

Global Models for Metabolism and Transcriptional Regulation in *Caldicellulosiruptor* Strains

Ying Zhang^{1*}(yingzhang@uri.edu), Ke Zhang,¹ Alex Arzamasov,² Dmitry A. Rodionov,² James R. Crosby,³ Ryan G. Bing,³ Diep M. N. Nguyen,⁴ Gabriel M. Rubinstein,⁴ Robert M. Kelly,³ and **Michael W. W. Adams⁴**

¹University of Rhode Island, Kingston, RI; ²Sanford-Burnham-Prebys Med. Discovery Institute, San Diego, CA; ³North Carolina State University, Raleigh, NC; ⁴University 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 3-hydroxypropionate as products. 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.

The thermophilic cellulolytic bacterium *Caldicellulosiruptor bescii* has extensive and highly diversified carbohydrate utilization machinery. A detailed reconstruction of this machinery including uptake mechanisms, biochemical transformations and transcriptional regulation is of key importance for the scope of our DOE-sponsored project involving metabolic engineering of *Caldicellulosiruptor* species for unpretreated lignocellulose conversion to bioproducts. Accurate functional assignment of carbohydrate utilization genes is challenging due to substantial variations of the respective pathways between species including frequent non-orthologous gene displacements and functionally divergent paralogs. A global phylogenomic analysis of 13 *Caldicellulosiruptor* and other thermophilic species in Clostridia positions our target organism at a distant branch from other widely studied model organisms. Therefore, homology-based annotations in existing genomic databases are often incomplete and imprecise. To address this challenge we combine a subsystems-based approach to pathway analysis (implemented in the SEED genomic platform) with *in silico* reconstruction of metabolic networks and transcriptional regulatory networks.

We have obtained a first draft of metabolic reconstruction of *C. bescii* that contains 530 metabolic genes, 787 metabolites (non-unique) and 695 metabolic reactions. The draft reconstruction is represented in a YAML format and curated with support of the PSAMM

software [1] to incorporate all published experimental data and inferences about enzyme activities and substrate specificities. Further, differential gene expression patterns on various carbohydrates will be integrated into the simulation of metabolic activities. To reconstruct transcriptional regulons in thirteen available *Caldicellulosiruptor* genomes we identified previously known regulons (e.g., arabinose regulon AraR) based on ortholog mapping and applied *ab initio* prediction of novel regulons (e.g., fucose regulon FucR) using available transcriptomics data for *C. bescii* and *C. saccharolyticus* grown on different mono- and polysaccharides. Reconstructed regulons provide an additional layer of genome context, helping to significantly improve the accuracy of functional annotations and metabolic reconstruction. For instance, the inferred AraR regulon includes a novel functional variant of arabinose isomerase gene, which is non-orthologous to previously characterized enzymes.

Our analysis also revealed substantial differences in sugar catabolic pathways between *Caldicellulosiruptor* and other previously studied bacteria. The repertoire of transporters and regulators involved in sugar catabolism in *Caldicellulosiruptor* demonstrate the most prominent differences in comparison with other taxa. For example, by analyzing the Rex regulon, an NADH-sensing transcriptional regulator, among the thirteen *Caldicellulosiruptor* genomes, we identified conserved core regulon including [FeFe]- and [NiFe]-hydrogenases, various ferredoxin-dependent oxidoreductases and enzymes involved in central carbon metabolism. Using the systems-based approach, we will further identify specificities of these carbon utilization enzymes and provide new information to further complete the reconstruction of metabolic and regulatory networks.

This study illustrates the power of the subsystems-based approach for comparative genomic reconstruction of metabolic and regulatory networks and this will be further extended for the assessment of biofuel production by *C. bescii*.

References

1. 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