Metabolic Drivers of Virulence in Clostridioides difficile

KEYWORDS: Clostridioides difficile, metabolic modeling, transcriptomics, virulence

1. Abstract

Background and Motivation
Clostridioides difficile, a significant cause of hospital-acquired infections, exhibits increasing antibiotic resistance and recurrent infection, emphasizing the need for novel treatment strategies. This study aims to identify metabolic pathways influencing virulence, which could serve as therapeutic targets, using genome-scale metabolic network reconstructions (GENREs).

Key Findings/Contributions
• Developed GENREs for a hypervirulent isolate (R20291) and a laboratory strain (630) of C. difficile.
• Identified the pentose phosphate pathway and the usage of cytidine and N-acetylneuraminate as key metabolic drivers affecting virulence.
• Discovered that different carbon source usage correlates with varied virulence factor expression.

Methods/Approach
• Constructed GENREs using genome annotations and validated them with experimental data on gene essentiality and carbon source utilization.
• Conducted growth simulations and single-gene deletion analyses demonstrating high accuracy.
• Integrated transcriptomic data to generate context-specific models and predict metabolic activities linked to virulence.

Limitations and Open Questions
• Some metabolite groups, especially nucleotides and carboxylic acids, were underrepresented in predictions, requiring further model refinement.
• The complex regulatory networks controlling virulence need to be incorporated to fully understand metabolic-pathogenicity interplay.

Significance and Implications
These findings highlight the potential of targeting specific metabolic pathways to manage C. difficile infections. GENREs provide a robust framework for discovering novel therapeutic targets and understanding pathogen metabolism, offering broader applications for combating antibiotic-resistant bacteria.

2. Definitions and Overview

2.1 Genome-Scale Metabolic Network Reconstruction (GENRE)

A genome-scale metabolic network reconstruction (GENRE) is a comprehensive representation of the metabolic processes in an organism derived from its genomic data. It details the biochemical reactions and pathways that occur within the organism, typically encompassing all known metabolic capabilities encoded by the genome.

3. Methods

3.1 GENRE Construction from Bacterial Genome Annotations

1. Genome Annotation:

• Begin with the annotated bacterial genome. Identify and catalog genes involved in metabolic processes, including those encoding enzymes.

2. Draft Reconstruction:
• Use databases like KEGG, MetaCyc, or UniProt to associate genes with metabolic reactions. Generate a preliminary metabolic network based on these associations.

3. Gap Filling:
• Identify missing reactions or pathways (gaps) necessary for metabolic functions. Fill gaps using computational methods or known biological information to ensure network connectivity and completeness.

4. Curation:
• Manually curate the draft reconstruction. Verify gene-reaction associations by cross-referencing primary literature and experimental data.
• Refine model components to correct errors or incorporate organism-specific metabolic features.

5. Network Validation:
• Validate the metabolic network with experimental data. Compare in silico predictions with observed growth characteristics under different conditions.

6. Model Optimization:
• Optimize the metabolic model using flux balance analysis (FBA) or other constraint-based methods to predict metabolic flux distributions.

7. Documentation and Iteration:
• Document all steps and decisions made during the reconstruction process. Iteratively refine the model as new data become available.

8. Software Tools:
• Utilize tools like COBRA Toolbox, ModelSEED, and MEMOTE for automated reconstruction, analysis, and validation of metabolic networks.

Reference: Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. **Nature Protocols**, 5(1), 93–121. DOI: https://doi.org/10.1038/nprot.2009.203

Monk, J., Nogales, J., & Palsson, B. Ø. (2014). Optimizing genome-scale network reconstructions. **Nature Biotechnology**, 32(5), 447–452. DOI: https://doi.org/10.1038/nbt.2870

Seaver, S. M. D., et al. (2021). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic reconstructions. **Nucleic Acids Research**, 49(D1), D570–D581. DOI: https://doi.org/10.1093/nar/gkaa746

3.2 Databases and Annotation Tools for GENREs

1. BioCyc Database Collection:
This resource includes the **Clostridioides difficile** Pathway/Genome Database, which integrates genome data with metabolic pathways, regulatory networks, and other biological information. It offers tools for genome browsing, metabolic map visualization, and analysis of omics data. (clostridium.biocyc.org)

2. MetaCyc:
A comprehensive database of metabolic pathways and enzymes, MetaCyc provides extensive information on chemical compounds, reactions, and metabolic pathways across all domains of life. (en.wikipedia.org)

3. BRENDA:

The BRaunschweig ENzyme DAtabase is a detailed repository of enzyme functional and molecular data. It includes information on enzyme properties, functions, and significance, facilitating the annotation of metabolic reactions in GENREs. (en.wikipedia.org)

4. KEGG (Kyoto Encyclopedia of Genes and Genomes):
KEGG is a collection of databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances. (en.wikipedia.org)

5. Reactome:
A free online database of biological pathways, Reactome is manually curated and authored by PhD-level biologists. (en.wikipedia.org)

6. Metannogen:
This Java-based software allows for the annotation of existing biological networks and the reconstruction of metabolic networks. It is optimized for compartmentalized metabolic networks and provides advanced features for manual annotation and cross-referencing. (bioinformatics.org)

3.3 Gene Essentiality Determination

Gene essentiality in **Clostridioides difficile** is experimentally determined using high-throughput transposon mutagenesis coupled with sequencing techniques such as Transposon-Directed Insertion Site Sequencing (TraDIS). This method involves creating a comprehensive library of transposon mutants, each with a unique insertion site, and then sequencing these sites to identify genes that are disrupted. Genes with no or significantly fewer insertions are considered essential, as their disruption likely impairs bacterial growth or survival. For instance, a study on the epidemic strain R20291 generated over 70,000 unique mutants and identified 404 essential genes required for in vitro growth. (pubmed.ncbi.nlm.nih.gov/25714712)

3.4 GENRE Validation and Optimization

To validate predictions from GENREs, experimental datasets such as gene essentiality data and carbon source utilization profiles are typically employed. By comparing in silico predictions with empirical data, researchers can assess the accuracy of the GENREs. For example, in silico simulations have been shown to recapitulate the majority (76%) of growth phenotypes observed experimentally, indicating a high degree of agreement between model predictions and actual bacterial behavior. (pnas.org/doi/full/10.1073/pnas.2119396119)

3.5 In Silico Single-Gene Deletion Algorithms

1. Flux Balance Analysis (FBA):
This linear programming-based method evaluates the impact of gene deletions by simulating metabolic flux distributions under steady-state conditions. By constraining the fluxes associated with the deleted gene to zero, FBA predicts changes in cellular growth and metabolic production. (en.wikipedia.org)

2. Mixed-Integer Linear Programming (MILP):
MILP extends FBA by incorporating integer constraints, enabling the identification of optimal gene deletion strategies for desired phenotypic outcomes. For instance, the gDel_minRN algorithm utilizes MILP to design minimal reaction networks that achieve growth-coupled production by determining gene deletions that repress the maximum number of reactions via gene-protein-reaction associations. (pubmed.ncbi.nlm.nih.gov/36809057)

3. Deep Learning-Based Approaches:
Recent advancements have introduced machine learning techniques to predict gene deletion strategies. The DeepGDel framework employs deep learning algorithms to integrate sequential gene

and metabolite data, facilitating the automatic prediction of gene deletions for growth-coupled production. (arxiv.org/abs/2504.06316)

4. Database-Enhanced Frameworks:
Methods like DBgDel enhance computational efficiency by mining existing databases for prior information on gene deletions, integrating this data to narrow the search space and accelerate the computation of gene deletion strategies. (arxiv.org/abs/2411.08077)

5. Synthetic Lethal Analysis:
Algorithms such as Fast-SL identify sets of reactions or genes whose simultaneous deletion is lethal to the organism. This approach is valuable for understanding gene interactions and designing targeted interventions. (arxiv.org/abs/1406.6557)

3.6 Integration of Transcriptomic Data into GENREs

Methods for Integrating Transcriptomic Data:

1. GIMME (Gene Inactivity Moderated by Metabolism and Expression):
This method utilizes gene expression data to identify active and inactive genes, minimizing the inclusion of low-expression reactions while maintaining a predefined objective function value. (ncbi.nlm.nih.gov/pmc/articles/PMC9029533)

2. iMAT (Integrative Metabolic Analysis Tool):
iMAT categorizes reactions into high, moderate, and low expression groups based on transcriptomic data. It aims to maximize high-expression reactions and minimize low-expression ones, thereby aligning the metabolic model with observed gene expression patterns. (ncbi.nlm.nih.gov/pmc/articles/PMC9029533)

3. tINIT (Transcriptional Integration of Tissue-Specific GEMs):
tINIT integrates gene expression data with GEMs to predict active metabolic reactions, considering both gene expression levels and regulatory interactions to estimate transcriptional activity within the metabolic network. (pmc.ncbi.nlm.nih.gov/articles/PMC10383060)

4. MADE (Metabolic Adjustment by Differential Expression):
MADE analyzes gene expression data across multiple conditions to identify statistically significant changes, determining high- and low-expression reactions to adapt the metabolic model accordingly. (ncbi.nlm.nih.gov/pmc/articles/PMC9029533)

5. E-Flux:
This approach constrains the upper bounds of reactions classified as low-expression based on a given threshold and expression data, effectively integrating transcriptomic information into the metabolic model. (ncbi.nlm.nih.gov/pmc/articles/PMC9029533)

Software Tools Implementing These Methods:

• IgemRNA: A novel tool that combines multiple pre-processing and transcriptome analysis methods, including various thresholding, gene mapping, and reaction constraining options. It is compatible with MATLAB-based software and optionally utilizes COBRA Toolbox 3.0 functionality. (ncbi.nlm.nih.gov/pmc/articles/PMC9029533)

• COMO: A user-friendly pipeline that integrates multi-omics data processing, context-specific metabolic model development, simulations, drug databases, and disease data to aid drug discovery. COMO can process bulk and single-cell RNA-seq, microarrays, and proteomics data to develop context-specific metabolic models. (arxiv.org/abs/2011.02103)

• SWIFTCORE: A tool designed for the context-specific reconstruction of genome-scale metabolic networks, facilitating the integration of transcriptomic data to generate tissue-specific models.

(bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-3440-y)

• IntLIM (Integration through Linear Modeling): An R package that evaluates phenotype-specific relationships between gene and metabolite levels, aiding in the integration of transcriptomic and metabolomic data. (bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2085-6)

3.7 Model Curation and Comparative Analysis

Reconstruction (GENRE) analysis of bacteria has enabled systematic investigation of the genetic and metabolic properties that contribute to downstream virulence phenotypes. With this in mind, we generated and extensively curated C. difficile GENREs.

GENREs were compared against in vitro gene essentiality and carbon utilization.

Measurement of GENRE performance is the comparison of predicted essential genes for growth in silico and those found to be essential experimentally through forward genetics screens.

analyze ensembles of genome-scale metabolic network reconstructions.

context-specific metabolic network analysis.

context-specific models of C. difficile metabolism by integrating transcriptomic data.

Through contextualization of each model using transcriptomes generated from in vitro hypervirulent isolate (strain [str.] R20291) and a historic strain (str. 630), validating both with in vitro and in vivo data sets. Growth simulations revealed signi■cant corre-

focused on the best-characterized hypervirulent isolate, str. R20291. However, to maximize the utility of the bulk of published C. difficile metabolic research, we elected to

and infection conditions, we discovered reliance on the pentose phosphate pathway

essentiality results for glycolysis and the pentose phosphate pathway across both the rough and smooth phase variant context-specific models.

4. Clostridioides difficile Strain Differences

Clostridioides difficile strain R20291, a hypervirulent PCR ribotype 027 isolate, exhibits distinct genomic and phenotypic characteristics compared to the laboratory strain 630 (ribotype 012).

4.1 Genomic Differences

• Additional Genes: R20291 possesses 234 genes absent in strain 630, contributing to differences in motility, antibiotic resistance, and toxicity.
(genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-9-r102)

• Unique Genetic Regions: R20291 contains five unique regions not found in strain 630, including a novel phage island, a two-component regulatory system, and transcriptional regulators.
(genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-9-r102)

4.2 Phenotypic Differences

• Motility: R20291 demonstrates enhanced motility compared to strain 630, likely due to the additional genes related to flagellar function.
(genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-9-r102)

• Toxin Production: R20291 produces higher levels of toxins A and B, contributing to its increased virulence. (genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-9-r102)

• Antibiotic Resistance: R20291 exhibits increased resistance to certain antibiotics, potentially due to the acquisition of additional resistance genes. (genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-9-r102)

5. Metabolic Pathways Influencing Virulence

5.1 Pentose Phosphate Pathway (PPP)

The pentose phosphate pathway (PPP) is a crucial metabolic route in bacterial cells, serving both anabolic and catabolic functions. It operates parallel to glycolysis and is divided into two phases:

1. Oxidative Phase: This phase oxidizes glucose-6-phosphate, producing ribulose-5-phosphate, carbon dioxide, and NADPH. NADPH is essential for reductive biosynthetic reactions, including fatty acid and nucleotide synthesis, and for maintaining cellular redox balance. (en.wikipedia.org)

2. Non-Oxidative Phase: This phase generates ribose-5-phosphate for nucleotide synthesis and intermediates like erythrose-4-phosphate and sedoheptulose-7-phosphate, which are precursors for aromatic amino acids and other biosynthetic pathways. (en.wikipedia.org)

In pathogenic bacteria, the PPP influences virulence by:

• Providing Biosynthetic Precursors: The PPP supplies ribose-5-phosphate for nucleotide synthesis and erythrose-4-phosphate for aromatic amino acids, essential for DNA replication, RNA transcription, and protein synthesis, all critical for bacterial proliferation and virulence. (pmc.ncbi.nlm.nih.gov/articles/PMC4540601)

• Generating NADPH: NADPH produced by the PPP is vital for biosynthetic processes and for counteracting oxidative stress, enabling pathogens to survive hostile environments within the host. (pmc.ncbi.nlm.nih.gov/articles/PMC4540601)

• Influencing Virulence Factor Production: In **Staphylococcus aureus**, disruptions in the PPP affect energy production and nucleotide synthesis, leading to impaired biofilm formation and reduced resistance to host defenses, thereby diminishing virulence. (pmc.ncbi.nlm.nih.gov/articles/PMC10368262)

While specific studies on **Clostridioides difficile** (C. difficile) are limited, the PPP's fundamental roles in metabolism and stress response suggest it likely contributes to the pathogen's virulence.

5.2 Cytidine Catabolism

In bacterial metabolic networks, cytidine catabolism primarily involves the conversion of cytidine to uridine, which is then further degraded. The key enzymes facilitating this process are cytidine deaminase and uridine phosphorylase.

• Cytidine Deaminase (CDA):
This enzyme catalyzes the hydrolytic deamination of cytidine, converting it into uridine and releasing ammonia. This reaction is crucial in pyrimidine metabolism, influencing both salvage and catabolic pathways. CDA is widely distributed among bacteria. (en.wikipedia.org)

• Uridine Phosphorylase:
Following the action of cytidine deaminase, uridine phosphorylase catalyzes the phosphorolysis of uridine, resulting in the production of uracil and ribose-1-phosphate. This step is integral to the further breakdown and utilization of pyrimidine nucleosides. (pubmed.ncbi.nlm.nih.gov/7600008)

These enzymatic activities have been observed in various bacterial species. For instance, in **Sphingomonas paucimobilis**, both cytidine deaminase and uridine phosphorylase are active,

facilitating the utilization of cytidine as a nitrogen source. (pubmed.ncbi.nlm.nih.gov/7600008)

Additionally, in **Escherichia coli**, the metabolism of cytidine involves its conversion to uridine by cytidine deaminase, followed by further degradation. (pmc.ncbi.nlm.nih.gov/articles/PMC11792562)

Utilization of N-acetylneuraminic acid and cytidine decreases sporulation in C.

5.3 N-Acetylneuraminate (Sialic Acid) Utilization

N-Acetylneuraminate, commonly known as sialic acid, is a nine-carbon sugar acid predominantly found as a terminal component of glycoproteins and glycolipids on mammalian cell surfaces. It plays a crucial role in cellular recognition processes, including pathogen-host interactions.

Clostridioides difficile (C. difficile) lacks the enzymatic machinery to cleave sialic acid directly from host mucins. Instead, it relies on other gut microbiota, such as Bacteroides thetaiotaomicron, to liberate free sialic acid from mucins. Once free, C. difficile can uptake and metabolize sialic acid through the nan operon, which encodes the necessary transport and catabolic enzymes. This operon includes genes such as nanE, nanA, nanT, and nanK, facilitating the conversion of sialic acid into intermediates like pyruvate and acetate, which feed into central metabolic pathways.
(journals.plos.org/article?id=10.1371/journal.ppat.1011034)

The ability to utilize sialic acid provides C. difficile with a competitive advantage during colonization, especially following antibiotic treatment that disrupts the normal gut microbiota.
(journals.plos.org/article?id=10.1371/journal.ppat.1011034)

To then assess the effect of N-acetylneuraminate and cytidine on sporulation in C.

6. Regulatory Networks Linking Metabolism and Virulence

Clostridioides difficile regulates virulence gene expression through several key regulatory networks that interact closely with metabolic pathways. The primary regulators include CcpA, CodY, Rex, PrdR, SigH, and Spo0A.

• CcpA (Catabolite Control Protein A):
Mediates carbon catabolite repression by responding to fructose-1,6-bisphosphate levels, binding to promoter regions of toxin genes (particularly tcdR) and repressing their expression in the presence of rapidly metabolizable sugars like glucose. (europepmc.org/articles/PMC4398617)

• CodY:
A global