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

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

Abstract

Clostridium difficile infection (CDI) has grown to be the most prevalent cause of hospital acquired infection in the United States. Susceptibility to CDI is induced by recent antibiotic exposure, which is known to alter the structure of the gut microbiome and to affect the availability of growth nutrients in the gut. We hypothesized that *C. difficile* is a generalist that adapts its physiology to the nutrients available within the gut. We orally challenged C57BL/6 mice that previously received one of three antibiotics with *C. difficile* and demonstrated that it was able to colonize the cecum within 18 hours of infection. However, levels of both spore and toxin production, which are known to be affected by nutrient availability, varied between each antibiotic treatment group. To more closely investigate the specific responses of *C. difficile* as it colonized the cecum, we performed *in vivo* transcriptional analysis of *C. difficile* from cecal content of infected mice. This approach revealed variation in expression of genes that drive life-cycle switches as well as metabolic pathways associated with catabolizing a variety of carbon sources such as carbohydrates, amino acids, and amino sugars. To assess which substrates *C. difficile* was most likely exploiting in each antibiotic-perturbed microbiome,

we developed a novel metabolite scoring algorithm within the genome-scale bipartite metabolic network of *C. difficile* that incorporated both network topology and transcript abundance to infer the likelihood that a given metabolite was acquired from the environment. Applying this approach, we found that *C. difficile* indeed occupies alternative nutrient niches across each antibiotic-perturbed microbiome and that the highlighted metabolites support significant growth, *in vitro*. Results from this analysis support the hypothesis that consumption of N-acetyl-D-glucosamine and Stickland fermentation substrates are central components of *C. difficile*'s metabolic strategy and pathogenesis. This work has implications for elucidating specifics of the nutrient niche of *C. difficile* during infection and may lead to the discovery of targeted measures to prevent *C. difficile* colonization including potential pre- or probiotic therapies.

Introduction

Infection by the Gram-positive, spore-forming bacterium *Clostridium difficile* has increased in both prevalence and severity across numerous countries during the last decade[1]. In the United States, *C. difficile* was estimated to have caused >500,000 infections and resulted in ~$4.8 billion worth of acute care costs in 2014[2]. *C. difficile* infection (CDI) causes an array of toxin-mediated symptoms ranging from abdominal pain and diarrhea to the more life-threatening conditions pseudomembraneous colitis and toxin megacolon. Prior treatment with antibiotics is the most common risk factor associated with susceptibility to CDI[3]. It has been shown that antibiotic therapy alters the structure and function of the gut microbiota making it susceptible to colonization by *C. difficile*[4]. This is referred to as colonization resistance, in which the gut microbiota inhibits the persistence or growth of a number of pathogenic bacteria. Colonization

resistance can be achieved by multiple mechanisms including competition for physical space or growth nutrients[5].

Mouse models have been an effective tool for studying the mechanisms of colonization resistance. Use of distinct antibiotic classes to vary the structure of the microbiota which have been shown to result in susceptibility *C. difficile* colonization[6–8]. In this case, the antibiotics chosen significantly impact the structure and diversity of the cecal microbiome (Fig. S1a & S1b). It has been further demonstrated that at 18 hours after being introduced to a cefoperazone treated mouse, *C. difficile* reached its maximum vegetative cell density in the cecum[9]. This provided a single timepoint to measure the largest population of metabolically active *C. difficile*. Building upon these results, others have shown that many of these antibiotic classes also alter the gut metabolome, increasing the concentrations of known *C. difficile* growth substrates[7,10–12]. Taken together these results are a strong indication that the healthy gut microbiota inhibits the growth of *C. difficile* through limitation of substrates it needs to grow. The ability of an intact gut community to exclude *C. difficile* colonization is suggestive of the nutrient-niche hypothesis in which an organism must be able to utilize a subset of available resources better than all competitors to colonize the intestine[13,14].

Based on its genome sequence and *in vitro* growth characteristics, *C. difficile* appears able to fill multiple nutrient niches. *C. difficile* has a relatively large and mosaic genome, it is amenable to a variety of growth substrates, and is able to colonize a diverse array of hosts suggesting that that it is a bacterial generalist[15–17]. The ability to metabolize a variety of substrates is important since these substrates affect the regulation of genes involved in *C. difficile*'s pathogenesis. For example, *in vitro* transcriptomic analysis

suggests that high concentrations of easily metabolized carbon sources, such as glucose or amino acids, inhibit toxin gene expression and sporulation[18,19]. These genes are regulated by DNA-binding sigma factors, such as the pleiotropic regulator *ccpA*, which are under the control of environmental nutrient concentrations, especially carbohydrates[20,21]. Downstream effects of this regulation likely have enormous impact on the lifestyle and metabolic strategy of *C. difficile* when colonizing across sensitive hosts.

Previous transcriptomic studies of *C. difficile* have mainly focused on transcription of virulence factors, *in vitro*[22,23], with some work characterizing transcription during colonization of germfree mice[24,25]. More relevant to nutrient acquisition, *C. difficile* up-regulated several phosphotransferase systems (PTS) and ABC transporters in germfree mice, alluding to metabolic adaptation to nutrient availability *in vivo*[25]. Although these analyses are informative, they are either primarily directed toward the expression of virulence factors or lack the context of the gut microbiota which *C. difficile* must compete against for substrates. Metabolomic analyses have also been used to more directly assay changes in bacterial metabolism as they relate to CDI[7,12]; however, these methods cannot focus on *C. difficile*-specific metabolites and more closely resemble echoes of metabolism, not currently active processes. In contrast to these approaches, *in vivo C. difficile* transcriptomic analysis from specific pathogen free (SPF) animals may provide unique insight into its active metabolic pathways in a more realistic model of infection. Integrating transcriptomic data with genome-scale metabolic modeling has previously aided in identifying the most active aspects of an organism's metabolism and

which substrates are preferred by the organism[26–28]. Applying these methods to study *C. difficile* colonization would allow us to directly test the nutrient-niche hypothesis.

Founded on the ability of *C. difficile* to grow on a diverse array of carbon sources and its ability to colonize a variety of communities, we hypothesized that it focuses its metabolism to fit the context of the community it is attempting colonize. To test this hypothesis, we employed a mouse model of infection to compare the response of *C. difficile* to the gut environment caused by different classes of antibiotics. The antibiotics used in this study included streptomycin (Fig. 1a), cefoperazone (Fig. 1b), and clindamycin (Fig. 1c). These antibiotics differentially affect the structure of the gut microbiota[8]. Each has also been shown to alter the gut metabolome relative to untreated animals[7,10,12]. As such, we predicted that *C. difficile* would encounter a unique subset of nutrients and competitors in each environment, which would necessitate distinct adaptive responses. To determine whether *C. difficile* is a generalist and differentially responds to each condition, we assayed for differences in the amount of sporulation and toxin activity phenotypes and used metabolic models built using *C. difficile* expression data. In each of the three antibiotic conditions we challenged with *C. difficile*, as well as in monoassociated germfree mice, we observed that *C. difficile* adapted its nutrient utilization profile to colonize to high levels and express its virulence factors.

Results

Insert Table 1 here

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

Due to the connection between metabolism, sporulation, and toxin production in *C. difficile*, we measured sporulation and toxin production at 18 hours post infection in each group. There was not a significant difference in the number of vegetative cells between any susceptible condition tested (Fig. 2a). All antibiotic treated (Table 1) specific pathogen free (SPF) and germfree (GF) animals were colonized to ~1×10^8 colony forming units (c.f.u.) per gram of content, while untreated SPF mice maintained colonization resistance to *C. difficile*. Despite having the same number of vegetative *C. difficile* cells, large differences were detected in the density of *C. difficile* spores. Significantly more spores (P = 0.005, 0.008, 0.003) were detected in ex-GF mice than in the antibiotic treated mice (Fig. 2b). The spore densities in both streptomycin and clindamycin-treated mice were also generally higher than that in cefoperazone-treated mice. There was significantly more toxin activity in ex-GF animals than any other colonized group (all P <= 0.001), but toxin titer also varied between antibiotic treatment groups (Fig. 2c). Although similar toxin activity was found in both the cefoperazone and clindamycin-treated groups, toxin titer was below the limit of detection in most streptomycin-treated animals. These results indicate that *C. difficile* was able to colonize different communities to a consistently high level, but that the density of spores and toxin titer varied by treatment.

C. difficile adapts the expression of genes for virulence and key sigma factors that are under the control of environmental nutrient concentrations. To more closely investigate the responses of *C. difficile* to colonizing distinct susceptible gut environments, we performed whole transcriptome analysis of *C. difficile* during infection of the antibiotic treatment models. We then narrowed our analysis to focus on genes

that control or code for functions that have been linked to nutrient concentrations in the intestines during CDI. After observing differences in spore load, we first examined transcription of the most highly expressed genes in the *C. difficile* sporulation pathway[29–32] (Fig. 3a). Across the four conditions where *C. difficile* colonized, we observed transcriptional profiles consistent with observed spore levels (Fig. 2b). The mice treated with cefoperazone had the lowest spore density and had the highest level of expression for the anti-sigma factors *spoVG* and *spoVS*. The produts of these genes are involved in suppressing expression of genes found later in the sporulation pathway[33]. Streptomycin-treated mice had the next highest density of spores and the highest expression of genes associated with sporulation activation (*spoIIAB*/*spoIIE*), but they also had relatively high levels of expression of *sspA* and *sspB*, which are genes that code for effectors that protect DNA from damage during dormancy. Next, in mice treated with clindamycin, *C. difficile* expressed genes associated with late stages of sporulation, including those for spore coat components (*cdeC*, *cotD*, and *cotJB2*), spore formation (*spoIVA*, *spoVB*, and *spoVFB*), and *sspA* and *sspB*. Finally, GF mice harbored the highest density of spores and those *C. difficile* primarily expressed the dormancy genes linked with the latest stages sporulation. Together these data demonstrate that *C. difficile* differentially expressed genes associated with sporulation that corresponded to the presence of spores in the cecum.

Expression of genes for quorum sensing and pathogenicity have been linked to changes in the nutrients that can be found in the environment of *C. difficile*. Both the *agr* locus and *luxS* gene are thought to be associated with inducing the expression of *C. difficile* virulence in several strains[34,35]. Considering the link between quorum sensing

158 genes and toxin production, we expected the expression of genes for quorum sensing
159 and toxin production and toxin titer to be concordant. Based on this model, we expected
160 GF mice to have the highest levels of expression of genes for toxin production (Fig. 3b)
161 and quorum sensing (Fig. 3c); however, these transcripts were not found in the GF
162 mice. We also observed the highest level of expression for quorum sensing genes in
163 cefoperazone-treated mice, but *tcdA* expression in these animals was not the highest
164 among the different treatment groups. Interestingly, the levels of expression for genes
165 associated with toxin production did not match the toxin titers observed in the animals.
166 These results suggest that the relationship between toxin titer and the expression of
167 genes for toxin production is even more complex than current models indicate.

168 We next focused on the regulators of metabolic pathways. Sigma factors are master
169 regulators and a subset have been shown to integrate signals from intra- and
170 extracellular nutrient concentrations[20,21,31,36]. The transcription of the global repressor
171 *codY* is responsive to intracellular concentrations of *C. difficile* energy sources[37].
172 Highest transcription for this gene was found in cefoperazone-treated and GF mice (Fig.
173 3d). The regulation networks of CodY and CcpA are highly interconnected, with the
174 expression of *ccpA* specifically linked to local concentration of rapidly metabolizable
175 carbon sources[38]. Cefoperazone-treated mice also exhibited increased transcription of
176 *ccpA*, but the GF condition did not follow the same pattern. CcpA acts directly on *spo0A*
177 (Fig. 3d), which positively regulates initiation of the sporulation pathway in *C. difficile*.
178 Transcripts for *spo0A* were highly abundant in all conditions tested except for
179 clindamycin-treated mice, where it was still moderately detectable. The sig-family of
180 sigma factors is under the control of spo0A and regulate different stages of sporulation.

The genes from this family with the highest total transcription (*sigA1*, *sigF*, *sigG*, *sigH*, and *sigK*) each demonstrated a unique pattern of expression between conditions. These results indicate that complete expression of sporulation likely integrates multiple levels of signaling and is more complex than a single metabolic switch. Both CcpA and Spo0A also control pathogenicity by regulating toxin production (Fig. 3d). We found expression of the toxin negative regulator *tcdC* in all of the antibiotic-treated groups, but no detectable transcripts for the positive toxin A/B regulator *tcdR* were seen in any treatment. In addition to its effects on sporulation and virulence, CcpA also regulates the expression of other sigma factors that generally mediate distinct forms of *C. difficile* metabolism as needed. These targets include *rex* (general fermentation regulator) and *prdR* (Stickland fermentation regulator) (Fig. 3d). Although the expression of both has been shown to be linked to environmental proline concentrations, *rex* integrates additional signals from the intracellular NADH/NAD+ ratio to also control carbohydrate fermentation. Low-level transcription of *prdR* was found across all conditions, however *C. difficile* expression the *rex* gene highly in both cefoperazone-treated and GF mice. Combined, the variable expression of these sigma factors support the hypothesis that *C. difficile* adapts expression metabolism to fit its needs between colonized environments.

Gene sets from multiple *C. difficile* metabolic pathways are differentially expressed between colonized environments. In the context of similar colonization between antibiotic-treated animals, differential expression of global metabolic control mechanisms that are under the control of specific nutrient concentrations suggests that *C. difficile* adapts to each environment when in competition with the resident microbiota.

To test this further, we quantified the total expression for all KEGG annotated genes in the *C. difficile* genome (Fig. S2a). We then focused on general differences in carbohydrate (Fig. S2b) and amino acid (Fig. S2c) metabolism in order to asses for apparent differences in the utilization of carbon sources by *C. difficile* across environments. However aside from overall lower expression of most gene families in GF mice, no other clear trends were evident at this broad level of analysis so we moved toward a more fine-scale resolution of annotation and focused on specific gene sets known to contribute to certain forms of *C. difficile* metabolism (Table S1). Also, to more effectively compare between colonized states, we calculated the percentage of total expression between antibiotic-treated conditions for each gene (Fig. 4). We then identified the condition in which each gene was most highly transcribed and adjusted the size of the corresponding point relative to the largest transcipt abundance (Fig. 4a). This demonstrated that genes involved in amino acid catabolism had the greatest amount of expression overall, relative to other gene sets. This category includes those enzymes involved in Stickland fermentation (*arg*, *fdh*, *grd*, and *prd* loci) as well as several general peptidases (*pep* family). These results indicated that catabolizing environmental amino acids may be important for the growth of *C. difficile* during infection.

To more clearly indentify associations of gene sets with each condition, we also analyzed each set separately. First, we found that the expression of genes associated with amino acid catabolism were expressed at nearly consistent levels across the conditions (Fig. 4b). This was in agreement with the high level of overall expression associated with these genes. Additionally, genes for the metabolism of the host-derived

amino sugars N-acetylglucosamine and N-acetylmannosamine were also expressed at consistent levels across each treatment group (*glm*, *nan*, *mur*, and *acd* loci) (Fig. 4c). Along similar lines with related molecules, a number of genes for certain monosaccharides entering (*gal*, *man*, *pmi*, and *tag* loci) and leading through glycolysis (*fba*, *fbp*, *gap*, and *pfk*), as well as catabolism of the polysaccharides trehalose and cellibiose (*treA* and *celG*) were expressed relatively evenly between each condition (Fig. 4d & 4e). Combined, these findings suggest that catabolism of amino acids and specific carbohydrates are likely core components of the *C. difficile* nutritional strategy during infection.

Aside from those gene sets that were equally expressed across conditions, there were also large scale differences in expression of certain pathways between groups of mice. We chose to assess sugar transport systems have been associated with adaptive expression of phosphotransferase systems (PTS) and ABC transporters with many known differences in substrate specificities[25]. Among the genes classified as PTS transporters (Fig. 4f) were overrepresented in both clindamycin and streptomycin-treated mice, while ABC sugar transporters (Fig. 4g) were overrepresented in the cefoparazone-treated mice. The most stark differences were seen in transcription for genes involved in sugar alcohol catabolism (Fig. 4h). Expression of these genes was entirely absent from clindamycin-treated mice and expression of genes for mannitol utilization (*mtl* operon) were overrepresented in cefoparazone-treated mice and expression of genes for sorbitol utilization (*srl* operon) were overrepresented in streptomycin-treated mice. Concordant patterns also emerged in genes associated with fermentation end steps (Fig. 4i) and polysaccharide degradation (Fig. 4e). Short chain

fatty acids (SCFAs) and alcohols are the end products of both carbohydrate and amino acid fermentation in *C. difficile* through separate pathways with shared terminal steps. Transcripts for genes involved in *C. difficile* butyrate/butanol metabolism (*ptb*, *buk1*, *cat2*, and *adhE*) were more abundant in clindamycin-treated mice (Fig. 4i). Additionally, alpha/beta-galactosidase genes (*aglB* and *bglA*) were also overrepresented in clindamycin-treated mice (Fig. 4e). Together these patterns suggested that polysaccharide fermentation occurred this condition. More subtle differences were seen in those gene associated with glycolysis (Fig. 4d). This category includes genes for not only the steps of glycolysis, but also several genes that mediate entry points of monosaccharides to glycolysis. Transcripts for several genes in this group (*eno*, *gapA*, *gpmI*, *tpi*, and *pyk*) were overrepresented in cefoparazone-treated mice, however *fruK* was overrepresented in streptomycin-treated mice which catalyzes the committed step of glycolysis. Overall, these results support the hypothesis that *C. difficile* is able to adapt its metabolism to fit the nutrient availability across different susceptible environments.

Structure of genome-scale bipartite metabolic model underscores known bacterial metabolism. To further investigate which metabolites were differentially utilized between conditions, we represented the metabolic network of *C. difficile* as a directed bipartite graph using the genome annotation. Enzymes and metabolites were represented by nodes and their interaction by the edge between the nodes (Fig. 5a). To validate our metabolic network, we calculated betweenness centrality (BC) and overall closeness centralization index (OCCI) for all enzyme and metabolite nodes in the bipartite metabolic network of *C. difficile* generated for this study (Table S2). In

biological terms, BC reflects the amount of influence a given hub has on the overall flow of metabolism through the network[39] and OCCI indicates those enzymes and substrates that are the most central components of the organism's metabolism[40]. For both enzymes and substrates, the 18 of top 20 nodes with the highest BC values were involved in glycolysis, fermentation, and amino acid synthesis. In agreement, almost all nodes with the largest OCCI values were involved in glycolysis and amino acid synthesis as well. Enzymes that scored highly in both metrics included pyruvate kinase, aspartate aminotransferase, and formate C-acetyltransferase while substrates consistently scoring most highly were pyruvate, acetyl-CoA, D-glyceraldehyde 3-phosphate. This indicated to us that the topology of the network reflects established bacterial physiology.

Metabolite importance algorithm reveals adaptive nutritional strategies of *C. difficile* during infection across distinct environments. Moving beyond a strictly topological analysis of the *C. difficile* metabolic network, we sought to utilize transcriptomic data to infer which metabolites *C. difficile* is most likely to obtain from its environment in each condition. To accomplish this we mapped normalized transcript abundances to the enzyme nodes in the network. Due to the coupling of transcription and translation in bacteria, we were able to use this information as a proxy for enzyme levels. The importance of each metabolite was measured as the log-transformed difference between the average transcript levels of enzymes that use the metabolite as a substrate and those that generate it as a product (Fig. 5b). A metabolite with a high importance score is most likely obtained from the environment because the expression of genes for enzymes that produce the metabolite are low. Then, using a Monte Carlo-

296 style simulation, we generated a random transcript abundance distribution for each
297 enzyme node to then calculate new metabolite importance scores for each iteration. We
298 then created a confidence interval of scores for each metabolite that would likely result
299 from random noise[41]. This provided a standard of comparison for actual importance
300 scores from single timepoint measurments, and ultimately allow for computing the
301 significance level that a given score has a high probability of being excluded from its
302 associated null hypothesis score distribution.

303 Applying these methods to the *C. difficile* transcriptomic data collected from the *in vivo*
304 CDI models, we sought to assess differential patterns of metabolite importance. We first
305 ranked the importance scores to identify the most important metabolites for each
306 treatment group (Table S3). To identify the core metabolites that are essential to *C.*
307 *difficile* in any condition, we compared the highest 50 scoring, significant metabolites
308 from each treatment group ($P < 0.05$) (Fig. 6a). The host derived amino sugar N-acetyl-
309 D-glucosamine was found to be consistently important. Components of the Stickland
310 fermentation pathway were also found to be important to *C. difficile* in all conditions
311 tested including proline, 3−hydroxybutanoyl−CoA, formate, and some selenium-
312 containing compounds[42–44]. This indicated that these metabolites may be an integral
313 component of the nutrient niche for *C. difficile* in any infection condition. Additionally,
314 acetate was found to be important in all conditions, but was just below the significance
315 cutoff in GF mice (Table S3). It has been shown that *C. difficile* metabolizes acetate for
316 use in glycolysis[45]. We directly tested the relative concentration of acetate in
317 cefoperazone-treated *C. difficile*-infected mice versus mock-infected mice. We found
318 that *C. difficile* colonization led to a significant decrease in the levels of acetate (Fig. S4)

319 suggesting that *C. difficile* was utilizing acetate in the cecum. These findings provided
320 validation for our metabolite importance algorithm as well as supporting known
321 elements of *C. difficile* metabolism.

322 Returning to our hypothesis that *C. difficile* adapts its metabolism to fit the surrounding
323 community, we identified those metabolites that were uniquely important to each
324 condition in which *C. difficile* colonized. We cross-referenced the top 25 positively
325 scoring, significant substrates ($P < 0.05$) between treatment groups to uncover the most
326 important patterns of nutrient utilization by *C. difficile* in each (Fig. 6b). Each group of
327 metabolites contained at least one known carbohydrate growth substrate of *C.*
328 *difficile*[7,46]. This included close analogs of D-fructose, mannitol, N-acetylneuraminic
329 acid, and salicin. Furthermore, in GF mice where no other competitors are present, our
330 model indicated that *C. difficile* was more likely to acquires several amino acids (lysine,
331 leucine, and isoleucine) from the environment instead of expending energy to produce
332 them itself. These data support the hypothesis that *C. difficile* may exploit alternative
333 nutrient sources between the susceptible environments it colonizes.

334 **Carbon sources sources predicted to be important using network-based**
335 **approach support *C. difficile* differential growth *in vitro*.** To validate the biological
336 relevance of substrates identified as uniquely important to *C. difficile* metabolism
337 through our network-based analysis, we tested whether *C. difficile* was able to utilize
338 each substrate for *in vitro* growth (Fig. 6c). This was performed using a modified defined
339 *C. difficile* minimal media[7], supplemented individually with the selected carbohydrates
340 implicated by high importance scores. Because *C. difficile* is auxotrophic for several
341 amino acids it was necessary to include amino acids in the minimal media; however,

since it can use amino acids for growth through Stickland fermentation the most effective negative control was growth in media lacking carbohydrates but containing amino acids (Max OD_{600} = 0.212).

N-acetyl-D-glucosamine important to *C. difficile* in each condition tested (Fig. 6b). When tested for improved growth, significantly more growth (Max OD_{600} = 0.774) was observed compared to no carbohydrate (+ amino acids) controls ($P < 0.001$). This provided evidence that N-acetyl-D-glucosamine, derived from the host mucus layer, may be a central component of the *C. difficile* nutritional niche during infection. Trehalose was also shown to be important in each condition and supported *C. difficile* significant growth ($P < 0.001$; Max OD_{600} = 0.559), but was more likely provided by the diet than from the host. Furthermore, at least one carbohydrate highlighted as distinctly more important in each of the antibiotic treatment groups provided high levels of *C. difficile* growth relative to control wells ($P < 0.001$). This included D-fructose (streptomycin; Max OD_{600} = 0.671), mannitol (cefoperazone; Max OD_{600} = 0.464), salicin (clindamycin; Max OD_{600} = 0.869), and N-acetylneuraminate (GF; Max OD_{600} = 0.439). Because it was not possible to test aminofructose directly, we instead chose to test fructose, an immediate breakdown byproduct of aminofructose catabolism. We also tested both starch and acetate for the ability to support *C. difficile* growth *in vitro*, but neither should any improvement over no carbohydrate control (Fig. S5). Maximum growth rate analysis for each carbohydrate also indicated potential hierarchy of growth nutrient preference (Table S4). The progression is as follows: D-fructose (slope = 0.089), N-acetyl-D-glucosamine (slope = 0.085), salicin (slope = 0.077), mannitol (slope = 0.044) / trehalose (slope = 0.044), and finally N-acetylneuriminate (slope = 0.024).

This suggested that *C. difficile* was most well-suited to metabolize the nutrient source that is most likely to be present in all susceptible mouse ceca.

Discussion

Collectively, our results support the hypothesis that *C. difficile* can adapt its metabolism to the available niche landscape across susceptible gut environments and give insight to the adaptive strategies that *C. difficile* can use to colonize diverse human microbiota. Data from both our *in vivo* and *in vitro* experiments demonstrate the plasticity of *C. difficile* to effectively change its metabolism to utilize alternative resources for growth. This may be the result of increased concentration of particular metabolites as a consequence of concordant decreases in the population of one or more competitors for those resources. These preliminary conclusions are further supported by previous mass spectrometry-based efforts analyzing the metabolome from mouse intestinal content treated under similar conditions to those used in the current study. These investigations revealed that several of the substrates predicted to be used by *C. difficile* in a given condition through metabolic modeling (Fig. 6a & 6b), are increased in the gastrointestinal tract of mice in the corresponding treatment group. One recent study found that cefoperazone treatment resulted in a 553-fold increase in mannitol concentration in the cecum of mice prior to *C. difficile* colonization[7]. Similar trends have also been demonstrated in streptomycin-treated conventional and GF mice[10,47]. Together these results provide evidence that our network-based approach accurately predicts which metabolites *C. difficile* adapts its metabolic strategy towards, most likely due to changes in availability.

In addition to uncovering adaptive strategies of *C. difficile*, our method is also able to identify consistent trends in metabolism across environments. The findings that N-acetyl-D-glucosamine and Stickland fermentations substrates were consistently among the highest scoring shared metabolite among all tested conditions strongly indicates that these metabolites are central to the nutritional strategy of *C. difficile* and may be utilized in and condition in which they are available. The metabolism of both substrate types provides not only carbon and energy to *C. difficile*, but are also a source for nitrogen which is a limited resource in the mammalian lower GI tract[48]. Apart from exploring differential patterns in known metabolism, our modeling approach also allowed for the identification of emergent properties for the metabolic strategy of *C. difficile* during infection. One interesting result is the appearance of CO_2, an apparent metabolic end product, in the list of shared important metabolites (Fig. 6a). While this may be a shortcoming of the annotation, one group has posited that *C. difficile* may actually be autotrophic under certain conditions and could explain the appearance of CO_2 in Fig. 6a[49]. Furthermore, oxygen appears to be significantly important in clindamycin-treated mice (Fig. 6b). Reactive oxygen species could be introduced to the gut through antibiotic-induced stress on host mitochondia[50]. Despite the fact that *C. difficile* is considered to be a strict anaerobe, it does possess the functionality to deal with oxidative stress[51]. What this highlights is that our method does not only identify growth substrates, it also reports any metabolits that is very likely being removed from the environment.

While our results are consistent with previously published work on the metabolism of *C. difficile*, there are potential limitations of this approach. Ultimately, the metabolite

410 importance calculation is dependent on correct and existing gene annotation. In this
411 regard it has been shown that the pathway annotations in KEGG are robust to missing
412 elements[52], however this does not completely eliminate the posibility for this type of
413 error. Due to the topology of the metabolic network, we were also unable to integrate
414 stoichiometry for each reaction which may effect rates of consumption or production. In
415 addition to computational limitations, our network-based approach simplifies several
416 aspects of bacterial metabolism. First, the importance algorithm operates under the
417 assumption that all detectable transcript is translated to effector protein. While this is not
418 completely accurate, since bacterial transcription and translation are physically coupled,
419 we were comfortable using normalized levels of transcription to infer approximate
420 amount of translation. Second, the metabolite importance scores do not account for the
421 amount of each metabolite that is actually available. Finally, the importance algorithm
422 only consider the transcription of those enzyme nodes immediately adjacent to the
423 metabolite node of interest. Although this does not negate any observations made in the
424 current study, it may be beneficial to incorporate the importance of other local
425 metabolites or subnetworks into the final score of each metabolite. In spite of these
426 assumptions, the method outlined here supports known elements of *C. difficile* biology
427 and future studies could employ metabolomic analysis to confirm the predictions made
428 here.

429 Based on the evidence presented, our results support the hypothesis that *C. difficile* is a
430 metabolic generalist and is able to catabolize alternative carbon sources across
431 susceptible gut environments. This may be due to an inability to outcompete a collection
432 of metabolic specialists in an intact community, and separate classes on antibiotics

differentially eliminate these populations and allow for *C. difficile* colonization. This concept may also potentially explain the success rate of fecal microbial transplant (FMT), in that wholesale installation of a diverse range of specialized metabolic strategies is enough to outcompete *C. difficile* from the majority of perturbed gut environments. Furthermore, our metabolic network platform may also prove informative for generating hypotheses through reverse ecology that could ultimately lead to uncovering new interaction between species that ultimately impact host health[53]. In conclusion, *C. difficile* is is able to optimize its nutritional strategy for each colonized gut environment. Our results implicate that further considerations are needed when attempting to design targeted prebiotic and probiotic therapies for the prevention or elimination of *C. difficile* from the human gut.

Methods

Animal care and antibiotic administration Adapted from the previously described model[54], six-to-eight week-old SPF C57BL/6 mice were obtained from a single breeding colony maintained at the University of Michigan for all experiments. Six-to-eight week-old GF C57BL/6 mice were obtained from a single breeding colony maintained at the University of Michigan and fed Laboratory Rodent Diet 5001 from LabDiet for all experiments. All animal protocols were approved by the University Committee on Use and Care of Animals at the University of Michigan and carried out in accordance with the approved guidelines. Specified SPF animals were administered one of three antibiotics; cefoperazone, streptomycin, or clindamycin (Table 1). Cefoperazone (0.5 mg/ml) and streptomycin (5.0 mg/ml) were administered in distilled drinking water *ad libitum* for 5 days with 2 days recovery with untreated distilled drinking water prior to

456 infection. Clindamycin (10 mg/kg) was given via intraperitoneal injection 24 hours before
457 time of infection.

458 ***C. difficile* infection and necropsy** *C. difficile* strain 630 spores were prepared from a
459 single large batch whose concentration was determined a week prior to challenge for all
460 experiments. On the day of challenge, 1×10^3 *C. difficile* spores were administered to
461 mice via oral gavage in phosphate-buffered saline (PBS) vehicle. Subsequent
462 quantitative plating for c.f.u. was performed to ensure correct dosage. Infection negative
463 control animals were given an oral garage of 100 µl PBS at the same time as those
464 mice administered *C. difficile* spores. 18 hours following infection, mice were euthanized
465 by carbon dioxide asphyxiation. Necropsy was then performed and cecal content was
466 split into three small aliquots (~100 µl). Two were flash frozen immediately for later DNA
467 extraction and toxin titer analysis respectively. The third aliquot was quickly moved to an
468 anaerobic chamber for c.f.u. quantification. The remaining content in the ceca (~1 ml)
469 was emptied into a stainless steel mortar in a dry ice/ethanol bath using 1 ml of sterile
470 PBS. This process was repeated for each mouse within a treatment group to pool
471 content into a single large sample (9 mice across 3 cages) to compensate for cage
472 effects as much as possible while maximizing sequencing depth. The content was then
473 finely ground and stored at -80° C for subsequent RNA extraction.

474 ***C. difficile* cultivation and quantification** Cecal samples were weighed and serially
475 diluted under anaerobic conditions (6% H, 20% CO_2, 74% N_2) with anaerobic PBS.
476 Differential plating was performed to quantify both *C. difficile* spores and vegetative cells
477 by plating diluted samles on CCFAE plates (fructose agar plus cycloserine, cefoxitin,
478 and erythromycin) at 37° C for 24 hours under anaerobic conditions[55]. It is important to

note that the germination agent taurocholate was omitted from these plates in order to only quantify vegetative cells. In parallel, undiluted samples were heated at 60° C for 30 minutes to eliminate vegetative cells and leave only spores[56]. 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[57]. Briefly, filtered-sterilized cecal content was serially diluted 1:5 in PBS. As a control for toxin-mediated cell rounding the cecal content was diluted a further 1:2 by the addition of an equal volume of goat anti-toxin serum (T5000; TechLab). Vero cells were grown to a confluent monolayer in DMEM (Dulbecco's Modified Eagle's medium), supplemented with 10% heat-inactivated fetal bovine serum and 1% penicillin-streptomycin. The cells then were transferred to a conical tube and centrifuged at 1,000 rpm for 5 minutes to pellet the cells. The old media was removed and the cells were re-suspended in fresh media to a final concentration of 1×10^5 cells per 90µL. 90µL of the cell suspension were seeded in each well of a 96-well plate and incubated at 37° C in a 5% CO_2 humidified incubator for 4 hours. Following the incubation, cecal samples were added to the Vero cells and the plate was incubated overnight at 37° C. Plates were viewed after 24 hours at 10x magnification for cell rounding. The cytotoxic titer was defined as the $\log_{10}$ transformed reciprocal of the highest dilution that produced rounding in 80% of the cells. A more detailed protocol with product information can be found at:

https://github.com/jlleslie/Intraspecific_Competition/blob/master/methods/Verocell_Toxin Activity_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[58]. 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[58]. 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)[59]. Curated sequences were clustered into operational taxonomic units (OTUs) using a 97% similarity cutoff with the average neighbor clustering algorithm. The number of sequences in each sample was rarefied to 2,500 per sample to minimize the effects of uneven sampling.

RNA extraction, shotgun library preparation, and sequencing To generate enough mRNA biomass contributed by *C. difficile*, we pooled cecal content from all mouse replicates into a single large isolation for each treatment group. Pooling was performed in a sterile stainless steel mortar resting in dry ice and a small amount of 100% ethanol. After all content for the given group was added, the sample was ground with a sterile pestle to a fine powder and scraped into a sterile 50 ml polypropylene conical tube.

524 Samples were stored at -80° C until the time of extraction. Immediately before RNA
525 extraction, 3 ml of lysis buffer (2% SDS, 16 mM EDTA and 200 mM NaCl) contained in
526 a 50 ml polypropylene conical tube was first heated for 5 minutes in a boiling water
527 bath[60]. The hot lysis buffer was added to the frozen and ground cecal content. The
528 mixture was boiled with periodic vortexing for another 5 minutes. After boiling, an equal
529 volume of 37° C acid phenol/chloroform was added to the cecal content lysate and
530 incubated at 37° C for 10 minutes with periodic vortexing. The mixture was the
531 centrifuged at 2,500 x g at 4° C for 15 minutes. The aqueous phase was then
532 transferred to a sterile tube and an equal volume of acid phenol/chloroform was added.
533 This mixture was vortexed and centrifuged at 2,500 x g at 4° for 5 minutes. The process
534 was repeated until aqueous phase was clear. The last extraction was performed with
535 chloroform/isoamyl alcohol to remove acid phenol. An equal volume of isopropanol was
536 added and the extracted nucleic acid was incubated overnight at -20° C. The following
537 day the sample was centrifuged at 12000 x g at 4° C for 45 minutes. The pellet was
538 washed with 0° C 100% ethanol and resuspended in 200 µl of RNase-free water.
539 Following the manufacturer's protocol, samples were then treated with 2 µl of Turbo
540 DNase for 30 minutes at 37° C. RNA samples were retrieved using the Zymo Quick-
541 RNA MiniPrep according the manufacturer's protocol. Completion of the reaction was
542 assessed using PCR for the V4 region of the 16S rRNA gene (Kozich, 2013). Quality
543 and integrity of RNA was assessed using the Agilent RNA 6000 Nano kit for total
544 prokaryotic RNA. The Ribo-Zero Gold rRNA Removal Kit Epidemiology was then used
545 to deplete prokaryotic and eukaryotic rRNA from the samples according the
546 manufacturer's protocol. Prior to library construction, quality and integrity as measured

again using the Agilent RNA 6000 Pico Kit. Stranded RNA-Seq libraries were made constructed with the TruSeq Total RNA Library Preparation Kit v2, both using the manufacturer's protocol. The Agilent DNA High Sensitivity Kit was used to measure concentration and fragment size distribution before sequencing. High-throughput sequencing was performed by the University of Michigan Sequencing Core in Ann Arbor, MI. For all groups, sequencing was repeated across 4 lanes of an Illumina HiSeq 2500 using the 2x50 bp chemistry.

Sequence curation, read mapping, and normalization. Raw transcript sequencing read curation was performed in a two step process. Residual 5' and 3' Illumina adapter sequences were trimmed using CutAdapt[35] on a per library basis. Reads were quality trimmed using Sickle (Joshi, 2011) on the default settings. An average of ~300,000,000 total reads (both paired and orphaned) remained after quality trimming. Mapping was accomplished using Bowtie2[61] and the default stringent settings. ~1,600,000 reads in sample each mapped to the annotated nucleotide gene sequences of *PeptoClostridium difficile* str. 630 from the KEGG: Kyoto Encyclopedia of Genes and Genomes[62]. Optical and PCR duplicates were then removed using Picard MarkDuplicates (http://broadinstitute.github.io/picard/), leaving ~150,000 reads per sample for final analysis. The remaining mappings were converted to idxstats format using Samtools[63] and the read counts per gene were tabulated. Discordant pair mappings were discarded and counts were then normalized to read length and gene length to give a per base report of gene coverage. Unless indicated otherwise, each collection of reads was then 1000-fold iteratively subsampled to 90% of the lowest sequence total within each analysis, and a median expression value for each gene was calculated.

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

Metabolite Importance Calculation. The substrate importance algorithm (Fig. 5a) favors metabolites that are more likely acquired from the environment (not produced

593 within the network), and will award them a higher score (Fig. 6b & 6c). The presumption
594 of our approach was that enzymes that were more highly transcribed were more likely to
595 utilize the substrates they act on due to coupled bacterial transcription and translation. If
596 a compound was more likely to be produced, the more negative the resulting score
597 would be. To calculate the importance of a given metabolite (m), we used rarefied
598 transcript abundances mapped to respective enzyme nodes. This was represented by t_o
599 and t_i to designate if an enzyme created or utilized m. The first step was to calculate the
600 average expression of enzymes for reactions that either created a given metabolite (i) or
601 consumed that metabolite (ii). For each direction, the sum of transcripts for enzymes
602 connecting to a metabolite were divided by the number of contributing edges (e_o or e_i) to
603 normalize for highly connected metabolite nodes. Next the raw metabolite importance
604 score was calculated by subtracting the creation value from the consumption value to
605 weight for metabolites that are likely acquired exogenously. The difference was $\log_2$
606 transformed for comparability between scores of individual metabolites. This resulted in
607 a final value that reflected the likelihood a metabolite was acquired from the
608 environment.Untransformed scores that already equaled to 0 were ignored and negative
609 values were accounted for by transformation of the absolute value then multiplied by -1.
610 These methods have been written into a single python workflow, along with supporting
611 reference files, and is presented as bigSMALL (Bacterlal Genome-Scale Metabolic
612 models for AppLied reverse ecoLogy) available in a public Github repository at
613 https://github.com/mjenior/bigsmall.

614 **Transcriptome Bootstrapping and Probability Distribution Comparison.** As
615 sequencing replicates of *in vivo* transcriptomes was not feasible, we applied a Monte

616 Carlo style simulation to distinguish calculated metabolite importances due to distinct
617 transcriptional patterns for the environment measured from those metabolites that were
618 constitutively important. We employed a 10,000-fold bootstrapping approach of
619 randomly reassigning transcript abundance for enzyme nodes and recalculating
620 metabolite importances. This approach was chosen over fitting a simulated
621 transcriptome to a negative binomial distribution because it created a more relevant
622 standard of comparison for lower coverage sequencing efforts. Using this method, each
623 substrate node accumulated a random probability distribution of importance scores
624 which were then used to calculate the median and confidence interval in order to
625 ultimately generate a p-value for each metabolite. This was a superior approach to
626 switch randomization since the connections of the network itself was created through
627 natural selection and any large-scale alterations would yield biologically uninformative
628 comparisons[64]. These calculations are also included within the standard bigSMALL
629 workflow presented above.

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

636 **Anaerobic *in vitro C. difficile* growth curves.** The carbon-free variation of *C. difficile*
637 Basal Defined Medium (NCMM) was prepared as previously described[7]. Individual
638 carbohydrate sources were added at a final concentration of 5 mg/mL and pair-wise

carbohydrate combinations were added at 2.5 mg/mL each (5 mg/mL total). A solution of the required amino acids was made separately and added when noted at identical concentrations to the same study. 245 µl of final media mixes were added to a 96-well sterile clear-bottom plate. A rich media growth control was also included, consisting of liquid Brain-Heart Infusion + 0.5% cysteine. All culturing and growth measurement were performed anaerobically in a Coy Type B Vinyl Anaerobic Chamber (3.0% H, 5.0% CO_2, 92.0% N, 0.0% O_2). *C. difficile* str. 630 was grown for 14 hours at 37° C in 3 ml BHI + 0.5% cysteine. Cultures were then centrifuged at 2000 rpm for 5 minutes and resulting pellets were then washed twice with sterile, anaerobic 1 × phosphate-buffered saline (PBS). Washed pellets were resuspended in 3 ml more PBS and 5 µl of prepped culture was added the each growth well of the plate containing aliquoted media. The plate was then placed in a Tecan Sunrise plate reader, heated to 37° C. Plates were incubated for 24 hours with automatic optical density readings at 600 nm taken every 30 minutes. OD_{600} values were normalized to readings from wells containing sterile media of the same type at equal time of incubation. Growth rates and other curve metrics were determined by differentiation analysis of the measured OD_{600} over time in R to obtain the slope at each time point.

Statistical methods. All statistical analyses were performed using R (v.3.2.0). Significant differences 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 acetate concentration were determined by Wilcoxon rank-abundance test with Holm-Bonferroni correction. Significant differences for growth curves compared to no carbohydrate

control (+ amino acids) were calculated using 2-way ANOVA with Holm-Bonferroni correction. Significance for metabolite importance scores was determined as described above.

Acknowledgements

The authors would like to acknowledge members of the University of Michigan Germfree Mouse Center, Sequencing Core, and Metabolomics Core for their assistance in experimental design, execution, and data collection. This work was supported by funding from the National Institutes of Health to PDS (R01GM099514, P30DK034933, U19AI09087, and U01AI124255), VBY (P30DK034933, U19AI09087, and U01AI124255), and a Translational Research Education Certificate grant to JLL (MICHR; UL1TR000433). Pooled and quality trimmed transcriptomic read data and experiment metadata are available through the NCBI Sequence Read Archive (SRA; PRJNA354635). Data processing steps for beginning from raw sequence data to the final manuscript are hosted at http://www.github.com/SchlossLab/Jenior_Modeling_NatMicro_2016.

Author Information

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

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

682 **Contributions** M.L.J. conceived, designed and performed experiments, analyzed data,
683 and drafted the manuscript. J.L.L. performed experiments and analyzed data. V.B.Y.
684 contributed to the manuscript. P.D.S. interpreted data and contributed the manuscript.

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

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

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	0.5 mg/ml
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	5.0 mg/ml
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	10 mg/kg body weight

687
688 **Table 1 l Antibiotics used during *C. difficile* infection models.**

Figure Legends

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

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

712 **Figure 3 | Select *C. difficile* gene set expression compared between treatment**
713 **group.** Relative abundances of *C. difficile* transcript for specific genes of interest,
714 rarefied to 90% of the total number of reads within each colonized condition shown. **(a)**
715 Transcription for select genes from the *C. difficile* sporulation pathway with the greatest
716 variation in expression between the conditions tested. **(b)** Relative abundances of
717 transcript for genes that encode effector proteins from the *C. difficile* pathogenicity
718 locus. **(c)** Transcript abundances for genes associated with quorum sensing in *C.*
719 *difficile*. **(d)** Transcript relative abundance of select sigma factors which expression or
720 activity is influenced by environmental metabolite concentrations. Asterisks (*) indicate
721 genes from which transcript was undetectable.

722 **Figure 4 | *C. difficile* expression of gene sets for carbon metabolism pathways**
723 **across antibiotic pretreatments. (a)** Ternary plot indicating the relative abundance of
724 transcripts for all *C. difficile* str. 630 genes across the three colonized antibiotic-treated
725 conditions (gray points). Raw transcript abundances were iteratively rarefied and the
726 median expression of each gene was calculated (~24x coverage). Each point
727 represents a unique gene from the annotated genome of *C. difficile* str. 630 with
728 position reflecting the ratio of transcription for that gene in all three antibiotic
729 pretreatments. Transcripts for genes that are over-represented in a single condition are
730 placed more proximal to the corner associated with that treatment group. Points placed
731 near the center are equally expressed across all of the conditions measured at 18 hours
732 post-infection. Points are colored based on inclusion in specific carbon metabolic
733 pathways, and point sizes within groups of interest were determined based on the
734 highest expression value for each gene from a single condition. **(b – i)** Groups from (a)

735 are shown individually, without abundance information, for ease of comparison. Genes
736 included in each group with normalized transcript abundances can be found in Table
737 S1, and refer to Fig. S3 for additional explanantion of figure interpretation.

738 **Figure 5 | Genome-scale bipartite metabolic modeling results using the**
739 **transcriptome of *C. difficile* str. 630 in each colonized environment. (a)** Largest
740 component from the bipartite genome-scale metabolic model of *C. difficile* str. 630. The
741 complete network contains 447 enzymes and 758 metabolites, with 2135 directed
742 edges. Size of enzyme nodes is relative to the number of normalized reads mapped to
743 the corresponding gene. The sizes shown reflect the transcriptome of *C. difficile* str. 630
744 during infection of cefoperazone-treated mice after 18 hours of infection. Below the
745 representative network is the metabolite importance algorithm separated into 3
746 components; (i) relative transcription of reactions consuming a metabolite, (ii) relative
747 transcription of reactions consuming a metabolite, and (iii) difference of consumption
748 and creation of the given metabolite. **(b)** The expanded window displays an example of
749 a single metabolite importance calculation based on local enzyme gene transcription.
750 White values in the red nodes represent the number of normalized transcript reads
751 mapping to the gene sequence for each enzyme node. Average expression of input and
752 output reactions surrounding metabolite **m** are calculated at then the difference of these
753 values found to get the relative importance of **m**. Log_2 transformation is then performed
754 for uniform comparison between metabolites.

755 **Figure 6 | Results from network-based metabolite importance calculation and *in***
756 ***vitro* growth with important carbohydrates.** Prior to importance calculation, transcript
757 abundances for each condition were evenly rarefied for even comparison across

758 colonized environments (~18x coverage). **(a)** Median shared significant metabolites
759 among the 50 highest scoring metabolites from each condition ($P < 0.05$). Median
760 importance scores and pooled random distribution were recalculated per metabolite
761 using the scores from each condition tested. **(b)** Distinctly important significant
762 metabolites from each treatment group ($P < 0.05$). The top 25 scoring metabolites from
763 each group was cross-referenced against each other group resulting in metabolites that
764 are differentially important between environments. **(c)** *in vitro* growth curves validating
765 identified growth nutrients from network analysis. One metabolite that is consistently
766 important to *C. difficile* and at least one metabolite indicated as distinctly important from
767 each group supported growth significantly more ($P < 0.001$) than no carbohydrate
768 control (+ amino acids, gray line). Only those carbon sources that significantly improved
769 *C. difficile* growth over control are displayed (remainder are located in Table S4).
770 Significant differences were calculated using 2-Way ANOVA with Holm-Bonferroni
771 correction.

772 **Supplementary Figure 1 | Analysis of bacterial community structure resulting**
773 **from antibiotic treatment.** Results from 16S rRNA gene amplicon sequencing from
774 bacterial communities of cecal content in both mock-infected and *C. difficile* 630-
775 infected animals 18 hours post-infection across pretreatment models. **(a)** Non-metric
776 multidimensional scaling (NMDS) ordination based on $Theta_{YC}$ distances for the gut
777 microbiome of all conventionally-raised mice used in these experiments (n = 63). All
778 treatment groups are significantly different from all other groups by AMOVA ($P < 0.001$).
779 **(b)** Inverse Simpson diversity for each cecal community from the mice in (a). Cecal

780 communities from mice not treated with any antibiotics are significantly more diverse
781 than any antibiotic-treated condition by Wilcoxon test ($P < 0.001$).

782 **Supplementary Figure 2 | Expression of specific KEGG gene families.** Abundances
783 of normalized transcriptomic reads from *C. difficile* str. 630 in each tested condition. **(a)**
784 All KEGG families. **(b)** Those sub-families within Carbohydrate metabolism. **(c)** Sub-
785 families within Amino acid metabolism.

786 **Supplementary Figure 3 | Additional explanation for Figure 4 interpretation.**
787 Relative abundance of transcription for *C. difficile* 630 genes during infection across the
788 3 antibiotic pretreatment models used during this study. Points that are located closer to
789 a corner are more highly transcribed in the condition associated with that corner
790 compared to the others. As this shows a 3-dimensional data set in 2 dimensions, there
791 is an amount of distortion proximal to each corner. Simply put for points that are nearer
792 to an edge, a greater percentage of their total transcription was contributed by *C.*
793 *difficile* colonizing those mice. **(a)** This point represents the transcription for a gene that
794 is overrepresented in cefoperazone-treated mice. **(b)** This point represents a gene in
795 which transcripts are equally detectable in all 3 conditions. **(c)** Transcripts for this gene
796 are only underrepresented in only cefoperazone-treated mice, and are equally
797 detectable in clindamycin and streptomycin-treated animals.

798 **Supplementary Figure 4 | *in vivo* acetate concentrations with GC-MS analysis**
799 **from cefoperazone-treated mouse cecal content.** 2 groups of 5 mice each were
800 pretreated with cefoperazone as outlined Fig. 1b. A single cage was infected with *C.*
801 *difficile* in the same fashion as other experiments described here. Animals were

802 necropsied at 18 hours post-infection and cecal content was flash frozen for later GC-
803 MS analysis. Significance was determined using Wilcoxon signed-rank test.

804 **Supplementary Figure 5 | Additional growth curves for additional carbon sources**
805 **and controls.** Significant metabolites from network analysis that did not provide
806 improved growth over no carbohydrate (+ amino acids) control. Also included is the
807 negative control of minimal media with no amino acids as well as *C. difficile* growth in
808 standard Brain-Heart Infusion broth.

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

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

813 **Supplementary Table 3 | All metabolites with significant important scores for *C.***
814 ***difficile* in each colonized condition.**

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

816 ## References

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

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

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

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

827 5. Lawley, T. D. & Walker, A. W. Intestinal colonization resistance. *Immunology* **138,** 1–
828 11 (2013).

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

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

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

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

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

841 11. Ferreyra, J. A. *et al.* Gut microbiota-produced succinate promotes *Clostridium*
842 *difficile* infection after antibiotic treatment or motility disturbance. *Cell Host and Microbe*
843 **16,** 770–777 (2014).

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

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

849 14. Wilson, K. H. & Perini, F. Role of competition for nutrients in suppression of
850 *Clostridium difficile* by the colonic microflora. *Infection and Immunity* **56,** 2610–2614
851 (1988).

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

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

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

859 18. Neumann-Schaal, M., Hofmann, J. D., Will, S. E. & Schomburg, D. Time-resolved
860 amino acid uptake of *Clostridium difficile* 630 Delta-erm and concomitant fermentation

861 product and toxin formation. *BMC Microbiology* 281 (2015). doi:10.1186/s12866-015-
862 0614-2

863 19. Nawrocki, K. L., Edwards, A. N., Daou, N., Bouillaut, L. & McBride, S. M. CodY-
864 dependent regulation of sporulation in *Clostridium difficile*. *Journal of Bacteriology* **198,**
865 2113–2130 (2016).

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

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

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

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

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

876 25. Kansau, I. *et al.* Deciphering adaptation strategies of the epidemic *Clostridium*
877 *difficile* 027 strain during infection through in vivo transcriptional analysis. *PLoS ONE*
878 **11,** (2016).

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

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

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

887 29. Putnam, E. E., Nock, A. M., Lawley, T. D. & Shen, A. SpoIVA and sipl are
888 *Clostridium difficile* spore morphogenetic proteins. *Journal of Bacteriology* **195,** 1214–
889 1225 (2013).

890 30. Underwood, S. *et al.* Characterization of the sporulation initiation pathway of
891 *Clostridium difficile* and its role in toxin production. *Journal of Bacteriology* **191,** 7296–
892 7305 (2009).

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

895 32. Saujet, L., Pereira, F. C., Henriques, A. O. & Martin-Verstraete, I. The regulatory
896 network controlling spore formation in *Clostridium difficile*. *FEMS Microbiology Letters*
897 **358,** 1–10 (2014).

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

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

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

905 36. Donnelly, M. L., Fimlaid, K. A. & Shen, A. Characterization of *Clostridium difficile*
906 spores lacking either SpoVAC or DPA Synthetase. *Journal of Bacteriology* JB.00986–15
907 (2016). doi:10.1128/JB.00986-15

908 37. Dineen, S. S., Villapakkam, A. C., Nordman, J. T. & Sonenshein, A. L. Repression
909 of *Clostridium difficile* toxin gene expression by CodY. *Molecular Microbiology* **66,** 206–
910 219 (2007).

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

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

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

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

920 42. Aboulnaga, E. H. *et al.* Effect of an oxygen-tolerant bifurcating butyryl coenzyme a
921 dehydrogenase/electron-transferring flavoprotein complex from *Clostridium difficile* on
922 butyrate production in *Escherichia coli*. *Journal of Bacteriology* **195,** 3704–3713 (2013).

923 43. Fonknechten, N. *et al. Clostridium sticklandii*, a specialist in amino acid
924 degradation:revisiting its metabolism through its genome sequence. *BMC genomics* **11,**
925 555 (2010).

926 44. Jackson, S., Calos, M., Myers, A. & Self, W. T. Analysis of proline reduction in the
927 nosocomial pathogen *Clostridium difficile*. *Journal of Bacteriology* **188,** 8487–8495
928 (2006).

929 45. Karlsson, S., Burman, L. G. & Åkerlund, T. Induction of toxins in *Clostridium difficile*
930 is associated with dramatic changes of its metabolism. *Microbiology* **154,** 3430–3436
931 (2008).

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

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

936 48. Fuller, M. F. & Reeds, P. J. Nitrogen cycling in the gut. *Annual review of nutrition* **18,**
937 385–411 (1998).

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

940 50. Li, C. H., Cheng, Y. W., Liao, P. L., Yang, Y. T. & Kang, J. J. Chloramphenicol
941 causes mitochondrial stress, decreases ATP biosynthesis, induces matrix
942 metalloproteinase-13 expression, and solid-tumor cell invasion. *Toxicological Sciences*
943 **116,** 140–150 (2010).

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

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

949 53. Levy, R. & Borenstein, E. Reverse Ecology: from systems to environments and
950 back. *Advances in experimental medicine and biology* **751,** 329–345 (2012).

951 54. Theriot, C. M. *et al.* Cefoperazone-treated mice as an experimental platform to
952 assess differential virulence of *Clostridium difficile* strains. *Gut microbes* **2,** 326–334
953 (2011).

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

957 56. Sorg, J. a. & Sonenshein, A. L. Inhibiting the initiation of *Clostridium difficile* spore
958 germination using analogs of chenodeoxycholic acid, a bile acid. *Journal of Bacteriology*
959 **192,** 4983–4990 (2010).

960 57. Leslie, J. L. *et al.* Persistence and toxin production by *Clostridium difficile* within
961 human intestinal organoids result in disruption of epithelial paracellular barrier function.
962 *Infection and Immunity* **83,** 138–145 (2015).

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

965 59. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for
966 rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and*
967 *Environmental Microbiology* **73,** 5261–5267 (2007).

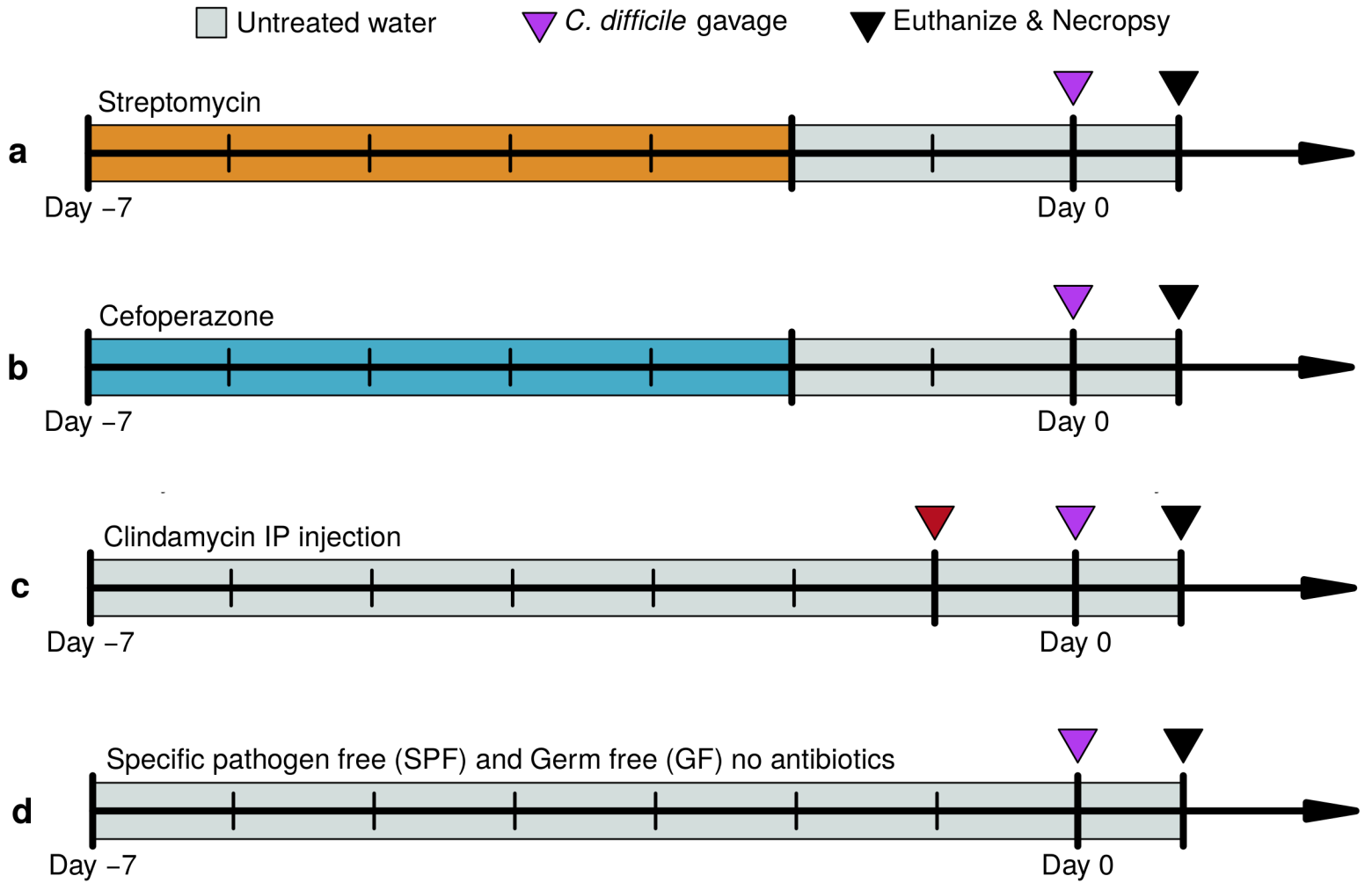
968 60. Lopez-Medina, E., Neubauer, M. M., Pier, G. B. & Koh, A. Y. RNA isolation of
969 *Pseudomonas aeruginosa* colonizing the murine gastrointestinal tract. *Journal of*
970 *visualized experiments : JoVE* 6–9 (2011). doi:10.3791/3293

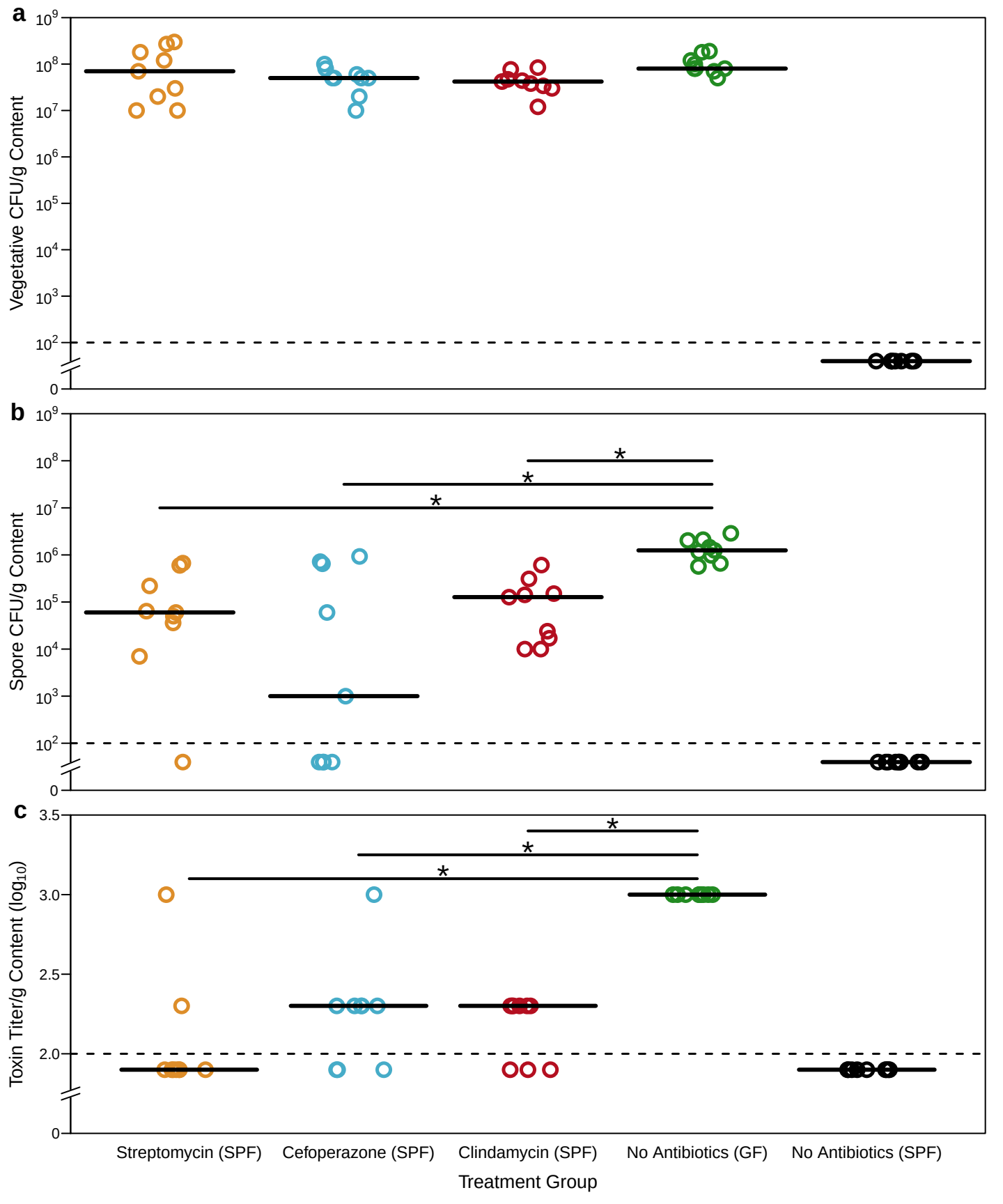
971 61. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-
972 efficient alignment of short DNA sequences to the human genome. *Genome Biol* 1–10
973 (2009). doi:gb-2009-10-3-r25 [pii]\r10.1186/gb-2009-10-3-r25

974 62. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. **27,** 29–34
975 (1999).

976 63. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,**
977 2078–2079 (2009).

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


Untreated water
C. difficile gavage
Euthanize & Necropsy
a
Streptomycin
Day −7
Day 0
b
Cefoperazone
Day −7
Day 0
c
Clindamycin IP injection
Day −7
Day 0
d
Specific pathogen free (SPF) and Germ free (GF) no antibiotics
Day −7
Day 0


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})
3.5
3.0
2.5
2.0
0
*
*
*
Streptomycin (SPF)
Cefoperazone (SPF)
Clindamycin (SPF)
No Antibiotics (GF)
No Antibiotics (SPF)
Treatment Group

Streptomycin (SPF)
Cefoperazone (SPF)
Clindamycin (SPF)
No antibiotics (GF)

a

Sporulation

Relative Transcript Abundance

30%
20%
10%
0%

spoIIAB *spoIIE* *spoVG* *spoVS* *cdeC* *cotD* *cotJB2* *spoIVA* *spoVB* *spoVFB* *sspA* *sspB*

Early

Late

b

Pathogenicity

1.5%
1.0%
0.5%
0%

tcdA *tcdB* *tcdE*

c

Quorum sensing

Relative Transcript Abundance

3.0%
2.0%
1.0%
0%

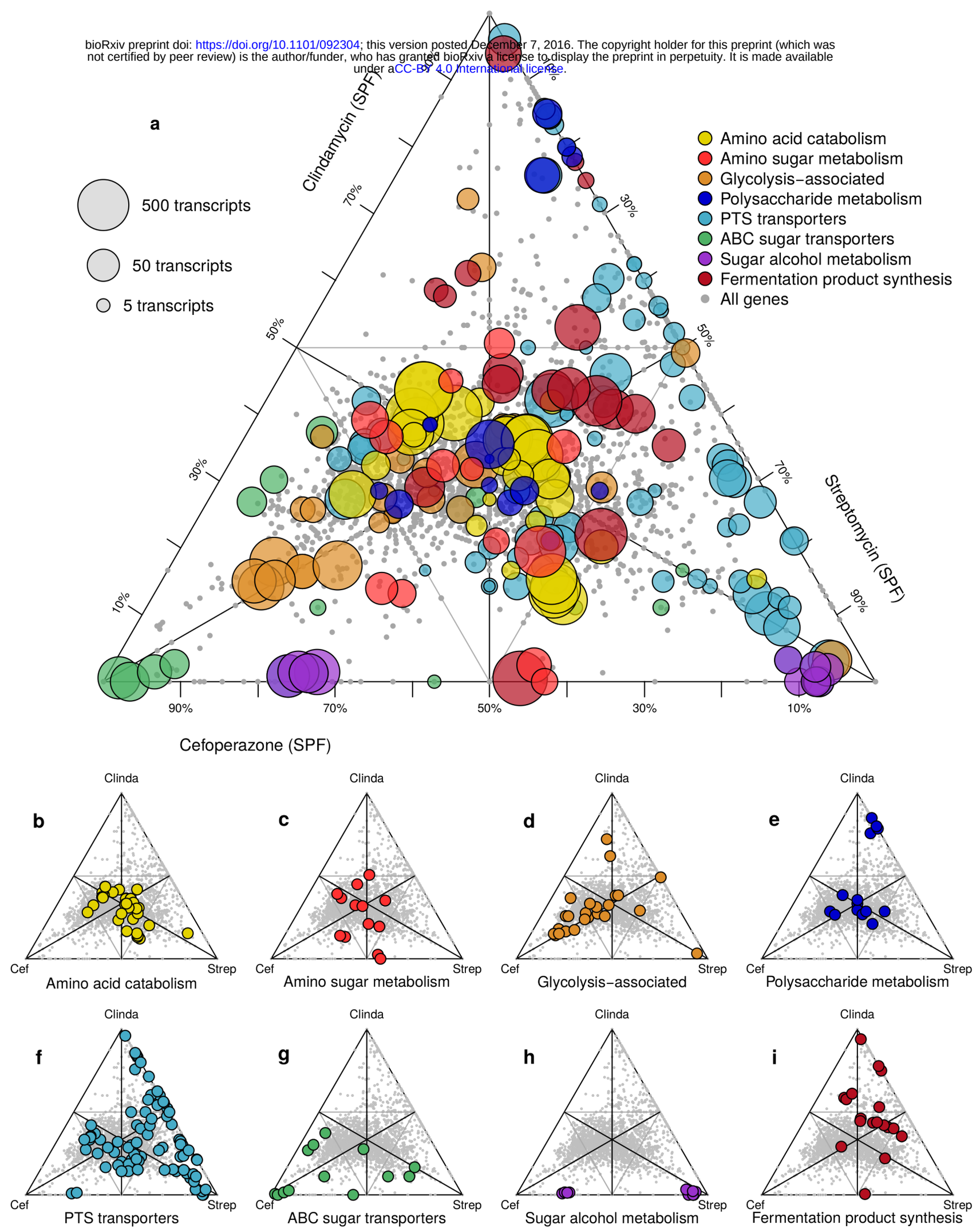
agrB *agrD* *luxS*

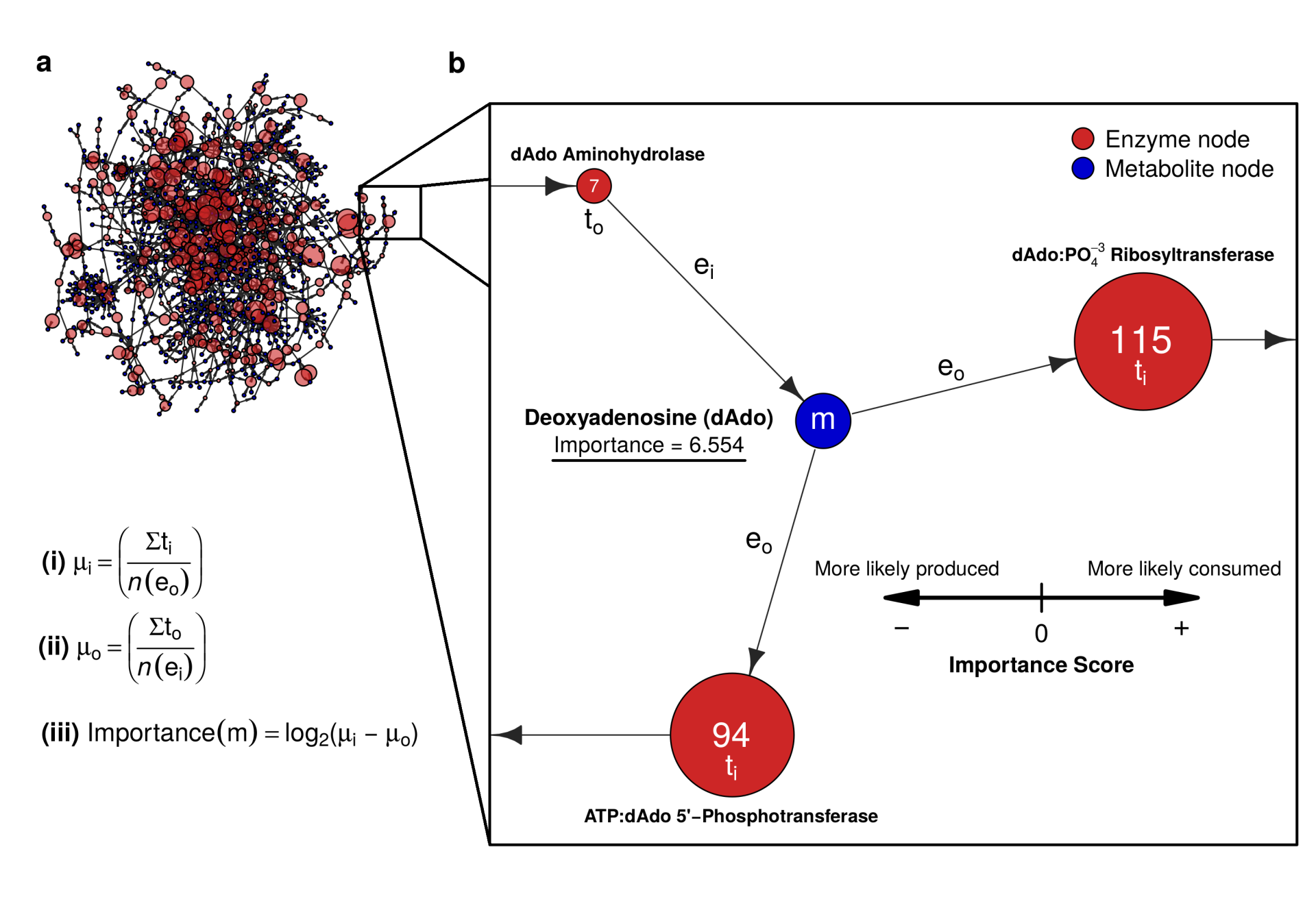
d

Sigma factors

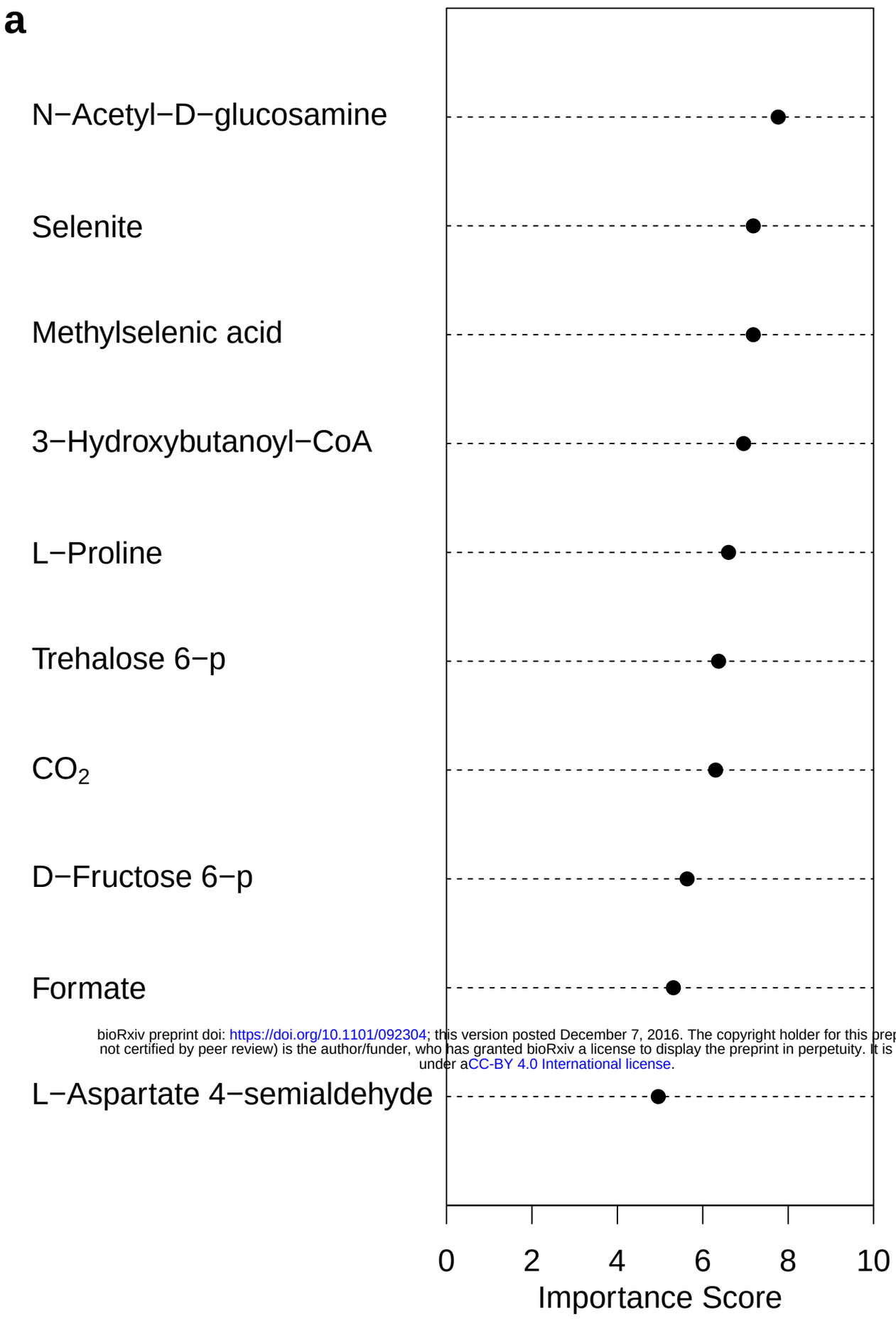
30%
20%
10%
0%

codY *ccpA* *cdtR* *tcdC* *tcdR* *spo0A* *sigA1* *sigF* *sigG* *sigH* *sigK* *rex* *prdR*

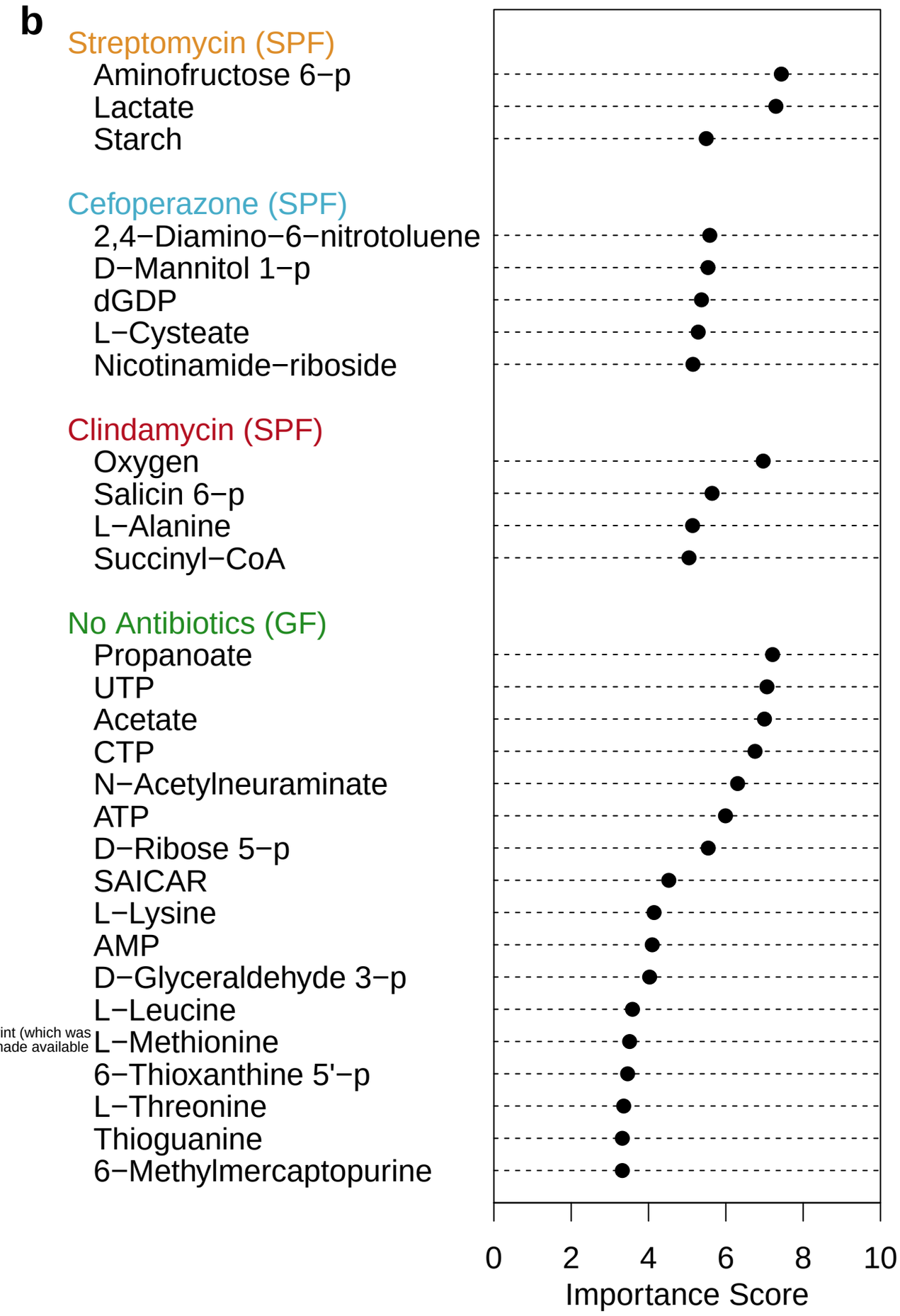



a
b
Enzyme node
Metabolite node
dAdo Aminohydrolase
7
t_o
e_i
dAdo:PO_4^{-3} Ribosyltransferase
115
t_i
e_o
Deoxyadenosine (dAdo)
Importance = 6.554
m
e_o
More likely produced
More likely consumed
–
0
+
Importance Score
94
t_i
ATP:dAdo 5'–Phosphotransferase
(i) $\mu_i = \left(\frac{\Sigma t_i}{n(e_o)}\right)$
(ii) $\mu_o = \left(\frac{\Sigma t_o}{n(e_i)}\right)$
(iii) $\text{Importance}(m) = \log_2(\mu_i - \mu_o)$

a



b



c