Coupling Metabolic Source Isotopic Pair Labeling And Genome Wide Association For Metabolite And Gene Annotation In Plants

Jeff Simpson,[1,2] Cole Wunderlich,[1,2] Rajdeep Khangura,[1,2] Brian Dilkes,[1,2] and **Clint Chapple,[1,2]* (chapple@purdue.edu)**

[1]Department of Biochemistry, Purdue University, West Lafayette, IN; [2]Purdue Center for Plant Biology, West Lafayette, IN

Project Goals: Improving our understanding of plant genomes and metabolomes is critical to unlocking higher plant productivity, developing new strategies to protect crops from stress, and identifying sources of new plant-based products. Progress towards these goals is currently limited by the fact that we do not know the identity of most plant metabolites, their biochemical origins, or the function of most of the genes involved in their synthesis and regulation. We are addressing these challenges using our recently developed stable isotope feeding/LC-MS/genome wide association (GWA) strategy to identify functional gene-metabolite relationships for metabolites that are derived from amino acids in Arabidopsis and sorghum. The aims of this work are to establish precursor-of-origin annotations for plant metabolites; use metabolite annotations to identify functional gene-metabolite relationships using GWA; and authenticate these associations by reverse genetics and biochemical assays of enzyme activity.

Enhanced sensitivity and throughput for analytical tools in metabolomics, principally liquid chromatography-mass spectrometry (LC-MS), permit the measurement of thousands of metabolites in complex plant extracts in a single analytical run. Parallel reductions in the cost of nucleic acid sequencing technologies and informatics has identified all of the genes encoded by many plant genomes and provided data about their expression. Despite these technical leaps, we cannot rationally design plant products because we do not know the functions of most genes in plant genomes nor the identity of most plant metabolites. We propose to provide fundamental understanding of metabolic gene functions and leverage “omics” datasets to annotate gene functions and identify metabolites.

Whole genome sequencing has identified millions of single nucleotide polymorphisms (SNPs) segregating in populations of many plant species, including Arabidopsis and sorghum. It is now possible to use these nucleotide polymorphisms to identify the genetic mechanism of natural variation in traits, including metabolite accumulation. The advantage of combing metabolite profiling and this Genome Wide Association (GWA) analyses lies in the ability of metabolite levels to report on enzyme function and genetic mapping to reveal gene-metabolite associations without prior knowledge that any such association exists.

We have developed a robust isotopic labeling/ LC-MS/ GWA approach to meet the challenges associated with identifying metabolites and annotating genes within plant metabolic pathways. We initially focused on identifying phenylalanine-derived metabolites in Arabidopsis, and genes

associated with their production. The experimental approach and bioinformatic pipeline is now being adapted to identify metabolites derived from any metabolic pathway for which isotopically labeled precursors are available. Briefly, metabolites were extracted from Arabidopsis stems fed [6-13C] ring-labeled Phe or unlabeled Phe and metabolite profiles were obtained by LC-MS. The chromatograms were aligned and queried for the presence of unlabeled and isotopically labeled "peak pairs" in 13C-Phe-fed samples to both detect and measure the relative abundances of unlabeled metabolites and their isotopologues. This method identified more than 500 phenylalanine-derived LC-MS features (the "phenylalanome"), many of which represent compounds that had not been described previously. Matches for these 500 phenylalanome mass features were identified in an unlabeled LC-MS dataset collected from an association mapping panel of Arabidopsis. GWA identified strong associations between these mass features and multiple genes encoding known enzymes in the phenylpropanoid pathway as well as enzymes with no validated functions. Many of these associations were confirmed by mutant analysis when the phe-derived metabolite was gained or lost in a mutant corresponding to the gene predicted by GWA. Thus, this approach can be used to simultaneously discover the pathway of origin for metabolites and to definitively link genes of unknown function to pathways, metabolic products, and reactions.

This work will synthesize our capacity to sequence plant genomes and classify metabolites based on their pathway of origin to provide functional annotation to genes associated with plant metabolism. Here we propose to classify metabolites produced from metabolic pathways that require an amino acid as their precursor, then utilize GWA and mutant resources in Arabidopsis and Sorghum to inform species-specific annotations for genes associated with plant metabolism.

This work is supported by the U.S. Department of Energy under Award DE-SC0020368.