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

Authors:

Matthew L Jenior[1]

Jhansi L Leslie[2]

Deborah A Powers[1]

Elizabeth M Garrett[6]

Kimberly A Walker[6]

Mary E Dickenson[1]

William A Petri Jr[2,4,5]

Rita Tamayo[6]

Jason A Papin*[1,2,3]

* denotes corresponding author

Affiliations:

1. Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA
2. Department of Medicine, Division of Infectious Diseases & International Health, University of Virginia, Charlottesville, VA, USA
3. Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA, USA
4. Department of Microbiology, Immunology and Cancer Biology, University of Virginia Health System, Charlottesville, Virginia, USA
5. Department of Pathology, University of Virginia Health System, Charlottesville, Virginia, USA.
6. Department of Microbiology & Immunology, University of North Carolina Chapel Hill School of Medicine, NC, USA

ABSTRACT

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

MAIN

The nosocomial bacterial pathogen *Clostridioides* (formerly *Clostridium*) *difficile* causes a toxin-mediated diarrheal illness and is now the leading cause of hospital-acquired infection in the United States [1,2]. Susceptibility to *C. difficile* infection (CDI) is most frequently preceded by exposure to antibiotic therapy [3]. 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 fermentation [9]. 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 efforts [13] and accurately predicting auxotrophies and competition/cooperation between species for growth substrates [14,15]. GENREs also create improved context for the interpretation of omics data [16], and have provided powerful utility for identification of novel drug and gene targets accelerating downstream laboratory testing [17]. 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,

GENREs must be revisited or remade entirely to improve the quality of the resultant metabolic predictions. As such, we began with the updated genome of the highly-characterized laboratory strain *C. difficile* str. 630 [24], first generating a *de novo* reconstruction followed by extensive literature-driven manual curation of catabolic pathways, metabolite transport, and a biomass objective function. We proceeded to use this reconstruction as a template to also create a curated GENRE for the more recently isolated hyper-virulent strain R20291 [25]. Predictions from both GENREs were subsequently compared against published *in vitro* gene essentiality and carbon utilization screens. These predictions indicated a high degree of agreement across experimental datasets. To then assess the application of our GENREs for *in situ* metabolic prediction, we integrated transcriptomic data collected from both *in vitro* and *in vivo* conditions into our models and assessed the emergent metabolic activities. Analysis of context-specific pathogen metabolism revealed conserved patterns of metabolism. Across states of increased virulence, both strains of *C. difficile* favored increased fermentation of amino acids and decreased capacity for glycolysis. These trends agreed with published phenotypes [10,26], and supported the advantage provided by GENREs for delineating complex metabolic networks and patterns of gene expression into more tractable experimental targets. Additionally *in vivo* gene essentiality highlighted specific aspects of nucleotide scavenging as critical for growth during infection and may provide preliminary targets for future inhibitor discovery. Overall, high-quality GENREs can greatly augment the discovery of novel therapeutics to treat CDI due to the connections between metabolic signals and colonization or virulence induction in *C. difficile*. Finally, the current study lays the groundwork for systems-level analyses of CDI-associated metabolism in the context of complex extracellular environments like the gut microbiome during infection.

RESULTS

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

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

genome curation (iCN900 [27]) [28]. Another GENRE was developed for str. 630Δerm (iHD992 [22]), a strain derived from str. 630 by serial passage until erythromycin resistance was lost [29]. Four additional *C. difficile* strain GENREs were generated as a part of an effort to generate numerous new reconstructions for members of the gut microbiota [30]; these reconstructions received only semi-automated curation performed without *C. difficile*-specific considerations.

To establish a baseline for the metabolic predictions possible with current *C. difficile* GENREs, we selected common criteria with large impacts on the quality of subsequent predictions for model performance (Fig. S1A). The first of these metrics is the level of consistency in the stoichiometric matrix [31–33], which reflects proper conservation of mass and that no metabolites are incorrectly created or destroyed during simulations. The next metric is a ratio for the quantity of metabolic reactions lacking gene-reaction rules to those possessing associated genes [34], which may indicate an overall confidence in the annotation of the reactions. These features reflect the importance of mass conservation and deliberate gene/reaction annotation which each drive confidence in downstream metabolic predictions, omics data integration, and likelihood for successful downstream experimentation. We found that each GENRE performed well in some categories, but unique challenges were found in each which made comparing simulation results across models challenging. For example, neither iMLTC806cdf nor iHD994 have any detectable gene annotations associated with the reactions they contain. A high degree of stoichiometric matrix inconsistency was detected across icdf834, iHD992, and iCN900; with iHD992 many intracellular metabolites were able to be generated without acquiring necessary precursors from the environment. These findings reinforced the value of proper biochemical constraints for GENREs to allow for improved fidelity to the target organism's *in situ* metabolism.

We went on to determine the cumulative MEMOTE quality score for each *C. difficile* GENRE (Fig. S1A). MEMOTE is a recent series of model quality assessment guidelines, agreed upon by the research community, and developed into a single platform to create an independent comparable quality metric across GENREs [35]. These percentages reflect a composite measurement of mass conservation, reaction constraint, and standardized component annotation that are necessary for carrying out reliable simulations [34]. The three oldest *C. difficile* reconstructions each scored <50%; conversely the most recent GENRE (iCN900) received a 74% cumulative MEMOTE score yet underperformed in the other metrics. Furthermore, the pre-curation draft *C. difficile* GENREs generated for this study scored similarly (~40%) to those automatically curated AGORA

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

Finally, we assessed key metabolic functionalities and established general principles of *C. difficile* physiology within each of the existing GENREs. First, we compared imputed doubling times of each GENRE, derived from the optimal biomass objective flux value simulated in rich media [36]. While not strictly a measurement of GENRE quality, this value may generally reflect the degree of functional predictions possible with a given GENRE based on its deviation from measured values of ~29 minutes under similar conditions [37]. This analysis uncovered that most GENREs indicated doubling times relatively close to the experimental measures, however iMLTC806cdf and iHD992 gave times under 5 minutes and iCN900 was well over 500 minutes (Fig. S1D). We also detected structural inconsistencies across several GENREs. For example, those GENREs acquired from the AGORA database possessed several intracellular metabolic products not adequately accounted for biologically (Table S1), as well as mitochondrial compartments despite being bacteria. Additionally, several key *C. difficile* metabolic pathways either were incomplete or absent from the curated models including multi-step Stickland fermentation, membrane-dependent ATP synthase, dipeptide and aminoglycan utilization, and a variety of saccharide fermentation pathways [38]. Overall, the existing *C. difficile* GENREs possessed numerous mass imbalances and annotation inconsistencies, lacked key functional capacities, and failed to phenotypically mimic *C. difficile* growth. These collective results motivated the generation of a new reconstruction for our intended analyses.

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

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

630 [28] and the published str. R20291 genome [25], both available on the Pathosystems Resource Integration Center database (PATRIC) [39]. Following a recent protocol for creating high-quality genome-scale models [40], and utilizing the ModelSEED framework and reaction database [41], we generated gap-filled scaffold reconstructions for both strains. Gap-filling refers to the automated process of identifying incomplete metabolic pathways due to an apparent lack of genetic evidence that are also necessary for *in silico* growth, and subsequent addition of the minimal functionality needed to achieve flux through these pathways [42]. The resultant scaffolds were stripped of reactions that were added due to gap-filling in order to be most reflective of original genomic content and partially reveal pathways in need of manual curation (Table S2). Additionally, to focus the reconstructions on bioconversion of metabolites, we removed genes that encoded enzymes involved in macromolecule synthesis (e.g. ribosomal genes). We subsequently performed complete translated proteome alignment between str. 630 and str. R20291, resulting in 684 homologous metabolic gene products and 22 and 33 unique gene products, respectively (Table S3). Among the distinctive features were additional genes for dipeptides import in str. 630 and glycogen import and utilization in str. R20291, which have both been linked to modulated levels of virulence across strains of *C. difficile* [43,44]. After resolving the dissimilarities between the strains by incorporating corresponding metabolism to each reconstruction, we moved on to extensive manual curation of both GENREs.

Metabolic Network Curation and Ensemble Gap-filling

Manual curation is required in order to ultimately produce high-quality GENREs and make meaningful biological predictions [45]. As such, we proceeded to manually incorporate 259 new reactions (with associated genes and metabolites) and altered the conditions of an additional 312 reactions already present within each GENRE prior to gap-filling (Table S2). Primary targets and considerations for the manual curation of the *C. difficile* GENREs included:

- Anaerobic glycolysis, fragmented TCA-cycle, and known molecular oxygen detoxification [38,46]
- Minimal media components and known auxotrophies [47–49]
- Aminoglycan and dipeptide catabolism [50–52]
- Many Stickland fermentation oxidative and reductive pathways (Table S2) [37,53–62]

- Carbohydrate fermentation and SCFA metabolism [53,63–65]
- Energy metabolite reversibility (e.g. ATP, GTP, FAD, etc. [32])
- Periplasmic-associated H+ gradient and ATP synthase
- Additional pathogenicity-associated metabolites (e.g. p-cresol [55] and ethanolamine [66])

Following the outlined manual additions, we created a customized biomass objective function with certain elements tailored to each strain of *C. difficile*. Our biomass objective function formulation was initially adapted from the well-curated GENRE of the close phylogenetic relative *Clostridium acetobutylicum* [67] with additional considerations for tRNA synthesis and formation of cell wall macromolecules, including teichoic acid and peptidoglycan (Table S2). Coefficients within the formulations of DNA replication, RNA replication, and protein synthesis component reactions were adjusted by genomic nucleotide abundances and codon frequencies in order to yield strain-specific biomass objective functions [68]. To successfully simulate growth, we next performed an ensemble-based pFBA gap-filling approach [69,70], utilizing the ModelSEED reaction bag modified to focus on Gram-positive anaerobic bacterial metabolism (see Materials & Methods). We performed gap-filling across six distinct and progressively more limited media conditions; complete medium, Brain-Heart Infusion (BHI [71]), *C. difficile* Defined Medium +/- glucose (CDM [49]), No Carbohydrate Minimal Medium (NCMM [5]), and Basal Defined Medium (BDM [47]) (Table S2). With each step new reactions found across an ensemble were collected and integrated into the draft reconstruction. A total of 68 new reactions allowed for robust growth across all conditions.

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

Gene essentiality results from new GENREs closely match experimental transposon screens

A standard measurement of GENRE performance is the comparison of predicted essential genes for growth *in silico* and those found to be essential experimentally through forward genetic screens [72]. This form of analysis moves past strict network quality criteria and into biologically tractable predictions. Many *C. difficile* strains have been historically difficult to manipulate genetically [73]; however, methods were recently developed and a large-scale transposon mutagenesis screen was published for str. R20291 [74]. As such, we first utilized the proteomic alignment from the previous section to determine those genes in str. 630 that possessed homologs within the str. R20291 dataset. We simulated single gene knockouts for all genes and evaluated for >1% optimal biomass objective flux in BHI medium after growth simulation [75] for both iCdR700 (Fig. 1A) and iCdG698 (Fig. 1B), cross-referencing the results with those in the published study. These comparisons revealed overall accuracies of 89.1% and 88.9%, with negative-predictive values as high as 90.0% for iCdR700 and 89.9% for iCdG698. These results demonstrated that our GENREs correctly predicted with high accuracy the same genes determined to be essential for laboratory growth.

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

To assess if GENRE requirements reflected the components of minimal medium derived experimentally, we identified the minimum subset of metabolites that our model required as an exogenous supply for growth. Importantly, the specific metabolite composition of *C. difficile* minimal medium has been defined across three separate laboratory studies [47–49]. Through *in silico* limitation of extracellular metabolites to only the experimentally determined requirements, followed by growth simulations with systematic omission of each component individually, we were able to determine the impact of each component on achieving some level of biomass flux (Fig. 1C). This analysis revealed that the majority of metabolites found to be essential during growth simulation have also been shown experimentally to be required for *in vitro* growth. In disagreement with two of the published studies, simulations indicated that neither iCdG698 (str. 630) nor iCdR700 (str. R20291) is auxotrophic for methionine. However, the published formulation of BDM where methionine is present found the amino acid to be largely growth-enhancing and not essential for small levels of growth [48]. Additionally, it has been demonstrated in the laboratory that *C. difficile* is able to harvest sufficient bioavailable sulfur from excess cysteine instead of methionine [49,76], further supporting a non-essential status for

this metabolite. In a similar fashion, pantothenate (vitamin B5) only appears to enhance growth rate *in vitro* and is not necessarily required to support slow growth rates. Finally, our results also indicated that iCdR700 was not auxotrophic for isoleucine relative to iCdG698, and indeed contained additional genes coding for synthesis of a precursor (3S)-3-methyl-2-oxopentanoate (*ilvC*, a ketol-acid reductoisomerase) which were not present in its counterpart GENRE (Table S3). Interestingly, increases in isoleucine consumption are associated with greater pathogenicity in some *C. difficile* strains [77], which may contribute to the hypervirulence of str. R20291. In summary, the *in silico* minimal requirements for iCdG698 and iCdR700 closely mirrored experimental results for both strains of *C. difficile* in addition to reconciling partially conflicting reports on experimentally-determined auxotrophies.

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

We next assessed additional carbon sources that impact the growth yield predictions for both GENREs. Utilizing previously published results for both *C. difficile* strains in a Carbon Source Utilization Screen [78], we simulated the degree to which each metabolite influenced growth yield in minimal medium. Importantly, *C. difficile* is auxotrophic for specific amino acids (e.g. proline; Fig 1C) that it is also able to catabolize through Stickland fermentation [79], so the diluting background medium must be supplemented with small concentrations of these metabolites. As such, the values are reported as the ratio of the final optical density for growth with the given metabolite versus low levels of growth observed in the background medium alone. Despite this calculation not being a direct comparison of utilization capability as in traditional Biolog analyses [80], it provides insight into an organism's metabolic preferences. We similarly calculated the influence of each metabolite on the optimal biomass flux at quasi-steady state of each model provided with the same background media conditions as the Biolog analysis (Fig. 2A). Across all of the 116 total metabolites that were in both the *in vitro* screen as well as the *C. difficile* GENREs, we identified significant predictive correlations in the amount of growth enhancement for iCdG698 (p-value < 0.001) and iCdR700 (p-value = 0.002) (Fig. 2B & 2C). This relationship was even more pronounced for carbohydrates and amino acids, primary carbon sources for *C. difficile* (Fig. S2). When these predictions were reduced to binary interpretations of either enhancement or non-enhancement of growth, we found that iCdG698 predicted 92.8% and iCdR700 predicted 96.6% true-positive enhancement calls (Fig. 2D). Importantly, this metric is the most valuable measure in this instance as it

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

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

Following GENRE validation, we sought to qualify the ability of each GENRE to predict *in situ* metabolic phenotypes across diverse experimental settings. As previously stated, GENREs have provided powerful platforms for the integration of transcriptomic data, creating greater context for the shifts observed between conditions and capturing the potential influence of pathways not obviously connected [81]. With this application in mind, we chose to generate context-specific models for both *in vitro* and *in vivo* experimental conditions characterized with RNA-Seq analysis utilizing a recently published unsupervised transcriptomic data integration method [82]. Briefly, this approach calculates the most cost-efficient usage of the metabolic network in order to achieve growth given the pathway investments indicated by the transcriptomic data. This process is in line with the concept that natural selection generally selects against wasteful production of cellular machinery and affords the ability to make much more fine-scale predictions of metabolic changes that *C. difficile* undergoes as it activates pathogenicity. The resultant patterns also reveal central elements within context-specific metabolism that could lead to targeted strategies for intentional downregulation of virulence factors through metabolic circuitry.

A recent study determined that phase variation, a reversible mechanism employed by many bacterial pathogens to generate phenotypic heterogeneity and maximize overall fitness of the population, also occurs in *C. difficile* str. R20291 and influences virulence expression [83]. One aspect of this phase variation manifests as a rough or smooth-edged colony morphology on solid agar; the morphologies can be propagated via subculture and are associated with distinct motility behaviors and altered virulence [84]. The colony morphology variants are generated through the phase variable (on/off) expression of the *cmrRST* genes. With this in mind, we sequenced transcriptomes from experimentally grown rough and smooth phase variants of *C. difficile* str. R20291 grown on solid BHI rich medium for 48 hours. Utilizing these data, we generated context-specific versions of iCdR700 in simulated rich media conditions. It has been previously shown that mutation of *cmr*-family genes does not significantly alter growth rate *in vitro* [84]. Growth simulation results predicted no

significant difference in optimal biomass flux values between phase variants (Fig. 3A), which agrees with previously published experimental growth rate measurements for *C. difficile* [37]. We then calculated essential genes in each variant model similar to the earlier analysis which identified 81 core genes essential in both contexts (Table S4), another 13 genes essential to growth for both variants, and 5 genes that were conditionally essential between the morphologies in BHI rich medium (Fig. 3C). The conditionally essential gene set restricted to the smooth variant included an N-acetylglucosamine PTS system as well as pyruvate kinase, which mediates the last step of glycolysis and a bulk of the ATP generation. Notably, at the transcriptional level, reads mapped to pyruvate kinase were detected at nearly identical levels between the rough and smooth isolates (Table S4). These results indicate that glycolytic enzymes may be more active in the smooth colony variants. The essentiality of N-acetylglucosamine transport in the context-specific model for the smooth phase was of interest as this variant has been previously shown to generate biofilms [84], in which N-acetylglucosamine is often a component [85]. We found that predicted exchange efflux of N-acetylglucosamine in the smooth variant was significantly greater than in rough (Fig. S3C). Conversely, in the rough context-specific model were multiple essential genes involved in Stickland fermentation (Fig. 3B). As with the pyruvate kinase gene, similar levels of transcription for these genes were also observed between smooth and rough variants (Table S4). These data were indicative of a potential trade-off between glycolysis and amino acid (Stickland) fermentation between smooth and rough phases respectively. In addition to genes that were critical for growth, we also identified those that were only required to achieve high growth yields in each context. This gene set included additional carbohydrate transporters in the smooth variant and multiple amino acid transporters in the rough variant (Table S4), further supporting differential utilization of glycolysis and Stickland fermentation across phases with highly dissimilar flux distributions of core metabolic pathways (Fig. S3), in spite of largely similar optimal growth rates (Fig. 3A).

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

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

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

Given laboratory media conditions (as used in the results described above) are much more easily defined, we also wanted to examine GENRE performance and prediction quality under more complex *in vivo* infection conditions. Another previously published study assessed the differential transcriptional activity of *C. difficile* str. 630 in the gut during infection in a mouse model pretreated with either streptomycin or clindamycin to induce sensitivity to colonization. These distinct treatments have different impacts on the structure of the gut microbiota [87] and allow for identical levels of pathogen colonization and vegetative cell load in the cecum. However, these different treatments result in highly dissimilar levels of sporulation (another phenotype linked to *C. difficile* virulence) where streptomycin is associated with undetectable spore CFUs and clindamycin with significantly higher levels [88]. The authors of this study also detected no significant difference in toxin activity

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

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

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

We also identified N-acetylneuraminate (NEu5Ac) as highly utilized in the lower sporulation context, a host-derived glycolysis substrate that *C. difficile* readily uses as a carbon source for growth [7]. This consumption was supported in the metabolomics screen where concentrations of this metabolite were significantly decreased following infection only in the lower spore condition (Fig. 4F). In agreement, this analysis also predicted that frequent byproducts of carbohydrate fermentation, acetate and formate, were more abundant in the lower sporulation context (Fig. 4D). Alternatively, both N-acetylglucosamine (GlcNAc) and N-acetylmannosamine (ManNAc) were predicted to be secreted at much larger rates from the low spore context-specific model (Fig. 4D, S5A, & S5B). Interestingly, these metabolites are integral components of biofilms [94], and *C. difficile* has been previously shown to generate these structures under certain circumstances [85]. We

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

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

DISCUSSION

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

between *C. difficile* virulence and metabolism. Furthermore, as much of bacterial pathogenicity is now being attributed to shifts in metabolism the analyses described here may provide large benefits to the identification of possible treatment targets in *C. difficile* and other recalcitrant pathogens [97]. In the current study, we develop and validate two highly-curated genome-scale metabolic network reconstructions for a well-described laboratory strain (str. 630) in addition to a more recently characterized hyper-virulent strain (str. R20291) of *C. difficile*. Both iCdG698 (str. 630) and iCdR700 (str. R20291) draw from numerous molecular and metabolic studies of *C. difficile* and Clostridial metabolism in order to accurately incorporate a large array of metabolic subsystems known to be present across strains of the pathogen. We further improved the quality of the models through careful curation of core metabolic strategies, including amino acid and carbohydrate fermentation, to ensure growth in all major defined growth conditions for *C. difficile*.

After the curation process was complete, we found a high degree of agreement between model predictions and validating experimental datasets. Both iCdG698 and iCdR700 indicate that the respective strains are able to catabolize amino acids as the sole carbon source through Stickland fermentation and require only those metabolites present in the experimentally determined minimal media to achieve growth. Additionally, close correlations of *in silic*o predictions with both gene essentiality and carbon source utilization screens supported that the GENREs accurately recapitulate *C. difficile* physiology and reconcile some previous inconsistencies in *C. difficile* metabolism literature. Following contextualization using *in situ* transcriptomic data, both GENREs were also able to demonstrate established complex metabolic phenotypes for both laboratory and infection conditions. These analyses collectively indicated a shift away from glycolytic metabolism, and toward amino acid fermentation, during periods of increased pathogenicity. These findings could lay the groundwork for novel approaches to curbing the expression of virulence factors by influencing environmental conditions to favor certain forms of metabolism over others. *In vivo* context-specific gene essentiality also predicted proline racemase to be critical for growth during infection, yet it was previously found to be dispensable in an animal model using a forward genetic screen [96]. While this result may indicate necessary future curation, it may also be attributable to the specific conditions of that infection and may vary across distinct host gut environments, leading to possible implications in personalized medicine.

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

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

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

MATERIALS & METHODS

C. difficile GENRE Construction

We utilized PATRIC reference genomes from *Clostridioides difficile* str. 630 and *Clostridioides difficile* str. R20291 as initial reconstruction templates for the automated ModelSEED pipeline [39,98,99]. The automated ModelSEED draft reconstruction was converted utilizing the Mackinac pipeline (https://github.com/mmundy42/mackinac) into a form more compatible with the COBRA toolbox [100]. Upon removal of GENRE components lacking genetic evidence (i.e. gap-filled), extensive manual curation was performed in accordance with best practices agreed upon by the community [101]. We subsequently performed ensemble gap-filling as previously described, utilizing a stoichiometrically consistent anaerobic, Gram-positive ModelSEED universal reaction collection curated for this purpose and available alongside code associated with this study. Next, we corrected reaction inconsistencies and incorrect physiological properties (e.g. ensured free water diffusion across compartments). Final transport reactions were then validated with TransportDB [102]. All formulas are mass and charged balanced at an assumed pH of 7.0 using the ModelSEED database in order to maintain a consistent and supported namespace to augment GENRE interpretability and future curation efforts. We then collected annotation data for all model components (genes, reactions, and metabolites) from SEED [101,103], KEGG [104], PATRIC, RefSeq [105], EMBL [106], and BiGG [107] databases and integrated it into the annotation field dictionary now supported in the most recent SBML version [108]. Complete MEMOTE quality reports for both *C. difficile* GENREs are also available in the GitHub repository associated with this study, and full pipelines for model generation are explicitly outlined in Jupyter notebooks hosted there as well. Download of either iCdG698 (str. 630) or iCdR700 (str. R20291) is possible from the studies' Github or the Papin lab website (https://bme.virginia.edu/csbl/Downloads1.html).

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

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

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

C. difficile str. R20291 growth was maintained in an anaerobic environment of 85% N_2, 5% CO_2, and 10% H_2. The strain was grown on BHI-agar (37 g/L Bacto brain heart infusion, 1.5% agar) medium at 37 °C for 48 hours to obtain isolated colonies. Rough and smooth colonies were chosen for propagation on BHI-agar to ensure colony morphology maintenance [84]. Basal Defined Medium (BDM) was formulated as previously published [47] with the addition of 1.5% agar for plates, and incubated for 48 hours at 37 °C to generate isolated colonies. Microscopy images were taken on an EVOS XL Core Cell Imaging System at 4x magnification.

RNA isolation, and transcriptome sequencing

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

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

Genomic and transcriptomic data processing

Alignment of *C. difficile* str. 630 and str. R20291 peptide sequences was performed using bidirectional BLASTp. RNA-Seq reads were first quality-trimmed with Sickle with a cutoff ≧Q30 [112]. Mapping curated reads to the respective *C. difficile* genome was then performed with Bowtie2 [113]. MarkDuplicates then removed optica/PCR duplicates (broadinstitute.github.io/picard/), and mappings were converted to idxstats format using SAMtools [114]. Abundances were then normalized to both read and target lengths. Transcriptomic integration and context-specific model generation were performed with RIPTiDe and maxfit_contextualize() on the default settings [82].

Statistical Methods

All statistical analysis was performed in R v3.2.0. Non-metric multidimensional scaling of Bray-Curtis dissimilarity and perMANOVA analyses accomplished using the vegan R package [115]. Significant differences for single reaction flux distributions and metabolite concentrations were determined by Wilcoxon signed-rank test.. Supervised machine-learning was accomplished with the implementation of AUC-Random Forest also in R [116]. All code associated with this study is available in the study-associated GitHub repository.

Data availability

Genomic and proteomic data for the strains *Clostridioides difficile* str. 630 (PATRIC ref. 272563.8) and *Clostridioides difficile* str. R20291 (PATRIC ref. 645463.3) was downloaded from the PATRIC database [98]. Transcriptomic data was downloaded in raw FASTQ format from the NCBI Sequence Read Archive (PRJNA415307 and PRJNA354635) and Gene Expression Omnibus (GSE158225) . Github repository for this

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

Author Contributions

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

JLL - Conceptualization. Data generation. Editing manuscript.

DAP - Data analysis. Editing manuscript.

EMG - Conceptualization. Data generation. Editing manuscript.

KAW - Data generation. Editing manuscript.

MED - Data analysis. Editing manuscript.

WAP - Supervision. Editing manuscript.

RT - Supervision. Data generation. Editing manuscript.

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

Acknowledgements

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

FIGURE & TABLE LEGENDS

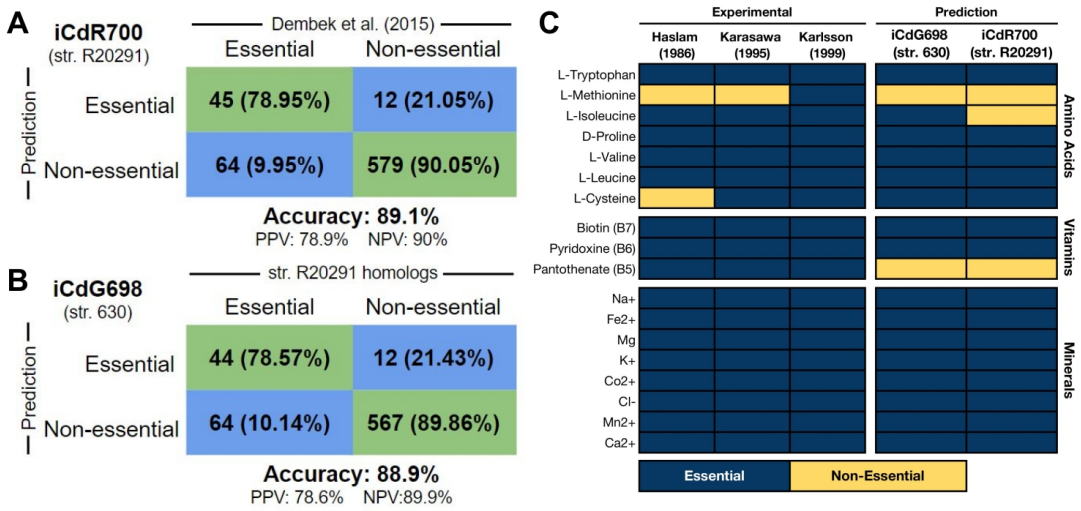


Figure 1) Gene and minimal growth substrate essentiality predictions closely match experimental results. (A) Gene essentiality results for iCdR700 (str. R20291) using the transposon mutagenesis screen results published in Dembek et al. 2015, and **(B)** gene essentiality for iCdG698 (str. 630) utilizing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

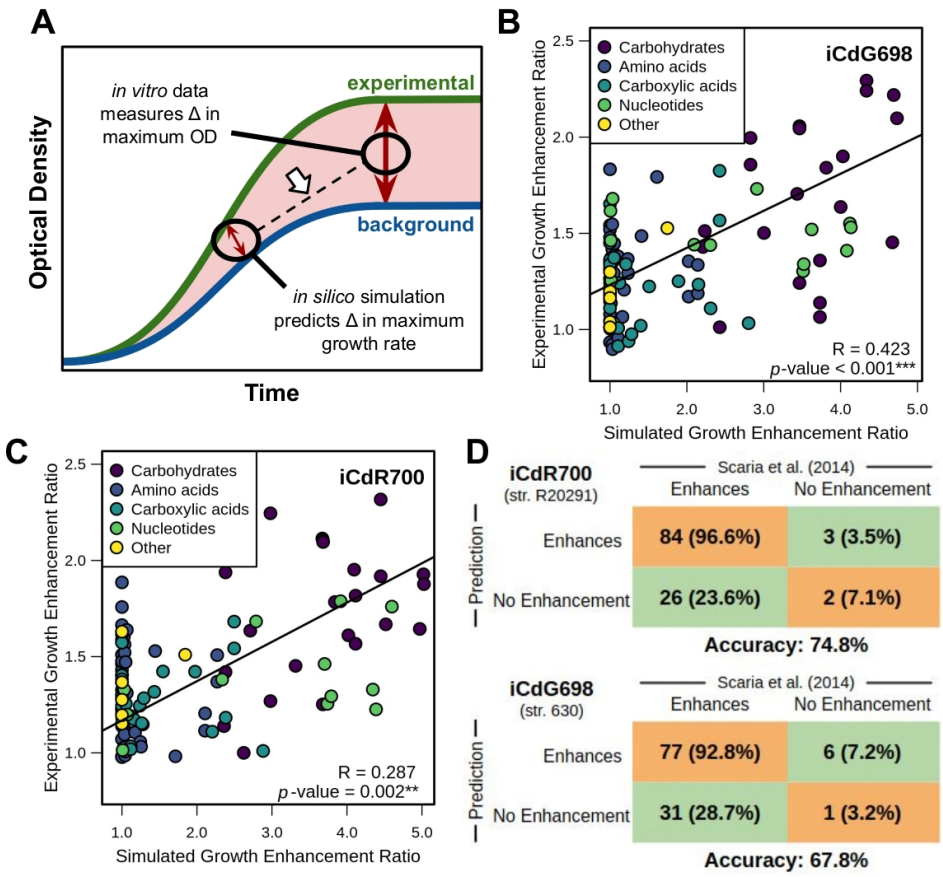


Figure 2) Carbon source growth enhancement predictions reflect laboratory measurements. Experimental analysis was performed for both str. 630 and str. R20291 in Scaria et al. 2014., and 115 metabolites were shared between the GENREs and the Biolog carbon source phenotypic screen. **(A)** Schematic of specific *in vitro* and *in silico* measurements being utilized. The arrow indicates the correlations made in subsequent panels. Ratios of overall *in vitro* growth enhancement by each metabolite were correlated with the corresponding results from growth simulations in the same media for **(B)** iCdG698 (str. 630) and **(C)** iCdR700 (str. R20291). Points are colored by their biochemical grouping, and significant relationships were determined by Spearman correlation. **(D)** Binary quantification for predictions in B & C respectively.

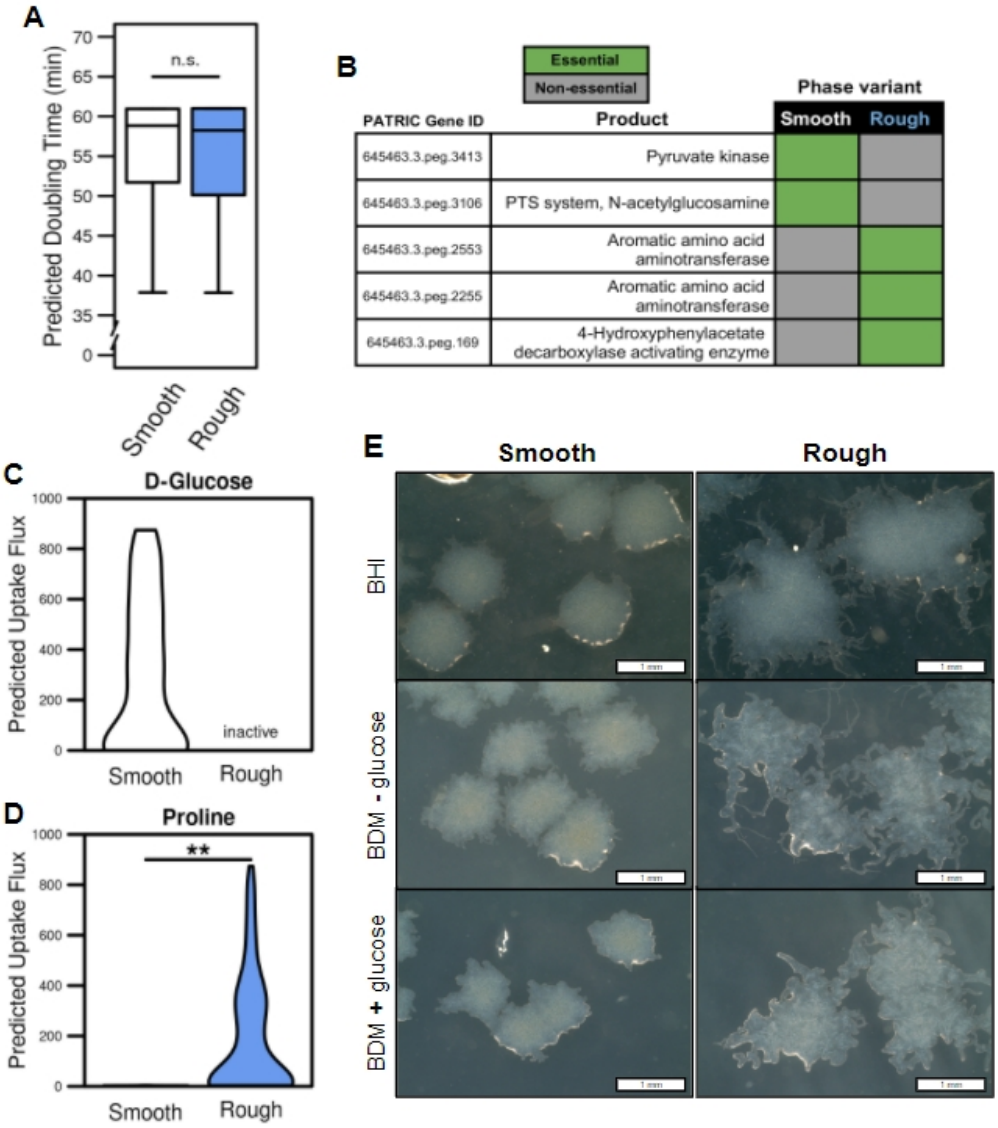


Figure 3) iCdR700 predicts concerted metabolic shifts during phase variation in str. R20291 grown *in vitro*. Transcriptomes were collected from rough or smooth colony morphology clones grown on BHI agar for 48 hours, and subsequently used to generate context-specific models. **(A)** Doubling times calculated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

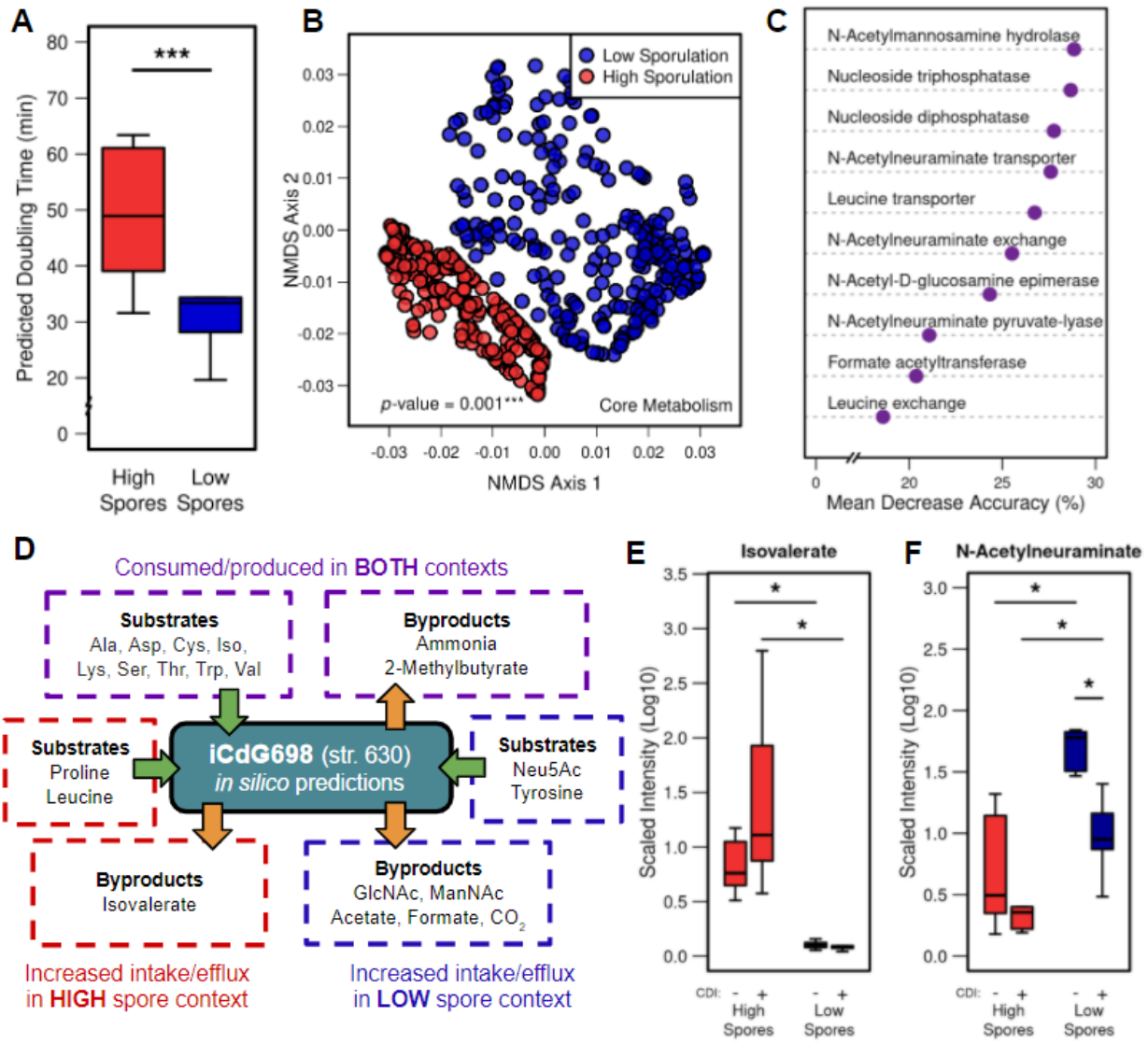


Figure 4) iCdG698 predicts distinct metabolic patterns associated with pathogenicity across *in vivo* str. 630 infections. Transcriptomic integration and predictions with iCdG698, 18 hours post-infection with str. 630 across infections with either high or low levels of sporulation were detected in the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

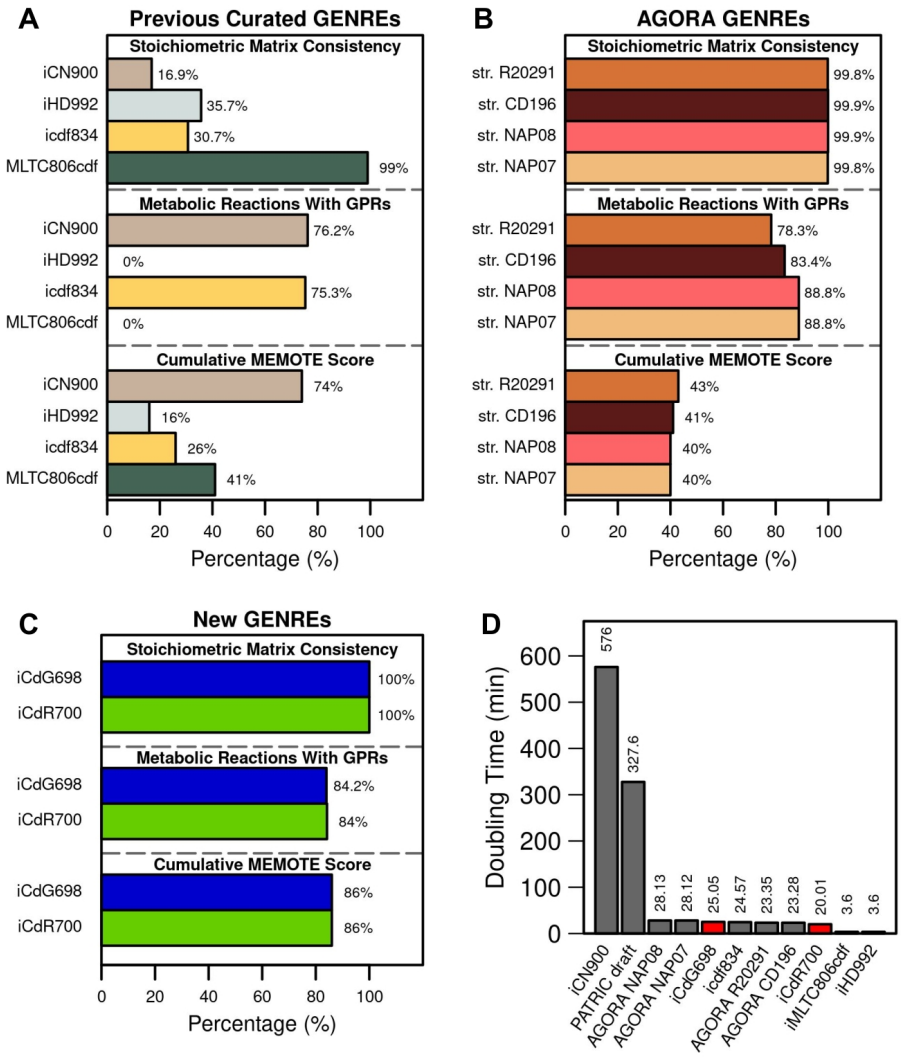


Figure S1) Selected quality metrics for *C. difficile* GENREs. Stoichiometric inconsistency describes consistent mass conservation across metabolic reactions. Assessing for metabolic reactions without gene-reaction rules (GPRs) excludes exchange reactions, transport reactions, and those reactions associated 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

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

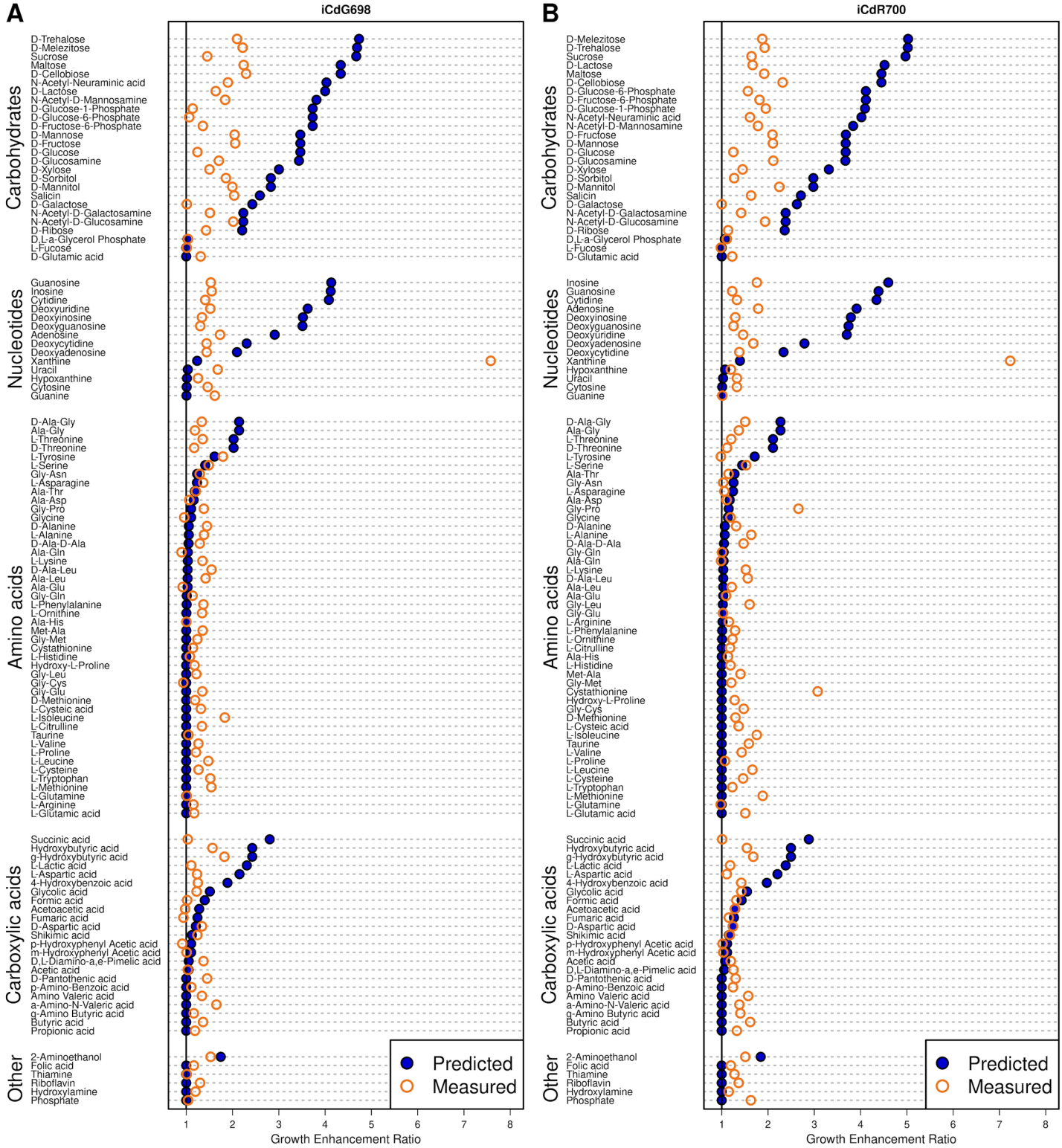


Figure S2) Specific shifts in simulated versus measured growth enhancement for each metabolite measured in the carbon source utilization screen. Metabolites are separated into groups by metabolite superfamily designation. Fold change for both *in vitro* and *in silico* measurements reflects growth enhancement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

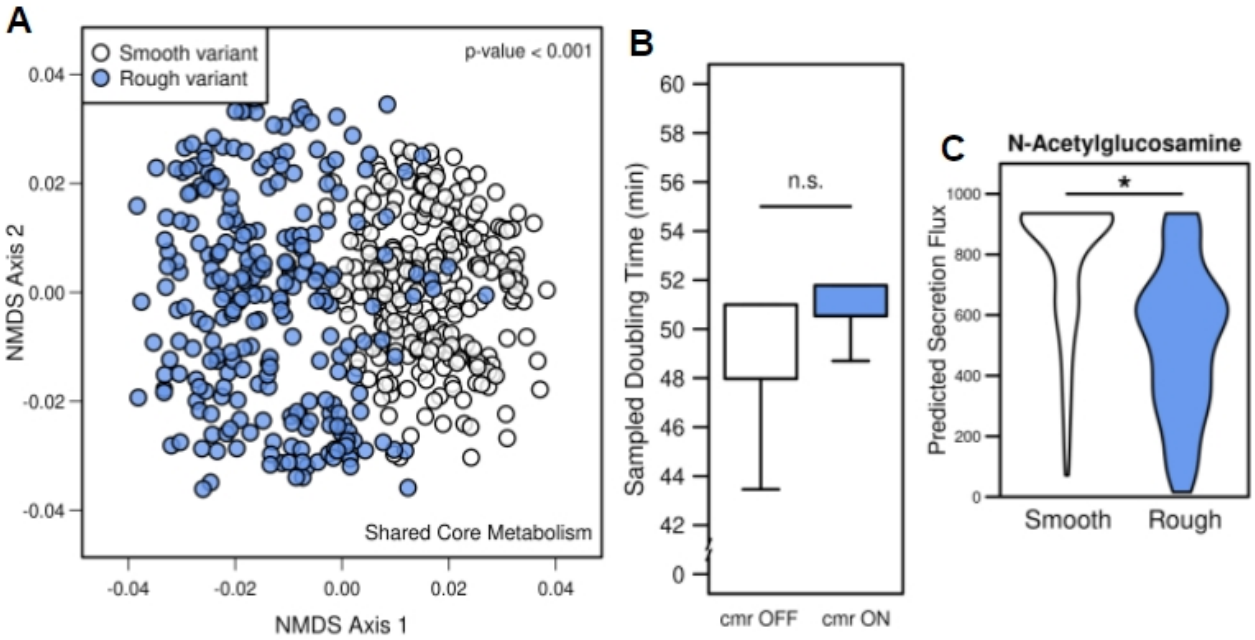


Figure S3) Change in growth simulations using *in vitro* transcriptomic data from naturally occurring or mutant phase-locked str. R20291 colony variants integrated into iCdR700. Rough versus Smooth variant transcriptomes integrated with RIPTiDe into iCdR700. **(A)** NMDS ordination of Bray-Curtis dissimilarities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

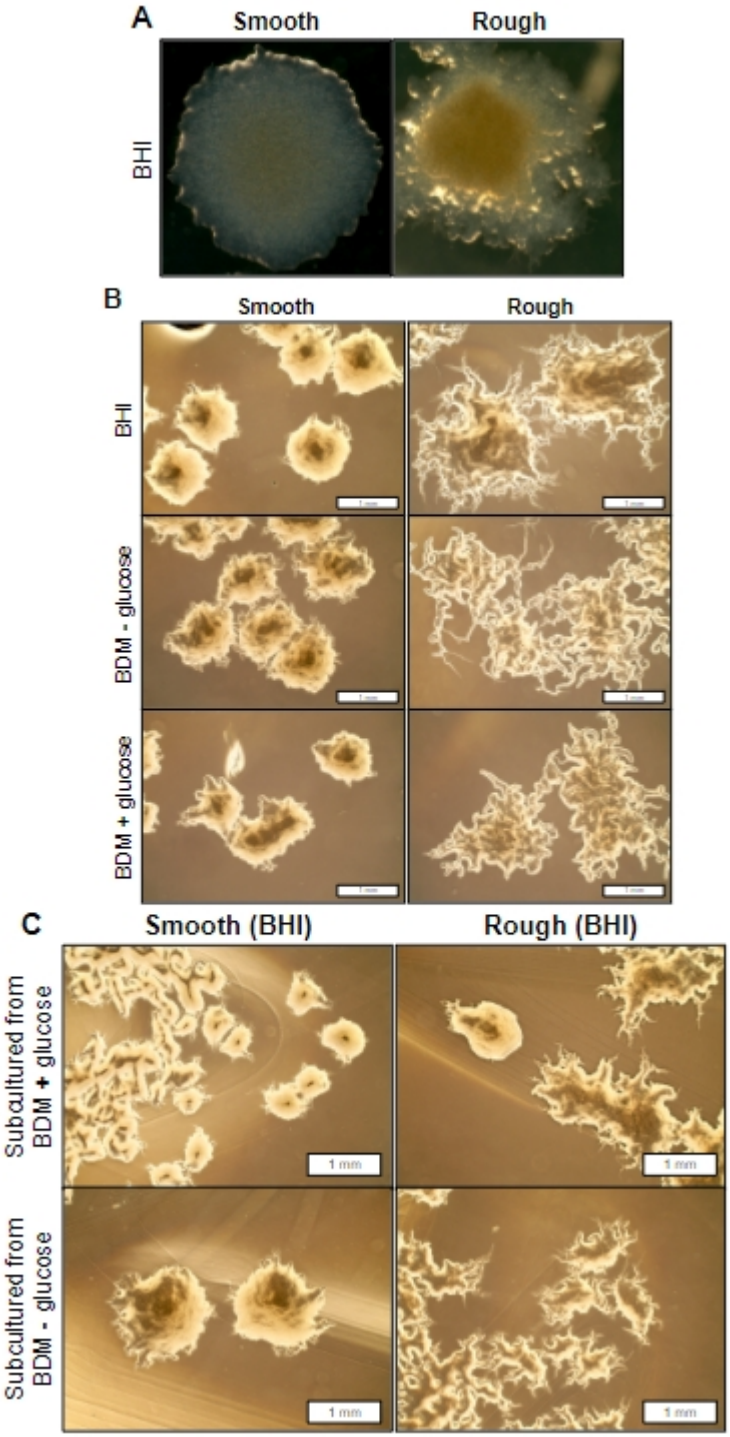


Figure S4) Additional microscopy of phase variant colony morphologies. (A) *C. difficile* str. R20291 phase variants progenitor colonies generated on solid BHIS agar following 48 hours of growth at 37° C under anaerobic conditions. These colonies were subcultured and utilized for all subsequent defined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

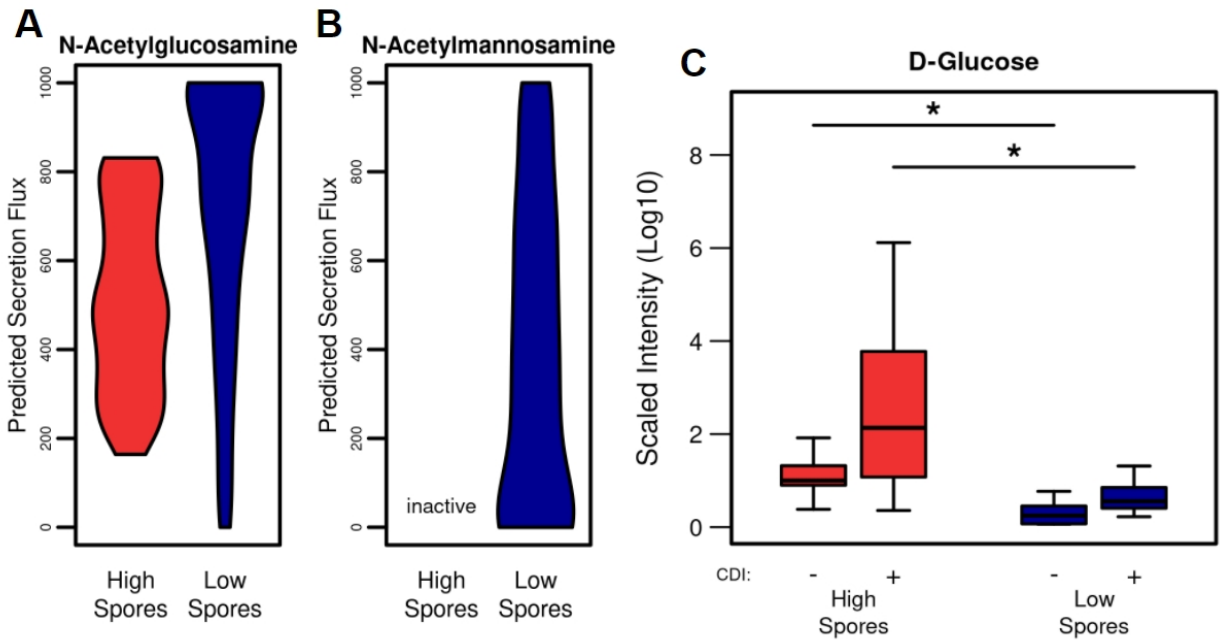


Figure S5) Context-specific growth simulation with iCdR698 predicts discordant carbohydrate-associated metabolism which agrees with *in vivo* measurements. (A) Exchange reaction flux associated with N-acetylglucosamine export for both high and low spore context-specific models (*p*-value = 0.067). Significant 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).

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

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

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

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

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

REFERENCES

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

11. (u.s.), C. F. D. C. A. P. & Centers for Disease Control and Prevention (U.S.). Antibiotic resistance threats in the United States, 2019. (2019) doi:10.15620/cdc:82532.

12. Merrigan, M. *et al.* Human hypervirulent *Clostridium difficile* strains exhibit increased sporulation as well as robust toxin production. *J. Bacteriol.* **192**, 4904–4911 (2010).

13. Hao, T. *et al.* The Genome-Scale Integrated Networks in Microorganisms. *Front. Microbiol.* **9**, 296 (2018).

14. Pacheco, A. R., Moel, M. & Segrè, D. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat. Commun.* **10**, 103 (2019).

15. Seif, Y. *et al.* Metabolic and genetic basis for auxotrophies in Gram-negative species. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 6264–6273 (2020).

16. Hadadi, N. *et al.* Mechanistic insights into bacterial metabolic reprogramming from omics-integrated genome-scale models. *NPJ Syst Biol Appl* **6**, 1 (2020).

17. Cesur, M. F., Siraj, B., Uddin, R., Durmuş, S. & Çakır, T. Network-Based Metabolism-Centered Screening of Potential Drug Targets in *Klebsiella pneumoniae* at Genome Scale. *Frontiers in Cellular and Infection Microbiology* vol. 9 (2020).

18. Jijakli, K. & Jensen, P. A. Metabolic Modeling of *Streptococcus mutans* Reveals Complex Nutrient Requirements of an Oral Pathogen. *mSystems* **4**, (2019).

19. Bosi, E. *et al.* Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3801–9 (2016).

20. Larocque, M., Chénard, T. & Najmanovich, R. A curated *C. difficile* strain 630 metabolic network: prediction of essential targets and inhibitors. *BMC Syst. Biol.* **8**, 117 (2014).

21. Kashaf, S. S., Angione, C. & Lió, P. Making life difficult for *Clostridium difficile*: augmenting the pathogen's metabolic model with transcriptomic and codon usage data for better therapeutic target characterization. *BMC Systems Biology* vol. 11 (2017).

22. Dannheim, H., Will, S. E., Schomburg, D. & Neumann-Schaal, M. *Clostridioides difficile* 630Δerm in silicoandin vivo- quantitative growth and extensive polysaccharide secretion. *FEBS Open Bio* vol. 7 602–615 (2017).

23. King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–22 (2016).

24. Monot, M. *et al.* Reannotation of the genome sequence of *Clostridium difficile* strain 630. *J. Med. Microbiol.* **60**, 1193–1199 (2011).

25. Stabler, R. A. *et al.* Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* **10**, R102 (2009).

26. Antunes, A. *et al.* Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. *Nucleic Acids Res.* **40**, 10701–10718 (2012).

27. Norsigian, C. J. *et al.* Systems biology analysis of the *Clostridioides difficile* core-genome contextualizes microenvironmental evolutionary pressures leading to genotypic and phenotypic divergence. *NPJ Syst Biol Appl* **6**, 31 (2020).

28. Dannheim, H. *et al.* Manual curation and reannotation of the genomes of *Clostridium difficile* 630Δerm and C. difficile 630. *Journal of Medical Microbiology* vol. 66 286–293 (2017).

29. Hussain, H. A., Roberts, A. P. & Mullany, P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Δerm) and demonstration that the conjugative transposon Tn916ΔE enters the genome of this strain at multiple sites. *Journal of Medical Microbiology* vol. 54 137–141 (2005).

30. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).

31. Gevorgyan, A., Poolman, M. G. & Fell, D. A. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics* **24**, 2245–2251 (2008).

32. Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B. & Lercher, M. J. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Comput. Biol.* **13**, e1005494 (2017).

33. Schellenberger, J., Lewis, N. E. & Palsson, B. Ø. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.* **100**, 544–553 (2011).

34. Ravikrishnan, A. & Raman, K. Critical assessment of genome-scale metabolic networks: the need for a unified standard. *Brief. Bioinform.* **16**, 1057–1068 (2015).

35. Lieven, C. *et al.* MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).

36. Oberhardt, M. A., Puchałka, J., dos Santos, V. A. P. M. & Papin, J. A. Reconciliation of Genome-Scale

Metabolic Reconstructions for Comparative Systems Analysis. *PLoS Computational Biology* vol. 7 e1001116 (2011).

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

49. Haslam, S. C. *et al.* Growth of *Clostridium difficile* and production of toxins A and B in complex and defined media. *J. Med. Microbiol.* **21**, 293–297 (1986).

50. Stiemsma, L. T., Turvey, S. E. & Finlay, B. B. An antibiotic-altered microbiota provides fuel for the enteric foe. *Cell research* vol. 24 5–6 (2014).

51. Olson, M. E., King, J. M., Yahr, T. L. & Horswill, A. R. Sialic acid catabolism in *Staphylococcus aureus*. *J. Bacteriol.* **195**, 1779–1788 (2013).

52. Engevik, M. A. *et al.* Human Clostridium difficile infection: altered mucus production and composition. *Am. J. Physiol. Gastrointest. Liver Physiol.* **308**, G510–24 (2015).

53. Nakamura, S., Nakashio, S., Yamakawa, K., Tanabe, N. & Nishida, S. Carbohydrate Fermentation by *Clostridium difficile*. *Microbiology and Immunology* vol. 26 107–111 (1982).

54. Kim, J., Darley, D. & Buckel, W. 2-Hydroxyisocaproyl-CoA dehydratase and its activator from *Clostridium difficile*. *FEBS J.* **272**, 550–561 (2005).

55. Selmer, T. & Andrei, P. I. p-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycyl radical enzyme catalysing the formation of p-cresol. *Eur. J. Biochem.* **268**, 1363–1372 (2001).

56. Bouillaut, L., Self, W. T. & Sonenshein, A. L. Proline-dependent regulation of *Clostridium difficile* Stickland metabolism. *J. Bacteriol.* **195**, 844–854 (2013).

57. Jackson, S., Calos, M., Myers, A. & Self, W. T. Analysis of proline reduction in the nosocomial pathogen *Clostridium difficile*. *J. Bacteriol.* **188**, 8487–8495 (2006).

58. Kim, J., Darley, D., Selmer, T. & Buckel, W. Characterization of (R)-2-Hydroxyisocaproate Dehydrogenase and a Family III Coenzyme A Transferase Involved in Reduction of L-Leucine to Isocaproate by *Clostridium difficile*. *Applied and Environmental Microbiology* vol. 72 6062–6069 (2006).

59. Britz, M. L. & Wilkinson, R. G. Leucine dissimilation to isovaleric and isocaproic acids by cell suspensions of amino acid fermenting anaerobes: the Stickland reaction revisited. *Can. J. Microbiol.* **28**, 291–300 (1982).

60. Kim, J., Hetzel, M., Boiangiu, C. D. & Buckel, W. Dehydration of (R)-2-hydroxyacyl-CoA to enoyl-CoA in the fermentation of α-amino acids by anaerobic bacteria. *FEMS Microbiology Reviews* vol. 28 455–468 (2004).

61. Yu, L., Blaser, M., Andrei, P. I., Pierik, A. J. & Selmer, T. 4-Hydroxyphenylacetate decarboxylases:

properties of a novel subclass of glycyl radical enzyme systems. *Biochemistry* **45**, 9584–9592 (2006).

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

76. Dubois, T. *et al.* Control of *Clostridium difficile* Physiopathology in Response to Cysteine Availability. *Infect. Immun.* **84**, 2389–2405 (2016).

77. Ikeda, D. *et al.* Effect of Isoleucine on Toxin Production by *Clostridium difficile* in a Defined Medium. *Zentralblatt für Bakteriologie* vol. 287 375–386 (1998).

78. Scaria, J. *et al.* Comparative nutritional and chemical phenome of *Clostridium difficile* isolates determined using phenotype microarrays. *Int. J. Infect. Dis.* **27**, 20–25 (2014).

79. Battaglioli, E. J. *et al. Clostridioides difficile* uses amino acids associated with gut microbial dysbiosis in a subset of patients with diarrhea. *Science Translational Medicine* vol. 10 eaam7019 (2018).

80. Oh, Y.-K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **282**, 28791–28799 (2007).

81. Blazier, A. S. & Papin, J. A. Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.* **3**, 299 (2012).

82. Jenior, M. L., Moutinho, T. J., Jr, Dougherty, B. V. & Papin, J. A. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput. Biol.* **16**, e1007099 (2020).

83. Anjuwon-Foster, B. R. & Tamayo, R. A genetic switch controls the production of flagella and toxins in *Clostridium difficile*. *PLoS Genet.* **13**, e1006701 (2017).

84. Garrett, E. M. *et al.* Phase variation of a signal transduction system controls *Clostridioides difficile* colony morphology, motility, and virulence. *PLoS Biol.* **17**, e3000379 (2019).

85. Dubois, T. *et al.* A microbiota-generated bile salt induces biofilm formation in *Clostridium difficile*. *NPJ Biofilms Microbiomes* **5**, 14 (2019).

86. Passmore, I. J. *et al.* Para-cresol production by *Clostridium difficile* affects microbial diversity and membrane integrity of Gram-negative bacteria. *PLoS Pathog.* **14**, e1007191 (2018).

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

88. Jenior, M. L., Leslie, J. L., Young, V. B. & Schloss, P. D. *Clostridium difficile* colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. doi:10.1101/092304.

89. Wilson, K. H. & Perini, F. Role of competition for nutrients in suppression of *Clostridium difficile* by the colonic microflora. *Infect. Immun.* **56**, 2610–2614 (1988).

90. Purcell, E. B., McKee, R. W., McBride, S. M., Waters, C. M. & Tamayo, R. Cyclic diguanylate inversely regulates motility and aggregation in *Clostridium difficile*. *J. Bacteriol.* **194**, 3307–3316 (2012).

91. Bordeleau, E., Fortier, L.-C., Malouin, F. & Burrus, V. c-di-GMP turn-over in *Clostridium difficile* is controlled by a plethora of diguanylate cyclases and phosphodiesterases. *PLoS Genet.* **7**, e1002039 (2011).

92. Antunes, A., Martin-Verstraete, I. & Dupuy, B. CcpA-mediated repression of *Clostridium difficile* toxin gene expression. *Mol. Microbiol.* **79**, 882–899 (2011).

93. Sivsammye, G. & Sims, H. V. Presumptive identification of *Clostridium difficile* by detection of p-cresol in prepared peptone yeast glucose broth supplemented with p-hydroxyphenylacetic acid. *Journal of Clinical Microbiology* vol. 28 1851–1853 (1990).

94. Zhang, B. & Powers, R. Analysis of bacterial biofilms using NMR-based metabolomics. *Future Med. Chem.* **4**, 1273–1306 (2012).

95. Maegawa, T. *et al.* Linkage between toxin production and purine biosynthesis in *Clostridium difficile*. *J. Med. Microbiol.* **51**, 34–41 (2002).

96. Wu, X. & Hurdle, J. G. The *Clostridium difficile* proline racemase is not essential for early logarithmic growth and infection. *Canadian Journal of Microbiology* vol. 60 251–254 (2014).

97. Raškevičius, V. *et al.* Genome scale metabolic models as tools for drug design and personalized medicine. *PLoS One* **13**, e0190636 (2018).

98. Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42**, D581–91 (2014).

99. Faria, J. P., Rocha, M., Rocha, I. & Henry, C. S. Methods for automated genome-scale metabolic model reconstruction. *Biochem. Soc. Trans.* **46**, 931–936 (2018).

100. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).

101. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 121 (2019).

102. Ren, Q., Chen, K. & Paulsen, I. T. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* **35**, D274–9 (2007).

103. Seaver, S. M. D. *et al.* The ModelSEED Database for the integration of metabolic annotations and the reconstruction, comparison, and analysis of metabolic models for plants, fungi, and microbes. doi:10.1101/2020.03.31.018663.

104. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* vol. 28 27–30 (2000).

105. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–5 (2007).

106. Baker, W. The EMBL Nucleotide Sequence Database. *Nucleic Acids Research* vol. 28 19–23 (2000).

107. Norsigian, C. J. *et al.* BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* **48**, D402–D406 (2020).

108. Hucka, M. *et al.* The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. *J. Integr. Bioinform.* **16**, (2019).

109. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).

110. Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489 (2010).

111. Keaty, T. C. & Jensen, P. A. gapsplit: Efficient random sampling for non-convex constraint-based models. doi:10.1101/652917.

112. Website. Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

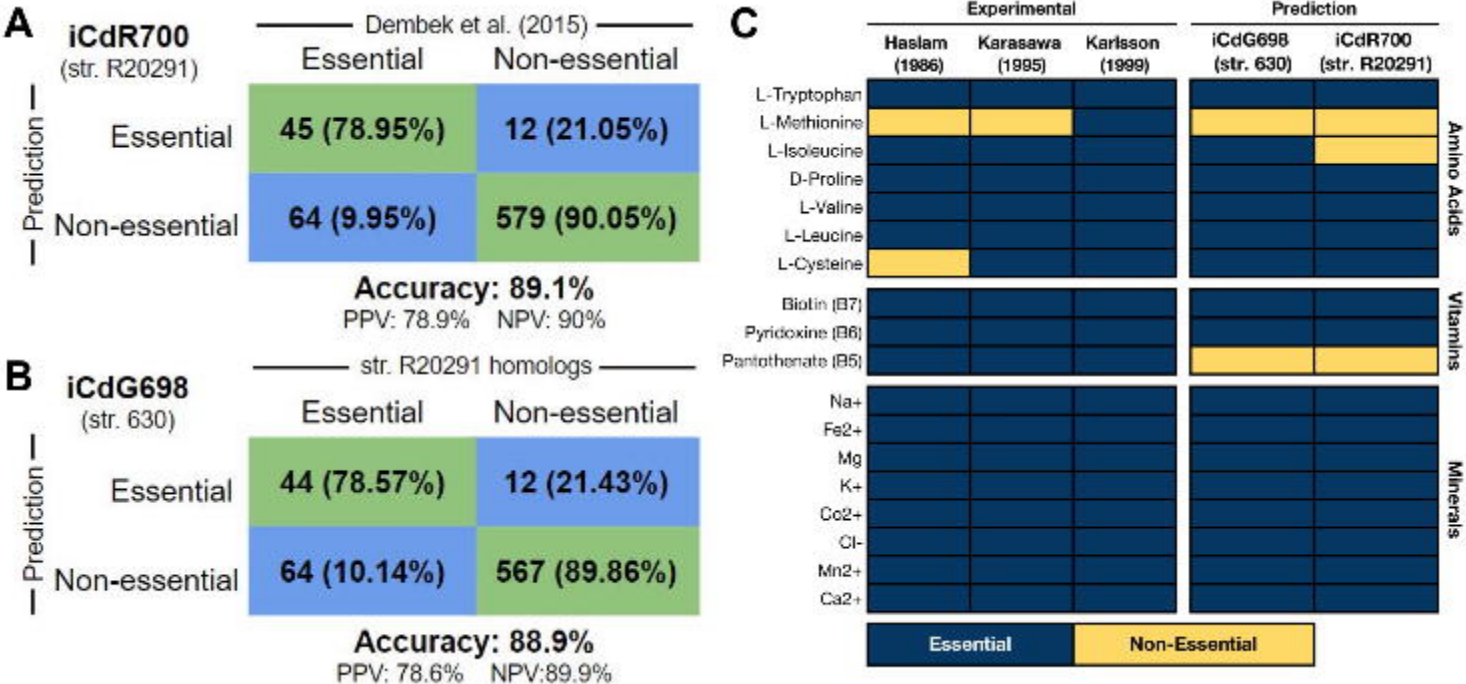
113. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

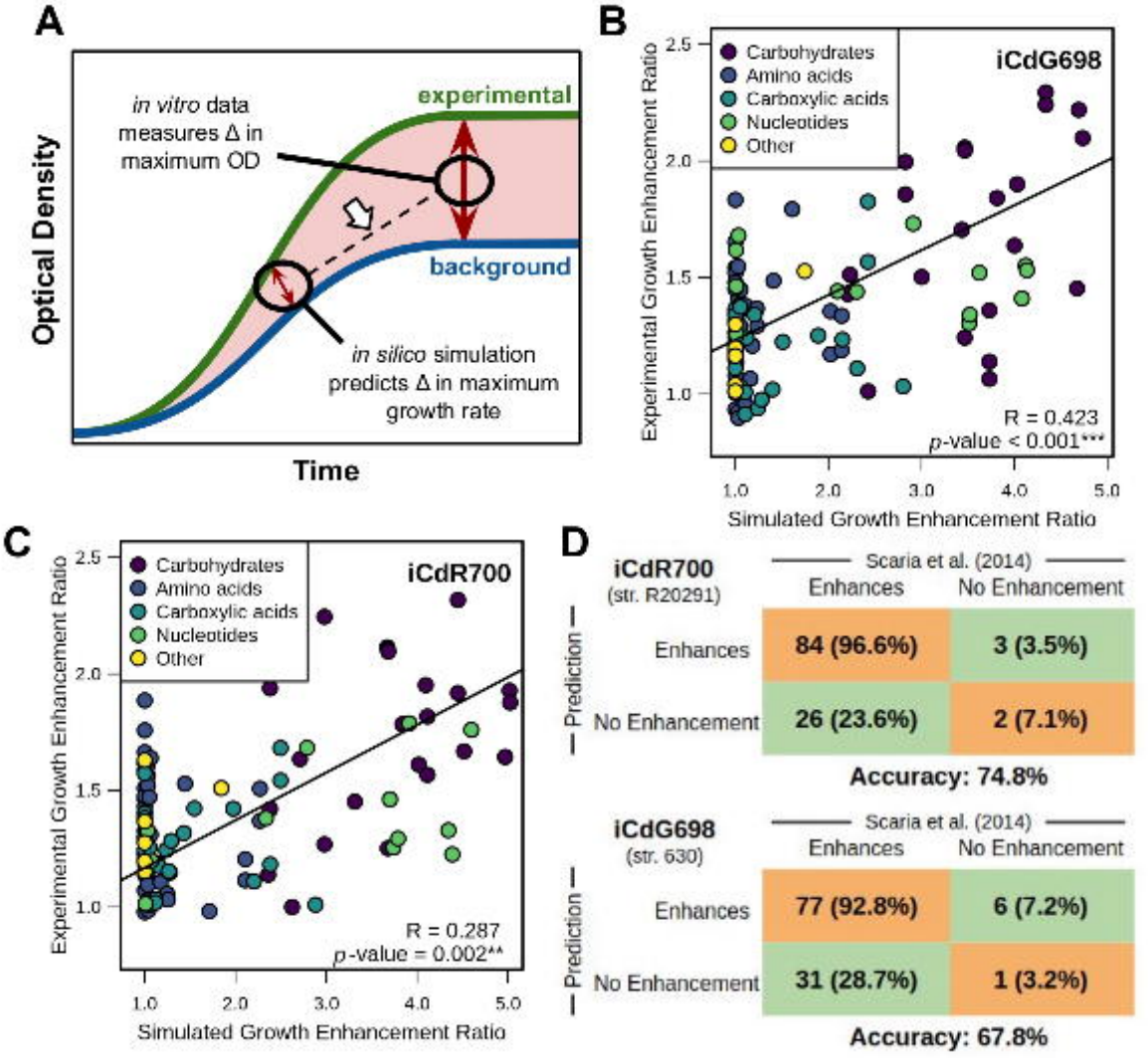
114. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

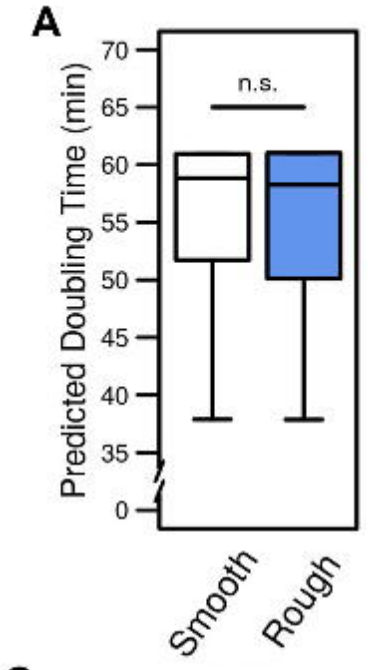
115. Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* vol. 14 927 (2003).

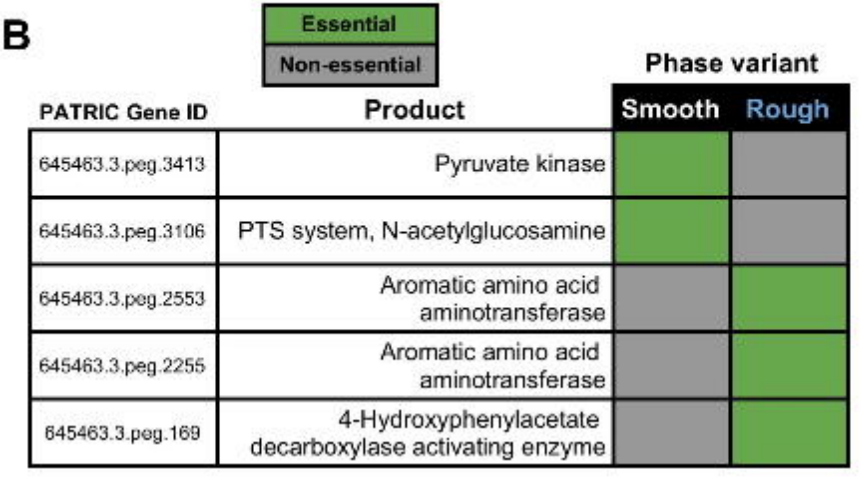
116. Janitza, S., Strobl, C. & Boulesteix, A.-L. An AUC-based permutation variable importance measure for

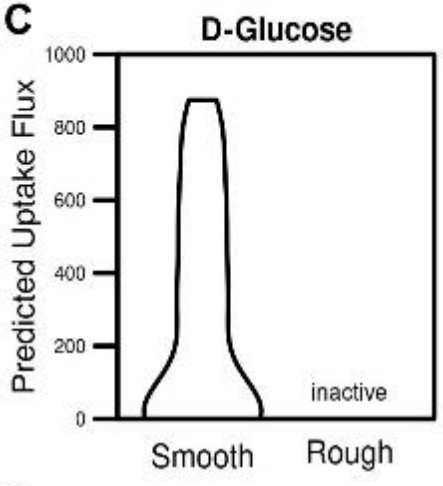
random forests. *BMC Bioinformatics* **14**, 119 (2013).

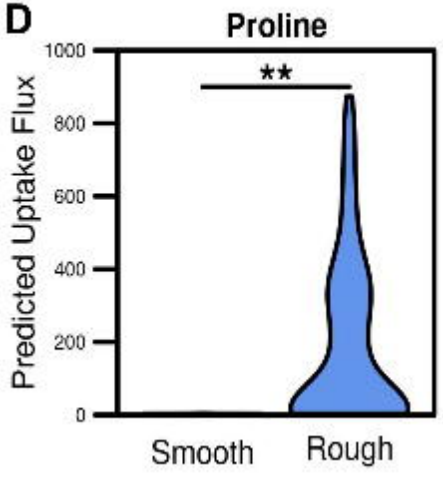

A
iCdR700
(str. R20291)
Dembek et al. (2015)
Essential
Non-essential
Prediction
Essential
Non-essential
45 (78.95%)
12 (21.05%)
64 (9.95%)
579 (90.05%)
Accuracy: 89.1%
PPV: 78.9% NPV: 90%
B
iCdG698
(str. 630)
str. R20291 homologs
Essential
Non-essential
Prediction
Essential
Non-essential
44 (78.57%)
12 (21.43%)
64 (10.14%)
567 (89.86%)
Accuracy: 88.9%
PPV: 78.6% NPV:89.9%
C
Experimental
Prediction
Haslam (1986)
Karasawa (1995)
Karlsson (1999)
iCdG698 (str. 630)
iCdR700 (str. R20291)
L-Tryptophan
L-Methionine
L-Isoleucine
D-Proline
L-Valine
L-Leucine
L-Cysteine
Amino Acids
Biotin (B7)
Pyridoxine (B6)
Pantothenate (B5)
Vitamins
Na+
Fe2+
Mg
K+
Co2+
Cl-
Mn2+
Ca2+
Minerals
Essential
Non-Essential

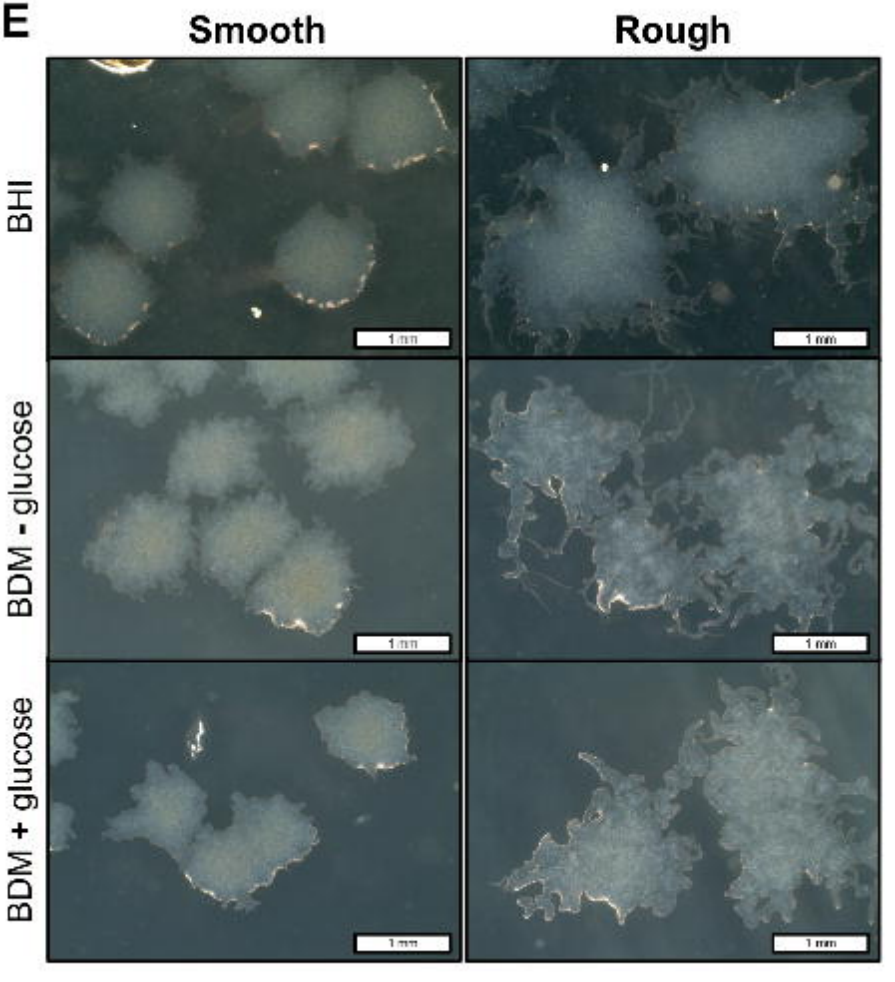

A
Optical Density
Time
experimental
background
in vitro data measures Δ in maximum OD
in silico simulation predicts Δ in maximum growth rate
B
iCdG698
Carbohydrates
Amino acids
Carboxylic acids
Nucleotides
Other
Experimental Growth Enhancement Ratio
Simulated Growth Enhancement Ratio
R = 0.423
p-value < 0.001***
C
iCdR700
Carbohydrates
Amino acids
Carboxylic acids
Nucleotides
Other
Experimental Growth Enhancement Ratio
Simulated Growth Enhancement Ratio
R = 0.287
p-value = 0.002**
D
iCdR700 (str. R20291)
Scaria et al. (2014)
Enhances
No Enhancement
Prediction
Enhances
No Enhancement
84 (96.6%)
3 (3.5%)
26 (23.6%)
2 (7.1%)
Accuracy: 74.8%
iCdG698 (str. 630)
Scaria et al. (2014)
Enhances
No Enhancement
Prediction
Enhances
No Enhancement
77 (92.8%)
6 (7.2%)
31 (28.7%)
1 (3.2%)
Accuracy: 67.8%

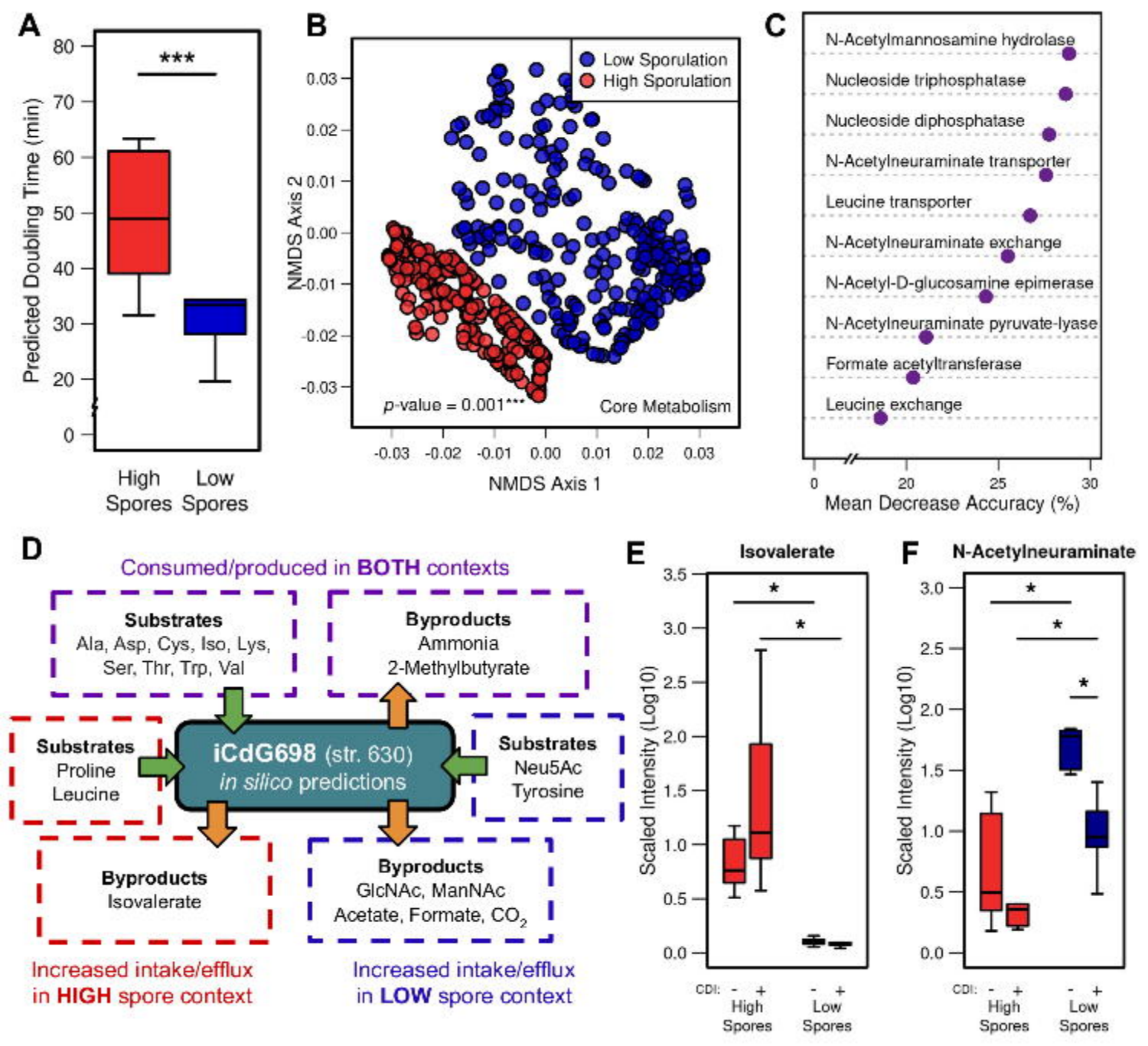

A
Predicted Doubling Time (min)
70
65
60
55
50
45
40
35
0
n.s.
Smooth
Rough


B
Essential
Non-essential
Phase variant
PATRIC Gene ID
Product
Smooth
Rough
645463.3.peg.3413
Pyruvate kinase
645463.3.peg.3106
PTS system, N-acetylglucosamine
645463.3.peg.2553
Aromatic amino acid aminotransferase
645463.3.peg.2255
Aromatic amino acid aminotransferase
645463.3.peg.169
4-Hydroxyphenylacetate decarboxylase activating enzyme


C
D-Glucose
Predicted Uptake Flux
1000
800
600
400
200
0
inactive
Smooth
Rough


D
Proline
Predicted Uptake Flux
1000
800
600
400
200
0
**
Smooth
Rough


E
Smooth
Rough
BHI
BDM - glucose
BDM + glucose
1 mm


A
Predicted Doubling Time (min)
***
High Spores
Low Spores
B
Low Sporulation
High Sporulation
NMDS Axis 2
NMDS Axis 1
p-value = 0.001***
Core Metabolism
C
N-Acetylmannosamine hydrolase
Nucleoside triphosphatase
Nucleoside diphosphatase
N-Acetylneuraminate transporter
Leucine transporter
N-Acetylneuraminate exchange
N-Acetyl-D-glucosamine epimerase
N-Acetylneuraminate pyruvate-lyase
Formate acetyltransferase
Leucine exchange
Mean Decrease Accuracy (%)
D
Consumed/produced in BOTH contexts
Substrates
Ala, Asp, Cys, Iso, Lys, Ser, Thr, Trp, Val
Byproducts
Ammonia
2-Methylbutyrate
Substrates
Proline
Leucine
iCdG698 (str. 630)
in silico predictions
Substrates
Neu5Ac
Tyrosine
Byproducts
Isovalerate
Byproducts
GlcNAc, ManNAc
Acetate, Formate, CO2
Increased intake/efflux in HIGH spore context
Increased intake/efflux in LOW spore context
E
Isovalerate
Scaled Intensity (Log10)
CDI: - + - +
High Spores
Low Spores
F
N-Acetylneuraminate
Scaled Intensity (Log10)
CDI: - + - +
High Spores
Low Spores