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

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

### Abstract

*Clostridium difficile* infection (CDI) has grown to be the most prevalent cause of hospital acquired infection in the United States. Susceptibility to CDI is induced by recent antibiotic exposure, which is known to alter the structure of the gut microbiome and to affect the availability of growth nutrients in the gut. We hypothesized that *C. difficile* is a generalist that adapts its physiology to the nutrients available within the gut. We orally challenged C57BL/6 mice that previously received one of four antibiotics with *C. difficile* and demonstrated that it was able to colonize the cecum within 18 hours of infection. However, levels of both spore and toxin production, which are known to be affected by nutrient availability, varied between each antibiotic treatment group. To more closely investigate the specific responses of *C. difficile* as it colonized the cecum, we performed *in vivo* transcriptional analysis of *C. difficile* from cecal content of infected mice. This approach revealed variation in expression of genes that drive life-cycle switches as well as metabolic pathways associated with catabolizing a variety of carbon sources such as carbohydrates, amino acids, and amino sugars. To assess which substrates *C. difficile* was most likely exploiting in each antibiotic-perturbed microbiome, we developed a novel metabolite scoring algorithm within the genome-scale bipartite metabolic network of *C. difficile* that incorporated both network topology and transcript abundance to infer the likelihood that a given metabolite was acquired from the environment. Applying this approach, we found that *C. difficile* indeed occupies alternative nutrient niches across each antibiotic-perturbed microbiome and that the highlighted metabolites support significant growth, *in vitro*. Results from this analysis support the hypothesis that consumption of N-acetyl-D-glucosamine and Stickland fermentation substrates are central components of *C. difficile*'s metabolic strategy and pathogenesis. This work has implications for elucidating specifics of the nutrient niche of *C. difficile* during infection and may lead to the discovery of targeted measures to prevent *C. difficile* colonization including potential pre- or 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 decade1. 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 20142. *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 CDI3. It has been shown that antibiotic therapy alters the structure and function of the gut microbiota making it susceptible to colonization by *C. difficile*4. This is referred to as colonization resistance, in which the gut microbiota inhibits the persistence or growth of a number of pathogenic bacteria. Colonization resistance can be achieved by multiple mechanisms including competition for physical space or growth nutrients5.

Mouse models have been an effective tool for studying the mechanisms of colonization resistance. Use of distinct antibiotic classes to vary the structure of the microbiota has been shown to result in varied levels of *C. difficile* colonization6–8. Additionally, 18 hours after being introduced to a cefoperazone treated mouse, *C. difficile* reached its maximum vegetative cell density in the cecum9. This provided a single timepoint to measure the largest population of metabolically active *C. difficile*. Building upon these results, others have shown that many of these antibiotic classes also alter the gut metabolome, increasing the concentrations of known *C. difficile* growth substrates7,10–12. 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 intestine13,14.

Based on its genome sequence and *in vitro* growth characteristics, *C. difficile* appears able to fill multiple nutrient niches. *C. difficile* has a relatively large and mosaic genome, it is amenable to a variety of growth substrates, and is able to colonize a diverse array of hosts suggesting that that it is a bacterial generalist15–17. The ability to metabolize a variety of substrates is important since these substrates affect the regulation of genes involved in *C. difficile*’s pathogenesis. For example, *in vitro* transcriptomic analysis suggests that high concentrations of easily metabolized carbon sources, such as glucose or amino acids, inhibit toxin gene expression and sporulation18,19. These genes are regulated by DNA-binding sigma factors, such as the pleiotropic regulator *ccpA*, which are under the control of environmental nutrient concentrations, especially carbohydrates20,21. Downstream effects of this regulation likely have enormous impact on the lifestyle and metabolic strategy of *C. difficile* when colonizing across sensitive hosts.

Previous transcriptomic studies of *C. difficile* have mainly focused on transcription of virulence factors, *in vitro*22,23, with some work characterizing transcription during colonization of germfree mice24,25. More relevant to nutrient acquisition, *C. difficile* up-regulated several phosphotransferase systems (PTS) and ABC transporters in germfree mice, alluding to metabolic adaptation to nutrient availability *in vivo*25. Although these analyses are informative, they are either primarily directed toward the expression of virulence factors or lack the context of the gut microbiota which *C. difficile* must compete against for substrates. Metabolomic analyses have also been used to more directly assay changes in bacterial metabolism as they relate to CDI7,12; however, these methods cannot focus on *C. difficile*-specific metabolites and more closely resemble echoes of metabolism, not currently active processes. In contrast to these approaches, *in vivo* *C. difficile* transcriptomic analysis from SPF animals may provide unique insight into its active metabolic pathways in a more realistic model of infection. Integrating transcriptomic data with genome-scale metabolic modeling has previously aided in identifying the most active aspects of an organism’s metabolism and which substrates are preferred by the organism26–28. Applying these methods to study *C. difficile* colonization would allow us to directly test the nutrient-niche hypothesis.

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 it focuses 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 different classes of antibiotics. The antibiotics used in this study included streptomycin (Fig. 1a), cefoperazone (Fig. 1b), and clindamycin (Fig. 1c). These antibiotics differentially affect the structure of the gut microbiota8. Each has also been shown to alter the gut metabolome relative to untreated animals7,10,12. As such, we predicted that *C. difficile* would encounter a unique subset of nutrients and competitors in each environment, which would necessitate distinct adaptive responses. To determine whether *C. difficile* is a generalist and differentially responds to each condition, we assayed for differences in the amount of sporulation and toxin activity phenotypes and used metabolic models built using *C. difficile* expression data. In each of the three antibiotic conditions we challenged with *C. difficile*, as well as in monoassociated germfree mice, we observed that *C. difficile* adapted its nutrient utilization profile to colonize to high levels and express its virulence factors.

### Results

**Levels of *C. difficile* sporulation and toxin activity vary between antibiotic-treated specific pathogen free and germfree mice.**

Due to the connection between metabolism, sporulation, and toxin production in *C. difficile*, we measured sporulation and toxin production at 18 hours post infection in each group (Fig. 1). There was not a significant difference in the number of vegetative cells between any susceptible condition tested (Fig. 2a). All antibiotic treated (Table 1) specific pathogen free (SPF) and germfree (GF) animals were colonized to ~1×108 colony forming units (c.f.u.) per gram of content, while untreated SPF mice maintained colonization resistance to *C. difficile*. Despite having the same number of vegetative *C. difficile* cells, large differences were detected in the density of *C. difficile* spores. Significantly more spores (*P* = 0.005, 0.008, 0.003) were detected in ex-germfree mice than in the antibiotic treated mice (Fig. 2b). The spore densities in both streptomycin and clindamycin-treated mice were also generally higher than that in cefoperazone-treated mice. There was significantly more toxin activity in ex-germfree animals than any other colonized group (*P* = 0.002, 0.002, 0.001), but toxin titer also varied between antibiotic treatment groups (Fig. 2c). Although similar toxin activity was found in both the cefoperazone and clindamycin-treated groups, toxin titer was below the limit of detection in most streptomycin-treated animals. These results indicate that *C. difficile* was able to colonize different communities to a consistently high level, but that the density of spores and toxin titer varied by treatment.

***C. difficile* adapts the expression of genes for virulence and key sigma factors that are under the control of environmental nutrient concentrations.** To more closely investigate the responses of *C. difficile* to colonizing distinct susceptible gut environments, we performed whole transcriptome analysis of *C. difficile* during infection of the antibiotic treatment models. We then narrowed our analysis to focus on genes that control or code for functions that have been linked to nutrient concentrations in the intestines during CDI. After observing differences in spore load, we first examined transcription of the most highly expressed genes in the *C. difficile* sporulation pathway29–32 (Fig. 3a). Across the four conditions where *C. difficile* colonized, we observed transcriptional profiles consistent with observed spore levels (Fig. 2b). The mice treated with cefoperazone had the lowest spore density and had the highest level of expression for the anti-sigma factors *spoVG* and *spoVS*. The procuts of these genes are involved in suppressing expression of genes found later in the sporulation pathway33. Streptomycin-treated mice had the next highest density of spores and the highest expression of genes associated with sporulation activation (*spoIIAB*/*spoIIE*), but they also had relatively high levels of expression of *sspA* and *sspB*, which are genes that code for effectors that protect DNA from damage during dormancy. Next, in mice treated with clindamycin, *C. difficile* expressed genes associated with late stages of sporulation, including those for spore coat components (*cdeC*, *cotD*, and *cotJB2*), spore formation (*spoIVA*, *spoVB*, and *spoVFB*), and *sspA* and *sspB*. Finally, GF mice harbored the highest density of spores and those *C. difficile* primarily expressed the dormancy genes linked with the latest stages sporulation. Together these data demonstrate that *C. difficile* differentially expressed genes associated with sporulation that corresponded to the presence of spores in the cecum.

Expression of genes for quorum sensing and pathogenicity have been linked to changes in the nutrients that can be found in the environment of *C. difficile*. Both the *agr* locus and *luxS* gene are thought to be associated with inducing the expression of *C. difficile* virulence in several strains34,35. Considering the link between quorum sensing genes and toxin production, we expected the expression of genes for quorum sensing and toxin production and toxin titer to be concordant. Based on this model, we expected GF mice to have the highest levels of expression of genes for toxin production (Fig. 3b) and quorum sensing (Fig. 3c); however, these transcripts were not found in the germfree mice. We also observed the highest level of expression for quorum sensing genes in cefoperazone-treated mice, but *tcdA* expression in these animals was not the highest among the different treatment groups. Interestingly, the levels of expression for genes associated with toxin production did not match the toxin titers observed in the animals. These results suggest that the relationship between toxin titer and the expression of genes for toxin production is even more complex than current models indicate.

We next focused on the regulators of metabolic pathways. Sigma factors are master regulators and a subset have been shown to integrate signals from intra- and extracellular nutrient concentrations20,21,31,36. The transcription of the global repressor *codY* is responsive to intracellular concentrations of GTP and certain amino acids, which are *C. difficile* energy sources37. Highest transcription for this gene was found in cefoperazone-treated and GF mice (Fig 3d). The regulation networks of CodyY and CcpA are highly interconnected, with the expression of *ccpA* specifically linked to local concentration of rapidly metabolizable carbon sources38. Cefoperazone-treated mice also exhibited increased transcription of *ccpA*, but the GF condition did not follow the same pattern. CcpA acts directly on *spo0A* (Fig. 3d), which positively regulates initiation of the sporulation pathway in *C. difficile*. Transcripts for *spo0A* were highly abundant in all conditions tested except for clindamycin-treated mice, where it was still moderately detectable. The sig-family of sigma factors is under the control of spo0A and regulate different stages of sporulation. The genes from this family with the highest total transcription (*sigA1*, *sigF*, *sigG*, *sigH*, and *sigK*) each demonstrated a unique pattern of expression between conditions. These results indicate that complete expression of sporulation likely integrates multiple levels of signaling and is more complex than a single metabolic switch. Both CcpA and Spo0A also control pathogenicity by regulating toxin production (Fig. 3d). We found expression of the toxin negative regulator *tcdC* in all of the antibiotic-treated groups, but no detectable transcripts for the positive toxin A/B regulator *tcdR* were seen in any treatment. In addition to its effects on sporulation and virulence, CcpA also regulates the expression of other sigma factors that generally mediate distinct forms of *C. difficile* metabolism as needed. These targets include *rex* (general fermentation regulator) and *prdR* (Stickland fermentation regulator) (Fig. 3d). Although the expression of both has been shown to be linked to environmental proline concentrations, *rex* integrates additional signals from the intracellular NADH/NAD+ ratio to also control carbohydrate fermentation. Low-level transcription of *prdR* was found across all conditions, however *C. difficile* expression the *rex* gene highly in both cefoperazone-treated and GF mice. Combined, the variable expression of these sigma factors support the hypothesis that *C. difficile* adapts expression metabolism to fit its needs between colonized environments.

**Gene sets from multiple *C. difficile* metabolic pathways are differentially expressed between colonized environments.** In the context of similar colonization between antibiotic-treated animals, differential expression of global metabolic control mechanisms that are under the control of specific nutrient concentrations suggests that *C. difficile* adapts to each environment when in competition with the resident microbiota. To test this further, we quantified the total expression for all KEGG annotated genes in the *C. difficile* genome (Fig. S1a). We then focused on general differences in carbohydrate (Fig. S1b) and amino acid (Fig. S1c) metabolism in order to asses for apparent differences in the utilization of carbon sources by *C. difficile* across environments. However aside from overall lower expression of most gene families in germfree mice, no other clear trends were evident at this broad level of analysis so we moved toward a more fine-scale resolution of annotation and focused on specific gene sets known to contribute to certain forms of *C. difficile* metabolism (Table S1). Also, to more effectively compare between colonized states, we calculated the percentage of total expression between antibiotic-treated conditions for each gene (Fig 4). We then identified the condition in which each gene was most highly transcribed and adjusted the size of the corresponding point relative to the largest transcipt abundance (Fig 4a). This demonstrated that genes involved in amino acid catabolism had the greatest amount of expression overall, relative to other gene sets. This category includes those enzymes involved in Stickland fermentation (*arg*, *fdh*, *grd*, and *prd* loci) as well as several general peptidases (*pep* family). These results indicated that catabolizing environmental amino acids may be important for the growth of *C. difficile* during infection.

To more clearly indetify associations of gene sets with each condition, we also analyzed each set separately. First, we found that the expression of genes associated with amino acid catabolism were expressed at nearly consistent levels across the conditions (Fig. 4b). This was in agreement with the high level of overall expression associated with these genes. Additionally, genes for the metabolism of the host-derived amino sugars N-acetylglucosamine and N-acetylmannosamine were also expressed at consistent levels across each treatment group (*glm*, *nan*, *mur*, and *acd* loci) (Fig. 4c). Along similar lines with related molecules, a number of genes for certain monosaccharides entering (*gal*, *man*, *pmi*, and *tag* loci) and leading through glycolysis (*fba*, *fbp*, *gap*, and *pfk*), as well as catabolism of the polysaccharides trehalose and cellibiose (*treA* and *celG*) were expressed relatively evenly between each condition (Fig. 4d & 4e). Combined, these findings suggest that catabolism of amino acids and specific carbohydrates are likely core components of the *C. difficile* nutritional strategy during infection.

Aside from those gene sets that were equally expressed across conditions, there were also large scale differences in expression of certain pathways between groups of mice. We chose to assess sugar transport systems since CodY has been associated with adaptive expression of phosphotransferase systems (PTS) and ABC transporters with many known differences in substrate specificities. Among the genes classified as PTS transporters (Fig. 4f) were overrepresented in both clindamycin and streptomycin-treated mice, while ABC sugar transporters (Fig. 4g) were overrepresented in the cefoparazone-treated mice. The most stark differences were seen in transcription for genes involved in sugar alcohol catabolism (Fig. 4h). Expression of these genes was entirely absent from clindamycin-treated mice and expression of genes for mannitol utilization (*mtl* operon) were overrepresented in cefoparazone-treated mice and expression of genes for sorbitol utilization (*srl* operon) were overrepresented in streptomycin-treated mice. Concordant patterns also emerged in genes associated with fermentation end steps (Fig. 4i) and polysaccharide degradation (Fig. 4e). Short chain fatty acids (SCFAs) and alcohols are the end products of both carbohydrate and amino acid fermentation in *C. difficile* through separate pathways with shared terminal steps. Transcripts for genes involved in *C. difficile* butyrate/butanol metabolism (*ptb*, *buk1*, *cat2*, and *adhE*) were more abundant in clindamycin-treated mice (Fig. 4i). Additionally, alpha/beta-galactosidase genes (*aglB* and *bglA*) were also overrepresented in clindamycin-treated mice (Fig. 4e). Together these patterns support polysaccharide fermentation in this condition. More subtle differences were seen in those gene associated with glycolysis (Fig. 4d). This category includes genes for not only the steps of glycolysis, but also several genes that mediate entry points of monosaccharides to this pathway. Transcripts for several genes in this group (*eno*, *gapA*, *gpmI*, *tpi*, and *pyk*) were overrepresented in cefoparazone-treated mice, however *fruK* was overrepresented in streptomycin-treated mice which catalyzes the committed step of glycolysis. Overall, these results support the hypothesis that *C. difficile* is able to adapt its metabolism to fit the nutrient availability across different susceptible environments.

**Structure of genome-scale bipartite metabolic model underscores known bacterial metabolism.** 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 annotation. Enzymes and metabolites were represented by nodes and their interaction by the edge between the nodes (Fig. 5a). To validate our metabolic network, we calculated betweenness centrality (BC) and overall closeness centralization index (OCCI) for all enzyme and metabolite nodes in the bipartite metabolic network of *C. difficile* generated for this study (Table S2). In biological terms, BC reflects the amount of influence a given hub has on the overall flow of metabolism through the network39 and OCCI indicates those enzymes and substrates that are the most central components of the organism’s metabolism40. For both enzymes and substrates, the 18 of top 20 nodes with the highest BC values were involved in glycolysis, fermentation, and amino acid synthesis. In agreement, almost all nodes with the largest OCCI values were involved in glycolysis and amino acid synthesis as well. Enzymes that scored highly in both metrics included pyruvate kinase, aspartate aminotransferase, and formate C-acetyltransferase while substrates consistently scoring most highly were pyruvate, acetyl-CoA, D-glyceraldehyde 3-phosphate. This indicated to us that the topology of the network reflects established bacterial physiology.

**Metabolite importance algorithm reveals adaptive nutritional strategies of *C. difficile* during infection across distinct environments** Moving beyond a strictly topological analysis of the *C. difficile* metabolic network, we sought to utilize transcriptomic data to infer which metabolites *C. difficile* is most likely obtaining from its environment in each condition. To accomplish this we mapped normalized transcript abundances to the enzyme nodes in the network. Due to the coupling of transcription and translation in bacteria, we were able to use this information as a proxy for enzyme levels. The importance of each metabolite was measured as the log-transformed difference between the average transcript levels of enzymes that use the metabolite as a substrate and those that generate it as a product (Fig. 5b). A metabolite with a high importance score is most likely obtained from the environment because the expression of genes for enzymes that produce the metabolite are low. Then, using a Monte Carlo-style simulation, we generated a random transcript abundance distribution for each enzyme node to then calculate new metabolite importance scores for each iteration. We were then able to create a confidence interval of scores for each metabolite that would likely result from random noise41. This provided a standard of comparison for actual importance scores from single timepoint measurments, and ultimately allow for computing the significance level that a given score has a high probability of being excluded from its associated null hypothesis score distribution.

Applying these methods to the *C. difficile* transcriptomic data collected from the *in vivo* CDI models, we sought to assess differential patterns of metabolite importance. We first ranked the importance scores to identify the most important metabolites for each treatment group (Table S3). To identify the core metabolites that are essential to *C. difficile* in any condition, we compared the highest 50 scoring, significant metabolites from each treatment group (*P* < 0.05) (Fig. 6a). The host derived amino sugar N-acetyl-D-glucosamine was found to be consistently important, indicating that it may be an integral component of the nutrient niche for *C. difficile* under any condition or that it is always available in the cecum. Additionally, acetate was found to be important in all conditions, but was just below the significance cutoff in germfree mice (Table S3). It has been shown that *C. difficile* produces acetate during fermentation and metabolizes it effectively to utilize in glycolysis42. We directly tested the relative concentration of acetate in cefoperazone-treated *C. difficile*-infected mice versus mock-infected mice. We found that *C. difficile* colonization led to a significant decrease in the levels of acetate (Fig. S3) suggesting that *C. difficile* was utilizing acetate in the cecum. These findings provided validation for our metabolite importance algorithm as well as supporting known elements of *C. difficile* metabolism.

Returning to our hypothesis that *C. difficile* adapts its metabolism to fit the surrounding community, we identified those metabolites that were uniquely important to each condition in which *C. difficile* colonized. We cross-referenced the top 25 positively scoring, significant substrates (*P* < 0.05) between treatment groups to uncover the most important patterns of nutrient utilization by *C. difficile* in each (Fig. 6b). Each group of metabolites contained at least one known carbohydrate growth substrate of *C. difficile*7,43. This included close analogs of D-fructose, mannitol, N-acetylneuraminic acid, and salicin. Furthermore, in GF mice where no other competitors are present, our model indicated that *C. difficile* was more likely to acquires several amino acids (lysine, leucine, and isoleucine) from the environment instead of expending energy to produce them itself. These data support the hypothesis that *C. difficile* may exploit alternative nutrient sources between the susceptible environments it colonizes.

**Carbon sources sources predicted to be important using network-based approach support *C. difficile* differential growth *in vitro*.** To validate the biological relevance of substrates identified as uniquely important to *C. difficile* metabolism through our network-based analysis, we tested whether *C. difficile* was able to utilize each substrate for *in vitro* growth (Fig. 6c). This was performed using a modified defined *C. difficile* minimal media7, supplemented individually with the selected carbohydrates implicated by high importance scores. Because *C. difficile* can use amino acids for growth through Stickland fermentation, but is auxotrophic for several amino acids, it was necessary to include amino acids in the minimal media. This meant that the most effective negative control was growth in media lacking carbohydrates but containing amino acids (Max OD600 = 0.212).

N-acetyl-D-glucosamine was shown to be important to *C. difficile* in each condition tested (Fig. 6b). When tested for improved growth, significantly more growth (Max OD600 = 0.774) was observed compared to no carbohydrate controls (*P* << 0.001). This provided evidence that N-acetyl-D-glucosamine, derived from the host mucus layer, may be a central component of the *C. difficile* nutritional niche during infection. Trehalose was also shown to be important in each condition and supported *C. difficile* significant growth (*P* << 0.001; Max OD600 = 0.559), but was more likely provided by the diet than from the host. Furthermore, at least one carbohydrate highlighted as distinctly more important in each of the antibiotic treatment groups provided high levels of *C. difficile* growth relative to control wells (*P* << 0.001). This included D-fructose (streptomycin; Max OD600 = 0.671), mannitol (cefoperazone; Max OD600 = 0.464), salicin (clindamycin; Max OD600 = 0.869), and N-acetylneuraminate (GF; Max OD600 = 0.439). Because it was not possible to test aminofructose directly, we instead chose to test fructose, an immediate breakdown byproduct of aminofructose catabolism. We also tested both starch and acetate for the ability to support *C. difficile* growth *in vitro*, but neither should any improvement over No Carbohydrate control (Fig. S4). Maximum growth rate analysis for each carbohydrate also indicated potential hierarchy of growth nutrient preference (Table S4). The progression is as follows: D-fructose (slope = 0.089), N-acetyl-D-glucosamine (slope = 0.085), salicin (slope = 0.077), mannitol (slope = 0.044) / trehalose (slope = 0.044), and finally N-acetylneuriminate (slope = 0.024). This suggested that *C. difficile* was most well-suited to metabolize the nutrient source that is most likely to be present in all susceptible mouse ceca.

### Discussion

Collectively, our results support the hypothesis that *C. difficile* can adapt its metabolism to the available niche landscape across susceptible gut environments. Mice treated with different antibiotics harbored distinct microbial communities and when *C. difficile* colonized these communities it altered its gene expression. Mapping this gene expression onto a metabolic network model for *C. difficile* allowed us to quantify the importance of different metabolites acquired from its environment. That the most important metabolites differed between antibiotic treatment groups indicated a shift in *C. difficile* metabolism. Furthermore, we validated these results by growing *C. difficile* in media where important metabolites were the sole carbohydrate and observed robust growth. These results give insight to the adaptive strategies that *C. difficile* can use to colonize diverse human microbiota.

By integrating multiple levels of data into a transcriptome-enabled, genome-scale model of *C. difficile* metabolism we were able to obtain a deeper understanding of the pathogen’s behavior in different contexts of infection. Comparative studies on gene content alone enabled us to identify core genomic elements and putative colonization factors44,45. Although these are important steps in directing future research, they lack an expression component to identify when these functions are likely to be produced by a cell. Transcriptomic analysis is more informative than a genome, since it allows one to quantify the differential expression of genes under distinct conditions and can uncover patterns of adaptation24. Transcriptomic analysis has traditionally been limited since the scope is typically the expression of single gene families or pathways without concern for how they interact with other pathways. Our network-based approach instead focused on uncovering those resources that *C. difficile* acquires from the environment using an integrative approach in which all reactions that are annotated to interact with a given metabolite are incorporated into the importance of that metabolite at the time point measured. In this way we were able to amplify the signal of demand for each metabolite and more reliably identify what resources a bacterium is importing from its environment. For example, a traditional transcriptomic analysis indicated that sorbitol is the most likely carbon source of *C. difficile* when infecting streptomycin-treated mice (Fig. 4d). However, upon a network-based analysis it became clear that other carbon sources that are more likely to be used for growth in this condition (Fig. 5b) and may be due to increased availability compared to other metabolites. To this point, when we analyzed sorbitol as a growth substrate during the *in vitro* experiments we saw that it indeed did not provide for increased growth over No Carbohydrate control and that the network-based method provided a more robust prediction of nutrient preference (Fig. S4).

Our hypothesis is supported by previous mass spectrometry-based efforts analyzing the metabolome from mouse intestinal content treated under similar conditions to those used in the current study. These investigations revealed that several of the substrates predicted to be used by *C. difficile* in a given condition through metabolic modeling (Fig. 6a & 6b), are increased in the gastrointestinal tract of mice in the corresponding treatment group. One recent study found that cefoperazone treatment resulted in a 553-fold increase in mannitol concentration in the cecum of mice prior to *C. difficile* colonization7. Similar trends of increased concentration for *C. difficile* growth substrates are also true in streptomycin-treated conventional (amino sugars and fructose10) and GF mice (N-acetylneurminate46). Interestingly, the capacity for bacteria to catabolize the amino sugar N-acetylneurminate is highly distributed among species that inhabit the mammalian gut47. This may mean that N-acetylneuraminate is strongly competed for by the healthy microbiota and *C. difficile* can only access it in the absence of all other competitors. Together, these results provide evidence that our network-based approach accurately predicts which metabolites *C. difficile* chooses to metabolize in a given environments and that these changes are most likely due to nutrient availability.

Additionally, the finding that N-acetyl-D-glucosamine is consistently the highest scoring shared metabolite among all tested conditions strongly indicated that it may be always utilized by *C. difficile* during infection. As it is likely constitutively available at some concentration, and is a source of both carbon and nitrogen (a potentially limited resource in the mammalian lower GI tract48) also provides evidence that this is true. A conserved strategy may also be evident in that numerous components of the Stickland fermentation pathway were found to be important to *C. difficile* in all conditions tested. Proline is one of the primary substrates the *C. difficile* uses to generate ATP form amino acids, and the first steps in the pathway initialized by a proline reductase which requires selenium as a cofactor49, eventually leading through the intermediate 3−hydroxybutanoyl−CoA ultimately toward the production of butyrate50. Since the diet that the mice are given is rich in both amino acids and supplemented with selenium, it is not unlikely that this pathway may be more energetically favorable in this instance than other forms of metabolism. Additionally, formate has been shown to greatly increase the efficiency of Stickland fermentation in Clostridia by decreasing intracellular H+ concentrations and allowing for increased production of short chain fatty acids51. These findings that there are also metabolites that *C. difficile* will utilize in any condition and are central to its nutritional strategy.

In addition to exploring differential patterns in known metabolism, our modeling approach also allowed for the identification of emergent properties for the metabolic strategy of *C. difficile* during infection (Fig. 6a). One interesting result is the appearance of CO2, an apparent metabolic end product, in the list of shared important metabolites. While this may be a shortcoming inherent in database-driven research, it could instead serve as evidence for yet unappreciated aspects of a bacterium’s metabolism. Along these lines, one group has posited that *C. difficile* may actually be autotrophic under certain conditions which could explain the appearance of CO2 in the conserved metabolite list between all conditions52. Furthermore, oxygen appears as important in clindamycin-treated mice. Despite the fact that *C. difficile* is considered a strict anaerobe, it has been shown to be able to resist oxidative stress through expression of multiple putative oxidoreductases53. Increased reactive oxygen species may be induced by residual antibiotic affect host or other bacterial cells**???**. It is important to note that the metabolites highlighted by our method are not necessarily growth substrates, only how probable it is that they are being removed from the environment which encompasses possible detoxification. As such, this platform may also prove informative for generating hypotheses through reverse ecology**???** that could ultimately lead to uncovering new edges of interaction between species.

While much of the results presented agree with previously published findings on the metabolism of *C. difficile*, there are some potential limitations of primarily database-driven methods. Ultimately, the metabolite importance calculation is dependent on gene annotation. Genes may be annotated incorrectly, possibly giving a skewed version of that organism’s overall metabolism. Concordantly, this also implies that our approach only accounts for those enzyme-mediated reactions we are currently aware of. For this reason, we instead rely on emergent properties from the combination established metabolic reactions contained in the chosen database. In this regard, other groups have shown the the wider pathway annotations in KEGG to be more robust to missing elements and well-suited the the type of analysis described here54. Additionally, due to the topology of the metabolic network where each unique metabolite only appears a single time, we were unable to integrate stoichiometry information for each reaction which may have a dramatic effect the rate of consumption/production. However, results from this analysis reflects known aspects of *C. difficile* catabolism and indicates that the lack of this information may not greatly impact our ability to identify important metabolites. Aside from limitations due to annotation, our network-based approach makes a set of basic simplifications of bacterial metabolism as well. First, this method also operates under the assumption that all detectable transcript is then translated to complete effector proteins. While this is not completely accurate, since bacterial transcription and translation are physically coupled, we were comfortable using normalized levels of transcription to infer approximate amount of translation. Second, the metabolite importance scores do not account for the amount of each metabolite that is available. Future studies could employ metabolomic analysis which would allow for confirmation that important metabolites are actually present for use. Finally, the importance algorithm only consider the transcription of those enzyme nodes immediately adjacent to the metabolite node of interest. Although this does not negate any observations made in the current study, it may be beneficial to incorporate the importance of other local metabolites or subnetworks into the final score of each metabolite. In spite of these assumptions, the method outlined here supports known elements of *C. difficile* biology and reveals the adaptive nature of the pathogen’s nutritional strategy during infection.

Based on the evidence presented, and in spite of the limitations discussed above, our results strongly support our hypothesis that *C. difficile* metabolizes alternative carbon sources across susceptible gut environments for growth. Together, our results demonstrate that our novel metabolic modeling approach highlights which substrates an organism is likely using at the time of sampling. Ultimately, these genome-scale metabolic models provide a framework to predict phenotype changes of an organism across environments. Our results also indicate that further considerations are needed when attempting to design more targeted probiotics for the prevention or elimination of *C. difficile* from the human gut. Different classes of antibiotics may each result in a distinct gut environment which *C. difficile* can exploit. Therefore, the best approach may be to consider each of these as a separate problem which different collections of bacteria can restore to a resistant state.

### Methods

**Animal care and antibiotic administration.** Adapted from the previously described model55, six-to-eight week-old specific pathogen free (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.** *C. difficile* strain 630 spores were prepared from a single large batch whose concentration was determined a week prior to challenge for all experiments. 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 c.f.u. was performed to ensure correct dosage. Infection negative control animals were given an oral garage 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. Necropsy was then performed and cecal content was split into three small aliquots (~100 μl). Two were flash frozen immediately for later DNA extraction and toxin titer analysis respectively. The third aliquot was quickly moved to an anaerobic chamber for c.f.u. quantification. The remaining content in the ceca (~1 ml) was emptied into a stainless steel mortar in a dry ice/ethanol bath using 1 ml of sterile PBS. This process was repeated for each mouse within a treatment group to pool content into a single large sample (9 mice across 3 cages) to compensate for cage effects as much as possible while maximizing sequencing depth. The 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 samles on CCFAE plates (fructose agar plus cycloserine, cefoxitin, and erythromycin) at 37° C for 24 hours under anaerobic conditions56. It is important to note that the germination agent taurocholate was omitted from these plates in order to only quantify vegetative cells. In parallel, undiluted samples were heated at 60° C for 30 minutes to eliminate vegetative cells and leave only spores57. 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 performed58. Briefly, filtered-sterilized cecal content was serially diluted 1:5 in PBS. As a control for toxin-mediated cell rounding the cecal content was diluted a further 1:2 by the addition of an equal volume of goat anti-toxin serum (T5000; TechLab). Vero cells were grown to a confluent monolayer in DMEM (Dulbecco's Modified Eagle's medium), supplemented with 10% heat-inactivated fetal bovine serum and 1% penicillin-streptomycin. The cells then were transferred to a conical tube and centrifuged at 1,000 rpm for 5 minutes to pellet the cells. The old media was removed and the cells were re-suspended in fresh media to a final concentration of 1×105 cells per 90μL. 90μL of the cell suspension were seeded in each well of a 96-well plate and incubated at 37° C in a 5% CO2 humidified incubator for 4 hours. Following the incubation, cecal samples were added to the Vero cells and the plate was incubated overnight at 37° C. Plates were viewed after 24 hours at 10x magnification for cell rounding. The cytotoxic titer was defined as the log10 transformed reciprocal of the highest dilution that produced rounding in 80% of the cells. A more detailed protocol with product information can be found at: <https://github.com/jlleslie/Intraspecific_Competition/blob/master/methods/Verocell_ToxinActivity_Assay.Rmd>

**RNA extraction, 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 bath59. 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 (Kozich, 2013). Quality and integrity of RNA was assessed 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 CutAdapt35 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 Bowtie260 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 Genomes61. 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 Samtools62 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 constructed using KEGG 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 (ex. 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 do not have annotations for the chemical compounds the interact with are discarded. Metabolites were then associated each enzyme in which they interact with, directionality, and reversibility of each biochemical conversion is also saved. This process is repeated for all enzymes in the given bacterial genome, with each enzyme and metabolite node only appearing once. The resulting data structure is an associative array of enzymes associated with lists of both categories of substrates (input and output), which can then be represented as a bipartite network. The final metabolic network of *C. difficile* str. 630 contains a total of 1205 individual nodes (447 enzymes and 758 substrates) with 2135 directed edges. Transcriptomic mapping data is 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. 5a) favors metabolites that are more likely acquired from the environment (not produced within the network), and will award them a higher score (Fig. 6b & 6c). The presumption of our approach is that enzymes that are more highly transcribed are more likely to utilize the substrates they act on due to coupled bacterial transcription and translation. If a compound is more likely to be produced, the more negative the resulting score will be. To calculate the importance of a given metabolite (m), we used rarefied transcript abundances mapped to respective enzyme nodes. This is represented by to and ti to designate if an enzyme creates or utilizes m. The first step is to calculate the average expression of enzymes for reactions that either create a given metabolite (i) and consume that metabolite (ii). For each direction, the sum of transcripts for enzymes connecting to a metabolite are divided by the number of contributing edges (eo or ei) to normalize for highly connected metabolite nodes. Next the raw metabolite importance score is calculated by subtracting the creation value from the consumption value to weight for metabolites that are likely acquired exogenously. The difference is Log2 transformed for comparability between scores of individual metabolites. This results in a final value that reflects the likelihood a metabolite is acquired from the environment. Untransformed scores that already equal to 0 are ignored and negative values are accounted for by transformation of the absolute value then multiplication 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 Bootstrapping 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 are constitutively important. Additionally, an approach of this type is a powerful tool for hypothesis generation and can provide evidence for future lines of investigation into the role of specific metabolites in the metabolism of bacteria in a given environment. 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 in order to ultimately generate a p-value for each metabolite. This is 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 comparisons63. These calculations are also included within the standard bigSMALL workflow presented above.

**Measuring *in vivo* concentrations of acetate.** Cecal content was flash frozen in liquid nitrogen at the time of necropsy and sent for short chain fatty acid analysis analysis using GC-MS (gas chromatography–mass spectrometry) as described in the targeted metabolomics section of Theriot et al., 2014. All assays were performed at the Michigan Regional Comprehensive Metabolomics Resource Core in Ann Arbor, MI.

**Anaerobic *in vitro* *C. difficile* growth curves.** The carbon-free variation of *C. difficile* Basal Defined Medium (NCMM) was prepared as previously described7. Individual carbohydrate sources were added at a final concentration of 5 mg/mL, also as outlined in the aforementioned study. Pair-wise carbohydrate combinations were added at a respective concentration of 2.5 mg/mL (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 aliquoted 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.

**Statistical Methods.** All statistical analyses were performed using R (v.3.2.0). Significant differences of CFU, toxin titer, and acetate concentration were determined by Wilcoxon test with Holm-Bonferroni correction. Significant differences for growth curves compared to No Carbohydrate control were calculated using 2-way ANOVA with Holm-Bonferroni correction. Significance for metabolite importance scores was determined as described above.

### Acknowledgements

The authors would like to acknowledge members of the University of Michigan germfree Mouse Center, Sequencing Core, and Metabolomics Core for their assistance in experimental design, execution, and data collection. This work was supported by funding from the National Institutes of Health to P. Schloss (R01GM099514, P30DK034933), the Enterics Research Investigators Network (ERIN; U19AI09087), and a Translational Research Education Certificate grant to Jhansi Leslie (MICHR; UL1TR000433). 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 sequence data to the final manuscript are hosted at <http://www.github.com/SchlossLab/Jenior_Modeling_NatMicro_2016>.

### Author Information

**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.

**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.

**Competing interests** The authors declare no competing interest.

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

### Figures

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

**Figure 1 | Experimental time lines for mouse model pretreatments and *C. difficile* infection.** Nine wild-type C57BL/6 mice were included in each pretreatment 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 at the time of infection. Sacrifice and necropsy was done 18 hours post-challenge and cecal content was then collected.

**Figure 2 | *C. difficile* sporulation and toxin activity quantification after 18 hours of infection.** ***(a)*** Vegetative *C. difficile* c.f.u. per gram of cecal content (n=9 mice per group). No significant differences were observed in between any group colonized by *C. difficile*. **(b)** *C. difficile* spore c.f.u. per gram of cecal content. Significantly more spores were detectable in GF mice compared to any of the antibiotic-treated SPF groups (*P* < 0.05). **(c)** Toxin titer from cecal content measured by activity in Vero cell rounding assay (n=9 mice per group). GF mice also displayed significantly more toxin activity than all other groups (*P* < 0.05). Untreated mice in a,b,c had no detectable *C. difficile* or toxin activity and were significantly different from all other groups in each assay (*P* << 0.01). Median values are shown for each group with significant differences calculated using Wilcoxon rank-sum test with the Holm-Bonferroni correction. Dotted lines denote the limit of detection for both assays, and undetectable points are shown just below the limit of detection for clarity.

**Figure 3 | Select *C. difficile* gene set expression compared between treatment group.** Relative abundances of *C. difficile* transcript for specific genes of interest, rarefied to 90% of the total number of reads within each colonized condition shown. **(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 4 | *C. difficile* expression of gene sets for carbon metabolism pathways across antibiotic pretreatments.** **(a)** Ternary plot indicating the relative abundance of transcripts for all *C. difficile* str. 630 genes across the three colonized antibiotic-treated conditions (gray points). Raw transcript abundances were iteratively rarefied and the median expression of each gene was calculated (~24x coverage). Each point represents a unique gene from the annotated genome of *C. difficile* str. 630 with position reflecting the ratio of transcription for that gene in all three antibiotic pretreatments. Transcripts for genes that are over-represented in a single condition are placed more proximal to the corner associated with that treatment group. Points placed near the center are equally expressed across all of the conditions measured at 18 hours post-infection. Points are colored based on inclusion in specific carbon metabolic pathways, and point sizes within groups of interest were determined based on the highest expression value for each gene from a single condition. **(b – i)** Groups from (a) are shown individually, without abundance information, for ease of comparison. Genes included in each group with normalized transcript abundances can be found in Table S1, and refer to Fig. S2 for additional explanantion of figure interpretation.

**Figure 5 | Genome-scale bipartite metabolic modeling results using the transcriptome of *C. difficile* str. 630 in each colonized environment.** **(a)** Largest component from the bipartite genome-scale metabolic model of *C. difficile* str. 630. The complete network contains 447 enzymes and 758 metabolites, with 2135 directed edges. Size of enzyme nodes is relative to the number of normalized reads mapped to the corresponding gene. The sizes shown reflect the transcriptome of *C. difficile* str. 630 during infection of cefoperazone-treated mice after 18 hours of infection. Below the representative network is the metabolite importance algorithm separated into 3 components; (i) relative transcription of reactions consuming a metabolite, (ii) relative transcription of reactions consuming a metabolite, and (iii) difference of consumption and creation of the given metabolite. **(b)** The expanded window displays an example of a single metabolite importance calculation based on local enzyme gene transcription. White values in the red nodes represent the number of normalized transcript reads mapping to the gene sequence for each enzyme node. Average expression of input and output reactions surrounding metabolite **m** are calculated at then the difference of these values found to get the relative importance of **m**. Log2 transformation is then performed for uniform comparison between metabolites.

**Figure 6 | Results from network-based metabolite importance calculation and *in vitro* growth with important carbohydrates.** Prior to importance calculation, transcript abundances for each condition were evenly rarefied for even comparison across colonized environments (~18x coverage). **(a)** Median shared significant metabolites among the 50 highest scoring metabolites from each condition (*P* < 0.05). Median importance scores and pooled random distribution were recalculated per metabolite using the scores from each condition tested. **(b)** Distinctly important significant metabolites from each treatment group (*P* < 0.05). The top 25 scoring metabolites from each group was cross-referenced against each other group resulting in metabolites that are differentially important between environments. **(c)** *in vitro* growth curves validating identified growth nutrients from network analysis. One metabolite that is consistently important to *C. difficile* and at least one metabolite indicated as distinctly important from each group supported growth significantly more (*P* << 0.001) than no carbohydrate control (+ amino acids, gray line). Only those carbon sources that significantly improved *C. difficile* growth over control are displayed (remainder are located in Table S4). Significant differences were calculated using 2-Way ANOVA with Holm-Bonferroni correction.

**Supplementary Figure 1 | Expression of specific KEGG gene families.** Abundances of normalized transcriptomic reads from *C. difficile* str. 630 in each tested condition. **(a)** All KEGG families. **(b)** Those sub-families within Carbohydrate metabolism. **(c)** Sub-families within Amino acid metabolism.

**Supplementary Figure 2 | Additional explanation for Figure 4 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-treated 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-treated mice, and are equally detectable in clindamycin and streptomycin-treated animals.

**Supplementary Figure 3 | *in vivo* acetate concentrations with GC-MS analysis from cefoperazone-treated mouse cecal content.** 2 groups of 5 mice each were pretreated with cefoperazone as outlined Fig. 1b. A single cage was infected with *C. difficile* in the same fashion as other experiments described here. Animals were necropsied at 18 hours post-infection and cecal content was flash frozen for later GC-MS analysis. Significance was determined using Wilcoxon signed-rank test.

**Supplementary Figure 4 | Additional growth curves for additional carbon sources and controls.** Significant metabolites from network analysis that did not provide improved growth over no carbohydrate ( + amino acids) control. Also included is the negative control of minimal media with no amino acids as well as *C. difficile* growth in standard Brain-Heart Infusion broth.

**Supplementary Table 1 | Sets of genes included in Figure 4 with normalized abundances and citations.**

**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 important scores for *C. difficile* in each colonized condition.**

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

### References

1. Lessa, F. C., Gould, C. V. & McDonald, L. C. 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 (2012).

2. Lessa, F. C. *et al.* Burden of Clostridium difficile Infection in the United States. *The New England journal of medicine* **372,** 825–834 (2015).

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

4. Young, V. B. & Schmidt, T. M. Antibiotic-Associated Diarrhea Accompanied by Large-Scale Alterations in the Composition of the Fecal Microbiota. *Journal of Clinical Microbiology* **42,** 1203–1206 (2004).

5. Lawley, T. D. & Walker, A. W. Intestinal colonization resistance. *Immunology* **138,** 1–11 (2013).

6. Chen, X. *et al.* A Mouse Model of Clostridium difficile-Associated Disease. *Gastroenterology* **135,** 1984–1992 (2008).

7. Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. (2014). doi:[10.1038/ncomms4114.Antibiotic-induced](https://doi.org/10.1038/ncomms4114.Antibiotic-induced)

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

9. Koenigsknecht, M. J. *et al.* Dynamics and establishment of Clostridium difficile infection in the murine gastrointestinal tract. *Infection and Immunity* **83,** 934–941 (2015).

10. Antunes, L. C. M. *et al.* Effect of antibiotic treatment on the intestinal metabolome. *Antimicrobial Agents and Chemotherapy* **55,** 1494–1503 (2011).

11. Ferreyra, J. A. *et al.* Gut microbiota-produced succinate promotes C. Difficile infection after antibiotic treatment or motility disturbance. *Cell Host and Microbe* **16,** 770–777 (2014).

12. Jump, R. L. P. *et al.* Metabolomics analysis identifies intestinal microbiota-derived biomarkers of colonization resistance in clindamycin-treated mice. *PLoS ONE* **9,** (2014).

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

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

15. Sebaihia, M. *et al.* The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature genetics* **38,** 779–786 (2006).

16. Tracy, B. P., Jones, S. W., Fast, A. G., Indurthi, D. C. & Papoutsakis, E. T. Clostridia: The importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. **23,** 364–381 (2012).

17. Songer, J. G. & Anderson, M. A. Clostridium difficile: An important pathogen of food animals. **12,** 1–4 (2006).

18. Neumann-Schaal, M., Hofmann, J. D., Will, S. E. & Schomburg, D. Time-resolved amino acid uptake of Clostridium difficile 630erm and concomitant fermentation product and toxin formation. *BMC Microbiology* 281 (2015). doi:[10.1186/s12866-015-0614-2](https://doi.org/10.1186/s12866-015-0614-2)

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

20. Antunes, A. *et al.* Global transcriptional control by glucose and carbon regulator CcpA in Clostridium difficile. **40,** 10701–10718 (2012).

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

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

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

24. Janoir, C. *et al.* Adaptive strategies and pathogenesis of clostridium difficile from In vivo transcriptomics. *Infection and Immunity* **81,** 3757–3769 (2013).

25. Kansau, I. *et al.* Deciphering adaptation strategies of the epidemic clostridium difficile 027 strain during infection through in vivo transcriptional analysis. *PLoS ONE* **11,** (2016).

26. Patil, K. R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of the Sciences of the United States of America* **102,** 2685–2689 (2005).

27. Borenstein, E., Kupiec, M., Feldman, M. W. & Ruppin, E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proceedings of the National Academy of Sciences* **105,** 14482–14487 (2008).

28. Suthers, P. F. *et al.* Genome-scale metabolic reconstruction Of mycoplasma genitalium, iPS189. *PLoS Computational Biology* **5,** (2009).

29. Putnam, E. E., Nock, A. M., Lawley, T. D. & Shen, A. SpoIVA and sipl are Clostridium difficile spore morphogenetic proteins. *Journal of Bacteriology* **195,** 1214–1225 (2013).

30. Underwood, S. *et al.* Characterization of the sporulation initiation pathway of Clostridium difficile and its role in toxin production. *Journal of Bacteriology* **191,** 7296–7305 (2009).

31. Fimlaid, K. A. *et al.* Global Analysis of the Sporulation Pathway of Clostridium difficile. *PLoS Genetics* **9,** (2013).

32. Saujet, L., Pereira, F. C., Henriques, A. O. & Martin-Verstraete, I. The regulatory network controlling spore formation in Clostridium difficile. *FEMS Microbiology Letters* **358,** 1–10 (2014).

33. Matsuno, K. & Sonenshein, A. L. Role of SpoVG in asymmetric septation in Bacillus subtilis. *Journal of Bacteriology* **181,** 3392–3401 (1999).

34. Lee, A. S. Y. & Song, K. P. LuxS/autoinducer-2 quorum sensing molecule regulates transcriptional virulence gene expression in Clostridium difficile. *Biochemical and Biophysical Research Communications* **335,** 659–666 (2005).

35. Martin, M. J. *et al.* The agr locus regulates virulence and colonization genes in clostridium difficile 027. *Journal of Bacteriology* **195,** 3672–3681 (2013).

36. Donnelly, M. L., Fimlaid, K. A. & Shen, A. Characterization of <i>Clostridium difficile</i> spores lacking either SpoVAC or DPA Synthetase. *Journal of Bacteriology* JB.00986–15 (2016). doi:[10.1128/JB.00986-15](https://doi.org/10.1128/JB.00986-15)

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

38. Martin-Verstraete, I., Peltier, J. & Dupuy, B. The regulatory networks that control Clostridium difficile toxin synthesis. **8,** (2016).

39. Potapov, A. P., Voss, N., Sasse, N. & Wingender, E. Topology of mammalian transcription networks. *Genome informatics. International Conference on Genome Informatics* **16,** 270–278 (2005).

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

41. McGill, R., Tukey, J. W. & Larsen, W. a. Variations of Box Plots. *The American Statistician* **32,** 12–16 (1978).

42. Karlsson, S., Burman, L. G. & Åkerlund, T. Induction of toxins in Clostridium difficile is associated with dramatic changes of its metabolism. *Microbiology* **154,** 3430–3436 (2008).

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

44. Janvilisri, T. *et al.* Microarray identification of Clostridium difficile core components and divergent regions associated with host origin. *Journal of Bacteriology* **191,** 3881–3891 (2009).

45. Scaria, J. *et al.* Comparative genomic and phenomic analysis of Clostridium difficile and <i>Clostridium sordellii</i>, two related pathogens with differing host tissue preference. *BMC genomics* **16,** 448 (2015).

46. Matsumoto, M. *et al.* Impact of intestinal microbiota on intestinal luminal metabolome. *Scientific reports* **2,** 233 (2012).

47. Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. Diversity of microbial sialic acid metabolism. *Microbiology and molecular biology reviews : MMBR* **68,** 132–53 (2004).

48. Fuller, M. F. & Reeds, P. J. Nitrogen cycling in the gut. *Annual review of nutrition* **18,** 385–411 (1998).

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

50. Aboulnaga, E. H. *et al.* 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 (2013).

51. Fonknechten, N. *et al.* Clostridium sticklandii, a specialist in amino acid degradation:revisiting its metabolism through its genome sequence. *BMC genomics* **11,** 555 (2010).

52. Köpke, M., Straub, M. & Dürre, P. Clostridium difficile Is an Autotrophic Bacterial Pathogen. *PLoS ONE* **8,** (2013).

53. Emerson, J. E., Stabler, R. A., Wren, B. W. & Fairweather, N. F. Microarray analysis of the transcriptional responses of Clostridium difficile to environmental and antibiotic stress. *Journal of Medical Microbiology* **57,** 757–764 (2008).

54. Green, M. L. & Karp, P. D. The outcomes of pathway database computations depend on pathway ontology. *Nucleic Acids Research* **34,** 3687–3697 (2006).

55. Theriot, C. M. *et al.* Cefoperazone-treated mice as an experimental platform to assess differential virulence of Clostridium difficile strains. *Gut microbes* **2,** 326–334 (2011).

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

57. Sorg, J. a. & Sonenshein, A. L. Inhibiting the initiation of <i>Clostridium difficile</i> spore germination using analogs of chenodeoxycholic acid, a bile acid. *Journal of Bacteriology* **192,** 4983–4990 (2010).

58. Leslie, J. L. *et al.* 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 (2015).

59. Lopez-Medina, E., Neubauer, M. M., Pier, G. B. & Koh, A. Y. RNA isolation of Pseudomonas aeruginosa colonizing the murine gastrointestinal tract. *Journal of visualized experiments : JoVE* 6–9 (2011). doi:[10.3791/3293](https://doi.org/10.3791/3293)

60. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 1–10 (2009). doi:[gb-2009-10-3-r25 [pii]\r10.1186/gb-2009-10-3-r25](https://doi.org/gb-2009-10-3-r25 [pii]\r10.1186/gb-2009-10-3-r25)

61. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. **27,** 29–34 (1999).

62. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

63. Basler, G., Ebenhöh, O., Selbig, J. & Nikoloski, Z. Mass-balanced randomization of metabolic networks. *Bioinformatics* **27,** 1397–1403 (2011).