Metabolic Source Isotopic Pair Labeling and Genome-Wide Association Are Complementary Tools for the Identification of Metabolite-Gene Associations in Plants

Jeffrey Simpson[1] (jsimpso1@purdue.edu), Cole Wunderlich[1], Brian Dilkes[1] and **Clint Chapple**[1]

[1]Purdue University, West Lafayette, IN

Project Goals: Improving our understanding of plant genomes and metabolomes is critical to understand the function of genes, unlock higher plant productivity, develop new strategies to protect crops from biotic and abiotic stress, and identify sources of new plant-based products. Progress towards these goals is 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 will address these challenges through our recently developed stable isotope feeding/LC-MS/genome wide association strategy. This will identify functional gene-metabolite relationships for metabolites that are derived from amino acids in Arabidopsis and sorghum and authenticate them using reverse genetics. When complete, these data will identify known and unknown metabolites within untargeted LC-MS analyses, and characterize the genes involved in their synthesis.

The optimal extraction of information from untargeted metabolomics analyses is a continuing challenge. Here we describe an approach that combines stable isotope labeling, LC-MS, and the development of a computational pipeline (named PODIUM- https://github.com/chapple-lab/podium) to automatically identify metabolites produced from a selected metabolic precursor. We identified the subset of the soluble metabolome generated from phenylalanine (Phe) in Arabidopsis thaliana, which we refer to as the Phe-derived metabolome (FDM) In addition to identifying Phe-derived metabolites present in a single wild-type reference accession, the FDM was established in nine enzymatic and regulatory mutants in the phenylpropanoid pathway. To identify genes associated with variation in Phe-derived metabolites in Arabidopsis, MS features collected by untargeted metabolite profiling of an Arabidopsis diversity panel were retrospectively annotated to the FDM and natural genetic variants responsible for differences in accumulation of FDM features were identified by genome-wide association. Large differences in Phe-derived metabolite accumulation and presence/absence variation of abundant metabolites were observed in the nine mutants as well as between accessions from the diversity panel. Many Phe-derived metabolites that accumulated in mutants also accumulated in non-Col-0 accessions and was associated to genes with known or suspected functions in the phenylpropanoid pathway as well as genes with no known functions. Overall, we show that cataloguing a biochemical pathway's products through isotopic labeling across genetic variants can substantially contribute to the identification of metabolites and genes associated with their biosynthesis.

Funding statement. This work was supported by the U.S. Department of Energy, Office of Science (BER), Grant DE-SC0020368.