Metabolic Drivers of Clostridioides difficile Virulence Uncovered

Summary

Summary Title:

Novel Metabolic Drivers of Clostridioides difficile Virulence Identified via Metabolic Network Analysis

Background and Motivation:

Clostridioides difficile is a leading cause of hospital-acquired infections, primarily due to its toxin-mediated diarrhea and increasing antibiotic resistance. Traditional antibiotic treatments can disrupt gut microbiota, providing a niche for C. difficile colonization. The research focuses on targeting C. difficile's metabolism to combat infections as an alternative to antibiotics. Genome-scale metabolic network reconstructions (GENREs) offer promising avenues for identifying metabolic processes linked to bacterial virulence.

Key Findings/Contributions:

• GENREs were constructed and curated for both a hypervirulent strain (R20291) and a historic laboratory strain (630) of C. difficile.
• These models were validated with in vitro and in vivo data, showing high predictive accuracy for carbon source usage and gene essentiality.
• Specific metabolic pathways, such as the pentose phosphate pathway and others involved in sporulation and biofilm formation, were identified as drivers of virulence.
• Experimental results confirmed that specific metabolites reduced virulence expression by influencing metabolic pathways.

Methods/Approach:

• GENRE Construction: Developed for two bacterial strains, involving genome annotation, proteome alignment, and manual curation of known metabolic pathways.
• Validation: Validation against in vitro gene essentiality and carbon utilization data; correlation analysis with experimental data showed high predictiveness.
• Context-specific Modeling: Integrated transcriptomic data for in vitro and infection conditions to tailor metabolic models and identify key metabolic shifts associated with virulence.

Limitations and Open Questions:

• While predictive, the models under-represent certain metabolite groups due to limited annotation.
• More complex regulatory networks need further exploration to fully understand metabolism-driven virulence.
• The study opens avenues for assessing the metabolic interactions with gut microbiota, warranting further investigation into probiotic interventions.

Significance and Implications:

• This study enhances the understanding of how metabolic pathways regulate bacterial virulence, offering potential novel therapeutic targets that do not rely on conventional antibiotics.
• The findings have broader implications for treating antibiotic-resistant C. difficile strains by unveiling specific metabolic vulnerabilities that can be therapeutically targeted.
• GENREs serve as robust tools for metabolic research, facilitating the discovery of metabolism-directed therapies in infectious diseases.

Clostridioides difficile is a leading cause of hospital-acquired infections, primarily due to its toxin-mediated diarrhea and increasing antibiotic resistance. Traditional antibiotic treatments can disrupt gut microbiota, providing a niche for C. difficile colonization. The research focuses on targeting C. difficile's metabolism to combat infections as an alternative to antibiotics. Genome-scale metabolic network reconstructions (GENREs) offer promising avenues for identifying metabolic processes linked to bacterial virulence.

GENRE Construction

Constructing a genome-scale metabolic network reconstruction (GENRE) for a bacterial pathogen like **Clostridioides difficile** involves several key steps and tools. Here is an overview:

1. Genome Annotation:
• Tools: Prodigal, RAST, or the NCBI Prokaryotic Genome Annotation Pipeline.
• Annotate the genome to identify genes and predict their functions.

2. Draft Model Construction:
• Tools: Pathway Tools, ModelSEED, or KBase.
• Use genome annotations to create an initial draft metabolic model by associating genes with metabolic reactions.

3. Metabolite and Reaction Identification:
• Identify and curate reactions and metabolites from databases such as KEGG, MetaCyc, and BRENDA.

4. Gap-Filling:
• Tools: GapFind and GapFill algorithms.
• Address gaps in the network by adding reactions that allow the model to produce essential biomass components.

5. Model Refinement and Curation:
• Manually curate the model using literature and experimental data to correct errors and add missing reactions.
• Validate and refine the model using data from experimental studies to ensure physiological relevance.

6. Network Validation and Simulation:
• Tools: COBRA Toolbox, OptFlux, or CellNetAnalyzer.
• Simulate growth conditions and compare predicted growth rates with experimental data.

7. Integration with Omics Data:
• Incorporate transcriptomic, proteomic, and metabolomic data to refine and contextualize the model.

8. Iterative Refinement:
• Continuous model validation and updates based on new experimental findings.

Key Resources:

• Oberhardt, M. A., Palsson, B. Ø., & Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. **Molecular Systems Biology**, 5, 320.
• Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. **Nature Protocols**, 5(1), 93-121.

Proteome Alignment

Proteome alignment plays a crucial role in the curation of genome-scale metabolic reconstructions (GENREs) by facilitating the accurate annotation of genes and their associated metabolic functions. This process involves aligning protein sequences to genomic data to identify gene-protein-reaction (GPR) associations, which are essential for constructing comprehensive and functional metabolic models.

One notable tool in this domain is miniprot, a protein-to-genome aligner designed for mapping protein sequences to complete genomes. Miniprot integrates advanced techniques such as k-mer sketching and SIMD-based dynamic programming, offering significant speed improvements over existing tools while maintaining comparable accuracy. Its efficiency makes it particularly suitable for annotating genes in non-model organisms, thereby enhancing the quality of GENREs. ([arxiv.org](https://arxiv.org/abs/2210.08052?utm_source=openai))

In addition to miniprot, several other software tools and algorithms are commonly employed for proteome alignment during GENRE curation:
• CoReCo: This algorithm facilitates the automatic reconstruction of metabolic models for related species by leveraging protein homology alignments. It utilizes KEGG as a reaction database and incorporates automatic gap filling using atom maps of reactions, producing functional models ready for simulation. ([en.wikipedia.org](https://en.wikipedia.org/wiki/Metabolic_network_modelling?utm_source=openai))
• RAVEN Toolbox: RAVEN performs genome-wide functional annotations using template models or KEGG as sources for protein homology alignments. It is widely used for semi-automatic reconstruction of metabolic models, aiding in the identification of GPR associations. ([journals.scholarsportal.info](https://journals.scholarsportal.info/html/22181989/v12i0001/nfp_gmmeiuobd.xml?utm_source=openai))
• Pathway Tools: This bioinformatics software enables the reconstruction and prediction of metabolic pathways, including reaction atom mappings and metabolic route searches. It contains MetaFlux gap filler that automatically identifies missing reactions, nutrients, and secretions, thereby enhancing the completeness of metabolic models. ([en.wikipedia.org](https://en.wikipedia.org/wiki/Metabolic_network_modelling?utm_source=openai))

Gene Essentiality Validation In Vitro

To validate gene essentiality in **Clostridioides difficile** (**C. difficile**) in vitro, several standard assays and methods are employed:

1. Transposon Mutagenesis and High-Throughput Sequencing (TraDIS):
• Method: A comprehensive transposon library is generated by inserting transposons randomly into the **C. difficile** genome. High-throughput sequencing identifies insertion sites.
• Interpretation: Genes lacking transposon insertions are considered essential, as disruptions in these genes likely prevent bacterial survival under the tested conditions. ([journals.asm.org](https://journals.asm.org/doi/full/10.1128/mbio.02383-14?utm_source=openai))

2. Targeted Gene Knockout with Complementation:
• Method: Specific genes are inactivated using techniques like the ClosTron system, which employs group II intron-mediated gene disruption. To confirm essentiality, a second functional copy of the gene is introduced elsewhere in the genome.
• Interpretation: If the bacterium remains viable only when the functional gene copy is present, the gene is deemed essential. ([pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC5154367/?utm_source=openai))

3. In Silico Metabolic Modeling:
• Method: Genome-scale metabolic network reconstructions (GENREs) are used to simulate metabolic fluxes. Computational analyses, such as Flux Balance Analysis (FBA), predict the impact of individual gene deletions on biomass production.
• Interpretation: Genes whose in silico deletion significantly reduces or halts biomass production are predicted to be essential. ([bmcsystbiol.biomedcentral.com](https://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-014-0117-z?utm_source=openai))

4. Flux Sampling with Graph Neural Networks (FluxGAT):

• Method: This approach integrates flux sampling data with graph neural networks to predict gene essentiality without relying on predefined cellular objective functions.
• Interpretation: Genes identified as essential through this unbiased computational method are considered critical for bacterial survival.
([arxiv.org](https://arxiv.org/abs/2403.18666?utm_source=openai))

Carbon Source Utilization Assessment

Assessing carbon source utilization in bacteria involves several techniques:

1. Phenotype Microarrays (PMs): These high-throughput platforms evaluate bacterial growth on numerous carbon sources simultaneously. Each well in a microplate contains a distinct carbon substrate, and bacterial growth is monitored, often using colorimetric indicators.

2. Biolog Phenotype Microarrays: A specific type of PM, Biolog plates are widely used to assess metabolic capabilities by measuring bacterial respiration in the presence of various carbon sources.

3. Simmons' Citrate Agar Test: This method determines a bacterium's ability to utilize citrate as its sole carbon source. Growth and a color change in the medium indicate positive utilization. ([en.wikipedia.org](https://en.wikipedia.org/wiki/Simmons%27$_{citrate}$agar?utm_source=openai))

4. Genome-Scale Metabolic Network Reconstructions (GENREs): These computational models predict metabolic capabilities based on genomic data, allowing in silico analysis of carbon source utilization. ([pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC8547418/?utm_source=openai))

For **Clostridioides difficile**, Biolog Phenotype Microarrays are predominantly used to assess carbon source utilization. Studies have employed these microarrays to profile the growth of various **C. difficile** strains on numerous carbon substrates, providing insights into their metabolic phenotypes. ([pmc.ncbi.nlm.nih.gov](https://pmc.ncbi.nlm.nih.gov/articles/PMC11343193/?utm_source=openai))

Statistical and Computational Validation of GENREs

In the validation of genome-scale metabolic network reconstructions (GENREs), several statistical and computational methods are employed to correlate model predictions with experimental data:

1. Flux Balance Analysis (FBA): This constraint-based approach predicts steady-state flux distributions by optimizing an objective function, typically biomass production. FBA predictions are validated by comparing in silico growth rates or growth/no-growth phenotypes across various substrates, growth conditions, or gene knockouts with experimental observations. ([ncbi.nlm.nih.gov](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10055486/?utm_source=openai))

2. Flux Variability Analysis (FVA): FVA assesses the range of possible fluxes through each reaction under given constraints, identifying reactions with low variability that may be critical to the organism. This analysis helps in understanding the flexibility and robustness of metabolic networks. ([en.wikipedia.org](https://en.wikipedia.org/wiki/Flux$_{balance}$analysis?utm_source=openai))

3. 13C-Metabolic Flux Analysis (13C-MFA): By tracing the distribution of 13C-labeled substrates through metabolic pathways, 13C-MFA provides detailed insights into intracellular fluxes. These experimentally determined fluxes are compared with model predictions to validate and refine GENREs. ([ncbi.nlm.nih.gov](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10055486/?utm_source=openai))

4. Bayesian Metabolic Flux Analysis: This probabilistic approach models the entire metabolic system, inferring flux distributions based on experimental data and prior knowledge. It accounts for uncertainties and reveals flux couplings, offering a comprehensive validation framework. ([arxiv.org](https://arxiv.org/abs/1804.06673?utm_source=openai))

5. Machine Learning Techniques: Advanced algorithms analyze discrepancies between model predictions and experimental data, identifying key metabolic fluxes and reaction pathways that influence model accuracy. For instance, machine learning has been used to highlight the importance of

hydrogen ion exchange and central metabolism branch points in determining model accuracy. ([embopress.org](https://www.embopress.org/doi/full/10.15252/msb.202311566?utm_source=openai))

6. Statistical Metrics: Metrics such as the area under the precision-recall curve (AUPRC) quantify the accuracy of model predictions. AUPRC is particularly useful in evaluating the correspondence between predicted and observed phenotypes, especially in high-throughput mutant fitness data. ([embopress.org](https://www.embopress.org/doi/full/10.15252/msb.202311566?utm_source=openai))

Context-Specific Metabolic Modeling

Context-specific metabolic modeling involves integrating transcriptomic data into genome-scale metabolic reconstructions (GENREs) to tailor the models to specific conditions or tissues. This integration helps refine the metabolic network by emphasizing reactions that are more likely active under given conditions, based on gene expression levels.

Key Algorithms Used

1. GIMME (Gene Inactivity Moderated by Metabolism and Expression):
This method incorporates transcriptomic data to minimize the discrepancies between gene expression data and the predicted metabolic activity. It selects a subset of reactions that fulfills a specific objective while ensuring that reactions associated with highly expressed genes are active.
Reference: Becker, S. A., & Palsson, B. O. (2008). "Context-Specific Metabolic Networks Are Consistent with Experiments," PLOS Computational Biology.

2. iMAT (Integrative Metabolic Analysis Tool):
iMAT uses gene expression data to categorize reactions into three groups: highly expressed, moderately expressed, and not expressed. The algorithm then finds an optimal configuration that maximizes the consistency between the model and high expression levels.
Reference: Zur, H., & Ruppin, E. (2007). "iMAT: An Integrative Metabolic Analysis Tool," Molecular Systems Biology.

3. INIT (Integrative Network Inference for Tissues):
This method uses a scoring system based on transcriptomic data to prioritize reactions, creating a tissue-specific model. It emphasizes the importance of tissue-specific reactions by integrating expression data with the metabolic model.
Reference: Jensen, P. A., Lutz, K. A., & Papin, J. A. (2011). "Tissue-specific modeling of human metabolism: an initial step toward drug discovery," BMC Systems Biology.

4. FASTCORE:
FASTCORE is an algorithm designed to extract context-specific metabolic networks quickly. It starts with a core set of reactions based on expression data and iteratively adds reactions to ensure metabolic functionality.
Reference: Vlassis, N., Pacheco, M. P., & Sauter, T. (2014). "Fast reconstruction of compact context-specific metabolic network models," PLoS Computational Biology.

Manual Curation of Metabolic Pathways

In the manual curation of metabolic pathways for bacterial genome-scale metabolic reconstructions (GENREs), several key databases and resources are routinely utilized:

1. KEGG (Kyoto Encyclopedia of Genes and Genomes): KEGG provides comprehensive information on metabolic pathways, enzymes, and chemical compounds across various organisms. It serves as a foundational resource for understanding metabolic networks and their components. ([en.wikipedia.org](https://en.wikipedia.org/wiki/KEGG?utm_source=openai))

2. MetaCyc: This database offers extensive data on metabolic pathways and enzymes, curated from a vast array of scientific literature. MetaCyc encompasses pathways from all domains of life, making it invaluable for annotating metabolic functions in bacterial GENREs.

([en.wikipedia.org](https://en.wikipedia.org/wiki/MetaCyc?utm_source=openai))

3. BiGG (Biochemical, Genetic, and Genomic): BiGG is a repository of genome-scale metabolic models, providing high-quality, manually curated models that include detailed information on reactions and metabolites. It facilitates the development and validation of bacterial metabolic reconstructions. ([onlinelibrary.wiley.com](https://onlinelibrary.wiley.com/doi/full/10.1007/s40484-017-0108-3?utm_source=openai))

4. BRENDA (Braunschweig Enzyme Database): BRENDA is a comprehensive enzyme information system, offering detailed data on enzyme functions, structures, and kinetics. It is instrumental in annotating enzymatic activities within metabolic pathways. ([en.wikipedia.org](https://en.wikipedia.org/wiki/BRENDA?utm_source=openai))

5. EcoCyc: Specifically focused on **Escherichia coli**, EcoCyc provides an in-depth database of its genome and metabolic pathways. It serves as a model for curating metabolic information in other bacterial species. ([en.wikipedia.org](https://en.wikipedia.org/wiki/EcoCyc?utm_source=openai))

6. Reactome: Reactome is a free, manually curated database of biological pathways, including metabolic processes. It offers detailed pathway diagrams and annotations, aiding in the understanding of complex metabolic networks. ([en.wikipedia.org](https://en.wikipedia.org/wiki/Reactome?utm_source=openai))

7. BioCyc Database Collection: This collection includes numerous organism-specific Pathway/Genome Databases (PGDBs), providing curated genome and metabolic pathway information for a wide range of bacteria. ([en.wikipedia.org](https://en.wikipedia.org/wiki/BioCyc_database_collection?utm_source=openai))

Pentose Phosphate Pathway (PPP)

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

1. Oxidative Phase: This phase generates NADPH by oxidizing glucose-6-phosphate to ribulose-5-phosphate. NADPH is essential for reductive biosynthetic reactions, such as fatty acid and nucleotide synthesis, and for maintaining cellular redox balance. ([en.wikipedia.org](