

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

Sarah Tomkovich¹, Joshua M.A. Stough¹, Lucas Bishop¹, Patrick D. Schloss^{1†}

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

1 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 source. We treated mice from 6 sources (2 University of Michigan colonies
5 and 4 commercial vendors) with clindamycin, followed by a *C. difficile* challenge and then measured
6 *C. difficile* colonization levels throughout the infection. The microbiota were profiled via 16S rRNA
7 gene sequencing to examine the variation across sources and alterations due to clindamycin
8 treatment and *C. difficile* challenge. While all mice were colonized 1-day post-infection, variation
9 emerged from days 3-7 post-infection with animals from some sources colonized with *C. difficile* for
10 longer and at higher levels. We identified bacteria that varied in relative abundance across sources
11 and throughout the experiment. Some bacteria were consistently impacted by clindamycin treatment
12 in all sources of mice including *Lachnospiraceae*, *Ruminococcaceae*, and *Enterobacteriaceae*. To
13 identify bacteria that were most important to colonization regardless of the source, we created
14 logistic regression models that successfully classified mice based on whether they cleared *C.*
15 *difficile* by 7 days post-infection using community composition data at baseline, post-clindamycin,
16 and 1-day post-infection. With these models, we identified 4 bacteria that were predictive of
17 whether *C. difficile* cleared. They varied across sources (*Bacteroides*), were altered by clindamycin
18 (*Porphyromonadaceae*), or both (*Enterobacteriaceae* and *Enterococcus*). Allowing for microbiota
19 variation across sources better emulates human inter-individual variation and can help identify
20 bacterial 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 is an established risk, there is variation in who becomes asymptotically colonized, develops
24 an infection, or has adverse infection outcomes. Mouse models of *C. difficile* infection (CDI) are
25 widely used to answer a variety of *C. difficile* pathogenesis questions. However, the inter-individual
26 variation between mice from the same breeding facility is less than what is observed in humans.
27 Therefore, we challenged mice from 6 different breeding colonies with *C. difficile*. We found that the
28 starting microbial community structures and *C. difficile* persistence varied by the source of mice.

29 Interestingly, a subset of the bacteria that varied across sources were associated with how long *C.*
30 *difficile* was able to colonize. By increasing the inter-individual diversity of the starting communities,
31 we were able to better model human diversity. This provided a more nuanced perspective of *C.*
32 *difficile* pathogenesis.

33 Introduction

34 Antibiotics are a common risk factor for *Clostridioides difficile* infections (CDIs) due to their effect on
35 the intestinal microbiota, but there is variation in who goes on to develop severe or recurrent CDIs
36 after exposure (1, 2). Additionally, asymptomatic colonization, where *C. difficile* is detectable, but
37 symptoms are absent, has been documented in infants and adults (3, 4). The intestinal microbiota
38 has been implicated in asymptomatic colonization (5, 6), susceptibility to CDIs (7), and adverse CDI
39 outcomes (9–12). However, it is not clear how much inter-individual microbiota variation contributes
40 to the range of outcomes observed after *C. difficile* exposure relative to other risk factors.

41 Mouse models of CDIs have been a great tool for understanding *C. difficile* pathogenesis (13).
42 The number of CDI mouse model studies has grown substantially since Chen et al. published
43 their C57BL/6 model in 2008, which disrupted the gut microbiota with antibiotics to enable *C.*
44 *difficile* colonization and symptoms such as diarrhea and weight loss (14). CDI mouse models
45 have been used to examine translationally relevant questions regarding *C. difficile*, including the
46 role of the microbiota and efficacy of potential therapeutics for treating CDIs (15). However,
47 variation in the microbiota between mice from the same breeding colony is much less than the
48 inter-individual variation observed between humans (16, 17). Studying CDIs in mice with a
49 homogeneous microbiota is likely to overstate the importance of individual mechanisms. Using
50 mice that have a more heterogeneous microbiota would allow researchers to identify and validate
51 more generalizable mechanisms responsible for CDI.

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

61 Results

62 **The variation in the microbiota is high between mice from different sources.** We obtained
63 C57BL/6 mice from 6 different sources: two colonies from the University of Michigan that were
64 split from each other in 2010 (the Young and Schloss lab colonies) and four commercial vendors:
65 the Jackson Laboratory, Charles River Laboratories, Taconic Biosciences, and Envigo (which was
66 formerly Harlan). These 4 vendors were chosen because they are commonly used for murine CDI
67 studies (26, 34–40). Two experiments were conducted, approximately 3 months apart.

68 We sequenced the 16S rRNA gene from fecal samples collected from these mice after they
69 acclimated to the University of Michigan animal housing environment. We first examined the
70 alpha diversity across the 6 sources of mice. There was a significant difference in the richness
71 (i.e. number of observed operational taxonomic units (OTUs)), but not Shannon diversity index
72 across the sources of mice ($P_{FDR} = 0.03$ and $P_{FDR} = 0.052$, respectively; Fig. 1A-B and Tables
73 S1-2). Next, we compared the community structure of mice (Fig. 1C). The source of mice and
74 the interactions between the source and cage effects explained most of the observed variation
75 between fecal communities (PERMANOVA combined $R^2 = 0.90$, $P < 0.001$; Fig. 1C and Table
76 S3). Mice that are co-housed tend to have similar gut microbiotas due to coprophagy (41). Since
77 mice within the same source were housed together, it was not surprising that the cage effect also
78 contributed to the observed community variation. There were some differences between the 2
79 experiments we conducted, as the experiment and cage effects significantly explained the observed
80 community variation for the Schloss and Young lab mouse colonies (Fig. S2A-B and Table S4).
81 However, most of the vendors also clustered by experiment (Fig. S2C-D, F), suggesting there was
82 some community variation between the 2 experiments within each source, particularly for Schloss,
83 Young, and Envigo mice (Fig. S2G-H). After finding differences at the community level, we next
84 identified the bacteria that varied between sources of mice. There were 268 OTUs with relative
85 abundances that were significantly different between the sources (Fig. 1D and Table S5). Though
86 we saw differences between experiments at the community level, there were no OTUs that were
87 significantly different between experiments within Schloss, Young, and Envigo mice (all $P > 0.05$).
88 By using mice from six sources we were able to increase the variation in the starting communities
89 to evaluate in a clindamycin-based CDI model.

90 **Clindamycin treatment renders all mice susceptible to *C. difficile* 630 colonization, but**
91 **clearance time varies across sources.** Clindamycin is frequently implicated with human CDIs
92 (42) and was part of the antibiotic treatment for the frequently cited 2008 CDI mouse model (14). We
93 have previously demonstrated mice are rendered susceptible to *C. difficile*, but clear the pathogen
94 within 9 days when treated with clindamycin alone (21, 43). All mice were treated with 10 mg/kg
95 clindamycin via intraperitoneal injection and one day later challenged with 10^3 *C. difficile* 630
96 spores (Fig. 2A). The day after infection, *C. difficile* was detectable in all mice at a similar level
97 (median CFU range: 2.2e+07-1.3e+08; $P_{FDR} = 0.15$), indicating clindamycin rendered all mice
98 susceptible regardless of source (Fig. 2B). However, between 3 and 7 days post-infection, we
99 observed variation in *C. difficile* levels across sources of mice (all $P_{FDR} \leq 0.019$; Fig. 2B and
100 Table S6). This suggested the source of mice was associated with *C. difficile* clearance. While the
101 colonization dynamics were similar between the two experiments, the Schloss mice took longer to
102 clear in the first experiment compared to the second and the Envigo mice took longer to clear in the
103 second experiment compared to the first (Fig. S2A-B). The change in the mice's weight significantly
104 varied across sources of mice with the most weight loss occurring two days post-infection (Fig.
105 2C and Table S7). There was also one Jackson and one Envigo mouse that died between 1- and
106 3-days post-infection during the second experiment. Mice obtained from Jackson, Taconic, and
107 Envigo tended to lose more weight, have higher *C. difficile* CFU levels and take longer to clear the
108 infection compared to the other sources of mice (although there was variation between experiments
109 with Schloss and Envigo mice). This was particularly evident 7 days post-infection (Fig. 2B-C,
110 Fig. S2C-D), when 57% of the mice were still colonized with *C. difficile* (Fig. S2E). By 9 days
111 post-infection the majority of the mice from all sources had cleared *C. difficile* with the exception of
112 1 Taconic mouse from the first experiment and 2 Envigo mice from the second experiment (Fig.
113 2B). Thus, clindamycin rendered all mice susceptible to *C. difficile* 630 colonization, regardless of
114 source, but there was significant variation in disease phenotype across the sources of mice.

115 **Clindamycin treatment alters bacteria in all sources, but a subset of bacterial differences**
116 **across sources persists.** Given the variation in fecal communities that we observed across
117 breeding colonies, we hypothesized that variation in *C. difficile* clearance would be explained by
118 community variation across the 6 sources of mice. As expected, clindamycin treatment decreased

119 the richness and Shannon diversity across all sources of mice (Fig. 3A-B). Interestingly, significant
120 differences in diversity metrics between sources emerged after clindamycin treatment, with Charles
121 River mice having higher richness and Shannon diversity than most of the other sources ($P_{FDR} <$
122 0.05; Fig 3A-B and Tables S1-2). The clindamycin treatment decreased the variation in community
123 structures between sources of mice. The source of mice and the interactions between source
124 and cage effects explained almost all of the observed variation between communities (combined
125 $R^2 = 0.99$, $P < 0.001$; Fig. 3C and Table S3). However, there were only 18 OTUs with relative
126 abundances that significantly varied between sources (Fig. 3D and Table S8). Next, we identified
127 the bacteria that shifted after clindamycin treatment, regardless of source by analyzing paired fecal
128 samples from mice that were collected at baseline and after clindamycin treatment. We identified
129 153 OTUs that were altered after clindamycin treatment in most mice (Fig. 3E and Table S9).
130 When we compared the list of significant clindamycin impacted bacteria with the bacteria that
131 varied between sources post-clindamycin, we found 4 OTUs that were shared between the lists
132 (*Enterobacteriaceae* (OTU 1), *Lachnospiraceae* (OTU 130), *Lactobacillus* (OTU 6), *Enterococcus*
133 (OTU 23); Fig. 3D-E and Tables S8-9). Importantly, some of the OTUs that varied between sources
134 also shifted with clindamycin treatment. For example, *Proteus* increased after clindamycin treatment
135 (Fig. 3D), but only in Taconic mice. *Enterococcus* was primarily found only in mice purchased from
136 commercial vendors and also increased in relative abundance after clindamycin treatment (Fig.
137 3D). These findings demonstrate that clindamycin had a consistent impact on the fecal bacterial
138 communities of mice from all sources and only a subset of the OTUs continued to vary between
139 sources.

140 **Microbiota variation between sources is maintained after *C. difficile* challenge.** One day
141 post-infection, significant differences in diversity metrics remained across sources ($P_{FDR} < 0.05$, Fig
142 4A-B and Tables S1-2). Although the Charles River mice had more diverse communities and were
143 also able to clear *C. difficile* faster than the other sources, diversity did not explain the observed
144 variation in *C. difficile* colonization across sources. The Young and Schloss mice had the lowest
145 diversity 1 day post-infection and were able to clear *C. difficile* earlier than Jackson, Taconic and
146 Envigo mice. The source of mice and the interactions between source and cage effects continued
147 to explain most of the observed community variation (combined $R^2 = 0.88$; $P < 0.001$; Fig. 4C

148 and Table S3). One day after *C. difficile* challenge, there were 44 OTUs with significantly different
149 relative abundances across sources (Fig. 4D and Table S10).

150 Throughout the experiment, the source of mice continued to be the dominant factor that explained
151 the observed variation across fecal communities (PERMANOVA $R^2 = 0.35$, $P < 0.001$) followed by
152 interactions between cage effects and the day of the experiment (Movie S1 and Table S11). Fecal
153 samples from the same source of mice continued to cluster closely to each other throughout the
154 experiment. By 7 days post-infection, when approximately 43% mice had cleared *C. difficile*, most
155 of the mice had not recovered to their baseline community structure (Fig. 4E). The distance to
156 the baseline community did not explain the variation in *C. difficile* clearance as the Schloss and
157 Young mice had mostly cleared *C. difficile*, but their communities were a greater distance from
158 baseline 7 days post-infection compared to the Jackson and Taconic mice that were still colonized.
159 In summary, mouse bacterial communities varied significantly between sources throughout the
160 course of the experiment and a consistent subset of bacteria remained different between sources
161 regardless of clindamycin and *C. difficile* challenge.

162 **Baseline, post-clindamycin, and post-infection community data can predict mice that will**
163 **clear *C. difficile* by 7 days post-infection.** After identifying taxa that varied between sources,
164 changed after clindamycin treatment, or both, we determined which taxa were influencing the
165 variation in *C. difficile* colonization at day 7 (Fig. 2B, Fig. S2C). We trained three L2-regularized
166 logistic regression models with either input bacterial community data from the baseline (day = -1),
167 post-clindamycin (day = 0), or post-infection (day = 1) timepoints of the experiment to predict *C.*
168 *difficile* colonization status on day 7 (Fig. S3A-B). All models were better at predicting *C. difficile*
169 colonization status on day 7 than random chance (all $P < 0.001$, Table S12). The model based on
170 the post-clindamycin (AUROC = 0.78) community OTU data performed significantly better than the
171 baseline (AUROC = 0.72) or the post-infection (AUROC = 0.67) models ($P_{FDR} < 0.001$ for pairwise
172 comparisons; Fig. S3C and Table S13). Thus, we were able to use bacterial relative abundance
173 data from the time of *C. difficile* challenge to differentiate mice that had cleared *C. difficile* before
174 day 7 from the mice still colonized with *C. difficile* at that timepoint. This result suggests that the
175 bacterial community's response to clindamycin treatment had the greatest influence on subsequent
176 *C. difficile* colonization dynamics.

177 To examine the bacteria that were driving each model's performance, we selected the 20 OTUs
178 that had the highest absolute feature weights in each of the 3 models (Table S14). First, we looked
179 at OTUs from the model with the best performance, which was based on the post-clindamycin
180 treatment (day 0) bacterial community data. Out of the 10 highest ranked OTUs, 7 OTUs were
181 associated with *C. difficile* colonization 7 days post-infection (*Bacteroides*, *Escherichia/Shigella*, 2
182 *Lachnospiraceae*, *Lactobacillus*, *Porphyromonadaceae*, and *Ruminococcaceae*), while 3 OTUs
183 were associated with clearance (*Enterobacteriaceae*, *Lachnospiraceae*, *Porphyromonadaceae*;
184 Fig. 5A). Next, we examined whether any of the top 20 ranked OTUs from the post-clindamycin
185 (day 0) model were also important in the other 2 classification models based on baseline (day
186 -1) and 1 day post-infection community data. We identified 6 OTUs that were important to the
187 post-clindamycin model and either the baseline or 1 day post-infection models (*Enterobacteriaceae*,
188 *Ruminococcaceae*, *Lactobacillus*, *Bacteroides*, *Porphyromonadaceae*, *Erysipelotrichaceae*; Table
189 S14). Thus, a subset of bacterial OTUs were important for determining *C. difficile* colonization
190 dynamics across multiple timepoints.

191 To determine whether the OTUs driving the classification models also varied between sources,
192 were altered by clindamycin treatment, or both, we identified the OTUs from each model that varied
193 between sources (Fig. 1D, 3D, 4D and Tables S5, S8, S10) or were impacted by clindamycin
194 treatment (Fig. 3E and Table S9; Fig. S4). Comparing the features important to the 3 models
195 identified 14 OTUs associated with source, 21 OTUs associated with clindamycin treatment, and
196 6 OTUs associated with both (Fig. 5B). Together, these results suggest that the initial bacterial
197 communities and their responses to clindamycin influenced the clearance of *C. difficile*.

198 Several OTUs that overlapped with our previous analyses appeared across at least 2 models
199 (*Bacteroides*, *Enterococcus*, *Enterobacteriaceae*, *Porphyromonadaceae*), so we examined how
200 the relative abundances of these OTUs varied over the course of the experiment (Fig. 6). Across
201 the 9 days post-infection, there was at least 1 timepoint when the relative abundances of these
202 OTUs significantly varied between sources (Table S15). Interestingly, there were no OTUs that
203 emerged as consistently enriched or depleted in mice that were colonized past 7 days post-infection,
204 suggesting that multiple bacteria influence *C. difficile* colonization dynamics.

205 **Discussion**

206 Applying our CDI model to 6 different sources of mice, allowed us to identify bacterial taxa that
207 were unique to different sources as well as taxa that were universally impacted by clindamycin.
208 We trained logistic regression models with baseline (day -1), post-clindamycin treatment (day 0),
209 and 1-day post-infection fecal community data that could predict whether mice cleared *C. difficile*
210 by 7 days post-infection better than random chance. We identified *Bacteroides*, *Enterococcus*,
211 *Enterobacteriaceae*, *Porphyromonadaceae* (Fig. 6) as candidate bacteria within these communities
212 that influenced variation in *C. difficile* colonization dynamics since these bacteria were all important
213 in the logistic regression models and varied by source, were impacted by clindamycin treatment,
214 or both. Overall, our results demonstrated clindamycin was sufficient to render mice from multiple
215 sources susceptible to CDI and only a subset of the inter-individual microbiota variation across
216 mice from different sources was needed to predict which mice could clear *C. difficile*.

217 Other studies have used mice from multiple sources to identify bacteria that either promote
218 colonization resistance or increase susceptibility to enteric infections (22, 23, 26–30). For example,
219 against *Salmonella* infections, *Enterobacteriaceae* and segmented filamentous bacteria have
220 emerged as protective (22, 27). We found *Enterobacteriaceae* increased in all sources of mice
221 after clindamycin treatment, facilitating *C. difficile* colonization. However, there was also variation
222 in *Enterobacteriaceae* relative abundance levels between sources that was associated with the
223 variation in *C. difficile* colonization dynamics across sources. Thus, bacteria may have differential
224 roles in determining susceptibility depending on the type of bacterial infection.

225 Differences in CDI mouse model studies have been attributed to intestinal microbiota variation
226 across sources. For example, researchers using the same clindamycin treatment and C57BL/6
227 mice had different *C. difficile* outcomes, one having sustained colonization (32), while the other had
228 transient colonization (18), despite both using *C. difficile* VPI 10643. Baseline differences in the
229 microbiota composition have been hypothesized to partially explain the differences in colonization
230 outcomes and overall susceptibility to *C. difficile* after treatment with the same antibiotic (13,
231 31). When we treated mice from 6 different sources with clindamycin and challenged them with
232 *C. difficile* 630, we found microbiota variation across sources impacted colonization outcomes,

233 but not susceptibility. A previous study with *C. difficile* identified an endogenous protective *C.*
234 *difficile* strain LEM1 that bloomed after antibiotic treatment in mice from Jackson or Charles
235 River Laboratories, but not Taconic that protected mice against the more toxigenic *C. difficile*
236 VPI10463 (26). Given that we obtained mice from the same vendors, we checked all mice for
237 endogenous *C. difficile* by plating stool samples that were collected after clindamycin treatment.
238 However, we did not identify any endogenous *C. difficile* strains prior to challenge, suggesting
239 there were no endogenous protective strains in the mice we received and other bacteria mediated
240 the variation in *C. difficile* colonization across sources. The *C. difficile* strain used could also
241 be contributing to the variation in *C. difficile* outcomes seen across different research groups.
242 For example, a group found differential colonization outcomes after clindamycin treatment, with
243 *C. difficile* 630 and M68 infections eventually becoming undetectable while strain BI-7 remained
244 detectable up to 70 days post-treatment (44). The bacterial perturbations induced by clindamycin
245 treatment have been well characterized and our findings agree with previous CDI mouse model work
246 demonstrating *Enterococcus* and *Enterobacteriaceae* were associated with *C. difficile* susceptibility
247 and *Porphyromonadaceae*, *Lachnospiraceae*, *Ruminococcaceae*, and *Turicibacter* were associated
248 with resistance (19, 21, 32, 33, 43–46). While we have demonstrated that susceptibility is uniform
249 across sources of mice after clindamycin treatment, there could be different outcomes for either
250 susceptibility or clearance in the case of other antibiotic treatments.

251 We found the time needed to naturally clear *C. difficile* varied across sources of mice implying that
252 at least in the context of the same perturbation, microbiota differences influence infection outcome.
253 More importantly, we were able to explain the variation observed across sources with a subset of
254 OTUs that were also important for predicting *C. difficile* colonization status 7 days post-infection.
255 Since all but 3 mice eventually cleared *C. difficile* 630 by 9 days post-infection and the model built
256 with the post-clindamycin (day 0) OTU relative abundance data had the best performance, our
257 results suggest clindamycin treatment had a larger role in determining *C. difficile* susceptibility and
258 clearance than the source of the mice.

259 Using mice from multiple sources successfully increased the inter-animal variation. One alternative
260 approach that has been used in some CDI studies is to associate mice with human microbiotas
261 (47–52). However, a major caveat to this method is the substantial loss of human microbiota

262 community members upon transfer to mice (53, 54). Additionally with the exception of 2 recent
263 studies (47, 48), most of these studies associated mice with just 1 types of human microbiota either
264 from a single donor or a single pool from multiple donors (49–52). This approach does not aid in
265 the goal of modeling the interpersonal variation seen in humans to understand how the microbiota
266 influences susceptibility to CDIs and adverse outcomes. Importantly, our study using mice from
267 6 different sources increased the variation between groups of mice compared to using 1 source
268 alone, to better reflect the inter-individual microbiota variation observed in humans.

269 Another motivation for associating mice with human microbiotas is to study the bacteria associated
270 with the disease in humans. Decreased *Bifidobacterium*, *Porphyromonas*, *Ruminococcaceae* and
271 *Lachnospiraceae* and increased *Enterobacteriaceae*, *Enterococcus*, *Lactobacillus*, and *Proteus*
272 have all been associated with human CDIs (7). Encouragingly, these populations were well
273 represented in our study, suggesting most of the mouse sources are suitable for gaining insights
274 into the bacteria influencing *C. difficile* colonization and infections in humans. An important
275 exception was *Enterococcus*, which was primarily absent from University of Michigan colonies and
276 *Proteus*, which was only found in Taconic mice. The fact that some CDI-associated bacteria were
277 only found in a subset of mice has important implications for future CDI mouse model studies, but
278 also models the natural patchiness of microbial populations in humans.

279 Other microbiota and host factors that were outside the scope of our current study may also
280 contribute to the differences in *C. difficile* colonization dynamics between sources of mice.
281 The microbiota is composed of viruses, fungi, and parasites in addition to bacteria, and these
282 non-bacterial members can also vary across sources of mice (55, 56). While our study focused
283 solely on the bacterial portion, viruses and fungi have also begun to be implicated in the context
284 of CDIs or FMT treatments for recurrent CDIs (35, 57–60). Beyond community composition, the
285 metabolic function of the microbiota also has a CDI signature (20, 46, 61, 62) and can vary across
286 mice from different sources (63). For example, microbial metabolites, particularly secondary
287 bile acids and butyrate production, have been implicated as important contributors to *C. difficile*
288 resistance (33, 44). Interestingly, butyrate has previously been shown to vary across mouse
289 vendors and mediated resistance to *Citrobacter rodentium* infection, a model of enterohemorrhagic
290 and enteropathogenic *Escherichia coli* infections (23). Evidence for immunological toning

291 differences in IgA and Th17 cells across mice from different vendors have also been documented
292 and (64, 65) could influence the host response to CDI (66, 67). The outcome after *C. difficile*
293 exposure depends on a multitude of factors, including age, diet, and immunity; all of which are also
294 influenced by the microbiota.

295 We have demonstrated that the ways baseline microbiotas from different mouse sources respond
296 to clindamycin treatment influence the length of time mice remained colonized with *C. difficile* 630.
297 To better understand the contribution of the microbiota to *C. difficile* pathogenesis and treatments,
298 using multiple sources of mice may yield more insights than a single source. Furthermore, for
299 studies wanting to examine the interplay between particular bacteria such as *Enterococcus* and *C.*
300 *difficile*, these results could serve as a resource for selecting mice to address the question. Using
301 mice from multiple sources helps model the interpersonal microbiota variation among humans to
302 aid our understanding of how the gut microbiota provides colonization resistance to CDIs.

303 **Acknowledgements**

304 This work was supported by the National Institutes of Health (U01AI124255). ST was supported by
305 the Michigan Institute for Clinical and Health Research Postdoctoral Translation Scholars Program
306 (UL1TR002240 from the National Center for Advancing Translational Sciences). We thank members
307 of the Schloss lab for feedback on planning the experiments and data presentation. In particular,
308 we thank Begüm Topçuoğlu for help with implementing logistic regression models, Ana Taylor for
309 help with media preparation and sample collection, and Nicholas Lesniak for his critical feedback
310 on the manuscript. We also thank members of Vincent Young's lab, particularly Kimberly Vendrov,
311 for guidance with the *C. difficile* infection mouse model and donating the mice. We also thank the
312 Unit for Laboratory Animal Medicine at the University of Michigan for maintaining our mouse colony
313 and providing the institutional support for our mouse experiments. Finally, we thank Kwi Kim, Austin
314 Campbell, and Kimberly Vendrov for their help in maintaining the Schloss lab's anaerobic chamber.

315 Materials and Methods

316 **(i) Animals.** All experiments were approved by the University of Michigan Animal Care and Use
317 Committee (IACUC) under protocol number PRO00006983. Female C57BL/7 mice were obtained
318 from 6 different sources: The Jackson Laboratory, Charles River Laboratories, Taconic Biosciences,
319 Envigo, and two colonies at the University of Michigan (the Schloss lab colony and the Young
320 lab colony). The Young lab colony was originally established with mice purchased from Jackson
321 in 2002, and the Schloss lab colony was established in 2010 with mice donated from the Young
322 lab. The 4 groups of mice purchased from vendors were allowed to acclimate to the University of
323 Michigan mouse facility for 13 days prior to starting the experiment. At least 4 female mice (age
324 5-10 weeks) were obtained per source and mice from the same source were primarily housed at a
325 density of 2 mice per cage. The experiment was repeated once, approximately 3 months after the
326 start of the first experiment.

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

330 **(iii) *C. difficile* infection model.** Mice were challenged with 10^3 spores of *C. difficile* strain 630
331 via oral gavage post-infection 1 day after clindamycin treatment as described previously (21). Mice
332 weights and stool samples were taken daily through 9 days post-challenge (Fig. 1A). Collected
333 stool was split for *C. difficile* quantification and 16S rRNA sequencing analysis. For *C. difficile*
334 quantification, stool samples were transferred to the anaerobic chamber, serially diluted in PBS,
335 plated on taurocholate-cycloserine-cefoxitin-fructose agar (TCCFA) plates, and counted after 24
336 hours of incubation at 37°C under anaerobic conditions. A sample from the day 0 timepoint
337 (post-clindamycin and prior to *C. difficile* challenge) was also plated on TCCFA to ensure mice
338 were not already colonized with *C. difficile* prior to infection. There were 3 deaths recorded over the
339 course of the experiment, 1 Taconic mouse died prior to *C. difficile* challenge and 1 Jackson and 1
340 Envigo mouse died between 1- and 3-days post-infection. Mice were categorized as cleared when
341 no *C. difficile* was detected in the first serial dilution (limit of detection: 100 CFU). Stool samples for
342 16S rRNA sequencing were snap frozen in liquid nitrogen and stored at -80°C until DNA extraction.

343 **(iv) 16S rRNA sequencing.** DNA was extracted from -80 °C stored stool samples using the DNeasy
344 Powersoil HTP 96 kit (Qiagen) and an EpMotion 5075 automated pipetting system (Eppendorf).
345 The V4 region was amplified for 16S rRNA with the AccuPrime Pfx DNA polymerase (Thermo
346 Fisher Scientific) using custom barcoded primers, as previously described (68). The ZymoBIOMICS
347 microbial community DNA standards was used as a mock community control (69) and water was
348 used as a negative control per 96-well extraction plate. The PCR amplicons were cleaned up
349 and normalized with the SequalPrep normalization plate kit (Thermo Fisher Scientific). Amplicons
350 were pooled and quantified with the KAPA library quantification kit (KAPA biosystems), prior to
351 sequencing using the MiSeq system (Illumina).

352 **(v) 16S rRNA gene sequence analysis.** mothur (v. 1.43) was used to process all sequences
353 (70) with a previously published protocol (68). Reads were combined and aligned with the SILVA
354 reference database (71). Chimeras were removed with the VSEARCH algorithm and taxonomic
355 assignment was completed with a modified version (v16) of the Ribosomal Database Project
356 reference database (v11.5) (72) with an 80% confidence cutoff. Operational taxonomic units (OTUs)
357 were assigned with a 97% similarity threshold using the opticlus algorithm (73). To account for
358 uneven sequencing across samples, samples were rarefied to 5,437 sequences 1,000 times for
359 alpha and beta diversity analyses, and a single time to generate relative abundances for model
360 training. PCoAs were generated based on θ_{YC} distances. Permutational multivariate analysis
361 of variance (PERMANOVA) was performed on mothur-generated θ_{YC} distance matrices with the
362 adonis function in the vegan package (74) in R (75).

363 **(vi) Classification model training and evaluation.** Models were generated based on mice that
364 were categorized as either cleared or colonized 7 days post-infection and had sequencing data
365 from the baseline (day -1), post-clindamycin (day 0), and post-infection (day 1) timepoints of the
366 experiment. Input bacterial community relative abundance data at the OTU level from the baseline,
367 post-clindamycin, and 1-day post-infection timepoints was used to generate 3 classification models
368 that predicted *C. difficile* colonization status 7 days post-infection. The L2-regularized logistic
369 regression models were trained and tested using the caret package (76) in R as previously
370 described (77) with the exception that we used 60% training and 40% testing data splits for the
371 cross-validation of the training data to select the best cost hyperparameter and the testing of

372 the held out test data to measure model performance. The modified training to testing ratio was
373 selected to accommodate the small number of samples in the dataset. Code was modified from
374 https://github.com/SchlossLab/ML_pipeline_microbiome to update the classification outcomes
375 and change the data split ratios. The modified repository to regenerate our modeling analysis is
376 available at https://github.com/tomkoset/ML_pipeline_microbiome.

377 **(vii) Statistical analysis.** All statistical tests were performed in R (v 4.0.2) (75). The Kruskal-Wallis
378 test was used to analyze differences in *C. difficile* CFU, mouse weight change, and alpha
379 diversity across sources with a Benjamini-Hochberg correction for testing multiple timepoints,
380 followed by pairwise Wilcoxon comparisons with Benjamini-Hochberg correction. For taxonomic
381 analysis and generation of logistic regression model input data, *C. difficile* (OTU 20) was removed.
382 Bacterial relative abundances that varied across sources at the OTU level were identified with the
383 Kruskal-Wallis test with Benjamini-Hochberg correction for testing all identified OTUs, followed by
384 pairwise Wilcoxon comparisons with Benjamini-Hochberg correction. The Wilcoxon rank sum test
385 was used to test for OTUs that differed between experiments within the Schloss, Young, and Envigo
386 sources with Benjamini-Hochberg correction for testing all identified OTUs. OTUs impacted by
387 clindamycin treatment were identified using the paired Wilcoxon signed rank test with matched
388 pairs of mice samples from day -1 and day 0. To determine whether classification models had
389 better performance (test AUROCs) than random chance (0.5), we used the one-sample Wilcoxon
390 signed rank test. To examine whether there was an overall difference in predictive performance
391 across the 3 classification models we used the Kruskal-Wallis test followed by pairwise Wilcoxon
392 comparisons with Benjamini-Hochberg correction for multiple hypothesis testing. The tidyverse
393 package (v 1.3.0) was used to wrangle and graph data (78).

394 **(viii) Code availability.** Code for all data analysis and generating this manuscript is available at
395 https://github.com/SchlossLab/Tomkovich_Vendor_XXXX_2020.

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

398 **References**

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

- 423 Alimentary Pharmacology & Therapeutics 44:715–727.
- 424 11. Pakpour S, Bhanvadia A, Zhu R, Amarnani A, Gibbons SM, Gurry T, Alm EJ, Martello LA. 2017.
425 Identifying predictive features of *Clostridium difficile* infection recurrence before, during, and after
426 primary antibiotic treatment. Microbiome 5.
- 427 12. Lee AA, Rao K, Limsrivilai J, Gilliland M, Malamet B, Briggs E, Young VB, Higgins PDR. 2020.
428 Temporal gut microbial changes predict recurrent *Clostridioides difficile* infection in patients with
429 and without ulcerative colitis. Inflammatory Bowel Diseases <https://doi.org/10.1093/ibd/izz335>.
- 430 13. Hutton ML, Mackin KE, Chakravorty A, Lyras D. 2014. Small animal models for the study of
431 *Clostridium difficile* disease pathogenesis. FEMS Microbiology Letters 352:140–149.
- 432 14. Chen X, Katchar K, Goldsmith JD, Nanthakumar N, Cheknis A, Gerding DN, Kelly CP. 2008. A
433 mouse model of *Clostridium difficile*-associated disease. Gastroenterology 135:1984–1992.
- 434 15. Best EL, Freeman J, Wilcox MH. 2012. Models for the study of *Clostridium difficile* infection.
435 Gut Microbes 3:145–167.
- 436 16. Baxter NT, Wan JJ, Schubert AM, Jenior ML, Myers P, Schloss PD. 2014. Intra- and
437 interindividual variations mask interspecies variation in the microbiota of sympatric peromyscus
438 populations. Applied and Environmental Microbiology 81:396–404.
- 439 17. Nagpal R, Wang S, Woods LCS, Seshie O, Chung ST, Shively CA, Register TC, Craft S,
440 McClain DA, Yadav H. 2018. Comparative microbiome signatures and short-chain fatty acids in
441 mouse, rat, non-human primate, and human feces. Frontiers in Microbiology 9.
- 442 18. Reeves AE, Theriot CM, Bergin IL, Huffnagle GB, Schloss PD, Young VB. 2011. The interplay
443 between microbiome dynamics and pathogen dynamics in a murine model of *Clostridium difficile*
444 infection 2:145–158.
- 445 19. Schubert AM, Sinani H, Schloss PD. 2015. Antibiotic-induced alterations of the murine gut
446 microbiota and subsequent effects on colonization resistance against *Clostridium difficile*. mBio 6.
- 447 20. Jenior ML, Leslie JL, Young VB, Schloss PD. 2017. *Clostridium difficile* colonizes alternative

- 448 nutrient niches during infection across distinct murine gut microbiomes. *mSystems* 2.
- 449 21. Jenior ML, Leslie JL, Young VB, Schloss PD. 2018. *Clostridium difficile* alters the structure and
450 metabolism of distinct cecal microbiomes during initial infection to promote sustained colonization.
451 *mSphere* 3.
- 452 22. Velazquez EM, Nguyen H, Heasley KT, Saechao CH, Gil LM, Rogers AWL, Miller BM, Rolston
453 MR, Lopez CA, Litvak Y, Liou MJ, Faber F, Bronner DN, Tiffany CR, Byndloss MX, Byndloss
454 AJ, Bäumler AJ. 2019. Endogenous Enterobacteriaceae underlie variation in susceptibility to
455 *Salmonella* infection. *Nature Microbiology* 4:1057–1064.
- 456 23. Osbelt L, Thiemann S, Smit N, Lesker TR, Schröter M, Gálvez EJC, Schmidt-Hohagen K, Pils
457 MC, Mühlen S, Dersch P, Hiller K, Schlüter D, Neumann-Schaal M, Strowig T. 2020. Variations in
458 microbiota composition of laboratory mice influence *Citrobacter rodentium* infection via variable
459 short-chain fatty acid production. *PLOS Pathogens* 16:e1008448.
- 460 24. Stough JMA, Dearth SP, Denny JE, LeCleir GR, Schmidt NW, Campagna SR, Wilhelm SW.
461 2016. Functional characteristics of the gut microbiome in C57BL/6 mice differentially susceptible to
462 *Plasmodium yoelii*. *Frontiers in Microbiology* 7.
- 463 25. Alegre M-L. 2019. Mouse microbiomes: Overlooked culprits of experimental variability. *Genome
464 Biology* 20.
- 465 26. Etienne-Mesmin L, Chassaing B, Adekunle O, Mattei LM, Bushman FD, Gewirtz AT. 2017.
466 Toxin-positive *Clostridium difficile* latently infect mouse colonies and protect against highly
467 pathogenic *C. difficile*. *Gut* 67:860–871.
- 468 27. Lai NY, Musser MA, Pinho-Ribeiro FA, Baral P, Jacobson A, Ma P, Potts DE, Chen Z, Paik D,
469 Soualhi S, Yan Y, Misra A, Goldstein K, Lagomarsino VN, Nordstrom A, Sivanathan KN, Wallrapp A,
470 Kuchroo VK, Nowarski R, Starnbach MN, Shi H, Surana NK, An D, Wu C, Huh JR, Rao M, Chiu IM.
471 2020. Gut-innervating nociceptor neurons regulate peyer's patch microfold cells and SFB levels to
472 mediate *Salmonella* host defense. *Cell* 180:33–49.e22.
- 473 28. Thiemann S, Smit N, Roy U, Lesker TR, Gálvez EJC, Helmecke J, Basic M, Bleich A, Goodman

- 474 474 AL, Kalinke U, Flavell RA, Erhardt M, Strowig T. 2017. Enhancement of IFNgamma production by
475 distinct commensals ameliorates *Salmonella*-induced disease. *Cell Host & Microbe* 21:682–694.e5.
- 476 476 29. Rolig AS, Cech C, Ahler E, Carter JE, Ottemann KM. 2013. The degree of *Helicobacter*
477 *pylori*-triggered inflammation is manipulated by preinfection host microbiota. *Infection and Immunity*
478 81:1382–1389.
- 479 479 30. Ge Z, Sheh A, Feng Y, Muthupalani S, Ge L, Wang C, Kurnick S, Mannion A, Whary MT, Fox
480 JG. 2018. *Helicobacter pylori*-infected C57BL/6 mice with different gastrointestinal microbiota have
481 contrasting gastric pathology, microbial and host immune responses. *Scientific Reports* 8.
- 482 482 31. Lawley TD, Young VB. 2013. Murine models to study *Clostridium difficile* infection and
483 transmission. *Anaerobe* 24:94–97.
- 484 484 32. Buffie CG, Jarchum I, Equinda M, Lipuma L, Gobourne A, Viale A, Ubeda C, Xavier J, Pamer
485 EG. 2011. Profound alterations of intestinal microbiota following a single dose of clindamycin results
486 in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infection and Immunity* 80:62–73.
- 487 487 33. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew
488 M, Viale A, Littmann E, Brink MRM van den, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint
489 NC, Xavier JB, Pamer EG. 2014. Precision microbiome reconstitution restores bile acid mediated
490 resistance to *Clostridium difficile*. *Nature* 517:205–208.
- 491 491 34. Spinler JK, Brown A, Ross CL, Boonma P, Conner ME, Savidge TC. 2016. Administration of
492 probiotic kefir to mice with *Clostridium difficile* infection exacerbates disease. *Anaerobe* 40:54–57.
- 493 493 35. Markey L, Shaban L, Green ER, Lemon KP, Mecsas J, Kumamoto CA. 2018. Pre-colonization
494 with the commensal fungus candida albicans reduces murine susceptibility to *Clostridium difficile*
495 infection. *Gut Microbes* 1–13.
- 496 496 36. McKee RW, Aleksanyan N, Garrett EM, Tamayo R. 2018. Type IV pili promote *Clostridium*
497 *difficile* adherence and persistence in a mouse model of infection. *Infection and Immunity* 86.
- 498 498 37. Yamaguchi T, Konishi H, Aoki K, Ishii Y, Chono K, Tateda K. 2020. The gut microbiome diversity

- 499 of *Clostridioides difficile*-inoculated mice treated with vancomycin and fidaxomicin. Journal of
500 Infection and Chemotherapy 26:483–491.
- 501 38. Stroke IL, Letourneau JJ, Miller TE, Xu Y, Pechik I, Savoly DR, Ma L, Sturzenbecker LJ,
502 Sabalski J, Stein PD, Webb ML, Hilbert DW. 2018. Treatment of *Clostridium difficile* infection
503 with a small-molecule inhibitor of toxin UDP-glucose hydrolysis activity. Antimicrobial Agents and
504 Chemotherapy 62.
- 505 39. Quigley L, Coakley M, Alemayehu D, Rea MC, Casey PG, O'Sullivan, Murphy E, Kiely B, Cotter
506 PD, Hill C, Ross RP. 2019. *Lactobacillus gasseri* APC 678 reduces shedding of the pathogen
507 *Clostridium difficile* in a murine model. Frontiers in Microbiology 10.
- 508 40. Mullish BH, McDonald JAK, Pechlivanis A, Allegretti JR, Kao D, Barker GF, Kapila D, Petrof
509 EO, Joyce SA, Gahan CGM, Glegola-Madejska I, Williams HRT, Holmes E, Clarke TB, Thursz
510 MR, Marchesi JR. 2019. Microbial bile salt hydrolases mediate the efficacy of faecal microbiota
511 transplant in the treatment of recurrent *Clostridioides difficile* infection. Gut 68:1791–1800.
- 512 41. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. 2015. How informative is the mouse for human
513 gut microbiota research? Disease Models & Mechanisms 8:1–16.
- 514 42. Guh AY, Kutty PK. 2018. *Clostridioides difficile* infection 169:ITC49.
- 515 43. Tomkovich S, Lesniak NA, Li Y, Bishop L, Fitzgerald MJ, Schloss PD. 2019. The proton
516 pump inhibitor omeprazole does not promote *Clostridioides difficile* colonization in a murine model.
517 mSphere 4.
- 518 44. Lawley TD, Clare S, Walker AW, Stares MD, Connor TR, Raisen C, Goulding D, Rad R,
519 Schreiber F, Brandt C, Deakin LJ, Pickard DJ, Duncan SH, Flint HJ, Clark TG, Parkhill J, Dougan
520 G. 2012. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy
521 resolves relapsing *Clostridium difficile* disease in mice. PLoS Pathogens 8:e1002995.
- 522 45. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, Mastroeni P, Scott P,
523 Raisen C, Mottram L, Fairweather NF, Wren BW, Parkhill J, Dougan G. 2009. Antibiotic treatment of
524 *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and

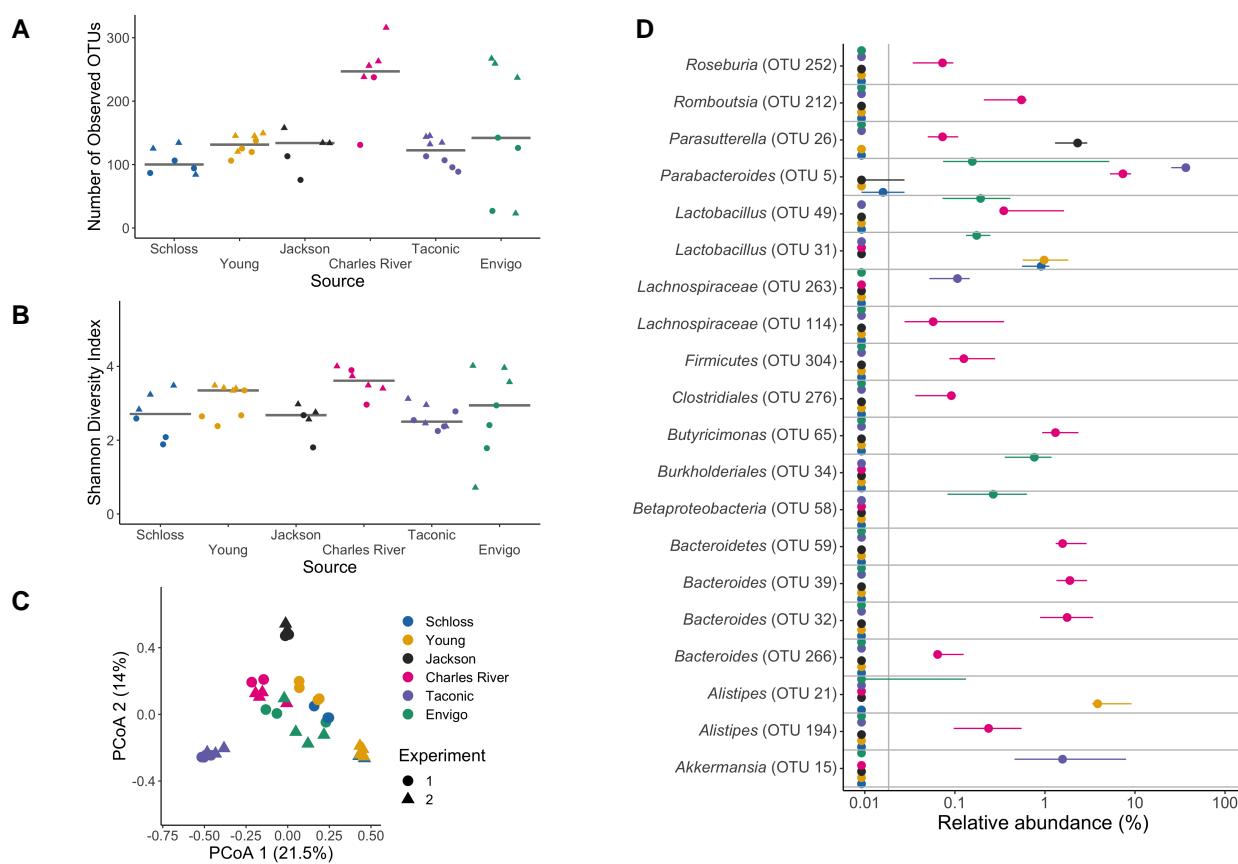
- 525 severe disease in immunocompromised hosts. *Infection and Immunity* 77:3661–3669.
- 526 46. Jump RLP, Polinkovsky A, Hurless K, Sitzlar B, Eckart K, Tomas M, Deshpande A, Nerandzic
527 MM, Donskey CJ. 2014. Metabolomics analysis identifies intestinal microbiota-derived biomarkers
528 of colonization resistance in clindamycin-treated mice. *PLoS ONE* 9:e101267.
- 529 47. Nagao-Kitamoto H, Leslie JL, Kitamoto S, Jin C, Thomsson KA, Gilliland MG, Kuffa P, Goto Y,
530 Jenq RR, Ishii C, Hirayama A, Seekatz AM, Martens EC, Eaton KA, Kao JY, Fukuda S, Higgins PDR,
531 Karlsson NG, Young VB, Kamada N. 2020. Interleukin-22-mediated host glycosylation prevents
532 *Clostridioides difficile* infection by modulating the metabolic activity of the gut microbiota. *Nature
533 Medicine* 26:608–617.
- 534 48. Battaglioli EJ, Hale VL, Chen J, Jeraldo P, Ruiz-Mojica C, Schmidt BA, Rekdal VM, Till LM, Huq
535 L, Smits SA, Moor WJ, Jones-Hall Y, Smyrk T, Khanna S, Pardi DS, Grover M, Patel R, Chia N,
536 Nelson H, Sonnenburg JL, Farrugia G, Kashyap PC. 2018. *Clostridioides difficile* uses amino acids
537 associated with gut microbial dysbiosis in a subset of patients with diarrhea. *Science Translational
538 Medicine* 10:eaam7019.
- 539 49. Robinson CD, Auchtung JM, Collins J, Britton RA. 2014. Epidemic *Clostridium difficile* strains
540 demonstrate increased competitive fitness compared to nonepidemic isolates. *Infection and
541 Immunity* 82:2815–2825.
- 542 50. Collins J, Auchtung JM, Schaefer L, Eaton KA, Britton RA. 2015. Humanized microbiota mice
543 as a model of recurrent *Clostridium difficile* disease. *Microbiome* 3.
- 544 51. Collins J, Robinson C, Danhof H, Knetsch CW, Leeuwen HC van, Lawley TD, Auchtung JM,
545 Britton RA. 2018. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature
546* 553:291–294.
- 547 52. Hryckowian AJ, Treuren WV, Smits SA, Davis NM, Gardner JO, Bouley DM, Sonnenburg JL.
548 2018. Microbiota-accessible carbohydrates suppress *Clostridium difficile* infection in a murine
549 model. *Nature Microbiology* 3:662–669.
- 550 53. Fouladi F, Glenny EM, Bulik-Sullivan EC, Tsilimigras MCB, Sioda M, Thomas SA, Wang Y, Djukic

- 551 Z, Tang Q, Tarantino LM, Bulik CM, Fodor AA, Carroll IM. 2020. Sequence variant analysis reveals
552 poor correlations in microbial taxonomic abundance between humans and mice after gnotobiotic
553 transfer. *The ISME Journal* <https://doi.org/10.1038/s41396-020-0645-z>.
- 554 54. Walter J, Armet AM, Finlay BB, Shanahan F. 2020. Establishing or exaggerating causality for
555 the gut microbiome: Lessons from human microbiota-associated rodents. *Cell* 180:221–232.
- 556 55. Rasmussen TS, Vries L de, Kot W, Hansen LH, Castro-Mejía JL, Vogensen FK, Hansen AK,
557 Nielsen DS. 2019. Mouse vendor influence on the bacterial and viral gut composition exceeds the
558 effect of diet. *Viruses* 11:435.
- 559 56. Mims TS, Abdallah QA, Watts S, White C, Han J, Willis KA, Pierre JF. 2020. Variability in
560 interkingdom gut microbiomes between different commercial vendors shapes fat gain in response
561 to diet. *The FASEB Journal* 34:1–1.
- 562 57. Stewart DB, Wright JR, Fowler M, McLimans CJ, Tokarev V, Amaniera I, Baker O, Wong H-T,
563 Brabec J, Drucker R, Lamendella R. 2019. Integrated meta-omics reveals a fungus-associated
564 bacteriome and distinct functional pathways in *Clostridioides difficile* infection. *mSphere* 4.
- 565 58. Ott SJ, Waetzig GH, Rehman A, Moltzau-Anderson J, Bharti R, Grasis JA, Cassidy L, Tholey A,
566 Fickenscher H, Seegert D, Rosenstiel P, Schreiber S. 2017. Efficacy of sterile fecal filtrate transfer
567 for treating patients with *Clostridium difficile* infection. *Gastroenterology* 152:799–811.e7.
- 568 59. Zuo T, Wong SH, Lam K, Lui R, Cheung K, Tang W, Ching JYL, Chan PKS, Chan MCW, Wu
569 JCY, Chan FKL, Yu J, Sung JJY, Ng SC. 2017. Bacteriophage transfer during faecal microbiota
570 transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut*
571 *gutjnl–2017–313952*.
- 572 60. Zuo T, Wong SH, Cheung CP, Lam K, Lui R, Cheung K, Zhang F, Tang W, Ching JYL, Wu JCY,
573 Chan PKS, Sung JJY, Yu J, Chan FKL, Ng SC. 2018. Gut fungal dysbiosis correlates with reduced
574 efficacy of fecal microbiota transplantation in *Clostridium difficile* infection. *Nature Communications*
575 9.
- 576 61. Robinson JI, Weir WH, Crowley JR, Hink T, Reske KA, Kwon JH, Burnham C-AD, Dubberke

- 577 577 ER, Mucha PJ, Henderson JP. 2019. Metabolomic networks connect host-microbiome processes to
578 human *Clostridioides difficile* infections. *Journal of Clinical Investigation* 129:3792–3806.
- 579 579 62. Fletcher JR, Erwin S, Lanzas C, Theriot CM. 2018. Shifts in the gut metabolome and *Clostridium*
580 *difficile* transcriptome throughout colonization and infection in a mouse model. *mSphere* 3.
- 581 581 63. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, Li X, Long H, Zhang J, Zhang D, Liu C, Fang
582 Z, Chou J, Glanville J, Hao Q, Kotowska D, Colding C, Licht TR, Wu D, Yu J, Sung JJY, Liang Q, Li
583 J, Jia H, Lan Z, Tremaroli V, Dworzynski P, Nielsen HB, Bäckhed F, Doré J, Chatelier EL, Ehrlich
584 SD, Lin JC, Arumugam M, Wang J, Madsen L, Kristiansen K. 2015. A catalog of the mouse gut
585 metagenome. *Nature Biotechnology* 33:1103–1108.
- 586 586 64. Fransen F, Zagato E, Mazzini E, Fosso B, Manzari C, Aidy SE, Chiavelli A, D'Erchia AM,
587 Sethi MK, Pabst O, Marzano M, Moretti S, Romani L, Penna G, Pesole G, Rescigno M. 2015.
588 BALB/c and C57BL/6 mice differ in polyreactive IgA abundance, which impacts the generation of
589 antigen-specific IgA and microbiota diversity. *Immunity* 43:527–540.
- 590 590 65. Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee
591 CA, Lynch SV, Tanoue T, Imaoka A, Itoh K, Takeda K, Umesaki Y, Honda K, Littman DR. 2009.
592 Induction of intestinal th17 cells by segmented filamentous bacteria. *Cell* 139:485–498.
- 593 593 66. Azrad M, Hamo Z, Tkawkho L, Peretz A. 2018. Elevated serum immunoglobulin a levels in
594 patients with *Clostridium difficile* infection are associated with mortality. *Pathogens and Disease* 76.
- 595 595 67. Saleh MM, Frisbee AL, Leslie JL, Buonomo EL, Cowardin CA, Ma JZ, Simpson ME, Scully KW,
596 Abhyankar MM, Petri WA. 2019. Colitis-induced th17 cells increase the risk for severe subsequent
597 *Clostridium difficile* infection. *Cell Host & Microbe* 25:756–765.e5.
- 598 598 68. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a
599 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the
600 MiSeq illumina sequencing platform. *Applied and Environmental Microbiology* 79:5112–5120.
- 601 601 69. Sze MA, Schloss PD. 2019. The impact of DNA polymerase and number of rounds of
602 amplification in PCR on 16S rRNA gene sequence data. *mSphere* 4.

- 603 70. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
604 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF.
605 2009. Introducing mothur: Open-source, platform-independent, community-supported software
606 for describing and comparing microbial communities. *Applied and Environmental Microbiology*
607 75:7537–7541.
- 608 71. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012.
609 The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools.
610 *Nucleic Acids Research* 41:D590–D596.
- 611 72. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR,
612 Tiedje JM. 2013. Ribosomal database project: Data and tools for high throughput rRNA analysis.
613 *Nucleic Acids Research* 42:D633–D642.
- 614 73. Westcott SL, Schloss PD. 2017. OptiClust, an improved method for assigning amplicon-based
615 sequence data to operational taxonomic units. *mSphere* 2.
- 616 74. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB,
617 Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2018. Vegan: Community ecology
618 package.
- 619 75. R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for
620 Statistical Computing, Vienna, Austria.
- 621 76. Kuhn M. 2008. Building predictive models in RUsing the caret Package. *Journal of Statistical*
622 *Software* 28.
- 623 77. Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. 2020. A framework for effective
624 application of machine learning to microbiome-based classification problems. *mBio* 11.
- 625 78. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes
626 A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D,
627 Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the
628 tidyverse. *Journal of Open Source Software* 4:1686.

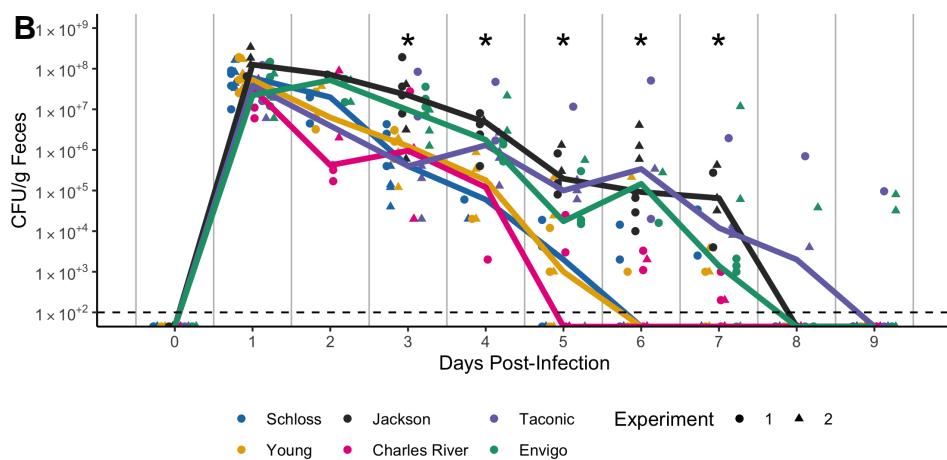
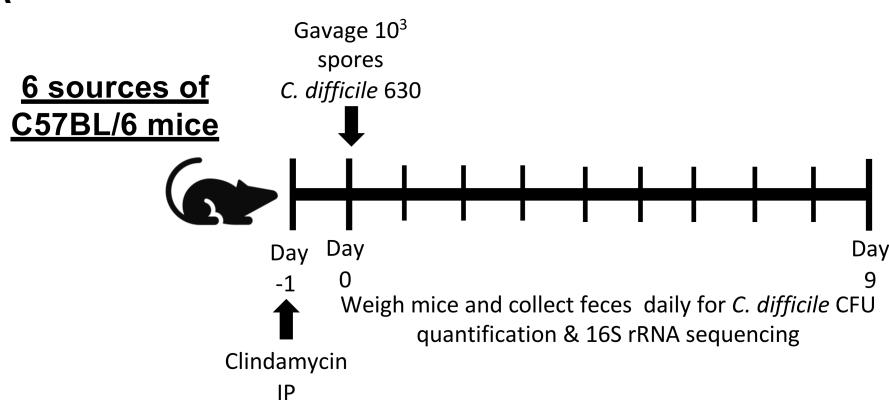
629 **Figures**



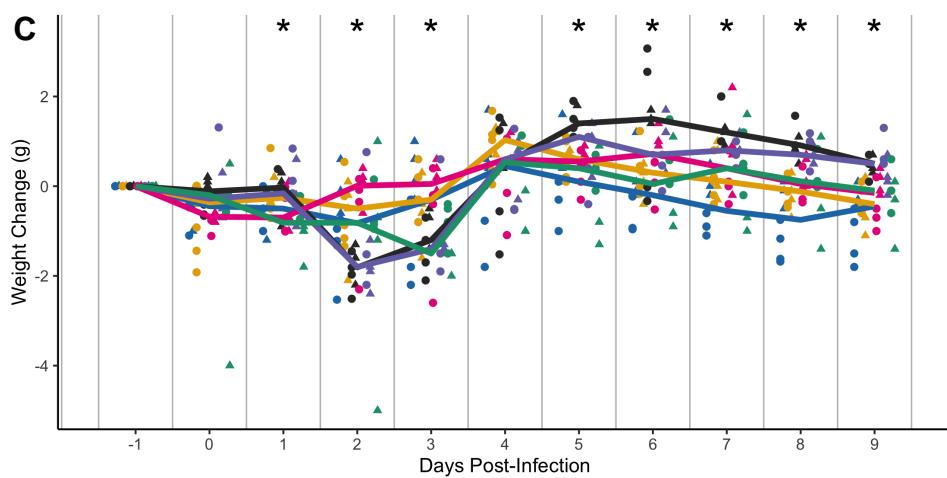
630

631 **Figure 1. Microbiota variation is high between mice from different sources.** A-B. Number of
 632 observed OTUs (A) and Shannon diversity index values (B) across sources of mice at baseline
 633 (day -1 of the experiment). Differences between sources were analyzed by Kruskal-Wallis test
 634 with Benjamini-Hochberg correction for testing each day of the experiment and the adjusted P
 635 value was < 0.05 for panel A (Table S1). None of the P values from pairwise Wilcoxon comparisons
 636 between sources were significant after Benjamini-Hochberg correction (Table S2). Gray lines
 637 represent the median values for each source of mice. C. Principal Coordinates Analysis (PCoA) of
 638 θ_{YC} distances of baseline stool samples. Source and the interaction between source and cage
 639 effects explained most of the variation (PERMANOVA combined $R^2 = 0.90$, $P < 0.001$; Table S3).
 640 For A-C: each symbol represents the value for a stool sample from an individual mouse, circles
 641 represent experiment 1 mice and triangles represent experiment 2 mice. D. The median (point) and
 642 interquartile range (colored lines) of the relative abundances for the 20 most significant OTUs out
 643 of the 268 OTUs that varied across sources at baseline (Table S5).

A



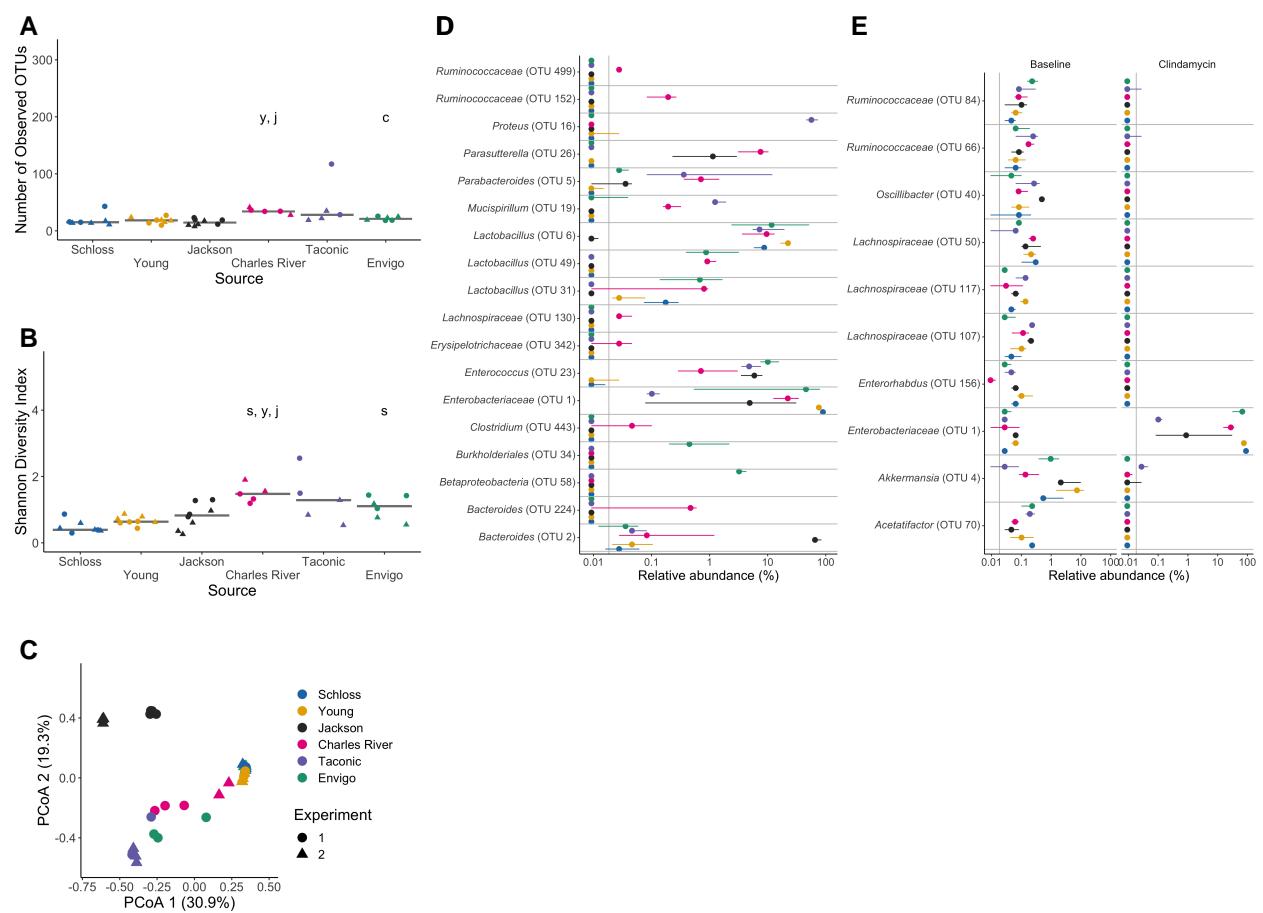
C



644

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

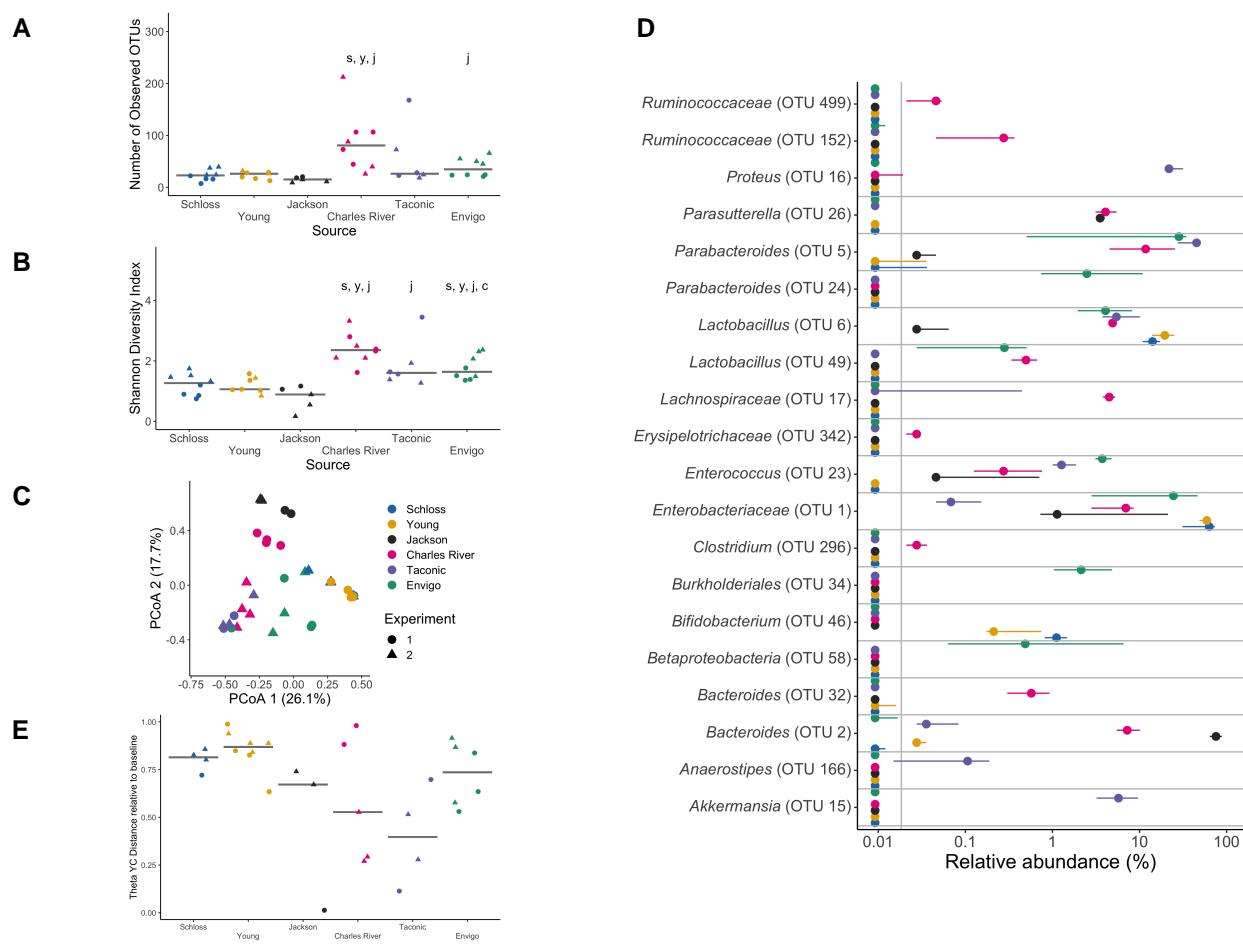
649 (N = 8), Taconic Biosciences (N = 8), and Envigo (N = 8). All mice were administered 10 mg/kg
650 clindamycin intraperitoneally (IP) 1 day before challenge with *C. difficile* 630 spores on day 0.
651 Mice were weighed and feces was collected daily through the end of the experiment (9 days
652 post-infection). Note: 3 mice died during course of experiment. 1 Taconic mouse prior to infection
653 and 1 Jackson and 1 Envigo mouse between 1- and 3-days post-infection. B. *C. difficile* CFU/gram
654 stool measured over time (N = 20-49 mice per timepoint) via serial dilutions. The black line
655 represents the limit of detection for the first serial dilution. CFU quantification data was not available
656 for each mouse due to early deaths, stool sampling difficulties, and not plating all of the serial
657 dilutions. C. Mouse weight change measured in grams over time (N = 45-49 mice per timepoint),
658 all mice were normalized to the weight recorded 1 day before infection. For B-C: timepoints
659 where differences between sources of mice were statistically significant by Kruskal-Wallis test
660 with Benjamini-Hochberg correction for testing across multiple days (Table S6 and Table S7) are
661 reflected by the asterisk above each timepoint (*, $P < 0.05$). Lines represent the median for each
662 source and circles represent individual mice from experiment 1 while triangles represent mice from
663 experiment 2.



664

665 **Figure 3. Clindamycin treatment alters bacteria in all sources, but a subset of bacterial**
 666 **differences across sources persists.** A-B. Number of observed OTUs (A) and Shannon diversity
 667 index values (B) across sources of mice after clindamycin treatment (day 0). Differences between
 668 sources were analyzed by Kruskal-Wallis test with Benjamini-Hochberg correction for testing each
 669 day of the experiment and the adjusted P value was < 0.05 (Table S1). Significant P values from
 670 the pairwise Wilcoxon comparisons between sources with Benjamini-Hochberg correction are
 671 displayed as the first initial of each group compared to the group that they are listed above (Table
 672 S2). C. PCoA of θ_{YC} distances from stools collected post-clindamycin. Source and the interaction
 673 between source and cage effects explained most of the variation observed post-clindamycin
 674 (PERMANOVA combined $R^2 = 0.99$, $P < 0.001$; Table S3). For A-C, each symbol represents a
 675 stool sample from an individual mouse, with circles representing experiment 1 mice and triangles
 676 representing experiment 2 mice. D. The median (point) and interquartile range (colored lines) of
 677 the relative abundances for the 18 OTUs (Table S8) that varied between sources after clindamycin

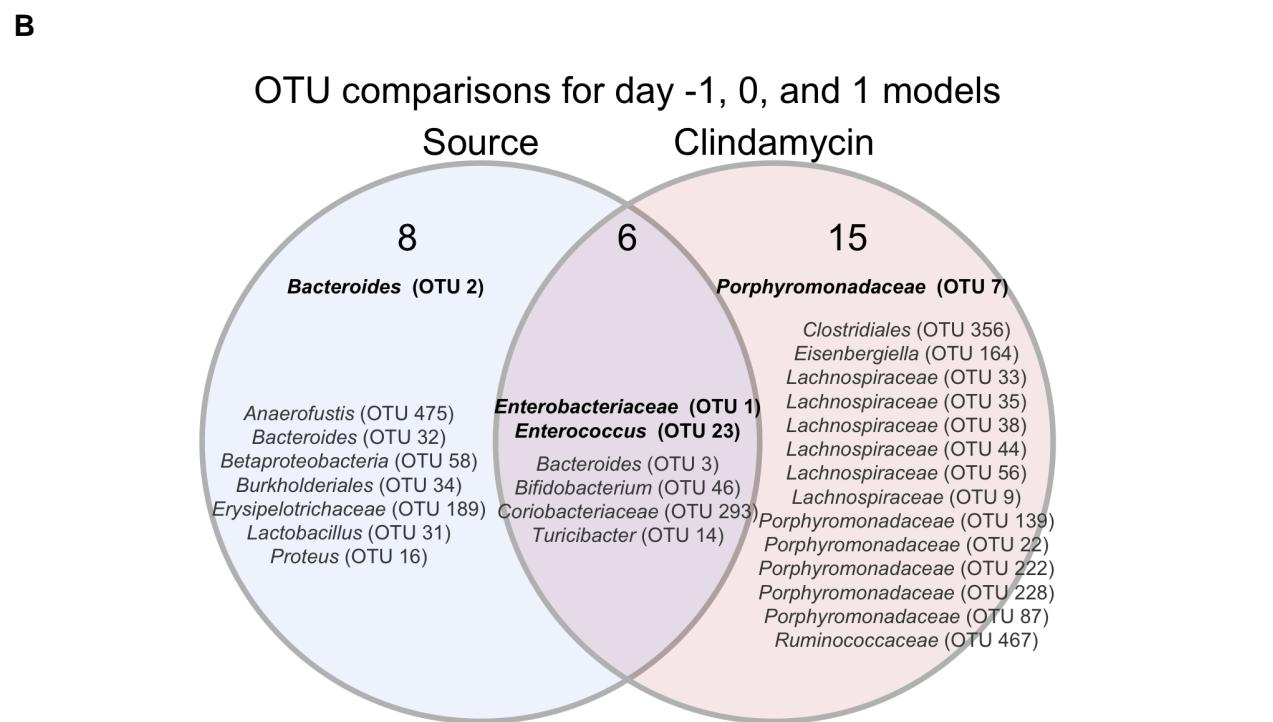
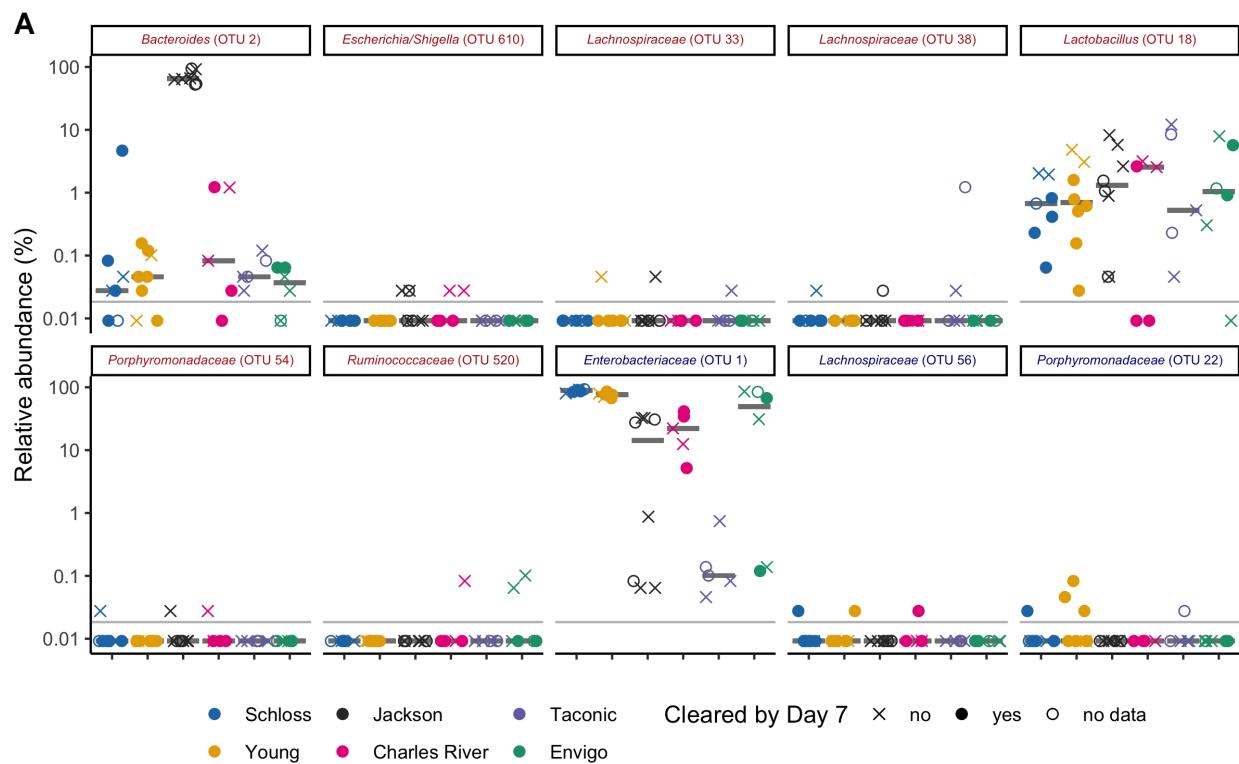
678 treatment (day 0). E. The median (point) and interquartile range (colored lines) of the top 10 most
679 significant OTUs out of 153 with relative abundances that changed because of the clindamycin
680 treatment (adjusted *P* value < 0.05). Data were analyzed by paired Wilcoxon signed rank test of
681 mice that had paired sequence data for baseline (day -1) and post-clindamycin (day 0) timepoints
682 ($N = 31$), with Benjamini-Hochberg correction for testing all identified OTUs (Table S9). The gray
683 vertical line indicates the limit of detection.



684

685 **Figure 4. Microbiota variation across sources is maintained after *C. difficile* challenge.** A-B.
 686 Number of observed OTUs (A) and Shannon diversity index values (B) across sources of mice
 687 1-day post-infection. Data were analyzed by Kruskal-Wallis test with Benjamini-Hochberg correction
 688 for testing each day of the experiment and the adjusted *P* value was < 0.05 (Table S1). Significant
 689 *P* values from the pairwise Wilcoxon comparisons between sources with Benjamini-Hochberg
 690 correction are displayed as the first initial of each group compared to the group that they are listed
 691 above (Table S2). PCoA of θ_{YC} distances of 1-day post-infection stool samples. Source and
 692 the interaction between source and cage effects explained most of the variation between fecal
 693 communities (PERMANOVA combined $R^2 = 0.88$, $P < 0.001$; Table S3). For A-C: each symbol
 694 represents the value for a stool sample from an individual mouse, circles represent experiment 1
 695 mice and triangles represent experiment 2 mice. D. The median (point) and interquartile range
 696 (colored lines) of the relative abundances for the top 20 most significant OTUs out of the 44 OTUs
 697 that varied between sources 1-day post-infection. The gray vertical line indicates the limit of

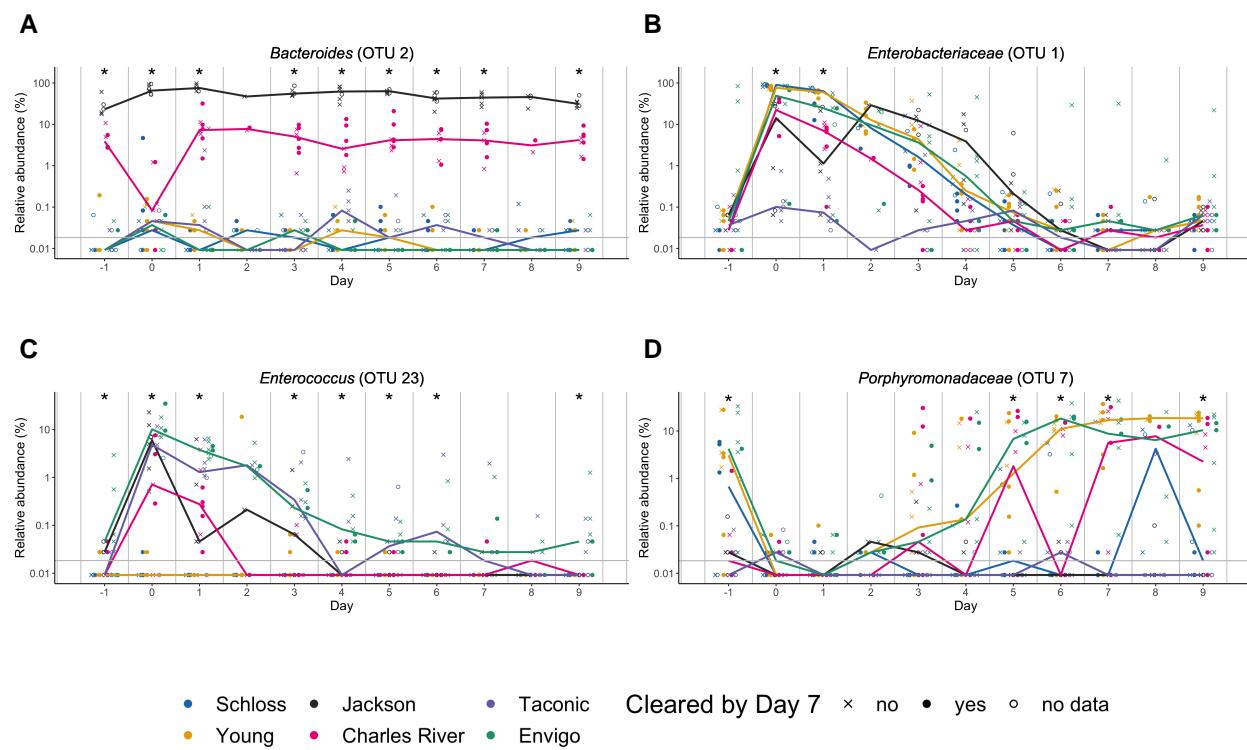
698 detection. For each timepoint OTUs with differential relative abundances across sources of mice
699 were identified by Kruskal-Wallis test with Benjamini-Hochberg correction for testing all identified
700 OTUs (Table S10). E. θ_{YC} distances of fecal samples collected 7-days post-infection relative to the
701 baseline (day -1) sample for each mouse. Each symbol represents an individual mouse. Gray lines
702 represent the median for each source.



703

704 **Figure 5. Bacteria that influenced whether mice cleared *C. difficile* by day 7. A.**
705 Post-clindamycin (day 0) relative abundance data for the 10 OTUs with the highest rankings

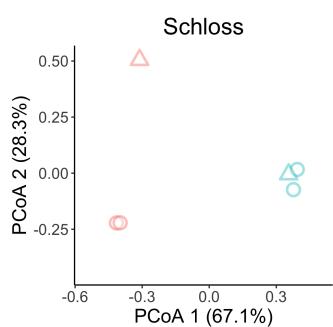
706 based on feature weights in the post-clindamycin (day 0) classification model. Red font represents
707 OTUs that correlated with *C. difficile* colonization and blue font represents OTUs that correlated
708 with clearance. Symbols represent the relative abundance data for an individual mouse. Gray
709 bars indicate the median relative abundances for each source. B. Venn diagram that combines
710 OTUs that were important to the day -1, 0, and 1 classification models (Fig. S4, Table S14) and
711 either overlapped with taxa that varied across sources at the same timepoint, were impacted by
712 clindamycin treatment, or both. Bold OTUs were important to more than 1 classification model.



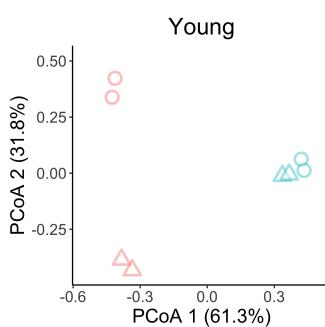
713

714 **Figure 6: OTUs associated with *C. difficile* colonization dynamics vary across sources**
715 **throughout the experiment.** A-D. Relative abundances of bold OTUs from Fig. 5B that were
716 important in at least two classification models are shown over time. A. *Bacteroides* (OTU 2), which
717 varied across sources throughout the experiment. B-C. *Enterobacteriaceae* (B) and *Enterococcus*
718 (C), which significantly varied across sources and were impacted by clindamycin treatment. D.
719 *Porphyromonadaceae* (OTU 7), which was significantly impacted by clindamycin treatment and
720 after examining relative abundance dynamics over the course of the experiment was found to
721 also significantly vary between sources of mice on days -1, 5, 6, 7, and 9 of the experiment.
722 Symbols represent the relative abundance data for an individual mouse. Colored lines indicate
723 the median relative abundances for each source. The gray horizontal line represents the limit of
724 detection. Timepoints where differences between sources of mice were statistically significant by
725 Kruskal-Wallis test with Benjamini-Hochberg correction for testing across multiple days (Table S15)
726 are identified by the asterisk above each timepoint (*, P < 0.05).

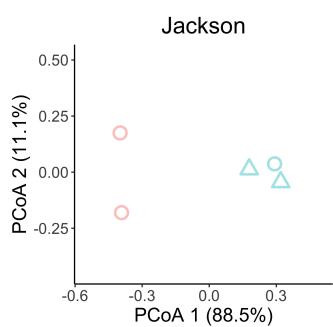
A



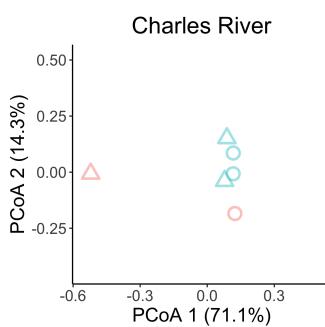
B



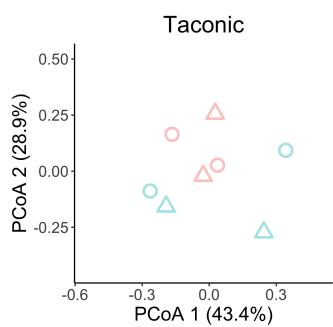
C



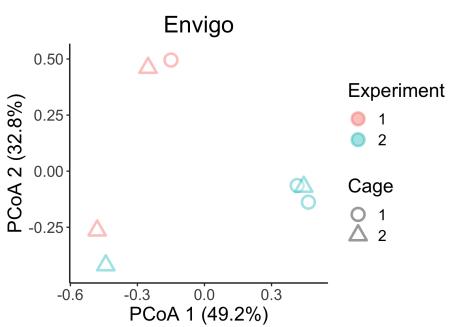
D



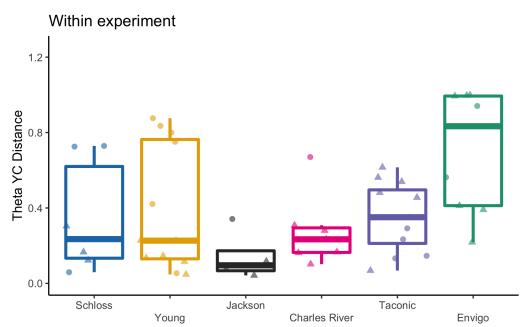
E



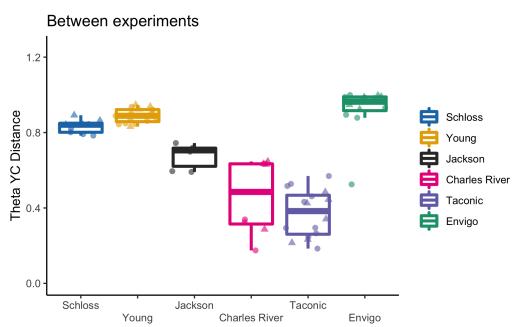
F



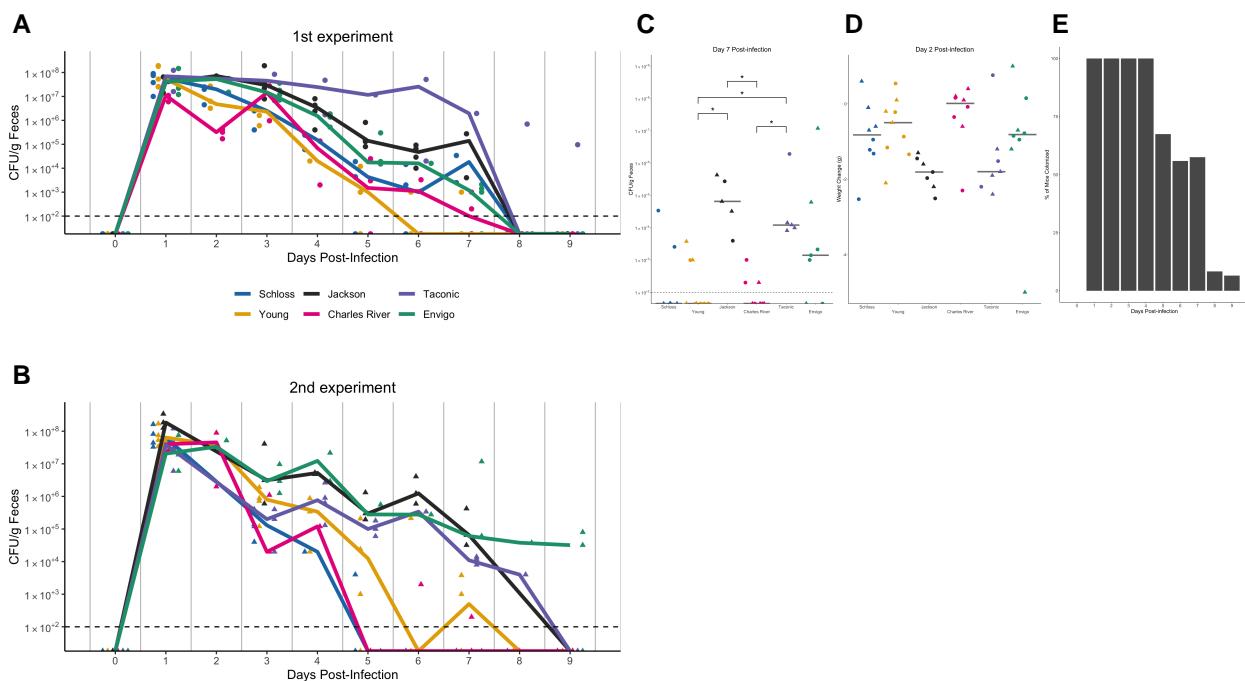
G



H



728 **Figure S1. Bacterial communities vary between experiments for some sources.** A-F. PCoA
729 of θ_{YC} distances for the baseline fecal bacterial communities within each source of mice. Each
730 symbol represents a stool sample from an individual mouse with color corresponding to experiment
731 and shape representing cage mates. Experiment number and cage effects explained most of the
732 observed variation for samples from the Schloss (PERMANOVA combined $R^2 = 0.99$; $P \leq 0.033$)
733 and Young (combined $R^2 = 0.95$; $P \leq 0.03$) mice (Table S4). G-H: Boxplots of the θ_{YC} distances of
734 the 6 sources of mice relative to mice within the same source and experiment (G) or mice within
735 the same source and between experiments (H) at baseline (day -1). Symbols represent individual
736 mouse samples: circles for experiment 1 and triangles for experiment 2.

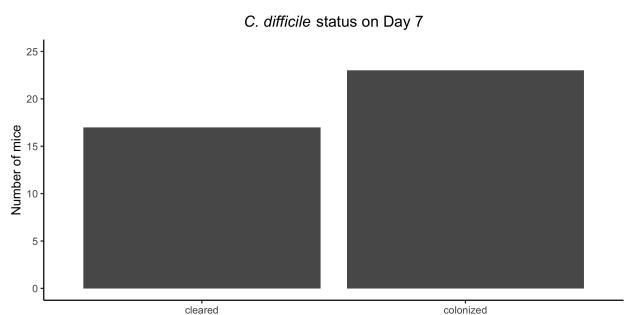


737

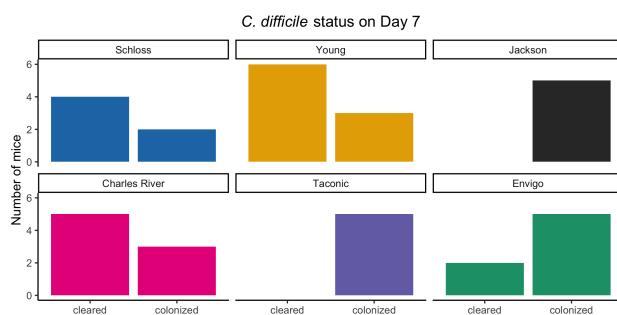
738 **Figure S2. *C. difficile* CFU variation across sources varies slightly between the 2**
 739 **experiments.** A-B. *C. difficile* CFU/gram of stool quantification over time for experiment 1 (A) and

740 2 (B). Experiments were conducted approximately 3 months apart. Lines represent the median
 741 CFU for each source, symbols represent individual mice and the black line represents the limit
 742 of detection. C. *C. difficile* CFU/gram stool 7-days post-infection across sources of mice with an
 743 asterisk for pairwise Wilcoxon comparisons with Benjamini-Hochberg correction where $P < 0.05$.
 744 D. Mouse weight change 2-days post-infection across sources of mice, no pairwise Wilcoxon
 745 comparisons were significant after Benjamini-Hochberg correction. For C-D: circles represent
 746 experiment 1 mice, triangles represent experiment 2 mice and gray lines indicate the median
 747 values for each group. E. Percent of mice that were colonized with *C. difficile* over the course of the
 748 experiment. Each day the percent is calculated based on the mice where *C. difficile* CFU was
 749 quantified for that particular day. Total N for each day: day 1 (N = 42), day 2 (N = 20), day 3 (N =
 750 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 =
 751 46).

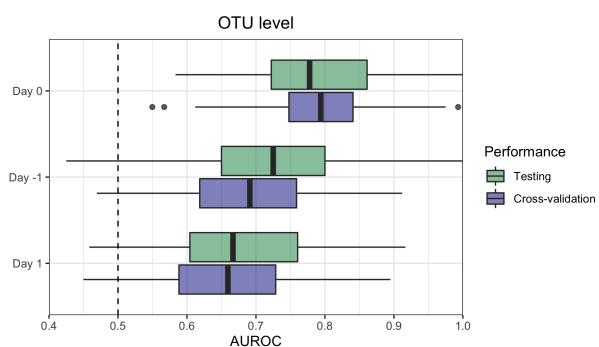
A



B

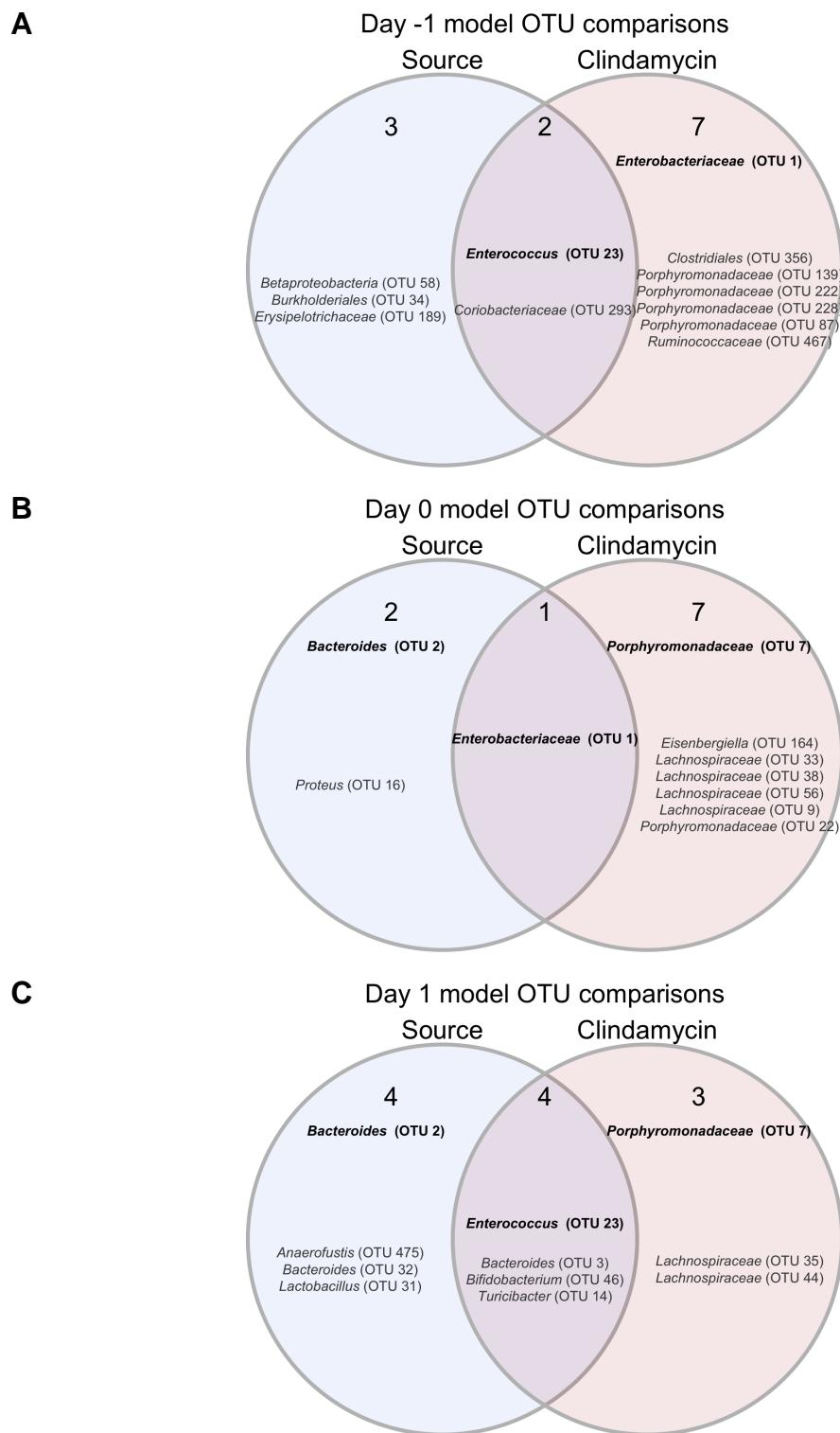


C



752

753 **Figure S3. Bacterial community composition before, after clindamycin perturbation, and**
754 **post-infection can predict *C. difficile* colonization status 7 days post-challenge.** A. Bar
755 graph visualizations of overall 7-days post-infection *C. difficile* colonization status that were used as
756 classification outcomes to build L2-regularized logistic regression models. Mice were classified as
757 colonized or cleared (not detectable at the limit of detection of 100 CFU) based on CFU g/stool data
758 from 7 days post-infection. B. *C. difficile* CFU status on Day 7 within each mouse source. N = 8-9
759 mice per group. C. L2-regularized logistic regression classification model area under the receiving
760 operator characteristic curve (AUROCs) to predict *C. difficile* CFU on day 7 post-infection (Fig. 2B,
761 Fig. S2C) based on the OTU community relative abundances at baseline (day -1), post-clindamycin
762 (day 0), and 1-day post-infection. All models performed better than random chance (AUROC =
763 0.5, all $P < 0.001$, Table S12) and the model built with post-clindamycin bacterial OTU relative
764 abundances had the best performance ($P_{FDR} < 0.001$ for all pairwise comparisons, Table S13).
765 See Table S14 for list of the 20 OTUs that were ranked as most important to each model.



766

767 **Figure S4. OTUs from classification models based on baseline, post-clindamycin treatment,**
768 **or 1-day post-infection community data vary by source, clindamycin treatment, or both. A-C.**

769 Venn diagrams of OTUs from the top 20 OTUs from the baseline (A), post-clindamycin treatment (B),
770 and 1-day post-infection (C) classification models (Table S14) that overlapped with OTUs that varied
771 across sources at the corresponding timepoint (Tables S5, 8, 10), were impacted by clindamycin
772 treatment (Table S9), or both. Bold OTUs were important to more than 1 classification model.

773 **Supplementary Tables and Movie**

774 **Movie S1. Large shifts in bacterial community structures occurred after clindamycin and**
775 ***C. difficile* infection.** PCoA of θ_{YC} distances animated from days -1 through 9 of the experiment.
776 Source was the variable that explained the most observed variation across fecal communities
777 (PERMANOVA source $R^2 = 0.35$, $P = 0.0001$, Table S11) followed by interactions between cage
778 effects and day of the experiment. Transparency of the symbol corresponds to the day of the
779 experiment, each symbol represents a sample from an individual mouse at a specific timepoint.
780 Circles represent mice from experiment 1 and triangles represent mice from expeirment 2.

781 **Tables S1-S15. Excel workbook of Tables S1-S15.**

782 **Table S1. Alpha diversity metrics Kruskal-Wallis statistical results.**

783 **Table S2. Alpha diversity metrics pairwise Wilcoxon statistical results.**

784 **Table S3. PERMANOVA results for mice at baseline (day -1), post-clindamycin (day 0), and**
785 **post-infection (day 1).**

786 **Table S4. PERMANOVA results for each source of mice at baseline (day -1).**

787 **Table S5. OTUs with relative abumdances that significantly vary between sources at**
788 **baseline (day -1).**

789 **Table S6. *C. difficile* CFU statistical results.**

790 **Table S7. Mouse weight change statistical results.**

791 **Table S8. OTUs with relative abundances that significantly vary between sources**
792 **post-clindamycin (day 0).**

793 **Table S9. OTUs with relative abundances that significantly changed after clindamycin**
794 **treatment.**

795 **Table S10. OTUs with relative abundances that significantly vary between sources 1-day**
796 **post-infection.**

797 **Table S11. PERMANOVA results for mice across all timepoints.**

798 **Table S12. Statistical results of L2-regularized logistic regression model performances**
799 **compared to random chance.**

800 **Table S13. Pairwise comparisons of L2-regularized logistic regression model performances.**

801 **Table S14. Top 20 most important OTUs for each of the 3 L2-regularized logistic regression**
802 **models based on OTU relative abundance data.**

803 **Table S15. OTUs with relative abundances that significantly varied between sources of mice**
804 **on at least 1 day of the experiment by Kruskal-Wallis test.**