Genome-scale metabolic and regulatory network reconstruction of *Caldicellulosiruptor bescii*

Ying Zhang[1]*(yingzhang@uri.edu), Ke Zhang, [1] Irina A. Rodionova,[2] Dmitry A. Rodionov,[2] James R. Crosby,[3] Ryan G. Bing,[3] Diep M. N. Nguyen,[4] Tania N. N. Tanwee,[4] Gabriel M. Rubinstein,[4] Robert M. Kelly,[3] and **Michael W. W. Adams[4]**

[1]University of Rhode Island, Kingston, RI; [2]Sanford-Burnham-Prebys Med. Discovery Institute, La Jolla, CA; [3]North Carolina State University, Raleigh, NC; [4]University of Georgia, Athens, GA;

Project Goals: We are using systems biology-guided approaches to develop a non-model, microbial metabolic engineering platform based on the most thermophilic lignocellulose-degrading organism known, *Caldicellulosiruptor bescii,* which grows optimally near 80°C. This work leverages recent breakthrough advances in the development of molecular genetic tools for this organism, complemented by a deep understanding of its metabolism and physiology gained over the past decade of study in the PIs' laboratories. We are applying the latest metabolic reconstruction and modeling approaches to optimize biomass to product conversion using switchgrass as the model plant and acetone and other fermentation products as targets. The over-arching goal is to demonstrate that a non-model microorganism, specifically an extreme thermophile, can be a strategic metabolic engineering platform for industrial biotechnology using a systems biology-based approach.

Caldicellulosiruptor bescii is an extremely thermophilic, strictly anaerobic, gram-positive bacterium. It is the most thermophilic cellulolytic bacterium known to date (T_{opt}=78~80°C, T_{max}=90°C), and it can use a wide range of simple and complex carbohydrates. Its ability to degrade plant biomass without enzymatic or chemical pretreatments and at a high optimum growth temperature offers several advantages for industrial applications. Engineered *C. bescii* strains has been shown to produce desired bioproducts, such as ethanol, from un-pretreated plant biomass through consolidated bioprocessing (CBP). However, efficient metabolic engineering requires in-depth understanding of its metabolic and transcriptional regulatory networks.

In this study, we applied a subsystems-based approach combining comparative genomics, transcriptional regulon prediction, and genome-scale modeling to reconstruct an integrated view of the metabolic and regulatory network of *C. bescii*. The complete genomes of thirteen species of *Caldicellulosiruptor* were used for ortholog mapping and comparative analysis. Functional gene assignments, genome context analysis, comparative analysis of orthologous genes and DNA upstream regions, gene co-occurrence analysis and protein similarity searches were performed in the SEED environment (1). We also used genome annotations from Swiss-Prot, KEGG, TCDB, and RegPrecise databases and published experimental data. The previously generated RNASeq datasets for whole-genome gene expression and transcriptional start sites obtained for *C. bescii* grown on five different carbon sources (glucose, xylose, cellobiose, xylan, cellulose) were used for validation of reconstructed transcriptional regulons and for refinement of transporter specificities. The curation of a genome-scale model (GEM) was done with the support of PSAMM software (2) to incorporate the

known and predicted metabolic functions of enzymes and transporters.

The global reconstruction of carbohydrate utilization includes almost 200 *C. bescii* genes, encoding enzymes, transporters and transcription factors (TFs), involved in more than 20 distinct pathways for the utilization of various carbohydrates. Using comparative genomics, we identified *ab initio* novel DNA-binding motifs and reconstructed regulatory networks for 24 TFs controlling individual sugar catabolic pathways. The global reconstruction of carbohydrate utilization is being integrated with a *C. bescii* GEM, which contains 721 metabolites and 772 metabolic reactions associated with 520 metabolic genes, covering 17% or the entire genome and 73% of the metabolic and carbohydrate transport genes in *C. bescii*. Besides the carbohydrate utilization pathways, the *C. bescii* GEM contains a diverse range of catabolic and anabolic pathways, including central carbon metabolism and the biosynthesis of proteins, nucleotides, lipids, vitamins, and cofactors. Biomass production of *C. bescii* is represented in the GEM using a biomass objective function that is carefully calibrated using experimental measurements of major cell components. Growth predictions made by the *C. bescii* GEM is validated through matching metabolic simulations to with growth measurements in batch and chemostat culture using defined media. This model serves as a stepping stone for the engineering of *C. bescii* strains to enable and enhance the yields of bio-based fuels and chemicals.

IMPORTANCE In this study, we built a predictive model for simulating the metabolism of the non-model organism, *C. bescii*. The simulation predictions made by simulations can provide potential directions for the more efficient metabolic engineering of *C. bescii*.

References

1. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42:D206-14 (doi: 10.1093/nar/gkt1226)

2. Steffensen JL, Dufault-Thompson K, Zhang Y (2016) PSAMM: A Portable System for the Analysis of Metabolic Models. PLoS Comput Biol 12. (doi:10.1371/journal.pcbi.1004732)

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019391