Analyses of Soybean Seed Transcriptomics and Metabolomics Data Using R: A Systems Biology Approach to Understanding Seed Composition

Ling Li\textsuperscript{1}, Wenxu Zhou\textsuperscript{2}, Wengang Zhou\textsuperscript{1}, Dan Nettleton\textsuperscript{3}, Mark Westgate\textsuperscript{4}, Basil Nikolau\textsuperscript{2}, and Eve Syrkin Wurtele\textsuperscript{1}

\textsuperscript{1}Department of Genetics, Development, and Cell Biology, \textsuperscript{2}Department of Biochemistry, Biophysics, and Molecular Biology, \textsuperscript{3}Department of Statistics, and \textsuperscript{4}Department of Agronomy, Iowa State University, Ames, IA 50011 [http://www.metnetdb.org]

ABSTRACT

In addition to being the propagule that ensures the dissemination of plants, seeds also provide one of the major products of agriculture. The biochemical storage reserves that are deposited within the seed during its development fall into three general categories: proteins, oils, and carbohydrates. These seed reserves are biosynthesized by the programmed expression of a metabolic network during seed development. In most commercial lines of soybean grown in the Midwestern states of the US, seeds are composed of 40% protein, 20% oil, 15% soluble carbohydrate, and 15% fiber (http://www.ars-grin.gov/Soyinfo/introduction_e.htm). There is considerable knowledge concerning the basic biochemical processes by which imported carbon and nitrogen is converted to the final products, protein, oil, and carbohydrate. However, there is a great deal to be learned concerning the molecular, biochemical and genetic mechanisms that regulate this complex metabolic network. Recent developments in genomics have started to provide the catalogue of genes that would be required for this process. We have taken advantage of microarray and metabolomics technology to identify the global gene expression profile that regulates the developmental and biochemical network, which determines final seed structure and composition. We have coupled this with bioinformatics analyses to gain insights as to the regulation of the biochemical program that determines soybean seed development.

Using R in the exploRase software, we have analyzed high dimensional transcriptomics and metabolomics data derived from seeds of soybean lines with high and low protein composition. These analyses have enabled the identification of genes and metabolites that may be important for soybean seed composition.

RESULTS

Affymetrix Signal values were natural log transformed and median centered within each GeneChip, and normalized log signals were analyzed separately for each probe set using a linear model analysis. An overall F-test was conducted to scan for any change in mean expression across the time points examined. The p-values from these F-tests were (FDR) at specified levels. In addition, the p-values for all pairwise comparisons between time-specific expression level means were computed in each gene-specific linear model analysis. The set of p-values for each comparison of one time to another were also converted to q-values as described above.

K-medoids cluster analysis was used to organize and visualize the expression patterns of the 2869 genes with q-values less than 0.01 in the overall F-test for change of expression over time.

Cluster analysis of 400 metabolites that alter accumulation during Evans seed development. These metabolites are grouped into eight clusters based on their accumulation patterns.

Expression of genes involved in starch metabolism, is consistent with starch accumulation during Evans seed development. Gene expression is viewed via MetaOMGraph from MetNet (http://www.metnetdb.org/MetNet_MetaOMGraph.htm).

Analyses using R

Legends

Metabolomics: Study of small-molecule metabolite profiles. Carbohydrate metabolism (starch, sugar); amino acid (AA); fatty acid (FA) metabolism. Anabolism (synthesis/elongation); catabolism (degradation/oxidation).

MetNet systems biology platform for plant 'omics was also used for data analysis.

Funding: Plant Science Institute, Iowa State University; Consortium for Plant Biotechnology Research, USDA.