**Target Journal:** NPJ Systems Biology & Applications

**Working Title:** Genome-scale metabolic network analysis of the bacterial pathogen *Clostridioides difficile* highlights conserved patterns of virulence-related core metabolic reprogramming

**Authors:**

Matthew L Jenior1

Deborah A Powers1

Jhansi L Leslie2

Mary E Dickenson1

William A Petri Jr2,4,5

Rita Tamayo6

Jason A Papin1,2,3

**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, NC, USA

**ABSTRACT**

*Clostridioides difficile* is a Gram-positive, sporulating anaerobe that has become the leading cause of hospital-acquired infection. Over the previous decade, numerous studies have brought the importance of metabolism to numerous aspects of *C. difficile* infection and pathogenesis to the forefront. Additionally, due to growing threats of antibiotic resistance and recurrent infection, targeting components of metabolism presents another possible means by which to combat this infection. In the past, genome-scale metabolic network analysis of bacteria has afforded the ability to systematically investigate how genetic and metabolic properties potentially contribute to observed phenotypes and predict the outcome of perturbations. These predictions ultimately create a platform for high-throughput identification and screening of potential therapeutic targets prior to laboratory testing. To accomplish these goals in *C. difficile*, we constructed highly-curated metabolic network reconstructions for a well-studied laboratory strain of the pathogen (str. 630) as well as a more recently characterized hyper-virulent isolate (str. R20291). These GENREs include key components of its core metabolism and nutrient acquisition systems to recapitulate metabolic behaviors within the complex milieu of the gut. Simulating the impact of single-gene deletions revealed accuracies of 89.1% and 88.9% for each GENRE respectively 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). Subsequently, *in vitro* and *in vivo* transcriptomic data integration with each GENRE predicted shifts in carbohydrate & amino acid metabolism which corresponded with differential virulence factor expression experimentally. Collectively, our results indicated that GENRE-based analyses of *C. difficile* are an effective means for understanding metabolism during infection and novel therapeutic target identification.

**INTRODUCTION**

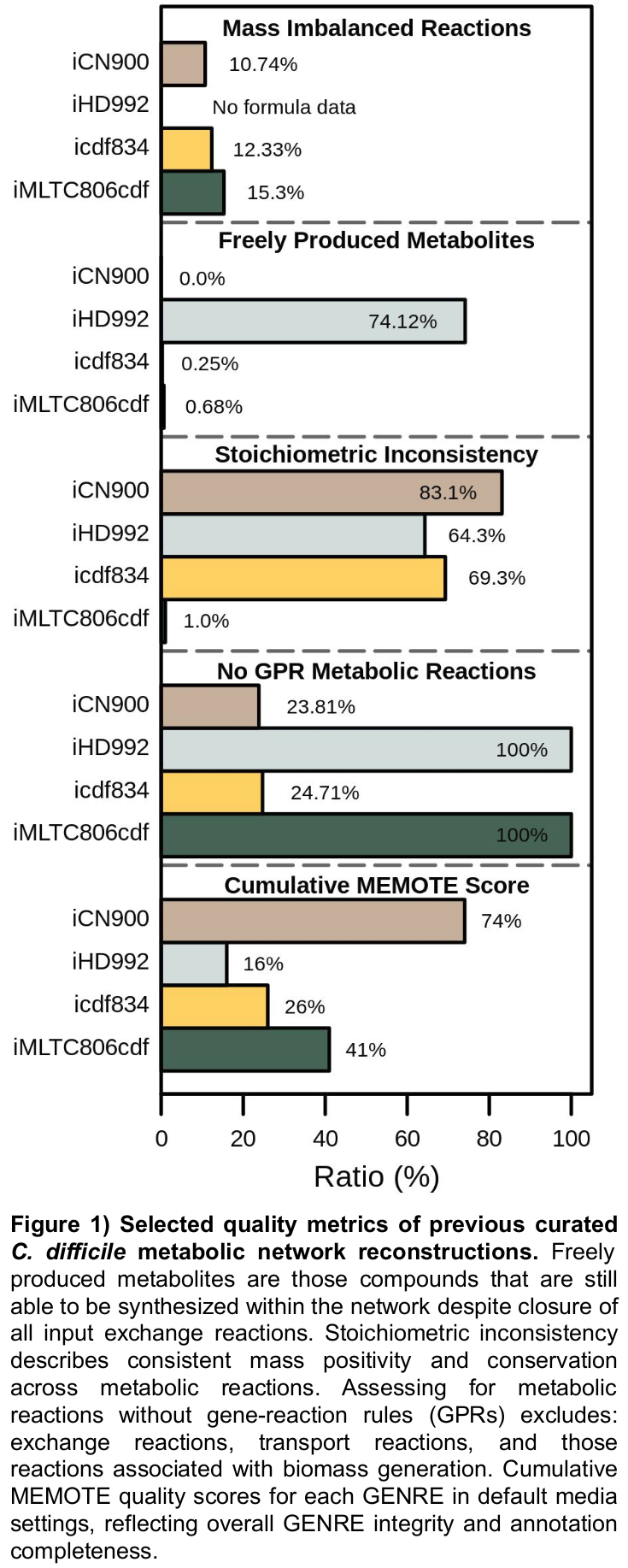
The nosocomial bacterial pathogen *Clostridioides* (formerly *Clostridium*) *difficile* which induces a toxin-mediated, diarrheal illness and is now the leading cause of hospital-acquired infection in the United States [[1,2]](https://paperpile.com/c/bcuLgy/1ulD+9xCv). Susceptibility to *C. difficile* infection (CDI) is most frequently preceded by exposure to antibiotic therapy for a prior infection [[3]](https://paperpile.com/c/bcuLgy/HLTf). While these drugs are often life-saving for primary bacterial infection, the collateral damage incurred to the healthy gut microbiota eliminates additional populations of species and alters the metabolic environment of the gut, leaving it susceptible to colonization by *C. difficile* [[4–6]](https://paperpile.com/c/bcuLgy/7Yq4+YZiK+UQVX). Recently, it was established that *C. difficile* adapts transcription of distinct catabolic pathways to conditions found in susceptible gut environments corresponding to the antibiotic pretreatment [[7,8]](https://paperpile.com/c/bcuLgy/u8hZ+oX75). These transcriptional shifts indicated that *C. difficile* must tightly 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 environments including carbohydrate and amino acid fermentation [[9]](https://paperpile.com/c/bcuLgy/taqp), however the relative utility of each metabolic strategy across given infections remains unknown. Furthermore, it has also found that the availability of nutrients including fermentable monosaccharides and certain amino acids drives expression of virulence in *C. difficile* [[9,10]](https://paperpile.com/c/bcuLgy/Yx4C+taqp). Given these findings, along with the increased prevalence of antibiotic resistance and hyper-virulence among *C. difficile* isolates [[11,12]](https://paperpile.com/c/bcuLgy/wPL5+HYlY), novel therapeutic strategies are desperately needed. 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 data on many observable metabolic phenotypes. These tools have achieved success in directing genetic engineering efforts [[13]](https://paperpile.com/c/bcuLgy/wfxs) or accurately predicting auxotrophies and competition/cooperation between species for growth substrates [[14,15]](https://paperpile.com/c/bcuLgy/InLc+1o8L). Most importantly, GENREs also create improved context for the interpretation of omics data [[16]](https://paperpile.com/c/bcuLgy/3KUg), and have provided powerful utility for identification of novel drug and gene targets accelerating downstream laboratory testing [[17]](https://paperpile.com/c/bcuLgy/3Gom). Leveraging these properties, several recent studies have found new possible metabolic targets within medically-relevant pathogens including *Klebsiella pneumophila*, *Staphylococcus aureus*, and *Streptococcus mutans* [[17–19]](https://paperpile.com/c/bcuLgy/3Gom+hrWV+Nl5s). Taken together, these principles make GENRE-based analyses a strong platform for analysis 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]](https://paperpile.com/c/bcuLgy/zJh0+hjrs+yzc8+slBm). However, analysis of these GENREs reinforced the necessity for carefully constructed stoichiometry and flux constraints in order to ensure that downstream predictions have the highest probability of accuracy. Additionally, as understanding of genome annotation and metabolic functionality increases, GENREs must be revisited or remade in order 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]](https://paperpile.com/c/bcuLgy/prFb), first generating a *de novo* reconstruction followed by extensive literature-driven manual curation of catabolic pathways, metabolite transport, and biomass objective function. We went on to use this reconstruction as a template to also create a curated GENRE for the more recently isolated hyper-virulent strain R20291 [[25]](https://paperpile.com/c/bcuLgy/ve1M). Both GENREs were subsequently validated against *in vitro* gene essentiality and carbon utilization screens, indicating a high degree of agreement for both datasets. To then assess the application of our GENREs for *in situ* metabolic prediction, we utilized transcriptomic data collected from both liquid media and active infection conditions to integrate into our models and assess the emergent metabolic activities. Analysis of context-specific pathogen metabolism revealed conserved patterns of metabolism across states of increased virulence favoring increased fermentation of amino acids and decreased capacity for glycolysis. These trends agreed with published phenotypes [[10,26]](https://paperpile.com/c/bcuLgy/eNfO+Yx4C), and supported the advantage provided by GENREs for delineating complex metabolic networks and patterns of gene expression into more tractable experimental targets. Overall, high-quality GENREs can greatly augment the discovery of novel therapeutic targets against CDI due to the established strong 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**

As so much of *C. difficile* pathogenicity is now being attributed to shifts in metabolism, effective GENRE-based analysis could likely provide enormous benefit to the identification of possible treatment targets in this and other recalcitrant pathogens. We began by first collecting and assessing the quality of existing *C. difficile* GENREs. The primary focus of curated *C. difficile* metabolic modeling efforts has been for 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]](https://paperpile.com/c/bcuLgy/zJh0)) and subsequent revision (icdf834 [[20,21]](https://paperpile.com/c/bcuLgy/zJh0+hjrs)), were followed by a recent *de novo* creation following updated genome curation (iCN900 [[23]](https://paperpile.com/c/bcuLgy/slBm)) [[27]](https://paperpile.com/c/bcuLgy/eEGc). Another GENRE developed for a variant of the str. 630 was for str. 630Δerm (iHD992 [[22]](https://paperpile.com/c/bcuLgy/yzc8)), which was generated following serial passage of the parent strain until erythromycin resistance was lost [[28]](https://paperpile.com/c/bcuLgy/ISZq). 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 (AGORA [[29]](https://paperpile.com/c/bcuLgy/JTGz)), however these did not receive more than semi-automated curation which was performed without *C. difficile*-specific considerations.

In order to establish a baseline for metabolic predictions possible with current *C. difficile* GENREs, we first decided on common criteria that would likely be most determinant of subsequent prediction quality for growth substrate utilization or essential network components. These metrics included the ratios: Mass imbalanced reactions, freely produced metabolites [[30,31]](https://paperpile.com/c/bcuLgy/V80t+zG2K), stoichiometric inconsistency [[32]](https://paperpile.com/c/bcuLgy/HxH1), and non-exchange/non-transport reactions lacking gene-reaction rules [[33]](https://paperpile.com/c/bcuLgy/C4sW) (Fig. 1). These features reflect the importance of mass conservation, flux consistency, 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 while no GENRE performed poorly in all categories, unique problems were found in each which made comparing simulation results across models impossible. For example, neither iMLTC806cdf and iHD994 have any detectable gene annotations associated with the reactions they contain. A high degree of stoichiometric inconsistency was detected across icdf834, iHD992, and iCN900; however only in iHD992 were the majority of intracellular metabolites able to be produced without any input metabolites 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 also determine the cumulative MEMOTE quality score for each *C. difficile* GENRE (Fig. 1). MEMOTE is a recent series of model quality assessment guidelines, agreed upon by the research community, and developed into a single platform in order to create an independent comparable quality metric across GENREs [[34]](https://paperpile.com/c/bcuLgy/t0ep). These percentages reflect a composite of mass conservation, reaction constraint, and standardized component annotation measures that are necessary for carrying out reliable simulations [[33]](https://paperpile.com/c/bcuLgy/C4sW). The three oldest reconstructions each score <50%, while 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 the newly established baseline standards of GENRE quality which dramatically impact downstream prediction.

Finally, we also assessed key metabolic functionalities as well as 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 [[35]](https://paperpile.com/c/bcuLgy/h3w7). 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 their deviation from measured values of ~29 minutes under similar conditions [[36]](https://paperpile.com/c/bcuLgy/caFq). This analysis uncovered that most GENREs were relatively close to the experimental measures, however iMLTC806cdf and iHD992 had times under 5 minutes and iCN900 was well over 500 minutes (Fig. S1C). We also detected structural inconsistencies across several GENREs. For example, those GENREs acquired from the AGORA database possessed numerous intracellular demand reactions 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 & aminoglycan utilization, and a variety of saccharide fermentation pathways [[37]](https://paperpile.com/c/bcuLgy/uAAr). Overall, the existing *C. difficile* GENREs possessed numerous mass imbalances & 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.

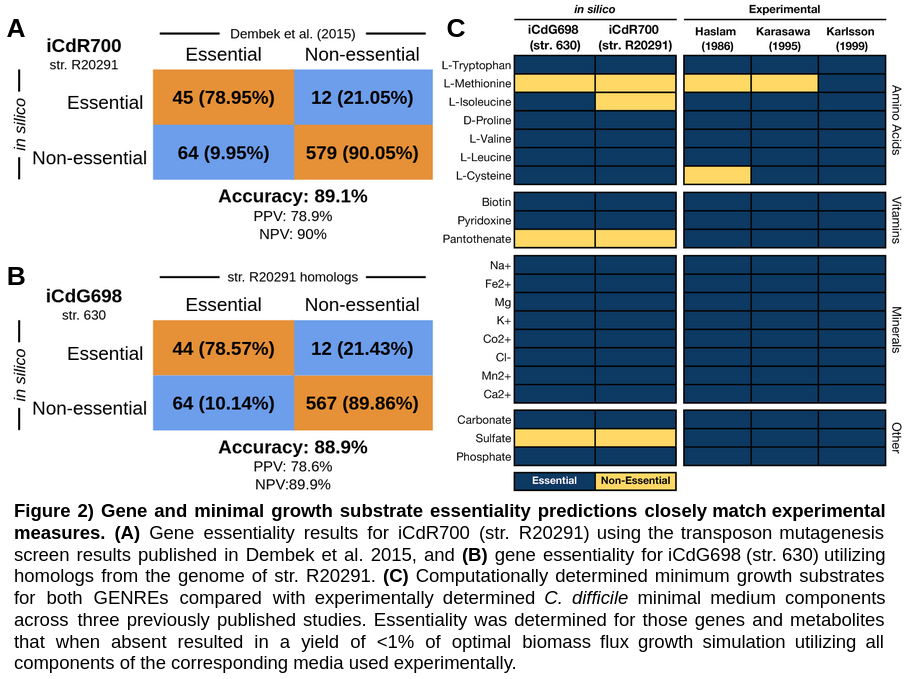
***C. difficile* Metabolic Network Scaffold Construction**

With the previously mentioned emergence of hypervirulent strains of *C. difficile*, it is important that future modeling efforts be equipped 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, in order to maximize utility for the bulk of published *C. difficile* metabolic research, we elected to simultaneously generate a reconstruction for the lab-adapted str. 630 in parallel. This afforded the ability to continuously cross-reference curations made between the models, and identify emergent differences specifically due to genomic content most readily. We began the reconstruction process by accessing the re-annotated genome of str. 630 [[27]](https://paperpile.com/c/bcuLgy/eEGc) and the published str. R20291 genome [[25]](https://paperpile.com/c/bcuLgy/ve1M), both available on the Pathosystems Resource Integration Center database (PATRIC) [[38]](https://paperpile.com/c/bcuLgy/vKjk). Following recent protocol for creating high-quality genome-scale models [[39]](https://paperpile.com/c/bcuLgy/Jq6y), and utilizing the automated ModelSEED framework and reaction database [[40]](https://paperpile.com/c/bcuLgy/BZUV), we generated scaffold reconstructions for both strains. The resultant scaffolds were stripped of those reactions added due to the gap-filling process, in order to be most reflective of original genomic content and partially reveal pathways in need of manual curation in order to complete (Table S2). Additionally, we removed genes that encoded enzymes involved in macromolecule synthesis, to focus the reconstructions more formally on bioconversion of metabolites (ex. ribosomal genes). We subsequently performed complete translated proteome alignment between str. 630 and str. R20291, resulting in 684 homologous metabolic gene products and 14 & 16 unique genes respectively (Table S2). After resolving these dissimilarities between the strains by incorporating corresponding metabolism to each reconstruction, we moved on to extensive manual curation of both GENREs.

**Curation of Metabolic Network and Ensemble Gap-filling**

Manual curation is required in order to ultimately produce high-quality GENREs and make meaningful biological predictions [[41]](https://paperpile.com/c/bcuLgy/Xq7J). 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 [[37,42]](https://paperpile.com/c/bcuLgy/uAAr+RwFV)
* Minimal media components and known auxotrophies [[43–45]](https://paperpile.com/c/bcuLgy/pMad+sF89+XmWT)
* Aminoglycan and dipeptide catabolism [[46–48]](https://paperpile.com/c/bcuLgy/U40N+0YqG+OtRS)
* As complete of Stickland fermentation oxidative and reductive pathways as possible [[36,49–58]](https://paperpile.com/c/bcuLgy/qHTh+vwkF+caFq+GEuQ+wAjo+Dqve+dTur+8gWd+fQVI+8xyh+xaYL)
* Carbohydrate fermentation and SCFA metabolism [[49,59–61]](https://paperpile.com/c/bcuLgy/qHTh+VH3H+CHq2+9VNK)
* Energy metabolite reversibility (ex. ATP, GTP, FAD, etc. [[30]](https://paperpile.com/c/bcuLgy/V80t))
* Periplasmic-associated H+ gradient and ATP synthase
* Additional pathogenicity-associated metabolites (ex. p-cresol [[51]](https://paperpile.com/c/bcuLgy/GEuQ) and ethanolamine [[62]](https://paperpile.com/c/bcuLgy/UwNL))

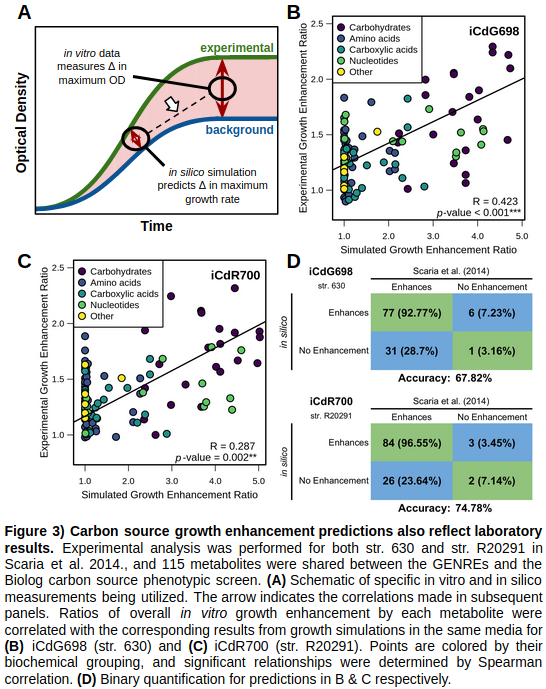
Following the outlined manual additions, we then created a customized biomass objective function with certain elements tailored to each strain of *C. difficile*. Our biomass objective function formulation was initially adapted from the well-curated GENRE of the close phylogenetic relative *Clostridium acetobutylicum* [[63]](https://paperpile.com/c/bcuLgy/UpBb) with additional considerations for tRNA synthesis and cell wall macromolecule formation, 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 ultimately yield strain-specific biomass objective functions [[64]](https://paperpile.com/c/bcuLgy/YvCC). In order to achieve growth, we next performed an ensemble-based pFBA gap-filling approach [[65,66]](https://paperpile.com/c/bcuLgy/xKdL+qODY), 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 media , Brain-Heart Infusion (BHI [[67]](https://paperpile.com/c/bcuLgy/vSN5)), *C. difficile* Defined Minimal Media +/- glucose (CDM [[45]](https://paperpile.com/c/bcuLgy/XmWT)), No Carbohydrate Minimal Media (NCMM [[5]](https://paperpile.com/c/bcuLgy/YZiK)), and *C. difficile* Defined Minimal Media (CDMM [[43]](https://paperpile.com/c/bcuLgy/pMad)) (Table S2). With each step collecting all new reactions found across members of an ensemble and integrating them into the draft reconstruction, resulting in a total of 68 new reactions that allowed for robust growth across all conditions.

Final steps of the curation process were focused on limiting directionality of reactions known to be irreversible, extensive balancing of remaining incorrect reaction stoichiometry, and adding annotation data for all network components. We repeated the assessments from Figure 1 and found that our GENREs had substantial improvements in all metrics including few, if any, flux or mass inconsistencies and now both received a cumulative MEMOTE score of 86% (Fig. S1A). 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 knockout studies**

A standard measurement of GENRE performance are comparisons of predicted essential genes for growth *in silico*, and those found to be essential experimentally through forward genetic screens [[68]](https://paperpile.com/c/bcuLgy/NP7Q). This form of analysis moves past strictly network quality criteria and into biologically tractable predictions. Many *C. difficile* strains have been historically difficult to manipulate genetically [[69]](https://paperpile.com/c/bcuLgy/OEVE), however methods were recently developed and a large-scale transposon mutagenesis screen was published for str. R20291 [[70]](https://paperpile.com/c/bcuLgy/Xn6K). 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 performed single gene knockouts for all genes and evaluated for >1% biomass flux in BHI media after growth simulation [[71]](https://paperpile.com/c/bcuLgy/e7oO) for both iCdR700 (Fig. 2A) and iCdG698 (Fig. 2B), cross-referencing the results with those in the published study. These comparisons revealed overall accuracies of 89.1% and 88.9%, with positive-predictive values of 78.9% and 78.6% respectively. Negative-predictive values were even higher with 90% for iCdR700 and 89.9% for iCdG698, indicating that our results for both GENREs closely matched their corresponding experimental datasets.

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

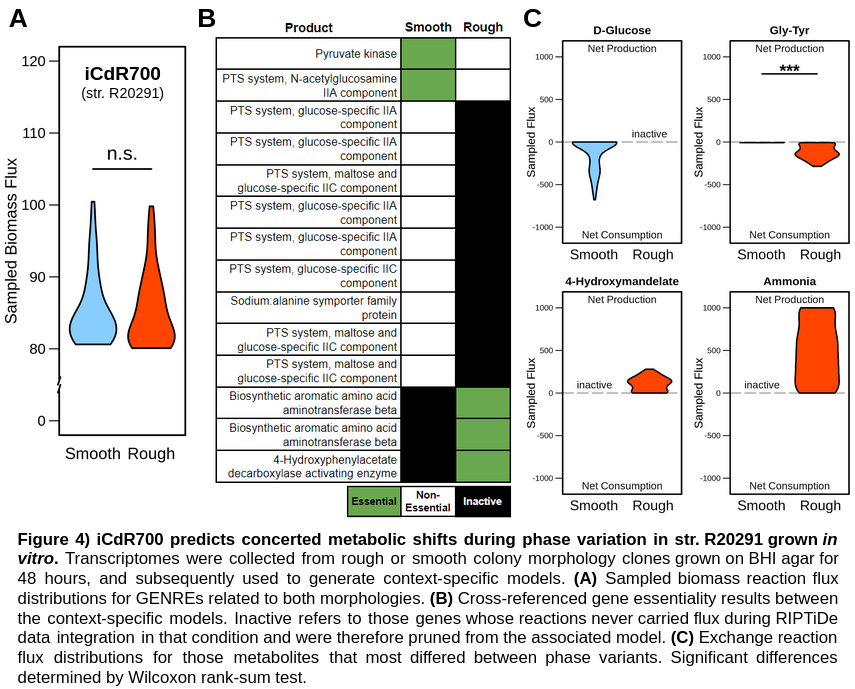
As a secondary minimal growth requirement check following gap-filling, we then identified the computationally-derived minimum subset of metabolites that our model required exogenous supply for growth. This task was accomplished through systematic *in silico* limitation followed by growth simulation for each component of the experimentally defined minimal media composition, also utilized in the final phase of gap-filling (Fig. 2C), and comparing the results across three separate studies that each published on minimal growth requirements for a diverse subset of *C. difficile* strains [[43–45]](https://paperpile.com/c/bcuLgy/XmWT+pMad+sF89). This analysis revealed that the majority of metabolites found to be essential during growth simulation, have also been shown to also be required for *in vitro* growth. Interestingly, neither str. 630 nor str. R20291 are auxotrophic for methionine, which was present in only one of the published formulations of CDMM where it was found to be largely growth-enhancing and not necessarily essential for only small levels of growth [[44]](https://paperpile.com/c/bcuLgy/sF89). Furthermore, it appeared that iCdR700 (str. R20291) lost the auxotrophy for isoleucine, which had some evidence at the genetic level (Table S2) and increases isoleucine consumption is associated with greater pathogenicity in some *C. difficile* strains [[72]](https://paperpile.com/c/bcuLgy/QLVL). Finally, as sulfate is added to the media alongside magnesium (MgSO4), it was not surprising that this was not essential as *C. difficile* is able to harvest sufficient bioavailable sulfur from excess cysteine [[45,73]](https://paperpile.com/c/bcuLgy/XmWT+UbbB). 

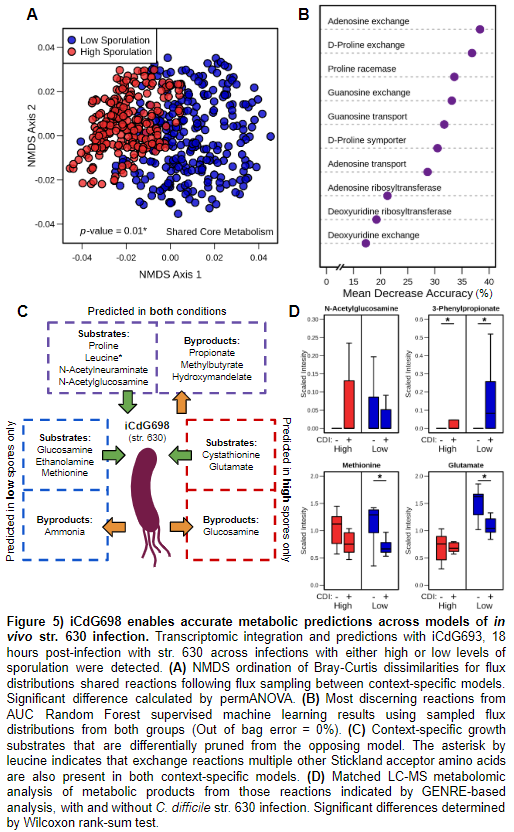
**Metabolite-specific growth enhancement also strongly correlates with *in vitro* results**

We moved on to assess additional carbon sources that impact the growth rate predictions for both GENREs. Utilizing previously published results for both *C. difficile* strains in a Carbon Source Utilization Screen [[74]](https://paperpile.com/c/bcuLgy/sgDP), we simulated the degree to which each metabolite influenced growth rate in minimal media. Importantly, *C. difficile* is auxotrophic for specific amino acids that it is also able to catabolize through Stickland fermentation [[75]](https://paperpile.com/c/bcuLgy/EfpG). As such, the values are reported as the ratio of final optical density for the given metabolite and background growth. Concordantly, we similarly calculated the influence of each metabolite on the optimal biomass flux at quasi-steady state of each model provided the same background media conditions as the Biolog analysis (Fig. 3A). 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 for iCdG698 (*p*-value < 0.001) and iCdR700 (*p*-value = 0.002) for the amount of growth enhancement (Fig. 3B & 3C). This relationship was even more pronounced for carbohydrates and amino acids, metabolites groups that their utilization pathways received the largest amounts of manual curation (Fig. S2). Furthermore, when these predictions were reduced to binary interpretations of either enhancement or non-enhancement, we found that iCdG698 allowed for 92.77% and iCdR700 allowed for 96.55% correct enhancement calls (Fig. 3D). 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 conserved patterns relating to virulence both *in vitro* and *in vivo***

Following GENRE validation, we then sought to quantify 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 potential influence of pathways not obviously connected [[76]](https://paperpile.com/c/bcuLgy/n6A6). With this 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 transcriptomic integration platform [[77]](https://paperpile.com/c/bcuLgy/XsXX). This process affords the ability to make much more fine-scale predictions of metabolic changes *C. difficile* undergoes as it activates pathogenicity, but also reveals conserved nodes across models that could lead to targeted strategies for intentional down-regulation of virulence factors through metabolic circuitry.

A recent study identified that phase variation, a common strategy in many bacterial pathogens that express phenotypic heterogeneity within a single population to maximize overall fitness, likely also alters pathogenicity expression in *C. difficile* str. R20291. This phenotype most readily manifested in a rough or smooth-edged colony morphology on solid agar, which is able to be propagated onto new media and is associated with varied expression of toxin and flagella virulence factors [[78]](https://paperpile.com/c/bcuLgy/Lo4H). Utilizing transcriptomic data isolated and sequenced from rough or smooth phase variants, we generated context-specific versions of iCdR700 in simulated rich media conditions. It has been previously shown that phase-locked mutants in *cmr*-family genes displayed no significant difference in growth rate [[78]](https://paperpile.com/c/bcuLgy/Lo4H). Indeed context-specific iCdR700 models predicted no significant difference in sampled biomass flux values (Fig. 4A; *p*-value = 0.583), reflecting no change in predicted biomass adjective flux between the two context-specific variants *in silico*. We then calculated essential genes in each variant model similar to the earlier analysis which identified 81 core genes essential to both conditions (Table S4), another 13 being essential to growth in BHI for both variants, and 14 genes that were conditionally essential between the morphologies (Fig. 4B). These differences indicated deviation in optimal metabolic activity, despite the flux distributions of core pathways appearing largely similar between variant-specific models (Fig S3). Conditionally-essential genes fall into three categories; Essential for minimal growth, non-essential for minimal growth, or entirely inactive in the context of interest. The smooth-only essential gene set included several carbohydrate transporters as well as pyruvate kinase, which mediates the last step of glycolysis and a bulk of the ATP generation. Additionally at the transcriptional level, reads mapped to pyruvate kinase were detected at nearly identical levels between the rough and smooth isolates (Table S4). Our results indicated that glycolysis was not only non-essential in the smooth colony variants, but also largely pruned from the network associated with the rough colony transcriptome. Conversely, in the rough context-specific model, genes and reactions linked to Stickland fermentation or amino acid transport appeared to be differentially essential (Fig. 4B). Transcriptional patterns similar to those found with pyruvate kinase were also seen in these genes (Table S4). Both trends for differential metabolic strategy were then reinforced when we compared sampled flux distributions for the associated exchange reactions for common substrates and byproducts of each respective pathway (Fig. 4C). Exchange reactions represent the edges of metabolic interaction of an organism with its environment and serve as both the input and output mechanism for their specific metabolite substrate to and from the system. An overall negative flux for an exchange reaction distribution reflects net import of that substrate into the network, while positive flux indicates export of a byproduct. We found that not only was glucose imported into the smooth variant model, this functionality was entirely inactive in the rough-associated model (Fig. 4C). Similarly, Gly-Tyr (a dipeptide of glycine and tyrosine) includes two known substrates of Stickland fermentation, while both 4-hydroxymandelate and ammonia are byproducts of this pathway [[79]](https://paperpile.com/c/bcuLgy/VSLN). It has been previously reported that this relationship between phase variant and metabolism may indeed occur in *C. difficile*, and our collective results from contextualized iCdR700 analysis support this to be the case.

As laboratory media conditions are much more easily defined, so 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 across separate classes of antibiotics that allow for identical levels of pathogen colonization, yet resulted in distinct expression of sporulation (another virulence factor) at the time point measured [[80]](https://paperpile.com/c/bcuLgy/bARQ). This characterization was paired with untargeted metabolomic analysis of intestinal content to correlate the transcriptional activity of metabolic pathways with changes in abundance of their respective substrates and byproducts following infection, making this dataset extremely valuable for our purposes. We subsequently overlaid these data onto our GENRE of str. 630 (iCdG698) in the same fashion as the preceding analysis, and first compared the combined activities of those reactions conserved between conditions in order to quantify differential use of core metabolism. 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. 5A). In agreement with the previous findings that *C. difficile* is able to adapt for distinct growth substrates [[80]](https://paperpile.com/c/bcuLgy/bARQ), 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 those conditions that favor increased sporulation support a lower diversity of potential metabolic strategies. Then using supervised machine learning, we identified those reactions which most readily delineate flux distributions from low and high context-specific models (Fig. 5B). This analysis revealed several reactions associated with proline fermentation, a preferred growth substrate of *C. difficile* which access to has been previously shown to negatively regulate *C. difficile* pathogenicity [[52]](https://paperpile.com/c/bcuLgy/wAjo).

We then moved to assess exchange reactions that were differentially pruned between the context-specific models (Fig. 5C) and compared them with changes in concentration of select associated metabolites (Fig. 5D). First, we looked toward those exchanges for carbon sources shared between conditions and found the host-derived aminoglycans N-acetylglucosamine and N-acetylneuraminic acid. Both are known to be *C. difficile* growth substrates, and are readily available in the intestines of these mice [[7]](https://paperpile.com/c/bcuLgy/u8hZ). Additionally, exchanges for the preferred Stickland donor proline and multiple acceptors (including leucine\*; Table S5) are present in both models, as well as multiple fermentation end products. Of the three byproducts listed, phenylpropionate is the closest analogous metabolite measured, and was found to be significantly increased in both groups (Fig. 5D; *p*-values < 0.001), reaffirming the importance of fermentative metabolism to *C. difficile*. We then contrasted groups to identify those exchange reactions that differ in presence or directionality between contexts. This comparison highlighted that ethanolamine was predicted to only be imported in the lower sporulation condition, which has also been shown to be associated with reduced *C. difficile* pathogenicity [[62]](https://paperpile.com/c/bcuLgy/UwNL). Methionine was predicted to also be utilized in the lower sporulation condition and although its concentrations decrease across both groups following infection, the only significant relationship is in the lower sporulation condition (*p*-value = 0.001). Alternatively, in the case of glutamate the model-based predictions are actually inverse to the metabolomic results, potentially highlighting possible contributions by other members of the gut microbiota or a target of future curation. Most interestingly, glucosamine appeared as a byproduct in the high sporulation-associated model while remaining a substrate in the low sporulation group (Fig. 5C). Indeed it has been previously shown that *C. difficile* may secrete this metabolite during biofilm formation under conditions that induce greater virulence expression [[81]](https://paperpile.com/c/bcuLgy/KVs5). This was potentially bolstered by the *in vivo* measurement on N-acetylglucosamine, the closest related molecule, where it actually increased in concentration in the higher sporulation setting (Fig. 5D). 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*.

To then examine the str. R20291 GENRE’s utility 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 24 genes that are essential only during infection (Table S5). Among the genes highlighted were numerous components of purine metabolism. These genes are highly expressed during infection and inhibition of specific enzymes within this pathway has been shown to downregulate toxin production [[6,82]](https://paperpile.com/c/bcuLgy/UQVX+hZwR). These findings demonstrated 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 are 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. 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 iCd700 (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 iCdG693 and iCdR700 were able to catabolize amino acids as the sole carbon source through Stickland fermentation and required 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. 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 in order to favor certain forms of metabolism over others. Context-specific essentiality results also uncovered conserved patterns of genes necessary for growth during infection, and provides a list of potential targets for future inhibitor screens.

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 iCdG693 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. Efforts in the immediate future 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 genomic 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 iCdG693 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. Systems-biology approaches have afforded an ability to assess fine-scale changes to metabolism of single species within complex environments that may have downstream implications on health and disease. As such, finding core metabolic properties in *C. difficile* strains may be key in identifying potential probiotic competitor strains or even molecular inhibitors of metabolic components.

**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 [[38,83,84]](https://paperpile.com/c/bcuLgy/NeV1+vKjk+eFPI). 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 [[85]](https://paperpile.com/c/bcuLgy/HkUi). 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 [[86]](https://paperpile.com/c/bcuLgy/z6SR). 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 (ex. ensured free water diffusion across compartments). Final transport reactions were then validated with TransportDB [[87]](https://paperpile.com/c/bcuLgy/yte3). 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 [[86,88]](https://paperpile.com/c/bcuLgy/z6SR+8s9e), KEGG [[89]](https://paperpile.com/c/bcuLgy/W6yQ), PATRIC, RefSeq [[90]](https://paperpile.com/c/bcuLgy/yhG1), EMBL [[91]](https://paperpile.com/c/bcuLgy/63Em), and BiGG [[92]](https://paperpile.com/c/bcuLgy/H3oY) databases and integrated it into the annotation field dictionary now supported in the most recent SBML version [[93]](https://paperpile.com/c/bcuLgy/QUCe). 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 or iCdR700 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 [[94]](https://paperpile.com/c/bcuLgy/hGyW). The techniques utilized included; Flux-balance analysis, flux-variability analysis [[95]](https://paperpile.com/c/bcuLgy/fcQw), Gapsplit flux-sampler [[96]](https://paperpile.com/c/bcuLgy/oRsJ), and minimal\_medium on exhaustive search settings. GENRE quality assessment tools were also developed in python and are fully available in project Github repository. MEMOTE quality reports were generated using the web-based implementation found at https://memote.io/.

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 [[97]](https://paperpile.com/c/bcuLgy/NI0k). Mapping curated reads to the respective C. difficile genome was then performed with Bowtie2 [[98]](https://paperpile.com/c/bcuLgy/ps56). MarkDuplicates then removed optica/PCR duplicates (broadinstitute.github.io/picard/), and mappings were converted to idxstats format using SAMtools [[99]](https://paperpile.com/c/bcuLgy/dsCg). Abundances were then normalized to both read and target lengths. Transcriptomic integration and context-specific model generation was performed with RIPTiDe [[77]](https://paperpile.com/c/bcuLgy/XsXX).

Statistical Methods

All statistical analysis was performed in R v3.2.0 using the vegan package [[100]](https://paperpile.com/c/bcuLgy/vg7W). Significant differences for single reaction flux distributions and metabolite concentrations were determined by Wilcoxon signed-rank test, and those for multi-reaction distributions were determined by permANOVA of Bray-Curtis dissimilarity. Supervised machine-learning was accomplished with the implementation of AUC-Random Forest also in R [[101]](https://paperpile.com/c/bcuLgy/JrzB). All code associated with this study is available in the study-associated GitHub repository.

Data sources and code 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 [[83]](https://paperpile.com/c/bcuLgy/NeV1). Transcriptomic data was downloaded in raw FASTQ format from the NCBI Sequence Read Archive (SRA: PRJNA415307 and SRA: PRJNA354635), or provided by the Tamayo Lab. 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 curation GENRE design, creation, validation, curation, and analysis. Visualization of data. Drafting manuscript.

DAP - GENRE validation, benchmarking, and curation. drafting and editing manuscript

JLL - GENRE design and benchmarking, drafting and editing manuscript

MED - GENRE validation and curation. drafting and editing manuscript

WAP - editing the manuscript

RT - in vitro C. difficile phase variation analysis, transcriptomic data generation, drafting and editing the manuscript

JAP - Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing

**Acknowledgements**

The authors would like to acknowledge Bonnie Dougherty, Laura Dunphy, and Dawson Payne for their input and feedback on modeling parameters and objective format. The authors have declared that no competing interests exist. This work was supported by funding from The U.S. National Institutes of Health (1R01AT010253) to JP. The funding agency had no role in study design, data collection/analysis, or preparation of the manuscript.

**FIGURE & TABLE LEGENDS**

**Figure 1) Selected quality metrics of previous curated *C. difficile* metabolic network reconstructions.** Freely produced metabolites are those compounds that are still able to be synthesized within the network despite closure of all input exchange reactions. Stoichiometric inconsistency describes consistent mass positivity and 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.

**Figure 2) Gene and minimal growth substrate essentiality predictions closely match experimental measures.** **(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% of optimal biomass flux growth simulation utilizing all components of the corresponding media used experimentally.

**Figure 3) Carbon source growth enhancement predictions also reflect laboratory results.** 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 4) 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)** Sampled biomass reaction flux distributions for GENREs related to both morphologies. **(B)** Cross-referenced gene essentiality results between the context-specific models. Inactive refers to those genes whose reactions never carried flux during RIPTiDe data integration in that condition and were therefore pruned from the associated model. **(C)** Exchange reaction flux distributions for those metabolites that most differed between phase variants. Significant differences determined by Wilcoxon rank-sum test.

**Figure 5) iCdG698 enables accurate metabolic predictions across models of *in vivo* str. 630 infection.** Transcriptomic integration and predictions with iCdG693, 18 hours post-infection with str. 630 across infections with either high or low levels of sporulation were detected. **(A)** NMDS ordination of Bray-Curtis dissimilarities for flux distributions shared reactions following flux sampling between context-specific models. Significant difference calculated by permANOVA. **(B)** Most discerning reactions from AUC Random Forest supervised machine learning results using sampled flux distributions from both groups (Out of bag error = 0%). **(C)** Context-specific growth substrates that are differentially pruned from the opposing model. The asterisk by leucine indicates that exchange reactions multiple other Stickland acceptor amino acids are also present in both context-specific models. **(D)** Matched LC-MS metabolomic analysis of metabolic products from those reactions indicated by GENRE-based analysis, with and without *C. difficile* str. 630 infection. Significant differences determined by Wilcoxon rank-sum test.

**Figure S1) Quality metrics for new and auto-curated AGORA *C. difficile* GENREs.** **(A)** Selected quality metrics reported for all GENREs. **(B)** Cumulative MEMOTE scores for remaining GENREs in default media. **(C)** Imputed doubling time in complete media, calculated as the reciprocal optimal biomass flux per unit time for all GENREs.

**Figure S2)** **Specific shifts in simulated versus measure growth enhancement for each metabolite measured in the carbon source utilization screen.** Separated by metabolite group designation. **(A)** iCdG698 and **(B)** iCdR700.

**Figure S3) Change in core flux distributions *in vitro* iCdR700.** NMDS ordination of Bray-Curtis dissimilarities between flux distributions of shared reactions of context-specific models.

**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 flux distributions *in vitro* with iCdR700 (str. R20291)

**Table S5)** Gene essentiality and reaction flux for predicted substrates/byproducts in iCdG698 (str. 630) *in vivo*.

**REFERENCES**

1. [Lessa FC, Winston LG, McDonald LC, Emerging Infections Program C. difficile Surveillance Team. Burden of Clostridium difficile infection in the United States. The New England journal of medicine. 2015. pp. 2369–2370.](http://paperpile.com/b/bcuLgy/1ulD)

2. [Bella SD, Di Bella S, Ascenzi P, Siarakas S, Petrosillo N, di Masi A. Clostridium difficile Toxins A and B: Insights into Pathogenic Properties and Extraintestinal Effects. Toxins. 2016. p. 134. doi:](http://paperpile.com/b/bcuLgy/9xCv)[10.3390/toxins8050134](http://dx.doi.org/10.3390/toxins8050134)

3. [Thomas C. Antibiotics and hospital-acquired Clostridium difficile-associated diarrhoea: a systematic review. Journal of Antimicrobial Chemotherapy. 2003. pp. 1339–1350. doi:](http://paperpile.com/b/bcuLgy/HLTf)[10.1093/jac/dkg254](http://dx.doi.org/10.1093/jac/dkg254)

4. [Antunes LCM, Han J, Ferreira RBR, Lolić P, Borchers CH, Finlay BB. Effect of antibiotic treatment on the intestinal metabolome. Antimicrob Agents Chemother. 2011;55: 1494–1503.](http://paperpile.com/b/bcuLgy/7Yq4)

5. [Theriot CM, Koenigsknecht MJ, Carlson PE Jr, Hatton GE, Nelson AM, Li B, et al. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. Nat Commun. 2014;5: 3114.](http://paperpile.com/b/bcuLgy/YZiK)

6. [Fletcher JR, Erwin S, Lanzas C, Theriot CM. Shifts in the Gut Metabolome and Clostridium difficile Transcriptome throughout Colonization and Infection in a Mouse Model. mSphere. 2018. doi:](http://paperpile.com/b/bcuLgy/UQVX)[10.1128/msphere.00089-18](http://dx.doi.org/10.1128/msphere.00089-18)

7. [Jenior ML, Leslie JL, Young VB, Schloss PD. Colonizes Alternative Nutrient Niches during Infection across Distinct Murine Gut Microbiomes. mSystems. 2017;2. doi:](http://paperpile.com/b/bcuLgy/u8hZ)[10.1128/mSystems.00063-17](http://dx.doi.org/10.1128/mSystems.00063-17)

8. [Jenior ML, Leslie JL, Young VB, Schloss PD. Clostridium difficile Alters the Structure and Metabolism of Distinct Cecal Microbiomes during Initial Infection To Promote Sustained Colonization. mSphere. 2018. doi:](http://paperpile.com/b/bcuLgy/oX75)[10.1128/msphere.00261-18](http://dx.doi.org/10.1128/msphere.00261-18)

9. [Hofmann JD, Otto A, Berges M, Biedendieck R, Michel A-M, Becher D, et al. Metabolic Reprogramming of Clostridioides difficile During the Stationary Phase With the Induction of Toxin Production. Frontiers in Microbiology. 2018. doi:](http://paperpile.com/b/bcuLgy/taqp)[10.3389/fmicb.2018.01970](http://dx.doi.org/10.3389/fmicb.2018.01970)

10. [Dineen SS, McBride SM, Sonenshein AL. Integration of metabolism and virulence by Clostridium difficile CodY. J Bacteriol. 2010;192: 5350–5362.](http://paperpile.com/b/bcuLgy/Yx4C)

11. [(u.s.) CFDCAP, Centers for Disease Control and Prevention (U.S.). Antibiotic resistance threats in the United States, 2019. 2019. doi:](http://paperpile.com/b/bcuLgy/wPL5)[10.15620/cdc:82532](http://dx.doi.org/10.15620/cdc:82532)

12. [Merrigan M, Venugopal A, Mallozzi M, Roxas B, Viswanathan VK, Johnson S, et al. Human hypervirulent Clostridium difficile strains exhibit increased sporulation as well as robust toxin production. J Bacteriol. 2010;192: 4904–4911.](http://paperpile.com/b/bcuLgy/HYlY)

13. [Hao T, Wu D, Zhao L, Wang Q, Wang E, Sun J. The Genome-Scale Integrated Networks in Microorganisms. Front Microbiol. 2018;9: 296.](http://paperpile.com/b/bcuLgy/wfxs)

14. [Pacheco AR, Moel M, Segrè D. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. Nat Commun. 2019;10: 103.](http://paperpile.com/b/bcuLgy/InLc)

15. [Seif Y, Choudhary KS, Hefner Y, Anand A, Yang L, Palsson BO. Metabolic and genetic basis for auxotrophies in Gram-negative species. Proc Natl Acad Sci U S A. 2020;117: 6264–6273.](http://paperpile.com/b/bcuLgy/1o8L)

16. [Hadadi N, Pandey V, Chiappino-Pepe A, Morales M, Gallart-Ayala H, Mehl F, et al. Mechanistic insights into bacterial metabolic reprogramming from omics-integrated genome-scale models. NPJ Syst Biol Appl. 2020;6: 1.](http://paperpile.com/b/bcuLgy/3KUg)

17. [Cesur MF, 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. 2020. doi:](http://paperpile.com/b/bcuLgy/3Gom)[10.3389/fcimb.2019.00447](http://dx.doi.org/10.3389/fcimb.2019.00447)

18. [Jijakli K, Jensen PA. Metabolic Modeling of Streptococcus mutans Reveals Complex Nutrient Requirements of an Oral Pathogen. mSystems. 2019;4. doi:](http://paperpile.com/b/bcuLgy/hrWV)[10.1128/mSystems.00529-19](http://dx.doi.org/10.1128/mSystems.00529-19)

19. [Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, Palsson BØ. Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. Proc Natl Acad Sci U S A. 2016;113: E3801–9.](http://paperpile.com/b/bcuLgy/Nl5s)

20. [Larocque M, Chénard T, Najmanovich R. A curated C. difficile strain 630 metabolic network: prediction of essential targets and inhibitors. BMC Syst Biol. 2014;8: 117.](http://paperpile.com/b/bcuLgy/zJh0)

21. [Kashaf SS, 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. 2017. doi:](http://paperpile.com/b/bcuLgy/hjrs)[10.1186/s12918-017-0395-3](http://dx.doi.org/10.1186/s12918-017-0395-3)

22. [Dannheim H, Will SE, Schomburg D, Neumann-Schaal M. Clostridioides difficile630Δerm in silicoandin vivo- quantitative growth and extensive polysaccharide secretion. FEBS Open Bio. 2017. pp. 602–615. doi:](http://paperpile.com/b/bcuLgy/yzc8)[10.1002/2211-5463.12208](http://dx.doi.org/10.1002/2211-5463.12208)

23. [King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 2016;44: D515–22.](http://paperpile.com/b/bcuLgy/slBm)

24. [Monot M, Boursaux-Eude C, Thibonnier M, Vallenet D, Moszer I, Medigue C, et al. Reannotation of the genome sequence of Clostridium difficile strain 630. J Med Microbiol. 2011;60: 1193–1199.](http://paperpile.com/b/bcuLgy/prFb)

25. [Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, et al. Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol. 2009;10: R102.](http://paperpile.com/b/bcuLgy/ve1M)

26. [Antunes A, Camiade E, Monot M, Courtois E, Barbut F, Sernova NV, et al. Global transcriptional control by glucose and carbon regulator CcpA in Clostridium difficile. Nucleic Acids Res. 2012;40: 10701–10718.](http://paperpile.com/b/bcuLgy/eNfO)

27. [Dannheim H, Riedel T, Neumann-Schaal M, Bunk B, Schober I, Spröer C, et al. Manual curation and reannotation of the genomes of Clostridium difficile 630Δerm and C. difficile 630. Journal of Medical Microbiology. 2017. pp. 286–293. doi:](http://paperpile.com/b/bcuLgy/eEGc)[10.1099/jmm.0.000427](http://dx.doi.org/10.1099/jmm.0.000427)

28. [Hussain HA, Roberts AP, 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. 2005. pp. 137–141. doi:](http://paperpile.com/b/bcuLgy/ISZq)[10.1099/jmm.0.45790-0](http://dx.doi.org/10.1099/jmm.0.45790-0)

29. [Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol. 2017;35: 81–89.](http://paperpile.com/b/bcuLgy/JTGz)

30. [Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. PLoS Comput Biol. 2017;13: e1005494.](http://paperpile.com/b/bcuLgy/V80t)

31. [Schellenberger J, Lewis NE, Palsson BØ. Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys J. 2011;100: 544–553.](http://paperpile.com/b/bcuLgy/zG2K)

32. [Gevorgyan A, Poolman MG, Fell DA. Detection of stoichiometric inconsistencies in biomolecular models. Bioinformatics. 2008;24: 2245–2251.](http://paperpile.com/b/bcuLgy/HxH1)

33. [Ravikrishnan A, Raman K. Critical assessment of genome-scale metabolic networks: the need for a unified standard. Brief Bioinform. 2015;16: 1057–1068.](http://paperpile.com/b/bcuLgy/C4sW)

34. [Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, et al. MEMOTE for standardized genome-scale metabolic model testing. Nat Biotechnol. 2020;38: 272–276.](http://paperpile.com/b/bcuLgy/t0ep)

35. [Oberhardt MA, Puchałka J, dos Santos VAPM, Papin JA. Reconciliation of Genome-Scale Metabolic Reconstructions for Comparative Systems Analysis. PLoS Computational Biology. 2011. p. e1001116. doi:](http://paperpile.com/b/bcuLgy/h3w7)[10.1371/journal.pcbi.1001116](http://dx.doi.org/10.1371/journal.pcbi.1001116)

36. [Neumann-Schaal M, Hofmann JD, Will SE, Schomburg D. Time-resolved amino acid uptake of Clostridium difficile 630Δerm and concomitant fermentation product and toxin formation. BMC Microbiol. 2015;15: 281.](http://paperpile.com/b/bcuLgy/caFq)

37. [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. 2019. doi:](http://paperpile.com/b/bcuLgy/uAAr)[10.3389/fmicb.2019.00219](http://dx.doi.org/10.3389/fmicb.2019.00219)

38. [Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2017;45: D535–D542.](http://paperpile.com/b/bcuLgy/vKjk)

39. [Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010;5: 93–121.](http://paperpile.com/b/bcuLgy/Jq6y)

40. [Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, Henry C. Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. Methods Mol Biol. 2013;985: 17–45.](http://paperpile.com/b/bcuLgy/BZUV)

41. [Mendoza SN, Olivier BG, Molenaar D, Teusink B. A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biology. 2019. doi:](http://paperpile.com/b/bcuLgy/Xq7J)[10.1186/s13059-019-1769-1](http://dx.doi.org/10.1186/s13059-019-1769-1)

42. [Janoir C, Denève C, Bouttier S, Barbut F, Hoys S, Caleechum L, et al. Adaptive strategies and pathogenesis of Clostridium difficile from in vivo transcriptomics. Infect Immun. 2013;81: 3757–3769.](http://paperpile.com/b/bcuLgy/RwFV)

43. [Karasawa T, Ikoma S, Yamakawa K, Nakamura S. A defined growth medium for Clostridium difficile. Microbiology. 1995;141 ( Pt 2): 371–375.](http://paperpile.com/b/bcuLgy/pMad)

44. [Karlsson S, Burman LG, Åkerlund T. Suppression of toxin production in Clostridium difficile VPI 10463 by amino acids. Microbiology. 1999;145 ( Pt 7): 1683–1693.](http://paperpile.com/b/bcuLgy/sF89)

45. [Haslam SC, Ketley JM, Mitchell TJ, Stephen J, Burdon DW, Candy DC. Growth of Clostridium difficile and production of toxins A and B in complex and defined media. J Med Microbiol. 1986;21: 293–297.](http://paperpile.com/b/bcuLgy/XmWT)

46. [Stiemsma LT, Turvey SE, Finlay BB. An antibiotic-altered microbiota provides fuel for the enteric foe. Cell research. 2014. pp. 5–6.](http://paperpile.com/b/bcuLgy/U40N)

47. [Olson ME, King JM, Yahr TL, Horswill AR. Sialic acid catabolism in Staphylococcus aureus. J Bacteriol. 2013;195: 1779–1788.](http://paperpile.com/b/bcuLgy/0YqG)

48. [Engevik MA, Yacyshyn MB, Engevik KA, Wang J, Darien B, Hassett DJ, et al. Human Clostridium difficile infection: altered mucus production and composition. Am J Physiol Gastrointest Liver Physiol. 2015;308: G510–24.](http://paperpile.com/b/bcuLgy/OtRS)

49. [Nakamura S, Nakashio S, Yamakawa K, Tanabe N, Nishida S. Carbohydrate Fermentation byClostridium difficile. Microbiology and Immunology. 1982. pp. 107–111. doi:](http://paperpile.com/b/bcuLgy/qHTh)[10.1111/j.1348-0421.1982.tb00159.x](http://dx.doi.org/10.1111/j.1348-0421.1982.tb00159.x)

50. [Kim J, Darley D, Buckel W. 2-Hydroxyisocaproyl-CoA dehydratase and its activator from Clostridium difficile. FEBS J. 2005;272: 550–561.](http://paperpile.com/b/bcuLgy/vwkF)

51. [Selmer T, Andrei PI. p-Hydroxyphenylacetate decarboxylase from Clostridium difficile. A novel glycyl radical enzyme catalysing the formation of p-cresol. Eur J Biochem. 2001;268: 1363–1372.](http://paperpile.com/b/bcuLgy/GEuQ)

52. [Bouillaut L, Self WT, Sonenshein AL. Proline-dependent regulation of Clostridium difficile Stickland metabolism. J Bacteriol. 2013;195: 844–854.](http://paperpile.com/b/bcuLgy/wAjo)

53. [Jackson S, Calos M, Myers A, Self WT. Analysis of proline reduction in the nosocomial pathogen Clostridium difficile. J Bacteriol. 2006;188: 8487–8495.](http://paperpile.com/b/bcuLgy/Dqve)

54. [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. 2006. pp. 6062–6069. doi:](http://paperpile.com/b/bcuLgy/dTur)[10.1128/aem.00772-06](http://dx.doi.org/10.1128/aem.00772-06)

55. [Britz ML, Wilkinson RG. Leucine dissimilation to isovaleric and isocaproic acids by cell suspensions of amino acid fermenting anaerobes: the Stickland reaction revisited. Can J Microbiol. 1982;28: 291–300.](http://paperpile.com/b/bcuLgy/8gWd)

56. [Kim J, Hetzel M, Boiangiu CD, Buckel W. Dehydration of (R)-2-hydroxyacyl-CoA to enoyl-CoA in the fermentation of α-amino acids by anaerobic bacteria. FEMS Microbiology Reviews. 2004. pp. 455–468. doi:](http://paperpile.com/b/bcuLgy/fQVI)[10.1016/j.femsre.2004.03.001](http://dx.doi.org/10.1016/j.femsre.2004.03.001)

57. [Yu L, Blaser M, Andrei PI, Pierik AJ, Selmer T. 4-Hydroxyphenylacetate decarboxylases: properties of a novel subclass of glycyl radical enzyme systems. Biochemistry. 2006;45: 9584–9592.](http://paperpile.com/b/bcuLgy/8xyh)

58. [de Vladar HP. Amino acid fermentation at the origin of the genetic code. Biol Direct. 2012;7: 6.](http://paperpile.com/b/bcuLgy/xaYL)

59. [Esquivel-Elizondo S, Ilhan ZE, Garcia-Peña EI, Krajmalnik-Brown R. Insights into Butyrate Production in a Controlled Fermentation System via Gene Predictions. mSystems. 2017. doi:](http://paperpile.com/b/bcuLgy/VH3H)[10.1128/msystems.00051-17](http://dx.doi.org/10.1128/msystems.00051-17)

60. [Ferreyra JA, Wu KJ, Hryckowian AJ, Bouley DM, Weimer BC, Sonnenburg JL. Gut microbiota-produced succinate promotes C. difficile infection after antibiotic treatment or motility disturbance. Cell Host Microbe. 2014;16: 770–777.](http://paperpile.com/b/bcuLgy/CHq2)

61. [Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. Environmental Microbiology. 2017. pp. 29–41. doi:](http://paperpile.com/b/bcuLgy/9VNK)[10.1111/1462-2920.13589](http://dx.doi.org/10.1111/1462-2920.13589)

62. [Nawrocki KL, Wetzel D, Jones JB, Woods EC, McBride SM. Ethanolamine is a valuable nutrient source that impacts Clostridium difficile pathogenesis. Environ Microbiol. 2018;20: 1419–1435.](http://paperpile.com/b/bcuLgy/UwNL)

63. [Senger RS, Papoutsakis ET. Genome-scale model for Clostridium acetobutylicum: Part I. Metabolic network resolution and analysis. Biotechnol Bioeng. 2008;101: 1036–1052.](http://paperpile.com/b/bcuLgy/UpBb)

64. [Lachance J-C, Lloyd CJ, Monk JM, Yang L, Sastry AV, Seif Y, et al. BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. PLoS Comput Biol. 2019;15: e1006971.](http://paperpile.com/b/bcuLgy/YvCC)

65. [Biggs MB, Papin JA. Managing uncertainty in metabolic network structure and improving predictions using EnsembleFBA. PLoS Comput Biol. 2017;13: e1005413.](http://paperpile.com/b/bcuLgy/xKdL)

66. [Medlock GL, Moutinho TJ, Papin JA. Medusa: Software to build and analyze ensembles of genome-scale metabolic network reconstructions. PLoS Comput Biol. 2020;16: e1007847.](http://paperpile.com/b/bcuLgy/qODY)

67. [Atlas R, Ronald Atlas. Handbook of Microbiological Media, Fourth Edition. 2010. doi:](http://paperpile.com/b/bcuLgy/vSN5)[10.1201/ebk1439804063](http://dx.doi.org/10.1201/ebk1439804063)

68. [Blazier AS, Papin JA. Reconciling high-throughput gene essentiality data with metabolic network reconstructions. PLoS Comput Biol. 2019;15: e1006507.](http://paperpile.com/b/bcuLgy/NP7Q)

69. [Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, et al. The ClosTron: Mutagenesis in Clostridium refined and streamlined. J Microbiol Methods. 2010;80: 49–55.](http://paperpile.com/b/bcuLgy/OEVE)

70. [Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, et al. High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. MBio. 2015;6: e02383.](http://paperpile.com/b/bcuLgy/Xn6K)

71. [O’Brien EJ, Monk JM, Palsson BO. Using Genome-scale Models to Predict Biological Capabilities. Cell. 2015;161: 971–987.](http://paperpile.com/b/bcuLgy/e7oO)

72. [Ikeda D, Karasawa T, Yamakawa K, Tanaka R, Namiki M, Nakamura S. Effect of Isoleucine on Toxin Production by Clostridium difficile in a Defined Medium. Zentralblatt für Bakteriologie. 1998. pp. 375–386. doi:](http://paperpile.com/b/bcuLgy/QLVL)[10.1016/s0934-8840(98)80174-6](http://dx.doi.org/10.1016/s0934-8840(98)80174-6)

73. [Dubois T, Dancer-Thibonnier M, Monot M, Hamiot A, Bouillaut L, Soutourina O, et al. Control of Clostridium difficile Physiopathology in Response to Cysteine Availability. Infect Immun. 2016;84: 2389–2405.](http://paperpile.com/b/bcuLgy/UbbB)

74. [Scaria J, Chen J-W, Useh N, He H, McDonough SP, Mao C, et al. Comparative nutritional and chemical phenome of Clostridium difficile isolates determined using phenotype microarrays. Int J Infect Dis. 2014;27: 20–25.](http://paperpile.com/b/bcuLgy/sgDP)

75. [Battaglioli EJ, Hale VL, Chen J, Jeraldo P, Ruiz-Mojica C, Schmidt BA, et al. Clostridioides difficile uses amino acids associated with gut microbial dysbiosis in a subset of patients with diarrhea. Science Translational Medicine. 2018. p. eaam7019. doi:](http://paperpile.com/b/bcuLgy/EfpG)[10.1126/scitranslmed.aam7019](http://dx.doi.org/10.1126/scitranslmed.aam7019)

76. [Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol. 2012;3: 299.](http://paperpile.com/b/bcuLgy/n6A6)

77. [Jenior ML, Moutinho TJ Jr, Dougherty BV, Papin JA. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. PLoS Comput Biol. 2020;16: e1007099.](http://paperpile.com/b/bcuLgy/XsXX)

78. [Garrett EM, Sekulovic O, Wetzel D, Jones JB, Edwards AN, Vargas-Cuebas G, et al. Phase variation of a signal transduction system controls Clostridioides difficile colony morphology, motility, and virulence. PLoS Biol. 2019;17: e3000379.](http://paperpile.com/b/bcuLgy/Lo4H)

79. [Parthasarathy A, Cross PJ, Dobson RCJ, Adams LE, Savka MA, Hudson AO. A Three-Ring Circus: Metabolism of the Three Proteogenic Aromatic Amino Acids and Their Role in the Health of Plants and Animals. Front Mol Biosci. 2018;5: 29.](http://paperpile.com/b/bcuLgy/VSLN)

80. [Jenior ML, Leslie JL, Young VB, Schloss PD. Clostridium difficile colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. doi:](http://paperpile.com/b/bcuLgy/bARQ)[10.1101/092304](http://dx.doi.org/10.1101/092304)

81. [Poquet I, Saujet L, Canette A, Monot M, Mihajlovic J, Ghigo J-M, et al. Clostridium difficile Biofilm: Remodeling Metabolism and Cell Surface to Build a Sparse and Heterogeneously Aggregated Architecture. Frontiers in Microbiology. 2018. doi:](http://paperpile.com/b/bcuLgy/KVs5)[10.3389/fmicb.2018.02084](http://dx.doi.org/10.3389/fmicb.2018.02084)

82. [Maegawa T, Karasawa T, Ohta T, Wang X, Kato H, Hayashi H, et al. Linkage between toxin production and purine biosynthesis in Clostridium difficile. J Med Microbiol. 2002;51: 34–41.](http://paperpile.com/b/bcuLgy/hZwR)

83. [Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. 2014;42: D581–91.](http://paperpile.com/b/bcuLgy/NeV1)

84. [Faria JP, Rocha M, Rocha I, Henry CS. Methods for automated genome-scale metabolic model reconstruction. Biochem Soc Trans. 2018;46: 931–936.](http://paperpile.com/b/bcuLgy/eFPI)

85. [Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc. 2019;14: 639–702.](http://paperpile.com/b/bcuLgy/HkUi)

86. [Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. Genome Biol. 2019;20: 121.](http://paperpile.com/b/bcuLgy/z6SR)

87. [Ren Q, Chen K, Paulsen IT. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res. 2007;35: D274–9.](http://paperpile.com/b/bcuLgy/yte3)

88. [Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN, 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:](http://paperpile.com/b/bcuLgy/8s9e)[10.1101/2020.03.31.018663](http://dx.doi.org/10.1101/2020.03.31.018663)

89. [Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000. pp. 27–30. doi:](http://paperpile.com/b/bcuLgy/W6yQ)[10.1093/nar/28.1.27](http://dx.doi.org/10.1093/nar/28.1.27)

90. [Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007;35: D61–5.](http://paperpile.com/b/bcuLgy/yhG1)

91. [Baker W. The EMBL Nucleotide Sequence Database. Nucleic Acids Research. 2000. pp. 19–23. doi:](http://paperpile.com/b/bcuLgy/63Em)[10.1093/nar/28.1.19](http://dx.doi.org/10.1093/nar/28.1.19)

92. [Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Res. 2020;48: D402–D406.](http://paperpile.com/b/bcuLgy/H3oY)

93. [Hucka M, Bergmann FT, Chaouiya C, Dräger A, Hoops S, Keating SM, et al. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. J Integr Bioinform. 2019;16. doi:](http://paperpile.com/b/bcuLgy/QUCe)[10.1515/jib-2019-0021](http://dx.doi.org/10.1515/jib-2019-0021)

94. [Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol. 2013;7: 74.](http://paperpile.com/b/bcuLgy/hGyW)

95. [Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. BMC Bioinformatics. 2010;11: 489.](http://paperpile.com/b/bcuLgy/fcQw)

96. [Keaty TC, Jensen PA. gapsplit: Efficient random sampling for non-convex constraint-based models. doi:](http://paperpile.com/b/bcuLgy/oRsJ)[10.1101/652917](http://dx.doi.org/10.1101/652917)

97. [Website. [cited 16 Jun 2020]. Available:](http://paperpile.com/b/bcuLgy/NI0k) [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.](about:blank)

98. [Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359.](http://paperpile.com/b/bcuLgy/ps56)

99. [Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.](http://paperpile.com/b/bcuLgy/dsCg)

100. [Dixon P. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 2003. p. 927. doi:](http://paperpile.com/b/bcuLgy/vg7W)[10.1658/1100-9233(2003)014[0927:vaporf]2.0.co;2](http://dx.doi.org/10.1658/1100-9233(2003)014%5B0927:vaporf%5D2.0.co;2)

101. [Janitza S, Strobl C, Boulesteix A-L. An AUC-based permutation variable importance measure for random forests. BMC Bioinformatics. 2013;14: 119.](http://paperpile.com/b/bcuLgy/JrzB)