*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. A compromised gut microbiota, typically through recent antibiotic exposure, is a prerequisite feature of *C. difficile* colonization susceptibility. This has been described for multiple antibiotic classes in which many result in distinct gut communities, each presenting individual metabolic challenges to *C. difficile*. We hypothesized that *C. difficile* must adapt its physiology to nutrients availability within the gut. Utilizing an *in vivo* model of CDI, we demonstrated that *C. difficile* highly colonized the cecum of mice that received one of three antibiotic pretreatments. We found levels of spore and toxin production varied between each antibiotic treatment group, and both processes are known to be regulated by specific nutrient concentrations. To more closely investigate specific responses of *C. difficile* during infection, we performed transcriptional analysis of *C. difficile* from cecal content of infected mice. This revealed variation in expression of life-cycle switches and catabolic pathways for a variety of carbon sources. In order to assess which substrates *C. difficile* was exploiting, we further characterized the systems with transcriptomic-enabled genome-scale metabolic modeling and untargeted metabolomic analysis. Through the development of a novel metabolite scoring algorithm, leveraging a genome-scale metabolic network architecture, we were able to infer that a given metabolite was acquired from the environment. Output from these models was then validated through comparison with untargeted metabolomic analysis of all infection conditions, which significantly correlated with those metabolites predicted to be important to *C. difficile* in each infection. Our results highlight conserved elements of *C. difficile*'s metabolic strategy across infections, including the consumption of N-acetyl-D-glucosamine and Stickland fermentation substrates. Additionally, these data support the hypothesis that *C. difficile* indeed occupies alternative nutrient niches by metabolizing separate carbohydrate sources in each infection and these distinctions track with disparity seen in pathogenicity.

### 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 hosts and ultimately cause disease. This work also provides evidence that *C. difficile* virulence may be driven by accessibility of specific carbohydrates utilized for growth during each infection. This work has implications for elucidating drivers of *C. difficile* pathogenesis and uncover specifics colonization resistance. This could lead to the discovery of targeted measures to prevent *C. difficile* colonization including potential pre- or probiotic therapies. Furthermore, the metabolite importance calculation workflow described here could provide a useful platform to enable the study of nutrient requirements of pathogen in the context of infection or 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). 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 nutrients (5).

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 which have been shown to result in susceptibility *C. difficile* colonization (6–8). In this case, the antibiotics chosen significantly impact the structure and diversity of the cecal microbiome (Fig. S1a & S1b). It has been further demonstrated that at 18 hours after being introduced to a cefoperazone treated mouse, *C. difficile* reached its maximum vegetative cell density in the cecum (9). 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 substrates (7, 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 intestine (13, 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 generalist (15–17). 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 sporulation (18, 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 carbohydrates (20, 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 mice (24, 25). More relevant to nutrient acquisition, *C. difficile* up-regulated several phosphotransferase systems (PTS) and ABC transporters in gnotobiotic 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 CDI (7, 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 specific pathogen free (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 organism (26–28). Applying these methods to study *C. difficile* colonization would allow us to directly test the nutrient-niche hypothesis.

Founded 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, cefoperazone, and clindamycin (Table 1; Fig. S1). These antibiotics differentially affect the structure of the gut microbiota (8). Each has also been shown to alter the gut metabolome relative to untreated animals (7, 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. Due to technical restrictions at each level of analysis, we employed a multi-omic analysis in order to accurately capture the underlying biological patterns. 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. Differences in resource consumption were also strongly supported by untargeted metabolomic analysis. Out data suggest that *C. difficile* is indeed capable of adapting components of it's nutrient niche to the environment it has colonized, while maintaining at least some focus on catabolism of amino acids and host-derived aminoglycans in each infected condition.

### 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. There was not a significant difference in the number of vegetative cells between any susceptible condition tested (Fig. 1A). 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-GF mice than in the antibiotic treated mice (Fig. 1B). 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-GF animals than any other colonized group (all *P* <= 0.001), but toxin titer also varied between antibiotic treatment groups (Fig. 1C). 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. As *C. difficile* represents <3.3% of the bacterial load in the cecum of any antibiotic-treated SPF mouse (Fig. S2C), and with rRNA depletion also eliminating >90% of each raw RNA isolation (29), it was necessary to pool samples within each treatment group. Moreover, this also required a high degree of sequencing depth per sample in order to yield sufficient quantities of reads that mapped to *C. difficile* str. 630 (4019 genes and putative ORFs; KEGG 2016) with 100% fidelity. Approximately 300 million raw paired-end reads were sequenced per pooled sample, with an average of ~167 thousand reads attributed to *C. dificile* following all curation steps (outlined in Methods). This corresponded to an average of ~42x coverage across the *C. difficile* str. 630 reference genome used. While this value may be low compared to some transcriptional studies of *in vitro* bacterial monoculture or *in vivo* gnotobiotic monoassociation (20, 24), it is not feasible to reach higher coverage in the context of a diverse bacterial community. These factors made the read abundances for *C. difficile* reported here were unavoidable, however these levels have been shown to be largely acceptable for our type of analysis due to the focus on large-scale differences in specific annotated genes (30). Furthermore, microarray-based gene expression analysis was also not a viable alternative as the amount of background transcription from other bacterial species would contribute greatly to background non-specific binding of orthologous transcript and eliminate any true *C. difficile* signal.

We initially focused 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 pathway (31–34) (Fig. 2A). Across the four conditions where *C. difficile* colonized, we observed transcriptional profiles consistent with observed spore levels (Fig. 1B). The mice treated with cefoperazone had the lowest spore density and had the highest level of expression for the anti-sigma factor *spoVG*. The products of these genes are involved in suppressing expression of genes found later in the sporulation pathway (35). 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*, *spoVS*,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 strains (36, 37). 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. 2B) and quorum sensing (Fig. 2C); however, these transcripts were not found in the GF 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 and that enzyme activity does not necessarily reflect the concentrations of transcript nor effector protein available in the cell.

Moving upstream, we next focused on the regulators of metabolic pathways. Specific sigma factors are master regulators and a subset have been shown to integrate signals from intra- and extracellular nutrient concentrations (20, 21, 33, 38). The transcription of the global repressor *codY* is responsive to intracellular concentrations of *C. difficile* energy sources (39). Highest transcription for this gene was found in cefoperazone-treated and GF mice (Fig. 2D). The regulation networks of CodY and CcpA are highly interconnected, with the expression of *ccpA* specifically linked to local concentration of rapidly metabolizable carbon sources (40). 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. 2D), 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. 2D). 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. 2D). 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. S2A). We then focused on general differences in carbohydrate (Fig. S2B) and amino acid (Fig. S2C) 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 GF 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. 3). 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 transcript abundance. 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 identify 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. 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). 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. 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 have been associated with adaptive expression of phosphotransferase systems (PTS) and ABC transporters with many known differences in substrate specificities (25). Among the genes classified as PTS transporters were overrepresented in both clindamycin and streptomycin-treated mice, while ABC sugar transporters were overrepresented in the cefoparazone-treated mice. The most stark differences were seen in transcription for genes involved in sugar alcohol catabolism. 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 and polysaccharide degradation. 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. Additionally, alpha/beta-galactosidase genes (*aglB* and *bglA*) were also overrepresented in clindamycin-treated mice. Together these patterns suggested that polysaccharide fermentation occurred this condition. More subtle differences were seen in those gene associated with glycolysis. This category includes genes for not only the steps of glycolysis, but also several genes that mediate entry points of monosaccharides to glycolysis. 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 interactions by directed connecting edges (Fig. 4A). 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 (41). 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 (42). Together, these metrics allow 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" (43). Interestingly, both ATP and GTP were not included in the top 10 of either list, despite still scoring relatively highly. Their exclusion may be a byproduct of the KEGG reference used for network construction (reaction\_mapformula.lst) which focuses on conversions of smaller molecules and mostly excludes the synthesis of large macromolecules (ex. DNA, RNA, Proteins, etc.) which require higher input energy. While may be a limitation of certain analyses, ours was not effected as the interest of this study was in those simple nutrients acquired from the environment. Together, these results indicated that the topology of the overall network reflects established bacterial physiology and amino acid catabolism is a central strategy of *C. difficile* str. 630 based on the structure of its metabolic network.

**Metabolite importance algorithm supports known trends and reveals adaptive nutritional strategies of *C. difficile* during infection across distinct environments.** Moving into analysis of the *C. difficile* metabolic network, we sought to utilize transcriptomic data to infer which metabolites *C. difficile* is most likely to obtain from its environment in each condition. To accomplish this we mapped normalized transcript abundances to the enzyme nodes in the network. As transcription and translation are coupled in bacteria, we hypothesized that we could incorporate whole transcriptome sequencing results into the metabolic model in order to impute active metabolism. Concordantly, 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 (44). Because of this, we were reasonably confident in utilizing normalized transcript abundance as a proxy for enzyme levels. The importance of each metabolite was measured as the log2-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. 4A & 4B). 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. Due to the fact that separate sequencing efforts for each individual mouse in each group was impossible, we instead adopted a Monte Carlo-style simulation in order to achieve some degree of statistical validation to our findings. To accomplish this, we first calculated metabolite importance scores for 10,000 iterations of random transcript abundance reassignment of all enzyme nodes in the network. Next we used these distributions to calculate a 95% confidence interval for each metabolite and create comparators that represent random noise (45). This ultimately allowed for computing the significance level that a given score had a high probability of being excluded from its associated null hypothesis score distribution (Fig. 4C). All downstream analysis of important metabolites was only performed using those substrates that met their respective significance cutoff (*P* < 0.05).

Applying these methods to the *C. difficile* transcriptomic data collected from the *in vivo* CDI models, we sought to identify patterns of growth substrate importance within each infected condition (Table S3). To first identify the core metabolites that are most essential to *C. difficile* in any condition, we cross-referenced the highest 50 scoring metabolites from each treatment group and recalculated the median importance across all conditions (Fig. 5A). The host derived amino sugar N-acetyl-D-glucosamine 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 utilized by *C. difficile*, entering into glycolysis. Additionally, components of the Stickland fermentation pathway were also found to be important to *C. difficile* in all conditions tested including proline, 3−hydroxybutanoyl−CoA, and formate (46–48). Both of these forms energy acquisition have been previously associated with *C. difficile* during mono-associated conditions (24), but never before in the context of a complex community of potential competitors. This indicated that these metabolites may be an integral component of the nutrient niche for *C. difficile* in any infection condition.

Moving on to assess differential patterns of metabolite importance across environments, we further focused the analysis and compared the highest 25 scoring metabolites in each infection condition to screen for those that are distinctly important to *C. difficile* in each respective environment (Fig. 5B). Interestingly, the resulting groups of metabolites all contained at least one known carbohydrate growth substrate of *C. difficile* (7, 14, 49). This included sorbitol, mannitol, salicin, and N-acetylneuraminic acid. While the first 3 are more likely introduced by the diet, N-acetylneuraminic acid is another amino sugar integrated into the host mucin layer (50). Previously, the concentration of this molecule was shown to increased through liberation by other bacterial species to the benefit of *C. difficile* (49), however our results suggest that *C. difficile* will utilize it for growth regardless of concentration. Furthermore, in GF mice where no other competitors are present, our model indicated that *C. difficile* was more likely to acquire numerous amino acids from the environment, presumably instead of going through the process of synthesis. These data support the hypothesis that *C. difficile* may exploit alternative nutrient sources between the susceptible environments it colonizes.

We went on to test each important possible growth substrates to ensure that growth phenotypes of our strain matched previously published results (Fig. 5C). This was performed using a modified defined *C. difficile* minimal media (7), supplemented individually with the selected carbohydrates implicated by high importance scores. Because *C. difficile* is auxotrophic for several amino acids, it was necessary to include amino acids in the minimal media despite the capability to achieve modest growth through Stickland fermentation of these substrates. This focused our analysis on carbohydrates and made the most effective negative control growth in media lacking carbohydrates but containing amino acids. Also since proline was identified as necessary for minimal media (51), so was therefore not tested for growth phenotype changes in this analysis despite its importance in shared modeling analysis (Fig. 5A). When tested for improved growth, at least one carbohydrate found to be important in all measured environments as well as among those indicated to be distinctly important within each condition. This included N-acetylglucosamine (shared; Max OD600 = 0.744), mannitol (cefoperazone; Max OD600 = 0.461), salicin (clindamycin; Max OD600 = 0.869), and N-acetylneuraminate (GF; Max OD600 = 0.375). Each of these were shown to allow for significantly more growth (*P* < 0.001) than in no carbohydrate controls (Max OD600 = 0.212). Galactitol was also tested as it was important in streptomycin-treated animals, however it demonstrated nearly identical growth patterns to no carbohydrate control (Table S4). These data suggest a potential hierarchy in carbohydrate preference and supported that *C. difficile* was well-suited to adapt its metabolism toward nutrient sources that are differentially available across susceptible mouse ceca.

**Metabolomic analysis supports that *C. difficile* consumes metabolites indicated as important by metabolic modeling and implicates proline as a critical component of its nutrient niche across models of infection.** In order to support conclusions from transcriptional network-based analysis we employed untargeted 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 (Metabolon, Durham NC). Using these results, we first wanted to assess variability in the datasets generated for this study with replicates to infer levels of consistency through transriptomic sequencing efforts. In summary, extremely low and consistent sample variance was found in across all 16S and metabolomic replicates (Fig. S5). As these results represent data that is essentially above and below transcript abundance in terms of biological hierarchy, this reassured our conclusions at the various levels of separate analysis. Additionally, by incorporating multiple sets of omics data describing the same system provided even greater confidence in the overall trends.

With respect the modeling results, we measured whether important metabolites from network analysis were increased in relative concentration in their corresponding susceptible environments by comparing untreated SPF mice with mock-infected susceptible groups at the same time point (Fig. S6). Looking first toward N-acetylglucosamine since it was the most important shared metabolite (Fig. 5B), we found that its relative concentration was actually significantly decreased in all susceptible conditions tested (Fig. S6A). A similar trend was also seen in salicylate (Fig. S5F; *P* < 0.05), a product of salacin cleavage (salicin was not included in the panel). Additionally we found that N-acetylneuraminate (Fig. S6B), mannitol/sorbitol (Fig. S6G), and galactitol (Fig. S6H) were significantly increased in only cefoperazone-treated SPF and GF mice (*P* < 0.05). Interestingly, only Stickland fermentation substrates (proline; Fig. S6C and trans-4-hydroxyproline; Fig. S6E) were for to be significantly increased in all susceptible environments tested (*P* < 0.05). Glycine (Fig. S6D) was found to also be consistently increased, but only achieved significant differences in cefoperazone-treated SPF and GF mice (*P* < 0.05).

We were then interested if the concentrations of important metabolites decreased during *C. difficile* infection compared to mock infection, implying their consumption by the pathogen in the infection conditions they were found to be important. Both groups of host-derived aminoglycans (N-acetylglucosamine/N−acetylgalactosamine Fig. 6A and N-acetylneuraminate Fig. 6F) were significantly decreased in the presence of *C. difficile* when mock-infected ceca demonstrated detectable levels of the metabolites. No significant relationship was seen in any of the other carbohydrates with distinctly high importance scores (galactitol; Fig. 6C or mannitol/sorbitol; Fig. 6D). The slight increase in the concentration of salicylate (Fig. 6E) may be interpreted as due to increased cleavage of salicin by *C. difficile*, but requires further investigation. As side note, acetate was found to be important in all conditions, but was just below the significance cutoff in GF mice (Table S3) and has been shown to be consumed by *C. difficile* during Stickland fermentation (52). Confirming our prediction, we found that *C. difficile* colonization indeed led to a significant decrease in the levels of acetate (Fig. S7C). Together, these findings provided some validation for our metabolite importance algorithm as a method for predicting metabolites that are most likely consumed by a bacteria in a given environment, not just growth substrates. Finally, in accordance with the previous analysis, we found that proline (Fig. 6A) and trans-4-hydroxyproline (Fig. S7B) were significantly decreased in every *C. difficile* colonized environment compared to mock infection (*P* < 0.05). Glycine was concordantly decreases in each condition following infection, but not all of these relationships are significant (Fig. S7A). These results most strongly support the consumption of amino acids is actually the primary energy source of *C. difficile* during infection.

**Integrated multi-omic analysis reveals strong predictive value for networking approach in identifying additional novel patterns of metabolism.** Moving beyond specific metabolite analysis, we wanted to determine how well metabolite importance scores related to the relative concentration of all metabolites in a given environment. To most effectively combine metabolite importance scores with the affiliated concentration data, we decided to calculate the median change in relative concentration for each metabolite measured between corresponding mock and *C. difficile*-infected conditions. This was accomplished by dividing the median log10-scaled intensity of each metabolite in mock-infected mice by its corresponding intensity in *C. difficile*-infected mice to give the ( ) median scaled intensity for each metabolite. In this way, larger values indicated those metabolites that decrease in concentration when *C. difficile* colonizes. Additionally this resulted in a single value we could combine with importance scores for each metabolite and perform a linear correlation analysis. Positive correlations would therefore indicate that the metabolites predicted to be consumed by *C. difficile* through transcript-informed metabolic networks have a concomitant decrease in the metabolome. This relationship would also imply that those metabolites with negative importance scores and increased concentration in infected animals are being produced by *C. difficile*.

We applied this approach first to a pooled analysis of all groups in order to maximize our ability to validate the predictive capacity of the modeling algorithm (Fig. 7A). This indeed revealed a significant positive correlation (*P* < 0.01) and supported our method for identifying metabolites that are most likely to be consumed by *C. difficile* based on transcriptional activity. Furthermore, this technique also allowed us to analyze outliers from the generalized linear model to assess if they are explained by other known patterns of *C. difficile* biology. In this instance, 16 outlier metabolites were present and fell into only 3 separate KEGG annotation categories; amino acids, carbohydrates, and nucleotides (Table S5). Also as predicted, those outlier metabolites that were important in their respective condition as well as diminished in the metabolome following infection (12 of 16) were all annotated as amino acid or carbohydrate growth substrates or growth enhancers of *C. difficile* (51). This included L-proline, L-threonine, N-acetyl-D-glucosamine, D-fructose, and galactitol. Alternatively, the metabolites that were more likely to be produced based on importance scores but decreased in the context of infection were the nucleotides adenine, deoxyadenosine, and xanthine. Although *C. difficile* has the capacity to synthesize each (52), it may also have the ability to catabolize these purines to participate downstream in Stickland reactions (53). The findings here may indicate that *C. difficile* utilizes both environmental and synthesized pools of the molecules at a faster rate than they are produced by the bacterium alone and may reinforce the centrality of amino acid fermentation to *C. difficile* *in vivo* metabolism.

We then directed the analysis toward each individual infection condition tested. In streptomycin-pretreatment (Fig. 7B) and cefoperazone-pretreatment (Fig. 7C) the correlations did not achieve significance, but remained positive. However, analysis of outliers revealed that in streptomycin-pretreatment both the host derived aminoglycan N-acetyl-D-glucosamine and L-proline were implicated as most likely to be utilized by *C. difficile* while only L-proline was the only growth substrate outlier in this way in cefoperazone-pretreated mice. Inversely, both clindamycin-pretreatment (Fig. 7D) gnotobiotic mice (Fig. 7E) demonstrated a significant correlation between importance scores and ( ) median scaled intensity (*P* < 0.01). Although clindamycin-pretreatment was associated with the most outlier metabolites, both groups also shared a similar trends of select carbohydrates and primarily metabolites involved in Stickland fermentation being being important and decreased in infected conditions. In fact across all groups proline was found to be both highly important and decreased in concentration after infection. In combination with the previously mentioned patterns of correlates, this further supports Stickland fermentation as a core metabolic strategy of *C. difficile* in the gut which agrees with previous *in vitro* findings (48). Together, these data support network-derived importance scores as well as our hypothesis that *C. difficile* adapts its portions of its metabolic strategy in distinct susceptible environments despite certain highly conserved elements.

### Discussion

Collectively, our results support the hypothesis that *C. difficile* can adapt its metabolism to the available niche landscape across susceptible gut environments and give insight to the adaptive strategies that *C. difficile* can use to colonize diverse human microbiota. All of our *in vivo*, *in vitro*, and *in silico* data demonstrate the plasticity of *C. difficile* to effectively change its metabolism to utilize alternative resources for growth. This may be the result of increased concentration of particular metabolites as a consequence of concordant decreases in the population of one or more competitors for those resources. In addition to uncovering adaptive strategies of *C. difficile*, our method is also able to identify consistent trends in metabolism across environments. The findings that N-acetyl-D-glucosamine and Stickland fermentation substrates were consistently among the highest scoring shared metabolite among all tested conditions strongly indicates that these metabolites are central to the nutritional strategy of *C. difficile* and may be utilized in and condition in which they are available. The metabolism of both substrate types provides not only carbon and energy to *C. difficile*, but are also a source for nitrogen which is a limited resource in the mammalian lower GI tract (54).

Apart from 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. One interesting result is the appearance of CO2, an apparent metabolic end product, in the list of shared important metabolites (Fig. 5A). 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 (55). This highlights that our method does not only identify growth substrates, but also reports any metabolites that are being removed from the environment. Although our modeling results are consistent with previously published work on the metabolism of *C. difficile*, there are potential limitations of this approach. 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 (56), 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. Conversely, a physical limitation of the current study may be the the lack of replication in transcriptomic sequencing. In the future it may be possible for RT-qPCR validation of gene expression values, but this was not within our capacity to integrate here. In spite of these assumptions, the method outlined here supports known elements of *C. difficile* biology and our subsequent *in vitro* and *in vivo* metabolomic analysis confirmed the predictions of previously unseen components made here. It is also important to note that these limitations, as well as the inherent complexity of infection in context of the gut microbiome, were motivations for employing multi-omic analysis that only when employed together were able to reveal a clearer image of *C. difficile*'s metabolism during infection.

Based on the evidence presented, our results support the hypothesis that *C. difficile* is a metabolic generalist and is able to catabolize alternative carbon sources across susceptible gut environments. This may be due to an inability to outcompete a collection of metabolic specialists in an intact community, and separate classes on antibiotics differentially eliminate these populations and allow for *C. difficile* colonization. Additionally, differential importance scores for carbohydrates in agreement with metabolomic results across distinct environments, and when considered with variable sporulation and toxin production, may indicate that *C. difficile* pathogenesis is directed by which carbohydrates are available for use. Furthermore, competition for nutrients may also potentially explain the success rate of fecal microbial transplant (FMT), in that wholesale installation of a diverse range of specialized metabolic strategies is enough to outcompete *C. difficile* from the majority of perturbed gut environments. This may be particularly true with competitors for proline as our data also suggests that overall *C. difficile* colonization resistance may be driven by the availability of fermentable amino acids. *C. difficile* may also have evolved to utilize the host-derived aminoglycan N-acetyl-D-glucosamine in all infection conditions as it is a consistently available nutrient source across susceptible hosts. Metabolic modeling output also significantly correlated with changes in the metabolome associated with *C. difficile* colonization, establish validity as a platform from analysis of nutrient utilization and potential preference/hierarchy elucidation. This significant relationship is even more important because modeling analysis was based exclusively on genome annotation without prior knowledge of nutrient availabilities nor taking into account the metabolic activities of the surrounding microbes in each condition. In light of these findings our metabolic network platform combining genomics with transcriptomics, may also prove informative for generating hypotheses through reverse ecology that could ultimately lead to uncovering new interactions between many species in the gut community that ultimately impact host health (57). In conclusion, *C. difficile* is is able to optimize its nutritional strategy for each colonized gut environment. Our results implicate that further considerations are needed when attempting to design targeted prebiotic and probiotic therapies for the prevention or elimination of *C. difficile* from the human gut.

### Methods

**Animal care and antibiotic administration** Adapted from the previously described model (58), 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** *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 conditions (59). 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 spores (60). 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 (61). 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>

**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 (62). 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 (62). 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) (63). 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 (64). 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 (37) 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 (65) 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 (66). 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 (67) 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 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 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. 4a) favors metabolites that are more likely acquired from the environment (not produced within the network), and will award them a higher score (Fig. 5b & 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 in order to ultimately generate a p-value for each metabolite. 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(68).

**Measuring *in vivo* concentrations of acetate.** Cecal contents were flash frozen in liquid nitrogen at the time of necropsy and subjected to short chain fatty acid quantification 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 described (7). 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. Finally, samples were 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 rank-abundance test with Benjamini-Hochberg correction. Undetectable points used the limit of detection for all statistical calculations. Significant differences for growth curves compared to no carbohydrate control (+ amino acids) were calculated using 2-way ANOVA with Benjamini-Hochberg correction. Significance for metabolite importance scores was determined as described above via comparison to Monte-Carlo simulations. Outlier metabolites in linear correlation analysis were identified with a 1.5 minimum cutoff for squared Studentized residual analysis of each individual generalized linear model.

### 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_2016>.

**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

**Figure 1 | *C. difficile* sporulation and toxin activity quantification after 18 hours of infection.** **(A)** Vegetative *C. difficile* c.f.u. per gram of cecal content. 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. 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). Dotted lines denote the limit of detection for both assays (LOD). Undetectable points are shown just below the LOD for clarity, however all medians and significant differences were calculated using the LOD. Significant differences calculated using Wilcoxon rank-sum test with the Benjamini-Hochberg correction.

**Figure 2 | 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 3 | *C. difficile* expression of gene sets for carbon metabolism pathways across antibiotic pretreatments.** Ternary plot indicating the relative abundance of transcripts for all *C. difficile* str. 630 genes across the three colonized antibiotic-treated condition s (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. Gene groups from are shown individually below 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. S3 for additional explanation of figure interpretation.

**Figure 4 | 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. Enzyme node sizes reflect the levels of detectable transcript from each gene. 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 producing a metabolite, and (III) difference of consumption and creation of the given metabolite. **(B)** The expanded window displays a partial example of D-fructose importance calculation based on local enzyme transcription. Values in the red nodes represent normalized transcript reads mapping to each enzyme node. Average expression of input and output reactions are used to calculate the relative importance of **m**. **(C)** Example Mont-Carlo simulation results for **m**. Shown here is the distribution resulting from 10000-fold Monte-Carlo simulation of transcript randomization for all enzymes in the metabolic network. The black solid line represents the median of the distribution, the red dotted lines denote the 95% confidence interval, and the blue line indicates the actual measured score for **m**.

**Figure 5 | 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 between growth curves were calculated using 2-Way ANOVA with Benjamini-Hochberg correction.

**Figure 6 | Change in *in vivo* concentrations following *C. difficile* infection of select metabolites indicated as important by metabolic modeling.** Comparison of relative concentrations of select metabolites from untargeted UPLC-MS analysis of *C. difficile* vs mock-infected mouse cecal content (n = 9 mice per group). **(A)** N−acetylglucosamine / N−acetylgalactosamine, **(B)** Proline, **(C)** Galactitol, **(D)** Mannitol / Sorbitol, **(E)** Salicylate, and **(F)** N−acetylneuraminate. Paired metabolites were quantified simultaneously as the only differ by chirality making differentiation impossible. Significant differences (*P* < 0.05) are denoted by a single asterisk (\*) and were determined using Wilcoxon signed-rank test with Benjamini-Hochberg correction when appropriate.

**Figure 7 | Correlation of change in metabolite concentration following *C. difficile* colonization with importance scores derived from transcript-informed metabolic modeling.** Spearman correlations for combined metabolome and metabolic model output in each condition with outlier metabolites identified through squared-residual analysis. ( ) median scaled intensity was calculated by dividing the median log10-scaled intensity of each metabolite from GS-MS in mock-infected mice by its corresponding intensity in *C. difficile*-infected mice (n = 9 per group). Larger values represent metabolites which decrease in concentration upon infection. These values are shown here relative to their calculated importance score in each condition separately using the algortihm outlined in Fig. 4B. **(A)** Results from all treatment groups pooled for single analysis (*P* < 0.05). Importance scores were first paired with their matching ( ) median scaled intensity within each treatment group prior to combining datasets. Outliers are labelled by their KEGG metabolite family annotation and were defined as those points with a ( ) median scaled intensity squared residual >1.5 with the cutoff area labeled in gray. Correlations and outliers were recalculated for each component condition as follows: **(B)** Streptomycin pretreatment, **(C)** Cefoperazone pretreatment, **(D)** Clindamycin pretreatment (*P* < 0.05), **(E)** and Gnotobiotic mice (*P* < 0.05).

**Supplementary Figure 1 | Experimental time lines 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 at the time of infection. Sacrifice 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 conventionally-raised mice used in these experiments (n = 63). All treatment groups are significantly different from all 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-treated condition by Wilcoxon test (*P* < 0.001). **(C)** Representation of reads contributed by *C. difficile* in each sequenced condition. Shown in each group is the quantity of normalized reads contributed by either *C. difficile* or all other reads within that sequencing effort. The percents listed at the top of each group is the proportion of the total community composed of *C. difficile*. No infected condition had >3.6% of the bacterial reads sequenced annotate as *C. difficile*. Significantly less were for *C. difficile* were detected in each condition, determined by Wilcoxon rank-sum test (*P* < 0.001). All mock-infected conditions had <0.042% of reads annotating as *C. difficile*, which has within the acceptable margin of error for our data curation pipeline. It should also be noted that mock-infected groups were culture (Fig. 1) and PCR negative (data not shown) for *C. difficile* str. 630.

**Supplementary Figure 3 | 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)** Sub-families within Carbohydrate metabolism. **(C)** Sub-families within Amino acid metabolism.

**Supplementary Figure 4 | 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 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 in order to demonstrate consistent measurements in multiple levels of data, and further support our approach to pooled transcriptomic sequncing. **(A)** OTU abundances from 16S rRNA gene sequencing, sample variances for each OTU were calculated individually prior to sumary statistic calculations. **(B)** Scaled intensities from untargeted metabolomic analysis, sample variances for each metabolite were calculated individually prior to sumary statistic calculations. In both groups of calculations all median sample variances are >1, indicating extremely low levels of variability between samples of the same type.

**Supplementary Figure 6 | Relative *in vivo* concentrations of metabolites highlighted through network-based analysis in uninfected mice** Select metabolite concentrations using normalized untargeted UPLC-MS quantification from cecal content of *C. difficile* susceptible mice compared to resistant controls. Samples were collected at the same 18 hour timepoint as the analyses performed in infection experiments (n = 9 mice per group). Highlighted carbohydrates included: **(A)** N−Acetylglucosamine / N−Acetylgalactosamine, **(B)** N−Acetylneuraminate, **(C)** Proline, **(D)** Glycine, **(E)** Salicylate, **(F)** Mannitol / Sorbitol, and **(G)** Galactitol. Significant differences between experimental, susceptible groups and resistant SPF mice (*P* < 0.05) are denoted by a single asterisk (\*) and were determined using Wilcoxon rank-abundance test with Benjamini-Hochberg correction.

**Supplementary Figure 7 | Change in *in vivo* concentrations following *C. difficile* infection of additional select metabolites indicated as important by metabolic modeling algorithm.** Comparison of relative concentrations of select metabolites from untargeted UPLC analysis of *C. difficile* vs mock-infected mouse cecal content. **(A)** Glycine and **(B)** Acetate. Relative concentrations of glycine were quantified as in Figure 6 with untargeted UPLC-MS analysis. Absolute concentrations of acetate were measured in a separate analysis using GC-MS. Significant differences (*P* < 0.05) are denoted by a single asterisk (\*) and were determined using Wilcoxon signed-rank test with Benjamini-Hochberg correction when appropriate.

**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.**

**Supplementary Table 5 | Outliers from generalized linear models of combined metabolome and metabolic network output analysis.**

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

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

6. **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.

7. **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.

8. **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**.

9. **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.

10. **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.

11. **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.

12. **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**.

13. **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.

14. **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.

15. **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.

16. **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.

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

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

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

20. **Antunes, A.**, **E. Camiade**, **M. Monot**, **E. Courtois**, **F. Barbut**, **N. V. Sernova**, **D. A. Rodionov**, **I. Martin-Verstraete**, and **B. Dupuy**. 2012. Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. U.S. Patent 21.

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

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

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

24. **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.

25. **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**.

26. **Patil, K. R.**, and **J. Nielsen**. 2005. 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.

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

28. **Suthers, P. F.**, **M. S. Dasika**, **V. S. Kumar**, **G. Denisov**, **J. I. Glass**, and **C. D. Maranas**. 2009. Genome-scale metabolic reconstruction Of mycoplasma genitalium, iPS189. PLoS Computational Biology **5**.

29. **O’Neil, D.**, **H. Glowatz**, and **M. Schlumpberge**. 2013. Ribosomal RNA depletion for efficient use of RNA-seq capacity. Current Protocols in Molecular Biology.

30. **Haas, B. J.**, **M. Chin**, **C. Nusbaum**, **B. W. Birren**, and **J. Livny**. 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC Genomics **13**:734.

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

32. **Underwood, S.**, **S. Guan**, **V. Vijayasubhash**, **S. D. Baines**, **L. Graham**, **R. J. Lewis**, **M. H. Wilcox**, and **K. Stephenson**. 2009. Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production. Journal of Bacteriology **191**:7296–7305.

33. **Fimlaid, K. A.**, **J. P. Bond**, **K. C. Schutz**, **E. E. Putnam**, **J. M. Leung**, **T. D. Lawley**, and **A. Shen**. 2013. Global Analysis of the Sporulation Pathway of *Clostridium difficile*. PLoS Genetics **9**.

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

35. **Matsuno, K.**, and **A. L. Sonenshein**. 1999. Role of SpoVG in asymmetric septation in *Bacillus subtilis*. Journal of Bacteriology **181**:3392–3401.

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

37. **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.

38. **Donnelly, M. L.**, **K. A. Fimlaid**, and **A. Shen**. 2016. Characterization of *Clostridium difficile* spores lacking either SpoVAC or DPA Synthetase. Journal of Bacteriology JB.00986–15.

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

40. **Martin-Verstraete, I.**, **J. Peltier**, and **B. Dupuy**. 2016. The regulatory networks that control *Clostridium difficile* toxin synthesis. U.S. Patent 5.

41. **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.

42. **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.

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

44. **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.

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

46. **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.

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

48. **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.

49. **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.

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

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

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

53. **Hartwich, K.**, **A. Poehlein**, and **R. Daniel**. 2012. The Purine-Utilizing Bacterium Clostridium acidurici 9a: A Genome-Guided Metabolic Reconsideration. PLoS ONE **7**.

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

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

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

57. **Levy, R.**, and **E. Borenstein**. 2012. Reverse Ecology: from systems to environments and back. Advances in experimental medicine and biology **751**:329–345.

58. **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.

59. **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.

60. **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.

61. **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.

62. **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.

63. **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.

64. **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.

65. **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.

66. **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.

67. **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.

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