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

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

Abstract

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

growth substrates. To assess which resources *C. difficile* exploited, we developed a genome-scale metabolic model with a transcriptomic-enabled metabolite scoring algorithm integrating network architecture. This platform identified nutrients *C. difficile* used preferentially between infections. These findings were validated through untargeted mass spectrometry of each microbiome. Our results supported the hypothesis that *C. difficile* inhabits alternative nutrient niches across cecal microbiomes with increased preference for nitrogen-containing carbon sources, particularly Stickland fermentation substrates and host-derived aminoglycans.

Importance

Infection by the bacterium *Clostridium difficile* causes an inflammatory diarrheal disease which can become life-threatening, and has grown to be the most prevalent nosocomial infection. Susceptibility to *C. difficile* infection is strongly associated with previous antibiotic treatment, which disrupts the gut microbiota and reduces its ability to prevent colonization. In this study we demonstrated that *C. difficile* altered pathogenesis between hosts pretreated with antibiotics from separate classes, as well as exploited different nutrient sources across these environments. Our metabolite importance calculation also provides a platform to study nutrient requirements of pathogens during the context of infection. Our results suggest that *C. difficile* colonization resistance is mediated by multiple groups of bacteria competing for several subsets of nutrients, and could explain why total reintroduction of competitors through fecal microbial transplant is the most effective treatment to date. This work could ultimately contribute to the identification of targeted measures that prevent or reduce *C. difficile* colonization including pre- and probiotic therapies.

Introduction

Infection by the Gram-positive, spore-forming bacterium *Clostridium difficile* has increased in both prevalence and severity across numerous countries during the last decade (1). In the United States, *C. difficile* was estimated to have caused >500,000 infections and resulted in ~$4.8 billion worth of acute care costs in 2014 (2). *C. difficile* infection (CDI) causes an array of toxin-mediated symptoms ranging from abdominal pain and diarrhea to the more severe conditions pseudomembraneous colitis and toxin megacolon. Prior treatment with antibiotics is the most common risk factor associated with development of CDI (3). Antibiotics likely contribute to susceptibility to CDI by disrupting the gut microbiota (4). In mouse models, multiple antibiotics can induce susceptibility to *C. difficile* colonization (5–7). Notably, each antibiotic resulted in unique gut bacterial communities that were receptive to high levels of *C. difficile* colonization. Others have also shown that antibiotics from multiple classes also alter the gut metabolome, increasing the concentrations of some *C. difficile* growth substrates (6, 8–10). The ability of an unaltered murine gut community to exclude *C. difficile* colonization supports the nutrient-niche hypothesis, which states that an organism must be able to utilize a subset of available resources better than all competitors to colonize the intestine (11, 12). Taken together these results are a strong indication that the healthy gut microbiota inhibits the growth of *C. difficile* by limiting the availability of the substrates it needs to grow.

Based on genomic and *in vitro* growth characteristics, *C. difficile* appears able to fill multiple nutrient niches. *C. difficile* has a relatively large and mosaic genome, it can utilize a variety of growth substrates, and possesses a diverse array host range (13–

15). These qualities are hallmarks of ecological generalists (16). *C. difficile* has also been shown to integrate signals from multiple forms of carbon metabolism to regulate its pathogenesis. *In vitro* transcriptomic analyses suggests that high concentrations of easily metabolized carbon sources, such as glucose or amino acids, inhibit toxin gene expression and sporulation (17, 18). Other studies have indicated that other aspects of *C. difficile* metabolism may be influenced through environmental nutrient concentration-sensitive global transcriptional regulators such as CodY and CcpA (19, 20). These previous analyses have mainly focused on *in vitro* growth (21, 22) or colonization of germfree mice (20, 23). Although these analyses are informative, they are either primarily directed toward the expression of pathogenicity factors or lack the context of the gut microbiota which *C. difficile* must compete against for substrates. Metabolomic investigations have also been used to assay changes in bacterial metabolism as they relate to CDI and have characterized the levels of germinants and growth substrate availability (6, 10); however, metabolomic approaches are unable to attribute a metabolite to specific organisms in the gut community. Thus metabolomics more closely represents the echoes of total community metabolism, not the currently active processes of any one population. It has thus far not been possible to study *C. difficile*'s metabolism *in vivo*.

To overcome these limitations, we implemented transcriptomic and untargeted metabolomic analyses of *C. difficile* and the surrounding environemnt to better understand the active metabolic pathways in a model of infection. Based on the ability of *C. difficile* to grow on a diverse array of carbon sources and its ability to colonize a variety of communities, we hypothesized that *C. difficile* adapts its metabolism to fit the

context of the environment it is attempting colonize. To test this hypothesis, we employed a mouse model of infection to compare the response of *C. difficile* to the gut environment caused by three antibiotics from distinct classes. By characterizing the transcriptome of *C. difficile* in these different communities and the metabolome of the respective environments, we were able to generate a systems model to directly test the nutrient-niche hypothesis.

Results

Levels of *C. difficile* sporulation and toxin activity vary among different microbiomes. Conventionally-reared SPF mice were treated with either streptomycin, cefoperazone, or clindamycin (Table 1 and Fig. S1). These antibiotics were selected because they each have distinct and significant impacts on the structure of the cecal microbiome (Fig. S2A and S2B). We challenged the antibiotic treated mice and germfree (ex-GF) mice with *C. difficile* stain 630 to understand the pathogen's physiology with and without other microbiota. This toxigenic strain of *C. difficile* was chosen for its moderate clinical severity in mouse models (24) and well-annotated genome (25). After infection, we measured sporulation and toxin production at 18 hours post inoculation. That time point corresponded with when another laboratory strain of *C. difficile* reached its maximum vegetative cell density in the cecum with limited sporulation (26). There was not a significant difference in the number of vegetative *C. difficile* cells in the ceca of mice pretreated with any of the three antibiotics (Fig. 1A). All antibiotic treated and ex-GF mice were colonized to ~1×10^8 colony forming units (cfu) per gram of cecal content, while untreated mice maintained colonization resistance to *C. difficile* (Table 1). Despite having the same number of vegetative *C. difficile* cells, more

spores were detected in ex-GF mice than in the antibiotic pretreated mice (P = 0.003, 0.004, and 0.003; Fig. 1B). There was also a significantly higher toxin titer in ex-GF animals than any other colonized group (all $P < 0.001$), with slight variation between antibiotic pretreatment groups (Fig. 1C). These results showed that *C. difficile* colonized different communities to consistently high levels. In addition, colonization in the context of different microbiomes resulted in moderate differences in the expression of *C. difficile* pathogenicity. To investigate the physiology of *C. difficile* when colonizing distinct susceptible gut environments, we performed whole transcriptome analysis of *C. difficile* growing in the cecum during infection.

***C. difficile* alters its gene expression pathways when colonizing distinct antibiotic-pretreated environments.** We first attempted to measure differential expression of specific genes associated with *in vivo* phenotype changes reported in previous studies. These included genes involved in sporulation, toxin production, quorum sensing, and metabolite-regulated sigma factors (Fig. S3). Despite large-scale differences between pretreatment groups, no clear trends were evident between gene expression and colonization, sporulation, or toxin production. This further indicated that *C. difficile* adapted its metabolism to the environment that it colonized. As such, we next focused on specific groups of genes known to contribute to *C. difficile* metabolism (Table S1 and Fig. 2A). Genes involved in amino acid catabolism, including those that encoded enzymes involved in Stickland fermentation and general peptidases, had the highest level of expression. Stickland fermentation refers to the coupled fermentation of amino acid pairs in which one is deaminated and the other is reduced to ultimately generate ATP (27). This suggested that *C. difficile* catabolized environmental amino

136 acids during infection, regardless of the structure of the surrounding community.
137 Although there were gene categories that were equally expressed across conditions in
138 spite of the community differences, there were patterns of expression for certain gene
139 families and specific genes that were distinct to each antibiotic pretreatment. In mice
140 pretreated with cefoperazone, *C. difficile* tended to have more expression of genes in
141 the ABC sugar transporter and sugar alcohol catabolism (e.g. mannitol) families and
142 fewer genes in the PTS transporter family than the other pretreatment groups. In mice
143 pretreated with clindamycin, *C. difficile* tended to have higher expression of genes from
144 disaccharide catabolism (e.g. (B)-galactosidases and trehalose/maltose/cellibiose
145 hydrolases), fermentation product metabolism (including consumption or production of
146 acetate, lactate, butyrate, succinate, ethanol, and butanol), and PTS transporter
147 families. Genes from the sugar alcohol catabolism and ABC sugar transporter families
148 were not highly expressed in the clindamycin-pretreated mice. Finally, in mice
149 pretreated with streptomycin, *C. difficile* had higher levels of expression of genes from
150 the sugar alcohol catabolism (e.g. sorbitol) and PTS transporter families. Combined,
151 these results suggested that while catabolism of amino acids and specific
152 carbohydrates are core components of the *C. difficile* nutritional strategy during
153 infection, *C. difficile* adapted its metabolism across different susceptible environments.

154 **Genome-scale metabolic model structure underscores known *C. difficile***
155 **physiology.** To further investigate which metabolites were differentially utilized between
156 conditions, we represented the metabolic network of *C. difficile* as a directed bipartite
157 graph using the genome and biochemical reaction annotations available in the KEGG
158 database. Enzymes and metabolites were represented by nodes, and their interactions

by directed connecting edges (Fig. 3A). The complete *C. difficile* network contained 447 enzymes and 758 metabolites, with 2135 directed edges. To validate our metabolic network, we analyzed network topology by calculating two metrics of centrality, betweenness centrality (BC) and closeness centrality (CC), to determine which nodes are critical to the structure of the metabolic network and if these patterns reflect known biology (Table S2). Both metrics utilize shortest paths, which refer to fewest possible number of network connections that lie between two given nodes. The BC of each node is the fraction of shortest paths that pass through that node and connect all other potential pairs of nodes. In biological terms, this refers to the amount of influence a given hub has on the overall flow of metabolism (28). Similarly, CC is the reciprocal sum of the lengths of shortest paths included in each node's BC. This value demonstrates how essential a given node is to the overall structure of the metabolic network (29). Metabolic network structural studies of *Escherichia coli* have found that metabolites with the highest centrality calculations are involved in fundamental processes in metabolism, namely glycolysis and the citrate acid cycle pathway (30). As such, these metrics allow for assessment of the degree to which a metabolic network accurately depicts established principles of bacterial metabolism.

Following application of both methods, we found 5 enzymes that were shared between the top 10 enzymes from BC and CC calculations (2-dehydro-3-deoxyphosphogluconate aldolase, aspartate aminotransferase, pyruvate-flavodoxin oxidoreductase, formate C-acetyltransferase, and 1-deoxy-D-xylulose-5-phosphate synthase). These enzymes primarily participate in core processes including glycolysis, the pentose phosphate pathway, or the citric acid cycle. Upon analysis of the other 15

high-scoring enzymes combined from BC and CC analyses, the majority were also components of the previously mentioned pathways, as well as several for the metabolism of amino acids (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" (30). Notably absent from the most well-connected metabolites were molecules like ATP or NADH. Their exclusion is likely a byproduct of the KEGG LIGAND reference used for network construction, which excludes cofactors from most biochemical reactions. While this may be a limitation of certain analyses, our study was not affected as the primary interest was in those substrates acquired from the environment. These results reflected the defined biological patterns of *C. difficile* and was therefore a viable platform to study metabolism of the pathogen.

Metabolite importance algorithm reveals adaptive nutritional strategies of *C. difficile* during infection of distinct environments. We next sought to include transcriptomic data with the metabolic model to infer which metabolites *C. difficile* most likely utilized from a given environment. To accomplish this we mapped normalized transcript abundances to the enzyme nodes in the network. The importance of each metabolite was measured as the $\log_2$-transformed difference in average transcript levels of enzymes that use the metabolite as a substrate and those that generate the metabolite as a product (Fig. 3B). A metabolite with a high importance score was more

likely obtained from the environment because the expression of genes for enzymes that produce the metabolite were low. *C. difficile* represented a small precentage of the community in each colonized environment (Fig. S2C), making it impossible to sequence the transcriptome of individual mice due to the depth required to sufficiently sequence the transcripts of *C. difficile*. Because of this, we generated a single transcriptome per condition using pooled mRNA from all mice within each pretreatment group (Table S5). We adopted a Monte Carlo-style simulation for iterative random transcriptome comparison to provide statistical validation of our network-based findings. This process generated random score distributions for all metabolite nodes in the network, which made it possible to calculate a confidence interval that represented random noise for each metabolite. This ultimately allowed for assessment of the probability that a given metabolite was excluded from the associated null distribution (Fig. 3C).

To identify the core metabolites that were most essential for *C. difficile* growth, regardless of the environment, we cross-referenced the 40 highest scoring metabolites from each treatment group (Fig. 4A). Aminoglycan N-acetylglucosamine (GlcNAc) was found to the have the highest median importance of all shared metabolites, which has been shown to be a readily available source of carbon and nitrogen which can be limiting in the gut (20). We went on to confirm that our strain of *C. difficile* could metabolize GlcNAc for growth (Fig. 4B) in *C. difficile* minimal media (31). The Stickland fermentation acceptor proline was also found to be important in all conditions tested (32). *C. difficile* is auxotrophic for not only proline, but also cysteine, leucine, isoleucine, tryptophan, and valine, which prevented testing for *in vitro* growth changes on proline despite providing for modest growth in the no carbohydrate control. Previous analysis of

C. difficile colonizing GF mice under mono-associated conditions indicated that *C. difficile* uses both sets of metabolites (20); however, use of these metabolites in the context of a complex community of potential competitors has not been observed previously. This analysis indicated that these metabolites might be an integral component of the nutrient niche for *C. difficile*.

We next focused on the highest scoring metabolites that were distinct to each infection condition (Fig. 4A). The resulting groups of metabolites all contained at least one known carbohydrate growth substrate of *C. difficile* (Fig. 4B) (6, 12, 33). These included sorbitol (streptomycin-pretreated), mannitol (cefoperazone-pretreated), and salicin (clindamycin-pretreated), which were likely introduced by the animal's diet. In the ex-GF mice, N-acetylneuraminate (Neu5Ac), which is an aminoglycan obtained from the animal's mucin layer, had a high importance score (34). The concentration of Neu5Ac was previously shown to be liberated by other bacterial species to the benefit of *C. difficile* (33). Considering that Neu5Ac had one of the higher importance scores in our ex-GF animals, this suggests that *C. difficile* does not require other bacteria to liberate it from the mucus layer. Furthermore, in ex-GF mice two forms of acetate were important which *C. difficile* is known to convert convert further into butyrate during fermentation (35). In streptomycin-pretreatment, galactitol scored highly and has been shown to be a sugar alcohol growth substrate for other *C. difficile* strains *in vivo* (23), but it did not improve *in vitro* growth in this study (Table S4). Succinyl-CoA was of distinct importance in clindamycin-pretreated mice, which is the direct precursor to succinate that is produced the gut microbiota and utilized by *C. difficile* for growth (9). Cefoperazone and clindamycin preatreatments were also associated with increased importance of ribose

analogs, which has been shown to be another carbohydrate substrate that *C. difficile* ferments for growth (36). Additionally, alanine was found to be important in this condition and is a preferred electron donor in Stickland fermentation (27). These data supported the hypothesis that *C. difficile* exploited alternative nutrient sources between the susceptible environments it colonized.

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

First, we compared the relative concentration of important metabolites in untreated SPF mice and antibiotic pretreated mice in the absence of CDI (Fig. 5). We found that the relative concentration of GlcNAc was actually significantly lower in all susceptible conditions (Fig. 5A; all $P < 0.001$). The Stickland fermentation acceptors proline (all $P < 0.05$) and hydroxyproline (all $P < 0.05$) were significantly higher in all susceptible environments tested (Fig. 5B and S7B). Conversely, the Stickland donor alanine was significantly lower across all susceptible conditions (Fig. 5D; all $P < 0.05$). Succinate was significantly higher in both streptomycin and clindamycin pretreated mice (Fig. 5E;

all $P < 0.05$). Among the cefoperazone-pretreated SPF and GF mice, we also found that mannitol/sorbitol (Fig. 5C), N-acetylneuraminate (Fig. 5F), and glycine (Fig. S6A) were significantly higher in cefoperazone-treated SPF and GF mice (all $P < 0.05$). These results supported the assertion that antibiotic treatment opened potential nutrient niches that *C. difficile* was able to exploit for its growth.

Second, we compared relative concentrations of important metabolites during CDI and mock-infection within each pretreatment group (Fig. 5). Both groups of host-derived aminoglycans, GlcNAc/GalNAc (Fig. 5A) and Neu5Ac (Fig. 5F), were significantly lower when in the presence of *C. difficile* in ex-GF mice ($P < 0.05$ and 0.01). In agreement with the previous results, we found that the Stickland acceptors proline (Fig. 5B) and hydroxyproline (Fig. S6C) were significantly lower in every *C. difficile* colonized environment (all $P < 0.05$). Glycine, another preferred Stickland acceptor, was lower in each condition following infection with significant change in cefoperazone-pretreated mice (Fig. S6D; $P < 0.05$). The Stickland donors leucine and isoleucine were significantly lower in all infected conditions except streptomycin-pretreated mice (Fig. S6A and S6B; all $P < 0.05$). Concentrations of alanine were also lower in all infected conditions compared to mock infection, however none of the changes met our threshold for significance (Fig. 5D). These results strongly supported the hypothesis that amino acids are a primary energy source of *C. difficile* during infection. A significant difference was seen for mannitol/sorbitol in ex-GF mice ($P < 0.01$), but not in cefoperazone-pretreated mice (Fig. 5C). Although a lower the concentration of succinate in both streptomycin and clindamycin pretreated mice was observed, neither was found to be significant. Overall, metabolomic analysis supported our metabolite importance

297 algorithm for predicting the metabolites utilized by *C. difficile* during different infection
298 conditions. Results from metabolic modeling combined with untargeted metabolomic
299 analysis also suggested a possible hierarchy of preferred growth substrates.

300 Discussion

301 Our results expand upon previous understanding of *C. difficile* metabolism during
302 infection by showing that not only does the pathogen adapt its metabolism to life inside
303 of a host (20, 23), but also to the context of the specific gut environment in which it finds
304 itself. Previous transcriptomic efforts to measure the response of *C. difficile* have
305 demonstrated *in vivo* changes in metabolism following colonization of GF mice. In this
306 study, we utilized a conventionally-reared mouse model of infection to compare the
307 response of *C. difficile* to colonization in the context of varied gut communities
308 generated by pretreatment with representatives from distinct classes of antibiotics. With
309 these models, we identified subtle differences in sporulation and toxin activity between
310 each antibiotic-pretreated condition. Transcriptomic sequencing of *C. difficile* across
311 colonized environments indicated complex expression patterns of genes in catabolic
312 pathways for a variety of carbon sources. Through integration of transcriptomic data
313 with genome-scale metabolic modeling, we were able to deconvolute these signals.
314 This allowed us to observe that *C. difficile* likely genenrated energy by metabolizing
315 specific alternative carbohydrates, carboxylic acids, and aminoglycans across colonized
316 conditions. We also found that Stickland fermentation substrates and products, as well
317 another host-derived amino glycan N-acetylglucosamine, were consistently among the
318 highest scoring shared metabolites which indicated that these metabolites were central
319 to the *in vivo* nutritional strategy of *C. difficile*. To confirm our modeling-based results we

320 employed untargeted mass spectrometry that demonstrated greater availability of many
321 metabolites highlighted by our algorithm in susceptible gut environs. Metabolomic
322 analysis further revealed differential reduction of important metabolites during CDI,
323 which suggested a hierarchy for the utilization of certain growth nutrients.

324 An explanation for the differences seen in metabolite importance and substrate
325 availability could be the concomitant lower population density of one or more
326 competitors for certain resources. Ex-GF mice, where no other microbial competitors
327 are present, provided a partially controlled system of resource competition. In this
328 condition, Neu5Ac was found to be the most important substrate and concentrations
329 Neu5Ac were significantly higher in susceptible mice. The concentrations of Neu5Ac
330 were concordantly lower in infected mice relative to mock-infected mice. The same
331 trend was also present in cefoperazone-pretreatment, which suggested that *C. difficile*
332 may be less competitive for this host-derived aminoglycan and may only have access
333 when certain competitors are no longer present. In the presence of a microbiota, *C.*
334 *difficile* population-level nutrient utilization patterns differed across each environment
335 tested. For example, past studies have concluded that specific PTS and ABC transport
336 systems are upregulated *in vivo* (20, 23), but our results indicate more complex
337 regulation with inverse expression of the respective systems between antibiotic
338 pretreatments (Fig. 2). In agreement with earlier research we found that *C. difficile* likely
339 fermented amino acids for energy during infection of GF mice in addition to aminoglycan
340 catabolism. Our results go on to support that this metabolic strategy was conserved
341 across all infection conditions tested. Several Stickland substrates had consistently
342 higher importance scores including alanine, leucine, and proline indeed dropped

343 concentration during infection (Table S3, Fig. 5A, and S6A). Fermentation of amino
344 acids provides not only carbon and energy, but are also a source of nitrogen which is a
345 limited resource in the mammalian lower gastrointestinal tract (37). This makes
346 Stickland fermentation a valuable metabolic strategy, and it stands to reason that *C.*
347 *difficile* would use this strategy across all environments it colonizes. This same principle
348 may also extend to host mucus layer derived aminoglycans as they are another source
349 of carbon and nitrogen which, despite augmented release by members of the microbiota
350 (38), would be present at some basal concentration regardless of other species'
351 intercession. Finally, we did find disagreement in some metabolite importance scores
352 and the difference in *in vivo* concentration of previously suggested *C. difficile* growth
353 substrates between mock infected and infected mice. This may indicate a nutrient
354 preference hierarchy during infection. Based on our results, we propose that amino
355 acids are prized above all other substrates, followed by aminoglycans, then
356 carbohydrates, sugar alcohols, or carboxylic acids depending on their availability in the
357 environment. Since the latter provide carbon and energy, but not nitrogen, it appears
358 that *C. difficile* metabolism strongly values nitrogen-containing carbon sources that fulfill
359 a larger proportion of its biological requirements but this requires additional investigation
360 to confirm.

361 Our systems approach to studying *C. difficile* behavior during the infection of suceptible
362 communities is novel because it combines multiple levels of biological data to identify
363 metabolic trends that would not be apparent by a single method. Only through
364 integrative multi-omic analysis of *C. difficile* infection employing genomics,
365 transcriptomics, and metabolomics were we able to uncover a much clearer image of *C.*

366 *difficile*'s nutrient niche space during infection in the context of complex microbial
367 communities. By virtue of our importance algorithm's reliance on network topology, the
368 signal contributed by those metabolites on the periphery of the network, which are more
369 likely to be imported from the environment, was amplified. This modeling approach may
370 also allow for the identification of emergent properties for the metabolism of *C. difficile*
371 during infection. One example could be the appearance of CO_2, an apparent metabolic
372 end product, in the list of shared important metabolites. Although this may be a
373 shortcoming of the genome annotation, one group has posited that *C. difficile* may
374 actually be autotrophic under certain conditions (39). These findings highlight that our
375 method not only identifies growth substrates, but reports all metabolites that are being
376 utilized for other processes.

377 Several factors limited our ability to generate and interpret transcriptomic and
378 metabolomic data. Most prominently, we were forced to pool the cecal contents of
379 multiple animals to generate a sufficient quantity of high quality RNA that would permit
380 us to sample the transcriptome of a rare member of the microbiome. Microarray-based
381 gene expression measurement was not a viable alternative to sequencing as the
382 amount of background orthologous transcription from other bacterial species would
383 contribute greatly to non-specific binding and bias the true *C. difficile* signal.
384 Metabolomic comparisons are also complicated by the fact that multiple organisms
385 contribute to the metabolite pool. The metabolic patterns of the other species in each
386 system (host and microbe) could instead be altered by pathogen colonization. As the
387 concentrations of metabolites in our untargeted assay were reported in relative terms, it
388 was difficult to discern whether the available biomass of *C. difficile* reaches a level to

create these differences on its own. Possible limitations of our modeling approach also exist, despite much of our results being consistent with previously published work on the metabolism of *C. difficile*. Ultimately, the metabolite importance calculation is dependent on correct and existing gene annotation. In this regard it has been shown that the pathway annotations in KEGG are robust to missing elements (40), however this does not completely eliminate the possibility for this type of error. Due to the topology of the metabolic network, we were also unable to integrate stoichiometry for each reaction which may effect rates of consumption or production. Reaction reversibility also varies depending on versions of enzymes possessed by each species. Incorrect directionality annotations may lead to mislabeling reactants or products and potentially lead to incorrect importance calculations. With additional manual curation of the *C. difficile* metabolic network, more species specific discoveries can eventually be made. Ultimately, the application of multiple methods to study the altered physiology of *C. difficile* in mock-infected and infected communities allowed us to validate our results based on known elements of *C. difficile* biology and to internally cross validate the novel results from our experiments. Ultimately, these results combine to underscore predictions of nutrient niche plasticity.

Our combined genomic, transcriptomic, and metabolomic analysis showed that when infecting diverse host-associated gut environments, *C. difficile* optimized its nutrient utilization profile to each gut environment and effectively colonize the host. These results have implications for the development of targeted measures to prevent *C. difficile* colonization through pre- or probiotic therapy. In the future, this systems-level

411 approach could be easily expanded to study the niche landscape of entire communities
412 of bacteria in response to antibiotic perturbation or pathogen colonization.

413 ## Materials and Methods

414 **Animal care and antibiotic administration.** Six-to-eight week-old GF C57BL/6 mice
415 were obtained from a single breeding colony maintained at the University of Michigan
416 and fed Laboratory Rodent Diet 5001 from LabDiet for all experiments. All animal
417 protocols were approved by the University Committee on Use and Care of Animals at
418 the University of Michigan and carried out in accordance with the approved guidelines.
419 Specified SPF animals were administered one of three antibiotics; cefoperazone,
420 streptomycin, or clindamycin (Table 1). Cefoperazone (0.5 mg/ml) and streptomycin (5.0
421 mg/ml) were administered in distilled drinking water *ad libitum* for 5 days with 2 days
422 recovery with untreated distilled drinking water prior to infection. Clindamycin (10 mg/kg)
423 was given via intraperitoneal injection 24 hours before time of infection. Adapted from a
424 previously described model (24).

425 ***C. difficile* infection and necropsy.** All *C. difficile* strain 630 spores were prepared
426 from a single large batch whose concentration was determined a week prior to
427 challenge. On the day of challenge, 1×10^3 *C. difficile* spores were administered to mice
428 via oral gavage in phosphate-buffered saline (PBS) vehicle. Subsequent quantitative
429 plating to enumerate the spores was performed to ensure correct dosage. Mock-
430 infected animals were given an oral gavage of 100 µl PBS at the same time as those
431 mice administered *C. difficile* spores. 18 hours following infection, mice were euthanized
432 by carbon dioxide asphyxiation and necropsied to obtain the cecal contents. Two 100 µl
433 aliquots were immediately flash frozen for later DNA extraction and toxin titer analysis,

434 respectively. A third 100 µl aliquot was quickly transferred to an anaerobic chamber for
435 quantification of *C. difficile* abundance. The remaining content in the ceca
436 (approximately 1 mL) was mixed with 1 mL of sterile PBS in a stainless steel mortar
437 housed in a dry ice and ethanol bath. The cecal contents of 9 mice from 3 cages was
438 pooled into the mortar. Pooling cecal contents was necessary so that there would be a
439 sufficient quantity of high quality rRNA-free RNA for deep sequencing. The pooled
440 content was then finely ground and stored at -80° C for subsequent RNA extraction.

441 ***C. difficile* cultivation and quantification.** Cecal samples were weighed and serially
442 diluted under anaerobic conditions (6% H, 20% CO_2, 74% N_2) with anaerobic PBS.
443 Differential plating was performed to quantify both *C. difficile* spores and vegetative cells
444 by plating diluted samples on CCFAE plates (fructose agar plus cycloserine (0.5%),
445 cefoxitin (0.5%), and erythromycin (0.2%)) at 37° C for 24 hours under anaerobic
446 conditions (41). It is important to note that the germination agent taurocholate was
447 omitted from these plates to quantify only vegetative cells. In parallel, undiluted samples
448 were heated at 60° C for 30 minutes to eliminate vegetative cells and leave only spores
449 (42). These samples were serially diluted under anaerobic conditions in anaerobic PBS
450 and plated on CCFAE with taurocholate (10%) at 37° C for 24 hours. Plating was
451 simultaneously done for heated samples on CCFAE to ensure all vegetative cells had
452 been eliminated.

453 ***C. difficile* toxin titer assay.** To quantify the titer of toxin in the cecum a Vero cell
454 rounding assay was performed as in (43). Briefly, filtered-sterilized cecal content was
455 serially diluted in PBS and added to Vero cells in a 96-well plate. Plates were viewed
456 after 24 hour incubation for cell rounding. A more detailed protocol with product

information can be found at:
https://github.com/SchlossLab/Jenior_Modeling_mBio_2017/blob/master/protocols/toxin_assay/Verocell_ToxinActivity_Assay.Rmd

16S rRNA gene sequencing. DNA was extracted from approximately 50 mg of cecal content from each mouse using the PowerSoil-htp 96 Well Soil DNA isolation kit (MO BIO Laboratories) and an epMotion 5075 automated pipetting system (Eppendorf). The V4 region of the bacterial 16S rRNA gene was amplified using custom barcoded primers and sequenced as described previously using an Illumina MiSeq sequencer (44). 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 (44). 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) (45). Curated sequences were clustered into operational taxonomic units (OTUs) using a 97% similarity cutoff with the average neighbor clustering algorithm. The number of sequences in each sample was rarefied to 2,500 per sample to minimize the effects of uneven sampling.

RNA extraction, shotgun library preparation, and sequencing. Pooled, flash-frozen samples were ground with a sterile pestle to a fine powder and scraped into a sterile 50 ml polypropylene conical tube. Samples were stored at -80° C until the time of extraction. Immediately before RNA extraction, 3 ml of lysis buffer (2% SDS, 16 mM EDTA and 200 mM NaCl) contained in a 50 ml polypropylene conical tube was first

heated for 5 minutes in a boiling water bath (46). The hot lysis buffer was added to the frozen and ground cecal content. The mixture was boiled with periodic vortexing for another 5 minutes. After boiling, an equal volume of 37° C acid phenol/chloroform was added to the cecal content lysate and incubated at 37° C for 10 minutes with periodic vortexing. The mixture was the centrifuged at 2,500 x g at 4° C for 15 minutes. The aqueous phase was then transferred to a sterile tube and an equal volume of acid phenol/chloroform was added. This mixture was vortexed and centrifuged at 2,500 x g at 4° for 5 minutes. The process was repeated until aqueous phase was clear. The last extraction was performed with chloroform/isoamyl alcohol to remove the acid phenol. An equal volume of isopropanol was added and the extracted nucleic acid was incubated overnight at -20° C. The following day the sample was centrifuged at 12000 x g at 4° C for 45 minutes. The pellet was washed with 0° C 100% ethanol and resuspended in 200 µl of RNase-free water. Samples were then treated with 2 µl of Turbo DNase for 30 minutes at 37° C. RNA samples were retrieved using the Zymo Quick-RNA MiniPrep. Completion of the DNase reaction was assessed using PCR for the V4 region of the 16S rRNA gene for 30 cycles (Kozich, 2013). Quality and integrity of RNA was measured using the Agilent RNA 6000 Nano kit for total prokaryotic RNA. The Ribo-Zero Gold rRNA Removal Kit Epidemiology was then used to deplete 16S and 18S rRNA from the samples. Prior to library construction, quality and integrity as measured again using the Agilent RNA 6000 Pico Kit. Stranded RNA-Seq libraries were made constructed with the TruSeq Total RNA Library Preparation Kit v2. The Agilent DNA High Sensitivity Kit was used to measure concentration and fragment size distribution before sequencing. High-throughput sequencing was performed by the University of

Michigan Sequencing Core in Ann Arbor, MI. For all groups, sequencing was repeated across 4 lanes of an Illumina HiSeq 2500 using the 2x50 bp chemistry.

Sequence curation, read mapping, and normalization. Read curation was performed in a two step process. First, residual 5' and 3' Illumina adapter sequences were trimmed using CutAdapt (47) on a per library basis. Reads were then quality trimmed using Sickle (Joshi, 2011) on the default settings. An average of ~261,000,000 total reads (both paired and orphaned) remained after quality trimming. Mapping was accomplished using Bowtie2 (48) and the default stringent settings. An average of ~6,880,000 reads in sample each mapped to the annotated nucleotide gene sequences of *Clostridioides difficile* 630 from the KEGG: Kyoto Encyclopedia of Genes and Genomes (49). Optical and PCR duplicates were then removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard/), leaving an average of ~167,000 reads per sample for final analysis (Table S5). The remaining mappings were converted to idxstats format using Samtools (50) 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 subsampled to the lowest sequence total across the libraries.

Reaction Annotation & Bipartite Network Construction. The metabolism of *C. difficile* stain 630 was represented as a directed bipartite graph with both enzymes and metabolites as nodes. Briefly, models were semi-automatically constructed using KEGG (2016 edition) ortholog (KO) gene annotations to which transcripts had been mapped. Reactions that each KEGG ortholog mediate were extracted from ko_reaction.list

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

Metabolite Importance Calculation. The substrate importance algorithm (Fig. 3a) favors metabolites that are more likely acquired from the environment (not produced within the network), and will award them a higher score (Fig. 4b & 6c). The presumption of our approach was that enzymes that were more highly transcribed were more likely to utilize the substrates they act on due to coupled bacterial transcription and translation. If a compound was more likely to be produced, the more negative the resulting score would be. To calculate the importance of a given metabolite (m), we used rarefied

549 transcript abundances mapped to respective enzyme nodes. This was represented by t_o
550 and t_i to designate if an enzyme created or utilized m. The first step was to calculate the
551 average expression of enzymes for reactions that either created a given metabolite (i) or
552 consumed that metabolite (ii). For each direction, the sum of transcripts for enzymes
553 connecting to a metabolite were divided by the number of contributing edges (e_o or e_i) to
554 normalize for highly connected metabolite nodes. Next the raw metabolite importance
555 score was calculated by subtracting the creation value from the consumption value to
556 weight for metabolites that are likely acquired exogenously. The difference was $\log_2$
557 transformed for comparability between scores of individual metabolites. This resulted in
558 a final value that reflected the likelihood a metabolite was acquired from the
559 environment. Untransformed scores that already equaled to 0 were ignored and
560 negative values were accounted for by transformation of the absolute value then
561 multiplied by -1. These methods have been written into a single python workflow, along
562 with supporting reference files, and is presented as bigSMALL v1.0 (BacterIal Genome-
563 Scale Metabolic models for AppLied reverse ecoLogy) available in a public Github
564 repository at https://github.com/mjenior/bigsmall.

565 **Transcriptome Randomization and Probability Distribution Comparison.** As
566 sequencing replicates of *in vivo* transcriptomes was not feasible, we applied a Monte
567 Carlo style simulation to distinguish calculated metabolite importances due to distinct
568 transcriptional patterns for the environment measured from those metabolites that were
569 constitutively important. We employed a 10,000-fold bootstrapping approach of
570 randomly reassigning transcript abundance for enzyme nodes and recalculating
571 metabolite importances. This approach was chosen over fitting a simulated

572 transcriptome to a negative binomial distribution because it created a more relevant
573 standard of comparison for lower coverage sequencing efforts. Using this method, each
574 substrate node accumulated a random probability distribution of importance scores
575 which were then used to calculate the median and confidence interval to generate a
576 probability for each metabolite importance score to be the result of more than chance.
577 This was a superior approach to switch randomization since the connections of the
578 network itself was created through natural selection and any large-scale alterations
579 would yield biologically uninformative comparisons (51).

580 **Anaerobic *in vitro C. difficile* growth curves.** The carbon-free variation of *C. difficile*
581 Basal Defined Medium (NCMM) was prepared as previously described (6). Individual
582 carbohydrate sources were added at a final concentration of 5 mg/mL and pair-wise
583 carbohydrate combinations were added at 2.5 mg/mL each (5 mg/mL total). A solution
584 of the required amino acids was made separately and added when noted at identical
585 concentrations to the same study. 245 µl of final media mixes were added to a 96-well
586 sterile clear-bottom plate. A rich media growth control was also included, consisting of
587 liquid Brain-Heart Infusion with 0.5% cysteine. All culturing and growth measurement
588 were performed anaerobically in a Coy Type B Vinyl Anaerobic Chamber (3.0% H, 5.0%
589 CO_2, 92.0% N, 0.0% O_2). *C. difficile* str. 630 was grown for 14 hours at 37° C in 3 mL
590 BHI with 0.5% cysteine. Cultures were then centrifuged at 2000 rpm for 5 minutes and
591 resulting pellets were washed twice with sterile, anaerobic phosphate-buffered saline
592 (PBS). Washed pellets were resuspended in 3 mL more PBS and 5 µl of prepped
593 culture was added the each growth well of the plate containing media. The plate was
594 then placed in a Tecan Sunrise plate reader. Plates were incubated for 24 hours at 37°

595 C with automatic optical density readings at 600 nm taken every 30 minutes. OD_{600}
596 values were normalized to readings from wells containing sterile media of the same type
597 at equal time of incubation. Growth rates and other curve metrics were determined by
598 differentiation analysis of the measured OD_{600} over time in R to obtain the slope at each
599 time point.

600 **Quantification of *in vivo* metabolite relative concentrations.** Metabolomic analysis
601 performed by Metabolon (Durham, NC), a brief description of their methods is as
602 follows. All methods utilized a Waters ACQUITY ultra-performance liquid
603 chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate
604 mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source
605 and Orbitrap mass analyzer at 35,000 mass resolution. Samples were dried then
606 reconstituted in solvents compatible to each of the four methods. The first, in acidic
607 positive conditions using a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm)
608 using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1%
609 formic acid (FA). The second method was identical to the first but was
610 chromatographically optimized for more hydrophobic compounds. The third approach
611 utilized a basic negative ion optimized conditions using a separate dedicated C18
612 column. Basic extracts were gradient eluted from the column using methanol and water,
613 however with 6.5mM Ammonium Bicarbonate at pH 8. Samples were then analyzed via
614 negative ionization following elution from a hydrophilic interaction chromatography
615 column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of
616 water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis
617 alternated between MS and data-dependent MS n scans using dynamic exclusion. The

scan range varied slighted between methods but covered 70-1000 m/z. Library matches for each compound were checked for each sample and corrected if necessary. Peaks were quantified using area under the curve.

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

Funding Information

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

Acknowledgements

The authors would like to acknowledge Charles Koumpouras for assistance with DNA extractions and metabolomic sample preparation. We would also like to acknowledge members of the University of Michigan Germfree Mouse Center, University of Michigan

641 Sequencing Core, and Metabolon for their assistance in experimental design, execution,
642 and data collection. Pooled and quality trimmed transcriptomic read data and
643 experiment metadata are available through the NCBI Sequence Read Archive (SRA;
644 PRJNA354635). Data processing steps for beginning from raw sequence data to the
645 final manuscript are hosted at
646 http://www.github.com/SchlossLab/Jenior_Modeling_mBio_2017. The authors would
647 additionally like to thank Geoffrey Hannigan Ph.D, Kaitlin Flynn Ph.D, and Nielsen
648 Baxter Ph.D. for their suggestions on manuscript drafts.

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

652 **Department of Internal Medicine/Infectious Diseases Division, University of**
653 **Michigan Medical Center, Ann Arbor, Michigan. Department of Microbiology and**
654 **Immunology, University of Michigan, Ann Arbor, Michigan.** Vincent B. Young M.D.
655 Ph.D.

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

661 **Corresponding author** Correspondence to Patrick D. Schloss

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

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

663

Figure Legends

Figure 1 | Gut environment context affects *C. difficile* sporulation and toxin activity. Quantification of spore cfu and toxin titer from cecal content of infected mice (n = 9 per group). **(A)** Vegetative *C. difficile* cfu per gram of cecal content (P = n.s.). **(B)** *C. difficile* spore cfu per gram of cecal content. **(C)** Toxin titer from cecal content measured by activity in Vero cell rounding assay. Dotted lines denote limits of detection (LOD). Values for undetectable points were imputed as half the LOD for calculation of significant differences. Significance ($P < 0.05$), denoted by single asterisk, was determined with Wilcoxon signed-rank test with Benjamini–Hochberg correction.

Figure 2 | *C. difficile* alters expression metabolic pathways between antibiotic pretreatment models. Each point in the ternary plot represents a unique gene from the annotated genome of *C. difficile* str. 630. Position reflects the ratio of median rarefied transcript abundance for that gene between the three colonized antibiotic pretreatment models. Genes from specific metabolic pathways of interest are labeled and transcription from all other genes are shown in gray. **(A)** Size of highlighted points is relative to the largest transcript abundance among the antibiotic pretreatments for each gene. Categories of metabolism are displayed separately in **(B-I)**. Genes, annotations, and normalized transcript abundances can be found in Table S1. Refer to Fig. S4 for additional figure interpretation.

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

687 Importance algorithm components: (I) average transcription of reactions consuming a
688 metabolite, (II) average transcription of reactions producing a metabolite, and (III)
689 difference of consumption and production. **(B)** The expanded window displays a partial
690 example of D-fructose importance calculation. Values in the red nodes represent
691 normalized transcript reads mapping to enzymes. **(C)** Example 10000-fold Mont-Carlo
692 simulation results corresponding to a significant importance score for **m**.

693 **Figure 4 | Metabolic network analysis reveals differential carbon source utilization**
694 **by *C. difficile* across infections.** Reported metabolites were calculated to have <2.5%
695 probability to be included in the associated random score distribution. All comparison.
696 All comparisons between groups was performed using the 40 highest scoring
697 metabolites from each condition and to further focus the analysis we reported only the
698 top five scored metabolites from any group listed.**(A)** Shared importance represents the
699 median score of metabolites that were consistently important among all infected
700 conditions. Below the conserved patterns, distinctly important metabolites for each
701 group are shown. **(B)** 18 hour *C. difficile* str. 630 *in vitro* growth validating substrates
702 from network analysis. All statistical comparison was performed relative to no
703 carbohydrate control (all $P < 0.001$). Significance was determined with one-way ANOVA
704 with Benjamini–Hochberg correction.

705 **Figure 5 | Untargeted *in vivo* metabolomics support network-based metabolite**
706 **importance scores and suggest nutrient preference hierarchy.** Paired metabolites
707 were quantified simultaneously as the only differ by chirality making differentiation
708 impossible. Black asterisks inside the panels represent significant differences between
709 mock and *C. difficile*-infected groups within separate treatment groups (all $P < 0.05$).

Gray asterisks along the top margin of each panel indicate significant difference from untreated SPF mice (all $P < 0.05$). Significance was determined with Wilcoxon signed-rank test with Benjamini–Hochberg correction.

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

Supplementary Figure 2 | Analysis of bacterial community structure resulting from antibiotic treatment. Results from 16S rRNA gene amplicon sequencing from bacterial communities of cecal content in both mock-infected and *C. difficile* 630-infected animals 18 hours post-infection across pretreatment models. **(A)** Non-metric multidimensional scaling (NMDS) ordination based on Theta_{YC} distances for the gut microbiome of all SPF mice used in these experiments (n = 36). All treatment groups are significantly different from each other groups by AMOVA ($P < 0.001$). **(B)** Inverse Simpson diversity for each cecal community from the mice in (A). Cecal communities from mice not treated with any antibiotics are significantly more diverse than any antibiotic-pretreated condition ($P < 0.001$). **(C)** Representation of 16S amplicon reads

733 contributed by *C. difficile* in each sequenced condition compared to the total bacterial
734 community. The percents listed at the top of each group is the proportion of the total
735 community represented by *C. difficile*. Significantly less were for *C. difficile* were
736 detected in each condition ($P < 0.001$).

737 **Supplementary Figure 3 | Select *C. difficile* gene set expression compared**
738 **between treatment group.** Relative abundances of *C. difficile* transcript for specific
739 genes of interest. **(A)** Transcription for select genes from the *C. difficile* sporulation
740 pathway with the greatest variation in expression between the conditions tested. **(B)**
741 Relative abundances of transcript for genes that encode effector proteins from the *C.*
742 *difficile* pathogenicity locus. **(C)** Transcript abundances for genes associated with
743 quorum sensing in *C. difficile*. **(D)** Transcript relative abundance of select sigma factors
744 which expression or activity is influenced by environmental metabolite concentrations.
745 Asterisks (*) indicate genes from which transcript was undetectable.

746 **Supplementary Figure 4 | Additional explanation for Figure 2 interpretation.**
747 Relative abundance of transcription for *C. difficile* 630 genes during infection across the
748 3 antibiotic pretreatment models used during this study. Points that are located closer to
749 a corner are more highly transcribed in the condition associated with that corner
750 compared to the others. As this shows a 3-dimensional data set in 2 dimensions, there
751 is an amount of distortion proximal to each corner. Simply put for points that are nearer
752 to an edge, a greater percentage of their total transcription was contributed by *C.*
753 *difficile* colonizing those mice. **(A)** This point represents the transcription for a gene that
754 is overrepresented in cefoperazone-pretreated mice. **(B)** This point represents a gene in
755 which transcripts are equally detectable in all 3 conditions. **(C)** Transcripts for this gene

are only underrepresented in only cefoperazone-pretreated mice, and are equally detectable in clindamycin and streptomycin-pretreated animals.

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

Supplementary Figure 6 | Change in *in vivo* concentrations of additional Stickland fermentation substrates. Comparison of concentrations for other Stickland fermentation substrates from *C. difficile*-infected and mock-infected mouse cecal content 18 hours post-infection. Labels in the top left corner of each panel indicate whether the amino acid is a Stickland donor or acceptor. Black asterisks inside the panels denote significant differences between mock and *C. difficile*-infected groups within separate treatment groups (all $P < 0.05$). Gray asterisks along the top margin of each panel indicate significant difference from untreated SPF mice (all $P < 0.05$).

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

778 **Supplementary Table 2 | Topology metrics for enzyme and metabolite nodes in**

779 **the *C. difficile* str. 630 metabolic network.**

780 **Supplementary Table 3 | All metabolites with significant importance scores in**

781 **each colonized condition.**

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

783 **Supplementary Table 5 | Transcriptomic read counts at each step of curation.**

References

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

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

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

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

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

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

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

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

810 9. **Ferreyra, J. A.**, **K. J. Wu**, **A. J. Hryckowian**, **D. M. Bouley**, **B. C. Weimer**, and **J. L.**
811 **Sonnenburg**. 2014. Gut microbiota-produced succinate promotes *Clostridium difficile*
812 infection after antibiotic treatment or motility disturbance. Cell Host and Microbe
813 **16**:770–777.

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

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

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

823 13. **Sebaihia, M.**, **B. W. Wren**, **P. Mullany**, **N. F. Fairweather**, **N. Minton**, **R. Stabler**,
824 **N. R. Thomson**, **A. P. Roberts**, **A. M. Cerdeño-Tárraga**, **H. Wang**, **M. T. G. Holden**,
825 **A. Wright**, **C. Churcher**, **M. a Quail**, **S. Baker**, **N. Bason**, **K. Brooks**, **T.**

826 **Chillingworth**, **A. Cronin**, **P. Davis**, **L. Dowd**, **A. Fraser**, **T. Feltwell**, **Z. Hance**, **S.**
827 **Holroyd**, **K. Jagels**, **S. Moule**, **K. Mungall**, **C. Price**, **E. Rabbinowitsch**, **S. Sharp**, **M.**
828 **Simmonds**, **K. Stevens**, **L. Unwin**, **S. Whithead**, **B. Dupuy**, **G. Dougan**, **B. Barrell**,
829 and **J. Parkhill**. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has
830 a highly mobile, mosaic genome. Nature genetics **38**:779–786.

831 14. **Tracy, B. P.**, **S. W. Jones**, **A. G. Fast**, **D. C. Indurthi**, and **E. T. Papoutsakis**.
832 2012. Clostridia: The importance of their exceptional substrate and metabolite diversity
833 for biofuel and biorefinery applications. Current Opinion in Biotechnology **23**:364–381.

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

836 16. **Gripp, E.**, **D. Hlahla**, **X. Didelot**, **F. Kops**, **S. Maurischat**, **K. Tedin**, **T. Alter**, **L.**
837 **Ellerbroek**, **K. Schreiber**, **D. Schomburg**, **T. Janssen**, **P. Bartholomäus**, **D.**
838 **Hofreuter**, **S. Woltemate**, **M. Uhr**, **B. Brenneke**, **P. Grüning**, **G. Gerlach**, **L. Wieler**, **S.**
839 **Suerbaum**, and **C. Josenhans**. 2011. Closely related *Campylobacter jejuni* strains from
840 different sources reveal a generalist rather than a specialist lifestyle. BMC Genomics
841 **12**:584.

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

845 18. **Nawrocki, K. L.**, **A. N. Edwards**, **N. Daou**, **L. Bouillaut**, and **S. M. McBride**. 2016.
846 CodY-dependent regulation of sporulation in *Clostridium difficile*. Journal of Bacteriology
847 **198**:2113–2130.

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

20. **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.

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

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

23. **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**.

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

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

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

873 27. **Jackson, S.**, **M. Calos**, **A. Myers**, and **W. T. Self**. 2006. Analysis of proline
874 reduction in the nosocomial pathogen *Clostridium difficile*. Journal of Bacteriology
875 **188**:8487–8495.

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

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

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

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

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

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

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

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

898 36. **Nakamura, S.**, **S. Nakashio**, **K. Yamakawa**, **N. Tanabe**, and **S. Nishida**. 1982.
899 Carbohydrate Fermentation by *Clostridium difficile*. Microbiology and Immunology
900 **26**:107–111.

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

903 38. **Marcobal, A.**, **A. M. Southwick**, **K. A. Earle**, and **J. L. Sonnenburg**. 2013. A
904 refined palate: Bacterial consumption of host glycans in the gut. Glycobiology **23**:1038–
905 1046.

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

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

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

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

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

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

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

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

930 47. **Martin, M. J.**, **S. Clare**, **D. Goulding**, **A. Faulds-Pain**, **L. Barquist**, **H. P. Browne**,
931 **L. Pettit**, **G. Dougan**, **T. D. Lawley**, and **B. W. Wren**. 2013. The *agr* locus regulates

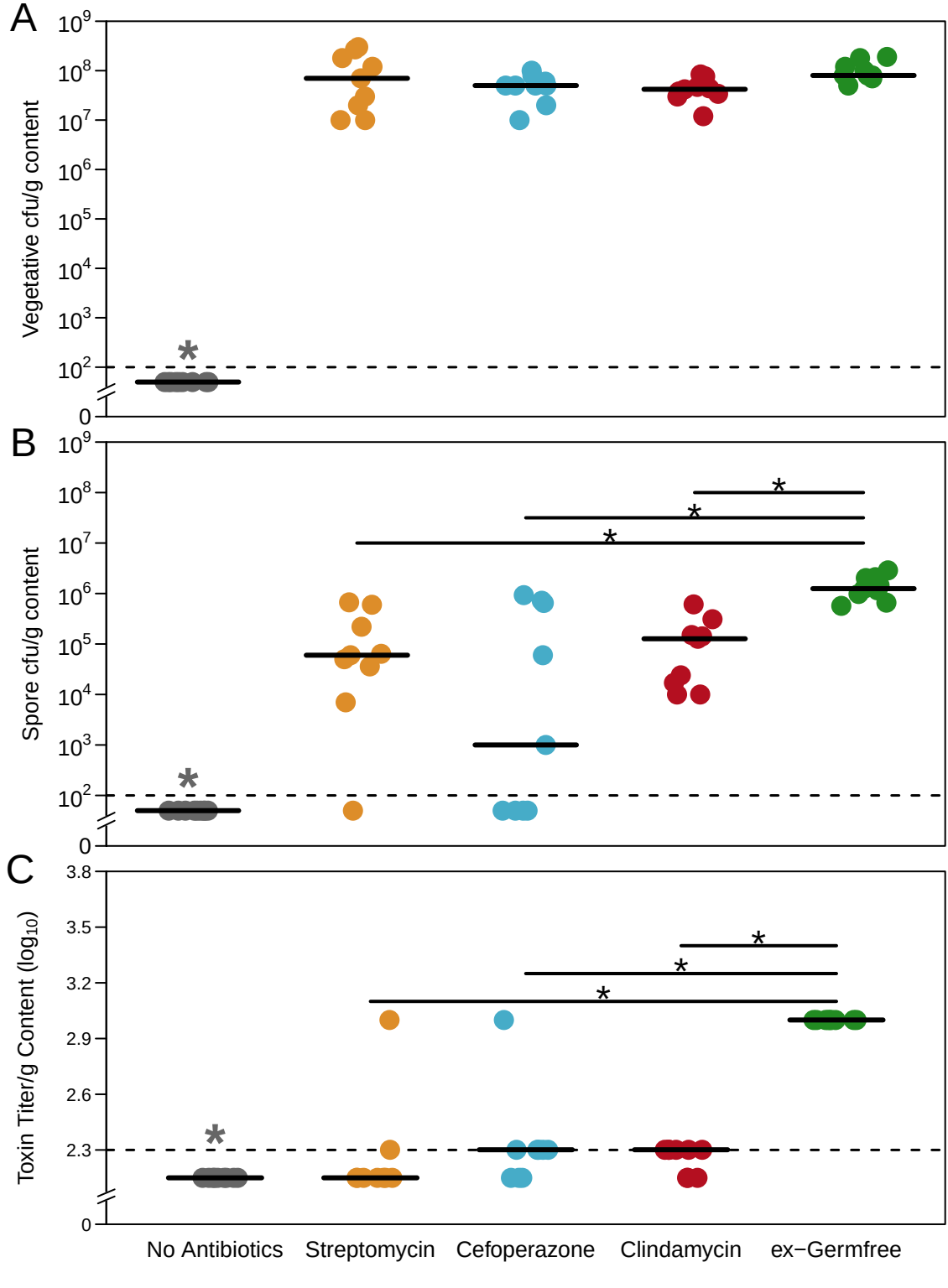
virulence and colonization genes in *Clostridium difficile* 027. Journal of Bacteriology **195**:3672–3681.

48. **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.

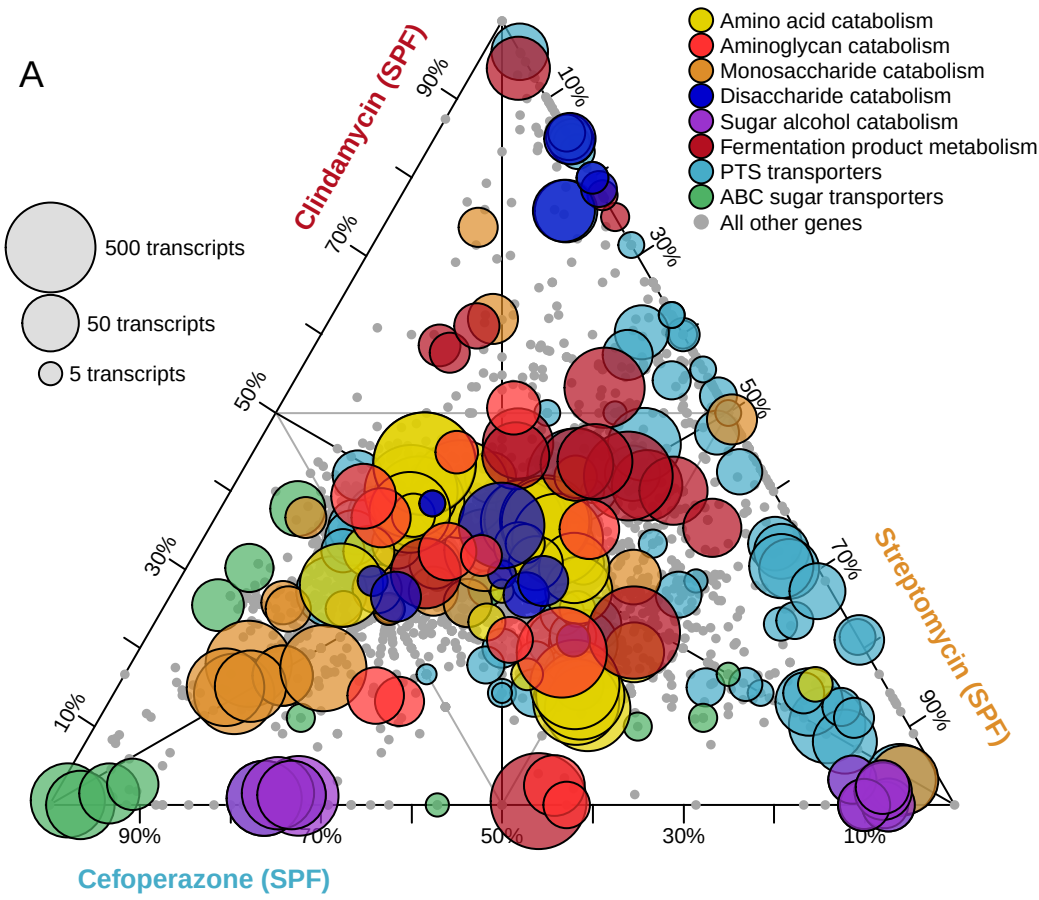
49. **Ogata, H.**, **S. Goto**, **K. Sato**, **W. Fujibuchi**, **H. Bono**, and **M. Kanehisa**. 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research **27**:29–34.

50. **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.

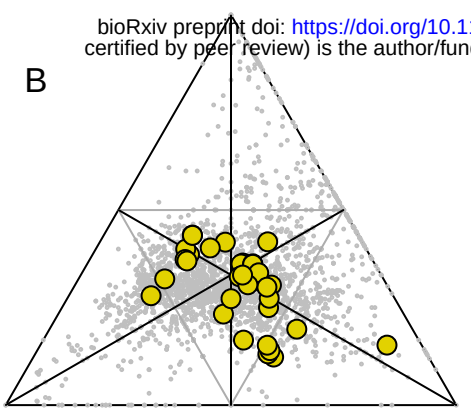
51. **Basler, G.**, **O. Ebenhöh**, **J. Selbig**, and **Z. Nikoloski**. 2011. Mass-balanced randomization of metabolic networks. Bioinformatics **27**:1397–1403.


A
Vegetative cfu/g content
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
0
*
B
Spore cfu/g content
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
0
*
*
*
*
C
Toxin Titer/g Content (log_{10})
3.8
3.5
3.2
2.9
2.6
2.3
0
*
*
*
*
No Antibiotics
Streptomycin
Cefoperazone
Clindamycin
ex−Germfree

A

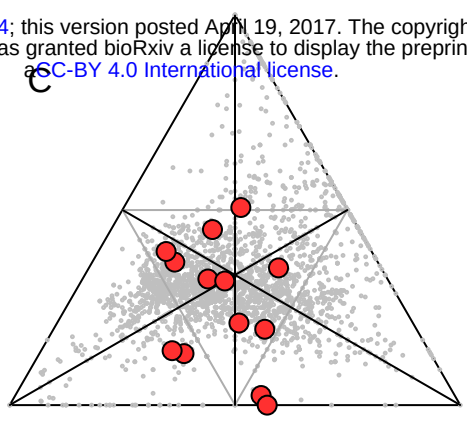


B



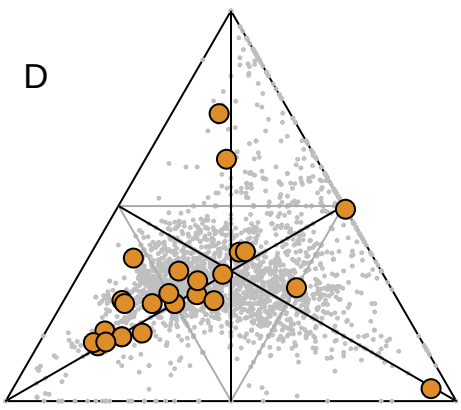
Amino acid catabolism

C



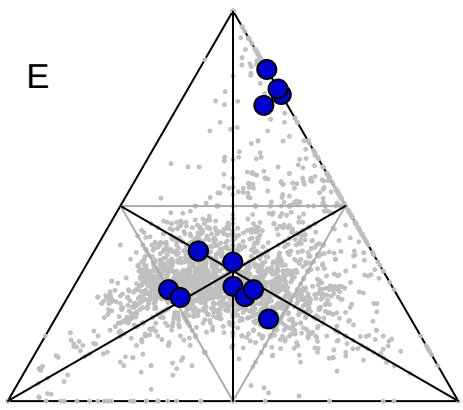
Aminoglycan catabolism

D



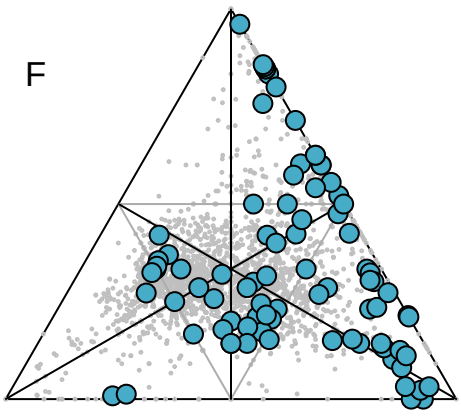
Monosaccharide catabolism

E



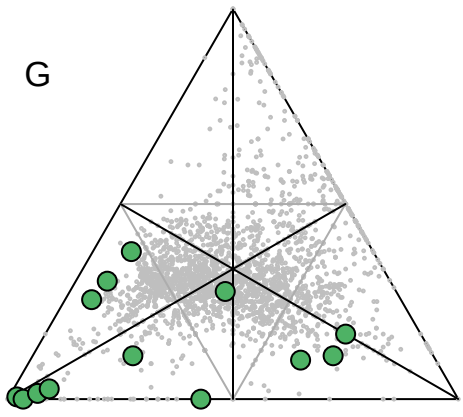
Disaccharide catabolism

F



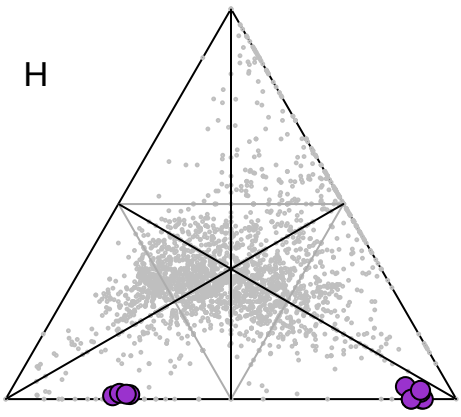
PTS transporters

G



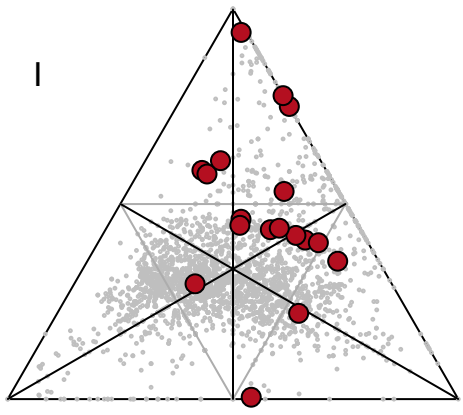
ABC sugar transporters

H



Sugar alcohol catabolism

I



Fermentation product metabolism