

## 12. Integration of Multi-Omic Data for Advanced Consolidated Bioprocesses

Intawat Nookaew<sup>1</sup> (nookaewi@ornl.gov), Dan Olsen,<sup>2</sup> Mirko Basen,<sup>3</sup> Manesh Shah,<sup>1</sup> Visanu Wanchai,<sup>1</sup> Cong Trinh<sup>4</sup>, Philip D. Hyatt,<sup>1</sup> Steve D. Brown,<sup>1</sup> Miriam L.L and,<sup>1</sup> Michael R Leuze,<sup>1</sup> Robert L. Hettich,<sup>1</sup> Robert M. Kelly<sup>5</sup>, Mike Adams,<sup>3</sup> Lee Lynd,<sup>2</sup> David W. Ussery,<sup>1</sup> and Paul Gilna<sup>1</sup>

<sup>1</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>Dartmouth College, Hanover, New Hampshire; <sup>3</sup>University of Georgia, Athens; <sup>4</sup>University of Tennessee, <sup>5</sup>North Carolina State University

<http://bioenergy.org>

**Project Goals:** The BioEnergy Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. BESC researchers provide enabling technologies in characterization, 'omics, modeling and data management in order to (1) understand chemical and structural changes within biomass and (2) to provide insights into biomass formation and conversion.

Within the BESC project, several multi-level transcriptome, proteome and metabolome data have been generated across different laboratories. In general, the aim of these studies is to better understand the cellular processes of BESC organisms used in CBP, which include *Clostridium thermocellum* and *Caldicellulosiruptor bescii*. We have applied computational and systems biology approaches to quantitatively integrate multi-level 'omic data from various experiments. These approaches provide the important frameworks and strategies for further improvement of the capabilities of the organisms for advancing biofuel production. We present three examples of collaborative efforts within BESC between experimentalists and computational biologists to accelerate the development BESC's CBP organisms.

First, we present a metabolic model for the well-curated central metabolism of *C. thermocellum* DSM 1313. The model was supported by available transcriptomics and proteomic data, used rigorous fermentation data to validate the model, and provided comprehensive metabolic flux analysis under various environmental and genetic perturbations. With this central metabolic network in hand, we can identify targets of genetic modifications to design optimal networks for enhanced ethanol production, and also hypothesize potential bottlenecks. The central metabolism has been extended to a genome-scale metabolic network (GeM) that will be strengthened by incorporating physiological data and 'omics data for network curation as well as metabolic flux constraints derived from experimental data from the laboratories across from BESC. The developed genome- scale metabolic model takes into account growth media optimized for *C. thermocellum* and predicts complex cellular phenotypes, such as utilization of lignocellulosic biomass for biofuel production. These predictions can lead to new metabolic engineering targets.

Second, *C. thermocellum* has been engineered for increased ethanol production by different sequential deletion strategies. After direct genetic engineering on identified metabolic targets was performed, the mutants were re-sequenced and non-targeted genetic mutations were found to occur. Therefore, in-depth analysis of genome re-sequencing of different lineages was used to uncover additional mutations that may play effect on changing on cellular physiology. A computational analysis pipeline tailor-made for genome re-sequencing was used to identify desired and undesired genetic changes, such as single nucleotide variation, transposon changes, and the presence of transformation markers. The genetic changes within

individual strains will be further evaluated with fermentation data as well as transcriptome data to evaluate the impact of complex mutations on ethanol production and undesired fermentation products. We found a non-targeted point mutation in a developed strain impacted on cofactor specificity of AdhE that is beneficial for ethanol production.

Third, we present a predicted transcript-based annotation of the *C. bescii* genome, based on transcript RNA-seq, 5'RACE RNA-seq and extracellular proteomics for a variety of growth conditions in different carbon sources to determine all transcribed genes, identity of small RNAs/CRISPR, and map the intensity of transcription along the chromosome. Those experimental data will capture the missing features from standard computational gene annotation. Global transcriptional landscapes of *C. bescii* can be used to design strategies for further strain improvements. Mapping of transcript expression levels and gene regulatory patterns will be identified through the integration of transcriptome, extracellular proteome using network analysis across different growth conditions to inform metabolic modeling and metabolic engineering efforts.

*The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.*