1 **Title:** Personalized *Clostridioides difficile* engraftment risk prediction and probiotic
2 therapy assessment in the human gut.

3 **Authors:** Alex Carr[1,2], Nitin S. Baliga[1,2,3,4], Christian Diener[1,5,*], and Sean M.
4 Gibbons[1,2,6,7,*]
5 **Affiliations:**
6
7 [1] Institute for Systems Biology, Seattle, WA, USA
8 [2] Molecular Engineering Program, University of Washington, Seattle, WA, USA
9 [3] Departments of Biology and Microbiology, University of Washington, Seattle, WA, USA
10 [4] Lawrence Berkeley National Lab, Berkeley, CA, USA
11 [5] Diagnostic and Research Institute of Hygiene, Microbiology and Environmental
12 Medicine, Medical University of Graz, Graz, Austria
13 [6] Departments of Bioengineering and Genome Sciences, University of Washington,
14 Seattle, WA, USA
15 [7] eScience Institute, University of Washington, Seattle, WA, USA
16 [*] correspondence can be addressed to cdiener@isbscience.org and
17 sgibbons@isbscience.org
18

19 **Abstract:**

20 *Clostridioides difficile* colonizes up to 30-40% of community-dwelling adults without
21 causing disease. *C. difficile* infections (CDIs) are the leading cause of antibiotic-
22 associated diarrhea in the U.S. and typically develop in individuals following disruption
23 of the gut microbiota due to antibiotic or chemotherapy treatments. Further treatment of
24 CDI with antibiotics is not always effective and can lead to antibiotic resistance and
25 recurrent infections (rCDI). The most effective treatment for rCDI is the reestablishment
26 of an intact microbiota via fecal microbiota transplants (FMTs). However, the success of
27 FMTs has been difficult to generalize because the microbial interactions that prevent
28 engraftment and facilitate the successful clearance of *C. difficile* are still only partially
29 understood. Here we show how microbial community-scale metabolic models (MCMMs)
30 accurately predicted known instances of *C. difficile* colonization susceptibility or
31 resistance *in vitro* and *in vivo*. MCMMs provide detailed mechanistic insights into the
32 ecological interactions that govern *C. difficile* engraftment, like cross-feeding or
33 competition involving metabolites like succinate, trehalose, and ornithine, which differ
34 from person to person. Indeed, three distinct *C. difficile* metabolic niches emerge from
35 our MCMMs, two associated with positive growth rates and one that represents non-
36 growth, which are consistently observed across 15,204 individuals from five
37 independent cohorts. Finally, we show how MCMMs can predict personalized
38 engraftment and *C. difficile* growth suppression for a probiotic cocktail (VE303)
39 designed to replace FMTs for the treatment rCDI. Overall, this powerful modeling
40 approach predicts personalized *C. difficile* engraftment risk and can be leveraged to
41 assess probiotic treatment efficacy. MCMMs could be extended to understand the
42 mechanistic underpinnings of personalized engraftment of other opportunistic bacterial
43 pathogens, beneficial probiotic organisms, or more complex microbial consortia.
44

Introduction

The human gut microbiome plays important roles in shaping host metabolism, in the development of chronic diseases, and in preventing opportunistic pathogen colonization and infection [1–3]. The metabolic versatility of gut bacteria allows for the stable coexistence of hundreds of commensal species within the gastrointestinal tract [4]. Some species extract energy and nutrition directly from indigestible dietary substrates, like plant fibers or recalcitrant proteins, while others subsist largely on host-derived mucosal glycans or on the vast array of metabolic byproducts produced by primary fiber, protein, and mucus degraders [5,6]. Saturation of these metabolic niches by commensal microbes can prevent colonization and engraftment by external microbes that may share a similar niche, including pathobionts [7–9].

Perturbations to the gut microbiome (e.g., antibiotic use or diarrhoeal events) provide a window of opportunity for pathobiont colonization [10], which could in turn lead to the development of disease following subsequent perturbations [11,12]. One such pathobiont, *Clostridioides difficile*, is the most common hospital acquired gastrointestinal infection in the U.S. [13,14]. *C. difficile* colonizes as much as 30-40% of community-dwelling adults without causing disease, lying in wait until the opportunity for infection arises [15,16]. During active *C. difficile* infection (CDI), antibiotic treatment can be effective in suppressing *C. difficile* growth, but antibiotics also disrupt the ecology of the commensal microbiota and potentiate reinfection if *C. difficile* is not completely cleared by the treatment [11,12]. Thus, an intact gut microbiota that prevents *C. difficile* colonization and engraftment is critical to the host's defense against CDIs [3]. This

68 understanding has led to the widespread use of fecal microbiota transplants (FMTs) as

69 a means of combating cases of recurrent CDI (rCDI), where antibiotic treatment proves

70 insufficient [17]. While the biology of *C. difficile* has been fairly well-characterized in the

71 context of disease, the pre-disease mechanisms of *C. difficile* colonization and

72 engraftment are still poorly understood, as are the factors that govern *C. difficile*

73 decolonization and FMT efficacy [10].

74   There are currently no mechanistically-grounded, generalizable approaches to

75 accurately predicting the engraftment of an exogenous bacterial taxon in the context of

76 a given microbiota. Previous work has leveraged machine learning (ML) to predict the

77 engraftment of FMT donor strains in FMT recipients [18]. While effective and relatively

78 accurate, this kind of quasi-black-box ML approach does not provide a means of

79 understanding the molecular mechanisms that facilitate or prevent engraftment. Here,

80 we present an alternative approach to engraftment prediction that leverages microbial

81 community-scale metabolic models (MCMMs), which provide detailed, mechanistic

82 information on the ecological interactions within individual microbiota that prevent or

83 facilitate engraftment, in addition to generating accurate engraftment predictions.

84   Genome-scale metabolic models and classical flux balance analysis (FBA) have

85 been invaluable tools for exploring how environmental conditions impact the metabolic

86 capacities of individual bacterial taxa grown *in vitro* [19]. However, extending these

87 methods to complex, multi-species communities has proved to be a challenge.

88 Recently, we developed an approach called cooperative tradeoff flux balance analysis

89 (ctFBA), which leverages microbiome compositional and dietary constraints to rapidly

90   estimate steady-state community-scale metabolic fluxes [20,21]. Here, we use publicly

91   available 16S amplicon and shotgun metagenomic data from *in vitro* and *in vivo* studies

92   with both known and unknown *C. difficile* colonization dynamics, along with our

93   community-scale metabolic modeling framework, called MICOM [20], to build and test

94   MCMMs to estimate *C. difficile* engraftment potential within a given microbiome and

95   dietary/environmental context. We present novel insights into how *C. difficile* can

96   occupy three discrete metabolic niches across commensal communities, what metabolic

97   interactions within gut bacterial communities promote or prevent colonization, and we

98   show how we can predict potential responders and non-responders to a defined

99   probiotic cocktail that has recently shown efficacy in the treatment of rCDI [22]. Overall,

100   MCMMs provide a novel path towards predicting *C. difficile* engraftment risk.

101   Furthermore, these models can be leveraged to design precision dietary or probiotic

102   interventions aimed at decolonizing individuals who are already carrying *C. difficile* and

103   preventing engraftment in those who are not. Finally, we suggest that MCMMs could

104   enable precision engineering of the gut microbiome through personalized engraftment

105   predictions for other pathobionts beyond *C. difficile*, probiotic bacterial strains, or for

106   entire microbial consortia (e.g., FMTs from different donors), in the context of a specific

107   diet.

108
109
110   **Results**
111

112   *Development of an in silico invasion assay to simulate C. difficile colonization*

113   To simulate the colonization of *C. difficile* we developed an *in silico* invasion assay that

114   leverages microbiome relative abundance data, manually curated genome-scale

115   metabolic models of gut bacteria from the AGORA database, constraints on the

116   diet/environment, and the MICOM modeling framework (Fig. 1A) [20,23]. Here, we used

117   several existing 16S amplicon and shotgun metagenomic sequencing data sets to

118   validate our approach [10,18,24–27]. Amplicon sequencing data is often limited to genus-

119   level resolution in the taxonomic classifications of amplicon sequence variants (ASVs).

120   Even with metagenomic sequencing data, species level resolution can be poor when

121   mapping to available species-level metabolic models. Generally, only the most

122   abundant or prevalent commensal bacterial species are well represented. Therefore, we

123   focused on genus-level MCMMs, to approximate the community metabolic context, for

124   our invasion assays (see Methods) [28]. However, we also evaluated the use of species-

125   level MCMMs using *in vitro* 16S amplicon and stool shotgun metagenomic data (Fig.

126   S1A). Briefly, strain-level metabolic models from AGORA were combined at the genus

127   or species level, to account for potential coexistence of multiple strains from a given

128   genus or species within an individual and to reduce potential bias from arbitrarily

129   selecting individual strain models. Using this approach ~75% of reads, on average,

130   could be mapped to a genus-level metabolic model within the AGORA database from

131   16S amplicon data, while ~90% of reads, on average, could be mapped using shotgun

132   metagenomic data. At the species level, only ~50% of reads could be mapped, on

133   average, using the same *in vivo* metagenomics data, while 100% of reads could be

134   mapped for *in vitro* communities profiles using 16S amplicon sequencing (i.e., because

135   type strains were used in these experiments), illustrating the inherent limitation of using

136 species level models for *in vivo* applications and further justifying our focus on genus-

137 level analyses (Fig. S1A). To simulate the invasion of *C. difficile* into these model

138 communities, a pan-species model of *Clostridioides*, representing all four common *C.*

139 *difficile* strains present in the AGORA database (including hypervirulent and non-

140 epidemic strains), was introduced at a relative abundance of 10% (see below for

141 justification for this percentage), while other community relative abundances were

142 decreased proportionally to approximate a minor perturbation in community-wide

143 biomass (Figure 1A). Growth simulations were then performed using a medium

144 representing an average European diet (i.e., a standard developed-world diet

145 appropriate to the cohorts studied here), with fluxes of metabolites known to be

146 absorbed in the small intestine decreased by 90%, as previously described for *in vivo*

147 samples [20]. To represent the anaerobic basal broth (ABB) used to culture the *in vitro*

148 communities, a curated medium was constructed using a previously described

149 representation of Luria-Bertani (LB) broth as a template and the known composition of

150 ABB (see Methods for further details) [27,29]. Growth rates were estimated using ctFBA,

151 as implemented in MICOM, which uses a regularization step and allows for a

152 suboptimal community growth rate in order to achieve a more realistic growth rate

153 distribution across the community [20,21]. Import and export fluxes were estimated using

154 parsimonious enzyme usage FBA (pFBA) [20].

155   Personalized MCMMs were constructed for each sample and the potential for *C.*

156 *difficile* engraftment was quantified as the model-inferred growth rate. ctFBA has a

157 single free parameter that needs to be chosen, the tradeoff between community-wide

158 growth rates and individual, taxon-specific growth rates. Assuming that most taxa

159   detected at appreciable abundances in a gut microbiome are actively growing *in vivo*

160   and *in vitro*, a trade-off value was selected by choosing the minimal deviation from

161   optimal community growth for which >70% of genera obtained non-zero growth rates on

162   average (Fig. S1B). We found that with a trade-off value of 0.8 (i.e., 80% of maximal

163   community biomass production) resulted in a mean fraction of genera with non-zero

164   growth of >70% using genus-level models constructed from 16S amplicon data. This

165   same non-zero growth fraction was achieved with a trade-off value of 0.9 for the genus-

166   level models constructed from metagenomics data. The 16S and metagenomics specific

167   tradeoff values were also used for the species level models derived from these data

168   types, respectively. At the species-level these values were associated with much higher

169   growth fractions (~90% and 95% for 16S amplicon and shotgun metagenomics,

170   respectively; Fig. S1B).

171   We first validated our approach using an *in vitro* 16S amplicon data set where *C.*

172   *difficile* was co-cultured with communities of human commensal gut bacteria at differing

173   levels of species richness [27]. We found that MCMM-inferred *C. difficile* biomass

174   (calculated using the MCMM-inferred *C. difficile* growth rates, assuming exponential

175   growth over a fixed interval, see Methods) accurately reflected a negative trend in

176   measured *C. difficile* abundance along a gradient of increasing species richness and

177   was significantly correlated with empirically measured *C. difficile* biomass after 48 hours

178   of growth (Fig. 1B; ordinary least squares (OLS) $R^2=0.45$, $p < 10^{-16}$). Similarly, we found

179   that the normalized *C. difficile* growth rate (normalized to the overall MCMM-inferred

180   community growth rate, see Methods) decreased as a function of community richness

181   and was a even stronger predictor of the empirically-observed *C. difficile* biomass after

182   48 hours of growth (Fig. 1C; ordinary least squares (OLS) R^2=0.5, p <10^{-16}). Biclustering

183   of the inferred species growth rates suggested that growth of *C. difficile* was largely

184   suppressed by *B. thetaiotaomicron, P. vulgatus*, and *E. lenta in vitro*, as indicated by the

185   strong negative associations between the growth rates of these species and the growth

186   rate of *C. difficile* (Pearson r <-0.45, p <0.01 for species)*.* Positive associations between

187   the growth rates of *C. difficile, C. scindens*, and *P. hiranonis* (Pearson r >0.8, p <10^{-6}

188   for all species) were inline with the observation that these species tend to share

189   overlapping metabolic niches. The previously-observed suppressive effects of *C.*

190   *scindens* and *P. hiranonis* on *C. difficile* were not apparent within our modeling

191   framework [27], likely due to the fact that pH and its effects on growth are not easily

192   captured by FBA models. However, while *in vitro* acidification is a strong environmental

193   force in anaerobic batch culture, we do not expect variation in pH to be as prominent in

194   the buffered environment of the human colon.

195   To further validate our approach *in vivo* we leveraged a time series with a known

196   *C. difficile* colonization event (Fig. 1E) [10,30]. We found that estimated *C. difficile* growth

197   rates were at or below the limit of solver accuracy (<10^{-6}, which effectively indicates an

198   absence of growth) in samples collected prior to colonization and comparable to growth

199   rates of other dominant genera in samples taken after the initial colonization event

200   (Welch's t-test t=-3.29, p=0.003, for comparison of $\log_{10}$ growth rates before and after

201   the known colonization event; Fig. 1E). Furthermore, we saw patchy engraftment

202   predictions in a second individual that was known to be colonized by *C. difficile* at a low

203   level (i.e., near the limit of detection) throughout a time series (Fig. 1F). We also

204   assessed the importance of propagule pressure [31] (i.e., the relative abundance at which

205 the invasive taxon is introduced into the models) and found that below 10% relative

206 abundance, agreement between growth rate estimates and measured abundances was

207 poor (Fig. S1C). Thus, propagule pressure plays an important role in predicted

208 engraftment success [27].

209 Based on these initial results (Fig. S1), we decided to use a fixed tradeoff value

210 of 0.8 and a *C. difficile* invasion fraction of 10% for all subsequent analyses of 16S

211 amplicon data (i.e., the highest value at which >70% of the community showed a

212 positive growth rate). Using the same heuristic, a tradeoff value of 0.9 was chosen for

213 the metagenomics analyses, along with a 10% *C. difficile* invasion fraction (Fig. S1B).

214 *in silico invasion assay accurately predicts C. difficile colonization potential in rCDI*

215 *patients pre- and post- FMT*

216 We applied our *in silico* invasion model to two separate datasets (16S amplicon and

217 metagenomic shotgun sequencing, respectively) of rCDI patients who received FMTs

218 and were subsequently followed over time [18,26]. These data provided an additional test

219 of MCMM performance and a means to explore the metabolic features associated with

220 community-scale colonization susceptibility or resistance across a larger population.

221 Given that all individuals in the rCDI cohorts had experienced multiple recurrent

222 infections, we expected pre-FMT microbiomes from these patients to be more

223 susceptible to invasion. Additionally, both of the original studies found that nearly all the

224 patient microbiomes transitioned to a compositional state that was much closer to the

225 healthy controls post-FMT than to their pre-FMT compositional states [18,26]. Thus, we

226 expected post-FMT samples would be less susceptible to invasion. For the sake of

227 comparison of results across studies, genus-level MCMMs were used for both analyses.

228 Across both data sets, rCDI patients had significantly higher MCMM-predicted *C.*

229 *difficile* growth rates prior to FMT treatment than they did post FMT treatment (Fig. 2A;

230 Welch's t-test *t*=-3.19, *p*=0.001 for comparison of pre-FMT vs. post-FMT). Predicted *C.*

231 *difficile* growth rates were negatively associated with Shannon diversity, albeit weakly

232 (Fig. 2B; OLS R^2=0.05, *p*=0.01), which is in line with prior empirical observations

233 indicating that lower diversity communities are more susceptible to *C. difficile*

234 colonization and the development of rCDI [9,27,32,33].

235 The community-scale import flux profile prior to *in silico* invasion was predictive

236 of *C. difficile* growth rate following invasion in Weingarden *et al.* (Fig. 2C). High-

237 dimensional community-scale import flux profiles were projected into a two-dimensional

238 space using the Uniform Manifold Approximation and Projection (UMAP) technique (Fig.

239 2C) [34]. The UMAP projection provides a visual means of identifying patterns in the high

240 dimensional import flux space. The closer points are to one another in this ordination

241 the more similar their import flux profiles are. Thus, clusters of points in the UMAP can

242 represent distinct metabolic environments across samples. The ordination plot indicated

243 that *C. difficile* appears to grow well in more than one metabolic environment, when

244 colonizing different individuals. Indeed, we saw that the predicted metabolic

245 environments occupied by *C. difficile* could vary within an individual over time (Fig. 2C).

246 For most patients in the Weingarden *et al.* cohort, there was a transition from

247 colonization-susceptibility pre-FMT to colonization-resistance post-FMT (Fig. 2A,C). We

248 next examined the different metabolic niches that *C. difficile* was able to exploit when

249 colonizing individuals across both rCDI-FMT cohorts, to better understand this

250 phenotypic plasticity.

251 *C. difficile is predicted to occupy three distinct metabolic niches within the human gut*

252 *microbiome*

253 To characterize *C. difficile* colonization-associated niches and identify the potential for

254 multiple metabolic strategies associated with its growth, we examined *C. difficile* import

255 fluxes with high variance ($\log_{10}$ flux variance >=4.5) across the rCDI-FMT cohorts. Using

256 this criteria the high variance metabolites identified across the 16S and metagenomic

257 data sets were highly consistent (Fig. 3). We identified 31 and 44 high-variance

258 metabolite fluxes from the 16S and metagenomics analyses, respectively. Of these, 27

259 were shared, 4 were unique to the Weingarden *et al.* (16S) analysis, and 17 were

260 unique to Iraniro *et al.* (metagenomics) analysis. Biclustering of the high variance import

261 flux data and an examination of how the apparent clusters associated with growth rates

262 revealed that *C. difficile* makes use of multiple metabolic strategies (Fig. 3). Three major

263 clusters were observed across patient samples and cohorts. We designate these three

264 clusters as "high growth", "moderate growth" and "no growth" (Fig. 3). The high growth

265 cluster in both cohorts included many of the pre-FMT samples and was characterized

266 by consistently high import fluxes for all the metabolites identified as most strongly

267 coupled to *C. difficile* growth across all models. The moderate growth cluster showed a

268 sparser metabolite consumption profile. For example, ornithine and fructose were

269 rapidly consumed in the high growth cluster, but showed almost no consumption in the

270 moderate growth cluster (Fig. 3). As expected, very few metabolites were consumed by

271 *C. difficile* above the zero-threshold of 10^{-6} in the no growth cluster (Fig. 3).

272 The metabolic strategies employed by *C. difficile* within the MCMMs showed

273 convergence with several observations from the literature. For example, we found that

274 metabolites known to promote growth of *C. difficile in vivo* (e.g., succinate, ornithine,

275 and trehalose) were preferentially utilized when available and were associated with

276 higher pathobiont growth rates [35–37]. In addition, the consumption of amino acids valine,

277 leucine, glycine, glutamate, glutamine, and proline were associated with higher *C.*

278 *difficile* growth rates in the MCMMs, implying that *C. difficile* employs Strickland

279 fermentation in one of its growth modes, which has been observed empirically [38].

280 Following up on these findings we examined how cooperative and competitive

281 interactions within MCMMs contributed to *C. difficile* colonization. To accomplish this,

282 we examined the import and export fluxes of metabolites associated with *C. difficile*

283 colonization (e.g., amino acids, ornithine, succinate, etc.). Genera that produced

284 metabolites consumed by *C. difficile* likely promote its growth, while those consuming *C.*

285 *difficile* growth-associated metabolites may be in direct competition. For ornithine and

286 succinate, we found that cooperative and competitive interactions are context-

287 dependent, varying across samples. The genus *Phocaeicola*, for instance, produces

288 ornithine in some samples, which is in turn consumed by *C. difficile*, while in other

289 contexts it consumes ornithine, competing with *C. difficile* (Fig. 4A). Meanwhile,

290 *Roseburia*, and *Faecalibacterium* compete with *C. difficile* for ornithine, but these

291 genera also produce succinate and leucine in some contexts, which *C. difficile*

292 consumes (Fig. 4A-B). Thus, the overall community context, rather than the presence or

293 absence of any single taxon, appears to be the most important factor in determining the

294 metabolic strategies used by *C. difficile* and can lead to emergent competitive or

295 cooperative interactions, which can either hinder or promote colonization.

296 Finally, we assessed whether or not compositional variation in the microbiome

297 alone could explain observed differences in MCMM-predicted *C. difficile* growth rates.

298   We found that compositional variation was a modest predictor of estimated *C. difficile*

299   growth rate (out-of-sample least absolute shrinkage and selection operator (LASSO)

300   regression R^2=0.37 for the Weingarden *et al.* rCDI-FMT cohort, see Methods).

301   Meanwhile, the import flux derived growth clusters (e.g., "high growth", "medium

302   growth", and "no growth" groups) could explain the vast majority of the variance in

303   predicted *C. difficile* growth rates (analysis of variance (ANOVA) R^2=0.94 for the CDI-

304   FMT cohort). However, this comparison between model-estimated growth rates and

305   growth clusters is a bit circular (i.e., both are derived from the model), so as mentioned

306   above, we also observed that MCMM-estimated *C. difficile* growth rates were able to

307   explain 50% of the variance in empirical *C. difficile* biomass measurements *in vitro* (Fig.

308   1C). In summary, these results suggest that compositional information alone is not

309   sufficient for consistently accurate engraftment predictions.

310

311   *Associations with C. difficile growth provide insights into the role of community context*

312   In order to assess the consistency of the *C. difficile* growth clusters, we leveraged four

313   independent *in vivo* 16S amplicon data sets, including the time series and the

314   Weingarden *et al.* rCDI-FMT studies presented above (Figs. 1-4), along with two large

315   cross-sectional cohorts (i.e., the American Gut and Arivale cohorts), covering a total of

316   14,862 individuals [24,30,39,40]. We evaluated growth and flux predictions generated across

317   all four data sets and found that *C. difficile* fell into the same three clusters that were

318   identified in the rCDI-FMT data sets, representing no growth, moderate growth, and

319   high growth (Fig. 5A).

320　　　　To further contextualize the metabolic niche of *C. difficile*, we integrated model

321　　outputs for all four 16S data sets. Specifically, we looked at import fluxes across all

322　　genera. Most genera formed unique clusters in the UMAP projection, suggesting that

323　　each genus had a single metabolic niche that was consistent across datasets (Fig. 5B).

324　　Within this community context, we found that *C. difficile* still fell into three very distinct

325　　clusters (Fig. 5B). Several genera that showed some of the strongest competitive and

326　　cooperative interactions with *C. difficile* across data sets, *Blautia*, *Faecalibacterium*,

327　　*Ruminococcus* and *Dorea*, clustered near one another in import flux space, with some

328　　overlap, indicating that these commensal taxa shared a similar metabolic niche (Fig.

329　　5B). Meanwhile, *Bacteroides* and *Phocaeicola*, two apparent *C. difficile* competitors,

330　　formed clusters largely separate from those formed by the other genera. Several of

331　　these taxa also occupied niche space near the *C. difficile* moderate growth cluster,

332　　supporting the potential for competitive interactions between *C. difficile* and these

333　　genera. The *C. difficile* high growth cluster was more separated from other commensal

334　　genera, suggesting that this growth mode is more specialized in the context of the

335　　surrounding community (Fig. 5B).

336　　　　　We next explored gut community diversity and predicted *C. difficile* growth rates

337　　across the four data sets. Specifically, we looked at Shannon diversity, which integrates

338　　species richness and evenness and is commonly used to quantify gut microbiome

339　　alpha-diversity. Lower Shannon diversity is commonly associated with disease states,

340　　like diarrhea, while higher diversity has generally been associated with diverse plant-

341　　based diets and overall better health [32]. However, constipated individuals generally

342　　have higher gut microbiome alpha-diversity as well, suggesting that there may be an

343   optimal range of alpha-diversity across healthy individuals [39,41]. Our initial analysis using

344   the rCDI-FMT cohort suggested a monotonically negative relationship between

345   predicted *C. difficile* growth rate and Shannon (Fig. 2B). However, the integrated data

346   sets, which spanned a wider range of diversity, showed a U-shaped relationship

347   between Shannon diversity and predicted *C. difficile* growth rate (Fig. 5C). Intermediate

348   levels of Shannon diversity were associated with the lowest predicted growth rates, on

349   average, with higher average growth at the upper and lower tails of the distribution (Fig.

350   5C). The relationship between Shannon diversity and predicted growth rate suggests

351   extremes in either direction on the diversity scale are, on average, more permissive to

352   *C. difficile* engraftment.

353   *Blood metabolites and clinical labs associated with MCMM-predicted C. difficile*
354   *colonization susceptibility*

355   We next sought to identify potential blood-based markers that were significantly

356   associated with MCMM-predicted *C. difficile* growth rates. Previous work has shown

357   that circulating blood metabolites can be leveraged to predict gut microbiome alpha-

358   diversity [39]. We identified several blood metabolites and clinical chemistries significantly

359   associated with MCMM-predicted *C. difficile* growth rates in the Arivale cohort, after

360   adjusting for common covariates (i.e., sex, age, and BMI) and correcting for multiple

361   tests (FDR *q* <0.05). These included two secondary bile acids, an unannotated

362   metabolite previously associated with the abundance of the family *Eggerthellacea*, and

363   several red blood cell-associated clinical chemistries (Fig. S1D)[42]. Unfortunately, while

364   significant, these blood-based markers, along with sex, age, and BMI, collectively

365   accounted for only ~5% of the variance in MCMM-predicted *C. difficile* growth rates.

366   Thus, it appears MCMM-based estimates of *C. difficile* engraftment, constrained by

367   fecal microbiome data, cannot be readily replaced with commonly measured clinical

368   chemistries or blood metabolites.

369

370   *MCMMs predicts engraftment heterogeneity of probiotic cocktail designed to treat rCDI*

371   As a final proof-of-concept for our modeling framework we simulated a probiotic

372   intervention using a previously validated probiotic cocktail designed to treat rCDI [22,43].

373   The probiotic, referred to as VE303, was composed of 8 commensal Clostridia strains

374   and shown to be effective at treating CDI in mice [43]. This probiotic was also shown to be

375   safe, well-tolerated, and effective in reducing rCDI incidence in humans [22]. An earlier

376   study in both mice and humans found that engraftment of VE303 strains was optimal

377   following antibiotic pretreatment [43]. With these facts in hand, we designed a simulated

378   intervention that mimicked the treatment found to be most effective by Dsouza *et al* [43].

379   We were only able to identify metabolic models for 6 of the 8 strains in VE303 in the

380   AGORA database [23]. We leveraged the Weingarden *et al.* rCDI-FMT dataset to test this

381   six-member probiotic cocktail, paired with *in silico* invasion by *C. difficile*. The probiotic

382   cocktail was introduced to patient samples, alongside 10% *C. difficile*, at a total relative

383   abundance of 50%, which was evenly distributed among the six strains. We also

384   simulated vancomycin treatment by reducing the abundance of *C. difficile* and all

385   commensal genera known to be impacted by vancomycin treatment [44] by 90%. We

386   found that a combined probiotic and antibiotic intervention most effectively suppressed

387 the growth of *C. difficile* in both the moderate and high *C. difficile* growth rate clusters

388 (Fig. 6A).

389   To better understand the mechanism of action of the probiotic cocktail, we

390 assessed the growth characteristics and the niche proximity of the probiotic strains in

391 relation to *C. difficile.* We found that suppression of *C. difficile* growth occurred more

392 readily when the average growth of the probiotic strains was high ($>10^{-4}$, Fig. 6B) and

393 when the average niche distance between the probiotic strains and *C. difficile* was low

394 (<25 import flux Euclidean distance, Fig. 6C). We also found that, relative to other

395 genera, several of the probiotics strains occupied niches closer to *C. difficile*, although

396 these niche distances could vary widely for each organism depending on their

397 community context (Fig. S2). Finally, we compared the import fluxes of the probiotic

398 strains and *C. difficile* for the metabolites identified as important for *C. difficile* growth

399 (Fig. 3). This analysis showed that, in addition to occupying niches similar to *C. difficile*,

400 several of the probiotic strains directly competed for metabolites important for *C. difficile*

401 growth, such as succinate, ornithine, and trehalose (Fig. 6D). Cumulatively these results

402 suggest that metabolic competition is the mechanism by which the probiotic cocktail

403 suppressed *C. difficile* growth, which is consistent with the emerging consensus in the

404 field [9,43,45]. Finally, we found that certain probiotic strains were more or less likely to

405 engraft in an individual (Fig. 6D), and that this engraftment/growth was generally

406 associated with *C. difficile* suppression (Fig. 6B), which indicates that MCMMs can be

407 leveraged to identify responders and non-responders prior to these kinds of probiotic

408 interventions.

409

410 **Discussion**

411

412 In this study, we provide a framework for predicting *C. difficile* engraftment risk in the

413 human gut microbiome using MCMMs. While we focus on *C. difficile*, due to its clinical

414 importance, this approach could be extended to other opportunistic bacterial pathogens,

415 probiotic organisms, or even entire communities, in the case of FMTs. We were able to

416 show how our approach predicts expected longitudinal and cross-sectional variation in

417 *C. difficile* colonization potential *in vitro* and *in vivo* (Figs. 1-2), using both shotgun

418 metagenomic and 16S amplicon sequencing data sets, and we provide insights into the

419 metabolic strategies leveraged by *C. difficile* in different ecological contexts (Figs. 3-4).

420 Our analysis not only recapitulates known metabolic associations with *C. difficile* growth

421 (e.g., consumption of trehalose, ornithine, and succinate; Fig. 3), it suggests additional

422 associations (e.g., importance of reduced sulfur compounds like cysteine, Stickland

423 fermentation reactants, and utilization of dietary sugars, like fructose; Fig. 3).

424 Additionally, we show that competition and cooperation with community members can

425 prevent or promote colonization of *C. difficile*, and that many of these associations are

426 highly context-dependent (Fig. 4).

427 Consistent with the idea that simple metrics of community structure and

428 composition alone are not effective predictors of colonization susceptibility, we found

429 that community compositional variation was a modest predictor of estimated *C. difficile*

430 growth rate and the relationship between alpha diversity and estimated *C. difficile*

431 growth rate was nonlinear (Fig. 5C). Not only did low diversity communities tend to be

432 more invasible, as might be expected due to putative non-saturation of the metabolic

433   niche space [9], but high diversity communities were also more prone to *C. difficile*

434   engraftment (Fig. 5C). In high diversity communities, successful colonization may be

435   due to changes in the interaction landscape, like a higher rate of niche-construction in

436   higher-diversity systems, which would be in line with the diversity-begets-diversity

437   hypothesis [46]. An intermediate range of alpha-diversity seems to be optimal for

438   mitigating *C. difficile* colonization potential (Figure 5C). Overall, these complex

439   mappings between community composition and pathobiont engraftment risk underscore

440   the necessity of systems-scale tools, like MCMMs, that are capable of synthesizing this

441   complexity.

442   Several genera were found to engage in cooperative and competitive interactions

443   with *C. difficile* across MCMMs. *Blautia*, *Faecalibacterium*, and *Dorea* were all shown to

444   benefit *C. difficile* through production of key metabolites that it consumes, like succinate

445   and leucine, but were also capable of competition for other metabolic resources (Fig. 4).

446   Meanwhile *Rumminococcus*, *Bacteroides* and *Phocaeicola* growth rates were often

447   negatively associated with *C. difficile* engraftment, competing for some of the same

448   metabolites that *C. difficile* consumed at high rates (Fig. 1D, 4). Contextualizing these

449   results through analysis of individual taxon import fluxes across studies we found that

450   *Blautia*, *Faecalibacterium*, *Phocaeicola, Rumminococcus,* and *Dorea* share similar

451   niches with one another, with *Phocaeicola being the most divergent from the others*. In

452   most cases, these niches did not overlap with *C. difficile*, but in a subset of individuals

453   all five occupied niches states overlapping or similar to *C. difficile* (Fig. 5B). Thus, while

454   we observed competition for some key metabolites, on a global scale, the majority of

455   the metabolic niche space used by *C. difficile* tends not to overlap with its apparent

456   competitors (Fig. 5B). These results highlight how flexible commensal gut bacteria are

457   in adapting their import fluxes to the communities in which they reside, which in turn

458   suggests why so many taxa are able to coexist. These results also suggest that

459   colonization resistance is an emergent feature of multi-species metabolic interactions,

460   and not strongly driven by any particular pairwise interaction.

461   In addition to developing a simulation framework to predict engraftment we

462   sought to identify blood-based clinical chemistries and blood metabolites that were

463   associated with MCMM-inferred *C. difficile* growth rate. We found three blood

464   metabolites that were independently associated with predicted *C. difficile* growth rates.

465   These included two secondary bile acids and an unannotated metabolite. One of the

466   secondary bile acids, isoursodeoxycholate, has previously been positively associated

467   with the abundance of *Bacteroides* [39], and was negatively associated with predicted *C.*

468   *difficile* growth rates. This result is in line with the apparent competition between

469   *Bacteroides* and *C. difficile* in our MCMMs. We also identified several clinical labs

470   negatively associated with predicted growth rate (Fig. S1D). However, together with

471   age, sex, and BMI, these features only accounted for ~5% of the variance in predicted

472   growth rates. Thus, while these features may be signatures for colonization

473   susceptibility in the blood, their clinical relevance is limited at this time.

474   Finally, we demonstrated that a probiotic intervention (i.e., 6/8 strains from

475   VE303), which recently showed positive efficacy results in a double-blinded, placebo-

476   controlled clinical trial for the treatment of rCDI [22], suppresses the growth of *C. difficile in*

477   *silico* in most people in the Weingarden *et al.* rCDI-FMT study (Fig. 6). We also showed

478   that the mechanism of action for this particular probiotic is likely competition for

479   metabolites essential for the growth of *C. difficile*, as many of the probiotic strains

480   occupy niches close to *C. difficile* and directly compete for these metabolites, such as

481   succinate and ornithine, in samples where growth suppression was observed (Fig. 6).

482   Furthermore, analysis of niche distances between *C. difficile* and other genera across

483   donors suggest selecting strains from *Blautia and Dorea* (e.g., including *B. producta and*

484   *D. longicatana,* from VE303), in addition to *Anaerostipes*, *Roseburia*, and

485   *Faecalibacterium*, could be leveraged to design better individual-specific probiotic

486   cocktails capable of suppressing *C. difficile* and rescuing VE303 non-responders (Figs.

487   S2 and 6B). These results illustrate how MCMMs may be powerful tools for assessing

488   the individual-specific efficacy of clinically-relevant probiotics, in addition to

489   understanding personalized pathobiont colonization susceptibility.

490   Future work should test this MCMM framework in the design of precision

491   interventions to prevent engraftment of *C. difficile* and other pathobionts, to design

492   precision probiotic interventions, and to improve the production of beneficial microbial

493   metabolites, like short-chain-fatty-acids, or reduce the production of undesired

494   metabolites, such as hydrogen sulfide or trimethylamine [47–49]. In summary, MCMMs

495   present a promising new path forward in engineering the ecological composition and

496   metabolic outputs of microbiota to prevent or treat disease.

497
498   **Methods**
499

500   *External Data Collection and Processing*
501
502   Data used in this study came from six sources. This included both cross-sectional and

503   time series 16S amplicon and shotgun metagenomics sequencing data from Hromada

504   *et al.*, David *et al.*, Weingarden, A. *et al.*, Iraniro *et al.,* the American Gut (McDonald, D.

505   *et al.*), and a former scientific wellness program run by Arivale, Inc.[18,24,26,27,30,40]. Publicly

506   available 16S amplicon sequence data and associated metadata were downloaded from

507   the sequence read archive (SRA). Additionally, de-identified 16S amplicon sequence

508   data, associated metadata, and paired blood-based clinical chemistries and

509   metabolomics were obtained for 2,687 research consenting individuals that were

510   formerly participants in the Arivale wellness program. Raw 16S amplicon sequence data

511   were processed using QIIME2 (v2020.11.1). In brief, the QIIME2 workflow consisted of

512   read demultiplexing using the command *qiime tools import*, and an associated manifest

513   table for each study describing read metadata followed by read quality assessment

514   using *qiime demux summarize*. Read quality assessment was used to determine

515   trimming parameters for subsequent denoising using the QIIME2 implementation of

516   DADA2 via the command, *qiime dada2 denoise-single* or *qiime dada2 denoise-paired*,

517   for single and paired reads respectively. The first 10 bases were trimmed from all reads

518   and reads were truncated to a length where median quality score was >20 (100-150

519   base pair for the data leveraged). Following denoising, data were reformatted into a

520   table format using the command *qiime metadata tabulate*, and representative sequence

521   taxonomy was inferred using a custom NCBI classifier with the command *qiime feature-*

522   *classifier classify-sklearn*. The NCBI classifier was trained using 16S 515f-806r V4

523   regions extracted from all available bacterial NCBI genomes. To train the classifier 515f-

524   806r regions were extracted from NCBI sequences using the command *qiime feature-*

525   *classifier extract-reads*, followed by the command *qiime feature-classifier fit-classifier-*

526   *naive-bayes* using the extracted V4 sequences and a table of known taxonomies. Pre-

527   processed data from Iraniro *et al.* were provided by the authors upon request and

528   included sample taxonomic relative abundance inferred using Metaphlan4 and

529   associated sample metadata. See Iraniro *et al.* for further details on their data

530   processing pipelines [18]. Sample taxonomic relative abundance and associated

531   metadata from Hromada *et al.* were obtained from their supplemental data. See

532   Hromada *et al.* for further details [27]. For source code and tables of processed data refer

533   to the Github repository listed below in Data and Source Code Availability.

534

535   *Model Construction and growth simulations*

536   To construct community level metabolic models sample specific taxonomic abundance

537   profiles inferred from 16S amplicon or shotgun metagenomic sequencing were

538   summarized at the genus or species level and mapped to genus or species level

539   metabolic models from the AGORA database (v1.03) using MICOM (v0.25.1). Taxa with

540   a relative abundance less than 0.1% were omitted from community models. An *in silico*

541   media previously designed to represent an average western diet was applied which

542   defined the bounds for metabolic imports by the model communities for all *in vivo*

543   analyses [20,23]. For the *in vitro* communities a custom medium was developed to reflect

544   the ABB medium used experimentally. The custom medium was developed using a

545   previously defined template designed for the representation of Luria-Bertani (LB)

546   medium, this contained metabolite concentration and fluxes for both yeast extract and

547   tryptone [29]. These details were used to convert the quantities of yeast extract and

548   peptone present in the experimental medium anaerobic basal broth (ABB) to fluxes. The

549   additional defined components were included with fluxes equal to the millimolar

550   concentration per hour. Finally, additional components were added to the constructed

551   medium so that all 13 species used in the *in vitro* communities could achieve a growth

552   rate of 10^{-3} or greater using a routine implemented in the MICOM function

553   *complete_db_medium.* Across all conditions, growth rates were inferred using

554   cooperative tradeoff flux balance analysis (ctFBA). In brief, this is a two-step

555   optimization scheme, where the first step finds the largest possible biomass production

556   rate for the full microbial community and the second step infers taxon-specific growth

557   rates and fluxes, while maintaining community growth within a fraction of the theoretical

558   maximum (i.e., the tradeoff parameter), thus balancing individual growth rates and the

559   community-wide growth rate[20]. For all models in the manuscript that leveraged 16S

560   amplicon data we used a tradeoff parameter of 0.8. A tradeoff parameter of 0.9 was

561   used for all the results derived from shotgun metagenomics data. These parameter

562   values were chosen by identifying the largest tradeoff value (corresponding to the

563   smallest deviation from maximal community biomass) that allowed >70% taxa to grow

564   (growth rate $>10^{-6}$). Import and export fluxes were estimated using parsimonious

565   enzyme usage FBA (pFBA) and a defined medium constructed to represent an average

566   European diet or ABB [20]. pFBA further constrained simulation results by requiring

567   genera to utilize the lowest overall flux through their networks to achieve maximal

568   growth [19]. For source code and tables of processed data refer to the Github repository

569   listed below in Data and Source Code Availability.

570

571   *Biomass estimation and growth rate normalization for in vitro communities*

572   To allow better comparison between model predictions and experimental results two

573   growth rate transformations were employed. These methods were implemented to

574   account for the fact that batch growth (in which cultures grow on a limited quantity of

575   nutrients) is not compatible with the steady state assumption employed by the modeling

576   framework (which assumes continuous supply of nutrients). Estimation of biomass was

577   leveraged to compare growth rates with experimentally measured end point optical

578   densities. This consisted of using a simple exponential growth model of the form $OD_f =$

579    $OD_i e^{kt}$, where OD_f is the final predicted biomass, OD_i is the initial biomass, k is the

580   MCMM inferred growth rate, and t is time in hours. The second transformation

581   employed was a growth rate normalization procedure, which accounted for differences

582   in community growth rates across samples. Growth rates were divided by the MCMM

583   inferred community growth rate, which was computed from the sum of relative growth

584   rates of each taxon in the system (e.g., the product of growth rate and relative

585   abundance), $k_{norm} = \frac{k}{\sum_i^n k_i r_i}$.

586   *Probiotic Intervention*

587   To model probiotic intervention a combination of strains previously shown to be effective

588   at suppressing the growth of *C. difficile* in mice were used [43]. Metabolic models for six of

589   the eight stains in the VE303 cocktail described by Dsouza *et al.* were identified in the

590   AGORA database and intervention was simulated by introducing them along with *C.*

591   *difficile* to individual samples. A total probiotic fraction of 50% was used, which was

592   evenly distributed among the six strains. This fraction was determined to be the most

593   effective at suppressing the growth of *C. difficile* growth *in silico* for the samples tested

594   (data not shown). Vancomycin treatment was simulated by reducing the abundance of

595   *C. difficile* and all genera known to be impacted by vancomycin by 90% [44]. Growth

596   simulations were performed as described above. For source code and tables of

597   processed data refer to the Github repository listed below in Data and Source Code

598   Availability.

599

600   *Statistical analysis*

601   Statistical analyses were performed using functions from the python scipy (v1.7.1),

602   seaborn (v0.11.2), sklearn-learn (v0.24.2), umap-learn (v0.5.1), and statsmodels

603   (v0.13.1) packages. Linear associations were performed using the *statsmodels* ordinary

604   least squares function *OLS*, and visualized using the *seaborn* function *regplot*. Least

605   absolute shrinkage and selection operator (LASSO) was performed using the *sklearn*

606   *Lasso* function and a training-test framework. Data were split into training and test sets

607   (70% of samples were randomly assigned to the training set) and model performance

608   was assessed across a range of regularization values. LASSO training and test set R^2

609   values were used to select the model with the best test set R^2 that did not overfit

610   training data (training R^2 >> test R^2). Analysis of variance (ANOVA) was performed

611   using the *statsmodels OLS* and *anova_lm* functions. UMAP dimensionality reduction

612   was performed using the *umap* function from the *umap* package and associated

613   methods with default parameters (i.e., *n_components=2*, *n_neighbors=15*,

614   *metric='euclidean'*, etc.). Biclustering was performed using the *seaborn* function

615   *clustermap* and the Ward clustering algorithm. Hexagonal binning and associated

616   histograms were generated using the *seaborn* function *jointplot.* Locally weighted

617 scatterplot smoothing (LOWESS) curves were generated using the *lowess* function from

618 *statsmodels* with default parameters. Additional statistical tests included the t-test and

619 Wilcox rank sum test implemented in scipy as *ttest_ind* and *wilcoxon* respectively. For

620 source code and tables of processed data refer to the Github repository listed below in

621 Data and Source Code Availability.

622

623 **Data and Source Code Availability**

624 Processed data tables and source code to reproduce the findings presented in this

625 manuscript can be found at https://github.com/Gibbons-Lab/cdiff_invasion. Raw 16S

626 amplicon sequence data from David *et al.*, Weingarden, A. *et al.*, the American Gut

627 (McDonald, D. *et al.*) can be downloaded using the sequence read archive (SRA)

628 accession numbers PRJEB6518, PRJEB19996, and PRJEB11419 respectively. Data

629 and metadata from Iraniro *et al.* were provided from the authors upon request. Data

630 from Hromada *et al.* were downloaded from the manuscript supplementary information.

631 Metadata were obtained from manuscript supplementary information, where available.

632 Qualified researchers can access the full Arivale deidentified dataset, including all raw

633 data, supporting the findings in this study for research purposes through signing a Data

634 Use Agreement (DUA). Inquiries to access the data can be made at data-

635 access@isbscience.org and will be responded to within 7 business days.

636
637

References:

1. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).

2. Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**, 1079–1094 (2015).

3. Schubert, A. M., Sinani, H. & Schloss, P. D. Antibiotic-Induced Alterations of the Murine Gut Microbiota and Subsequent Effects on Colonization Resistance against Clostridium difficile. *MBio* **6**, e00974 (2015).

4. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2016).

5. Bell, A. & Juge, N. Mucosal glycan degradation of the host by the gut microbiota. *Glycobiology* **31**, 691–696 (2021).

6. Baxter, N. T. *et al.* Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary Interventions with Three Fermentable Fibers. *MBio* **10**, (2019).

7. Gibbons, S. M. Defining Microbiome Health through a Host Lens. *mSystems* **4**, (2019).

8. Cheng, A. G. *et al.* Design, construction, and in vivo augmentation of a complex gut

666   microbiome. *Cell* **185**, 3617–3636.e19 (2022).

667   9.   Spragge, F. *et al.* Microbiome diversity protects against pathogens by nutrient

668   blocking. *Science* **382**, eadj3502 (2023).

669   10.   VanInsberghe, D. *et al.* Diarrhoeal events can trigger long-term Clostridium difficile

670   colonization with recurrent blooms. *Nat Microbiol* **5**, 642–650 (2020).

671   11.   Shields, K., Araujo-Castillo, R. V., Theethira, T. G., Alonso, C. D. & Kelly, C. P.

672   Recurrent Clostridium difficile infection: From colonization to cure. *Anaerobe* **34**,

673   59–73 (2015).

674   12.   Song, J. H. & Kim, Y. S. Recurrent Clostridium difficile Infection: Risk Factors,

675   Treatment, and Prevention. *Gut Liver* **13**, 16–24 (2019).

676   13.   Dubberke, E. R. & Olsen, M. A. Burden of Clostridium difficile on the healthcare

677   system. *Clin. Infect. Dis.* **55 Suppl 2**, S88–92 (2012).

678   14.   Banaei, N., Anikst, V. & Schroeder, L. F. Burden of Clostridium difficile infection in

679   the United States. *The New England journal of medicine* vol. 372 2368–2369

680   (2015).

681   15.   Crobach, M. J. T. *et al.* Understanding Clostridium difficile Colonization. *Clin.*

682   *Microbiol. Rev.* **31**, (2018).

683   16.   Baron, S. W. *et al.* Screening of carriers in an urban academic medical center:

684   Understanding implications of disease. *Infect. Control Hosp. Epidemiol.* **41**, 149–

685   153 (2020).

686   17.   Khoruts, A., Staley, C. & Sadowsky, M. J. Faecal microbiota transplantation for

687   Clostridioides difficile: mechanisms and pharmacology. *Nat. Rev. Gastroenterol.*

688   *Hepatol.* **18**, 67–80 (2021).

689   18. Ianiro, G. *et al.* Variability of strain engraftment and predictability of microbiome

690   composition after fecal microbiota transplantation across different diseases. *Nat.*

691   *Med.* **28**, 1913–1923 (2022).

692   19. Lewis, N. E. *et al.* Omic data from evolved E. coli are consistent with computed

693   optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390 (2010).

694   20. Diener, C., Gibbons, S. M. & Resendis-Antonio, O. MICOM: Metagenome-Scale

695   Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems* **5**,

696   (2020).

697   21. Diener, C. & Gibbons, S. M. More is Different: Metabolic Modeling of Diverse

698   Microbial Communities. *mSystems* e0127022 (2023).

699   22. Louie, T. *et al.* VE303, a Defined Bacterial Consortium, for Prevention of Recurrent

700   Clostridioides difficile Infection: A Randomized Clinical Trial. *JAMA* (2023)

701   doi:10.1001/jama.2023.4314.

702   23. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for

703   773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).

704   24. McDonald, D. *et al.* American Gut: an Open Platform for Citizen Science

705   Microbiome Research. *mSystems* **3**, (2018).

706   25. Wilmanski, T. *et al.* Gut microbiome pattern reflects healthy ageing and predicts

707   survival in humans. *Nat Metab* **3**, 274–286 (2021).

708   26. Weingarden, A. *et al.* Dynamic changes in short- and long-term bacterial

709   composition following fecal microbiota transplantation for recurrent Clostridium

710   difficile infection. *Microbiome* vol. 3 Preprint at https://doi.org/10.1186/s40168-015-

711   0070-0 (2015).

712   27. Hromada, S. *et al.* Negative interactions determine Clostridioides difficile growth in

713   synthetic human gut communities. *Mol. Syst. Biol.* **17**, e10355 (2021).

714   28. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data

715   science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

716   29. Marinos, G., Kaleta, C. & Waschina, S. Defining the nutritional input for genome-

717   scale metabolic models: A roadmap. *PLoS One* **15**, e0236890 (2020).

718   30. David, L. A. *et al.* Host lifestyle affects human microbiota on daily timescales.

719   *Genome Biol.* **15**, R89 (2014).

720   31. Lockwood, J. L., Cassey, P. & Blackburn, T. The role of propagule pressure in

721   explaining species invasions. *Trends Ecol. Evol.* **20**, 223–228 (2005).

722   32. Chang, J. Y. *et al.* Decreased Diversity of the Fecal Microbiome in Recurrent

723   Clostridium difficile—Associated Diarrhea. *J. Infect. Dis.* **197**, 435–438 (2008).

724   33. Pakpour, S. *et al.* Identifying predictive features of Clostridium difficile infection

725   recurrence before, during, and after primary antibiotic treatment. *Microbiome* **5**, 148

726   (2017).

727   34. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and

728   Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).

729   35. Ferreyra, J. A. *et al.* Gut microbiota-produced succinate promotes C. difficile

730   infection after antibiotic treatment or motility disturbance. *Cell Host Microbe* **16**,

731   770–777 (2014).

732   36. Pruss, K. M. *et al.* Oxidative ornithine metabolism supports non-inflammatory C.

733   difficile colonization. *Nat Metab* **4**, 19–28 (2022).

734   37. Buckley, A. M. *et al.* Trehalose-Induced Remodelling of the Human Microbiota

735   Affects Clostridioides difficile Infection Outcome in an In Vitro Colonic Model: A

736   Pilot Study. *Front. Cell. Infect. Microbiol.* **11**, 670935 (2021).

737 38. Bouillaut, L., Self, W. T. & Sonenshein, A. L. Proline-dependent regulation of

738   Clostridium difficile Stickland metabolism. *J. Bacteriol.* **195**, 844–854 (2013).

739 39. Wilmanski, T. *et al.* Blood metabolome predicts gut microbiome α-diversity in

740   humans. *Nat. Biotechnol.* **37**, 1217–1228 (2019).

741 40. Manor, O. *et al.* Health and disease markers correlate with gut microbiome

742   composition across thousands of people. *Nat. Commun.* **11**, 5206 (2020).

743 41. Tian, H., Chen, Q., Yang, B., Qin, H. & Li, N. Analysis of Gut Microbiome and

744   Metabolite Characteristics in Patients with Slow Transit Constipation. *Dig. Dis. Sci.*

745   **66**, 3026–3035 (2021).

746 42. Bar, N. *et al.* A reference map of potential determinants for the human serum

747   metabolome. *Nature* **588**, 135–140 (2020).

748 43. Dsouza, M. *et al.* Colonization of the live biotherapeutic product VE303 and

749   modulation of the microbiota and metabolites in healthy volunteers. *Cell Host*

750   *Microbe* **30**, 583–598.e8 (2022).

751 44. Nazzal, L. *et al.* Effect of Vancomycin on the Gut Microbiome and Plasma

752   Concentrations of Gut-Derived Uremic Solutes. *Kidney Int Rep* **6**, 2122–2133

753   (2021).

754 45. Horrocks, V., King, O. G., Yip, A. Y. G., Marques, I. M. & McDonald, J. A. K. Role of

755   the gut microbiota in nutrient competition and protection against intestinal pathogen

756   colonization. *Microbiology* **169**, (2023).

757 46. Madi, N., Vos, M., Murall, C. L., Legendre, P. & Shapiro, B. J. Does diversity beget

758   diversity in microbiomes? *Elife* **9**, (2020).

759   47. Liu, Y. & Dai, M. Trimethylamine N-Oxide Generated by the Gut Microbiota Is

760   Associated with Vascular Inflammation: New Insights into Atherosclerosis.

761   *Mediators Inflamm.* **2020**, 4634172 (2020).

762   48. Dalile, B., Van Oudenhove, L., Vervliet, B. & Verbeke, K. The role of short-chain

763   fatty acids in microbiota-gut-brain communication. *Nat. Rev. Gastroenterol.*

764   *Hepatol.* **16**, 461–478 (2019).

765   49. Murros, K. E. Hydrogen Sulfide Produced by Gut Bacteria May Induce Parkinson's

766   Disease. *Cells* **11**, (2022).

767
768   **Figures:**
769
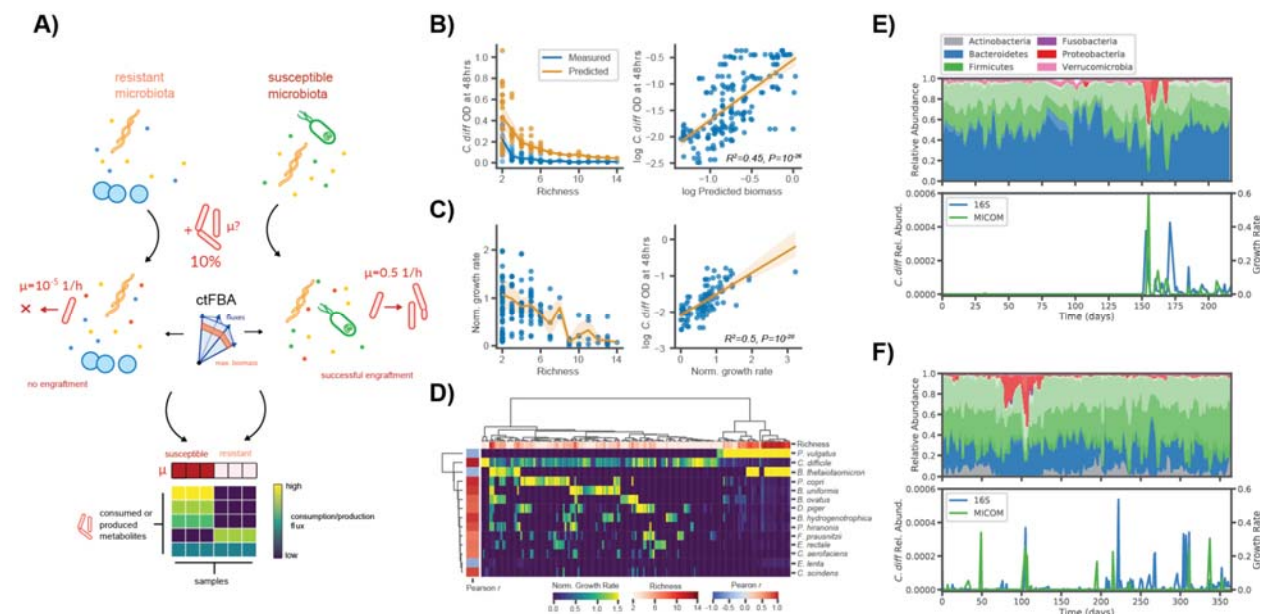
770
771   **Figure 1. *in silco* invasion assay accurately predicts *C. difficile* engraftment *in***

772   ***vitro* and *in vivo*.** (**A**) Schematic illustrating the *in silico* invasion assay workflow

773   leveraged in this study. Personalized microbial community-scale metabolic models

774   (MCMMs) are supplemented with 10% of a pan-genus *Clostridioides* model to simulate

775   an invasion event and ctFBA was used to predict *C. difficile* engraftment and metabolic

776   fluxes. (**B**) Measured *C. difficile* biomass at 48 hours from *Hromada et al.* and predicted

777   biomass, computed from MCMM-predicted *C. difficile* growth rate and an exponential

778   growth model, across a gradient of *in vitro* community richness. Mean trend lines and

779   standard deviation are displayed for both measured (blue) and predicted (yellow)

780   biomass. Relationship between $\log_{10}$ predicted biomass and $\log_{10}$ biomass. Ordinary

781   least squares fit and 95% confidence interval are displayed, as well as regression R^2

782   and *p*-value (**C**) Normalized *C. difficile* growth rate along a community richness gradient

783   and the relationship between normalized *C. difficile* growth rate and measured $\log_{10}$

784   biomass. Mean trend line and standard deviation are displayed for the relationship

785   between normalized *C. difficile* growth rate and community richness. Ordinary least

786   squares fit and 95% confidence interval are displayed, as well as regression R^2 and *p*-

787   value, for the relationship between normalized *C. difficile* growth rate and measured

788    $\log_{10}$ biomass . (**D**) Biclustered normalized growth rates for species across conditions.

789   Species normalized growth rate is displayed using blue-to-yellow heatmap coloring for

790   each sample. Sample community richness is displayed using white-to-red heatmap

791   coloring on the top row of the plot. Pearson correlation coefficients between *C. difficile*

792   normalized growth rate and the normalized growth rate of each other species are

793   displayed using blue-to-red heatmap coloring in the leftmost column of the plot.

794   Colorbars at the bottom of the blot indicate the scales for the various measures. (**E**)

795   Donor A time series taken from David *et al.* displaying daily fluctuations in microbiome

796   composition over a period of several months. Composition is displayed, colored by

797   phylum-level annotations (different shading indicates taxonomic families). At day 150,

798 Donor A experienced a diarrheal event and was subsequently colonized by *C. difficile*.

799 Estimates of *C. difficile* relative abundance from 16S sequencing and MICOM-predicted

800 *C. difficile* growth rates are displayed. (**F**) Time series from Donor B from the same

801 study, who was apparently colonized by *C. difficile* (at very low relative abundances,

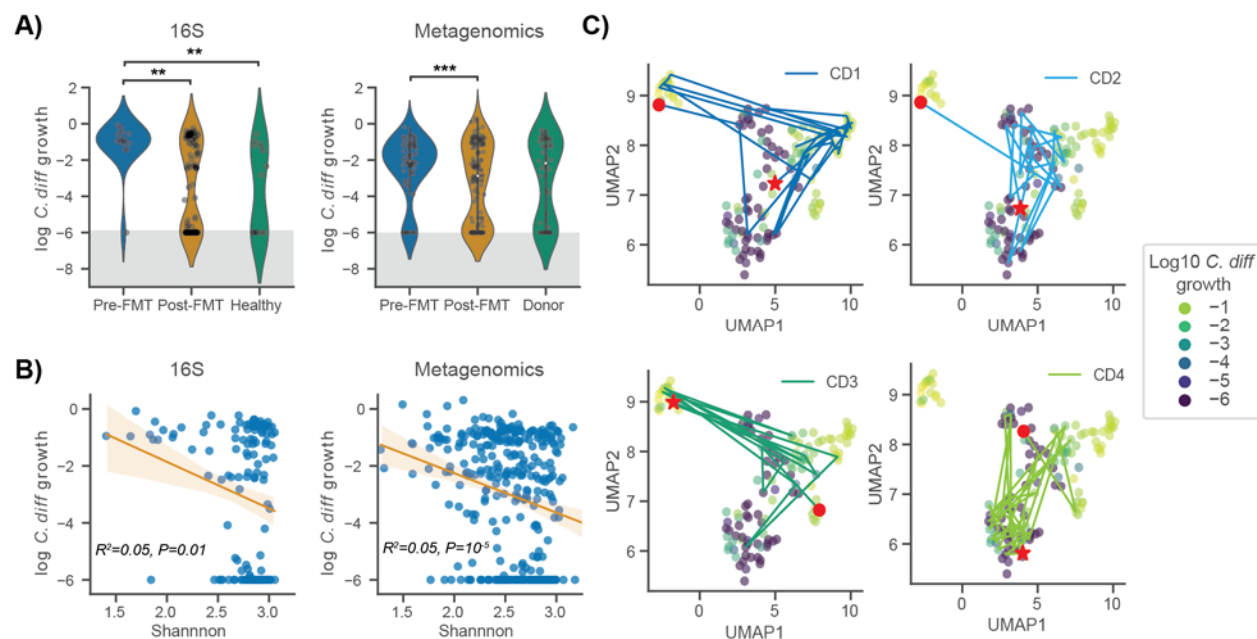802 near the limit of detection) throughout the sampling period.

803

804
805

806 **Figure 2. *C. difficile* growth rate predictions capture importance of community**

807 **context for patient recovery from CDI**. (**A**) Violin plots displaying predicted *C. difficile*

808 $\log_{10}$ growth rate distributions across patient disease status categories. Data from

809 Weingarden *et al.* and Ianiro *et al.* were leveraged to validate the approach using both

810 16S amplicon and metagenomic sequencing. Gray shading indicates the numerical

811 accuracy of the solver (values below 10^{-6} can not be distinguished from zero and are

812 considered negligible). Bars indicate comparisons for which differences were significant

813 using Welch's t-test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. (**B**) Relationship between

814 predicted *C. difficile* $\log_{10}$ growth rate and Shannon diversity across datasets. Ordinary

815   least squares fit and 95% confidence interval are displayed, as well as regression R^2

816   and *p*-value. (**C**) Two-dimensional representation of community import fluxes prior to *in*

817   *silico* invasion using UMAP colored by $\log_{10}$ growth rates of *C. difficile* following *in silico*

818   invasion. Patient trajectories are displayed, each with a red circle representing the

819   patient's starting point (prior to FMT), and a red star representing the patient's end point

820   (post recovery).

821

822
823
824
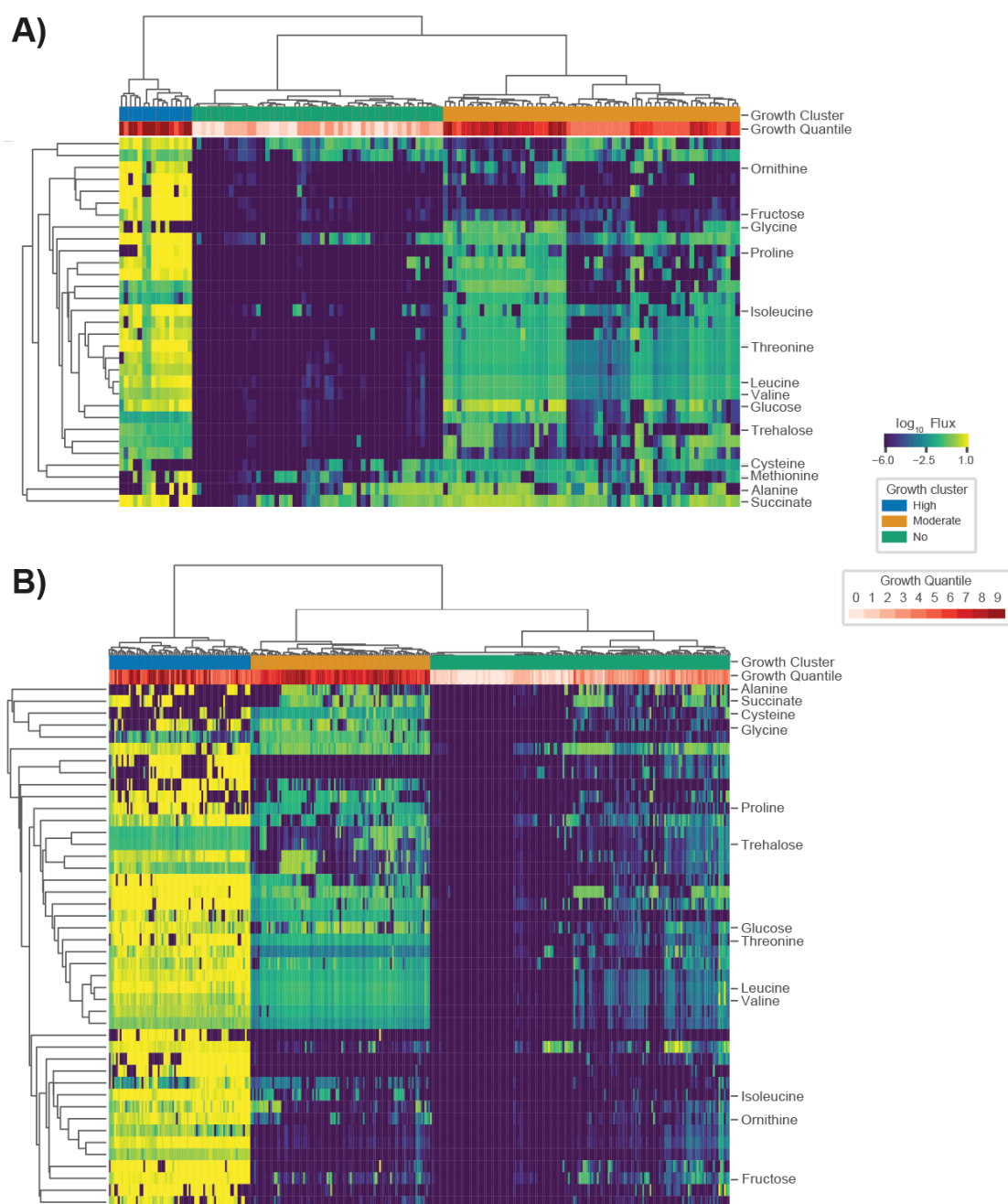
Figure 3. *C. difficile* occupies multiple metabolic niches across communities. (A)

Biclustered *C. difficile* $\log_{10}$ import fluxes from Weingarden *et al.*, where each row is the

import flux of a particular metabolite and each column is a patient sample. (**B**)

Biclustered *C. difficile* $\log_{10}$ import fluxes from Ianiro *et al.*, where each row is the import

flux of a particular metabolite and each column is a patient sample. In each heatmap,

830   imports for which the log variance across samples was >=4.5 are displayed. Fluxes

831   across samples are displayed using blue-to-yellow heatmap coloring. *C. difficile* $\log_{10}$

832   growth rate quantiles are displayed in white-to-red heatmap coloring for each patient

833   sample in the top row of each plot. Additionally, three coarse grain growth clusters are

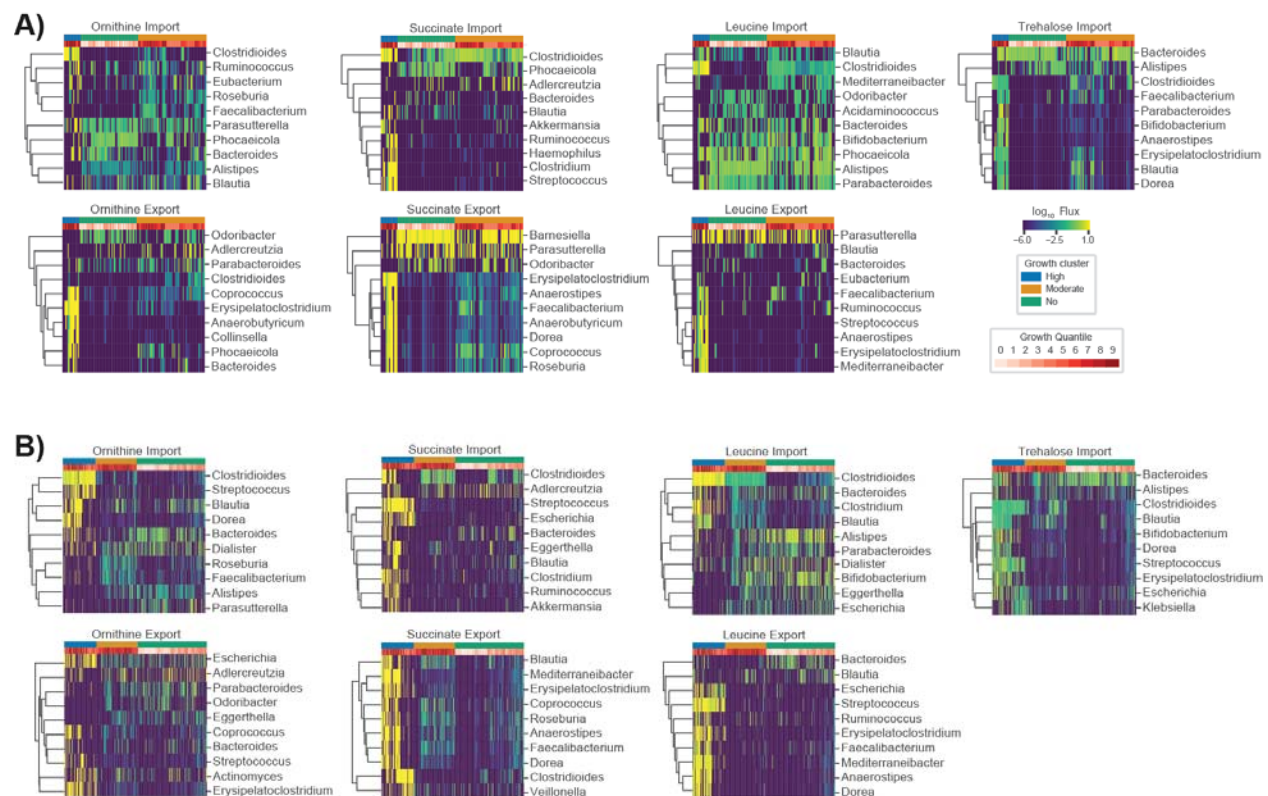834   noted. These growth clusters represent "high growth", "moderate growth", and "no

835   growth" phenotypes.

836

837   **Figure 4. Import and export fluxes of key metabolites across communities**

838   **highlight the context-dependency of key *C. difficile* competitors and cooperators.**

839   (**A**) Biclustered community import and export fluxes of specific metabolites associated

840   with *C. difficile* colonization from Weingarden *et al.*, where each row is a genus and

841   each column is a patient sample. (**B**) Biclustered community import and export fluxes of

842   specific metabolites associated with *C. difficile* colonization from Ianiro *et al.*, where

843   each row is a genus and each column is a patient sample. For each heatmap the top 10

844   genera with the highest mean import or export flux across samples are displayed.

845   Fluxes across samples are displayed using blue-to-yellow heatmap coloring. *C. difficile*

846     $\log_{10}$ growth rate quantiles are displayed in white-to-red heatmap coloring for each

847   patient sample in the top row of each plot. Additionally, three coarse grain growth

848   clusters are noted. These growth clusters represent "high growth", "moderate growth",
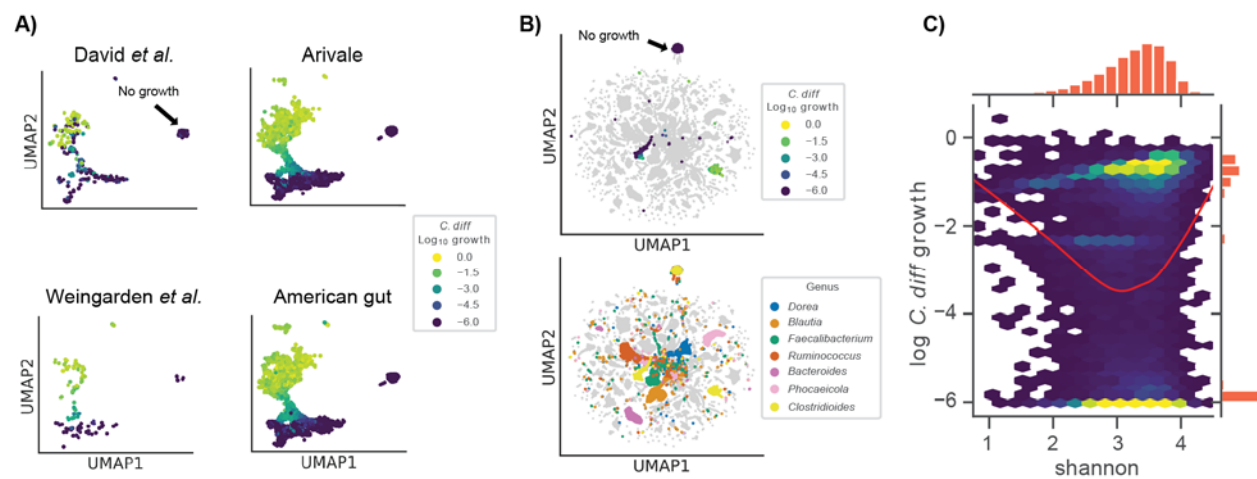
849   and "no growth" phenotypes.

850

851

852   **Figure 5. Growth niches across large four independent cohorts challenged *in***

853   ***silico* with *C. difficile*.** (**A**) Two dimensional representation of $\log_{10}$ *C. difficile* import

854   fluxes using UMAP across four independent data sets. Colors denote *C. difficile* growth

855   rate ranging from low (blue) to high (yellow). The position of the no growth cluster is

856   indicated. (**B**) Two dimensional representation of $\log_{10}$ genus import fluxes using UMAP

857   across all datasets. Top panel displays $\log_{10}$ *C. difficile* growth rate within the context of

858   all other genera. Bottom panel colors *C. difficile* along with six genera of interest: *Dorea*,

859   *Blautia*, *Faecalibacterium*, *Ruminococcus*, *Bacteroides*, and *Phocaeicola*. The position

860   of the no growth cluster is indicated. (**C**) Two-dimensional hexagonal binning of $\log_{10}$ *C.*

861   *difficile* growth rate and community alpha diversity (Shannon index). Red trend line

862   indicates a LOWESS fit to the $\log_{10}$ *C. difficile* growth rate and community Shannon
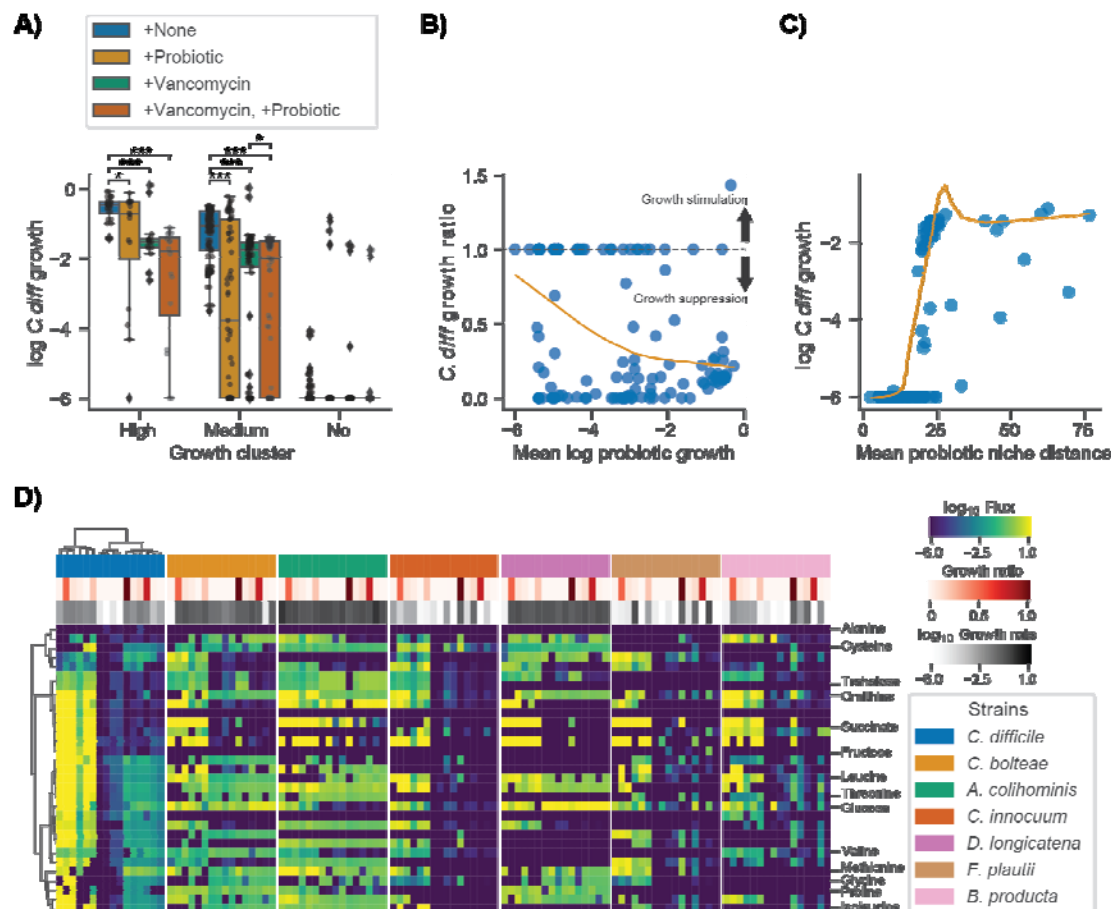
863   diversity data.

864

Figure 6. Simulated probiotic intervention effectively suppresses *C. difficile*

866   **growth *in silico*.** (**A**) Box plots displaying $\log_{10}$ *C. difficile* growth rate across growth

867   clusters and simulated interventions. Growth clusters are those identified by bi-

868   clustering of *C. difficile* import fluxes using the Weingarden data (Figure 3). Conditions

869   include +None (no intervention control), +Probiotic (introduction of 6 strain probiotic

870   previously identified as an effective treatment for CDI at a total relative abundance of

871   50% equally distributed across the strains), +Vancomycin (90% reduction of *C. difficile*

872 relative abundance as well as all genera known to be impacted by Vancomycin), and

873 +Vancomycin, +Probiotic (introduction of 6 strain probiotic in combination with simulated

874 vancomycin treatment). Bars indicate comparisons for which differences were

875 significant using the Wilcoxon signed-rank test. *, p <0.05; **, p <0.01; ***, p <0.001.

876 (**B**) Relationship between *C. difficile* growth ratio and mean $\log_{10}$ probiotic growth rate.

877 *C. difficile* growth ratio is the growth rate of samples in the +Vancomycin, +Probiotic

878 intervention relative to +None. Values below 1 indicate growth suppression by the

879 probiotic and values above 1 indicate growth stimulation. The dashed line marks the

880 value at which no effect is observed (1). Orange trend line indicates a LOWESS fit to

881 the *C. difficile* growth ratio and mean $\log_{10}$ probiotic growth rate. (**C**) Relationship

882 between $\log_{10}$ *C. difficile* growth rate and mean probiotic niche distance. Niche distance

883 was calculated using the euclidean distance of $\log_{10}$ import flux vectors of each

884 probiotic strain relative to *C. difficile* on a per sample basis. Orange trend line indicates

885 a LOWESS fit to the $\log_{10}$ *C. difficile* growth rate and mean probiotic niche distance. (**D**)

886 Biclustered $\log_{10}$ import fluxes for *C. difficile* and probiotic strains for samples previously

887 identified as "high growth", where each row is the import flux of a particular metabolite

888 and each column is a patient sample. Imports displayed are those previously identified

889 as important for *C. difficile*. Color bars indicate sample *C. difficile* growth ratio and strain

890 specific $\log_{10}$ growth rate. Ordering of samples and metabolites is the same across

891 heatmaps and based on biclustering of *C. difficile* data.
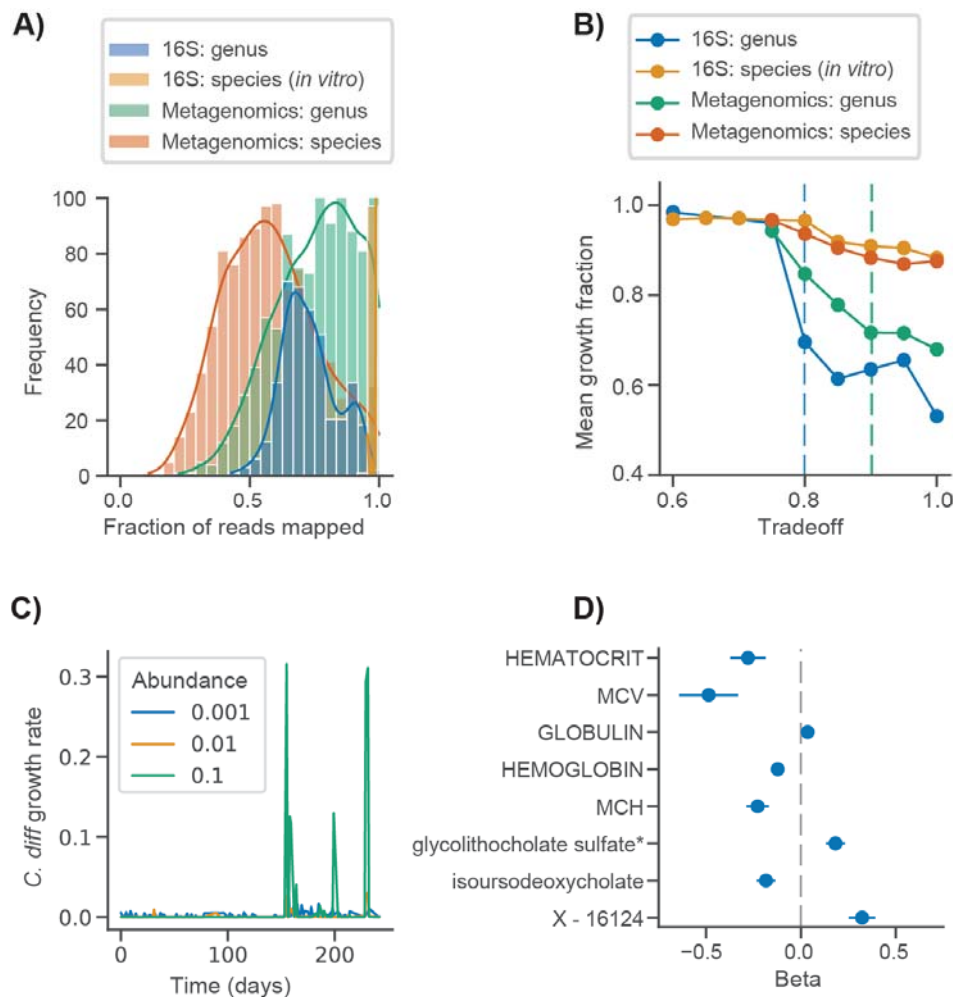
892

893

Figure S1. Development of *in silico C. difficile* invasion assay. (**A**) Histograms displaying the fraction of reads mapped at the genus and species levels for the David *et al.* and Hromade *et al.* 16S amplicon data and Ianiro *et al.* metagenomics data using an NCBI reference and metabolic models from the AGORA database. (**B**) Mean growth fraction across samples, datasets, and taxonomic mappings (e.g., fraction of taxa with estimated growth rate $>10^{-6}$) as a function of model tradeoff value. Dashed lines indicated the tradeoff values chosen for subsequent analyses. The blue dashed line indicates the value used for 16S data and green dashed line denotes the value used for metagenomics data. Tradeoff values were chosen such that mean growth fractions of ~0.7 were achieved using genus level mapping. (**C**) Relationship between *C. difficile*

905 invasion abundance and growth rate for one of the two David *et al.* time series. (**D**)

906 Association coefficients for estimated *C. difficile* log growth rate, blood metabolite

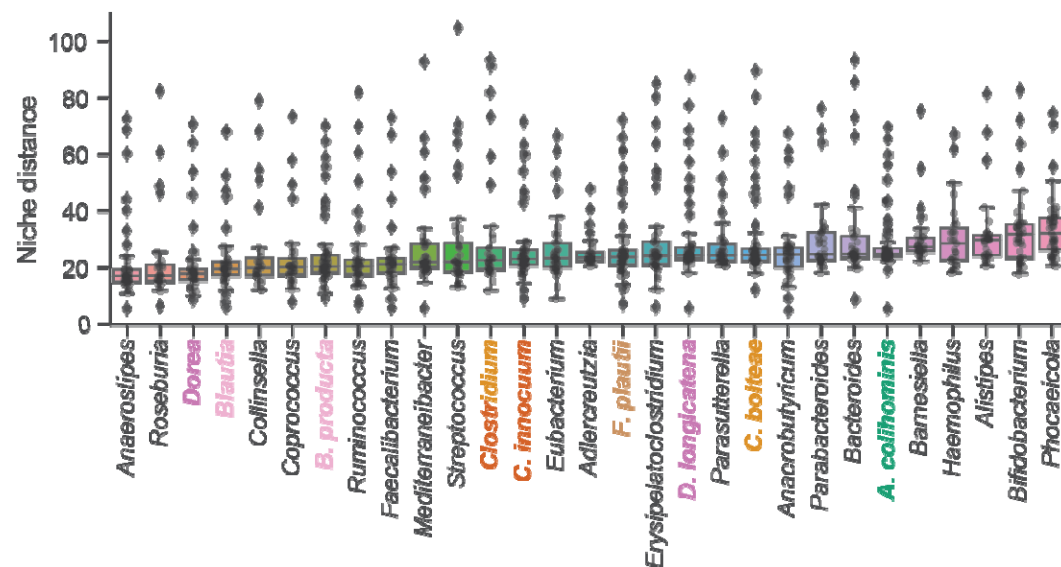907 concentrations, and clinical labs for the Arivale cohort.

908

909

Figure S2. Probiotic strains and associated genera have niche distances close to

C. difficile relative to unrelated genera. Niche distances of strains and genera,

represented as the euclidean distance between flux vectors, relative to *C. difficile*

across CDI-FMT cohort samples for which the *C. difficile* growth rate $>10^{-6}$. Genera and

strains are ordered by the median niche distance. Probiotic strains and associated

genera are colored, consistent with the legend in Fig. 5D. *C. bolteae* and *C. innocuum*

are both members of the genus *Clostridium*.