Curated *Clostridioides difficile* 630 Database in BioCyc.org

Amanda Mackie[2], Ingrid M. Keseler[1], Ian Paulsen[2], David S. Weiss[3], Peter D. Karp (pkarp@ai.sri.com)[1]

[1]SRI International, Menlo Park, CA, USA; [2]Macquarie University, Sydney, Australia; [3]University of Iowa, Iowa City, IA, USA

cdifficile.biocyc.org

Overview

BioCyc.org is a web portal that contains databases for 18,000 sequenced microbes, including 15 *Clostridioides difficile* strains. BioCyc supports a variety of use cases including information seeking, Omics data analysis, and comparative analysis.

A curated database resource for *Clostridioides difficile* 630 was included in BioCyc in 2015. Enzymatic functions predicted in the genome annotation were computationally integrated into a predicted metabolic pathway network, and predicted operons were included in the database. The database was further improved by manual curation of the predicted metabolic and regulatory network, and by the addition of summaries, literature references and Gene Ontology terms for experimentally studied gene products.

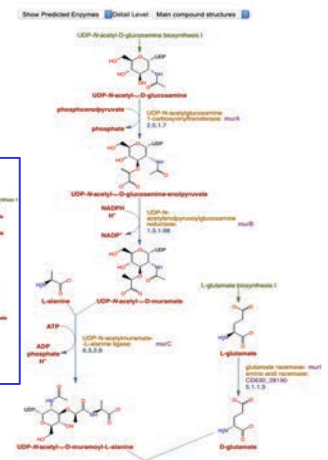| Manually Curated Experimental Evidence	Number
Gene products	131
Transcription units	17
Regulatory interactions	40
Gene Ontology terms	226

The BioCyc Database Collection

The BioCyc website provides an extensive set of bioinformatics tools for searching and analyzing these databases and leveraging them for analysis of omics datasets. Genome-related tools include a genome browser, sequence search and alignment tools, and extraction of sequence regions. Pathway-related tools include pathway diagrams, navigation of searchable and zoomable organism-specific metabolic map diagrams, and a tool for searching for metabolic routes that connect metabolites of interest. Operons, regulatory sites, and the full organism regulatory network can be displayed. Comparative analysis tools enable comparisons of genome organization, of orthologs, and of pathway complements. Omics data analysis tools support enrichment analysis and painting of transcriptomics and metabolomics data onto individual pathways and the full metabolic map diagrams. The Omics Dashboard tool enables interactive exploration of omics datasets through a hierarchy of cellular systems.

Learn how to use BioCyc through Webinars.

Information about each gene product is summarized on the individual **gene/protein pages**. Evidence code icons show whether experimental information is available and link to the primary literature. The Operations menu on the right enables access to gene/protein sequences and to a number of comparison operations, including alignments with orthologs in other organisms.
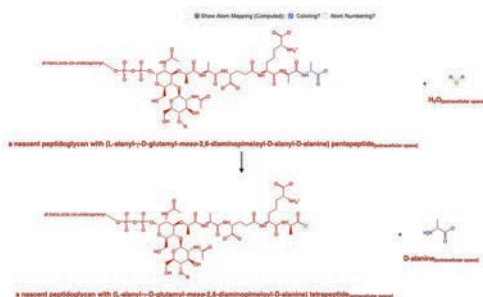
Reaction pages contain chemical structures and atom mapping (if available).
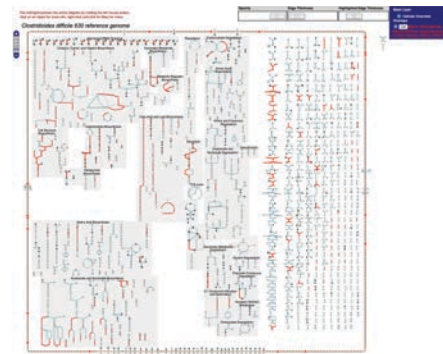

Metabolic pathways can be displayed at different levels of resolution. This partial pathway shows structures of the main compounds, EC numbers of the reactions, and the enzymes catalyzing them. The inset shows the entire pathway at the minimal detail level. Objects on the page are clickable; click here to open this pathway, then click on the name of a metabolite, or on a reaction arrow, to navigate to the compound and reaction pages.

The **Cellular Overview** diagram shows the metabolic network of *C. difficile* 630. Lines depict reactions, and dots depict metabolites. Biosynthetic pathways are shown to the left, and catabolic pathways are to their right. Here we have highlighted in red orthologs of **essential genes** identified in *C. difficile* R20291 by Dembek et al., MBio 6(2):e02383 (2015). Click here to open the online diagram; it can be zoomed interactively and can be searched using a menu on the right side of the web page. Other types of data, such as gene expression, proteomics or metabolomics data, can be painted onto this overview.
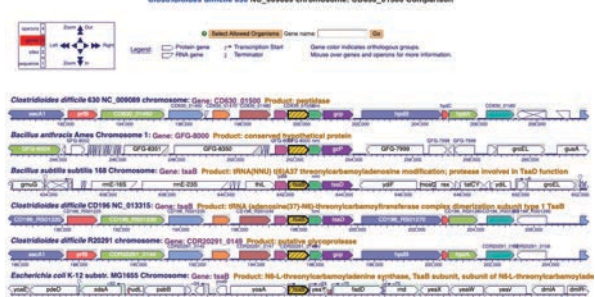
Orthologs in different organisms can be aligned within their genomic context in the **Multi-Genome Browser**. This display is color-coded by gene orthology. A grey background indicates predicted transcription units and a light green background indicates experimentally verified TUs. Shown here are orthologs of a gene that is essential for growth in *C. difficile* R20291. The orthologous gene product in *E. coli* functions in threonylcarbamoyladenosine biosynthesis (a tRNA modification). Clicking on a gene navigates to its gene/protein page.

References

BioCyc: Briefings in Bioinformatics 20(4): 1085-193 (2017)
Pathway Tools: Briefings in Bioinformatics 2019 Dec 8;bbz104 (2019)