

The initial gut microbiota and response to antibiotic perturbation influence *Clostridioides difficile* clearance 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 V4 region of the 16S rRNA gene from fecal samples collected from these mice
69 after they acclimated to the University of Michigan animal housing environment. We first examined
70 the 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 ($P_{\text{FDR}} = 0.03$ and $P_{\text{FDR}} = 0.052$, respectively; Fig. 1A-B and Tables
73 S1-2). Next, we compared the community structure of mice (Fig. 1C). The source of mice and the
74 interactions between the source and cage effects explained most of the observed variation between
75 fecal communities (PERMANOVA combined $R^2 = 0.90$, $P < 0.001$; Fig. 1C and Table S3). Mice
76 that are co-housed tend to have similar gut microbiotas due to coprophagy (41). Since mice within
77 the same source were housed together, it was not surprising that the cage effect also contributed
78 to the observed community variation. There were some differences between the 2 experiments
79 we conducted, as the experiment and cage effects significantly explained the observed community
80 variation for the Schloss and Young lab mouse colonies (Fig. S1A-B and Table S4). However,
81 most of the vendors also clustered by experiment (Fig. S1C-D, F), suggesting there was some
82 community variation between the 2 experiments within each source, particularly for Schloss, Young,
83 and Envigo mice (Fig. S1G-H). After finding differences at the community level, we next identified
84 the bacteria that varied between sources of mice. There were 268 OTUs with relative abundances
85 that were significantly different between the sources at baseline (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 at baseline (all
88 $P > 0.05$). By using mice from six sources we were able to increase the variation in the starting
89 communities 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, thus colonization is transient when treated with clindamycin alone (21, 43). All mice
95 were treated with 10 mg/kg clindamycin via intraperitoneal injection and one day later challenged
96 with 10^3 *C. difficile* 630 spores (Fig. 2A). The day after infection, *C. difficile* was detectable in all
97 mice at a similar level (median CFU range: 2.2e+07-1.3e+08; $P_{FDR} = 0.15$), indicating clindamycin
98 rendered all mice susceptible regardless of source (Fig. 2B). However, between 3 and 7 days
99 post-infection, we observed variation in *C. difficile* levels across sources of mice (all $P_{FDR} \leq 0.019$;
100 Fig. 2B and Table S6). This suggested the source of mice was associated with *C. difficile* clearance.
101 While the colonization dynamics were similar between the two experiments, the Schloss mice took
102 longer to clear *C. difficile* in the first experiment compared to the second and the Envigo mice
103 took longer to clear *C. difficile* in the second experiment compared to the first (Fig. S2A-B). The
104 change in the mice's weight significantly varied across sources of mice with the most weight loss
105 occurring two days post-infection (Fig. 2C and Table S7). There was also one Jackson and one
106 Envigo mouse that died between 1- and 3-days post-infection during the second experiment. Mice
107 obtained from Jackson, Taconic, and Envigo tended to lose more weight, have higher *C. difficile*
108 CFU levels and take longer to clear the infection compared to the other sources of mice (although
109 there was variation between experiments with Schloss and Envigo mice). This was particularly
110 evident 7 days post-infection (Fig. 2B-C, Fig. S2C-D), when 57% of the mice were still colonized
111 with *C. difficile* (Fig. S2E). By 9 days post-infection the majority of the mice from all sources had
112 cleared *C. difficile* with the exception of 1 Taconic mouse from the first experiment and 2 Envigo
113 mice from the second experiment (Fig. 2B). Thus, clindamycin rendered all mice susceptible to
114 *C. difficile* 630 colonization, regardless of source, but there was significant variation in disease
115 phenotype across the sources of mice.

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

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

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

148 to explain most of the observed community variation (combined $R^2 = 0.88$; $P < 0.001$; Fig. 4C
149 and Table S3). One day after *C. difficile* challenge, there were 44 OTUs with significantly different
150 relative abundances across sources (Fig. 4D and Table S10).

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

163 **Baseline, post-clindamycin, and post-infection community data can predict mice that will**
164 **clear *C. difficile* by 7 days post-infection.** After identifying taxa that varied between sources,
165 changed after clindamycin treatment, or both, we determined which taxa were influencing the
166 variation in *C. difficile* colonization at day 7 (Fig. 2B, Fig. S2C). We trained three L2-regularized
167 logistic regression models with either input bacterial community data from the 6 sources of mice
168 at the baseline (day = -1), post-clindamycin (day = 0), or post-infection (day = 1) timepoints of the
169 experiment to predict *C. difficile* colonization status on day 7 (Fig. S3A-B). All models were better
170 at predicting *C. difficile* colonization status on day 7 than random chance (all $P < 0.001$, Table
171 S12). The model based on the post-clindamycin (AUROC = 0.78) community OTU data performed
172 significantly better than the baseline (AUROC = 0.72) or the post-infection (AUROC = 0.67) models
173 ($P_{FDR} < 0.001$ for pairwise comparisons; Fig. S3C and Table S13). Thus, we were able to use
174 bacterial relative abundance data from the time of *C. difficile* challenge to differentiate mice that had
175 cleared *C. difficile* before day 7 from the mice still colonized with *C. difficile* at that timepoint. This
176 result suggests that the bacterial community's response to clindamycin treatment had the greatest

177 influence on subsequent *C. difficile* colonization dynamics.

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

193 To determine whether the OTUs driving the classification models also varied between sources,
194 were altered by clindamycin treatment, or both, we identified the OTUs from each model that varied
195 between sources (Fig. 1D, 3D, 4D and Tables S5, S8, S10) or were impacted by clindamycin
196 treatment (Fig. 3E and Table S9; Fig. S4). Comparing the features important to the 3 models
197 identified 14 OTUs associated with source, 21 OTUs associated with clindamycin treatment, and
198 6 OTUs associated with both (Fig. 5B). Together, these results suggest that the initial bacterial
199 communities and their responses to clindamycin influenced the clearance of *C. difficile*.

200 Several OTUs that overlapped with our previous analyses appeared across at least 2 models
201 (*Bacteroides*, *Enterococcus*, *Enterobacteriaceae*, *Porphyromonadaceae*), so we examined how
202 the relative abundances of these OTUs varied over the course of the experiment (Fig. 6). Across
203 the 9 days post-infection, there was at least 1 timepoint when the relative abundances of these
204 OTUs significantly varied between sources (Table S15). Interestingly, there were no OTUs that

205 emerged as consistently enriched or depleted in mice that were colonized past 7 days post-infection,
206 suggesting that multiple bacteria influence *C. difficile* colonization dynamics.

207 **Discussion**

208 Applying our CDI model to 6 different sources of mice, allowed us to identify bacterial taxa that were
209 unique to different sources as well as taxa that were universally impacted by clindamycin. We trained
210 L2-regularized logistic regression models with baseline (day -1), post-clindamycin treatment (day 0),
211 and 1-day post-infection fecal community data that could predict whether mice cleared *C. difficile*
212 by 7 days post-infection better than random chance. We identified *Bacteroides*, *Enterococcus*,
213 *Enterobacteriaceae*, *Porphyromonadaceae* (Fig. 6) as candidate bacteria within these communities
214 that influenced variation in *C. difficile* colonization dynamics since these bacteria were all important
215 in the logistic regression models and varied by source, were impacted by clindamycin treatment,
216 or both. Overall, our results demonstrated clindamycin was sufficient to render mice from multiple
217 sources susceptible to CDI and only a subset of the inter-individual microbiota variation across
218 mice from different sources was needed to predict which mice could clear *C. difficile*.

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

227 Differences in CDI mouse model studies have been attributed to intestinal microbiota variation
228 across sources. For example, researchers using the same clindamycin treatment and C57BL/6
229 mice had different *C. difficile* outcomes, one having sustained colonization (32), while the other had
230 transient colonization (18), despite both using *C. difficile* VPI 10643. Baseline differences in the
231 microbiota composition have been hypothesized to partially explain the differences in colonization

232 outcomes and overall susceptibility to *C. difficile* after treatment with the same antibiotic (13, 31).
233 When we treated mice from 6 different sources with clindamycin and challenged them with *C.*
234 *difficile* 630, we found microbiota variation across sources impacted colonization outcomes, but not
235 susceptibility. A previous study with *C. difficile* identified an endogenous protective *C. difficile* strain
236 LEM1 that bloomed after antibiotic treatment in mice from Jackson or Charles River Laboratories,
237 but not Taconic that protected mice against the more toxigenic *C. difficile* VPI10463 (26). Given
238 that we obtained mice from the same vendors, we checked all mice for endogenous *C. difficile*
239 by plating stool samples that were collected after clindamycin treatment. However, we did not
240 identify any endogenous *C. difficile* strains prior to challenge, suggesting there were no endogenous
241 protective strains in the mice we received and other bacteria mediated the variation in *C. difficile*
242 colonization across sources. The *C. difficile* strain used could also be contributing to the variation in
243 *C. difficile* outcomes seen across different research groups. For example, a group found differential
244 colonization outcomes after clindamycin treatment, with *C. difficile* 630 and M68 infections eventually
245 becoming undetectable while strain BI-7 remained detectable up to 70 days post-treatment (44).
246 One study limitation is that we only used female mice. Sex has been shown to influence microbiota
247 variation in mice (45), so we used female mice to reduce this confounding variable and also
248 match the sex used in previous CDI studies that administered clindamycin to mice (32, 33, 44,
249 46). The bacterial perturbations induced by clindamycin treatment have been well characterized
250 and our findings agree with previous CDI mouse model work demonstrating *Enterococcus* and
251 *Enterobacteriaceae* were associated with *C. difficile* susceptibility and *Porphyromonadaceae*,
252 *Lachnospiraceae*, *Ruminococcaceae*, and *Turicibacter* were associated with resistance (19, 21,
253 32, 33, 43, 44, 46, 47). While we have demonstrated that susceptibility is uniform across sources
254 of mice after clindamycin treatment, there could be different outcomes for either susceptibility or
255 clearance in the case of other antibiotic treatments.

256 We found the time needed to naturally clear *C. difficile* varied across sources of mice implying that
257 at least in the context of the same perturbation, microbiota differences influence infection outcome.
258 More importantly, we were able to explain the variation observed across sources with a subset of
259 OTUs that were also important for predicting *C. difficile* colonization status 7 days post-infection.
260 Since all but 3 mice eventually cleared *C. difficile* 630 by 9 days post-infection and the model built

261 with the post-clindamycin (day 0) OTU relative abundance data had the best performance, our
262 results suggest clindamycin treatment had a larger role in determining *C. difficile* susceptibility and
263 clearance than the source of the mice.

264 Using mice from multiple sources successfully increased the inter-animal variation. One alternative
265 approach that has been used in some CDI studies is to associate mice with human microbiotas
266 (48–53). However, a major caveat to this method is the substantial loss of human microbiota
267 community members upon transfer to mice (54, 55). Additionally, with the exception of 2 recent
268 studies (48, 49), most of these studies associated mice with just 1 type of human microbiota either
269 from a single donor or a single pool from multiple donors (50–53). This approach does not aid in
270 the goal of modeling the interpersonal variation seen in humans to understand how the microbiota
271 influences susceptibility to CDIs and adverse outcomes. Importantly, our study using mice from
272 6 different sources increased the variation between groups of mice compared to using 1 source
273 alone, to better reflect the inter-individual microbiota variation observed in humans.

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

284 Other microbiota and host factors that were outside the scope of our current study may also
285 contribute to the differences in *C. difficile* colonization dynamics between sources of mice.
286 The microbiota is composed of viruses, fungi, and parasites in addition to bacteria, and these
287 non-bacterial members can also vary across sources of mice (56, 57). While our study focused
288 solely on the bacterial portion, viruses and fungi have also begun to be implicated in the context

289 of CDIs or FMT treatments for recurrent CDIs (35, 58–61). Beyond community composition, the
290 metabolic function of the microbiota also has a CDI signature (20, 47, 62, 63) and can vary across
291 mice from different sources (64). For example, microbial metabolites, particularly secondary
292 bile acids and butyrate production, have been implicated as important contributors to *C. difficile*
293 resistance (33, 44). Interestingly, butyrate has previously been shown to vary across mouse
294 vendors and mediated resistance to *Citrobacter rodentium* infection, a model of enterohemorrhagic
295 and enteropathogenic *Escherichia coli* infections (23). Evidence for immunological toning
296 differences in IgA and Th17 cells across mice from different vendors have also been documented
297 (65, 66) and could influence the host response to CDI (67, 68), particularly relevant for *C. difficile*
298 strains that induce more severe disease than *C. difficile* 630. The outcome after *C. difficile* exposure
299 depends on a multitude of factors, including genetics, age, diet, and immunity; all of which also
300 influence the microbiota.

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

309 **Acknowledgements**

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

321 **Materials and Methods**

322 **(i) Animals.** All experiments were approved by the University of Michigan Animal Care and Use
323 Committee (IACUC) under protocol number PRO00006983. Female C57BL/7 mice were obtained
324 from 6 different sources: The Jackson Laboratory, Charles River Laboratories, Taconic Biosciences,
325 Envigo, and two colonies at the University of Michigan (the Schloss lab colony and the Young
326 lab colony). The Young lab colony was originally established with mice purchased from Jackson
327 in 2002, and the Schloss lab colony was established in 2010 with mice donated from the Young
328 lab. The 4 groups of mice purchased from vendors were allowed to acclimate to the University of
329 Michigan mouse facility for 13 days prior to starting the experiment. At least 4 female mice (age
330 5-10 weeks) were obtained per source and mice from the same source were primarily housed at a
331 density of 2 mice per cage. The experiment was repeated once, approximately 3 months after the
332 start of the first experiment.

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

336 **(iii) *C. difficile* infection model.** Mice were challenged with 10^3 spores of *C. difficile* strain 630
337 via oral gavage post-infection 1 day after clindamycin treatment as described previously (21). Mice
338 weights and stool samples were taken daily through 9 days post-infection (Fig. 1A). Collected
339 stool was split for *C. difficile* quantification and 16S rRNA sequencing analysis. For *C. difficile*
340 quantification, stool samples were transferred to the anaerobic chamber, serially diluted in PBS,
341 plated on taurocholate-cycloserine-cefoxitin-fructose agar (TCCFA) plates, and counted after 24
342 hours of incubation at 37°C under anaerobic conditions. A sample from the day 0 timepoint
343 (post-clindamycin and prior to *C. difficile* challenge) was also plated on TCCFA to ensure mice
344 were not already colonized with *C. difficile* prior to infection. There were 3 deaths recorded over the
345 course of the experiment, 1 Taconic mouse died prior to *C. difficile* challenge and 1 Jackson and 1
346 Envigo mouse died between 1- and 3-days post-infection. Mice were categorized as cleared when
347 no *C. difficile* was detected in the first serial dilution (limit of detection: 100 CFU). Stool samples for
348 16S rRNA sequencing were snap frozen in liquid nitrogen and stored at -80°C until DNA extraction.

349 **(iv) 16S rRNA sequencing.** DNA was extracted from -80 °C stored stool samples using the DNeasy
350 Powersoil HTP 96 kit (Qiagen) and an EpMotion 5075 automated pipetting system (Eppendorf).
351 The V4 region was amplified for 16S rRNA with the AccuPrime Pfx DNA polymerase (Thermo
352 Fisher Scientific) using custom barcoded primers, as previously described (69). The ZymoBIOMICS
353 microbial community DNA standards was used as a mock community control (70) and water was
354 used as a negative control per 96-well extraction plate. The PCR amplicons were cleaned up
355 and normalized with the SequalPrep normalization plate kit (Thermo Fisher Scientific). Amplicons
356 were pooled and quantified with the KAPA library quantification kit (KAPA biosystems), prior to
357 sequencing using the MiSeq system (Illumina).

358 **(v) 16S rRNA gene sequence analysis.** mothur (v. 1.43) was used to process all sequences
359 (71) with a previously published protocol (69). Reads were combined and aligned with the SILVA
360 reference database (72). Chimeras were removed with the VSEARCH algorithm and taxonomic
361 assignment was completed with a modified version (v16) of the Ribosomal Database Project
362 reference database (v11.5) (73) with an 80% confidence cutoff. Operational taxonomic units (OTUs)
363 were assigned with a 97% similarity threshold using the opticlus algorithm (74). Based on the mock
364 communities, our overall sequencing error rate was 0.0112% and all water controls had less than
365 1000 sequences (range: 18-875). To account for uneven sequencing across samples, samples
366 were rarefied to 5,437 sequences 1,000 times for alpha and beta diversity analyses, and a single
367 time to generate relative abundances for model training. PCoAs were generated based on the
368 Yue and Clayton measure of dissimilarity (θ_{YC}) distances (75). Permutational multivariate analysis
369 of variance (PERMANOVA) was performed on mothur-generated θ_{YC} distance matrices with the
370 adonis function in the vegan package (76) in R (77).

371 **(vi) Classification model training and evaluation.** Models were generated based on mice that
372 were categorized as either cleared or colonized 7 days post-infection and had sequencing data
373 from the baseline (day -1), post-clindamycin (day 0), and post-infection (day 1) timepoints of the
374 experiment. Input bacterial community relative abundance data at the OTU level from the baseline,
375 post-clindamycin, and 1-day post-infection timepoints was used to generate 3 classification models
376 that predicted *C. difficile* colonization status 7 days post-infection. The L2-regularized logistic
377 regression models were trained and tested using the caret package (78) in R as previously

described (79) with the exception that we used 60% training and 40% testing data splits for testing of the held out test data to measure model performance and repeated 2.5-fold cross-validation of the training data to select the best cost hyperparameter. The modified training to testing ratio was selected to accommodate the small number of samples in the dataset. Code was modified from https://github.com/SchlossLab/ML_pipeline_microbiome to update the classification outcomes and change the data split ratios. The modified repository to regenerate our modeling analysis is available at https://github.com/tomkosev/ML_pipeline_microbiome.

(vii) Statistical analysis. All statistical tests were performed in R (v 4.0.2) (77). The Kruskal-Wallis test was used to analyze differences in *C. difficile* CFU, mouse weight change, and alpha diversity across sources with a Benjamini-Hochberg correction for testing multiple timepoints, followed by pairwise Wilcoxon comparisons with Benjamini-Hochberg correction. For taxonomic analysis and generation of logistic regression model input data, *C. difficile* (OTU 20) was removed. Bacterial relative abundances that varied across sources at the OTU level were identified with the Kruskal-Wallis test with Benjamini-Hochberg correction for testing all identified OTUs, followed by pairwise Wilcoxon comparisons with Benjamini-Hochberg correction. The Wilcoxon rank sum test was used to test for OTUs that differed between experiments within the Schloss, Young, and Envigo sources with Benjamini-Hochberg correction for testing all identified OTUs. OTUs impacted by clindamycin treatment were identified using the paired Wilcoxon signed rank test with matched pairs of mice samples from day -1 and day 0. To determine whether classification models had better performance (test AUROCs) than random chance (0.5), we used the one-sample Wilcoxon signed rank test. To examine whether there was an overall difference in predictive performance across the 3 classification models we used the Kruskal-Wallis test followed by pairwise Wilcoxon comparisons with Benjamini-Hochberg correction for multiple hypothesis testing. The tidyverse package (v 1.3.0) was used to wrangle and graph data (80).

(viii) Code availability. Code for all data analysis and generating this manuscript is available at https://github.com/SchlossLab/Tomkovich_Vendor_mSphere_2020.

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

406 **References**

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

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

- 456 nutrient niches during infection across distinct murine gut microbiomes. *mSystems* 2.
- 457 21. Jenior ML, Leslie JL, Young VB, Schloss PD. 2018. *Clostridium difficile* alters the structure and
458 metabolism of distinct cecal microbiomes during initial infection to promote sustained colonization.
459 *mSphere* 3.
- 460 22. Velazquez EM, Nguyen H, Heasley KT, Saechao CH, Gil LM, Rogers AWL, Miller BM, Rolston
461 MR, Lopez CA, Litvak Y, Liou MJ, Faber F, Bronner DN, Tiffany CR, Byndloss MX, Byndloss
462 AJ, Bäumler AJ. 2019. Endogenous Enterobacteriaceae underlie variation in susceptibility to
463 *Salmonella* infection. *Nature Microbiology* 4:1057–1064.
- 464 23. Osbelt L, Thiemann S, Smit N, Lesker TR, Schröter M, Gálvez EJC, Schmidt-Hohagen K, Pils
465 MC, Mühlen S, Dersch P, Hiller K, Schlüter D, Neumann-Schaal M, Strowig T. 2020. Variations in
466 microbiota composition of laboratory mice influence *Citrobacter rodentium* infection via variable
467 short-chain fatty acid production. *PLOS Pathogens* 16:e1008448.
- 468 24. Stough JMA, Dearth SP, Denny JE, LeCleir GR, Schmidt NW, Campagna SR, Wilhelm SW.
469 2016. Functional characteristics of the gut microbiome in C57BL/6 mice differentially susceptible to
470 *Plasmodium yoelii*. *Frontiers in Microbiology* 7.
- 471 25. Alegre M-L. 2019. Mouse microbiomes: Overlooked culprits of experimental variability. *Genome
472 Biology* 20.
- 473 26. Etienne-Mesmin L, Chassaing B, Adekunle O, Mattei LM, Bushman FD, Gewirtz AT. 2017.
474 Toxin-positive *Clostridium difficile* latently infect mouse colonies and protect against highly
475 pathogenic *C. difficile*. *Gut* 67:860–871.
- 476 27. Lai NY, Musser MA, Pinho-Ribeiro FA, Baral P, Jacobson A, Ma P, Potts DE, Chen Z, Paik D,
477 Soualhi S, Yan Y, Misra A, Goldstein K, Lagomarsino VN, Nordstrom A, Sivanathan KN, Wallrapp A,
478 Kuchroo VK, Nowarski R, Starnbach MN, Shi H, Surana NK, An D, Wu C, Huh JR, Rao M, Chiu IM.
479 2020. Gut-innervating nociceptor neurons regulate peyer's patch microfold cells and SFB levels to
480 mediate *Salmonella* host defense. *Cell* 180:33–49.e22.
- 481 28. Thiemann S, Smit N, Roy U, Lesker TR, Gálvez EJC, Helmecke J, Basic M, Bleich A, Goodman

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

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

- 532 46. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, Mastroeni P, Scott P,
533 Raisen C, Mottram L, Fairweather NF, Wren BW, Parkhill J, Dougan G. 2009. Antibiotic treatment of
534 *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and
535 severe disease in immunocompromised hosts. *Infection and Immunity* 77:3661–3669.
- 536 47. Jump RLP, Polinkovsky A, Hurless K, Sitzlar B, Eckart K, Tomas M, Deshpande A, Nerandzic
537 MM, Donskey CJ. 2014. Metabolomics analysis identifies intestinal microbiota-derived biomarkers
538 of colonization resistance in clindamycin-treated mice. *PLoS ONE* 9:e101267.
- 539 48. Nagao-Kitamoto H, Leslie JL, Kitamoto S, Jin C, Thomsson KA, Gilliland MG, Kuffa P, Goto Y,
540 Jenq RR, Ishii C, Hirayama A, Seekatz AM, Martens EC, Eaton KA, Kao JY, Fukuda S, Higgins PDR,
541 Karlsson NG, Young VB, Kamada N. 2020. Interleukin-22-mediated host glycosylation prevents
542 *Clostridioides difficile* infection by modulating the metabolic activity of the gut microbiota. *Nature
543 Medicine* 26:608–617.
- 544 49. Battaglioli EJ, Hale VL, Chen J, Jeraldo P, Ruiz-Mojica C, Schmidt BA, Rekdal VM, Till LM, Huq
545 L, Smits SA, Moor WJ, Jones-Hall Y, Smyrk T, Khanna S, Pardi DS, Grover M, Patel R, Chia N,
546 Nelson H, Sonnenburg JL, Farrugia G, Kashyap PC. 2018. *Clostridioides difficile* uses amino acids
547 associated with gut microbial dysbiosis in a subset of patients with diarrhea. *Science Translational
548 Medicine* 10:eaam7019.
- 549 50. Robinson CD, Auchtung JM, Collins J, Britton RA. 2014. Epidemic *Clostridium difficile* strains
550 demonstrate increased competitive fitness compared to nonepidemic isolates. *Infection and
551 Immunity* 82:2815–2825.
- 552 51. Collins J, Auchtung JM, Schaefer L, Eaton KA, Britton RA. 2015. Humanized microbiota mice
553 as a model of recurrent *Clostridium difficile* disease. *Microbiome* 3.
- 554 52. Collins J, Robinson C, Danhof H, Knetsch CW, Leeuwen HC van, Lawley TD, Auchtung JM,
555 Britton RA. 2018. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature
556* 553:291–294.
- 557 53. Hryckowian AJ, Treuren WV, Smits SA, Davis NM, Gardner JO, Bouley DM, Sonnenburg JL.

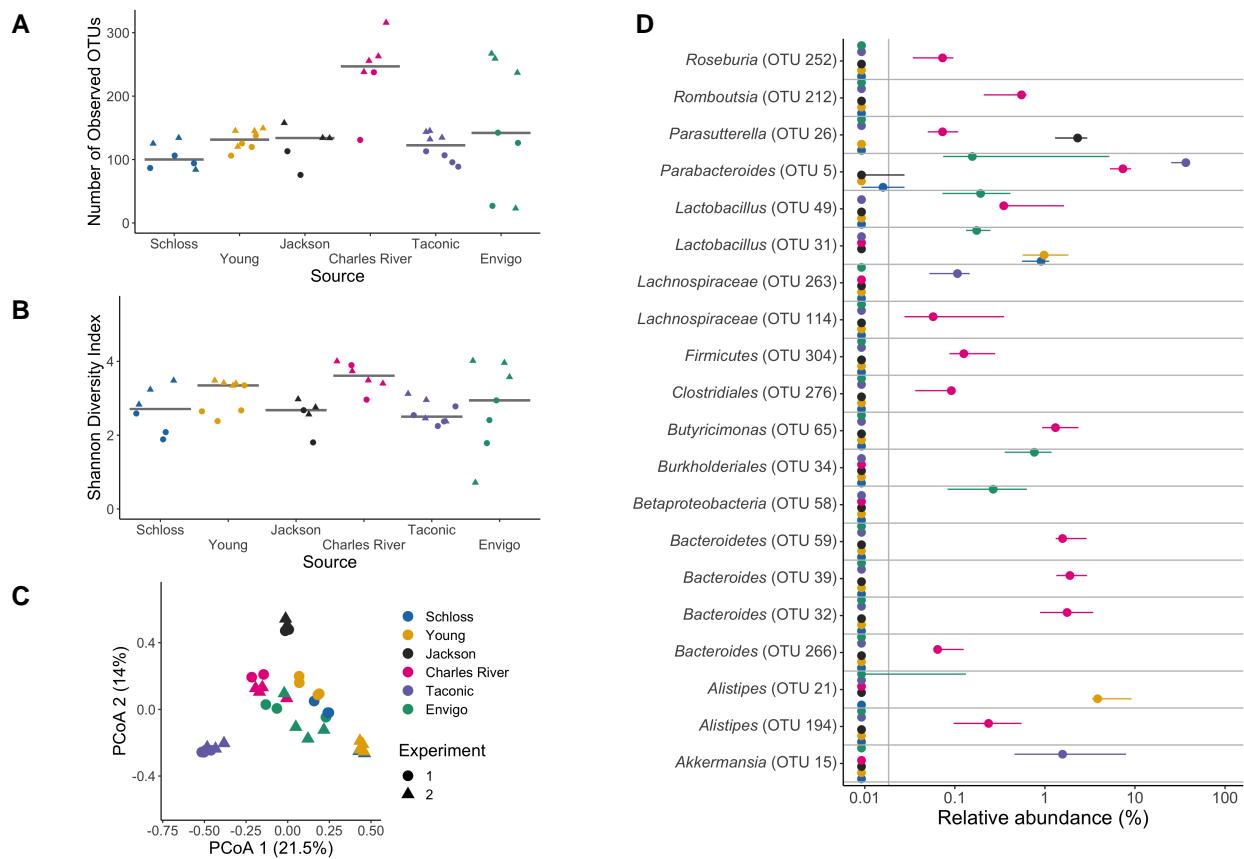
- 558 2018. Microbiota-accessible carbohydrates suppress *Clostridium difficile* infection in a murine
559 model. *Nature Microbiology* 3:662–669.
- 560 54. Fouladi F, Glenny EM, Bulik-Sullivan EC, Tsilimigras MCB, Sioda M, Thomas SA, Wang Y, Djukic
561 Z, Tang Q, Tarantino LM, Bulik CM, Fodor AA, Carroll IM. 2020. Sequence variant analysis reveals
562 poor correlations in microbial taxonomic abundance between humans and mice after gnotobiotic
563 transfer. *The ISME Journal* <https://doi.org/10.1038/s41396-020-0645-z>.
- 564 55. Walter J, Armet AM, Finlay BB, Shanahan F. 2020. Establishing or exaggerating causality for
565 the gut microbiome: Lessons from human microbiota-associated rodents. *Cell* 180:221–232.
- 566 56. Rasmussen TS, Vries L de, Kot W, Hansen LH, Castro-Mejía JL, Vogensen FK, Hansen AK,
567 Nielsen DS. 2019. Mouse vendor influence on the bacterial and viral gut composition exceeds the
568 effect of diet. *Viruses* 11:435.
- 569 57. Mims TS, Abdallah QA, Watts S, White C, Han J, Willis KA, Pierre JF. 2020. Variability in
570 interkingdom gut microbiomes between different commercial vendors shapes fat gain in response
571 to diet. *The FASEB Journal* 34:1–1.
- 572 58. Stewart DB, Wright JR, Fowler M, McLimans CJ, Tokarev V, Amaniera I, Baker O, Wong H-T,
573 Brabec J, Drucker R, Lamendella R. 2019. Integrated meta-omics reveals a fungus-associated
574 bacteriome and distinct functional pathways in *Clostridioides difficile* infection. *mSphere* 4.
- 575 59. Ott SJ, Waetzig GH, Rehman A, Moltzau-Anderson J, Bharti R, Grasis JA, Cassidy L, Tholey A,
576 Fickenscher H, Seegert D, Rosenstiel P, Schreiber S. 2017. Efficacy of sterile fecal filtrate transfer
577 for treating patients with *Clostridium difficile* infection. *Gastroenterology* 152:799–811.e7.
- 578 60. Zuo T, Wong SH, Lam K, Lui R, Cheung K, Tang W, Ching JYL, Chan PKS, Chan MCW, Wu
579 JCY, Chan FKL, Yu J, Sung JJY, Ng SC. 2017. Bacteriophage transfer during faecal microbiota
580 transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut*
581 *gutjnl*–2017–313952.
- 582 61. Zuo T, Wong SH, Cheung CP, Lam K, Lui R, Cheung K, Zhang F, Tang W, Ching JYL, Wu JCY,
583 Chan PKS, Sung JJY, Yu J, Chan FKL, Ng SC. 2018. Gut fungal dysbiosis correlates with reduced

- 584 efficacy of fecal microbiota transplantation in *Clostridium difficile* infection. Nature Communications
585 9.
- 586 62. Robinson JI, Weir WH, Crowley JR, Hink T, Reske KA, Kwon JH, Burnham C-AD, Dubberke
587 ER, Mucha PJ, Henderson JP. 2019. Metabolomic networks connect host-microbiome processes to
588 human *Clostridioides difficile* infections. Journal of Clinical Investigation 129:3792–3806.
- 589 63. Fletcher JR, Erwin S, Lanzas C, Theriot CM. 2018. Shifts in the gut metabolome and *Clostridium*
590 *difficile* transcriptome throughout colonization and infection in a mouse model. mSphere 3.
- 591 64. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, Li X, Long H, Zhang J, Zhang D, Liu C, Fang
592 Z, Chou J, Glanville J, Hao Q, Kotowska D, Colding C, Licht TR, Wu D, Yu J, Sung JJY, Liang Q, Li
593 J, Jia H, Lan Z, Tremaroli V, Dworzynski P, Nielsen HB, Bäckhed F, Doré J, Chatelier EL, Ehrlich
594 SD, Lin JC, Arumugam M, Wang J, Madsen L, Kristiansen K. 2015. A catalog of the mouse gut
595 metagenome. Nature Biotechnology 33:1103–1108.
- 596 65. Fransen F, Zagato E, Mazzini E, Fosso B, Manzari C, Aidy SE, Chiavelli A, D'Erchia AM,
597 Sethi MK, Pabst O, Marzano M, Moretti S, Romani L, Penna G, Pesole G, Rescigno M. 2015.
598 BALB/c and C57BL/6 mice differ in polyreactive IgA abundance, which impacts the generation of
599 antigen-specific IgA and microbiota diversity. Immunity 43:527–540.
- 600 66. Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee
601 CA, Lynch SV, Tanoue T, Imaoka A, Itoh K, Takeda K, Umesaki Y, Honda K, Littman DR. 2009.
602 Induction of intestinal th17 cells by segmented filamentous bacteria. Cell 139:485–498.
- 603 67. Azrad M, Hamo Z, Tkawkho L, Peretz A. 2018. Elevated serum immunoglobulin a levels in
604 patients with *Clostridium difficile* infection are associated with mortality. Pathogens and Disease 76.
- 605 68. Saleh MM, Frisbee AL, Leslie JL, Buonomo EL, Cowardin CA, Ma JZ, Simpson ME, Scully KW,
606 Abhyankar MM, Petri WA. 2019. Colitis-induced th17 cells increase the risk for severe subsequent
607 *Clostridium difficile* infection. Cell Host & Microbe 25:756–765.e5.
- 608 69. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a
609 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the

- 610 MiSeq illumina sequencing platform. *Applied and Environmental Microbiology* 79:5112–5120.
- 611 70. Sze MA, Schloss PD. 2019. The impact of DNA polymerase and number of rounds of
612 amplification in PCR on 16S rRNA gene sequence data. *mSphere* 4.
- 613 71. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
614 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF.
615 2009. Introducing mothur: Open-source, platform-independent, community-supported software
616 for describing and comparing microbial communities. *Applied and Environmental Microbiology*
617 75:7537–7541.
- 618 72. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012.
619 The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools.
620 *Nucleic Acids Research* 41:D590–D596.
- 621 73. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR,
622 Tiedje JM. 2013. Ribosomal database project: Data and tools for high throughput rRNA analysis.
623 *Nucleic Acids Research* 42:D633–D642.
- 624 74. Westcott SL, Schloss PD. 2017. OptiClust, an improved method for assigning amplicon-based
625 sequence data to operational taxonomic units. *mSphere* 2.
- 626 75. Yue JC, Clayton MK. 2005. A similarity measure based on species proportions. *Communications
627 in Statistics - Theory and Methods* 34:2123–2131.
- 628 76. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB,
629 Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2018. Vegan: Community ecology
630 package.
- 631 77. R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for
632 Statistical Computing, Vienna, Austria.
- 633 78. Kuhn M. 2008. Building predictive models in RUsing the caret Package. *Journal of Statistical
634 Software* 28.

- 635 79. Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. 2020. A framework for effective
636 application of machine learning to microbiome-based classification problems. *mBio* 11.
- 637 80. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes
638 A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D,
639 Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the
640 tidyverse. *Journal of Open Source Software* 4:1686.

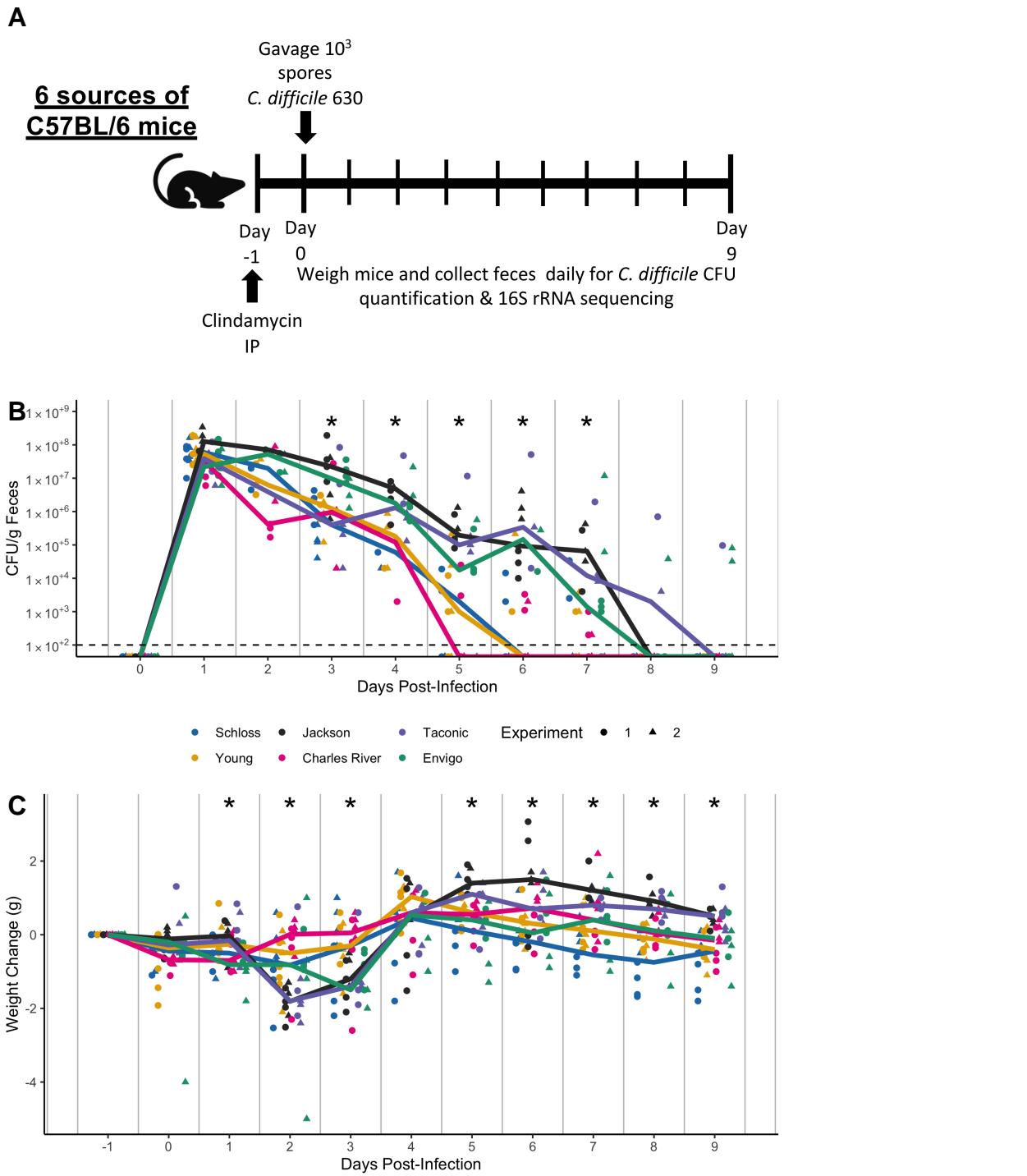
641 **Figures**



642

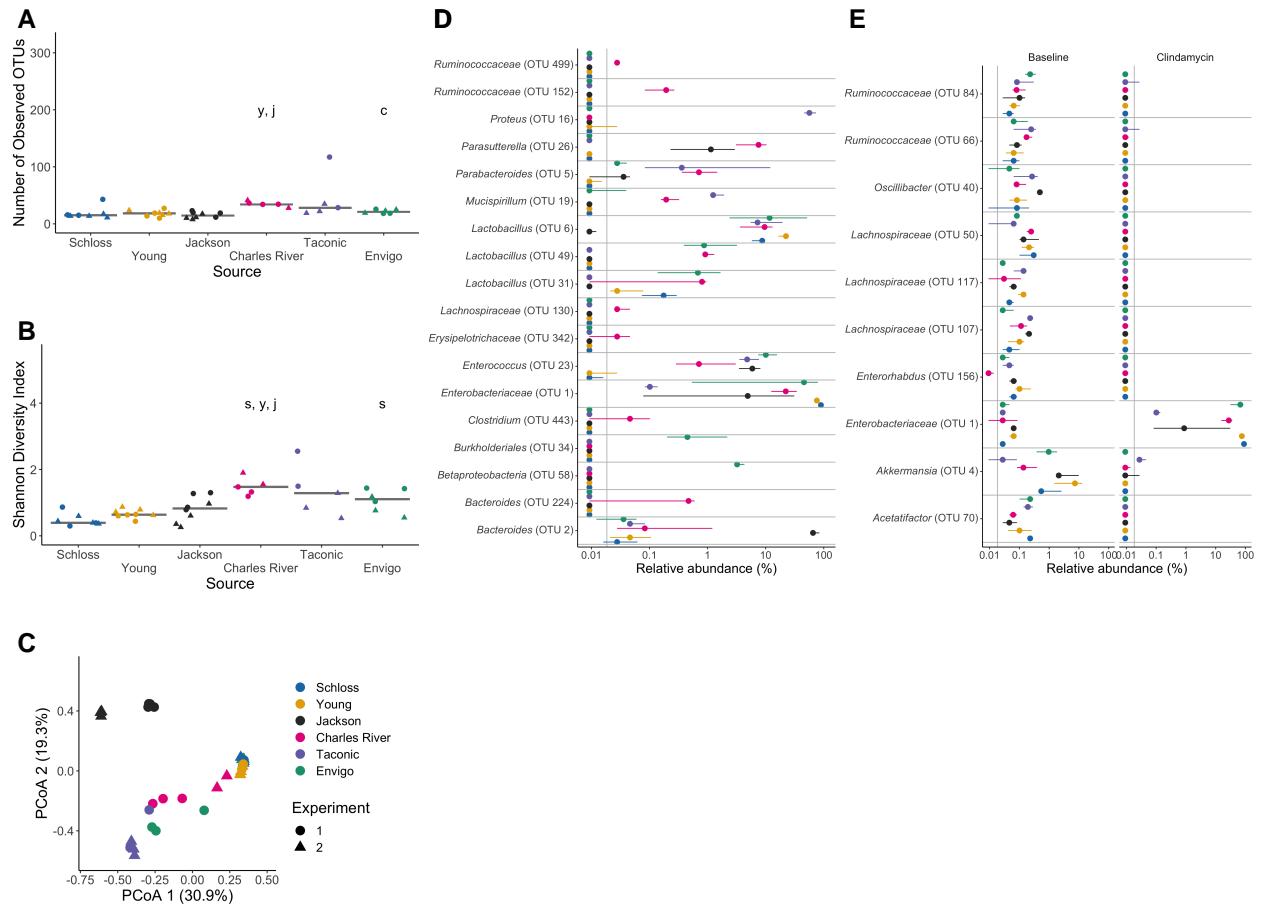
643 **Figure 1. Microbiota variation is high between mice from different sources.** A-B. Number of
 644 observed OTUs (A) and Shannon diversity index values (B) across sources of mice at baseline
 645 (day -1 of the experiment). Differences between sources were analyzed by Kruskal-Wallis test
 646 with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

656 Benjamini-Hochberg correction (Table S5).



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

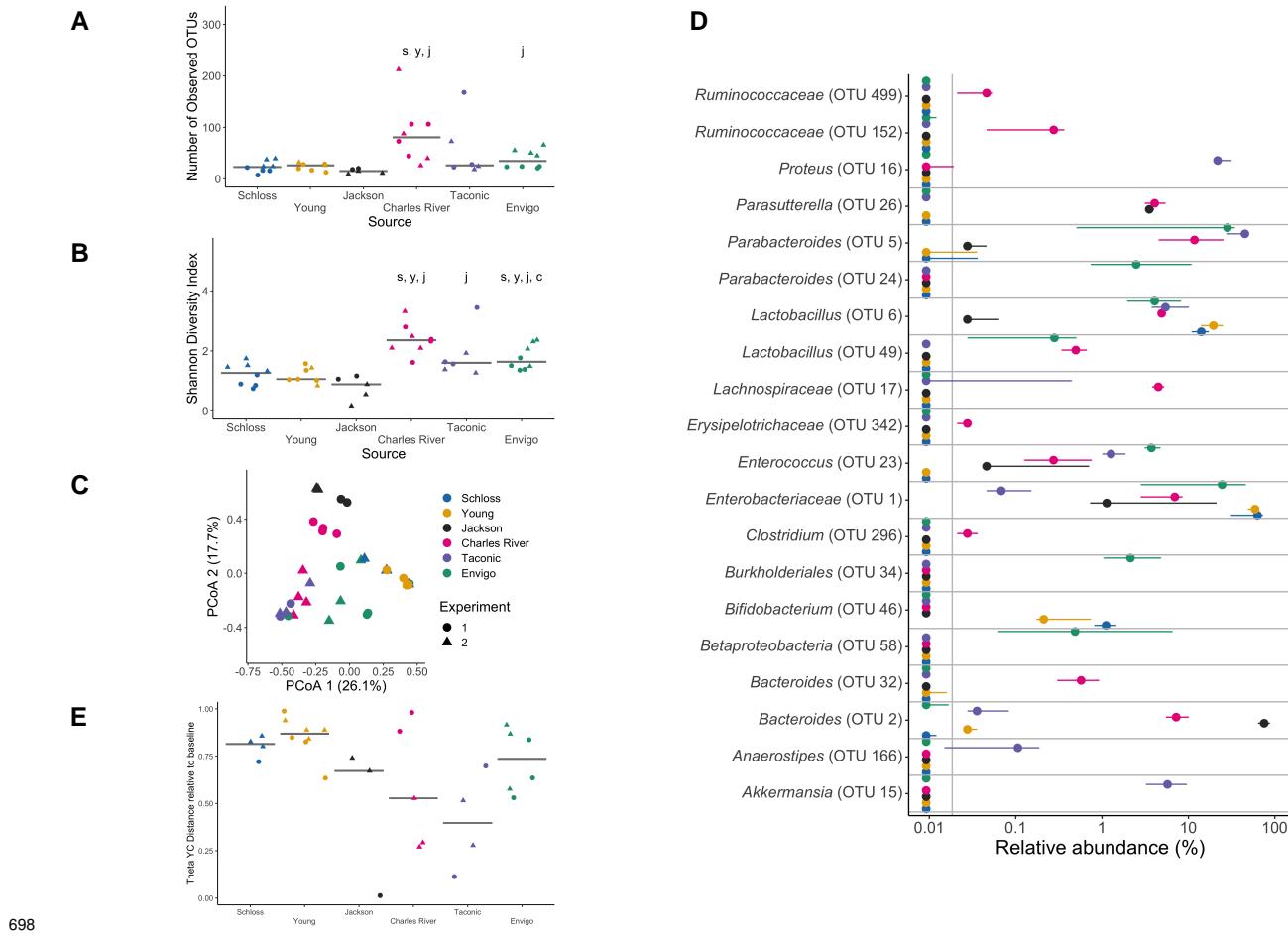
= 8), Taconic Biosciences (N = 8), and Envigo (N = 8). Mice that were ordered from commercial vendors acclimated to the University of Michigan mouse facility for 13 days prior to antibiotic administration. All mice were administered 10 mg/kg clindamycin intraperitoneally (IP) 1 day before challenge with *C. difficile* 630 spores on day 0. Mice were weighed and feces was collected daily through the end of the experiment (9 days post-infection). Note: 3 mice died during course of experiment. 1 Taconic mouse prior to infection and 1 Jackson and 1 Envigo mouse between 1- and 3-days post-infection. B. *C. difficile* CFU/gram stool measured over time (N = 20-49 mice per timepoint) via serial dilutions. The black line represents the limit of detection for the first serial dilution. CFU quantification data was not available for each mouse due to early deaths, stool sampling difficulties, and not plating all of the serial dilutions. C. Mouse weight change measured in grams over time (N = 45-49 mice per timepoint), all mice were normalized to the weight recorded 1 day before infection. For B-C: timepoints where differences between sources of mice were statistically significant by Kruskal-Wallis test with Benjamini-Hochberg correction for testing across multiple days (Table S6 and Table S7) are reflected by the asterisk above each timepoint (*, $P < 0.05$). Lines represent the median for each source and circles represent individual mice from experiment 1 while triangles represent mice from experiment 2.



678

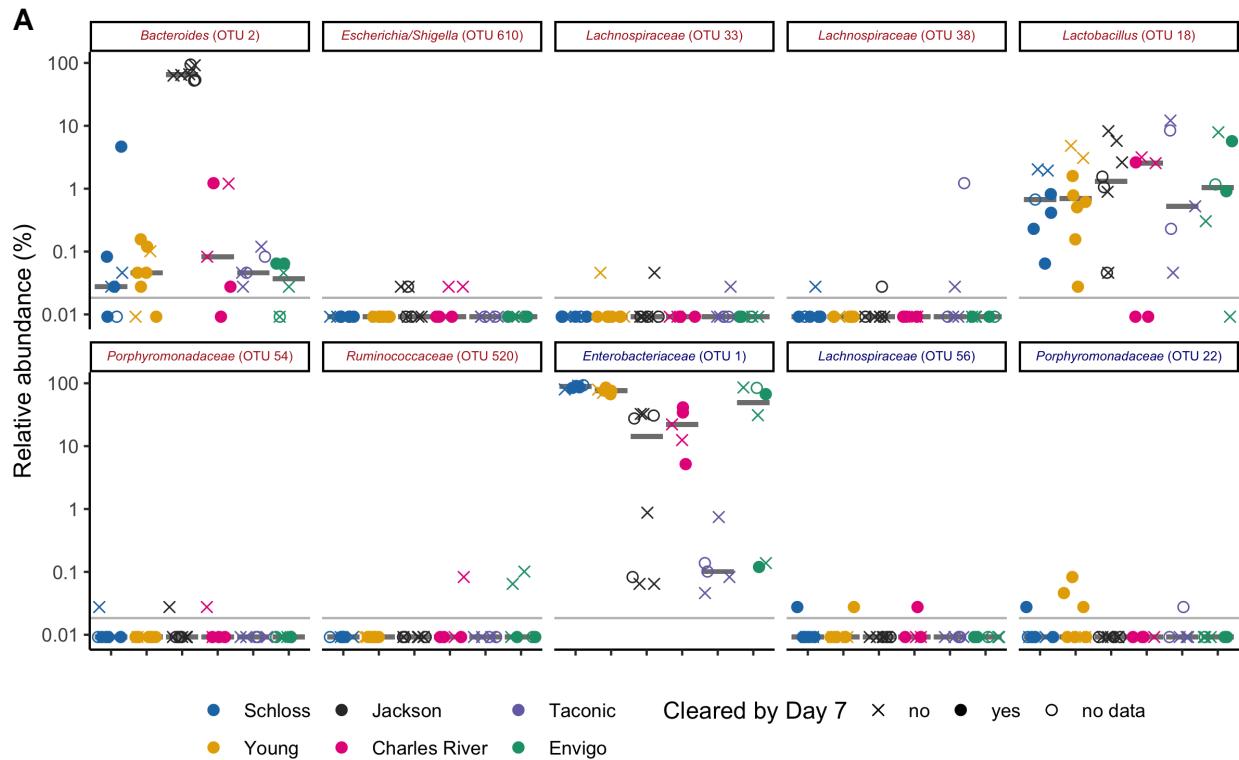
679 **Figure 3. Clindamycin treatment alters bacteria in all sources, but a subset of bacterial**
 680 **differences across sources persists.** A-B. Number of observed OTUs (A) and Shannon diversity
 681 index values (B) across sources of mice after clindamycin treatment (day 0). Differences between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

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



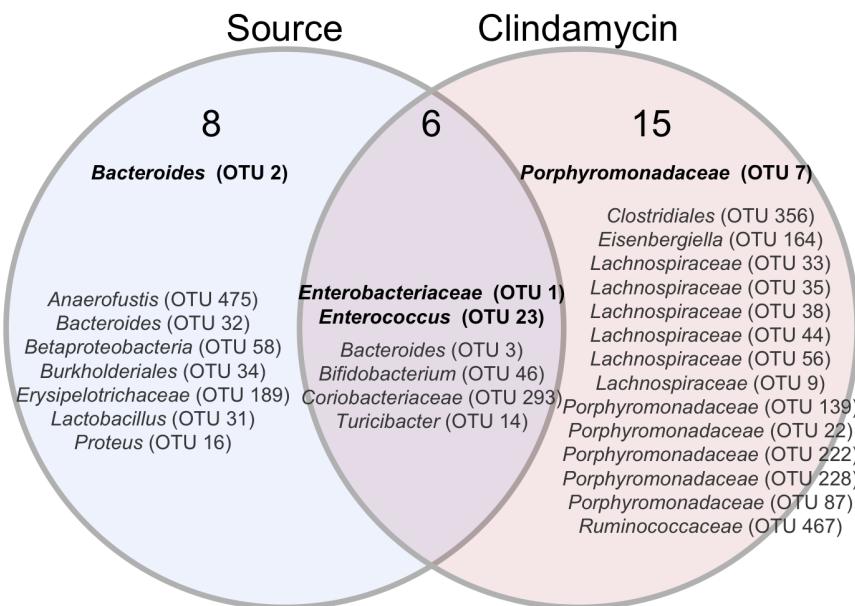
698 **Figure 4. Microbiota variation across sources is maintained after *C. difficile* challenge.** A-B.
 699 Number of observed OTUs (A) and Shannon diversity index values (B) across sources of mice
 700 1-day post-infection. Data were analyzed by Kruskal-Wallis test with Benjamini-Hochberg correction
 701 for testing each day of the experiment and the adjusted *P* value was < 0.05 (Table S1). Significant
 702 *P* values from the pairwise Wilcoxon comparisons between sources with Benjamini-Hochberg
 703 correction are displayed as the first initial of each group compared to the group that they are listed
 704 above (Table S2). PCoA of θ_{YC} distances of 1-day post-infection stool samples. Source and
 705 the interaction between source and cage effects explained most of the variation between fecal
 706 communities (PERMANOVA combined $R^2 = 0.88$, $P < 0.001$; Table S3). For A-C: each symbol
 707 represents the value for a stool sample from an individual mouse, circles represent experiment 1
 708 mice and triangles represent experiment 2 mice. D. The median (point) and interquartile range
 709 (colored lines) of the relative abundances for the top 20 most significant OTUs out of the 44 OTUs
 710 that varied between sources 1-day post-infection. The gray vertical line indicates the limit of

⁷¹² detection. For each timepoint OTUs with differential relative abundances across sources of mice
⁷¹³ were identified by Kruskal-Wallis test with Benjamini-Hochberg correction for testing all identified
⁷¹⁴ OTUs (Table S10). E. θ_{YC} distances of fecal samples collected 7-days post-infection relative to the
⁷¹⁵ baseline (day -1) sample for each mouse. Each symbol represents an individual mouse. Gray lines
⁷¹⁶ represent the median for each source.



B

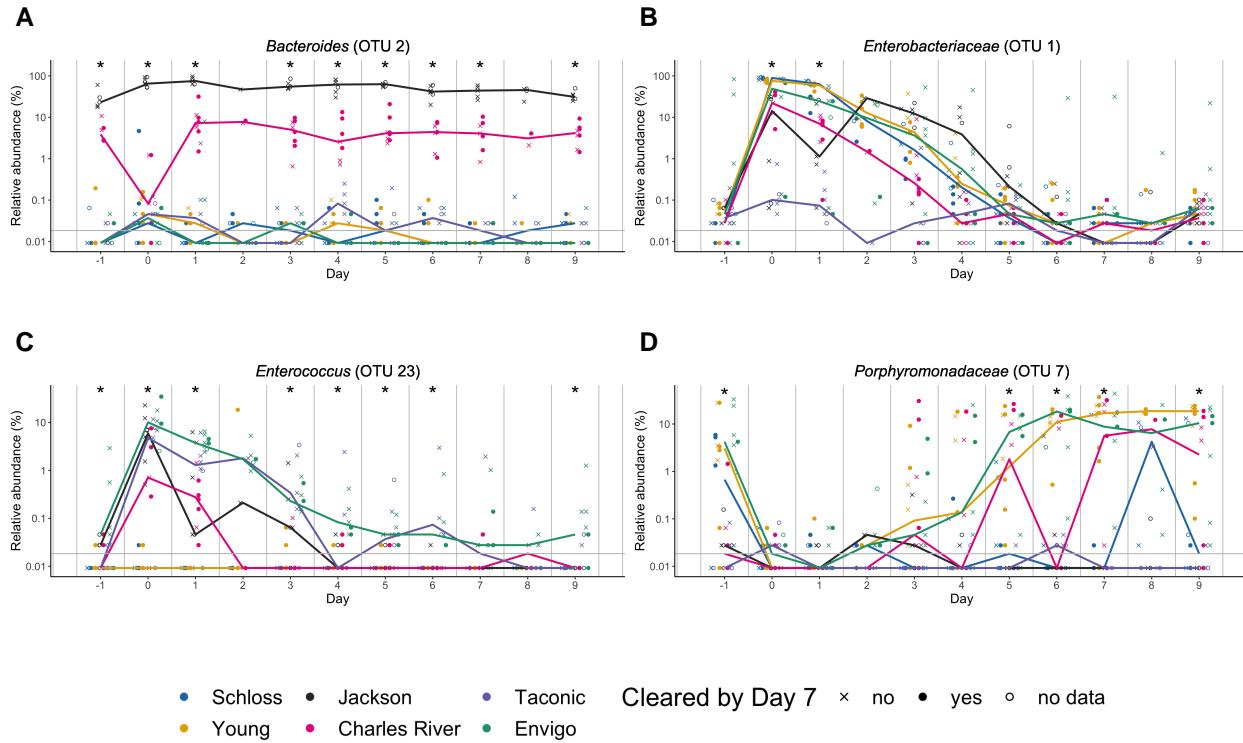
OTU comparisons for day -1, 0, and 1 models



717

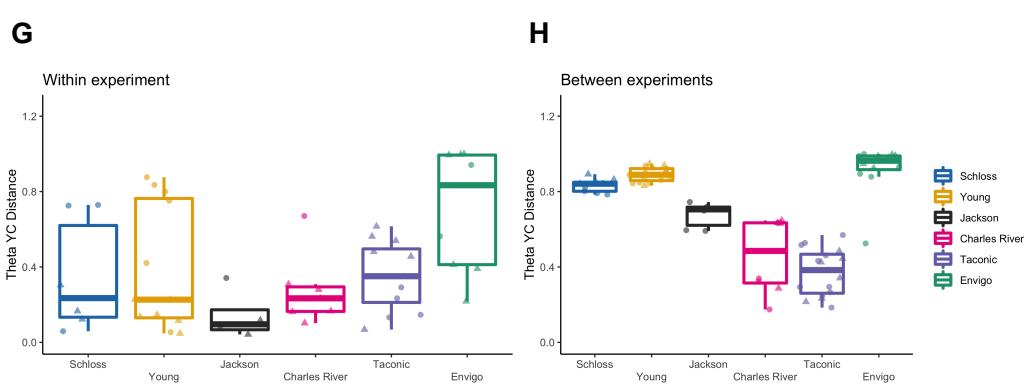
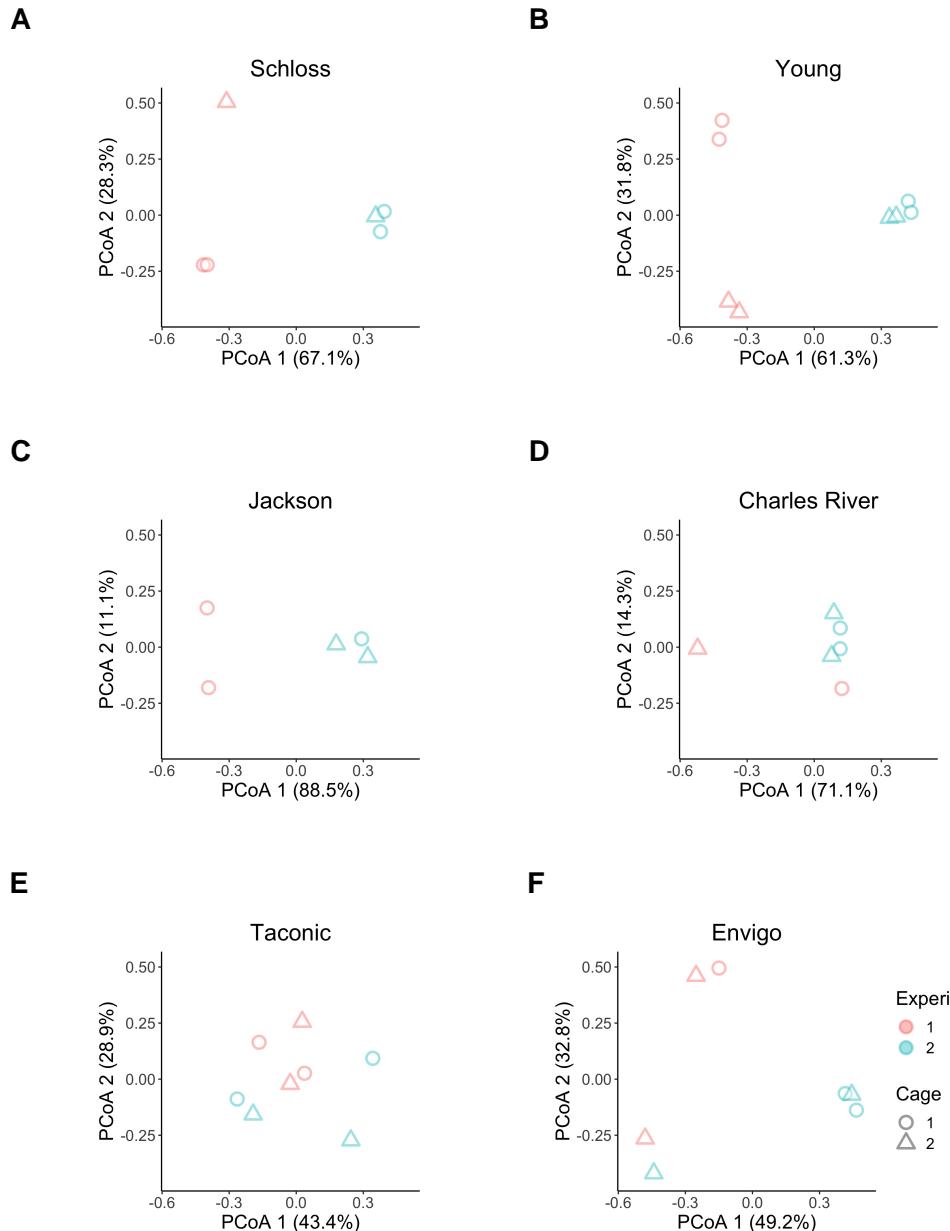
718 **Figure 5. Bacteria that influenced whether mice cleared *C. difficile* by day 7.** A.
 719 Post-clindamycin (day 0) relative abundance data for the 10 OTUs with the highest rankings based

720 on feature weights in the post-clindamycin (day 0) classification model. Red font represents OTUs
721 that correlated with *C. difficile* colonization and blue font represents OTUs that correlated with
722 clearance. Symbols represent the relative abundance data for an individual mouse. Gray bars
723 indicate the median relative abundances for each source. The gray horizontal lines indicate the
724 limit of detection. B. Venn diagram that combines OTUs that were important to the day -1, 0, and
725 1 classification models (Fig. S4, Table S14) and either overlapped with taxa that varied across
726 sources at the same timepoint, were impacted by clindamycin treatment, or both. Bold OTUs were
727 important to more than 1 classification model.

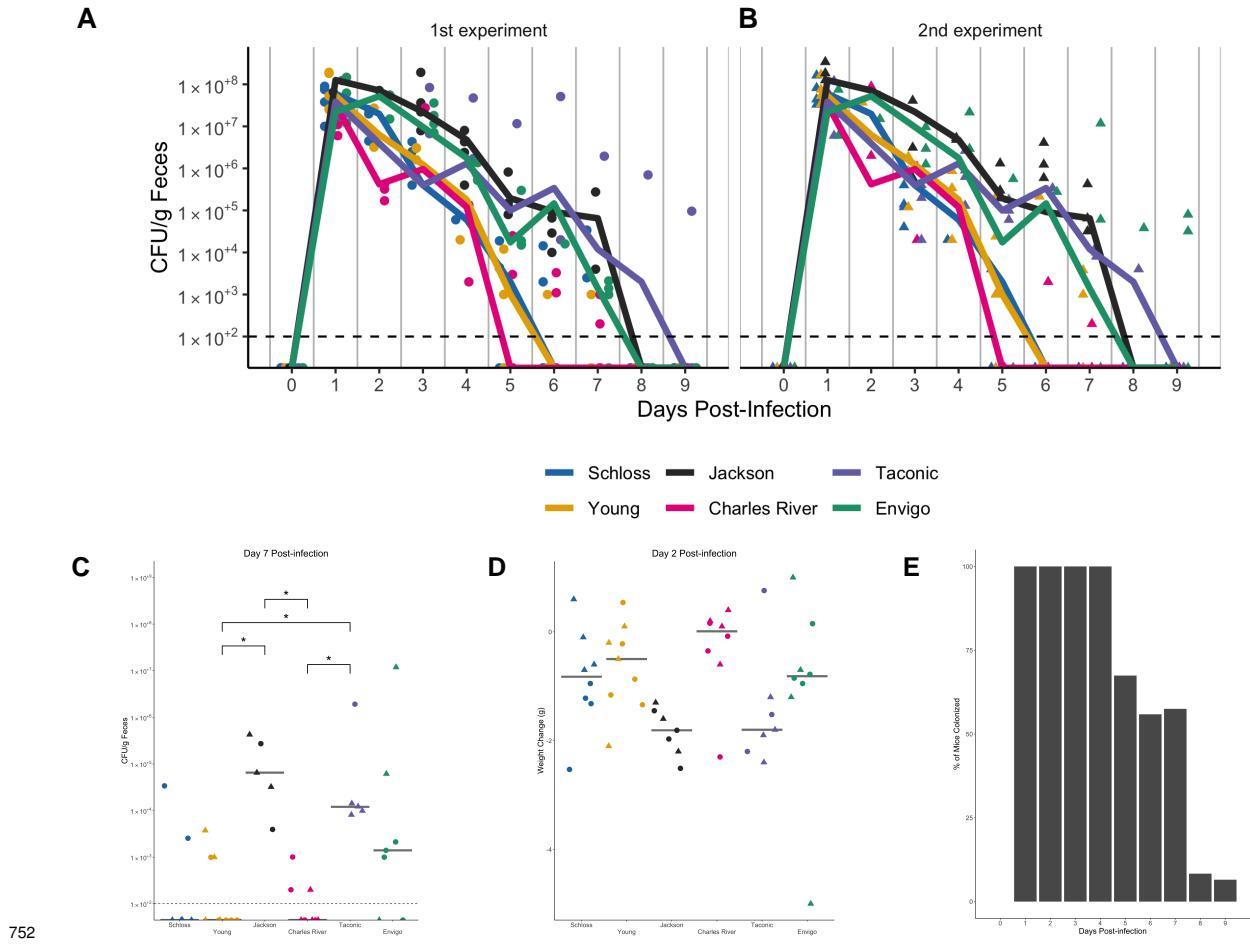


728

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

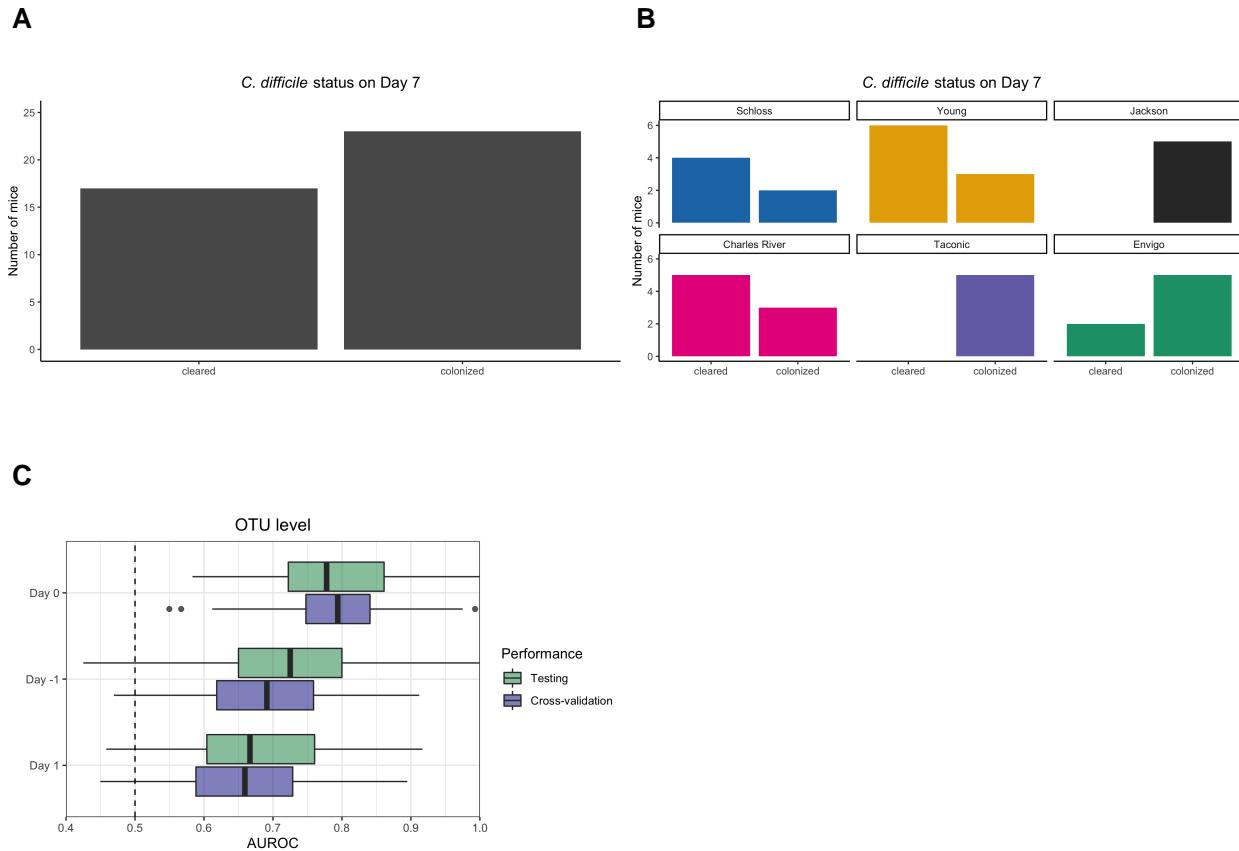


743 **Figure S1. Bacterial communities vary between experiments for some sources.** A-F. PCoA
744 of θ_{YC} distances for the baseline fecal bacterial communities within each source of mice. Each
745 symbol represents a stool sample from an individual mouse with color corresponding to experiment
746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

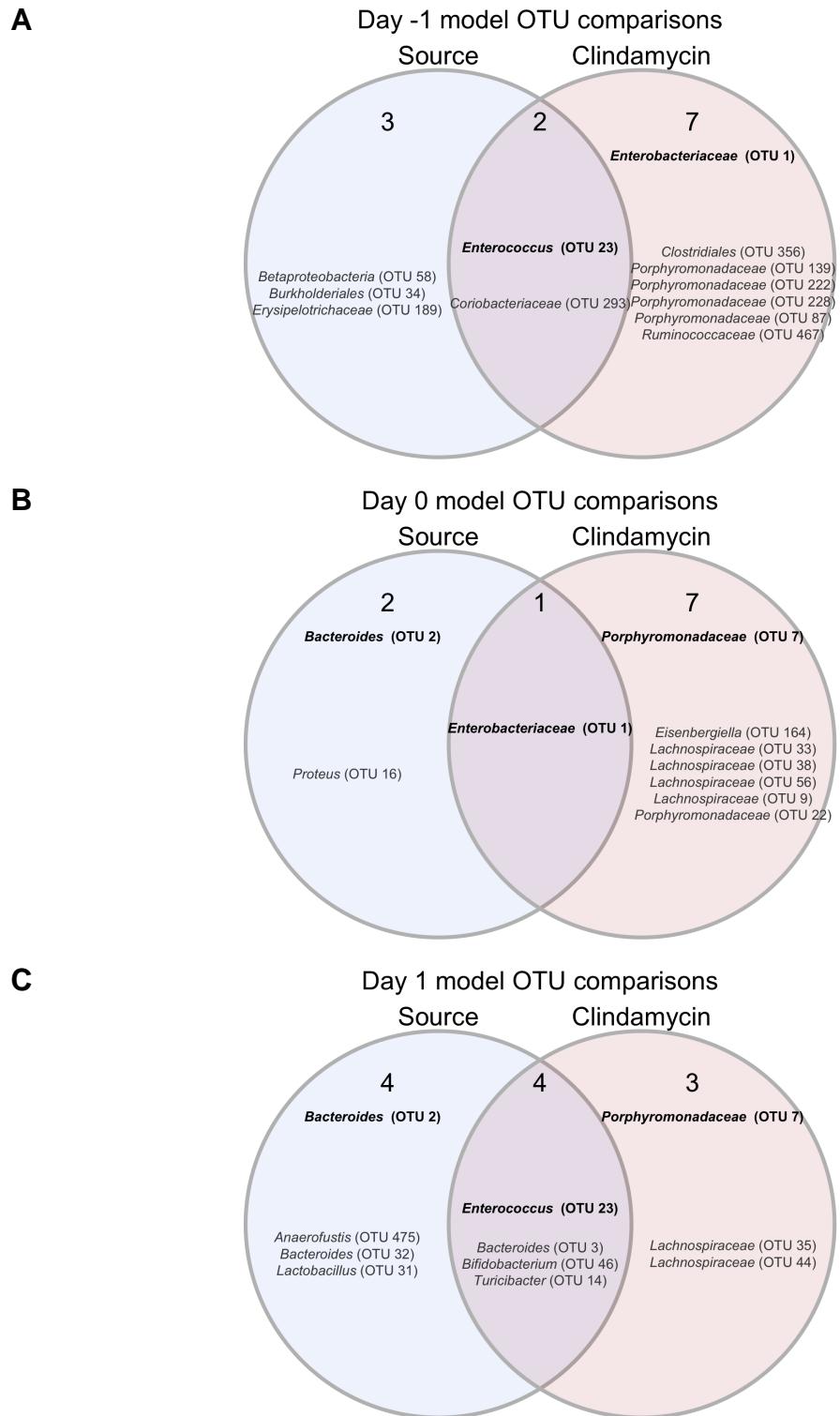


752 **Figure S2.** *C. difficile* CFU variation across sources varies slightly between the 2
 753 experiments. A-B. *C. difficile* CFU/gram of stool quantification over time for experiment 1 (A) and
 754 2 (B). Experiments were conducted approximately 3 months apart. Lines represent the median
 755 CFU for each source, symbols represent individual mice and the black line represents the limit
 756 of detection. C. *C. difficile* CFU/gram stool 7-days post-infection across sources of mice with an
 757 asterisk for pairwise Wilcoxon comparisons with Benjamini-Hochberg correction where $P < 0.05$.
 758 D. Mouse weight change 2-days post-infection across sources of mice, no pairwise Wilcoxon
 759 comparisons were significant after Benjamini-Hochberg correction. For C-D: circles represent
 760 experiment 1 mice, triangles represent experiment 2 mice and gray lines indicate the median
 761 values for each group. E. Percent of mice that were colonized with *C. difficile* over the course of the
 762 experiment. Each day the percent is calculated based on the mice where *C. difficile* CFU was
 763 quantified for that particular day. Total N for each day: day 1 (N = 42), day 2 (N = 20), day 3 (N =
 764 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 =
 765 10).

766 46).



768 **Figure S3. Bacterial community composition before, after clindamycin perturbation, and**
 769 **post-infection can predict *C. difficile* colonization status 7 days post-infection.** A. Bar graph
 770 visualizations of overall 7-days post-infection *C. difficile* colonization status that were used as
 771 classification outcomes to build L2-regularized logistic regression models. Mice were classified as
 772 colonized or cleared (not detectable at the limit of detection of 100 CFU) based on CFU g/stool data
 773 from 7 days post-infection. B. *C. difficile* CFU status on Day 7 within each mouse source. N = 8-9
 774 mice per group. C. L2-regularized logistic regression classification model area under the receiving
 775 operator characteristic curve (AUROCs) to predict *C. difficile* CFU on day 7 post-infection (Fig. 2B,
 776 Fig. S2C) based on the OTU community relative abundances at baseline (day -1), post-clindamycin
 777 (day 0), and 1-day post-infection. All models performed better than random chance (AUROC =
 778 0.5, all $P < 0.001$, Table S12) and the model built with post-clindamycin bacterial OTU relative
 779 abundances had the best performance ($P_{FDR} < 0.001$ for all pairwise comparisons, Table S13).
 780 See Table S14 for list of the 20 OTUs that were ranked as most important to each model.



781

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

⁷⁸⁴ Venn diagrams of OTUs from the top 20 OTUs from the baseline (A), post-clindamycin treatment (B),
⁷⁸⁵ and 1-day post-infection (C) classification models (Table S14) that overlapped with OTUs that varied
⁷⁸⁶ across sources at the corresponding timepoint (Tables S5, 8, 10), were impacted by clindamycin
⁷⁸⁷ treatment (Table S9), or both. Bold OTUs were important to more than 1 classification model.

788 **Supplementary Tables and Movie**

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

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

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

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

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

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

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

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

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

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

808 **Table S9. OTUs with relative abundances that significantly changed after clindamycin**
809 **treatment.**

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

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

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

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

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

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