Nutrient Niche Space of *Clostridium difficile* Across Susceptible Microbiomes and the Impact of Infection on Metabolism of the Murine Cecal Microbiota

by

Matthew Lee Jenior

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

Doctoral committee:

Associate Professor Patrick D. Schloss, chair
Associate Professor Gregory J. Dick
Assistant Professor Nicole M. Koropatkin
Professor Thomas M. Schmidt
Professor Vincent B. Young

“Sometimes science is more art than science Morty, a lot of people don’t get that.”

- Rick Sanchez C-137

mljenior@umich.edu
ORCID: 0000-0003-1884-3543

ACKNOWLEDGEMENTS

Numerous people have had a tremendous impact on my life and the direction of my scientific career. First and foremost I must my thank my collaborator in research and life, Jhansi Leslie. Your passion and dedication for science is a constant source of inspiration and motivation. I am fortunate to have you in my life and look forward to continuing on our path together. My mentor Pat Schloss has been a constant pillar of support during my time in graduate school and has always been ready to help when I needed it the most. His influence has not only made me love the field I have decided to join, but his rigor as a scientist has made a large impact on how I view the work that we do and is something I hope to carry with me. I also must deeply thank the other members of my doctoral committee. Drs. Greg Dick, Nicole Koropatkin, Tom Schmidt, and Vince Young have each helped shape my ability to ask and answer scientific questions. Our meetings have each been a great experience and have guided me to put forth the best possible work that I could.

Many friends, family, colleagues, and fellow trainees along this process have had I large influence on my life and I would be remiss to no mention them as well. My parents have always been as supportive as anyone could possibly be though the ups and downs of my education, and I could really never thank them enough for all of the things, both tangible and intangible, they have given me throughout my life. My good friends the Baxters, the Szes, Zack Benet, and Corey Turner have all be extremely

close and supportive friends over the past 6 years (or beyond) and I now that we will remain close for years to come. Many current and former lab mates including Niel Baxter, Joe Zackular, Kathryn Iverson, Alyx Schubert, Hamide Sanini, Geoffrey Hannigan, Marc Sze, and Kaitlin Flynn. You have all been incredibly supportive throughout our time together and I will miss sharing a lab space with you. Several members of the Young Lab have also been instrumental in my growth as a scientist. Drs. Casey Theriot, Anna Seekatz, and Mark Koenigsknecht have each provided valuable advice for science and careers whenever I would ask. Several past mentors from my time as an undergraduate have also had a strong influence on the on the direction of my scientific career. Drs. Bradley Goodner, Dennis Taylor, and Louis Oliphant helped me find my passions in science, and these discoveries have directly impacted my thesis work. I would not be the scientist I am today without their teachings or the learning environment present at Hiram College. I would also like to thank the members of the UM Germ-free animal facility; Sara Poe, Natalie Anderson, and Chris Vowlles have been some of the best people I have ever worked with and always enjoyable to talk to. And of course my dog Penny, for giving me a reason to take a break from writing to go play some fetch.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The nosocomial pathogen *Clostridium difficile* causes an antibiotic-associated diarrheal disease, and largest single cause of hospital-acquired infection as well as gastroenteritis-associated death in the United States. The connection with prior antibiotic therapy is due to the collateral damage induced by these drugs on the community of indigenous bacteria that reside along the gastrointestinal tract. In its healthy state, the gut microbiota prevents the establishment of *C. difficile* in the gut through the intrinsic property known as colonization resistance. Following a perturbation, like exposure to antibiotics, this community becomes susceptible to colonization by the pathogen and subsequent disease. Most antibiotic classes have been associated with *C. difficile* infection (CDI) susceptibility; many leading to distinct community structures with unique metabolic profiles stemming from variation in bacterial targets of action. Additionally, a subset of these antibiotics are more closely associated with recurrent or persistent infection. In this thesis I demonstrate that certain susceptible gut communities are more permissive of long-term *C. difficile* colonization, and that the pathogen has a disproportionate effect on the metabolic activity of communities where persistence occurs. Taxonomic analyses of altered gene expression revealed that this effect consistently impact minority bacterial genera of the community across infection groups. In order to measure the adaptive capacity of *C. difficile* to these diverse environments, I also generated a genomic/transcriptomic-enabled metabolic modeling

platform to assess the differences in nutrient preference of *C. difficile* across different community contexts. This revealed the pathogen inhabited distinct nutrient niche spaces across susceptible gut environments. My dissertation work has strong implications in future research of targeted pre- and probiotic therapies that mitigate primary or established *C. difficile* colonization from the gastrointestinal tract.

CHAPTER ONE

Introduction

The Gut Microbiota and Impacts on Health

Humans are inhabited by trillions of indigenous microorganisms, in a complex ecosystem known as the microbiome. These communities of microbes are composed of diverse representatives from across the tree of life including bacteria, archaea, fungi, and viruses. Collections of these organisms have colonized every exterior surface of the body, but by far the most dense populations are found within the lower gastrointestinal tract. A recent estimate stated that there are most likely a nearly identical number of microbial cells living associated with humans as there are human cells in the entire body (1). Species surveys of this environment have revealed that more than 1,000 unique taxonomic groups can potentially be found here (2). Additionally, the largest percentage of this biomass has also been shown to be represented by bacteria. One study also found that there are approximately 10 trillion bacteria per gram of content in the distal colon (3). The species diversity of this consortia is only eclipsed by its own genetic diversity, in that it has been estimated that the gut microbiota is in possession of 150-fold more genes than are in the human genome (4). This extreme genetic diversity lends itself to an equally large metabolic capacity not present in human cells. Among these are examples of complex polysaccharide and mucin degradation (5),

methanogenesis (6), sulfate-reduction (7), and multiple forms of distinct anaerobic fermentation (8) to highlight a subset of functions.

These organisms have been present within us throughout evolution of both groups and as a result, each has developed processes that are interdependent (9). As early as the mid-1800's, the famed microbiologist Theodor Escherich began to recognize the possible importance of our resident microbes in our well-being (10). Only recently have we begun to appreciate mechanistic impacts that the microbiome plays in both health and disease. In terms of positive effects on the host, a number of groups have characterized instances of vitamin biosynthesis (11), enhanced nutrient acquisition (12), education of immunity (13), and resistance to colonization by pathogens (14). Disruption of normal community structure and function, or dysbiosis, has also been strongly linked to multiple disease states thus far. The most common of these effects has been connections to chronic inflammatory states which have been shown to exacerbate the progression of conditions like colorectal cancer (15) and inflammatory bowel disease (16). Connections have also been made to neurodevelopment (17) and degeneration (18), which could dramatically impact host cognition and behavior. Also, further connections to metabolic disorders like Type-2 diabetes and obesity (19, 20) have been made. There is still an incredible number of unanswered questions concerning the interactions of these communities of organisms and how this interspecies metabolic communication contributes to host physiology.

For the purpose of discussion throughout the subsequent chapters I would like to establish definitions for several terms I use frequently, at least in the context of this

thesis. First, microbiome refers to the complete collection of biotic and abiotic factors contained in the microenvironment of interest, this case the mammalian gut. The root of this word is biome, referring to the community of organisms and the environmental factors that drive their assembly in a habitat (21), modified to specify the micro- scale. Next, microbiota refers strictly the microscopic organisms that inhabit a given microbiome, biota being the root word which defines the member organisms of a biome. Within the microbiota, the metagenome is the pooled gene content of all organisms in the community. The operator here genome, is the full genetic content of a single organism, and the modifier meta- abstracts this to concept to beyond one genome into all of the genetic material present in the microbiota. Similarly, metatranscriptome is total active transcription of this assemblage under certain conditions at a given time, and is intrinsically dependent on the metagenome. Both genome and transcriptome have specific definitions at the individual organism scale, and are simply expanded to encompass more organisms simultaneously when applied to the larger microbiota. Finally, metabolome describes all distinct organic and inorganic metabolites in the environment of a microbiome at a particular time. As there has been debate at length to the proper definition of some of these terms, it is critical to the presentation of my dissertation work to understand the distinctions I make when describing my findings involving multiple layers of community-level biology at the same time.

Clostridium difficile Infection and Experimental Models

The etiological agent of *Clostridium difficile* infection, or CDI, is the bacterial pathogen *Clostridium difficile*. Susceptibility to colonization by *C. difficile* has been closely linked with previous antibiotic exposure. This trait is primarily mediated by the gut microbiota which will be discussed more at length in later chapters. CDI susceptibility has been associated with most antibiotics in the clinic with clindamycin, cephalosporins, fluoroquinolones being associated with increased reports of persistent infection (22). *C. difficile* produces two main toxins encoded by the genes *tcdA* and *tcdB* which are the mediators of disease. Briefly, these toxins enter the cytosol of host cells and glycosylates small GTP-binding proteins (Rho, Rac, and Cdc42), inactivating them. The consequence of this actin condensation, leading to cell-rounding, and ultimately cell death (23). Certain strains have also been found to be in possession of a binary toxin which causes actin depolymerization through ADP-ribosylation, and has been associated with increased CDI mortality (24). Through the actions of these toxins, CDI has been associated with a range of symptoms beginning with mild diarrhea, moving into severe inflammatory diarrhea, or even into the most severe symptoms of pseudomembranous colitis and toxic megacolon (25). In addition to toxin production, *C. difficile* has a complex bacterial life-cycle and possesses a spore phase that is thought to be the transmissive form of the pathogen (26). These endospores can persist for extended periods in the environment, and are highly aerotolerant, chemically tolerant, and UV resistant making them extremely difficult to remove from hospital surfaces (27). Life-cycle switches as well as control of virulence expression in *C. difficile* are the result

of signals integrated from the availability of certain metabolites in the environment it is colonizing (28). These differences in nutrient accessibility and the changes they induce in *C. difficile* physiology have implications in the overall microbial ecology of CDI in the gut and will be discussed in following chapters.

C. difficile itself is a Gram-positive, motile anaerobe that was first isolated by Hall and Toole in 1935 from healthy infant stool and was then known as Bacillus difficilus for it's fastidious *in vitro* growth (29). In spite of colonizing the infants asymptomatically, the authors were surprised to find that this species was highly pathogenic in both guinea pigs and rabbits mediated through production of a toxin. Complications associated with *C. difficile* infections became prevalent shortly after the inception of antibiotic use in the clinic in the 1950's. Physician's noted patients with increased incidence of pseudomembranous colitis following treatment with several classes of antibiotics which was at the time assumed to be due to *Staphylococcus aureus* infection (30). It was concluded later that a separate infectious agent must be responsible when clindamycin treated patients began to exhibit diarrhea and pseudomembranous colitis, with no culturable *S. aureus* (31). The first clinical recognition of infection by *C. difficile* came in 1977, when Bartlett et al. replicated the clindamycin-associated diarrhea in a mouse model and reisolated outgrown *C. difficile* from these animals (32). Over the next several decades, incidence of CDI world-wide has grown exponentially with outbreaks occurring on multiple continents, disproportionately impacting the elderly (33). Today, the prevalence of CDI in the United States has increased to nearly epidemic levels with an estimated half of a million infections, with ~29,000 leading to death of the patient

(34). The threat of CDI is compounded by the concomitant rise in the overall virulence of clinical isolates, potentially leading to more individuals reporting symptoms but with the same number of total infected (35). The situation is made even worse by growing detection rates of antibiotic resistant strains (36). These factors resulted in approximately $4.8 billion in acute health care costs in the US alone for 2015 (34). The urgency of this healthcare crisis necessitates the research of novel treatment options to stem the tide of infection and disease caused by *C. difficile*

As mentioned earlier, several animal models of CDI have been developed over the past several decades and have proven to be invaluable research tools for understanding the role of the gut microbiota in limiting this infection. The first *in vivo* experimental model of CDI was the Syrian hamster, which was easily colonized by *C. difficile* following pretreatment with clindamycin (37). Early reports failed to recognize *C. difficile* as the cause of the lethal diarrheal disease experienced by these animals, only later was the agent identified (38). This system proved invaluable for assessing the damage to host epithelia induced by the toxins (39), however the lethality of this model was not reflective of the disease course in humans. Gnotobiotic mice also provided a useful research platform (40), but were inherently susceptible to the infection and did not provide a realistic model of human infection. In 2008, a method of sensitizing conventionally-reared mice to *C. difficile* colonization was discovered. This model included pretreatment with 5 antibiotics in the drinking water, followed by an intraperitoneal injection of clindamycin prior to infection (41). This discovery was a boon for *C. difficile* researchers as the course of disease in mice more closely resembled

what was seen in humans. This model allowed for not only pathogenesis in a milder disease phenotype, but for studies of the normal murine microbiota in the context of CDI. Subsequent studies identified singly administered antibiotics that were sufficient to allow for *C. difficile* colonization and disease in a mouse model, providing additional tractability for studying differentially perturbed yet equally susceptible gut microbiomes (42–44). During experimentation for my thesis work I employed three separate antibiotic pretreatment regimes in conventional C57BL/6J mice, previously defined by these groups.

Colonization Resistance to *C. difficile* and Niche Theory

Colonization resistance refers to the intrinsic ability of the gut microbiota to inhibit the colonization and growth of invasive species, namely pathogens (45). The first description of antibiotic-associated infections, eventually linked to the microbiota, was made in 1943 but was incorrectly attributed to a delayed toxicity of penicillin (46). The general concept of protection via resident microbes was originally described experimentally in 1955 by Rolf Freter in reference to the ability of the gut microbes to prevent lethal *Cholera* infection in a guinea pig model (47). In the following year Freter made the significant finding that not only did pretreatment with streptomycin sensitize mice and guinea pigs to infection by Shigella, but that feeding these animals a streptomycin-resistant non-toxigenic strain of *Escherichia coli* prevent subsequent *Shigella* colonization (48). In 1962, it began to be referred to as “antibiotic-induced susceptibility” when it was found that pretreatment with streptomycin also reduced the

infectious dose of *Salmonella enterica typhimurium* in a mouse model by over 10,000 fold (49). The term Colonization Resistance itself came in 1971 when another group noted that antibiotic pretreatment not only increased susceptibility to immediate infection by *E. coli*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa*, but that the resistant state of the gut microbiota returned over a period of approximately two weeks following cessation of antibiotics (50). Freter's group later also characterized this phenomena in a gnotobiotic mouse model were pre-colonized by intestinal microbiota from conventionally-reared mice for *C. difficile* infection (51). Then in 1988, it was shown that colonization resistance, at least to certain pathogens including *C. difficile*, was mediated by competition for growth nutrients through continuous flow reactor experiments, where sterilized intestinal content from resistant mice would not grow the pathogen without supplementation with glucose (52).

As a pioneer in this field, Freter defined what would later be referred to as the nutrient-niche hypothesis which offered an explanation as to how the healthy gut microbiota mediated the exclusion of pathogenic bacterial species (53). The basic ecological framework was based on a series of three assumptions: 1. There are a finite number of niches in the gut, 2. A species must be better than all other competitors at acquiring at least a single nutrient source, and 3. That a species must reproduce faster that the washout rate of the intestine (54). Over the years, this concept has been supported by additional work in competitive exclusion of carbohydrate-based nutrient niches *in vivo* by a non-pathogenic strain of *E. coli* against pathogenic strains of the bacterium (55, 56). Additionally, it was later shown that this same non-pathogenic *E. coli*

outcompetes *S. typhimurium* for iron, and eliminates the infection from the gut of mice (57). These data provide support that in the case of colonization resistance to at least certain species which include *C. difficile*, competition for nutrients was an effective means of exclusion. As a side note, further mechanisms for colonization resistance to other bacterial pathogens have been demonstrated including bacteriocin/microcin-based antagonism (58), modulation of host immunity (59), or indirect antagonism through bioconversion of specific metabolites (60) among factors that may determine resistance to various infections.

Subsequent studies of *C. difficile* colonization resistance specifically have been met with more mixed results in attempts to identify specific competitors in the gut. Moreover, it has also been shown that separate classes of antibiotics that cause CDI susceptibility in a mouse model, each lead to distinct gut metabolomes with highly divergent nutrient profiles available in each (61–63). The fact that *C. difficile* colonizes each of these environments effectively promotes the hypothesis that it is a metabolic generalist, as the pathogen possesses many traits that define these organisms. In microbial ecology, an ecological generalist is a species that can colonize a wide variety of habitats, a trait which also necessitates a large enough spectrum of nutrient utilization capabilities to support this behavior (64). *C. difficile* has been found to infect a large variety of animal species including numerous small and large mammals as well as birds (65). *C. difficile* has also been demonstrated to utilize a large range of substrates for growth including simple sugars, complex polysaccharides, amino acids and peptides (66). This ability is likely due to its large and variable genomic capacity (67), which is

another trait of many bacterial generalists that are able to adapt to diverse and fluctuating environments (68). These features are not congruent with the nutrient-niche hypothesis in which species that persistently colonize require discrete specialization. As such, although this theory has been largely explanatory of competition for nutrients by members of the gut microbiota against specialized pathogenic species, it appears to require revision in the case of a pathogen with a large possible nutrient niche space. Furthermore, Freter's hypothesis relies on an additional assumption of even mixing of species and nutrients through the gut environment. This is not the reality as specific biogeographic habitats in the gut and stochastic processes drive an unequal distribution of competitors and their respective niche spaces across most of the intestinal environment (69). I propose a more expanded theory of the nutrient niche landscape in the gut that is more reminiscent of current niche theory from macroecology, and I explore this in Chapters 2 and 3.

The concept of organismal niches as single, discrete units is more in line with the first widely accepted definition of a niche from Joseph Grinnell in 1917 (70). This description strictly included the behaviors and habitat requirements that allowed a species to persist in the environment for which they are adapted. This theory allowed for competitive exclusion as well as "empty" niches, and was famously supported by George Gause in the formation of Gause's Law which stated that two species competing for the same limited resource cannot coexist (71). This concept is the best analogy for the existing nutrient-niche hypothesis framework for the gut. However, the definition of a niche was then broadened in 1927 by Charles Elton to now encompass

the ability of an organism to change behavior to actively adapt to it's habitat and manipulate elements of it in order to maintain colonization, thus allowing for more than one species at least partially inhabiting the same niche (72). Elton's theory brought forth the foundation for the concepts of an organism's fundamental versus its realized niche. Fundamental niche refers to all aspects of an environment that an organism is able to colonize in the absence of all competitors, while realized niche are the subset of these components that are actually utilized when in actual competition with other organisms. This concept is reflective of findings discussed on Chapter 3. Finally in 1957 George Hutchinson proposed an additional evolution to the concept of a niche, which brought it to our modern understanding of the distribution and interaction of realized niches in nature (73). His description now included an n-dimensional hypervolume representing the total "space" that organisms can inhabit in a boundaried habitat (or biotope) where each axis is a distinct resource or environment condition. Along each axis is the full spectrum within which that condition or resource can vary. In this framework, separate organisms inhabit a segregated portion of mutli-dimensional space based on their capabilities and current behavior within the overall niche space. This definition of a niche is far more explanatory of species co-existence with niche overlaps as competing species can partially segregate from each other within their respective areas of the hypervolume (74). One could imagine the scenario of antibiotic treatment to this community diminishing the volume of some species and their niche space in the gut environment biotope. Remaining species with previous segregation, but shared niche qualities would then reorient to inhabit various portions of vacated space. One could

then imagine an additional species, or pathogen, that is introduced and now able to colonize the restructured biotope. Hypervolumes would be vacated differentially across distinct antibiotic treatments, necessitating generalism to colonize successfully, as is case with *C. difficile*.

Direction of Thesis Work

The focus of my thesis has been on analyzing metabolic properties of the microbes that resides in the gastrointestinal tract, specifically the bacterial component of this community in the setting of infection by *C. difficile*. My research has revolved around the central hypothesis that distinct antibiotic classes result in microbiomes with divergent metabolic challenges that *C. difficile* must adapt to or manipulate in order to succeed across conditions. By leveraging the mouse model of infection previously described, I utilized distinct classes of antibiotic pretreatments to characterize the behavior of the gut microbiome, including *C. difficile*, during infection when compared to uninfected mice. In the subsequent chapters I discuss two sides of this interplay; changes at the community level in response to infection, and metabolic adaptation of *C. difficile* to varied susceptible gut microbiome. More specifically, in Chapter 2 I address shifts in both the metatranscriptomes and metabolomes of infected mice as a product of *C. difficile* colonization and the potential ramifications of these changes in long-term colonization. Then in Chapter 3 I cover the metabolic adaptations of *C. difficile* to each distinct microbiome and how that may relate to disease, using genome-scale metabolic modeling techniques integrated with transcriptomic analysis. Together, these research

directions have implications for the ecology of infection in the mammalian gut which I discuss in the final chapter as well as future directions which may ultimately reveal novel treatment approaches to *C. difficile* infection.

References

1. **Sender, R.**, **S. Fuchs**, **R. Milo**. 2016. Revised estimates for the number of human and bacteria cells in the body. PLOS Biology **14**:036103.

2. **Lozupone, C. A.**, **J. I. Stombaugh**, **J. I. Gordon**, **J. K. Jansson**, and **R. Knight**. 2012. Diversity, stability and resilience of the human gut microbiota. Nature **489**:220–230.

3. **Savage, D. C.** 1977. Microbial Ecology of the Gastrointestinal Tract. Annual Review of Microbiology **31**:107–133.

4. **Qin, J.**, **R. Li**, **J. Raes**, **M. Arumugam**, **S. Burgdorf**, **C. Manichanh**, **T. Nielsen**, **N. Pons**, **T. Yamada**, **D. R. Mende**, **J. Li**, **J. Xu**, **S. Li**, **D. Li**, **J. Cao**, **B. Wang**, **H. Liang**, **H. Zheng**, **Y. Xie**, **J. Tap**, **P. Lepage**, **M. Bertalan**, **J.-m. Batto**, **T. Hansen**, **D. L. Paslier**, **A. Linneberg**, **H. B. Nielsen**, **E. Pelletier**, **P. Renault**, **Y. Zhou**, **Y. Li**, **X. Zhang**, **S. Li**, **N. Qin**, and **H. Yang**. 2010. A human gut microbial gene catalog established by metagenomic sequencing. Nature **464**:59–65.

5. **Flint, H. J.**, **K. P. Scott**, **S. H. Duncan**, **P. Louis**, and **E. Forano**. 2012. Microbial degradation of complex carbohydrates in the gut. **3**:289–306.

6. **Gaci, N.**, **G. Borrel**, **W. Tottey**, **P. W. O'Toole**, and **J. F. Brugère**. 2014. Archaea and the human gut: New beginning of an old story **20**:16062–16078.

7. **Rey, F. E.**, **M. D. Gonzalez**, **J. Cheng**, **M. Wu**, **P. P. Ahern**, and **J. I. Gordon**. 2013. Metabolic niche of a prominent sulfate-reducing human gut bacterium. Proceedings of the National Academy of Sciences **110**:13582–13587.

8. **Bernalier-Donadille, A.** 2010. Fermentative metabolism by the human gut microbiota. Gastroenterologie Clinique et Biologique **34**:S16–S22.

9. **Schnorr, S. L.**, **K. Sankaranarayanan**, **C. M. Lewis**, and **C. Warinner**. 2016. Insights into human evolution from ancient and contemporary microbiome studies **41**:14–26.

10. **Escherich, T. H.** 1989. The intestinal bacteria of the neonate and breast-fed infant. Reviews of Infectious Diseases **11**:352–356.

11. **Magnúsdóttir, S.**, **D. Ravcheev**, **V. De Crécy-Lagard**, and **I. Thiele**. 2015. Systematic genome assessment of B-vitamin biosynthesis suggests cooperation among gut microbes. Frontiers in Genetics **6**.

12. **Bäckhed, F.**, **H. Ding**, **T. Wang**, **L. V. Hooper**, **G. Y. Koh**, **A. Nagy**, **C. F. Semenkovich**, and **J. I. Gordon**. 2004. The gut microbiota as an environmental factor that regulates fat storage. Proceedings of the National Academy of Sciences of the United States of America **101**:15718–23.

13. **Wu, H.-J.**, and **E. Wu**. 2012. The role of gut microbiota in immune homeostasis and autoimmunity. Gut Microbes **3**:4–14.

14. **Buffie, C. G.**, and **E. G. Pamer**. 2013. Microbiota-mediated colonization resistance against intestinal pathogens. Nature Reviews Immunology **13**:790–801.

15. **Zackular, J. P.**, **N. T. Baxter**, **K. D. Iverson**, **W. D. Sadler**, **J. F. Petrosino**, **G. Y. Chen**, and **P. D. Schloss**. 2013. The gut microbiome modulates colon tumorigenesis. mBio **4**.

16. **Sartor, R. B.**, and **S. K. Mazmanian**. 2012. Intestinal Microbes in Inflammatory Bowel Diseases. The American Journal of Gastroenterology Supplements **1**:15–21.

17. **Hsiao, E. Y.**, **S. W. McBride**, **S. Hsien**, **G. Sharon**, **E. R. Hyde**, **T. McCue**, **J. A. Codelli**, **J. Chow**, **S. E. Reisman**, **J. F. Petrosino**, **P. H. Patterson**, and **S. K. Mazmanian**. 2013. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell **155**:1451–1463.

18. **Sampson, T. R.**, **J. W. Debelius**, **T. Thron**, **S. Janssen**, **G. G. Shastri**, **Z. E. Ilhan**, **C. Challis**, **C. E. Schretter**, **S. Rocha**, **V. Gradinaru**, **M. F. Chesselet**, **A. Keshavarzian**, **K. M. Shannon**, **R. Krajmalnik-Brown**, **P. Wittung-Stafshede**, **R. Knight**, and **S. K. Mazmanian**. 2016. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. Cell **167**:1469–1480.e12.

19. **Devaraj, S.**, **P. Hemarajata**, and **J. Versalovic**. 2013. The human gut microbiome and body metabolism: Implications for obesity and diabetes **59**:617–628.

20. **Sanz, Y.**, **M. Olivares**, **Á. Moya-Pérez**, and **C. Agostoni**. 2015. Understanding the role of gut microbiome in metabolic disease risk. Pediatric Research **77**:236–244.

21. **Whittaker, R. H.** 1975. Communities and ecosystems. Communities and ecosystems.

22. **Gerding, D.** 2004. Clindamycin, Cephalosporins, Fluoroquinolones, and *Clostridium difficile*–Associated Diarrhea: This Is an Antimicrobial Resistance Problem. Clin Infect Dis **38**:646–648.

23. **Voth, D. E.**, and **J. D. Ballard**. 2005. *Clostridium difficile* toxins: Mechanism of action and role in disease **18**:247–263.

24. **Gerding, D. N.**, **S. Johnson**, **M. Rupnik**, and **K. Aktories**. 2014. *Clostridium difficile* binary toxin CDT: mechanism, epidemiology, and potential clinical importance **5**:15–27.

25. **Bartlett, J. G.**, and **D. N. Gerding**. 2008. Clinical Recognition and Diagnosis of Clostridium difficile Infection. Clinical Infectious Diseases **46**:S12–S18.

26. **Deakin, L. J.**, **S. Clare**, **R. P. Fagan**, **L. F. Dawson**, **D. J. Pickard**, **M. R. West**, **B. W. Wren**, **N. F. Fairweather**, **G. Dougan**, and **T. D. Lawley**. 2012. The *Clostridium difficile* spo0A gene is a persistence and transmission factor. Infection and Immunity **80**:2704–2711.

27. **Lawley, T. D.**, **S. Clare**, **L. J. Deakin**, **D. Goulding**, **J. L. Yen**, **C. Raisen**, **C. Brandt**, **J. Lovell**, **F. Cooke**, **T. G. Clark**, and **G. Dougan**. 2010. Use of purified *Clostridium difficile* spores to facilitate evaluation of health care disinfection regimens. Applied and Environmental Microbiology **76**:6895–6900.

28. **Bouillaut, L.**, **T. Dubois**, **A. L. Sonenshein**, and **B. Dupuy**. 2015. Integration of metabolism and virulence in *Clostridium difficile*. Research in Microbiology **166**:375–383.

29. **Hall, I.**, and **E. O'Toole**. 1935. Intestinal flora in new-born infants: With a description of a new pathogenic anaerobe, Bacillus difficilis. American Journal of Diseases of Children **49**:390–402.

30. **Altemeier, W.**, **R. Hummel**, and **E. Hill**. 1963. Staphylococcal Enterocolitis Following Antibiotic Therapy. Annals of Surgery **157**:847–857.

31. **Tedesco, F.**, **R. Barton**, and **D. Alpers**. 1974. Clindamycin-Associated Colitis. A Prospective Study. Annals of Internal Medicine **81**:429–433.

32. **Bartlett, J. G.**, **a B. Onderdonk**, **R. L. Cisneros**, and **D. L. Kasper**. 1977. Clindamycin-associated colitis due to a toxin-producing species of Clostridium in hamsters. The Journal of infectious diseases **136**:701–705.

33. **Depestel, D. D.**, and **D. M. Aronoff**. 2013. Epidemiology of *Clostridium difficile* infection. Journal of pharmacy practice **26**:464–75.

34. **Lessa, F. C.**, **Y. Mu**, **W. M. Bamberg**, **Z. G. Beldavs**, **G. K. Dumyati**, **J. R. Dunn**, **M. M. Farley**, **S. M. Holzbauer**, **J. I. Meek**, **E. C. Phipps**, **L. E. Wilson**, **L. G. Winston**, **J. A. Cohen**, **B. M. Limbago**, **S. K. Fridkin**, **D. N. Gerding**, and **L. C. McDonald**. 2015. Burden of *Clostridium difficile* Infection in the United States. New England Journal of Medicine **372**:825–834.

35. **Quesada-Gomez, C.**, **D. Lopez-Urena**, **L. Acuna-Amador**, **M. Villalobos-Zuniga**, **T. Du**, **R. Freire**, **C. Guzman-Verri**, **M. D. M. Gamboa-Coronado**, **T. D. Lawley**, **E. Moreno**, **M. R.Mulvey**, **G. A. D. C. Brito**, **E. Rodriguez-Cavallini**, **C. Rodriguez**, and **E. Chaves-Olartea**. 2015. Emergence of an outbreak-associated *Clostridium difficile* variant with increased virulence. Journal of Clinical Microbiology **53**:1216–1226.

36. **CDC**. 2013. Antibiotic Resistance Threats in the United States, 2013 | Antibiotic/Antimicrobial Resistance | CDC.

37. **Small, J.** 1968. Fatal enterocolitis in hamsters given lincomycin hydrochloride. Lab Anim Care **18**:411–420.

38. **Bartlett, J. G.**, **N. Moon**, **T. W. Chang**, **N. Taylor**, and **A. B. Onderdonk**. 1978. Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. Gastroenterology **75**:778–782.

39. **Price, a B.**, **H. E. Larson**, and **J. Crow**. 1979. Morphology of experimental antibiotic-associated enterocolitis in the hamster: a model for human pseudomembranous colitis and antibiotic-associated diarrhoea. Gut **20**:467–475.

40. **Wilson, K. H.**, **J. N. Sheagren**, **R. Freter**, **L. Weatherbee**, and **D. Lyerly**. 1986. Gnotobiotic Models for Study of the Microbial Ecology of *Clostridium difficile* and Escherichia coli. Journal of Infectious Diseases **153**:547–551.

41. **Chen, X.**, **K. Katchar**, **J. D. Goldsmith**, **N. Nanthakumar**, **A. Cheknis**, **D. N. Gerding**, and **C. P. Kelly**. 2008. A Mouse Model of *Clostridium difficile*-Associated Disease. Gastroenterology **135**:1984–1992.

42. **Theriot, C. M.**, **C. C. Koumpouras**, **P. E. Carlson**, **I. I. Bergin**, **D. M. Aronoff**, and **V. B. Young**. 2011. Cefoperazone-treated mice as an experimental platform to assess differential virulence of *Clostridium difficile* strains. Gut microbes **2**:326–334.

43. **Lawley, T. D.**, **S. Clare**, **A. W. Walker**, **D. Goulding**, **R. A. Stabler**, **N. Croucher**, **P. Mastroeni**, **P. Scott**, **C. Raisen**, **L. Mottram**, **N. F. Fairweather**, **B. W. Wren**, **J. Parkhill**, and **G. Dougan**. 2009. Antibiotic treatment of *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. Infection and Immunity **77**:3661–3669.

44. **Schubert, A. M.**, **H. Sinani**, and **P. D. Schloss**. 2015. Antibiotic-induced alterations of the murine gut microbiota and subsequent effects on colonization resistance against *Clostridium difficile*. mBio **6**.

45. **Lawley, T. D.**, and **A. W. Walker**. 2013. Intestinal colonization resistance. Immunology **138**:1–11.

46. **Hamre, D.** 1943. The toxicity of penicillin as prepared for clinical use. The American journal of the medical sciences **206**:642.

47. **Rolf Freter**. 1955. The Fatal Enteric Cholera Infection in the Guinea Pig, Achieved by Inhibition of Normal Enteric Flora. The Journal of Infectious Diseases **97**:57–65.

48. **Freter, R.** 1956. Experimental enteric Shigella and Vibrio infections in mice and guinea pigs. The Journal of experimental medicine **104**:411–418.

49. **Bohnhoff, M.**, and **C. P. Miller**. 1962. Enhanced susceptibility to salmonella infection in streptomycin-treated mice. Journal of Infectious Diseases **111**:117–127.

50. **Waaij, D. van der**, **J. M. Berghuis-de Vries**, and **J. E. C. Lekkerkerk-van der Wees**. 1971. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. The Journal of hygiene **69**:405–11.

51. **Wilson, K. H.**, **J. N. Sheagren**, **R. Freter**, **L. Weatherbee**, and **D. Lyerly**. 1986. Gnotobiotic Models for Study of the Microbial Ecology of *Clostridium difficile* and Escherichia coli. Journal of Infectious Diseases **153**:547–551.

52. **Wilson, K. H.**, and **F. Perini**. 1988. Role of competition for nutrients in suppression of *Clostridium difficile* by the colonic microflora. Infection and Immunity **56**:2610–2614.

53. **Freter, R.**, **H. Brickner**, **J. Fekete**, **M. M. Vickerman**, and **K. E. Carey**. 1983. Survival and implantation of Escherichia coli in the intestinal tract. Infection and Immunity **39**:686–703.

54. **Freter, R.**, **H. Brickner**, **M. Botney**, **D. Cleven**, and **A. Aranki**. 1983. Mechanisms That Control Bacterial Populations in Continuous-Flow Culture Models of Mouse Large Intestinal Flora. Infection and Immunity **39**:676–685.

55. **Chang, D.-E.**, **D. J. Smalley**, **D. L. Tucker**, **M. P. Leatham**, **W. E. Norris**, **S. J. Stevenson**, **A. B. Anderson**, **J. E. Grissom**, **D. C. Laux**, **P. S. Cohen**, and **T. Conway**. 2004. Carbon nutrition of Escherichia coli in the mouse intestine. Proceedings of the National Academy of Sciences **101**:7427–7432.

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

57. **Deriu, E.**, **J. Z. Liu**, **M. Pezeshki**, **R. A. Edwards**, **R. J. Ochoa**, **H. Contreras**, **S. J. Libby**, **F. C. Fang**, and **M. Raffatellu**. 2013. Probiotic bacteria reduce salmonella typhimurium intestinal colonization by competing for iron. Cell Host and Microbe **14**:26–37.

58. **Sassone-Corsi, M.**, **S.-P. Nuccio**, **H. Liu**, **D. Hernandez**, **C. T. Vu**, **A. A. Takahashi**, **R. A. Edwards**, and **M. Raffatellu**. 2016. Microcins mediate competition among Enterobacteriaceae in the inflamed gut. Nature **540**:280–283.

59. **Inagaki, H.**, **T. Suzuki**, **K. Nomoto**, and **Y. Yoshikai**. 1996. Increased susceptibility to primary infection with Listeria monocytogenes in germfree mice may be due to lack of accumulation of L- selectin+ CD44+ T cells in sites of inflammation. Infection and Immunity **64**:3280–3287.

60. **Buffie, C. G.**, **V. Bucci**, **R. R. Stein**, **P. T. McKenney**, **L. Ling**, **A. Gobourne**, **D. No**, **H. Liu**, **M. Kinnebrew**, **A. Viale**, **E. Littmann**, **M. R. M. van den Brink**, **R. R. Jenq**, **Y. Taur**, **C. Sander**, **J. R. Cross**, **N. C. Toussaint**, **J. B. Xavier**, and **E. G. Pamer**. 2014. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. Nature **517**:205–208.

61. **Antunes, L. C. M.**, **J. Han**, **R. B. R. Ferreira**, **P. Lolić**, **C. H. Borchers**, and **B. B. Finlay**. 2011. Effect of antibiotic treatment on the intestinal metabolome. Antimicrobial Agents and Chemotherapy **55**:1494–1503.

62. **Jump, R. L. P.**, **A. Polinkovsky**, **K. Hurless**, **B. Sitzlar**, **K. Eckart**, **M. Tomas**, **A. Deshpande**, **M. M. Nerandzic**, and **C. J. Donskey**. 2014. Metabolomics analysis identifies intestinal microbiota-derived biomarkers of colonization resistance in clindamycin-treated mice. PLoS ONE **9**.

63. **Theriot, C. M.**, **M. J. Koenigsknecht**, **P. E. Carlson**, **G. E. Hatton**, **A. M. Nelson**, **B. Li**, **G. B. Huffnagle**, **J. Z. Li**, and **V. B. Young**. 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. Nature Communications **5**:3114.

64. **Monard, C.**, **S. Gantner**, **S. Bertilsson**, **S. Hallin**, and **J. Stenlid**. 2016. Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. Scientific Reports **6**:37719.

65. **Keessen, E. C.**, **W. Gaastra**, and **L. J. A. Lipman**. 2011. *Clostridium difficile* infection in humans and animals, differences and similarities **153**:205–217.

66. **Scaria, J.**, **J. W. Chen**, **N. Useh**, **H. He**, **S. P. McDonough**, **C. Mao**, **B. Sobral**, and **Y. F. Chang**. 2014. Comparative nutritional and chemical phenome of *Clostridium difficile* isolates determined using phenotype microarrays. International Journal of Infectious Diseases **27**:20–25.

67. **Sebaihia, M.**, **B. W. Wren**, **P. Mullany**, **N. F. Fairweather**, **N. Minton**, **R. Stabler**, **N. R. Thomson**, **A. P. Roberts**, **A. M. Cerdeño-Tárraga**, **H. Wang**, **M. T. Holden**, **A. Wright**, **C. Churcher**, **M. A. Quail**, **S. Baker**, **N. Bason**, **K. Brooks**, **T. Chillingworth**, **A. Cronin**, **P. Davis**, **L. Dowd**, **A. Fraser**, **T. Feltwell**, **Z. Hance**, **S. Holroyd**, **K. Jagels**, **S. Moule**, **K. Mungall**, **C. Price**, **E. Rabbinowitsch**, **S. Sharp**, **M. Simmonds**, **K. Stevens**, **L. Unwin**, **S. Whithead**, **B. Dupuy**, **G. Dougan**, **B. Barrell**, and **J. Parkhill**. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nature Genetics **38**:779–786.

68. **Bentkowski, P.**, **C. Van Oosterhout**, and **T. Mock**. 2015. A model of genome size evolution for prokaryotes in stable and fluctuating environments. Genome Biology and Evolution **7**:2344–2351.

69. **Pereira, F. C.**, and **D. Berry**. 2017. Microbial nutrient niches in the gut **19**:1366–1378.

70. **Grinnell, J.** 1917. The Niche-Relationships of the California Thrasher. The Auk **34**:427–433.

71. **Gause, G. F.** 1932. Experimental studies on the struggle for existence. Journal of Experimental Biology **9**:389–402.

72. **Elton, C.** 1927. Animal Ecology. Animal Ecology 207.

73. **Hutchinson, G. E.** 1957. Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology **22**:415–427.

74. **Colwell, R. K.**, and **T. F. Rangel**. 2009. Hutchinson's duality: The once and future niche. Proceedings of the National Academy of Sciences **106**:19651–19658.

CHAPTER TWO

CLOSTRIDIUM DIFFICILE INFECTION DIFFERENTIALLY ALTERS THE STRUCTURE AND METABOLIC ACTIVITY OF DISTINCT INTESTINAL MICROBIOMES TO PROMOTE SUSTAINED COLONIZATION

Summary

Clostridium difficile has grown to be the most common cause of hospital-acquired infection over the last decade in the United States. A problematic feature of this infection is that initial susceptibility to colonization by the pathogen is closely linked to previous antibiotic therapy. This connection is due to the impact that antibiotics have on the indigenous bacterial community in the gastrointestinal tract which, in a healthy state, possesses colonization resistance to *C. difficile*. However, perturbed communities are susceptible to colonization where the pathogen can subsequently multiply and produce toxin, leading to inflammatory diarrheal disease. Furthermore, certain patients are unable to clear the pathogen and remain persistently colonized. In this study, we explored the effect of *C. difficile* colonization on community-level gene expresion and metabolism using a murine model of antibiotic treatment and infection. We characterized multiple susceptible communities utilizing metagenome-enable metatranscriptomics supplemented by untargeted metabolomic mass-spectrometry across multiple classes of antibiotic treatment to begin to understand the depth of metabolic crosstalk and effect that *C. difficile* has on infected communities of bacteria. Our results demonstrate that the metabolic activity of microbiomes that *C. difficile* is

more likely to colonize for longer periods of time are differentially impacted 18 hours post-infection. These changes in activity are reflected in the metabolic environment of the cecum and indicate a restructured nutrient-niche landscape particularly for those involving Stickland fermentation substrates and certain carbohydrates. This work underscores potential mechanisms by which the pathogen alters the ecology of the GI tract to promote persistence.

Introduction

One of the many beneficial functions provided by the indigenous gut bacterial community is its ability to prevent infection by pathogens (1). This attribute, termed colonization resistance, is one of the main mechanisms of protection from the gastrointestinal pathogen *Clostridium difficile* (2–4). *Clostridium difficile* is the etiological agent of *Clostridium difficile* infection (CDI), a toxin-mediated diarrheal disease that has dramatically increased in prevalence over the last 10 years and results in an estimated 453,000 infections with 29,000 deaths in the US annually (5). Antibiotics are a major risk factor for CDI and are thought to increase susceptibility to CDI as they disrupt the gut bacterial community structure, but it is still unclear what specific changes to the microbiota contribute to this sensitivity (6, 7). Associations between the membership and functional capacity of the microbiota as measured by the metabolic output suggest that antibiotics increase susceptibility by altering the nutrient milieu in the gut to one that favors *C. difficile* metabolism (8–10). An ongoing theory is that *C. difficile* colonization resistance is driven by competition for growth nutrients by an intact community of

metabolic specialists and has been born out through animal model experimentation of the past several decades (11–13). This line of reasoning has been carried through to the downstream restoration of colonization resistance in that, although community structure may not return to its precise original state, the functional/metabolic capacity rebounds and may be able to outcompete *C. difficile* for resources and clear the infection (10, 14). While most classes of antibiotics have been associated with initial susceptibility to CDI, fluoroquinolones, clindamycin, and cephalosporins in particular are linked to increased risk of recurrent or persistent *C. difficile* infection (15–17). This raises questions about the groups differentially impacted both directly or indirectly by certain treatments and what this ultimately means for these environments during infection by permitting long-term colonization.

Leveraging distinct antibiotic treatment regimes in a murine model of CDI (18), we and others previously shown that *C. difficile* adapts it's catabolism to distinct cecal microbiomes that resulted from separate classes of antibiotics (19). This is supported by the hypothesis that each differentially sensitized gut environment possesses an alternative nutrient niche landscape and *C. difficile* is able to adjust its metabolism accordingly. Although it is well established that *C. difficile* is able to colonize these communities effectively, it is yet to be determined whether these differences in the metabolic capacity of communities following antibiotic treatment correlate with prolonged *C. difficile* colonization. Defining the functional status of the resident microbiota for any disease has been difficult and has led to a limited understanding of specific species interactions that occur with *C. difficile* during infection. To address this

fundamental question we employed a conventionally-reared murine model of CDI in the context of treatment by distinct antibiotic classes and assessed the effect of CDI on the microbiome utilizing paired metagenomic-enabled metatransciptomics and untargeted metabolomics. This approach allowed us to not only characterize the metabolic output of the community, but also which subgroups of bacteria were differentially active under these conditions. Our data supports that *C. difficile* colonization indeed alters community-level gene expression, and that this degree of change was reflected in the metabolome of these communities. Furthermore, in spite of shifts in species abundance, the metatrascriptome and metabolome changed very little in treatment groups that cleared the infection. This work highlights that a better appreciation of the effects of CDI on the gut microbiota may be needed to develop more successful targeted therapies that eliminate *C. difficile* after persistent colonization.

Experimental Procedures

Animal care and antibiotic administration

For a more detailed description of the procedure, refer to (19). In short, an equal number of male and female conventionally-reared six-to-eight week-old C57BL/6 mice in each experimental group were administered one of three antibiotics; cefoperazone, streptomycin, or clindamycin (As described in Table 2.1) before oral *C. difficile* infection. All animal protocols were approved by the University Committee on Use and Care of Animals at the University of Michigan and carried out in accordance with the approved guidelines.

Antibiotic	Class	Target	Activity	Administration	Dosage
Cefoperazone	Cephalosporin (3rd generation)	Primarily Gram-positive bacteria, with increased activity against Gram-negative bacteria	Irreversibly crosslink bacterial transpeptidases to peptidoglycan and prevents cell wall synthesis	Drinking water Ad libitum for 5 days, 2 days untreated drinking water prior to infection	0.5 mg/ml drinking water
Streptomycin	Aminoglycoside	Active against most Gram-negative aerobic and facultative anaerobic bacilli	Protein synthesis inhibitor through binding the 30S portion of the 70S ribosomal subunit	Drinking water Ad libitum for 5 days, 2 days untreated drinking water prior to infection	5.0 mg/ml drinking water
Clindamycin	Lincosamide	Primarily active against Gram-positive bacteria, most anaerobic bacteria, and some mycoplasma	Protein synthesis inhibition through binding to the 23s portion of the 50S ribosomal subunit	Intraperitoneal injection 24 hours prior to infection	10 mg/kg body weight

Table 2.1 | Antibiotics used during *C. difficile* infection models.

C. difficile infection and necropsy

For a more detailed description of the procedure, refer to (19). On the day of challenge, ~1×10^3 *C. difficile* strain 630 spores were administered to mice via oral gavage in phosphate-buffered saline (PBS) vehicle. Mock-infected animals were given an oral gavage of 100 µl PBS at the same time as those mice administered *C. difficile* spores. 18 hours following infection, mice were euthanized by CO_2 asphyxiation and necropsied to obtain the cecal contents. Aliquots were immediately flash frozen for later DNA extraction and toxin titer analysis. A third aliquot was transferred to an anaerobic chamber for quantification of *C. difficile* abundance. The remaining content in the ceca was mixed in a stainless steel mortar housed in a dry ice and ethanol bath. Cecal contents from all mice within each treatment group were pooled into the mortar prior to grinding to a fine powder. The ground content was then stored at -80° C for subsequent RNA extraction. For 10-day colonization studies, fresh stool was collected from infected mice each day beginning on the day of infection. Mice were monitored for overt signs of disease and were euthanized after the final stool collection.

C. difficile cultivation and quantification

For a more detailed description of the procedure, refer to (19). Briefly, cecal samples were weighed and serially diluted under anaerobic conditions with anaerobic PBS. Differential plating was performed to quantify both *C. difficile* spores and vegetative cells by plating diluted samples on CCFAE plates (fructose agar plus cycloserine, cefoxitin, and erythromycin) at 37° C for 24 hours under anaerobic

conditions (42). In parallel, undiluted samples were heated at 60° C for 30 minutes to eliminate vegetative cells and leave only spores (29). These samples were serially diluted under anaerobic conditions in anaerobic PBS and plated on CCFAE with taurocholate at 37° C for 24 hours. Plating was simultaneously done for heated samples on CCFAE to ensure all vegetative cells had been eliminated. CFU quantification for 10-day colonization experiments was performed from stoll using TCCFAE to measure total *C. difficile* load in these animals over time.

C. difficile toxin titer assay

To quantify the titer of toxin in the cecum a Vero cell rounding assay was performed as described elsewhere (19, 43). Briefly, filtered-sterilized cecal content was serially diluted in PBS and added to Vero cells in a 96-well plate. Plates were viewed after 24 hour incubation for cell rounding.

DNA/RNA extraction and sequencing library preparation

DNA for downstream shotgun metagenomic and 16S rRNA gene sequencing was extracted from approximately 50 mg of cecal content from each mouse using the PowerSoil-htp 96 Well Soil DNA isolation kit (MO BIO Laboratories) and an epMotion 5075 automated pipetting system (Eppendorf). The V4 region of the bacterial 16S rRNA gene was amplified using custom barcoded primers (44). Equal molar rations of raw isolated DNA within each treatment group were then pooled and ~2.5 ng of material was used to generate shotgun libraries with a modified 10-cycle Nextera XT genomic library

construction protocol (Illumina). This was done to mimic the pooling strategy necessary for metatranscriptomic library preparation. Final libraries were pooled at equal molar ratios and stored at -20° C. For RNA extraction, a more detailed description of the procedure can be found in (19). Briefly, immediately before RNA extraction, 3 ml of lysis buffer (2% SDS, 16 mM EDTA and 200 mM NaCl) contained in a 50 ml polypropylene conical tube was heated for 5 minutes in a boiling water bath (45). The hot lysis buffer was added to the frozen and ground cecal content. The mixture was boiled with periodic vortexing for another 5 minutes. After boiling, an equal volume of 37° C acid phenol/chloroform was added to the cecal content lysate and incubated at 37° C for 10 minutes with periodic vortexing. The mixture was the centrifuged at 2,500 x g at 4° C for 15 minutes. The aqueous phase was then transferred to a sterile tube and an equal volume of acid phenol/chloroform was added. This mixture was vortexed and centrifuged at 2,500 x g at 4° for 5 minutes. The process was repeated until aqueous phase was clear. The last extraction was performed with chloroform/isoamyl alcohol to remove acid phenol. An equal volume of isopropanol was added and the extracted nucleic acid was incubated overnight at -20° C. The following day the sample was centrifuged at 12000 x g at 4° C for 45 minutes. The pellet was washed with 0° C 100% ethanol and resuspended in 200 µl of RNase-free water. Following the manufacturer's protocol, samples were then treated with 2 µl of Turbo DNase for 30 minutes at 37° C. RNA samples were retrieved using the Zymo Quick-RNA MiniPrep according the manufacturer's protocol. The Ribo-Zero Gold, immediately before RNA extraction, 3 ml of lysis buffer (2% SDS, 16 mM EDTA and 200 mM NaCl) contained in a 50 ml

polypropylene conical tube was heated for 5 minRNA Removal Kit Epidemiology was then used to deplete prokaryotic and eukaryotic rRNA from the samples according the manufacturer's protocol. Stranded RNA-Seq libraries were made constructed with the TruSeq Total RNA Library Preparation Kit v2, both using the manufacturer's protocol. Average length of amplicon libraries for both DNA and cDNA sequencing was performed using an Agilent BioAnalyzer with High Sensativity DNA Analysis kits. Completed libraries were pooled in equal molar ratios within their respective groups and stored at -20° C until time of sequencing.

High-throughput sequencing and raw read curation. Sequencing of 16S rRNA gene amplicon libraries was performed using an Illumina MiSeq sequencer as described previously (44). The 16S rRNA gene sequences were curated using the mothur software package (v1.36) as described in (19). Shotgun metagenomic sequencing was performed in 2 phases. Libraries from mock-infected communities, that were also to be utilized for contig assembly, were sequenced using an Illumina HiSeq 2500 on 2x250 paired-end settings and was repeated across 2 lanes to normalize for inter-run variation. *C. difficile*-infected metagenomic libraries were sequenced with an Illumina NextSeq 300 with 2x150 settings across 2 runs to also normalize for inter-run variation. These efforts resulted in an average of 280 million paired raw reads per sample. Metatranscriptomic sequencing was performed on an Illumina HiSeq 2500 with 2x50 settings and was repeated across 4 lanes for normalization and to obtain necessary coverage (46). This gave an average of 380 million raw cDNA per library. Both metagenomic and metatranscriptomic sequencing was performed at the University of

Michigan Sequencing Core. Raw sequencing read curation for both metagenomic and metatranscriptomic datasets was performed in a two step process. Residual 5' and 3' Illumina adapter sequences were trimmed using CutAdapt (47) on a per library basis. Reads were quality trimmed using Sickle (48) with a quality cutoff of Q30. This resulted in approximately 270 million reads per library (both paired and orphaned) for both metagenomic and metatranscriptomic sequencing. Actual read abundances for individual metagenomic and metatranscriptomic sequencing efforts can be found in Table 2.2.

Metatranscriptome	**Raw reads pairs[1]**	**Quality trimmed (% remaining)[2]**	
Streptomycin – *C. difficile* infected	307082142	253185782 (82.45%)	
Streptomycin – mock infected	270291248	226737019 (83.89%)	
Cefoperazone – *C. difficile* infected	376175754	322153373 (85.64%)	
Cefoperazone – mock infected	342590768	289481340 (84.5%)	
Clindamycin – *C. difficile* infected	265973934	223661860 (84.09%)	
Clindamycin – mock infected	261472804	216933273 (82.97%)	
No Antibiotics – mock infected	397936700	329411775 (82.78%)	
Superscipts:			
1=Total individual reads directly from sequencer – all paired			
2= Read totals following adapter, quality trimming, and residual in silico rRNA removal - paired and orphaned			
3=Reads totals that mapped to metagenomic assembly after optical and PCR duplicate removal			
Metagenome	**Raw reads pairs[1]**	**Quality trimmed (% remaining)[2]**	
Streptomycin – *C. difficile* infected[5]	407440672	378758657 (92.96%)	
Streptomycin – mock infected[4]	395429292	394499193 (99.77%)	
Cefoperazone – *C. difficile* infected[5]	392694214	364820685 (92.9%)	
Cefoperazone – mock infected[4]	178456950	178133857 (99.82%)	
Clindamycin – *C. difficile* infected[5]	348426748	317183590 (91.03%)	
Clindamycin – mock infected[4]	118787094	118502295 (99.76%)	
No Antibiotics – mock infected[4]	116784668	116560872 (99.81%)	
Superscipts:			
1=Total individual reads directly from sequencer – all paired			
2= Read totals following adapter, quality trimming, and residual in silico rRNA removal - paired and orphaned			
3=Reads totals that mapped to metagenomic assembly after optical and PCR duplicate removal			
4= Sequenced with HiSeq 2500 – 2 x 250			
5=Sequenced with NextSeq – 2 x 150			
Assembly	**Total contigs**	**N50**	**Sequences > 1 kb**
Streptomycin – mock infected	1317910	255	2479
Cefoperazone – mock infected	2539357	598	158781
Clindamycin – mock infected	811592	400	3913
No Antibiotics – mock infected	496917	1434	83535

Table 2.2 | High-throughput sequencing read counts and metagenomic assembly quality. Raw and curated read abundances for both metagenomic and metatranscriptomic sequencing efforts. Raw read curation steps are outlined in Materials & Methods. Metagenomic contig summary statistics reflect the quality of assembly for each group.

Metagenomic contig assembly and gene annotation

Metagenomic contigs were assembled using Megahit (49) with the following settings; minimum kmer size of 87, maximum kmer size of 127, and a kmer step size of 10. Prodigal was utilized to to identify putative gene sequences, and were screened for a minimum length of 250 nucleotides. These sequences were translated to amino acids and peptides were annotated against the KEGG protein database (50) using Diamond implementation of BLASTp (51). Peptide-level gene annotations were assigned to the corresponding nucleotide sequence, and genes failing to find a match in KEGG were preserved as unannotated genes. Final nucleotide fasta files with KEGG annotations were then utilized in the construction of Bowtie2 (52) mapping databases from downstream analyses.

DNA/cDNA read mapping and normalization

Mapping was accomplished using Bowtie2 (52) and the default stringent settings. Optical and PCR duplicates were then removed using Picard MarkDuplicates (http://roadinstitute.github.io/picard/). The remaining mappings were converted to idxstats format using Samtools (53) and the read counts per gene were tabulated. Discordant pair mappings were discarded and counts were then normalized to read length and gene length to give a per base report of gene coverage. Unless indicated otherwise, each collection of reads was then 1000-fold iteratively subsampled to 90% of the lowest sequence total within each analysis, and a median expression value for each gene was calculated.

Quantification of *in vivo* metabolite relative concentrations

For a more detailed description of the procedure, refer to (19). Metabolomic analysis performed by Metabolon (Durham, NC), a brief description of their methods is as follows. All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer at 35,000 mass resolution. Samples were dried then reconstituted in solvents compatible to each of the four methods. The first, in acidic positive conditions using a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). The second method was identical to the first but was chromatographically optimized for more hydrophobic compounds. The third approach utilized a basic negative ion optimized conditions using a separate dedicated C18 column. Basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. Samples were then analyzed via negative ionization following elution from a hydrophilic interaction chromatography column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS n scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Library matches for each compound were checked for each sample and corrected if necessary. Peaks were quantified using area under the curve.

Statistical methods

All statistical analyses were performed using R (v.3.2.0). Significant differences between community structure of treatment groups from 16S rRNA gene sequencing were determined with AMOVA in the mothur software package. Significant differences of Inv. Simpson diversity, cfu, toxin titer, and metabolite concentrations were determined by Wilcoxon signed-rank test with Benjamini-Hochberg correction. Undetectable points used half the limit of detection for cfu and toxin statistical calculations. LEfSe analysis with OTU data was performed with a cutoff that each OTU must appear in all samples from their respective groups to be considered true signal (n = 9). Random forest was performed using the implementation in R (54), with the informative threshold of MDAs greater than the absolute value of the lowest MDA defined by (55). Distances of outlier points from center line during metatranscriptomic comparisons was accomplished using 2-dimensional linear geometry.

Results

Distinct antibiotic treatments are associated with different patterns of clearance following primary infection. Conventionally-reared specific pathogen free mice were treated with one of three different antibiotics to sensitize the animals to *C. difficile* colonization. The selected antibiotics were streptomycin, cefoperazone, and clindamycin (Table 2.1 & Fig. 2.1). Each drug was chosen not only due to its ability to reduce *C. difficile* colonization resistance in a mouse model (18), but also for their distinct and significant impacts on the structure and diversity of the cecal microbiota (all

$p < 0.001$; Fig. 2.4A) as well as differences in ability of the community to recover and clear infection (Fig. 2.2B). Selection of this toxigenic *C. difficile* strain was based on its moderate clinical severity in mouse models (20), previous studies of *in vitro* metabolism (21), and well-annotated genome (22). Briefly, mice were treated with the respective antibiotic and were subsequently orally gavaged ~1×10^3 *C. difficile* str. 630 spores (Fig. 2.2A). We then monitored for disease over the following 10 days and cultured *C. difficile* from stool to quantify colony forming units (cfu) per gram over time. The day after infection in each antibiotic treatment model, we observed equal high *C. difficile* colonization, however over the following 8 days only clindamycin treated mice cleared the infection (Fig. 2.2B).

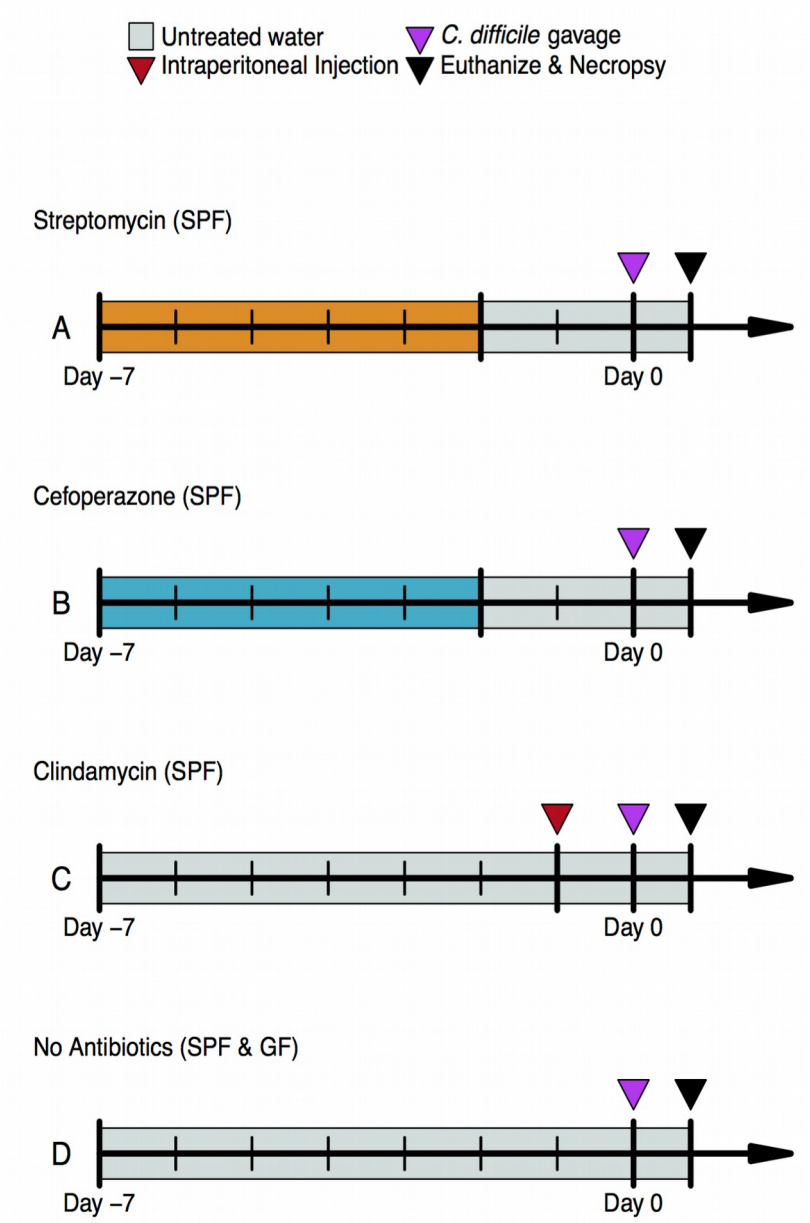

Figure 2.1 | Experimental timelines for mouse model pretreatments and *C. difficile* infection. 9 wild-type C57BL/6 mice across 3 cages were included in each treatment group. **(A)** Streptomycin or **(B)** cefoperazone administered *ad libitum* in drinking water for 5 days with 2 days recovery with untreated drinking water before infection, **(C)** a single clindamycin intraperitoneal injection one day prior to infection, or **(D)** no antibiotic pretreatment (for both SPF control and GF mice). If no antibiotics were administered in the drinking water, mice were given untreated drinking water for the duration of the experiment beginning 7 days prior to infection. At the time of infection, mice were challenged with 1×10^3 *C. difficile* str. 630 spores. Euthanization and necropsy was done 18 hours post-challenge and cecal content was then collected.

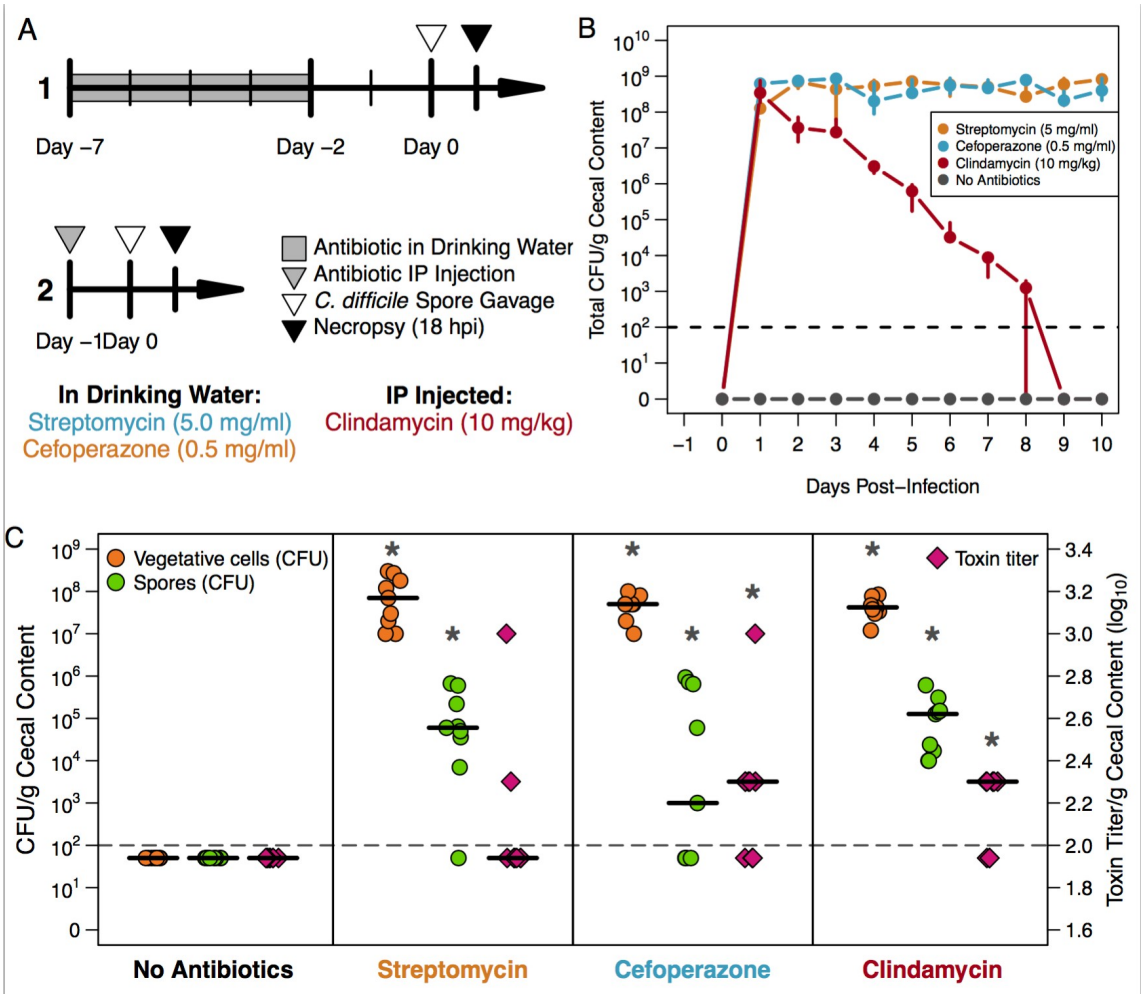

Figure 2.2 | Experimental models of *C. difficile* infection and distinct virulence patterns. (A) Experimental timelines of antibiotic treatment and infection mouse models. **(B)** *C. difficile* 630 CFU in stool of infected mice following each antibiotic treatment regime over 10 days of infection. Median and interquartile range are shown for each time point. The dotted line represents the limit of detection. **(C)** Quantification of *C. difficile* cfu and toxin titer in cecal content across antibiotic treatment models alter 18 hours of infection. Black lines indicate median values. Gray asterisks indicate significant difference from no antibiotic controls by Wilcoxon rank-sum test with the Benjamini-Hochberg correction (all $p < 0.001$).

C. difficile differentially expresses virulence factors across separate antibiotic treatments. With the differential clearance results between antibiotic treatments, we hypothesized that each community presented separate metabolic challenges to *C. difficile* which could explain the clearance trends. It had been previously demonstrated that *C. difficile* virulence factor expression is regulated by availability of certain nutrients in the environment (21), so we first sought to measure sporulation and toxin production. Continuing with the same pretreatment and primary infection protocols used in the 10-day colonization experiments, we chose to focus our analysis on 18-hours post-infection to assess behavior of *C. difficile* directly prior to the beginning of clearance. This end point corresponded with a previous study where *C. difficile* reached maximum cecal vegetative cell density with few detectable spores (23). Moreover, we also elected to take all further measurements from cecal content because it is more likely to be a site of active bacterial metabolism compared to stool. This also allowed for assessment of functional differences in the microbiota were apparent between antibiotic treatments early during infection and would correlate with the downstream clearance phenotypes. At 18 hours after infection, there were no significant differences in the number of vegetative cells between any antibiotic-treatment tested. All susceptible mice were colonized to ~1×10^8 vegetative cfu per gram of cecal content, while untreated mice maintained *C. difficile* colonization resistance (Fig. 2.2C). We also measured both sporulation and toxin activity as activation of both processes has been linked to environmental concentrations of specific growth nutrients (21). Despite having similar amounts of vegetative *C. difficile* cells, varying levels of both spore cfu and toxin titer

were observed across each of the antibiotic treatments. These results showed that *C. difficile* colonized different antibiotic-treated mice to consistently high levels, and the distinct treatments corresponded with moderate differences in the expression of *C. difficile* virulence factors.

It has not yet been established whether *C. difficile* colonization impacts the structure of the gut microbiota during infection in susceptible mice. In order to evaluate changes in bacterial population abundances in response to perturbation and *C. difficile* colonization, we sequenced the V4 region of the 16S rRNA gene from the cecal content of both mock and *C. difficile*-infected mice across antibiotic treatment models. To focus out analysis specifically on the surrounding bacterial communities, we ignored all *C. difficile* rRNA gene sequences. We confirmed that each antibiotic treatment significantly impacted both cecal community structure (Bray-Curtis distances) and diversity (inverse-Simpson) compared to untreated control mice (all $p < 0.001$, Fig. 2.4A & 2.3B). We then confirmed that each antibiotic treatment induced distinct shifts in the community structure which were significantly different (all $p < 0.001$, Fig. 2.4C). The composition of streptomycin-treated communities was more variable between cages, but was generally enriched for members of phylum Bacteroidetes (Fig. 2.3A). Cefoperazone and clindamycin-treated cecal communities were consistently dominated by the families Lactobacillaceae and Enterobacteriaceae respectively (Fig. 2.3A). Despite variation in the community structures generated by streptomycin treatment, those communities were colonized evenly (Fig. 2.2B, 2.2C, & 2.3). Initial *C. difficile* colonization levels were consistent in spite of significantly different community structures.

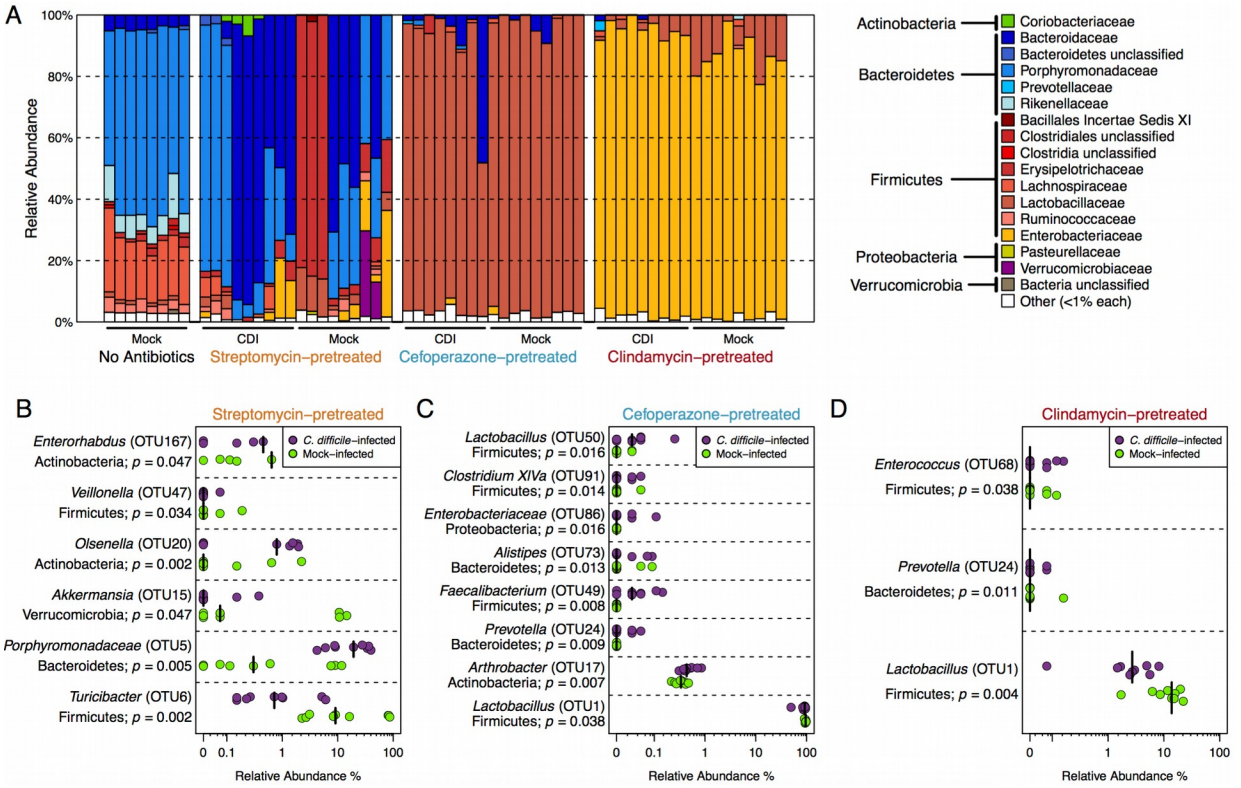

Figure 2.3 | Impact of *C. difficile* colonization on other bacterial populations abundances in the gut microbiota. (A) Relative abundance of family-level taxonomic classification for OTUs in each treatment group. **(B-D)** Discriminating OTUs with LEfSe analysis between Mock and *C. difficile*-infected communities within each treatment group, p-values from LEfSe are shown on the left with taxonomic information (all $p < 0.05$). Relative abundance of the respective OTUs from each mouse along with medians are shown for each treatment group.

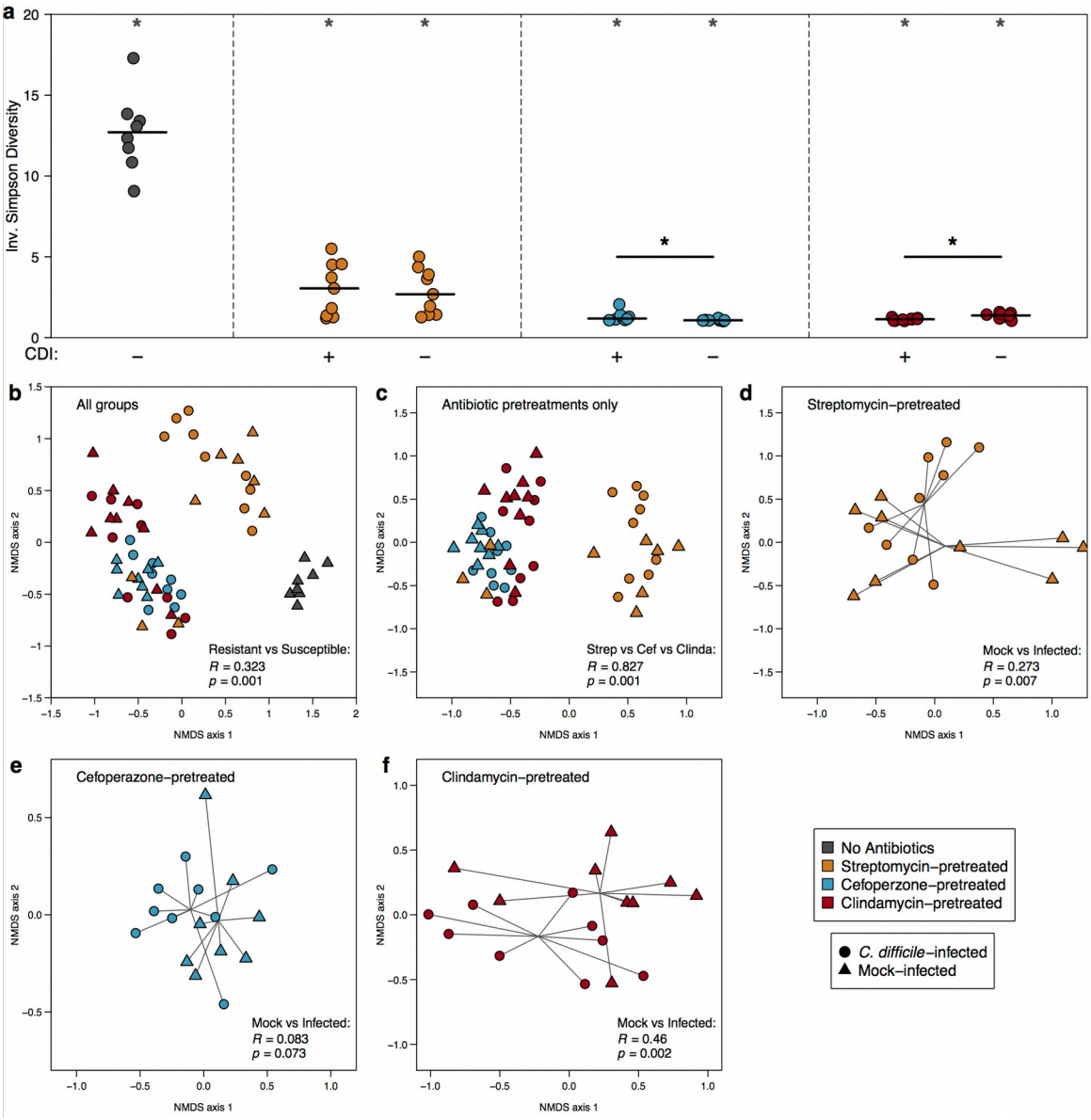

Figure 2.4 | Impact of antibiotic treatment and *C. difficile* infection on cecal community structure. (A) Inverse-Simpson diversity of cecal communities from all treatment groups. Gray stars indicate significant difference from no antibiotic controls (all $p < 0.001$). Black stars denote within treatment group significant difference between mock and *C. difficile*-infected communities. Differences were calculated using Wilcoxon rank-sum test with Benjamini-Hochberg correction. NMDS ordinations of Bray-Curtis distances comparing the groups labeled in the bottom left of each plotting area. **(B)** Antibiotic-treated compared to Untreated controls and **(C)** comparison between only antibiotic treatment groups. **(D-F)** Within antibiotic treatment comparisons for the effect of *C. difficile* colonization on community structure. Significant differences and correlation coefficients for ordination analyses were found using ANOVA.

Next, we measured the effect that *C. difficile* colonization had on overall community structure and composition for each of the antibiotic groups. We found that the structure of both streptomycin and clindamycin-treated infected communities were significantly different from their corresponding uninfected controls (streptomycin p = 0.014, Fig. 2.4D; clindamycin p = 0.003, Fig. 2.4F). Cefoperazone-treated communities did not significantly differ from their control group (Fig. 2.4E). We then sought to identify specific OTUs that significantly differed in abundance between mock and *C. difficile*-infected communities within each treatment group. Utilizing LEfSe differential abundant feature detection (24), we identified 16 OTUs that discriminated between infected and uninfected communities in at least one antibiotic treatment (Fig. 2.3B-2D); however, these OTUs were generally near the limit of detection and had an inconsistent abundance profile across pretreatment groups.

Distinct antibiotic classes lead to alternative markers of *C. difficile* colonization susceptibility. Several groups have demonstrated that treatment with antibiotics not only alters the structure of the resident microbiota, but also has a dramatic impact of the intestinal metabolome (8–10). To test the metabolic features of the altered communities, we performed untargeted metabolomic analysis on separate aliquots of the cecal contents that were also utilized in the 16S rRNA gene sequencing. A total of 727 distinct metabolites were identified through a combination of several liquid chromatography and mass spectrometry techniques. First, we characterized the differences between the metabolomes of the mock-infected communities to measure the impact of antibiotic treatment that generated *C. difficile*-susceptible conditions (Fig. 2.5). All of the antibiotic

treatments significantly altered the cecal metabolome compared to untreated, *C. difficile*-resistant mice as quantified by Bray-Curtis dissimilarity ($p < 0.001$; Fig. 2.5a). When metabolites were mapped to KEGG pathways, it was clear that the differences between resistant and susceptible metabolomes were the result of widespread physiological effects (Fig. 2.6). Similar to the differences between resistant and susceptible states, the patterns of specific metabolite concentrations diverged depending on the antibiotic treatment (Fig. 2.6). These results demonstrated that each antibiotic treatment lead to distinct susceptible metabolomic structures.

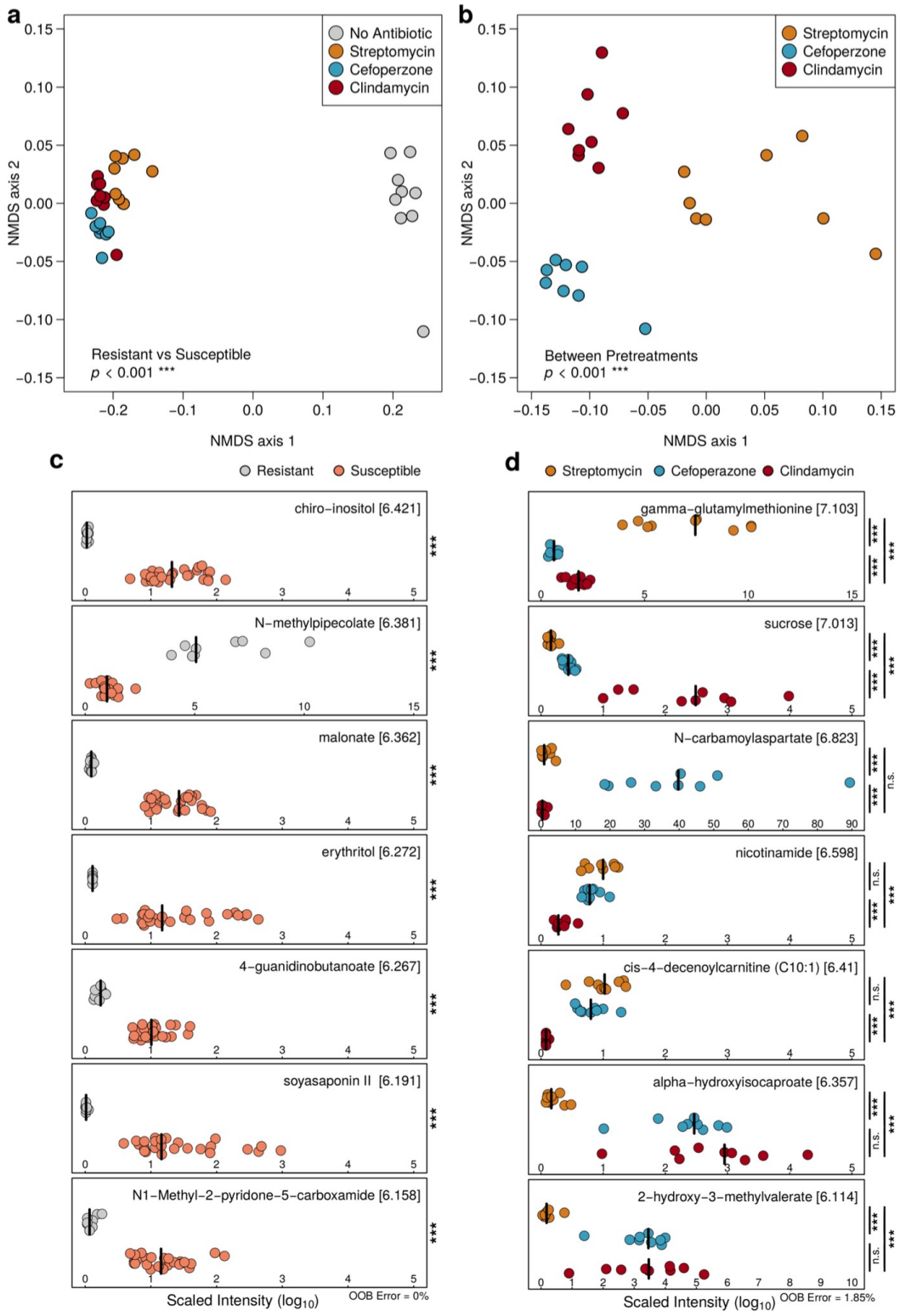

Figure 2.5 | Metabolite markers of *C. difficile* infection susceptibility.

Figure 2.5 | Metabolite markers of *C. difficile* infection susceptibility. Results from UPLC/MS metabolomic analysis of cecal content from mice also used in previous analyses. Only mock-infected metabolomic results were included this analysis to identify markers of susceptibility. **(A-B)** NMDS ordinations of Bray-Curtis distances comparing the cecal metabolome dissimilarity of mice receiving no treatment or one of the three distinct classes of antibiotics. (A) Metabolomes of resistant mice are significantly different from antibiotic treated animals ($p < 0.001$). (B) Antibiotic treated metabolomes and also significantly distinct from one another ($p < 0.001$). Significant differences for NMDS analyses were calculated with AMOVA. **(C-D)** Scaled intensities of metabolites with highest mean decrease in accuracy (MDA) from random forest feature selection discriminating groups from (A-B). MDA is labeled in brackets beside each metabolite name and out-of-bag error from internal cross-validation is labeling along the bottom axis. Asterisks along the right axis indicate significant difference by Wilcoxon rank-sum test. (C) Relative concentrations of metabolites that distinguish cecal content from mice resistant to *C. difficile* colonization from susceptible animals. (D) Metabolites concentrations for those molecules that are able to differentiate antibiotic treatment groups. Multiple comparisons were accounted for using the Benjamini-Hochberg correction (*** < 0.001, ** <= 0.01, * <= 0.05).

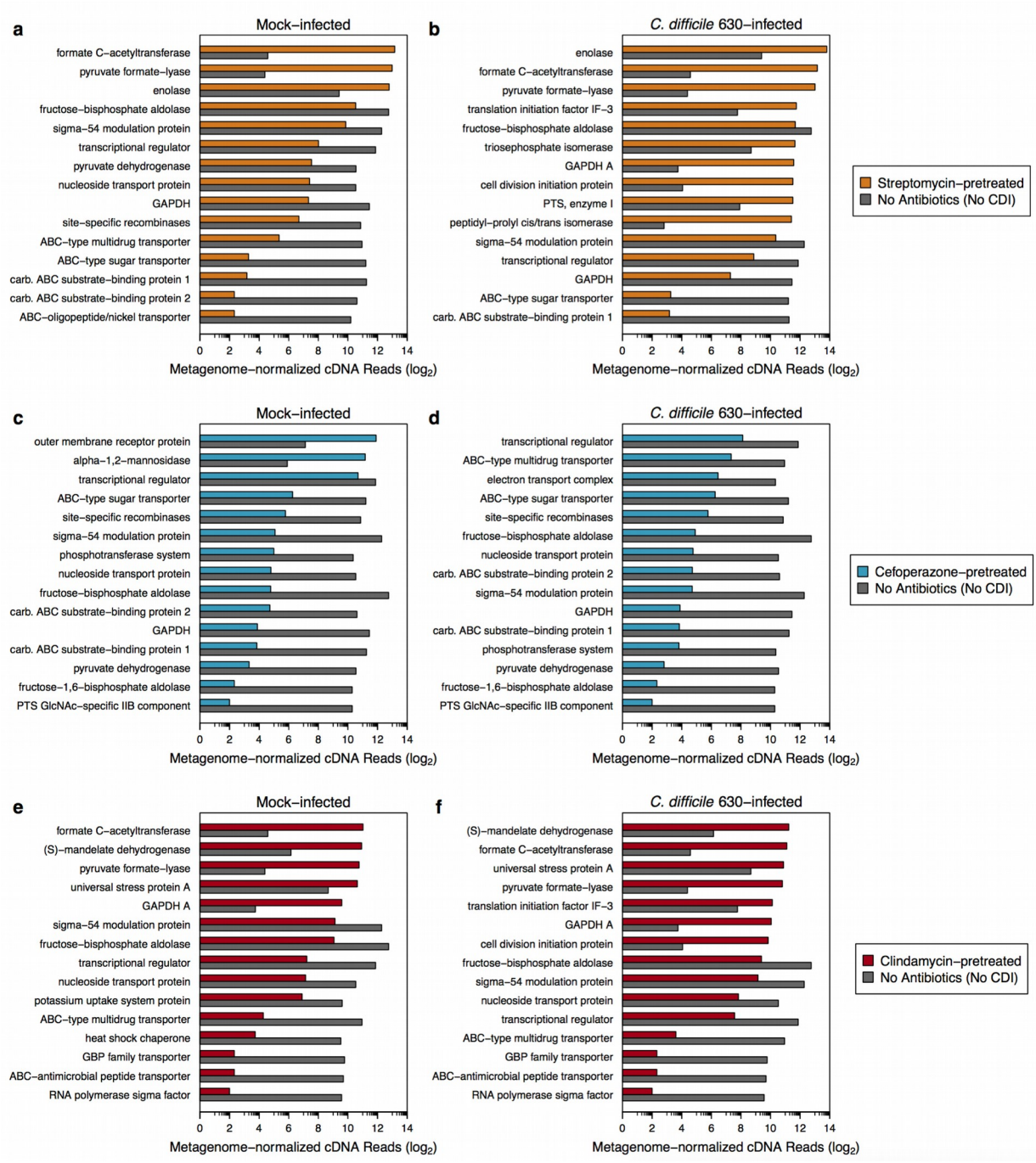

Figure 2.6 | Gene-level expression for each community compared to the level expressed in resistant mice for each gene.

Figure 2.6 | Gene-level expression for each community compared to the level expressed in resistant mice for each gene. Murine cecal metatransctipromic sequencing results. Both reads and genes attributed to *C. difficile* were removed prior to analysis to focus on the changes in the community caused by infection. Shown are genes in each groups with the largest disparity from a metatranscriptome from a community resistant to *C. difficile* colonization. Panels are displayed as follows; Streptomycin treatment: **(A)** mock infection, **(B)** *C. difficile* infection. Cefoperazone treatment: **(C)** mock infection, **(D)** *C. difficile* infection. Clindamycin treatment: **(E)** mock infection, **(F)** *C. difficile* infection.

We were then interested in identifying those metabolites with the highest degree of change between resistant and susceptible groups groups. Due to the large number of individual metabolites with significant differences in relative concentration between groups, we employed a machine learning approach using random forest (25) to highlight those metabolites that most easily differentiated the antibiotic groups. In order to further limit the analysis to only the most informative metabolite, we ranked metabolites in order of highest Mean Decrease Accuracy (the amount in which their removal negatively impacts correct sample classification) and reported the top 7 in each analysis (Fig. 2.5C). These lists primarily included precursors or intermediates of carbohydrate fermentation such as chiro−inositol, malonate, erythritol, 4−guanidinobutanoate, and soyasaponin II, the majority of which were increased in susceptible conditions. Decreases in these metabolites *in vivo* have previously been associated with a reduction in the normal levels of polysaccharide fermentation present in the intact microbiota (9). Furthermore, N−methylpipecolate was the only metabolite we identified that consistently decrease during antibiotic treatment. This molecule is byproduct of amino acid catabolism and is typically recycled under normal conditions (26). These data support that amino acid catabolism is disrupted at a community-level following antibiotic treatment and may suggest open nutrient niches for *C. difficile*, a known fermenter of peptides (27), to colonize. We then applied this technique to identify metabolites that distinguish antibiotic treatment groups from each other(Fig. 2.5d). Interestingly, several additional carbohydrate or amino acid catabolism byproducts were highlighted by these means including hydroxyisocaproate, methylvalerate,

glutamylmethionine, and N−carbamoylaspartate (9). These data suggested that the populations responsible for normal chains of fermentation may be differentially effected by distinct classes of antibiotics. It also supported that the distinct antibiotics allowed for different forms of nutrient catabolism to survive the treatment and potentially implied alternative profiles of metabolic competition for *C. difficile* to cope with upon colonization. Additionally, sucrose is a *C. difficile* growth substrate (28) which further promoted the hypothesis of differentially vacated nutrient niches due to antibiotic treatment that are now accessible to *C. difficile*. After following this unsupervised approach for identification of susceptibility markers, we also measured differences in relative concentration of metabolites previously connected to potential *C. difficile* colonization susceptibility (10). In this way, we assessed differences in bile acids, deeply connected to the life cycle of *C. difficile* (29) and whose bioconversion by the microbiota has been implicated as a driver of colonization resistance (30). We found that there was no persistently increased bile acid across the chosen antibiotic treatments (Fig. 2.7A). A similar trend was also seen in amino acids (Fig. 2.7B). This suggested that despite varying efficiencies, the fact that *C. difficile* can recognize a subset of these molecules in any of the observed contexts appears to be sufficient to allow for sufficient germination and outgrowth to occur. For carbohydrates (Fig. 2.7B), we found that several were significantly increased across all antibiotic groups which included arabitol/xylitol, ribitol, and sucrose. Together, our results supported that each susceptible environment was distinguishable from other groups with its own subset of enriched *C. difficile* growth substrates. This could be an indication that particular

competitors were eliminated during antibiotic treatment, or those community members normally responsible for the consumption of these metabolites have altered their metabolic program to exploit alternative nutrient sources.

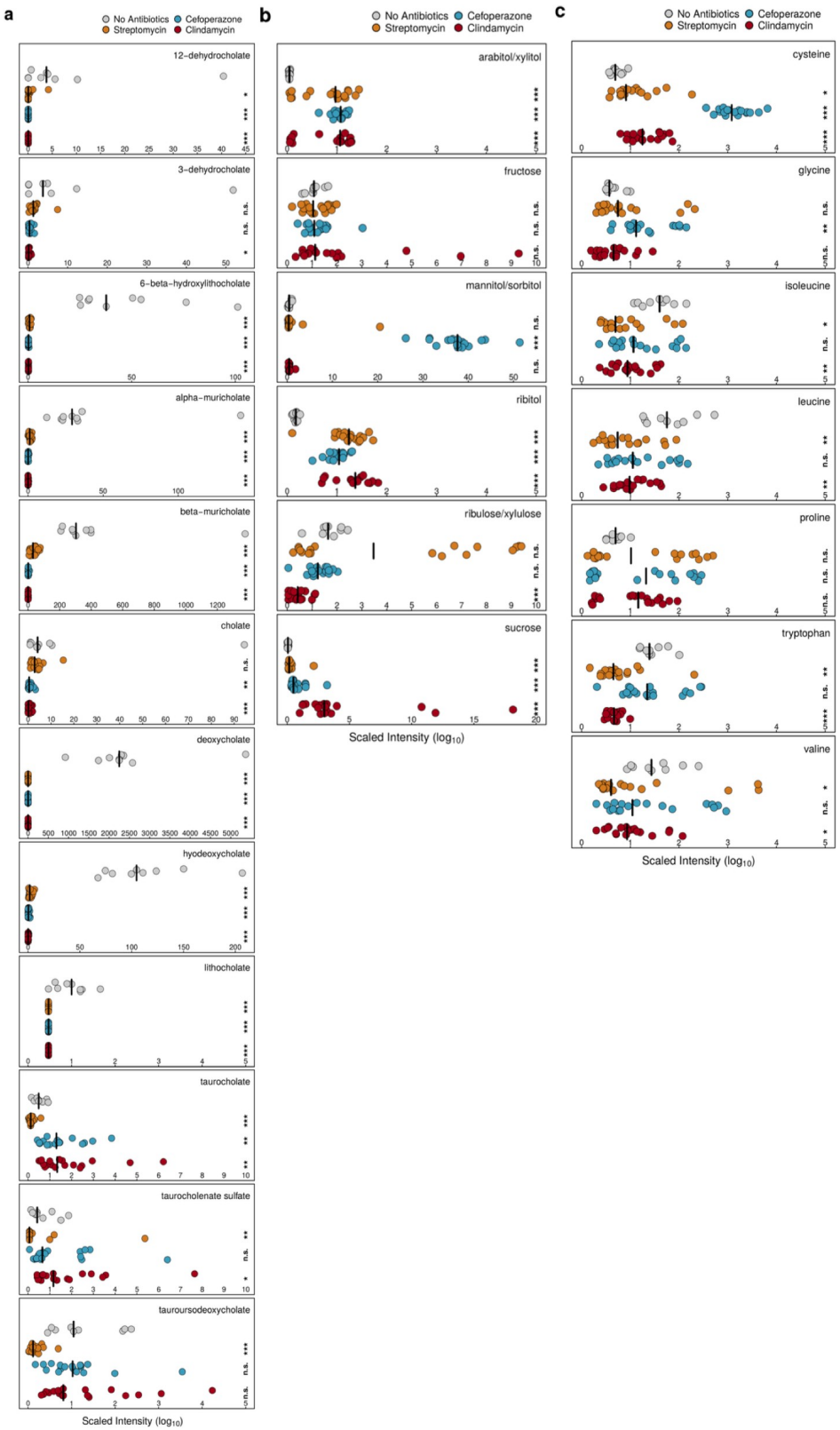

Figure 2.7 | Relative concentrations of select metabolite groups in each group.

Figure 2.7 | Relative concentrations of select metabolite groups in each group. Metabolites included in this analysis were chosen based on their previously published links to *C. difficile* physiology or susceptibility to infection. Groups are as follows; **(A)** Bile acids, **(B)** Carbohydrates, and **(C)** Amino Acids. Significant differences were determined by Wilcoxon rank-sum test with Benjamini-Hochberg correction.

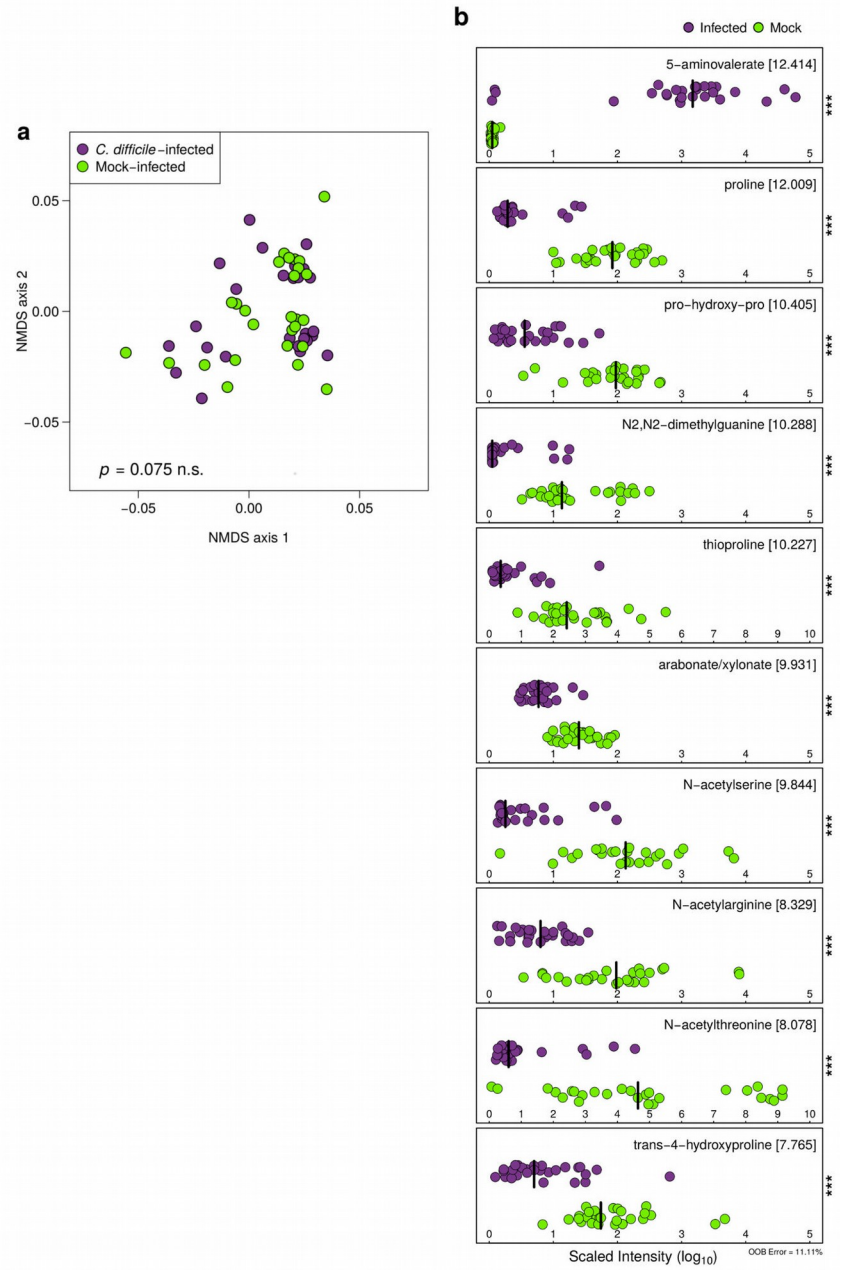

Figure 2.8 | Effect of infection on the cecal metabolome across treatment groups. Pooled analysis of antibiotic treated animals only. **(A)** NMDS ordination of Bray-Curtis distances differentiating mock and *C. difficile*-infected metabolomes (p = 0.075). **(B)** Random forest classification results for metabolites that effectively distinguish infected and uninfected conditions. Shown are relative concentrations of the top 10 metabolites with the highest mean decrease in accuracy from internal cross-validation. Shown in the top right corner of each panel are the metabolite names and mean decrease accuracy of each. Significant differences in concentration between mock and *C. difficile*-infected groups were determined by Wilcoxon rank-sum test with Benjamini-Hochberg correction.

***C. difficile* colonization alters each susceptible cecal metabolome distinctly.**

Following the changes to the cecal metabolome in response to antibiotic treatment, we assessed the degree to which *C. difficile* colonization altered the cecal metabolome of susceptible animals. We hypothesized that the introduction of a new competitor, *C. difficile*, would impact the metabolome either through signatures of its own metabolism or by causing a shift in the metabolism of other members in the surrounding community. First, in a similar approach to identifying susceptibility markers, we observed the Bray-Curtis dissimilarities of mock-infected and *C. difficile*-infected metabolomes within each antibiotic treatment group separately. First we performed this analysis at a global level and compared cecal metabolomes from all mice across treatment groups, and were unable to detect a consistent difference between groups in this way (p = 0.075; Fig. 2.8A). We moved on to individual antibiotic groups and found that both streptomycin (p = 0.039) and cefoperazone (p = 0.016) treated metabolomes deviated significantly from that of mock infection (Fig. 2.9A - Fig. 2.9B). However, clindamycin treated cecal metabolomes were not significantly altered by the presence of the pathogen (p = 0.127; Fig. 2.9C). These results diverged from what was seen in the paired OTU relative abundance results where instead the community structure of cefoperazone was unchanged and clindamycin was significantly different (p = 0.003; Fig. 2.6F, & Fig. 2.10C). Interestingly, streptomycin-treated microbiomes were significantly altered by *C. difficile* infection at both the OTU and metabolomic levels (Fig. 2.9A). These data indicated that large shifts of populations in the cecal microbiota was not implicitly associated with concordant shifts in the metabolome. This supported the

hypothesis that divergent community structures can ultimately share a convergent metabolic output despite changes to community structure and membership.

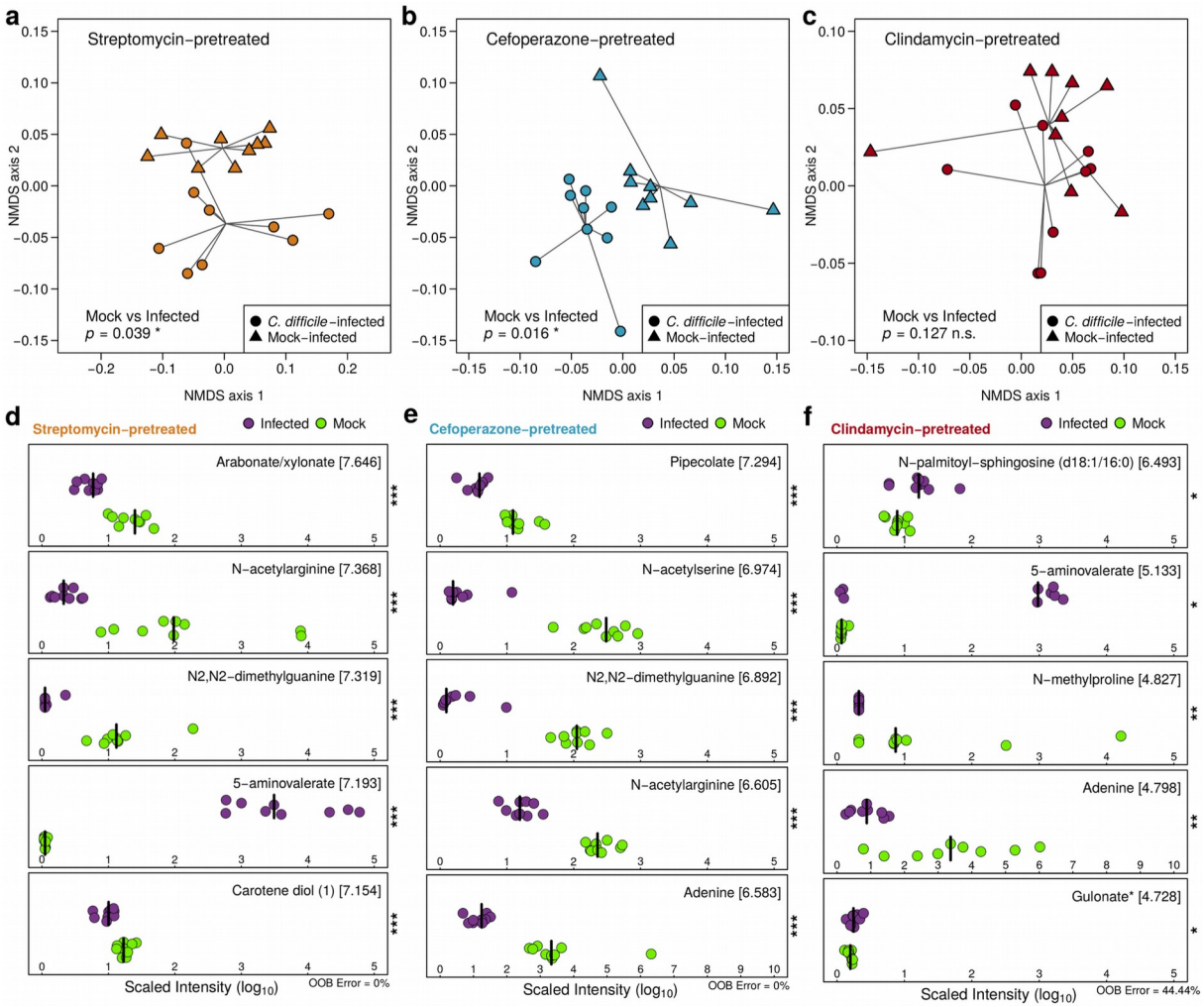

Figure 2.9 | *C. difficile* infection has differential effects on the cecal metabolome. (a-c) NMDS ordinations of Bray-Curtis distances comparing infection groups within each antibiotic treatment; **(A)** Streptomycin, **(B)** Cefoperazone, and **(C)** Clindamycin. Significant differences were determined by AMOVA. **(D-F)** Relative concentrations of metabolites with highest Mean Decrease Accuracy (MDA) from Random Forest feature selection to discriminate between Mock and *C. difficile*-infected conditions within each antibiotic treatment. Respective MDA is labeled in brackets next to each metabolite name and out of bag error from internal cross-validation of each model is labeled under the plotting area. (d) Streptomycin treatment, (e) Cefoperazone treatment, and (f) Clindamycin treatment. Significant differences are labeled along the right axis and were calculated with Wilcoxon rank-sum test with Benjamini-Hochberg correction (*** < 0.001, ** <= 0.01, * <= 0.05).

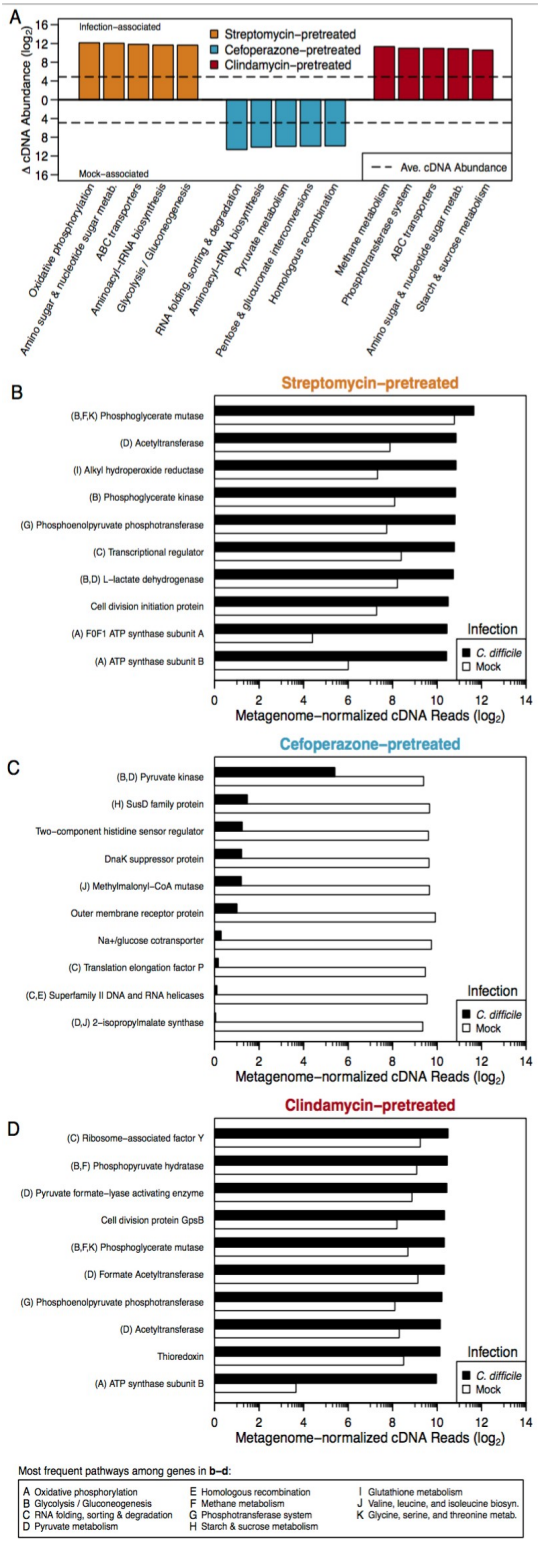

Figure 2.10 | *C. difficile* infection alters community-level select pathway and gene expression of the gut microbiota across perturbed communities.

Figure 2.10 | *C. difficile* infection alters community-level select pathway and gene expression of the gut microbiota across perturbed communities. Metatranscriptomic (cDNA) read abundances associated with each gene were normalized to their associated metaganomic (DNA) coverage, resulting in values that reflect upregulation. **(A)** Differences in read abundance for top 5 KEGG pathways with the largest amount of change between Mock and *C. difficile*-infected states within each antibiotic treatment. Values above the center line were expressed more during *C. difficile* infection, and those below the line were expressed more in Mock-infected animals. Dotted lines indicate average expression of pathways associated with each condition. **(B-D)** cDNA read abundances of the top 10 genes with the largest differences in expression within each indicated antibiotic treatment group. Shown are the expression levels for the genes displayed on the left during Mock (white) or *C. difficile* (black) infection. Gene names and member pathways indicated by letter codes along the left axis (pathway legend below).

We then sought to identify changes those metabolites that were potential markers of *C. difficile* infection. To accomplish this task, we applied the same random forest that was used earlier to differentiate infected and uninfected metabolomes overall and within each treatment group separately. We were able to distinguish those microbiomes infected with *C. difficile* from those that were not, and reported the 10 metabolites with the greatest MDA (Fig. 2.8B). The strongest single predictor and only metabolite among the top 10 that was increased during infection was 5-aminovalerate. This amino acid analog is a known byproduct of D-proline fermentation in *C. difficile* (31). *C. difficile* is able to catabolize proline along with glycine through a set of paired biochemical reactions known as Stickland fermentation (32). Additional prominent signatures of infection across our metabolomic datasets were the significant decreases to the concentration of 4 individual proline-containing amino acids which were each highly abundant in the absence of infection (all $p < 0.001$). These combined results demonstrated that while distinct metabolic challenges may exist in each susceptible metabolome, Stickland fermentation could perhaps be a preferred energy acquisition pathway for *C. difficile in vivo*. In agreement with the ordination analysis, random forest was only able to reliably classify infected mice in streptomycin and cefoperazone treatment while clindamycin maintained a high out-of-bag error (OOB = 44.44%). Metabolites that distinguished infection conditions within each treatment group had a high level of variation with only a few shared metabolites including acetylarginine, dimethylguanine, and adenine however none were in top 5 of all groups. Despite a moderate amount of conserved metabolic signatures across infections these data

support our hypothesis that not only does each antibiotic treatment create different microbiota community structures with different metabolic potentials, but also that the metabolism of these distinct communities responds to *C. difficile* colonization in a unique manner. However, it is not possible from these data to distinguish changes to the metabolome that were a result of altered community metabolism, altered host metabolism, or from *C. difficile* directly.

C. difficile colonization induced shifts in the expression of several catabolic pathways and nutrient acquisition systems across susceptible communities. Our combined 16S rRNA gene sequencing and metabolomic results demonstrated that antibiotic treatments resulted in distinct bacterial communities which likely led to the altered metabolite profiles in each; however, it does not preclude the possibility that the host or *C. difficile* itself (only during infection) were responsible for the metabolomic differences. We hypothesized that the altered metabolic function of the microbiota as a product of antibiotic perturbation drove the changes seen in the metabolomes of the respective antibiotic treatment groups. To gain a more specific understanding how the microbiota is shaping the metabolic environemtn in each treatment group, we employed a metagenomic-enabled metatranscriptomic shotgun sequencing approach with paired DNA and RNA samples collected from the cecal content of the mice used in the previous analyses. Metagenomic reads from mock-infected cecal communities were assembled into contigs and putative genes were identified resulting in 234,124 (streptomycin), 83,096 (cefoperazone), and 35,977 (clindamycin) potential genes in each metagenome. Streptomycin treatment resulted in a significantly more diverse

community (Fig. 2.4A) than other groups, so a larger detectable metagenome was expected. Putative genes were then annotated according to KEGG and the subset of genes the were successfully annotated with function were utilized for the next analysis. DNA and cDNA reads from both infected and uninfected conditions were then mapped to the gene catalog that corresponded to their antibiotic treatment group of origin. The resulting abundances were normalized to both sequencing read length and target gene length to yield a per base mapping abundance. Finally, after equal subsampling, metatranscriptomic read abundances for each gene were divided by their corresponding metagenomic coverage in order to normalize for overrepresented genes and species. Therefore, final expression values represent the level of expression upregulation for each gene outside of those from which transcript is abundant but is only expressed at low levels from genes which are highly abundant.

Utilizing the fully normalized metatranscriptomes, we first focused on differences in gene expression in broadly defined pathways and gene categories for each antibiotic treatment with and without *C. difficile*-infection. Based on the metabolomic results, we hypothesized that pathways with the greatest differences would include those involved in the metabolism of carbohydrates and amino acids. We then calculated the difference in cDNA abundance for each pathway between infected and uninfected conditions, represented as delta-cDNA abundance. To highlight the largest dfferences, we limited the analysis to the top 5 KEGG pathways with the most change between mock and *C. difficile*-infected conditions within each treatment (Fig. 2.10A). In streptomycin and clindamycin treatments, greater expression of KEGG pathways was observed in the *C.*

difficile-infected metatranscriptomes. Both groups displayed large changes in amino sugar metabolism and ABC transporters, however other distinct carbon metabolism pathways were upregulated in each. Glycolysis and oxidative phosphorylation were overrepresented in streptomycin treated mice while starch/sucrose metabolism and PTS systems were more abundant associated with clindamycin treated mice. Together these shifts suggest that communities differentially adapt carbon metabolism pathways in response to colonization of *C. difficile*. Conversely, the largest differences seen in cefoperazone treated mice were over-expressed in the absence of *C. difficile* infection. These pathways included three separate pathways for the replication or manipulation of genetic material (RNA Processing, tRNA Sythesis, & Homologous Recombination). Instead, many genes involved in anaerobic glucose metabolism and select ABC-transporters were upregulated. These results indicate that the cecal microbiota of infected mice shifts its metabolism toward catabolizing simple carbohydrates. We then moved on to perform a more fine-scale analysis of changes at the gene level, by selecting the genes in each antibiotic treatment group with the largest disparity in normalized cDNA abundance between mock and *C. difficile*-infected groups (Fig. 2.10B - D). In agreement with pathway-level differences in expression, the majority of genes with the greatest difference between mock and infected mice belonged to pathways highlighted in Fig. 2.10A with three additional pathways relating to amino acid metabolism (I: Glutathione metabolism, J: Valine/Leucine/Isoleucine metabolism, & K: Glycine/Serine/Threonine metabolism). Numerous genes for transport of simple carbohydrates and glycolysis were also differentially overrepresented under infected

conditions across treatment groups when compared to mock infection of untreated, resistant communities (Fig. 2.10). These data expand on our interpretation of pathway-level analysis and support our hypothesis that *C. difficile* colonization leads to changes in community-level expression of genes for nutrient acquisition and catabolism.

C. difficile colonization corresponds with large-scale changes in expression of genes from specific bacterial taxa. Because not all bacterial taxa share identical metabolic capabilities, we hypothesized that specific subsets of bacteria were differentially affected by the presence or metabolic activity of *C. difficile*. We sought to delineate the transcriptomic contributions of separate bacterial taxa within each metatranscriptomic dataset. To accomplish this we utilized the genus level taxonomic information associated with each KEGG annotation to identify which group likely contributed a given gene to the metagenome. Many genes in the KEGG database are annotated as hypothetical or uncharacterized but still possess a taxonomic annotation. This resulted in substantially more genes from the total being conserved for analysis in each group. With these data, we narrowed the focus onto transcription for genera that represented >0.01% of genes receiving taxonomic annotations in any of the metagenomic assemblies. We then directly compared the normalized cDNA abundances for each gene between infected and uninfected states for each antibiotic treatment (Fig. 2.5A-2.5C). Coordinates were determined by the relative expression of each gene in mock and *C. difficile* infection for x and y axes respectively. This causes genes with equal transcription in both conditions being compared to be strongly correlated and positioned proximal to the central diagonal line. As such, we applied

linear correlation and a squared residual cutoff to define those genes that are most strongly upregulated in either condition, and finally calculated the mean distance of outliers in each group from the center line (represented in arbitrary units or AU). This resulted in 2473 outliers at an average distance of 2.545 AU associated with streptomycin, 2930 outliers at an average distance 3.854 AU with cefoperazone, and only 727 outliers at an average distance of 2.414 AU in clindamycin treatment. Overall, the clindamycin treatment was associated with the fewest gene expression outliers.

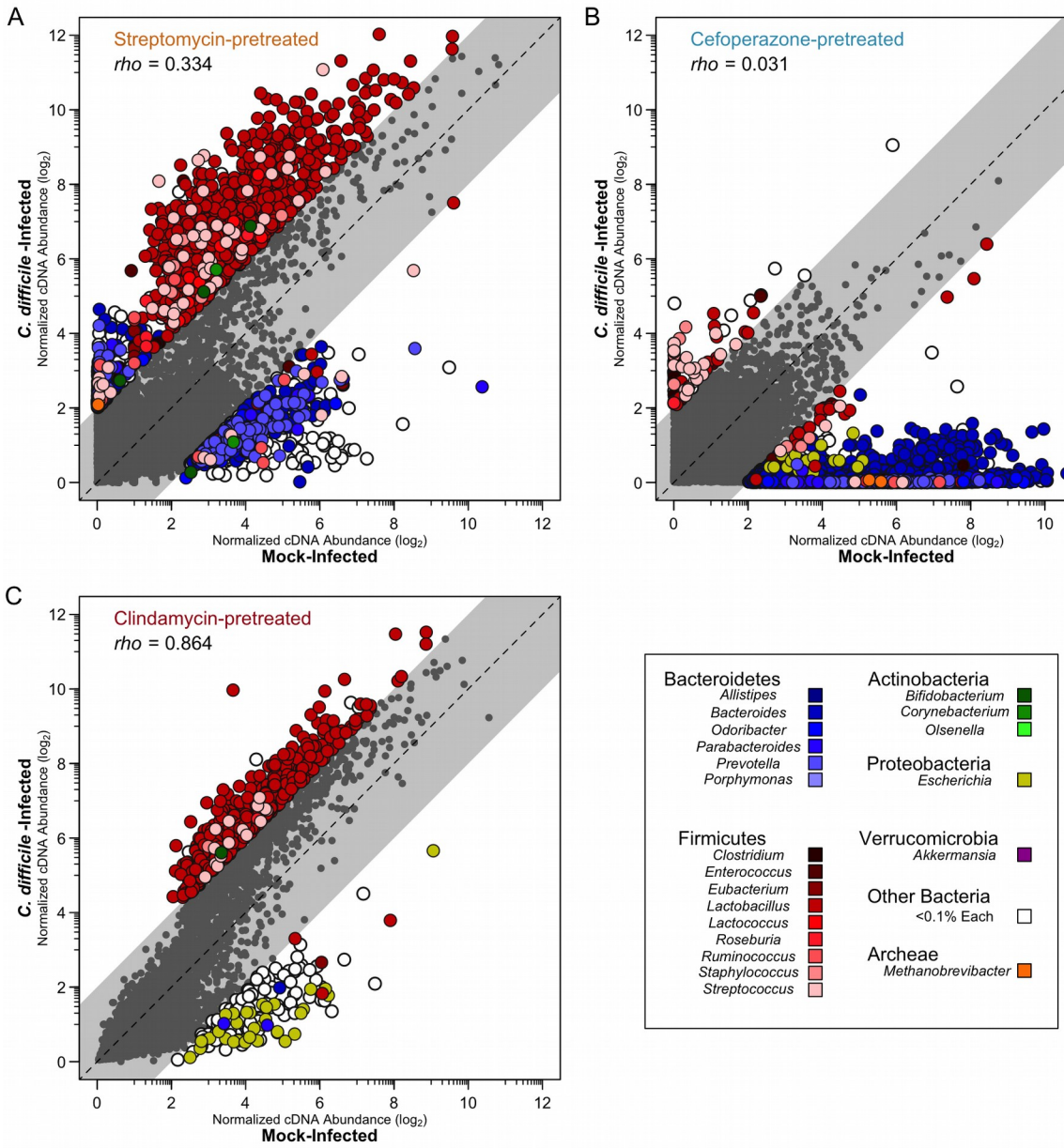

Figure 2.11 | *C. difficile* colonization alters gene expression of taxonomic groups differentially between antibiotic treatments. Each point represents a unique gene from the respective metagenomic assembly. Coordinates were determined by the log2-transformed expression level of each gene between *C. difficile*-infected and mock-infected conditions. Metatranscriptomic read abundances were normalized to their associated metagenomic coverage. Colored indicate genus of origin, and gray areas denote genes with consistent expression between conditions and outliers to this region were determined by least squares regression analysis with a minimum residual value of 2. Antibiotic treatments; **(A)** Streptomycin-treated, **(B)** Cefoperazone-treated, and **(C)** Clindamycin-treated.

Incorporation of the genes-level taxonomic information for each transcript revealed that outlier genes were contributed by underrepresented genera. First, in streptomycin-treated mice, the most prominent differences were in 937 genes belonging to Lactobacillus that were upregulated with *C. difficile* infection (Fig. 2.11A). Next, in cefoperazone treatment 2290 genes belonging to Bacteroides were upregulated in mock infected mice (Fig. 2.11B). A consistent trend in streptomycin and cefoperazone treatments was an overrepresentation of highly expressed genes from genera belonging to Bacteroidetes during mock infection. The metatransciptomes within both of these treatment conditions poorly correlated between mock and infected conditions, indicating a high degree of change induced by *C. difficile* colonization ($r = 0.0334$ & $r = 0.031$). Finally, in clindamycin treated mice the largest difference in transcription was for 510 Lactobacillus genes during *C. difficile* infected mice (Fig. 2.11C). Infected and uninfected metatranscriptomes associated with this antibiotic correlated the more strongly than either other treatment ($r = 0.862$), further supporting that *C. difficile* colonization had a low impact on transcription of the cecal microbiota. These data support that *C. difficile* may differentially modify the transcriptional activity of separate microbial taxa based on the context of the community in which it is colonizing. This could have implications in altering the ecosystem of the gut to promote persistence and ultimately negatively affect the ability of the community to clear infection.

Discussion

Our results demonstrate that distinct intestinal ecosystems are impacted differently by *C. difficile* colonization and that these changes to community metabolism could have implications for the ability of the pathogen to persist in those environments. Furthermore, discordant modifications to OTU abundances, transcriptional activity of the microbiota, and cecal metabolome shortly after infection suggest that *C. difficile* manipulates the niche landscape of the intestinal tract in order to better maintain long-term colonization. This hypothesis is best supported by the disparity of community-level phenotypes between clindamycin and cefoperazone/streptomycin treatment models. Only in the clindamyin treatment model were animals able to go on to clear detectable *C. difficile* colonization within 18 hours. Unlike the other conditions, clindamycin-treated communities were significantly altered in community structure following *C. difficile* colonization, but not at the metabolomic level (Fig. 2.4F). This disconnect between community structure differences and overall metabolic output was explained by the low levels of change in the metatranscriptomes of these communities when compared to the other treatment groups in response to infection (Fig. 2.9D). Collectively these results advance the idea that in order for *C. difficile* to maintain colonization for longer periods of time, it must partition desired niche spaces in the context of a given microbiome, and that the activities of certain subsets may be more readily reshaped than others. Instances of active nutrient niche restructuring in the gut have been documented previously for prominent symbiotic bacterial species in gnotobiotic mice (33), but never

before in a model of infection with a conventional community of microbes. Interestingly, taxonomic groups most highly represented as outliers in the normalized metatranscriptomes of each antibiotic treatments were non-dominant species of each respective cecal community by 16S rRNA gene sequencing (on average <5% of community; Fig. 2.3A). These data give the impression that *C. difficile* may "attack the loser", meaning those populations more targeted by the antibiotic treatment and in the midst of recovery, in order to have the highest probability of success in the gut environment it is currently colonizing. Previous studies have found that rare taxonomic groups, even those at a low abundance as a result of a spontaneous perturbation, may have disproportion effects on the metabolic environment of the community at large (34). For example, this strategy has been observed in temperate lakes where conditionally rare microbes were found to be far more metabolically active than highly abundant taxa (35), and this concept would likely apply to bacterial groups recovering population density following cessation of antibiotic treatment. As such, *C. difficile* may preferentially seek to compete with these organisms to ultimately affect greater change to the entire ecosystem and open a long-lasting nutrient niche. While this hypothesis requires further exploration to adequately support, it provides an ecological framework for future research questions concerning the interactions of *C. difficile* with susceptible communities in the gut.

This study is one of the first *in vivo* observations of a medically relevant bacterial pathogen altering metabolic activity of a host-associated community to potentially promote colonization. Another group had previously identified potential metabolite

markers of *C. difficile* infection in patient feces (36), but were not able to make connections to changes to community metabolism that were afforded to us by paired untargeted metabolomic analyasis and metatranscriptomic sequencing. In a recent study, one group found that a tick-vectored bacterial pathogen alters the ability of the resident microbiota of the tick by interrupting proper biofilm formation and allowing lasting colonization (37). In both cases the pathogen modifies aspects of the microbiota it is colonizing, however in the case of *C. difficile* the interaction appears to more centered on access to nutrients than a persistent spatial niche. While we acknowledge that this study may not elucidate the specific mechanism by which this interaction occurs, the combined systems analysis strengthens each individual level of observation and only when employed together does a clearer definition of *C. difficile*-related microbial ecology in the gut emerge. This research lays the groundwork for a more rationale consideration of the metabolic functionalities of bacterial taxa to consider when attempting to rebuild *C. difficile* colonization resistance across differentially perturbed gut environments.

In spite of consistent signals across multiple levels of -omics datasets, possible shortcomings to our interpretation of the presented data do exist. First, as with all transcriptomic studies, the relative level of mRNA detected for a given gene does not necessarily reflect the amount of functional protein made by a cell. This topic is discussed in greater depth in Chapter 3. Furthermore, interpretation of timing may also be an issue since a large influx of transcript for a specific product may signal an initial upregulation before subsequent translation has been able to occur. This also omits

consideration of any post-translational modifications that are required for ultimately functional enzymes. In terms of metabolomics, alternative possible interpretations of the data also exist. For example, I considered metabolites that did not change in concentration between uninfected and infected conditions as unimpacted by changes in bacterial metabolism induced by *C. difficile* colonization. However, this may instead indicate that the metabolism of *C. difficile* itself may simply replace the level present in the uninfected community. Such instances would not be detectable through untargeted mass spectrometry alone, however the combination of methods utilized here present a much more unified description of the system than any of the component techniques alone.

Several groups have attempted to identify single bacterial species or limited strain consortia that are able to replicate this effect, but each has been met with incomplete restoration of colonization resistance or function through yet unexplored means (30, 38–40). The effect we observed of *C. difficile* colonization on community metabolic activity could also be linked to pathogen strain and may offer explanation to the propensity of some strains to persist over others where toxin activity could play a role (41). Moreover, the current work contributes to the existing concept that the healthy gut microbiota maintains colonization resistance to *C. difficile* by outcompeting the pathogen for preferred nutrient niche space. Ultimately, our results suggest that each susceptible and subsequently infected microbiome may be unique and require specific microbes or functionalities to restore colonization resistance to *C. difficile* in that specific context. Conversely, colonization resistance against *C. difficile* may be the result of

contributions by distinct subcommunities of bacteria across each unique resistant gut community. As the microbiome is so intimately connected to colonization resistance against the bacterium, it has become imperative to understand what factors allow some gut environments to be persistently colonized while others are not. This research lays the groundwork for future studies to assess context dependent restoration of *C. difficile* colonization resistance and what factors are able to interfere with the ability of *C. difficile* to modify gut ecology in order to promote clearance.

References

1. **Vollaard, E. J.**, and **H. A. L. Clasener**. 1994. Colonization resistance. U.S. Patent 3.

2. **Freter, R.** 1955. The Fatal Enteric Cholera Infection in the Guinea Pig, Achieved by Inhibition of Normal Enteric Flora. The Journal of Infectious Diseases **97**:57–65.

3. **Fekety, R.**, **J. Silva**, **R. Toshniwal**, **M. Allo**, **J. Armstrong**, **R. Browne**, **J. Ebright**, and **G. Rifkin**. 1979. Antibiotic-associated colitis: Effects of antibiotics on clostridium difficile and the disease in hamsters. Reviews of Infectious Diseases **1**:386–397.

4. **Britton, R. A.**, and **V. B. Young**. 2012. Interaction between the intestinal microbiota and host in Clostridium difficile colonization resistance. Trends in microbiology **20**:313–9.

5. **Lessa, F. C.**, **Y. Mu**, **W. M. Bamberg**, **Z. G. Beldavs**, **G. K. Dumyati**, **J. R. Dunn**, **M. M. Farley**, **S. M. Holzbauer**, **J. I. Meek**, **E. C. Phipps**, **L. E. Wilson**, **L. G. Winston**, **J. A. Cohen**, **B. M. Limbago**, **S. K. Fridkin**, **D. N. Gerding**, and **L. C. McDonald**. 2015. Burden of Clostridium difficile Infection in the United States. New England Journal of Medicine **372**:825–834.

6. **Antonopoulos, D. A.**, **S. M. Huse**, **H. G. Morrison**, **T. M. Schmidt**, **M. L. Sogin**, and **V. B. Young**. 2009. Reproducible community dynamics of the

gastrointestinal microbiota following antibiotic perturbation. Infection and Immunity **77**:2367–2375.

7. **Buffie, C. G.**, **I. Jarchum**, **M. Equinda**, **L. Lipuma**, **A. Gobourne**, **A. Viale**, **C. Ubeda**, **J. Xavier**, and **E. G. Pamer**. 2012. Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. Infection and Immunity **80**:62–73.

8. **Antunes, L. C. M.**, **J. Han**, **R. B. R. Ferreira**, **P. Loli**, **C. H. Borchers**, and **B. B. Finlay**. 2011. Effect of antibiotic treatment on the intestinal metabolome. Antimicrobial Agents and Chemotherapy **55**:1494–1503.

9. **Jump, R. L. P.**, **A. Polinkovsky**, **K. Hurless**, **B. Sitzlar**, **K. Eckart**, **M. Tomas**, **A. Deshpande**, **M. M. Nerandzic**, and **C. J. Donskey**. 2014. Metabolomics analysis identifies intestinal microbiota-derived biomarkers of colonization resistance in clindamycin-treated mice. PLoS ONE **9**.

10. **Theriot, C. M.**, **M. J. Koenigsknecht**, **P. E. Carlson**, **G. E. Hatton**, **A. M. Nelson**, **B. Li**, **G. B. Huffnagle**, **J. Z Li**, and **V. B. Young**. 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. Nature communications **5**:3114.

11. **Wilson, K. H.**, and **F. Perini**. 1988. Role of competition for nutrients in suppression of Clostridium difficile by the colonic microflora. Infection and Immunity **56**:2610–2614.

12. **Sambol, S. P.**, **M. M. Merrigan**, **J. K. Tang**, **S. Johnson**, and **D. N. Gerding**. 2002. Colonization for the Prevention of Clostridium difficile Disease in Hamsters. The Journal of infectious diseases **186**:14–16.

13. **Perez-Cobas, A. E.**, **A. Artacho**, **S. J. Ott**, **A. Moya**, **M. J. Gosalbes**, and **A. Latorre**. 2014. Structural and functional changes in the gut microbiota associated to Clostridium difficile infection. Frontiers in Microbiology **5**.

14. **Zaura, E.**, **B. W. Brandt**, **M. J. T. de Mattos**, **M. J. Buijs**, **M. P. M. Caspers**, **M. U. Rashid**, **A. Weintraub**, **C. E. Nord**, **A. Savell**, **Y. Hu**, **A. R. Coates**, **M. Hubank**, **D. A. Spratt**, **M. Wilson**, **B. J. F. Keijser**, and **W. Crielaard**. 2015. Same Exposure but two radically different responses to antibiotics: Resilience of the salivary microbiome versus long-term microbial shifts in feces. mBio **6**.

15. **Thomas, C.**, **M. Stevenson**, and **T. V. Riley**. 2003. Antibiotics and hospital-acquired Clostridium difficile-associated diarrhoea: A systematic review **51**:1339–1350.

16. **Brown, K. A.**, **N. Khanafer**, **N. Daneman**, and **D. N. Fisman**. 2013. Meta-analysis of antibiotics and the risk of community-associated Clostridium difficile infection. Antimicrobial Agents and Chemotherapy **57**:2326–2332.

17. **Bignardi, G.** 1998. Risk factors for Clostridium difficile infection. Journal of Hospital Infection **40**:1–15.

18. **Schubert, A. M.**, **H. Sinani**, and **P. D. Schloss**. 2015. Antibiotic-induced alterations of the murine gut microbiota and subsequent effects on colonization resistance against Clostridium difficile. mBio **6**.

19. **Jenior, M. L.**, **J. L. Leslie**, **V. B. Young**, and **P. D. Schloss**. 2017. Clostridium difficile colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. bioRxiv.

20. **Theriot, C. M.**, **C. C. Koumpouras**, **P. E. Carlson**, **I. I. Bergin**, **D. M. Aronoff**, and **V. B. Young**. 2011. Cefoperazone-treated mice as an experimental platform to assess differential virulence of Clostridium difficile strains. Gut microbes **2**:326–334.

21. **Bouillaut, L.**, **T. Dubois**, **A. L. Sonenshein**, and **B. Dupuy**. 2015. Integration of metabolism and virulence in Clostridium difficile. Research in Microbiology **166**:375–383.

22. **Eijk, E. van**, **S. Anvar**, **H. P. Browne**, **W. Leung**, **J. Frank**, **A. M. Schmitz**, **A. P. Roberts**, and **W. Smits**. 2015. Complete genome sequence of the Clostridium difficile laboratory strain 630erm reveals differences from strain 630, including translocation of the mobile element CTn5. BMC Genomics **16**:31.

23. **Koenigsknecht, M. J.**, **C. M. Theriot**, **I. L. Bergin**, **C. A. Schumacher**, **P. D. Schloss**, and **V. B. Young**. 2015. Dynamics and establishment of Clostridium difficile infection in the murine gastrointestinal tract. Infection and Immunity **83**:934–941.

24. **Segata, N.**, **J. Izard**, **L. Waldron**, **D. Gevers**, **L. Miropolsky**, **W. S. Garrett**, and **C. Huttenhower**. 2011. Metagenomic biomarker discovery and explanation. Genome Biology **12**:R60.

25. **Liaw, a**, and **M. Wiener**. 2002. Classification and Regression by randomForest. R news **2**:18–22.

26. **Goyer, A.**, **T. L. Johnson**, **L. J. Olsen**, **E. Collakova**, **Y. Shachar-Hill**, **D. Rhodes**, and **A. D. Hanson**. 2004. Characterization and Metabolic Function of a

Peroxisomal Sarcosine and Pipecolate Oxidase from Arabidopsis. Journal of Biological Chemistry **279**:16947–16953.

27. **Fonknechten, N.**, **S. Chaussonnerie**, **S. Tricot**, **A. Lajus**, **J. R. Andreesen**, **N. Perchat**, **E. Pelletier**, **M. Gouyvenoux**, **V. Barbe**, **M. Salanoubat**, **D. Le Paslier**, **J. Weissenbach**, **G. N. Cohen**, and **A. Kreimeyer**. 2010. Clostridium sticklandii, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. BMC genomics **11**:555.

28. **Nakamura, S.**, **S. Nakashio**, **K. Yamakawa**, **N. Tanabe**, and **S. Nishida**. 1982. Carbohydrate Fermentation by Clostridium difficile. Microbiology and Immunology **26**:107–111.

29. **Sorg, J. A.**, and **A. L. Sonenshein**. 2010. Inhibiting the initiation of Clostridium difficile spore germination using analogs of chenodeoxycholic acid, a bile acid. Journal of Bacteriology **192**:4983–4990.

30. **Buffie, C. G.**, **V. Bucci**, **R. R. Stein**, **P. T. McKenney**, **L. Ling**, **A. Gobourne**, **D. No**, **H. Liu**, **M. Kinnebrew**, **A. Viale**, **E. Littmann**, **M. R. M. van den Brink**, **R. R. Jenq**, **Y. Taur**, **C. Sander**, **J. R. Cross**, **N. C. Toussaint**, **J. B. Xavier**, and **E. G. Pamer**. 2014. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature **517**:205–208.

31. **Neumann-Schaal, M.**, **J. D. Hofmann**, **S. E. Will**, and **D. Schomburg**. 2015. Time-resolved amino acid uptake of Clostridium difficile 630erm and concomitant fermentation product and toxin formation. BMC Microbiology 281.

32. **Bouillaut, L.**, **W. T. Self**, and **A. L. Sonenshein**. 2013. Proline-dependent regulation of Clostridium difficile stickland metabolism. Journal of Bacteriology **195**:844–854.

33. **Mahowald, M. A.**, **F. E. Rey**, **H. Seedorf**, **P. J. Turnbaugh**, **R. S. Fulton**, **A. Wollam**, **N. Shah**, **C. Wang**, **V. Magrini**, **R. K. Wilson**, **B. L. Cantarel**, **P. M. Coutinho**, **B. Henrissat**, **L. W. Crock**, **A. Russell**, **N. C. Verberkmoes**, **R. L. Hettich**, and **J. I. Gordon**. 2009. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. Proceedings of the National Academy of Sciences **106**:5859–5864.

34. **Jousset, A.**, **C. Bienhold**, **A. Chatzinotas**, **L. Gallien**, **A. Gobet**, **V. Kurm**, **K. Küsel**, **M. C. Rillig**, **D. W. Rivett**, **J. F. Salles**, **M. G. A. van der Heijden**, **N. H. Youssef**, **X. Zhang**, **Z. Wei**, and **W. H. G. Hol**. 2017. Where less may be more: how the rare biosphere pulls ecosystems strings. The ISME Journal.

35. **Shade, A.**, **S. E. Jones**, **J. Gregory Caporaso**, **J. Handelsman**, **R. Knight**, **N. Fierer**, and **J. A. Gilbert**. 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. mBio **5**.

36. **Rojo, D.**, **M. J. Gosalbes**, **R. Ferrari**, **A. E. Pérez-Cobas**, **E. Hernández**, **R. Oltra**, **J. Buesa**, **A. Latorre**, **C. Barbas**, **M. Ferrer**, and **A. Moya**. 2015. Clostridium difficile heterogeneously impacts intestinal community architecture but drives stable metabolome responses. The ISME Journal **9**:2206–2220.

37. **Abraham, N. M.**, **L. Liu**, **B. L. Jutras**, **A. K. Yadav**, **S. Narasimhan**, **V. Gopalakrishnan**, **J. M. Ansari**, **K. K. Jefferson**, **F. Cava**, **C. Jacobs-Wagner**, and **E. Fikrig**. 2017. Pathogen-mediated manipulation of arthropod microbiota to promote infection. Proceedings of the National Academy of Sciences 201613422.

38. **Reeves, A. E.**, **M. J. Koenigsknecht**, **I. L. Bergin**, and **V. B. Young**. 2012. Suppression of Clostridium difficile in the gastrointestinal tracts of germfree mice inoculated with a murine isolate from the family Lachnospiraceae. Infection and Immunity **80**:3786–3794.

39. **Lawley, T. D.**, **S. Clare**, **A. W. Walker**, **M. D. Stares**, **T. R. Connor**, **C. Raisen**, **D. Goulding**, **R. Rad**, **F. Schreiber**, **C. Brandt**, **L. J. Deakin**, **D. J. Pickard**, **S. H. Duncan**, **H. J. Flint**, **T. G. Clark**, **J. Parkhill**, and **G. Dougan**. 2012. Targeted Restoration of the Intestinal Microbiota with a Simple, Defined Bacteriotherapy Resolves Relapsing Clostridium difficile Disease in Mice. PLoS Pathogens **8**.

40. **Petrof, E. O.**, **G. B. Gloor**, **S. J. Vanner**, **S. J. Weese**, **D. Carter**, **M. C. Daigneault**, **E. M. Brown**, **K. Schroeter**, and **E. Allen-Vercoe**. 2013. Stool substitute transplant therapy for the eradication of Clostridium difficile infection: 'RePOOPulating' the gut. Microbiome **1**:3.

41. **Adlerberth, I.**, **H. Huang**, **E. Lindberg**, **N. Åberg**, **B. Hesselmar**, **R. Saalman**, **C. E. Nord**, **A. E. Wold**, and **A. Weintraubb**. 2014. Toxin-Producing clostridium difficile strains as long-term gut colonizers in healthy infants. Journal of Clinical Microbiology **52**:173–179.

42. **Wilson, K. H.**, **M. J. Kennedy**, and **F. R. Fekety**. 1982. Use of sodium taurocholate to enhance spore recovery on a medium selective for Clostridium difficile. Journal of Clinical Microbiology **15**:443–446.

43. **Leslie, J. L.**, **S. Huang**, **J. S. Opp**, **M. S. Nagy**, **M. Kobayashi**, **V. B. Young**, and **J. R. Spence**. 2015. Persistence and toxin production by Clostridium difficile within human intestinal organoids result in disruption of epithelial paracellular barrier function. Infection and Immunity **83**:138–145.

44. **Kozich, J. (. of M.**, and **P. Schloss**. 2013. 16S Sequencing with the Illumina MiSeq Personal Sequencer. University of Michigan Health System SOP **3.1**:1–16.

45. **Lopez-Medina, E.**, **M. M. Neubauer**, **G. B. Pier**, and **A. Y. Koh**. 2011. RNA isolation of Pseudomonas aeruginosa colonizing the murine gastrointestinal tract. Journal of visualized experiments : JoVE 6–9.

46. **Franzosa, E. A.**, **X. C. Morgan**, **N. Segata**, **L. Waldron**, **J. Reyes**, **A. M. Earl**, **G. Giannoukos**, **M. R. Boylan**, **D. Ciulla**, **D. Gevers**, **J. Izard**, **W. S. Garrett**, **A. T. Chan**, and **C. Huttenhower**. 2014. Relating the metatranscriptome and metagenome of the human gut. Proceedings of the National Academy of Sciences **111**:E2329–E2338.

47. **Martin, M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet **17**:10.

48. **Joshi, N.**, and **J. Fass**. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle. 2011.

49. **Li, D.**, **C. M. Liu**, **R. Luo**, **K. Sadakane**, and **T. W. Lam**. 2014. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics **31**:1674–1676.

50. **Ogata, H.**, **S. Goto**, **K. Sato**, **W. Fujibuchi**, **H. Bono**, and **M. Kanehisa**. 1999. KEGG: Kyoto encyclopedia of genes and genomes. U.S. Patent 1.

51. **Buchfink, B.**, **C. Xie**, and **D. H. Huson**. 2015. Fast and sensitive protein alignment using DIAMOND. Nature methods **12**:59–60.

52. **Langmead, B.**, and **S. L. Salzberg**. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods **9**:357–9.

53. **Li, H.**, **B. Handsaker**, **A. Wysoker**, **T. Fennell**, **J. Ruan**, **N. Homer**, **G. Marth**, **G. Abecasis**, and **R. Durbin**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.

54. **Breiman, L.** 2001. Random forests. Machine Learning **45**:5–32.

55. **Segal, M. R.** 2004. Machine Learning Benchmarks and Random Forest Regression. Biostatistics 1–14.

CHAPTER THREE

CLOSTRIDIUM DIFFICILE COLONIZES ALTERNATIVE NUTRIENT NICHES DURING INFECTION ACROSS DISTINCT MURINE GUT MICROBIOMES

Summary

Clostridium difficile is the largest single cause of hospital-acquired infection in the United States. A major risk factor for *Clostridium difficile* infection (CDI) is prior exposure to antibiotics, as they disrupt the gut bacterial community which protects from *C. difficile* colonization. Multiple antibiotic classes have been associated with CDI susceptibility; many leading to distinct community structures stemming from variation in bacterial targets of action. These microbiomes present separate metabolic challenges to *C. difficile*, therefore we hypothesized that the pathogen adapts its physiology to available nutrients within different gut environments. Utilizing an *in vivo* CDI model, we demonstrated *C. difficile* highly colonized ceca of mice pretreated with any of three antibiotics from distinct classes. Levels of *C. difficile* spore formation and toxin activity varied between animals based on the antibiotic administered. These physiologic processes in *C. difficile* are partially regulated by environmental nutrient concentrations. To investigate metabolic responses of the bacterium *in vivo*, we performed transcriptomic analysis of *C. difficile* from ceca of infected mice across pretreatments. This revealed heterogeneous expression in numerous catabolic pathways for diverse growth substrates. To assess which resources *C. difficile* exploited, we developed a

genome-scale metabolic model with a transcriptomic-enabled metabolite scoring algorithm integrating network architecture. This platform identified nutrients *C. difficile* used preferentially between infections, which were validated through untargeted mass spectrometry of each microbiome. These data supported the hypothesis that *C. difficile* inhabits alternative nutrient niches across cecal microbiomes with increased preference for nitrogen-containing carbon sources, particularly Stickland fermentation substrates and host-derived glycans. Our metabolite score calculation also provides a platform to study nutrient requirements of pathogens during the context of infection. Our results suggest that *C. difficile* colonization resistance is mediated by multiple groups of bacteria competing for several subsets of nutrients, and could explain why total reintroduction of competitors through fecal microbial transplant is the most effective treatment to date. This work could ultimately contribute to the identification of targeted, context-dependent measures that prevent or reduce *C. difficile* colonization including pre- and probiotic therapies.

Introduction

Infection by the Gram-positive, spore-forming bacterium *Clostridium difficile* has increased in both prevalence and severity across numerous countries during the last decade (1). In the United States, *C. difficile* was estimated to have caused >500,000 infections and resulted in ~$4.8 billion worth of acute care costs in 2014 (2). *C. difficile* infection (CDI) causes an array of toxin-mediated symptoms ranging from abdominal pain and diarrhea to the more severe conditions pseudomembraneous colitis and toxic

megacolon. Prior treatment with antibiotics is the most common risk factor associated with development of CDI (3). Antibiotics likely contribute to susceptibility to CDI by disrupting the gut microbiota (4). In mouse models, multiple antibiotics can induce susceptibility to *C. difficile* colonization (5–7). Notably, each antibiotic resulted in unique gut bacterial communities that were receptive to high levels of *C. difficile* colonization. Others have also shown that antibiotics from multiple classes also alter the gut metabolome, increasing the concentrations of some *C. difficile* growth substrates (6, 8–10). The ability of an unaltered murine gut community to exclude *C. difficile* colonization supports the nutrient-niche hypothesis, which states that an organism must be able to utilize a subset of available resources better than all competitors to colonize the intestine (11, 12). Taken together these results are a strong indication that the healthy gut microbiota inhibits the growth of *C. difficile* by limiting the availability of the substrates it needs to grow.

Based on genomic and *in vitro* growth characteristics, *C. difficile* appears able to adapt to a variety nutrient niches (13). *C. difficile* has a relatively large and mosaic genome, it can utilize a variety of growth substrates, and possesses a diverse array host range (6, 14–16). These qualities are hallmarks of ecological generalists (17). *C. difficile* has also been shown to integrate signals from multiple forms of carbon metabolism to regulate its pathogenesis. *in vitro* transcriptomic analyses suggests that high concentrations of easily metabolized carbon sources, such as glucose or amino acids, inhibit toxin gene expression and sporulation (18, 19). Other studies have indicated that other aspects of *C. difficile* metabolism may be influenced through

environmental nutrient concentration-sensitive global transcriptional regulators such as CodY and CcpA (20, 21). These previous analyses have mainly focused on *in vitro* growth (22, 23) or colonization of germfree mice (14, 21). Although these analyses are informative, they are either primarily directed toward the expression of pathogenicity factors or lack the context of the gut microbiota which *C. difficile* must compete against for substrates. Metabolomic investigations have also been used to assay changes in bacterial metabolism as they relate to CDI and have characterized the levels of germinants and growth substrate availability (6, 10); however, metabolomic approaches are unable to attribute a metabolite to specific organisms in the gut community. Thus metabolomics more closely represents the echoes of total community metabolism, not the currently active processes of any one population. It has thus far not been possible to study *C. difficile*'s metabolism *in vivo*.

To overcome these limitations, we implemented transcriptomic and untargeted metabolomic analyses of *C. difficile* and the surrounding environemnt to better understand the active metabolic pathways in a model of infection. Based on the ability of *C. difficile* to grow on a diverse array of carbon sources and its ability to colonize a variety of communities, we hypothesized that *C. difficile* adapts its metabolism to fit the context of the environment it is attempting colonize. To test this hypothesis, we employed a mouse model of infection to compare the response of *C. difficile* to the gut environment caused by three antibiotics from distinct classes. By characterizing the transcriptome of *C. difficile* in these different communities and the metabolome of the

respective environments using paired samples from the same groups of mice, we were able to generate a systems model to directly test the nutrient-niche hypothesis.

Experimental Procedures

Animal care and antibiotic administration

Six-to-eight week-old GF C57BL/6 mice were obtained from a single breeding colony maintained at the University of Michigan and fed Laboratory Rodent Diet 5001 from LabDiet for all experiments. All animal protocols were approved by the University Committee on Use and Care of Animals at the University of Michigan and carried out in accordance with the approved guidelines. Specified SPF animals were administered one of three antibiotics; cefoperazone, streptomycin, or clindamycin (Table 2.1). Cefoperazone (0.5 mg/ml) and streptomycin (5.0 mg/ml) were administered in distilled drinking water ad libitum for 5 days with 2 days recovery with untreated distilled drinking water prior to infection. Clindamycin (10 mg/kg) was given via intraperitoneal injection 24 hours before time of infection. Adapted from a previously described model (24).

C. difficile infection and necropsy

All *C. difficile* strain 630 spores were prepared from a single large batch whose concentration was determined a week prior to challenge. On the day of challenge, 1×103 *C. difficile* spores were administered to mice via oral gavage in phosphate-buffered saline (PBS) vehicle. Subsequent quantitative plating to enumerate the spores was performed to ensure correct dosage. Mock-infected animals were given an oral

gavage of 100 µl PBS at the same time as those mice administered *C. difficile* spores. 18 hours following infection, mice were euthanized by carbon dioxide asphyxiation and necropsied to obtain the cecal contents. Two 100 µl aliquots were immediately flash frozen for later DNA extraction and toxin titer analysis, respectively. A third 100 µl aliquot was quickly transferred to an anaerobic chamber for quantification of *C. difficile* abundance. The remaining content in the ceca (approximately 1 mL) was mixed with 1 mL of sterile PBS in a stainless steel mortar housed in a dry ice and ethanol bath. The cecal contents of 9 mice from 3 cages was pooled into the mortar. Pooling cecal contents was necessary so that there would be a sufficient quantity of high quality rRNA-free RNA for deep sequencing. The pooled content was then finely ground and stored at -80° C for subsequent RNA extraction.

C. difficile cultivation and quantification

Cecal samples were weighed and serially diluted under anaerobic conditions (6% H, 20% CO2, 74% N2) with anaerobic PBS. Differential plating was performed to quantify both *C. difficile* spores and vegetative cells by plating diluted samples on CCFAE plates (fructose agar plus cycloserine (0.5%), cefoxitin (0.5%), and erythromycin (0.2%)) at 37° C for 24 hours under anaerobic conditions (42). It is important to note that the germination agent taurocholate was omitted from these plates to quantify only vegetative cells. In parallel, undiluted samples were heated at 60° C for 30 minutes to eliminate vegetative cells and leave only spores (43). These samples were serially diluted under anaerobic conditions in anaerobic PBS and plated on

CCFAE with taurocholate (10%) at 37° C for 24 hours. Plating was simultaneously done for heated samples on CCFAE to ensure all vegetative cells had been eliminated.

C. difficile toxin titer assay

To quantify the titer of toxin in the cecum a Vero cell rounding assay was performed as in (44). Briefly, filtered-sterilized cecal content was serially diluted in PBS and added to Vero cells in a 96-well plate. Plates were blinded and viewed after 24 hour incubation for cell rounding. A more detailed protocol with product information: https://github.com/SchlossLab/Jenior_Modeling_mSystems_2017/blob/master/protocols/toxin_assay/Verocell_ToxinActivity_Assay.Rmd

16S rRNA gene sequencing and read curation

DNA was extracted from approximately 50 mg of cecal content from each mouse using the PowerSoil-htp 96 Well Soil DNA isolation kit (MO BIO Laboratories) and an epMotion 5075 automated pipetting system (Eppendorf). The V4 region of the bacterial 16S rRNA gene was amplified using custom barcoded primers and sequenced as described previously using an Illumina MiSeq sequencer (45). All 63 samples were sequenced on a single sequencing run. The 16S rRNA gene sequences were curated using the mothur software package (v1.36), as described previously (45). In short, paired-end reads were merged into contigs, screened for quality, aligned to SILVA 16S rRNA sequence database, and screened for chimeras. Sequences were classified using a naive Bayesian classifier trained against a 16S rRNA gene training set provided by the

Ribosomal Database Project (RDP) (46). Curated sequences were clustered into operational taxonomic units (OTUs) using a 97% similarity cutoff with the average neighbor clustering algorithm. The number of sequences in each sample was rarefied to 2,500 per sample to minimize the effects of uneven sampling.

RNA extraction, shotgun library preparation, and sequencing

Pooled, flash-frozen samples were ground with a sterile pestle to a fine powder and scraped into a sterile 50 ml polypropylene conical tube. Samples were stored at -80° C until the time of extraction. Immediately before RNA extraction, 3 ml of lysis buffer (2% SDS, 16 mM EDTA and 200 mM NaCl) contained in a 50 ml polypropylene conical tube was first heated for 5 minutes in a boiling water bath (47). The hot lysis buffer was added to the frozen and ground cecal content. The mixture was boiled with periodic vortexing for another 5 minutes. After boiling, an equal volume of 37° C acid phenol/chloroform was added to the cecal content lysate and incubated at 37° C for 10 minutes with periodic vortexing. The mixture was the centrifuged at 2,500 x g at 4° C for 15 minutes. The aqueous phase was then transferred to a sterile tube and an equal volume of acid phenol/chloroform was added. This mixture was vortexed and centrifuged at 2,500 x g at 4° for 5 minutes. The process was repeated until aqueous phase was clear. The last extraction was performed with chloroform/isoamyl alcohol to remove the acid phenol. An equal volume of isopropanol was added and the extracted nucleic acid was incubated overnight at -20° C. The following day the sample was centrifuged at 12000 x g at 4° C for 45 minutes. The pellet was washed with 0° C 100%

ethanol and resuspended in 200 µl of RNase-free water. Samples were then treated with 2 µl of Turbo DNase for 30 minutes at 37° C. RNA samples were retrieved using the Zymo Quick-RNA MiniPrep. Completion of the DNase reaction was assessed using PCR for the V4 region of the 16S rRNA gene for 30 cycles (Kozich, 2013). Quality and integrity of RNA was measured using the Agilent RNA 6000 Nano kit for total prokaryotic RNA. The Ribo-Zero Gold rRNA Removal Kit Epidemiology was then used to deplete 16S and 18S rRNA from the samples. Prior to library construction, quality and integrity as measured again using the Agilent RNA 6000 Pico Kit. Stranded RNA-Seq libraries were made constructed with the TruSeq Total RNA Library Preparation Kit v2. The Agilent DNA High Sensitivity Kit was used to measure concentration and fragment size distribution before sequencing. High-throughput sequencing was performed by the University of Michigan Sequencing Core in Ann Arbor, MI. For all groups, sequencing was repeated across 4 lanes of an Illumina HiSeq 2500 using the 2x50 bp chemistry.

cDNA read curation, mapping, and normalization

Raw read curation was performed in a two step process. First, residual 5' and 3' Illumina adapter sequences were removed using CutAdapt (48) on a per library basis. Reads were then quality trimmed using Sickle (Joshi, 2011) on the default settings. An average of ~261,000,000 total reads (both paired and orphaned) remained after quality trimming. Mapping was accomplished using Bowtie2 (49) and the default stringent settings allowing for 0 mismatches again target reference genes. An average of ~6,880,000 reads in sample each mapped to the annotated nucleotide gene sequences

of Clostridioides difficile 630 from the KEGG: Kyoto Encyclopedia of Genes and Genomes (50). Optical and PCR duplicates were then removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard/), leaving an average of ~167,000 reads per sample for final analysis. The remaining mappings were converted to idxstats format using Samtools (51) and the read counts per gene were tabulated. Discordant pair mappings were discarded and counts were then normalized to read length and gene length to give a per base report of gene coverage. Each collection of reads was then subsampled to 90% of the lowest sequence total across the libraries resulting in even quantities of normalized read abundances in each group to be utilized in downstream analysis. This method was chosen as normalization to housekeeping genes would artificially remove their contributions to metabolic flux and reduce the information provided by our metabolite score calculations within our metabolic modeling approach.

Reaction Annotation & Bipartite Network Construction

The metabolism of *C. difficile* stain 630 was represented as a directed bipartite graph with both enzymes and metabolites as nodes. Briefly, models were semi-automatically constructed using KEGG (2016 edition) ortholog (KO) gene annotations to which transcripts had been mapped. Reactions that each KEGG ortholog mediate were extracted from ko_reaction.list located in /kegg/genes/ko/. KOs that do not mediate simple biochemical reactions (e.g. mediate interactions of macromolecules) were omitted. Metabolites linked to each reaction were retrieved from

reaction_mapformula.lst file located in /kegg/ligand/reaction/ from the KEGG release. Those reactions that did not have annotations for the chemical compounds the interact with are discarded. Metabolites were then associated with each enzyme and the directionality and reversibility of each biochemical conversion was also saved. This process was repeated for all enzymes in the given bacterial genome, with each enzyme and metabolite node only appearing once. The resulting data structure was an associative array of enzymes associated with lists of both categories of substrates (input and output), which could then be represented as a bipartite network. The final metabolic network of *C. difficile* strain 630 contained a total of 1205 individual nodes (447 enzymes and 758 substrates) with 2135 directed edges. Transcriptomic mapping data was then re-associated with the respective enzyme nodes prior to scoring calculations. Betweenness-centrality and overall closeness centralization indices were calculated using the igraph R package found at http://igraph.org/r/.

Metabolite Score Calculation

The substrate scoring algorithm (Fig. 3.6A) favors metabolites that are more likely acquired from the environment (not produced within the network), and will award them a higher score (Fig. 3.6B & 3.4A). The presumption of our approach was that enzymes that were more highly transcribed were more likely to utilize the substrates they act on due to coupled bacterial transcription and translation. If a compound was more likely to be produced, the more negative the resulting score would be. To calculate the score of a given metabolite (m), we used rarefied transcript abundances mapped to

respective enzyme nodes. This was represented by to and ti to designate if an enzyme created or utilized m. The first step was to calculate the average expression of enzymes for reactions that either created a given metabolite (i) or consumed that metabolite (ii). For each direction, the sum of transcripts for enzymes connecting to a metabolite were divided by the number of contributing edges (eo or ei) to normalize for highly connected metabolite nodes. Next the raw metabolite score was calculated by subtracting the creation value from the consumption value to weight for metabolites that are likely acquired exogenously. The difference was log2 transformed for comparability between scores of individual metabolites. This resulted in a final value that reflected the likelihood a metabolite was acquired from the environment. Untransformed scores that already equaled to 0 were ignored and negative values were accounted for by transformation of the absolute value then multiplied by -1. These methods have been written into a single python workflow, along with supporting reference files, and is presented as bigSMALL v1.0 (BacterIal Genome-Scale Metabolic models for AppLied reverse ecoLogy) available in a public Github repository at https://github.com/mjenior/bigsmall.

Transcriptome Randomization and Probability Distribution Comparison

As sequencing replicates of *in vivo* transcriptomes was not feasible, we applied a Monte Carlo style simulation to distinguish calculated metabolite scores due to distinct transcriptional patterns for the environment measured from those metabolites that were constitutively scored at the extremes of the scale. We employed a 10,000-fold

bootstrapping approach of randomly reassigning transcript abundance for enzyme nodes and recalculating metabolite scores. This approach was chosen over fitting a simulated transcriptome to a negative binomial distribution because it created a more relevant standard of comparison for lower coverage sequencing efforts. Using this method, each substrate node accumulated a random probability distribution of metabolite scores which were then used to calculate the median and confidence interval to generate a probability for each metabolite score to be the result of more than chance. This was a superior approach to switch randomization since the connections of the network itself was created through natural selection and any large-scale alterations would yield biologically uninformative comparisons (52).

Anaerobic in vitro *C. difficile* growth curves

The carbon-free variation of *C. difficile* Basal Defined Medium (NCMM) was prepared as previously described (6). Individual carbohydrate sources were added at a final concentration of 5 mg/mL and pair-wise carbohydrate combinations were added at 2.5 mg/mL each (5 mg/mL total). A solution of the required amino acids was made separately and added when noted at identical concentrations to the same study. 245 µl of final media mixes were added to a 96-well sterile clear-bottom plate. A rich media growth control was also included, consisting of liquid Brain-Heart Infusion with 0.5% cysteine. All culturing and growth measurement were performed anaerobically in a Coy Type B Vinyl Anaerobic Chamber (3.0% H, 5.0% CO_2, 92.0% N, 0.0% O_2). *C. difficile* str. 630 was grown for 14 hours at 37° C in 3 mL BHI with 0.5% cysteine. Cultures were

then centrifuged at 2000 rpm for 5 minutes and resulting pellets were washed twice with sterile, anaerobic phosphate-buffered saline (PBS). Washed pellets were resuspended in 3 mL more PBS and 5 µl of prepped culture was added the each growth well of the plate containing media. The plate was then placed in a Tecan Sunrise plate reader. Plates were incubated for 24 hours at 37° C with automatic optical density readings at 600 nm taken every 30 minutes. OD_{600} values were normalized to readings from wells containing sterile media of the same type at equal time of incubation. Growth rates and other curve metrics were determined by differentiation analysis of the measured OD_{600} over time in R to obtain the slope at each time point.

Quantification of in vivo metabolite relative concentrations

Metabolomic analysis performed by Metabolon (Durham, NC), a brief description of their methods is as follows. All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer at 35,000 mass resolution. Samples were dried then reconstituted in solvents compatible to each of the four methods. The first, in acidic positive conditions using a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). The second method was identical to the first but was chromatographically optimized for more hydrophobic compounds. The third approach utilized a basic negative ion optimized conditions using a separate dedicated C18

column. Basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. Samples were then analyzed via negative ionization following elution from a hydrophilic interaction chromatography column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS n scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Library matches for each compound were checked for each sample and corrected if necessary. Peaks were quantified using area under the curve.

Statistical methods

All statistical analyses were performed using R (v.3.2.0). Significant differences between community structure of treatment groups from 16S rRNA gene sequencing were determined with AMOVA in the mothur software package. Significant differences of Inv. Simpson diversity, cfu, toxin titer, and metabolite concentrations were determined by Wilcoxon signed-rank test with Benjamini-Hochberg correction. Undetectable points used half the limit of detection for all statistical calculations. Significant differences for growth curves compared to no carbohydrate control (+ amino acids) were calculated using 1-way ANOVA with Benjamini-Hochberg correction.

Results

Levels of *C. difficile* sporulation and toxin activity vary among different microbiomes

Conventionally-reared SPF mice were treated with either streptomycin, cefoperazone, or clindamycin (Table 2.1 and Fig. 2.1). These antibiotics were selected because they each have distinct and significant impacts on the structure of the cecal microbiome (Fig. 3.1A and 3.1B). We challenged the antibiotic treated mice and germfree (ex-GF) mice with *C. difficile* stain 630 to understand the pathogen's physiology with and without other microbiota. This toxigenic strain of *C. difficile* was chosen for its moderate clinical severity in mouse models (24) and well-annotated genome (25). After infection, we measured sporulation and toxin production at 18 hours post inoculation. That time point corresponded with when another laboratory strain of *C. difficile* reached its maximum vegetative cell density in the cecum with limited sporulation (26). There was not a significant difference in the number of vegetative *C. difficile* cells in the ceca of mice pretreated with any of the three antibiotics (Fig. 3.2A). All antibiotic treated and ex-GF mice were colonized to ~1×10^8 colony forming units (cfu) per gram of cecal content, while untreated mice maintained colonization resistance to *C. difficile* (Fig. 3.2A). Despite having the same number of vegetative *C. difficile* cells, more spores were detected in ex-GF mice than in the antibiotic pretreated mice (p = 0.003, 0.004, and 0.003; Fig. 3.2B). There was also a significantly higher toxin titer in ex-GF animals than any other colonized group (all $p < 0.001$), with slight variation

between antibiotic pretreatment groups (Fig. 3.2C). These results showed that *C. difficile* colonized different communities to consistently high levels, but had subtle variation in sporulation and toxin activity between distinct antibiotic-pretreated environments. As activation of both traits has been linked to recognition of distinct nutrient source concentrations in the environment, we hypothesized that *C. difficile* was utilizing different growth substrates across the conditions tested. To investigate the physiology of *C. difficile* when colonizing distinct susceptible gut environments, we performed whole transcriptome analysis of *C. difficile* from the cecal content of the same mice used in cfu and toxin titer analyses.

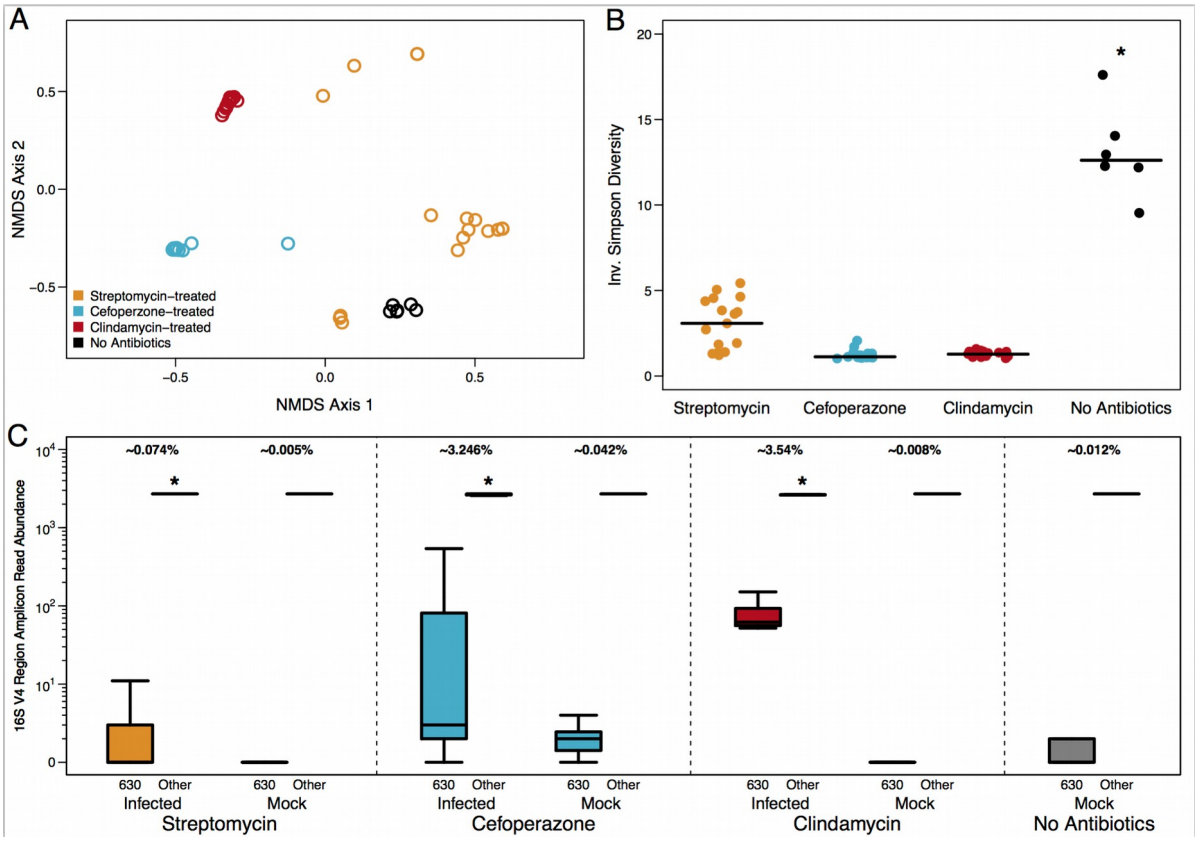

Figure 3.1 | Analysis of bacterial community structure resulting from antibiotic treatment. Results from 16S rRNA gene amplicon sequencing from bacterial communities of cecal content in both mock-infected and *C. difficile* 630-infected animals 18 hours post-infection across pretreatment models. **(A)** Non-metric multidimensional scaling (NMDS) ordination based on ThetaYC distances for the gut microbiome of all SPF mice used in these experiments (n = 36). All treatment groups are significantly different from each other groups by AMOVA ($p < 0.001$). **(B)** Inverse Simpson diversity for each cecal community from the mice in (A). Cecal communities from mice not treated with any antibiotics are significantly more diverse than any antibiotic-pretreated condition ($p < 0.001$). **(C)** Representation of 16S amplicon reads contributed by *C. difficile* in each sequenced condition compared to the total bacterial community. The percents listed at the top of each group is the proportion of the total community represented by *C. difficile*. Significantly less were for *C. difficile* were detected in each condition ($p < 0.001$).

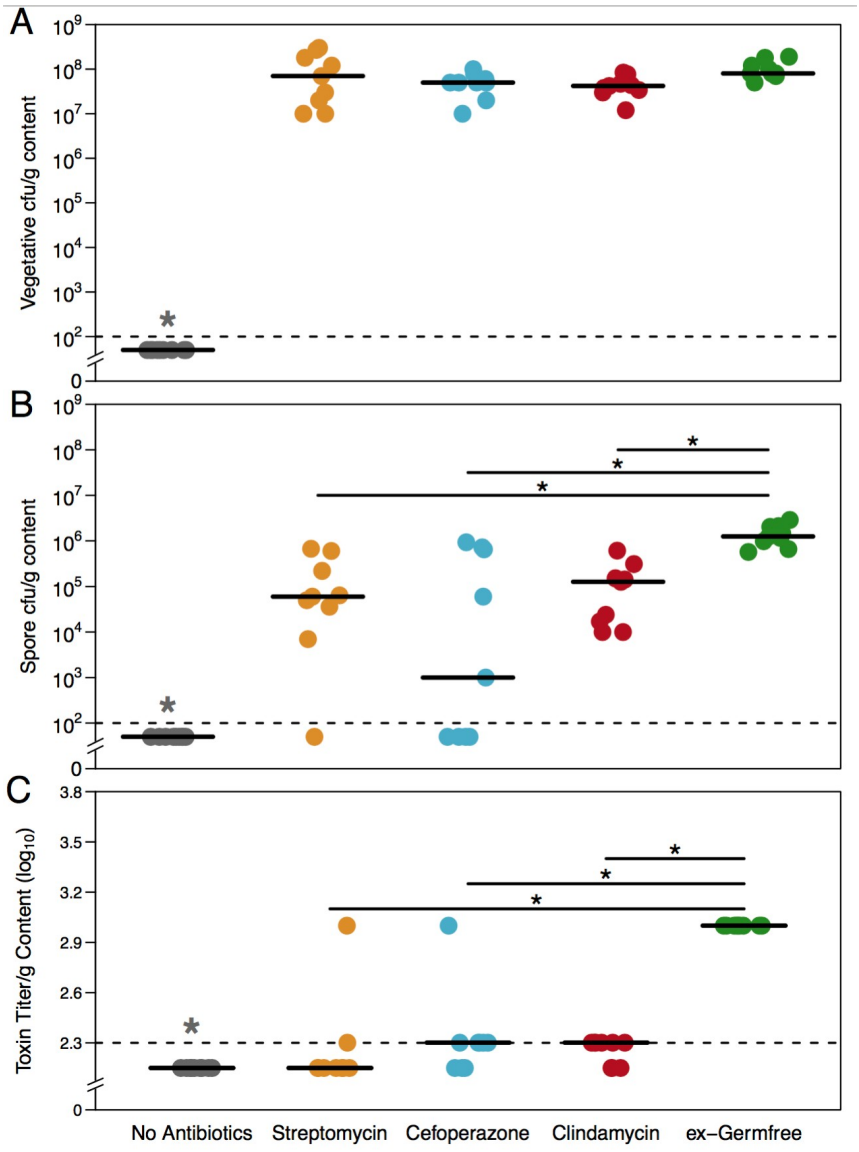

Figure 3.2 | Gut environment context affects *C. difficile* sporulation and toxin activity. Quantification of spore cfu and toxin titer from cecal content of infected mice (n = 9 per group). **(A)** Vegetative *C. difficile* cfu per gram of cecal content (p = n.s.). **(B)** *C. difficile* spore cfu per gram of cecal content. **(C)** Toxin titer from cecal content measured by activity in Vero cell rounding assay. Dotted lines denote limits of detection (LOD). Values for undetectable points were imputed as half the LOD for calculation of significant differences. Significance (p < 0.05), denoted by single asterisk, was determined with Wilcoxon signed-rank test with Benjamini–Hochberg correction.

C. difficile alters its gene expression pathways when colonizing distinct antibiotic-pretreated environments

Utilizing aliquots from the same mice in the previous assays, we attempted to measure differential expression of specific genes associated with *in vivo* phenotype changes reported in previous studies. Microarray-based gene expression measurement was not a viable alternative to sequencing as the amount of background orthologous transcription from other bacterial species would contribute greatly to non-specific binding and bias the true *C. difficile* signal, therefore we employed an RNA-Seq based approach to quantify transcription. As *C. difficile* represented a small percentage of the community in each colonized environment (Fig. 3.1C), making it impossible to sequence the transcriptome of individual mice due to the depth required to sufficiently sample the transcripts of *C. difficile*. This required the generation of a single transcriptome per condition using pooled mRNA from all mice within each pretreatment group. Following sequencing, read curation, and stringent mapping to *C. difficile* str. 630 genes (Materials & Methods) we implemented two steps of abundance normalization to compare expression between groups. Transcript abundances for each target gene were first corrected to both read length and target gene length, which resulted in an average per-base expression level for each. Adjusted values were then down-sampled to the same total read abundance for each mapping effort, allowing for even comparison between the conditions. Additionally, before proceeding with the analysis we did and assessed variation in expression of select bacterial housekeeping genes across treatment groups (Fig. 3.3A). Due to the heterogeneity of *C. difficile* reference genes across strains (27),

DNA gyrase subunit A (GyrA), threonyl-tRNA synthetase (ThrS), and ATP-dependent Clp protease (ClpP) were chosen because of their conservation across bacterial phyla and have been commonly utilized as standards of comparison for numerous transcriptional studies (14, 28, 29). Consistent expression for each of the housekeeping genes was observed across treatments, which supported that our results were more likely to be a true reflection of *C. difficile* expression *in vivo*.

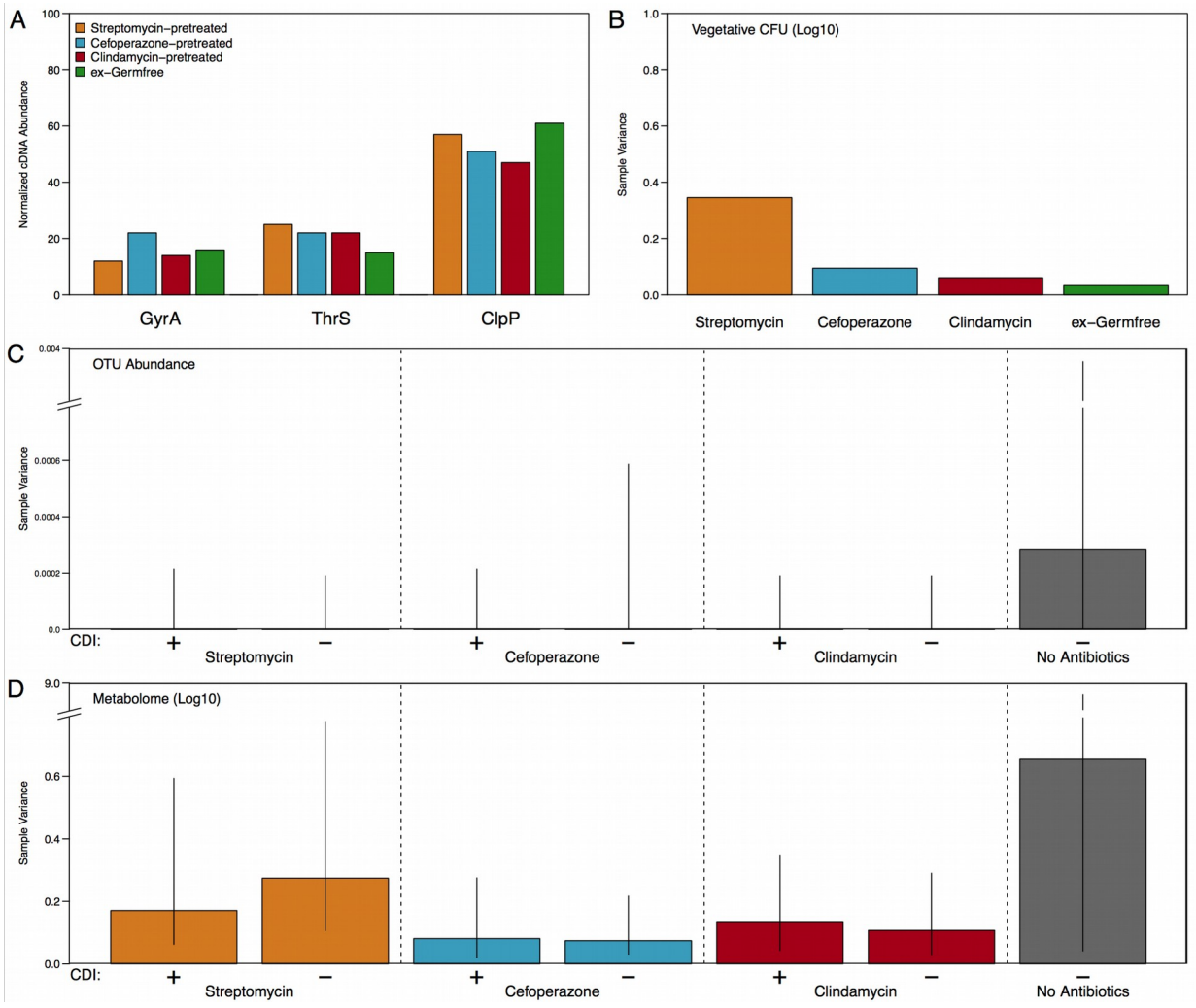

Figure 3.3 | Levels of within-group variation across datasets generated for this study. (A) Normalized transcript abundance of select housekeeping and central metabolism genes. (I) Housekeeping genes; DNA gyrase subunit A (GyrA), threonyl-tRNA synthetase (ThrS), and ATP-dependent Clp protease (ClpP).(II) Genes in separate metabolic pathways that contribute to input substrate score; enolase, glycine reductase (GrdA), and D-proline reductase (PrdA). **(B)** Median sample variance for vegetative *C. difficile* cfu from each colonized condition. **(C)** Median and interquartile range of the sample variance of OTU abundances from 16S rRNA gene sequencing, sample variances for each OTU were calculated individually prior to summary statistic calculations. **(D)** Median and interquartile range of the sample variance of Scaled intensities from untargeted metabolomic analysis, sample variances for each metabolite were in the same fashion as with OTU abundances. Data (other than transcriptomic results) was collected from the same nine animals per group were (n = 9).

Our initial transcriptomic analysis focused on genes involved in sporulation, toxin production, quorum sensing, and metabolite-regulated sigma factors (Fig. 3.4). Despite large-scale differences between pretreatment groups, no clear trends were evident between gene expression and colonization, sporulation, or toxin production. This further indicated that *C. difficile* adapted its metabolism to the environment that it colonized. As such, we next focused on specific groups of genes known to contribute to *C. difficile* metabolism (Fig. 3.6A). Genes involved in amino acid catabolism, including those that encoded enzymes involved in Stickland fermentation and general peptidases, had the highest level of expression. Stickland fermentation refers to the coupled fermentation of amino acid pairs in which one is deaminated and the other is reduced to ultimately generate ATP (30). This suggested that *C. difficile* catabolized environmental amino acids during infection, regardless of the structure of the surrounding community. Although there were gene categories that were equally expressed across conditions in spite of the community differences, there were patterns of expression for certain gene families and specific genes that were distinct to each antibiotic pretreatment. In mice pretreated with cefoperazone, *C. difficile* tended to have more expression of genes in the ABC sugar transporter and sugar alcohol catabolism (e.g. mannitol) families and fewer genes in the PTS transporter family than the other pretreatment groups. In mice pretreated with clindamycin, *C. difficile* tended to have higher expression of genes from disaccharide catabolism (e.g. beta-galactosidases and trehalose/maltose/cellibiose hydrolases), fermentation product metabolism (including consumption or production of acetate, lactate, butyrate, succinate, ethanol, and butanol), and PTS transporter

families. Genes from the sugar alcohol catabolism and ABC sugar transporter families were not highly expressed in the clindamycin-pretreated mice. Finally, in mice pretreated with streptomycin, *C. difficile* had higher levels of expression of genes from the sugar alcohol catabolism (e.g. sorbitol) and PTS transporter families. Combined, these results suggested that while catabolism of amino acids and specific carbohydrates are core components of the *C. difficile* nutritional strategy during infection, *C. difficile* adapted its metabolism across different susceptible environments.

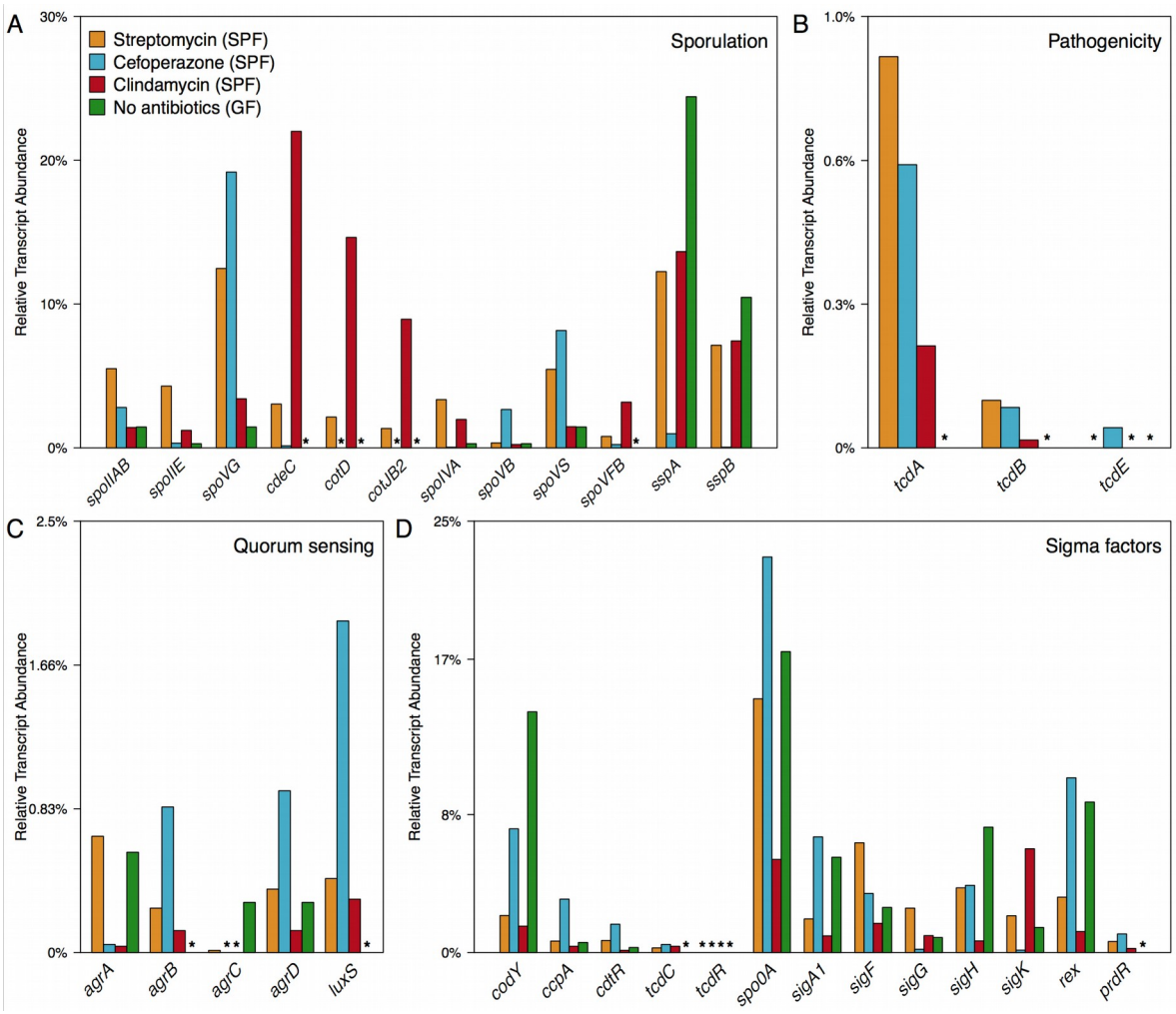

Figure 3.4 | Select *C. difficile* gene set expression compared between treatment group. Relative abundances of *C. difficile* transcript for specific genes of interest. **(A)** Transcription for select genes from the *C. difficile* sporulation pathway with the greatest variation in expression between the conditions tested. **(B)** Relative abundances of transcript for genes that encode effector proteins from the *C. difficile* pathogenicity locus. **(C)** Transcript abundances for genes associated with quorum sensing in *C. difficile*. **(D)** Transcript relative abundance of select sigma factors which expression or activity is influenced by environmental metabolite concentrations. Asterisks (*) indicate genes from which transcript was undetectable.

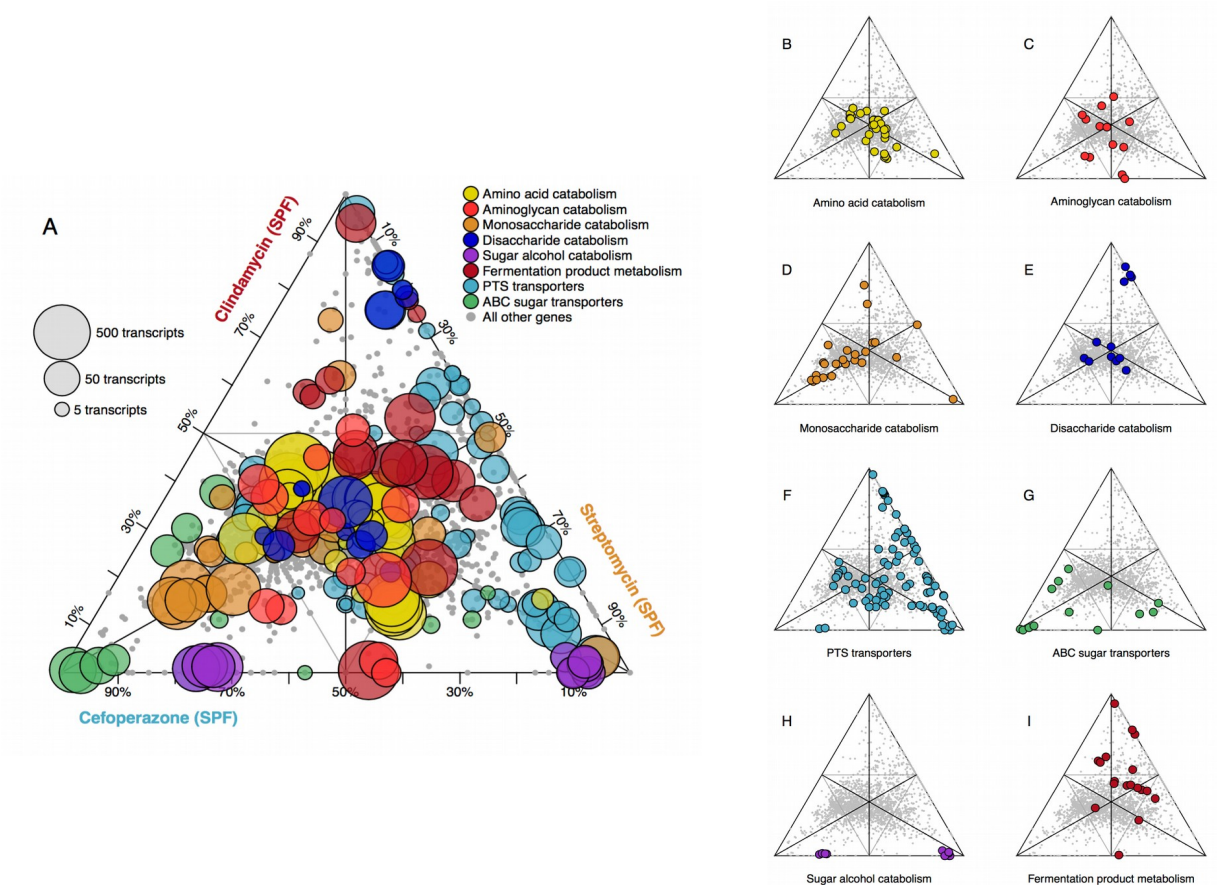

Figure 3.5 | *C. difficile* alters expression metabolic pathways between antibiotic pretreatment models. Each point in the ternary plot represents a unique gene from the annotated genome of *C. difficile* str. 630. Position reflects the ratio of median rarefied transcript abundance for that gene between the three colonized antibiotic pretreatment models. Genes from specific metabolic pathways of interest are labeled and transcription from all other genes are shown in gray. **(A)** Size of highlighted points is relative to the largest transcript abundance among the antibiotic pretreatments for each gene. Categories of metabolism are displayed separately in **(B-I)**.

Genome-scale metabolic model structure underscores known *C. difficile* physiology

To further investigate which metabolites were differentially utilized between conditions, we created a generalizeable tool to generate *de novo* genome-enabled directed, bipartite metabolic models of bacterial species using KEGG gene and biochemical reaction annotations. We implemented this for *C. difficile* str. 630 shown in Fig. 3.6A, with enzymes and metabolites were represented by nodes, and their interactions by directed connecting edges. The *C. difficile* str. 630 network we created contained a total of 447 enzymes and 758 metabolites, with 2135 directed edges. To validate our metabolic network, we analyzed network topology by calculating two metrics of centrality, betweenness centrality (BC) and closeness centrality (CC), to determine which nodes are critical to the structure of the metabolic network and if these patterns reflect known biology. Both metrics utilize shortest paths, which refer to fewest possible number of network connections that lie between two given nodes. The BC of each node is the fraction of shortest paths that pass through that node and connect all other potential pairs of nodes. In biological terms, this refers to the amount of influence a given hub has on the overall flow of metabolism (31). Similarly, CC is the reciprocal sum of the lengths of shortest paths included in each node's BC. This value demonstrates how essential a given node is to the overall structure of the metabolic network (32). Metabolic network structural studies of Escherichia coli have found that metabolites with the highest centrality calculations are involved in fundamental processes in metabolism, namely glycolysis and the citrate acid cycle pathway (33). As

such, these metrics allow for assessment of the degree to which a metabolic network accurately depicts established principles of bacterial metabolism.

Following application of both methods, we found 5 enzymes that were shared between the top 10 enzymes from BC and CC calculations (2-dehydro-3-deoxyphosphogluconate aldolase, aspartate aminotransferase, pyruvate-flavodoxin oxidoreductase, formate C-acetyltransferase, and 1-deoxy-D-xylulose-5-phosphate synthase). These enzymes primarily participate in core processes including glycolysis, the pentose phosphate pathway, or the citric acid cycle. Upon analysis of the other 15 high-scoring enzymes combined from BC and CC analyses, the majority were also components of the previously mentioned pathways, as well as several for the metabolism of amino acids. Similarly, the intersection of those substrates with high both BC and CC values revealed 6 metabolites as central nodes to the metabolism of *C. difficile* (pyruvate, acetyl-CoA, 2-oxoglutarate, D-4-hydroxy-2-oxoglutarate, D-glyceraldehyde 3-phosphate, and L-glutamate). Not only are these members of glycolysis and the citric acid cycle, but pyruvate, acetyl-CoA, and L-glutamate contribute to numerous intracellular pathways as forms of biological "currency" (33). Notably absent from the most well-connected metabolites were molecules like ATP or NADH. Their exclusion is likely a byproduct of the KEGG LIGAND reference used for network construction, which excludes cofactors from most biochemical reactions. While this may be a limitation of certain analyses, our study was not affected as the primary interest was in those substrates acquired from the environment. These results reflected the

defined biological patterns of *C. difficile* and was therefore a viable platform to study metabolism of the pathogen.

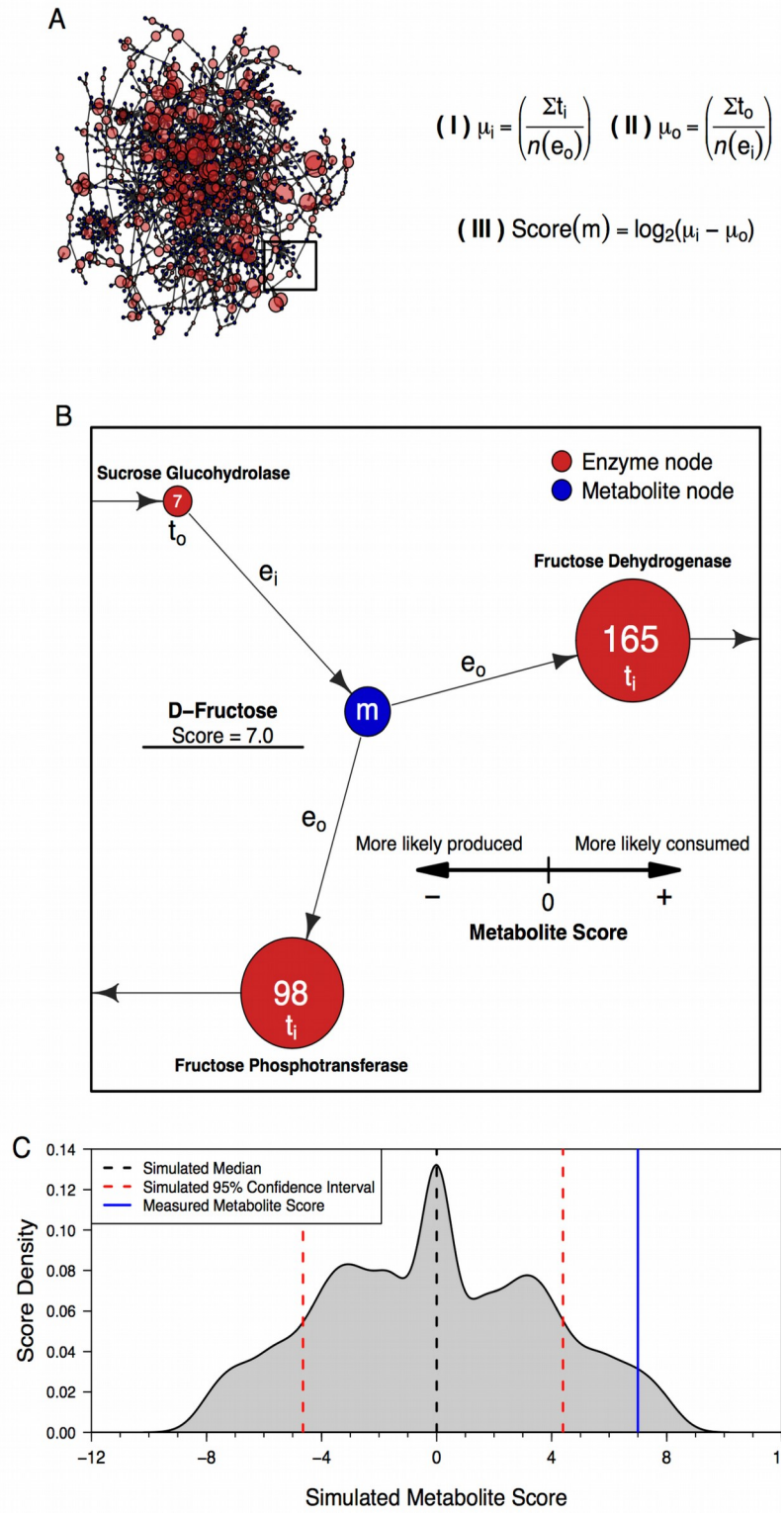

Figure 3.6 | *C. difficile* str. 630 genome-enabled bipartite metabolic network architecture and transcriptomic-enabled metabolite score calculation.

Figure 3.6 | *C. difficile* str. 630 genome-enabled bipartite metabolic network architecture and transcriptomic-enabled metabolite score calculation. (A) Largest component from the bipartite GEM of *C. difficile* str. 630. Enzyme node sizes reflect the levels of detectable transcript from each gene. Metabolite score algorithm components: (I) average transcription of reactions consuming a metabolite, (II) average transcription of reactions producing a metabolite, and (III) difference of consumption and production. **(B)** The expanded window displays a partial example of D-fructose score calculation. Values in the red nodes represent normalized transcript reads mapping to enzymes. **(C)** Example 10000-fold Mont-Carlo simulation results corresponding to a significant metabolite score for **m**.

Metabolite score algorithm reveals adaptive nutritional strategies of *C. difficile* during infection of distinct environments

We next sought to include the transcriptomic results into the metabolic model to infer which metabolites *C. difficile* most likely utilized from a given environment. To accomplish this we mapped normalized transcript abundances to the enzyme nodes in the network. Similar approaches have been previously successful in demonstrating that transcript abundance data can be utilized through the lense of genome-scale metabolic networks to accurately predict microbial metabolic responses to environmental pertubation and identify reporter metabolites of changes (34). In our system, the score of each metabolite was measured as the log2-transformed difference in average transcript levels of enzymes that use the metabolite as a substrate and those that generate the metabolite as a product (Fig. 3.6B). A metabolite with a high score was more likely obtained from the environment because the expression of genes for enzymes that produce the metabolite were low. It is important to note here that molecules that are more likely produced in our model are not necessarily likely to be released to the environment. Our models do not include the synthesis of large macromolecules (ie. long polypeptides or cytoskeleton) and should therefore only be utilized to consider input metabolites to a network. Due to the previously mentioned limited technical replication of sequencing efforts, we adopted a Monte Carlo-style simulation for iterative random transcriptome comparison to provide statistical validation of our network-based findings. This process generated random score distributions for each metabolite node in the network, which made it possible to calculate a confidence

interval that represented random noise for each metabolite. This ultimately allowed for assessment of the probability that a given metabolite was excluded from the associated null distribution (Fig. 3.6C).

To identify the core metabolites that were most essential for *C. difficile* growth, regardless of the environment, we cross-referenced the 40 highest scoring metabolites from each treatment group (Fig. 3.7A). Aminoglycan N-acetylglucosamine (GlcNAc) was found to the have the highest median score of all shared metabolites, which has been shown to be a readily available source of carbon and nitrogen which can be limiting in the gut (21). We went on to confirm that our strain of *C. difficile* could metabolize GlcNAc for growth (Fig. 3.7B) in *C. difficile* minimal media (35). The Stickland fermentation acceptor proline was also found to be scored highly in all conditions tested (36). *C. difficile* is auxotrophic for not only proline, but also cysteine, leucine, isoleucine, tryptophan, and valine, which prevented testing for in vitro growth changes on proline despite providing for modest growth in the no carbohydrate control. Previous analysis of *C. difficile* colonizing GF mice under mono-associated conditions indicated that *C. difficile* uses both sets of metabolites (21); however, use of these metabolites in the context of a complex community of potential competitors has not been observed previously. This analysis indicated that these metabolites might be an integral component of the nutrient niche for *C. difficile*.

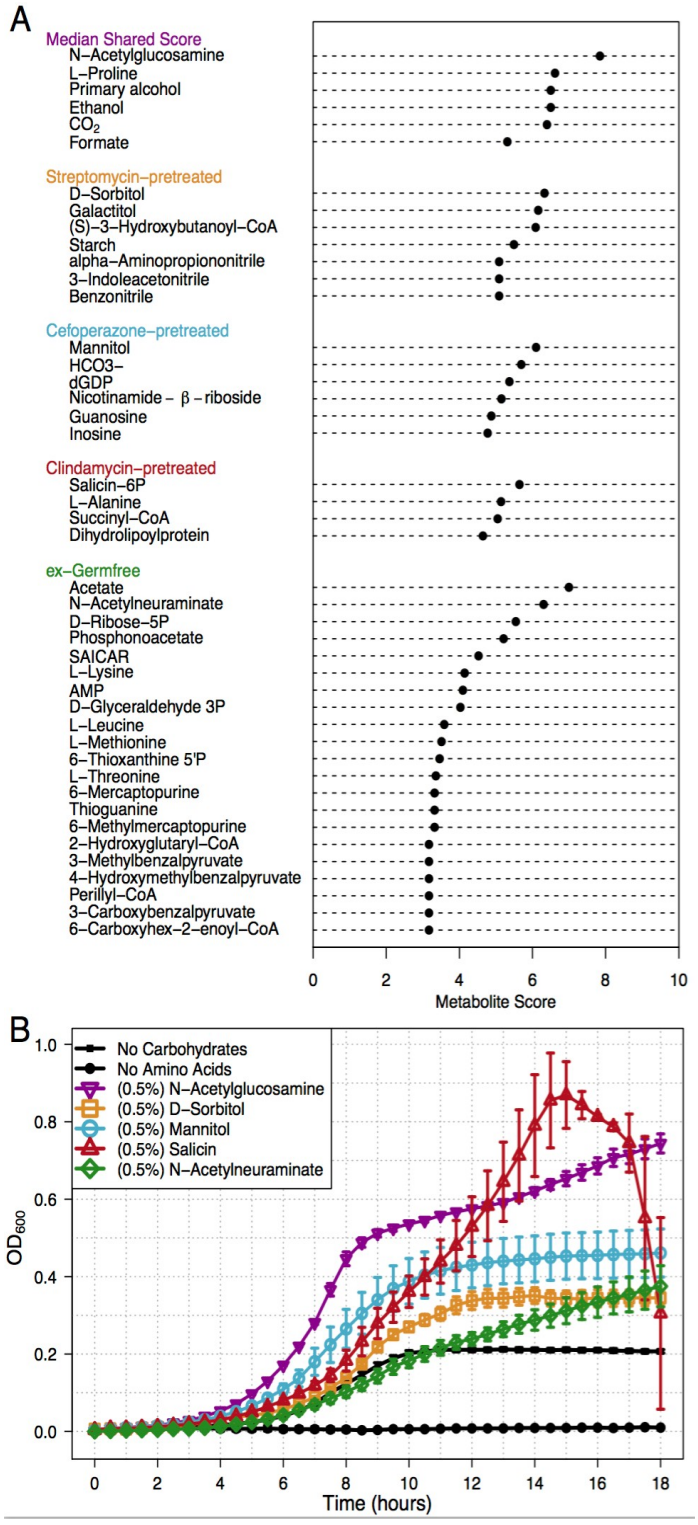

Figure 3.7 | Metabolic network analysis reveals differential carbon source utilization by *C. difficile* across infections.

Figure 3.7 | Metabolic network analysis reveals differential carbon source utilization by *C. difficile* across infections. Reported metabolite scores were calculated to have <2.5% probability to be included in the associated random score distribution. Analysis was performed using the 40 highest scoring metabolites from each condition. (A) Shared metabolite score represents the median score of metabolites that were consistently scored highly among all infected conditions. Below the conserved patterns, are shown the distinct metabolites for each group's subset. (B) 18 hour *C. difficile* str. 630 *in vitro* growth validating substrates from network analysis. All statistical comparison was performed relative to no carbohydrate control (all $p < 0.001$). Significance was determined with one-way ANOVA with Benjamini–Hochberg correction.

In vivo metabolomic analysis supports that *C. difficile* consumes metabolites indicated by metabolic modeling

To further validate the results of our metabolic model, we tested the effect of *C. difficile* on the metabolite pool in additional aliquots of cecal content from the antibiotic-treated and GF mice used in all previous analyses. We employed non-targeted ultra-performance liquid chromatography and mass spectrometry (UPLC-MS) to measure the relative *in vivo* concentrations of metabolites in the conditions investigated, with special attention to those highlighted by large metabolite scores. We tested whether the susceptible communities had significantly different concentrations of each metabolite relative to untreated SPF mice and whether the presence of *C. difficile* affected the metabolite composition.

First, we compared the relative concentration of highly scored metabolites in untreated SPF mice and antibiotic pretreated mice in the absence of CDI (Fig. 3.8). We found that the relative concentration of GlcNAc was actually significantly lower in all susceptible conditions (Fig. 3.8A; all $p < 0.001$). The Stickland fermentation acceptors proline (all $p < 0.05$) and hydroxyproline (all $p < 0.05$) were significantly higher in all susceptible environments tested (Fig. 3.8B & Fig. 2.9 B). Conversely, the Stickland donor alanine was significantly lower across all susceptible conditions (Fig. 3.9D; all $p < 0.05$). Succinyl-CoA was score most highly in clindamycin pretreatment, which is the direct precursor to succinate by succinyl-CoA transferases (37). Succinate has been shown to support *C. difficile* growth in vivo through a synergistic relationship with at least one other bacterial species (9). As succinyl-CoA was not measured in our

metabolimic assay, we instead found that succinate was indeed significantly higher in clindamycin pretreated mice (Fig. 3.9E; all $p < 0.05$). Among the cefoperazone-pretreated SPF and GF mice, we also found that mannitol/sorbitol (Fig. 3.9C), N-acetylneuraminate (Fig. 3.9F), and glycine (Fig. 3.3A) were significantly higher in cefoperazone-treated SPF and GF mice (all $p < 0.05$). These results supported the assertion that antibiotic treatment opened potential nutrient niches that *C. difficile* was able to exploit for its growth.

Second, we compared relative concentrations of high scoring metabolites during CDI and mock-infection within each pretreatment group (Fig. 3.8 & Fig. 2.9). Both groups of host-derived glycans, GlcNAc/GalNAc (Fig. 3.9A) and Neu5Ac (Fig. 3.9F), were significantly lower when in the presence of *C. difficile* in ex-GF mice ($p < 0.05$ and 0.01). In agreement with the previous results, we found that the Stickland acceptors proline (Fig. 3.9B) and hydroxyproline (Fig. 3.3C) were significantly lower in every *C. difficile* colonized environment (all $p < 0.05$). Glycine, another preferred Stickland acceptor, was lower in each condition following infection with significant change in cefoperazone-pretreated mice (Fig. 3.3D; $p < 0.05$). The Stickland donors leucine and isoleucine were significantly lower in all infected conditions except streptomycin-pretreated mice (Fig. 3.3A-B; all $p < 0.05$). Concentrations of alanine were also lower in all infected conditions compared to mock infection, however none of the changes met our threshold for significance (Fig. 3.9D). These results strongly supported the hypothesis that amino acids are a primary energy source of *C. difficile* during infection. A significant difference was seen for mannitol/sorbitol in ex-GF mice ($p < 0.01$), but not

in cefoperazone-pretreated mice (Fig. 3.9C). Although a lower the concentration of succinate in both streptomycin and clindamycin pretreated mice was observed, neither was found to be significant. Overall, metabolomic analysis supported our metabolite score algorithm for predicting the metabolites utilized by *C. difficile* during different infection conditions. Results from metabolic modeling combined with untargeted metabolomic analysis also suggested a possible hierarchy of preferred growth substrates.

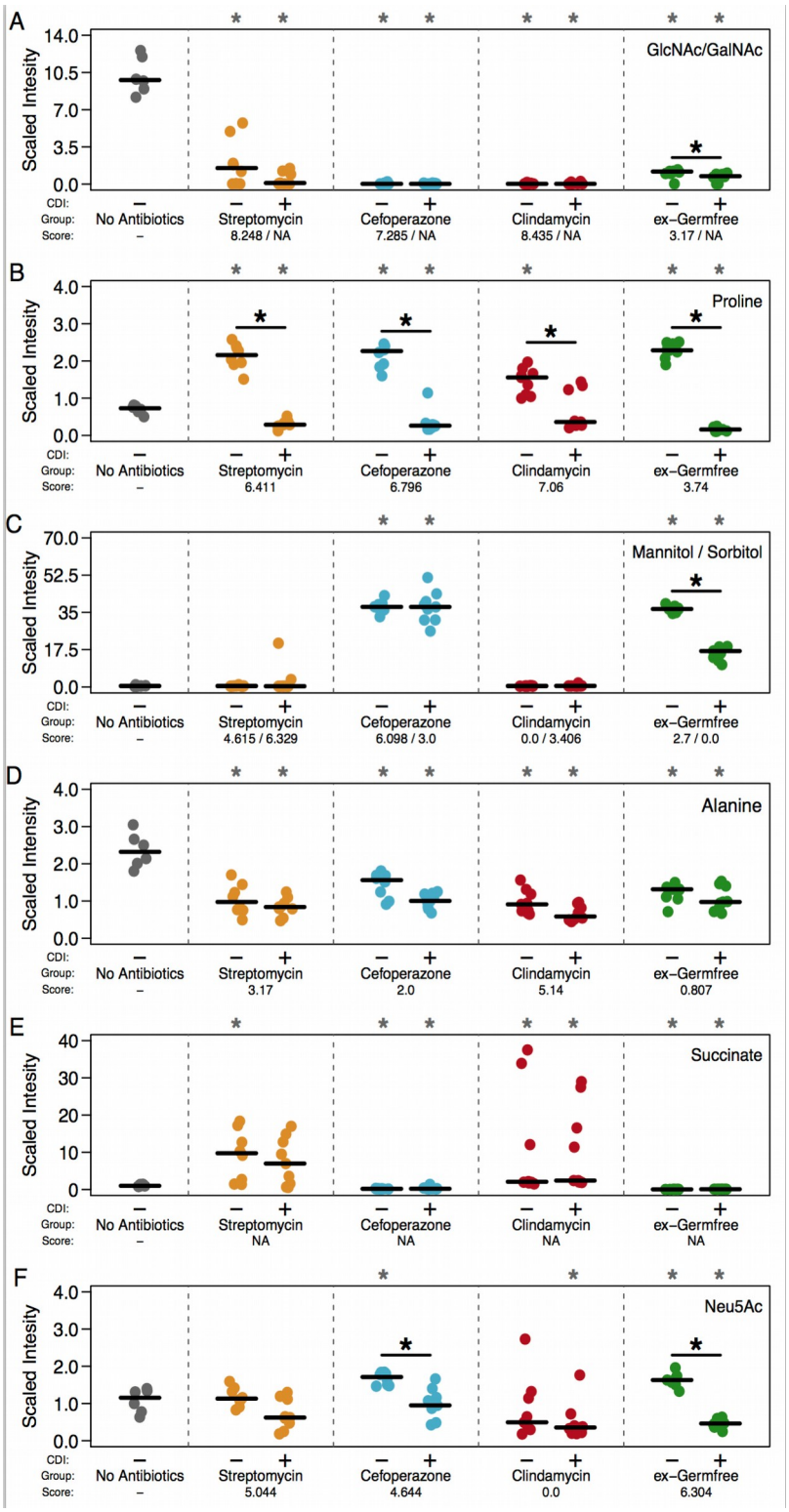

Figure 3.8 | Untargeted *in vivo* metabolomics support network-based metabolite scores and suggest nutrient preference hierarchy.

Figure 3.8 | Untargeted *in vivo* metabolomics support network-based metabolite scores and suggest nutrient preference hierarchy. Paired metabolites were quantified simultaneously as the only differ by chirality making differentiation impossible. CDI status and *C. difficile* metabolite scores during infection are indicated below each panel. NAs denote metabolites that were not included in our metabolic model of *C. difficile* str. 630. Black asterisks inside the panels represent significant differences between mock and *C. difficile*-infected groups within separate treatment groups (all $p < 0.05$). Gray asterisks along the top margin of each panel indicate significant difference from untreated SPF mice (all $p < 0.05$). Significance was determined with Wilcoxon signed-rank test with Benjamini–Hochberg correction.

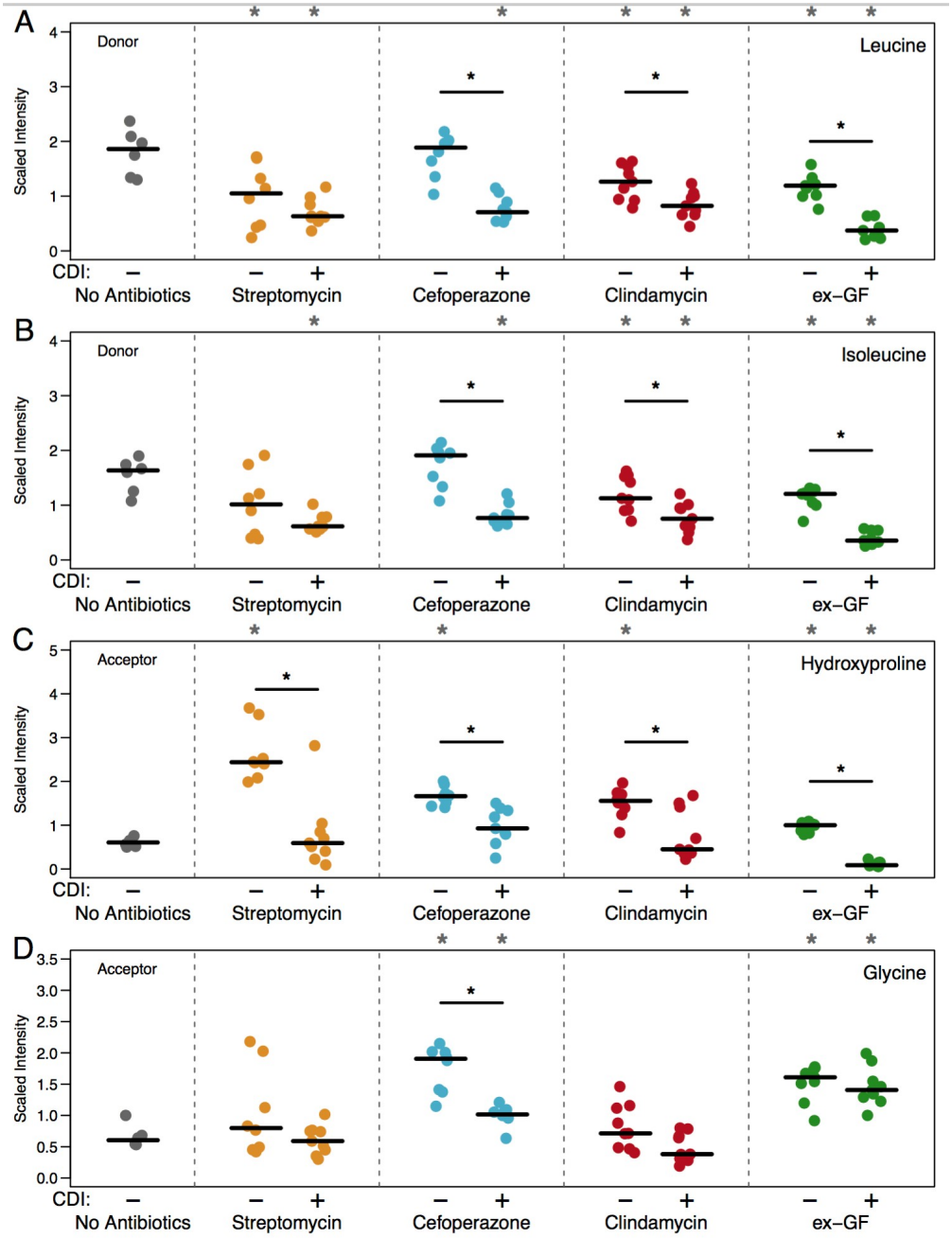

Figure 3.9 | Change in *in vivo* concentrations of additional Stickland fermentation substrates. Comparison of concentrations for other Stickland fermentation substrates from *C. difficile*-infected and mock-infected mouse cecal content 18 hours post-infection. Labels in the top left corner of each panel indicate whether the amino acid is a Stickland donor or acceptor. Black asterisks inside the panels denote significant differences between mock and *C. difficile*-infected groups within separate treatment groups (all $p < 0.05$). Gray asterisks along the top margin of each panel indicate significant difference from untreated SPF mice (all $p < 0.05$).

Discussion

Our results expand upon previous understanding of *C. difficile* metabolism during infection by showing that not only does the pathogen adapt its metabolism to life inside of a host (14, 21), but also to the context of the specific gut environment in which it finds itself. Previous transcriptomic efforts to measure the response of *C. difficile* have demonstrated *in vivo* changes in metabolism following colonization of GF mice. In this study, we utilized a conventionally-reared mouse model of infection to compare the response of *C. difficile* to colonization in the context of varied gut communities generated by pretreatment with representatives from distinct classes of antibiotics. With these models, we identified subtle differences in sporulation and toxin activity between each antibiotic-pretreated condition. Transcriptomic sequencing of *C. difficile* across colonized environments indicated complex expression patterns of genes in catabolic pathways for a variety of carbon sources. Through integration of transcriptomic data with genome-scale metabolic modeling, we were able to deconvolute these signals. This allowed us to observe that *C. difficile* likely genenrated energy by metabolizing specific alternative carbohydrates, carboxylic acids, and host-derived glycans across colonized conditions. We also found that Stickland fermentation substrates and products, as well another host-derived amino glycan N-acetylglucosamine, were consistently among the highest scoring shared metabolites which indicated that these metabolites were central to the *in vivo* nutritional strategy of *C. difficile*. To confirm our modeling-based results we employed untargeted mass spectrometry that demonstrated

greater availability of many metabolites highlighted by our algorithm in susceptible gut environs. Metabolomic analysis further revealed differential reduction of highly scored metabolites during CDI, which suggested a hierarchy for the utilization of certain growth nutrients.

An explanation for the differences seen in metabolite score and substrate availability could be the concomitant lower population density of one or more competitors for certain resources. Ex-GF mice, where no other microbial competitors are present, provided a partially controlled system of resource competition. In this condition, Neu5Ac was found to be the highest scored substrate and concentrations Neu5Ac were significantly higher in susceptible mice. The concentrations of Neu5Ac were concordantly lower in infected mice relative to mock-infected mice. The same trend was also present in cefoperazone-pretreatment, which suggested that *C. difficile* may be less competitive for this host-derived aminoglycan and may only have access when certain competitors are no longer present. In the presence of a microbiota, *C. difficile* population-level nutrient utilization patterns differed across each environment tested. For example, past studies have concluded that specific PTS and ABC transport systems are upregulated *in vivo* (14, 21), but our results indicate more complex regulation with inverse expression of the respective systems between antibiotic pretreatments (Fig. 3.5). In agreement with earlier research we found that *C. difficile* likely fermented amino acids for energy during infection of GF mice in addition to aminoglycan catabolism. Our results go on to support that this metabolic strategy was conserved across all infection conditions tested. Several Stickland substrates had

consistently higher metabolite scores including alanine, leucine, and proline indeed dropped concentration during infection (Fig. 3.9A, and S5). Fermentation of amino acids provides not only carbon and energy, but are also a source of nitrogen which is a limited resource in the mammalian lower gastrointestinal tract (38). This makes Stickland fermentation a valuable metabolic strategy, and it stands to reason that *C. difficile* would use this strategy across all environments it colonizes. This same principle may also extend to host mucus layer derived glycans as they are another source of carbon and nitrogen which, despite augmented release by members of the microbiota (39), would be present at some basal concentration regardless of other species' intercession. Finally, we did find disagreement in some metabolite scores and the difference in in vivo concentration of previously suggested *C. difficile* growth substrates between mock infected and infected mice. This may indicate a nutrient preference hierarchy during infection. Based on our results, we propose that amino acids are prized above all other substrates, followed by host glycans, then carbohydrates, sugar alcohols, or carboxylic acids depending on their availability in the environment. Since the latter provide carbon and energy, but not nitrogen, it appears that *C. difficile* metabolism strongly values nitrogen-containing carbon sources that fulfill a larger proportion of its biological requirements but this requires additional investigation to confirm.

Our systems approach to studying *C. difficile* metabolism during the infection of susceptible communities combines multiple levels of biological data to identify metabolic trends that would not be apparent by a single method. Only through integrative multi-omic analysis of *C. difficile* infection employing genomics, transcriptomics, and

metabolomics were we able to uncover a much clearer image of *C. difficile*'s nutrient niche space during infection in the context of complex microbial communities. By the nature of our algorithm's reliance on network topology, the signal contributed by those metabolites on the periphery of the network, which are more likely to be imported from the environment, was amplified. This approach could be especially useful for identifying edges of competition for nutrients between colonizing pathogens and indigenous communties of bacteria, as is the case with *C. difficile*. Our modeling platform may also allow for the identification of emergent properties for the metabolism of *C. difficile* during infection. One example could be the appearance of CO2 and formate, apparent metabolic end products, in the list of shared metabolites which scored highly across conditions. Although this may be a shortcoming of the genome or database annotation, one group has posited that *C. difficile* may actually be autotrophic under certain conditions and require both of these substrates to undergo this process (40). These findings highlight that our method not only identifies growth substrates, but reports all metabolites that are being utilized for other processes.

Several factors limited our ability to generate and interpret transcriptomic and metabolomic data. Most prominently, we were forced to pool the cecal contents of multiple animals to generate a sufficient quantity of high quality RNA that would permit us to sample the transcriptome of a rare member of the microbiome. Due to possible biological variation between samples that could be masked by this approach, we quantified within-group sample variation for vegetative CFU, 16S rRNA gene abundance, and untargeted mass-spectrometry (Fig. 3.3C-D). This revealed extremely

low variablity in each treatment group at multiple levels of biology, and since these data were collected using matched cecal samples, we were more confident that our transcriptomic results reflected reality. Metabolomic comparisons were also complicated by the fact that multiple organisms contribute to the metabolite pool. The metabolic patterns of the other species in each system (host and microbe) could instead be altered by pathogen colonization. As the concentrations of metabolites in our untargeted assay were reported in relative terms, it was difficult to discern whether the available biomass of *C. difficile* reaches a level to create these differences on its own. Possible limitations of our modeling approach also existed, despite much of our results being consistent with previously published work and our own untargeted metabolomic analysis. Ultimately, the metabolite score calculation is dependent on correct and existing gene annotation. In this regard it has been shown that the pathway annotations in KEGG are robust to missing elements (41), however this does not completely eliminate the possibility for this type of error. Due to the topology of the metabolic network, we were also unable to integrate stoichiometry for each reaction which may effect rates of consumption or production. Reaction reversibility also varies depending on versions of enzymes possessed by each species. Since our algorithm favorably weights those metabolites closer to the network periphery, incorrect directionality annotations may lead to mislabeling reactants or products and potentially lead to incorrect metabolite score calculations. With additional manual curation of the *C. difficile* metabolic network, more species specific discoveries can eventually be made. Even with this possibility, the application of multiple methods to study the altered physiology

of *C. difficile* in mock-infected and infected communities allowed us to validate our results based on known elements of *C. difficile* biology and to internally cross validate the novel results from our experiments. Ultimately, these results combine to underscore predictions of nutrient niche plasticity.

Our combined genomic, transcriptomic, and metabolomic analysis showed that when infecting diverse host-associated gut environments, *C. difficile* optimized its nutrient utilization profile to each gut environment and effectively colonize the host. Focusing on previously established metabolic capabilities of the pathogen, we identify that these forms of metabolism are differentially important to *C. difficile* when colonizing distinct environments. These results have implications for the development of targeted measures to prevent *C. difficile* colonization through pre- or probiotic therapy. In the future, this systems-level approach could be easily expanded to study the niche landscape of entire communities of bacteria in response to antibiotic perturbation or pathogen colonization.

References

1. **Lessa, F. C.**, **C. V. Gould**, and **L. C. McDonald**. 2012. Current status of *Clostridium difficile* infection epidemiology. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **55 Suppl 2**:S65–70.

2. **Lessa, F. C.**, **Y. Mu**, **W. M. Bamberg**, **Z. G. Beldavs**, **G. K. Dumyati**, **J. R. Dunn**, **M. M. Farley**, **S. M. Holzbauer**, **J. I. Meek**, **E. C. Phipps**, **L. E. Wilson**, **L. G. Winston**, **J. a Cohen**, **B. M. Limbago**, **S. K. Fridkin**, **D. N. Gerding**, and **L. C. McDonald**. 2015. Burden of *Clostridium difficile* Infection in the United States. The New England Journal of Medicine **372**:825–834.

3. **Leffler, D. A.**, and **J. T. Lamont**. 2015. *Clostridium difficile* Infection. New England Journal of Medicine **372**:1539–1548.

4. **Britton, R. A.**, and **V. B. Young**. 2014. Role of the intestinal microbiota in resistance to colonization by *Clostridium difficile*. Gastroenterology **146**:1547–1553.

5. **Chen, X.**, **K. Katchar**, **J. D. Goldsmith**, **N. Nanthakumar**, **A. Cheknis**, **D. N. Gerding**, and **C. P. Kelly**. 2008. A Mouse Model of *Clostridium difficile*-Associated Disease. Gastroenterology **135**:1984–1992.

6. **Theriot, C. M.**, **M. J. Koenigsknecht**, **P. E. C. Jr**, **G. E. Hatton**, **A. M. Nelson**, **B. Li**, **G. B. Huffnagle**, **J. Li**, and **V. B. Young**. 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. Nat Commun **3114**.

7. **Schubert, A. M.**, **H. Sinani**, and **P. D. Schloss**. 2015. Antibiotic-induced alterations of the murine gut microbiota and subsequent effects on colonization resistance against *Clostridium difficile*. mBio **6**.

8. **Antunes, L. C. M.**, **J. Han**, **R. B. R. Ferreira**, **P. Loli**, **C. H. Borchers**, and **B. B. Finlay**. 2011. Effect of antibiotic treatment on the intestinal metabolome. Antimicrobial Agents and Chemotherapy **55**:1494–1503.

9. **Ferreyra, J. A.**, **K. J. Wu**, **A. J. Hryckowian**, **D. M. Bouley**, **B. C. Weimer**, and **J. L. Sonnenburg**. 2014. Gut microbiota-produced succinate promotes *Clostridium difficile* infection after antibiotic treatment or motility disturbance. Cell Host and Microbe **16**:770–777.

10. **Jump, R. L. P.**, **A. Polinkovsky**, **K. Hurless**, **B. Sitzlar**, **K. Eckart**, **M. Tomas**, **A. Deshpande**, **M. M. Nerandzic**, and **C. J. Donskey**. 2014. Metabolomics analysis identifies intestinal microbiota-derived biomarkers of colonization resistance in clindamycin-treated mice. PLoS ONE **9**.

11. **Freter, R.**, **H. Brickner**, **M. Botney**, **D. Cleven**, and **A. Aranki**. 1983. Mechanisms that control bacterial populations in continuous-flow culture models of mouse large intestinal flora. Infection and Immunity **39**:676–685.

12. **Wilson, K. H.**, and **F. Perini**. 1988. Role of competition for nutrients in suppression of *Clostridium difficile* by the colonic microflora. Infection and Immunity **56**:2610–2614.

13. **Sebaihia, M.**, **B. W. Wren**, **P. Mullany**, **N. F. Fairweather**, **N. Minton**, **R. Stabler**, **N. R. Thomson**, **A. P. Roberts**, **A. M. Cerdeño-Tárraga**, **H. Wang**, **M. T. G. Holden**, **A. Wright**, **C. Churcher**, **M. a Quail**, **S. Baker**, **N. Bason**, **K. Brooks**, **T. Chillingworth**, **A. Cronin**, **P. Davis**, **L. Dowd**, **A. Fraser**, **T. Feltwell**,

Z. Hance, **S. Holroyd**, **K. Jagels**, **S. Moule**, **K. Mungall**, **C. Price**, **E. Rabbinowitsch**, **S. Sharp**, **M. Simmonds**, **K. Stevens**, **L. Unwin**, **S. Whithead**, **B. Dupuy**, **G. Dougan**, **B. Barrell**, and **J. Parkhill**. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nature genetics **38**:779–786.

14. **Kansau, I.**, **A. Barketi-Klai**, **M. Monot**, **S. Hoys**, **B. Dupuy**, **C. Janoir**, and **A. Collignon**. 2016. Deciphering adaptation strategies of the epidemic *Clostridium difficile* 027 strain during infection through in vivo transcriptional analysis. PLoS ONE **11**.

15. **Songer, J. G.**, and **M. A. Anderson**. 2006. *Clostridium difficile*: An important pathogen of food animals. Anaerobe **12**:1–4.

16. **Janvilisri, T.**, **J. Scaria**, **A. D. Thompson**, **A. Nicholson**, **B. M. Limbago**, **L. G. Arroyo**, **J. G. Songer**, **Y. T. Gröhn**, and **Y. F. Chang**. 2009. Microarray identification of *Clostridium difficile* core components and divergent regions associated with host origin. Journal of Bacteriology **191**:3881–3891.

17. **Gripp, E.**, **D. Hlahla**, **X. Didelot**, **F. Kops**, **S. Maurischat**, **K. Tedin**, **T. Alter**, **L. Ellerbroek**, **K. Schreiber**, **D. Schomburg**, **T. Janssen**, **P. Bartholomäus**, **D. Hofreuter**, **S. Woltemate**, **M. Uhr**, **B. Brenneke**, **P. Grüning**, **G. Gerlach**, **L. Wieler**, **S. Suerbaum**, and **C. Josenhans**. 2011. Closely related *Campylobacter jejuni* strains from different sources reveal a generalist rather than a specialist lifestyle. BMC Genomics **12**:584.

18. **Neumann-Schaal, M.**, **J. D. Hofmann**, **S. E. Will**, and **D. Schomburg**. 2015. Time-resolved amino acid uptake of *Clostridium difficile* 630 Delta-erm and concomitant fermentation product and toxin formation. BMC Microbiology 281.

19. **Nawrocki, K. L.**, **A. N. Edwards**, **N. Daou**, **L. Bouillaut**, and **S. M. McBride**. 2016. CodY-dependent regulation of sporulation in *Clostridium difficile*. Journal of Bacteriology **198**:2113–2130.

20. **Dineen, S. S.**, **S. M. McBride**, and **A. L. Sonenshein**. 2010. Integration of Metabolism and Virulence by *Clostridium difficile* CodY. Journal of Bacteriology **192**:5350–5362.

21. **Janoir, C.**, **C. Denève**, **S. Bouttier**, **F. Barbut**, **S. Hoys**, **L. Caleechum**, **D. Chapetón-Montes**, **F. C. Pereira**, **A. O. Henriques**, **A. Collignon**, **M. Monot**, and **B. Dupuy**. 2013. Adaptive strategies and pathogenesis of *Clostridium difficile* from *in vivo* transcriptomics. Infection and Immunity **81**:3757–3769.

22. **Matamouros, S.**, **P. England**, and **B. Dupuy**. 2007. *Clostridium difficile* toxin expression is inhibited by the novel regulator TcdC. Molecular Microbiology **64**:1274–1288.

23. **Antunes, A.**, **I. Martin-Verstraete**, and **B. Dupuy**. 2011. CcpA-mediated repression of *Clostridium difficile* toxin gene expression. Molecular Microbiology **79**:882–899.

24. **Theriot, C. M.**, **C. C. Koumpouras**, **P. E. Carlson**, **I. I. Bergin**, **D. M. Aronoff**, and **V. B. Young**. 2011. Cefoperazone-treated mice as an experimental platform to assess differential virulence of *Clostridium difficile* strains. Gut microbes **2**:326–334.

25. **Monot, M.**, **C. Boursaux-Eude**, **M. Thibonnier**, **D. Vallenet**, **I. Moszer**, **C. Medigue**, **I. Martin-Verstraete**, and **B. Dupuy**. 2011. Reannotation of the genome sequence of *Clostridium difficile* strain 630. Journal of Medical Microbiology **60**:1193–1199.

26. **Koenigsknecht, M. J.**, **C. M. Theriot**, **I. L. Bergin**, **C. A. Schumacher**, **P. D. Schloss**, and **V. B. Young**. 2015. Dynamics and establishment of *Clostridium difficile* infection in the murine gastrointestinal tract. Infection and Immunity **83**:934–941.

27. **Metcalf, D.**, **S. Sharif**, and **J. Weese**. 2010. Evaluation of candidate reference genes in *Clostridium difficile* for gene expression normalization. Anaerobe **16**:439–443.

28. **Gendron, N.**, **H. Putzer**, and **M. Grunberg-Manago**. 1994. Expression of both Bacillus subtilis threonyl-tRNA synthetase genes is autogenously regulated. Journal of Bacteriology **176**:486–494.

29. **Sjögren, L.**, and **A. Clarke**. 2011. Assembly of the Chloroplast ATP-Dependent Clp Protease in Arabidopsis Is Regulated by the ClpT Accessory Proteins. The Plant Cell **23**:322–332.

30. **Jackson, S.**, **M. Calos**, **A. Myers**, and **W. T. Self**. 2006. Analysis of proline reduction in the nosocomial pathogen *Clostridium difficile*. Journal of Bacteriology **188**:8487–8495.

31. **Potapov, A. P.**, **N. Voss**, **N. Sasse**, and **E. Wingender**. 2005. Topology of mammalian transcription networks. Genome informatics. International Conference on Genome Informatics **16**:270–278.

32. **Koschutzki, D.**, and **F. Schreiber**. 2008. Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene Regulation and Systems Biology **2008**:193–201.

33. **Ma, H. W.**, and **A. P. Zeng**. 2003. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics **19**:1423–1430.

34. **Patil, K. R.**, and **J. Nielsen**. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proceedings of the National Academy of Sciences of the United States of America **102**:2685–9.

35. **Karasawa, T.**, **S. Ikoma**, **K. Yamakawa**, and **S. Nakamura**. 1995. A defined growth medium for *Clostridium difficile*. Microbiology **141**:371–375.

36. **Aboulnaga, H.**, **O. Pinkenburg**, **J. Schiffels**, **A. El-Refai**, **W. Buckel**, and **T. Selmer**. 2013. Effect of an oxygen-tolerant bifurcating butyryl coenzyme a dehydrogenase/electron-transferring flavoprotein complex from *Clostridium difficile* on butyrate production in *Escherichia coli*. Journal of Bacteriology **195**:3704–3713.

37. **Van Grinsven, K.**, **S. Rosnowsky**, **S. Van Weelden**, **S. Pütz**, **M. Van Der Giezen**, **W. Martin**, **J. Van Hellemond**, **A. Tielens**, and **K. Henze**. 2008. Acetate:succinate CoA-transferase in the hydrogenosomes of Trichomonas vaginalis: Identification and characterization. Journal of Biological Chemistry **283**:1411–1418.

38. **Fuller, M. F.**, and **P. J. Reeds**. 1998. Nitrogen cycling in the gut. Annual review of nutrition **18**:385–411.

39. **Marcobal, A.**, **A. M. Southwick**, **K. A. Earle**, and **J. L. Sonnenburg**. 2013. A refined palate: Bacterial consumption of host glycans in the gut. Glycobiology **23**:1038–1046.

40. **Köpke, M.**, **M. Straub**, and **P. Dürre**. 2013. *Clostridium difficile* Is an Autotrophic Bacterial Pathogen. PLoS ONE **8**.

41. **Green, M. L.**, and **P. D. Karp**. 2006. The outcomes of pathway database computations depend on pathway ontology. Nucleic Acids Research **34**:3687–3697.

42. **Wilson, K. H.**, **M. J. Kennedy**, and **F. R. Fekety**. 1982. Use of sodium taurocholate to enhance spore recovery on a medium selective for *Clostridium difficile*. Journal of Clinical Microbiology **15**:443–446.

43. **Sorg, J. a.**, and **A. L. Sonenshein**. 2010. Inhibiting the initiation of *Clostridium difficile* spore germination using analogs of chenodeoxycholic acid, a bile acid. Journal of Bacteriology **192**:4983–4990.

44. **Leslie, J. L.**, **S. Huang**, **J. S. Opp**, **M. S. Nagy**, **M. Kobayashi**, **V. B. Young**, and **J. R. Spence**. 2015. Persistence and toxin production by *Clostridium difficile* within human intestinal organoids result in disruption of epithelial paracellular barrier function. Infection and Immunity **83**:138–145.

45. **Kozich, J.**, **S. Westcott**, **N. Baxter**, **S. Highlander**, and **P. Schloss**. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol **79**:5112–5120.

46. **Wang, Q.**, **G. M. Garrity**, **J. M. Tiedje**, and **J. R. Cole**. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology **73**:5261–5267.

47. **Lopez-Medina, E.**, **M. M. Neubauer**, **G. B. Pier**, and **A. Y. Koh**. 2011. RNA isolation of *Pseudomonas aeruginosa* colonizing the murine gastrointestinal tract. Journal of visualized experiments : JoVE 6–9.

48. **Martin, M. J.**, **S. Clare**, **D. Goulding**, **A. Faulds-Pain**, **L. Barquist**, **H. P. Browne**, **L. Pettit**, **G. Dougan**, **T. D. Lawley**, and **B. W. Wren**. 2013. The *agr* locus regulates virulence and colonization genes in *Clostridium difficile* 027. Journal of Bacteriology **195**:3672–3681.

49. **Langmead, B.**, **C. Trapnell**, **M. Pop**, and **S. L. Salzberg**. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 1–10.

50. **Ogata, H.**, **S. Goto**, **K. Sato**, **W. Fujibuchi**, **H. Bono**, and **M. Kanehisa**. 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research **27**:29–34.

51. **Li, H.**, **B. Handsaker**, **A. Wysoker**, **T. Fennell**, **J. Ruan**, **N. Homer**, **G. Marth**, **G. Abecasis**, and **R. Durbin**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**:2078–2079.

52. **Basler, G.**, **O. Ebenhöh**, **J. Selbig**, and **Z. Nikoloski**. 2011. Mass-balanced randomization of metabolic networks. Bioinformatics **27**:1397–1403.

53. **Bonett, D. G.**, and **R. M. Price**. 2002. Statistical inference for a linear function of medians: confidence intervals, hypothesis testing, and sample size requirements. Psychological methods **7**:370–383.

CHAPTER FOUR

Discussion

The perturbed gut ecosystem following cessation of antibiotic treatment has a diminished capacity to prevent the colonization of the bacterial pathogen *Clostridium difficile*. This has been demonstrated to be primarily mediated through the intercession of the gut microbiota. Healthy communities of bacteria in the gastrointestinal tract are able to prevent *C. difficile* colonization. This process is thought to be mediated through a number of proposed mechanisms including exclusion from spatial niches, direct antagonism through products like bacteriocins, indirect effects through cross-talk with the host's immune system, and competition for desired subsets of growth nutrients (1). Several studies have supported the hypothesis that the elimination of preferred nutrient niches by a resistant microbiota may be the strongest explanation of this relationship (2–4). With the reduction in competitor populations due to the antibiotic, *C. difficile* gains access to those nutrients that would otherwise be inaccessible, allowing the bacterium to colonize, outgrow, and ultimately cause it’s hallmark disease. Additionally, each susceptible environment presents slightly varying nutrient milieus that the colonizing pathogen must be equipped to exploit and be able to adapt its metabolism accordingly. Following colonization across these varied microbiomes, we also detected variation in the metabolic impact infection has on each community and that these differences

correlate with downstream clearance. Collectively, my results support the theoretical framework that *C. difficile* is in possession of reasonable metabolic plasticity, despite certain conserved aspects across environments. This plasticity allows it to colonize highly varied community structures and can alter the ecosystem in ways that allow for long-term colonization of the host.

The Ecosystem of the Gut is Altered as a Consequence of *C. difficile* Colonization

In Chapter 2, I presented and discussed my findings that upon colonization of distinct gut environments, *C. difficile* differentially altered the ecosystem of the gastrointestinal tract across antibiotic pretreatment models. The degree to which community-level transcriptional activity and metabolic output are changed positively correlates with the duration of *C. difficile* colonization. Furthermore, we observed a disparity in the amount that community structure is modified by *C. difficile* infection and changes in the metatranscriptome and subsequently the metabolome. These alterations to the microbiome may ultimately play into the ability of the pathogen to persist in some communities over others. The implication of these results is that there are multiple subcategories of *C. difficile* susceptibility. While a community may not be resistant to initial colonization it may still be able to eliminate pathogen colonization over a shorter time period. This “secondary” colonization resistance is characterized by recalcitrance to shifting overall metabolism to accommodate that of *C. difficile*. The specific factors or taxa that drive this newly identified process are yet to be mechanistically tested and present an exciting future research opportunity. Generally speaking, the data outlined in

Chapter 2 illustrates that the importance of function in the gut microbiota outweighs that of structure for the clearance of *C. difficile* colonization.

C. difficile Differentially Adapts Metabolism to Distinct Environments and the Ramifications of this for Disease

In Chapter 3, I focused my analysis more acutely on the metabolism of *C. difficile* and how the pathogen shifts nutrient utilization to succeed across different perturbed communities. By implementing a transcriptome-based metabolic modeling approach for assessing differences in environmentally-acquired resources, I was able to uncover signatures of *C. difficile* inhabiting distinct nutrient niche spaces across susceptible environments. As discussed in Chapter 1, *C. difficile* should be considered a bacterial generalist. Each of the antibiotic pretreated environments present separate challenges to *C. difficile*, not the least of which being the acquisition of growth nutrients which are equally likely be divergent between host species. Previous research from our group has demonstrated that *C. difficile* is capable of colonizing numerous distinct, but equally susceptible gut environments (5). Furthermore, the untargeted metabolic data presented in Chapter 2 reconfirmed that these pretreatment conditions resulted in significantly different metabolite profiles also in agreement with previous studies (4, 6, 7). Congruent with the principle that *C. difficile* can colonize large nutrient niche space, another group has demonstrated that it is able to utilize a large range of substrates for growth including simple sugars, complex polysaccharides, amino acids and peptides (8). The capability of *C. difficile* to respond to different environments appropriately stems

from a reasonably large and variable genomic capacity (9). This study revealed numerous metabolic capacities to adapt to the diversity of available nutrients within each susceptible gut environment. These patterns were ultimately reflected in results from the metabolic modeling approach I designed and implemented.

Numerous other groups have made mechanistic connections between the expression of virulence traits and the composition of the nutrient milieu *C. difficile* has colonized (6, 10, 11), which are at least partially observed in my work in Chapter 3. Aside from growth nutrients, *C. difficile* also requires particular metabolites to be present in the environment before it is able to proceed through its life-cycle and even commence the process of colonization. *C. difficile*, like many other subgroups of Clostridia, is able to form environmentally-resistant, dormant endospore which is thought to be the transmissive form of the pathogen (12). Prior to colonization, *C. difficile* must germinate from this spore phase into the metabolically-active vegetative phase of the life-cycle. This switch requires multiple environmental factors to signal an appropriate environment for *C. difficile* to colonize. Among these necessary metabolites are specific bile acids (primary germinants) and amino acids (secondary co-germinants). Conjugated bile acids (ie. taurocholate) have been shown to be the most potent germination signals. The deconjugated forms, while still being potent germination signals, are lethal to vegetative *C. difficile* at physiologic concentrations (13). We observed that the bile acid profiles between the antibiotic classes I chose for these studies were inconsistent with one another (Chapter 2), and suggests that many different combinations of bile acids may be sufficient to strictly inform *C. difficile* spores that they are in the GI tract of a

potential host. I hypothesize that the recognition by spores of co-germinants (ie. proline and glyine (14)) is actually a mechanism by which *C. difficile* guages the availability to nutrients niches in the gut as these amino acids are also substrates for Stickland fermentation. Together, this work supports the prediction that competition for nutrients may extend beyond active metabolism, and low levels of potential metabolic competition may recognized by *C. difficile* even before it leaves the spore.

Potential Mechanism of the Metabolic Interaction

In the previous chapters I commented that *C. difficile* may subvert host or community biology to better enable sustained colonization. This would not be the first instance to be reported of a bacterial pathogen hijacking the environment of the gut to promote its own metabolism. Studies of *Salmonella enterica* serotype *Typhimurium* infection found that the pathogen induces inflammation in the murine epithelial layer to induce the release of reactive-oxygen species into the lumen of the GI tract (15). As *Salmonella* is resistant to the bactericidal properties of these molecules, this action provides a terminal electron acceptor for the pathogen to utilize during aerobic respiration while simultaneously subverting host immunity to kill bacterial competitors for growth nutrients. In our dataset we identified a consistent trend across antibiotic class pretreatments for the significant increase of 5-aminovalerate in cecal content following infection by *C. difficile* (Fig. 4.1). This molecule is the byproduct of Stickland fermentation in *C. difficile* (16) and lysine degradation in many species (17). Concordantly, the 5-aminovalerate concentrations in the metabolomes we measured

are a nearly perfect inverse correlation with D-proline concentrations (Chapter 2), which strongly indicated that the increase was most likely due to the contribution of *C. difficile* metabolism. This molecule is especially interesting in the context of CDI for several reasons. Firstly, 5-aminovalerate has a combinatorial bactericidal effect when combined with certain reactive oxygen species (18). Due to the inflammatory effect of toxin on the host, it is highly likely that these additional molecules would be introduced to the gut lumen during infection. This may be a mechanism by which *C. difficile* actively antagonizes potential competitors. It should also be noted that toxin production is repressed by increased environmental nutrient concentrations (10), and decreased expression may indicate environments that *C. difficile* indirectly recognizes as the absence of competition for those substrates. Secondly, *C. difficile* possesses a version of proline reductase that may be capable of converting 5-aminovalerate back to D-proline when combined with lipoate in an energetically unfavorable reaction (19, 20). This action could not only detoxify the bactericidal effect but also allow for substrate recycling through Stickland fermentation. Finally, 5-aminovalerate has been shown to have direct effects on host physiology. It has been demonstrated that 5-aminovalerate is a GABA antagonist in both rat brains and guinea pig intestine (21). The functional effects of this receptor differ likely along the biogeography of the GI tract and their antagonism has been shown to increase gut motility (22), potentially promoting further increased dissemination following sporulation. Another possible host-interaction is that this molecule has also been shown to have anticoagulant activity (23) and could contribute to the “leaky gut” phenotype induced by *C. difficile* toxin, which in turn could

allow for increased access to nutrients entering the lumen from the bloodstream or increased bacterial antigen leaving the gastrointestinal tract. These combined factors lead to the hypothesis that *C. difficile* creates this molecule to subvert the inflammation caused by toxin production and other host processes to obtain a competitive advantage over specific community members. This is may not be the exclusive means by which *C. difficile* instigates change to allow persistence, however it should be noted that clindamycin pretreatment was the only group that did not see a significant increase in 5-aminovalerate following infection (Fig. 4.1). Likewise, whether or not this mechanism contributes to long-term colonization and is not solely active shortly after primary infection is also yet to be determined.

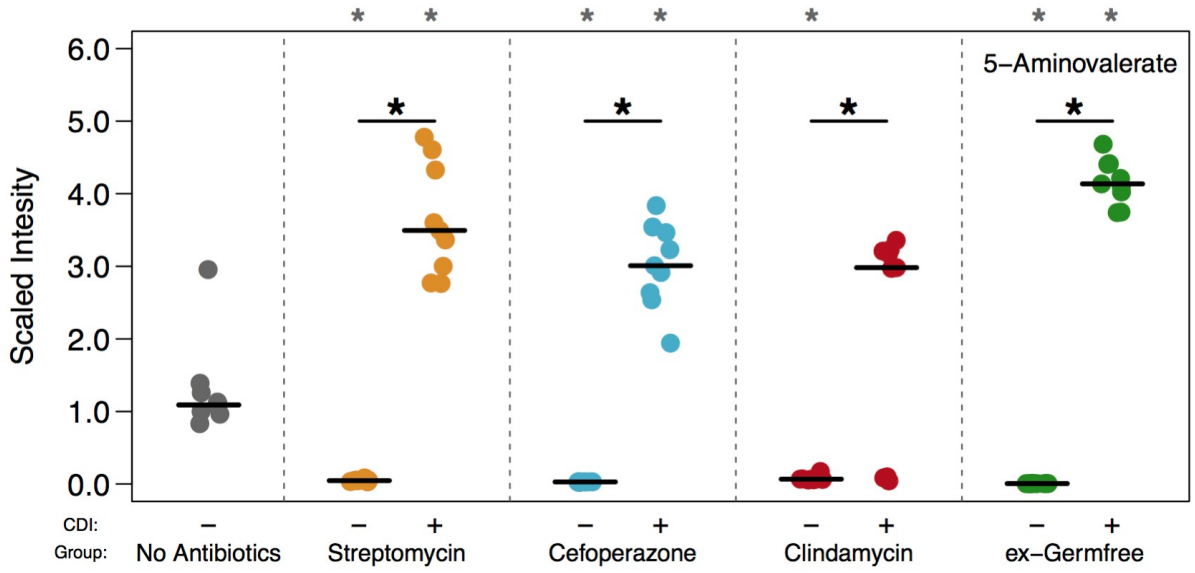

Figure 4.1 | Relative concentrations of 5-aminovalerate in each of the pretreatment and infections groups. Black asterisks denote significant differences between infected and uninfected groups within each susceptible condition. Gray asterisks along the top axis indicate significant difference from untreated mouse cecal metabolome. Significant differences were calculated using Wilcoxon rank-sum test with Benjamini-Hochberg correction ($p < 0.05$).

Both symptomatic and asymptomatic persistent colonization of *C. difficile* in patients presents difficulties for mitigating illness and spread of the pathogen in healthcare facilities (24). As antibiotic therapy has been shown to frequently lead to a cycle of recurrent infection (25) and antibiotic resistance in *C. difficile* is becoming increasingly prevalent (26), alternative approaches to eliminate persistent *C. difficile* colonization are desperately needed. Along these lines, fecal microbial transplant (FMT), has been identified as the most effective treatment to ongoing *C. difficile* colonization that we have identified thus far with >90% reported efficacy (27). A possible mechanism by which this clearance is mediated during this treatment is through immediate re-installation competitors for nutrients from a healthy community that are more adept than the generalist *C. difficile* for nutrient niche space in the gut. However, it has also been shown that the organisms that are constituents of the initial transferred sample, these members do not maintain colonization (28, 29). In the context of the research presented in the previous chapters, it is plausible that this process may instead temporarily re-introduce taxa that are able to metabolize the antagonistic molecules produced by *C. difficile*, temporarily eliminating this advantage. It is at this point that more specialized groups for given resources can outcompete *C. difficile*, and without this access to growth nutrients it can longer compete with the intestinal washout rate and is removed from the system. This is likely one of many contributing methods by which the gut microbiome resists colonization by *C. difficile*, and through an increased understanding of these interactions we could one day designed target biological

therapies to eliminate *C. difficile* infection or prevent colonization without the continued involvement of antibiotic therapy.

Future Research Avenues

In addition to following the line of questioning stemming from the 5-aminovalerate findings, there also remains large portions of planned research I was unable to complete. The multi-omic data that I have generated to characterize these experimental systems as well as the bioinformatic tools I have generated to integrate these datasets have led to a number of potential future research projects. The most prominent of these directions would be expansion of the metabolic modeling algorithms to the community-scale integrating both metagenomic and metatranscriptomic data. The first step in the approach that would likely have the high accuracy for modeling metabolism of individual bacterial groups in the cecum would be to assemble discrete genomic units on which to map transcript as opposed to the entire pool of genetic diversity. This task could be accomplished utilizing tools such as CONCOCT which cluster assembled contigs into groups of sequences that are similar in both metagenomic read coverage and kmer composition (30). The logic being that sequences that share kmer frequencies, a well established signature of individual species (31), and have similar number of reads mapping to them, are more likely to share an origin of a single species or one of its close relatives. I had completed this for each of the distinct metagenomes from mock-infected animals (to avoid assembling and mapping cDNA to *C. difficile* genes), and annotated the genes in each cluster according to the KEGG protein database. I added

an additional binning step where I collapsed clusters with >50% of annotated genes originating from the same genus level classification. This was done to compensate for possible error in clustering sequences that actually belonged together as well as eliminate very small clusters of genes that in reality may be constituents of larger groups. Final gene clusters were then assembled into Contig-Enabled Models (or CEMs), exactly analogous to what was performed for *C. difficile* in chapter 3, and multiple metrics of network centrality were calculated for each. Models with extremely low average connectivity were pruned away as it was unlikely that those gene collection were a true reflection of the metabolism of that organism or group. Metatranscriptomic reads were then mapped to gene nodes in each of the networks across the respective communities, and using the previously described algorithm in chapter 3, metabolite importance scores for all community members was calculated. An additional algorithm was devised to gauge the putative metabolic competition or synergy between individual groups and at a community-level, but with normalizing for species abundance and without assessing for population density effects. A depiction of the basic algorithm structure as well as one calculated interaction between *C. difficile* and a *Lactobacillus* CEM and then an assessment of most consumed or produced metabolites across all the CEMs in the cefoperazone pretreated community is shown in Fig. 4.2. The resulting scores could then be correlated with changes in the metabolome between infected and uninfected conditions. While this approach appeared to find an amount of tractable signal for competition between certain species and *C. difficile*, it also amplified some signals contributed by potential annotation artifacts. This was due to the nature of the

metabolite importance algorithm disproportionately weighting metabolites on the periphery of the network. As stated in chapter 3, this more readily identified input metabolites to the networks, however if reactions were annotated as reversible or functioned differently in the organism of interest compared to the KEGG database entry, metabolic output could easily be labeled as inputs during unsupervised network construction. The only solution to this would be that prior to metabolite importance calculations, manual curation of reactions included in each network is necessary. Future studies of inter-species competition for nutrients in complex communities of bacteria could greatly benefit by utilizing the computational platform I have developed following its completion to identify indicator metabolites as edges of strongest competition or synergy that may affect community assemblages.

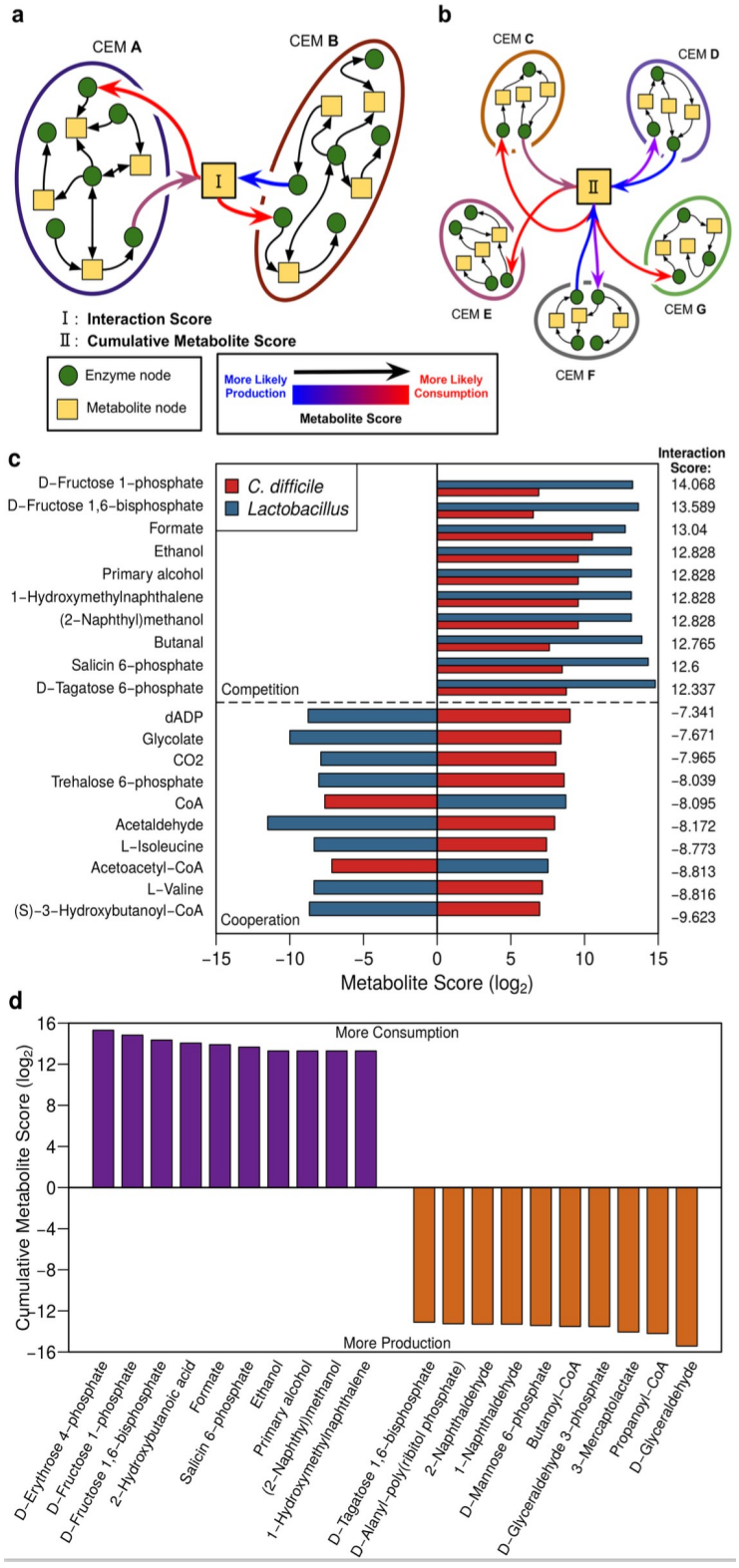

Figure 4.2 | Community-level modeling approach and example of metabolic interactions.

Figure 4.2 | Community-level modeling approach and example of metabolic interactions. (A) Schematic of interaction calculations between individual contig-enabled models (CEMs). Color of arrow corresponds to the individual metabolite scores for each CEM for metabolites I or II. An interaction score reflects the sum of the component $\log_2$ metabolite scores from each CEM for all metabolites they share a score for, since not all metabolites are in the network of all organisms. **(B)** Schematic for calculating cumulative metabolite scores across the cecal microbiota. The score for each individual metabolite from each CEM in the community was cumulatively added to yield final demand or consumption probabilities. The following examples utilized normalized metagenomic and metatranscriptomic read abundances from cefoperazone-pretreated, infected mice. The presented scores were part of the 90th percentile of highest or lowest scores scores for all organisms included in the calculations. **(C)** Putative interaction of *C. difficile* and Lactobacillus CEMs in the infected community. In instances when both scores were positive it implied competition and scores on both sides indicate possible cooperation. **(D)** Community-level metabolic crosstalk calculations of all CEMs in the community. Values above the central line are more likely to be consumed by the community, and those below may be produced overall.

Concluding Remarks

C. difficile is a pathogen of opportunity that exploits the susceptibilities of gut microbial communities to propagate and cause disease. Little is known about the specific interactions that *C. difficile* has with members of the gut microbiota. These likely include those that prevent *C. difficile* colonization in healthy communities, or those in the perturbed ecosystems that determine whether *C. difficile* colonizes long-term or is eliminated. In past studies, parsing the metabolic contribution of discrete groups in a community of organisms to the overall ecology has been difficult. With the advent of high-throughput sequencing, improved mass-spectrometry techniques, and methods for integrating multiple levels of high-dimensional biological data it is becoming easier to decipher these interactions. With continued research, we will one day understand the complex interplay and develop effective strategies for limiting *C. difficile* colonization in humans without risk of recurrent disease or acquired antibiotic resistances.

References

1. **Britton, R. A.**, and **V. B. Young**. 2012. Interaction between the intestinal microbiota and host in *Clostridium difficile* colonization resistance. Trends in microbiology **20**:313–9.

2. **Wilson, K. H.**, and **F. Perini**. 1988. Role of competition for nutrients in suppression of *Clostridium difficile* by the colonic microflora. Infection and Immunity **56**:2610–2614.

3. **Robinson, C. D.**, **J. M. Auchtung**, **J. Collins**, and **R. A. Britton**. 2014. Epidemic *Clostridium difficile* strains demonstrate increased competitive fitness compared to nonepidemic isolates. Infection and Immunity **82**:2815–2825.

4. **Theriot, C. M.**, **M. J. Koenigsknecht**, **P. E. Carlson**, **G. E. Hatton**, **A. M. Nelson**, **B. Li**, **G. B. Huffnagle**, **J. Z. Li**, and **V. B. Young**. 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. Nature Communications **5**:3114.

5. **Schubert, A. M.**, **H. Sinani**, and **P. D. Schloss**. 2015. Antibiotic-induced alterations of the murine gut microbiota and subsequent effects on colonization resistance against *Clostridium difficile*. mBio **6**.

6. **Antunes, L. C. M.**, **J. Han**, **R. B. R. Ferreira**, **P. Lolić**, **C. H. Borchers**, and **B. B. Finlay**. 2011. Effect of antibiotic treatment on the intestinal metabolome. Antimicrobial Agents and Chemotherapy **55**:1494–1503.

7. **Jump, R. L. P.**, **A. Polinkovsky**, **K. Hurless**, **B. Sitzlar**, **K. Eckart**, **M. Tomas**, **A. Deshpande**, **M. M. Nerandzic**, and **C. J. Donskey**. 2014. Metabolomics analysis identifies intestinal microbiota-derived biomarkers of colonization resistance in clindamycin-treated mice. PLoS ONE **9**.

8. **Scaria, J.**, **J. W. Chen**, **N. Useh**, **H. He**, **S. P. McDonough**, **C. Mao**, **B. Sobral**, and **Y. F. Chang**. 2014. Comparative nutritional and chemical phenome of *Clostridium difficile* isolates determined using phenotype microarrays. International Journal of Infectious Diseases **27**:20–25.

9. **Sebaihia, M.**, **B. W. Wren**, **P. Mullany**, **N. F. Fairweather**, **N. Minton**, **R. Stabler**, **N. R. Thomson**, **A. P. Roberts**, **A. M. Cerdeño-Tárraga**, **H. Wang**, **M. T. Holden**, **A. Wright**, **C. Churcher**, **M. A. Quail**, **S. Baker**, **N. Bason**, **K. Brooks**, **T. Chillingworth**, **A. Cronin**, **P. Davis**, **L. Dowd**, **A. Fraser**, **T. Feltwell**, **Z. Hance**, **S. Holroyd**, **K. Jagels**, **S. Moule**, **K. Mungall**, **C. Price**, **E. Rabbinowitsch**, **S. Sharp**, **M. Simmonds**, **K. Stevens**, **L. Unwin**, **S. Whithead**, **B. Dupuy**, **G. Dougan**, **B. Barrell**, and **J. Parkhill**. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nature Genetics **38**:779–786.

10. **Dineen, S. S.**, **A. C. Villapakkam**, **J. T. Nordman**, and **A. L. Sonenshein**. 2007. Repression of *Clostridium difficile* toxin gene expression by CodY. Molecular Microbiology **66**:206–219.

11. **Bouillaut, L.**, **T. Dubois**, **A. L. Sonenshein**, and **B. Dupuy**. 2015. Integration of metabolism and virulence in *Clostridium difficile*. Research in Microbiology **166**:375–383.

12. **Abt, M. C.**, **P. T. McKenney**, and **E. G. Pamer**. 2016. *Clostridium difficile* colitis: pathogenesis and host defence. Nature Publishing Group **14**:609–620.

13. **Theriot, C. M.**, **A. A. Bowman**, and **V. B. Young**. 2016. Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for *Clostridium difficile* Spore Germination and Outgrowth in the Large Intestine. mSphere **1**:e00045–15.

14. **Sorg, J. A.**, and **A. L. Sonenshein**. 2008. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. Journal of bacteriology **190**:2505–12.

15. **Winter, S. E.**, **P. Thiennimitr**, **M. G. Winter**, **B. P. Butler**, **D. L. Huseby**, **R. W. Crawford**, **J. M. Russell**, **C. L. Bevins**, **L. G. Adams**, **R. M. Tsolis**, **J. R. Roth**, and **A. J. Bäumler**. 2010. Gut inflammation provides a respiratory electron acceptor for Salmonella. Nature **467**:426–429.

16. **Neumann-Schaal, M.**, **J. D. Hofmann**, **S. E. Will**, and **D. Schomburg**. 2015. Time-resolved amino acid uptake of *Clostridium difficile* 630erm and concomitant fermentation product and toxin formation. BMC Microbiology **15**:281.

17. **Chou, H. T.**, **M. Hegazy**, and **C. D. Lu**. 2010. L-lysine catabolism is controlled by L-arginine and ArgR in Pseudomonas aeruginosa PAO1. Journal of Bacteriology **192**:5874–5880.

18. **Ko, K. C.**, **B. Wang**, **P. C. Tai**, and **C. D. Derby**. 2008. Identification of potent bactericidal compounds produced by escapin, an L-amino acid oxidase in the ink of the sea hare Aplysia californica. Antimicrobial Agents and Chemotherapy **52**:4455–4462.

19. **Kabisch, U. C.**, **A. Gräntzdörffer**, **A. Schierhorn**, **K. P. Rücknagel**, **J. R. Andreesen**, and **A. Pich**. 1999. Identification of D-proline reductase from Clostridium sticklandii as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a proprotein. Journal of Biological Chemistry **274**:8445–8454.

20. **Ogata, H.**, **S. Goto**, **K. Sato**, **W. Fujibuchi**, **H. Bono**, and **M. Kanehisa**. 1999. KEGG: Kyoto encyclopedia of genes and genomes **27**:29–34.

21. **Kristiansen, U.**, **A. Hedegaard**, **C. Herdeis**, **T. M. Lund**, **B. Nielsen**, **J. J. Hansen**, **E. Falch**, **H. Hjeds**, and **P. Krogsgaarden-Larsen**. 1992. Hydroxylated Analogues of 5-Aminovaleric Acid as 4-Aminobutyric AcidB Receptor Antagonists: Stereostructure-Activity Relationships. Journal of Neurochemistry **58**:1150–1159.

22. **Hyland, N. P.**, and **J. F. Cryan**. 2010. A gut feeling about GABA: Focus on GABAB receptors. Frontiers in Pharmacology **OCT**.

23. **Cole, K. R.**, and **F. J. Castellino**. 1984. The binding of antifibrinolytic amino acids to kringle-4-containing fragments of plasminogen. Archives of Biochemistry and Biophysics **229**:568–575.

24. **Cole, S. A.**, and **T. J. Stahl**. 2015. Persistent and Recurrent *Clostridium difficile* Colitis. Clinics in colon and rectal surgery **28**:65–9.

25. **Kelly, C. P.**, and **J. T. LaMont**. 2008. *Clostridium difficile*–more difficult than ever. The New England journal of medicine **359**:1932–1940.

26. **Huang, H.**, **A. Weintraub**, **H. Fang**, and **C. E. Nord**. 2009. Antimicrobial resistance in *Clostridium difficile*. U.S. Patent 6.

27. **Rohlke, F.**, and **N. Stollman**. 2012. Fecal microbiota transplantation in relapsing *Clostridium difficile* infection. Therapeutic advances in gastroenterology **5**:403–420.

28. **Lawley, T. D.**, **S. Clare**, **A. W. Walker**, **M. D. Stares**, **T. R. Connor**, **C. Raisen**, **D. Goulding**, **R. Rad**, **F. Schreiber**, **C. Brandt**, **L. J. Deakin**, **D. J. Pickard**, **S. H. Duncan**, **H. J. Flint**, **T. G. Clark**, **J. Parkhill**, and **G. Dougan**. 2012. Targeted Restoration of the Intestinal Microbiota with a Simple, Defined Bacteriotherapy Resolves Relapsing *Clostridium difficile* Disease in Mice. PLoS Pathogens **8**.

29. **Seekatz, A. M.**, **J. Aas**, **C. E. Gessert**, **T. A. Rubin**, **D. M. Saman**, **J. S. Bakken**, and **V. B. Young**. 2014. Recovery of the gut microbiome following fecal microbiota transplantation. mBio **5**.

30. **Alneberg, J.**, **B. S. Bjarnason**, **I. de Bruijn**, **M. Schirmer**, **J. Quick**, **U. Z. Ijaz**, **L. Lahti**, **N. J. Loman**, **A. F. Andersson**, and **C. Quince**. 2014. Binning metagenomic contigs by coverage and composition. Nature Methods **11**:1144–1146.

31. **Sandberg, R.**, **G. Winberg**, **C. I. Bränden**, **A. Kaske**, **I. Ernberg**, and **J. Cöster**. 2001. Capturing whole-genome characteristics in short sequences using a naive Bayesian classifier. Genome Research **11**:1404–1409.