High-Throughput Determination of a Subcellular Metabolic Network Map of Plants

Jason Thomas[1*](jthomas@carnegiescience.edu), Suryatapa Ghosh Jha[1], Kevin Radja[1,3],Tallyta Silva[2], William Dwyer[1], Justin Krupp[1], Charles Hawkins[1], Angela Xu[1], Bo Xue[1], Christine Aquino[1], Markita P. Landry[3], Jenny C. Mortimer[2], David W. Ehrhardt[1], **Seung Y. Rhee[1]**

[1]Carnegie Institution for Science, Stanford, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; University of California, Berkeley, CA, [3]; Virginia Tech, Blacksburg, VA

Project Goals: The goal of this project is to build an integrated pipeline to characterize metabolic interactions and pathways at a cellular level, using a combination of computational prediction, metabolic network modeling, and high-throughput experimental testing. This pipeline will be divided into three stages in order to develop a high-resolution subcellular map of small molecule metabolism in Sorghum and Brachypodium: a) generating localization predictions using bioinformatic algorithms, b) testing those predictions using nanotechnology mediated transformation of fluorescently tagged target proteins and high-sensitivity confocal imaging, and c) using the experimental data to generate new compartmentalized metabolic network models as well as refining existing pathway models. This project will initiate the creation of a repository for subcellular locations of metabolic enzymes, yielding important insight into the structure and function of metabolic networks in model systems as well as economically important crop species.

Advances in our understanding of plant metabolism have underpinned many traits that contribute towards improving plant productivity. To identify, by predictive modeling and experimentation, and engineer desirable metabolic traits, such as maximizing biomass production under suboptimal conditions or reallocation of biomass from carbohydrates to lipids, we must decode the complex metabolic networks. Subcellular compartmentalization of metabolic reactions through the locations of enzymes is critical to understanding, modeling, and engineering plant metabolism. Yet, localization of the majority of the predicted enzymes in *S. bicolor* and *B. distachyon* is not yet known. The paucity of experimentally validated information in most plants, especially in the DOE flagship bioenergy plants, severely limits scientists and engineers to assess the performance and translatability of computational tools and resources.

In this project we have developed an integrated pipeline that combines computational prediction, metabolic network modeling, and high-throughput experimental testing using state of the art technologies in live confocal imaging, nanomaterial-mediated plant transformation with target metabolic enzymes, and metabolic network modeling. The enzyme localization prediction pipeline, named Compartmentalization Of METabolic networks (COMET), collects annotation data from existing metabolic pathway databases. For sorghum, the database SorghumbicolorCyc was created by using the E2P2 software to predict enzymatic function of the proteins in the sorghum genome sequence, then using the Pathologic and SAVI software to call the presence in sorghum of metabolic pathways from the Metacyc reference database. COMET implements a novel network-based classifier, MetaboLoc, to infer compartmentalization of metabolic pathways based on information available from GFP experiments and high-quality predictions from DeepLoc, an existing sequence-based classifier. When compared against DeepLoc, our classifier MetaboLoc achieved near equal performance in several of the subcellular compartments surveyed and

even outperformed the sequence-based classifier in the vacuole and Golgi apparatus. By circumventing reliance on sequence information, MetaboLoc enables the COMET pipeline to predict localization for over 95% of the total reactome, regardless of the species. The final compartmentalized network can serve as input for downstream analyses like metabolic domain enrichment or comparison between species

To validate COMET predictions, 48 candidate metabolic genes in *Arabidopsis thaliana* were chosen for subcellular localization. To achieve this, Gateway cloning technology was used to fuse these candidates with mCherry fluorescent tags, and subsequently expressed in *Nicotiana benthamiana* leaf cells using transient, Agrobacterium-mediated plant transformation techniques. Current efforts are focused on utilizing high resolution confocal live imaging (EMCCD spinning disk and Leica HyD point scanning) to determine the locations of these fusion proteins *in planta*, and to confirm these subcellular localization results using known organellar markers. The dataset collected from these validations will be used for developing the network maps and refine current models and to apply the methods to *S. bicolor* and *B. distachyon.*.

Overall, this project aims to holistically decipher the complexity of plant metabolic networks in order to engineer pathways to tackle impending challenges regarding climate change, food security, and the availability of sustainable energy sources.

This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0020366