

# **The initial gut microbiota and response to antibiotic perturbation influence *Clostridioides difficile* colonization in mice**

Sarah Tomkovich<sup>1</sup>, Joshua M.A. Stough<sup>1</sup>, Lucas Bishop<sup>1</sup>, Patrick D. Schloss<sup>1†</sup>

† To whom correspondence should be addressed: pschloss@umich.edu

<sup>1</sup> Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109

1   **Abstract**

2   The gut microbiota has a key role in determining susceptibility to *Clostridioides difficile* infections  
3   (CDIs). However, much of the mechanistic work examining CDIs in mouse models use animals  
4   obtained from a single university colony or vendor. We treated mice from 6 different sources (2  
5   University of Michigan colonies and 4 vendors) with a single clindamycin dose, followed by a *C.*  
6   *difficile* challenge 1 day later and then measured *C. difficile* colonization levels through 9 days  
7   post-infection. The microbiota was profiled via 16S rRNA gene sequencing to examine the variation  
8   across sources and alterations due to clindamycin treatment and *C. difficile* challenge. While all  
9   sources of mice were colonized 1-day post-infection, variation emerged from days 3-7 post-infection  
10   with animals from some sources colonized with *C. difficile* for longer and at higher levels. We  
11   identified bacteria that varied in relative abundance across sources and throughout the experiment.  
12   Some bacteria were consistently impacted by clindamycin treatment in all sources of mice including  
13   *Lachnospiraceae*, *Ruminococcaceae*, and *Enterobacteriaceae*. To identify bacteria that were most  
14   important to colonization regardless of the source, we created logistic regression models that  
15   successfully classified mice based on whether they cleared *C. difficile* by 7 days post-infection using  
16   baseline, post-clindamycin, and post-infection community composition data. With these models,  
17   we identified 4 bacteria that varied across sources (*Bacteroides*), were altered by clindamycin  
18   (*Porphyromonadaceae*), or both (*Enterobacteriaceae* and *Enterococcus*). Microbiota variation  
19   across sources better emulates human interindividual variation and can help identify bacterial  
20   drivers of phenotypic variation in the context of CDIs.

21   **Importance**

22   *Clostridioides difficile* is a leading nosocomial infection. Although perturbation to the gut microbiota  
23   has been established as a key risk factor, there is variation in who becomes asymptotically  
24   colonized, develops an infection, or has an infection with adverse outcomes. *C. difficile* infection  
25   (CDI) mouse models are widely used to answer a variety of *C. difficile* pathogenesis questions.  
26   However, the inter-individual variation between mice is less than what is observed in humans,  
27   particularly if just one source of mice is used. In this study, we administered clindamycin to mice  
28   from 6 different breeding colonies and challenged them with *C. difficile*. Interestingly, only a subset

29 of the bacteria that vary across sources were associated with how long *C. difficile* was able to  
30 colonize. Future studies examining the interplay between the microbiota and *C. difficile* should  
31 consider using mice from multiple sources to better reflect human interindividual variation.

32 **Introduction**

33 Antibiotics are a common risk factor for *Clostridioides difficile* infections (CDIs), but there is variation  
34 in who goes on to develop severe or recurrent CDIs after exposure (1, 2). Additionally, asymptomatic  
35 colonization, where *C. difficile* is detectable, but symptoms are absent has been documented  
36 in infants and adults (3, 4). The intestinal microbiome has been implicated in asymptomatic  
37 colonization (5, 6), susceptibility to CDIs (7), and adverse CDI outcomes (9–12).

38 Mouse models of CDIs have been a great tool for understanding *C. difficile* pathogenesis (13). The  
39 number of CDI mouse model studies has grown substantially since Chen et al. published their  
40 C57BL/6 model in 2008, which disrupted the gut microbiota with antibiotics to enable *C. difficile*  
41 colonization and symptoms such as diarrhea and weight loss (14). CDI mouse models have been  
42 used to examine translationally relevant questions regarding *C. difficile*, including the role of the  
43 microbiota and efficacy of potential therapeutics for treating CDIs (15). However, variation in the  
44 microbiome between mice from the same breeding colony is much less than the variation observed  
45 between humans (16, 17). Additionally, studying the contribution of the microbiota to a particular  
46 disease phenotype in one set of lab mice after the same perturbation could yield a number of  
47 findings of which only a fraction may be driving the phenotype.

48 In the past, our group has attempted to introduce more microbiome variation into the CDI mouse  
49 model by using a variety of antibiotic treatments (18–21). An alternative approach to maximize  
50 microbiome variation is to use mice from multiple sources (22, 23). Microbiome differences between  
51 different mouse vendors have been well documented and shown to influence susceptibility to a  
52 variety of diseases (24, 25), including enteric infections (22, 23, 26–30). Additionally, research  
53 groups have observed different CDI outcomes in mice despite using similar models and the  
54 microbiome has been proposed as one factor potentially mediating CDI susceptibility and outcomes  
55 (13, 18, 21, 31–33). Here we examined how variations in the baseline microbiome and responses  
56 to clindamycin treatment in C57BL/6 mice from six different sources influenced susceptibility to *C.*  
57 *difficile* colonization and the time needed to clear the infection.

58 **Results**

59 **Microbiota variation is high between mice from different sources.** To test how the microbiotas  
60 of mice from different sources impact colonization dynamics after clindamycin exposure, we utilized  
61 C57BL/6 mice from 6 different sources: two colonies from the University of Michigan that were split  
62 from each other 10 years ago (the Young and Schloss lab colonies) and four commercial source:  
63 the Jackson Laboratory, Charles River Laboratories, Taconic Biosciences, and Envigo (which was  
64 formerly Harlan). These 4 vendors were chosen because they represent commonly used vendors  
65 for CDI studies in mice (26, 34–40). We used 16S rRNA gene sequencing to characterize the fecal  
66 bacterial communities across sources at baseline.

67 Since antibiotics and other risk factors of CDIs are associated with decreased microbiota diversity  
68 (41), we first examined alpha diversity across the 6 sources of mice. Initially, there was a significant  
69 difference in the number of observed operational taxonomic units (OTUs), but not Shannon diversity  
70 index ( $P_{FDR} = 0.03$  and  $P_{FDR} = 0.052$ , respectively) across sources of mice (Fig. 1A-B and Table  
71 S1). Next, we compared the community structure of mice from the 6 sources over the course of  
72 the experiment using principal coordinate analysis (PCoA) of the  $\theta_{YC}$  distances. Permutational  
73 multivariate analysis of variance (PERMANOVA) analysis revealed source and cage interactions  
74 explained most of the observed variation across fecal communities explained most of the observed  
75 community variation (combined  $R^2 = 0.90$ ,  $P = 0.0001$ , Fig. 3 and Table S2). Mice that are  
76 co-housed have been shown to have similar gut microbiotas due to coprophagy (42) and since  
77 mice within the same source were housed together, it is not surprising that cage also contributed to  
78 the observed microbiota variation. Since we conducted two experiments approximately 3 months  
79 apart, we also compared baseline communities across the 2 experiments. Experiment and cage  
80 significantly explained the observed variation only for the Schloss and Young lab mouse colonies  
81 (Fig. S2A-B and Table S2). However, most of the vendors also clustered by experiment (Fig. S2C-D,  
82 F), suggesting there was some community variation between the 2 experiments within each source,  
83 particularly for Schloss, Young, and Envigo mice (Fig. S2G-H). After finding differences at the  
84 community level, we next identified the bacteria that initially varied across sources of mice. There  
85 were 268 OTUs with relative abundances that varied across sources (Fig. 1D and Table S3). By  
86 using mice from six sources we were able to increase the number of microbiota communities to test

87 with the clindamycin *C. difficile* colonization mouse model.

88 **Clindamycin treatment renders all mice susceptible to *C. difficile* 630 colonization, but**  
89 **clearance time varies across sources.** After the acclimation period, all mice were treated with  
90 10 mg/kg clindamycin via intraperitoneal injection and one day later challenged with  $10^3$  *C. difficile*  
91 630 spores (Fig. 2A). Clindamycin was chosen because we have previously demonstrated mice  
92 are rendered susceptible, but consistently cleared *C. difficile* within 9 days (21, 43), clindamycin  
93 is frequently implicated with human CDIs (44), and is also part of the antibiotic treatment for the  
94 frequently cited 2008 CDI mouse model (14). The day after infection, *C. difficile* was detectable in all  
95 mice at a similar level (median CFU range: 2.2e+07-1.3e+08;  $P_{FDR} = 0.15$ ), indicating clindamycin  
96 rendered all mice susceptible regardless of source (Fig. 2B). Interestingly, variation in *C. difficile*  
97 CFU levels across sources of mice emerged from days 3-7 post-infection (all  $P_{FDR} \leq 0.019$ ; Fig.  
98 2B and Table S4), suggesting mouse source is associated with *C. difficile* clearance. We conducted  
99 two experiments and while the colonization dynamics were similar across most sources of mice,  
100 there was some variation between the 2 experiments, particularly for the Schloss and Envigo mice  
101 (Fig. S2A-B). Although *C. difficile* 630 causes mild symptoms in mice compared to other *C. difficile*  
102 strains (45), we also saw that weight change significantly varied across sources of mice with the  
103 most weight lost two days post-infection (Fig. 2C and Table S5). Importantly, there was also one  
104 Jackson and one Envigo mouse that died between 1- and 3-days post-infection during the second  
105 experiment. Interestingly, mice obtained from Jackson, Taconic, and Envigo tended to lose more  
106 weight, have higher *C. difficile* CFU levels and take longer to clear the infection compared to the  
107 other sources of mice (although there was variation between experiments with Schloss and Envigo  
108 mice), which was particularly evident 7 days post-infection (Fig. 2B-C, Fig. S2C-D), when 57% of  
109 the mice were still colonized with *C. difficile* (Fig. S2E). By 9 days post-infection the majority of the  
110 mice from all sources had cleared *C. difficile* (Fig. 2C) with the exception of 1 Taconic mouse from  
111 the first experiment and 2 Envigo mice from the second experiment. Thus, clindamycin rendered  
112 all mice susceptible to *C. difficile* 630 colonization, regardless of source, but there was significant  
113 variation in disease phenotype across the sources of mice.

114 **Clindamycin treatment alters bacteria in all sources, but a subset of bacterial differences**  
115 **across sources persists.** Given the well known variation in mouse microbiomes across breeding

colonies (25), we hypothesized that variation in *C. difficile* clearance would be explained by microbiota variation across the 6 sources of mice. We used 16S rRNA sequencing to characterize microbiota variation across sources after clindamycin treatment and throughout the rest of the experiment. PERMANOVA analysis revealed source was the dominant factor that explained the observed variation across fecal communities ( $R^2 = 0.35$ ,  $P = 0.0001$ ) followed by interactions between cage (mice from the same sources were housed together, primarily at a density of 2 mice per cage) and day of the experiment (Movie S1 and Table S2). Since the majority of the perturbations happened over the initial days of the experiment, we decided to focus on the communities after clindamycin treatment (day 0) and post-infection (day 1).

As expected, clindamycin treatment decreased richness and Shannon diversity across all sources of mice (Fig. 3A-B). Interestingly, significant differences in diversity metrics across sources ( $P_{FDR} < 0.05$ ) emerged after clindamycin treatment, with Charles River mice having higher richness and Shannon diversity than most of the other sources (Fig 3A-B and Table S1). The variation in community structures was also diminished by clindamycin treatment (Fig. 3C and Table S2). Next we identified the bacteria that shifted after clindamycin treatment, regardless of source by analyzing all mice that had sequence data from fecal samples collected at baseline and after clindamycin treatment. We identified 153 OTUs that were altered after clindamycin treatment (Fig. 3D and Table S6) in most mice. Despite the significant impact of clindamycin, there were still 18 OTUs (Fig. 3E and Table S3) with relative abundances that significantly varied across sources. Interestingly, when we compared the list of significant clindamycin impacted bacteria with the bacteria that varied across sources post-clindamycin, we found 4 OTUs (*Enterobacteriaceae* (OTU 1), *Lachnospiraceae* (OTU 130), *Lactobacillus* (OTU 6), *Enterococcus* (OTU 23)) overlapped (Fig. 3D-E and Tables S3, S6). These findings demonstrate that clindamycin has a consistent impact on the fecal bacterial communities of mice from all sources and only a subset of the OTUs continued to vary across sources.

**Microbiota variation across sources is maintained after *C. difficile* challenge.** Post-infection, significant differences in diversity metrics were maintained across sources ( $P_{FDR} < 0.05$ , Fig 4A-B and Table S1). Although the Charles River mice had more diverse microbiotas, the Young and Schloss lab mice were also able to clear *C. difficile* faster than some of the other sources, suggesting

145 microbiota diversity alone does not explain the observed variation in *C. difficile* colonization across  
146 sources. Source and cage interaction continued to explain most of the observed community variation  
147 (combined  $R^2 = 0.88$ , respectively;  $P = 0.0001$ ; Fig. 4D and Table S2). After *C. difficile* challenge,  
148 there were 44 OTUs (Fig. 4D and Table S3) with significantly different relative abundances across  
149 sources. In spite of the experimental perturbations that occurred, there were 12 OTUs with  
150 relative abundances that consistently varied across sources at baseline, post-clindamycin, and  
151 post-infection. Importantly, some of the OTUs that consistently varied across sources also shifted  
152 with clindamycin treatment. For example, *Proteus* increased after clindamycin treatment, but only in  
153 Taconic mice. *Enterococcus* was primarily found only in mice purchased from commercial vendors  
154 and also increased after clindamycin treatment. In summary, mouse bacterial communities varied  
155 significantly according to source throughout the course of the experiment and a consistent subset of  
156 bacterial taxa remained different across sources regardless of clindamycin and *C. difficile* challenge.

157 **Baseline, post-clindamycin, and post-infection community data can predict mice that will**  
158 **clear *C. difficile* by 7 days post-infection.** After identifying taxa that varied by source, changed  
159 after clindamycin treatment, or both, we next wanted to determine which taxa were influencing the  
160 variation in *C. difficile* colonization at day 7 (Fig. 2D, Fig. S2C). We trained three L2-regularized  
161 logistic regression models with input bacterial community data from the baseline, post-clindamycin,  
162 and post-infection timepoints of the experiment to predict *C. difficile* colonization status on day 7 (Fig.  
163 S3A-B). All models were better at predicting *C. difficile* colonization status on day 7 than random  
164 chance (all  $P \leq 1.6e-17$ ; Table S7). Interestingly, the model based on the post-clindamycin (day 0)  
165 community OTU data performed significantly better than the other models with an area under the  
166 receiving operator characteristic curve (AUROC) of 0.78 ( $P_{FDR} \leq 3.3e-06$  for pairwise comparisons;  
167 Table S7). Thus, we were able to use bacterial relative abundance data alone to differentiate mice  
168 that had cleared *C. difficile* before day 7 from the mice still colonized with *C. difficile* at that timepoint.  
169 Interestingly, the model built with OTU relative abundance data post-clindamycin treatment had the  
170 best performance, suggesting how the bacterial community responds to clindamycin treatment has  
171 the greatest influence on subsequent *C. difficile* colonization dynamics.

172 Next, to examine the bacteria that were driving each model's performance, we pulled out the top 20  
173 OTUs that had the highest absolute feature weights in each of the 6 models (Table S8). First, we

174 looked at OTUs from the model with the best performance that was based on the post-clindamycin  
175 treatment bacterial community data. While most of the 20 OTUs had low relative abundances  
176 on day 0, *Enterobacteriace*, *Bacteroides* and *Proteus* had high relative abundances in at least  
177 one source of mice and significantly varied across sources (Fig. 5A). Next, the top 20 taxa from  
178 each model were compared to the list of taxa that varied across source at the same timepoint  
179 (Fig. 1D, 3E, 4D and Table S3) and the taxa that were altered by clindamycin treatment (Fig. 3D  
180 and Table S6). We found a subset of OTUs that were important to the model and overlapped with  
181 bacteria that varied by either source, clindamycin treatment, or both (Fig. S4). Combining the  
182 overall results for the 3 models identified 14 OTUs associated with source, 21 OTUs associated  
183 with clindamycin treatment, and 6 OTUs associated with both (Fig. 5B). Several OTUs (*Bacteroides*  
184 (*OTU 2*), *Enterococcus* (*OTU 23*), *Enterobacteriaceae* (*OTU 1*), *Porphyromonadaceae* (*OTU 7*))  
185 appeared across at least 2 models, so we examined how the relative abundances of these OTUs  
186 varied over the course of the experiment (Fig. 6). Throughout the experiment, there was at least  
187 1 timepoint where relative abundances of these OTUs significantly varied across sources (Table  
188 S9). Interestingly, there were no OTUs that emerged as consistently enriched or depleted in mice  
189 that were colonized past 7 days post-infection with *C. difficile* 630, suggesting multiple bacteria  
190 influence the time needed to clear the infection. Together, these results suggest the initial bacterial  
191 communities and their responses to clindamycin influence the clearance of *C. difficile*.

## 192 Discussion

193 By running our CDI model with mice from 6 different sources, we were able to identify bacterial  
194 taxa that were unique to sources throughout the experiment as well as taxa that were universally  
195 impacted by clindamycin. We trained L2 logistic regression models with baseline, post-clindamycin  
196 treatment, and post-infection fecal community data that could predict whether mice cleared *C.*  
197 *difficile* by 7 days post-infection better than random chance. We identified *Bacteroides* (*OTU*  
198 *2*), *Enterococcus* (*OTU 23*), *Enterobacteriaceae* (*OTU 1*), *Porphyromonadaceae* (*OTU 7*) (Fig.  
199 6) as candidate bacteria within these communities that were influencing variation in *C. difficile*  
200 colonization dynamics since these bacteria were all important in the logistic regression models and  
201 varied by source, were impacted by clindamycin treatment, or both. Overall, our results demonstrate

202 clindamycin is sufficient to render mice from multiple sources susceptible to CDI and only a subset  
203 of the interindividual microbiota variation across mice from different sources was associated with  
204 the time needed to clear *C. difficile*.

205 Other studies have taken similar approaches by using mice from multiple sources to identify bacteria  
206 that either promote colonization resistance or increase susceptibility to enteric infections (22, 23,  
207 26–30). For example, in the context of *Salmonella* infections, *Enterobacteriaceae* and segmented  
208 filamentous bacteria have emerged as protective (22, 27). A previous study with *C. difficile* identified  
209 an endogenous protective *C. difficile* strain LEM1 that bloomed after antibiotic treatment in mice  
210 from Jackson or Charles River Laboratories, but not Taconic that protected mice against the more  
211有毒的 *C. difficile* VPI10463 (26). Given that we obtained mice from the same vendors, we  
212 checked all mice for endogenous *C. difficile* by plating stool samples that were collected after  
213 clindamycin treatment. However, we did not identify any endogenous *C. difficile* strains prior to  
214 challenge, suggesting there were no endogenous protective strains in the mice we received and  
215 other bacterial taxa mediated the variation in *C. difficile* colonization across sources.

216 Although all mice were susceptible to *C. difficile* colonization, mice from Jackson, Taconic, and  
217 Envigo mice tended to remain colonized through at least 7 days post-infection. We identified  
218 a subset of bacteria that were important in predicting whether a mouse was still colonized with  
219 *C. difficile* 7 days post-infection. These results suggest a subset of the bacterial community is  
220 responsible for determining the length of time needed to clear *C. difficile* colonization.

221 Differences in CDI mouse model studies have been attributed to different intestinal microbiotas  
222 of mice from different sources. For example, groups using the same clindamycin treatment and  
223 C57BL/6 mice had different *C. difficile* outcomes, one having sustained colonization (32), while  
224 the other had transient (18), despite both using *C. difficile* VPI 10643. Baseline differences in the  
225 microbiota composition have been hypothesized to partially explain the differences in colonization  
226 outcomes and overall susceptibility to *C. difficile* after treatment with the same antibiotic (13,  
227 31). The bacterial perturbations induced by clindamycin treatment have been well characterized  
228 and our findings agree with previous CDI mouse model work demonstrating *Enterococcus* and  
229 *Enterobacteriaceae* were associated with *C. difficile* susceptibility and *Porphyromonadaceae*,

230 *Lachnospiraceae*, *Ruminococcaceae*, and *Turicibacter* were associated with resistance (19, 21, 32,  
231 33, 43, 46–48). While we have demonstrated that susceptibility is uniform across sources of mice  
232 after clindamycin treatment, there could be different outcomes for either susceptibility or clearance  
233 in the case of other antibiotic treatments. The *C. difficile* strain used could also be contributing to  
234 the variation in *C. difficile* outcomes seen across different research groups. For example, a group  
235 found differential colonization outcomes after clindamycin treatment, with *C. difficile* 630 and M68  
236 eventually becoming undetectable while BI-7 remained detectable up to 70 days post-treatment (47).  
237 We found the time needed to naturally clear *C. difficile* varied across sources of mice implying that  
238 at least in the context of the same perturbation, microbiota differences seemed to influence infection  
239 outcome more than susceptibility. More importantly, we were able to reduce the variation observed  
240 across sources to identify a subset of OTUs that were also important for predicting *C. difficile*  
241 colonization status 7 days post-infection. Since all but 3 mice eventually cleared *C. difficile* 630 by 9  
242 days post-infection and the model built with the post-clindamycin OTU relative abundance data had  
243 the best performance, our results suggest clindamycin treatment had a large role in determining *C.*  
244 *difficile* susceptibility and clearance than the source of the mice.

245 Our approach successfully increased the diversity of murine bacterial communities tested in  
246 our clindamycin *C. difficile* model. One alternative approach that has been used in some CDI  
247 studies (49–54) is to associate mice with human microbiotas. However, a major caveat to this  
248 method is the substantial loss of human microbiota community members upon transfer to mice  
249 (55, 56). Additionally with the exception of 2 recent studies (49, 50), most of the CDI mouse model  
250 studies to date associated mice with just 1 types of human microbiota either from a single donor  
251 or a single pool from multiple donors (51–54), which does not aid in the goal of modeling the  
252 interpersonal variation seen in humans to understand how the microbiota influences susceptibility to  
253 CDIs and adverse outcomes. Importantly, our study using mice from 6 different sources increased  
254 the variation between groups of mice compared to using 1 source alone, to better reflect the  
255 inter-individual microbiota variation observed in humans. Encouragingly, decreased *Bifidobacterium*,  
256 *Porphyromonas*, *Ruminococcaceae* and *Lachnospiraceae* and increased *Enterobacteriaceae*,  
257 *Enterococcus*, *Lactobacillus*, and *Proteus* have all been associated with human CDIs (7) and were  
258 well represented in our study, suggesting most of the mouse sources are suitable for gaining insights

259 into microbiota associated factors influencing *C. difficile* colonization and infections in humans. An  
260 important exception was *Enterococcus*, which was primarily absent from the mice from University  
261 of Michigan colonies and *Proteus*, which was only found in Taconic mice. Importantly, the fact that  
262 some CDI-associated bacteria were only found in a subset of mice has important implications for  
263 future CDI mouse model studies.

264 There are several limitations to our work. The microbiome is composed of viruses, fungi, and  
265 parasites in addition to bacteria, and these non-bacterial members can also vary across mouse  
266 sources (57, 58). While our study focused solely on the bacterial portion, viruses and fungi have  
267 also begun to be implicated in the context of CDIs or FMT treatments for recurrent CDIs (35, 59–62).  
268 Beyond community composition, the metabolic function of the microbiota also has a CDI signature  
269 (20, 48, 63, 64) and can vary across mice from different sources (65). For example, microbial  
270 metabolites, particularly secondary bile acids and butyrate production, have been implicated as  
271 important contributors to *C. difficile* resistance (33, 47). Although, we only looked at composition,  
272 *Ruminococcaceae* and *Lachnospiraceae* both emerged as important taxa for classifying day 7 *C.*  
273 *difficile* colonization status and metagenomes from these bacteria have been shown to contain  
274 the bile acid-inducible gene cluster necessary for secondary bile acid formation and ability to  
275 produce butyrate (52, 66). Interestingly, butyrate has previously been shown to vary across vendors  
276 and mediated resistance to *Citrobacter rodentium* infection, a model of enterohemorrhagic and  
277 enteropathogenic *Escherichia coli* infections (23). Evidence for immunological toning differences  
278 in IgA and Th17 cells across mice from different vendors have also been documented and (67,  
279 68) may also influence response to CDI, particularly in the context of severe CDIs (69, 70). The  
280 outcome after *C. difficile* exposure depends on a multitude of factors, including age, diet, and  
281 immunity; all of which are also influenced by the microbiota.

282 We have demonstrated that the ways baseline microbiotas from different mouse sources respond  
283 to clindamycin treatment influences the length of time mice remained colonized with *C. difficile* 630.  
284 For those interested in dissecting the contribution of the microbiome to *C. difficile* pathogenesis  
285 and treatments, using multiple sources of mice may yield more insights than a single model alone.  
286 Furthermore, for studies wanting to examine the interplay between a particular bacterial taxon such  
287 as *Enterococcus* and *C. difficile*, these results could serve as a resource for selecting which mice

288 to order to address the question. Using mice from multiple sources helps model the interpersonal  
289 microbiota variation among humans to aid our understanding of how the gut microbiota contributes  
290 to CDIs.

291 **Acknowledgements**

292 This work was supported by the National Institutes of Health (U01AI124255). ST was supported by  
293 the Michigan Institute for Clinical and Health Research Postdoctoral Translation Scholars Program  
294 (UL1TR002240). We thank members of the Schloss lab for feedback on planning the experiments  
295 and data presentation, as well as code tutorials and feedback through Code club. In particular, we  
296 want to thank Begüm Topçuoğlu for help with implementing L2 logistic regression models using her  
297 pipeline, Ana Taylor for help with media preparation and sample collection, and Nicholas Lesniak for  
298 his critical feedback on the manuscript. We also thank members of Vincent Young's lab, particularly  
299 Kimberly Vendrov, for guidance with the *C. difficile* infection mouse model and donating the mice.  
300 We also want to thank the Unit for Laboratory Animal Medicine at the University of Michigan for  
301 maintaining our mouse colony and providing the institutional support for our mouse experiments.  
302 Finally, we thank Kwi Kim, Austin Campbell, and Kimberly Vendrov for their help in maintaining the  
303 Schloss lab's anaerobic chamber.

304 **Materials and Methods**

305 **(i) Animals.** All experiments were approved by the University of Michigan Animal Care and Use  
306 Committee (IACUC) under protocol number PRO00006983. Female C57BL/7 mice were obtained  
307 from 6 different sources: The Jackson Laboratory, Charles River Laboratories, Taconic Biosciences,  
308 Envigo, and two colonies at the University of Michigan (the Schloss lab colony and the Young lab  
309 colony). The Young lab colony was originally established with mice purchased from Jackson, and  
310 the Schloss lab colony was established 10 years ago with mice donated from the Young lab. The 4  
311 groups of mice purchased from vendors were allowed to acclimate to the University of Michigan  
312 mouse facility for 13 days prior to starting the experiment. At least 4 female mice (age 5-10 weeks)  
313 were obtained per source and mice from the same source were primarily housed at a density of 2  
314 mice per cage. The experiment was repeated once, approximately 3 months after the start of the  
315 first experiment.

316 **(ii) Antibiotic treatment.** After the 13-day acclimation period and 1 day prior to challenge (Fig.  
317 1A), all mice received 10 mg/kg clindamycin (filter sterilized through a 0.22 micron syringe filter  
318 prior to administration) via intraperitoneal injection.

319 **(iii) *C. difficile* infection model.** Mice were challenged with  $10^3$  spores of *C. difficile* strain 630  
320 via oral gavage post-infection 1 day after clindamycin treatment as described previously (21). Mice  
321 weights and stool samples were taken daily through 9 days post-challenge. Collected stool was  
322 split for *C. difficile* CFU quantification and 16S rRNA sequencing analysis. *C. difficile* quantification  
323 stool samples were transferred to the anaerobic chamber, serially diluted in PBS, plated on  
324 taurocholate-cycloserine-cefoxitin-fructose agar (TCCFA) plates, and counted after 24 hours of  
325 incubation at 37°C under anaerobic conditions. A sample from the day 0 timepoint (post-clindamycin  
326 and prior to *C. difficile* challenge) was also plated on TCCFA to ensure mice were not already  
327 colonized with *C. difficile* prior to infection. There were 3 deaths recorded over the course of the  
328 experiment, 1 Taconic mouse died prior to *C. difficile* challenge and 1 Jackson and 1 Envigo mouse  
329 died between 1- and 3-days post-infection. Mice were categorized as cleared when no *C. difficile*  
330 was detected in the first serial dilution (limit of detection: 100 CFU). Stool samples for 16S rRNA  
331 sequencing were snap frozen in liquid nitrogen and stored at -80°C until DNA extraction.

332 **(iv) 16S rRNA sequencing.** DNA was extracted from -80 °C stored stool samples using the DNeasy  
333 Powersoil HTP 96 kit (Qiagen) and an EpMotion 5075 automated pipetting system (Eppendorf).  
334 The V4 region was amplified for 16S rRNA with the AccuPrime Pfx DNA polymerase (Thermo  
335 Fisher Scientific) using custom barcoded primers, as previously described (71). The ZymoBIOMICS  
336 microbial community DNA standards was used as a mock community control (72) and water was  
337 used as a negative control per 96-well extraction plate. The PCR amplicons were cleaned up  
338 and normalized with the SequalPrep normalization plate kit (Thermo Fisher Scientific). Amplicons  
339 were pooled and quantified with the KAPA library quantification kit (KAPA biosystems), prior to  
340 sequencing using the MiSeq system (Illumina).

341 **(v) 16S rRNA gene sequence analysis.** mothur (v. 1.43) was used to process all sequences  
342 (73) with a previously published protocol (71). Reads were combined and aligned with the SILVA  
343 reference database (74). Chimeras were removed with the VSEARCH algorithm and taxonomic  
344 assignment was completed with a modified version (v16) of the Ribosomal Database Project  
345 reference database (v11.5) (75) with an 80% cutoff. Operational taxonomic units (OTUs) were  
346 assigned with a 97% similarity threshold using the opticlus algorithm (76). To account for uneven  
347 sequencing across samples, samples were rarefied to 5,437 sequences 1,000 times for alpha and  
348 beta diversity analyses. PCoAs were generated based on  $\theta_{YC}$  distances. Permutational multivariate  
349 analysis of variance (PERMANOVA) was performed on mothur-generated  $\theta_{YC}$  distance matrices  
350 with the adonis function in the vegan package (77) in R (78).

351 **(vi) Classification model training and evaluation.** Models were generated based on mice that  
352 were categorized as either cleared or colonized 7 days post-infection and had sequencing data  
353 from the baseline (day -1), post-clindamycin (day 0), and post-infection (day 1) timepoints of  
354 the experiment. Input bacterial community relative abundance data at the OTU level from the  
355 baseline, post-clindamycin, and post-infection timepoints was used to generate 6 classification  
356 models that predicted *C. difficile* colonization status 7 days post-infection. The L2-regularized  
357 logistic regression models were trained and tested using the caret package (79) in R as previously  
358 described (80) with the exception that we used 60% training and 40% testing data splits for the  
359 cross-validation of the training data to select the best cost hyperparameter and the testing of  
360 the held out test data to measure model performance. The modified training to testing ratio was

361 selected to accommodate the small number of samples in the dataset. Code was modified from  
362 [https://github.com/SchlossLab/ML\\_pipeline\\_microbiome](https://github.com/SchlossLab/ML_pipeline_microbiome) to update the classification outcomes and  
363 change the data split ratios. The modified repository to regenerate this analysis is available at  
364 [https://github.com/tomkosev/ML\\_pipeline\\_microbiome](https://github.com/tomkosev/ML_pipeline_microbiome).

365 **(vii) Statistical analysis.** All statistical tests were performed in R (v 3.5.2) (78). The Kruskal-Wallis  
366 test was used to analyze differences in *C. difficile* CFU, mouse weight change, and alpha  
367 diversity across sources with a Benjamini-Hochberg correction for testing multiple timepoints,  
368 followed by pairwise Wilcoxon comparisons with Benjamini-Hochberg correction. For taxonomic  
369 analysis and generation of logistic regression model input data, *C. difficile* (OTU 20) was removed.  
370 Bacterial relative abundances that varied across sources at the OTU level were identified with the  
371 Kruskal-Wallis test with Benjamini-Hochberg correction for testing all identified OTUs, followed  
372 by pairwise Wilcoxon comparisons with Benjamini-Hochberg correction. OTUs impacted by  
373 clindamycin treatment were identified using the Wilcoxon signed rank test with matched pairs  
374 of mice samples for day -1 and day 0. To determine whether classification models had better  
375 performance (test AUROCs) than random chance (0.5), we used the one-sample Wilcoxon signed  
376 rank test. To examine whether there was an overall difference in predictive performance across the  
377 6 classification models we used the Kruskal-Wallis test followed by pairwise Wilcoxon comparisons  
378 with Benjamini-Hochberg correction for multiple hypothesis testing. The tidyverse package was  
379 used to wrangle and graph data (v 1.3.0) (81).

380 **(viii) Code availability.** Code for all data analysis and generating this manuscript is available at  
381 [https://github.com/SchlossLab/Tomkovich\\_Vendor\\_XXXX\\_2020](https://github.com/SchlossLab/Tomkovich_Vendor_XXXX_2020).

382 **(ix) Data availability.** The 16S rRNA sequencing data have been deposited in the National Center  
383 for Biotechnology Information Sequence Read Archive (BioProject Accession no. PRJNA608529).

384 **References**

- 385 1. Teng C, Reveles KR, Obodozie-Ofoegbu OO, Frei CR. 2019. *Clostridium difficile* infection  
386 risk with important antibiotic classes: An analysis of the FDA adverse event reporting system.  
387 International Journal of Medical Sciences 16:630–635.
- 388 2. Kelly C. 2012. Can we identify patients at high risk of recurrent *Clostridium difficile* infection?  
389 Clinical Microbiology and Infection 18:21–27.
- 390 3. Zacharioudakis IM, Zervou FN, Pliakos EE, Ziakas PD, Mylonakis E. 2015. Colonization with  
391 toxinogenic *C. difficile* upon hospital admission, and risk of infection: A systematic review and  
392 meta-analysis. American Journal of Gastroenterology 110:381–390.
- 393 4. Crobach MJT, Vernon JJ, Loo VG, Kong LY, Péchiné S, Wilcox MH, Kuijper EJ. 2018.  
394 Understanding *Clostridium difficile* colonization. Clinical Microbiology Reviews 31.
- 395 5. Zhang L, Dong D, Jiang C, Li Z, Wang X, Peng Y. 2015. Insight into alteration of gut microbiota  
396 in *Clostridium difficile* infection and asymptomatic *c. difficile* colonization. Anaerobe 34:1–7.
- 397 6. VanInsberghe D, Elsherbini JA, Varian B, Poutahidis T, Erdman S, Polz MF. 2020. Diarrhoeal  
398 events can trigger long-term *Clostridium difficile* colonization with recurrent blooms. Nature  
399 Microbiology 5:642–650.
- 400 7. Mancabelli L, Milani C, Lugli GA, Turroni F, Cocconi D, Sinderen D van, Ventura M. 2017.  
401 Identification of universal gut microbial biomarkers of common human intestinal diseases by  
402 meta-analysis. FEMS Microbiology Ecology 93.
- 403 8. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. 2017. Meta-analysis of gut microbiome  
404 studies identifies disease-specific and shared responses. Nature Communications 8.
- 405 9. Seekatz AM, Rao K, Santhosh K, Young VB. 2016. Dynamics of the fecal microbiome in patients  
406 with recurrent and nonrecurrent *Clostridium difficile* infection. Genome Medicine 8.
- 407 10. Khanna S, Montassier E, Schmidt B, Patel R, Knights D, Pardi DS, Kashyap PC. 2016. Gut  
408 microbiome predictors of treatment response and recurrence in primary *Clostridium difficile* infection.

- 409 Alimentary Pharmacology & Therapeutics 44:715–727.
- 410 11. Pakpour S, Bhanvadia A, Zhu R, Amarnani A, Gibbons SM, Gurry T, Alm EJ, Martello LA. 2017.  
411 Identifying predictive features of *Clostridium difficile* infection recurrence before, during, and after  
412 primary antibiotic treatment. Microbiome 5.
- 413 12. Lee AA, Rao K, Limsrivilai J, Gilliland M, Malamet B, Briggs E, Young VB, Higgins PDR. 2020.  
414 Temporal gut microbial changes predict recurrent *Clostridioides difficile* infection in patients with  
415 and without ulcerative colitis. Inflammatory Bowel Diseases.
- 416 13. Hutton ML, Mackin KE, Chakravorty A, Lyras D. 2014. Small animal models for the study of  
417 *Clostridium difficile* disease pathogenesis. FEMS Microbiology Letters 352:140–149.
- 418 14. Chen X, Katchar K, Goldsmith JD, Nanthakumar N, Cheknis A, Gerding DN, Kelly CP. 2008. A  
419 mouse model of *Clostridium difficile*-associated disease. Gastroenterology 135:1984–1992.
- 420 15. Best EL, Freeman J, Wilcox MH. 2012. Models for the study of *Clostridium difficile* infection.  
421 Gut Microbes 3:145–167.
- 422 16. Baxter NT, Wan JJ, Schubert AM, Jenior ML, Myers P, Schloss PD. 2014. Intra- and  
423 interindividual variations mask interspecies variation in the microbiota of sympatric peromyscus  
424 populations. Applied and Environmental Microbiology 81:396–404.
- 425 17. Nagpal R, Wang S, Woods LCS, Seshie O, Chung ST, Shively CA, Register TC, Craft S,  
426 McClain DA, Yadav H. 2018. Comparative microbiome signatures and short-chain fatty acids in  
427 mouse, rat, non-human primate, and human feces. Frontiers in Microbiology 9.
- 428 18. Reeves AE, Theriot CM, Bergin IL, Huffnagle GB, Schloss PD, Young VB. 2011. The interplay  
429 between microbiome dynamics and pathogen dynamics in a murine model of *Clostridium difficile*  
430 infection 2:145–158.
- 431 19. Schubert AM, Sinani H, Schloss PD. 2015. Antibiotic-induced alterations of the murine gut  
432 microbiota and subsequent effects on colonization resistance against *Clostridium difficile*. mBio 6.
- 433 20. Jenior ML, Leslie JL, Young VB, Schloss PD. 2017. *Clostridium difficile* colonizes alternative

- 434 nutrient niches during infection across distinct murine gut microbiomes. *mSystems* 2.
- 435 21. Jenior ML, Leslie JL, Young VB, Schloss PD. 2018. *Clostridium difficile* alters the structure and  
436 metabolism of distinct cecal microbiomes during initial infection to promote sustained colonization.  
437 *mSphere* 3.
- 438 22. Velazquez EM, Nguyen H, Heasley KT, Saechao CH, Gil LM, Rogers AWL, Miller BM, Rolston  
439 MR, Lopez CA, Litvak Y, Liou MJ, Faber F, Bronner DN, Tiffany CR, Byndloss MX, Byndloss  
440 AJ, Bäumler AJ. 2019. Endogenous Enterobacteriaceae underlie variation in susceptibility to  
441 *Salmonella* infection. *Nature Microbiology* 4:1057–1064.
- 442 23. Osbelt L, Thiemann S, Smit N, Lesker TR, Schröter M, Gálvez EJC, Schmidt-Hohagen K, Pils  
443 MC, Mühlen S, Dersch P, Hiller K, Schlüter D, Neumann-Schaal M, Strowig T. 2020. Variations in  
444 microbiota composition of laboratory mice influence *Citrobacter rodentium* infection via variable  
445 short-chain fatty acid production. *PLOS Pathogens* 16:e1008448.
- 446 24. Stough JMA, Dearth SP, Denny JE, LeCleir GR, Schmidt NW, Campagna SR, Wilhelm SW.  
447 2016. Functional characteristics of the gut microbiome in C57BL/6 mice differentially susceptible to  
448 *Plasmodium yoelii*. *Frontiers in Microbiology* 7.
- 449 25. Alegre M-L. 2019. Mouse microbiomes: Overlooked culprits of experimental variability. *Genome*  
450 *Biology* 20.
- 451 26. Etienne-Mesmin L, Chassaing B, Adekunle O, Mattei LM, Bushman FD, Gewirtz AT. 2017.  
452 Toxin-positive *Clostridium difficile* latently infect mouse colonies and protect against highly  
453 pathogenic *C. difficile*. *Gut* 67:860–871.
- 454 27. Lai NY, Musser MA, Pinho-Ribeiro FA, Baral P, Jacobson A, Ma P, Potts DE, Chen Z, Paik D,  
455 Soualhi S, Yan Y, Misra A, Goldstein K, Lagomarsino VN, Nordstrom A, Sivanathan KN, Wallrapp A,  
456 Kuchroo VK, Nowarski R, Starnbach MN, Shi H, Surana NK, An D, Wu C, Huh JR, Rao M, Chiu IM.  
457 2020. Gut-innervating nociceptor neurons regulate peyer's patch microfold cells and SFB levels to  
458 mediate *Salmonella* host defense. *Cell* 180:33–49.e22.
- 459 28. Thiemann S, Smit N, Roy U, Lesker TR, Gálvez EJ, Helmecke J, Basic M, Bleich A, Goodman

- 460 AL, Kalinke U, Flavell RA, Erhardt M, Strowig T. 2017. Enhancement of IFNgamma production by  
461 distinct commensals ameliorates *Salmonella*-induced disease. *Cell Host & Microbe* 21:682–694.e5.
- 462 29. Rolig AS, Cech C, Ahler E, Carter JE, Ottemann KM. 2013. The degree of *Helicobacter*  
463 *pylori*-triggered inflammation is manipulated by preinfection host microbiota. *Infection and Immunity*  
464 81:1382–1389.
- 465 30. Ge Z, Sheh A, Feng Y, Muthupalani S, Ge L, Wang C, Kurnick S, Mannion A, Whary MT, Fox  
466 JG. 2018. *Helicobacter pylori*-infected C57BL/6 mice with different gastrointestinal microbiota have  
467 contrasting gastric pathology, microbial and host immune responses. *Scientific Reports* 8.
- 468 31. Lawley TD, Young VB. 2013. Murine models to study *Clostridium difficile* infection and  
469 transmission. *Anaerobe* 24:94–97.
- 470 32. Buffie CG, Jarchum I, Equinda M, Lipuma L, Gobourne A, Viale A, Ubeda C, Xavier J, Pamer  
471 EG. 2011. Profound alterations of intestinal microbiota following a single dose of clindamycin results  
472 in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infection and Immunity* 80:62–73.
- 473 33. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew  
474 M, Viale A, Littmann E, Brink MRM van den, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint  
475 NC, Xavier JB, Pamer EG. 2014. Precision microbiome reconstitution restores bile acid mediated  
476 resistance to *Clostridium difficile*. *Nature* 517:205–208.
- 477 34. Spinler JK, Brown A, Ross CL, Boonma P, Conner ME, Savidge TC. 2016. Administration of  
478 probiotic kefir to mice with *Clostridium difficile* infection exacerbates disease. *Anaerobe* 40:54–57.
- 479 35. Markey L, Shaban L, Green ER, Lemon KP, Mecsas J, Kumamoto CA. 2018. Pre-colonization  
480 with the commensal fungus candida albicans reduces murine susceptibility to *Clostridium difficile*  
481 infection. *Gut Microbes* 1–13.
- 482 36. McKee RW, Aleksanyan N, Garrett EM, Tamayo R. 2018. Type IV pili promote *Clostridium*  
483 *difficile* adherence and persistence in a mouse model of infection. *Infection and Immunity* 86.
- 484 37. Yamaguchi T, Konishi H, Aoki K, Ishii Y, Chono K, Tateda K. 2020. The gut microbiome diversity

- 485 of *Clostridioides difficile*-inoculated mice treated with vancomycin and fidaxomicin. Journal of  
486 Infection and Chemotherapy 26:483–491.
- 487 38. Stroke IL, Letourneau JJ, Miller TE, Xu Y, Pechik I, Savoly DR, Ma L, Sturzenbecker LJ,  
488 Sabalski J, Stein PD, Webb ML, Hilbert DW. 2018. Treatment of *Clostridium difficile* infection  
489 with a small-molecule inhibitor of toxin UDP-glucose hydrolysis activity. Antimicrobial Agents and  
490 Chemotherapy 62.
- 491 39. Quigley L, Coakley M, Alemayehu D, Rea MC, Casey PG, O'Sullivan, Murphy E, Kiely B, Cotter  
492 PD, Hill C, Ross RP. 2019. *Lactobacillus gasseri* APC 678 reduces shedding of the pathogen  
493 *Clostridium difficile* in a murine model. Frontiers in Microbiology 10.
- 494 40. Mullish BH, McDonald JAK, Pechlivanis A, Allegretti JR, Kao D, Barker GF, Kapila D, Petrof  
495 EO, Joyce SA, Gahan CGM, Glegola-Madejska I, Williams HRT, Holmes E, Clarke TB, Thursz  
496 MR, Marchesi JR. 2019. Microbial bile salt hydrolases mediate the efficacy of faecal microbiota  
497 transplant in the treatment of recurrent *Clostridioides difficile* infection. Gut 68:1791–1800.
- 498 41. Ross CL, Spinler JK, Savidge TC. 2016. Structural and functional changes within the gut  
499 microbiota and susceptibility to *Clostridium difficile* infection. Anaerobe 41:37–43.
- 500 42. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. 2015. How informative is the mouse for human  
501 gut microbiota research? Disease Models & Mechanisms 8:1–16.
- 502 43. Tomkovich S, Lesniak NA, Li Y, Bishop L, Fitzgerald MJ, Schloss PD. 2019. The proton  
503 pump inhibitor omeprazole does not promote *Clostridioides difficile* colonization in a murine model.  
504 mSphere 4.
- 505 44. Guh AY, Kutty PK. 2018. *Clostridioides difficile* infection 169:ITC49.
- 506 45. Theriot CM, Koumpouras CC, Carlson PE, Bergin II, Aronoff DM, Young VB. 2011.  
507 Cefoperazone-treated mice as an experimental platform to assess differential virulence of  
508 *Clostridium difficile* strains. Gut Microbes 2:326–334.
- 509 46. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, Mastroeni P, Scott P,

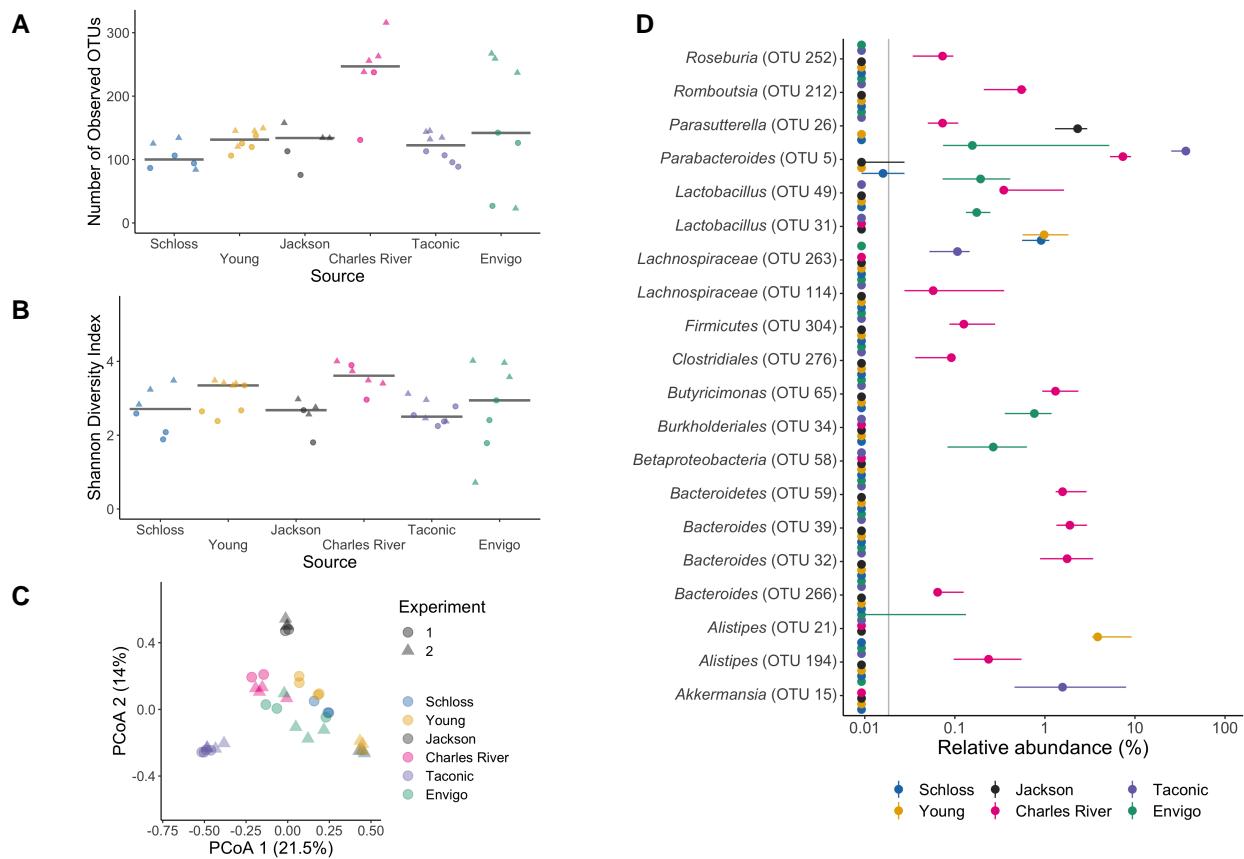
- 510 Raisen C, Mottram L, Fairweather NF, Wren BW, Parkhill J, Dougan G. 2009. Antibiotic treatment of  
511 *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and  
512 severe disease in immunocompromised hosts. *Infection and Immunity* 77:3661–3669.
- 513 47. Lawley TD, Clare S, Walker AW, Stares MD, Connor TR, Raisen C, Goulding D, Rad R,  
514 Schreiber F, Brandt C, Deakin LJ, Pickard DJ, Duncan SH, Flint HJ, Clark TG, Parkhill J, Dougan  
515 G. 2012. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy  
516 resolves relapsing *Clostridium difficile* disease in mice. *PLoS Pathogens* 8:e1002995.
- 517 48. Jump RLP, Polinkovsky A, Hurless K, Sitzlar B, Eckart K, Tomas M, Deshpande A, Nerandzic  
518 MM, Donskey CJ. 2014. Metabolomics analysis identifies intestinal microbiota-derived biomarkers  
519 of colonization resistance in clindamycin-treated mice. *PLoS ONE* 9:e101267.
- 520 49. Nagao-Kitamoto H, Leslie JL, Kitamoto S, Jin C, Thomsson KA, Gilliland MG, Kuffa P, Goto Y,  
521 Jenq RR, Ishii C, Hirayama A, Seekatz AM, Martens EC, Eaton KA, Kao JY, Fukuda S, Higgins PDR,  
522 Karlsson NG, Young VB, Kamada N. 2020. Interleukin-22-mediated host glycosylation prevents  
523 *Clostridioides difficile* infection by modulating the metabolic activity of the gut microbiota. *Nature  
524 Medicine* 26:608–617.
- 525 50. Battaglioli EJ, Hale VL, Chen J, Jeraldo P, Ruiz-Mojica C, Schmidt BA, Rekdal VM, Till LM, Huq  
526 L, Smits SA, Moor WJ, Jones-Hall Y, Smyrk T, Khanna S, Pardi DS, Grover M, Patel R, Chia N,  
527 Nelson H, Sonnenburg JL, Farrugia G, Kashyap PC. 2018. *Clostridioides difficile* uses amino acids  
528 associated with gut microbial dysbiosis in a subset of patients with diarrhea. *Science Translational  
529 Medicine* 10:eaam7019.
- 530 51. Robinson CD, Auchtung JM, Collins J, Britton RA. 2014. Epidemic *Clostridium difficile* strains  
531 demonstrate increased competitive fitness compared to nonepidemic isolates. *Infection and  
532 Immunity* 82:2815–2825.
- 533 52. Collins J, Auchtung JM, Schaefer L, Eaton KA, Britton RA. 2015. Humanized microbiota mice  
534 as a model of recurrent *Clostridium difficile* disease. *Microbiome* 3.
- 535 53. Collins J, Robinson C, Danhof H, Knetsch CW, Leeuwen HC van, Lawley TD, Auchtung JM,

- 536 Britton RA. 2018. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. Nature  
537 553:291–294.
- 538 54. Hryckowian AJ, Treuren WV, Smits SA, Davis NM, Gardner JO, Bouley DM, Sonnenburg JL.  
539 2018. Microbiota-accessible carbohydrates suppress *Clostridium difficile* infection in a murine  
540 model. *Nature Microbiology* 3:662–669.
- 541 55. Fouladi F, Glenny EM, Bulik-Sullivan EC, Tsilimigras MCB, Sioda M, Thomas SA, Wang Y, Djukic  
542 Z, Tang Q, Tarantino LM, Bulik CM, Fodor AA, Carroll IM. 2020. Sequence variant analysis reveals  
543 poor correlations in microbial taxonomic abundance between humans and mice after gnotobiotic  
544 transfer. *The ISME Journal*.
- 545 56. Walter J, Armet AM, Finlay BB, Shanahan F. 2020. Establishing or exaggerating causality for  
546 the gut microbiome: Lessons from human microbiota-associated rodents. *Cell* 180:221–232.
- 547 57. Rasmussen TS, Vries L de, Kot W, Hansen LH, Castro-Mejía JL, Vogensen FK, Hansen AK,  
548 Nielsen DS. 2019. Mouse vendor influence on the bacterial and viral gut composition exceeds the  
549 effect of diet. *Viruses* 11:435.
- 550 58. Mims TS, Abdallah QA, Watts S, White C, Han J, Willis KA, Pierre JF. 2020. Variability in  
551 interkingdom gut microbiomes between different commercial vendors shapes fat gain in response  
552 to diet. *The FASEB Journal* 34:1–1.
- 553 59. Stewart DB, Wright JR, Fowler M, McLimans CJ, Tokarev V, Amaniera I, Baker O, Wong H-T,  
554 Brabec J, Drucker R, Lamendella R. 2019. Integrated meta-omics reveals a fungus-associated  
555 bacteriome and distinct functional pathways in *Clostridioides difficile* infection. *mSphere* 4.
- 556 60. Ott SJ, Waetzig GH, Rehman A, Moltzau-Anderson J, Bharti R, Grasis JA, Cassidy L, Tholey A,  
557 Fickenscher H, Seegert D, Rosenstiel P, Schreiber S. 2017. Efficacy of sterile fecal filtrate transfer  
558 for treating patients with *Clostridium difficile* infection. *Gastroenterology* 152:799–811.e7.
- 559 61. Zuo T, Wong SH, Lam K, Lui R, Cheung K, Tang W, Ching JYL, Chan PKS, Chan MCW, Wu  
560 JCY, Chan FKL, Yu J, Sung JJY, Ng SC. 2017. Bacteriophage transfer during faecal microbiota  
561 transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut*

- 562 gutjnl–2017–313952.
- 563 62. Zuo T, Wong SH, Cheung CP, Lam K, Lui R, Cheung K, Zhang F, Tang W, Ching JYL, Wu JCY,  
564 Chan PKS, Sung JJY, Yu J, Chan FKL, Ng SC. 2018. Gut fungal dysbiosis correlates with reduced  
565 efficacy of fecal microbiota transplantation in *Clostridium difficile* infection. Nature Communications  
566 9.
- 567 63. Robinson JI, Weir WH, Crowley JR, Hink T, Reske KA, Kwon JH, Burnham C-AD, Dubberke  
568 ER, Mucha PJ, Henderson JP. 2019. Metabolomic networks connect host-microbiome processes to  
569 human *Clostridioides difficile* infections. Journal of Clinical Investigation 129:3792–3806.
- 570 64. Fletcher JR, Erwin S, Lanzas C, Theriot CM. 2018. Shifts in the gut metabolome and *Clostridium*  
571 *difficile* transcriptome throughout colonization and infection in a mouse model. mSphere 3.
- 572 65. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, Li X, Long H, Zhang J, Zhang D, Liu C, Fang  
573 Z, Chou J, Glanville J, Hao Q, Kotowska D, Colding C, Licht TR, Wu D, Yu J, Sung JJY, Liang Q, Li  
574 J, Jia H, Lan Z, Tremaroli V, Dworzynski P, Nielsen HB, Bäckhed F, Doré J, Chatelier EL, Ehrlich  
575 SD, Lin JC, Arumugam M, Wang J, Madsen L, Kristiansen K. 2015. A catalog of the mouse gut  
576 metagenome. Nature Biotechnology 33:1103–1108.
- 577 66. Vital M, Rud T, Rath S, Pieper DH, Schlüter D. 2019. Diversity of bacteria exhibiting bile  
578 acid-inducible 7alpha-dehydroxylation genes in the human gut. Computational and Structural  
579 Biotechnology Journal 17:1016–1019.
- 580 67. Fransen F, Zagato E, Mazzini E, Fosso B, Manzari C, Aidy SE, Chiavelli A, D'Erchia AM,  
581 Sethi MK, Pabst O, Marzano M, Moretti S, Romani L, Penna G, Pesole G, Rescigno M. 2015.  
582 BALB/c and C57BL/6 mice differ in polyreactive IgA abundance, which impacts the generation of  
583 antigen-specific IgA and microbiota diversity. Immunity 43:527–540.
- 584 68. Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee  
585 CA, Lynch SV, Tanoue T, Imaoka A, Itoh K, Takeda K, Umesaki Y, Honda K, Littman DR. 2009.  
586 Induction of intestinal th17 cells by segmented filamentous bacteria. Cell 139:485–498.
- 587 69. Azrad M, Hamo Z, Tkhawkho L, Peretz A. 2018. Elevated serum immunoglobulin a levels in

- 588 patients with *Clostridium difficile* infection are associated with mortality. *Pathogens and Disease* 76.
- 589 70. Saleh MM, Frisbee AL, Leslie JL, Buonomo EL, Cowardin CA, Ma JZ, Simpson ME, Scully KW,  
590 Abhyankar MM, Petri WA. 2019. Colitis-induced th17 cells increase the risk for severe subsequent  
591 *Clostridium difficile* infection. *Cell Host & Microbe* 25:756–765.e5.
- 592 71. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a  
593 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the  
594 MiSeq illumina sequencing platform. *Applied and Environmental Microbiology* 79:5112–5120.
- 595 72. Sze MA, Schloss PD. 2019. The impact of DNA polymerase and number of rounds of  
596 amplification in PCR on 16S rRNA gene sequence data. *mSphere* 4.
- 597 73. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,  
598 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF.  
599 2009. Introducing mothur: Open-source, platform-independent, community-supported software  
600 for describing and comparing microbial communities. *Applied and Environmental Microbiology*  
601 75:7537–7541.
- 602 74. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012.  
603 The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools.  
604 *Nucleic Acids Research* 41:D590–D596.
- 605 75. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR,  
606 Tiedje JM. 2013. Ribosomal database project: Data and tools for high throughput rRNA analysis.  
607 *Nucleic Acids Research* 42:D633–D642.
- 608 76. Westcott SL, Schloss PD. 2017. OptiClust, an improved method for assigning amplicon-based  
609 sequence data to operational taxonomic units. *mSphere* 2.
- 610 77. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB,  
611 Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2018. Vegan: Community ecology

- 612 package.
- 613 78. R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for  
614 Statistical Computing, Vienna, Austria.
- 615 79. Kuhn M. 2008. Building predictive models inRUsing thecaretPackage. Journal of Statistical  
616 Software 28.
- 617 80. Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. 2020. A framework for effective  
618 application of machine learning to microbiome-based classification problems. mBio 11.
- 619 81. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes  
620 A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D,  
621 Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the  
622 tidyverse. Journal of Open Source Software 4:1686.

623 **Figures**

624

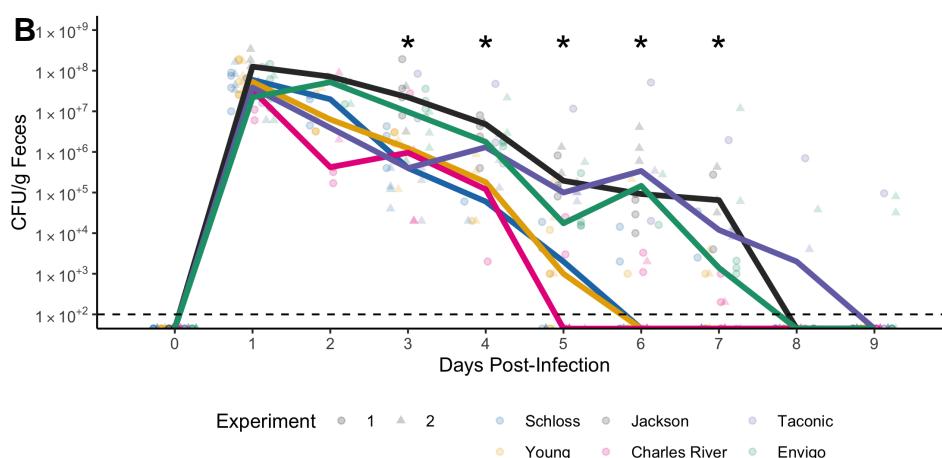
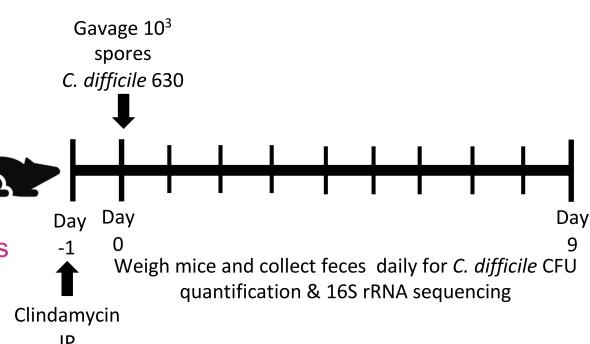
**Figure 1. Microbiota variation is high between mice from different sources.** A-B. Number of observed OTUs (A) and Shannon diversity index values (B) across sources of mice at baseline (day -1) of the experiment. Differences across sources were analyzed by Kruskal-Wallis test with Benjamini-Hochberg correction for testing each day of the experiment and the adjusted  $P$  value was  $< 0.05$  for panel A (Table S1). None of the  $P$  values from pairwise Wilcoxon comparisons between sources were significant after Benjamini-Hochberg correction (Table S1). Gray lines represent the median values for each source of mice. C. Principal Coordinates Analysis (PCoA) of  $\theta_{YC}$  distances of baseline stool samples. Permutational multivariate analysis of variance (PERMANOVA) analysis demonstrated that source and the interaction between source and cage explained most of the variation (combined  $R^2 = 0.90$ ,  $P = 0.0001$ , see Table S2). For A-C: each symbol representing the value for a stool sample from an individual mouse, circles represent experiment 1 mice and triangles represent experiment 2 mice. D. Plots highlighting the median (point) and interquartile range (colored lines) of the relative abundances for the top 20 bacteria out of the 268 OTUs that

638 varied across sources at baseline (Table S3).

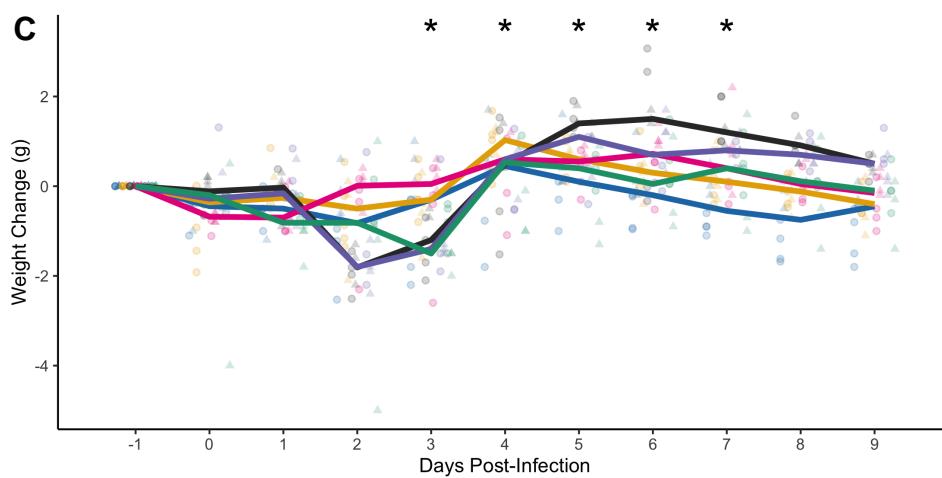
**A**

**Sources of C57BL/6 mice:**

1. Schloss Lab Colony at University of Michigan
2. Young Lab Colony at University of Michigan
3. The Jackson Laboratory
4. Charles River Laboratories
5. Taconic Biosciences
6. Envigo



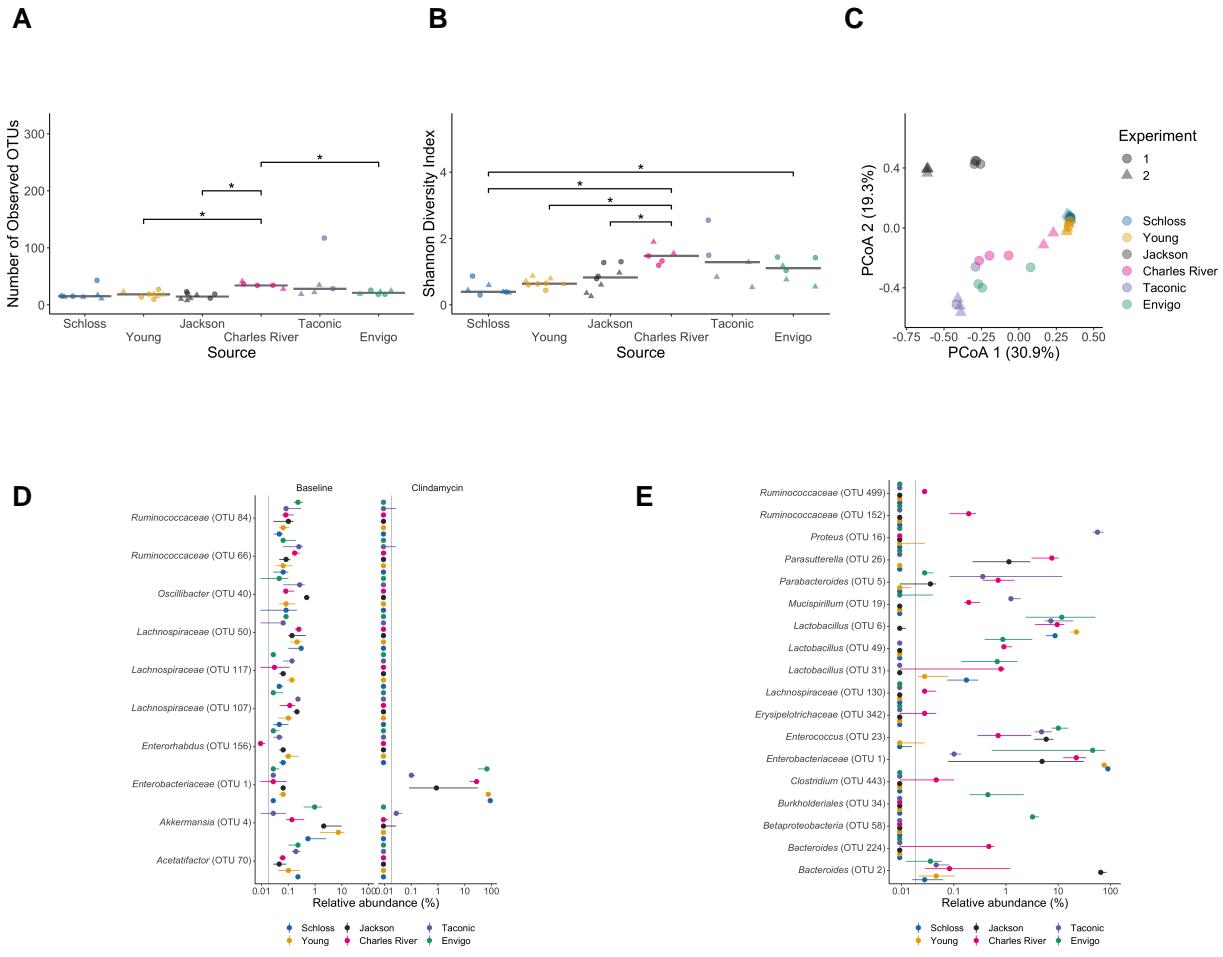
**C**



639

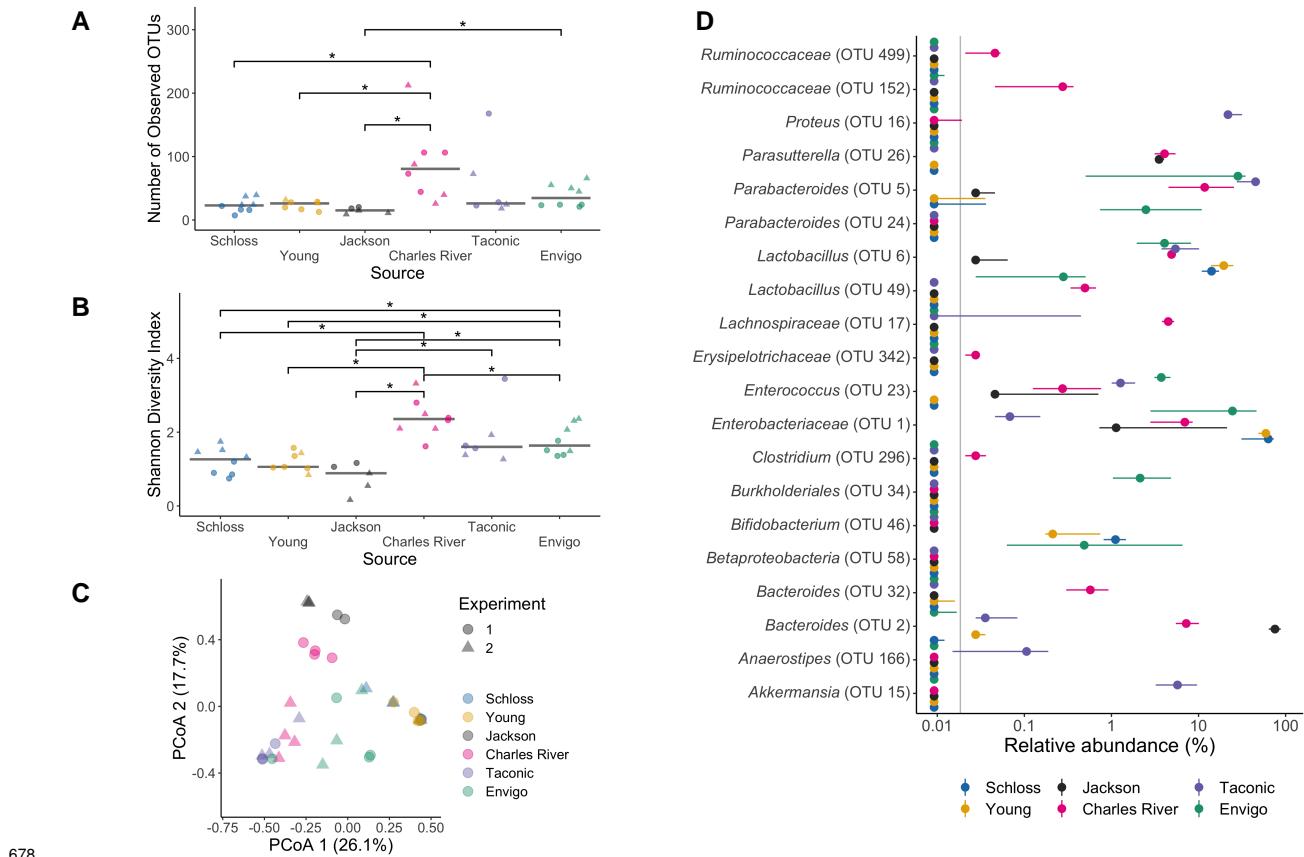
Figure 2. Clindamycin is sufficient to promote *C. difficile* colonization in all mice, but clearance time varies across sources of C57BL/6 mice. A. Setup of the experimental timeline. Mice for the experiments were obtained from 6 different sources: the Schloss ( $N = 8$ ) and Young lab ( $N = 9$ ) colonies at the University of Michigan, the Jackson Laboratory ( $N = 8$ ), Charles River

644 Laboratory (N = 8), Taconic Biosciences (N = 8), and Envigo (N = 8). All mice were administered  
645 10 mg/kg clindamycin intraperitoneally (IP) 1 day before challenge with *C. difficile* 630 spores on  
646 day 0. Mice were weighed and feces was collected daily through the end of the experiment (9  
647 days post-infection). Note: 3 mice died during course of experiment. 1 Taconic mouse prior to  
648 infection and 1 Jackson and 1 Envigo mouse between 1- and 3-days post-infection. B. *C. difficile*  
649 CFU/gram stool measured over time (N = 20-49 mice per timepoint) via serial dilutions. The  
650 black line represents the limit of detection for the first serial dilution. CFU quantification data was  
651 not available for each mouse due to early deaths, stool sampling difficulties, and not plating all  
652 of the serial dilutions. C. Mouse weight change measured in grams over time (N = 45-49 mice  
653 per timepoint), all mice were normalized to the weight recorded 1 day before infection. For B-C:  
654 timepoints where differences across sources of mice were statistically significant by Kruskal-Wallis  
655 test with Benjamini-Hochberg correction for testing across multiple days (Table S4 and Table S5)  
656 are reflected by the asterisk(s) above each timepoint (\*,  $P < 0.05$ ). Lines represent the median for  
657 each source and circles represent individual mice from experiment 1 while triangles represent mice  
658 from experiment 2.



660 **Figure 3. Clindamycin treatment alters bacteria in all sources, but a subset of bacterial**  
 661 **differences across sources persists.** A-B. Number of observed OTUs (A) and Shannon diversity  
 662 index values (B) across sources of mice after clindamycin treatment (day 0). Data were analyzed  
 663 by Kruskal-Wallis test with Benjamini-Hochberg correction for testing each day of the experiment  
 664 and the adjusted  $P$  value was  $< 0.05$ . Significant  $P$  values from the pairwise Wilcoxon comparisons  
 665 between sources with Benjamini-Hochberg correction are shown (Table S1). C. PCoA of  $\theta_{YC}$   
 666 distances from stools collected post-clindamycin. PERMANOVA analysis demonstrated that  
 667 source and the interaction between source and cage explained most of the variation observed  
 668 post-clindamycin (combined  $R^2 = 0.99$ ,  $P = 0.0001$ , see Table S2). For A-C, each symbol represents  
 669 a stool sample from an individual mouse, with circles representing experiment 1 mice and  
 670 triangles representing experiment 2 mice. D. Plots highlighting the median (point) and interquartile  
 671 range (colored lines) of the top 10 OTUs out of 153 with relative abundances that changed after  
 672 clindamycin treatment (adjusted  $P$  value  $< 0.05$ ). Data were analyzed by Wilcoxon signed rank test

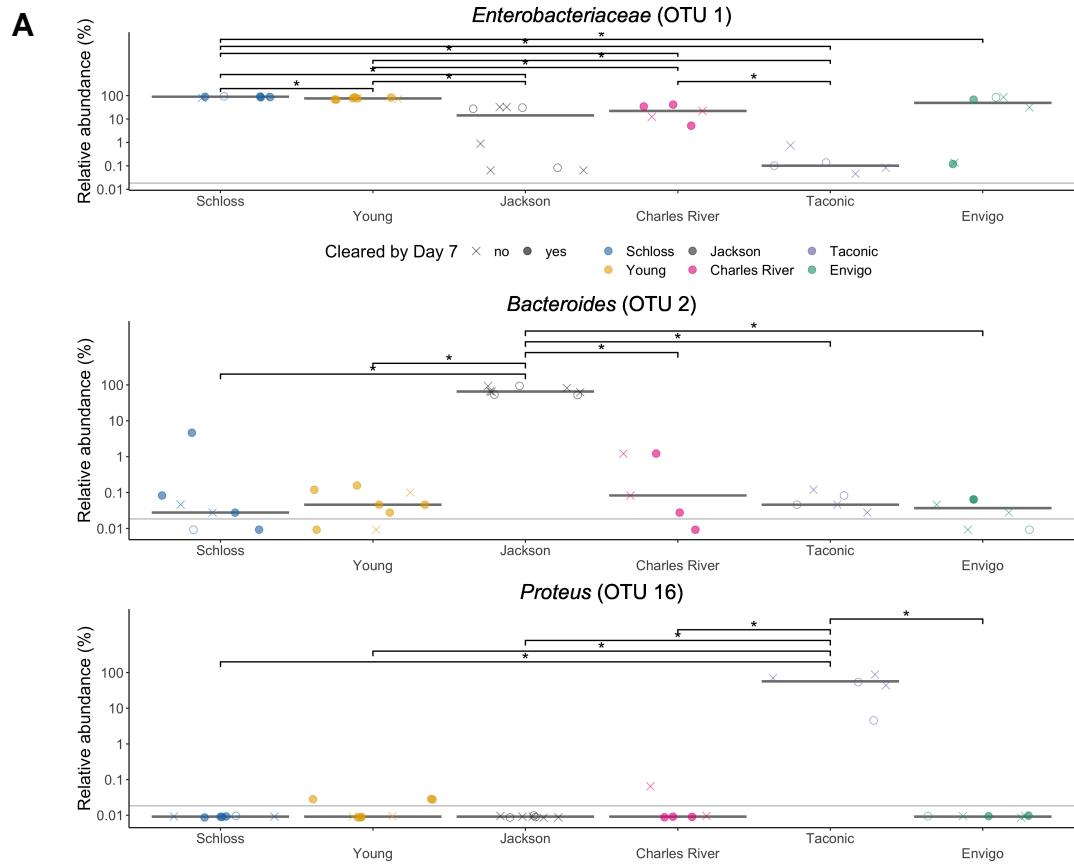
673 limited to mice that had paired sequence data for day -1 and 0 ( $N = 31$ ), with Benjamini-Hochberg  
674 correction for testing all identified OTUs (Table S6). The gray vertical line indicates the limit of  
675 detection. E. Plots highlighting the median (point) and interquartile range (colored lines) of the  
676 relative abundances for the 18 OTUs that varied across sources after clindamycin treatment (Table  
677 S3).



679 **Figure 4. Microbiota variation across sources is maintained after *C. difficile* challenge.**

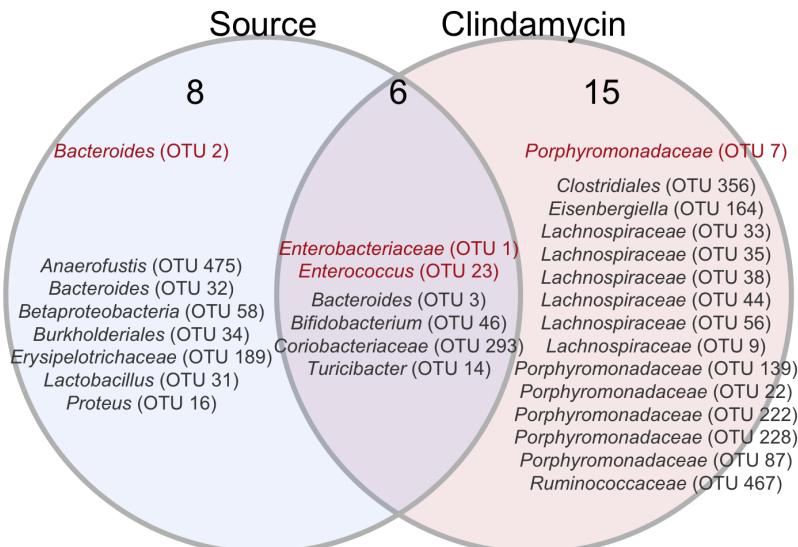
680 A-B. Number of observed OTUs (A) and Shannon diversity index values (B) across sources of  
 681 mice post-infection (day 1). Data were analyzed by Kruskal-Wallis test with Benjamini-Hochberg  
 682 correction for testing each day of the experiment and the adjusted *P* value was < 0.05. Significant  
 683 *P* values from the pairwise Wilcoxon comparisons between sources with Benjamini-Hochberg  
 684 correction are shown (Table S1). PCoA of  $\theta_{YC}$  distances of post-infection stool samples.  
 685 PERMANOVA analysis demonstrated that source and the interaction between source and cage  
 686 explained most of the variation (combined  $R^2 = 0.88$ ,  $P = 0.0001$ , Table S2). For A-C: each symbol  
 687 representing the value for a stool sample from an individual mouse, circles represent experiment  
 688 1 mice and triangles represent experiment 2 mice. D. Plots highlighting the median (point) and  
 689 interquartile range (colored lines) of the relative abundances for the top 20 bacteria out of the  
 690 44 OTUs that varied across sources post-infection. The gray vertical line indicates the limit of  
 691 detection. For each timepoint OTUs with differential relative abundances across sources of mice  
 692 were identified by Kruskal-Wallis test with Benjamini-Hochberg correction for testing all identified

<sub>693</sub> OTUs (Table S3).

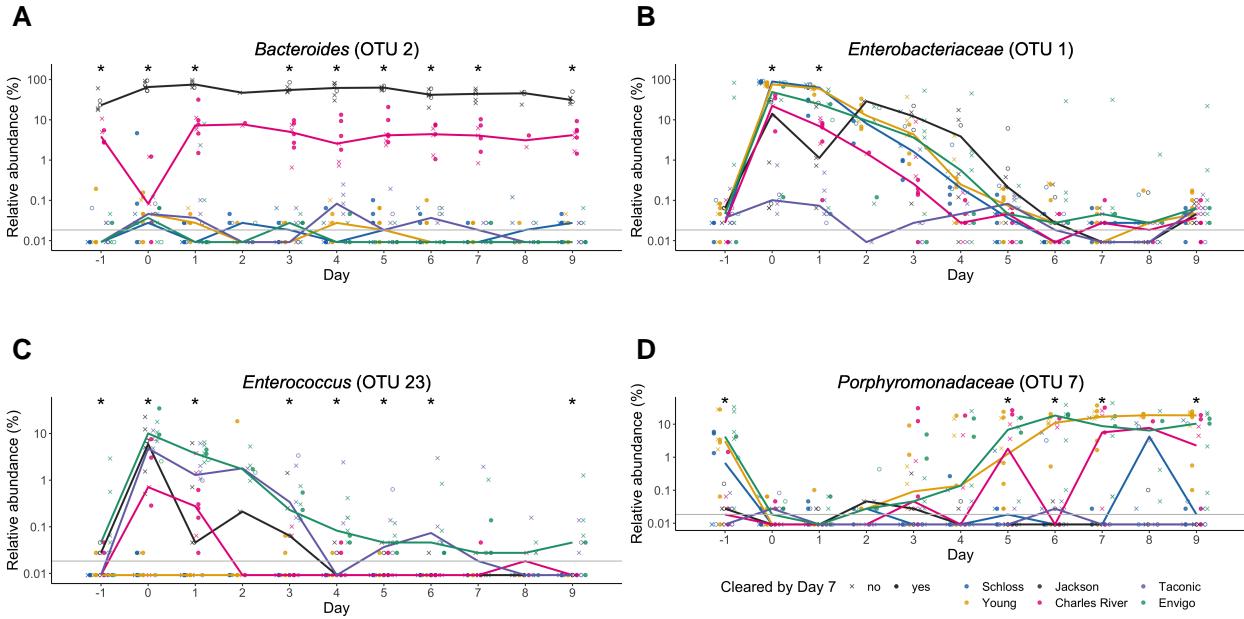


**B**

### Key taxa comparisons for day -1, 0, and 1 models

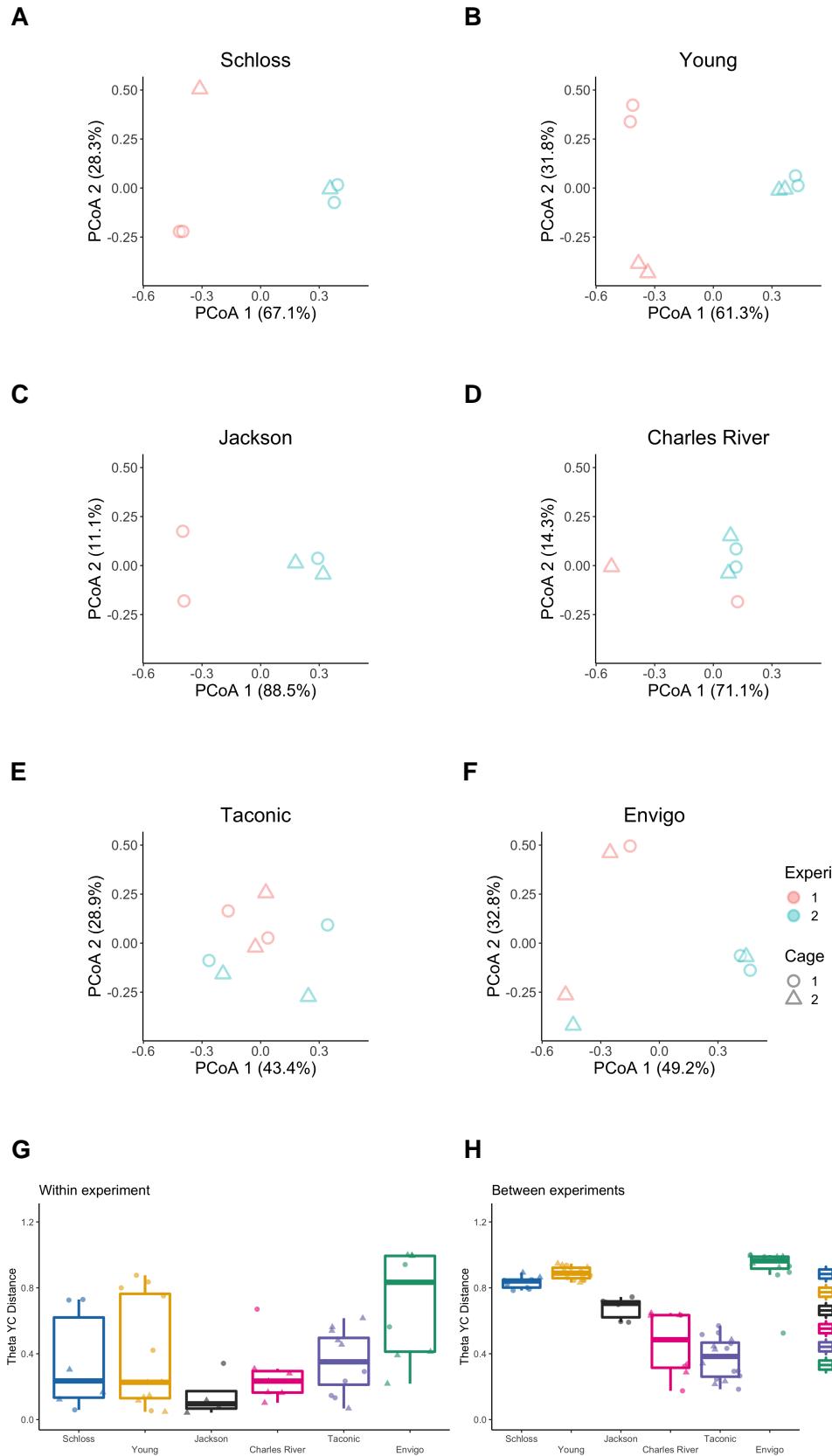


695 **5. Bacteria that influence whether mice cleared *C. difficile* by day 7.** A. Baseline relative  
696 abundance data for 3 of the OTUs from the classification model based on day 0 OTU relative  
697 abundances that significantly varied across sources of mice and had high relative abundances in  
698 the community. Symbols represent the relative abundance data for an individual mouse, circles  
699 represent mice that cleared *C. difficile* by day 7, X-shapes represent mice that were still colonized  
700 with *C. difficile*, and open circles represent mice that did not have *C. difficile* CFU counts for day 7  
701 post-infection. Gray lines indicate the median relative abundances for each source. Asterisks are  
702 shown for pairwise Wilcoxon comparisons with Benjamini-Hochberg correction where  $P < 0.05$ . B.  
703 Venn diagram that combines Fig. S4 summaries of OTUs that were important to the day -1, 0, and  
704 1 classification models (Table S8) and either overlapped with taxa that varied across sources at  
705 the same timepoint, were impacted by clindamycin treatment, or both. See Fig. S4 for separate  
706 comparisons of taxa from the day -1, 0, and 1 classification models. Red OTUs were important to  
707 more than 1 classification model.

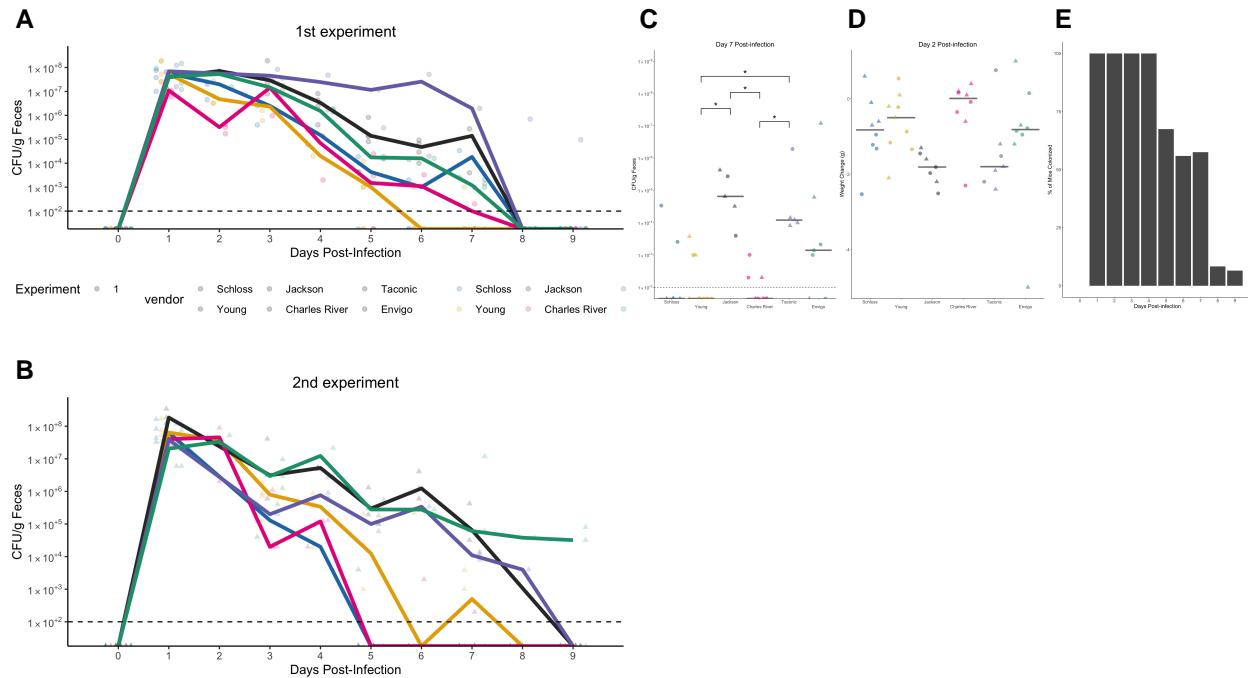


708

709 **Figure 6: Key OTUs vary across sources throughout the experiment.** A-D. Relative  
 710 abundances of red OTUs from Fig. 5A that were important for at least two classification models are  
 711 shown over time. A. *Bacteroides* (OTU 2), which varied across sources throughout the experiment.  
 712 B-C. *Enterobacteriaceae* (B) and *Enterococcus* (C), which significantly varied across sources and  
 713 were impacted by clindamycin treatment. D. *Porphyromonadaceae* (OTU 7), which was significantly  
 714 impacted by clindamycin treatment and examining relative abundance dynamics over the course of  
 715 the experiment, revealed timepoints where relative abundances also significantly varied across  
 716 sources of mice. Symbols represent the relative abundance data for an individual mouse, circles  
 717 represent mice that cleared *C. difficile* by day 7, X-shapes represent mice that were still colonized  
 718 with *C. difficile*, and open circles represent mice that did not have *C. difficile* CFU counts for day 7  
 719 post-infection. Colored lines indicate the median relative abundances for each source. The gray  
 720 horizontal line represents the limit of detection. Timepoints where differences across sources of  
 721 mice were statistically significant by Kruskal-Wallis test with Benjamini-Hochberg correction for  
 722 testing across multiple days (Table S9) are identified by the asterisk(s) above each timepoint (\*, P <  
 723 0.05).



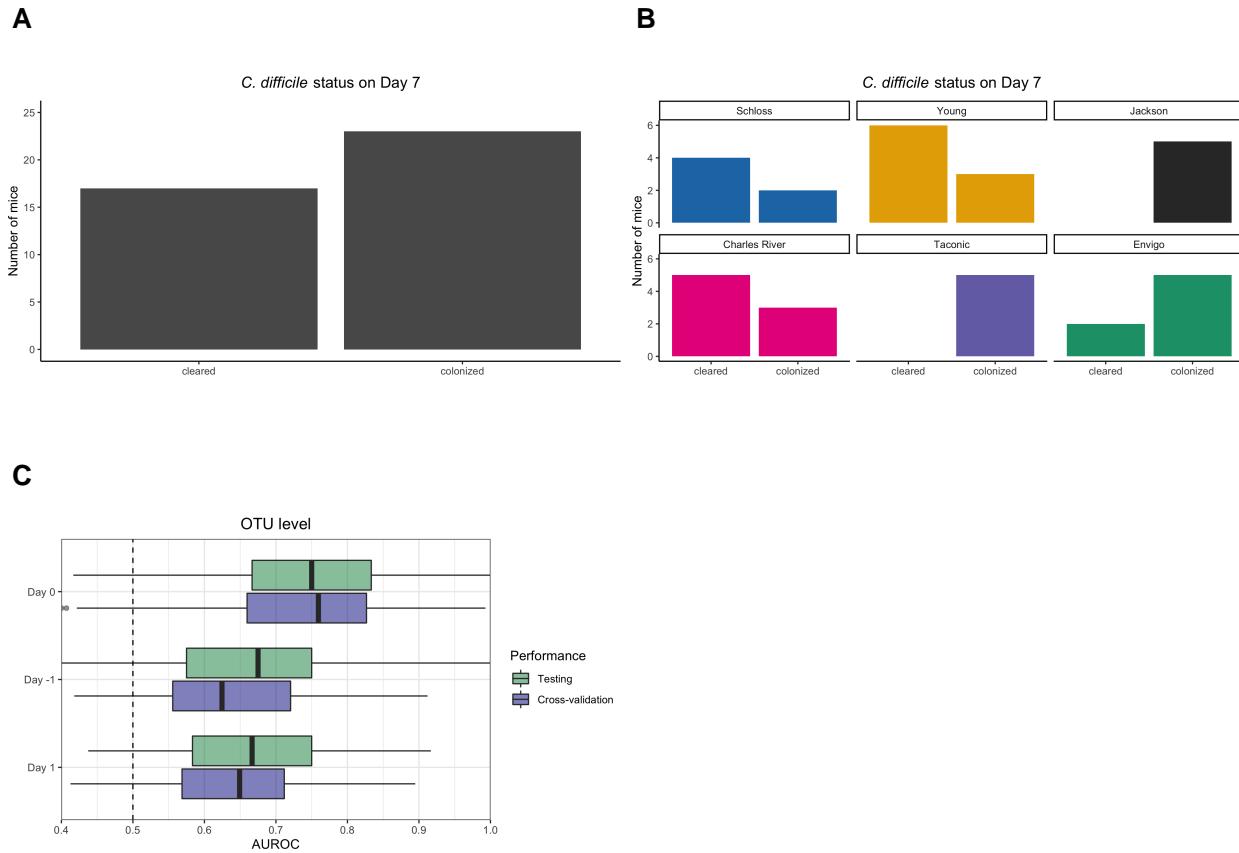
725 **Bacterial communities vary between experiments for some sources.** A-F: PCoA of  $\theta_{YC}$   
726 distances for the baseline fecal bacterial communities within each source of mice. Each symbol  
727 represents a stool sample from an individual mouse with color corresponding to experiment and  
728 shape representing cage mates. PERMANOVA was performed within each group to examine the  
729 contributions of experiment and cage to observed variation. Experiment number and cage only  
730 significantly explained observed variation for mice from the Schloss (combined  $R^2 = 0.99$ ;  $P \leq$   
731 0.033) and Young (combined  $R^2 = 0.95$ ;  $P \leq 0.029$ ) lab colonies (Table S2). G-H: Boxplots of the  
732  $\theta_{YC}$  distances of the 6 sources of mice relative to mice within the same source and experiment  
733 (G) or mice within the same source and between experiments (H) at baseline. Symbols represent  
734 individual mouse samples: circles for experiment 1 and triangles for experiment 2.



735

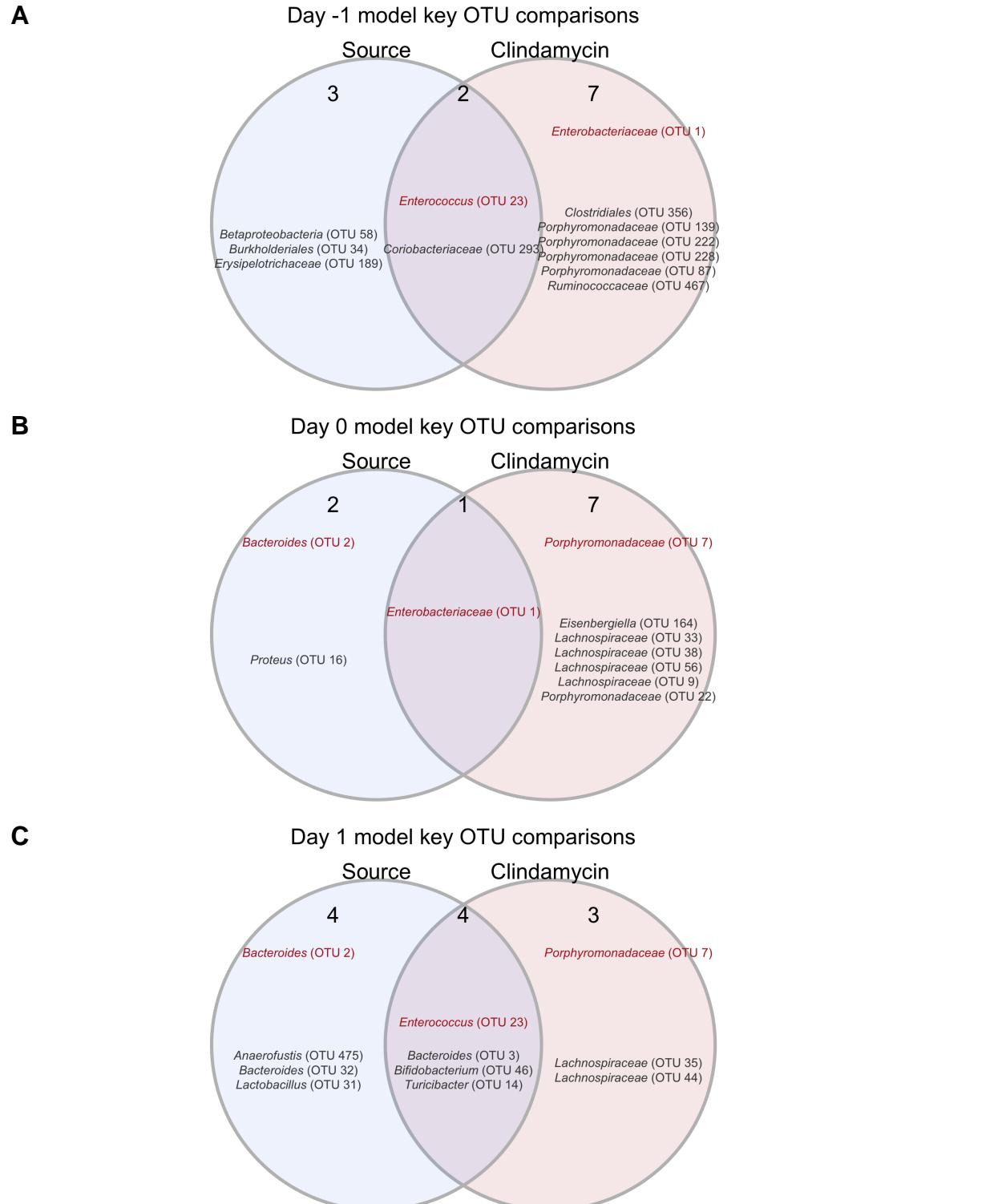
736 **Figure S2. *C. difficile* CFU variation across sources varies slightly between the 2**  
 737 **experiments.** A-B. *C. difficile* CFU/gram of stool quantification over time for experiment 1 (A) and

738 2 (B). Experiments were conducted approximately 3 months apart. Lines represent the median  
 739 CFU for each source, symbols represent individual mice and the black line represents the limit  
 740 of detection. C. *C. difficile* CFU/gram stool on day 7 post-infection across sources of mice with  
 741 asterisks for pairwise Wilcoxon comparisons with Benjamini-Hochberg correction where  $P < 0.05$ .  
 742 D. Mouse weight change 2 days post-infection across sources of mice, no pairwise Wilcoxon  
 743 comparisons were significant after Benjamini-Hochberg correction. For C-D: circles represent  
 744 experiment 1 mice, triangles represent experiment 2 mice and gray lines indicate the median  
 745 values for each group. E. Percent of mice that were colonized with *C. difficile* over the course of the  
 746 experiment. Each day the percent is calculated based on the mice where *C. difficile* CFU was  
 747 quantified for that particular day. Total N for each day: day 1 (N = 42), day 2 (N = 20), day 3 (N =  
 748 39), day 4 (N = 29), day 5 (N = 43), day 6 (N = 34), day 7 (N = 40), day 8 (N = 36), and day 9 (N =  
 749 46).



750

751 **Figure S3. Bacterial community composition before, after clindamycin perturbation, and**  
 752 **post-infection can predict *C. difficile* colonization status 7 days post-challenge.** A. Bar  
 753 graph visualizations of overall day 7 *C. difficile* colonization status that were used as classification  
 754 outcomes to build L2-regularized logistic regression models. Mice were classified as colonized or  
 755 cleared (not detectable at the limit of detection of 100 CFU) based on CFU g/stool data from 7 days  
 756 post-infection. B. *C. difficile* CFU status on Day 7 within each mouse source. N = 5-9 mice per  
 757 group. C. L2-regularized logistic regression classification model area under the receiving operator  
 758 characteristic curve (AUROCs) to predict *C. difficile* CFU on day 7 post-infectoin (Fig. 1D, Fig. S3)  
 759 based on the OTU community relative abundances at baseline (day -1), post-clindamycin (day 0),  
 760 and post-infection (day 1). All models performed better than random chance (AUROC = 0.5; all  
 761  $P \leq 1.6e-17$ ; Table S7) and the model built with post-clindamycin treated bacterial OTU relative  
 762 abundances had the best performance ( $(P_{FDR} \leq 3.3e-06$  for all pairwise comparisons; Table S7).  
 763 For list of the 20 OTUs that were ranked as most important to each model, see Table S8.



Figure

764  
765 **S4. OTUs from classification models based on baseline, post-clindamycin treatment, or**  
766 **post-infection community data vary by source, clindamycin treatment, or both.** A-C. Venn  
767 diagrams of top 20 important OTUs from baseline (A), post-clindamycin treatment (B), and

<sup>768</sup> post-infection (C) classification models (Table S8) that overlapped with OTUs that varied across  
<sup>769</sup> sources at the corresponding timepoint, were impacted by clindamycin treatment, or both. Bold  
<sup>770</sup> OTUs signify OTUs that were important to more than 1 classification model.

771 **Supplementary Tables and Movie**

772 All supplemental material is available at: [https://github.com/SchlossLab/Tomkovich\\_Vendor\\_XXXX\\_](https://github.com/SchlossLab/Tomkovich_Vendor_XXXX_)  
773 2020/submission.

774 **Movie S1. Large shifts in bacterial community structure occurred after clindamycin and**  
775 ***C. difficile* infection.** PCoA of  $\theta_{YC}$  distances animated from 0 through 9 days post-infection.  
776 PERMANOVA analysis indicated source was the variable that explained the most observed variation  
777 across fecal communities (source  $R^2 = 0.35$ ,  $P = 0.0001$ ) followed by interactions between cage  
778 and day of the experiment. Transparency of the circle corresponds to the day of the experiment,  
779 each circle represents a sample from an individual mouse at a specific timepoint. See Table S2  
780 for PERMANOVA results). Circles represent mice from experiment 1 and triangles represent mice  
781 from expeirment 2.

782 **Table S1. Diversity metrics Kruskal-Wallis and pairwise Wilcoxon statistical results.**

783 **Table S2. PERMANOVA results for mice at specific timepoints and across all timepoints.**

784 **Table S3. OTUs with relative abudances that significantly vary across sources at baseline,**  
785 **post-clindamycin, or post-infection timepoints.**

786 **Table S4. *C. difficile* CFU statistical results.**

787 **Table S5. Mouse weight change statistical results.**

788 **Table S6. OTUs with relative abudances that significantly changed after clindamycin**  
789 **treatment.**

790 **Table S7. Statistical results of L2-regularized logistic regression model performances.**

791 **Table S8. Top 20 most important OTUs for each of the 3 L2-regularized logistic regression**  
792 **models based on OTU relative abundance data.**

793 **Table S9. OTUs with relative abudances that significantly varied across sources of mice on**  
794 **at least 1 day of the experiment by Kruskal-Wallis test.**