

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 university colony or vendor. We treated mice from 6 different sources
5 (2 University of Michigan colonies and 4 vendors) with a single clindamycin dose, followed by
6 a *C. difficile* challenge 1 day later and then measured *C. difficile* colonization levels through 9
7 days post-infection. The microbiota were profiled via 16S rRNA gene sequencing to examine the
8 variation across sources and alterations due to clindamycin treatment and *C. difficile* challenge.
9 While all sources of mice were colonized 1-day post-infection, variation emerged from days 3-7
10 post-infection with animals from some sources colonized with *C. difficile* for longer and at higher
11 levels. We identified bacteria that varied in relative abundance across sources and throughout the
12 experiment. Some bacteria were consistently impacted by clindamycin treatment in all sources of
13 mice including *Lachnospiraceae*, *Ruminococcaceae*, and *Enterobacteriaceae*. To identify bacteria
14 that were most important to colonization regardless of the source, we created logistic regression
15 models that successfully classified mice based on whether they cleared *C. difficile* by 7 days
16 post-infection using baseline, post-clindamycin, and 1-day post-infection community composition
17 data. With these models, we identified 4 bacteria that varied across sources (*Bacteroides*), were
18 altered by clindamycin (*Porphyromonadaceae*), or both (*Enterobacteriaceae* and *Enterococcus*).
19 Microbiota variation across sources better emulates human inter-individual variation and can help
20 identify 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 an
24 infection, or has an infection with adverse outcomes. *C. difficile* infection (CDI) mouse models 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). The
42 number of CDI mouse model studies has grown substantially since Chen et al. published their
43 C57BL/6 model in 2008, which disrupted the gut microbiota with antibiotics to enable *C. difficile*
44 colonization and symptoms such as diarrhea and weight loss (14). CDI mouse models have been
45 used to examine translationally relevant questions regarding *C. difficile*, including the role of the
46 microbiota and efficacy of potential therapeutics for treating CDIs (15). However, variation in the
47 microbiota between mice from the same breeding colony is much less than the inter-individual
48 variation observed between humans (16, 17). Studying CDIs in mice with a homogenous microbiota
49 is likely to overstate the importance of individual mechanisms. Using mice that have a more
50 heterogenous microbiota would allow researchers to identify and validate more generalizable
51 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 variation in CDI outcomes despite using similar murine models (13, 18, 21, 31–33). Here
58 we examined how variations 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 sources:
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 alpha
70 diversity across the 6 sources of mice. There was a significant difference in the richness (i.e. number
71 of observed operational taxonomic units (OTUs)), but not Shannon diversity index ($P_{FDR} = 0.03$
72 and $P_{FDR} = 0.052$, respectively) across the sources of mice (Fig. 1A-B and Tables S1-2). Next,
73 we compared the community structure of mice (Fig. 1C). Interactions between the source and
74 cage effects explained most of the observed variation between fecal communities (PERMANOVA
75 combined $R^2 = 0.90$, $P < 0.001$, Fig. 1C and Table S3). Mice that are co-housed tend to have
76 similar gut microbiotas due to coprophagy (41). Since mice within the same source were housed
77 together, it is not surprising that the cage effect also contributed to the observed microbiota variation.
78 There were some differences between the 2 experiments we conducted, as the experiment and
79 cage effects significantly explained the observed community variation for the Schloss and Young
80 lab mouse colonies (Fig. S2A-B and Table S4). However, most of the vendors also clustered
81 by experiment (Fig. S2C-D, F), suggesting there was some community variation between the 2
82 experiments within each source, particularly for Schloss, Young, and Envigo mice (Fig. S2G-H).
83 After finding differences at the community level, we next identified the bacteria that varied between
84 sources of mice. There were 268 OTUs with relative abundances that were significantly different
85 between the sources (Fig. 1D and Table S5). Though we saw differences between experiments
86 at the community level, there were no OTUs that were significantly different between experiments
87 within Schloss, Young, and Envigo mice (Data not shown). By using mice from six sources we were
88 able to increase the number of microbiota communities to treat with clindamycin and challenge with
89 *C. difficile*.

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).
93 We have previously demonstrated mice are rendered susceptible to *C. difficile*, but cleared the
94 pathogen within 9 days when treated with clindamycin alone (21, 43). All mice were treated with 10
95 mg/kg clindamycin via intraperitoneal injection and one day later challenged with 10^3 *C. difficile*
96 630 spores (Fig. 2A). The day after infection, *C. difficile* was detectable in all mice at a similar
97 level (median CFU range: 2.2e+07-1.3e+08; $P_{FDR} = 0.15$), indicating clindamycin rendered all
98 mice susceptible regardless of source (Fig. 2B). However, between 3 and 7 days post-infection,
99 we 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 mouse source 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 1st experiment and the Envigo mice took longer to clear in the 2nd experiment (Fig.
103 S2A-B). The change in the mice's weight significantly varied across sources of mice with the most
104 weight lost two days post-infection (Fig. 2C and Table S7). There was also one Jackson and one
105 Envigo mouse that died between 1- and 3-days post-infection during the second experiment. Mice
106 obtained from Jackson, Taconic, and Envigo tended to lose more weight, have higher *C. difficile*
107 CFU levels and take longer to clear the infection compared to the other sources of mice (although
108 there was variation between experiments with Schloss and Envigo mice). This was particularly
109 evident 7 days post-infection (Fig. 2B-C, Fig. S2C-D), when 57% of the mice were still colonized
110 with *C. difficile* (Fig. S2E). By 9 days post-infection the majority of the mice from all sources had
111 cleared *C. difficile* (Fig. 2C) with the exception of 1 Taconic mouse from the first experiment and
112 2 Envigo mice from the second experiment. Thus, clindamycin rendered all mice susceptible to
113 *C. difficile* 630 colonization, regardless of source, but there was significant variation in disease
114 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 mouse microbiotas that we observed across
117 breeding colonies, we hypothesized that variation in *C. difficile* clearance would be explained by
118 microbiota 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 ($P_{FDR} < 0.05$) emerged after clindamycin treatment,
121 with Charles River mice having higher richness and Shannon diversity than most of the other sources
122 (Fig 3A-B and Tables S1-2). The clindamycin treatment decreased the variation in community
123 structures between sources of mice. Source and cage effects explained almost all of the observed
124 variation between communities (combined $R^2 = 0.99$, $P < 0.001$; Fig. 3C and Table S3). However,
125 there were only 18 OTUs (Fig. 3D and Table S8) with relative abundances that significantly varied
126 between sources. Next we identified the bacteria that shifted after clindamycin treatment, regardless
127 of source by analyzing paired fecal samples from mice that were collected at baseline and after
128 clindamycin treatment. We identified 153 OTUs that were altered after clindamycin treatment
129 in most mice (Fig. 3E and Table S9). When we compared the list of significant clindamycin
130 impacted bacteria with the bacteria that varied between sources post-clindamycin, we found 4 OTUs
131 (*Enterobacteriaceae* (OTU 1), *Lachnospiraceae* (OTU 130), *Lactobacillus* (OTU 6), *Enterococcus*
132 (OTU 23)) overlapped (Fig. 3D-E and Tables S8-9). Importantly, some of the OTUs that varied
133 between sources also shifted with clindamycin treatment. For example, *Proteus* increased after
134 clindamycin treatment (Fig. 3D), but only in Taconic mice. *Enterococcus* was primarily found only
135 in mice purchased from commercial vendors and also increased after clindamycin treatment (Fig.
136 3D). These findings demonstrate that clindamycin had a consistent impact on the fecal bacterial
137 communities of mice from all sources and only a subset of the OTUs continued to vary between
138 sources.

139 **Microbiota variation between sources is maintained after *C. difficile* challenge.** One day
140 post-infection, significant differences in diversity metrics remained across sources ($P_{FDR} < 0.05$,
141 Fig 4A-B and Tables S1-2). Although the Charles River mice had more diverse microbiotas and
142 were also able to clear *C. difficile* faster than some of the other sources. Microbiota diversity did not
143 explain the observed variation in *C. difficile* colonization across sources considering the Young and
144 Schloss mice had the lowest diversity after clindamycin treatment and were able to clear *C. difficile*
145 earlier than Jackson, Taconic and Envigo mice. Source and the interactions between source and
146 cage effects continued to explain most of the observed community variation (combined $R^2 = 0.88$;
147 $P < 0.001$; Fig. 4D and Table S3). One day after *C. difficile* challenge, there were 44 OTUs (Fig. 4D

148 and Table S10) with significantly different relative abundances across sources.

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

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

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

190 To determine whether the OTUs driving the classification models also varied between sources, were
191 altered by clindamycin treatment, or both, we examined whether the top 20 ranked OTUs from each
192 model overlapped with OTUs that varied between sources (Fig. 1D, 3D, 4D and Tables S5, S8, S10)
193 or were impacted by clindamycin treatment (Fig. 3E and Table S9). We found a subset of OTUs
194 that were important to the baseline (day -1), post-clindamycin (day 0), and 1 day post-infection
195 models and overlapped with bacteria that varied between sources, were altered by clindamycin
196 treatment, or both (Fig. S4). Combining the overall comparison results for the 3 models identified
197 14 OTUs associated with source, 21 OTUs associated with clindamycin treatment, and 6 OTUs
198 associated with both (Fig. 5B). Several OTUs (*Bacteroides*, *Enterococcus*, *Enterobacteriaceae*,
199 *Porphyromonadaceae*) that overlapped with our previous analyses appeared across at least 2
200 models, so we examined how the relative abundances of these OTUs varied over the course of
201 the experiment (Fig. 6). Throughout the experiment, there was at least 1 timepoint where relative
202 abundances of these OTUs significantly varied between sources (Table S15). Interestingly, there
203 were no OTUs that emerged as consistently enriched or depleted in mice that were colonized
204 past 7 days post-infection, suggesting multiple bacteria influence *C. difficile* colonization dynamics.

205 Together, these results suggest the initial bacterial communities and their responses to clindamycin
206 influence the clearance of *C. difficile*.

207 Discussion

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

220 Other studies have taken similar approaches by using mice from multiple sources to identify bacteria
221 that either promote colonization resistance or increase susceptibility to enteric infections (22, 23,
222 26–30). For example, in the context of *Salmonella* infections, *Enterobacteriaceae* and segmented
223 filamentous bacteria have emerged as protective (22, 27). A previous study with *C. difficile* identified
224 an endogenous protective *C. difficile* strain LEM1 that bloomed after antibiotic treatment in mice
225 from Jackson or Charles River Laboratories, but not Taconic that protected mice against the more
226 toxicogenic *C. difficile* VPI10463 (26). Given that we obtained mice from the same vendors, we
227 checked all mice for endogenous *C. difficile* by plating stool samples that were collected after
228 clindamycin treatment. However, we did not identify any endogenous *C. difficile* strains prior to
229 challenge, suggesting there were no endogenous protective strains in the mice we received and
230 other bacterial taxa mediated the variation in *C. difficile* colonization across sources.

231 Differences in CDI mouse model studies have been attributed to intestinal microbiota variation

232 across sources. For example, groups using the same clindamycin treatment and C57BL/6 mice
233 had different *C. difficile* outcomes, one having sustained colonization (32), while the other had
234 transient (18), despite both using *C. difficile* VPI 10643. Baseline differences in the microbiota
235 composition have been hypothesized to partially explain the differences in colonization outcomes
236 and overall susceptibility to *C. difficile* after treatment with the same antibiotic (13, 31). The
237 bacterial perturbations induced by clindamycin treatment have been well characterized and
238 our findings agree with previous CDI mouse model work demonstrating *Enterococcus* and
239 *Enterobacteriaceae* were associated with *C. difficile* susceptibility and *Porphyromonadaceae*,
240 *Lachnospiraceae*, *Ruminococcaceae*, and *Turicibacter* were associated with resistance (19, 21, 32,
241 33, 43–46). While we have demonstrated that susceptibility is uniform across sources of mice after
242 clindamycin treatment, there could be different outcomes for either susceptibility or clearance in
243 the case of other antibiotic treatments. The *C. difficile* strain used could also be contributing to
244 the variation in *C. difficile* outcomes seen across different research groups. For example, a group
245 found differential colonization outcomes after clindamycin treatment, with *C. difficile* 630 and M68
246 infections eventually becoming undetectable while strain BI-7 remained detectable up to 70 days
247 post-treatment (45). We found the time needed to naturally clear *C. difficile* varied across sources
248 of mice implying that at least in the context of the same perturbation, microbiota differences
249 seemed to influence infection outcome more than susceptibility. More importantly, we were able to
250 reduce the variation observed across sources to identify a subset of OTUs that were also important
251 for predicting *C. difficile* colonization status 7 days post-infection. Since all but 3 mice eventually
252 cleared *C. difficile* 630 by 9 days post-infection and the model built with the post-clindamycin (day 0)
253 OTU relative abundance data had the best performance, our results suggest clindamycin treatment
254 had a larger role in determining *C. difficile* susceptibility and clearance than the source of the mice.

255 Our approach successfully increased the diversity of murine bacterial communities tested in
256 our clindamycin *C. difficile* model. One alternative approach that has been used in some CDI
257 studies (47–52) is to associate mice with human microbiotas. However, a major caveat to this
258 method is the substantial loss of human microbiota community members upon transfer to mice
259 (53, 54). Additionally with the exception of 2 recent studies (47, 48), most of the CDI mouse model
260 studies to date associated mice with just 1 types of human microbiota either from a single donor

261 or a single pool from multiple donors (49–52), which does not aid in the goal of modeling the
262 interpersonal variation seen in humans to understand how the microbiota influences susceptibility to
263 CDIs and adverse outcomes. Importantly, our study using mice from 6 different sources increased
264 the variation between groups of mice compared to using 1 source alone, to better reflect the
265 inter-individual microbiota variation observed in humans. Encouragingly, decreased *Bifidobacterium*,
266 *Porphyromonas*, *Ruminococcaceae* and *Lachnospiraceae* and increased *Enterobacteriaceae*,
267 *Enterococcus*, *Lactobacillus*, and *Proteus* have all been associated with human CDIs (7) and
268 were well represented in our study, suggesting most of the mouse sources are suitable for gaining
269 insights into the bacteria influencing *C. difficile* colonization and infections in humans. An important
270 exception was *Enterococcus*, which was primarily absent from University of Michigan colonies and
271 *Proteus*, which was only found in Taconic mice. Importantly, the fact that some CDI-associated
272 bacteria were only found in a subset of mice has important implications for future CDI mouse model
273 studies.

274 There are several limitations to our work. The microbiota is composed of viruses, fungi, and
275 parasites in addition to bacteria, and these non-bacterial members can also vary across sources of
276 mice (55, 56). While our study focused solely on the bacterial portion, viruses and fungi have also
277 begun to be implicated in the context of CDIs or FMT treatments for recurrent CDIs (35, 57–60).
278 Beyond community composition, the metabolic function of the microbiota also has a CDI signature
279 (20, 46, 61, 62) and can vary across mice from different sources (63). For example, microbial
280 metabolites, particularly secondary bile acids and butyrate production, have been implicated as
281 important contributors to *C. difficile* resistance (33, 45). Although, we only looked at composition,
282 *Ruminococcaceae* and *Lachnospiraceae* both emerged as important taxa for classifying day 7 *C.*
283 *difficile* colonization status and metagenomes from these bacteria have been shown to contain the
284 bile acid-inducible gene cluster necessary for secondary bile acid formation and ability to produce
285 butyrate (50, 64). Interestingly, butyrate has previously been shown to vary across mouse vendors
286 and mediated resistance to *Citrobacter rodentium* infection, a model of enterohemorrhagic and
287 enteropathogenic *Escherichia coli* infections (23). Evidence for immunological toning differences in
288 IgA and Th17 cells across mice from different vendors have also been documented and (65, 66)
289 could influence the host response to CDI (67, 68). The outcome after *C. difficile* exposure depends

290 on a multitude of factors, including age, diet, and immunity; all of which are also influenced by the
291 microbiota.

292 We have demonstrated that the ways baseline microbiotas from different mouse sources respond
293 to clindamycin treatment influences the length of time mice remained colonized with *C. difficile*
294 630. For those interested in dissecting the contribution of the microbiota to *C. difficile* pathogenesis
295 and treatments, using multiple sources of mice may yield more insights than a single model alone.
296 Furthermore, for studies wanting to examine the interplay between a particular bacteria such as
297 *Enterococcus* and *C. difficile*, these results could serve as a resource for selecting which mice to
298 order to address the question. Using mice from multiple sources helps model the interpersonal
299 microbiota variation among humans to aid our understanding of how the gut microbiota contributes
300 to CDIs.

301 **Acknowledgements**

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

314 **Materials and Methods**

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

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

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

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

351 **(v) 16S rRNA gene sequence analysis.** mothur (v. 1.43) was used to process all sequences
352 (71) with a previously published protocol (69). Reads were combined and aligned with the SILVA
353 reference database (72). Chimeras were removed with the VSEARCH algorithm and taxonomic
354 assignment was completed with a modified version (v16) of the Ribosomal Database Project
355 reference database (v11.5) (73) with an 80% cutoff. Operational taxonomic units (OTUs) were
356 assigned with a 97% similarity threshold using the opticlus algorithm (74). To account for uneven
357 sequencing across samples, samples were rarefied to 5,437 sequences 1,000 times for alpha and
358 beta diversity analyses. PCoAs were generated based on θ_{YC} distances. Permutational multivariate
359 analysis of variance (PERMANOVA) was performed on mothur-generated θ_{YC} distance matrices
360 with the adonis function in the vegan package (75) in R (76).

361 **(vi) Classification model training and evaluation.** Models were generated based on mice that
362 were categorized as either cleared or colonized 7 days post-infection and had sequencing data
363 from the baseline (day -1), post-clindamycin (day 0), and post-infection (day 1) timepoints of
364 the experiment. Input bacterial community relative abundance data at the OTU level from the
365 baseline, post-clindamycin, and post-infection timepoints was used to generate 6 classification
366 models that predicted *C. difficile* colonization status 7 days post-infection. The L2-regularized
367 logistic regression models were trained and tested using the caret package (77) in R as previously
368 described (78) with the exception that we used 60% training and 40% testing data splits for the
369 cross-validation of the training data to select the best cost hyperparameter and the testing of
370 the held out test data to measure model performance. The modified training to testing ratio was

371 selected to accommodate the small number of samples in the dataset. Code was modified from
372 https://github.com/SchlossLab/ML_pipeline_microbiome to update the classification outcomes
373 and change the data split ratios. The modified repository to regenerate this analysis is available at
374 https://github.com/tomkosev/ML_pipeline_microbiome.

375 **(vii) Statistical analysis.** All statistical tests were performed in R (v 4.0.2) (76). The Kruskal-Wallis
376 test was used to analyze differences in *C. difficile* CFU, mouse weight change, and alpha
377 diversity across sources with a Benjamini-Hochberg correction for testing multiple timepoints,
378 followed by pairwise Wilcoxon comparisons with Benjamini-Hochberg correction. For taxonomic
379 analysis and generation of logistic regression model input data, *C. difficile* (OTU 20) was removed.
380 Bacterial relative abundances that varied across sources at the OTU level were identified with the
381 Kruskal-Wallis test with Benjamini-Hochberg correction for testing all identified OTUs, followed
382 by pairwise Wilcoxon comparisons with Benjamini-Hochberg correction. OTUs impacted by
383 clindamycin treatment were identified using the Wilcoxon signed rank test with matched pairs
384 of mice samples for day -1 and day 0. To determine whether classification models had better
385 performance (test AUROCs) than random chance (0.5), we used the one-sample Wilcoxon signed
386 rank test. To examine whether there was an overall difference in predictive performance across the
387 6 classification models we used the Kruskal-Wallis test followed by pairwise Wilcoxon comparisons
388 with Benjamini-Hochberg correction for multiple hypothesis testing. The tidyverse package (v 1.3.0)
389 was used to wrangle and graph data (79).

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

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

394 **References**

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

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

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

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

- 495 of *Clostridioides difficile*-inoculated mice treated with vancomycin and fidaxomicin. Journal of
496 Infection and Chemotherapy 26:483–491.
- 497 38. Stroke IL, Letourneau JJ, Miller TE, Xu Y, Pechik I, Savoly DR, Ma L, Sturzenbecker LJ,
498 Sabalski J, Stein PD, Webb ML, Hilbert DW. 2018. Treatment of *Clostridium difficile* infection
499 with a small-molecule inhibitor of toxin UDP-glucose hydrolysis activity. Antimicrobial Agents and
500 Chemotherapy 62.
- 501 39. Quigley L, Coakley M, Alemayehu D, Rea MC, Casey PG, O'Sullivan, Murphy E, Kiely B, Cotter
502 PD, Hill C, Ross RP. 2019. *Lactobacillus gasseri* APC 678 reduces shedding of the pathogen
503 *Clostridium difficile* in a murine model. Frontiers in Microbiology 10.
- 504 40. Mullish BH, McDonald JAK, Pechlivanis A, Allegretti JR, Kao D, Barker GF, Kapila D, Petrof
505 EO, Joyce SA, Gahan CGM, Glegola-Madejska I, Williams HRT, Holmes E, Clarke TB, Thursz
506 MR, Marchesi JR. 2019. Microbial bile salt hydrolases mediate the efficacy of faecal microbiota
507 transplant in the treatment of recurrent *Clostridioides difficile* infection. Gut 68:1791–1800.
- 508 41. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. 2015. How informative is the mouse for human
509 gut microbiota research? Disease Models & Mechanisms 8:1–16.
- 510 42. Guh AY, Kutty PK. 2018. *Clostridioides difficile* infection 169:ITC49.
- 511 43. Tomkovich S, Lesniak NA, Li Y, Bishop L, Fitzgerald MJ, Schloss PD. 2019. The proton
512 pump inhibitor omeprazole does not promote *Clostridioides difficile* colonization in a murine model.
513 mSphere 4.
- 514 44. Lawley TD, Clare S, Walker AW, Goulding D, Stabler RA, Croucher N, Mastroeni P, Scott P,
515 Raisen C, Mottram L, Fairweather NF, Wren BW, Parkhill J, Dougan G. 2009. Antibiotic treatment of
516 *Clostridium difficile* carrier mice triggers a supershedder state, spore-mediated transmission, and
517 severe disease in immunocompromised hosts. Infection and Immunity 77:3661–3669.
- 518 45. 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 46. Jump RLP, Polinkovsky A, Hurless K, Sitzlar B, Eckart K, Tomas M, Deshpande A, Nerandzic
523 MM, Donskey CJ. 2014. Metabolomics analysis identifies intestinal microbiota-derived biomarkers
524 of colonization resistance in clindamycin-treated mice. PLoS ONE 9:e101267.
- 525 47. Nagao-Kitamoto H, Leslie JL, Kitamoto S, Jin C, Thomsson KA, Gilliland MG, Kuffa P, Goto Y,
526 Jenq RR, Ishii C, Hirayama A, Seekatz AM, Martens EC, Eaton KA, Kao JY, Fukuda S, Higgins PDR,
527 Karlsson NG, Young VB, Kamada N. 2020. Interleukin-22-mediated host glycosylation prevents
528 *Clostridioides difficile* infection by modulating the metabolic activity of the gut microbiota. Nature
529 Medicine 26:608–617.
- 530 48. Battaglioli EJ, Hale VL, Chen J, Jeraldo P, Ruiz-Mojica C, Schmidt BA, Rekdal VM, Till LM, Huq
531 L, Smits SA, Moor WJ, Jones-Hall Y, Smyrk T, Khanna S, Pardi DS, Grover M, Patel R, Chia N,
532 Nelson H, Sonnenburg JL, Farrugia G, Kashyap PC. 2018. *Clostridioides difficile* uses amino acids
533 associated with gut microbial dysbiosis in a subset of patients with diarrhea. Science Translational
534 Medicine 10:eaam7019.
- 535 49. Robinson CD, Auchtung JM, Collins J, Britton RA. 2014. Epidemic *Clostridium difficile* strains
536 demonstrate increased competitive fitness compared to nonepidemic isolates. Infection and
537 Immunity 82:2815–2825.
- 538 50. Collins J, Auchtung JM, Schaefer L, Eaton KA, Britton RA. 2015. Humanized microbiota mice
539 as a model of recurrent *Clostridium difficile* disease. Microbiome 3.
- 540 51. Collins J, Robinson C, Danhof H, Knetsch CW, Leeuwen HC van, Lawley TD, Auchtung JM,
541 Britton RA. 2018. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. Nature
542 553:291–294.
- 543 52. Hryckowian AJ, Treuren WV, Smits SA, Davis NM, Gardner JO, Bouley DM, Sonnenburg JL.
544 2018. Microbiota-accessible carbohydrates suppress *Clostridium difficile* infection in a murine
545 model. Nature Microbiology 3:662–669.
- 546 53. Fouladi F, Glenny EM, Bulik-Sullivan EC, Tsilimigras MCB, Sioda M, Thomas SA, Wang Y, Djukic

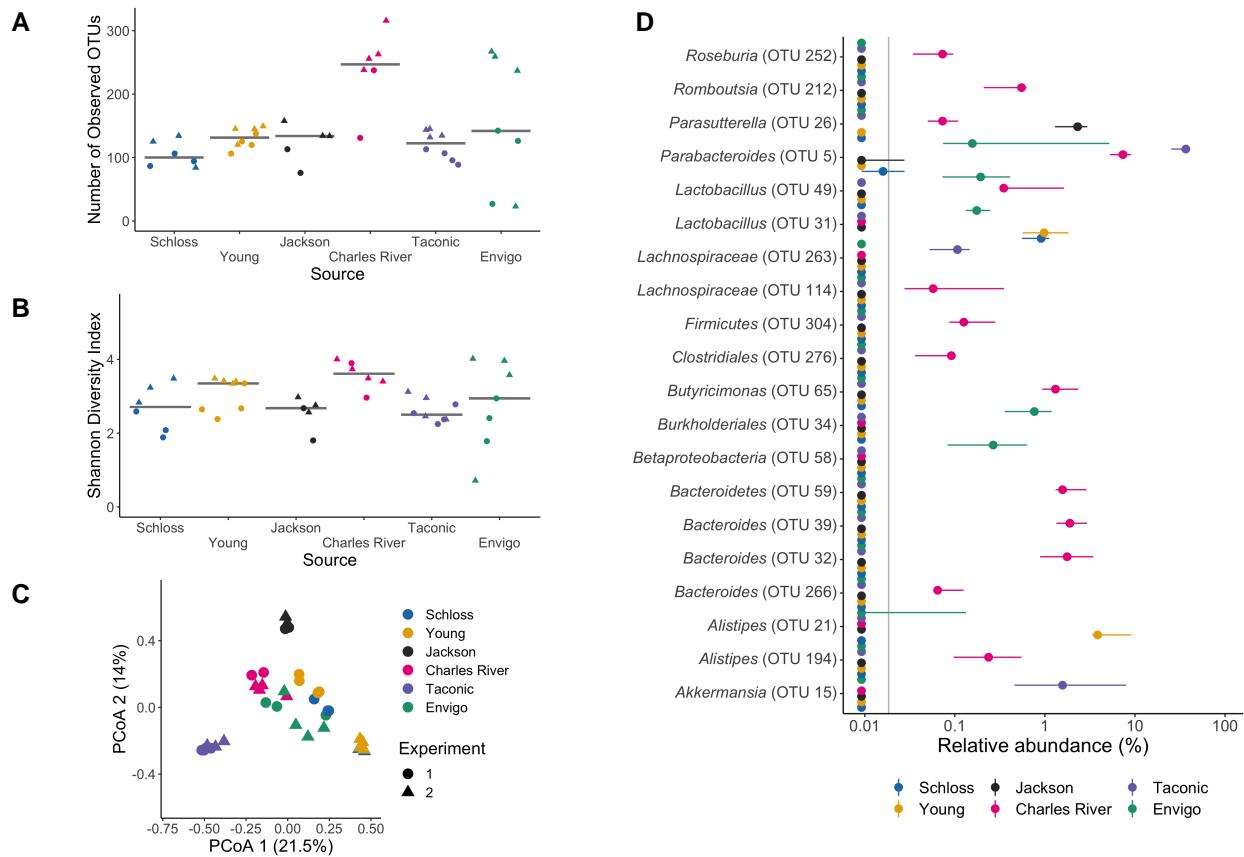
- 547 Z, Tang Q, Tarantino LM, Bulik CM, Fodor AA, Carroll IM. 2020. Sequence variant analysis reveals
548 poor correlations in microbial taxonomic abundance between humans and mice after gnotobiotic
549 transfer. *The ISME Journal* <https://doi.org/10.1038/s41396-020-0645-z>.
- 550 54. Walter J, Armet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

- 573 ER, Mucha PJ, Henderson JP. 2019. Metabolomic networks connect host-microbiome processes to
574 human *Clostridioides difficile* infections. *Journal of Clinical Investigation* 129:3792–3806.
- 575 62. Fletcher JR, Erwin S, Lanzas C, Theriot CM. 2018. Shifts in the gut metabolome and *Clostridium*
576 *difficile* transcriptome throughout colonization and infection in a mouse model. *mSphere* 3.
- 577 63. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, Li X, Long H, Zhang J, Zhang D, Liu C, Fang
578 Z, Chou J, Glanville J, Hao Q, Kotowska D, Colding C, Licht TR, Wu D, Yu J, Sung JJY, Liang Q, Li
579 J, Jia H, Lan Z, Tremaroli V, Dworzynski P, Nielsen HB, Bäckhed F, Doré J, Chatelier EL, Ehrlich
580 SD, Lin JC, Arumugam M, Wang J, Madsen L, Kristiansen K. 2015. A catalog of the mouse gut
581 metagenome. *Nature Biotechnology* 33:1103–1108.
- 582 64. Vital M, Rud T, Rath S, Pieper DH, Schlüter D. 2019. Diversity of bacteria exhibiting bile
583 acid-inducible 7alpha-dehydroxylation genes in the human gut. *Computational and Structural*
584 *Biotechnology Journal* 17:1016–1019.
- 585 65. Fransen F, Zagato E, Mazzini E, Fosso B, Manzari C, Aidy SE, Chiavelli A, D'Erchia AM,
586 Sethi MK, Pabst O, Marzano M, Moretti S, Romani L, Penna G, Pesole G, Rescigno M. 2015.
587 BALB/c and C57BL/6 mice differ in polyreactive IgA abundance, which impacts the generation of
588 antigen-specific IgA and microbiota diversity. *Immunity* 43:527–540.
- 589 66. Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee
590 CA, Lynch SV, Tanoue T, Imaoka A, Itoh K, Takeda K, Umesaki Y, Honda K, Littman DR. 2009.
591 Induction of intestinal th17 cells by segmented filamentous bacteria. *Cell* 139:485–498.
- 592 67. Azrad M, Hamo Z, Tkawkho L, Peretz A. 2018. Elevated serum immunoglobulin a levels in
593 patients with *Clostridium difficile* infection are associated with mortality. *Pathogens and Disease* 76.
- 594 68. Saleh MM, Frisbee AL, Leslie JL, Buonomo EL, Cowardin CA, Ma JZ, Simpson ME, Scully KW,
595 Abhyankar MM, Petri WA. 2019. Colitis-induced th17 cells increase the risk for severe subsequent
596 *Clostridium difficile* infection. *Cell Host & Microbe* 25:756–765.e5.
- 597 69. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a
598 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the

- 599 MiSeq illumina sequencing platform. *Applied and Environmental Microbiology* 79:5112–5120.
- 600 70. Sze MA, Schloss PD. 2019. The impact of DNA polymerase and number of rounds of
601 amplification in PCR on 16S rRNA gene sequence data. *mSphere* 4.
- 602 71. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
603 Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF.
604 2009. Introducing mothur: Open-source, platform-independent, community-supported software
605 for describing and comparing microbial communities. *Applied and Environmental Microbiology*
606 75:7537–7541.
- 607 72. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012.
608 The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools.
609 *Nucleic Acids Research* 41:D590–D596.
- 610 73. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR,
611 Tiedje JM. 2013. Ribosomal database project: Data and tools for high throughput rRNA analysis.
612 *Nucleic Acids Research* 42:D633–D642.
- 613 74. Westcott SL, Schloss PD. 2017. OptiClust, an improved method for assigning amplicon-based
614 sequence data to operational taxonomic units. *mSphere* 2.
- 615 75. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB,
616 Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2018. Vegan: Community ecology
617 package.
- 618 76. R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for
619 Statistical Computing, Vienna, Austria.
- 620 77. Kuhn M. 2008. Building predictive models in RUsing the caret Package. *Journal of Statistical*
621 *Software* 28.
- 622 78. Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. 2020. A framework for effective
623 application of machine learning to microbiome-based classification problems. *mBio* 11.

624 79. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes
625 A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D,
626 Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the
627 tidyverse. *Journal of Open Source Software* 4:1686.

628 **Figures**



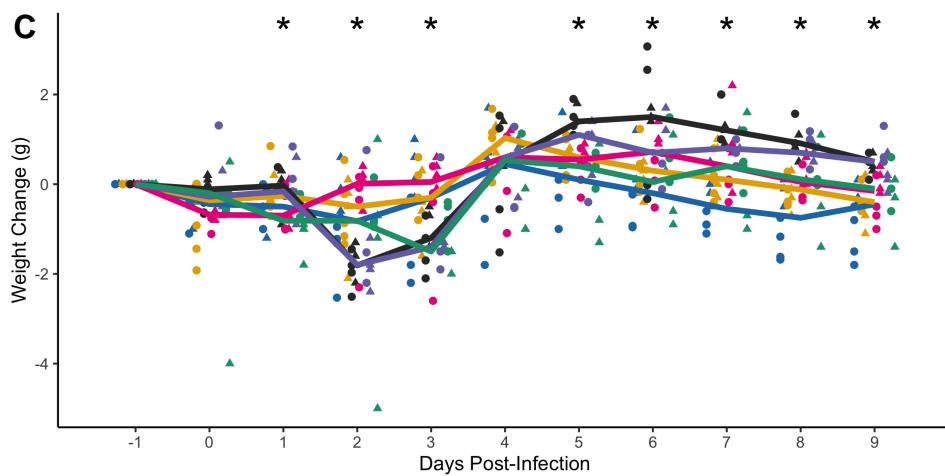
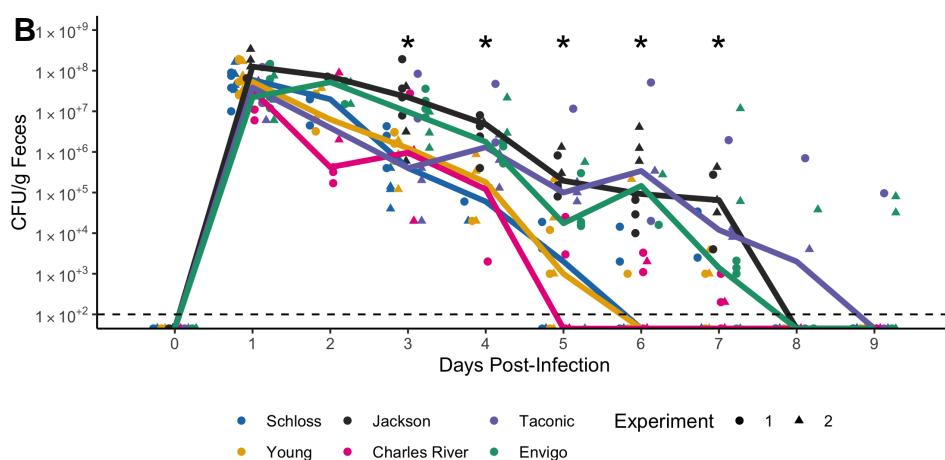
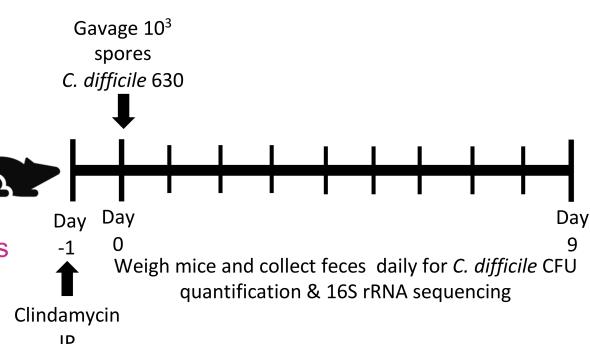
629

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

A

Sources of C57BL/6 mice:

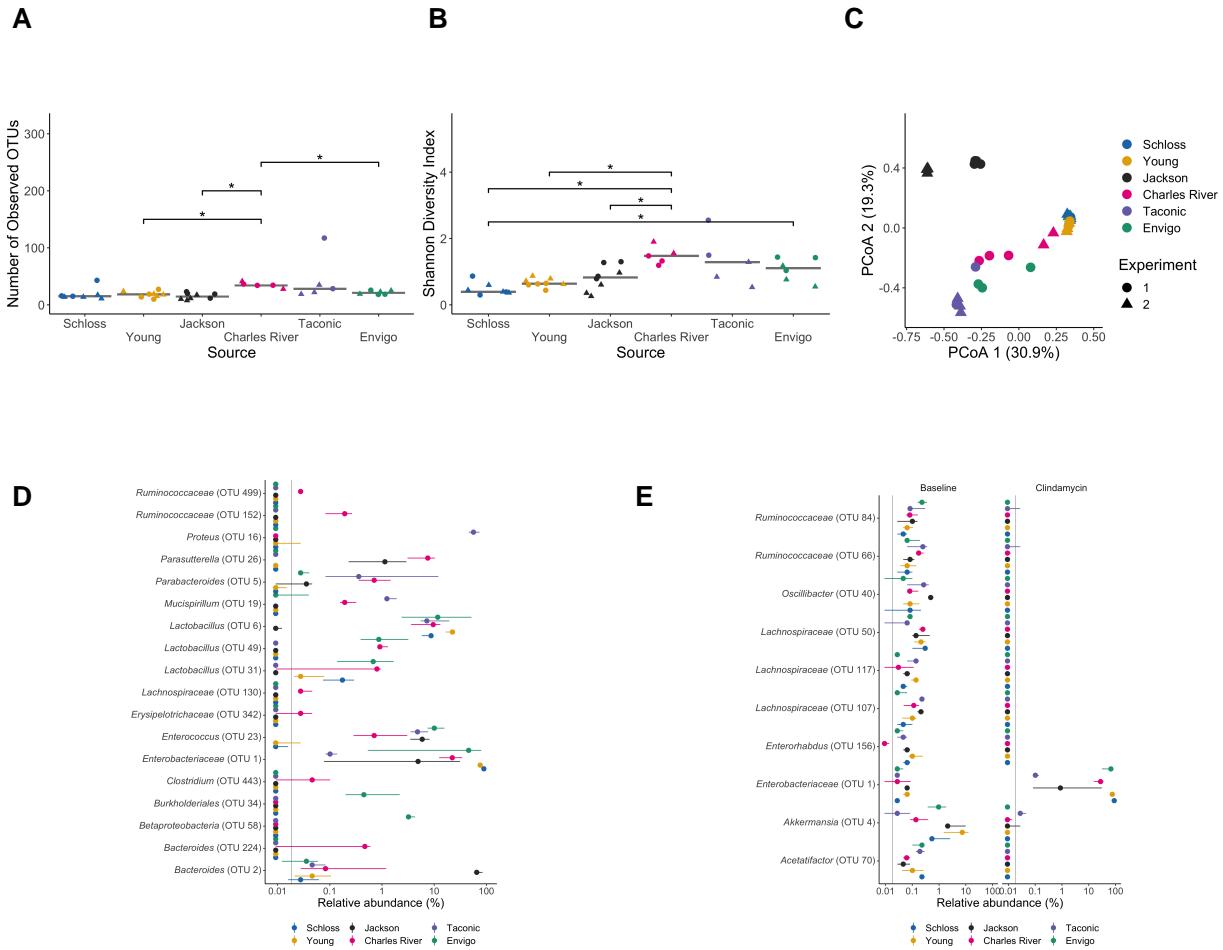
1. Schloss Lab Colony at University of Michigan
2. Young Lab Colony at University of Michigan
3. The Jackson Laboratory
4. Charles River Laboratories
5. Taconic Biosciences
6. Envigo



643

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

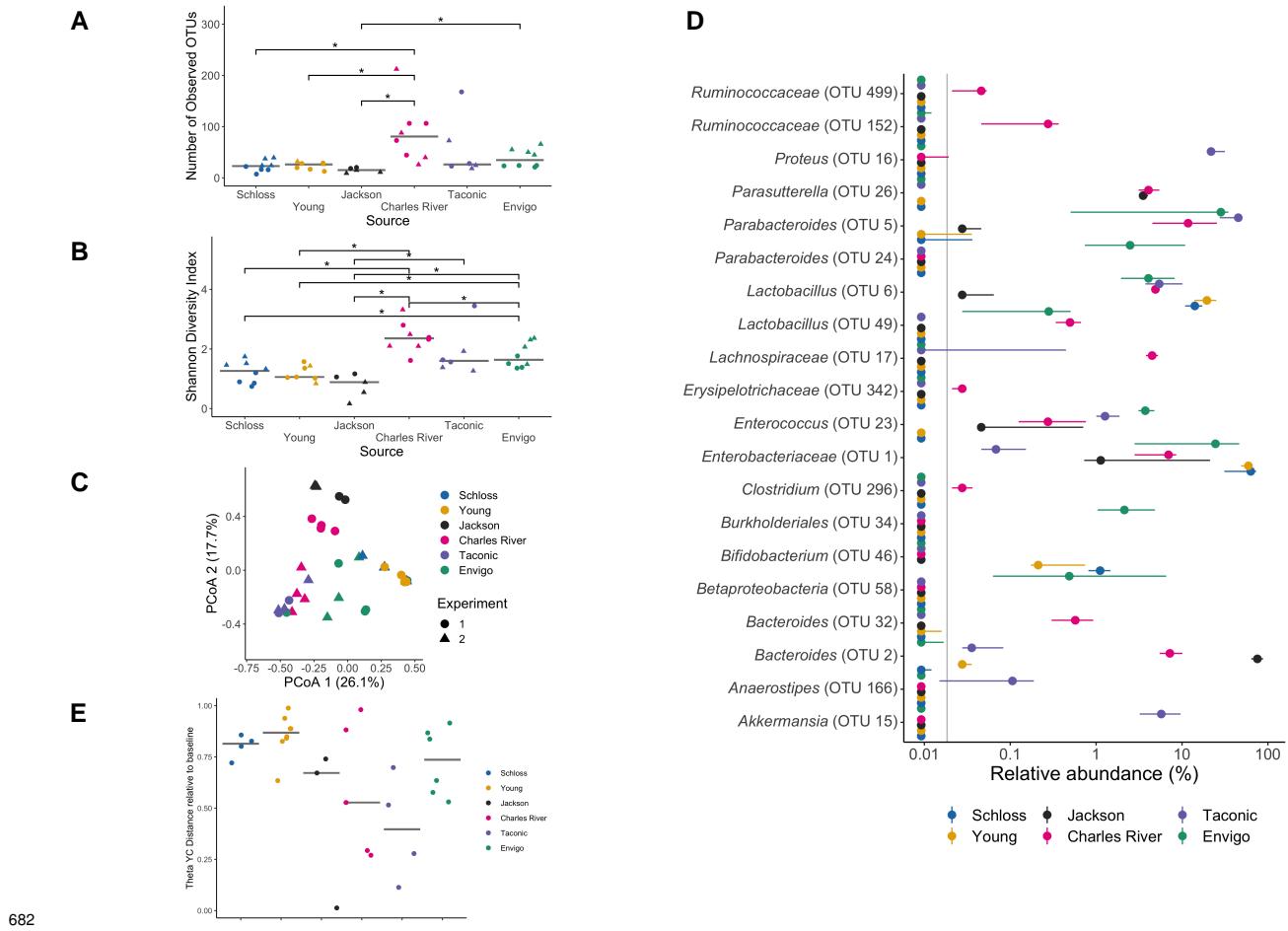
648 (N = 8), Taconic Biosciences (N = 8), and Envigo (N = 8). All mice were administered 10 mg/kg
649 clindamycin intraperitoneally (IP) 1 day before challenge with *C. difficile* 630 spores on day 0.
650 Mice were weighed and feces was collected daily through the end of the experiment (9 days
651 post-infection).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



663

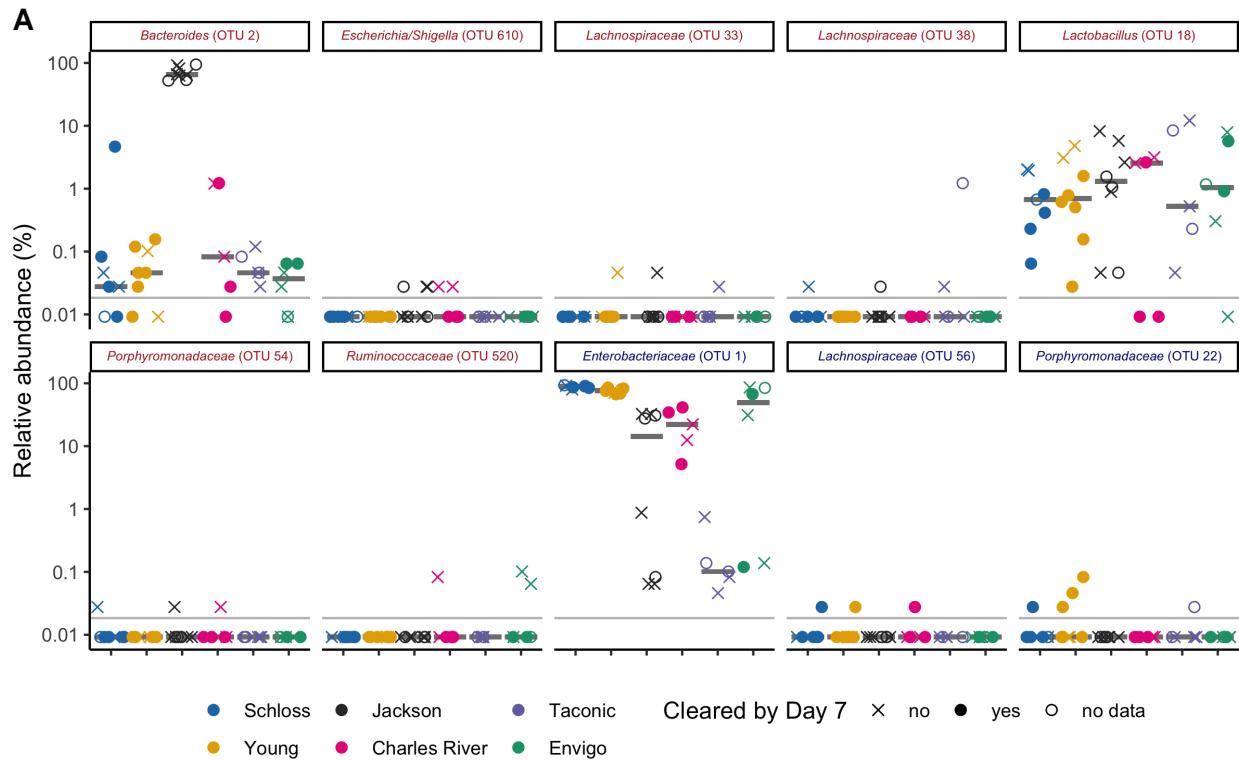
664 **Figure 3. Clindamycin treatment alters bacteria in all sources, but a subset of bacterial**
 665 **differences across sources persists.** A-B. Number of observed OTUs (A) and Shannon diversity
 666 index values (B) across sources of mice after clindamycin treatment (day 0). Differences between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

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



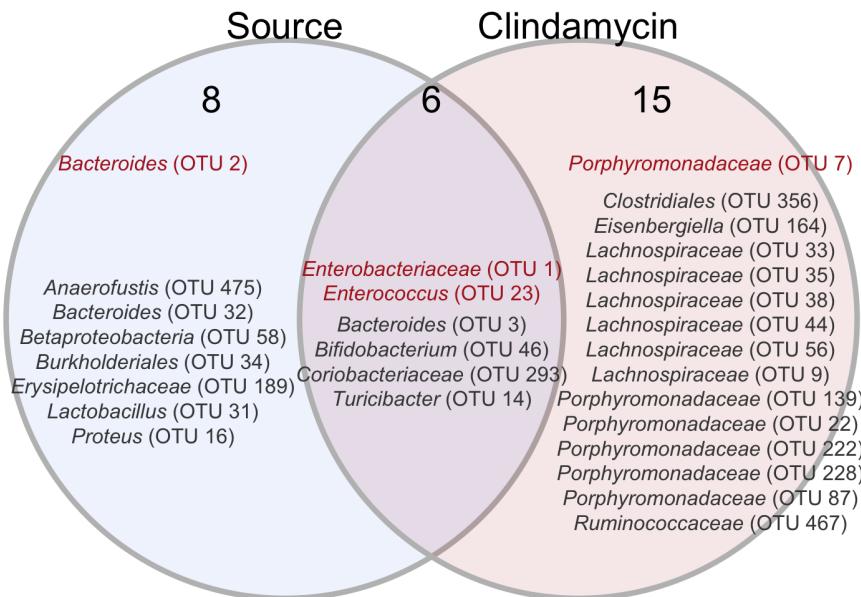
682 **Figure 4. Microbiota variation across sources is maintained after *C. difficile* challenge.** A-B.
683 Number of observed OTUs (A) and Shannon diversity index values (B) across sources of mice
684 1-day post-infection. Data were analyzed by Kruskal-Wallis test with Benjamini-Hochberg correction
685 for testing each day of the experiment and the adjusted *P* value was < 0.05 (Table S1). Significant
686 *P* values from the pairwise Wilcoxon comparisons between sources with Benjamini-Hochberg
687 correction are shown (Table S2). PCoA of θ_{YC} distances of 1-day post-infection stool samples.
688 Source and the interaction between source and cage effects explained most of the variation
689 between fecal communities (PERMANOVA combined $R^2 = 0.88$, $P < 0.001$, Table S3). For A-C:
690 each symbol represents the value for a stool sample from an individual mouse, circles represent
691 experiment 1 mice and triangles represent experiment 2 mice. D. Plots highlighting the median
692 (point) and interquartile range (colored lines) of the relative abundances for the top 20 bacteria out
693 of the 44 OTUs that varied between sources 1-day post-infection. The gray vertical line indicates
694 the limit of detection. For each timepoint OTUs with differential relative abundances across sources

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



B

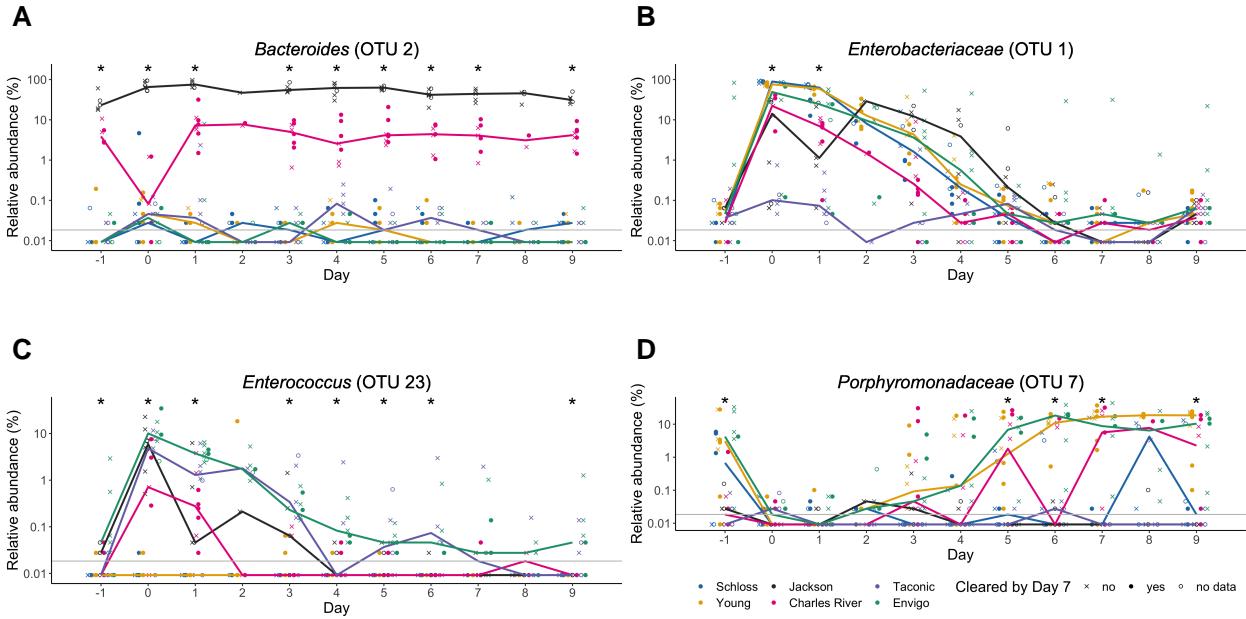
OTU comparisons for day -1, 0, and 1 models



700

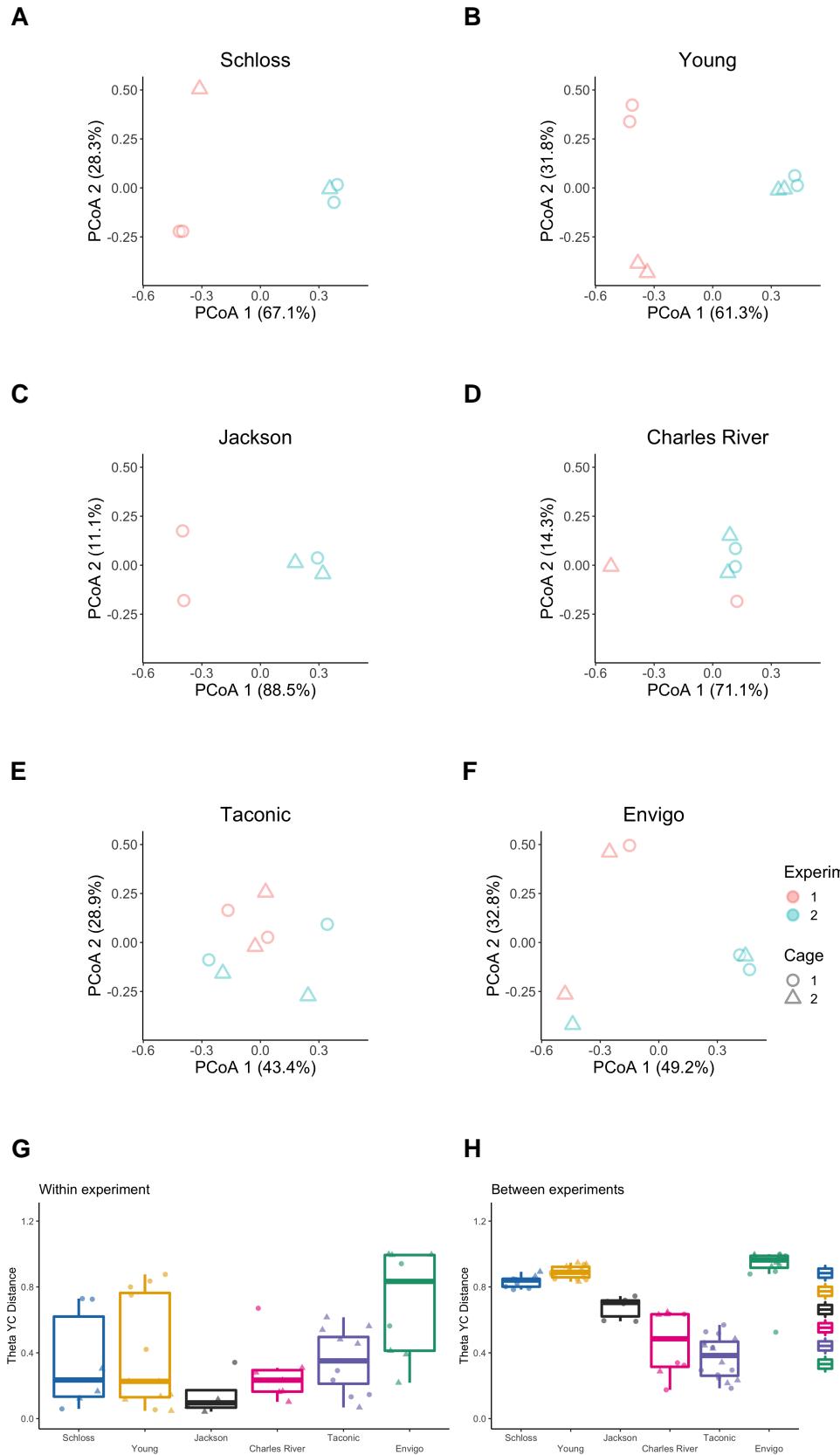
701 **Figure 5. Bacteria that influence whether mice cleared *C. difficile* by day 7.** A. Baseline (day
702 0) relative abundance data for the 10 OTUs with the highest rankings based on feature weights in

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

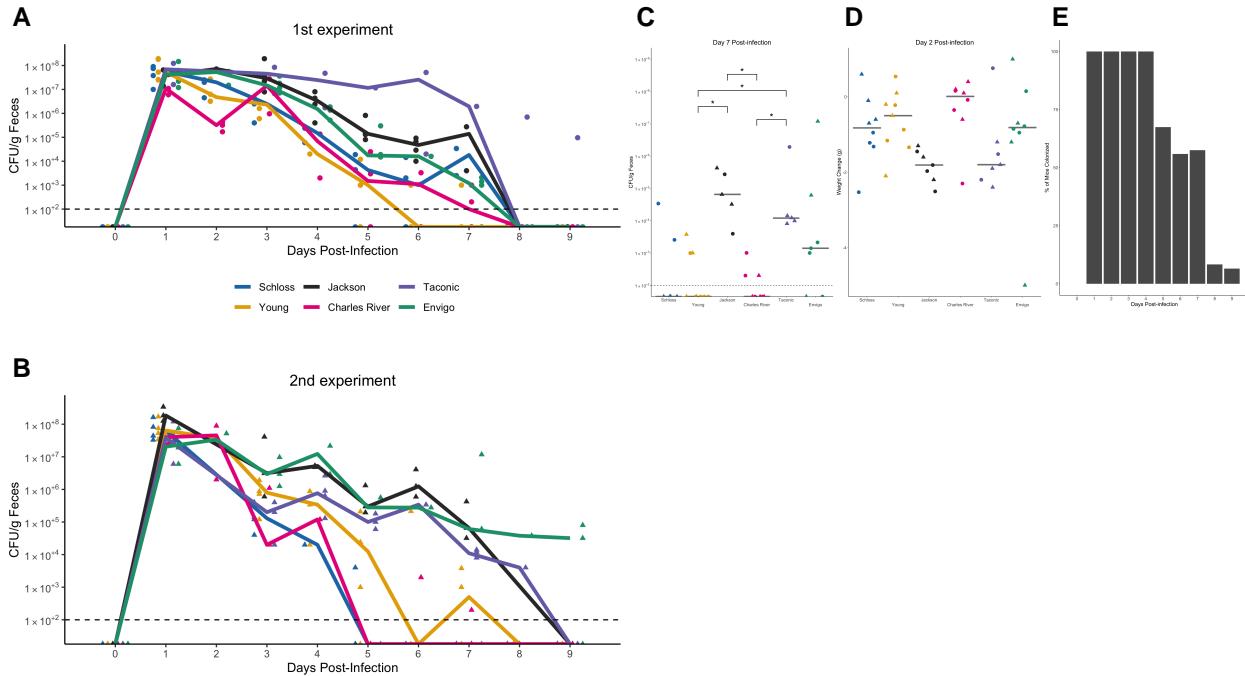


710

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

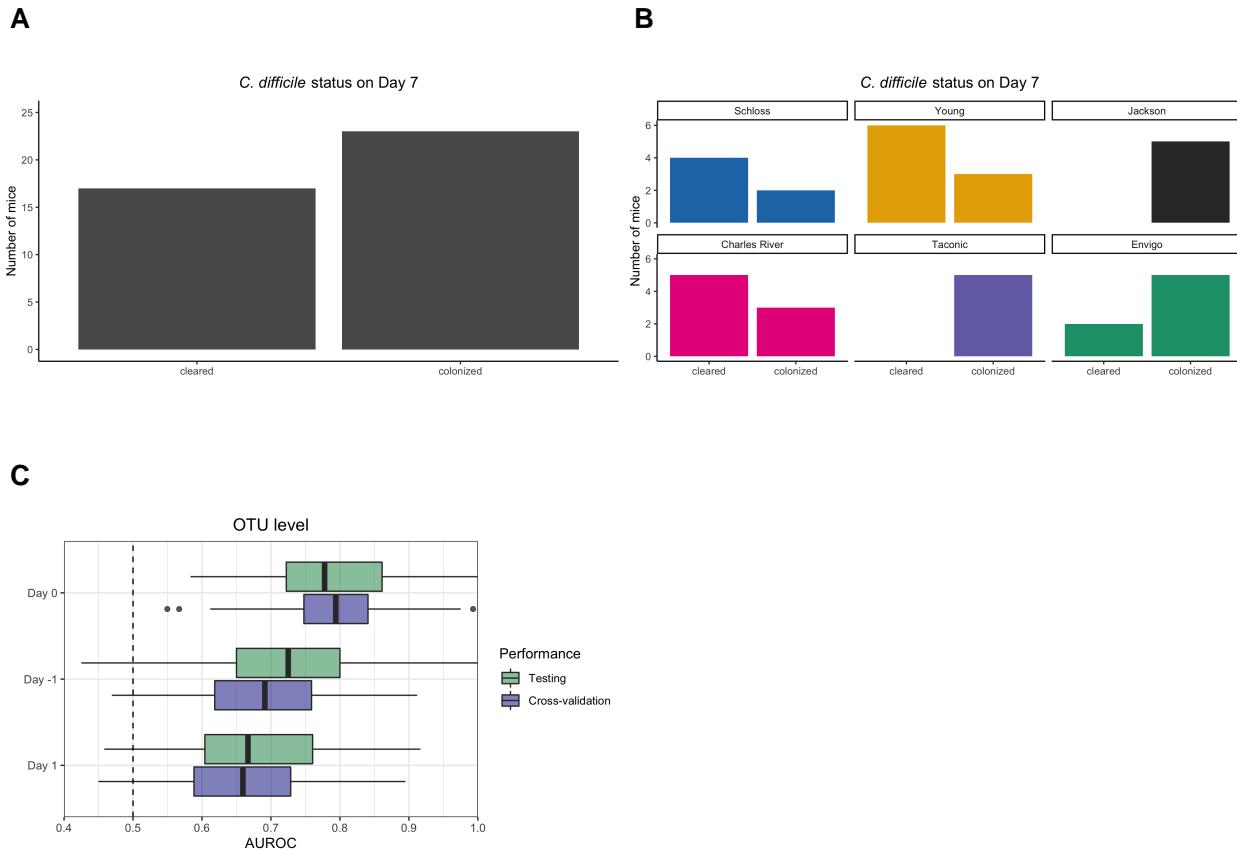
**Figure S1.**

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



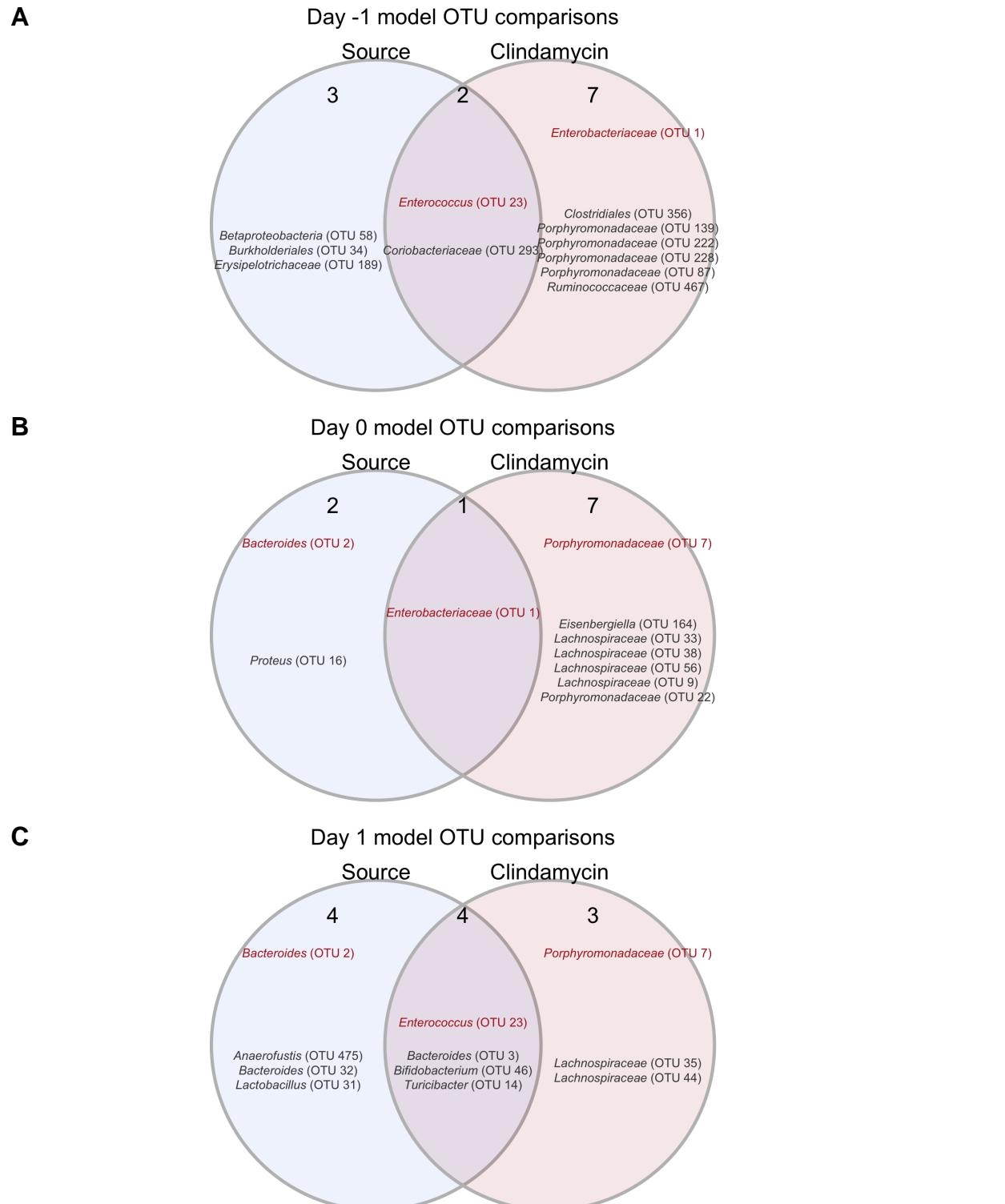
734

735 **Figure S2. *C. difficile* CFU variation across sources varies slightly between the 2**
 736 **experiments.** A-B. *C. difficile* CFU/gram of stool quantification over time for experiment 1 (A) and
 737 (B). Experiments were conducted approximately 3 months apart. Lines represent the median
 738 CFU for each source, symbols represent individual mice and the black line represents the limit
 739 of detection. C. *C. difficile* CFU/gram stool 7-days post-infection across sources of mice with an
 740 asterisk for pairwise Wilcoxon comparisons with Benjamini-Hochberg correction where $P < 0.05$.
 741 D. Mouse weight change 2-days post-infection across sources of mice, no pairwise Wilcoxon
 742 comparisons were significant after Benjamini-Hochberg correction. For C-D: circles represent
 743 experiment 1 mice, triangles represent experiment 2 mice and gray lines indicate the median
 744 values for each group. E. Percent of mice that were colonized with *C. difficile* over the course of the
 745 experiment. Each day the percent is calculated based on the mice where *C. difficile* CFU was
 746 quantified for that particular day. Total N for each day: day 1 (N = 42), day 2 (N = 20), day 3 (N =
 747 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 =
 748 46).



749

750 **Figure S3. Bacterial community composition before, after clindamycin perturbation, and**
 751 **post-infection can predict *C. difficile* colonization status 7 days post-challenge.** A. Bar
 752 graph visualizations of overall 7-days post-infection *C. difficile* colonization status that were used as
 753 classification outcomes to build L2-regularized logistic regression models. Mice were classified as
 754 colonized or cleared (not detectable at the limit of detection of 100 CFU) based on CFU g/stool data
 755 from 7 days post-infection. B. *C. difficile* CFU status on Day 7 within each mouse source. N = 8-9
 756 mice per group. C. L2-regularized logistic regression classification model area under the receiving
 757 operator characteristic curve (AUROCs) to predict *C. difficile* CFU on day 7 post-infection (Fig. 2B,
 758 Fig. S2C) based on the OTU community relative abundances at baseline (day -1), post-clindamycin
 759 (day 0), and 1-day post-infection. All models performed better than random chance (AUROC = 0.5,
 760 all $P \leq 1.6e-17$, Table S12) and the model built with post-clindamycin treated bacterial OTU relative
 761 abundances had the best performance ($P_{FDR} \leq 3.3e-06$ for all pairwise comparisons, Table S13).
 762 For list of the 20 OTUs that were ranked as most important to each model, see Table S14.



Figure

763
764 **S4. OTUs from classification models based on baseline, post-clindamycin treatment, or**
765 **post-infection community data vary by source, clindamycin treatment, or both.** A-C. Venn
766 diagrams of OTUs from the top 20 OTUs from the baseline (A), post-clindamycin treatment (B), and

⁷⁶⁷ post-infection (C) classification models (Table S14) that overlapped with OTUs that varied across
⁷⁶⁸ sources at the corresponding timepoint (Tables S5, 8, 10), were impacted by clindamycin treatment
⁷⁶⁹ (Table S9), or both. Red OTUs signify OTUs that were important to more than 1 classification
⁷⁷⁰ model.

771 **Supplementary Tables and Movie**

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

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

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

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

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

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

785 **Table S5. OTUs with relative abudances that significantly vary between sources at baseline**
786 **(day -1).**

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

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

789 **Table S8. OTUs with relative abudances that significantly vary between sources**
790 **post-clindamycin (day 0).**

791 **Table S9. OTUs with relative abudances that significantly changed after clindamycin**
792 **treatment.**

793 **Table S10. OTUs with relative abudances that significantly vary between sources**
794 **post-infection (day 1).**

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

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

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

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

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