

1 **Title:** Genome-scale metabolic reconstruction analysis of *Clostridioides difficile* identifies conserved patterns of
2 virulence-related metabolic reprogramming
3

4 **Authors:**

5 Matthew L Jenior¹

6 Jhansi L Leslie²

7 Deborah A Powers¹

8 Elizabeth M Garrett⁶

9 Kimberly A Walker⁶

10 Mary E Dickenson¹

11 William A Petri Jr^{2,4,5}

12 Rita Tamayo⁶

13 Jason A Papin^{*1,2,3}

14 * denotes corresponding author

15

16 **Affiliations:**

17 1. Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA

18 2. Department of Medicine, Division of Infectious Diseases & International Health, University of Virginia,
19 Charlottesville, VA, USA

20 3. Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA, USA

21 4. Department of Microbiology, Immunology and Cancer Biology, University of Virginia Health System,
22 Charlottesville, Virginia, USA

23 5. Department of Pathology, University of Virginia Health System, Charlottesville, Virginia, USA.

24 6. Department of Microbiology & Immunology, University of North Carolina Chapel Hill School of Medicine,
25 NC, USA

26

27 **ABSTRACT**

3
28 *Clostridioides difficile* is a Gram-positive, sporulating anaerobe that has become the leading cause of
29 hospital-acquired infections. Over the previous decade, many studies have demonstrated the importance of
30 metabolism in numerous aspects of *C. difficile* biology from initial colonization to regulation of virulence factors.
31 Additionally, due to growing threats of antibiotic resistance and recurrent infection, targeting components of
32 metabolism presents a novel possible approach to combat this infection. In the past, genome-scale metabolic
33 network analysis of bacteria has enabled systematic investigation of the genetic and metabolic properties that
34 potentially contribute to downstream phenotypes as well as prediction of outcome from perturbations to these
35 pathways. These predictions ultimately create a platform for high-throughput identification and screening of
36 metabolic targets prior to laboratory testing. To accomplish these goals in *C. difficile*, we constructed highly-
37 curated genome-scale metabolic network reconstructions (GENREs) for a well-studied laboratory strain of the
38 pathogen (str. 630) as well as a more recently characterized hyper-virulent isolate (str. R20291). These
39 computational modeling platforms account for key components of *C. difficile* core metabolism and nutrient
40 acquisition systems to recapitulate metabolic behaviors within the complex milieu of the gut. Simulating the
41 impact of single-gene deletions resulted in accuracies of ~89.9% for both GENREs compared with transposon
42 mutant libraries. Further analysis of both strains also revealed significant correlations between *in silico* and
43 experimentally measured growth in carbon source utilization screens (*p*-values ≤ 0.002), with positive
44 predictive values of ~95.0%. Subsequently, we generated context-specific models by integrating
45 transcriptomic data from *C. difficile* grown *in vitro* or during *in vivo* infection. Simulations also predicted the
46 consistent inverse patterns of carbohydrate and amino acid catabolism that corresponded with differential
47 virulence factor expression measured experimentally. Collectively, our results indicate that GENRE-based
48 analyses of *C. difficile* are an effective means for gaining novel insight into metabolism as it relates to
49 pathogenesis and provides a platform for the identification of novel therapeutic targets.

50

51 **MAIN**

52 The nosocomial bacterial pathogen *Clostridioides* (formerly *Clostridium*) *difficile* causes a toxin-
53 mediated diarrheal illness and is now the leading cause of hospital-acquired infection in the United States ^{1,2}.
54 Susceptibility to *C. difficile* infection (CDI) is most frequently preceded by exposure to antibiotic therapy ³.
55 While these drugs are life-saving they also diminish the abundance of other bacteria in the microbiota, altering

the metabolic environment of the gut, and leaving it susceptible to colonization by *C. difficile*^{4–6}. Recently, it was established that *C. difficile* adapts transcription of distinct catabolic pathways to the unique conditions found in susceptible gut environments following different antibiotic pretreatments^{7,8}. These transcriptional shifts indicated that *C. difficile* must coordinate differential metabolic activity in order to effectively compete across dissimilar gut environments for successful infection. In spite of these differences, there are known core elements of *C. difficile* metabolism across different environments including carbohydrate and amino acid fermentation⁹. However, the relative utility of each metabolic strategy across given infections remains unknown. Furthermore, it is also understood that the availability of nutrients including fermentable monosaccharides and certain amino acids influences expression of virulence genes in *C. difficile*^{9,10}. Given these findings, along with the increased prevalence of antibiotic resistance and hyper-virulence among *C. difficile* isolates^{11,12}, novel therapeutic strategies are desperately needed and targeting or altering these central nodes of metabolism may be an effective means of targeted therapy without continued exposure to antibiotics.

Genome-scale metabolic network reconstructions (GENREs) are mechanistic frameworks and mathematical formalizations of metabolic reactions encoded in the genome of a target organism, which are subsequently constrained by known biological and physical parameters. GENREs can serve as a knowledge base for metabolic capability of a given organism, as well as a platform for functional simulation and prediction for the impact of genotype on many observable metabolic phenotypes. These tools have achieved success in directing genetic engineering efforts¹³ and accurately predicting auxotrophies and competition/cooperation between species for growth substrates^{14,15}. GENREs also create improved context for the interpretation of omics data¹⁶, and have provided powerful utility for identification of novel drug and gene targets accelerating downstream laboratory testing¹⁷. Leveraging these properties, several recent studies have found new possible metabolic targets for medically-relevant pathogens including *Klebsiella pneumoniae*, *Staphylococcus aureus*, and *Streptococcus mutans*^{17–19}. Taken together, these principles make GENRE-based analyses a strong platform for analysis of and target identification in *C. difficile* metabolism.

A few previous efforts have been made to create GENREs for well characterized strains for *C. difficile*, each with varied objectives and corresponding predictive qualities^{20–23}. Analysis of these GENREs reinforced the necessity for carefully constructed stoichiometry and flux constraints to ensure that downstream predictions have the highest accuracy. As understanding of genome annotation and metabolic functionality increases,

84 GENREs must be revisited or remade entirely to improve the quality of the resultant metabolic predictions. As
85 such, we began with the updated genome of the highly-characterized laboratory strain *C. difficile* str. 630²⁴,
86 first generating a *de novo* reconstruction followed by extensive literature-driven manual curation of catabolic
87 pathways, metabolite transport, and a biomass objective function. We proceeded to use this reconstruction as
88 a template to also create a curated GENRE for the more recently isolated hyper-virulent strain R20291²⁵.
89 Predictions from both GENREs were subsequently compared against published *in vitro* gene essentiality and
90 carbon utilization screens. These predictions indicated a high degree of agreement across experimental
91 datasets. To then assess the application of our GENREs for *in situ* metabolic prediction, we integrated
92 transcriptomic data collected from both *in vitro* and *in vivo* conditions into our models and assessed the
93 emergent metabolic activities. Analysis of context-specific pathogen metabolism revealed conserved patterns
94 of metabolism. Across states of increased virulence, both strains of *C. difficile* favored increased fermentation
95 of amino acids and decreased capacity for glycolysis. These trends agreed with published phenotypes^{10,26}, and
96 supported the advantage provided by GENREs for delineating complex metabolic networks and patterns of
97 gene expression into more tractable experimental targets. Additionally *in vivo* gene essentiality highlighted
98 specific aspects of nucleotide scavenging as critical for growth during infection and may provide preliminary
99 targets for future inhibitor discovery. Overall, high-quality GENREs can greatly augment the discovery of novel
100 therapeutics to treat CDI due to the connections between metabolic signals and colonization or virulence
101 induction in *C. difficile*. Finally, the current study lays the groundwork for systems-level analyses of CDI-
102 associated metabolism in the context of complex extracellular environments like the gut microbiome during
103 infection.

104

105 **RESULTS**106 **Current State of *C. difficile* Genome-scale Metabolic Modeling Efforts**

107 We began by collecting and assessing the quality of existing *C. difficile* GENREs. The primary focus of
108 curated *C. difficile* metabolic modeling efforts has been on the first fully sequenced strain of *C. difficile*, str. 630.
109 A high degree of additional genomic and phenotypic characterization was later performed for this isolate,
110 making it an ideal candidate for representative GENRE creation. The first reconstruction effort (iMLTC806cdf
111²⁰) and subsequent revision (icdf834^{20,21}), were followed by a recent *de novo* creation following updated

112 genome curation (iCN900²⁷)²⁸. Another GENRE was developed for str. 630Δerm (iHD992²²), a strain derived
113 from str. 630 by serial passage until erythromycin resistance was lost²⁹. Four additional *C. difficile* strain
114 GENREs were generated as a part of an effort to generate numerous new reconstructions for members of the
115 gut microbiota³⁰; these reconstructions received only semi-automated curation performed without *C. difficile*-
116 specific considerations.

117 To establish a baseline for the metabolic predictions possible with current *C. difficile* GENREs, we
118 selected common criteria with large impacts on the quality of subsequent predictions for model performance
119 (Fig. S1A). The first of these metrics is the level of consistency in the stoichiometric matrix³¹⁻³³, which reflects
120 proper conservation of mass and that no metabolites are incorrectly created or destroyed during simulations.
121 The next metric is a ratio for the quantity of metabolic reactions lacking gene-reaction rules to those
122 possessing associated genes³⁴, which may indicate an overall confidence in the annotation of the reactions.
123 These features reflect the importance of mass conservation and deliberate gene/reaction annotation which
124 each drive confidence in downstream metabolic predictions, omics data integration, and likelihood for
125 successful downstream experimentation. We found that each GENRE performed well in some categories, but
126 unique challenges were found in each which made comparing simulation results across models challenging.
127 For example, neither iMLTC806cdf nor iHD994 have any detectable gene annotations associated with the
128 reactions they contain. A high degree of stoichiometric matrix inconsistency was detected across icdf834,
129 iHD992, and iCN900; with iHD992 many intracellular metabolites were able to be generated without acquiring
130 necessary precursors from the environment. These findings reinforced the value of proper biochemical
131 constraints for GENREs to allow for improved fidelity to the target organism's *in situ* metabolism.

132 We went on to determine the cumulative MEMOTE quality score for each *C. difficile* GENRE (Fig. S1A).
133 MEMOTE is a recent series of model quality assessment guidelines, agreed upon by the research community,
134 and developed into a single platform to create an independent comparable quality metric across GENREs³⁵.
135 These percentages reflect a composite measurement of mass conservation, reaction constraint, and
136 standardized component annotation that are necessary for carrying out reliable simulations³⁴. The three oldest
137 *C. difficile* reconstructions each scored <50%; conversely the most recent GENRE (iCN900) received a 74%
138 cumulative MEMOTE score yet underperformed in the other metrics. Furthermore, the pre-curation draft *C.*
139 *difficile* GENREs generated for this study scored similarly (~40%) to those automatically curated AGORA

140 models (Fig. S1B). Our results from MEMOTE analysis indicated the current *C. difficile* GENREs do not meet
141 some of the recent established standards which is likely to reduce the accuracy of downstream metabolic
142 predictions.

143 Finally, we assessed key metabolic functionalities and established general principles of *C. difficile*
144 physiology within each of the existing GENREs. First, we compared imputed doubling times of each GENRE,
145 derived from the optimal biomass objective flux value simulated in rich media³⁶. While not strictly a
146 measurement of GENRE quality, this value may generally reflect the degree of functional predictions possible
147 with a given GENRE based on its deviation from measured values of ~29 minutes under similar conditions³⁷.
148 This analysis uncovered that most GENREs indicated doubling times relatively close to the experimental
149 measures, however iMLTC806cdf and iHD992 gave times under 5 minutes and iCN900 was well over 500
150 minutes (Fig. S1D). We also detected structural inconsistencies across several GENREs. For example, those
151 GENREs acquired from the AGORA database possessed several intracellular metabolic products not
152 adequately accounted for biologically (Table S1), as well as mitochondrial compartments despite being
153 bacteria. Additionally, several key *C. difficile* metabolic pathways either were incomplete or absent from the
154 curated models including multi-step Stickland fermentation, membrane-dependent ATP synthase, dipeptide
155 and aminoglycan utilization, and a variety of saccharide fermentation pathways³⁸. Overall, the existing *C.*
156 *difficile* GENREs possessed numerous mass imbalances and annotation inconsistencies, lacked key functional
157 capacities, and failed to phenotypically mimic *C. difficile* growth. These collective results motivated the
158 generation of a new reconstruction for our intended analyses.

159

160 ***C. difficile* Metabolic Network Scaffold Construction**

161 The existence of hypervirulent strains of *C. difficile* that have unique metabolism and virulence factors
162 highlights the importance of equipping future modeling efforts to study and identify novel targets within these
163 isolates. With this in mind, we focused on the most well-characterized hypervirulent isolate, str. R20291.
164 However, to maximize the utility of the bulk of published *C. difficile* metabolic research, we elected to generate
165 a reconstruction for the lab-adapted str. 630 in parallel. This focus afforded the ability to continuously cross-
166 reference curations between the models and to more readily identify emergent differences that are specifically
167 due to genomic content. We began the reconstruction process by accessing the re-annotated genome of str.

168 630²⁸ and the published str. R20291 genome²⁵, both available on the Pathosystems Resource Integration
169 Center database (PATRIC)³⁹. Following a recent protocol for creating high-quality genome-scale models⁴⁰,
170 and utilizing the ModelSEED framework and reaction database⁴¹, we generated gap-filled scaffold
171 reconstructions for both strains. Gap-filling refers to the automated process of identifying incomplete metabolic
172 pathways due to an apparent lack of genetic evidence that are also necessary for *in silico* growth, and
173 subsequent addition of the minimal functionality needed to achieve flux through these pathways⁴². The
174 resultant scaffolds were stripped of reactions that were added due to gap-filling in order to be most reflective of
175 original genomic content and partially reveal pathways in need of manual curation (Table S2). Additionally, to
176 focus the reconstructions on bioconversion of metabolites, we removed genes that encoded enzymes involved
177 in macromolecule synthesis (e.g. ribosomal genes). We subsequently performed complete translated proteome
178 alignment between str. 630 and str. R20291, resulting in 684 homologous metabolic gene products and 22 and
179 33 unique gene products, respectively (Table S3). Among the distinctive features were additional genes for
180 dipeptides import in str. 630 and glycogen import and utilization in str. R20291, which have both been linked to
181 modulated levels of virulence across strains of *C. difficile*^{43,44}. After resolving the dissimilarities between the
182 strains by incorporating corresponding metabolism to each reconstruction, we moved on to extensive manual
183 curation of both GENREs.

184

185 Metabolic Network Curation and Ensemble Gap-filling

186 Manual curation is required in order to ultimately produce high-quality GENREs and make meaningful
187 biological predictions⁴⁵. As such, we proceeded to manually incorporate 259 new reactions (with associated
188 genes and metabolites) and altered the conditions of an additional 312 reactions already present within each
189 GENRE prior to gap-filling (Table S2). Primary targets and considerations for the manual curation of the *C.*
190 *difficile* GENREs included:

191

- 192 • Anaerobic glycolysis, fragmented TCA-cycle, and known molecular oxygen detoxification^{38,46}
193 • Minimal media components and known auxotrophies^{47–49}
194 • Aminoglycan and dipeptide catabolism^{50–52}
195 • Many Stickland fermentation oxidative and reductive pathways (Table S2)^{37,53–62}

14

- 196 ● Carbohydrate fermentation and SCFA metabolism^{53,63–65}
- 197 ● Energy metabolite reversibility (e.g. ATP, GTP, FAD, etc.³²)
- 198 ● Periplasmic-associated H⁺ gradient and ATP synthase
- 199 ● Additional pathogenicity-associated metabolites (e.g. p-cresol⁵⁵ and ethanolamine⁶⁶)
- 200

201 Following the outlined manual additions, we created a customized biomass objective function with
202 certain elements tailored to each strain of *C. difficile*. Our biomass objective function formulation was initially
203 adapted from the well-curated GENRE of the close phylogenetic relative *Clostridium acetobutylicum*⁶⁷ with
204 additional considerations for tRNA synthesis and formation of cell wall macromolecules, including teichoic acid
205 and peptidoglycan (Table S2). Coefficients within the formulations of DNA replication, RNA replication, and
206 protein synthesis component reactions were adjusted by genomic nucleotide abundances and codon
207 frequencies in order to yield strain-specific biomass objective functions⁶⁸. To successfully simulate growth, we
208 next performed an ensemble-based pFBA gap-filling approach^{69,70}, utilizing the ModelSEED reaction bag
209 modified to focus on Gram-positive anaerobic bacterial metabolism (see Materials & Methods). We performed
210 gap-filling across six distinct and progressively more limited media conditions; complete medium, Brain-Heart
211 Infusion (BHI⁷¹), *C. difficile* Defined Medium +/- glucose (CDM⁴⁹), No Carbohydrate Minimal Medium (NCMM
212⁵), and Basal Defined Medium (BDM⁴⁷) (Table S2). With each step new reactions found across an ensemble
213 were collected and integrated into the draft reconstruction. A total of 68 new reactions allowed for robust
214 growth across all conditions.

215 Final steps of the curation process were focused on limiting the directionality of reactions known to be
216 irreversible, extensive balancing of the remaining incorrect reaction stoichiometries, and adding annotation
217 data for all network components. We repeated the assessments that were performed for the earlier
218 reconstructions and found that our GENREs had substantial improvements in all metrics including few, if any,
219 flux or mass inconsistencies and now each received a cumulative MEMOTE score of 86% (Fig. S1C). The new
220 reconstructions were designated iCdG698 (str. 630) and iCdR700 (str. R20291). For a precise recounting of
221 computational steps refer to Materials & Methods. We then set out to validate model behaviors against actual
222 experimental data.

224 **Gene essentiality results from new GENREs closely match experimental transposon screens**

225 A standard measurement of GENRE performance is the comparison of predicted essential genes for
226 growth *in silico* and those found to be essential experimentally through forward genetic screens⁷². This form of
227 analysis moves past strict network quality criteria and into biologically tractable predictions. Many *C. difficile*
228 strains have been historically difficult to manipulate genetically⁷³; however, methods were recently developed
229 and a large-scale transposon mutagenesis screen was published for str. R20291⁷⁴. As such, we first utilized
230 the proteomic alignment from the previous section to determine those genes in str. 630 that possessed
231 homologs within the str. R20291 dataset. We simulated single gene knockouts for all genes and evaluated for
232 >1% optimal biomass objective flux in BHI medium after growth simulation⁷⁵ for both iCdR700 (Fig. 1A) and
233 iCdG698 (Fig. 1B), cross-referencing the results with those in the published study. These comparisons
234 revealed overall accuracies of 89.1% and 88.9%, with negative-predictive values as high as 90.0% for
235 iCdR700 and 89.9% for iCdG698. These results demonstrated that our GENREs correctly predicted with high
236 accuracy the same genes determined to be essential for laboratory growth.

237

238 **Predicted growth substrate utilization profiles mirror *in vitro* screening results**

239 To assess if GENRE requirements reflected the components of minimal medium derived
240 experimentally, we identified the minimum subset of metabolites that our model required as an exogenous
241 supply for growth. Importantly, the specific metabolite composition of *C. difficile* minimal medium has been
242 defined across three separate laboratory studies^{47–49}. Through *in silico* limitation of extracellular metabolites to
243 only the experimentally determined requirements, followed by growth simulations with systematic omission of
244 each component individually, we were able to determine the impact of each component on achieving some
245 level of biomass flux (Fig. 1C). This analysis revealed that the majority of metabolites found to be essential
246 during growth simulation have also been shown experimentally to be required for *in vitro* growth. In
247 disagreement with two of the published studies, simulations indicated that neither iCdG698 (str. 630) nor
248 iCdR700 (str. R20291) is auxotrophic for methionine. However, the published formulation of BDM where
249 methionine is present found the amino acid to be largely growth-enhancing and not essential for small levels of
250 growth⁴⁸. Additionally, it has been demonstrated in the laboratory that *C. difficile* is able to harvest sufficient
251 bioavailable sulfur from excess cysteine instead of methionine^{49,76}, further supporting a non-essential status for

252 this metabolite. In a similar fashion, pantothenate (vitamin B5) only appears to enhance growth rate *in vitro* and
253 is not necessarily required to support slow growth rates. Finally, our results also indicated that iCdR700 was
254 not auxotrophic for isoleucine relative to iCdG698, and indeed contained additional genes coding for synthesis
255 of a precursor (3S)-3-methyl-2-oxopentanoate (*ilvC*, a ketol-acid reductoisomerase) which were not present in
256 its counterpart GENRE (Table S3). Interestingly, increases in isoleucine consumption are associated with
257 greater pathogenicity in some *C. difficile* strains⁷⁷, which may contribute to the hypervirulence of str. R20291.
258 In summary, the *in silico* minimal requirements for iCdG698 and iCdR700 closely mirrored experimental results
259 for both strains of *C. difficile* in addition to reconciling partially conflicting reports on experimentally-determined
260 auxotrophies.

261

262 **Metabolite-specific growth enhancement strongly correlates with *in vitro* results**

263 We next assessed additional carbon sources that impact the growth yield predictions for both GENREs.
264 Utilizing previously published results for both *C. difficile* strains in a Carbon Source Utilization Screen⁷⁸, we
265 simulated the degree to which each metabolite influenced growth yield in minimal medium. Importantly, *C.*
266 *difficile* is auxotrophic for specific amino acids (e.g. proline; Fig 1C) that it is also able to catabolize through
267 Stickland fermentation⁷⁹, so the diluting background medium must be supplemented with small concentrations
268 of these metabolites. As such, the values are reported as the ratio of the final optical density for growth with the
269 given metabolite versus low levels of growth observed in the background medium alone. Despite this
270 calculation not being a direct comparison of utilization capability as in traditional Biolog analyses⁸⁰, it provides
271 insight into an organism's metabolic preferences. We similarly calculated the influence of each metabolite on
272 the optimal biomass flux at quasi-steady state of each model provided with the same background media
273 conditions as the Biolog analysis (Fig. 2A). Across all of the 116 total metabolites that were in both the *in vitro*
274 screen as well as the *C. difficile* GENREs, we identified significant predictive correlations in the amount of
275 growth enhancement for iCdG698 (*p*-value < 0.001) and iCdR700 (*p*-value = 0.002) (Fig. 2B & 2C). This
276 relationship was even more pronounced for carbohydrates and amino acids, primary carbon sources for *C.*
277 *difficile* (Fig. S2). When these predictions were reduced to binary interpretations of either enhancement or non-
278 enhancement of growth, we found that iCdG698 predicted 92.8% and iCdR700 predicted 96.6% true-positive
279 enhancement calls (Fig. 2D). Importantly, this metric is the most valuable measure in this instance as it

20

280 indicates that each GENRE possesses the machinery for catabolizing a given metabolite. Collectively, these
281 data strongly indicated that both GENREs were well-suited for prediction of growth substrate utilization in either
282 strain of *C. difficile*.

283

284 **Context-specific metabolism reveals inverse metabolic patterns relating to virulence *in vitro***

285 Following GENRE validation, we sought to qualify the ability of each GENRE to predict *in situ* metabolic
286 phenotypes across diverse experimental settings. As previously stated, GENREs have provided powerful
287 platforms for the integration of transcriptomic data, creating greater context for the shifts observed between
288 conditions and capturing the potential influence of pathways not obviously connected⁸¹. With this application in
289 mind, we chose to generate context-specific models for both *in vitro* and *in vivo* experimental conditions
290 characterized with RNA-Seq analysis utilizing a recently published unsupervised transcriptomic data
291 integration method⁸². Briefly, this approach calculates the most cost-efficient usage of the metabolic network in
292 order to achieve growth given the pathway investments indicated by the transcriptomic data. This process is in
293 line with the concept that natural selection generally selects against wasteful production of cellular machinery
294 and affords the ability to make much more fine-scale predictions of metabolic changes that *C. difficile*
295 undergoes as it activates pathogenicity. The resultant patterns also reveal central elements within context-
296 specific metabolism that could lead to targeted strategies for intentional downregulation of virulence factors
297 through metabolic circuitry.

298 A recent study determined that phase variation, a reversible mechanism employed by many bacterial
299 pathogens to generate phenotypic heterogeneity and maximize overall fitness of the population, also occurs in
300 *C. difficile* str. R20291 and influences virulence expression⁸³. One aspect of this phase variation manifests as
301 a rough or smooth-edged colony morphology on solid agar; the morphologies can be propagated via
302 subculture and are associated with distinct motility behaviors and altered virulence⁸⁴. The colony morphology
303 variants are generated through the phase variable (on/off) expression of the *cmrRST* genes. With this in mind,
304 we sequenced transcriptomes from experimentally grown rough and smooth phase variants of *C. difficile* str.
305 R20291 grown on solid BHI rich medium for 48 hours. Utilizing these data, we generated context-specific
306 versions of iCdR700 in simulated rich media conditions. It has been previously shown that mutation of *cmr*-
307 family genes does not significantly alter growth rate *in vitro*⁸⁴. Growth simulation results predicted no

308 significant difference in optimal biomass flux values between phase variants (Fig. 3A), which agrees with
309 previously published experimental growth rate measurements for *C. difficile*³⁷. We then calculated essential
310 genes in each variant model similar to the earlier analysis which identified 81 core genes essential in both
311 contexts (Table S4), another 13 genes essential to growth for both variants, and 5 genes that were
312 conditionally essential between the morphologies in BHI rich medium (Fig. 3C). The conditionally essential
313 gene set restricted to the smooth variant included an N-acetylglucosamine PTS system as well as pyruvate
314 kinase, which mediates the last step of glycolysis and a bulk of the ATP generation. Notably, at the
315 transcriptional level, reads mapped to pyruvate kinase were detected at nearly identical levels between the
316 rough and smooth isolates (Table S4). These results indicate that glycolytic enzymes may be more active in
317 the smooth colony variants. The essentiality of N-acetylglucosamine transport in the context-specific model for
318 the smooth phase was of interest as this variant has been previously shown to generate biofilms⁸⁴, in which N-
319 acetylglucosamine is often a component⁸⁵. We found that predicted exchange efflux of N-acetylglucosamine in
320 the smooth variant was significantly greater than in rough (Fig. S3C). Conversely, in the rough context-specific
321 model were multiple essential genes involved in Stickland fermentation (Fig. 3B). As with the pyruvate kinase
322 gene, similar levels of transcription for these genes were also observed between smooth and rough variants
323 (Table S4). These data were indicative of a potential trade-off between glycolysis and amino acid (Stickland)
324 fermentation between smooth and rough phases respectively. In addition to genes that were critical for growth,
325 we also identified those that were only required to achieve high growth yields in each context. This gene set
326 included additional carbohydrate transporters in the smooth variant and multiple amino acid transporters in the
327 rough variant (Table S4), further supporting differential utilization of glycolysis and Stickland fermentation
328 across phases with highly dissimilar flux distributions of core metabolic pathways (Fig. S3), in spite of largely
329 similar optimal growth rates (Fig. 3A).

330 The trends for the opposing metabolic strategies were reinforced when we compared sampled flux
331 distributions for the associated exchange reactions for the most common substrates of each respective
332 pathway, glucose and proline. We found not only that the model predicted that glucose was imported in the
333 smooth variant, but that this functionality was entirely inactive in the rough-associated model (Fig. 3C).
334 Alternatively, proline was utilized significantly more in the rough variant-specific model (Fig. 3D), and unlike
335 glucose import could not be entirely pruned from the opposing model as *C. difficile* is a proline auxotroph. It

336 has been previously reported that this relationship between colony morphology phase variant and metabolism
337 may occur in *C. difficile*⁸⁶, and our collective results from contextualized iCdR700 analysis support discordant
338 utilization of glycolysis or Stickland fermentation that may relate to phase variation. Based on these data, we
339 hypothesized that access to easily catabolized carbohydrates influences colony morphology due to phase
340 variation in *C. difficile*. To test this hypothesis, single colonies of either rough or smooth, grown anaerobically
341 for 48 hours on BHIS agar (Fig. S4A), were subcultured onto BDM (Materials & Methods) agar plates both with
342 and without 2 mg/ml glucose (Fig. 3E & S4B). Following anaerobic incubation for 48 hours we found that rough
343 variants maintained their morphology across both media, with the rough phenotype even exacerbated on the
344 minimal medium. However, while the smooth variant largely maintained its colony morphology upon subculture
345 onto BDM + glucose, the colonies became much more analogous to their rough counterparts when glucose
346 was absent. Further subculture of each altered morphology from minimal media back onto rich BHI medium
347 also appeared to support consistent switching between the respective morphologies (Fig. S4C). Our data
348 suggest that the absence of glucose provided a fitness advantage for variants that preferentially use Stickland
349 metabolism, selecting for the rough variant. Furthermore, these results are consistent with the hypothesis that
350 carbohydrates availability impacts phase variation in *C. difficile*, influencing the virulence-associated metabolic
351 state and that environmental stress due to limited nutrients may be a key factor in driving the shift between
352 phases.

353

354 **Predicted metabolism during infection also supports differential strategies relating to altered virulence**

355 Given laboratory media conditions (as used in the results described above) are much more easily
356 defined, we also wanted to examine GENRE performance and prediction quality under more complex *in vivo*
357 infection conditions. Another previously published study assessed the differential transcriptional activity of *C.*
358 *difficile* str. 630 in the gut during infection in a mouse model pretreated with either streptomycin or clindamycin
359 to induce sensitivity to colonization. These distinct treatments have different impacts on the structure of the gut
360 microbiota⁸⁷ and allow for identical levels of pathogen colonization and vegetative cell load in the cecum.
361 However, these different treatments result in highly dissimilar levels of sporulation (another phenotype linked to
362 *C. difficile* virulence) where streptomycin is associated with undetectable spore CFUs and clindamycin with
363 significantly higher levels⁸⁸. The authors of this study also detected no significant difference in toxin activity

364 between the groups. These experiments included paired, untargeted metabolomic analysis of intestinal content
365 to correlate the transcriptional activity of metabolic pathways with changes in the abundance of their respective
366 substrates and byproducts following infection. This analysis was performed for each antibiotic with both mock-
367 infected and *C. difficile*-colonized groups to extract the specific impact of the infection on the gut metabolome,
368 making this dataset extremely valuable for our purposes. Similar to the previous analysis, we overlaid these
369 data onto our GENRE of str. 630 (iCdG698) and compared predicted doubling times, which were calculated
370 from biomass objective flux in the sampled context-specific flux distributions (Fig. 4A). This comparison
371 revealed a significantly faster growth rate in the slower sporulation context (*p*-value << 0.001), reflecting a
372 potential focus on continued growth instead of spore formation and egress possibly due to preferred
373 environmental conditions. To then quantify differential use of core metabolism, we compared the activity of
374 those reactions conserved between conditions. We accomplished this analysis through unsupervised machine
375 learning (Non-Metric Multidimensional Scaling) of Bray-Curtis dissimilarity for sampled flux distributions of all
376 shared reactions (Fig. 4B). In agreement with the previous findings that *C. difficile* is able to adapt to distinct
377 growth substrates ⁸⁸, we found a significant difference (*p*-value = 0.001) between the activity of core
378 metabolism between high and low sporulation states. Additionally, within-group dissimilarities indicated that
379 much more variation was found within the low sporulation group, potentially indicating that conditions favoring
380 increased sporulation also support a lower diversity of potential metabolic strategies.

381 To support the unsupervised findings we implemented a supervised machine learning approach where
382 we identified those reactions which most readily delineate flux distributions from low and high spore context-
383 specific models, and reported the importance of each reaction to the overall classification success (Fig. 4C).
384 The most prominent signals highlighted by this approach were differences in the catabolism of the host-derived
385 mucus-associated aminoglycans N-acetylmannosamine, N-acetylneuraminate, and N-acetylglucosamine which
386 have been shown to be readily fermented by *C. difficile* and play a role in determining virulence factor
387 expression ^{26,89}. Additionally, multiple nucleoside phosphatase reactions which both contribute to maintenance
388 of intracellular phosphorylated guanosine which has also been associated with determining virulence
389 phenotype expression ^{90,91}. Taken together, these results support that environmental conditions that favor
390 increased glycolytic activity in *C. difficile* are inversely associated with virulence expression which agrees with
391 previous reports for the control of glucose over toxin expression ⁹².

392 We next cross-referenced exchange reactions that were differentially active across the high sporulation
393 and low sporulation context-specific models (Fig. 4D), and compared changes in the concentration of
394 associated metabolites from a paired untargeted metabolomics screen (Fig. 4E). This analysis predicted
395 multiple Stickland fermentation substrates to be utilized at similar rates across both contexts. We found that
396 proline was imported at higher rates in low spore-associated simulations (Table 4C; Table S5). This amino acid
397 was also detected in significantly higher concentrations only in mock infection, supporting consumption by *C.*
398 *difficile*⁷. These data agreed with findings from the previous section that amino acid catabolism may be
399 associated with higher expression of certain virulence factors, despite previous reports that extracellular proline
400 concentrations inversely correlated with expression of *C. difficile* toxin *in vitro*⁹. Leucine was also predicted to
401 be imported at higher rates in this context, and its associated Stickland byproduct isovalerate was predicted to
402 be produced only in the high spore model (Table S5). This trend agreed with *in vivo* metabolomic
403 measurements where isovalerate concentrations were significantly increased only in the context of higher
404 spores (Fig 4E). Conversely, another known Stickland substrate tyrosine was predicted to be utilized more in
405 the low spore-associated model (Fig. 4D). The byproduct of tyrosine fermentation, p-cresol, was also predicted
406 to be secreted more readily in the context of lower sporulation (Table S5). This largely *C. difficile*-specific
407 metabolic byproduct may be associated with its pathogenicity^{86,93}, and reinforces a potential link between
408 virulence expression. Collectively these results further support that while Stickland fermentation is a core
409 metabolic strategy in *C. difficile*, this pathway is differentially utilized under conditions that favor altered
410 virulence factor expression.

411 We also identified N-acetylneuraminate (NEu5Ac) as highly utilized in the lower sporulation context, a
412 host-derived glycolysis substrate that *C. difficile* readily uses as a carbon source for growth⁷. This
413 consumption was supported in the metabolomics screen where concentrations of this metabolite were
414 significantly decreased following infection only in the lower spore condition (Fig. 4F). In agreement, this
415 analysis also predicted that frequent byproducts of carbohydrate fermentation, acetate and formate, were more
416 abundant in the lower sporulation context (Fig. 4D). Alternatively, both N-acetylglucosamine (GlcNAc) and N-
417 acetylmannosamine (ManNAc) were predicted to be secreted at much larger rates from the low spore context-
418 specific model (Fig. 4D, S5A, & S5B). Interestingly, these metabolites are integral components of biofilms⁹⁴,
419 and *C. difficile* has been previously shown to generate these structures under certain circumstances⁸⁵. We

420 then found that during infection concentrations of D-glucose significantly increase (Fig. SC), potentially
421 indicating a lack of consumption by *C. difficile*. These combined results may indicate that increased reliance on
422 glycolysis may be associated with reduced sporulation but increased biofilm formation, supporting a complex
423 metabolic regulation of distinct aspects of *C. difficile* virulence.

424 To then examine the utility of the str. R20291 GENRE for identifying potential gene targets that may be
425 exploited to inhibit metabolism of the pathogen *in vivo*, we performed a similar *in silico* gene essentiality screen
426 as in the preceding section. We subsequently cross-referenced our results to limit our focus to those genes
427 that are only essential *in vivo* and shared across high and low sporulation-favoring conditions. This analysis
428 uncovered 35 genes that are essential only during infection (Table S5). Among the genes highlighted were
429 many components of nucleotide metabolism including pyrimidine synthesis regulator PyrR and adenylate
430 kinase. These genes are highly expressed during infection and inhibition of specific enzymes within this
431 pathway has been shown to downregulate toxin production^{6,95}. Furthermore, proline racemase, which is an
432 important part of Stickland fermentation in *C. difficile* and has been previously linked to virulence expression *in*
433 *vitro*⁹⁶, was also essential in both infection conditions. Alternatively, when we identified those genes that were
434 discordantly essential between the conditions we found that additional genes in the higher sporulation context
435 related to Stickland fermentation of glycine and proline; including glycine reductase and pyrroline-5-carboxylate
436 reductase (Table S5). These results further highlight the relationship between Stickland fermentation and
437 increased *C. difficile* sporulation. Additionally, these findings support that the GENREs were effective
438 mechanisms for identifying targetable metabolic components in *C. difficile* to limit colonization or pathogenicity.
439

440 DISCUSSION

441 The control for much of *C. difficile*'s physiology and pathogenicity is subject to a coalescence of
442 metabolic signals from both inside and outside of the cell. Historically, *C. difficile* research has suffered from a
443 shortage of molecular tools and high-quality predictive models for highlighting new potential therapies. Over
444 the previous decade, GENREs have become powerful tools for connecting genotype with phenotype, and
445 provided platforms for defining novel metabolic targets in biotechnology and improving interpretability of high-
446 dimensional omics data. These factors make GENRE-based analyses extremely promising for directing and
447 accelerating identification of possible therapeutic targets as well as a deeper understanding of the connections

448 between *C. difficile* virulence and metabolism. Furthermore, as much of bacterial pathogenicity is now being
449 attributed to shifts in metabolism the analyses described here may provide large benefits to the identification of
450 possible treatment targets in *C. difficile* and other recalcitrant pathogens⁹⁷. In the current study, we develop
451 and validate two highly-curated genome-scale metabolic network reconstructions for a well-described
452 laboratory strain (str. 630) in addition to a more recently characterized hyper-virulent strain (str. R20291) of *C.*
453 *difficile*. Both iCdG698 (str. 630) and iCdR700 (str. R20291) draw from numerous molecular and metabolic
454 studies of *C. difficile* and Clostridial metabolism in order to accurately incorporate a large array of metabolic
455 subsystems known to be present across strains of the pathogen. We further improved the quality of the models
456 through careful curation of core metabolic strategies, including amino acid and carbohydrate fermentation, to
457 ensure growth in all major defined growth conditions for *C. difficile*.

458 After the curation process was complete, we found a high degree of agreement between model
459 predictions and validating experimental datasets. Both iCdG698 and iCdR700 indicate that the respective
460 strains are able to catabolize amino acids as the sole carbon source through Stickland fermentation and
461 require only those metabolites present in the experimentally determined minimal media to achieve growth.
462 Additionally, close correlations of *in silico* predictions with both gene essentiality and carbon source utilization
463 screens supported that the GENREs accurately recapitulate *C. difficile* physiology and reconcile some previous
464 inconsistencies in *C. difficile* metabolism literature. Following contextualization using *in situ* transcriptomic data,
465 both GENREs were also able to demonstrate established complex metabolic phenotypes for both laboratory
466 and infection conditions. These analyses collectively indicated a shift away from glycolytic metabolism, and
467 toward amino acid fermentation, during periods of increased pathogenicity. These findings could lay the
468 groundwork for novel approaches to curbing the expression of virulence factors by influencing environmental
469 conditions to favor certain forms of metabolism over others. *In vivo* context-specific gene essentiality also
470 predicted proline racemase to be critical for growth during infection, yet it was previously found to be
471 dispensable in an animal model using a forward genetic screen⁹⁶. While this result may indicate necessary
472 future curation, it may also be attributable to the specific conditions of that infection and may vary across
473 distinct host gut environments, leading to possible implications in personalized medicine.

474 While the majority of validation data did agree with GENRE predictions, several areas of possible
475 expansion and curation are present in both GENREs. First, the scope of total genes included in iCdG698 and

476 iCdR700 may be more limited than previous reconstructions; however, we elected to focus on those gene sets
477 where the greatest amount of evidence and annotation data could be found to maximize confidence in
478 functionality included here. Future efforts could be directed at increasing the genomic coverage each GENRE
479 contains. Concordantly, both GENREs consistently underpredict the impact of some metabolite groups,
480 primarily nucleotides and carboxylic acids (Fig. S2), which could be due to the absent annotation of the
481 relevant cellular machinery. Furthermore, more complex regulatory networks ultimately determine final
482 expression of virulence factors and these may be needed additions in the future to truly understand the
483 interplay of metabolism and pathogenicity in *C. difficile*. In spite of these potential shortcomings, both iCdG698
484 and iCdR700 produced highly accurate metabolic predictions for their respective strains, and are strong
485 candidate platforms for directing future studies of *C. difficile* metabolic pathways. Additionally, the
486 contextualized growth simulation results indicated an inverse relationship between glycolysis and Stickland
487 fermentation with respect to expression of pathogenicity. Our results indicated that fermentation of specific
488 amino acids may be more associated with increased expression of *C. difficile* virulence factors. These changes
489 also seem to be predicated on a degree of environmental nutrient stress as the switch in phase was only
490 induced across formulations of minimal medium.

491 Systems-biology approaches have enabled the assessment of fine-scale changes to metabolism of
492 single species within complex environments that may have downstream implications on health and disease.
493 Overall, the combined *in vitro*- and *in vivo*-based results demonstrated that our GENREs are effective
494 platforms for gleaning additional understanding from omics datasets, outside of the standard analyses. Both
495 GENREs were able to accurately predict complex metabolic phenotypes when provided context-specific omic
496 data, and ultimately underscores the metabolic plasticity of *C. difficile*. The reciprocal utilization of glycolysis
497 and amino acid fermentation indeed support regimes of distinct metabolic programming associated with *C.*
498 *difficile* pathogenicity. With this in mind, finding core metabolic properties in *C. difficile* strains may be key in
499 identifying potential probiotic competitor strains or even molecular inhibitors of metabolic components. The
500 current study is an example of the strength that systems-level analyses have in contributing to more rapid
501 advancements in biological understanding, and in the future the metabolic network reconstructions presented
502 here are well-suited to accelerate research efforts toward the discovery of more targeted therapies.

504 **MATERIALS & METHODS**505 C. difficile GENRE Construction

506 We utilized PATRIC reference genomes from *Clostridioides difficile* str. 630 and *Clostridioides difficile*
507 str. R20291 as initial reconstruction templates for the automated ModelSEED pipeline ^{39,98,99}. The automated
508 ModelSEED draft reconstruction was converted utilizing the Mackinac pipeline (<https://github.com/mmunday42/mackinac>) into a form more compatible with the COBRA toolbox ¹⁰⁰. Upon removal of GENRE components
509 lacking genetic evidence (i.e. gap-filled), extensive manual curation was performed in accordance with best
510 practices agreed upon by the community ¹⁰¹. We subsequently performed ensemble gap-filling as previously
511 described, utilizing a stoichiometrically consistent anaerobic, Gram-positive ModelSEED universal reaction
512 collection curated for this purpose and available alongside code associated with this study. Next, we corrected
513 reaction inconsistencies and incorrect physiological properties (e.g. ensured free water diffusion across
514 compartments). Final transport reactions were then validated with TransportDB ¹⁰². All formulas are mass and
515 charged balanced at an assumed pH of 7.0 using the ModelSEED database in order to maintain a consistent
516 and supported namespace to augment GENRE interpretability and future curation efforts. We then collected
517 annotation data for all model components (genes, reactions, and metabolites) from SEED ^{101,103}, KEGG ¹⁰⁴,
518 PATRIC, RefSeq ¹⁰⁵, EMBL ¹⁰⁶, and BiGG ¹⁰⁷ databases and integrated it into the annotation field dictionary
519 now supported in the most recent SBML version ¹⁰⁸. Complete MEMOTE quality reports for both *C. difficile*
520 GENREs are also available in the GitHub repository associated with this study, and full pipelines for model
521 generation are explicitly outlined in Jupyter notebooks hosted there as well. Download of either iCdG698 (str.
522 630) or iCdR700 (str. R20291) is possible from the studies' Github or the Papin lab website
523 (<https://bme.virginia.edu/csbl/Downloads1.html>).
525

526 Growth simulations, flux-based analyses, and GENRE quality assessment

527 All modeling analyses were carried out using the COBRA toolbox implemented in python ¹⁰⁹. The
528 techniques utilized included: flux-balance analysis, flux-variability analysis ¹¹⁰, gapsplit flux-sampler ¹¹¹, and
529 minimal_medium on exhaustive search settings. GENRE quality assessment tools were also developed in
530 python and are fully available in the project GitHub repository. MEMOTE quality reports were generated using
531 the web-based implementation found at <https://memote.io/>.

532

533 C. difficile str. R20291 in vitro growth and microscopy

534 C. difficile str. R20291 growth was maintained in an anaerobic environment of 85% N₂, 5% CO₂, and
535 10% H₂. The strain was grown on BHI-agar (37 g/L Bacto brain heart infusion, 1.5% agar) medium at 37 °C for
536 48 hours to obtain isolated colonies. Rough and smooth colonies were chosen for propagation on BHI-agar to
537 ensure colony morphology maintenance ⁸⁴. Basal Defined Medium (BDM) was formulated as previously
538 published ⁴⁷ with the addition of 1.5% agar for plates, and incubated for 48 hours at 37 °C to generate isolated
539 colonies. Microscopy images were taken on an EVOS XL Core Cell Imaging System at 4x magnification.

540

541 RNA isolation, and transcriptome sequencing

542 For RNA isolation, rough and smooth isolates were subcultured in BHIS broth (37 g/L Bacto brain heart
543 infusion, 5 g/L yeast extract) overnight (16-18 h) at 37 °C, then 5 µL of the cultures were spotted on BHIS agar
544 (1.5% agar). After 24 h, the growth was collected and suspended in 1:1 ethanol:acetone for storage at -20 °C
545 until subsequent RNA isolation. Cells stored in ethanol:acetone were pelleted by centrifugation and washed in
546 TE (10 mM Tris, 1 mM EDTA, pH 7.6) buffer. Cell pellets were suspended in 1 mL Trisure reagent. Silica-glass
547 beads (0.1 mm) were added and cells were disrupted using bead beating (3800 rpm) for 1.5 minutes. Nucleic
548 acids were extracted using chloroform, purified by precipitation in isopropanol followed by washing the cold
549 70% ethanol, and suspended in nuclease-free water. Samples were submitted to Genewiz, LLC (South
550 Plainfield, NJ, USA) for quality control analysis, DNA removal, library preparation, and sequencing. RNA
551 sample quantification was done using a Qubit 2.0 fluorometer (Life Technologies), and RNA quality was
552 assessed with a 4200 TapeStation (Agilent Technologies). The Ribo Zero rRNA Removal Kit (Illumina) was
553 used to deplete rRNA from the samples. RNA sequencing library preparation was done using the NEBNext
554 Ultra RNA Library Prep Kit for Illumina (NEB) according to the manufacturer's protocol. Sequencing libraries
555 were checked using the Qubit 2.0 Fluorometer. The libraries were multiplexed for clustering on one lane of the
556 Illumina HiSeq flow cell. The samples were sequenced using a 2 x 150 Paired End configuration on an Illumina
557 HiSeq 2500 instrument. Image analyses and base calling were done using the HiSeq Control Software. The
558 resulting raw sequence data files (.bcl) were converted to the FASTQ format and de-multiplexed with bcl2fastq
559 2.17 software (Illumina). One mismatch was permitted for index sequence identification. Data were analyzed

560 using CLC Genomics Workbench v. 20 (Qiagen). Reads were mapped to the *C. difficile* R20291 genome
561 (FN545816.1) using the software's default scoring penalties for mismatch, deletion, and insertion differences.
562 All samples yielded over 22 million total reads, with over 20 million mapped to the reference (> 93% of total
563 reads, and > 90% reads in pairs). Transcript reads for each gene were normalized to the total number of reads
564 and gene length (expressed as reads per kilobase of transcript per million mapped reads [RPKM]). Raw and
565 processed sequence files are available at the NCBI GEO database under (Accession number Pending)

566

567 Genomic and transcriptomic data processing

568 Alignment of *C. difficile* str. 630 and str. R20291 peptide sequences was performed using bidirectional
569 BLASTp. RNA-Seq reads were first quality-trimmed with Sickle with a cutoff $\geq Q30$ ¹¹². Mapping
570 curated reads to the respective *C. difficile* genome was then performed with Bowtie2¹¹³. MarkDuplicates then
571 removed optica/PCR duplicates (broadinstitute.github.io/picard/), and mappings were converted to idxstats
572 format using SAMtools¹¹⁴. Abundances were then normalized to both read and target lengths. Transcriptomic
573 integration and context-specific model generation were performed with RIPTiDe and maxfit_contextualize() on
574 the default settings⁸².

575

576 Statistical Methods

577 All statistical analysis was performed in R v3.2.0. Non-metric multidimensional scaling of Bray-Curtis
578 dissimilarity and perMANOVA analyses accomplished using the vegan R package¹¹⁵. Significant differences
579 for single reaction flux distributions and metabolite concentrations were determined by Wilcoxon signed-rank
580 test.. Supervised machine-learning was accomplished with the implementation of AUC-Random Forest also in
581 R¹¹⁶. All code associated with this study is available in the study-associated GitHub repository.

582

583 Data availability

584 Genomic and proteomic data for the strains *Clostridioides difficile* str. 630 (PATRIC ref. 272563.8) and
585 *Clostridioides difficile* str. R20291 (PATRIC ref. 645463.3) was downloaded from the PATRIC database⁹⁸.
586 Transcriptomic data was downloaded in raw FASTQ format from the NCBI Sequence Read Archive
587 (PRJNA415307 and PRJNA354635) and Gene Expression Omnibus (GSE158225) . Github repository for this

588 study, with all programmatic code and GENREs described here, can be found at:
589 https://github.com/mjenior/Jenior_CdifficileGENRE_2020.

590

591 **Author Contributions**

592 MLJ - Conceptualization. Data generation and analysis. Drafting manuscript.

593 JLL - Conceptualization. Data generation. Editing manuscript.

594 DAP - Data analysis. Editing manuscript.

595 EMG - Conceptualization. Data generation. Editing manuscript.

596 KAW - Data generation. Editing manuscript.

597 MED - Data analysis. Editing manuscript.

598 WAP - Supervision. Editing manuscript.

599 RT - Supervision. Data generation. Editing manuscript.

600 JP - Funding acquisition. Supervision. Drafting and editing manuscript.

601

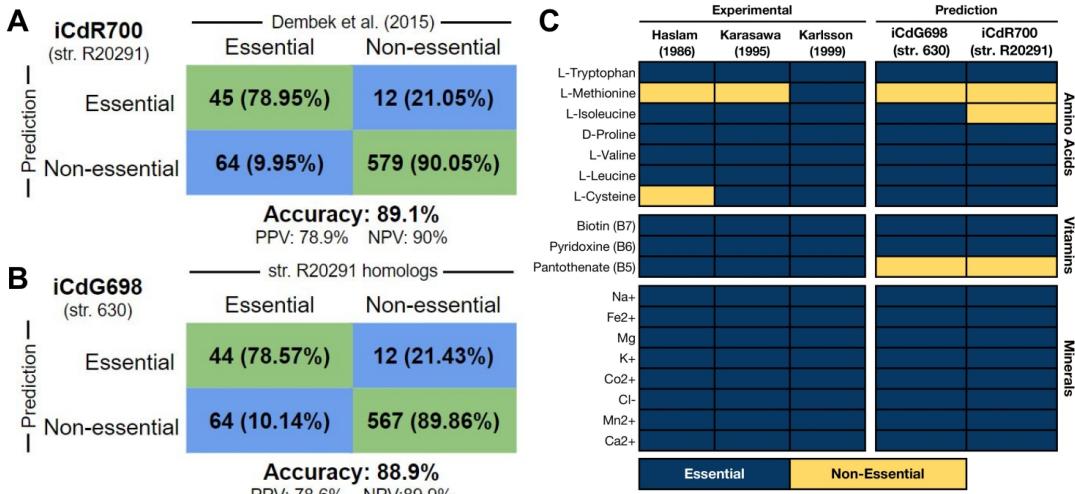
602 **Acknowledgements**

603 The authors would like to acknowledge Bonnie Dougherty, Laura Dunphy, and Dawson Payne for their
604 input and feedback on modeling parameters and biomass objective function formatting. The authors have
605 declared that no competing interests exist. This work was supported by funding from The U.S. National
606 Institutes of Health awards R01AT010253 to JP and R01AI143638 to RT. The funding agency had no role in
607 study design, data collection/analysis, or preparation of the manuscript.

608

609 **FIGURE & TABLE LEGENDS**

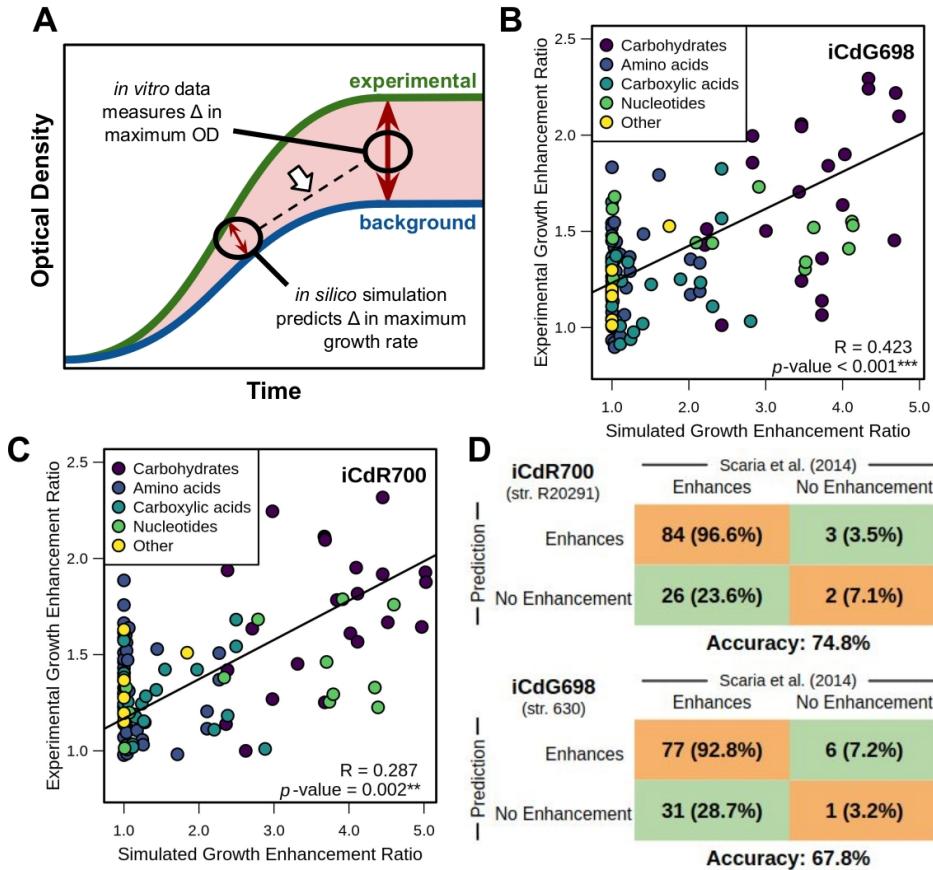
610



611

612 **Figure 1) Gene and minimal growth substrate essentiality predictions closely match experimental**
 613 **results. (A)** Gene essentiality results for iCdR700 (str. R20291) using the transposon mutagenesis screen
 614 results published in Dembek et al. 2015, and **(B)** gene essentiality for iCdG698 (str. 630) utilizing homologs
 615 from the genome of str. R20291. **(C)** Computationally determined minimum growth substrates for both
 616 GENREs compared with experimentally determined *C. difficile* minimal medium components across three
 617 previously published studies. Essentiality was determined for those genes and metabolites that when absent
 618 resulted in a yield of <1.0% of optimal biomass flux during growth simulation utilizing components of the
 619 corresponding media used experimentally. Additional trace minerals required for bacterial growth can be found
 620 in Table S2.

621

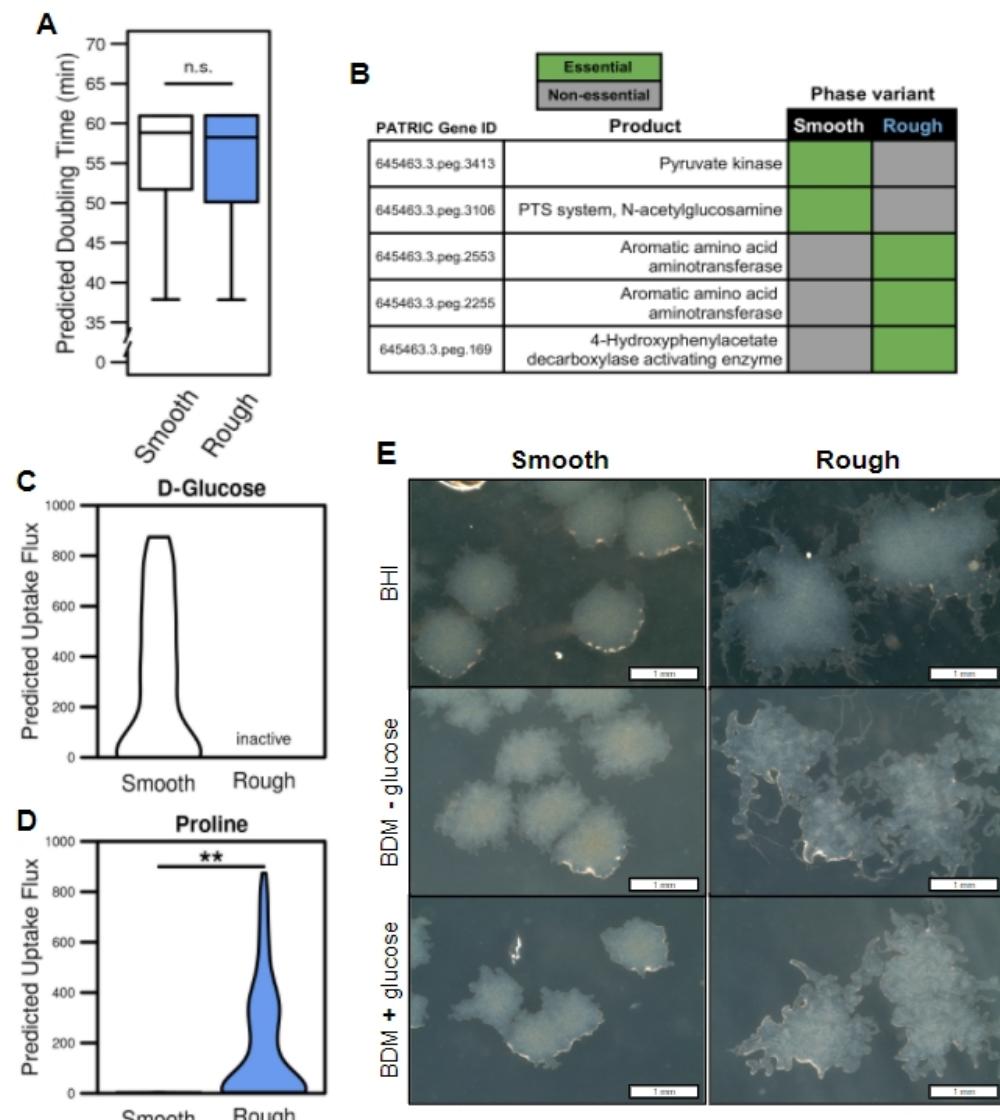


622

623 **Figure 2) Carbon source growth enhancement predictions reflect laboratory measurements.**

624 Experimental analysis was performed for both str. 630 and str. R20291 in Scaria et al. 2014., and 115
 625 metabolites were shared between the GENREs and the Biolog carbon source phenotypic screen. **(A)**
 626 Schematic of specific *in vitro* and *in silico* measurements being utilized. The arrow indicates the correlations
 627 made in subsequent panels. Ratios of overall *in vitro* growth enhancement by each metabolite were correlated
 628 with the corresponding results from growth simulations in the same media for **(B)** iCdG698 (str. 630) and **(C)**
 629 iCdR700 (str. R20291). Points are colored by their biochemical grouping, and significant relationships were
 630 determined by Spearman correlation. **(D)** Binary quantification for predictions in B & C respectively.

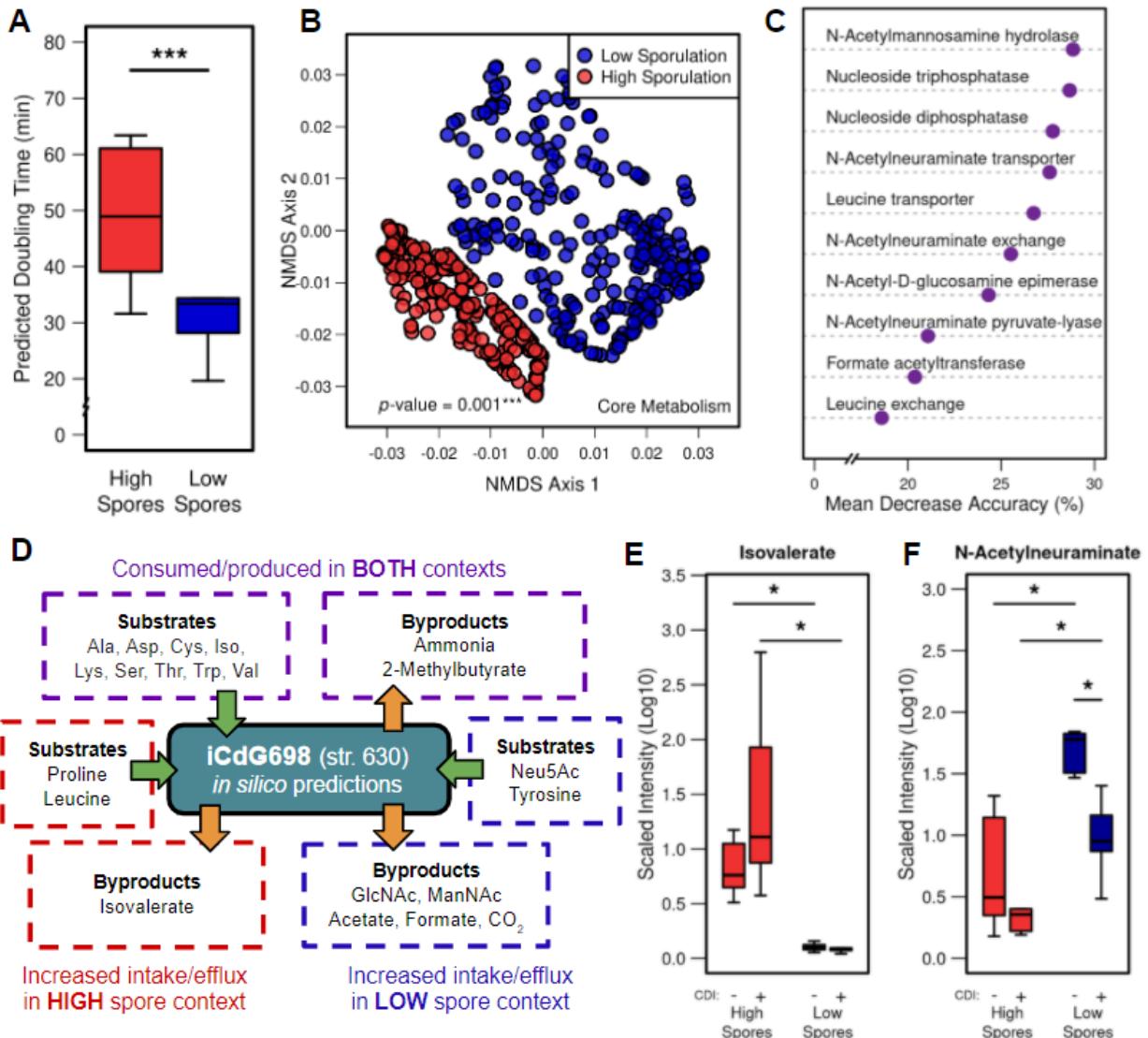
631



632

633 **Figure 3) iCdR700 predicts concerted metabolic shifts during phase variation in str. R20291 grown *in***
 634 ***vitro*.** Transcriptomes were collected from rough or smooth colony morphology clones grown on BHI agar for
 635 48 hours, and subsequently used to generate context-specific models. **(A)** Doubling times calculated from
 636 sampled biomass objective fluxes in each context-specific mode (p -value = 0.221). **(B)** Cross-referenced
 637 gene essentiality results between the context-specific models with $\geq 80\%$ optimal biomass
 638 generation. Importing exchange reaction absolute flux between phase variants for **(C)** D-glucose
 639 and **(D)** proline (** p -value = 0.007). Inactive label denotes reactions pruned during RIPTiDe transcriptome
 640 contextualization. All significant differences determined by Wilcoxon rank-sum test. **(E)** Colony morphologies
 641 resulting from smooth and rough variants of *C. difficile* str. R20291 grown on either BHI or BDM +/- glucose (2
 642 mg/ml) after 48 hours of growth (Phase contrast 20/40, 4X magnification). Defined medium colonies were then
 643 subcultured onto BHI medium for an additional 24 hours as indicated.

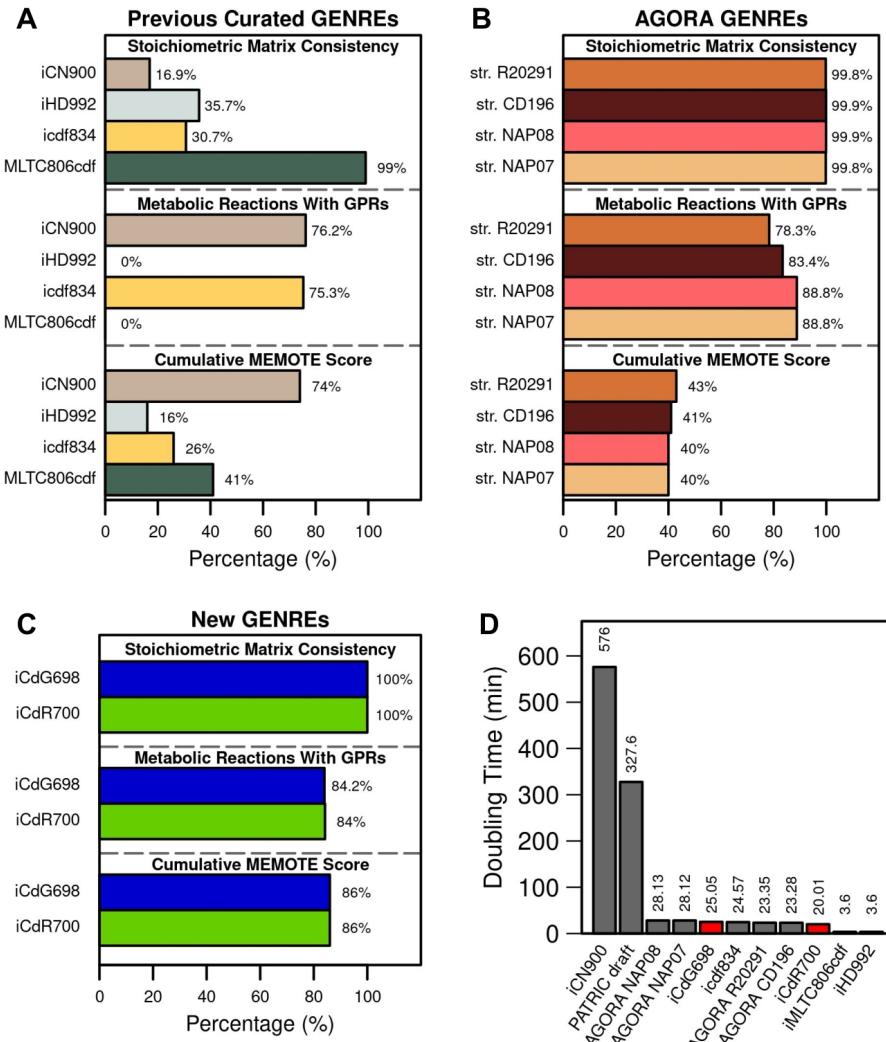
50



646 **Figure 4) iCdG698 predicts distinct metabolic patterns associated with pathogenicity across *in vivo* str.
647 630 infections.** Transcriptomic integration and predictions with iCdG698, 18 hours post-infection with str. 630
648 across infections with either high or low levels of sporulation were detected in the cecum. **(A)** Doubling times
649 calculated from sampled biomass objective fluxes in each context-specific model. Significant differences
650 determined by Wilcoxon rank-sum test (* p-value << 0.001). **(B)** NMDS ordination of Bray-Curtis dissimilarities
651 for flux distributions shared reactions following sampling of context-specific models. Significant difference
652 calculated by PERMANOVA. **(C)** Mean decrease accuracy for most discerning reactions from AUC Random
653 Forest supervised machine learning results using sampled flux distributions from both groups (Out of bag error
654 = 0%). **(D)** A subset of context-specific metabolite consumption or production predictions. Asterisks indicate
655 those metabolites that appear in both context-specific models, but flux through the associated exchange

reaction is significantly greater in the context shown (Table S5). **(E & F)** Liquid-chromatography mass spectrometry analysis from cecal content of mice with and without *C. difficile* str. 630 infection in antibiotic pretreatment groups that resulted in either high or low cecal spore CFUs for metabolites highlighted by growth simulation analysis: (E) Isovalerate and (F) N-Acetylneuraminate. Significant differences determined by Wilcoxon rank-sum test with Benjamini-Hochberg correction (* p-values ≤ 0.05).

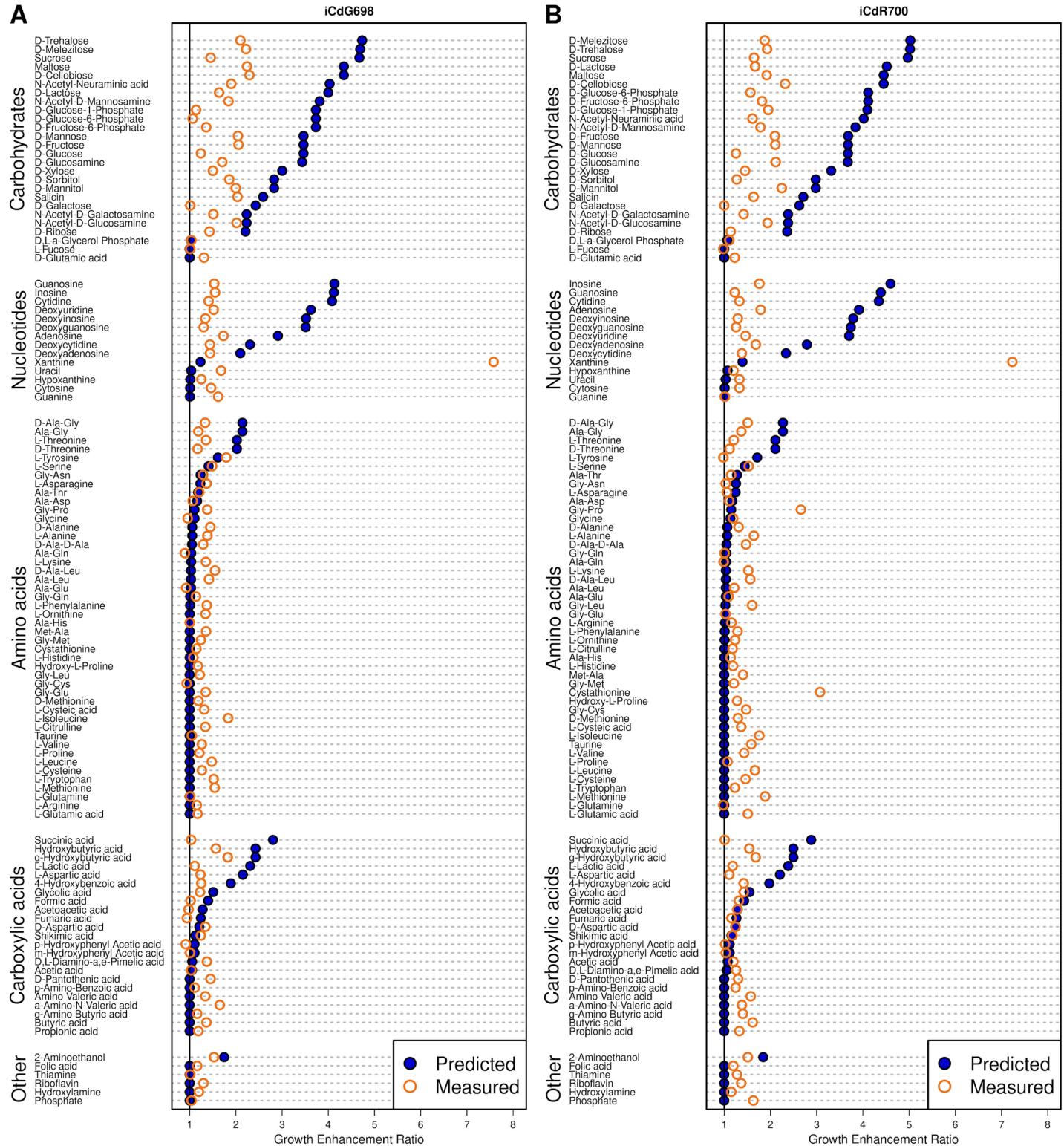
661



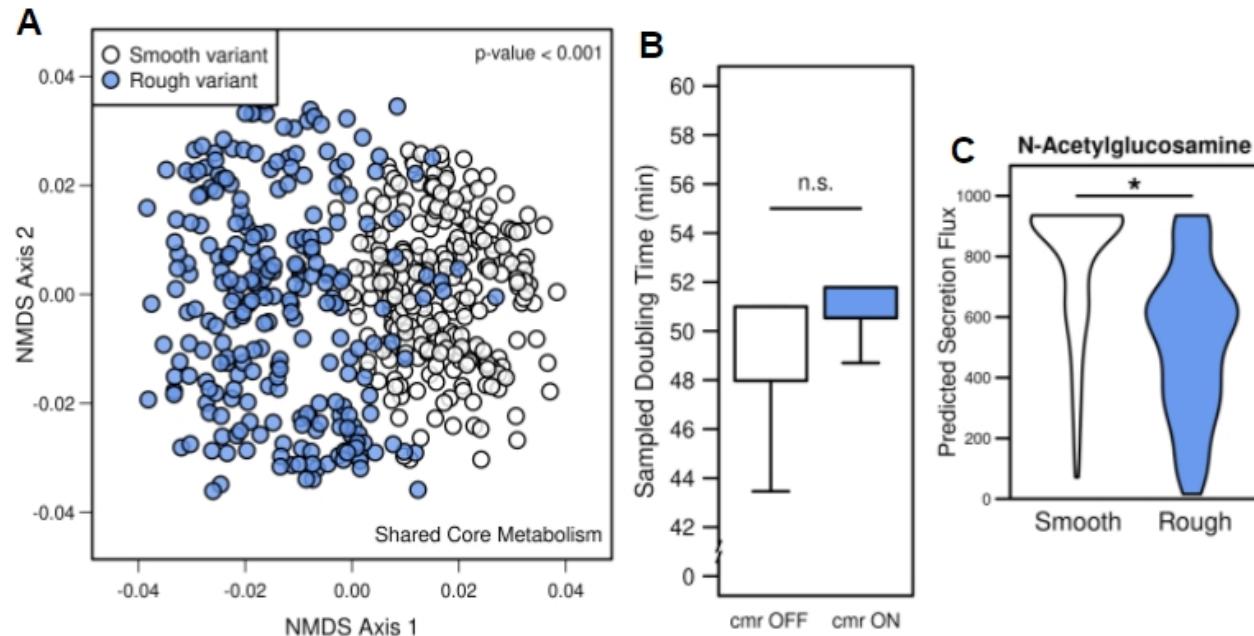
662

Figure S1) Selected quality metrics for *C. difficile* GENREs. Stoichiometric inconsistency describes consistent mass conservation across metabolic reactions. Assessing for metabolic reactions without gene-reaction rules (GPRs) excludes exchange reactions, transport reactions, and those reactions associated with biomass generation. Cumulative MEMOTE quality scores for each GENRE in default media settings, reflecting overall GENRE integrity and annotation completeness. **(A)** Quality assessments for previously published and

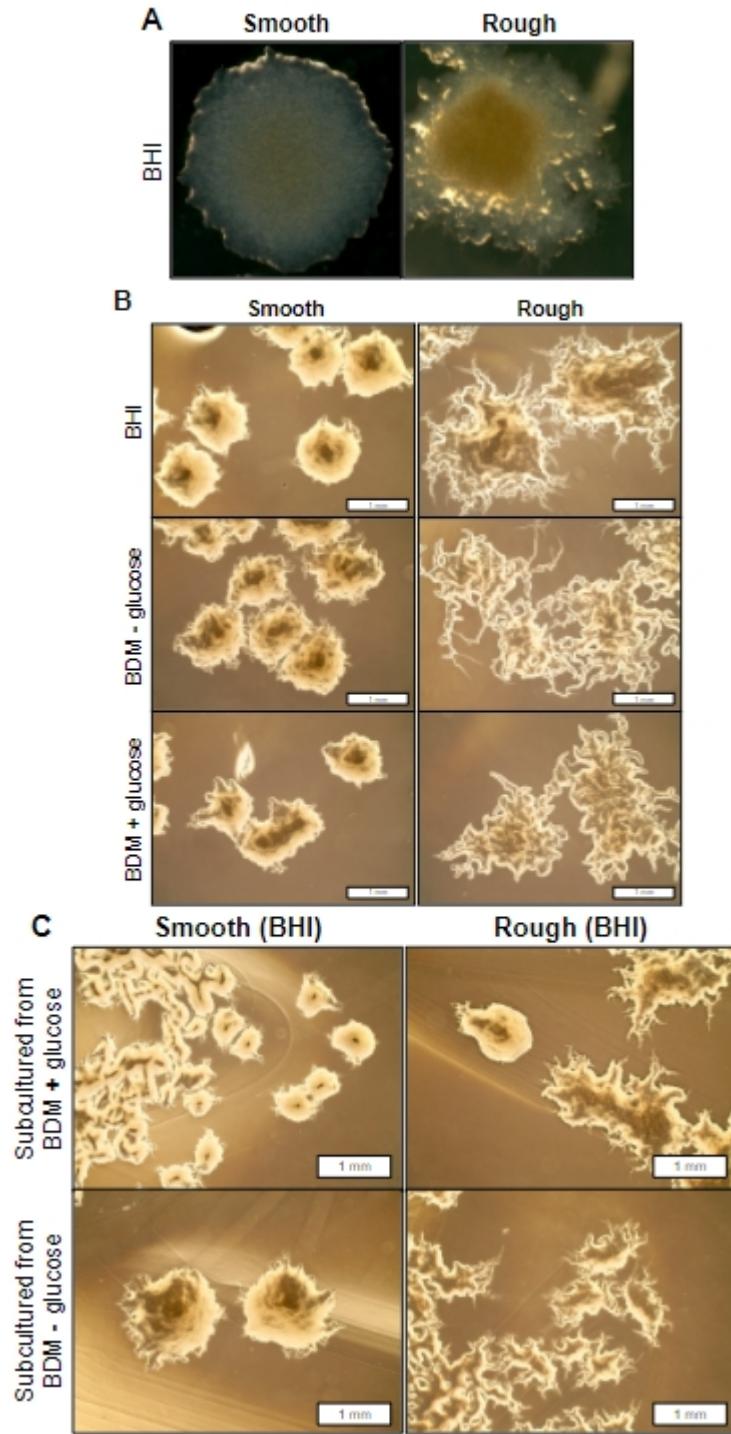
668 manually curated *C. difficile* GENREs, **(B)** *C. difficile* strain GENREs contained in the AGORA database of
 669 auto-curated reconstructions, or **(C)** the two new GENREs described in the current study. **(D)** Imputed doubling
 670 time in complete media, calculated as the reciprocal optimal biomass flux per unit time for all GENREs. Bars
 671 for previous GENREs are colored gray and bars for the new GENREs (*iCdG698* and *iCdR700*) are colored red.
 672



674 **Figure S2) Specific shifts in simulated versus measured growth enhancement for each metabolite**
 675 **measured in the carbon source utilization screen.** Metabolites are separated into groups by metabolite
 676 superfamily designation. Fold change for both *in vitro* and *in silico* measurements reflects growth enhancement
 677 for each metabolite relative to background (Fig. 2A). Results for both **(A)** iCdG698 and **(B)** iCdR700 are
 678 shown, and discrete Spearman correlation coefficients are listed for each category.
 679



681 **Figure S3) Change in growth simulations using *in vitro* transcriptomic data from naturally occurring or**
 682 **mutant phase-locked str. R20291 colony variants integrated into iCdR700.** Rough versus Smooth variant
 683 transcriptomes integrated with RIPTiDe into iCdR700. **(A)** NMDS ordination of Bray-Curtis dissimilarities
 684 between flux sampled distributions of shared reactions of context-specific models. Significant difference
 685 calculated by PERMANOVA (** p-value < 0.001). Transcriptomic data from cmr operon mutants (described
 686 previously) was also utilized to generate context-specific models for phase-locked isolates. **(B)** Following the
 687 same trend as phase-favoring colony variants, optimal biomass objective flux from each context-specific model
 688 was not significantly different. **(C)** Exchange reaction flux associated with N-acetylglucosamine export for both
 689 context-specific models (* p-value = 0.015). Significant difference determined by Wilcoxon rank-sum test.



690

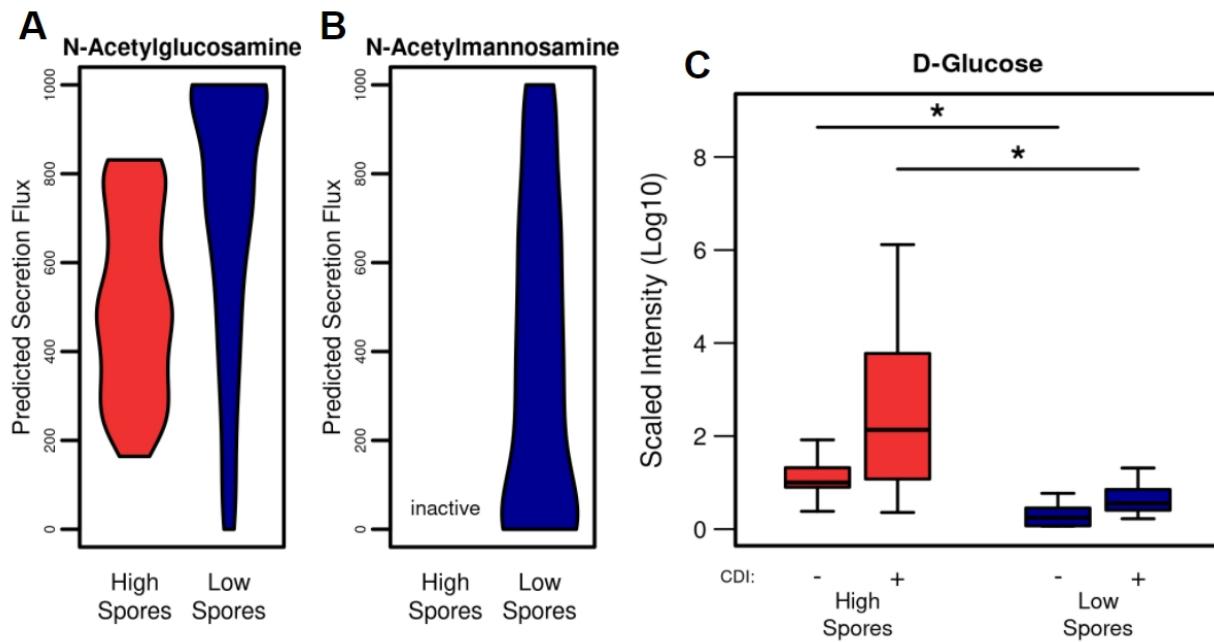
691 **Figure S4) Additional microscopy of phase variant colony morphologies. (A)** *C. difficile* str. R20291
 692 phase variants progenitor colonies generated on solid BHIS agar following 48 hours of growth at 37° C under
 693 anaerobic conditions. These colonies were subcultured and utilized for all subsequent defined minimal medium
 694 experiments. **(B)** Additional phase contrast (4/10) microscopy images of identical colonies from Fig. 3C (4X
 695 magnification). **(C)** Subcultured colonies from the indicated conditions in Fig. 3C onto BHI rich agar medium,
 696 incubated at 37° C for 48 hours anaerobically.

60

61

697

698



699

Figure S5) Context-specific growth simulation with iCdR698 predicts discordant carbohydrate-associated metabolism which agrees with *in vivo* measurements. **(A)** Exchange reaction flux associated with N-acetylglucosamine export for both high and low spore context-specific models (p -value = 0.067). Significant difference determined by Wilcoxon rank-sum test. **(B)** Exchange reaction flux associated with N-mannosamine export for the low spore context-specific models, this reaction was pruned in the high spore context. **(C)** Cecal concentrations of glucose across measured contexts. Matched LC-MS metabolomic analysis of relative D-glucose concentrations from cecal content of mice with and without *C. difficile* str. 630 infection in antibiotic pretreatment groups that resulted in either high or low cecal spore CFUs. Significant differences determined by Wilcoxon rank-sum test with Benjamini-Hochberg correction (* p -values ≤ 0.05).

709

Table S1) Topology summary statistics for *C. difficile* GENREs from AGORA and those generated here.

711

Table S2) GENRE creation steps, Biomass formulation, Gap-filling media compositions, and GENRE statistics.

713

Table S3) *C. difficile* 630 and R20291 PATRIC protein sequence alignment results.

715

62

716 **Table S4)** Differential transcription and exchange fluxes for iCdR700 (str. R20291) with *in vitro* transcriptome.

717

718 **Table S5)** Differential gene essentiality and exchange fluxes for iCdG698 (str. 630) with *in vivo* transcriptome.

719

720 **REFERENCES**

- 721 1. Lessa, F. C., Winston, L. G., McDonald, L. C. & Emerging Infections Program *C. difficile* Surveillance
722 Team. Burden of Clostridium difficile infection in the United States. *The New England journal of medicine*
723 vol. 372 2369–2370 (2015).
- 724 2. Bella, S. D. et al. Clostridium difficile Toxins A and B: Insights into Pathogenic Properties and
725 Extraintestinal Effects. *Toxins* vol. 8 134 (2016).
- 726 3. Thomas, C. Antibiotics and hospital-acquired *Clostridium difficile*-associated diarrhoea: a systematic
727 review. *Journal of Antimicrobial Chemotherapy* vol. 51 1339–1350 (2003).
- 728 4. Antunes, L. C. M. et al. Effect of antibiotic treatment on the intestinal metabolome. *Antimicrob. Agents*
729 *Chemother.* **55**, 1494–1503 (2011).
- 730 5. Theriot, C. M. et al. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase
731 susceptibility to *Clostridium difficile* infection. *Nat. Commun.* **5**, 3114 (2014).
- 732 6. Fletcher, J. R., Erwin, S., Lanzas, C. & Theriot, C. M. Shifts in the Gut Metabolome and *Clostridium*
733 *difficile* Transcriptome throughout Colonization and Infection in a Mouse Model. *mSphere* vol. 3 (2018).
- 734 7. Jenior, M. L., Leslie, J. L., Young, V. B. & Schloss, P. D. *Clostridium difficile* Colonizes Alternative Nutrient
735 Niches during Infection across Distinct Murine Gut Microbiomes. *mSystems* **2**, (2017).
- 736 8. Jenior, M. L., Leslie, J. L., Young, V. B. & Schloss, P. D. *Clostridium difficile* Alters the Structure and
737 Metabolism of Distinct Cecal Microbiomes during Initial Infection To Promote Sustained Colonization.
738 *mSphere* vol. 3 (2018).
- 739 9. Hofmann, J. D. et al. Metabolic Reprogramming of *Clostridioides difficile* During the Stationary Phase With
740 the Induction of Toxin Production. *Frontiers in Microbiology* vol. 9 (2018).
- 741 10. Dineen, S. S., McBride, S. M. & Sonenshein, A. L. Integration of metabolism and virulence by *Clostridium*
742 *difficile* CodY. *J. Bacteriol.* **192**, 5350–5362 (2010).

- 743 11. (u.s.), C. F. D. C. A. P. & Centers for Disease Control and Prevention (U.S.). Antibiotic resistance threats
744 in the United States, 2019. (2019) doi:10.15620/cdc:82532.
- 745 12. Merrigan, M. et al. Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as
746 robust toxin production. *J. Bacteriol.* **192**, 4904–4911 (2010).
- 747 13. Hao, T. et al. The Genome-Scale Integrated Networks in Microorganisms. *Front. Microbiol.* **9**, 296 (2018).
- 748 14. Pacheco, A. R., Moel, M. & Segrè, D. Costless metabolic secretions as drivers of interspecies interactions
749 in microbial ecosystems. *Nat. Commun.* **10**, 103 (2019).
- 750 15. Seif, Y. et al. Metabolic and genetic basis for auxotrophies in Gram-negative species. *Proc. Natl. Acad.
751 Sci. U. S. A.* **117**, 6264–6273 (2020).
- 752 16. Hadadi, N. et al. Mechanistic insights into bacterial metabolic reprogramming from omics-integrated
753 genome-scale models. *NPJ Syst Biol Appl* **6**, 1 (2020).
- 754 17. Cesur, M. F., Siraj, B., Uddin, R., Durmuş, S. & Çakır, T. Network-Based Metabolism-Centered Screening
755 of Potential Drug Targets in *Klebsiella pneumoniae* at Genome Scale. *Frontiers in Cellular and Infection
756 Microbiology* vol. 9 (2020).
- 757 18. Jijakli, K. & Jensen, P. A. Metabolic Modeling of *Streptococcus mutans* Reveals Complex Nutrient
758 Requirements of an Oral Pathogen. *mSystems* **4**, (2019).
- 759 19. Bosi, E. et al. Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-
760 specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3801–9 (2016).
- 761 20. Larocque, M., Chénard, T. & Najmanovich, R. A curated *C. difficile* strain 630 metabolic network:
762 prediction of essential targets and inhibitors. *BMC Syst. Biol.* **8**, 117 (2014).
- 763 21. Kashaf, S. S., Angione, C. & Lió, P. Making life difficult for *Clostridium difficile*: augmenting the pathogen's
764 metabolic model with transcriptomic and codon usage data for better therapeutic target characterization.
765 *BMC Systems Biology* vol. 11 (2017).
- 766 22. Dannheim, H., Will, S. E., Schomburg, D. & Neumann-Schaal, M. *Clostridioides difficile* 630Δerm in
767 silicoandin vivo- quantitative growth and extensive polysaccharide secretion. *FEBS Open Bio* vol. 7 602–
768 615 (2017).
- 769 23. King, Z. A. et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale
770 models. *Nucleic Acids Res.* **44**, D515–22 (2016).

- 771 24. Monot, M. et al. Reannotation of the genome sequence of *Clostridium difficile* strain 630. *J. Med.*
772 *Microbiol.* **60**, 1193–1199 (2011).
- 773 25. Stabler, R. A. et al. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains
774 provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* **10**, R102 (2009).
- 775 26. Antunes, A. et al. Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium*
776 *difficile*. *Nucleic Acids Res.* **40**, 10701–10718 (2012).
- 777 27. Norsigian, C. J. et al. Systems biology analysis of the *Clostridioides difficile* core-genome contextualizes
778 microenvironmental evolutionary pressures leading to genotypic and phenotypic divergence. *NPJ Syst Biol*
779 *Appl* **6**, 31 (2020).
- 780 28. Dannheim, H. et al. Manual curation and reannotation of the genomes of *Clostridium difficile* 630Δerm and
781 *C. difficile* 630. *Journal of Medical Microbiology* vol. 66 286–293 (2017).
- 782 29. Hussain, H. A., Roberts, A. P. & Mullany, P. Generation of an erythromycin-sensitive derivative of
783 *Clostridium difficile* strain 630 (630Δerm) and demonstration that the conjugative transposon Tn916ΔE
784 enters the genome of this strain at multiple sites. *Journal of Medical Microbiology* vol. 54 137–141 (2005).
- 785 30. Magnúsdóttir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the
786 human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
- 787 31. Gevorgyan, A., Poolman, M. G. & Fell, D. A. Detection of stoichiometric inconsistencies in biomolecular
788 models. *Bioinformatics* **24**, 2245–2251 (2008).
- 789 32. Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B. & Lercher, M. J. Erroneous energy-generating
790 cycles in published genome scale metabolic networks: Identification and removal. *PLoS Comput. Biol.* **13**,
791 e1005494 (2017).
- 792 33. Schellenberger, J., Lewis, N. E. & Palsson, B. Ø. Elimination of thermodynamically infeasible loops in
793 steady-state metabolic models. *Biophys. J.* **100**, 544–553 (2011).
- 794 34. Ravikrishnan, A. & Raman, K. Critical assessment of genome-scale metabolic networks: the need for a
795 unified standard. *Brief. Bioinform.* **16**, 1057–1068 (2015).
- 796 35. Lieven, C. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**,
797 272–276 (2020).
- 798 36. Oberhardt, M. A., Puchałka, J., dos Santos, V. A. P. M. & Papin, J. A. Reconciliation of Genome-Scale

- 799 Metabolic Reconstructions for Comparative Systems Analysis. *PLoS Computational Biology* vol. 7
800 e1001116 (2011).
- 801 37. Neumann-Schaal, M., Hofmann, J. D., Will, S. E. & Schomburg, D. Time-resolved amino acid uptake of
802 *Clostridium difficile* 630Δerm and concomitant fermentation product and toxin formation. *BMC Microbiol.*
803 **15**, 281 (2015).
- 804 38. Neumann-Schaal, M., Jahn, D. & Schmidt-Hohagen, K. Metabolism the Difficile Way: The Key to the
805 Success of the Pathogen *Clostridioides difficile*. *Frontiers in Microbiology* vol. 10 (2019).
- 806 39. Wattam, A. R. et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis
807 Resource Center. *Nucleic Acids Res.* **45**, D535–D542 (2017).
- 808 40. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction.
809 *Nat. Protoc.* **5**, 93–121 (2010).
- 810 41. Devold, S. et al. Automated genome annotation and metabolic model reconstruction in the SEED and
811 Model SEED. *Methods Mol. Biol.* **985**, 17–45 (2013).
- 812 42. Satish Kumar, V., Dasika, M. S. & Maranas, C. D. Optimization based automated curation of metabolic
813 reconstructions. *BMC Bioinformatics* **8**, 212 (2007).
- 814 43. Girinathan, B. P., Braun, S., Sirigireddy, A. R., Espinola-Lopez, J. & Govind, R. Correction: Importance of
815 Glutamate Dehydrogenase (GDH) in *Clostridium difficile* Colonization *in vivo*. *PLoS One* **11**, e0165579
816 (2016).
- 817 44. Bakker, D. et al. The HtrA-like protease CD3284 modulates virulence of *Clostridium difficile*. *Infect.*
818 *Immun.* **82**, 4222–4232 (2014).
- 819 45. Mendoza, S. N., Olivier, B. G., Molenaar, D. & Teusink, B. A systematic assessment of current genome-
820 scale metabolic reconstruction tools. *Genome Biology* vol. 20 (2019).
- 821 46. Janoir, C. et al. Adaptive strategies and pathogenesis of *Clostridium difficile* from *in vivo* transcriptomics.
822 *Infect. Immun.* **81**, 3757–3769 (2013).
- 823 47. Karasawa, T., Ikoma, S., Yamakawa, K. & Nakamura, S. A defined growth medium for *Clostridium difficile*.
824 *Microbiology* **141 (Pt 2)**, 371–375 (1995).
- 825 48. Karlsson, S., Burman, L. G. & Åkerlund, T. Suppression of toxin production in *Clostridium difficile* VPI
826 10463 by amino acids. *Microbiology* **145 (Pt 7)**, 1683–1693 (1999).

- 827 49. Haslam, S. C. et al. Growth of *Clostridium difficile* and production of toxins A and B in complex and
828 defined media. *J. Med. Microbiol.* **21**, 293–297 (1986).
- 829 50. Stiemstra, L. T., Turvey, S. E. & Finlay, B. B. An antibiotic-altered microbiota provides fuel for the enteric
830 foe. *Cell research* vol. 24 5–6 (2014).
- 831 51. Olson, M. E., King, J. M., Yahr, T. L. & Horswill, A. R. Sialic acid catabolism in *Staphylococcus aureus*. *J.*
832 *Bacteriol.* **195**, 1779–1788 (2013).
- 833 52. Engevik, M. A. et al. Human *Clostridium difficile* infection: altered mucus production and composition. *Am.*
834 *J. Physiol. Gastrointest. Liver Physiol.* **308**, G510–24 (2015).
- 835 53. Nakamura, S., Nakashio, S., Yamakawa, K., Tanabe, N. & Nishida, S. Carbohydrate Fermentation by
836 *Clostridium difficile*. *Microbiology and Immunology* vol. 26 107–111 (1982).
- 837 54. Kim, J., Darley, D. & Buckel, W. 2-Hydroxyisocaproyl-CoA dehydratase and its activator from *Clostridium*
838 *difficile*. *FEBS J.* **272**, 550–561 (2005).
- 839 55. Selmer, T. & Andrei, P. I. p-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycyl
840 radical enzyme catalysing the formation of p-cresol. *Eur. J. Biochem.* **268**, 1363–1372 (2001).
- 841 56. Bouillaut, L., Self, W. T. & Sonenshein, A. L. Proline-dependent regulation of *Clostridium difficile* Stickland
842 metabolism. *J. Bacteriol.* **195**, 844–854 (2013).
- 843 57. Jackson, S., Calos, M., Myers, A. & Self, W. T. Analysis of proline reduction in the nosocomial pathogen
844 *Clostridium difficile*. *J. Bacteriol.* **188**, 8487–8495 (2006).
- 845 58. Kim, J., Darley, D., Selmer, T. & Buckel, W. Characterization of (R)-2-Hydroxyisocaproate Dehydrogenase
846 and a Family III Coenzyme A Transferase Involved in Reduction of L-Leucine to Isocaproate by
847 *Clostridium difficile*. *Applied and Environmental Microbiology* vol. 72 6062–6069 (2006).
- 848 59. Britz, M. L. & Wilkinson, R. G. Leucine dissimilation to isovaleric and isocaproic acids by cell suspensions
849 of amino acid fermenting anaerobes: the Stickland reaction revisited. *Can. J. Microbiol.* **28**, 291–300
850 (1982).
- 851 60. Kim, J., Hetzel, M., Boiangiu, C. D. & Buckel, W. Dehydration of (R)-2-hydroxyacyl-CoA to enoyl-CoA in
852 the fermentation of α-amino acids by anaerobic bacteria. *FEMS Microbiology Reviews* vol. 28 455–468
853 (2004).
- 854 61. Yu, L., Blaser, M., Andrei, P. I., Pierik, A. J. & Selmer, T. 4-Hydroxyphenylacetate decarboxylases:

- 855 properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* **45**, 9584–9592 (2006).
- 856 62. de Vladar, H. P. Amino acid fermentation at the origin of the genetic code. *Biol. Direct* **7**, 6 (2012).
- 857 63. Esquivel-Elizondo, S., Ilhan, Z. E., Garcia-Peña, E. I. & Krajmalnik-Brown, R. Insights into Butyrate
- 858 Production in a Controlled Fermentation System via Gene Predictions. *mSystems* vol. 2 (2017).
- 859 64. Ferreyra, J. A. et al. Gut microbiota-produced succinate promotes *C. difficile* infection after antibiotic
- 860 treatment or motility disturbance. *Cell Host Microbe* **16**, 770–777 (2014).
- 861 65. Louis, P. & Flint, H. J. Formation of propionate and butyrate by the human colonic microbiota.
- 862 *Environmental Microbiology* vol. 19 29–41 (2017).
- 863 66. Nawrocki, K. L., Wetzel, D., Jones, J. B., Woods, E. C. & McBride, S. M. Ethanolamine is a valuable
- 864 nutrient source that impacts *Clostridium difficile* pathogenesis. *Environ. Microbiol.* **20**, 1419–1435 (2018).
- 865 67. Senger, R. S. & Papoutsakis, E. T. Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic
- 866 network resolution and analysis. *Biotechnol. Bioeng.* **101**, 1036–1052 (2008).
- 867 68. Lachance, J.-C. et al. BOFdat: Generating biomass objective functions for genome-scale metabolic
- 868 models from experimental data. *PLoS Comput. Biol.* **15**, e1006971 (2019).
- 869 69. Biggs, M. B. & Papin, J. A. Managing uncertainty in metabolic network structure and improving predictions
- 870 using EnsembleFBA. *PLoS Comput. Biol.* **13**, e1005413 (2017).
- 871 70. Medlock, G. L., Moutinho, T. J. & Papin, J. A. Medusa: Software to build and analyze ensembles of
- 872 genome-scale metabolic network reconstructions. *PLoS Comput. Biol.* **16**, e1007847 (2020).
- 873 71. Atlas, R. & Ronald Atlas. Handbook of Microbiological Media, Fourth Edition. (2010)
- 874 doi:10.1201/ebk1439804063.
- 875 72. Blazier, A. S. & Papin, J. A. Reconciling high-throughput gene essentiality data with metabolic network
- 876 reconstructions. *PLoS Comput. Biol.* **15**, e1006507 (2019).
- 877 73. Heap, J. T. et al. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *J. Microbiol. Methods*
- 878 **80**, 49–55 (2010).
- 879 74. Dembek, M. et al. High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile*.
- 880 *MBio* **6**, e02383 (2015).
- 881 75. O'Brien, E. J., Monk, J. M. & Palsson, B. O. Using Genome-scale Models to Predict Biological
- 882 Capabilities. *Cell* **161**, 971–987 (2015).

- 883 76. Dubois, T. et al. Control of *Clostridium difficile* Physiopathology in Response to Cysteine Availability.
884 *Infect. Immun.* **84**, 2389–2405 (2016).
- 885 77. Ikeda, D. et al. Effect of Isoleucine on Toxin Production by *Clostridium difficile* in a Defined Medium.
886 *Zentralblatt für Bakteriologie* vol. 287 375–386 (1998).
- 887 78. Scaria, J. et al. Comparative nutritional and chemical phenotype of *Clostridium difficile* isolates determined
888 using phenotype microarrays. *Int. J. Infect. Dis.* **27**, 20–25 (2014).
- 889 79. Battaglioli, E. J. et al. *Clostridioides difficile* uses amino acids associated with gut microbial dysbiosis in a
890 subset of patients with diarrhea. *Science Translational Medicine* vol. 10 eaam7019 (2018).
- 891 80. Oh, Y.-K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-scale reconstruction of
892 metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J.
893 Biol. Chem.* **282**, 28791–28799 (2007).
- 894 81. Blazier, A. S. & Papin, J. A. Integration of expression data in genome-scale metabolic network
895 reconstructions. *Front. Physiol.* **3**, 299 (2012).
- 896 82. Jenior, M. L., Moutinho, T. J., Jr, Dougherty, B. V. & Papin, J. A. Transcriptome-guided parsimonious flux
897 analysis improves predictions with metabolic networks in complex environments. *PLoS Comput. Biol.* **16**,
898 e1007099 (2020).
- 899 83. Anjuwon-Foster, B. R. & Tamayo, R. A genetic switch controls the production of flagella and toxins in
900 *Clostridium difficile*. *PLoS Genet.* **13**, e1006701 (2017).
- 901 84. Garrett, E. M. et al. Phase variation of a signal transduction system controls *Clostridioides difficile* colony
902 morphology, motility, and virulence. *PLoS Biol.* **17**, e3000379 (2019).
- 903 85. Dubois, T. et al. A microbiota-generated bile salt induces biofilm formation in *Clostridium difficile*. *NPJ
904 Biofilms Microbiomes* **5**, 14 (2019).
- 905 86. Passmore, I. J. et al. Para-cresol production by *Clostridium difficile* affects microbial diversity and
906 membrane integrity of Gram-negative bacteria. *PLoS Pathog.* **14**, e1007191 (2018).
- 907 87. Schubert, A. M., Sinani, H. & Schloss, P. D. Antibiotic-Induced Alterations of the Murine Gut Microbiota
908 and Subsequent Effects on Colonization Resistance against *Clostridium difficile*. *MBio* **6**, e00974 (2015).
- 909 88. Jenior, M. L., Leslie, J. L., Young, V. B. & Schloss, P. D. *Clostridium difficile* colonizes alternative nutrient
910 niches during infection across distinct murine gut microbiomes. doi:10.1101/092304.

- 911 89. Wilson, K. H. & Perini, F. Role of competition for nutrients in suppression of *Clostridium difficile* by the
912 colonic microflora. *Infect. Immun.* **56**, 2610–2614 (1988).
- 913 90. Purcell, E. B., McKee, R. W., McBride, S. M., Waters, C. M. & Tamayo, R. Cyclic diguanylate inversely
914 regulates motility and aggregation in *Clostridium difficile*. *J. Bacteriol.* **194**, 3307–3316 (2012).
- 915 91. Bordeleau, E., Fortier, L.-C., Malouin, F. & Burrus, V. c-di-GMP turn-over in *Clostridium difficile* is
916 controlled by a plethora of diguanylate cyclases and phosphodiesterases. *PLoS Genet.* **7**, e1002039
917 (2011).
- 918 92. Antunes, A., Martin-Verstraete, I. & Dupuy, B. CcpA-mediated repression of *Clostridium difficile* toxin gene
919 expression. *Mol. Microbiol.* **79**, 882–899 (2011).
- 920 93. Sivsammye, G. & Sims, H. V. Presumptive identification of *Clostridium difficile* by detection of p-cresol in
921 prepared peptone yeast glucose broth supplemented with p-hydroxyphenylacetic acid. *Journal of Clinical
922 Microbiology* vol. 28 1851–1853 (1990).
- 923 94. Zhang, B. & Powers, R. Analysis of bacterial biofilms using NMR-based metabolomics. *Future Med.
924 Chem.* **4**, 1273–1306 (2012).
- 925 95. Maegawa, T. et al. Linkage between toxin production and purine biosynthesis in *Clostridium difficile*. *J.
926 Med. Microbiol.* **51**, 34–41 (2002).
- 927 96. Wu, X. & Hurdle, J. G. The *Clostridium difficile* proline racemase is not essential for early logarithmic
928 growth and infection. *Canadian Journal of Microbiology* vol. 60 251–254 (2014).
- 929 97. Raškevičius, V. et al. Genome scale metabolic models as tools for drug design and personalized
930 medicine. *PLoS One* **13**, e0190636 (2018).
- 931 98. Wattam, A. R. et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids
932 Res.* **42**, D581–91 (2014).
- 933 99. Faria, J. P., Rocha, M., Rocha, I. & Henry, C. S. Methods for automated genome-scale metabolic model
934 reconstruction. *Biochem. Soc. Trans.* **46**, 931–936 (2018).
- 935 100. Heirendt, L. et al. Creation and analysis of biochemical constraint-based models using the COBRA
936 Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).
- 937 101. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale
938 metabolic models. *Genome Biol.* **20**, 121 (2019).

- 939 102.Ren, Q., Chen, K. & Paulsen, I. T. TransportDB: a comprehensive database resource for cytoplasmic
940 membrane transport systems and outer membrane channels. *Nucleic Acids Res.* **35**, D274–9 (2007).
- 941 103.Seaver, S. M. D. et al. The ModelSEED Database for the integration of metabolic annotations and the
942 reconstruction, comparison, and analysis of metabolic models for plants, fungi, and microbes.
943 doi:10.1101/2020.03.31.018663.
- 944 104.Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* vol. 28 27–30
945 (2000).
- 946 105.Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant
947 sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–5 (2007).
- 948 106.Baker, W. The EMBL Nucleotide Sequence Database. *Nucleic Acids Research* vol. 28 19–23 (2000).
- 949 107.Norsigian, C. J. et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the
950 phylogenetic tree. *Nucleic Acids Res.* **48**, D402–D406 (2020).
- 951 108.Hucka, M. et al. The Systems Biology Markup Language (SBML): Language Specification for Level 3
952 Version 2 Core Release 2. *J. Integr. Bioinform.* **16**, (2019).
- 953 109.Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction
954 and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).
- 955 110.Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**,
956 489 (2010).
- 957 111.Keaty, T. C. & Jensen, P. A. gapsplit: Efficient random sampling for non-convex constraint-based models.
958 doi:10.1101/652917.
- 959 112.Website. Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for
960 FastQ files (Version 1.33) [Software]. Available at <https://github.com/najoshi/sickle>.
- 961 113.Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359
962 (2012).
- 963 114.Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- 964 115.Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* vol. 14
965 927 (2003).
- 966 116.Janitza, S., Strobl, C. & Boulesteix, A.-L. An AUC-based permutation variable importance measure for

967 random forests. *BMC Bioinformatics* **14**, 119 (2013).