

Plant Metabolomics

For BOT 6516

Introduction

Modern metabolomics began about ten years ago and yet many continue to question the relative performance of this area of technology in advancing plant biology. However, it appears that metabolomics has become an important tool in plant biology research over the last several years in a variety of different areas of plant biology.

Metabolomics is part of modern systems biology. As such, systems biology aims to understand and model whole plant systems, their underlying molecular, biochemical, metabolic and physiological processes and how all of this is coordinated by highly integrated regulatory networks. The major challenge in the post-genomic era is to understand how interactions among the molecules in a cell, tissue, organ, organism and ecosystem contribute to and determine form and function.

Recent advances in high-throughput techniques and technology has fundamentally changed how metabolic processes are studied. Previously, most analytical methods were **targeted** to a narrow group of compounds (metabolites) usually on the basis of separation technology for a specific chemical class of compounds. However, the advent of **non-targeted** analytical methods breaches this limitation, and now many different metabolites of different metabolic origins and chemical properties can be measured simultaneously from a single sample extract.

Consequently, the amount of data generated in metabolomics studies is huge and the integration of large multi-variant type data can be very complicated. This change in how analytical approaches are conducted has eliminated one limitation and created a new limitation regarding global metabolic and metabolite studies. The major bottleneck is no longer the generation of necessary data sets, but analysis of the data sets to uncover meaningful biological interpretation to achieve the desired insight into plant systems.

Given the continued development of advanced analytical approaches and new bioinformatic tools, the number of useful applications for metabolomics continues to grow. Some examples of the value of metabolite and metabolomic analyses are only now becoming recognized and a reality. In one application, it is well known that the quality of crop plants and their derived products is frequently a direct function of their metabolite content. In particular, non-targeted metabolomics approaches offer the potential to extend knowledge and understanding of complex quality traits, even in the absence of the lack of information on the related biochemistry. Moreover, metabolomics is being used to uncover the biochemical bases of plant quality traits, and how genetic and environmental factors biochemically influence these traits. **Perhaps you can find some key examples in the literature include traits such as taste, disease resistance, and nutritional value.**

Given that plants are sessile, they have evolved an arsenal of chemical defenses against herbivores, pathogens and abiotic stresses as well as agents of seed dispersal and signaling. Secondary metabolites like glucosinolates are constituents of the plant's defense weaponry. Chemical defenses can include either 'direct defenses', when they directly impede herbivores or impede disease progression. They can also serve as 'indirect defenses', when plants attract natural enemies of herbivores by volatile organic compound release or providing herbivore enemy food sources. Most defense systems can be induced upon herbivore attack or pathogen infection. There is evidence that the effectiveness of plant defenses is co-determined by the levels of primary metabolites within the plant. This is because the building blocks, such as amino acids of secondary metabolites, originate from primary metabolism. Therefore plant metabolomics can be a valuable technique for ecologists studying plant-insect interactions, because both primary and secondary metabolites can be analyzed simultaneously. Because of the versatility of current analytical technology, active secondary compounds that usually occur in quantities that are several orders of magnitude smaller than those of primary compounds can be still be characterized because of the high dynamic range of sensitivity of the analytical platforms...

There is a widely-held belief that primary metabolic pathways of plants are known and completely understood, and that the 'textbook' pathways can be relied on as a framework for metabolomics research. This assumption is not well justified. Researchers continue to find surprises and novelty: new pathways, new enzymes, new functions for 'old' enzymes, mis-assigned enzymes, and unexpected metabolic plasticity. For example, a new pathway for the conversion of transitory leaf starch to cytosolic sugars of carbon metabolism has been discovered, that was not suspected a mere six years ago. The central role of maltose as an early primary metabolite has been revealed, along with a cytosolic heteroglycan complex about which very little is known, is implicated in this pathway. Research on fatty acid metabolism has revealed a new role for peroxisomal citrate synthase. In knock out lines for glyoxysomal malate synthase, an alternative pathway comes into play and the metabolites of this pathway can be detected by metabolomic approaches and point to a novel pathway. Recent findings concerning peroxisomal malate dehydrogenase indicate that it is not required by the glyoxylate cycle as widely believed. Also peroxisomal malate dehydrogenase knock out lines seem to grow well in ambient air.

Methodology

Although there is no general analytical platform that can be used in metabolomics to define, identify and quantify all metabolites in a biological sample, mass spectrometry (MS) has become the standard instrumentation for metabolomics analysis. Some reasons for this include its low detection limits, general robustness for reliability, stability and relative user friendliness. GC/MS is widely used for profiling metabolites from plant extracts that are volatile or that can be made volatile by derivatization. However, the importance of also including LC/MS analysis to cover more of the metabolome is obvious and warranted in order to approach the definition of "true" metabolomics analysis.

Both GC/MS and LC/MS generate large amounts of data, which sometimes is informative when the true identity of the peaks is known. Presently, the majority of mass fragments obtained by MS from plant extracts remain unknown. Thus a major limitation and a major effort is the identification of all of the mass fragments of plant metabolites. However, in large multi-variant datasets, important information of even unknown metabolites can help to contribute to the understanding of biological function. The resultant metabolite fingerprint database can be used to classify plant lines and perhaps diagnostic metabolic phenotypes by PCA, HCA and PLS-DA cluster analysis. Automated comparison of PCA loadings and contribution plots can indicate metabolites and pathways that can be targeted for more selective analytical regimes.

Omics

Plant metabolomics is one part of an “omics” universe and revolution in modern biology and can complement genomics (DNA sequences of genomes), transcriptomics (RNA expression levels of transcribed genes), proteomics (the protein complement of an organelle, a cell or organism), phosphoproteomics (phosphoproteins of a cell), lipidomics (lipid components of a membrane organelle, cell or organism and interactomics (the interactions of cellular biochemical elements) to list some of the major types.

Modern genetics-based plant biochemical pathway analysis has benefited from fast and accurate methods for metabolite measurement. Previously such ‘brute force’ mutant screening traditionally relied on specific compound-class analytical methods such as thin layer chromatography, gas chromatography and liquid chromatography. The advent of mass spectrometry-based methods has allowed genetic dissection of a wider variety of pathways. The recent development of ‘omic’ scale technologies now permits us to consider the feasibility of examining specific mutation-induced changes in hundreds or perhaps even thousands of molecules at a time. Model organism plant functional genomics resources such as large sequence-indexed insertion mutant collections are helping to blur the distinction between ‘forward’ and ‘reverse’ genetic screens and helping to accelerate progress in understanding the metabolome.

The model plant *Arabidopsis thaliana* is small and has a rapid growth cycle. Researchers have created collections of ecotypes, knockout mutants, and transgenic lines and a fully sequenced genome that makes this it ideal system for large-scale metabolomic analysis with respect to gene function analysis. Recent studies have demonstrated that there is a transcriptional regulatory component of core metabolic and biological processes. This transcription regulatory matrix acts to control of a wider set of transporters, enzymes, cofactor and substrate producing proteins and regulatory molecules that may represent an integrated common task.

Integration

Systems biology is the study of biology as an integrated system of genetic, protein, metabolite, cellular and pathway events that are in constant flux and highly interdependent. Due to the availability of advanced instrumentation it is now possible to generate very complex data sets as

systems biology approaches become a possibility. Researchers have recently focused on the integration of complementary data sets generated by transcriptomics, metabolite profiling, proteomics, enzyme assays and phenotyping approaches. Thus integrated analysis at the operational level of “omics” can provide clues for identification of gene function and the precise information about gene-to-metabolite and/or metabolite-to-metabolite networks never before possible.

In proof of concept studies, nutritional stress of nitrogen and sulfur results in global changes of the plant metabolome that could be correlated with the modulation of gene expression, indicating the presence of gene-to-metabolite networks, in particular, in glucosinolate biosynthesis. From other studies, comprehensive gene expression and metabolite profiles of anthocyanin overproducing *Arabidopsis* lines have revealed the function of novel genes responsible for modification and storage of anthocyanins. Often the function of candidate genes has been identified by analysis of the T-DNA insertion lines and the enzymatic activities of recombinant proteins. By expanding this strategy to include transcriptome analyses and public databases helps to unravel previously unknown networks. Using co-expression analyses, a working model of co-expressed genes in flavonoid pathway has been constructed. With the potential to link metabolomics data to complementary data from other analytical sources, plant research has never been in a better position to associate specific metabolites or metabolic profiles with other important characteristics such as bioactivity, phenotype, health-promoting capacity, anti-oxidant levels etc.

Perhaps the two most meaningful and experimentally accessible variables of metabolic networks are protein and metabolite concentrations. However, it is important to remember that direction and rate of fluxes within pathways and the network further depend on reaction thermodynamics and kinetic properties of the participating enzyme(s). Regulation frequently includes allosteric regulation and protein modification. The vitally important metabolically critical intracellular fluxes are the functional output of integrated biochemical and genetic interactions within complex metabolic networks that are pivotal for understanding of network operation. In the context of directly measurable concentrations of metabolites and proteins, fluxes are mostly nonmeasurable and must be studied by more targeted approaches. For this reason, quantification of intracellular fluxes has long lagged behind our capability to track global metabolite, mRNA or protein concentration changes.

Dataset Analyses

Systems biology generated data can have a multiway-, multiset- or multilevel structure or combinations of any and all of these and more. These high complex data structures require special models to summarize and visualize the data in readily accessible and useful information. Moreover, there is an increasing awareness that time and spatially resolved metabolomics is crucial for systems biology. Yet most metabolomic based studies have not adequately addressed resolution of subcellular, tissue or organ spatial distributions of metabolites. Therefore this calls for new data analysis and bioinformatics tools.

One example is MetNet, an emerging open-source software platform for exploration of disparate experimental data types and regulatory and metabolic networks in the context of Arabidopsis systems biology. The MetNet platform features graph visualizations, it can graph theoretic computations for determining biological distances, make interactive displays, and even unique multivariate displays. It includes a statistical analysis tool, graph modeling using an open source statistical analysis language, R, and versatile text mining.

Additional bioinformatics tools have been developed that integrate classical statistical analyses into dedicated graphical user interfaces. Some examples of bioinformatic and analysis tools include: 1, development of a knowledge-driven plant ontology (MapMan - BIN system) 2, visual integration of expression and metabolite data sets using the pathway visualization tool (MapMan), 3, development of a text-mining tool for updating and transfer plant ontologies (PlantMiner), 4, identification of functional hotspots and functional modules in expression data using functional statistics and 4, identification of trait-related lead genes by pattern matching strategies in industrial-scale metabolite profiling data sets.

Commentary summary based on presentations from:

4th International Conference on Plant Metabolomics

7th April - 10th April 2006

Reading, Berkshire, UK