*Clostridium difficile* colonizes alternative nutrient niches during infection across distinct murine gut communities

**Authors:** Matthew L. Jenior, Jhansi L. Leslie, Vincent B. Young, and Patrick D. Schloss\*

### Abstract

*Clostridium difficile* infection (CDI) has become the largest single cause of hospital-acquired infection in the United States. CDI susceptibility is most frequently associated with previous antibiotic exposure, which disrupt the gut bacterial community. This has been described for multiple antibiotic classes which result in distinct gut communities, each presenting separate metabolic challenges to *C. difficile*. We hypothesized that *C. difficile* adapts its physiology to nutrient availability within differentially susceptible gut environments. Utilizing an *in vivo* model of CDI, we demonstrated that *C. difficile* highly colonized the cecum of mice receiving one of three individual antibiotic pretreatments. We found levels of spore and toxin production varied between each group, both processes partially regulated by environmental nutrient concentrations. To more closely investigate metabolic responses of *C. difficile* during infection, we performed transcriptomic analysis of the pathogen from cecal content of infected mice. This revealed expression variation in numerous catabolic pathways for various carbon sources. To assess which substrates *C. difficile* was exploiting, we developed a transcriptomic-enabled genome-scale metabolic model of *C. difficile* and a metabolite scoring algorithm that leveraged network architecture. With this platform, we identified carbon sources used by *C. difficile* asymmetrically between infection models. These findings were validated through untargeted mass spectrometry analysis from each condition. Our results supported the hypothesis that *C. difficile* indeed metabolized alternative carbon sources across colonized environments. These data also highlighted conserved elements of *C. difficile*'s metabolic strategy, specifically consumption of host-derived aminoglycans and Stickland fermentation substrates.

### Importance

In this study we demonstrate that not only does *C. difficile* alter pathogenesis between differentially sensitized hosts, but also exploits separate nutrient niches across environments. Our results support that *C. difficile* possesses a highly plastic nutrient niche space, allowing it to successfully infect distinct communities. This work also provides evidence that *C. difficile* virulence may be driven by accessibility of specific carbon sources during infection. This work could lead to the discovery of targeted measures to prevent *C. difficile* colonization including potential pre- or probiotic therapies. Our metabolite importance calculation workflow also provides a platform to the study of nutrient requirements of pathogens in the context of infection or even patterns of substrate utilization in communities of bacteria.

### 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 life-threatening conditions pseudomembraneous colitis and toxin megacolon. Prior treatment with antibiotics is the most common risk factor associated with susceptibility to CDI (3), by disrupting the gut microbiota (4). Multiple distinct antibiotic have been shown to induce susceptibility *C. difficile* to colonization in a mouse model (5–7). Each resulted in unique gut bacterial communities that were all equally receptive to *C. difficile* colonization. Others have also shown that many of select antibiotic classes also alter the gut metabolome, increasing the concentrations of some *C. difficile* growth substrates (6, 8–10). Taken together these results are a strong indication that the healthy gut microbiota inhibits the growth of *C. difficile* through limitation of substrates it needs to grow. The ability of an intact gut community to exclude *C. difficile* colonization is suggestive of the nutrient-niche hypothesis in which an organism must be able to utilize a subset of available resources better than all competitors to colonize the intestine (11, 12).

Based on genomic and *in vitro* growth characteristics, *C. difficile* appears able to fill multiple nutrient niches. *C. difficile* has a relatively large and mosaic genome, it can utilize a variety of growth substrates, and possesses a diverse array host range (13–15). These qualities are hallmarks of a ecological generalist for bacteria. *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 (16, 17). Other studies further indicate that numerous additional aspects of *C. difficile* metabolism may be influenced by way of environmental nutrient concentration-sensative global transcriptional regulators such as CodY and CcpA (18, 19). However, the previous analyses have mainly focused on *in vitro* growth (20, 21) or colonization of germfree mice (19, 22). 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 a attribute a metabolite to specific organisms in the milieu. 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 used *in vivo* transcriptomic and metabolomic analysis of *C. difficile* in specific pathogen free (SPF) mice to better understand the active metabolic pathways in a more realistic 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 community 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 distinct classes of antibiotics. By characterizing the transcriptome of *C. difficile* in these different communities and the metabolome of the respective environments, we were able to generate a systems model to directly test the nutrient-niche hypothesis.

### Results

**Levels of *C. difficile* sporulation and toxin activity vary among different communities.** Conventionally-reared SPF mice were treated with one of three different antibiotics. These included streptomycin, cefoperazone, and clindamycin (Table 1 and Fig. S1), which were selected because they each significantly impact the structure and diversity of the cecal microbiome uniquely (Fig. S2A and S2B). In addition, we mono-associated germfree (ex-GF) mice with *C. difficile* to understand its physiology in the absence of other microbiota. We measured sporulation and toxin production at 18 hours post infection because that corresponds to the time when *C. difficile* reaches its maximum vegetative cell density in the cecum with limited sporulation (23). There was not a significant difference in the number of vegetative cells between any susceptible condition tested (Fig. 1A). All antibiotic treated (Table 1) and ex-GF mice were colonized to ~1×108 colony forming units (cfu) per gram of cecal content, while untreated mice maintained colonization resistance to *C. difficile*. Despite having the same number of vegetative *C. difficile* cells, largely more spores were detected in ex-GF mice than in the antibiotic pretreated mice (*P* = 0.003, 0.004, and 0.003; Fig. 1B). There was also significantly more toxin activity in ex-GF animals than any other colonized group (all *P* < 0.001), with slight variation between antibiotic pretreatment groups (Fig. 1C). These results indicated that *C. difficile* colonized different communities to a consistently high level and that the antibiotic perturbed communities had moderate contrast in levels of spores and toxin. To investigate the physiology of *C. difficile* when colonizing distinct susceptible gut environments, we performed whole transcriptome analysis of *C. difficile* growing in the cecum during infection.

\*\* *C. difficile* alters its gene expression pathways when colonizing different communities.\*\* We first attempted to measure differential expression of specific genes associated with *in vivo* phenotype changes reported in previous studies. These included genes involved in sporulation, the pathogenecity locus, quorum sensing, and metabolite-regulated sigma factors (Fig. S3); however despite large-scale differences between pretreatment groups, no clear trends between gene expression and colonization, sporulation, or toxin production were evident. It became evident that colonization of each treatment group resulted in different levels of expression for each group of genes. This further indicated that *C. difficile* may adapt its metabolism to the environment it is colonizing.

We next focused on specific gene families known to contribute to *C. difficile* metabolism (Table S1 and Fig. 2A). 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 (24). This indicated that catabolizing environmental amino acids was generally occurring in *C. difficile* during infection, regardless of the structure of the surrounding community. Although there were families 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. Mice pretreated with cefoperazone 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. *C. difficile* in clindamycin-pretreated mice tended to have more expression of genes from Disaccharide catabolism (e.g. (  B )-galactosidases and trehalose/maltose/cellibiose hydrolases), Fermentation product metabolism (including consumption or productiono 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 clindamycin-pretreated mice. Finally, in streptomycin pretreatment, *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 catabolism of amino acids and specific carbohydrates are core components of the *C. difficile* nutritional strategy during infection. These results also supported the hypothesis that *C. difficile* adapted its metabolism across different susceptible environments.

**Genome-scale metabolic model structure underscores known *C. difficile* biology.** To further investigate which metabolites were differentially utilized between conditions, we represented the metabolic network of *C. difficile* as a directed bipartite graph using the genome and biochemical reaction annotations available in KEGG (2016). Enzymes and metabolites were represented by nodes, and their interactions by directed connecting edges (Fig. 3A). The complete network of *C. difficile* str. 630 that we created contains 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 assess for those nodes which are critical to the structure of the metabolic network and if these patterns reflect known patterns in *C. difficile* or bacterial metabolism (Table S2). BC is the quantity of shortest paths connecting all other member nodes of a network that pass through a given node. In biological terms, this refers to the amount of influence a given hub has on the overall flow of metabolism through the network (25). Similarly, CC is instead a calculation for the reciprocal sum of the lengths of those shortest paths quantified by the BC. This value demonstrates how essential a given node is the the overall structure of the metabolic network (26). Together, these metrics allowed for the assessment of how much a network reflects known principles of highly central biological processes.

Combining both analyses to find those nodes that are not only central control points, but also important to the structure of the entire network, we found 5 of the top 10 enzymes according to both BC and CC were most critical to metabolism based on topology (2-dehydro-3-deoxyphosphogluconate aldolase, aspartate aminotransferase, pyruvate-flavodoxin oxidoreductase, formate C-acetyltransferase, and 1-deoxy-D-xylulose-5-phosphate synthase). Many of these enzymes participate in core processes including glycolysis, the pentose phosphate pathway, or the citric acid cycle. Upon analysis of the other 15 high-scoring enzymes, the majority are also distinct components of these pathways as well as several for the metabolism of amino acids (Table S2) 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" (27). 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 collectively validated our model of *C. difficile* str. 630 as a platform for studying metabolism.

**Metabolite importance algorithm reveals adaptive nutritional strategies of *C. difficile* during infection of distinct environments.** We next sought to include transcriptomic data with the metabolic model to infer which metabolites *C. difficile* most likely extracted from a given environment. To accomplish this we mapped normalized transcript abundances to the enzyme nodes in the network. The importance 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 it as a product (Fig. 3B). A metabolite with a high importance score was more likely obtained from the environment because the expression of genes for enzymes that produce the metabolite were low. Given that separate sequencing efforts for each individual mouse in each group was impossible, we adopted a Monte Carlo-style simulation for statistical validation of our findings. This process generated random score distributions for all metabolite nodes 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. 3C).

To identify the core metabolites that were most essential for *C. difficile* growth, regardless of the environment, we cross-referenced 40 highest scoring metabolites from each treatment group and identified the most consistenly high scoring metabolites (Fig. 4A). The host-derived aminoglycan N-acetylglucosamine (GlcNAc) was found to the have the highest median importance 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 (19). We went on to confirm that our strain of *C. difficile* str. 630 could metabolize GlcNAc for growth (Fig. 4B) in *C. difficile* minimal media (28). The Stickland fermentation acceptor proline was also found to be important in all conditions tested (29). *C. difficile* is auxotrophic for several amino acids including proline, which prevented testing it directly for *in vitro* growth changes despite providing for modest growth in no carbohydrate control. Previous analysis of *C. difficile* colonizing GF mice under mono-associated conditions has indicated that *C. difficile* uses both sets of metabolites (19); 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 may be an integral component of the nutrient niche for *C. difficile*.

We next focused on the highest scoring metabolites that were distinct to each infection condition (Fig. 4A). The resulting groups of metabolites all contained at least one known carbohydrate growth substrate of *C. difficile* (6, 12, 30) (Fig.4B). These included sorbitol (streptomycin-pretreated), mannitol (cefoperazone-pretreated), and salicin (clindamycin-pretreated), which are likely introduced by the animal's diet and N-acetylneuraminate (ex-Germfree), which is an aminoglycan obtained from the animal's mucin layer (31). The concentration of N-acetylneuraminate (Neu5Ac) was previously shown to be liberated by other bacterial species to the benefit of *C. difficile* (30). Considering that Neu5Ac had one of the higher importance scores in our ex-GF animals, this suggests that *C. difficile* does not require other bacteria to liberate it from the mucus layer. Furthermore, in ex-GF mice two forms of acetate were important which *C. difficile* is known to convert convert further into butyrate during fermentation (32). Galactitol also scored highly in Streptomycin-pretreatment and has been shown to be a sugar alcohol growth substrate for other *C. difficile* strains *in vivo* (22), but did not improve *in vitro* growth in this study (Table S4). Succinyl-CoA was of distinct importance in clindamycin-pretreated mice, which is the direct precursor to succinate that is produced the gut microbiota and utilized by *C. difficile* for growth (9). Cefoperazone and clindamycin preatreatments were also associated with increased importance of ribose analogs, which has been shown to be another carbohydrate substrate that *C. difficile* ferments for growth (33). Additionally, alanine was also found to be important in this condition and is a preferred electron donating amino acid in Stickland fermentation (24). These data supported the hypothesis that *C. difficile* exploited alternative nutrient sources between the susceptible environments it colonized.

***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 antibiotic-treated and GF mice. We used 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 importance 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 measured whether the relative concentration of important metabolites had a higher concentration in the susceptible environment relative to untreated SPF mice (Fig. 5). We found that the relative concentration of N-acetylglucosamine was actually significantly lower in all susceptible conditions (Fig. 5A; 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. 5B and S7B). Succinate was significantly increased in both streptomycin and clindamycin pretreated mice (Fig. 5D; all *P* < 0.05). Among the cefoperazone-pretreated SPF and GF mice, we also found that mannitol/sorbitol (Fig. 5C), N-acetylneuraminate (Fig. 5E), and glycine (Fig. S6A) were significantly higher in cefoperazone-treated SPF and GF mice (all *P* < 0.05). These results supported the assertion that antibiotic treatment opened nutrient niches that *C. difficile* was able to exploit for its growth.

Second, we measured the change in the concentrations of important metabolites during *C. difficile* infection compared to mock infected mice (Fig. 5). Both groups of host-derived aminoglycans, GlcNAc/GalNAc (Fig. 5A) and Neu5Ac (Fig. 5E), were significantly lower when 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. 5B) and hydroxyproline (Fig. S6C) 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-pretreatment (Fig. S6D; *P* < 0.05). The Stickland donors leucine and isoleucine were significantly decreased in all infected conditions except streptomycin-pretreatment (Fig. S6A and S6B; all *P* < 0.05). Concentrations of alanine were unchanged across any condition compared to untreated SPF mice (data not shown). 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-pretreatment (Fig. 5C). While a decrease in concentration of succinate in both streptomycin and clindamycin pretreatment were observed, neither relationship was found to be significant. Overall, metabolomic analysis supported our metabolite importance algorithm for predicting the metabolites utilized by *C. difficile* during different infection conditions. Results from metabolic modeling combined with UPLC-MS also suggested a possible preference hierarchy for growth substrates in *C. difficile*.

### Discussion

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 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 and found that *C. difficile* likely utilized specific alternative carbohydrates, carboxylic acids, and aminoglycans across colonized conditions for energy. 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 modeling-based we employed untargeted mass spectrometry that demonstrated increased availability of many metabolites highlighted by our algorithm in susceptible gut environs. Metabolomic analysis further revealed differential concentration reductions in important metabolites during *C. difficile* infection that were suggestive of a preference hierarchy for certain growth nutrients.

Our results expand upon previous understanding of *C. difficile* metabolism during infection by supporting that not only does the pathogen adapt it's metabolism to life inside of a host (19, 22), but also to the context of the specific gut environment in which it finds itself. By this rationale, differences in metabolite importance and substrate availability can be explained by the concomitant decrease in 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 competiton. In this condition, N-acetylneuraminate in these mice was found to be most highly important, and concentrations were also significantly increased in susceptible mice and concordantly decreased following infection. The same trend in the metabolome was also present in cefoperazone-pretreatment, which suggested that *C. difficile* is less competitive for this host-derived aminoglycan and may really have access when certain competition is at a minimum. 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 PTS transport systems are simply upregulated *in vivo* [reference], but our results indicate more complex regulation in concert with ABC sugar transporters inversely expressed between antibiotic pretreatments (Fig. 2). In agreement with earlier research we found that *C. difficile* likely preferred to ferment amino acids for energy during infection of GF mice, however we also found this metabolic strategy to be conserved across all infection conditions tested. *C. difficile* also demonstrated redundancy in this pathway several Stickland substrates had consistently high importance scores including alanine, leucine, and proline (Table S3) supported by drops in concentration during infection (Fig. 5A and S6A). 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 GI tract (34). This makes Stickland fermentation a valuable metabolic strategy, and stands to reason *C. difficile* would conserve across all environments it colonizes. This same principle may also extend to host mucus layer derived aminoglycans as they are another source of both carbon and nitrogen that, despite augmented release by members of the microbiota (35), would be present at some basal concentration regardless of other species' intercession. Finally, we did find disparity in some metabolite importance and change in *in vivo* concentration of certain previously characterized *C. difficile* growth substrates. This may indicate a nutrient preference hierarchy during infection in which amino acids are prized above all other substrates, followed by aminoglycans, then some permutation of carbohydrates, sugar alcohols, and carboxylic acids. 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 elucidate.

Our approach is novel to the study of *C. difficile* behavior during infection because it combines multiple levels of biological data to identify metabolic trends that would not be apparent in a single dataset. 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 virtue of our importance algorithm's reliance on network topology, signal contributed by those metabolites on the periphery, and more likely to be imported from the environment, is amplified. We have also made the pipeline generalizable to all species of bacteria whose gene content can be annotated using KEGG. Additionally, this modeling approach also may allow for the identification of emergent properties for the metabolism of *C. difficile* during infection. One example could be the appearance of CO2, an apparent metabolic end product, in the list of shared important metabolites. While this may be a shortcoming of the annotation, one group has posited that *C. difficile* may actually be autotrophic under certain conditions and could explain our results (36). These findings highlight that our method not only identifies growth substrates, but also reports metabolites that are being removed from the environment for any reason.

It is important to note that some of our methods were limited by a few inescapable constraints that were addressed as we progressed through experimentation and analysis. As *C. difficile* represented less than 3.3% of the bacterial load in the ceca of the antibiotic-treated mice (Fig. S2C), a high degree of sequencing depth per sample was required to yield sufficient quantities of reads contributed by *C. difficile*. This prohibited our ability to sequence each experimental replicate individually and made it necessary to pool samples within each treatment group to give to sufficient high quality RNA for library construction. Microarray-based gene expression measurement was also not a viable alternative as the amount of background orthologous transcription from other bacterial species would contribute greatly to non-specific binding and eliminate true *C. difficile* signal. Despite a single transcriptomic sequencing effort per condition, several models of bacterial protein expression determinants also suggest that intracellular concentration of mRNA is the strongest predictor for abundance of the corresponding protein products (37). We were therefore confident in utilizing normalized transcript abundance as a proxy for enzyme levels. Metabolomic comparisons are also complicated by the fact that multiple organisms contribute to the metabolite pool including host and other microbes. Possible limitations of our modeling approach also exist, despite much of our results being consistent with previously published work on the metabolism of *C. difficile*. Ultimately, the metabolite importance 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 (38), 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. Incorrect directionality annotations may lead to mislabeling reactants or products and potentially lead to incorrect importance calculations. With additional manual curation of the *C. difficile* metabolic network, more species specific discoveries can eventually be made.

In spite of these assumptions, our method does underscore known elements of *C. difficile* biology and subsequent analyses supported predictions of nutrient niche plasticity. In conclusion, this combined genomic + transcriptomic + metabolomic analysis showed that when infecting diverse host-associated gut environments, *C. difficile* optimized its nutrient utilization profile to each gut environment and effectively colonized. These results have implications for the research 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.

### Materials and Methods

**Animal care and antibiotic administration** Adapted from the previously described model (39), six-to-eight week-old SPF C57BL/6 mice were obtained from a single breeding colony maintained at the University of Michigan for all experiments. 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 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.

***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 for cfu 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 content 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, cefoxitin, and erythromycin) at 37° C for 24 hours under anaerobic conditions (40). 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 (41). 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.

***C. difficile* toxin titer assay** To quantify the titer of toxin in the cecum a Vero cell rounding assay was performed as in (42). 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. A more detailed protocol with product information can be found at: <https://github.com/SchlossLab/Jenior_Modeling_mBio_2017/blob/master/protocols/toxin_assay/Verocell_ToxinActivity_Assay.Rmd>

**16S rRNA gene sequencing** 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 (43). All 63 samples were sequenced on a single sequencing run.

**Sequence curation** The 16S rRNA gene sequences were curated using the mothur software package (v1.36), as described previously (43). 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) (44). 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** To generate enough mRNA biomass contributed by *C. difficile*, we pooled cecal content from all mouse replicates into a single large isolation for each treatment group. Pooling was performed in a sterile stainless steel mortar resting in dry ice and a small amount of 100% ethanol. After all content for the given group was added, the sample was 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 (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. Completion of the 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 prokaryotic and eukaryotic rRNA from the samples according the manufacturer's protocol. 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, both using the manufacturer's protocol. 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.

**Sequence curation, read mapping, and normalization.** Raw transcript sequencing read curation was performed in a two step process. Residual 5’ and 3’ Illumina adapter sequences were trimmed using CutAdapt (46) on a per library basis. Reads were quality trimmed using Sickle (Joshi, 2011) on the default settings. An average of ~300,000,000 total reads (both paired and orphaned) remained after quality trimming. Mapping was accomplished using Bowtie2 (47) and the default stringent settings. ~1,600,000 reads in sample each mapped to the annotated nucleotide gene sequences of *PeptoClostridium difficile* str. 630 from the KEGG: Kyoto Encyclopedia of Genes and Genomes (48). Optical and PCR duplicates were then removed using Picard MarkDuplicates (<http://broadinstitute.github.io/picard/>), leaving ~150,000 reads per sample for final analysis. The remaining mappings were converted to idxstats format using Samtools (49) 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.

**Reaction Annotation & Bipartite Network Construction.** The metabolism of *C. difficile* str. 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 str. 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 substrate importance calculations. Betweenness-centrality and overall closeness centralization indices were calculated using the igraph R package found at <http://igraph.org/r/>.

**Metabolite Importance Calculation.** The substrate importance algorithm (Fig. 3a) favors metabolites that are more likely acquired from the environment (not produced within the network), and will award them a higher score (Fig. 4b & 6c). 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 importance 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 importance 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 (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 importances due to distinct transcriptional patterns for the environment measured from those metabolites that were constitutively important. We employed a 10,000-fold bootstrapping approach of randomly reassigning transcript abundance for enzyme nodes and recalculating metabolite importances. 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 importance scores which were then used to calculate the median and confidence interval to generate a probability for each metabolite importance 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(50).

**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 + 0.5% cysteine. All culturing and growth measurement were performed anaerobically in a Coy Type B Vinyl Anaerobic Chamber (3.0% H, 5.0% CO2, 92.0% N, 0.0% O2). *C. difficile* str. 630 was grown for 14 hours at 37° C in 3 mL BHI + 0.5% cysteine. Cultures were then centrifuged at 2000 rpm for 5 minutes and resulting pellets were then washed twice with sterile, anaerobic 1 × 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, heated to 37° C. Plates were incubated for 24 hours with automatic optical density readings at 600 nm taken every 30 minutes. OD600 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 OD600 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.

### Funding Information

This work was supported by funding from the National Institutes of Health to PDS (R01GM099514, P30DK034933, U19AI09087, and U01AI124255), VBY (P30DK034933, U19AI09087, and U01AI124255), a Translational Research Education Certificate grant to JLL (MICHR; UL1TR000433), and was partially supported by a predoctoral fellowship from the Cellular Biotechnology Training Program to MLJ (T32GM008353).

### Acknowledgements

The authors would like to acknowledge Charles Koumpouras for assistance with DNA extractions and metabolomic sample preparation. We would also like to acknowledge members of the University of Michigan Germfree Mouse Center, University of Michigan Sequencing Core, and Metabolon for their assistance in experimental design, execution, and data collection. Pooled and quality trimmed transcriptomic read data and experiment metadata are available through the NCBI Sequence Read Archive (SRA; PRJNA354635). Data processing steps for beginning from raw sequence data to the final manuscript are hosted at <http://www.github.com/SchlossLab/Jenior_Modeling_mBio_2017>. The authors would additionally like to thank Geoffrey Hannigan Ph.D and Nielsen Baxter Ph.D. for their suggestions on manuscript drafts.

**Author Affiliations** **Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan.** Matthew L. Jenior, Jhansi L. Leslie, & Patrick D. Schloss Ph.D.

**Department of Internal Medicine & Infectious Diseases, University of Michigan Medical Center, Ann Arbor, Michigan.** Vincent B. Young M.D. Ph.D.

**Author Contributions** M.L.J. conceived, designed and performed experiments, analyzed data, and drafted the manuscript. J.L.L. performed experiments and analyzed data. V.B.Y. contributed to the manuscript. P.D.S. interpreted data and contributed the manuscript. The authors declare no conflicts of interest.

**Corresponding author** Correspondence to [Patrick D. Schloss](pdschloss@umich.edu)

### Figure Legends

**Table 1 | Antibiotics used during *C. difficile* murine infection models.**

**Figure 1 | 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.

**Figure 2 | *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)**. Genes, annotations, and normalized transcript abundances can be found in Table S1. Refer to Fig. S4 for additional figure interpretation.

**Figure 3 | *C. difficile* str. 630 genome-scale bipartite metabolic network architecture and transcriptomic-enabled metabolite importance calculation.** **(A)** Largest component from the bipartite genome-scale metabolic model of *C. difficile* str. 630. Enzyme node sizes reflect the levels of detectable transcript from each gene. Importance 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 importance 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 importance score for **m**.

**Figure 4 | Metabolic network analysis reveals differential carbon source utilization by *C. difficile* across infections.** Reported metabolites were calculated to have <2.5% probability to be included in the associated random score distribution. All comparison. All comparisons between groups was performed using the 40 highest scoring metabolites from each condition and to further focus the analysis we reported only the top five scored metabolites from any group listed.**(A)** Shared importance represents the median score of metabolites that were consistently important among all infected conditions. Below the conserved patterns, distinctly important metabolites for each group are shown. **(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.

**Figure 5 | Untargeted *in vivo* metabolomics support network-based metabolite importance scores and suggest nutrient preference hierarchy.** Paired metabolites were quantified simultaneously as the only differ by chirality making differentiation impossible. 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.

**Supplementary Figure 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 1×103 *C. difficile* str. 630 spores. Euthanization and necropsy was done 18 hours post-challenge and cecal content was then collected.

**Supplementary Figure 2 | 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).

**Supplementary Figure 3 | 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.

**Supplementary Figure 4 | Additional explanation for Figure 2 interpretation.** Relative abundance of transcription for *C. difficile* 630 genes during infection across the 3 antibiotic pretreatment models used during this study. Points that are located closer to a corner are more highly transcribed in the condition associated with that corner compared to the others. As this shows a 3-dimensional data set in 2 dimensions, there is an amount of distortion proximal to each corner. Simply put for points that are nearer to an edge, a greater percentage of their total transcription was contributed by *C. difficile* colonizing those mice. **(A)** This point represents the transcription for a gene that is overrepresented in cefoperazone-pretreated mice. **(B)** This point represents a gene in which transcripts are equally detectable in all 3 conditions. **(C)** Transcripts for this gene are only underrepresented in only cefoperazone-pretreated mice, and are equally detectable in clindamycin and streptomycin-pretreated animals.

**Supplementary Figure 5 | Within-group median sample variance for community-level data with replication.** Shown are the median and interquartile range of the sample variance for all fields in each experimental group. This was done to demonstrate consistent measurements in multiple levels of data (n = 9 per group). **(A)** OTU abundances from 16S rRNA gene sequencing, sample variances for each OTU were calculated individually prior to summary statistic calculations. **(B)** Scaled intensities from untargeted metabolomic analysis, sample variances for each metabolite were calculated individually prior to summary statistic calculations. In both groups of calculations all median sample variances are >1.0, indicating low levels of variability between samples of the same type.

**Supplementary Figure 6 | 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).

**Supplementary Table 1 | Sets of genes included in Figure 2 with normalized abundances and annotations.**

**Supplementary Table 2 | Topology metrics for enzyme and metabolite nodes in the *C. difficile* str. 630 metabolic network.**

**Supplementary Table 3 | All metabolites with significant importance scores in each colonized condition.**

**Supplementary Table 4 | Growth curve analysis for each carbon source.**

### 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.

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. **Tracy, B. P.**, **S. W. Jones**, **A. G. Fast**, **D. C. Indurthi**, and **E. T. Papoutsakis**. 2012. Clostridia: The importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. U.S. Patent 3.

15. **Songer, J. G.**, and **M. A. Anderson**. 2006. *Clostridium difficile*: An important pathogen of food animals. U.S. Patent 1.

16. **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.

17. **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.

18. **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.

19. **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.

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

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

22. **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**.

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. **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.

25. **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.

26. **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.

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

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

29. **Aboulnaga, E. 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.

30. **Ng, K. M.**, **J. a Ferreyra**, **S. K. Higginbottom**, **J. B. Lynch**, **P. C. Kashyap**, **S. Gopinath**, **N. Naidu**, **B. Choudhury**, **B. C. Weimer**, **D. M. Monack**, and **J. L. Sonnenburg**. 2013. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. Nature **502**:96–9.

31. **Almagro-Moreno, S.**, and **E. F. Boyd**. 2009. Insights into the evolution of sialic acid catabolism among bacteria. BMC Evol Biol **9**:118.

32. **Vital, M.**, **A. C. Howe**, and **J. M. Tiedje**. 2014. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. mBio **5**.

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

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

35. **Marcobal, A.**, **A. M. Southwick**, **K. A. Earle**, and **J. L. Sonnenburg**. 2013. A refined palate: Bacterial consumption of host glycans in the gut. U.S. Patent 9.

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

37. **Guimaraes, J. C.**, **M. Rocha**, and **A. P. Arkin**. 2014. Transcript level and sequence determinants of protein abundance and noise in Escherichia coli. Nucleic Acids Research **42**:4791–4799.

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

39. **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.

40. **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.

41. **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.

42. **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.

43. **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.

44. **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.

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. **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.

47. **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.

48. **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.

49. **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.

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