Linking Genes to Specialized Metabolic Pathways in Plants

Plants produce a tremendous diversity of specialized metabolites (SMs) to interact with and manage their environment. A major challenge hindering efforts to tap this seemingly boundless source of pharmacopeia is the identification of SM pathways and their constituent genes. Given the well-established observation that the genes comprising a SM pathway are co-regulated in response to specific environmental conditions, we hypothesized that genes from a given SM pathway would form tight associations (modules) with each other in gene co-expression networks, facilitating their identification. To evaluate this hypothesis, we used 10 global co-expression datasets—each a meta-analysis of hundreds to thousands of expression experiments—across eight plant model organisms to identify hundreds of modules of co-expressed genes for each species. In support of our hypothesis, 15.3–52.6% of modules contained two or more known SM biosynthetic genes (e.g., cytochrome P450s, terpene synthases, and chalcone synthases), and module genes were enriched in SM functions (e.g., glucoside and flavonoid biosynthesis). Moreover, modules recovered many experimentally validated SM pathways in these plants, including all six known to form biosynthetic gene clusters (BGCs). In contrast, genes predicted based on physical proximity on a chromosome to form plant BGCs were no more co-expressed than the null distribution for neighboring genes. These results not only suggest that most predicted plant BGCs do not represent genuine SM pathways but also argue that BGCs are unlikely to be a hallmark of plant specialized metabolism. We submit that global gene co-expression is a rich, but largely untapped, data source for discovering the genetic basis and architecture of plant natural products, which can be applied even without knowledge of the genome sequence.