*Clostridium difficile* infection differentially alters the structure and metabolic activity of distinct murine cecal microbiomes

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

### Abstract

### Introduction

*Clostridium difficile* is a

What c. difficile is and its associated infection are

Associations with multiple antibiotics in the clinic

different clinical outcomes

able to colonize different mouse models - cite alyx

differences in virulence factor expression

differences in gene expression - cite last paper

leading into possible nutrient niche differences - also last paper

why do i think the exchange is nutrient based?

leads to questions about differential responses on the commuinity to infection - ratios of competition or synergy could have downstream effects on infection outcome

our approach utilizing metagenome-enabled metatranscritpomics - to analyse potentially active processes in members of the microbiota - binning contigs to get discreet genomes to analyse transcriptomes of individual taxa

modeling the metabolism of complex communities of bacteria using next-generation sequencing data

initial metatranscriptomic analysis revealed that *C. difficile* colonization indeed altered community-level gene expression, particularly for pathways involved in the acquisition and catabolism of certain nutrient sources.

competition or independence?

### Results

***C. difficile* differentially expresses virulence factors across separate classes of antibiotic pretreatment.** Prior to oral *C. difficile* str. 630 spore challenge, conventionally-reared SPF mice were pretreated with one of three different antibiotics to sensitize the animals to *C. difficile* colonization. The selected antibiotics were streptomycin, cefoperazone, and clindamycin (Table S1). Each drug was chosen not only due to its ability to reduce *C. difficile* colonization resistance in a mouse model1, but also for their distinct and significant impacts on the structure and diversity of the cecal microbiota (all *p* < 0.001; Fig. S1a). Selection of this toxigenic *C. difficile* strain was based on its moderate clinical severity in mouse models**???**, previous studies of *in vitro* metabolism2, and well-annotated genome3. Briefly, mice were pretreated with the respective antibiotic as indicated in Fig. 1a and were subsequently orally gavaged ~1×103 *C. difficile* str. 630 spores before necropsy at 18-hours post-infection. This end point corresponded with a previous study where another laboratory strain of *C. difficile* reached maximum cecal vegetative cell density with few detectable spores4.

At this final time-point, we first quantified vegetative *C. difficile* colony forming units (cfu) per gram of cecal content (Fig. 1b). There was no significant difference in the number of vegetative cells between any antibiotic-pretreatment tested and all susceptible mice were colonized to ~1×108 vegetative cfu per gram of cecal content, while untreated mice maintained *C. difficile* colonization resistance. We moved on to measure both sporulation and toxin activity as activation of both processes has been linked to environmental concentrations of specific growth nutrients [@]. Despite having the similar amounts of vegetative *C. difficile* cells, varying levels of both spore cfu and toxin titer were found across each of the antibiotic pretreatment models tested. These results showed that *C. difficile* colonized different antibiotic-pretreated mice to consistently high levels, and the distinct treatments corresponded with moderate differences in the expression of *C. difficile* virulence factors.

***C. difficile* colonization differentially corresponded with structural differences in cecal bacterial community between antibiotic pretreatments.** Although it had been established that treatment with antibiotics alters the structure of the intestinal microbiota5 and that different classes of antibiotics exert distinct changes on the gut bacterial communities of mice1, it has not yet been determined if *C. difficile* colonization impacts the structure of the gut microbiota early during infection. In order to evaluate changes in bacterial population abundances in response to perturbation and *C. difficile* colonization, we performed 16S rRNA gene V4 region sequencing from cecal content of both mock and *C. difficile*-infected mice across antibiotic pretreatment models. To focus analysis specifically on the surrounding bacterial communities, we removed all sequences contributed by *C. difficile* prior to downstream analysis. Following data curation, we first reconfirmed that each antibiotic treatment significantly impacted both cecal community structure (ThetYC distances) and diversity (inverse-Simpson) compared to untreated control mice (all *p* < 0.001, Fig. S1a & S1b). We then ensured that each antibiotic class induced disntict shifts in community which significantly differed from one another (all *p* < 0.001, Fig. S1c). The composition of streptomycin-pretreated communities was more variable between cages, but was generally enriched for members of phylum Bacteroidetes (Fig. 2a). Cefoperazone and clindamycin-pretreated cecal were consistently dominated by the families Lactobacillaceae and Enterobacteriaceae respectively (Fig. 2a).

Next, we measured the effect that *C. difficile* colonization had on overall community structure and composition for each of the antibiotic groups. We found that structure both streptomycin and clidamycin-pretreated communities were significantly different from their within-group uninfected controls (*p* = 0.014, Fig. S1d; *p* = 0.003, Fig. S1f). Infected cefoperazone-pretreated communities did not significantly differ from their control group (Fig. S1e). We then sought to identify specific OTUs that significantly differed in abundance between mock and *C. difficile* infection within each pretreatment group. To accomplish this, we utilized LEfSe differential abundant feature detection6 for each perturbed cecal community comparing mock and *C. difficile*-infected conditions. In order to improve the ability to distinguish between potentially informative OTUs and differences that arise from cage effects, we screened for OTUs that appear in at least nine samples for each antibiotic pretreatment separately (n=9 mice per experimental group). This analysis revealed a total of 16 OTUs that discriminated between infected and uninfected communities in at least one antibiotic pretreatment (Fig. 2b-2d). The most frequent differences were seen in the abundance of OTUs at very low relative abundance. In these instances, these OTUs were generally only detectable in either mock of *C. difficile*-infected states. The type of change associated with infection was also not constant in that some OTUs would increase in relative abundance following *C. difficile* colonization while others instead decreased. Additionally, we performed a corrected Wilcoxon rank-sum test for the relative abundance of each discriminating OTU across pretreatments to assess for consistent changes in abundance. In streptomycin pretreatment, only *Porphyromonadaceae* (OTU5) was significantly increased during infection while *Turicibacter* (OTU6) was significantly decreased (*p* < 0.05, Fig. 1d). Cefoperzone-pretreatment was associated with significant increases in both *Faecalibacterium* (OTU49) and *Arthrobacter* (OTU17) with a decrease in *Lactobacillus* (OTU1) during infection. Finally, the only significantly different OTU in clindamycin pretreatment was *Lactobacillus* (OTU1) which increased with infection. Interstingly, not only was *Lactobacillus* (OTU1) the only OTU which discriminated infection groups within multiple antibiotic pretreatments (cefoperazone & clindamycin), but its relatively abundance significantly increased in both. These results indicated that certain taxonomic groups may be more prominently effected by *C. difficile* colonization than others and implies possible direct or indirect interactions.

**Distinct antibiotic classes lead to alternative markers of *C. difficile* colonization susceptibility.** Several previous groups have demonstrated that treatment with antibiotics not only alters the structure of the resident microbiota, but also has a dramatic impact of the intestinal metabolome as well**???**,7,8. We hypothesized that the distinct classes of antibiotics used in the previous experiments would differentially impact the cecal metabolome as a function of their effect on the microbiota from mice with a nearly identical starting community. To address this, we performed untargeted metabolomic analysis on separate aliquots of cecal content also utilized in 16S rRNA gene sequencing in an effort to identify possible conserved markers of susceptibility. A total of 727 distinct metabolites were identified through a combination of several liquid chromatography and mass spectrometry techniques and pathway membership was annotated using the KEGG compound database (Materials and Methods). We chose to focus initially of differences between untreated and antibioti-pretreated, mock-infected metabolomes to avoid the contribution of *C. difficile* and strictly measure the composition of susceptible cecal environments (Fig. 3). With this approach, we observed that any of the antibiotic pretreatments significantly altered the cecal metabolome compared to untreated, *C. difficile*-resistant mice as quantified by Bray-Curtis dissimilarity (*p* < 0.001; Fig. 3a). Additionally, differences between resistant and susceptible metabolomes were apparent across numerous metabolic pathways (Fig. S4). Next, we assessed the degree of difference between pretreatment groups and found that each antibiotic resulted in a unique cecal metabolome (*p* < 0.001; Fig. 3b). Similar to the differences between resistant and susceptible states, the patterns of specific metabolite concentrations were divergent based on the class of antibiotic pretreatment (Fig. S5). These results supported that not only is an intestinal environment that is susceptible to *C. difficile* colonization fundamentally different from a resistant state, it also supported that each antibiotic pretreatment lead to distinct susceptible metabolomic structures.

We were then interested in identifying those metabolites with the highest degree of change between the previously defined groups. Due to the large number of individual metabolites with significant differences in relative concentration between groups, we employed a machine learning approach using random forest9 to highlight those metabolites that most easily differentiate groups. In order to further limit the analysis to only the most informative metabolite, we ranked metabolites in order of highest Mean Decrease Accuracy (the amount in which their removal negatively impacts correct sample classification) and reported the top 7 in each analysis. When applied to the comparison of resistant versus susceptible cecal metabolomes, this approach was not only able to correctly classify resistnat and susceptible states (OOB = 0%), but also revealed several metabolites that significantly increased in relative concentration following administration of antibiotics (Fig. 3c). This primarily included precursors or intermediates of carbohydrate fermentation such as chiro−inositol, malonate, erythritol, 4−guanidinobutanoate, and soyasaponin II [several citations...]. This signified a reduction in the normal levels of polysaccharide fermentation present in the intact microbiota. Furthermore, N−methylpipecolate was the only metabolite identified by these means to consistently decrease during antibiotic pretreatment. This molecule has been established to be a byproduct of amino acid catabolism and is typically recycled under normal conditions10. These data support that amino acid catabolism is disrupted at a community-level following antibiotic treatment and may suggest open nutrient niches for *C. difficile*, a known fermenter of peptides11, to colonize. We then applied this technique to identify metabolites that distinguish antibiotic pretreatment groups (Fig. 3d). Interestingly, several additional carbohydrate or amino acid fermentation byproducts were highlighted by these means including hydroxyisocaproate, methylvalerate, glutamylmethionine, and N−carbamoylaspartate [more citations...]. These data suggested that the populations responsible for normal chains of fermentation may be differentially effected by distinct classes of antibiotics. It also supported that the distinct antibiotics allowed for different forms of nutrient catabolism to survive the treatment and potentially implied alternative profiles of metabolic competition for *C. difficile* to cope with upon colonization. Additionally, sucrose and carnitine are *C. difficile* growth substrates [@] which further promoted the hypothesis of differentially vacated nutrient niches due to antibiotic treatment that are now accessible to *C. difficile*. After following this unsupervised approach for identification of susceptibility markers, we also measured differences in relative concentration of metabolites previously connected to potential *C. difficile* colonization susceptibility []. In this way, we assessed differences in bile acids, deeply connected to the life cycle of *C. difficile*12 and whose bioconversion by the microbiota has been implicated as a driver of colonization resistance13. We found that there was no persistently increased bile acid across the chosen antibiotic pretreatments (Fig. S2a). A similar trend was also seen in amino acids (Fig. S2b). This suggested that despite varying efficiencies, the fact that *C. difficile* can recognize a subset of these molecules in any of the observed contexts appears to be sufficient to allow for sufficient germination and outgrowth to occur. For carbohydrates (Fig. S2b), we found that several were significantly increased across all antibiotic groups which included arabitol/xylitol, ribitol, and sucrose. Together, our results supported that each susceptible environment was distinguishable from other groups with its own subset of enriched *C. difficile* growth substrates. This could be an indication that particular competitors were eliminated during antibiotic treatment, or those community members normally responsible for the consumption of these metabolites have altered their metabolic program to exploit alternative nutrient sources.

***C. difficile* colonization alters the cecal metabolome.** Following the changes to the cecal metabolome in response to antibiotic pretreatment, we then assessed the degree to which *C. difficile* colonization alters the cecal metabolome of susceptible animals. We hypothesized that the introduction of a new competitor, *C. difficile*, would impact the metabolome either through signatures of its own metabolism or causing a shift in the metabolism of other members in the surrounding community. First, in a similar approach to identifying susceptibility markers, we observed the global Bray-Curtis dissimilarities of mock-infected and *C. difficile*-infected metabolomes within each antibiotic pretreatment group separately. First we performed this analysis at a global level and compared cecal metabolomes from all mice across pretreatment groups, and were unable to detect a consistent difference between groups in this way (*p* = 0.075; Fig. S3a). We moved on to individual antibiotic groups and found that both streptomycin (*p* = 0.039) and cefoperazone (*p* = 0.016) pretreated metabolomes deviated significantly from that of mock infection (Fig. 4a-4b). However, clindamycin pretreated cecal metabolomes were not significantly altered by the presence of the pathogen (*p* = 0.127; Fig. 4c). These results diverged from what was seen in the paired OTU relative abundance results where instead the community structure of cefoperazone was unchanged and clindamycin was significantly different (*p* = 0.003; Fig. S4e, S4f, & 4c). Interestingly, streptomycin-pretreated microbiomes were significantly altered by *C. difficile* infection at both the OTU and metabolomic levels (Fig. S4d & 4a). These data indicated that large shifts of populations in the cecal microbiota was not implicitly associated with concordant shifts in the metabolome. This supported the hypothesis that divergent community structures can ultimately share a convergent metabolic output despite changes to community structure and membership.

We then sought to identify changes those metabolites that were the most effective markers of *C. difficile* infection. To accomplish this task, we applied the same machine learning techniques as in Figure 3 to differentiate infected and uninfected metabolomes overall and within each pretreatment group separately. With this approach we were able to accurately differentiated those microbiomes infected with *C. difficile* and those that were not (Fig. S3b; OOB = 11.11%). By far the strongest signature of infection across our metabolomic datasets were the significant decreases to the concentration of 4 individual proline-containing amino acids which were each highly abundant in the absence of infection (all *p* < 0.001). *C. difficile* is known to catabolize proline along with glycine through a set of paired biochemical reactions known as Stickland fermentation14. This supported that while distinct metabolic challenges may exist in each susceptible metabolome, Stickland fermentation could perhaps be *C. difficile*'s preferred carbon acquisition pathway *in vivo*. In agreement with the ordination analysis from Figure 4a-4c, random forest was only able to reliably classify infected mice in streptomycin and cefoperazone pretreatment while clindamycin maintained a high out-of-bag error (OOB = 44.44%). Metabolites that distinguished infection conditions within each pretreatment group had a high level of variation with only a few shared metabolites including 5-aminovalerate, acetylarginine, dimethylguanine, and adenine however none were in top 5 of all groups. Despite a moderate amount of conserved metabolic signatures across infections these data support our hypothesis that not only does each antibiotic pretreatment create different microbiota community structures with different metabolic potentials, but also that the metabolism of these distinct communities responds to *C. difficile* colonization in a unique fashion. However, it is not possible from these data to distinguish changes to the metabolome that were a result of altered community metabolism, altered host metabolism, or from *C. difficile* directly.

***C. difficile* colonization may induce shifts the expression of several catabolic pathways and nutrient acquisition systems across the cecal microbiota** Our combined 16S and metabolomic results demonstrated that antibiotic pretreatments resulted in distinct bacterial communities which likely lead to the altered metabolite profiles in each, however it does not preclude the possibility that the host or *C. difficile* itself (only during infection) were responsible for the metabolomic differences. We hypothesized that the altered metabolic function of the microbiota as a product of antibiotic perturbation drove the changes seen in the metabolomes of the respective antibiotic pretreatment groups. In order the most closely analyze which bacterial groups differentially responded to infection, we employed a metagenomic-enabled metatranscriptomic shotgun sequencing approach with paired DNA and RNA samples collected from the cecal content of the mice used in the previous analyses. Metagenomic reads from mock-infected cecal communities were respectively assembled into contigs and putative genes were identified resulting in 234,124 (streptomycin), 83,096 (cefoperazone), and 35,977 (clindamycin) potential genes in each metagenome. Streptomycin pretreatment resulted in a significantly more diverse community (Fig. S1a) than other groups, so a larger detectable metagenome was expected. Putative genes were then annotated according to KEGG and the subset of genes the were successfully annotated with function were utilized for the next analysis; 15.51% (streptomycin), 19.47% (cefoperazone), and 34.78% (clindamycin). Despite utilizing only mock-infected metagenomes for assembly, we still removed genes in which the top alignment was to *C. difficile* prior to read mapping. This was done to avoid skewed results between uninfected and infected metatranscriptomes where *C. difficile* was present to contribute to the transcript pool. Quality-trimmed DNA and cDNA reads from both infected and uninfected conditions were then mapped to the gene catalog that corresponded to their antibiotic pretreatment group of origin. The resulting abundances were normalized to both sequencing read length and target gene length to yeild a per base mapping abundance. Finally, after equal subsampling, metatranscriptomic read abundances for each gene were divided by their corresponding metagenomic coverage in order to normalize for overrepresented genes and species. Therefore, final expression values represent the level of expression upregulation for each gene outside of those from which transcript is abundant but is only expressed at low levels from genes which are highly abundant.

Utilizing the fully normalized metatranscriptomes, we first focused on expressional differences in broadly defined pathways and gene catagories for each antibiotic pretreatment with and without *C. difficile*-infection. We hypothesized that pathways with the greatest amount of change would be involved in the metabolism of carbohydrates and amino acids to reflect differences seen at the metabolomic level. To observe these differences we quantified expression of all genes in each community that were annotated as part of a given pathway, regardless of taxonomic information. We then calculated the difference in cDNA abundance for each pathway between infected and uninfected conditions, represented as delta-cDNA abundance. To highlight the largest dfferences, we limited the analysis to the top 5 KEGG pathways with the most change between mock and *C. difficile*-infected conditions within each pretreatment (Fig. 5a). Average expression was nearly identical in either experimental or control, which demonstrated that similar levels of overall transcription was occurring in either state in spite of the large differences shown. In streptomycin and clindamycin pretreatments, greater expression of KEGG pathways was observed in the *C. difficile*-infected metatranscriptomes. Both groups displayed large changes in amino sugar metabolism and ABC transporters, however other distinct carbon metabolism pathways were upregulated in each. Glycolysis and oxidative phosphorylation are overrepresented in streptomycin pretreatment while starch/sucorse metabolism and PTS systems were more abundant associated with clindamycin. Together these shifts suggest that these communities differentially adapt carbon metabolism pathways in response to colonization of the invasive species *C. difficile*. Conversely, the largest differences seen in cefoperazone pretreatment were overexpressed in the absence of *C. difficile* infection. These pathways included three separate pathways for the replication or manipulation of genetic material (RNA Processing, tRNA Sythesis, & Homologous Recombination). Instead, many genes involved in anaerobic glucose metabolism and select ABC-transporters are upregulated (Table S3). These results support that the cecal microbiota of infected mice shifts metabolism toward metabolizing simple carbohydrates which could be the result of competition with *C.difficile* for otherwise uncontested resources. We then moved on to perform a more fine-scale analysis of changes at the individual gene annotation level selecting the genes in each antibiotic pretreatment group with the largest disparity in normalized cDNA abundance between mock and *C. difficile*-infected groups (Fig. 5b-5d). In agreement with pathway-level differences in expression, the majority genes with high degrees of difference belonged to pathways highlighted in Fig. 5a with three additional pathways relating to amino acid metabolism (I: Glutathione metabolism, J: Valine/Leucine/Isoleucine metabolism, & K: Glycine/Serine/Threonine metabolism). Additionally, numerous genes for transport of simple carbohydrates and glycolysis are differentially overrepresented under infected conditions across pretreatment groups when compared to mock infection of untreated, resistant communities (Fig S6). These data expand on our interpretation of pathway-level analysis and support our hypothesis that *C. difficile* colonization leads to changes in community-level expression of genes for nutrient acquisition and catabolism.

***C. difficile* colonization corresponds with large-scale changes in expression of specific bacterial taxa between perturbed cecal environments** As not all bacterial taxa share identical metabolic capabilities, we hypothesized that specific subsets of bacteria were differentially effected by the presence or metabolic activity of *C. difficile* due to increased competition for resources which in turn determined changes to the metabolome. With this in mind, we aspired to delineate the transcriptomic contributions of separate bacterial taxa to each metatranscriptome. To accomplish this we utilized the genus level taxonomic information associated with each KEGG annotation to identify which group likely contributed a given gene to the metagenome. Many genes in the KEGG database are annotated as hypothetical or uncharacterized but still possess a taxonomic annotation. This resulted in substantially more genes from the total being conserved for analysis in each group; 19.15% (streptomycin), 25.1% (cefoperazone), and 44.45% (clindamycin). With these data, we narrowed the focus onto transcription for genera that represented >0.01% of genes receiving taxonomic annotations in any of the three metagenomic assemblies. We then directly compared the normalized cDNA abundances for each gene between infected & uninfected states for each antibiotic pretreatment and colored points based on genus classifications (Fig. 3a-3c). Coordinates were determined by the relative expression of each gene in mock and *C. difficile* infection for x and y axes respectively. This causes genes with equal transcription in both conditions being compared to be strongly correlated and positioned proximal to the central diagonal line. As such, we then applied linear correlation and a squared residual cutoff to define those genes that are most strongly upregulated in either condition.

These data revealed distinct changes in expression for underrepresented genera for each antibiotic pretreatment between infections conditions. First, in streptomycin-pretreated mice, the most prominent differences were in 937 genes belonging to *Lactobacillus* were upregulated in *C. difficile* infection (Fig. 6a). Next, in cefoperazone pretreatment 2290 genes belonging to *Bacteroides* were more upregulated in mock infection (Fig. 6b). A consistent trend in streptomycin and cefoperazone pretreatments was an overrepresentation of expression from genera belonging to Bacteroidetes during mock infection. The metatransciptomes for both of these pretreatment poorly correlated with one another, indicating a high degree of change induced by *C. difficile* colonization (*r* = 0.0334 & *r* = 0.031). Finally, in clindamycin pretreatment the largest difference in transcription was for 510 *Lactobacillus* genes during *C. difficile* infection (Fig. 6c). Infected and uninfected metatranscriptomes associated with this antibiotic correlated the more strongly than either other pretreatment (*r* = 0.862), supporting that *C. difficile* colonization had a low impact on transcription of the cecal microbiota. This infection model has been shown to go on to clear *C. difficile* infection witin 7 days, as opposed to long-term colonization, which suggests that communities where *C. difficile* is less able to impact resident members' activity regain colonization resistance more quickly (Fig. S5....). These results from both cefoperazone and clindamycin pretreatments particularly corresponded with LEfSe analysis (Fig. 1e & 1f) where a *Lactobacillus* OTU was shared and significantly altered in abundance by the presence of *C. difficile*. Interestingly, taxonomic groups most highly represented as outliers in the normalized metatranscriptomes of all tested antibiotic pretretreatments were actually minority members of the respective cecal communities as determined 16S rRNA gene abundance (Fig. 1c). On average, *Bacteroides* was the most commonly altered taxa during infection with mean of 441 genes appearing as expression outliers across the three pretreatment goups. As a side note, the presence of Archeael sequences in our shotgun sequencing data was not surprising as the universal 16S rRNA gene V4 region primers we used do not reliably amplify Archeae. Together, our results support that *C. difficile* may differentially modify the transcriptional activity of separate microbial taxa based on the context of the community in which it is colonizing.

**Correlating changes in the metabolome and metatranscriptome** After observing separate bacterial taxa differentially expressing transcrtipt based on antibiotic pretreatment and infection status, we then hypothesized that the changes we saw at the metabolomic level were due to the activity of specific bacterial taxa that were differentially effected by *C. difficile* over others. To examine this we correlated the expression of specific metabolic pathways in select bacterial tax with the changes in relative concentration of metabolites associated with those pathways in each of the pretreatment groups tested.

### Discussion

Collectively, our results support the hypothesis that

First, we identify distinct changes in bacterial community structure in response to antibiotic pretreatment as well as *C. difficile* colonization. Results summary and implication

This study is the first of it's kind to combine multiple forms of high-dimensional biological data to characterize the responses of resident microbial communities to infection. The intimate connection of *C. difficile* we acknowledge that colonization could be driven more by the increase in relative abundance or activity of synergistic species or processes with *C. difficile*, however previous work\_\_\_\_\_ supports competition as the most likely primary contributor... (can, but shouldnt look at synergy in these models) Novelty of study possible limitations

and contributes the the concept that the healthy gut microbiota maintains colonization resistance to *C. difficile* by outcompeting the pathogen for preferred nutrient niche space. closing words and future directions This work has implications

The finding that c diff uses amino acids in each environment agrees with previous research in that....

Our results support the hypothesis that not only does the nutrient niche of *C. difficile* during infection adjust to the susceptible community composition, but also that certain members of the microbiota are differntially impacted by infection.

Colonization resistance to c difficile may be the result of the contribution by distinct subcommunities of bacteria across each unique community. Conversely, these results also offer support to the nutrient niche plasticity of *C. difficile* as it colonizes differentially sensitized gut ecosystems.

specific pathways turn on/off following infection effecting metabolites and have implications for the nutrient niche space of *C. difficile*.

### Materials and Methods

**Animal care and antibiotic administration** For a more detailed description of the procedure, refer to Jenior et al. (2017)**???**. In short, conventionally-reared six-to-eight week-old C57BL/6 mice were administered one of three antibiotics; cefoperazone, streptomycin, or clindamycin (As described in Table S1) before oral *C. difficile* infection. All animal protocols were approved by the University Committee on Use and Care of Animals at the University of Michigan and carried out in accordance with the approved guidelines.

***C. difficile* infection and necropsy.** For a more detailed description of the procedure, refer to Jenior et al. (2017)**???**. On the day of challenge, 1×103 *C. difficile* spores were administered to mice via oral gavage in phosphate-buffered saline (PBS) vehicle. Mock-infected animals were given an oral gavage of 100 μl PBS at the same time as those mice administered *C. difficile* spores. 18 hours following infection, mice were euthanized and necropsied to obtain the cecal contents. Aliquots were immediately flash frozen for later DNA extraction and toxin titer analysis. A third aliquot was transferred to an anaerobic chamber for quantification of *C. difficile* abundance. The remaining content in the ceca was mixed in a stainless steel mortar housed in a dry ice and ethanol bath. Cecal contents from all mice within each treatment group were pooled into the mortar prior to grinding to a fine powder. The ground content was then stored at -80° C for subsequent RNA extraction.

***C. difficile* cultivation and quantification** For a more detailed description of the procedure, refer to Jenior et al. (2017)**???**. Cecal samples were weighed and serially diluted under anaerobic conditions with anaerobic PBS. Differential plating was performed to quantify both *C. difficile* spores and vegetative cells by plating diluted samples on CCFAE plates (fructose agar plus cycloserine, cefoxitin, and erythromycin) at 37° C for 24 hours under anaerobic conditions**???**. In parallel, undiluted samples were heated at 60° C for 30 minutes to eliminate vegetative cells and leave only spores12. These samples were serially diluted under anaerobic conditions in anaerobic PBS and plated on CCFAE with taurocholate at 37° C for 24 hours. Plating was simultaneously done for heated samples on CCFAE to ensure all vegetative cells had been eliminated.

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

**DNA/RNA extraction and sequencing library preparation** DNA for downstream shotgun metagenomic and 16S rRNA gene sequencing was extracted from approximately 50 mg of cecal content from each mouse using the PowerSoil-htp 96 Well Soil DNA isolation kit (MO BIO Laboratories) and an epMotion 5075 automated pipetting system (Eppendorf). The V4 region of the bacterial 16S rRNA gene was amplified using custom barcoded primers15. Equal molar rations of raw isolated DNA within each treatment group were then pooled and ~2.5 ng of material was used to generate shotgun libraries with a modified 10-cycle Nextera XT genomic library construction protocol (Illumina). This was done to mimic the pooling strategy necessary for metatranscriptomic library preparation. Final libraries were pooled at equal molar ratios and stored at -20° C. For RNA extraction, a more detailed description of the procedure can be found in**???**. Briefly, immediately before RNA extraction, 3 ml of lysis buffer (2% SDS, 16 mM EDTA and 200 mM NaCl) contained in a 50 ml polypropylene conical tube was heated for 5 minutes in a boiling water bath**???**. The hot lysis buffer was added to the frozen and ground cecal content. The mixture was boiled with periodic vortexing for another 5 minutes. After boiling, an equal volume of 37° C acid phenol/chloroform was added to the cecal content lysate and incubated at 37° C for 10 minutes with periodic vortexing. The mixture was the centrifuged at 2,500 x g at 4° C for 15 minutes. The aqueous phase was then transferred to a sterile tube and an equal volume of acid phenol/chloroform was added. This mixture was vortexed and centrifuged at 2,500 x g at 4° for 5 minutes. The process was repeated until aqueous phase was clear. The last extraction was performed with chloroform/isoamyl alcohol to remove acid phenol. An equal volume of isopropanol was added and the extracted nucleic acid was incubated overnight at -20° C. The following day the sample was centrifuged at 12000 x g at 4° C for 45 minutes. The pellet was washed with 0° C 100% ethanol and resuspended in 200 μl of RNase-free water. Following the manufacturer's protocol, samples were then treated with 2 μl of Turbo DNase for 30 minutes at 37° C. RNA samples were retrieved using the Zymo Quick-RNA MiniPrep according the manufacturer's protocol. The Ribo-Zero Gold, immediately before RNA extraction, 3 ml of lysis buffer (2% SDS, 16 mM EDTA and 200 mM NaCl) contained in a 50 ml polypropylene conical tube was heated for 5 minRNA Removal Kit Epidemiology was then used to deplete prokaryotic and eukaryotic rRNA from the samples according the manufacturer's protocol. Stranded RNA-Seq libraries were made constructed with the TruSeq Total RNA Library Preparation Kit v2, both using the manufacturer's protocol. Average length of amplicon libraries for both DNA and cDNA sequencing was performed using an Agilent BioAnalyzer with High Sensativity DNA Analysis kits. Completed libraries were pooled in equal molar ratios within their respective groups and stored at -20° C until time of sequencing.

**High-throughput sequencing and raw read curation** Sequencing of 16S rRNA gene amplicon libraries was performed using an Illumina MiSeq sequencer as described previously15. The 16S rRNA gene sequences were curated using the mothur software package (v1.36) as described in**???**. Shotgun metagenomic sequencing was performed in 2 phases. Libraries from mock-infected communities, that were also to be utilized for contig assembly, were sequenced using an Illumina HiSeq 2500 on 2x250 paired-end settings and was repeated across 2 lanes to normalize for inter-run variation. *C. difficile*-infected metagenomic libraries were sequenced with an Illumina NextSeq 300 with 2x150 settings across 2 runs to also normalize for inter-run variation. These efforts resulted in an average of 280 million paired raw reads per sample. Metatranscriptomic sequencing was performed on an Illumina HiSeq 2500 with 2x50 settings and was repeated across 4 lanes for normalization and to obtain necessary coverage16. This gave an average of 380 million raw cDNA per library. Both metagenomic and metatranscriptomic sequencing was performed at the University of Michigan Sequencing Core. Raw sequencing read curation for both metagenomic and metatranscriptomic datasets was performed in a two step process. Residual 5’ and 3’ Illumina adapter sequences were trimmed using CutAdapt**???** on a per library basis. Reads were quality trimmed using Sickle17 with a quality cutoff of Q30. This resulted in approximately 270 million reads per library (both paired and orphaned) for both metagenomic and metatranscriptomic sequencing. Actual read abundances for individual metagenomic and metatranscriptomic sequencing efforts can be found in Table S2.

**Metagenomic contig assembly and gene annotation.** Metagenomic contigs were assembled using Megahit**???** with the following settings; minimum kmer size of 87, maximum kmer size of 127, and a kmer step size of 10. Progigal was utilized to to identify putative gene sequences, and were screen for a minimum length of 250 nucleotides. These sequences were translated to amino acids and peptides were annotated against the KEGG protein database18 using Diamond implementation of BLASTp19. Peptide-level gene annotations were assigned to the corresponding nucleotide sequence, and genes failing to find a match in KEGG were preserved as unannotated genes. Final nucleotide fasta files with KEGG annotations were then utilized in the construction of Bowtie220 mapping databases from downstream analyses.

**DNA/cDNA read mapping and normalization** Mapping was accomplished using Bowtie220 and the default stringent settings. Optical and PCR duplicates were then removed using Picard MarkDuplicates (<http://roadinstitute.github.io/picard/>). The remaining mappings were converted to idxstats format using Samtools21 and the read counts per gene were tabulated. Discordant pair mappings were discarded and counts were then normalized to read length and gene length to give a per base report of gene coverage. Unless indicated otherwise, each collection of reads was then 1000-fold iteratively subsampled to 90% of the lowest sequence total within each analysis, and a median expression value for each gene was calculated.

**Quantification of *in vivo* metabolite relative concentrations.** For a more detailed description of the procedure, refer to Jenior et al. (2017)**???**. Metabolomic analysis performed by Metabolon (Durham, NC), a brief description of their methods is as follows. All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer at 35,000 mass resolution. Samples were dried then reconstituted in solvents compatible to each of the four methods. The first, in acidic positive conditions using a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). The second method was identical to the first but was chromatographically optimized for more hydrophobic compounds. The third approach utilized a basic negative ion optimized conditions using a separate dedicated C18 column. Basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. Samples were then analyzed via negative ionization following elution from a hydrophilic interaction chromatography column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS n scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Library matches for each compound were checked for each sample and corrected if necessary. Peaks were quantified using area under the curve.

**Statistical methods.** All statistical analyses were performed using R (v.3.2.0). Significant differences between community structure of treatment groups from 16S rRNA gene sequencing were determined with AMOVA in the mothur software package. Significant differences of Inv. Simpson diversity, cfu, toxin titer, and metabolite concentrations were determined by Wilcoxon signed-rank test with Benjamini-Hochberg correction. Undetectable points used half the limit of detection for cfu and toxin statistical calculations.

### 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).

### Acknowledgments

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

**Author Affiliations** **Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan.** Matthew L. Jenior, Jhansi L. Leslie, & Patrick D. Schloss 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. 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 | Experimental models of *C. difficile* infection and distinct virulence patterns.** **(a)** Experimental timelines of antibiotic pretreatment and infection mouse models. **(b)** Quantification of *C. difficile* c.f.u. and toxin titer across antibiotic pretreatment models. Gray asterisks indicate significant difference from no antibiotic controls by Wilcoxon rank-sum test with the Benjamini-Hochberg correction (all *p* < 0.001).

**Figure 2 | Impact of *C. difficile* colonization on other bacterial populations abundances in the gut microbiota.** **(a)** Relative abundance of family-level taxonomic classification for OTUs in each treatment group. **(b-d)** Discriminating OTUs with LEfSe analysis between Mock and *C. difficile*-infected communities within each pretreatment group, *p*-values from LEfSe are shown on the left with taxonomic information (all *p* < 0.05). Relative abundance of the respective OTUs from each mouse along with medians are shown for each pretreatment group. Black asterisks indicate significant difference in relative abundance determined by Wilcoxon rank-sum test with Benjamini-Hochberg correction (*p* < 0.05).

**Figure 3 | Metabolite markers of *C. difficile* infection susceptibility.** Results from UPLC/MS metabolomic analysis of cecal content from mice also used in previous analyses. Only mock-infected metabolomic results were included this analysis to identify markers of susceptibility. **(a-b)** NMDS ordinations of Bray-Curtis distances comparing the cecal metabolome dissimilarity of mice receiving no pretreatment or one of the three distinct classes of antibiotics. (a) Metabolomes of resistant mice are significantly different from antibiotic pretreated animals (*p* < 0.001). (b) Antibiotic pretreated metabolomes and also significantly distinct from one another (*p* < 0.001). Significant differences for NMDS analyses were calculated with AMOVA. **(c-d)** Scaled intensities of metabolites with highest mean decrease in accuracy (MDA) from random forest feature selection discriminating groups from (a-b). MDA is labeled in brackets beside each metabolite name and out-of-bag error from internal cross-validation is labeling along the bottom axis. Asterisks along the right axis indicate significant difference by Wilcoxon rank-sum test. (c) Relative concentrations of metabolites that distinguish cecal content from mice resistant to *C. difficile* colonization from susceptible animals. (d) Metabolites concentrations for those molecules that are able to differentiate antibiotic pretreatment groups. Multiple comparisons were accounted for using the Benjamini-Hochberg correction (\*\*\* < 0.001, \*\* <= 0.01, \* <= 0.05).

**Figure 4 | *C. difficile* infection has differential effects on the cecal metabolome.** **(a-c)** NMDS ordinations of Bray-Curtis distances comparing infection groups within each antibiotic pretreatment; (a) Streptomycin, (b) Cefoperazone, and (c) Clindamycin. Significant differences were determined by AMOVA. **(d-f)** Relative concentrations of metabolites with highest Mean Decrease Accuracy (MDA) from Random Forest feature selection to discriminate between Mock and *C. difficile*-infected conditions within each antibiotic pretreatment. Respective MDA is labeled in brackets next to each metabolite name and out of bag error from internal cross-validation of each model is labeled under the plotting area. (d) Streptomycin pretreatment, (e) Cefoperazone pretreatment, and (f) Clindamycin pretreatment. Significant differences are labeled along the right axis and were calculated with Wilcoxon rank-sum test with Benjamini-Hochberg correction (\*\*\* < 0.001, \*\* <= 0.01, \* <= 0.05).

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

**Figure 6 | *C. difficile* colonization alters gene expression of taxonomic groups differentially between antibiotic pretreatements.** Each point represents a unique gene from the respective metagenomic assembly. Coordinates were determined by the log2-transformed expression level of each gene between *C. difficile*-infected and mock-infected conditions. Metatranscriptomic read abundances were normalized to their associated metagenomic coverage. Colored indicate genus of origin, and gray areas denote genes with consistent expression between conditions and outliers to this region were determined by least squares regression analysis with a minimum residual value of 2. Antibiotic pretreatments **(a)** Streptomycin-pretreated, **(b)** Cefoperazone-pretreated, and **(c)** Clindamycin-pretreated.

**Supplementary Figure 1 | Impact of antibiotic pretreatment and *C. difficile* infection on cecal community structure** **(a)** Inverse-Simpson diversity of cecal communities from all treatment groups. Gray stars indicate significant difference from no antibiotic controls (all *p* < 0.001). Black stars denote within treatment group significant difference between mock and *C. difficile*-infected communities. Differences were calculated using Wilcoxon rank-sum test with Benjamini-Hochberg correction. NMDS ordinations of Bray-Curtis distances comparing the groups labeled in the bottom left of each plotting area. **(b)** Antibiotic-pretreated compared to Untreated controls and **(c)** comparison between only antibiotic pretreatment groups. **(d-f)** Within antibiotic pretreatment comparisons for the effect of *C. difficile* colonization on community structure. Significant differences and correlation coefficients for ordination analyses were found using ANOSIM.

**Supplementary Figure 2 | Relative concentrations of bile acids in each antibiotic pretreatment group** Significant differences were determined by Wilcoxon rank-sum test with Benjamini-Hochberg correction.

**Supplementary Figure 3 | Global effect of infection on the cecal metabolome**

**Supplementary Figure 4 | Heatmap comparing Resistant and Susceptible mice for all metabolites** shown is the median of each group

**Supplementary Figure 5 | Heatmap comparing antibiotic pretreated groups** shown is the median of each group

**Supplementary Figure 6 | Gene-level expression for each community compared to the level expressed in resistant mice for each gene**

**Supplementary Table 1 | Antibiotic pretreatment regimes** **Supplementary Table 2 | High-throughput sequencing read counts and metagenomic assembly efficiency** **Supplementary Table 3 | Normalized cDNA abundances for each antibiotic pretreated group** **Supplementary Table 4 | Median scaled intensity for each metabolite for all antibiotic pretreatment groups** **Supplementary Table 5 | Results from Random Forest feature selection for metabolomic results**

### References

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

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

3. Eijk, E. van *et al.* Complete genome sequence of the Clostridium difficile laboratory strain 630erm reveals differences from strain 630, including translocation of the mobile element CTn5. *BMC Genomics* **16,** 31 (2015).

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

5. Robinson, C. J. & Young, V. B. Antibiotic administration alters the community structure of the gastrointestinal microbiota. *Gut Microbes* **1,** 279–284 (2010).

6. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biology* **12,** R60 (2011).

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

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

9. Liaw, a & Wiener, M. Classification and Regression by randomForest. *R news* **2,** 18–22 (2002).

10. Goyer, A. *et al.* Characterization and Metabolic Function of a Peroxisomal Sarcosine and Pipecolate Oxidase from Arabidopsis. *Journal of Biological Chemistry* **279,** 16947–16953 (2004).

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

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

13. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517,** 205–208 (2014).

14. Bouillaut, L., Self, W. T. & Sonenshein, A. L. Proline-dependent regulation of Clostridium difficile stickland metabolism. *Journal of Bacteriology* **195,** 844–854 (2013).

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

16. Franzosa, E. A. *et al.* Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences* **111,** E2329–E2338 (2014).

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

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

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

20. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9,** 357–9 (2012).

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