **53**(2): p. 156-165.
7. Goodrich, Julia K., et al., *Human Genetics Shape the Gut Microbiome.* Cell, 2014. **159**(4): p. 789-799.
8. Kurilshikov, A., et al., *Host Genetics and Gut Microbiome: Challenges and Perspectives.* Trends Immunol, 2017. **38**(9): p. 633-647.
9. Greene, L.K., et al., *A role for gut microbiota in host niche differentiation.* Isme j, 2020. **14**(7): p. 1675-1687.
10. Wolter, M., et al., *Leveraging diet to engineer the gut microbiome.* Nat Rev Gastroenterol Hepatol, 2021. **18**(12): p. 885-902.
11. Tanes, C., et al., *Role of dietary fiber in the recovery of the human gut microbiome and its metabolome.* Cell Host Microbe, 2021. **29**(3): p. 394-407.e5.

12. Walker, A.W., et al., *Dominant and diet-responsive groups of bacteria within the human colonic microbiota.* Isme j, 2011. **5**(2): p. 220-30.

13. Rastall, R.A., et al., *Structure and function of non-digestible carbohydrates in the gut microbiome.* Benef Microbes, 2022. **13**(2): p. 95-168.

14. Tuncil, Y.E., et al., *Subtle Variations in Dietary-Fiber Fine Structure Differentially Influence the Composition and Metabolic Function of Gut Microbiota.* mSphere, 2020. **5**(3).

15. Puhlmann, M.-L. and W.M. de Vos, *Intrinsic dietary fibers and the gut microbiome: Rediscovering the benefits of the plant cell matrix for human health.* Frontiers in Immunology, 2022. **13**.

16. Augustin, L.S.A., et al., *Dietary fibre consensus from the International Carbohydrate Quality Consortium (ICQC).* Nutrients, 2020. **12**(9).

17. Liu, Y., et al., *Substrate Use Prioritization by a Coculture of Five Species of Gut Bacteria Fed Mixtures of Arabinoxylan, Xyloglucan, β-Glucan, and Pectin.* Appl Environ Microbiol, 2020. **86**(2).

18. Kaur, A., et al., *Physical Inaccessibility of a Resistant Starch Shifts Mouse Gut Microbiota to Butyrogenic Firmicutes.* Mol Nutr Food Res, 2019. **63**(7): p. e1801012.

19. Villa, M.M., et al., *Interindividual Variation in Dietary Carbohydrate Metabolism by Gut Bacteria Revealed with Droplet Microfluidic Culture.* mSystems, 2020. **5**(3).

20. Lee, W.-J. and P.T. Brey, *How Microbiomes Influence Metazoan Development:Insights from History and Drosophila Modeling of Gut-Microbe Interactions.* Annual Review of Cell and Developmental Biology, 2013. **29**(1): p. 571-592.

21. Lesperance, D.N.A. and N.A. Broderick, *Gut Bacteria Mediate Nutrient Availability in Drosophila Diets.* Appl Environ Microbiol, 2020. **87**(1).

22. Tamaki, F.K., et al., *Physiology of digestion and the molecular characterization of the major digestive enzymes from Periplaneta americana.* J Insect Physiol, 2014. **70**: p. 22-35.

23. Eisner, T., *The digestion and absorption of fats in the foregut of the cockroach, Periplaneta americana.* J. Exp. Zool, 1955(130): p. 159-181.

24. Engel, P. and N.A. Moran, *The gut microbiota of insects – diversity in structure and function.* FEMS Microbiology Reviews, 2013. **37**(5): p. 699-735.

25. Ryan, C.A., *Protease Inhibitors in Plants: Genes for Improving Defenses Against Insects and Pathogens.* Annual Review of Phytopathology, 1990. **28**(Volume 28, 1990): p. 425-449.

26. Schauer, C., C.L. Thompson, and A. Brune, *The bacterial community in the gut of the cockroach Shelfordella lateralis reflects the close evolutionary relatedness of cockroaches and termites.* Appl Environ Microbiol, 2012. **78**(8): p. 2758-67.

27. Tinker, K.A. and E.A. Ottesen, *The Core Gut Microbiome of the American Cockroach, Periplaneta americana, Is Stable and Resilient to Dietary Shifts.* Appl Environ Microbiol, 2016. **82**(22): p. 6603-6610.

28. Cruden, D.L. and A.J. Markovetz, *Microbial ecology of the cockroach gut.* Annu Rev Microbiol, 1987. **41**: p. 617-43.

29. Jahnes, B.C. and Z.L. Sabree, *Nutritional symbiosis and ecology of host-gut microbe systems in the Blattodea.* Current Opinion in Insect Science, 2020. **39**: p. 35-41.

30. Zurek, L. and B.A. Keddie, *Contribution of the colon and colonie bacterial flora to metabolism and development of the American cockroach Periplaneta americana (L).* Journal of Insect Physiology, 1996. **42**(8): p. 743-748.

31. Ayayee, P.A., et al., *The role of gut microbiota in the regulation of standard metabolic rate in female Periplaneta americana.* PeerJ, 2018. **6**: p. e4717.

32. Vera-Ponce de León, A., et al., *Microbiota Perturbation or Elimination Can Inhibit Normal Development and Elicit a Starvation-Like Response in an Omnivorous Model Invertebrate.* mSystems, 2021. **6**(4): p. e0080221.

33. Ayayee, P.A., T. Larsen, and Z. Sabree, *Symbiotic essential amino acids provisioning in the American cockroach, Periplaneta americana (Linnaeus) under various dietary conditions.* PeerJ, 2016. **4**: p. e2046.

34. Cochran, D.G., D.E. Mullins, and K.J. Mullins, *Cytological Changes in the Fat Body of the American Cockroach, Periplaneta americana, in Relation to Dietary Nitrogen Levels.* Annals of the Entomological Society of America, 1979. **72**(2): p. 197-205.

35. Sabree, Z.L., S. Kambhampati, and N.A. Moran, *Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont.* Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19521-6.

36. Schauer, C., C. Thompson, and A. Brune, *Pyrotag sequencing of the gut microbiota of the cockroach Shelfordella lateralis reveals a highly dynamic core but only limited effects of diet on community structure.* PLoS One, 2014. **9**(1): p. e85861.

37. Lampert, N., A. Mikaelyan, and A. Brune, *Diet is not the primary driver of bacterial community structure in the gut of litter-feeding cockroaches.* BMC Microbiol, 2019. **19**(1): p. 238.

38. Bertino-Grimaldi, D., et al., *Bacterial community composition shifts in the gut of Periplaneta americana fed on different lignocellulosic materials.* Springerplus, 2013. **2**: p. 609.

39. Pérez-Cobas, A.E., et al., *Diet shapes the gut microbiota of the omnivorous cockroach Blattella germanica.* FEMS Microbiol Ecol, 2015. **91**(4).

40. Zhu, J., et al., *Diet Influences the Gut Microbial Diversity and Olfactory Preference of the German Cockroach Blattella germanica.* Curr Microbiol, 2022. **80**(1): p. 23.

41. Okarter, N. and R.H. Liu, *Health benefits of whole grain phytochemicals.* Crit Rev Food Sci Nutr, 2010. **50**(3): p. 193-208.

42. Tuncil, Y.E., et al., *Delayed utilization of some fast-fermenting soluble dietary fibers by human gut microbiota when presented in a mixture.* Journal of Functional Foods, 2017. **32**: p. 347-357.

43. Koropatkin, N.M., E.A. Cameron, and E.C. Martens, *How glycan metabolism shapes the human gut microbiota.* Nature Reviews Microbiology, 2012. **10**(5): p. 323-335.

44. Piper, M.D., et al., *A holidic medium for Drosophila melanogaster.* Nat Methods, 2014. **11**(1): p. 100-5.

45. Talyuli, O.A., et al., *The use of a chemically defined artificial diet as a tool to study Aedes aegypti physiology.* J Insect Physiol, 2015. **83**: p. 1-7.

46. Gonzales, K.K., et al., *The Effect of SkitoSnack, an Artificial Blood Meal Replacement, on Aedes aegypti Life History Traits and Gut Microbiota.* Sci Rep, 2018. **8**(1): p. 11023.

47. Majumder, R., et al., *Artificial Larval Diet Mediates the Microbiome of Queensland Fruit Fly.* Front Microbiol, 2020. **11**: p. 576156.

48. Noland, J.L., J.H. Lilly, and C.A. Baumann, *Vitamin Requirements of the Cockroach Blatella germanica (L.).* Annals of the Entomological Society of America, 1949. **42**(2): p. 154-164.

49. Haydak, M.H., *Influence of the Protein Level of the Diet on the Longevity of Cockroaches.* Annals of The Entomological Society of America, 1953. **46**: p. 547-560.

50. Hamilton, R.L. and C. Schal, *Effects of Dietary Protein Levels on Reproduction and Food Consumption in the German Cockroach (Dictyoptera: Blattellidae).* Annals of the Entomological Society of America, 1988. **81**(6): p. 969-976.

51. Tinker, K.A. and E.A. Ottesen, *Phylosymbiosis across Deeply Diverging Lineages of Omnivorous Cockroaches (Order Blattodea).* Appl Environ Microbiol, 2020. **86**(7).

52. Tinker, K.A. and E.A. Ottesen, *Differences in Gut Microbiome Composition Between Sympatric Wild and Allopatric Laboratory Populations of Omnivorous Cockroaches.* Front Microbiol, 2021. **12**: p. 703785.

53. Caporaso, J.G., et al., *Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**(Suppl 1): p. 4516-22.

54. Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina amplicon data.* Nat Methods, 2016. **13**(7): p. 581-3.

55. RStudioTeam, *RStudio: Integrated Development for R*. 2020.

56. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic Acids Res, 2013. **41**(Database issue): p. D590-6.

57. Oksanen, J.F., et al., *vegan: Community Ecology Package.* R package version 2.5-6, 2019(https://CRAN.R-project.org/package=vegan).

58. Legendre, P. and M. De Cáceres, *Beta diversity as the variance of community data: dissimilarity coefficients and partitioning.* Ecology Letters, 2013. **16**(8): p. 951-963.

59. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

60. Conway, J.R., A. Lex, and N. Gehlenborg, *UpSetR: an R package for the visualization of intersecting sets and their properties.* Bioinformatics, 2017. **33**(18): p. 2938-2940.

61. Friedman, J. and E.J. Alm, *Inferring Correlation Networks from Genomic Survey Data.* PLOS Computational Biology, 2012. **8**(9): p. e1002687.

62. Csardi, G. and T. Nepusz, *The igraph software package for complex network research.* InterJournal, 2006. **Complex Systems**: p. 1695.

63. Csárdi, G., et al., *igraph: Network Analysis and Visualization in R*. 2024.

64. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.

65. Dukes, H.E., K.A. Tinker, and E.A. Ottesen, *Disentangling hindgut metabolism in the American cockroach through single-cell genomics and metatranscriptomics.* Front Microbiol, 2023. **14**: p. 1156809.

66. Zhao, F., et al., *Dietary Proteins Rapidly Altered the Microbial Composition in Rat Caecum.* Current Microbiology, 2017. **74**(12): p. 1447-1452.

67. Akimbekov, N.S., et al., *Vitamin D and the host-gut microbiome: A brief overview.* Acta Histochem Cytochem, 2020. **53**(3): p. 33-42.

68. Zhu, Y., et al., *Meat, dairy and plant proteins alter bacterial composition of rat gut bacteria.* Sci Rep, 2015. **5**: p. 15220.

69. Waldron, K.W., M.L. Parker, and A.C. Smith, *Plant Cell Walls and Food Quality.* Compr Rev Food Sci Food Saf, 2003. **2**(4): p. 128-146.

70. Merzendorfer, H. and L. Zimoch, *Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases.* Journal of Experimental Biology, 2003. **206**(24): p. 4393-4412.

71. Vera-Ponce de León, A., et al., *Cultivable, Host-Specific Bacteroidetes Symbionts Exhibit Diverse Polysaccharolytic Strategies.* Appl Environ Microbiol, 2020. **86**(8).

72. Salgado, J.F.M., et al., *Unveiling lignocellulolytic potential: a genomic exploration of bacterial lineages within the termite gut.* Microbiome, 2024. **12**(1): p. 201.

73. Israeli-Ruimy, V., et al., *Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions.* Scientific Reports, 2017. **7**(1): p. 42355.

74. Drula, E., et al., *The carbohydrate-active enzyme database: functions and literature.* Nucleic Acids Res, 2022. **50**(D1): p. D571-d577.

75. Sun, Y., et al., *Gut Firmicutes: Relationship with dietary fiber and role in host homeostasis.* Crit Rev Food Sci Nutr, 2023. **63**(33): p. 12073-12088.

76. El Kaoutari, A., et al., *The abundance and variety of carbohydrate-active enzymes in the human gut microbiota.* Nat Rev Microbiol, 2013. **11**(7): p. 497-504.

77. Sung, H.G., et al., *Effects of Methylcellulose on Cellulolytic Bacteria Attachment and Rice Straw Degradation in the In vitro Rumen Fermentation.* Asian-Australas J Anim Sci, 2013. **26**(9): p. 1276-81.

78. Rasmussen, M.A., et al., *Inhibitory Effects of Methylcellulose on Cellulose Degradation by Ruminococcus flavefaciens.* Appl Environ Microbiol, 1988. **54**(4): p. 890-7.

79. Hespell, R.B. and M.A. Cotta, *Degradation and utilization by Butyrivibrio fibrisolvens H17c of xylans with different chemical and physical properties.* Appl Environ Microbiol, 1995. **61**(8): p. 3042-50.

80. Coen, J.A. and B.A. Dehority, *Degradation and utilization of hemicellulose from intact forages by pure cultures of rumen bacteria.* Appl Microbiol, 1970. **20**(3): p. 362-8.

81. Dehority, B.A., *Rate of isolated hemicellulose degradation and utilization by pure cultures of rumen bacteria.* Appl Microbiol, 1967. **15**(5): p. 987-93.

82. Gírio, F.M., et al., *Hemicelluloses for fuel ethanol: A review.* Bioresource Technology, 2010. **101**(13): p. 4775-4800.

83. Zhang, M., et al., *Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes.* Proc Natl Acad Sci U S A, 2014. **111**(35): p. E3708-17.

84. Centanni, M., et al., *Differential growth of bowel commensal Bacteroides species on plant xylans of differing structural complexity.* Carbohydrate Polymers, 2017. **157**: p. 1374-1382.

85. Dodd, D., R.I. Mackie, and I.K. Cann, *Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes.* Mol Microbiol, 2011. **79**(2): p. 292-304.

86. Kmezik, C., et al., *A polysaccharide utilization locus from the gut bacterium Dysgonomonas mossii encodes functionally distinct carbohydrate esterases.* J Biol Chem, 2021. **296**: p. 100500.

87. Pereira, G.V., et al., *Degradation of complex arabinoxylans by human colonic Bacteroidetes.* Nature Communications, 2021. **12**(1): p. 459.

88. Mohand-Oussaid, O., et al., *The extracellular xylan degradative system in Clostridium cellulolyticum cultivated on xylan: evidence for cell-free cellulosome production.* J Bacteriol, 1999. **181**(13): p. 4035-40.

89. Hershko Rimon, A., et al., *Novel clostridial cell-surface hemicellulose-binding CBM3 proteins.* Acta Crystallogr F Struct Biol Commun, 2021. **77**(Pt 4): p. 95-104.

90. Leth, M.L., et al., *Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut.* Nature Microbiology, 2018. **3**(5): p. 570-580.

91. Baker, J.T., et al., *Friend or Foe? Impacts of Dietary Xylans, Xylooligosaccharides, and Xylanases on Intestinal Health and Growth Performance of Monogastric Animals.* Animals (Basel), 2021. **11**(3).

92. Jefferson, A. and K. Adolphus, *The Effects of Intact Cereal Grain Fibers, Including Wheat Bran on the Gut Microbiota Composition of Healthy Adults: A Systematic Review.* Frontiers in Nutrition, 2019. **6**.

93. Broekaert, W.F., et al., *Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides.* Crit Rev Food Sci Nutr, 2011. **51**(2): p. 178-94.

94. Jana, U.K., N. Kango, and B. Pletschke, *Hemicellulose-Derived Oligosaccharides: Emerging Prebiotics in Disease Alleviation.* Front Nutr, 2021. **8**: p. 670817.

95. Smith, M.M. and J. Melrose, *Xylan Prebiotics and the Gut Microbiome Promote Health and Wellbeing: Potential Novel Roles for Pentosan Polysulfate.* Pharmaceuticals (Basel), 2022. **15**(9).

96. Berger, K., et al., *Xylooligosaccharides Increase Bifidobacteria and Lachnospiraceae in Mice on a High-Fat Diet, with a Concomitant Increase in Short-Chain Fatty Acids, Especially Butyric Acid.* J Agric Food Chem, 2021. **69**(12): p. 3617-3625.

97. Zhou, J., et al., *Dietary supplemental xylooligosaccharide modulates nutrient digestibility, intestinal morphology, and gut microbiota in laying hens.* Anim Nutr, 2021. **7**(1): p. 152-162.

98. Liu, B., et al., *Competition Between Butyrate Fermenters and Chain-Elongating Bacteria Limits the Efficiency of Medium-Chain Carboxylate Production.* Frontiers in Microbiology, 2020. **11**.

99. Chung, W.S.F., et al., *Impact of carbohydrate substrate complexity on the diversity of the human colonic microbiota.* FEMS Microbiology Ecology, 2019. **95**(1): p. fiy201.

100. DePoy, A.N., et al., *Microbial transcriptional responses to host diet maintain gut microbiome homeostasis in the American cockroach.* bioRxiv, 2024: p. 2024.10.31.621369.

101. Miyata, R., et al., *Influence of feed components on symbiotic bacterial community structure in the gut of the wood-feeding higher termite Nasutitermes takasagoensis.* Biosci Biotechnol Biochem, 2007. **71**(5): p. 1244-51.

102. Dong, H.L., et al., *Differences in gut microbiota between silkworms (Bombyx mori) reared on fresh mulberry (Morus alba var. multicaulis) leaves or an artificial diet.* RSC Adv, 2018. **8**(46): p. 26188-26200.

103. Xie, B.H., et al., *Analysis of gut microbiota of ladybug beetle (Harmonia axyridis) after feeding on different artificial diets.* BMC Microbiol, 2024. **24**(1): p. 5.

104. Gohl, P., C.M.R. LeMoine, and B.J. Cassone, *Diet and ontogeny drastically alter the larval microbiome of the invertebrate model Galleria mellonella.* Canadian Journal of Microbiology, 2022. **68**(9): p. 594-604.

105. Powell, J.E., et al., *The microbiome and gene expression of honey bee workers are affected by a diet containing pollen substitutes.* PLOS ONE, 2023. **18**(5): p. e0286070.

106. Allaway, D., et al., *Rapid reconstitution of the fecal microbiome after extended diet-induced changes indicates a stable gut microbiome in healthy adult dogs.* Appl Environ Microbiol, 2020. **86**(13).

107. Martínez-Mota, R., et al., *Natural diets promote retention of the native gut microbiota in captive rodents.* Isme j, 2020. **14**(1): p. 67-78.

108. Gurry, T., et al., *Predictability and persistence of prebiotic dietary supplementation in a healthy human cohort.* Sci Rep, 2018. **8**(1): p. 12699.

109. Hernot, D.C., et al., *In Vitro Digestion Characteristics of Unprocessed and Processed Whole Grains and Their Components.* Journal of Agricultural and Food Chemistry, 2008. **56**(22): p. 10721-10726.

110. Cardona, F., et al., *Benefits of polyphenols on gut microbiota and implications in human health.* J Nutr Biochem, 2013. **24**(8): p. 1415-22.

111. Parkar, S.G., T.M. Trower, and D.E. Stevenson, *Fecal microbial metabolism of polyphenols and its effects on human gut microbiota.* Anaerobe, 2013. **23**: p. 12-9.

112. Adom, K.K., M.E. Sorrells, and R.H. Liu, *Phytochemicals and antioxidant activity of milled fractions of different wheat varieties.* J Agric Food Chem, 2005. **53**(6): p. 2297-306.

113. Sultana, S., et al., *A Review of the Phytochemistry and Bioactivity of Clover Honeys (Trifolium spp.).* Foods, 2022. **11**(13): p. 1901.

114. Lin, J., Y. Gu, and K. Bian, *Bulk and Surface Chemical Composition of Wheat Flour Particles of Different Sizes.* Journal of Chemistry, 2019. **2019**: p. 5101684.

115. Fernando, W.M., et al., *The influence of environmental factors on the adhesion of combinations of probiotics to rice fibre fractions.* World J Microbiol Biotechnol, 2012. **28**(6): p. 2293-302.

116. Abell, G.C.J., et al., *Phylotypes related to Ruminococcus bromii are abundant in the large bowel of humans and increase in response to a diet high in resistant starch.* FEMS Microbiology Ecology, 2008. **66**(3): p. 505-515.

117. Sun, Y., Y. Su, and W. Zhu, *Microbiome-Metabolome Responses in the Cecum and Colon of Pig to a High Resistant Starch Diet.* 2016. **7**(779).

118. Ze, X., et al., *Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon.* The ISME Journal, 2012. **6**(8): p. 1535-1543.

119. Rangarajan, A.A., et al., *Ruminococcus bromii enables the growth of proximal Bacteroides thetaiotaomicron by releasing glucose during starch degradation.* Microbiology (Reading), 2022. **168**(4).

120. Patterson, M.A., M. Maiya, and M.L. Stewart, *Resistant Starch Content in Foods Commonly Consumed in the United States: A Narrative Review.* J Acad Nutr Diet, 2020. **120**(2): p. 230-244.

121. Dockman, R. and E. Ottesen, *Synthetic diets containing a single polysaccharide disrupt gut microbial community structure and microbial interaction networks in the American cockroach.* bioRxiv [preprint], 2024: p. 2024.05.15.594388.

CHAPTER 4

NICHE SPECIALIZATION AND CROSS-FEEDING INTERACTIONS IN OMNIVOROUS INSECT GUT MICROBIOTA[3]

[3] Dockman, R.L. To be submitted to a peer-reviewed journal.

Abstract

The gut microbiome is critical to host health, with dysbiotic communities linked to numerous gastrointestinal, immune, and cognitive dysfunctions. Dietary fiber plays a key role in influencing both the stability of the gut community as well as production of key gut metabolites. Using the American cockroach (*Periplaneta americana*) as a model omnivore, we use synthetic diets to identify how complex gut microbial communities respond to two of the most abundant plant polysaccharides, xylan and cellulose. To do so, we fed cockroaches synthetic diets containing one of these fibers or a mix of both in differing ratios. Through both 16S rRNA gene profiling and RNA-seq, we show that mixed fibers enrich for organisms characteristic of the source fibers as well as additional organisms only enriched by mixed-fiber diets. Through an organism-centric pangenome approach, we identify the impact of these fibers on gut microbiome activity. We found that gut communities responded strongly to xylan, with *Bacteroidota* belonging to *Bacteroides*, *Dysgonomonas*, and *Parabacteroides* producing xylan-active CAZymes at high levels. Multiple groups of *Bacillota* also responded strongly to a xylan diet, but appeared to act as cross-feeding secondary degraders, producing primarily xylosidases and transcripts associated with xylose utilization. In contrast, cellulose diets were associated with higher transcriptional activity among *Fibrobacterota*, which are typically a minor component of the cockroach gut microbiome but were the primary producers of CAZymes associated with cellulose and cellobiose degradation. These experiments provide new insight into gut microbial metabolism of these complex plant polysaccharides. Further, they highlight the utility of the cockroach model and synthetic diets to answer fundamental questions about gut microbial responses to different polysaccharides alone and in combination.

Introduction

Across the animal kingdom, beneficial microbiota within an organism's gastrointestinal tract modulate host heath by conferring immune and metabolic advantages to their host [1]. The precise composition of the microbial community is influenced by host factors, such as its habitat determining exposure to potential gut microbes and its anatomy filtering out unsuitable symbionts, as well as interactions between resident microbiota [2-4]. Among these factors, the most malleable characteristic is host diet: gut microbiota rely on the host to consume the nutrients necessary for their own growth [5, 6]. There is considerable interest in leveraging dietary components, particularly polysaccharides, to manipulate gut microbiome composition for desired human health outcomes [7-9]. However, major challenges remain in deciphering and predicting microbial responses to different dietary components. Culture and *in vitro* models cannot fully replicate the complex cross-talk within and between the gut microbial community and the host, while dietary manipulation in mammalian models is limited by the complex nutritional needs of the host. Therefore, isolating microbial responses within a complex community to a single substrate is difficult, and observations from *in vitro* culture*s* studies may not holistically capture the culture's behavior in a gut environment.

Of all of the macronutrients an omnivore consumes, fiber is the most relevant to the gut microbiome. While the host has first access to dietary components and can therefore extract desirable nutrients before food reaches gut microbiota, most host omnivores do not encode the cellulases or glycoside hydrolases required to break down fibers. Instead, these fibers pass directly to the gut, serving as an important source of carbon for the microbial community [10, 11]. The importance of fiber in maintaining a diverse and beneficial gut population has been well established in mammalian omnivores. Diets deficient in fiber disrupt microbial distribution

throughout the gut [12, 13], lower diversity [14], and stimulate increased catabolism of host glycans, damaging mucus barriers protecting the gut [15, 16]. Additionally, complex fibers serve as attachment loci for microbial adherence [17-19] and are fermented into valuable byproducts that benefit the host [20] [21, 22].

The term “fiber” is a broad classification encompassing any plant-derived carbohydrate that is recalcitrant to human enzymatic digestion, such as cellulose, hemicelluloses (mannan, beta-glucan, xylan), pectin, and lignin [23, 24]. These polysaccharides can be highly heterogenous, varying in solubility, sugar composition, or branching structures, all of which influence their accessibility to gut microbiota [25, 26]. Further, when placed in the context of natural foods, cell wall polysaccharides are tightly intertwined to stabilize the plant cell, which complicates their degradation and requires concerted microbial action to dismantle [24, 27, 28]. To fully understand the influence of fiber on the gut microbiome, it is vital to observe microbial responses to different fiber types. However, studying individual fiber sources as the sole carbon source is unethical in mammalian systems, confounding interpretation of results.

Omnivorous cockroaches such as the American cockroach (*Periplaneta americana*) are quickly emerging as model organisms for host-microbiome interactions that reflect the complexity of mammalian microbial communities while overcoming their dietary limitations [29]. The gut environment of cockroaches is analogous to mammals in that it can be divided into three functionally distinct sections: the foregut (crop and proventriculus) for initial amylase activity and mechanical digestion; the midgut (gastric caeca and ventriculus) for primary host digestive activity; the hindgut (colon) for waste consolidation and microbial fermentation [30]. Within the hindgut, cockroach gut microbiota are compositionally similar to the communities found in humans and mice, consisting primarily of the bacterial phyla *Bacteroidota*, *Bacillota*,

and *Desulfobacterota* [31-34]. While these bacteria are commonly found in gut communities, the niche they fill is influenced by what their host consumes. *Bacteroidota* isolates are frequently identified as primary fermenters that are capable of digesting complex polysaccharides, notably host-indigestible fibers, into their constituent molecules for either individual use or to release into the environment for consumption by secondary fermenters, such as *Bacillota* [11, 35-39]. In addition to their roles in oligosaccharide scavenging, members of *Bacillota* are also associated with amino acid fermentation. Fermentation by these two phyla converts indigestible polysaccharides into metabolic byproducts such as short chain fatty acids, lactate, formate, and hydrogen, which can be absorbed by the host as vital energy sources or passed on to the collective activity of hydrogen-consuming *Desulfobacterota* and methanogenic *Archaea* [40-42].

Previously, we have found that single fiber synthetic diets induce alterations in the gut microbiome taxonomic composition of the American cockroach despite its resilience to drastic changes in whole-food diet compositions [32, 43]. In particular, microcrystalline cellulose and xylan generated the largest differences in community structure, with the xylan-fed gut community displaying reduced diversity and fragmented co-correlation networks [43]. Cellulose and xylan are the two most abundant biopolymers found on Earth, and are therefore highly attractive subjects for fiber-based research. Cellulose is comprised of repeating glucose units bound by β(1-4) glycosidic bonds, forming a linear structure that serves as the scaffolding and backbone of plant cell walls. In contrast, xylan contains a backbone of xylose units bound by β(1-4) glycosidic bonds that can be decorated with branches of arabinose, fucose, methyl groups, acetylation, glucuronic acid, or galacturonic acid, among other sugars, serving with pectin as fortification of the cellulose scaffold [44, 45]. Both polysaccharides are indigestible to humans and cockroaches, restricting the portion of digestible carbohydrates that can be replaced in

mammalian diets. Insect models do not have the same restrictions, making *P. americana* an ideal platform in which to study these fibers more deeply.

Here, we fed American cockroaches synthetic diets featuring microcrystalline cellulose, xylan, or a mixture of both as the carbon source to identify how these chemically distinct fibers influence the functional landscape of a diverse gut community. Through 16S rRNA gene sequencing paired with metatranscriptomic profiling, we observed that dominant community members exhibit large differences in their abundance and activity when confronted with these two fibers, individually and in varying mixtures. These results highlight the flexible roles played by members of a highly complex gut community living within an unfussy omnivore and provide a framework for evaluating microbial responses to distinct dietary fibers.

Materials and Methods

Insects

Our *Periplaneta americana* colony has been maintained in captivity at the University of Georgia for over a decade. Mixed age and sex stock insects are maintained at room temperature in glass aquarium tanks with wood chip bedding and cardboard tubes for shelter in a 12:12 light:dark cycle. Water via cellulose sponge fit to a Tupperware reservoir and dog chow (Purina ONE chicken & rice formula, approximately 26% protein, 16% fat, and 3% fiber) are provided to stock colonies *ad libitum*.

Diet Composition

Synthetic diets were prepared as described in [43] and briefly summarized here. The synthetic diets contained, as percent dry weight, 0.5% Vanderzant vitamin mix (catalog #: 903244, MP Biomedicals, USA), 3% Wesson salt mix (catalog #: 902851, MP Biomedicals), 8% peptone (catalog #: J636, Amresco, VWR International, USA), 17% casein (catalog #: C3400,

Sigma-Aldrich,USA), and 1% cholesterol (catalog #: 0433, VWR). The remaining 70.5% consisted of either xylan from corn core (catalog #TCX0078, TCI Chemicals, USA), microcrystalline cellulose (MCC; 51um particle size; catalog #435236, Sigma-Aldrich), or a mixture of these carbohydrates in ratios of 1:1, 3:1, or 6:1. Throughout, ratio diets will referred to by the percentage of the majority-fiber. Diets with 6:1 ratios will be referred to as 86% "cellulose" or "86% xylan", 3:1 will be "75% cellulose" or "75%" xylan, and 1:1 will be referred to as "50% mix". For each diet, dry components were suspended in sufficient volumes of diH_2O to form a homogenous dough, shaped into pellets, then dehydrated at 65°C until solid. Food pellets were stored at -20°C until use.

Experimental Design

As described in [43] and [32], healthy mixed-sex *P. americana* adults were selected from the stock colony and moved to experimental tanks containing bleach-sterilized pebbles and polyvinyl chloride tubes for footing and shelter, with synthetic diets and water provided *ad libitum* in glass petri dishes. Dietary experiments lasted for two weeks, during which food and water were replaced daily and visible debris or oothecae were removed as needed.

All insects were sacrificed for sample collection upon completion of the two-week treatment period via cold-immobilization, decapitation, and dissection. The insects were dissected on individual clean culture plates with fine tipped forceps to remove exoskeleton and fat body adhered to the gut wall. The cleaned gut was then transferred to an aluminum dish on dry ice for freezing and divided into foregut, midgut, and hindgut sections. Hindgut sections to be used for DNA extraction were collected into 500uL phosphate-buffered saline (1X PBS), disrupted with a sterile pestle to suspend gut-attached bacteria, and stored at –20°C until DNA extraction.

To construct host transcriptomes and hindgut metatranscriptomes, five samples were collected from each individual insect: hindgut lumen, midgut lumen, hindgut tissue, midgut tissue, and fat body. The insects were dissected as described above, with the addition of gut-adhered visceral fat body collection while cleaning the gut surface. All samples for RNA sequencing were stored in 100uL RNAlater (Invitrogen, Thermo Fisher, USA) immediately upon removal from the dry ice. Fat body samples were directly deposited during the dissection. Midgut and hindgut sections were transferred to a sterile 1.5mL tube and pressed against the tube wall with a pestle to coax out gut lumen contents without pulverizing the tissues, after which the sample was vortexed in 100uL RNAlater. The pestle was again used to lightly wring out remaining lumen contents from the gut tissue, then to transfer the empty guts into fresh 1.5mL tubes containing RNAlater. Samples destined for RNA-seq were stored at -80°C until use.

Extraction and Sequencing of 16S rRNA Gene Libraries

DNA was extracted from 200µL aliquots of individual samples using the EZNA Bacterial DNA Kit (Omega Biotek, Norcross, GA, USA) as described in detail in [43]. The extracted DNA, suspended in 50uL of EZNA elution buffer, was checked for quality and quantitated using the Take3 plate for BioTek plate readers (Agilent, USA) prior to barcoding and Illumina-compatible library preparation.

The V4 region of the 16S rRNA gene was amplified and appended with barcoded primers from 5-10ng DNA per sample via 2-step polymerase chain reaction (PCR) using 0.02U/L Q5 Hot Start high-fidelity DNA polymerase (New England BioLabs, Ipswich, MA, USA), described in [32, 33, 46]. The final products were checked via gel electrophoresis and cleaned as outlined in the Omega Biotek Cycle Pure kit, then pooled at equimolar proportions for sequencing. The

prepared library was sequenced with 250 base pair paired-end Illumina MiSeq sequencing at the Georgia Genomics and Bioinformatics Core at the University of Georgia.

RNA Extraction

Both host and microbial RNA were extracted using the EZNA Total RNA Kit II (Omega-Biotek), which combines the phenol/guanidine isothiocyanate-based RNA-Solv Reagent (Omega-Biotek) with column-based nucleic acid purification. While both sample types shared most steps in RNA extraction, they differ in homogenization method. For host transcriptomes, half of each gut wall sample (split lengthwise) or approximately 15mg fat body was homogenized directly in 1mL RNA-Solv with an Ultra-Turrax rotor-stator homogenizer (Janke & Kunkel, Germany). The homogenized samples were phase-separated with 200uL chloroform (VWR) and the aqueous phase column-purified as described in the kit instructions. The final product was eluted into 50uL nuclease-free water.

For microbial RNA extraction, 50uL of each sample was vortexed with 200uL ice-cold RNase-free PBS, then centrifuged at 5,000g for five minutes to pellet. The supernatant was removed, and the pellet resuspended in 50uL of a 30mg/mL lysozyme/RNase-free TE buffer solution with 4uL Superase-IN RNase inhibitor (Invitrogen) via 30s of vortexing. Samples were incubated in a shaking incubator (Eppendorf, USA) for 10 minutes at 30°C and 350rpm, then transferred to a screw-top 2mL tube containing 50mg 0.1mm silica beads (BioSpec, USA) and 1 mL RNA-Solv. Cell lysis was performed via four cycles of bead beating at max speed on a vortex for 30s followed by 30s on ice, after which samples were phase-separated with 200uL chloroform and RNA isolated as above.

Isolated RNA from all sample types were incubated for 20 min at 37°C with 1uL TURBO DNase enzyme (Invitrogen) and 5uL TURBO DNase reaction buffer to destroy any

contaminating DNA, then purified using the column-based EZNA MicroElute RNA Clean-Up Kit (Omega-Biotek). Microbial RNA was eluted once in 15uL nuclease-free water, while host samples were eluted twice for 30uL total of purified RNA per sample. The RNA content was quantitated via the Take3 plate, then assessed for quality with the Agilent Bioanalyzer, and samples were stored at -80°C until library preparation.

RNA Library Construction

Host tissue mRNA was isolated with the NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB) according to kit instructions, while prokaryotic gut lumen RNA was ribodepleted with biotinylated probes constructed from the extracted DNA of pooled MCC-fed and xylan-fed cockroach guts based on the methods described in [34, 47]. Briefly, T7-promoter appended primers were used for PCR to amplify genes corresponding to both bacterial and archaeal 16S and 23S rRNA, in addition to cockroach 18S, 28S, and ITS regions (**Supplemental Table 4.1**). The purified PCR products were then transcribed with the AmpliScribe T7-Flash Biotin-RNA Transcription kit (Lucigen Corp, Middleton, WI) and cleaned with the MEGAclear Transcription Clean-up Kit (Invitrogen). These probes were hybridized with total prokaryotic RNA in saline-sodium citrate (SSC) buffer and formamide, then captured with streptavidin magnetic beads (NEB) which had been cleaned with 0.1N NaOH and SSC buffer ahead of time. The mRNA remaining in the supernatant was cleaned with the EZNA RNA clean-up kit prior to library preparation.

Host and microbial mRNA libraries were constructed using the NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (NEB), following instructions to obtain 300 bp inserts flanked by Illumina-compatible index primers (NEBNext® Multiplex Oligos for Illumina®). Following size and quality verification with the Bioanalyzer, the barcoded samples

were pooled together and sent to Novogene (Sacramento, CA, USA) for Illumina NovaSeq sequencing.

16S rRNA Bioinformatics

The amplicon data collected in this study was processed in R (version 4.2.1) using the R package DADA2 (version 1.24.0), with Amplicon Sequence Variants (ASVs) previously obtained from cockroach gut 16S rRNA gene sequencing included as a priors table [43, 48, 49]. Taxonomy was assigned using DADA2 and the ARB Silva v138 classifier to the species level, followed by filtering to remove sequences matching eukaryotic (chloroplast, mitochondria) or endosymbiotic *Blattabacterium* DNA [48, 50].

Alpha and beta diversity analyses were performed on rarefied count tables (10,790 reads) with the R package vegan (version 2.6-4) [51]. Alpha diversity was calculated for Shannon index (*diversity()*), number of observed ASVs, and Pielou's evenness (calculated as Shannon/log(Observed)) and evaluated for significance with Wilcoxon rank sum test (pairwise comparisons; *pairwise.wilcox.test()*) and Kruskal-Wallis test (multi-group comparisons; *kruskal.test()*). Bray-Curtis dissimilarity and Sørensen index matrices were determined with *vegdist()* then ordinated using non-metric multidimensional scaling (NMDS; function *metaMDS*). Overall ordination quality was assessed with PERMANOVA (function *adonis2()*), and *envfit()* was used to fit "dietary xylan percent" to the plots.

Differential abundance analysis of the ASVs was conducted using DESeq2 (version 1.36) with parameters 'fitType = "local"' and 'design = ~ Diet' [52]. Pairwise result tables were extracted for all diet comparisons and filtered to retain enriched ASVs with an adjusted p-value < 0.05 and baseMean > 100. Overlap in ASVs enriched by single- and mixed-fiber diets was

determined by designating ASVs enriched in any of the four ratio diets as enriched by mixed-fiber, then comparing these ASVs to those enriched on the two single-fiber diets.

Metatranscriptome Processing and Pangenome Construction

The paired-end metatranscriptome sequencing data were processed on the UGA computing cluster using both established programs and custom in-house perl scripts. Joint Genome Institute (JGI) programs BBduk and BBSplit (https://jgi.doe.gov/data-and-tools/software-tools/bbtools/) were used for adapter removal and contamination filtering respectively; BBSplit indexes contained the genome and plasmid of the blattid endosymbiont, *Blattabacterium* [53], and host sequences were flagged and removed based on alignment to the *P. americana* reference genome available on NCBI (BioProject PRJNA1098420) [54]. The remaining reads were screened for rRNA contamination with SortMeRNA and paired mRNA reads were merged using the JGI program BBMerge [55, 56]. Any paired reads that could not be merged using BBMerge were combined with a 10-N spacer sequence using an in-lab script. The merged consensus reads were repaired and filtered with a BBMap shell script (repair.sh), then translated with prodigal [57] for compatibility with DIAMOND blastp [58]. Alignment of the translated reads was performed against the nonredundant bacterial and archaea RefSeq protein databases [59, 60], in addition to a custom database comprising single cell genomes derived from the *P. americana* hindgut [34]. Results from these two DIAMOND runs were combined with an in-lab Perl script to maintain higher scoring hits and improve specificity of our annotations to our host organism. **Supplemental Table 4.2** reports per-sample reads remaining following each processing step.

To minimize the impact of the limited and partial reference genomes available for many cockroach gut microbiota, we mapped transcripts to approximately genus-level pangenomes for

key cockroach gut taxa as described in [61]. Pangenomes were constructed with anvi'o, which aligns and classifies input proteins from user-selected genomes to determine gene clusters that share structural and functional features [62-66]. Gene clusters were annotated with clusters of orthologous genes (COGs) [67, 68], KEGG orthologs [69-71], and carbohydrate-active enzymes (CAZymes) [72-74] during the anvi'o workflow. The final summary file generated in the anvi'o pipeline contains the computed gene clusters that are annotated with protein accessions from the reference genomes, COGs, KOs, and CAZymes. In-lab Perl scripts were used to match hits in the DIAMOND blastp output to the gene clusters and build pangenome count tables for further analysis.

Twelve previously constructed pangenomes (*Bacteroides, Parabacteroides, Dysgonomonas, Alistipes, Odoribacter, Paludibacteraceae, Clostridiaceae, Oscillospiraceae group 1, Oscillospiraceae group 2, Enterococcaceae, Desulfovibrio,* and *Desulfosarcina)* that were designed based on cockroach gut members [75] were used in this study, as well as five new pangenomes: *Lachnospiraceae A, Lachnospiraceae B, Lachnospiraceae C, Fusobacterium*, and *Fibrobacterota.* Since RNAseq can poorly differentiate *Bacillota* members, co-occurrence patterns observed in this experiment were considered when constructing these pangenomes. *Lachnospiraceae* groups were determined based on Spearman correlations calculated for the taxonomic abundance of *Lachnospiraceae* genera with absolute abundances > 1% across ratio diets. Genomes that were used as references to create these pangenomes are listed in **Supplemental Table 4.3.**

Metatranscriptome analyses

Before analyzing transcripts in the context of pangenomes, the overall taxonomic and structural composition of the metatranscriptomes was assessed based on the top-scoring

DIAMOND hits (based on highest bit scores) per read. Since many reads had multiple top-scoring hits, the count of taxonomic identities per read was split to allow each possibility to be considered, thus creating a weighted count table that acknowledged all possible microbes present. These weighted counts were rarefied to the smallest sample size (4,738,232) and used as input into the vegan R package for diversity calculations (alpha: *diversity();* beta: *vegdist()*), ordination (*metaMDS()*), and statistical analyses (*adonis2(), kruskal.test(), pairwise.wilcox.test()*) as described for the 16S rRNA amplicon analysis.

Pangenomes were analyzed for diet-related community structure using gene cluster count tables. Enrichment was determined using DESeq2, after which the number of significant gene clusters were extracted for each pairwise comparison and visualized as a heatmap. Principal component (PCA) and redundancy analyses (RDA) were performed on count tables after transformation with the DESeq2 variance stabilizing transformation function *vst()*. The function *rda()* from the vegan package was used for both; when no model is selected, the function generates unconstrained PCs. Constrained models used for RDA include percent xylan (numeric), whether the diet contains xylan in any amount ("Contains xylan"), whether the diet contains cellulose in any amount ("Contains cellulose"), and "Diet". Component values were obtained with vegan::*summary()* and model significance was assessed with ANOVA (function *anova()*).

The carbohydrate degrading capacity of the pangenomes was analyzed based on gene clusters with CAZyme assignments. Gene clusters sharing the same CAZyme assignment were aggregated together; if a gene cluster had more than one associated CAZyme, it was overcounted to consider both assignments. Substrate specificity for the various CAZyme families was annotated using a mapping file "dbCAN-sub.substrate.mapping.xls" available from the dbCAN2

server [72], with a subset replicated in **Supplemental Table 4.4**. Patterns of CAZyme expression associated with xylan or cellulose was visualized as a heatmap of overall read abundance rather than scaling by samples or CAZyme family.

Metabolic enrichment was determined via DESeq2 for the pangenomes using the individual KEGG orthologs assigned to gene clusters as well as gene clusters collapsed into their level 3 assignments for carbohydrate and amino acid metabolism pathways [69-71]. Enrichment in total pathway expression between the 100% xylan and cellulose diets was displayed as a color-block heatmap. For comparison across the ratio diets, pathways were visualized with heat map bubble plots sized by relative abundance of the entire sample transcript count and color scaled by relative abundance within the pangenome. Pangenome enrichment of individual enzymatic steps were annotated onto KEGG metabolic maps using the R package pathview [76].

Results

Community composition of ratio diets

In order to identify gut community responses to cellulose and xylan, we fed adult cockroaches synthetic diets featuring the hemicellulose xylan and microcrystalline cellulose as the carbohydrate portion with a fixed base of peptone, casein, micronutrients, and cholesterol as described in previous work [43]. In addition to diets featuring purely xylan or cellulose as the carbohydrate source, we combined the two fibers in ratios of (xylan:MCC) 6:1, 3:1, 1:3, and 1:6 by weight. These six diets were fed to mixed-sex adult cockroaches (n=10/diet) for two weeks, after which we dissected out the hindgut and sequenced the 16S rRNA gene profile of the gut microbiome. In total, 2,715,590 quality- and contaminant-filtered reads were obtained from 60 samples with an average read depth of 45,259 (range: 7,656 – 130,911) [48]. Taxonomic

assignment via the SILVA (v138.1) 16S rRNA database classified most reads as *Bacillota* or *Bacteroidota*, followed by *Desulfobacterota* (**Supplemental Figure S4.1**).

We examined the impact of dietary composition on the alpha and beta diversity of the cockroach hindgut microbiome. In general, host diet influenced the alpha diversity of the samples in terms of Shannon index (Kruskal-Wallis, $p < 0.001$), richness (Kruskal-Wallis, $p < 0.01$) and evenness (Kruskal-Wallis, $p < 0.001$) **(Figure 4.1A-C)**. Interestingly, deeper investigation revealed no direct difference between any of the mixed-fiber or 100% cellulose diet configurations in these measures. In contrast, cockroaches given 100% xylan had notably lower alpha diversity than insects fed 100% or 86% cellulose in all three measures and lower Shannon index and evenness scores than those given 75% cellulose or even 75% xylan (Wilcoxon paired tests, $p < 0.05$ each; **Figure 4.1A-C),** suggesting that the addition of cellulose into the diet increases and stabilizes the microbial diversity within a gut community. Beta diversity analyses showed a similarly strong correlation between dietary fiber choice and the gut microbiome composition for abundance-weighted Bray-Curtis (PERMANOVA: $R2 = 0.271$, $p. < 0.001$; **Figure 4.1D)** and abundance-independent Sørensen index (PERMANOVA: $R2 = 0.205$, $p. < 0.001$; **Figure 4.1E**) NMDS ordinations. In general, the cellulose-derived communities clustered closer together at the fringe of the ordination space, while the mixed and 100% xylan communities dispersed away from cellulose in an overlapping, gradient-like manner. Mapping dietary xylan percentage to the ordination plots with envfit confirmed that there is indeed an additive effect of increasing xylan proportion to community structure similarity on both abundance-weighted ($r=0.627$, $p < 0.001$; **Figure 4.1D**) and unweighted ($r=0.358$, $p < 0.001$; **Figure 4.1E**) measures, rather than clear clustering of the individual mixed-fiber or 100% xylan diets.

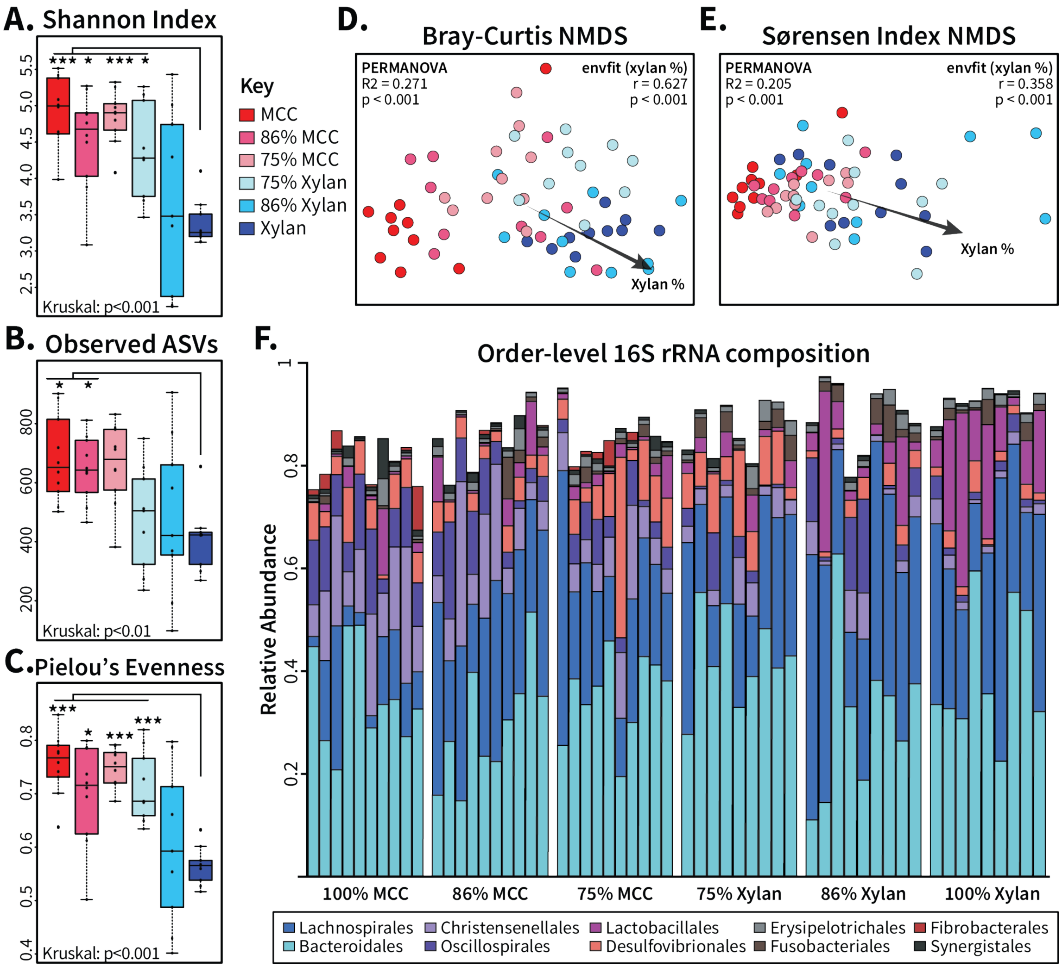


Figure 4.1: Community composition of gut microbiota from ratio diets show dose-dependent shift from cellulose-fed communities to xylan-fed communities. Samples were rarefied to 10,790 reads prior to calculation of **(A)** Shannon index**, (B)** richness via observed ASVs, and **(C)** Pielou's evenness and visualized by box plots. Significance was determined by Kruskal-Wallis test and Wilcoxon test for group and pairwise comparisons, respectively. Ordinations of the rarefied samples were obtained with beta diversity measures **(D)** Bray-Curtis dissimilarity and **(E)** Sørensen index. PERMANOVA was applied to determine the overall significance of diet in ordination structure with "percent xylan" fit to the data and analyzed using envfit. **(F)** ASV count tables were aggregated at the order taxonomic level to calculate relative abundance, and the top ten most abundant bacterial orders were visualized. MCC: microcrystalline cellulose; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

The differences in diversity we observed between diet treatments were associated with shifts in the taxonomic composition of the cockroach gut microbiota. At the order level (**Figure 4.1F**), the gut microbiota in the xylan and cellulose diets formed distinct compositions from each

other, with cellulose associated with higher *Christensenelalles* and *Oscillospirales* while xylan enriched for *Lachnospirales* and *Lactobacillales*. When cellulose and xylan were both provided in the diets, taxonomic signatures from the single-source fibers were still evident in the community compositions; all mixed diets retained the higher *Lachnospirales* abundance in xylan relative to cellulose, for example, and the two *Bacillota* orders enriched in cellulose appeared more frequently in mixed conditions than in xylan. We further investigated the particular ASVs driving these observations with DESeq2, using a baseMean cut-off of 100 and adjusted $p < 0.05$ to limit our analysis to taxa with the largest influence over the community structure, termed diet-responsive ASVs. In total, 84 ASVs were identified as significant in at least one pairwise comparison between the various diets (**Supplementary DESeq2 Result Table)** of which most (68) were found to differ specifically between cellulose and xylan (**Supplemental Figure S4.2A**). As expected, the 100% cellulose and xylan diets shared more ASVs with their corresponding fiber-majority diets than their fiber-minority diets, and showed substantial overlap with the mixed fibers in enriched ASVs (**Supplemental Figure S4.2B).** The 100% fiber diets only enriched for one organism each (cellulose: *ChristensenellaceaeR7.5;* xylan: *Bacteroidales.0)* that was not also enriched by one of the mixed fiber diets. In contrast, when considered together, the mixed fiber diets enriched for ASVs spread across *Clostridia* (8), *Bacteroidales* (6), and *Desulfovibrio* (1) that were not shared by the 100% fiber diets (**Supplemental Figure S4.2B**). ASVs corresponding to *Lachnospiraceae* and *Alistipes* were most commonly enriched by the mixed diets. *Lachnospiraceae*, a diverse group that is poorly defined by 16S, contained members enriched in both the xylan (**Supplemental Figure S4.2C)** and cellulose (**Supplemental Figure S4.2D**) diets as well, while only one *Alistipes* was enriched by xylan. Xylan enriched for more organisms in general than cellulose, but many were only

enriched in comparison to one other diet, often cellulose. Cellulose enriched for fewer organisms, but most were significantly higher in cellulose than in the 100%, 86%, and 75% xylan diets, highlighting the different ways these two fibers influence the gut community.

Taxonomic composition of metatranscriptomes

We selected cockroaches fed one of five synthetic diets containing as the carbohydrate source xylan, cellulose, or a mix of both fibers at 6:1 or 1:1 ratios for the construction of 30 (n=6/diet) individual paired-end hindgut microbiota metatranscriptome libraries. In total, 4,552,148,064 hindgut lumen reads were recovered (**Supplemental Table 4.2**). After filtering for contamination and read quality, a total of 614,662,256 consensus pairs (range: 6,013,230-37,692,726 per sample) were obtained. Translation and alignment of these pairs to non-redundant bacterial and archaeal RefSeq protein databases and cockroach-derived single cell genomes produced 484,022,581 successfully annotated proteins total (78.7% of paired reads, 4,740,617-31,943,483 per sample).

In general, the taxonomic composition of metatranscriptomic reads reflected phylum-level distribution patterns found in 16S amplicon data from the same fiber sources (**Supplemental Figure S4.1**). Compared to xylan, the 100% cellulose diet produced metatranscriptomes with higher alpha diversity **(Supplemental Figure S4.3A-C)** and formed denser clusters during NMDS ordinations of Bray-Curtis dissimilarity (**Supplemental Figure S4.3D**). While these broad measures of community functional composition reflected the 16S gene composition, several organismal groups were found to have activity levels in the metatranscriptomes that differed from their abundance predicted in 16S sequencing. For example, at the order-level (**Figure 4.2A**), *Bacteroidales* activity was strongly associated with increasing xylan percentage, with an average relative abundance of 63% compared to 35%

observed in the cellulose diet; 16S amplicon sequencing captured a much smaller difference of 39% in xylan vs 34% in cellulose. *Lactobacillales,* which comprised an average 14% of xylan 16S amplicon reads, showed little activity in the metatranscriptomes, staying under 1% of transcripts for the single fiber diets.

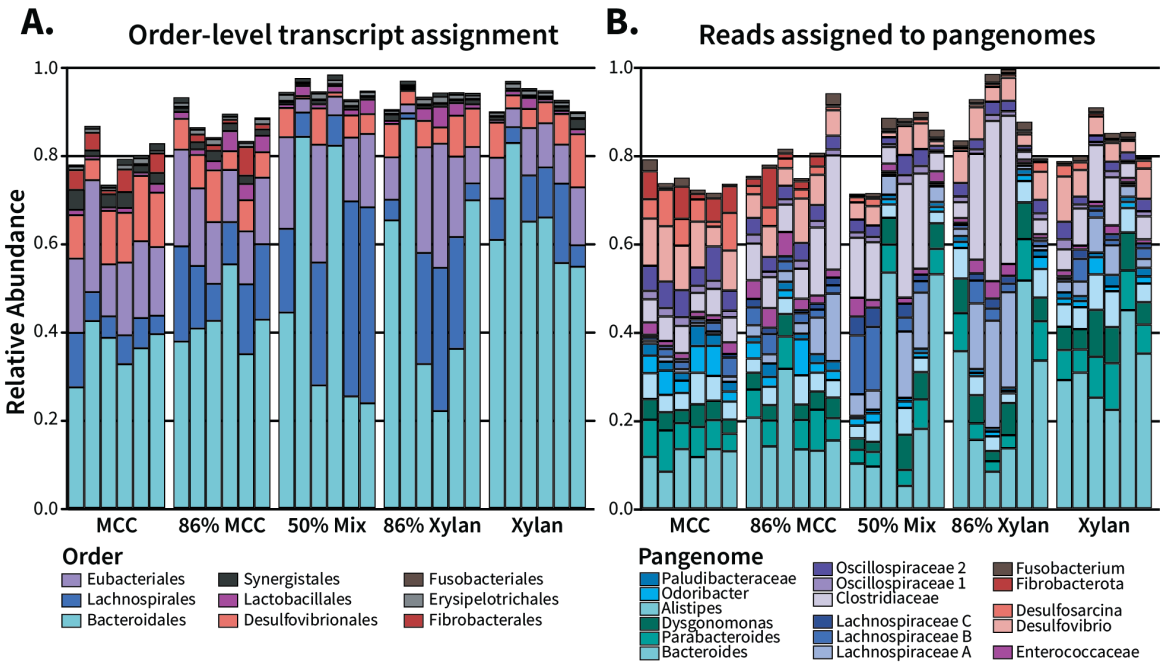


Figure 4.2: Pangenome assignment captures the most abundant taxonomic groups identified in RNA annotation. (A) Count tables tabulating the taxonomic composition of weighted RefSeq and single-cell genome hits were collapsed at the order level for relative abundance calculation. Relevant orders with high abundance in the previous 16S rRNA gene experiments were selected for visualization. (**B**) Relative abundance of metatranscriptomic reads matched to the 17 pangenomes.

Pangenome Overall Comparisons

To facilitate transcriptional analyses of cockroach-associated microbes, we used an in-house pipeline to map transcripts to 17 pangenomes representing key taxa within the cockroach gut microbiota [34, 61, 62]. These include 12 previously constructed pangenomes (*Bacteroides, Parabacteroides, Dysgonomonas, Alistipes, Odoribacter, Paludibacteraceae, Clostridiaceae, Oscillospiraceae group 1, Oscillospiraceae group 2, Enterococcaceae, Desulfovibrio,* and

Desulfosarcina) and five additional pangenomes (*Lachnospiraceae A, Lachnospiraceae B, Lachnospiraceae C, Fusobacterium*, and *Fibrobacterota*) that were newly constructed for this analysis. Altogether, between 66.5% and 94.7% of translated reads per sample were assigned to gene clusters within a pangenome (**Figure 4.2B).** The largest portion of reads mapped to *Bacteroides* followed by *Clostridiaceae, Lachnospiraceae A,* and *Desulfovibrio,* while *Fibrobacterota* had the fewest (**Table 4.1**). Most reads that mapped to the pangenomes were associated with KEGG orthologs (KOs), with functional annotation rates ranging from 68.85% to 85.95% of assigned reads per pangenome. Additionally, we identified gene clusters with carbohydrate active enzymes (CAZymes), to which between 1.44% and 11.27% of reads were mapped per pangenome (**Table 4.1**).

Table 4.1: Reads mapped to pangenome gene clusters and the percent that were assigned KEGG or CAZyme annotations.

Pangenome	Reads mapped	Reads with KO (%)	Unique KO terms	Reads with CAZyme (%)	Unique CAZymes
Alistipes	20992939	78.47	1897	3.41	193
Bacteroides	110224278	74.31	1572	7.26	172
Dysgonomonas	25551436	75.98	1902	11.27	207
Odoribacter	8493413	77.83	1535	3.17	84
Paludibacteraceae	5653004	82.97	1465	4.91	144
Parabacteroides	28204621	73.52	1757	8.36	201
Lachnospiraceae A	42841465	78.57	2807	2.99	197
Lachnospiraceae B	26114106	84.87	2763	4.04	253
Lachnospiraceae C	16067332	85.95	2854	2.84	213
Enterococcaceae	8488768	68.85	2398	7.06	127
Clostridiaceae	65006931	81.07	3473	3.18	255
Oscillospiraceae 1	5007584	74.43	1820	6.13	94
Oscillospiraceae 2	13510807	81.8	2743	3.21	139
Desulfovibrio	26612778	81.67	2935	2	107
Desulfosarcina	6531593	79.45	2438	1.44	63
Fibrobacterota	3380195	75.79	1249	4.84	82
Fusobacterium	4780018	79.89	1757	1.39	43

Principal components ordination of the pangenome transcriptomes showed that communities from the 100% cellulose diet formed a distinct cluster in all *Bacillota* (**Supplemental Figure S4.5B)** and in the *Bacteroidota* genera *Bacteroides, Parabacteroides,*

and *Alistipes* **(Supplemental Figure S4.5A).** The remaining MCC-*Bacteroidota* populations displayed some overlap with the majority cellulose diet, while *Fibrobacterota*, which stratified the most variation along the first PC (64.5%), clustered the 86% and 100% cellulose diets together far away from other samples **(Supplemental Figure S4.5C)**. This aligns well with pairwise analyses of differential gene expression between diets, which frequently found that the largest number of differentially expressed gene clusters were identified in pairwise analyses between cellulose alone vs. diets containing any amount of xylan (**Supplemental Figure S4.4).**

Table 4.2: Redundancy analysis of pangenome gene cluster expression.

Model factor:	Diet	% xylan	Contains xylan	Contains cellulose	Unconstrained PC1
Bacteroides	29.9%	26.3%	21.7%	6.9% n.s.	38.8%
Parabacteroides	16.8%	13.7%	12.5%	4.8% n.s.	28.2%
Dysgonomonas	15.9%	13.2%	12.6%	5.4% n.s.	25.2%
Alistipes	15.5%	12.6%	12.2%	4.5% n.s.	24.1%
Odoribacter	18.5%	16.4%	11.4%	5.5% n.s.	30.7%
Paludibacteraceae	16.8%	12.5%	10.9%	4.7% n.s.	24.9%
Clostridiaceae	25.4%	16.9%	23%	4.3% n.s.	31.1%
Lachnospiraceae A	28%	18.1%	26.7%	4.6% n.s.	31.1%
Lachnospiraceae B	17.6%	13%	16.2%	4.4% n.s.	21.5%
Lachnospiraceae C	19.3%	13.7%	17.7%	4.7% n.s.	22.5%
Oscillospiraceae 2	17.2%	13.8%	13.8%	4.6% n.s.	21.4%
Oscillospiraceae 1	18.4%	14.3%	14.3%	4.7% n.s.	22.4%
Enterococcaceae	15.9%	10.6%	15.2%	7.2%	21.2%
Fibrobacterota	53.7%	46.9%	22.7%	12.6%	64.5%
Desulfovibrio	11.2%	9.1%	7.8%	4.8% n.s.	19.3%
Desulfosarcina	15%	11.7%	8.2%	3.6% n.s.	35.9%
Fusobacterium	13%	10.9%	9.7%	5.1% n.s.	17.8%

All components are significantly different by ANOVA unless marked. n.s.: no significance

Redundancy analysis (RDA) from the R package vegan was applied to identify how well transcriptome variance was explained by dietary factors [51]. Predictably, stratifying by "diet" as a factor explained the most variation for all pangenomes but especially for *Fibrobacterota* (53.7%) and *Bacteroides* (29.85%), although constrained ordination did not capture the entirety of variation explained by unconstrained PCA (**Supplemental Figure S4.5)**. We also tested models based on the presence vs absence of the fibers (e.g. 100% cellulose vs all other diets) and

a numeric model of percent xylan. We observed phylum-dependent differences in how the models fit each pangenome; all *Bacillota* were more strongly separated (or equally separated for the *Oscillospiraceae* groups) by the addition of xylan in any amount to a diet, while all other pangenomes were better described by the numeric percent of xylan the diets contained (**Table 4.2**). Despite this difference, all pangenomes shared an interesting pattern: the presence or absence of xylan explained far more variation than that of cellulose, which only explained significant variation in *Enterococcaceae* (ANOVA: $F = 2.19$, $p < 0.01$) and *Fibrobacterota* (ANOVA: $F = 4.03$, $p < 0.05$). Overall, these results indicate there is a greater impact of xylan on the activity of gut microbiota that manifests in a phylum-specific manner.

Changes in metabolic gene expression across diets

To examine the metabolic impacts of diets containing xylan or cellulose as their only carbohydrate source, we analyzed differential expression of transcripts assigned to KEGG orthologs both at the individual KO level as well as aggregated by pathway. Across most taxa, xylan enriched for KOs involved in pathways of galactose degradation and pentose interconversions, while cellulose enriched for C5-branched dibasic acid metabolism **(Figure 4.3A)** and histidine metabolism (**Figure 4.3B)**. However, a notable trend was that many of the pangenomes within the *Bacillota* upregulated pathways associated with amino acid metabolism in cockroaches fed a cellulose diet, while many of the pathways upregulated in xylan fed cockroaches were associated with carbohydrate metabolism. *Clostridiaceae* showed the strongest pattern, increasing eight carbohydrate pathways on the xylan diet and eight amino acid pathways on the cellulose diet **(Figure 4.3),** followed by *Oscillospiraceae 1*. The other *Bacillota* enriched for a mix of pathways, but still showed a general tendency to increase expression of pathways associated with protein processing on cellulose and carbohydrates on xylan. *Desulfovibrio*

upregulated several pathways (C5-branched dibasic acid, lysine, tryptophan, arginine, and proline degradation) only on the cellulose diet but other organisms displayed altered behavior on both diets.

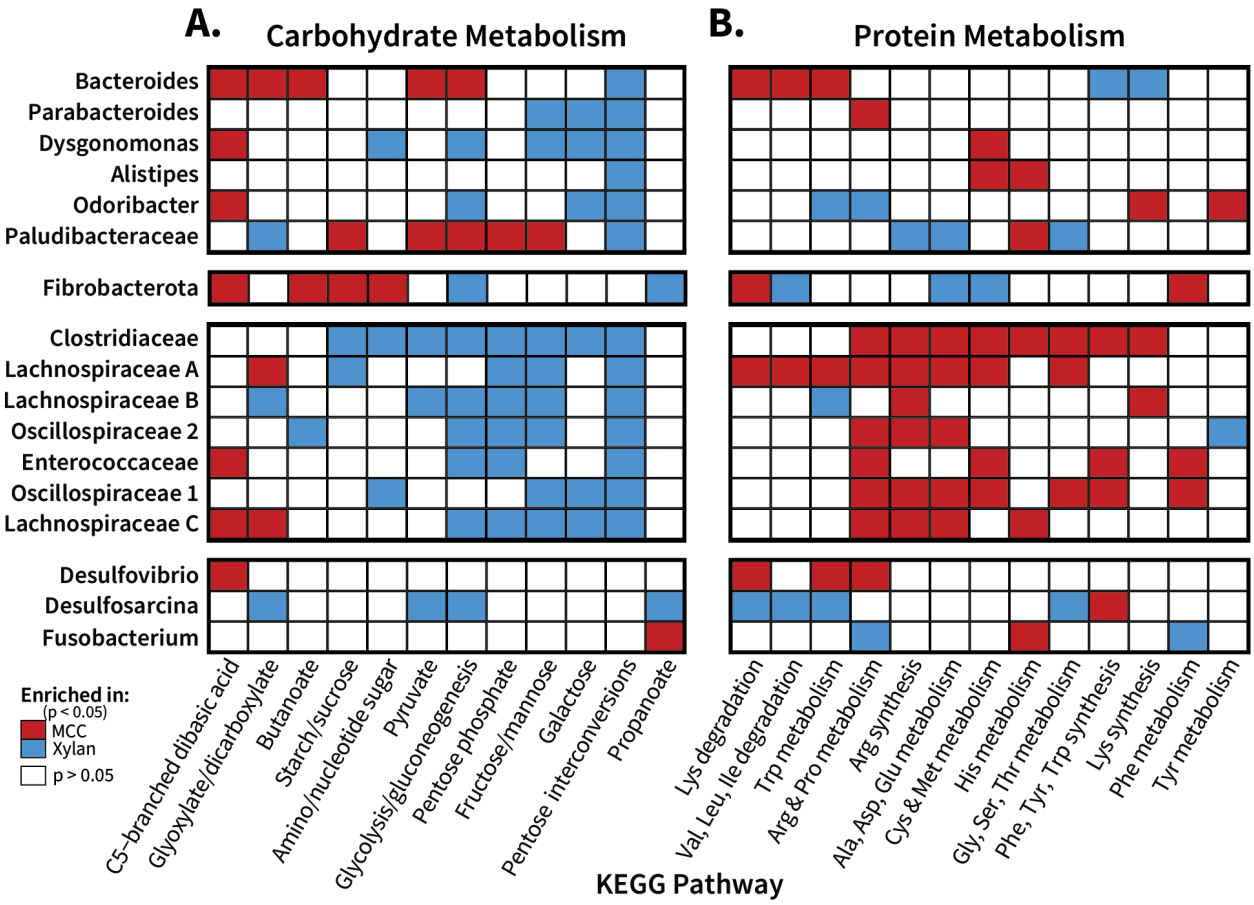


Figure 4.3: Pangenome organisms display fiber-dependent shifts in major metabolic pathways. Enrichment of (**A**) carbohydrate and (**B**) protein metabolic KEGG pathways was assessed for the pangenome transcriptomes of 100% xylan- and 100% MCC-fed insects. Gene clusters assigned relevant KOs were aggregated together based on pathway membership, which was then analyzed with DESeq2 at the pathway level to determine functional enrichment (adjusted $p < 0.05$) between the two fiber sources. All orthologs listed for the individual pathways were included into the overall pathway count, even if they appeared in multiple metabolic pathways.

In addition to pairwise comparisons between the single-fiber diets, we also explored changes in transcriptional activity for each pangenome across the ratio diets. Pangenome gene

expression per ratio diets of KEGG pathways associated with carbohydrate metabolism and amino acid metabolism are displayed in **Supplemental Figures S4.7** and **S4.8,** respectively. Here, heatmap color indicates the relative abundance of transcripts assigned to each pathway within the individual pangenome's transcriptome, while size indicates relative abundance of these transcripts out of all transcripts per sample/diet.

Across the ratios, *Bacteroides* transcripts associated with both carbohydrate (**Supplemental Figure S4.6A**) and protein (**Supplemental Figure S4.7A**) increased in overall abundance with increasing xylan content, consistent with the increase in total reads assigned to these organisms (**Figure 4.2**)**.** However, when expressed in terms of relative abundance within the pangenome, *Bacteroides* transcribed more carbohydrate-associated genes in the cellulose diet, with the proportion of transcripts decreasing as xylan percent increased. Only pentose interconversion genes were both upregulated by *Bacteroides* on the xylan diet while also increasing in total abundance as dietary xylan percentage increased (**Figure 4.3**). Other *Bacteroidota* members showed less linear shifts in the overall abundance of carbohydrate metabolic pathways (**Supplemental Figure S4.6A**); of these, *Dysgonomonas* overall activity shared the positive correlation with xylan percent observed in *Bacteroides*, but differed in enrichment patterns. *Dysgonomonas* enriched for five pathways on xylan (nucleotide sugar; glycolysis; fructose/mannose; galactose; pentose interconversions) and transcribed its only cellulose-enriched pathway (C5-dibasic acid) at low levels both overall and within its transcriptome in all diet treatments. *Parabacteroides* and *Alistipes* maintained similar activity levels across all diets, although *Parabacteroides* did enrich for three sugar-related pathways on the xylan diet while *Alistipes* enriched only for pentose conversions. In contrast, *Odoribacter* and *Paludibacteraceae* displayed their highest activity levels on 100% cellulose, which dropped

sharply once xylan percent reached 50%. *Paludibacteraceae* concurrently downregulated its expression in five pathways as xylan percent increased, matching its overall abundance shifts.

Amino acid metabolism pathways were generally transcribed at lower levels than the carbohydrate pathways among *Bacteroidota*, with less variation in expression observed over the diet gradient for most of the pangenomes (**Supplemental Figure S4.7A**). *Bacteroides* tryptophan metabolism showed a distinct switch from degradation on the 100% and 86% cellulose diet to synthesis on the 50% mix and xylan-majority diets, with a similar pattern in lysine metabolism. Among the other *Bacteroiodota* pangenomes, *Odoribacter* and *Paludibacteraceae* differentially expressed four pathways each, but generally amino acid pathway transcription remained consistent across the diet gradient.

Pangenomes within *Bacillota* exhibited both the highest overall activity and expression of carbohydrate pathway genes in mixed diets rather than either pure diet, particularly *Clostridiaceae, Lachnospiraceae,* and (to a lesser degree) *Enterococcaceae* (**Supplemental Figure S4.6B**). Carbohydrate metabolism in this group was largely enriched by xylan, but they displayed a clear preference for a mix of fibers over the 100% xylan diet. The *Oscillospiraceae* groups shifted their relative transcriptional activity based on diet, with carbohydrate pathways only enriched by xylan, but their absolute abundance of these pathways remained stable across the ratio diets with the exception of pentose interconversions and fructose/mannose metabolism in *Oscillospiraceae* group 1. Interestingly, while *Enterococcaceae* followed the general *Bacillota* trend of higher transcriptional activity on mixed diets, the opposite was true for galactose metabolism, where it had higher overall and relative transcription levels on both single fibers than it did on any mixed diet. Potentially, this group targets host peritrophic matrix glycans that

contain galactose during the single-fiber diets, while on mixed diets *Enterococcaceae* takes advantage of the larger pool of metabolic intermediates instead of degrading host glycans.

Bacillota pangenomes showed similar patterns of overall transcript abundance for pathways associated with amino acid metabolism as they did carbohydrate metabolism, with the mixed diets resulting in higher overall activity levels for *Clostridiaceae* and *Lachnospiraceae* (**Supplemental Figure S4.7B)**. However, in terms of relative abundance within the pangenome, they transcribed pathways associated with protein metabolism at higher levels on the cellulose diet than any of the xylan containing diets. This trend is particularly marked for arginine and proline metabolism genes, which declined significantly as xylan was added to the diet for all *Bacillota* excluding *Lachnospiraceae B*.

Of the other phyla investigated, *Fibrobacterota* showed the clearest relationship between overall transcript abundance and cellulose in pathway expression of both carbohydrate (**Supplemental Figure S4.6C**) and amino acid metabolism (**Supplemental Figure S4.7C**). *Desulfobacterota* showed similar preferences for cellulose, although their expression levels decreased more gradually and remained higher in total over the gradient if diets than *Fibrobacterota*. *Fusobacterium* showed minimal alterations in either absolute abundance or relative transcriptional activity. In general, pangenomes showed phylum-distinct correlations with diet in carbohydrate and amino acid metabolism, indicating that *Bacillota* benefit more from a mixed-fiber environment while *Bacteroidota* possess a more linear relationship with the proportion of their choice fiber source.

Expression of genes for carbohydrate active enzymes

To evaluate the carbohydrate-degrading capabilities of the cockroach microbiome, we evaluated CAZyme diversity and abundance across the pangenomes. Overall, cockroach gut microbiota expressed an array of CAZyme families, with each pangenome encoding between 43 (*Fusobacterium*) and 255 (*Clostridiaceae*) unique CAZyme families (**Table 4.1)**. Glycoside hydrolases (GH) comprised the majority of CAZyme classes identified, followed by glycosyl transferases and polysaccharide lyases (**Supplemental Figure S4.8A).** In general, the diversity of CAZymes expressed did not correlate with transcript abundance, either within the metatranscriptome or as a fraction of transcripts within a given pangenome. By far, *Bacteroides* produced the greatest overall abundance of CAZyme transcripts*,* followed by the Bacteroidota groups *Parabacteroides* and *Dysgonomonas.* In contrast, while *Clostridiaceae* and *Lachnospiraceae B* **(***Bacillota***)** encoded the largest number of CAZyme families (**Supplemental Figure S4.8A)** and CAZyme-annotated gene clusters (**Supplemental Figure S4.8B)**, they produced fewer CAZyme transcripts, both in total and as a proportion of their transcriptional activity **(Figure 4.4A, Table 4.1)**. However, not all Bacteroidota expressed CAZymes at a high level; CAZymes made up <5% of total transcriptional activity among *Alistipes, Odoribacter* ,and *Paludibacteraceae* **(Table 4.1)**. While all *Bacteroidota* and nearly all *Bacillota* transcribed more glycoside hydrolases than any other class, the *Desulfobacterota* pangenomes and *Fusobacterium* instead produced higher levels of glycosyl transferases (**Figure 4.4A**), consistent with a focus instead on carbohydrate modification rather than degradation. *Oscillospiraceae* group 2 deviated from the *Bacillota*, transcribing S-layer homology domains at higher levels than glycoside hydrolases. While not directly carbohydrate active, these proteins are often associated in *Bacillota* with CAZyme attachment to the cell wall.

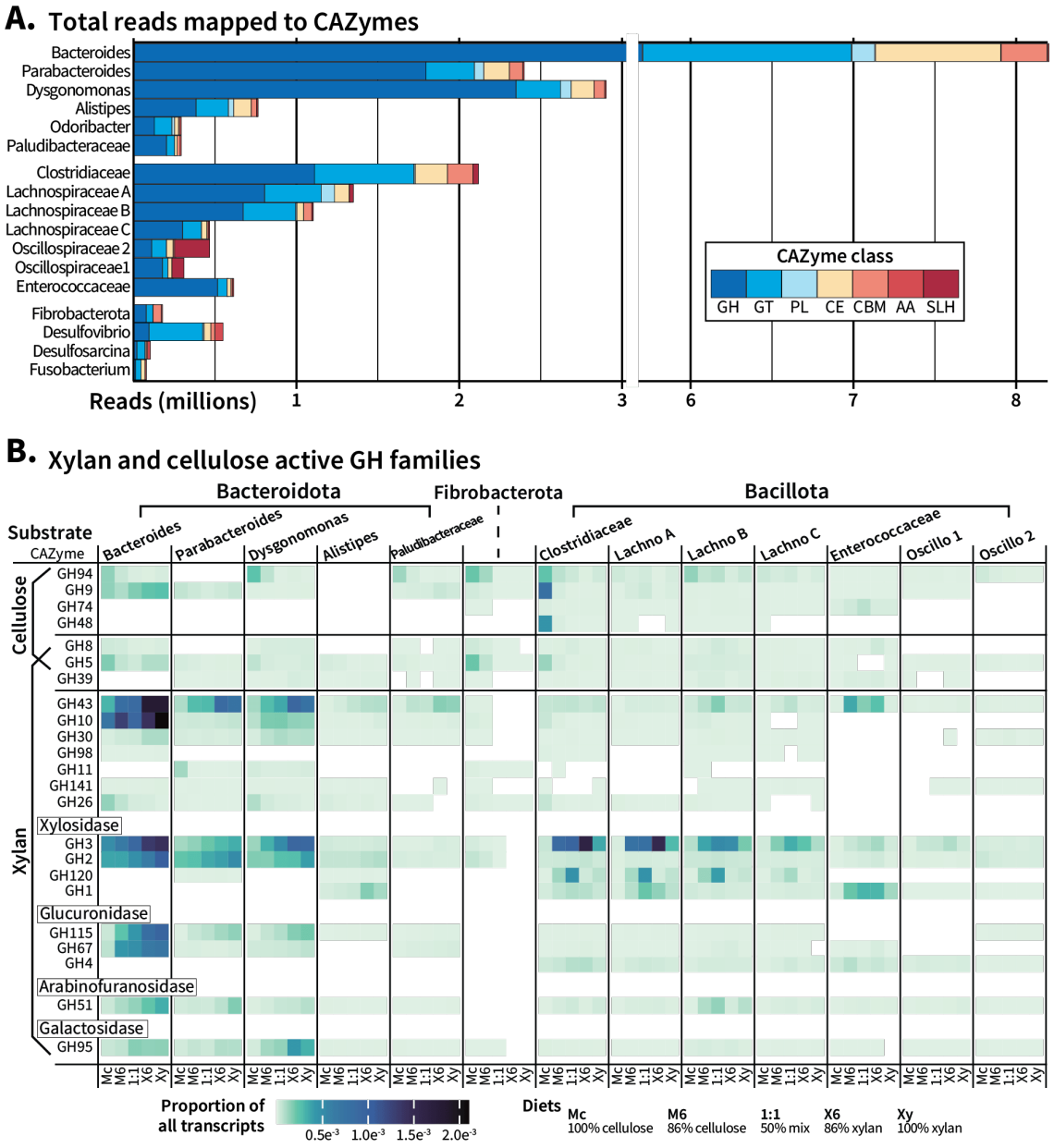


Figure 4.4: Glycoside hydrolase expression in pangenomes across the spectrum of fiber ratios. (A) The absolute abundance of CAZyme classes in each pangenome was calculated based on gene cluster transcripts (x 1,000,000) annotated within each class. **(B)** Gene clusters assigned to glycoside hydrolases (GH) associated with cellulose or xylan were aggregated and visualized with a heatmap scaled by proportion of these GH families out of all transcripts within the diets. Mc: 100% microcrystalline cellulose (MCC) diet; M6: 86% MCC diet; 1:1: 50% MCC/xylan diet; Xy: 100% xylan diet; X6: 86% xylan diet; Lachno: *Lachnospiraceae*; Oscillo: *Oscillospiraceae;* GH: glycoside hydrolase; GT: glycosyl transferase; PL: polysaccharide lyase; CE: carbohydrate esterase; AA: auxiliary activity; CBM: carbohydrate binding module; SLH: S-layer homology domain

To determine how the pangenomes responded to cellulose, xylan, and mixtures of the two as the primary dietary carbon source, we focused on CAZymes known to target these fibers, with special attention given to GH expression (**Figure 4.4B)**. We identified GH families in the pangenomes that were identified as cellulases (GH94, GH9, GH74, GH48), xylanases (GH98, GH43, GH30, GH11, GH10), or contained activities for both (GH8, GH5, GH39) according to a substrate specificity file (**Supplemental Table 4.4**) downloaded from the dbCAN2 server [72].

Of the cellulases, GH94 was highly transcribed by members of *Bacteroidota* (*Bacteroides, Dysgonomonas, Paludibacteraceae*), *Fibrobacterota,* and *Bacillota* (*Clostridiaceae, Lachnospiraceae B)* on the 100% cellulose diet, with expression levels decreasing as the dietary cellulose:xylan ratio decreased. GH9 was highly expressed in *Bacteroides,* but did not seem to consistently correlate with dietary cellulose levels in the *Bacteroidota;* only *Clostridiaceae* produced more of this enzyme in the cellulose diet. While GH8, 5, and 39 were classified as processing both fibers, only GH5 expression differed across the diets. GH5 was transcribed heavily by *Fibrobacterota* and *Clostridiaceae* on the cellulose diet, with *Bacteroides* expressing gene clusters with this CAZyme at both the 100% cellulose and xylan diets; the GH5 family contains several subfamilies with targets including arabinan, chitin, and β-mannan that may be the source of this variation (**Supplemental Table 4.4)**. Since xylan is a complex and heterogenous polysaccharide with wide variance in the side chain composition depending on its source, we divided xylan-associated glycoside hydrolases based on their most probable targets (**Figure 4.4B)**. Breakdown of xylan requires the successive and cooperative activity of glycosidases that remove sugar residues, endo-xylanases that cleave inner xylose-xylose bonds, and exo-xylanases or xylosidases that remove terminal xylose or xylooligosaccharides.

CAZyme families associated with side chain removal include hydrolases targeting glucuronic acid (GH115, GH67, GH4), arabinose (GH51), and galactose (GH95) side chains. Most of these enzymes were produced at high levels by *Bacteroides, Parabacteroides,* and *Dysgonomonas,* with the exception of GH4, which was absent in all *Bacteroidota.* Of the two glucuronidases present in this phylum, enzymes from family GH115 increased for these three pangenomes as xylan increased, while GH67 expression only responded to dietary xylan in *Bacteroides.* Of the other sugars, *Dysgonomonas* transcribed the most galactosidase (GH95), which was positively correlated with dietary xylan, but only marginally increased in arabinofuranosidase activity. Rather, *Parabacteroides* and *Bacteroides* produced steeper diet-associated gradients of GH51 family enzyme expression. In contrast, *Bacillota* pangenomes in general did not express CAZymes associated with hydrolyzation of these side chains at high levels or in response to dietary xylan. *Lachnospiraceae* B upregulated GH51 family enzymes in the mixed diets while *Clostridiaceae* and *Enterococcaceae* increased GH4 family enzymes targeting glucuronic acid, but the shifts in transcription for these enzymes were subtle and their expression was generally low in *Bacillota* pangenomes independent of diet.

Endo-xylanases are found in CAZyme families GH141, GH98, GH43, GH30, GH26, GH11, and GH10. Of these, GH43 stands out with the strongest correlations to dietary xylan percent, with all *Bacteroidota* members showing gradual increases in expression as percent xylan increased. In *Bacteroides,* GH10 shows similar upregulation correlated with diet, while *Dysgonomonas* appears to maintain more sustained levels once xylan is added to the diet in any amount. Curiously, GH11 and GH26 presented opposite patterns than expected, with higher levels found on the cellulose diet that suggest these enzymes are targeting glucans rather than xylan. In *Bacillota*, GH43 expression is increased on the mixed diets in *Enterococcaceae*,

Lachnospiraceae B, and *Lachnospiraceae C,* but otherwise xylanases in this phylum are expressed at low levels, if at all.

Finally, xylosidases are associated with the removal of terminal xylose during breakdown of shorter, lower complexity xylo-oligosaccharides. *Bacteroidota* primarily expressed xylosidases from families GH3 and GH2 in the characteristic gradient-like fashion, with *Dysgonomonas* producing more xylosidases than they did xylanases. *Bacteroides* and *Parabacteroides* produced high levels of GH3 xylosidases but expression of these enzymes remained below their GH43 xylanase expression. *Bacillota,* unlike with xylanases, did express GH3 xylosidases at high levels and significantly upregulated their expression in response to increasing dietary xylan content, which dropped once cellulose was absent in the diet. In addition to GH3, *Clostridiaceae* and *Lachnospiraceae* produced GH120 xylosidases, which consistently peaked in expression at the 50% mixed diet then dropped once xylan was the majority fiber.

Overall, based on CAZyme profiles obtained for these pangenomes, *Bacteroidota* in the cockroach gut microbiome are responsible for primary xylan deconstruction as well as the majority of side chain removal while *Bacillota* liberate terminal xylose residues.

Phylum level differences in core carbon metabolic activity

To decipher the specific contributions of cockroach gut microbiota to xylan and cellulose degradation and how these different groups utilized the metabolic products, we compared phylum-level expression of KEGG orthologs catalyzing steps involved in carbohydrate catabolism and central carbon metabolism (**Figure 4.5**). The KEGG modules for glycolysis (M00001), pentose phosphate pathway (M00004), and the Entner-Doudoroff shunt (M00308) were incorporated into a single reaction map with curated KOs related to glucuronic-

arabinoxylan and microcrystalline cellulose degradation. For KEGG ortholog IDs and their fiber-based enrichment for the individual pangenomes, please see **Supplemental Figure 4.10**.

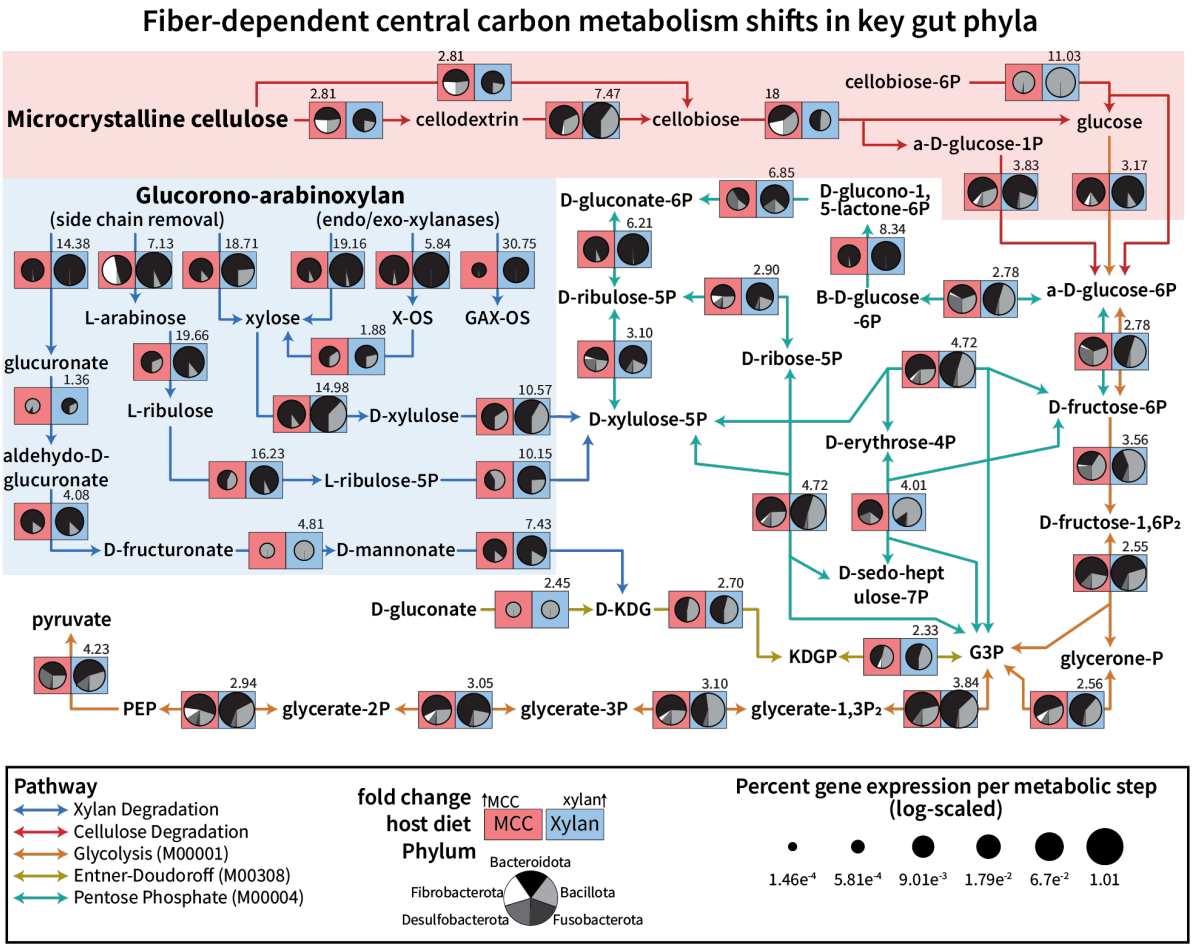


Figure 4.5: Key cockroach phyla demonstrate flexibility in central carbon metabolism when shifting between degradation of xylan and cellulose. Pangenome gene clusters containing KEGG orthologs involved in glycolysis, the Entner-Doudoroff pathway, and the pentose phosphate pathway, as well as curated KOs involved in xylan and cellulose degradation (see **Supplemental Figure 10**), were converted based on enzymatic step to their relative abundance of total transcripts and averaged by diet. Enzymatic steps within the pangenomes were summed together by phylum, and the total contribution of *Bacteroidota*, *Bacillota*, *Desulfobacterota, Fusobacterota,* and *Fibrobacterota* to each step was visualized with pie charts. The size of the pie charts indicates the overall transcriptional percent that step comprises in the 100% cellulose (red box) and 100% xylan (blue box) diets, with the number indicating the fold increase in relative abundance in the direction of either cellulose (left corner) or xylan (right corner). Pathways are indicated with arrow color, and the transparent boxes are included to focus attention on xylan degradation steps (blue) and cellulose degradation steps (red). MCC = microcrystalline cellulose; G3P = glyceraldehyde-3-phosphate; PEP = phosphoenolpyruvate

Cellulose is associated with significant increases in few KEGG orthologs, which can be split into three main activity types: endoglucanase (K01179), beta-glucosidase (K05349, K05350), and cellobiose phosphorylase (K00702). Endoglucanase activity, which produces cellodextrins or cellobiose from crystalline cellulose, was higher in the cellulose diet than xylan with much of the increase driven by *Fibrobacterota* activity, although several *Bacteroidota* and *Bacillota* pangenomes did show cellulose-based enrichment as well on this step. Cellobiose phosphorylase activity increased 18 fold on the cellulose diet, again driven by *Fibrobacterota* and supported by enrichment of fiber-degrading pangenomes. Orthologs for beta-glucosidases that may cleave cellodextrins into a glucose plus cellobiose were included, but only *Fibrobacterota* and *Oscillospiraceae* group 2 enriched for one of these KOs on cellulose; K05349 was enriched by xylan in six *Bacillota* organisms and three *Bacteroidota*, explaining the larger portion of transcripts found in the hemicellulose diet compared to cellulose (**Supplemental Figure S4.9**). Glycolysis genes were higher in the xylan diet than cellulose, but this may be due to the high activity and abundance of *Bacteroides* biasing the overall numbers. Notably, *Desulfobacterota* glycolytic activity was correlated strongly with the cellulose diet, but neither of those pangenomes could degrade cellulose or cellobiose on their own.

During xylan degradation, endo-xylanase activity was almost exclusively performed by *Bacteroidota* (K01181, K15924, K01198), while *Bacillota* showed a large proportion of xylosidase activity (K01811), consistent with the CAZyme profiles presented in **Figure 4.4**. *Bacillota* then converted xylose to D-xylulose (xylose isomerase; K01805) and phosphorylated the intermediate (xylulokinase; K00854) for funneling into the pentose phosphate pathway for central carbon metabolism. *Bacteroidota* did cleave xylose residues off to feed into central metabolism as well, but additionally demonstrated increased arabinose (K01209, K15921) and

glucoronate (K01235) removal, with the arabinose feeding into the pentose phosphate pathway through conversion to L-ribulose (K01804), L-ribulose-5P (K00853), and finally D-xylulose-5P (K03077). *Bacillota* pangenomes did enrich for arabinose cleavage and conversion (**Supplemental Figure S4.9**), but to a lesser degree than *Bacteroidota* and less so than they did xylose. Once D-xylulose-5P is obtained, these two phyla demonstrated different strategies; *Bacteroidota* increased their transcription of enzymes guiding the pentose into glycolysis as glucose, while *Bacillota* seemed to prefer transcribing higher pentose phosphate pathway genes that lead to intermediary glyceraldehyde-3-phosphate.

None of the pangenomes encoded a complete Entner-Doudoroff shunt, but glucuronate removed from glucuronoxylan could be converted to 2-keto-3-deoxy-gluconate (KDG) to form the intermediate glyceraldehyde-3-phosphate (G3P). While glucuronate conversion steps were inconsistently expressed by these pangenomes, both *Bacillota* and *Bacteroidota* contribute similar proportions of transcripts to this shunt and increased their transcription on the xylan diet, suggesting they may leverage this pathway as needed. Taken together, these results shed light on the strategies leveraged by a complex community of gut microbiota to degrade dietary fibers and incorporate the component sugars into their central carbon metabolism.

Discussion

Dietary fiber is an important driver of gut microbiome structure, and diets containing purified fibers enable the individual contributions of microbial community members to polysaccharide degradation to be teased apart even within complex gut systems. This work sought to identify the mechanisms behind large-scale compositional shifts observed previously in cockroach gut microbiota when exposed to synthetic diets containing purified fibers as the sole carbohydrate source [43]. To do this, we fed cockroaches synthetic diets containing as the

carbohydrate source xylan, microcrystalline cellulose, or a mixture of both polysaccharides in differing ratios to detect fiber-dependent microbial responses and determine how mixed fibers influence these dynamics. We tested the impact of these diets on the hindgut microbiome composition using 16S rRNA gene amplicon sequencing, followed by functional characterization of the microbiota through metatranscriptomic analysis. Leveraging an organism-centric pangenome approach, we characterized the metabolic strategies enriched in these organisms by the individual fibers and discovered phylum-dependent responses to the different ratios of xylan:cellulose in mixed fiber diets.

Compositional Overview

As previously observed [43], the hindgut microbiota from cockroaches fed xylan or cellulose differ in diversity and taxonomic composition of 16S rRNA profiles; xylan-fed communities contained lower alpha diversity (**Figure 4.1A-C**) than cellulose-fed communities and enriched for *Lactobacillales* and *Lachnospirales* in contrast to the *Christensenelalles* and *Oscillospirales* increased by the cellulose diet (**Figure 4.1F**). Expanding on these results using mixed-fiber diets, we found that diets containing mixtures of dietary xylan and cellulose select for a gradient in gut microbiome composition that reflects the influence of both fibers (**Figure 4.1**). However, the relationship between the dietary xylan:cellulose ratio and gut community composition was not linear. Mixed diets showed substantially higher alpha diversity than the 100% xylan regardless of their own xylan content (**Figure 4.1A-C**). In contrast, xylan content strongly influenced beta diversity measures, particularly those that consider taxon abundance, with clear separation of 100% cellulose diets from even low-xylan mixed diets (**Figure 4.1D-E**).

We compared these 16S rRNA results with the taxonomic compositions of our metatranscriptomes (**Figure 4.2A, S3**), which we calculated based on top-scoring hits obtained

from read alignment to both the nonredundant RefSeq protein database and cockroach-associated single cell genomes [34]. Alpha and beta diversity of metatranscriptome compositions reflected the core differences observed between the 100% xylan and 100% cellulose diets (**Supplemental Figure S4.3**): xylan communities were characterized by low alpha diversity and clearly separated from the cellulose community during Bray-Curtis ordination. We did observe some differences between the two datasets in their relative abundance of major bacterial orders (**Figure 4.2A**), partially due to the different taxonomies utilized by the different data processing methods (SILVA taxonomy for 16S rRNA gene vs NCBI taxonomy for metatranscriptomes) in the case of *Christensenellales* and *Oscillospirales*. Other differences came from *Lactobacillales,* which were identified as abundant in the 16S dataset but not the metatranscriptome, and in *Bacteroidales,* which displayed substantially higher activity in the metatranscriptome than expected based on 16S rRNA gene abundance. While the large increase in *Bacteroidales* abundance in the metatranscriptomes indicates that this group is highly transcriptionally active, it is unclear whether the diminished *Lactobacillales* population reflects a real population with low transcriptional activity, or if there are differences in the taxonomic assignment of transcripts vs 16S rRNA amplicons. Similar loss of *Lactobacillaceae* in metagenome results despite apparent 16S rRNA presence has been reported in the cockroach [34], suggesting there may be misannotation of the amplicon sequences.

Interestingly, the mixed-fiber metatranscriptomes seemed to have stronger majority-fiber affiliations in their alpha and beta diversity outcomes than they did in the amplicon dataset. We simplified the two 75% fiber diets in the 16S dataset to a single 50% mix for metatranscriptomics so this diet does not have a direct 16S comparison, but within the metatranscriptomes this mixture produced wildly variable community diversity measures (**Supplemental Figure 4.3**).

Whether this variability is an artifact of our samples or due to destabilizing effects of the fiber source is difficult to day; literature on the functional implications of mixing purified fiber sources in the complex microbiome of a living host is limited [77], while *in vitro* studies produce divergent results. Previous work modeling responses of human donor microbiota found in one case that individual fiber complexity was more impactful than mixing multiple (3 or 6) fiber sources [25] and in another, that multiple fibers increased diversity moreso than their component fibers [78].

Pangenome Activity

A major challenge in metatranscriptome analyses lies in transcript mapping and reference genome selection. This is a particular challenge for less-studied models such as the cockroach, where relatively few gut taxa have been cultured and sequenced. To overcome this challenge, we used an approach that mapped transcripts to pangenomes constructed from cockroach-originated single cell genomes [34] and closely related reference genomes from other environments. [61, 75]. In all diets, most reads were successfully assigned to one of 17 pangenomes of cockroach gut symbionts (**Figure 4.2B**), allowing us to analyze the fiber-dependent functional behavior of these organisms via their individual "transcriptomes", as well as identify phylum-wide patterns when faced with a gradient of fibers.

Further, this approach allows us to discern true differences that arise in the microbiota due to dietary fiber type. While the frequency of ASVs enriched per each genus may suggest a preference towards xylan or cellulose, abundant ASVs enriched by the two different fibers in the 16S rRNA dataset notably overlap in in taxonomic origin (**Supplemental Figure 4.2**). Major organisms with ASVs enriched by both fibers include Bacteroidota members *Bacteroides, Dysgonomonadaceae,* and *Parabacteroides,* while *Alistipes* and *Odoribacter* had members

enriched by xylan or the mixed fibers, all of which were investigated using our pangenome approach and discussed below. *Odoribacter* in particular shows increased activity on cellulose in metatranscriptomic analysis despite its enrichment by xylan in the 16S data set. Enriched *Bacillota* in the 16S data on both fiber sources include *Lachnospiraceae* members, *Ruminococcus*, and *Christensenellaceae,* although many more members were enriched by xylan than cellulose. *Ruminococcus* and *Christensenellaceae* are included in the *Oscillospiraceae* pangenome groups but ended up being less important in *Bacillota* activity; rather, *Clostridiaceae*, which were not represented in the 16S, are the dominant *Bacillota* group in the metatranscriptome. *Desulfobacterota* and *Fibrobacterota* were enriched by cellulose in both 16S abundance and in transcriptomic activity. Through the pangenomes, we are able to confirm the dynamics suggested by 16S rRNA gene abundance, while clarifying apparent discrepancies that arose, highlighting the value of this approach.

Bacteroidota

Bacteroidota are among the most abundant organisms found in omnivorous gut communities, including omnivorous cockroaches [32, 35, 43, 79, 80]. Here, we analyzed six *Bacteroidota* pangenomes found abundantly in the cockroach hindgut: *Bacteroides, Parabacteroides, Dysgonomonas, Alistipes, Odoribacter,* and *Paludibacteraceae.* Among these organisms, we found *Bacteroides* to be especially active regardless of dietary fiber source (**Figure 4.2B, Table 4.1**), consistent with their proposed role as major fiber-degrading players in the cockroach gut [34, 35]. In human gut systems, *Bacteroides* are particularly well-known for their ability to break down an array of fibers including xylan [81, 82], starch [83], host glycans [84, 85], cellulose [35], and pectin [86] due to their large assortment of encoded carbohydrate-active enzymes (CAZymes) and the use of polysaccharide utilization loci (PULs) to maximize

substrate compatibility. We observed the same CAZyme diversity and expression here, with *Bacteroides* transcribing 175 unique CAZyme families (**Supplemental Figure S4.8A**) at very high levels (**Figure 4.4A**), vastly outnumbering CAZyme transcription by all other organisms. Two other *Bacteroidota* groups, *Dysgonomonas* and *Parabacteroides,* were the second and third largest source of CAZyme transcripts, consistent with a role as primary fiber degraders. *Bacteroides, Dysgonomonas,* and *Parabacteroides* transcriptional profiles suggest a major role in metabolism of dietary xylan. These three *Bacteroidota* groups expressed CAZyme families that are active on common xylan branches, including orthologs responsible for cleaving and subsequently processing arabinose residues, and also displayed primary chain xylanase activity, which together argue towards their ability to utilize complex xylan structures (**Figure 4.4B**). Ordination analysis of transcriptional profiles from all three groups showed a strong response to diet (**Supplemental Figure S4.5A**), including a strong significant response to dietary xylan percentage and, to a lesser extent, presence/absence of xylan, but not presence/absence of cellulose. This, together with their high expression of xylan degradation genes and much weaker expression of cellulolytic genes suggests a strong preference for xylan rather than cellulose as a growth substrate. Given these results, we hypothesize that these groups are the primary xylan degraders in the cockroach gut. The persistence of these organisms in cockroaches fed a pure cellulose diet plus their enrichment of cellulase and cellobiose phosphorylase transcripts suggest that they may be able to metabolize cellulose if necessary. However, inferring this definitively is difficult given their persistence during long-term host starvation [32] and their ability to leverage host glycans for sustenance [84].

Bacillota

While *Bacillota* are commonly identified in omnivorous gut communities, they are usually classified as secondary fermenters rather than primary degraders [34, 80, 87]. Our results are consistent with a role as secondary fermenters of xylan breakdown products; *Bacillota* organisms were generally associated with dietary xylan content in their 16S rRNA gene abundance with more enriched ASVs than in cellulose (**Supplemental Figures S4.1, S4.2**), and pangenome transcriptional activity enriched overall by xylan presence (**Figure 4.2, Supplemental Figure S4.6, S4.8**) but primarily produced xylosidases rather than endo-xylanases or glycosidases active on side chains (**Figure 4.4B**). This was correlated with xylan-enriched transcription of xylose isomerase, xylulokinase, and pentose phosphate pathway orthologs feeding into G3P production (**Figure 4.5, S4.10**), illustrating the xylose-utilizing potential of Firmicutes [88, 89]. Interestingly, *Bacillota* transcriptional profiles were better explained by models considering absence/presence of xylan rather than dietary xylan percentage, consistent with a large metabolic response to dietary xylan at even the lowest abundance tested. *Bacillota* displayed clear preferences for xylan as a carbohydrate source, upregulating between 3-8 carbohydrate degradation pathways per organism, while on the cellulose diet they enriched instead for numerous amino acid metabolic pathways (**Figure 4.3),** suggesting a switch to utilization of dietary amino acids/proteins in cellulose diets, rather than secondary fermentation of cellulosic substrates. There are cellulose-active *Bacillota* found in herbivorous communities, such as *Ruminococcus* species [90, 91] or cellulosome-producing *Clostridium* [92], but the *Clostridiaceae* here showed higher preference for xylan (**Figure 4.2B, Supplemental Figure S4.6, S4.8**), and while *Oscillospiraceae* 2 was slightly more abundant on cellulose than xylan (**Supplemental Figure S4.6**) it shared the KEGG pathway enrichment patterns observed in the

other *Bacillota* pangenomes (**Figure 4.3**). Generally, *Bacillota* pangenomes had relatively similar metabolic activity allocations across the different xylan-containing diets, but surprisingly, displayed distinctly higher overall transcriptional activity in mixed-fiber diets (**Supplemental Figure S4.6, S4.8**). Although the apparent switch to amino acid-degrading vs. carbohydrate metabolic genes in cellulose fed cockroaches suggests that they may not be as actively involved in cellulose fermentation as xylan, they may still benefit from intermediates being released by cellulose-degrading organisms, either as catabolic substrates or as anabolic precursors for biomass production [93].

Fibrobacterota

Unexpectedly, we discovered a significant increase in *Fibrobacterota* transcripts in the 100% cellulose diet. This small phylum *contains* well-characterized members of ruminant and herbivore monogastric gut communities that are rarely identified or studied in omnivorous animals [94, 95], although we did identify a small presence of this group in our previous 16S rRNA amplicon survey of *P. americana [43]*. *Fibrobacterota* are remarkably specialized towards cellulose metabolism, encoding a unique suite of genes that facilitate fiber degradation by releasing cellulases and hemicellulases into the surrounding area prior to transporting cellodextrins into the cell for degradation [96-98]. In line with this, we did observe cellulose-associated CAZyme transcription (**Figure 4.4B**), in addition to high expression of cellulase ortholog transcription in the cellulose diet (**Figure 4.5**). Of the pangenomes assessed, *Fibrobacterota* were one of only 2 organisms where cellulose presence/absence significantly impacted overall transcriptional profile of that organism (**Table 4.2**), and unconstrained ordinations showed strong separation of transcriptional activity under high-cellulose (86% and 100%) diets. The sharp shift in transcriptional activity in diets with 50% or greater xylan content

suggest that gut microbial xylan metabolism may interfere with cellulose degradation by Fibrobacterota. These orga*nisms* are known to be highly sensitive to catabolite repression, and environmental pH [99]; the bloom in *Bacillota*, for example, likely altered altered gut pH or metabolite profiles [100, 101].

When comparing pangenome activity patterns in carbohydrate metabolic pathways across diet ratios (**Supplemental Figure S4.6**), we noticed that four additional organisms displayed their highest activity on cellulose: the *Desulfobacterota* pangenomes *Desulfosarcina* and *Desulfovibrio*, and *Bacteroidota* members *Paludibacteraceae* and *Odoribacter*. With the exception of *Desulfovibrio*, these were among the lower activity organisms in our metatranscriptomes, with little fiber-degrading activity themselves (**Table 4.1**: <5% of expressed gene clusters labeled as CAZymes), leading us to consider that *Fibrobacterota* may be supporting their higher growth in this diet. One characteristic of *Fibrobacterota* in rumen and gut communities is its role in cross-feeding; these organisms commonly provide neighbors with cellobiose and hemicellulose-derived oligo- and monosaccharides produced by their activity for the sole purpose of increasing cellulose accessibility, and they release metabolic waste products succinate and formate [102-105]. We did observe that *Fibrobacterota* enriched transcription increased transcription of succinate dehydrogenase in the cellulose diet (K00241, K00240, K00239; **Supplemental Figure S4.9, S4.11**), suggesting a similar function in cockroach-derived species. *Desulfobacterota* have previously been established as hydrogen and formate cross-feeders associated with cellulose degrading communities, strengthening the possibility of a relationship with cellulose-degrading *Fibrobacterota* [106-108]. *Paludibacteraceae* expressed higher GH94 family CAZymes on the cellulose diet, which typically has cellobiose phosphorylase activity (**Figure 4.4D**) and this organism showed similar substantial drops in

carbohydrate activity as seen in *Fibrobacterota* as cellulose percent decreased (**Supplemental Figure S4.6**). This suggests that *Paludibacteraceae* may benefit from the cellobiose liberated by *Fibrobacterota.* Finally, *Odoribacter* has not been previously reported to be an effective fiber degrader, but it has been hypothesized to consume succinate in mouse models [109, 110], and may be utilizing succinate produced by *Fibrobacterota*. While *Fibrobacterota* is an uncommon member of omnivore microbiomes, these findings suggest they may play similar roles in the cockroach gut as they do in herbivores, mediating cross-feeding interactions that increase and maintain diversity in complex gut systems.

Conclusions

In conclusion, through 16S rRNA gene profiling combined with RNA-seq, we discovered that xylan and cellulose differentially influence the community structure and function of the American cockroach hindgut microbiome. A noteworthy observation from this study was that the composition predicted by 16S rRNA gene sequencing was not necessarily reflected in the transcriptional activity of the gut community. In particular, *Bacteroides* was far more active than estimated by its amplicon abundance, and their transcriptional profile was consistent with a role as primary degraders of xylan despite the fact that they declined in relative abundance as measured by 16S rRNA gene assays of xylan-fed cockroaches. In contrast, *Lachnospirales* exhibited a very strong, significant relationship with xylan as measured by ASV abundance, but transcriptional profiles suggest a secondary role as fermenters of xylosides. In addition, they displayed far higher transcriptional activity in the mixed-fiber diet than on pure xylan. These results stress the importance of combining functional analysis with amplicon surveys to fully grasp the dynamics occurring within a complex gut community.

Through our organism-centric pangenome approach, we identified *Bacteroidota* as likely to be responsible for most xylan degradation in the cockroach gut, with members *Bacteroides*, *Parabacteroides,* and *Dysgonomonas* producing high levels of diverse CAZymes that increased with dietary xylan percent. By mixing these fibers in different ratios, we discovered that *Bacillota* members were enriched by xylan but flourished in abundance and activity when given mixed fibers, benefiting from their own metabolic activity as well as the actions of other fiber-degrading organisms. Finally, we observed surprisingly high *Fibrobacterota* cellulolytic activity on the cellulose diet. *Fibrobacterota* is an uncommon organism in omnivorous gut microbiomes; the ability of the cockroach to host these organisms may contribute to their famous ability to survive long-term on nutrient-poor diets. These results underscore the power of metatranscriptomic approaches combined with synthetic diets to assess gut microbial responses to purified or mixed dietary fibers, and highlight the utility of the omnivorous cockroach as a model to better understand complex microbial interactions in the guts of omnivores.

References

1. Bäckhed, F., et al., *Host-bacterial mutualism in the human intestine.* Science, 2005. **307**(5717): p. 1915-20.
2. Kurilshikov, A., et al., *Host Genetics and Gut Microbiome: Challenges and Perspectives.* Trends Immunol, 2017. **38**(9): p. 633-647.
3. Rangarajan, A.A., et al., *Ruminococcus bromii enables the growth of proximal Bacteroides thetaiotaomicron by releasing glucose during starch degradation.* Microbiology (Reading), 2022. **168**(4).
4. Greene, L.K., et al., *A role for gut microbiota in host niche differentiation.* Isme j, 2020. **14**(7): p. 1675-1687.
5. Wolter, M., et al., *Leveraging diet to engineer the gut microbiome.* Nat Rev Gastroenterol Hepatol, 2021. **18**(12): p. 885-902.
6. David, L.A., et al., *Diet rapidly and reproducibly alters the human gut microbiome.* Nature, 2014. **505**(7484): p. 559-563.
7. Islamova, Z.I., et al., *Comparative Assessment of the Prebiotic Activity of Some Pectin Polysaccharides.* Pharmaceutical Chemistry Journal, 2017. **51**(4): p. 288-291.
8. Scott, K.P., et al., *Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro.* FEMS Microbiol Ecol, 2014. **87**(1): p. 30-40.
9. Broekaert, W.F., et al., *Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides.* Crit Rev Food Sci Nutr, 2011. **51**(2): p. 178-94.
10. Rastall, R.A., et al., *Structure and function of non-digestible carbohydrates in the gut microbiome.* Benef Microbes, 2022. **13**(2): p. 95-168.

11. Flint, H.J., et al., *Microbial degradation of complex carbohydrates in the gut.* Gut Microbes, 2012. **3**(4): p. 289-306.

12. Riva, A., et al., *A fiber-deprived diet disturbs the fine-scale spatial architecture of the murine colon microbiome.* Nature Communications, 2019. **10**(1): p. 4366.

13. Erica, D.S. and L.S. Justin, *Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates.* Cell Metabolism, 2014.

14. Sonnenburg, E.D., et al., *Diet-induced extinctions in the gut microbiota compound over generations.* Nature, 2016. **529**(7585): p. 212-5.

15. Desai, M.S., et al., *A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility.* Cell, 2016. **167**(5): p. 1339-1353.e21.

16. Eric, C.M., N. Mareike, and S.D. Mahesh, *Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier.* Nature Reviews Microbiology, 2018.

17. Bulut, N., et al., *Matrix-entrapped fibers create ecological niches for gut bacterial growth.* Scientific Reports, 2023. **13**(1): p. 1884.

18. Kaur, A., et al., *Physical Inaccessibility of a Resistant Starch Shifts Mouse Gut Microbiota to Butyrogenic Firmicutes.* Mol Nutr Food Res, 2019. **63**(7): p. e1801012.

19. Leitch, E.C., et al., *Selective colonization of insoluble substrates by human faecal bacteria.* Environ Microbiol, 2007. **9**(3): p. 667-79.

20. Delannoy-Bruno, O., et al., *Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans.* Nature, 2021. **595**(7865): p. 91-95.

21. Ara, K., et al., *From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites.* Cell, 2016.

22. Canani, R.B., et al., *Potential beneficial effects of butyrate in intestinal and extraintestinal diseases.* World J Gastroenterol, 2011. **17**(12): p. 1519-28.

23. Abreu y Abreu, A.T., et al., *Dietary fiber and the microbiota: A narrative review by a group of experts from the Asociación Mexicana de Gastroenterología.* Revista de Gastroenterología de México (English Edition), 2021. **86**(3): p. 287-304.

24. Augustin, L.S.A., et al., *Dietary fibre consensus from the International Carbohydrate Quality Consortium (ICQC).* Nutrients, 2020. **12**(9).

25. Chung, W.S.F., et al., *Impact of carbohydrate substrate complexity on the diversity of the human colonic microbiota.* FEMS Microbiology Ecology, 2019. **95**(1): p. fiy201.

26. Rogowski, A., et al., *Glycan complexity dictates microbial resource allocation in the large intestine.* Nature Communications, 2015. **6**(1): p. 7481.

27. Puhlmann, M.-L. and W.M. de Vos, *Intrinsic dietary fibers and the gut microbiome: Rediscovering the benefits of the plant cell matrix for human health.* Frontiers in Immunology, 2022. **13**.

28. Holland, C., et al., *Plant Cell Walls: Impact on Nutrient Bioaccessibility and Digestibility.* Foods, 2020. **9**(2).

29. Engel, P. and N.A. Moran, *The gut microbiota of insects – diversity in structure and function.* FEMS Microbiology Reviews, 2013. **37**(5): p. 699-735.

30. Ma, H., et al., *Morphology and three-dimensional reconstruction of the digestive system of Periplaneta americana.* J Med Entomol, 2009. **46**(1): p. 165-8.

31. Cruden, D.L. and A.J. Markovetz, *Microbial ecology of the cockroach gut.* Annu Rev Microbiol, 1987. **41**: p. 617-43.

32. Tinker, K.A. and E.A. Ottesen, *The Core Gut Microbiome of the American Cockroach, Periplaneta americana, Is Stable and Resilient to Dietary Shifts.* Appl Environ Microbiol, 2016. **82**(22): p. 6603-6610.

33. Tinker, K.A. and E.A. Ottesen, *Differences in Gut Microbiome Composition Between Sympatric Wild and Allopatric Laboratory Populations of Omnivorous Cockroaches.* Front Microbiol, 2021. **12**: p. 703785.

34. Dukes, H.E., K.A. Tinker, and E.A. Ottesen, *Disentangling hindgut metabolism in the American cockroach through single-cell genomics and metatranscriptomics.* Front Microbiol, 2023. **14**: p. 1156809.

35. Vera-Ponce de León, A., et al., *Cultivable, Host-Specific Bacteroidetes Symbionts Exhibit Diverse Polysaccharolytic Strategies.* Appl Environ Microbiol, 2020. **86**(8).

36. Pereira, G.V., et al., *Degradation of complex arabinoxylans by human colonic Bacteroidetes.* Nat Commun, 2021. **12**(1): p. 459.

37. Reichardt, N., et al., *Phylogenetic distribution of three pathways for propionate production within the human gut microbiota.* The ISME Journal, 2014. **8**(6): p. 1323-1335.

38. Sorbara, M.T., et al., *Functional and Genomic Variation between Human-Derived Isolates of Lachnospiraceae Reveals Inter- and Intra-Species Diversity.* Cell Host Microbe, 2020. **28**(1): p. 134-146.e4.

39. Salyers, A.A., et al., *Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon.* Appl Environ Microbiol, 1977. **33**(2): p. 319-22.

40. Smith, N.W., et al., *Hydrogen cross-feeders of the human gastrointestinal tract.* Gut Microbes, 2019. **10**(3): p. 270-288.

41. Wang, T., et al., *Evidence for a multi-level trophic organization of the human gut microbiome.* PLOS Computational Biology, 2019. **15**(12): p. e1007524.

42. Catlett, J.L., et al., *Metabolic Synergy between Human Symbionts Bacteroides and Methanobrevibacter.* Microbiol Spectr, 2022. **10**(3): p. e0106722.

43. Dockman, R.L. and E.A. Ottesen, *Purified fibers in chemically defined synthetic diets destabilize the gut microbiome of an omnivorous insect model.* Frontiers in Microbiomes, 2024. **Volume 3 - 2024**.

44. Pollet, A., J.A. Delcour, and C.M. Courtin, *Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families.* Crit Rev Biotechnol, 2010. **30**(3): p. 176-91.

45. Smith, M.M. and J. Melrose, *Xylan Prebiotics and the Gut Microbiome Promote Health and Wellbeing: Potential Novel Roles for Pentosan Polysulfate.* Pharmaceuticals (Basel), 2022. **15**(9).

46. Tinker, K.A. and E.A. Ottesen, *Phylosymbiosis across Deeply Diverging Lineages of Omnivorous Cockroaches (Order Blattodea).* Appl Environ Microbiol, 2020. **86**(7).

47. Stewart, F.J., E.A. Ottesen, and E.F. DeLong, *Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics.* The ISME Journal, 2010. **4**(7): p. 896-907.

48. Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina amplicon data.* Nat Methods, 2016. **13**(7): p. 581-3.

49. RStudioTeam, *RStudio: Integrated Development for R*. 2020.

50. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic Acids Res, 2013. **41**(Database issue): p. D590-6.

51. Oksanen, J.F., et al., *vegan: Community Ecology Package.* R package version 2.5-6, 2019(https://CRAN.R-project.org/package=vegan).

52. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

53. Sabree, Z.L., S. Kambhampati, and N.A. Moran, *Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont.* Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19521-6.

54. Dockman, R.L., et al., *Genome Report: Improved chromosome-level genome assembly of the American cockroach, Periplaneta americana.* bioRxiv, 2025.

55. Kopylova, E., L. Noé, and H. Touzet, *SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data.* Bioinformatics, 2012. **28**(24): p. 3211-3217.

56. Bushnell, B., J. Rood, and E. Singer, *BBMerge – Accurate paired shotgun read merging via overlap.* PLOS ONE, 2017. **12**(10): p. e0185056.

57. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification.* BMC Bioinformatics, 2010. **11**(1): p. 119.

58. Buchfink, B., K. Reuter, and H.-G. Drost, *Sensitive protein alignments at tree-of-life scale using DIAMOND.* Nature Methods, 2021. **18**(4): p. 366-368.

59. Sayers, E.W., et al., *Database resources of the National Center for Biotechnology Information.* Nucleic Acids Res, 2024. **52**(D1): p. D33-d43.

60. Haft, D.H., et al., *RefSeq and the prokaryotic genome annotation pipeline in the age of metagenomes.* Nucleic Acids Res, 2024. **52**(D1): p. D762-d769.

61. DePoy, A.N., et al., *Microbial transcriptional responses to host diet maintain gut microbiome homeostasis in the American cockroach.* bioRxiv, 2024: p. 2024.10.31.621369.

62. Eren, A.M., et al., *Community-led, integrated, reproducible multi-omics with anvi'o.* Nature Microbiology, 2021. **6**(1): p. 3-6.

63. Delmont, T.O. and A.M. Eren, *Linking pangenomes and metagenomes: the Prochlorococcus metapangenome.* PeerJ, 2018. **6**: p. e4320.

64. Benedict, M.N., et al., *ITEP: An integrated toolkit for exploration of microbial pan-genomes.* BMC Genomics, 2014. **15**(1): p. 8.

65. van Dongen, S. and C. Abreu-Goodger, *Using MCL to extract clusters from networks.* Methods Mol Biol, 2012. **804**: p. 281-95.

66. Eddy, S.R., *Accelerated Profile HMM Searches.* PLOS Computational Biology, 2011. **7**(10): p. e1002195.

67. Galperin, M.Y., et al., *COG database update 2024.* Nucleic Acids Res, 2025. **53**(D1): p. D356-d363.

68. Tatusov, R.L., E.V. Koonin, and D.J. Lipman, *A genomic perspective on protein families.* Science, 1997. **278**(5338): p. 631-7.

69. Kanehisa, M., et al., *Data, information, knowledge and principle: back to metabolism in KEGG.* Nucleic Acids Research, 2014. **42**(D1): p. D199-D205.

70. Kanehisa, M., et al., *KEGG: new perspectives on genomes, pathways, diseases and drugs.* Nucleic Acids Research, 2017. **45**(D1): p. D353-D361.

71. Aramaki, T., et al., *KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold.* Bioinformatics, 2020. **36**(7): p. 2251-2252.

72. Zhang, H., et al., *dbCAN2: a meta server for automated carbohydrate-active enzyme annotation.* Nucleic Acids Research, 2018. **46**(W1): p. W95-W101.

73. Cantarel, B.L., et al., *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.* Nucleic Acids Res, 2009. **37**(Database issue): p. D233-8.

74. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013.* Nucleic Acids Res, 2014. **42**(Database issue): p. D490-5.

75. Dukes, H.E., *Uncovering microbial activities and characteristics underlying cockroach gut microbiome function and homeostasis*, in *Microbiology*. 2023, University of Georgia.

76. Luo, W. and C. Brouwer, *Pathview: an R/Bioconductor package for pathway-based data integration and visualization.* Bioinformatics, 2013. **29**(14): p. 1830-1.

77. Sawicki, C.M., et al. *Dietary Fiber and the Human Gut Microbiota: Application of Evidence Mapping Methodology*. Nutrients, 2017. **9**, DOI: 10.3390/nu9020125.

78. Cantu-Jungles, T.M., et al., *Systematically-designed mixtures outperform single fibers for gut microbiota support.* Gut Microbes, 2025. **17**(1): p. 2442521.

79. Schauer, C., C.L. Thompson, and A. Brune, *The bacterial community in the gut of the cockroach Shelfordella lateralis reflects the close evolutionary relatedness of cockroaches and termites.* Appl Environ Microbiol, 2012. **78**(8): p. 2758-67.

80. Mahowald, M.A., et al., *Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla.* Proc Natl Acad Sci U S A, 2009. **106**(14): p. 5859-64.

81. Zhang, M., et al., *Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes.* Proc Natl Acad Sci U S A, 2014. **111**(35): p. E3708-17.

82. Tuncil, Y.E., et al., *Subtle Variations in Dietary-Fiber Fine Structure Differentially Influence the Composition and Metabolic Function of Gut Microbiota.* mSphere, 2020. **5**(3).

83. Foley, M.H., D.W. Cockburn, and N.M. Koropatkin, *The Sus operon: a model system for starch uptake by the human gut Bacteroidetes.* Cell Mol Life Sci, 2016. **73**(14): p. 2603-17.

84. Martens, E.C., H.C. Chiang, and J.I. Gordon, *Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont.* Cell Host Microbe, 2008. **4**(5): p. 447-57.

85. Chen, L., et al., *Increased mucin-degrading bacteria by high protein diet leads to thinner mucus layer and aggravates experimental colitis.* J Gastroenterol Hepatol, 2021. **36**(10): p. 2864-2874.

86. Luis, A.S., et al., *Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides.* Nat Microbiol, 2018. **3**(2): p. 210-219.

87. Yutin, N. and M.Y. Galperin, *A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia.* Environmental Microbiology, 2013. **15**(10): p. 2631-2641.

88. Gu, Y., et al., *Reconstruction of xylose utilization pathway and regulons in Firmicutes.* BMC Genomics, 2010. **11**: p. 255.

89. Leth, M.L., et al., *Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut.* Nature Microbiology, 2018. **3**(5): p. 570-580.

90. Bule, P., et al., *Assembly of Ruminococcus flavefaciens cellulosome revealed by structures of two cohesin-dockerin complexes.* Sci Rep, 2017. **7**(1): p. 759.

91. Israeli-Ruimy, V., et al., *Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions.* Scientific Reports, 2017. **7**(1): p. 42355.

92. Mohand-Oussaid, O., et al., *The extracellular xylan degradative system in Clostridium cellulolyticum cultivated on xylan: evidence for cell-free cellulosome production.* J Bacteriol, 1999. **181**(13): p. 4035-40.

93. Koropatkin, N.M., E.A. Cameron, and E.C. Martens, *How glycan metabolism shapes the human gut microbiota.* Nature Reviews Microbiology, 2012. **10**(5): p. 323-335.

94. Raut, M.P., et al., *Deciphering the unique cellulose degradation mechanism of the ruminal bacterium Fibrobacter succinogenes S85.* Scientific Reports, 2019. **9**(1): p. 16542.

95. Neumann, A.P., C.A. McCormick, and G. Suen, *Fibrobacter communities in the gastrointestinal tracts of diverse hindgut-fermenting herbivores are distinct from those of the rumen.* Environ Microbiol, 2017. **19**(9): p. 3768-3783.

96. Suen, G., et al., *The complete genome sequence of Fibrobacter succinogenes S85 reveals a cellulolytic and metabolic specialist.* PLoS One, 2011. **6**(4): p. e18814.

97. Jun, H.S., et al., *Outer membrane proteins of Fibrobacter succinogenes with potential roles in adhesion to cellulose and in cellulose digestion.* J Bacteriol, 2007. **189**(19): p. 6806-15.

98. Burnet, M.C., et al., *Evaluating Models of Cellulose Degradation by Fibrobacter succinogenes S85.* PLoS One, 2015. **10**(12): p. e0143809.

99. Roger, V., et al., *Effects of Physicochemical Factors on the Adhesion to Cellulose Avicel of the Ruminal Bacteria Ruminococcus flavefaciens and Fibrobacter succinogenes subsp. succinogenes.* Appl Environ Microbiol, 1990. **56**(10): p. 3081-7.

100. Haindl, R., S. Schick, and U. Kulozik *Influence of Cultivation pH on Composition, Diversity, and Metabolic Production in an In Vitro Human Intestinal Microbiota*. Fermentation, 2021. **7**, DOI: 10.3390/fermentation7030156.

101. Ng Katharine, M., et al., *Single-strain behavior predicts responses to environmental pH and osmolality in the gut microbiota.* mBio, 2023. **14**(4): p. e00753-23.

102. Froidurot, A., et al., *Fibrobacter sp. HC4, a newly isolated strain, demonstrates a high cellulolytic activity as revealed by enzymatic measurements and in vitro assay.* Applied and Environmental Microbiology, 2024. **90**(8): p. e00514-24.

103. Fakih, I., et al., *Dynamic genome-based metabolic modeling of the predominant cellulolytic rumen bacterium Fibrobacter succinogenes S85.* mSystems, 2023. **8**(3): p. e01027-22.

104. Li, Q., J.A. Siles, and I.P. Thompson, *Succinic acid production from orange peel and wheat straw by batch fermentations of Fibrobacter succinogenes S85.* Appl Microbiol Biotechnol, 2010. **88**(3): p. 671-8.

105. Wells, J.E., et al., *Cellodextrin efflux by the cellulolytic ruminal bacterium Fibrobacter succinogenes and its potential role in the growth of nonadherent bacteria.* Appl Environ Microbiol, 1995. **61**(5): p. 1757-62.

106. Martins, M., C. Mourato, and I.A. Pereira, *Desulfovibrio vulgaris Growth Coupled to Formate-Driven H2 Production.* Environ Sci Technol, 2015. **49**(24): p. 14655-62.

107. Kuever, J., F.A. Rainey, and F. Widdel, *Desulfosarcina*, in *Bergey's Manual of Systematics of Archaea and Bacteria*. 2015. p. 1-7.

108. Chidthaisong, A. and R. Conrad, *Pattern of non-methanogenic and methanogenic degradation of cellulose in anoxic rice field soil.* FEMS Microbiology Ecology, 2000. **31**(1): p. 87-94.

109. Huber-Ruano, I., et al., *Orally administered Odoribacter laneus improves glucose control and inflammatory profile in obese mice by depleting circulating succinate.* Microbiome, 2022. **10**(1): p. 135.

110. Fernández-Veledo, S. and J. Vendrell, *Gut microbiota-derived succinate: Friend or foe in human metabolic diseases?* Reviews in Endocrine and Metabolic Disorders, 2019. **20**(4): p. 439-447.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

The American cockroach (*Periplaneta americana)* is a tractable, hardy, and experimentally malleable model organism that hosts a complex gut microbiota and shares lineages of dominant microbes with humans [1-8]. My work has developed high-quality resources for studying this remarkable insect and demonstrated the utility of synthetic diets for observing microbial responses within a highly complex, human-relevant microbiome to precise dietary changes.

The cockroach has a long history of research that continues to grow [9-11]. Its viability as a modern model organism is supported by its susceptibility to genetic manipulation [12-18] and dietary flexibility [1, 19], and uncovering the genetic basis of its survivalist nature may inform on the evolution of other insects [20]. However, the genomes produced prior to my work were insufficient: one genome was fragmented due to the sequencing technology at the time and contained no public annotations [18], while claims about the other available genome did not line up with karyotypical evidence of chromosome structure in the American cockroach [21, 22]. In my second chapter, I describe the chromosomally resolved *P. americana* genome that I, with the help of the USDA-ARS Hilo, Hawaii team under Dr. Scott Geib, sequenced and assembled. Comparing my assembly to the previous *P. americana* genomes as well as available termite and cockroach genomes, I confirmed that my genome included a chromosome that was missing from the former *P. americana* genome and identified orthologs shared by other Blattodea.

Additionally, I found that this cockroach's genome was enriched in genes related to immune and digestive function relative to termites, factors that could be related to the protective societal structures of termites. The highlight of this work is the creation and publication of a valuable resource for the scientific community; currently, this assembly serves as the reference genome on NCBI for the species and includes publicly available annotations from the NCBI Gnomon pipeline. Future directions for this work will include the use of RNA inhibition (RNAi) to investigate the impact of cockroach immune activities on gut microbiome assembly and response to perturbation.

Of all dietary components, dietary fiber is perhaps the most important to gut microbiota due to the host's inability to enzymatically degrade fiber; consequently, fiber is associated with a robust gut microbiome that crumbles when fiber is not available [23-28]. In my third chapter, I developed synthetic diets that were customizable so that the dietary impact on the microbiome from different sources of fiber could be investigated. I discovered that, unlike the nutritionally biased whole food diets that were previously found to have no impact on the cockroach microbiome composition [1], these synthetic diets did generate strong shifts in the microbial community. Through several rounds of dietary testing and 16S rRNA sequencing, I confirmed that the source of the variation was from the polysaccharide source itself rather than the simple sugar composition, and this effect was independent of unbalancing the diet in other nutritional capacities. Dietary xylan was the largest source of variation compared to the other fibers, leading to blooms in Bacillota that destabilized the community structure as seen in network analysis. From this study, we have a blueprint for creating artificial diets with different purified nutritional compounds. I used six complex carbohydrate sources in these diets, but they represent just a fraction of the many polysaccharide structures that can be isolated from natural sources or

synthesized; xylan and pectin are especially interesting carbohydrates due to their widely variable substitution patterns across sources. These diets minimize dietary variation and allow for thorough investigation of the nutrient of interest without necessitating the large nutrient volumes that are required to generate a response in mammalian model systems.

One limitation of the synthetic diets is that they are not designed to study microbial response to intrinsic fiber, which is the structure of cell wall structural polysaccharides that may be part of why the cockroach microbiome remained so stable previously. Other researchers have created matrix-embedded culture systems to mimic the natural structures formed by plant cells, which may be an interesting approach in the future of this lab [29-32].

Naturally, fiber is derived from plant cell walls that are composed of a cellulose scaffold with hemicellulose (xylan), lignin, and pectin interweaved throughout, providing the cell with structural stability while also physically obstructing enzymatic access to the individual polysaccharides. Within this cell matrix, the polysaccharides themselves are complex as well; xylan varies across its many sources in degree of polymerization, sugar residues, and side chain complexity, while cellulose is simplistic in its sugar composition but forms crystalline structures that are difficult to access for hydrolysis. Overall, my work, through molecular biology techniques and bioinformatic analysis, explores the dynamic, complex gut community found in an unexpected omnivorous insect. As xylan and cellulose generated starkly different microbial community compositions in my previous work outlined in chapter 3, microbial responses to these fibers were further explored in chapter 4. Synthetic diets containing xylan and cellulose mixed in different ratios revealed that differences in the gut microbiota abundance and activity were correlated with the fiber proportion, although this manifested differently depending on the microbial group. I found that *Clostridiaceae* and *Lachnospiraceae* bloomed in the mixed-fiber

diets, when before this experiment we assumed *Lachnospiraceae* to prefer xylan; many other researchers likely draw similar erroneous conclusions in fiber studies, since it seems that mixing pure fibers into different ratios for administration to a living host with a complex microbiome is rarely, if ever, done. *Bacteroides,* another group I found to be xylan-enriched in my previous work, produced a more predictable and gradual increase as the percent of xylan increased. These divergent dynamics across the fiber gradients indicate the alternative roles of gut microbiota in fiber degradation that cannot be captured in a single comparison. Metabolic pathway analysis identified further how core cockroach gut bacteria changed their strategy in response to the fiber source; while isolating the carbohydrate source to purified fibers destabilizes gut community structure, the metabolic flexibility I observed likely is maintaining the community composition during natural or complex diets.

Future Directions

Future research building off the work I have completed in this lab has many exciting possible directions. Regarding the carbohydrates I investigated in my third chapter, plans are in place to recreate the remaining synthetic diets for metatranscriptomic profiling in complement with my fourth chapter studies. Several collaborations are underway to expand on the surprising disruptions xylan caused in the gut community structure. Xylan-derived xylo-oligosaccharides are frequently studied as prebiotics, yet lowering the diversity and stability of gut microbiota as seen in my work is in opposition to the goal of prebiotic use. The structure of the exact xylan that I used in my experiments is unclear, but through a collaboration with Dr. Ian Black at the UGA Complex Carbohydrate Research Center (CCRC) we found that there may be more co-purified carbohydrates in the xylan mixture than the data sheets reported, and the ratio of internal to terminal xylose is lower than expected. Ongoing work with structurally complex and chemically

defined xylan varieties, which are being synthesized by Dr. Breeanna Urbanowicz at the UGA CCRC, will help clarify the roles cockroach microbiota play in xylan degradation and refine our knowledge of the interplay between xylan residues and microbial activity within a complex community. Further, efforts are underway using this labeled xylan to determine the spatial distribution of xylan branch removal and chain hydrolysis throughout the digestive tract, in concert with linkage analysis performed by Dr. Black tracking polysaccharide structure alterations as the carbohydrate proceeds through the gut. Long-term goals aim to investigate different pectin varieties in a similar manner.

In addition to the work related to xylan, experiments with Dr. Black will identify the precise glycomic composition of the peritrophic matrix. Substantial evidence supports the role of mucins secreted by the intestine in supporting gut bacteria in mammals, but less is known about insect peritrophic matrices [24, 33-35]. Starved cockroaches have a similar gut composition to those fed whole foods [1], implying the gut microbiota are surviving on host glycans. However, the exact structure of the cockroach peritrophic matrix has not been defined, restricting our ability to associate microbial enzymatic activity with host glycan degradation. Obtaining this structural data will allow for us to confirm similar roles of insect gut microbiota in host glycan degradation as found in mammalian microbiota and better understand how these microbes maintain their population during nutritional stress or starvation.

Other mechanisms microbes use to survive host starvation can be investigated using my research as a starting point. My initial interest when developing the synthetic diets was to understand how the gut microbiome of *P. americana* relates to uric acid, the nitrogenous waste product of protein metabolism, that collect in the fat body [36-42]. Using RNAi combined with low- and high-protein synthetic diets, I was able to modify xanthine dehydrogenase and uricase

production by the cockroach and identify phenotypic shifts in urate storage. What this means for the cockroach gut microbiome is still unclear. If the host is unable to release uric acid for *Blattabacterium* to subsequently convert into amino acids, there may be a larger impact on gut microbiome stability during periods of nutritional stress. Perhaps, stable isotope probing will allow for nitrogen to be traced from the fat body to the gut microbiome through raising germ free nymphs on labeled proteins, such as spirulina, and observing microbial incorporation of these isotopes when introduced into the nymph.

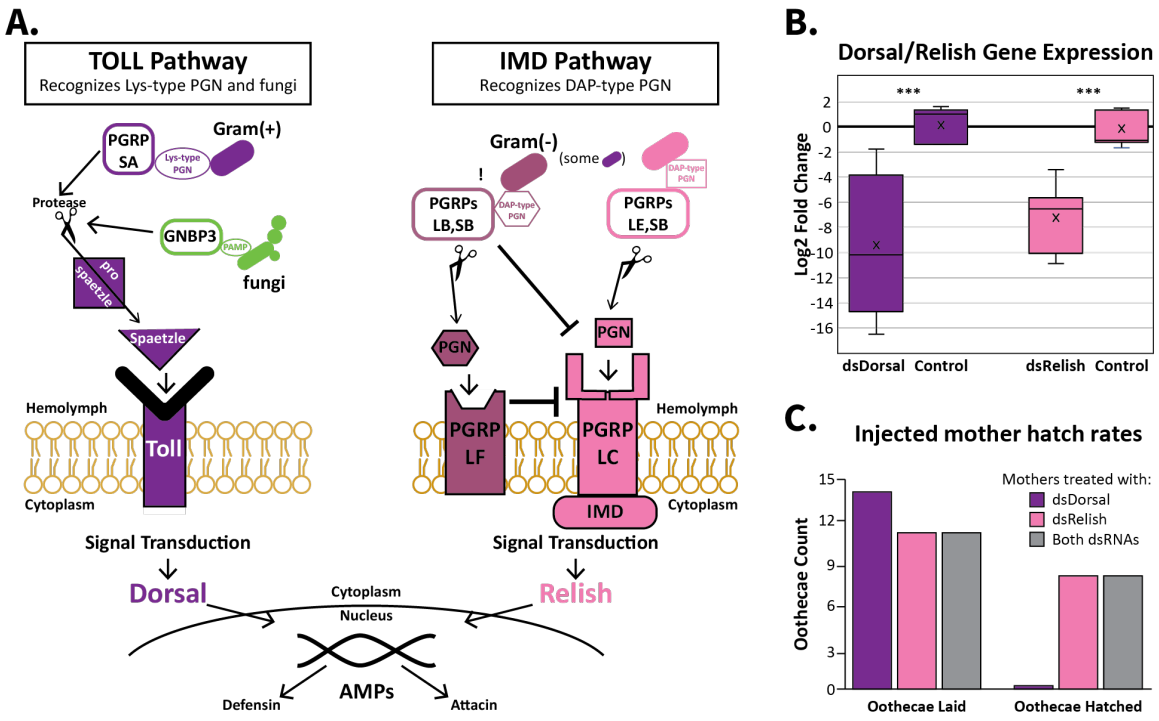


Figure 5.1: Immune pathway regulation in the American cockroach. (A) depicts the two primary insect immune pathways and the proteins involved in their regulation. (B) Boxplot of fold change in gene expression for Dorsal and Relish from RNAi experiments in adult insects. (C) Number of oothecae obtained during RNAi experiments and the oothecae that successfully emerged.

Finally, future work based on my development of RNAi protocols in *P. americana* may further characterize the relationship between the insect immune system and gut microbial colonization. During my work, I have successfully knocked down expression of transcription

factors Dorsal and Relish in adult insects, but no changes in gut community composition were detected. Possibly, since adults no longer molt and therefore do not shed their fore- and hindguts, the immune system ceases to be a key regulatory entity for maintaining gut microbiota. This hypothesis can be explored through a combination of germ-free insect technique and RNAi targeting these transcription factors, as well as additional immunity-associated genes annotated in my genome assembly. Interestingly, it may even be possible to obtain knockdown nymphs without administering dsRNA to them directly. During my RNAi experiments with Dorsal, which happens to be a protein necessary for embryonic development, I discovered that oothecae dropped by injected females failed to develop, indicating that RNAi-induced gene knockdown in cockroaches may be heritable. Investigating how inhibition of immune-related genes and transcription factors influences the initial colonization of germ-free nymphs may provide a fascinating snapshot of immune-mediated host-microbe interactions that dictate the gut microbial population.

Overall, my work characterizing the host and the microbiome of a complex gut community, through dietary modification, molecular biology techniques, and bioinformatic analysis, uncovered unique microbial dynamics in response to pure and mixed fibers that have not been thoroughly investigated in a living host. This work, combined with the genome I assembled for *Periplaneta americana*, will support the growth of this insect as a remarkable model organism, and increase our knowledge of microbial interactions underlying the structure of omnivorous microbiomes.

References

1. Tinker, K.A. and E.A. Ottesen, *The Core Gut Microbiome of the American Cockroach, Periplaneta americana, Is Stable and Resilient to Dietary Shifts.* Appl Environ Microbiol, 2016. **82**(22): p. 6603-6610.

2. Cruden, D.L. and A.J. Markovetz, *Microbial ecology of the cockroach gut.* Annu Rev Microbiol, 1987. **41**: p. 617-43.

3. Vera-Ponce de León, A., et al., *Cultivable, Host-Specific Bacteroidetes Symbionts Exhibit Diverse Polysaccharolytic Strategies.* Appl Environ Microbiol, 2020. **86**(8).

4. Bignell, D.E., *An experimental study of cellulose and hemicellulose degradation in the alimentary canal of the American cockroach.* Canadian Journal of Zoology, 1977. **55**: p. 579-589.

5. Huttenhower, C., et al., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. **486**(7402): p. 207-214.

6. *The Integrative Human Microbiome Project.* Nature, 2019. **569**(7758): p. 641-648.

7. Fackelmann, G., et al., *Gut microbiome signatures of vegan, vegetarian and omnivore diets and associated health outcomes across 21,561 individuals.* Nature Microbiology, 2025. **10**(1): p. 41-52.

8. Kurilshikov, A., et al., *Large-scale association analyses identify host factors influencing human gut microbiome composition.* Nat Genet, 2021. **53**(2): p. 156-165.

9. Roeder, K.D., *The effect of potassium and calcium on the nervous system of the cockroach, Periplaneta americana.* J Cell Comp Physiol, 1948. **31**(3): p. 327-38.

10. Bodenstein, D., *Studies on nerve regeneration in Periplaneta americana.* J Exp Zool, 1957. **136**(1): p. 89-115.

11. Troppmann, B., B. Walz, and W. Blenau, *Pharmacology of serotonin-induced salivary secretion in Periplaneta americana.* J Insect Physiol, 2007. **53**(8): p. 774-81.

12. Ciudad, L., X. Bellés, and M.D. Piulachs, *Structural and RNAi characterization of the German cockroach lipophorin receptor, and the evolutionary relationships of lipoprotein receptors.* BMC Mol Biol, 2007. **8**: p. 53.

13. Wang, K., et al., *Variation in RNAi efficacy among insect species is attributable to dsRNA degradation in vivo.* Insect Biochem Mol Biol, 2016. **77**: p. 1-9.

14. Lin, Y.H., et al., *Oral delivery of dsRNA lipoplexes to German cockroach protects dsRNA from degradation and induces RNAi response.* Pest Manag Sci, 2017. **73**(5): p. 960-966.

15. Huang, J.H., et al., *Practical Use of RNA Interference: Oral Delivery of Double-stranded RNA in Liposome Carriers for Cockroaches.* J Vis Exp, 2018(135).

16. Hennenfent, A., et al., *RNA interference supports a role for Nanchung–Inactive in mechanotransduction by the cockroach, Periplaneta americana, tactile spine.* Invertebrate Neuroscience, 2020. **20**(1): p. 1.

17. Li, L., et al., *Applications of RNA Interference in American Cockroach.* J Vis Exp, 2021(178).

18. Li, S., et al., *The genomic and functional landscapes of developmental plasticity in the American cockroach.* Nature Communications, 2018. **9**(1): p. 1008.

19. Dockman, R.L. and E.A. Ottesen, *Purified fibers in chemically defined synthetic diets destabilize the gut microbiome of an omnivorous insect model.* Frontiers in Microbiomes, 2024. **Volume 3 - 2024**.

20. *THE AMERICAN COCKROACH. Edited by William J. Bell and K. G. Adiyodi.* Physiological Entomology, 1982. **7**(1): p. 117-118.

21. John, B. and K.R. Lewis, *Chromosome structure in Periplaneta americana.* Heredity, 1960. **15**(1): p. 47-54.

22. Wang, L., et al., *Genome assembly and annotation of Periplaneta americana reveal a comprehensive cockroach allergen profile.* Allergy, 2023. **78**(4): p. 1088-1103.

23. Tanes, C., et al., *Role of dietary fiber in the recovery of the human gut microbiome and its metabolome.* Cell Host Microbe, 2021. **29**(3): p. 394-407.e5.

24. Flint, H.J., et al., *Microbial degradation of complex carbohydrates in the gut.* Gut Microbes, 2012. **3**(4): p. 289-306.

25. Jefferson, A. and K. Adolphus, *The Effects of Intact Cereal Grain Fibers, Including Wheat Bran on the Gut Microbiota Composition of Healthy Adults: A Systematic Review.* Frontiers in Nutrition, 2019. **6**.

26. Abreu y Abreu, A.T., et al., *Dietary fiber and the microbiota: A narrative review by a group of experts from the Asociación Mexicana de Gastroenterología.* Revista de Gastroenterología de México (English Edition), 2021. **86**(3): p. 287-304.

27. Riva, A., et al., *A fiber-deprived diet disturbs the fine-scale spatial architecture of the murine colon microbiome.* Nature Communications, 2019. **10**(1): p. 4366.

28. Desai, M.S., et al., *A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility.* Cell, 2016. **167**(5): p. 1339-1353.e21.

29. Augustin, L.S.A., et al., *Dietary fibre consensus from the International Carbohydrate Quality Consortium (ICQC).* Nutrients, 2020. **12**(9).

30. Holland, C., et al., *Plant Cell Walls: Impact on Nutrient Bioaccessibility and Digestibility.* Foods, 2020. **9**(2).

31. Puhlmann, M.-L. and W.M. de Vos, *Intrinsic dietary fibers and the gut microbiome: Rediscovering the benefits of the plant cell matrix for human health.* Frontiers in Immunology, 2022. **13**.

32. Bulut, N., et al., *Matrix-entrapped fibers create ecological niches for gut bacterial growth.* Scientific Reports, 2023. **13**(1): p. 1884.

33. Martens, E.C., H.C. Chiang, and J.I. Gordon, *Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont.* Cell Host Microbe, 2008. **4**(5): p. 447-57.

34. Glover, J.S., T.D. Ticer, and M.A. Engevik, *Characterizing the mucin-degrading capacity of the human gut microbiota.* Scientific Reports, 2022. **12**(1): p. 8456.

35. Chen, L., et al., *Increased mucin-degrading bacteria by high protein diet leads to thinner mucus layer and aggravates experimental colitis.* J Gastroenterol Hepatol, 2021. **36**(10): p. 2864-2874.

36. Park, M.S., P. Park, and M. Takeda, *Roles of fat body trophocytes, mycetocytes and urocytes in the American cockroach, Periplaneta americana under starvation conditions: an ultrastructural study.* Arthropod Struct Dev, 2013. **42**(4): p. 287-95.

37. Mullins, D.E. and D.G. Cochran, *Nitrogen excretion in cockroaches: uric Acid is not a major product.* Science, 1972. **177**(4050): p. 699-701.

38. Mullins, D.E. and D.G. Cochran, *Nitrogen metabolism in the american cockroach—I. An examination of positive nitrogen balance with respect to uric acid stores.* Comparative Biochemistry and Physiology Part A: Physiology, 1975. **50**(3): p. 489-500.

39. Mullins, D.E. and D.G. Cochran, *Nitrogen metabolism in the American cockroach: an examination of whole body and fat body regulation of cations in response to nitrogen balance.* J Exp Biol, 1974. **61**(3): p. 557-70.

40. Cochran, D.G., D.E. Mullins, and K.J. Mullins, *Cytological Changes in the Fat Body of the American Cockroach, Periplaneta americana, in Relation to Dietary Nitrogen Levels.* Annals of the Entomological Society of America, 1979. **72**(2): p. 197-205.

41. COCHRAN, D.G., *URIC ACID ACCUMULATION IN YOUNG AMERICAN COCKROACH NYMPHS.* 1979. **25**(2): p. 153-157.

42. Latorre, A., et al., *Of Cockroaches and Symbionts: Recent Advances in the Characterization of the Relationship between Blattella germanica and Its Dual Symbiotic System.* Life (Basel), 2022. **12**(2).

APPENDIX A

Chapter 2 Supplemental Information

File uploaded separately:

AppendixASupplementalFile.xlsx

Supplemental File S2.1: GO annotations and orthogroups for genes annotated in this genome (*P. americana*.)

Table S2.1: Table summarizing repeats identified by RepeatModeler in *P. americana* genome.

RepeatModeler and RepeatMasker Output

	Elements (n)	Length (Mb)	Percent (%)
Total Retroelements	**2364225**	**647.91**	**20.05**
SINEs	656509	112.1	3.47
Penelope	134060	33.05	1.02
LINEs	1615375	505.28	15.63
CRE/SLACS	4803	1.86	0.06
L2/CR1/Rex	431509	124.43	3.85
R1/LOA/Jockey	63736	26.12	0.81
R2/R4/NeSL	110731	36.15	1.12
RTE/Bov-B	596056	201.7	6.24
L1/CIN4	105530	37.16	1.15
LTR elements	92341	30.53	0.94
BEL/Pao	26445	5.55	0.17
Ty1/Copia	4703	1.41	0.04
Gypsy/DIRS1	54482	22.9	0.71
Retroviral	0	0	0
Total DNA Transposons	**2918518**	**785.38**	**24.3**
hobo-Activator	799615	254.82	7.88
Tc1-IS630-Pogo	1257068	304.69	9.43
MULE-MuDR	33964	9.84	0.3
PiggyBac	40357	10.56	0.33
Tourist/Harbinger	54352	17.53	0.54
Other	105971	22.03	0.68
Rolling-circles	235926	62.23	1.93
Unclassified	2276913	486.43	15.05
Small RNA	625209	106.1	3.28
Satellites	3110	4.41	0.14
Simple repeats	1768812	125.36	3.88
Low complexity	161181	10.23	0.32

Table S2.2: OrthoFinder statistics

	Overall	BGer	CFor	CSec	DPun	PAm2022	PAMFEO	ZNev
Genes (n)	132597	22330	12983	13170	28416	26568	16749	12381
Orthogroups (n)	16711	8982	9566	10110	11686	7579	10975	10065
Genes in OGs (n)	114385	17026	12341	12765	21389	22805	15963	12096
Genes in OGs (%)	86.3	76.2	95.1	96.9	75.3	85.8	95.3	97.7
Unassigned genes (n)	18212	5304	642	405	7027	3763	786	285
Unassigned genes (%)	13.7	23.8	4.9	3.1	24.7	14.2	4.7	2.3
Species-specific OGs (n)	3272	881	73	80	771	1220	198	49
Genes in species-specific OGs (n)	15270	3907	270	336	3453	6382	744	178
Genes in species-specific OGs (%)	11.5	17.5	2.1	2.6	12.2	24	4.4	1.4

OG: orthogroup; BGer: *Blattella germanica*; CFor: *Coptotermes formosanus*; CSec: *Cryptotermes secundus;* DPun: *Diploptera puntata;* PAm: *Periplaneta americana; ZNev: Zootermopsis nevadensis; PAMFEO: P. americana* assembly from this paper

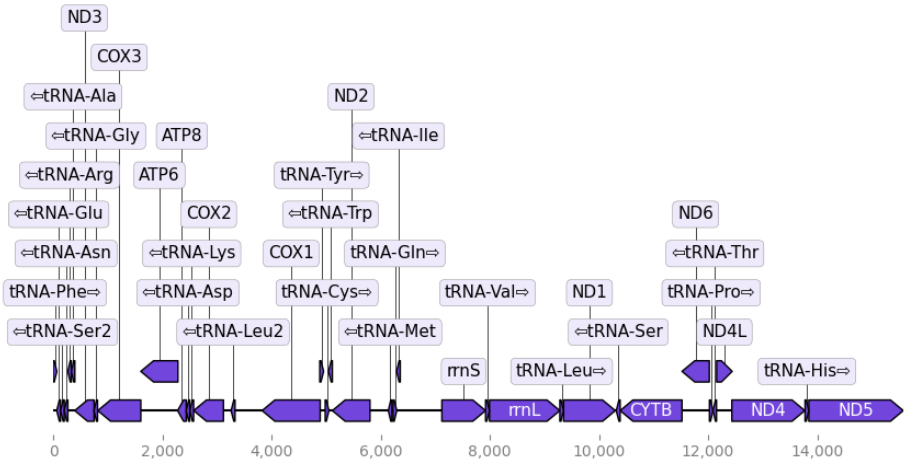


Figure S2.1: Mitogenome annotation output from MitoHiFi

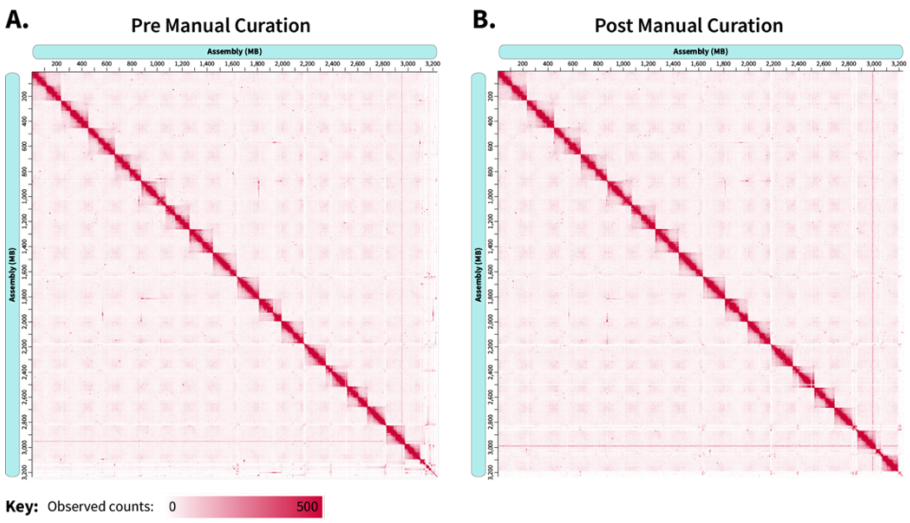


Figure S2.2: Hi-C contact maps of the assembled genome **(A)** before and **(B)** after manual curation using Juicebox.

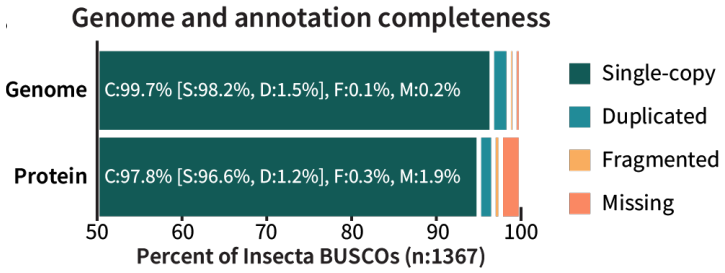


Figure S2.3: BUSCOs identified in this *P. americana* genome assembly and protein annotation.

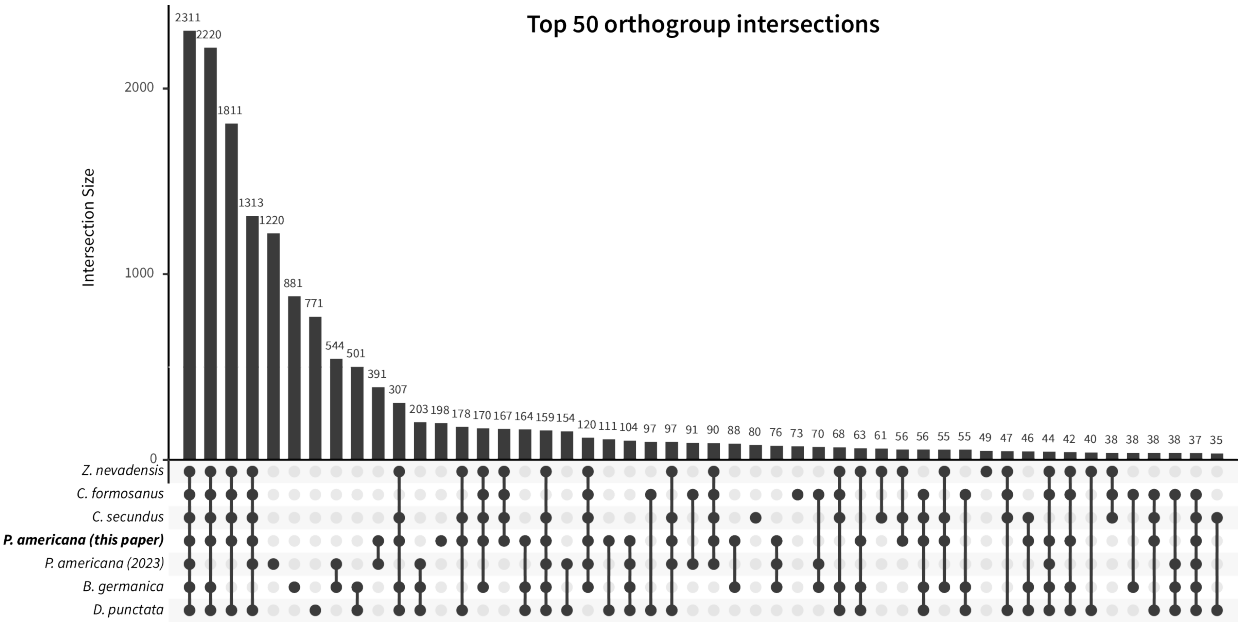


Figure S2.4: UpSet plot of orthogroup overlap between Blattodea species when previous *P. americana* assembly is included.

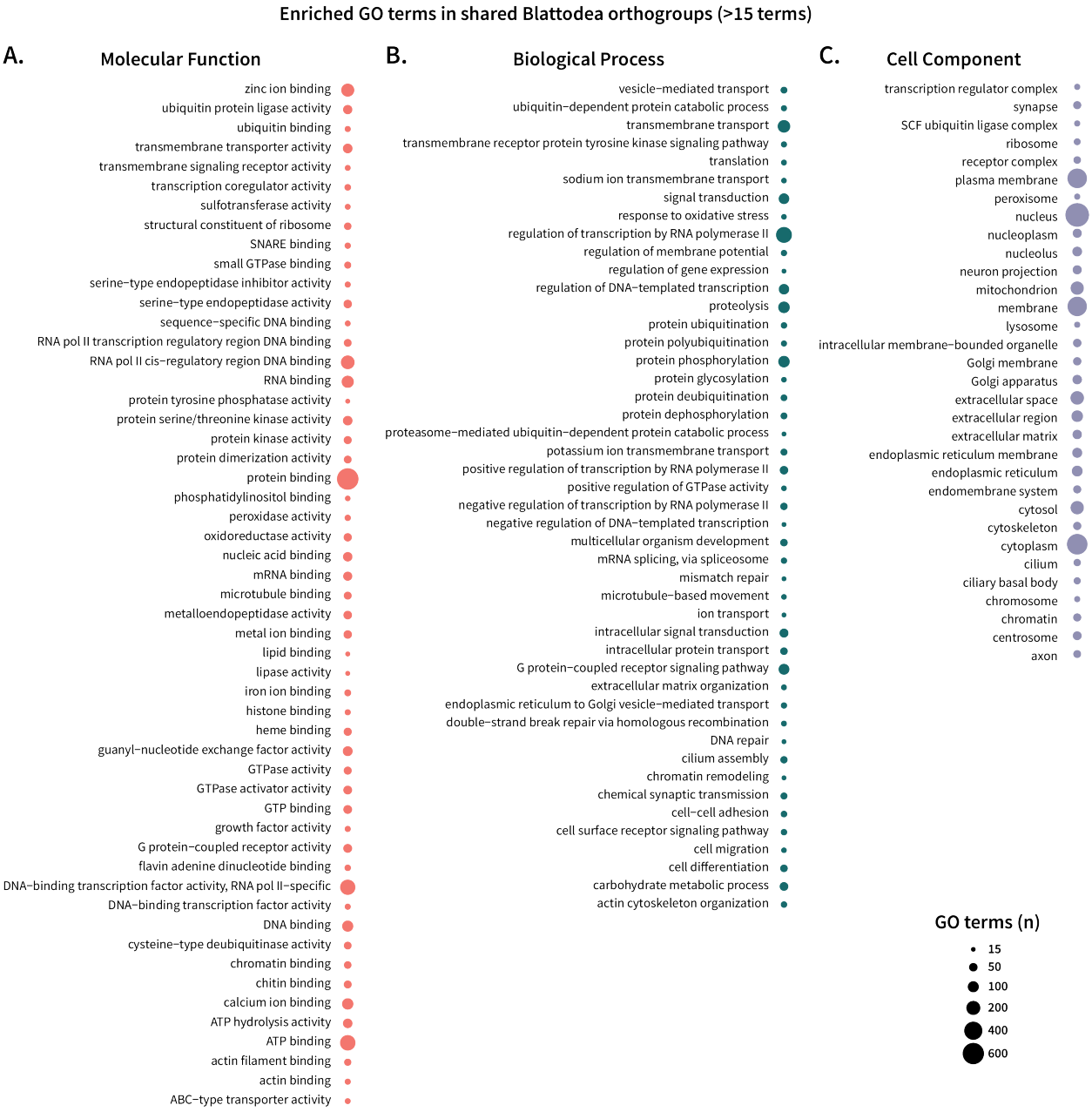


Figure S2.5: GO terms enriched in orthogroups shared by all Blattodea species. All genes included in the shared orthogroups in Figure 5D that had GO assignments were included for enrichment analysis by clusterProfiler. GO terms with abundances of at least 15 genes are reported for **(A)** molecular functions, **(B)** biological processes, and **(C)** cellular components.

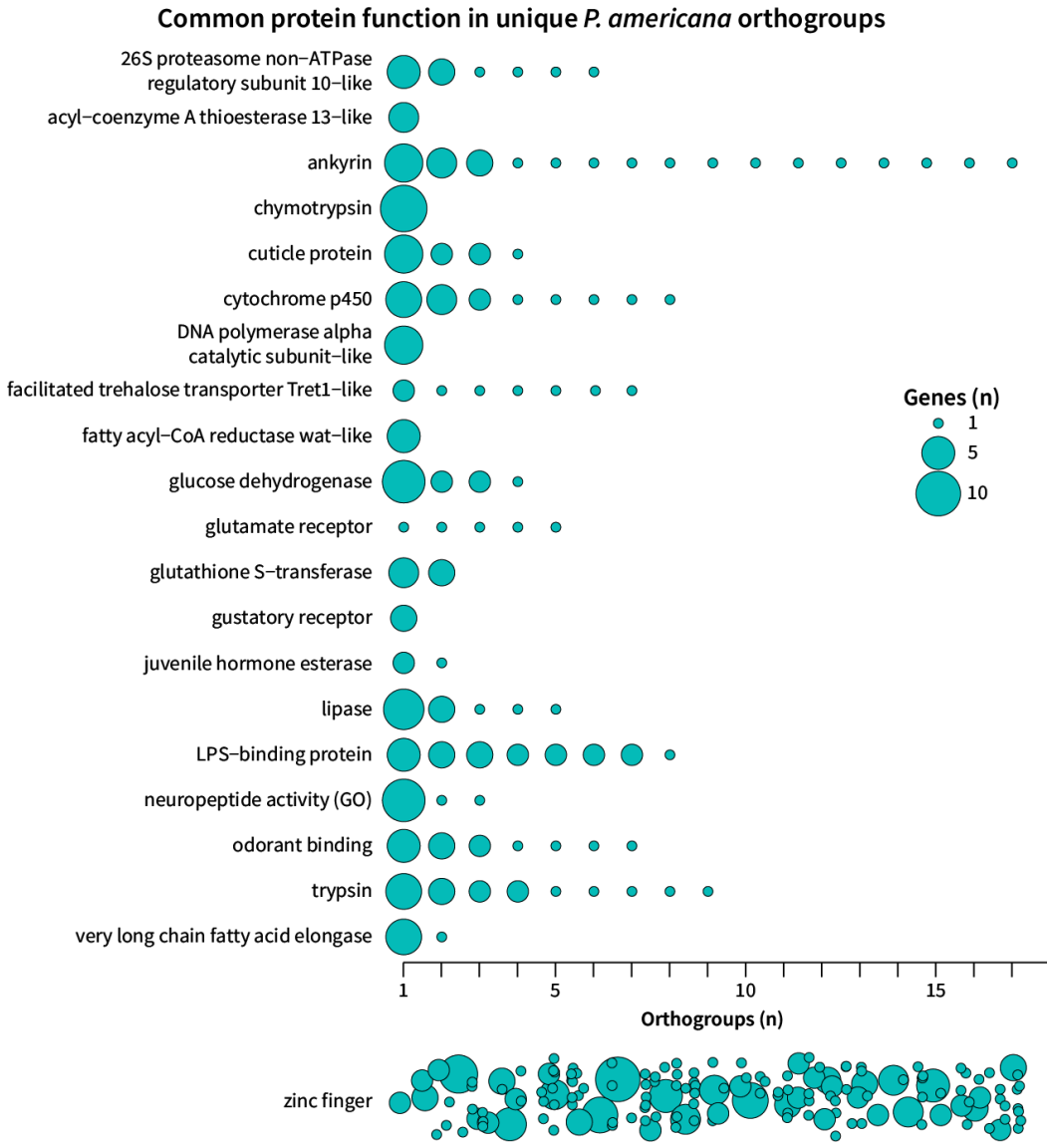


Figure S2.6: The names or GO terms of gene loci belonging to the 589 *P. americana*-unique orthogroups were collapsed to identify common functions in these genes and visualized with a bubble plot. Each bubble indicates a separate orthogroup per protein/function and is scaled by the number of genes sharing the associated name/GO term within the orthogroup. Proteins characterized by zinc finger motifs are plotted separately due to their large number of orthogroups (157 orthogroups, 248 genes).

APPENDIX B

Chapter 3 Supplemental Information[4]

[4] Dockman, R.L. and E.A. Ottesen. 2024. *Frontiers in Microbiomes*. Volume 3. Reprinted here with permission of the publisher.

Supplemental Table S3.1: Cohort sizes for experiments used in these studies.

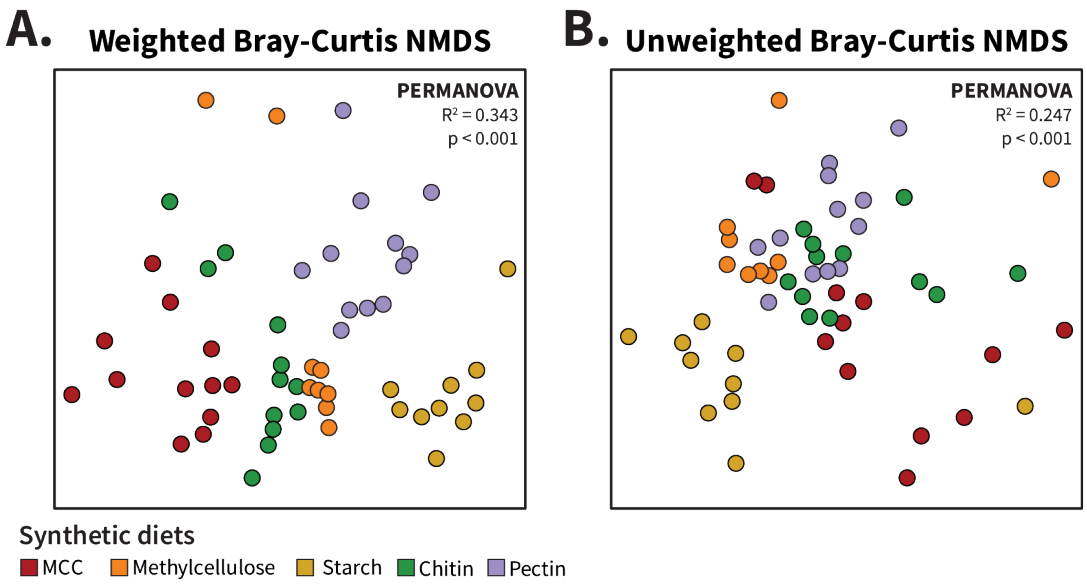
Cohort	Diet	Individuals dissected	Sequenced total (male/female)	Raw reads (total / average)	Quality reads (total / average)
1 *S: 2/2020* *E: 3/2020*	MCC-1	12	12 (6 / 6)	1,029,110 / 85,759	823,707 / 68,642
	Chitin	12	12 (9 / 3)	310,657 / 25,888	2726,57 / 22,721
	Pectin	12	12 (9 / 3)	231,967 / 19,331	203,208 / 16,934
	Xylan-1	12	12 (5 / 7)	568,859 / 47,405	417,805 / 34,817
	Methylcellulose	12	12 (8 / 4)	587,125 / 48,927	462,819 / 38,568
2 *S: 11/2020* *E: 11/2020*	Starch*	12	10 (9 / 1)	488,500 / 48,850	296,205 / 29,621
	MCC-2	12	10 (8 / 2)	613,777 / 61,378	340,395 / 34,040
	Xylan-2	12	10 (9 / 1)	1,078,499 / 107,850	736,808 / 73,681
3 *S: 1/2022* *E: 2/2022*	Xylan P-	12	8 (4 / 4)	556,764 / 69,596	444,612 / 55,577
	Xylan V-	12	8 (4 / 4)	692,806 / 86,601	556,169 / 69,521
	MCC P-	12	8 (4 / 4)	680,675 / 85,084	582,030 / 72,754
	MCC V-	12	8 (4 / 4)	575,169 / 71,896	494,232 / 61,779
	Xylose	12	8 (4 / 4)	655,253 / 81,907	556,133 / 69,517
	Cellobiose	12	8 (4 / 4)	527,794 / 65,974	409,609 / 51,201
	Glucose	12	8 (4 / 4)	450,316 / 56,290	374,747 / 46,843
4# *S: 8/2023* *E: 9/2023*	MCC-3	12	10	370,518 / 37,052	278,299 / 27,830
	Xylan-3	12	10	1,277,470 / 127,747	940,769 / 94,077
	Xylan-3-raw	12	10	1,306,817 / 130,682	914,751 / 91,475
	MCC-Tuna	12	10	1,753,300 / 175,330	1,386,390 / 138,639
	Xylan-Tuna	12	10	652,675 / 65,268	436,507 / 43,651
	Dog chow	12	10	257,678 / 25,768	190,726 / 19,073
Totals	*21 diets*	*252*	*206*	14,665,729 / 72,599	10,845,921 / 55,284

* The starch diet was analyzed with cohort 1
# sex was not recorded for cohort 4

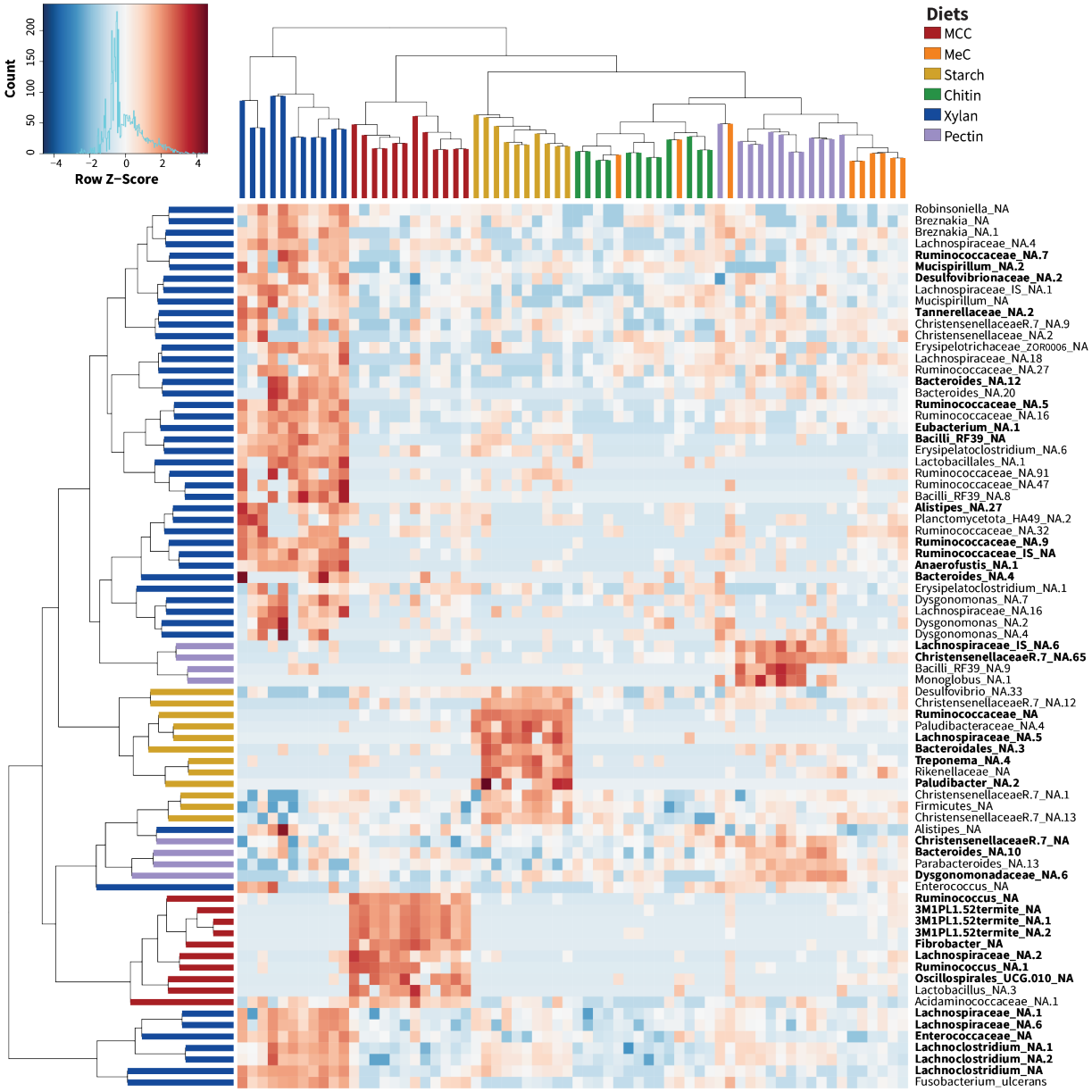
File uploaded separately:

AppendixBSupplementalFile.xlsx

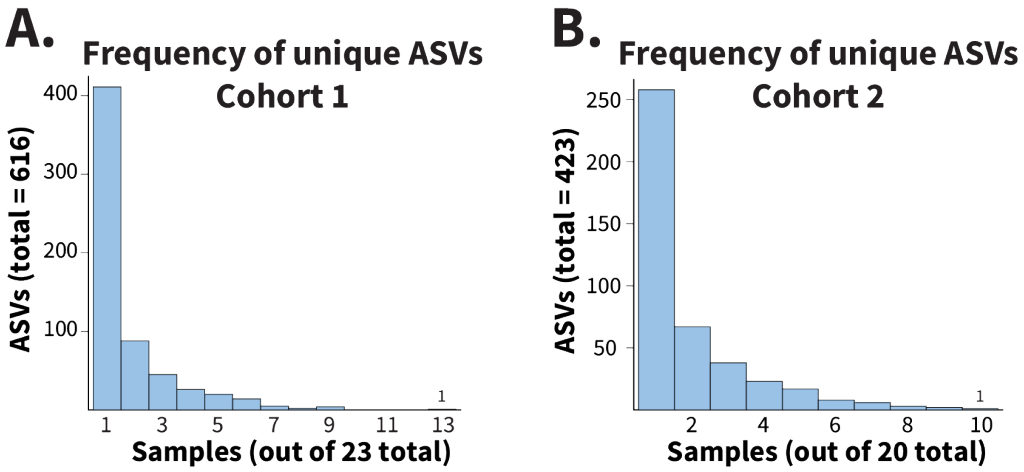
Supplemental File S3.1: UpSet sets generated for the synthetic and whole food diets.



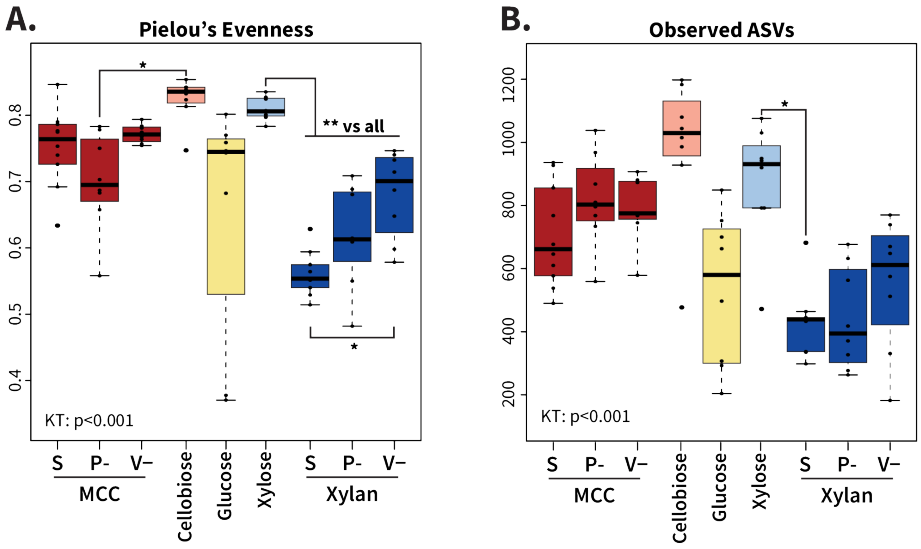
Supplemental Figure S3.1: NMDS ordination analysis of synthetic diet samples excluding xylan-fed cockroaches. As in Figure 1, samples were rarefied a constant depth of 7924 sequences for alpha and beta diversity calculations. Non-metric multidimensional scaling (NMDS) was used to plot (A) weighted and (B) unweighted Bray-Curtis dissimilarity of gut communities from synthetic diets, excluding those fed the xylan diet. Multivariate analysis was performed using PERMANOVA.



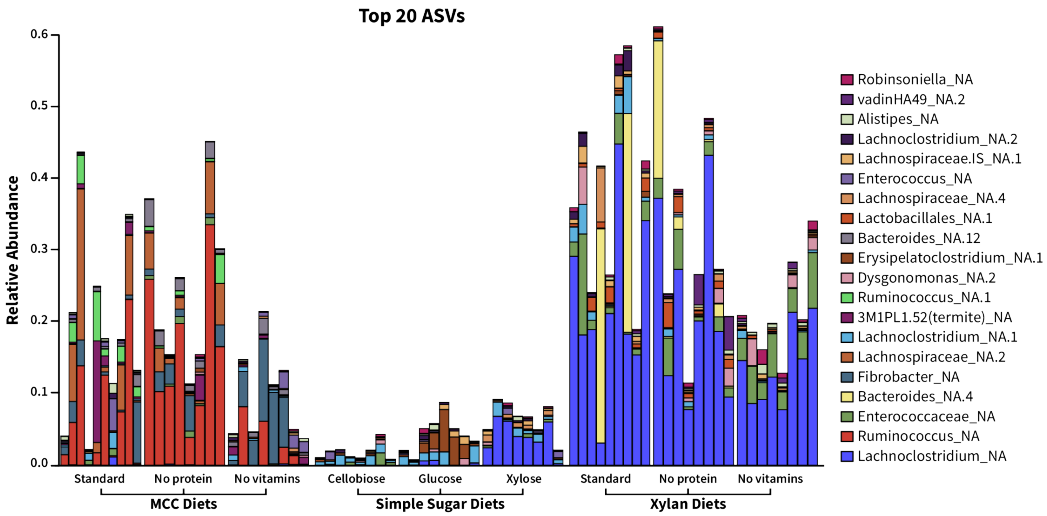
Supplemental Figure S3.2: Diet-characteristic ASVs determined by DESeq2. Pairwise DESeq2 analysis was performed on filtered (present in 5 or more samples) raw counts for all synthetic diets (n=66), and ASVs significant in one diet vs all others at adjusted $p < 0.05$ with a baseMean higher than 10 were selected (76 total). Variance stabilized transformed counts obtained from DESeq2 were scaled by row to generate z-scores for plotting. Dendrograms are colored by diet per sample (column) and diet the ASVs are associated with (row). Bolded names indicate the ASV belongs to "Set 1" as presented in Figure 4.



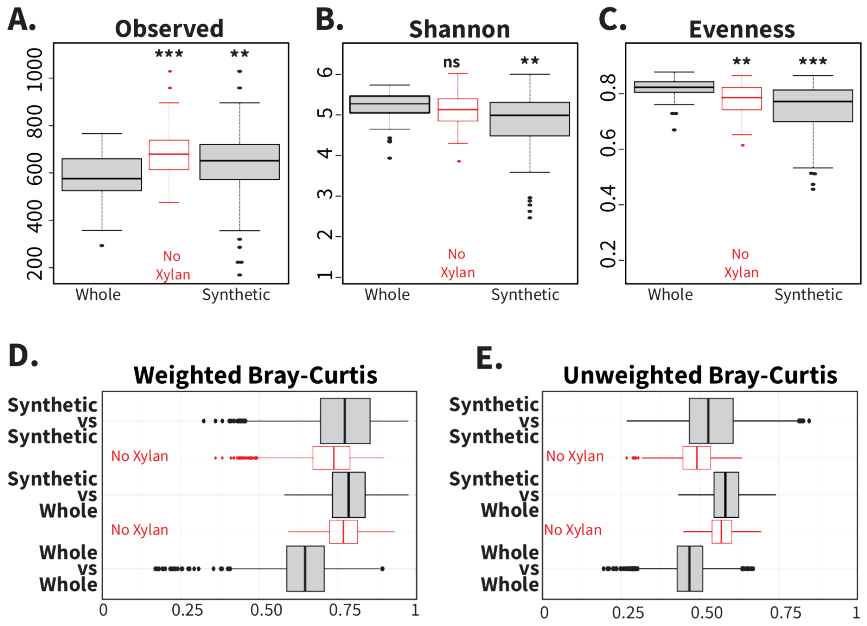
Supplemental Figure S3.3: Most unique ASVs in the Xylan/MCC replicate experiments are singletons. Unique ASVs from the rarefied (A) Cohort 1 and (B) Cohort 2 data were assessed for frequency of occurrence using the histogram function in R. Samples were rarefied to 9685 ASVs



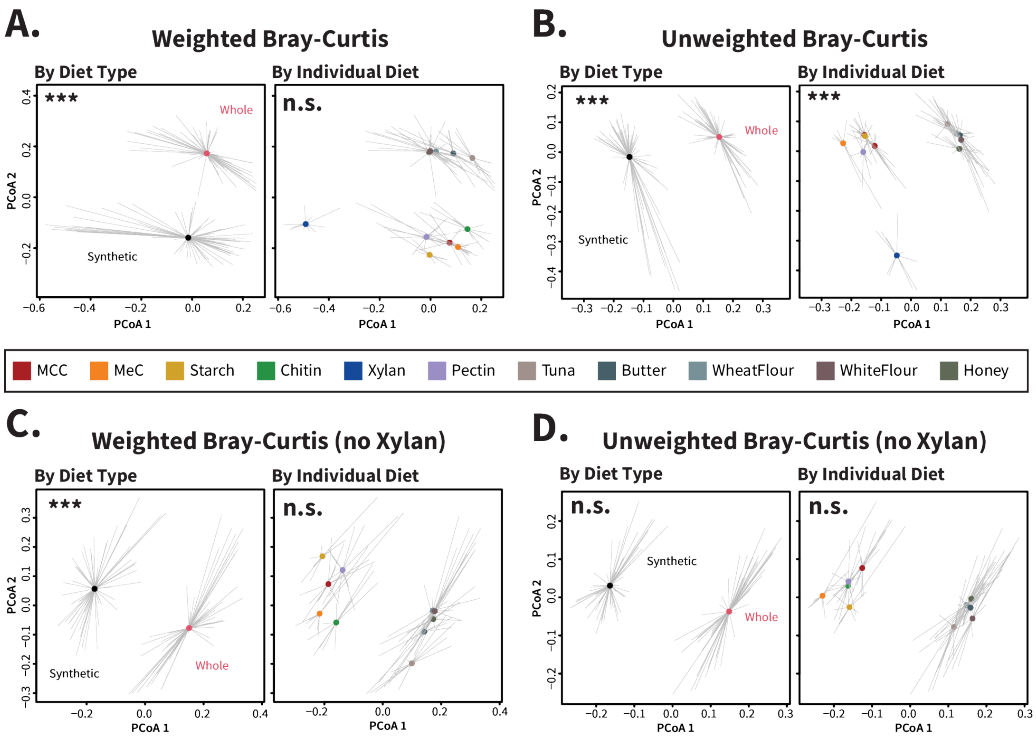
Supplemental Figure S3.4: Additional alpha diversity measures of standard, deficient, and simple-sugar diets. Samples were rarefied a constant depth of 12274 sequences for alpha diversity calculations (A) Pielou's evenness and (B) number of observed ASVs. MCC: microcrystalline cellulose; S: standard diet; P-: protein-deficient; V-: vitamin-deficient. *p < 0.05; **p<0.01



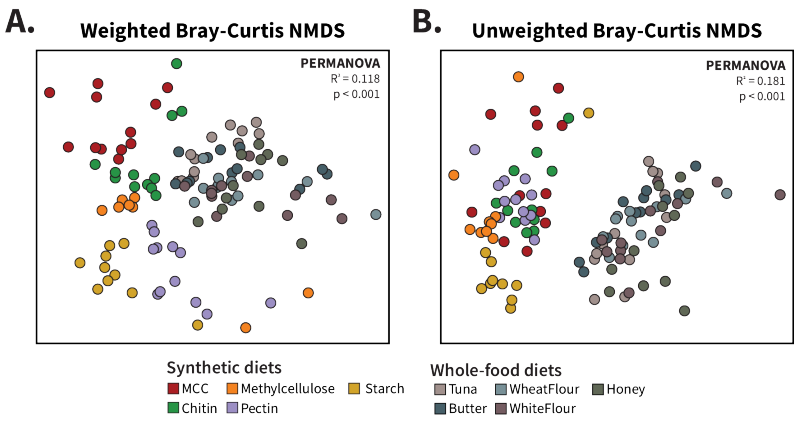
Supplemental Figure S3.5: Relative abundance of top ASVs in xylan/MCC diet variations. ASV tables were converted to proportion tables and sorted from most to least abundant ASV across standard, deficient, and simple-sugar variations of the xylan and MCC diets. The top 20 ASVs were included for visualization.



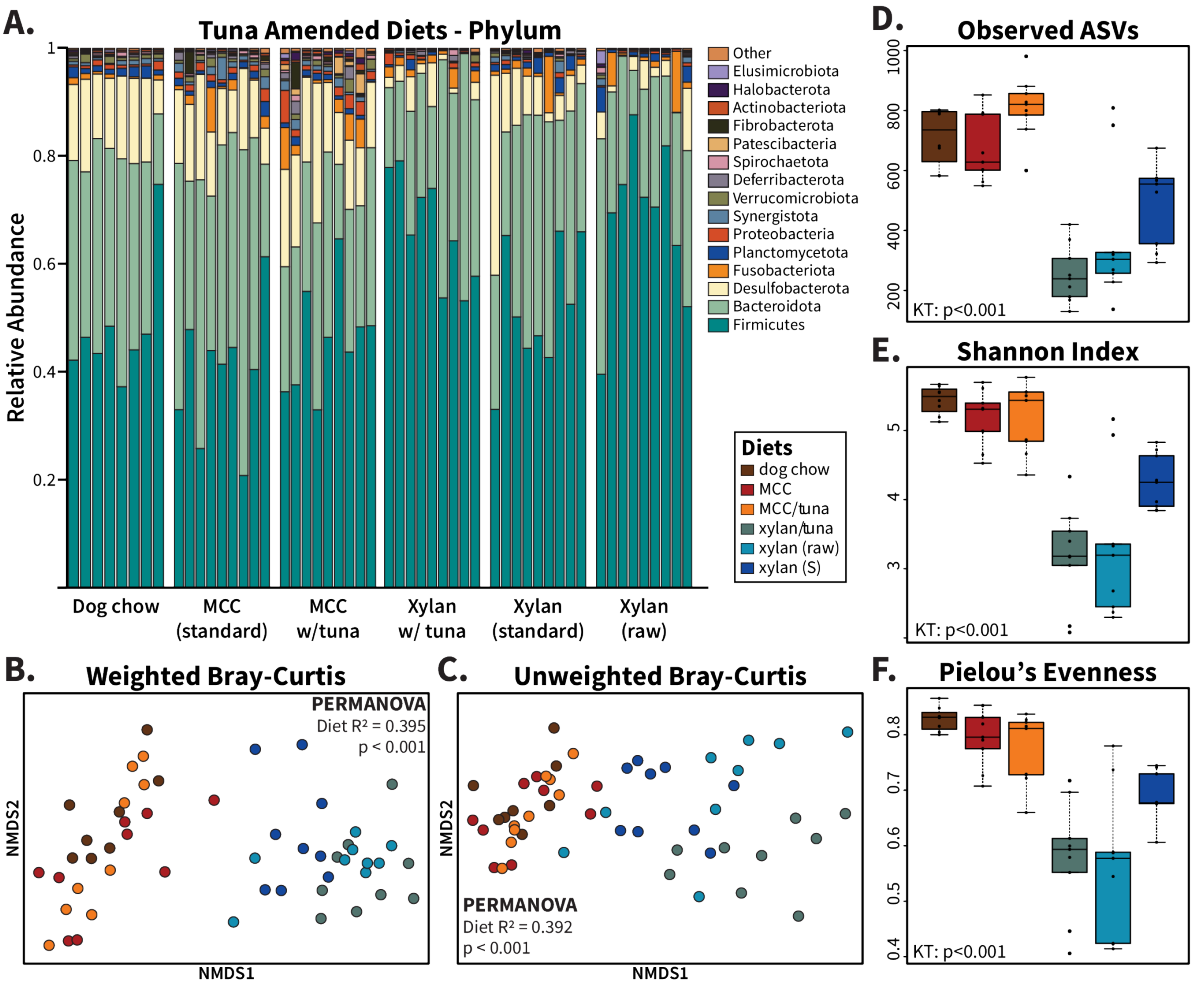
Supplemental Figure S3.6: Overview of alpha and beta diversity differences between whole food and synthetic diets. Raw sequence data from Tinker and Ottesen (2016) were reprocessed using the methods in this experiment to generate comparable ASVs. All samples were rarefied to 7924 ASVs for alpha and beta diversity analysis. Boxplots show (A) observed ASVs, (B) Shannon index, and (C) Pielou's evenness for each diet type, with red boxes representing the synthetic diets minus xylan-fed samples. Wilcoxon rank-sum test was used for pairwise statistical analysis between diet types. Beta diversity boxplots of (D) weighted and (E) unweighted Bray-Curtis dissimilarity, with red boxes representing synthetic diets excluding xylan. ** =p <0.01; *** =p<0.001, ns =no significance



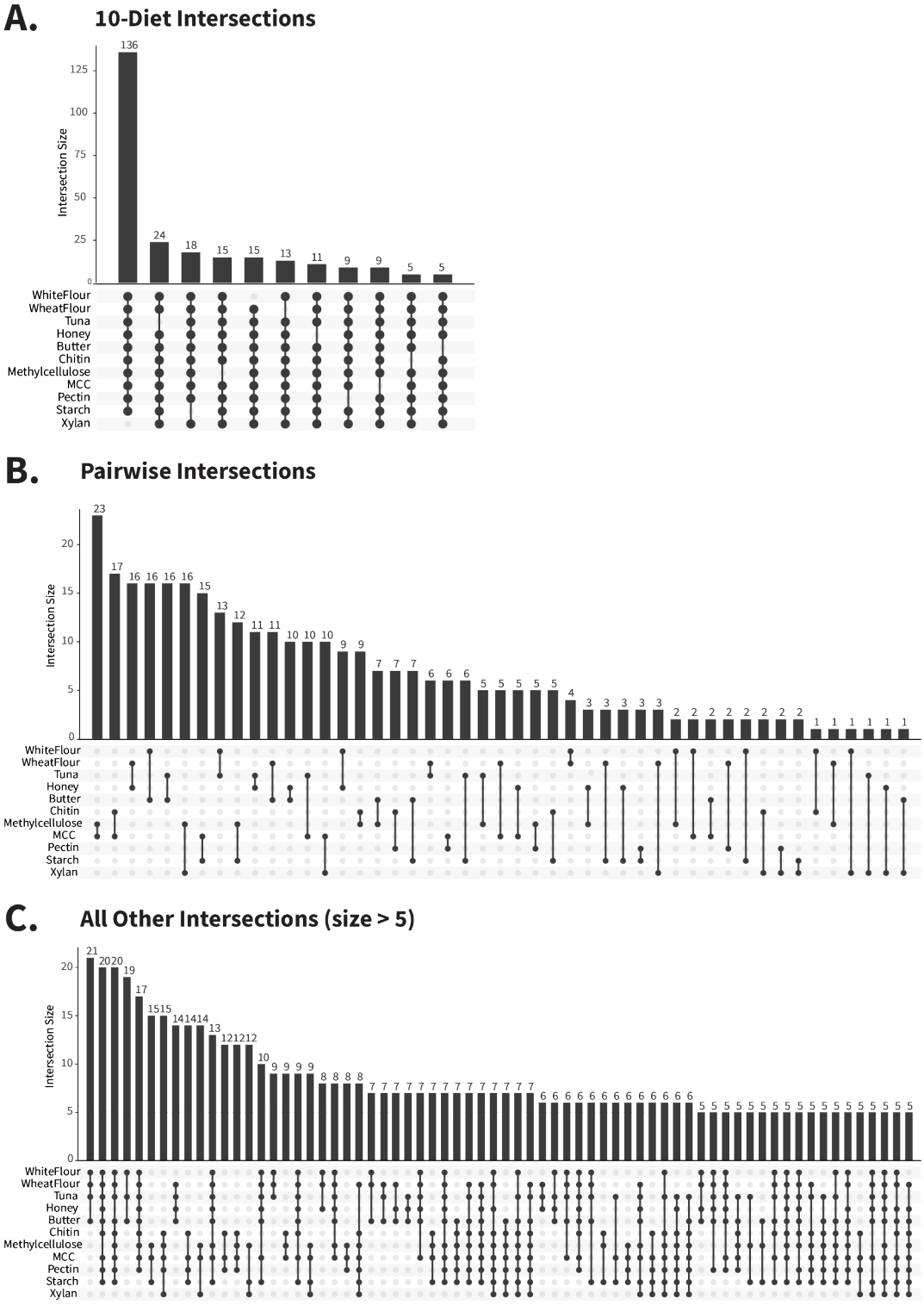
Supplemental Figure S3.7: Beta dispersion analysis of whole food and synthetic diets, with and without xylan-fed samples. The dispersion of (A,C) weighted and (B,D) unweighted Bray-Curtis dissimilarities was assessed for both diet group and individual diets, both (A,B) with and (C,D) without xylan-fed samples included in analysis. All samples were rarefied to 7924 ASVs for beta dispersion analysis, and ANOVA statistical testing was used to evaluate if dispersion differed between groups/diets. *** =p<0.001, ns =no significance



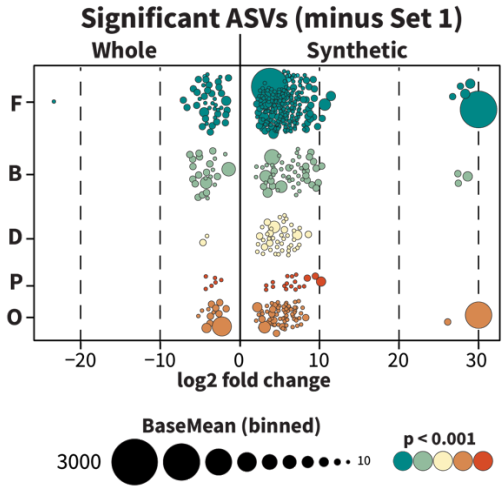
Supplemental Figure S3.8: Differences in beta diversity between whole food and synthetic diets without xylan-fed samples. NMDS ordinations of (A) weighted and (B) unweighted Bray-Curtis distances between cockroach hindgut samples from this study ("Synthetic") and data from Tinker and Ottesen (2016) for cockroaches fed tuna, butter, honey, wheat flour, and white flour ("Whole Food Diets"), excluding xylan-fed samples. Samples were rarefied to 7924 ASVs for beta diversity analysis. R2 and p-values for diet type comparisons were calculated using PERMANOVA.



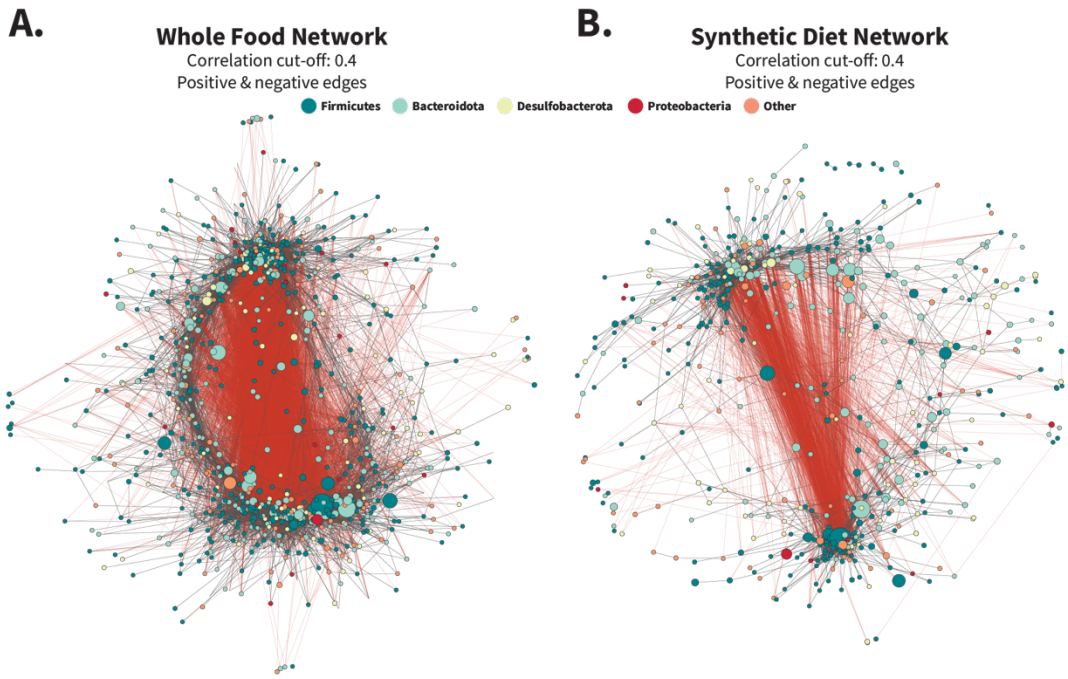
Supplemental Figure S3.9: Complex proteins did not reduce the influence of fiber source on gut microbiome composition. Samples were rarefied a constant depth of 12274 sequences for alpha and beta diversity calculations, while relative abundance was used to visualize phylum-level composition. (A) depicts phyla relative abundance. NMDS ordinations were generated for (B) weighted and (C) unweighted Bray-Curtis dissimilarity and assessed for significance with PERMANOVA. Alpha diversity measures (D) observed ASVs, (E) Shannon index, and (F) Pielou's evenness are plotted with boxplots and statistics were calculated using Kruskal test. MCC: microcrystalline cellulose; S: standard diet; KT: Kruskal test



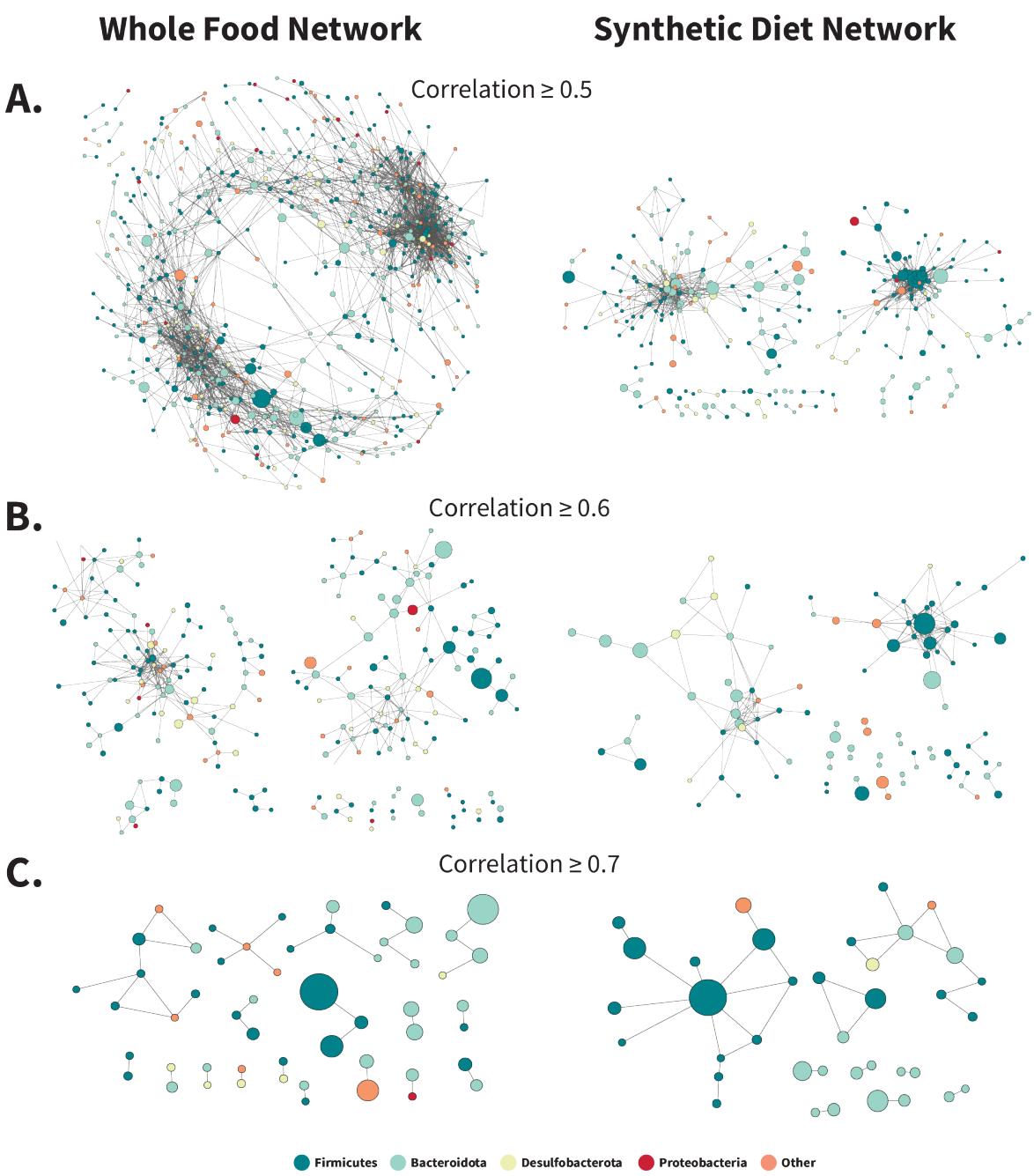
Supplemental Figure S3.10: Additional UpSet plots. The rarefied samples (7924 depth) from whole food and synthetic diets were used to generate UpSet plots. (A) depicts sets where ASVs are found in all but one of the diets. (B) depicts sets found in only two diets. (C) contains remaining sets not yet displayed with at least 5 ASVs. Additional sets were not included.



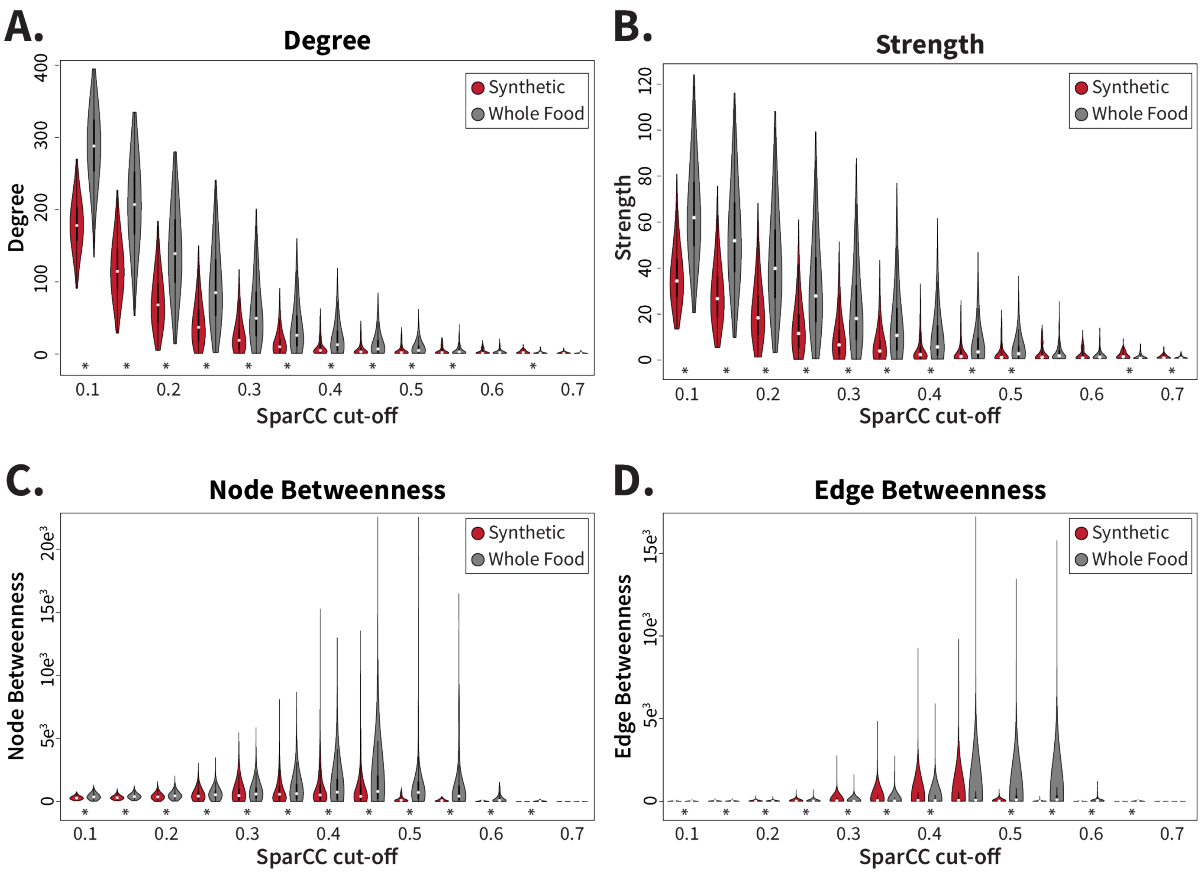
Supplemental Figure S3.11: Significantly enriched ASVs between whole food and synthetic diets, excluding Set 1 ASVs. Raw sequence count tables of whole food and synthetic diets were filtered to include ASVs present in at least 5 samples (out of 125 total) and analyzed using DESeq2 with diet type as the design factor. ASVs identified as significant ($p < 0.001$), excluding Set 1, are visualized.



Supplemental Figure S3.12: Whole food and synthetic networks including negative edges. Networks were calculated by SparCC from filtered count tables for (A) whole food and (B) synthetic diets separately to create two distinct networks. Count tables were filtered to include only ASVs present in at least 25% of the samples per diet set, resulting in 976 ASVs for whole food diets and 700 for synthetic. Networks were further pruned to remove edges with absolute values smaller than 0.4 before exporting to Cytoscape. Negative edges were included during initial layout generation with the edge-weighted spring embedded layout method.



Supplemental Figure S3.13: Whole food and synthetic networks at higher correlation cut-off levels. Networks generated for Figures 6 and S11 were filtered in Cytoscape by edge weight to visualize networks at (A) 0.5, (B) 0.6, and (C) 0.7 SparCC correlation values, with isolated nodes removed if they lost all adjacent edges. Data on edges, nodes, and connected components at these cut-off levels are presented in Figure 6C-E.



Supplemental Figure S3.14: Whole food diets encourage more interconnected network formation than synthetic diets. SparCC networks were generated for whole food and synthetic diet types using raw ASV counts excluding uncommon taxa (< 25% of samples in diet type). The networks were analyzed at increasing levels of positive associations to compare (A) degrees per node, (B) strength of co-correlations per node, (C) node betweenness, and (D) edge betweenness. Statistics were calculated using the Wilcoxon rank-sum test. * = $p<0.05$ Red: Synthetic; Grey: Whole Food.

APPENDIX C

Chapter 4 Supplemental Information

Supplemental Tables

Supplemental Table 4.1: Primers for generating antisense rRNA probes. DNA was extracted from two pooled cockroaches that had been fed synthetic diets containing either xylan or microcrystalline cellulose. Microbial and host-derived rRNA gene regions were amplified with the T7 promoter region appended to the reverse barcode using the primers and annealing temperatures listed in this table.

Region	Primer	Sequence	Anneal (C)
Archaea 16S	21F	TCCGGTTGATCCYGCCGG	70
	1492R	GCCAGTGAATTG-**T7**-GGGGYYACCTTGTTACGACTT	
Archaea 23S	189F	ASAGGGTGAHARYCCCGTA	70
	2490R	GCCAGTGAATTG-**T7**-GGCTGTCTCRCGACGGTCTRAACCCA	
Bacteria 16S	27F	AGAGTTTGATCCTGGCTCAG	39
	1492R	GCCAGTGAATTG-**T7**-GGACGGCTACCTTGTTACGACTT	
Bacteria 23S	189F	GAASTGAAACATCTHAGTA	39
	2490R	GCCAGTGAATTG-**T7**-GGCGACATCGAGGTGCCAAAC	
Eukaryote 18S	1F	ACCTGGTTGATCCTGCCAG	55
	1520R	AATTA-**T7**-ATTCYGCAGGTTCACCTAC	
Eukaryote 28S	26F	ACCCGCYGAAYTTAAGCATA	55
	3126R	AATTA-**T7**-ATTCTGRYTTAGAGGCGTTCAG	
Cockroach ITS	18S_1907F	CCTGCGGAAGGATCATTAAC	60
	28S_-2R	GCCAGTGAATTG-**T7**-GGCTTAAATTCAGCGGGTAGTCTC	

T7: 5′ TAATACGACTCACTATAG 3′

Supplemental Table 4.2: RNA read tracking through filtering steps.

	R1		R2		Both R1 and R2			
Sample	**initial reads**	**mRNA reads**	**initial reads**	**mRNA reads**	**paired reads**	**filtered reads**	**merged pairs**	**aligned with DIAMOND**
XT1_HgL1	83167708	25239059	83643619	25592007	49090264	45249236	22624618	18085988
XT1_HgL2	47251549	13717591	47902798	13999554	26871502	22956298	11478149	9256938
XT1_HgL3	79007875	25474355	79996235	25923374	49611730	38464086	19232043	15116932
XT1_HgL4	68303715	22841708	69863332	23089066	44978238	41017224	20508612	17093359
XT1_HgL5	62065220	25785807	62951191	25990256	50609950	48830212	24415106	20288972
XT1_HgL7	91950695	37851719	94496533	38063774	74326902	71794094	35897047	26883520
XT2_HgL1	71418991	25076230	72106314	25243676	48799680	47469354	23734677	19161622
XT2_HgL3	52469393	19524666	52759634	19687352	38619820	37445000	18722500	14284306
XT2_HgL4	83984458	39400506	84688749	39821140	77030924	75385452	37692726	31943483
XT2_HgL5	79415147	36860338	80928173	37043447	72835484	70430846	35215423	29295230
XT2_HgL6	67709119	29534363	69153831	29887045	57954520	55633270	27816635	23655953
XT2_HgL7	73025327	29103049	74984102	29950680	55444638	53052316	26526158	18876312
XT3_HgL1	46175011	16233127	49777036	16559143	31718522	30124226	15062113	14115915
XT3_HgL3	60384706	28281236	63055728	28448960	55805312	53821444	26910722	22514244
XT3_HgL4	89610533	30366146	90682727	30824303	59864898	54306166	27153083	23595116
XT3_HgL5	107781761	28649132	109234848	28848385	55647064	54225038	27112519	21560649
XT3_HgL9	73214318	11597279	73430946	12162830	22157234	16997708	8498854	7144408
XT3_HgL10	69422560	30198829	69952461	30316467	59390396	51683960	25841980	21365461
XT4_HgL6	54613882	13862109	54815672	14181923	25969036	24460134	12230067	8382119
XT4_HgL7	83302304	24855380	84860431	25080399	48623788	45523768	22761884	16206823
XT4_HgL8	75189344	16448440	75715652	16750454	31794802	28956358	14478179	10279389
XT4_HgL9	50606869	13376743	52100115	13695932	25613404	23274734	11637367	8419598
XT4_HgL10	102161190	32863172	102894236	33007141	64572796	62291220	31145610	23975956
XT4_HgL11	95711249	25753521	96394620	25776625	49656214	44609196	22304598	13918250
XT5_HgL3	65446044	18930397	66687129	19144507	37124530	34316272	17158136	13187727
XT5_HgL4	38712542	12923662	40478389	13043323	24976818	23244390	11622195	7978854
XT5_HgL5	81916125	7545610	82014303	7931598	13903986	13154186	6577093	4875506
XT5_HgL6	53985913	16666582	54300297	16780659	32513704	30386132	15193066	11545523
XT5_HgL7	115045888	8710681	115193918	9554599	15689626	12026460	6013230	4740617
XT5_HgL10	136881495	12508722	137154114	13039750	23307506	18195732	9097866	6273811

Supplemental Table 4.3 RefSeq accessions used to generate *Lachnospiraceae* pangenome groups.

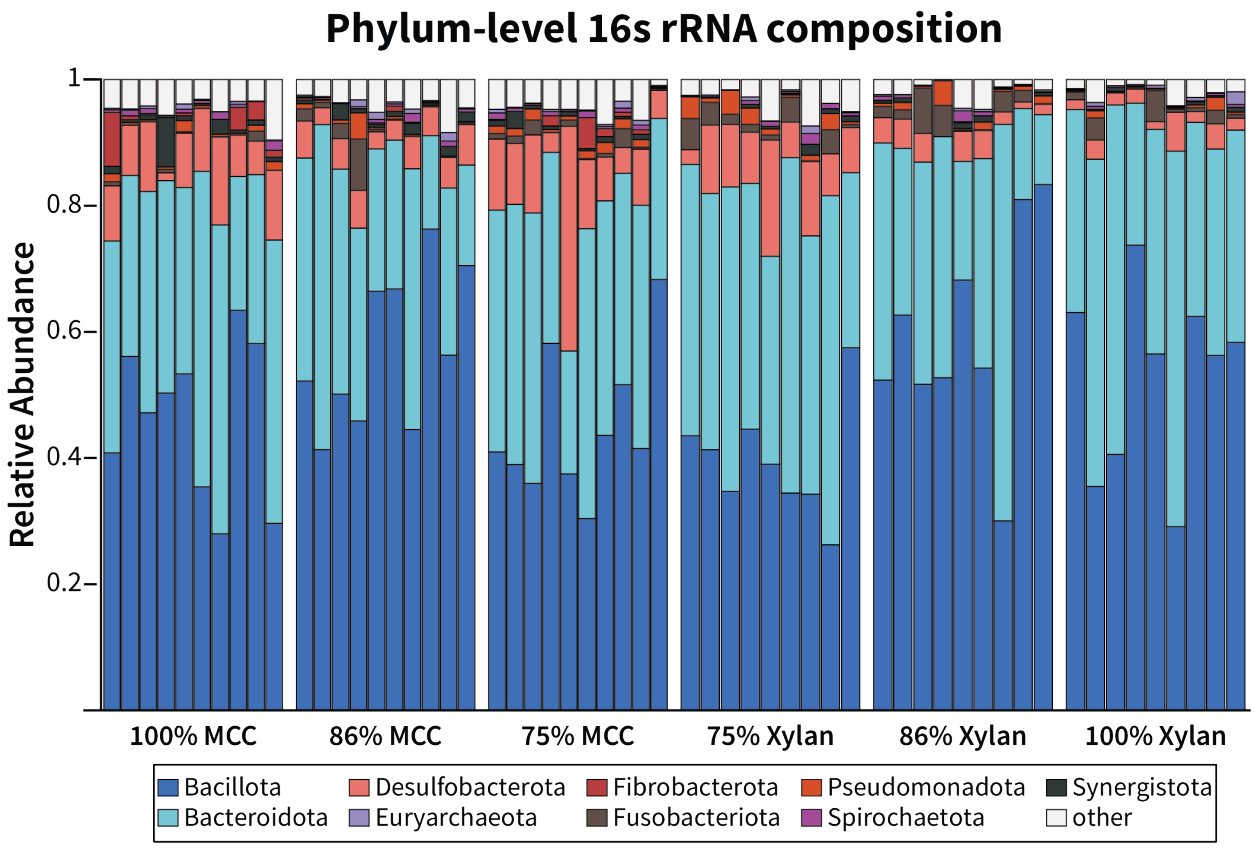
Genera	Accessions
Lachnospiraceae A	
Aequitasia	GCF_024160205|GCF_024721305|GCF_024160185|GCF_024721265
Brotaphodocola	GCF_020686985|GCF_003478505|GCF_003477935|GCF_003480315|GCF_003480105|GCF_003481985|GCF_003481825
Enterocloster	GCF_000158075|GCF_025149125|GCF_000233455|GCF_001078435|GCF_000234155|GCF_000371605|GCF_000371585|GCF_000371565|GCF_000371545|GCF_000371505|GCF_000371485|GCF_000424325|GCF_000436455|GCF_949390495|GCF_949400645|GCF_949474175|GCF_949523935|GCF_949531885|GCF_949542115|GCF_000371405|GCF_001078445|GCF_009696375|GCF_022771995|GCF_015549035|GCF_005845215|GCF_015556325|GCF_003473545|GCF_000954015|GCF_013282095|GCF_003434055|GCF_003467385|GCF_011317135|GCF_013304305|GCF_020554935|GCF_030373645|GCF_003466005|GCF_902385905|GCF_002959675|GCF_003437595|GCF_003458165|GCF_003458625|GCF_003464745|GCF_013112035|GCF_015547745|GCF_015551555|GCF_015556085|GCF_015557875|GCF_015558345|GCF_016889665|GCF_020555225|GCF_020555685|GCF_020559295|GCF_020563055|GCF_020736955|GCF_021771435|GCF_022137355|GCF_022138045|GCF_022138265|GCF_023008325|GCF_024460255|GCF_024462735|GCF_024622735|GCF_027668515|GCF_027671045|GCF_030839815|GCF_030844385|GCF_902375545|GCF_905197355|GCF_958411475|GCF_958422245|GCF_958451465|GCF_959023705|GCF_959605845|GCF_000371725|GCF_000371705|GCF_000371685|GCF_000371665|GCF_000371645|GCF_000154365|GCF_001078425|GCF_949494455|GCF_003433945|GCF_015548445|GCF_018785395|GCF_020554185|GCF_020736995|GCF_027662505|GCF_027666385|GCF_900115855|GCF_902364255|GCF_001405335|GCF_005844705|GCF_006538465|GCF_012273195|GCF_013299965|GCF_013304125|GCF_013304155|GCF_013304185|GCF_013304205|GCF_013304225|GCF_013304245|GCF_013304255|GCF_013304275|GCF_015152355|GCF_015548065|GCF_015669475|GCF_018381395|GCF_019012735|GCF_020297485|GCF_021532095|GCF_027663445|GCF_028210245|GCF_028210555|GCF_028210575|GCF_028210655|GCF_900100685|GCF_900108895|GCF_900113155|GCF_900447015|GCF_902374585|GCF_910586265|GCF_948511265|GCF_948522125|GCF_948524135|GCF_948582345|GCF_948587995|GCF_948603615|GCF_948681025|GCF_948901585|GCF_948908805|GCF_948919595|GCF_948932345|GCF_948952455|GCF_949013715|GCF_949081035|GCF_949121295|GCF_958414435|GCF_958435425|GCF_958436665|GCF_958453805|GCF_959023645|GCF_003024655|GCF_020564305|GCF_900102595|GCF_902364025|GCF_905194475|GCF_934882645|GCF_958421775|GCF_018368145|GCF_020554465|GCF_022440645|GCF_020709355|GCF_020709765|GCF_020554865|GCF_900540675|GCF_934402005|
Hungatella	GCF_003201875|GCF_027661945|GCF_027662645|GCF_027662785|GCF_027696135|GCF_932750805|GCF_001405675|GCF_001405995|GCF_003435045|GCF_003437645|GCF_003437905|GCF_003439535|GCF_003466285|GCF_003468235|GCF_003475805|GCF_009721605|GCF_015553645|GCF_015555845|GCF_018379215|GCF_018785385|GCF_022834955|GCF_022834975|GCF_024464175|GCF_025149285|GCF_027671645|GCF_902362405|GCF_902363795|GCF_905204275|GCF_937936815|GCF_959027765|GCF_959598815|GCF_000371445|GCF_000433395|GCF_000160095|GCF_014288005|GCF_022782185|GCF_022784085|GCF_014288035|GCF_024460495|
Lachnoanaerobaculum	GCF_003862475|GCF_030008055|GCF_003862485|GCF_905371455|GCF_001552975|GCF_000185385|GCF_000257705|GCF_018372015|GCF_000287675|GCF_017565785|GCF_000512995|GCF_000296385|GCF_003254255|GCF_003589745|GCF_902387945|GCF_937890375|GCF_938015485
Lachnoclostridium	GCF_900078195|GCF_000733755|GCF_000018685|GCF_000703105|GCF_000702985|GCF_905197605
Lacrimispora	GCF_000687555|GCF_002797975|GCF_000526995|GCF_000421505|GCF_000144625|GCF_900105615|GCF_003432035|GCF_000526575|GCF_003833015|GCF_007115105|GCF_900155545|GCF_009696365|GCF_017084465|GCF_003609635|GCF_900205965|GCF_000732605|GCF_900185635|GCF_016906045|GCF_900105215|GCF_900461315|GCF_026723765
Lachnospiraceae B	
Agathobacter	GCF_000020605|GCF_002735305|GCF_001406815
Kineothrix	GCF_000732725|GCF_004345255|GCF_030863805
Acetatifactor	GCF_003478095|GCF_003480225|GCF_009695995|GCF_014337175|GCF_024623325|GCF_025567015|GCF_900248245|GCF_910584235|GCF_910585425|GCF_910585615|GCF_910588225|GCF_943193215|GCF_947643695|GCF_947654235|GCF_948475165|GCF_948475395|GCF_948482205|GCF_948492135|GCF_948495795|GCF_950096775|GCF_950097205
Anaerobium	GCF_900096945
Anaerocolumna	GCF_009931695|GCF_014218335|GCF_018917405|GCF_030913705|GCF_900142215|GCF_900205915|GCF_947653275|GCF_000702945|GCF_014202875|GCF_014218355|GCF_029689925|GCF_900115365|GCF_900143645|GCF_902479815
Anaeromicropila	GCF_016591975
Anaerosacchariphilus	GCF_003363435
Butyrivibrio	GCF_025148445|GCF_023206215|GCF_900129945|GCF_900143205|GCF_900101605|GCF_000424465|GCF_900115735|GCF_000145035|GCF_000622085|GCF_900103635|GCF_015056685|GCF_017433805|GCF_017635265|GCF_017940685|GCF_024699075|GCF_000424145|GCF_900104155|GCF_000420825|GCF_000703165|GCF_000424285|GCF_000420845|GCF_003625485|GCF_000423945|GCF_000421405|GCF_000526935|GCF_900102515|GCF_900116875|GCF_000621565|GCF_000424385|GCF_900116865|GCF_000424585|GCF_000621605|GCF_000702265|GCF_000424545|GCF_900108105|GCF_000420945|GCF_000423925|GCF_000424265|GCF_003625475|GCF_000420865|GCF_900112195|GCF_010906925|GCF_015057245|GCF_900218035|GCF_000424305|GCF_000424005|GCF_000703005|GCF_015057185|GCF_017471675|GCF_017557435|GCF_017934625|GCF_900100545|GCF_900107545|GCF_000621865|GCF_900113655|GCF_900104335|GCF_008935055|GCF_900141825|GCF_947169445
Cockroach SCG	ButyrivibrioOttesenSCG.928.D06
Eisenbergiella	GCF_001722575|GCF_001717125|GCF_001717135|GCF_001722555|GCF_001881565|GCF_001722585|GCF_001722635|GCF_001722655|GCF_009696275|GCF_003435265|GCF_003435485|GCF_003478085|GCF_027682485|GCF

	_021769355|GCF_022781285|GCF_027680505|GCF_905205275|GCF_900243045|GCF_902385915|GCF_902471245|GCF_945899955|GCF_937926915|GCF_945830375|GCF_945863685|GCF_947507165
Gallintestinimicrobium	GCF_025567525.1
Roseburia	GCF_025567465|GCF_000225345|GCF_900537995|GCF_000174195|GCF_014287435|GCF_009695765|GCF_014287515|GCF_003612565|GCF_001940165|GCF_014297335|GCF_014287635.1
Suilimivivens	GCF_003612395|GCF_009917485|GCF_025567405.1
Lachnospiraceae C	
Lachnospira	GCF_000146185|GCF_000424105|GCF_000702205|GCF_003458705|GCF_009680455|GCF_014287955|GCF_020564355|GCF_900103815
Mediterraneibacter	GCF_000153925|GCF_001487105|GCF_014287475|GCF_016902345|GCF_020687545|GCF_025152405|GCF_900120155|GCF_934309345|GCF_934309415|GCF_934330165|GCF_944377805|GCF_003574295
Anaerosporobacter	GCF_012070565|GCF_900142955
Anaerostipes	GCF_002270485|GCF_005280655|GCF_014467075|GCF_016586355|GCF_018381315|GCF_018918155|GCF_018982945|GCF_025567365|GCF_030296915
Blautia	GCF_001689125|GCF_002222595|GCF_002270465|GCF_003287895|GCF_013304445|GCF_014287615|GCF_900120295|GCF_900461125|GCF_947654165
Coprococcus	GCF_003482105|GCF_025149915|GCF_025567285|GCF_025567345|GCF_019734885|GCF_902381825
Cuneatibacter	GCF_004216775
Dorea	GCF_025150245
Extibacter	GCF_001185345|GCF_004345005|GCF_008281175
Hominisplanchenecus	GCF_020687205|GCF_943193015
Cockroach SCG	Lachnospiraceae bacterium OttesenSCG-928-E19|Lachnospiraceae bacterium OttesenSCG-928-J05
Lachnotalea	GCF_003201285|GCF_008830185|GCF_900184995
Metalachnospira	GCF_018918145
Pseudolachnospira	GCF_022867805
Robinsoniella	GCF_000797495
Sellimonas	GCF_001280875|GCF_019754295|GCF_027924685

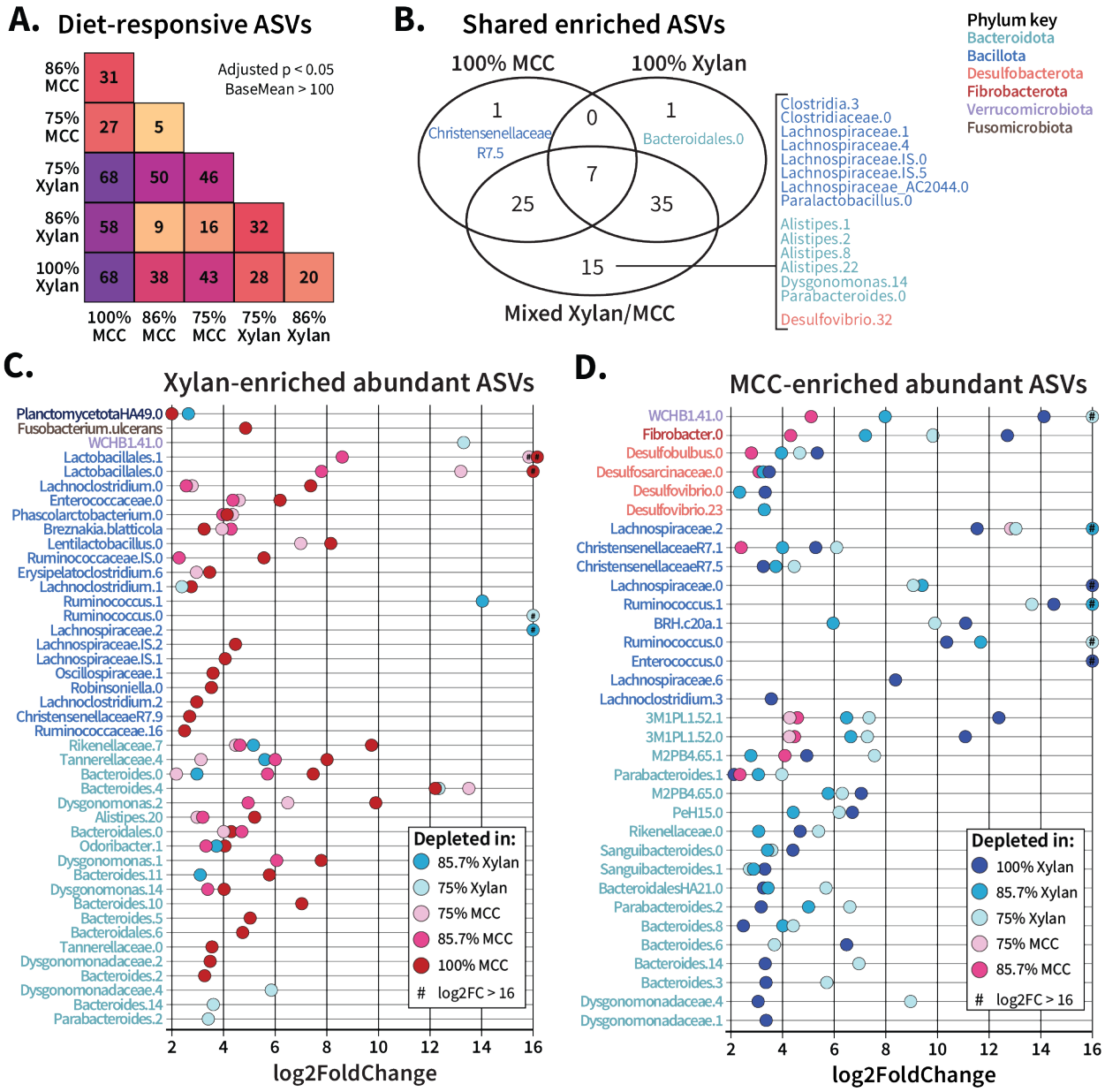
Supplemental Table 4.4: Glycoside hydrolases associated with cellulose or xylan degradation. Adapted from dbCAN2 substrate chart. (ref)

Substrate	Family	Description	Additional substrates
Xylan	GH1	β-xylosidase	β-fucosides, β-galactan, β-glucan, β-glucuronan, β-mannan
	GH10	endo-1,3-β-xylanase; endo-1,4-β-xylanase; arabinoxylan-specific endo-β-1,4-xylanase	β-glucan
	GH11	exo-1,4-β-xylosidase; endo-β-1,4-xylanase	
	GH115	xylan α-1,2-glucuronidase	
	GH120	β-xylosidase	
	GH141	xylanase	pectin
	GH2	β-xylosidase	α-mannan, arabinan, β-galactan, β-glucan, β-glucuronan, β-mannan, chitosan, host glycan, pectin
	GH26	β-1,3-xylanase	β-glucan, β-mannan
	GH3	xylan 1,4-β-xylosidase	arabinan, β-glucan, chitin, host glycan, peptidoglycan
	GH30	β-xylosidase; endo-β-1,4-xylanase; glucuronoarabinoxylan endo-β-1,4-xylanase	β-fucosides, β-glucan, β-glucuronan, host glycan
	GH4	α-glucuronidase	β-glucan, pectin, starch, sucrose
	GH43	β-1,3-xylosidase; β-xylosidase; α-L-arabinofuranosidase; xylanase	arabinan, β-galactan
	GH51	endo-β-1,4-xylanase	arabinan, β-glucan
	GH67	xylan α-1,2-glucuronidase; α-glucuronidase	
	GH95	α-L-galactosidase	xyloglucan, host glycan, pectin
	GH98	endo-β-1,4-xylanase	host glycan
Cellulose	GH48	reducing end-acting cellobiohydrolase; endo-β-1,4-glucanase	chitin
	GH74	endoglucanase	xyloglucan
	GH9	endoglucanase; exo-β-1,4-glucanase / cellodextrinase; cellobiohydrolase	β-glucan, chitosan
	GH94	cellobiose/cellodextrin/cellobionic acid phosphorylase	β-glucan, chitin
Xylan & Cellulose	GH39	exo-β-1,4-glucanase / cellodextrinase β-xylosidase; α-L-arabinofuranosidase	β-galactan, β-glucan, host glycan
	GH5	exo-β-1,4-glucanase / cellodextrinase; cellulose β-1,4-cellobiosidase arabinoxylan-specific endo-β-1,4-xylanase	arabinan, β-glucan, β-glycan, β-mannan, chitin, chitosan
	GH8	cellulase reducing-end-xylose releasing exo-oligoxylanase; endo-1,4-β-xylanase	β-glucan, chitosan

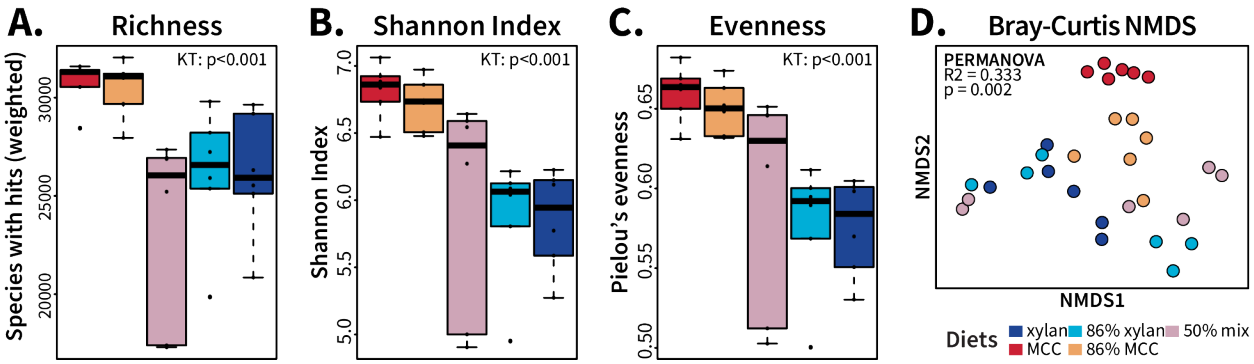
Figures



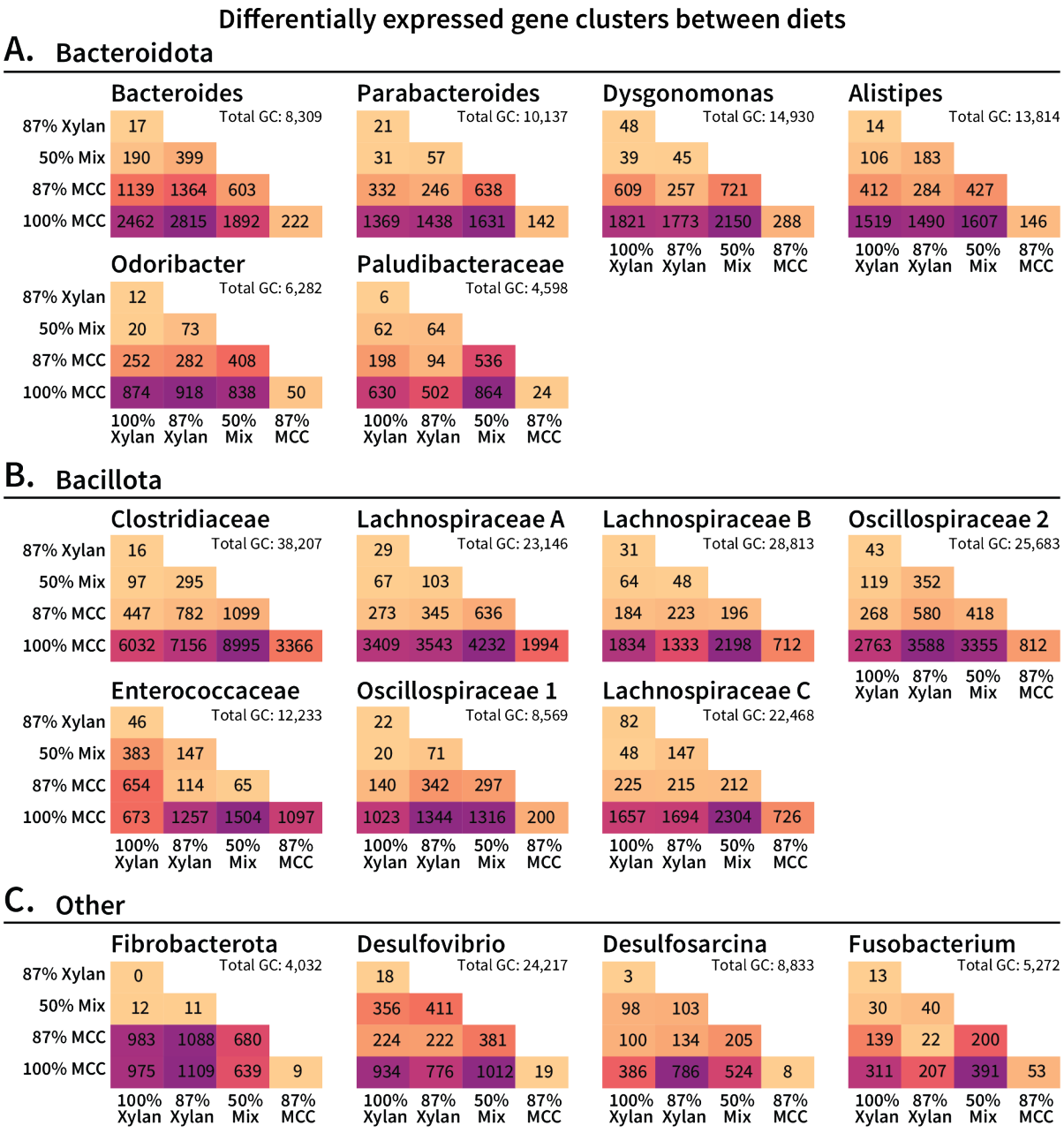
Supplemental Figure S4.1: Phylum-level community composition of cockroaches fed xylan differ from those fed cellulose. ASV count tables were aggregated at the phylum taxonomic level and converted to proportions. Phyla that comprised at least 1% of one sample were kept for visualization, while low abundant phyla were collapsed into "Other". MCC = microcrystalline cellulose



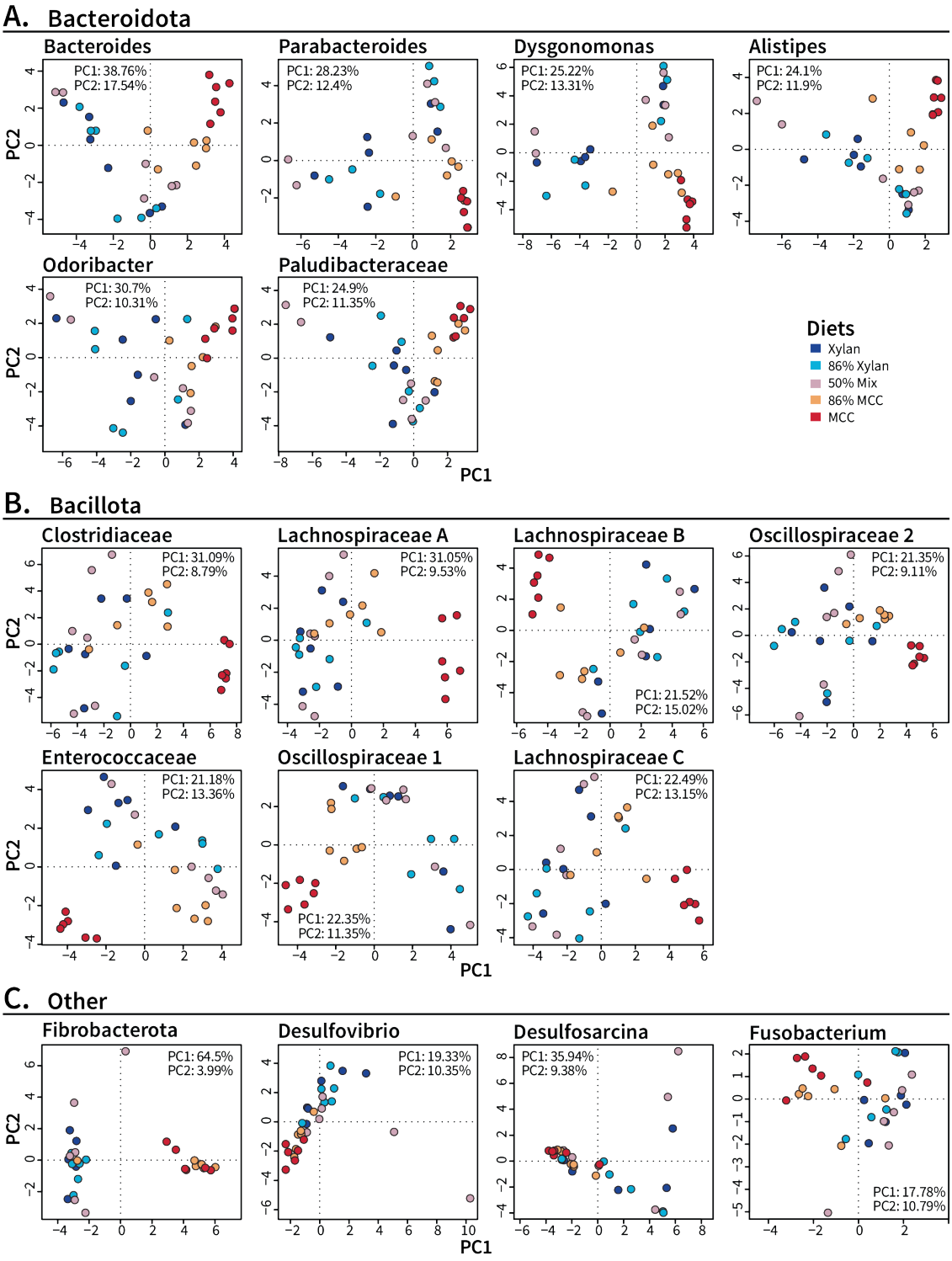
Supplemental Figure S4.2: Diet-based enrichment of ASVs on mixed and pure fiber diets. Enrichment of ASVs on the ratio diets was assessed with DESeq2, and pairwise results were extracted. (**A**) The total number of abundant (baseMean > 100) significantly differentially abundant ASVs identified in pairwise comparisons were plotted as a heatmap matrix, with color scaled within the total heatmap. (**B**) ASVs found to be enriched in 100% xylan, 100% cellulose, and any of the ratio diets were compared as a Venn diagram to identify overlap; since the ratio diets were aggregated, some ASVs appeared as significant in all three circles. (**C**) All abundant ASVs enriched by xylan were plotted against their log2 fold change vs the other five diets (pairwise comparison indicated by color). (**D**) All abundant ASVs enriched by cellulose were plotted against their log2 fold change vs the other five diets. MCC = microcrystalline cellulose; # =log2FC > 16



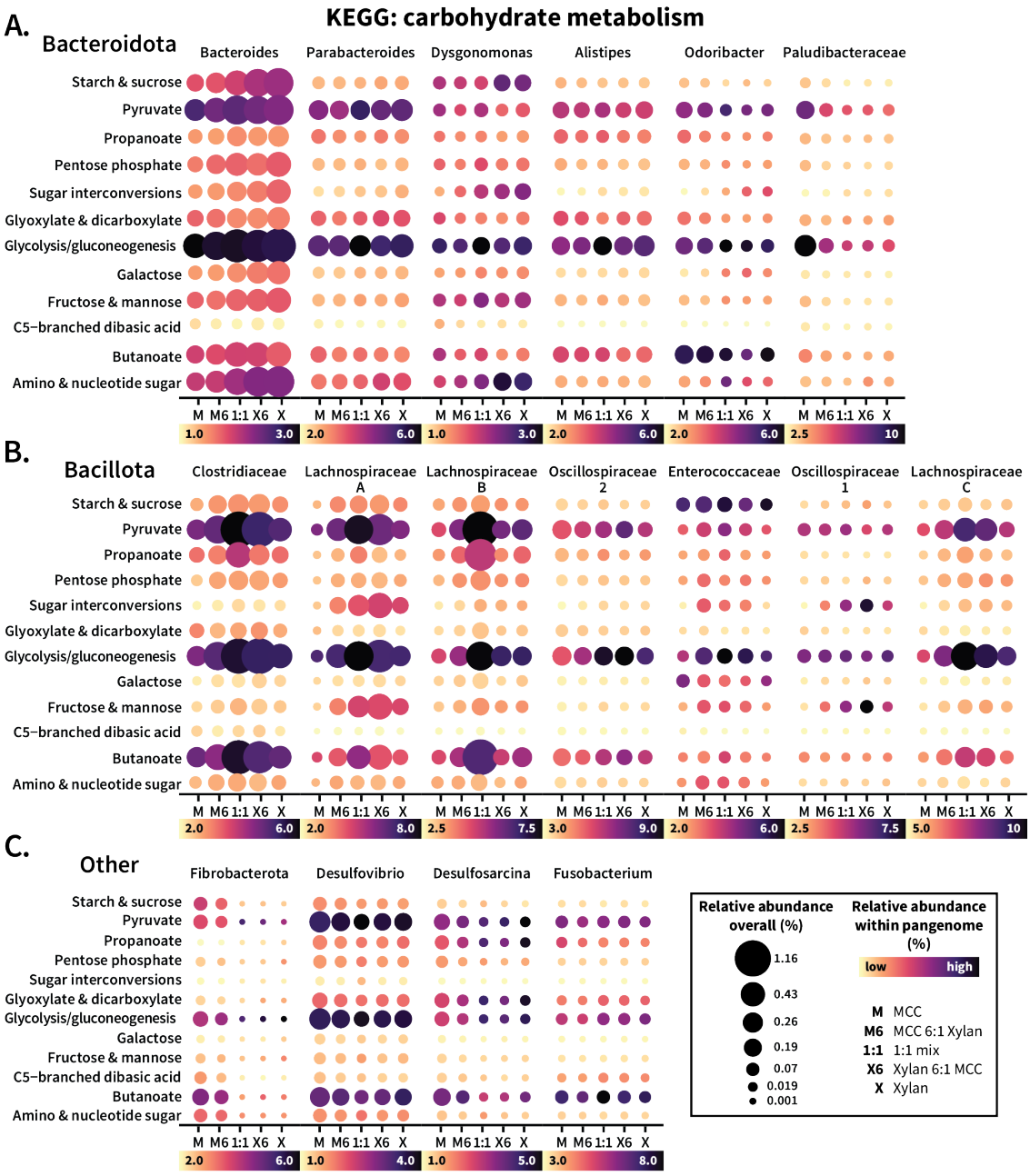
Supplemental Figure S4.3: Alpha and beta diversity of taxonomic composition of metatranscriptomes. The total counts of taxonomic identity assigned to the transcripts within each metatranscriptome was rarefied to 4,740,617 prior to comparing the alpha diversity measures (A) richness, (B) Shannon index, and (C) evenness as well as (D) Bray-Curtis dissimilarity of the different fiber diets. Significance of the alpha diversity tests was determined with Kruskal-Wallis test. Bray-Curtis dissimilarity ordination was produced by NMDS and evaluated for significance with PERMANOVA.



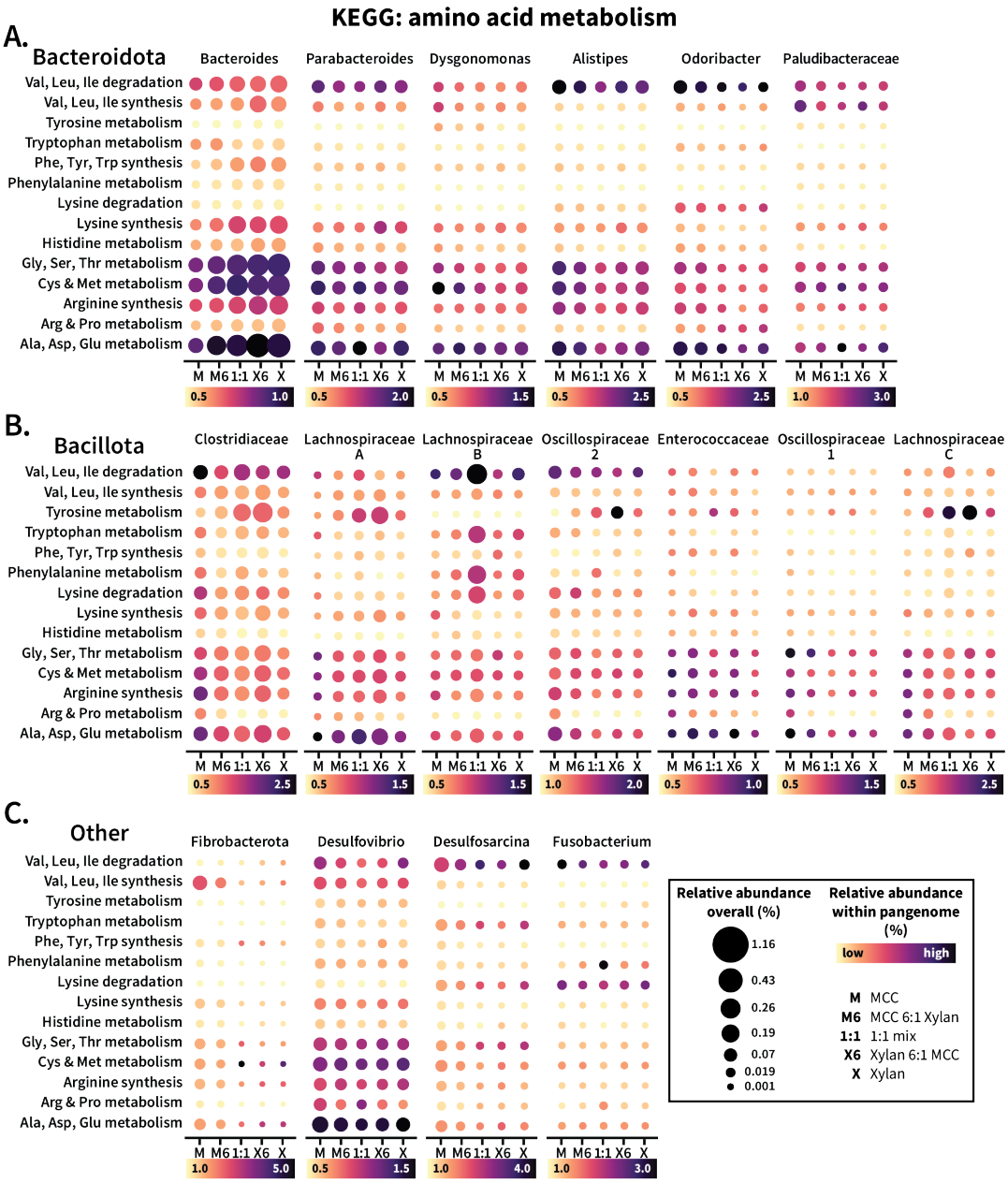
Supplemental Figure S4.4: Pangenome-specific differentially expressed gene clusters between fiber diets. Gene cluster count matrices were analyzed with DESeq2, fit to local sample dispersion. Pairwise results were pulled out using "contrast", and the total number of differentially expressed gene clusters with padj > 0.05 were summed for plotting. Heatmap color is scaled within each pangenome. GC: gene cluster.



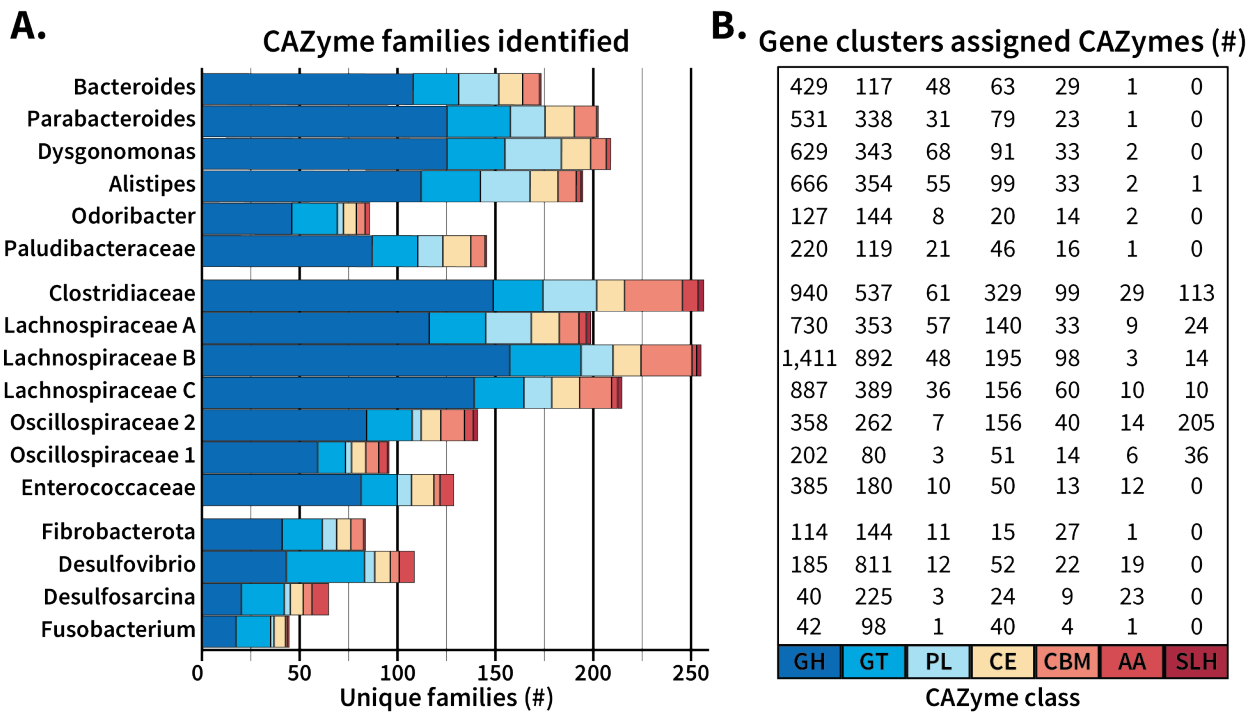
Supplemental Figure S4.5: Principal Component Analysis of gene cluster expression within each pangenome. Gene cluster expression for pangenomes from (**A**) *Bacteroidota,* (**B**) *Bacillota,* and (**C**) other pyla was normalized using the variance-stabilizing transformation provided with DESeq2 prior to unconstrained principal component calculations and subsequent ordination.



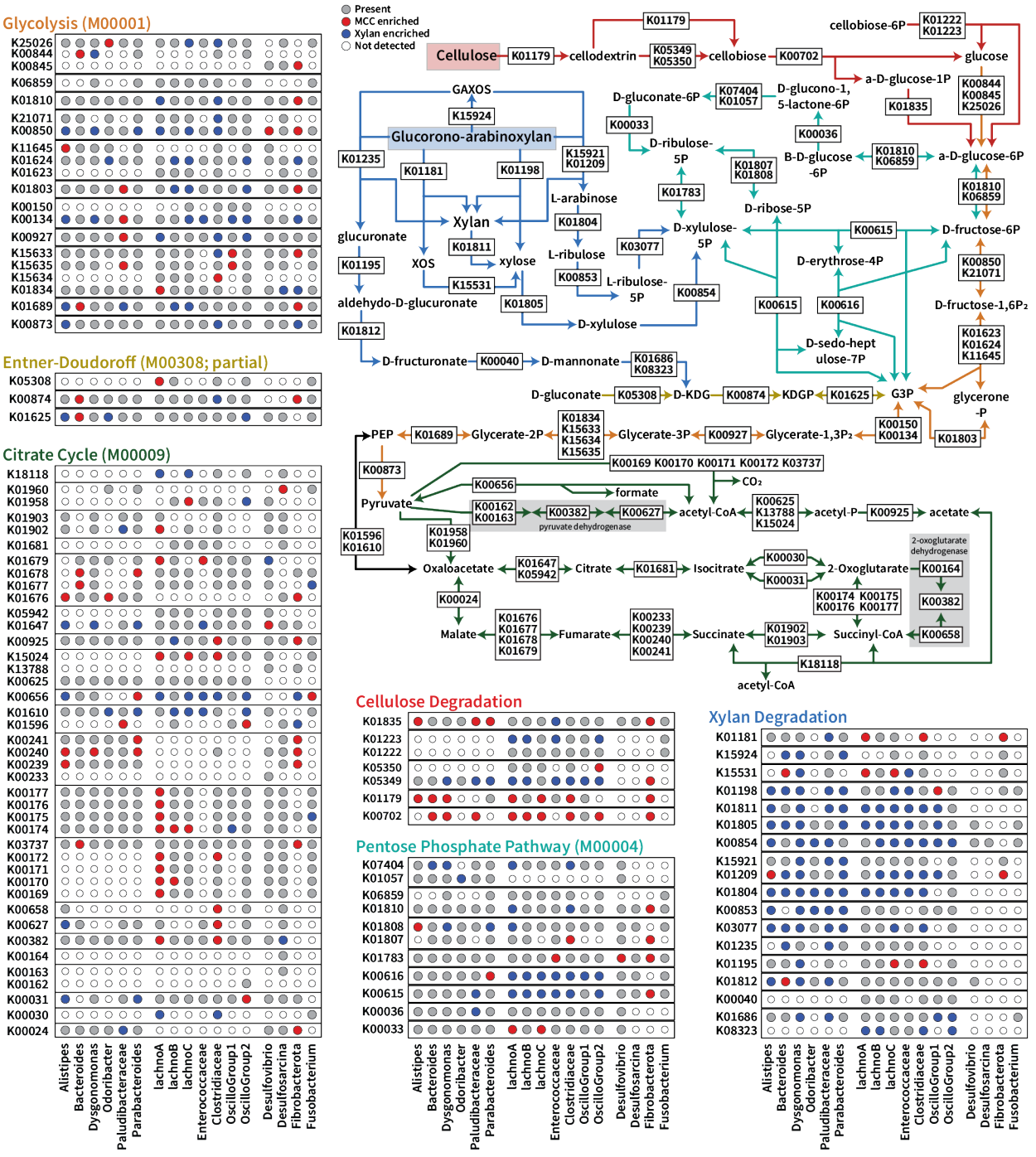
Supplemental Figure S4.6: Patterns of carbohydrate-degrading metabolic pathways across a gradient of fiber ratios. Pangenomes from (**A**) *Bacteroidota*, (**B**) *Bacillota*, and (**C**) other phyla were analyzed for KEGG carbohydrate metabolic pathway enrichment across the pure- and mixed-fiber diets. Gene clusters annotated with KEGG orthologs belonging to relevant metabolic pathway were summed together, divided by both pangenome transcripts (color) and total transcripts (size) per sample, then averaged by diet for plotting. Size is scaled based on total transcriptional abundance and therefore uses the same key for each pangenome, while color is scaled within each pangenome according to the key beneath each plot. M: 100% microcrystalline cellulose; M6: 86% cellulose; 1:1: 50% xylan and 50% cellulose; X6: 86% xylan; X: 100% xylan.



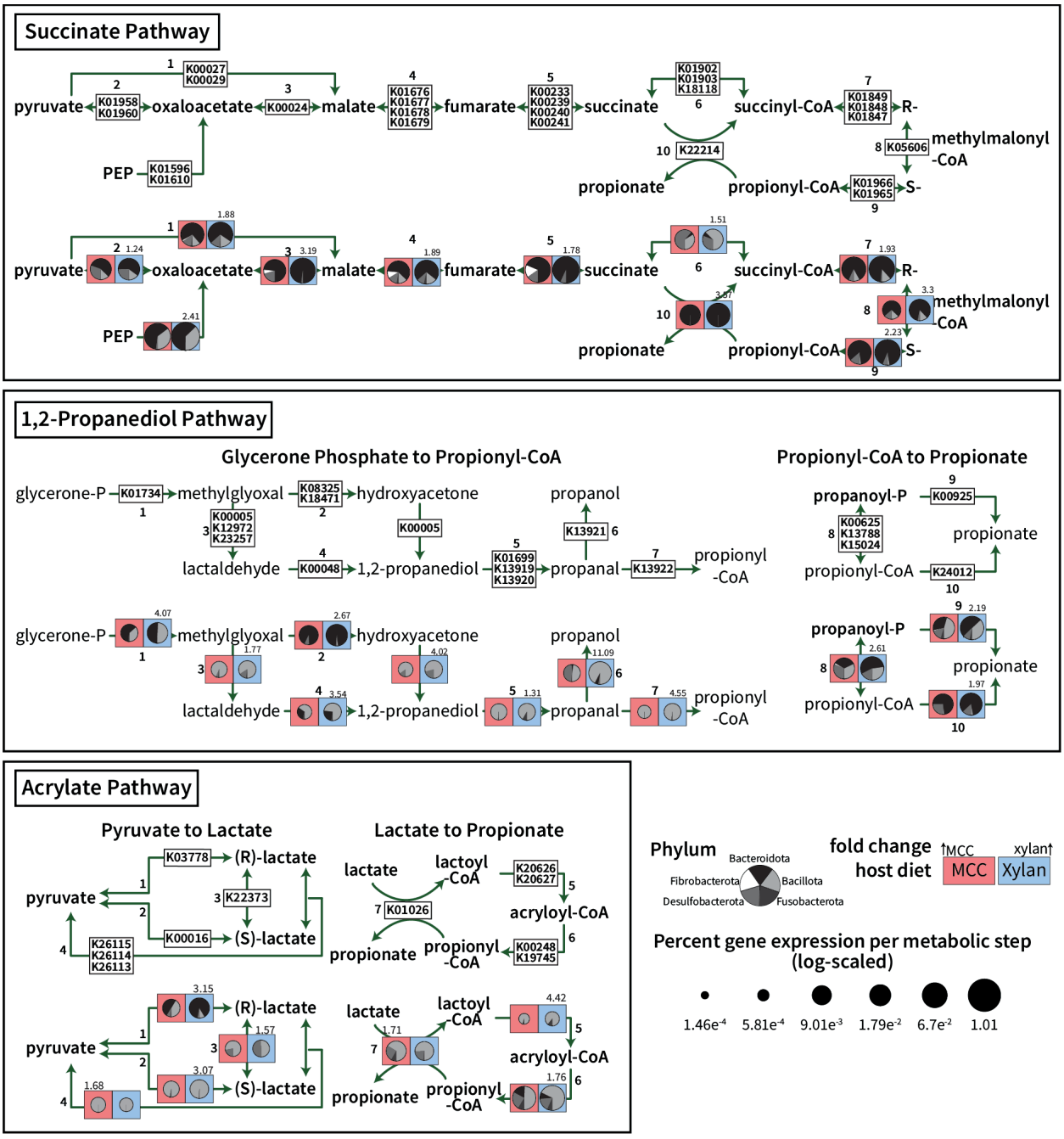
Supplemental Figure S4.7: Patterns of amino acid processing metabolic pathways across a gradient of fiber ratios. Pangenomes from (**A**) *Bacteroidota*, (**B**) *Bacillota*, and (**C**) other phyla were analyzed for KEGG amino acid metabolic pathway enrichment across the pure- and mixed-fiber diets. Gene clusters annotated with KEGG orthologs belonging to each relevant metabolic pathway were summed together, divided by both pangenome transcripts (color) and total transcripts (size) per sample, then averaged by diet for plotting. Size is scaled based on total transcriptional abundance and therefore uses the same key for each pangenome, while color is scaled within each pangenome according to the key beneath each plot. M: 100% microcrystalline cellulose; M6: 86% cellulose; 1:1: 50% xylan and 50% cellulose; X6: 86% xylan; X: 100% xylan.



Supplemental Figure S4.8: CAZyme family type and distribution expressed by the pangenomes. (A) The total number of unique CAZyme families found in the pangenomes assessed is plotted, colored by CAZyme class. **(B)** The total number of gene clusters annotated as each class. GH: glycoside hydrolase; GT: glycosyl transferase; PL: polysaccharide lyase; CE: carbohydrate esterase; AA: auxiliary activity; CBM: carbohydrate binding module; SLH: S-layer homology domain



Supplemental Figure S4.9: KEGG map and signficance by step of central carbon metabolism in cellulose and xylan degradation.



Supplemental Figure S4.10: Propionate production pathways in response to fiber. Three pathways of propionate production are presented, with pie charts created as described in Figure S4.5.