Acclimation and gene expression in drought-stressed pine

Author to whom all correspondence should be sent: , Department of Plant Pathology, Physiology and Weed Science, 103 PMB, 435 Old Glade Rd, Virginia Tech, Blacksburg, VA 24061; Phone 540 231 7701; Fax 540 231 5755; email grene@vt.edu.

Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine.

Jonathan I. Watkinson, Allan A. Sioson, Cecilia Vasquez-Robinet, Maulik Shukla, Margaret Ellis, Lenwood S. Heath, Naren Ramakrishnan, Boris Chevone, Layne T. Watson, Leonel van Zyl, Ulrika Egertsdotter, Ronald R. Sederoff, and Ruth Grene

Department of Plant Pathology, Physiology and Weed Science, Virginia Tech, Blacksburg, VA 24061(J. I. W., C. V.-R., B. C., R. G.); Computer Science Department, Virginia Tech, Blacksburg, VA 24061(A. A. S., M. S., M. E., L. S. H., N. R., L. T. W.); Forest Biotechnology Group, Department of Forestry, North Carolina State University, Raleigh, NC 27695-7247 (L. v. Z., R. R. S.); and Institute of Paper Science and Technology, 500 10th Street, N.W., Atlanta, GA 30318 (U. E.)

This work is supported in part by National Science Foundation Grants EIA-0103660 (Next Generation Software) and IBN-0219332 (Information Technology Research) to Heath, Ramakrishnan, Watson, and Grene
Corresponding author ; email grene@vt.edu; fax 540 231 5755

Abstract

Since the product of a single gene can influence many aspects of plant growth and development, it is necessary to understand how gene products act in concert and upon each other to effect adaptive changes to stressful conditions. We conducted experiments to improve our understanding of loblolly pine response to drought stress, and to further develop Expresso, a Next Generation Software computational system. Water was withheld from rooted plantlets of *Pinus taeda* to a water potential of -1Mpa for mild stress and -1.7Mpa for severe stress. Net photosynthesis was measured for each level of stress. RNA was isolated from needles and used in hybridizations against a microarray consisting of 2180 cDNA clones selected from 5, pine EST libraries to include the 3' end of representative genes associated with stress responsiveness and stress protection. The expression data was first analyzed based on the linear mixed model of Wolfinger et al. (2001) and then redescribed through Inductive Logic Programming (ILP). The change in RNA transcript profiles of *P. taeda* due to drought stress was correlated with physiological data reflecting photosynthetic acclimation to mild stress or photosynthetic failure during severe stress. Genes encoding specific chaperones, enzymes from the aromatic acid and flavonoid biosynthetic pathways, and from carbon metabolism showed distinctive responses associated with stress treatment. Five genes shown to have different transcript levels in response to either mild or severe stress were chosen for further analysis using Real Time PCR (RT-PCR). The RT-PCR results were in good agreement with those obtained on microarrays.

Introduction

Drought stress can limit plant growth resulting in reduced crop yields. Many attempts to analyze plant responses to drought have focused on single genes. However, the product of a single gene can function in or act upon multiple response pathways, influencing many aspects of plant growth and development. It is necessary, therefore, to understand how gene products act in concert and upon each other to effect adaptive changes to stressful conditions. We have used microarrays and Expresso, a Next Generation Software, computational system to analyze the changes in gene transcript profiles in loblolly pine in response to drought stress.

Microarrays have emerged as a prominent tool in the analysis of large scale gene expression. The spotting of cDNA onto glass slides allows for ascertaining the expression profile, as revealed by steady state transcript levels, of thousands of genes at a single time. The technique promises to reveal networks of genes that contribute to the same biological response and provide new information on the functions of unknown genes (Somerville and Somerville, 1999). Microarrays have been used in plants to identify changes in transcript levels of genes associated with drought stress (Seki et al., 2001; Ozturk et al., 2002; Heath et al., 2002), cold stress (Seki et al., 2001; Fowler and Thomashow, 2002), salt stress (Kawasaki et al., 2001; Ozturk et al., 2002), pathogen interaction (Schiedler et al., 2001), and during developmental programs (Ruan et al., 1998; Girke et al., 2000).

Microarrays present a challenge to researchers both in terms of their design and implementation (Kerr and Churchill, 2001) as well as in subsequent storage and analysis of data. Such challenges require automated computational assistance. With the

voluminous amounts of data generated from each microarray experiment, effective and efficient access to this information becomes critical. As new experiments bring additional insights into the genomic response of plants to specific conditions, it is advantageous to reanalyze previous experiments in light of these new insights. For example, Perez-Amador et al. (2001) used cluster analysis on 47 Arabidopsis gene expression profiles stored in the Stanford Microarray Database, an online database, and discovered novel expression patterns for several genes identified as having increased transcript levels in a mutant. Since the results of one microarray experiment can be used to plan subsequent experiments, access to data during design is essential.

We are designing a computational system - Expresso (Alscher et al., 2001; Heath et al., 2002; Sioson et al., 2003) - to address multiple phases of the microarray experiment lifecycle, including experimental design, microarray design and printing, data acquisition, image analysis, statistical analysis, and data mining. In contrast to many existing tools, Expresso is meant to systematize methodological issues in microarray experiments. Alscher et al. 2001 identifies the three key design principles behind Expresso: (i) model-based design and management of experiments (i.e., models of every step of the microarray experiment are used to avoid costly design errors associated with probe selection, probe adhesion to the solid support, and hybridization dynamics); (ii) algorithms for 'closing the experimental loop' (e.g., adjustments to the design can be made based on analyzed results, detected problems, and modified aims of the experimental biologist); and (iii) a lightweight data management system that allows graceful changes to the underlying schema of the data. For instance, new experiments can have attributes and aspects that were absent in previous experiments. The Expresso

system thus does not dictate a rigid experimental framework. The data management system is meant to support a variety of analysis algorithms; currently a number of attribute-value based clustering algorithms (e.g., Koyuturk et al., 2002) and the relational technique of inductive logic programming (Muggleton, 1999; Muggleton and Feng, 1990) are supported.

To date, the primary use of Expresso has been to study drought stress responses in loblolly pine, the predominant timber species in the southeastern US, covering about 13.4 million hectares (Schultz, 1999). Loblolly pine is also an important timber crop in Africa, Asia, and South America, where tree growth can exceed that of trees grown in the US. The faster growth of trees in these locales depends upon the use of suitable seed stock and, even then, improvements can be made to optimize growth in each locale (Schultz, 1999). Tree improvement involves traits such as better wood quality and higher density, greater disease resistance, and improved growth under various environmental conditions (Schultz, 1999). Drought stress can limit tree growth and alter wood quality (Lev-Yadun and Sederoff, 2000). It is necessary to understand responses to drought stress to achieve the development of crops with increased resistance to drought. Loblolly pine constitutes an excellent system to model drought stress in softwood timber species and gymnosperms in general. Furthermore, the use of a gymnosperm will add considerable knowledge to comparative genomic studies, extending what is known about gene function from angiosperms to gymnosperms.

Drought stress has been correlated with expression changes in many plant genes. These include the heat shock proteins (HSPs; Ristic et al., 1998), late embryogenic abundant proteins (LEAs; Iuchi et al., 1996), and aquaporins. Some of the more well-

characterized responses include genes whose products regulate expression of drought responsive genes, the dehydration responsive element binding proteins (DREB; Liu et al., 1998) and ABA responsive element binding proteins (ABREB; Chu et al., 2000; Kang et al., 2001). Other genes involved in drought stress are those associated with lipid signaling such as the phospholipase gene families (Katagiri et al., 2001; Sang et al., 2002) and those associated with detoxification of reactive oxygen species (Alscher et al., 1997).

Results obtained using microarrays have broadened the analysis of gene expression in response to drought and confirmed the results of individual gene studies. For example, Seki et al. (2001) reported 44 drought-inducible genes in Arabidopsis using microarrays and identified putative targets of DREB1A. Ozturk et al. (2002) noted a change in expression using microarrays for several barley genes identified as being drought responsive, including a LEA, a dehydrin, and a water channel protein. A water channel protein was also one of those identified by Seki et al (2001) in Arabidopsis, suggesting at least some commonality between results from different species. We have previously used Expresso to analyze changes in transcript profiles of loblolly pine in response to drought stress. We designed and printed an array consisting of 384 cDNAs chosen by us as being associated with stress responses. Using the 384 gene microarray, 72 genes showed increased transcript levels after four cycles of mild drought stress (Heath et al., 2002). This group of genes did not show increased transcript levels as a result of severe drought stress, making them candidate genes for drought tolerance mechanisms. The data mining technique of inductive logic programming (ILP) incorporated in Expresso associated gene expression with membership in putative functional categories. From the ILP results, we identified functional categories of genes that responded to the stress,

including 'heat' (HSPs and LEAs) as well as 'membrane transport proteins' (aquaporins and dehydrins). Other categories affected by the stress were 'cell wall related' and 'lignin biosynthesis'. Also identified as being stress regulated were a group of genes categorized as isoflavone reductases. Babiychuk et al. (1995) identified these genes as being involved in protection of cells from oxidative stress, which has been correlated with drought stress imposition. These enzymes are also related to those involved in lignin biosynthesis and in the synthesis of biotic, plant-defense compounds (Gang et al., 1999), suggesting that they are involved in multiple stress responsive pathways.

We have completed a second series of experiments to improve our understanding of loblolly pine response to drought stress, and to further exercise Expresso. The arrays consist of 2178 clones selected from 5 existing loblolly pine cDNA libraries (Stasolla et al., 2003). This paper focuses on the use of data mining algorithms in Expresso to analyze this newer set of data and to identify significant changes in gene expression.

Materials and Methods

Plant Material

Rooted plantlets of *Pinus taeda* from the Atlantic Coastal Plain were produced clonally by Dr. Barry Goldfarb at North Carolina State University (NCSU). These were transported to Virginia Tech and grown in a mixture of peat moss, perlite, and sand (1:1:1) under natural daylight in a greenhouse. Supplemental lighting (mercury vapor and high-pressure sodium lamps) was used for 6 h in the evening to maintain 16 h daylength from July to the termination of the experiment in September. Temperature was maintained at 26°C± 4.5 during the day and 18°C± 3.0 at night. Plants were watered as

needed and fertilized once a week with half-strength Hoagland's solution. For drought stress experiments, water was withheld until a desired water potential had been reached: -1Mpa for mild stress and -1.7Mpa for severe stress. Trees were dried down to the desired water potential and then re-watered. Needles and stems were collected 24 hours after re-watering. The stress was then repeated by again withholding water. Imposition of each level of drought stress was repeated 3 times with harvesting of needles and stems 24 hours after re-watering (Fig. 1). Water potential of needles was measured in the morning (pre-dawn) using a Plant Water Status Console (Model 3005, Soilmoisture Equipment Corp. Santa Barbara, CA). Net photosynthesis was measured at the time of peak drought stress for each level of stress at light saturation on a LiCor 6400 (Lincoln, Nebraska).

Microarrays

A set of 2178 clones was selected from 5 pine cDNA libraries to include the 3' end of representative genes from all 15 functional categories assigned by MIPS to Arabidopsis (Stasolla et al., 2003). Clone amplification, cleaning, and spotting was carried out at NCSU (Stasolla et al., 2003). Each clone was replicated 4 times on a slide. The clones selected to be on the array were placed within a hierarchy of functional categories designed by us to reflect processes that are affected by water deficit and the mechanisms employed to protect these processes.

RNA Extraction and Hybridization

RNA was extracted from needles according to the method of Chang et al. (1993) as modified in our laboratory. Briefly, after addition of the extraction buffer to the tissue sample, the tissue was homogenized for one minute with a Polytron tissue homogenizer at full speed. After centifugation (9000xg for 10 min), the supernatant was transferred to a new tube and 1/5th volume of 5% CTAB was added (5% CTAB, 0.7M NaCl). This was heated to 65°C for 20 min. An equal volume of chloroform:isoamyl alcohol (IAA; 24:1) was added and the sample was centrifuged. The supernatant was re-extracted with an equal volume of chloroform:IAA (24:1). The sample was centrifuged and the supernatant transferred to a new tube. The RNA was precipitated over night with 1/2 volume 10 M LiCl. Each pair of RNAs to be compared (treated vs. control, for each time point) were reverse transcribed and labeled with Cy3 and Cy5 dyes (Stasolla et al., 2003). Reciprocal labeling of each comparison was implemented to control for variation due to the dye and resulted in 8 replicates per clone per comparison. Hybridizations were carried out according to Stasolla et al. (2003). A total of 24 slides was used for the microarray analysis. A modified loop design (Kerr and Churchill, 2001) was used to formulate comparisons between treated and control samples. For each degree of stress, 12 slides were used which compared treated samples with two control samples using reciprocal labeling. For example, mild treated 1 was hybridized to mild control 1 and mild control 4; mild treated 2 was compared to mild control 2 and mild control 1. This resulted in a total of 16 replicate spots per treatment.

Real Time PCR

Real Time PCR was used to confirm the data obtained from the microarray experiments. The relative abundance of five ESTs, chalcone isomerase (07 H08), naringenin-2-oxo dioxygenase (Flavonone-3-hydroxylase, NXSI_063_D01), Pine LEA group 2 (NXCI_002_G10), Pine LEA group 3 (PC14G04), and HSP18 (ST40F04), shown to be differentially expressed in the second and/or third cycle of either mild or severe stress in the microarray experiments was tested. The same total RNA samples as were used for the microarrays were analysed. The concentration of total RNA was measured using the RiboGreen RNA Quantitation Reagent And Kit (Molecular Probes). First strand cDNA was reverse transcribed from 300 ng of total RNA using TaqMan Reverse Transcription Reagents (Applied Biosystems) according to manufacturers' instructions. Gene specific primers were designed by the Primer Express 1.0 program (Applied Biosystems). The relative transcript abundance was monitored on an Applied Biosystems 7700 Sequencer using SYBR Green PCR Master Mix (Applied Biosystems). The adenosine kinase amplicon was used as an internal control for normalisation.

Analysis

Hybridized slides were scanned using a ScanArray 5000 (Perkin Elmer, Boston, MA) and image analysis was performed using QuantArray software. The raw and background subtracted data were first analyzed based on the linear mixed model of Wolfinger et al. (2001). The Wolfinger method uses a two-phase analysis to remove global effects and to estimate the interaction between gene and treatment. In the first stage, the values of the major factors are estimated by fitting the intensity data to an ANOVA normalization model

$$y = \mu + T + A + D + P + AP + \varepsilon .$$

The μ accounts for the overall average intensity value across the various factors in the experiment. The effects *T*, *A*, *D*, *P*, and *AP* represent the variation in treatments (2 levels, treated and control), arrays (6 arrays), dyes (2 dyes, Cy3 and Cy5), printing pins (4 pins), and array-pin interactions. The *y* values are the logarithms of the intensity signals. In the second phase of the analysis, the residual values that result from the first stage are used to estimate the interaction between the gene and treatment at $\alpha = 0.05$, Bonferroni corrected significance level. The gene model used to achieve this is

$$\varepsilon = G + GT + GA + GD + GS(A) + \gamma .$$

Here, *G*, *GT*, *GA*, *GD*, and *GS*(*A*) are the expected mean of the residuals, gene-treatment interaction, gene-array interaction, gene-dye interaction, gene-pin interaction, gene-array-pin interaction, gene-spot interaction, where *S* is nested in *A*, respectively. For each of the 2178 genes on the arrays, a hypothesis test between control and treated is constructed. All of the effects in both stages are assumed to be multiplicative (so their logarithms are additive), and the residuals ε and γ are assumed to be normally distributed. Different model and distribution assumptions, or a different confidence level would result in different expression assessments. As suggested by Wolfinger et al. (2001) to assure an experiment-wide false positive rate of 0.05, the p-value cut-off is set at the Bonferroni value of 0.05/2180 = 3e-05. To reflect this on the database, the SAS computed p-values of expression fold change for each gene are multiplied by 2178 so that the analysis p-value cut-off could still be set at 0.05. Genes with positive (negative) estimated fold

change with corrected p-value less than or equal to 0.05 are assessed as positively (negatively) expressed. The remaining genes are assessed as unchanged.

Genes whose raw intensity signals reached the maximum 16-bit intensity value of 65535 are detected through visual inspection for systematic errors in scatter plots of estimated fold changes versus negative of the corresponding p-values. Systematic errors due to these "maxed out" genes appeared as spikes in the scatter plots. Since the intensity signals for these genes were not accurately recorded, they are removed from the dataset considered in the next analysis step.

Following the analysis of variance, the data was redescribed via inductive logic programming (ILP). Inductive logic programming is a data mining technique for finding relationships (Dzeroski and Lavrac, 2001) and redescription specifies the nature of these relationships. Signature patterns of gene expression changes across experiments are also derived from ILP.

Data Mining by Redescription

Data mining functionality in Expresso is based on a collection of 15 relational database tables, implemented using the Postgres database management system. These tables summarize information such as experimental conditions (the different stress conditions), cDNA details (accession number, annotation, and putative functional categories), water potential and photosynthesis measurements, and expression levels. The Expresso query interface allows the biologist to perform Structured Query Language (SQL) queries on these tables, such as: "which genes belong to the 'trafficking' category?" SQL queries satisfy the *closure* property, meaning the results of one query can

be used as the input to the next query. For instance, after determining which genes belong to the trafficking category, a second SQL query can be performed to determine their expression level in 'mild cycle 2.'

Such sequences of queries form the basis of the 'data mining by redescription' technique. The above sequence of queries constitutes a 'redescription operation' because the biologist begins by defining a set of genes using one aspect (functional category), and then restates the same set in terms of a different aspect (expression level).

Since the number of possible redescriptions that can be performed is huge (owing to the numerous pairs of aspects that can be chosen and the multiple cardinalities of the underlying sets), Expresso incorporates a data mining technique (called inductive logic programming or ILP) to search through the space of possible redescriptions and to automatically identify the most interesting redescriptions. Using ILP, the biologist merely instructs Expresso as: "redescribe from categories to expression level" and ILP searches for statements that make relationships from *some* category to *some* expression level in *some* experimental condition. Pruning strategies are used to focus on the most interesting redescriptions (this statement will be qualified shortly). For instance, since the functional categories are organized in a hierarchy, we can exploit the subsumption properties of membership in this hierarchy to narrow down our possibilities.

After a successful redescription, a reverse-redescription can be attempted. For instance, suppose we mine that "trafficking genes are up-expressed in mild cycle 2." We can then try to see if the relationship holds in reverse, i.e., "are the genes that are up-expressed in mild cycle 2 present in the trafficking category?" Or, we can continue the redecription into a third aspect (e.g., expression level in a different condition). Each of

these requests is just a different “schema” for the redescription operation. To summarize, the biologist has the ability to instruct Expresso in ways such as “redescribe from category to expression level”, “redescribe from expression level to expression level”, or “redescribe from expression level to expression level to category.” Redescription results are summarized in the form of “rules”. For instance, one rule that redescribes from expression level to expression level is:

~expression(X,severe,’+’) :- expression(X,mild,’+’).

This rule, represented here in predicate logic form, was reported in our earlier paper (Heath et al. 2002). Read from the right to left, it states: “if gene X was up-expressed in the mild stress condition, then it was not up-expressed in the severe-stress condition.” The symbol “~” denotes logical negation and the symbol “:-” denotes logical implication. We found this rule to hold in 69 out of 72 cases, yielding a very high degree of confidence. In other words, of the 72 genes that were up-expressed in the mild stress condition, 69 of them were indeed either down-expressed or not expressed in the severe stress condition.

Expresso allows the biologist to specify the criteria for mining these rules. In the experiments reported here, we require that every implication have a strength of at least 80%; the results are summarized in Figs. 5 and 6 as cartoons. It is important that the criteria of 80% *not* be interpreted as a measure of statistical significance. All expression levels are assessed in Expresso using a 5% Bonferroni corrected significance threshold for the Wolfinger test and positing the Wolfinger model and distribution assumptions; the

80% criteria influences the amount of confirming instances that the (subsequent) data mining algorithm must encounter, before it can generalize data into rules. Since we use the rules in a descriptive rather than a predictive fashion, they are meant to be used as a summarizer of the data assessments and for suggesting resultant (and testable) biological hypotheses. (Note that the rules describe the computed assessments based on modeling assumptions and an arbitrary confidence level threshold and not the expression data itself.)

Results

Drought Stress and Physiological Measurements

Trees reached the desired water potential for each cycle of stress at 3-4 days after withholding water for mild stress and at 6-7 days after withholding water for severe stress. Trees were subjected to three cycles of stress (Fig. 1). Control trees were watered normally throughout the experiment. Trees grown under mild stress showed little alteration in growth and, as did control trees, continued to produce new flushes of growth. Trees grown under severe stress had fewer new flushes of growth compared to controls, indicating that severe stress limited tree growth. Photosynthetic measurements were taken at the time of peak drought stress to identify the effect of the stress on photosynthesis. Both mild and severe stress levels led to reduced photosynthesis during the first cycle of stress with a much greater photosynthetic reduction in trees grown under severe stress (Table I). In subsequent cycles of stress, trees under mild stress showed recovery of photosynthetic rate to levels at or approaching those of controls, a result we considered to be photosynthetic acclimation to the stress. Trees grown under severe

stress remained much below control levels, a result we noted as photosynthetic failure during the stress.

Microarrays

The two stage, linear mixed model (Wolfinger et al., 2001) used by Expresso is intended to remove global effects from the computed intensity values. It also allows the estimation of significant changes in gene transcript levels below the standard 2.5 fold increase (Fig. 2). All reported results are significant at α=0.05. Table II shows the number of genes that responded by cycle and stress level. Between mild cycle 1 and 2, there was an increase in the number of genes showing regulation (from 127 to 188 positively affected clones and from 82 to 146 negatively affected clones), which correlates with photosynthetic acclimation. Trees grown under severe stress did not show such an increase and the number of genes showing a positive response remained consistent across all three cycles of severe stress. Figure 3 shows the genes that are regulated in both mild and severe stress. Generally, less than 20% of genes showing a response are shared between mild and severe stress cycles, which suggests that different response pathways are initiated for the different levels of stress. Many of the genes identified as being positively regulated were noted in our previous experiment (Heath et al., 2002) and include rubisco binding protein, aquaporin and isoflavone reductase (for a complete list of genes that responded, access our website at: http://bioinformatics.cs.vt.edu/~expresso/research_gene_expression_pine2.php

The results of the six normalization ANOVA models show that the array, treatment, pin and pin-array effects are significant in all cycles of mild and severe

treatment experiments ($p < 0.0001$). The dye effect, however, is significant in all mild treatment cycles ($p < 0.0001$) and in severe cycle 2 but not in severe cycle 1 ($p=0.3297$) and severe cycle 3 ($p=0.2038$). In microarray experiments, global effects in the intensity signals are expected as shown by the results. The normalization model employed in Expresso is successful in detecting and removing these effects from the logarithm of the intensity values.

Assessment of 2178 gene models per experiment shows that at $\alpha = 0.05$, significant gene-treatment interaction happens in 59.27 % to 74.31% of the gene models. Similarly, significant gene-array interaction, and gene-dye interactions happen in 75% to 87.71% and 63.35% to 80.92% of the gene models respectively. Table III shows the number of equations where the considered effects are significant at the same significance level. The results show that most genes interact with treatment, arrays, and dyes significantly. The gene model, therefore, proved adequate in detecting gene interactions with effects other than treatment and removing those from the gene-treatment effects we are interested in.

To aid in data mining by redescription, clones were assigned to the functional categories described in Heath et al. (2002). Though the clones were selected based on the MIPS functional categories, we used the gene annotations to re-categorize the genes in a hierarchy more reflective of stress response characteristics, e.g., the categories were based on plant responses to stress and the processes protected by the stress responses (Fig. 4). Notice that the categories have a hierarchical structure; if a clone X is in a given category C, it is also a member of all parent categories of C. In addition, genes can be placed in multiple categories. Inductive logic programming (ILP; Muggleton, 1999) was

applied to the statistically significant changes in gene expression levels for the various treatments together with the functional categories to obtain relationships between gene category and treatment. Signature patterns of changes in transcript levels relating to the functional categories were found using redescription (Figs. 5 and 6). An example of rules which shows that transcript levels of genes whose products participate in trafficking between the endoplasmic reticulum and the Golgi were elevated during acclimation (mild cycle 2 and 3) is:

expression(X,mc2,'+') :- category(X,trafficking).

expression(X,mc3,'+') :- category(X,trafficking).

Genes falling into the categories of 'nitrogen and sulfur metabolism', 'respiratory electron transport', and 'cell membranes' were also identified by ILP as showing significant changes in transcript levels in response to drought stress. We further focused on three groups of genes that were identified using ILP. One group, which can be found in the intersection of the categories 'Gene Expression' and 'ROS and Stress', includes 'protection and repair' genes such as the chaperones. These responded positively during acclimation. The second group of genes is included in the subcategories of 'carbon metabolism' and showed a positive response during all three cycles of mild stress with highest expression during acclimation. The third group falls under the category of 'phenylpropanoid metabolism' and showed a positive response during the first and third cycles of mild stress.

Protection and Repair Genes (Chaperones)

Genes encoding the chaperones can be found in the subcategories of protection and repair, heat, and cold within 'ROS and Stress' (Fig. 4). In trees grown under mild stress, more genes within 'heat' and 'protection and repair' had increased transcript levels compared to severe stress (Figs. 5 and 6). However, chaperones and other genes correlated with cold stress were higher in trees grown under severe stress. A more detailed analysis of some of the HSP and LEA genes identified as having altered transcript levels revealed that LEA group 2 homologs were associated more with mild stress whereas LEA group 3 were associated more with severe stress (Fig. 7). Figure 7 also shows that specific clones of HSP 70 and HSP 90 respond differently to mild and severe stress. For example an HSP 70, DNA k type homolog (clone NXSI_117_C08, contig 8005) and an HSP 90 homolog (NXSI_116_B04) show increased transcript levels during mild stress and no response during severe stress. Conversely, a mitochondrial HSP 70 (NXCI_022_G01) and a different HSP 90 (NXNV_149_E10, contig 5815) show increased transcript levels during severe stress and no response during mild stress.

Carbon metabolism

Genes in the category of 'Carbon Metabolism' were identified by ILP as having increased transcript levels in trees grown under mild stress. We focused on the genes associated with core carbon metabolism (Fig. 8). Genes that showed positive regulation in mild stress include pyruvate kinase, which leads to the formation of pyruvate from phosphoenolpyruvate, and pyruvate dehyrdrogenase, which feeds carbon from glycolysis into the TCA cycle. Other genes that show increased transcript levels in mild stress

include those in the oxidative pentose phosphate pathway such as ribose-5-phosphate epimerase, transaldolase and transketolase. Genes associated with the reductive pentose phosphate pathway such as RUBISCO and 2 plastidic isoforms of glyceraldehyde-3-phosphate show increased transcript levels in severe stress.

Phenylpropanoid Pathway and Aromatic Amino Acids

ILP produced rules concerning expression of genes involved in aromatic amino acid biosynthesis and in the phenylpropanoid pathway (Fig. 9). As there is a link between one branch of the aromatic amino acid pathway and the phenylpropanoid pathway, these groups were examined together. Genes associated with the entry of carbon into these pathways, DAHP synthase and 3-dehydroquinate synthase, had higher transcript levels in trees grown under mild stress compared to severe stress. Phenylalanine ammonia lyase (PAL) did not show a response in mild stress whereas cinnamate-4-hydroxlase had positive expression in cycles 1 and 3 of mild stress. Nearly all the genes involved with flavonoid biosynthesis that were represented on the array showed an increase under mild stress. Genes associated with lignin generally showed no response to mild or severe drought stress. Of the 12 laccase clones present on the array, only 4 showed a response with one isoform of laccase showing a positive response in mild stress, while another showed a positive response in severe (Fig. 9).

Real Time PCR

Five genes shown to have different transcript levels in response to either mild or severe stress were chosen for further analysis using Real Time PCR (RT PCR). Two

genes encoding LEA proteins were found to have similar expression profiles using RT PCR as compared with the microarrays. Pine LEA g3 (P14G04) was shown to be up-expressed in all 3 cycles of severe stress using microarray analysis (Fig. 7). A similar result was seen with the RT PCR where there was a 2.5 fold increase in transcript level during the second cycle of severe stress and a 1.7 fold increase in transcript level during the third cycle of severe stress (Table IV). The results generated for pLEAg2 by RT PCR were also similar to those generated by microarray analysis (Table IV and Fig. 7) with an increase in transcript levels during the third cycle of mild stress but no change during severe stress. The flavonoid genes chalcone isomerase and naraningenin-2-oxo dioxygenase (3-flavonone hydroxylase) were also found to have similar transcript profiles during drought stress when analyzed using RT PCR compared to the results obtained from microarrays (Table IV and Fig. 9).

Discussion

The analysis of drought response in loblolly pine using microarrays has implications for many aspects of the timber industry, including crop improvement. Results obtained with pine serve as a model to expand the knowledge of angiosperm gene analysis into gymnosperms. We have previously demonstrated that 72 of 384 genes show increased transcript abundance in loblolly pine grown under mild stress whereas 69 of those genes showed no response in severe stress (Heath et al., 2002). These expression patterns were correlated with acclimation to the stress. To further analyze the response of loblolly pine to drought, we used a 2178 gene array. We also took photosynthetic measurements at the time of tissue harvest. Pines grown under mild stress showed an

initial reduction in photosynthetic rate, with recovery to near control levels in subsequent drought cycles. The recovery in photosynthetic rate is an expression of photosynthetic acclimation to the drought stress.

The change in RNA transcript profiles of *P. taeda* due to drought stress can be correlated with physiological data reflecting photosynthetic acclimation to mild stress or photosynthetic failure during imposition of severe stress. The number of clones showing either positive or negative changes in transcript levels increases from 5.8% to 8.6% between mild cycle 1 and 2, suggesting that the behavior of these genes is correlated with acclimation. The near doubling of the number of genes showing negative changes (3.8% to 6.7 %) suggests that negative gene expression shares importance with positive gene expression in acclimation to the imposed stress. There was no comparable change in trees grown under severe water deficit where changes in genes showing altered transcript profiles remained at 6%. However, in both mild and severe stress the genes with altered transcript profiles dropped in the third cycle of stress.

The large number (approximately 80% of genes showing changes in transcript profiles) of genes unique to each stress level suggests that the plant is able to sense the degree of stress and activate different response pathways to cope with that stress. These different response pathways contain related genes; however, clones with the same annotation seem to be differentially expressed members of a multi-gene family. For example, one clone of laccase is positively regulated during mild stress and negatively regulated during severe, whereas another laccase gene is negatively regulated during mild stress and positively regulated in severe. Other examples include genes that encode annexin, a Ca^{2+} binding protein involved in diverse cellular functions (Gerke and Moss,

2002); ADP-ribosylation factor, a GTPase involved in vesicle budding and membrane trafficking (Takeuchi et al., 2002); and myo inositol-1-phosphate synthase (see supplemental data at http://bioinformatics.cs.vt.edu/~expresso/research_gene_expression_pine2.php). These results suggest that microarrays can differentiate among some members of a multi-gene family and that, in at least some cases, microarrays can be used to uncover the different function each member of a multi-gene family might have in plant physiology.

The expression of genes encoding chaperones have been associated with responses to stress imposition (Xu et al., 1996; Lee and Vierling, 2000; Zhang et al., 2000). We previously identified chaperones as being positively regulated in response to drought and suggested that they aided in adaptation to drought stress by protecting processes within the cell (Heath et al., 2002). Lee and Vierling (2000) suggest that one protection afforded by HSPs is manifested through the interaction of small HSPs with denatured proteins, preventing aggregation and promoting refolding of denatured proteins. Through ILP we have found that different chaperones were expressed in either mild or severe stress (Fig. 7). We generated a phylogenetic tree to determine how the *P. taeda* LEAs are related to each other and to LEAs from other species (Fig. 10). Two pine clones associate with group 2 LEAs (dehydrins) from cotton, Arabidopsis, and white spruce. Studies of dehydrin in cowpea point to a possible function in membrane protection (Ismail et al., 1999). Furthermore, Iuchi et al. (1996) found that group 2 LEAs were highly expressed under drought stress in a drought-tolerant cowpea. The group 2 LEAs from pine had higher transcript levels during mild stress suggesting that they are important for acclimation, possibly by stabilizing membranes. Several pine clones show

close homology to group 3 LEAs from Arabidopsis and cotton. A group 3 LEA from Barley (HVA1) confers tolerance to water and salt stress when expressed in rice (Xu et al., 1996). The pine group 3 LEAs had increased transcript levels in severe stress. We previously identified HSPs and LEAs as being responsive to drought stress and our results suggested that different homologs responded to the different levels of stress (Heath et al., 2002). The results of this study corroborate our previous experiment showing that the changes in expression of LEAs and HSPs are important to drought stress reponses in general. However, some, such as the group 2 LEAs, are more important during acclimation, whereas the group 3 LEAs may be more important to protection during severe stress. The LEAs that responded under severe stress have also been associated with cold responses (Dong et al., 2002) and ILP indicated that other cold associated genes showed higher transcript levels in severe stress than in mild. This suggests that severe drought stress activates different response pathways. It also indicates that the plant can sense the degree of stress and respond accordingly.

ILP found several rules indicating significant changes in carbon metabolism in response to mild drought stress. Our data suggest that carbon metabolism has a role in acclimation to drought stress. Using cluster analysis, Scheideler et al. (2001) found that genes categorized in carbon metabolism tended to have higher transcript levels in Arabidopsis treated with *Pseudomonas syringae* pv. *tomato*. They suggested that transcript regulation of genes encoding enzymes involved in carbon metabolism could direct the flow of carbon into specific pathways necessary to the response. That appears to be the case here, where at least two genes encoding enzymes involved in carbon metabolism, which could provide carbon skeletons to other metabolic processes, showed

increased transcript levels during mild stress. One of these genes encodes pyruvate kinase, which may be involved in providing carbon skeletons for amino acid biosynthesis. Since it is also an ATP producing enzyme, it may be necessary to provide reducing equivalents for other biosynthetic reactions. Another gene encodes pyruvate dehydrogenase, which is important for shuttling carbon from glycolysis into the TCA cycle. None of the other enzymes of the TCA cycle show increased transcript levels, which suggests that there is no net increase in flow through the TCA cycle. It is important to note, however, that regulation of these enzymes could be post-transcriptional. Since the TCA cycle provides carbon skeletons for many biosynthetic reactions, the increased transcript level of pyruvate dehydrogenase could be providing carbon for diverse uses within the plant. The maintenance and possible increase of general carbon metabolic pathways could be reflective of the acclimation process during mild drought stress. Trees grown under severe stress had reduced transcript levels of genes associated with reductive pentose phosphate pathway, suggesting that little photosynthate was being generated, which is in agreement with the photosynthetic data. Under severe stress, there is a general increase in mitochondrial electron transport (supplemental data) suggesting that the plant is compensating for drought-induced lack of reducing equivalents from the light reactions.

RT PCR analysis of the genes discussed above corroborated the results generated by the microarray data. Three of the genes within the functional category of 'Chaperones' were analyzed. As with microarray analysis, the group 2 LEAs were more responsive during photosynthetic acclimation in trees grown under mild water deficit. The group 3 LEAs, however, had higher transcript levels in severe stress suggesting a protective role

during photosynthetic failure. The two genes in 'flavonoids' that were selected for RT PCR analysis also had a similar expression profile with the two techniques. Therefore, the results of the RT PCR analysis serve to confirm the results of the microarray analysis and add support to the conclusions drawn from the microarray data.

Other rules found by ILP suggest that nitrogen and sulfur metabolism and phenylpropanoid metabolism are important to acclimation to drought stress. Within nitrogen and sulfur metabolism, an increase in transcripts encoding enzymes of aromatic amino acid biosynthesis was found. An increase in transcript levels of several genes in the oxidative pentose phosphate pathway under mild stress (Fig. 8) suggests that carbon is being shuttled to aromatic amino acid biosynthesis. A similar increase is not seen in severe stress cycles. Since there is no overall positive trend in the genes of lignin biosynthesis, we suggest that carbon from aromatic amino acids is being channeled through the phenylpropanoid pathway into flavonoids. A rise in transcript levels of some genes important to flavonoid biosynthesis is seen under mild stress. Flavonoids have been implicated in stress responses (Winkel-Shirley, 2002), and we previously noted that isoflavone reductase was positively regulated in trees grown under mild stress but not trees grown under severe stress (Heath et al., 2002). An increase in dihydroflavonol-4-reductase transcripts was noted during osmotic, ionic, and heat stress in yeast (Garay-Arroyo, 1999). Dihydroflavonol-4-reductase also showed an increase in transcript levels during dehydration in a drought resistant cow pea (*Vigna unguiculata*; Iuchi et al., 1996) suggesting that dihydroflavonols are important to resistance to the stress. Flavonone-3-hydroxylase (naringenin-2-oxo dioxygenase) is the first enzyme in the production of dihydroflavonols and showed increased transcript levels in all 3 cycles of mild stress.

Since photosynthetic acclimation only occurred in mild stress, it is possible that the flow of carbon to the flavonoids and specifically the dihydroflavonols is important to photosynthetic acclimation to mild drought stress.

Another gene associated with nitrogen and sulfur metabolism whose expression was affected by mild stress is arginine decarboxylase. This gene product is involved in polyamine biosynthesis and was noted as being drought responsive by Ozturk et al (2002). Arginine decarboxylase had higher transcript levels in the third cycle of mild stress when the plants have acclimated. Arginine decarboxylase along with s-adenosylmethionine (SAM) decarboxylase had lower transcript levels in trees grown under severe stress and were identified by ILP as being negatively regulated during severe cycles 1 and 2 (supplemental data). Other researchers have correlated SAM decarboxylase with drought stress, suggesting that polyamine biosynthesis is important to drought stress (Li and Chen, 2000; Li and Chen, 2000b). Spermine and spermidine, two polyamines in the SAM decarboxylase pathway, were found to induce elongation growth, increase photosynthetic capacity and reduce membrane leakage in drought stressed jack pine (*Pinus banksiana*; Rajasekeran and Blake 1999) pointing to a direct role of polyamines in cellular responses to drought stress. The negative regulation of arginine decarboxylase and SAM decarboxylase during severe drought stress may suggest that synthesis of the polyamines is reduced and that the polyamines are not functional in the protection of cellular processes during severe stress. However, SAM decarboxylase is important to ethylene synthesis, since it diverts SAM away from the ethylene biosynthetic pathway. The down-regulation of SAM decarboxylase may therefore reflect an increase in ethylene production in trees grown under severe stress.

Ethylene has been associated with many stress responses in plants (e.g., Chen et al., 2002), as well as with abscission and senescence (Taylor and Whitelaw, 2001). Since ethylene is associated with senescence, the trees grown under severe stress may be switching to a senescence program. Two senescence-related genes were found to have higher transcript levels in trees at the third cycle of severe stress (supplemental data).

Many of the same genes present on our 384 element arrays (Alscher et al., 2002) were present on the arrays used here allowing us to corroborate our previous data. The data presented here reflects changes in transcript profiles of loblolly pine over several cycles of stress, and it can be seen that the changes in transcript profiles of loblolly pine as discovered through the use of microarrays are both quantitative and qualitative. A large number of genes showed changes in transcript levels due to mild stress, and there was an increase in these genes with acclimation. Many of the genes identified have been previously correlated with plant stress responses including LEAs and HSPs. However, we found that certain homologs of the same gene respond differently to the two levels of drought stress. Again, this indicates that the plant is able to sense the degree of stress and may activate different response pathways accordingly. Many of the other genes identified in this study encode proteins associated with 'house-keeping' functions (pyruvate dehydrogenase and pyruvate kinase). This suggests that alteration of 'normal' metabolism can play a significant role in the stress response, perhaps in directing carbon or energy into stress-response pathways. The results presented here have been used in probe selection and design of our next microarray experiment.

One of the major advantages of data mining, as employed in Expresso, is the support for incorporating *partial information* into the analysis. Such partial information

can take many forms: results from previous years' experiments, prior data about which genes are implicated in which physiological processes, and information from the literature. Data mining techniques should have the ability to flexibly incorporate such partial information and build from them, rather than function as obscure, isolated algorithms. The inductive logic programming technique described in this paper is only one of the many approaches that can harness partial information. We are now supporting the construction and mining of multimodal networks, to help piece together parts of biological pathways. This approach uses algorithms from graph theory and probabilistic modeling to support the mining of such networks. Together, these approaches help address some of the grand challenges in bioinformatics data mining.

Literature Cited

Alscher R, Chevone B, Heath L and Ramakrishnan N (2001) Expresso: A problem solving environment for bioinformatics: Finding answers with microarray technology. In Tentner A. (ed.) *Proceedings of the High Perfomance Computing Symposium, Advanced Simulation Technologies Conference*. Society for Computer Simulation International, Seatttle, pp 63-69.

Alscher RG, Donahue JL and Cramer CL (1997) Reactive oxygen species and antioxidants: Relationships in green cells. Physiol. Plantarum 100: 224-233.

Babiychuk E, Kushnir S, Belles-Boix E, Van Montagu M and Inze D (1995) Arabidopsis thaliana NADPH oxidoreductase homologs confer tolerance of yeasts towards the thiol-oxidizing drug diamide. J. Biol. Chem. 270: 26224-26231.

Chang S, Puryear J and Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Molec. Biol. Reporter 11: 113-116.

Chen WQ, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou GZ, Whitham SA, Budworth PR, Tao Y, Xie ZY, Chen X, Lam S, Kreps JA, Harper JF, Si-Ammour A, Mauch-Mani B, Heinlein M, Kobayashi K, Hohn T, Dangl JL, Wang X, Zhu T (2002) Expression profile matrix of Arabidopsis transcription factor

genes suggests their putative functions in response to environmental stresses. Plant Cell 14: 559-574.

Dong CN, Danyluk J, Wilson KE, Pecock T, Hunner NPA and Sarhan F (2002) Cold-regulated cereal chloroplast late embryogenesis abundant-like proteins: molecular characterization and functional analysis. Plant Physiol. 129: 1368-1381.

Dzeroski S and Lavrac N (editors). 2001. Relational Data Mining, Springer, Berlin.

Fowler S and Thomashow MF (2002) Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. Plant Cell 14: 1675-1690.

Gang DR, Kasahara H, Vander Mijnsbrugge K, Bauw G, Boerjan W, Van Montagu M, Davin LB and Lewis NG (1999) Evolution of plant defense mechanisms - Relationships of phenylcoumaran benzylic ether reductases to pinoresinol-lariciresinol and isoflavone reductases. J. Biol. Chem. 274: 7516-7527.

Garay-Arroyo A and Covarrubias AA (1999) Three genes whose expression is induced by stress in *Saccharomyces cerevisiae.* Yeast. 15: 879-892.

Girke T, Todd J, Ruuska S, White J, Benning C and Ohlrogge J (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol. 124: 1570-1581

Heath LS, Ramakrishnan N, Sederoff RR, Whetten RW, Chevone BI, Struble CI, Jouenne VY, Chen D, van Zyl L and Grene R (2002) Studying the functional genomics of stress reponses in loblolly pine with the Expresso microarray experiment management system. Comp. Funct. Genom. 3: 226-243.

Iuchi S, YamaguchiShinozaki K, Urao T, Terao T and Shinosaki K (1996) Novel drought inducible genes in the highly drought-tolerant cowpea: cloning of cDNAs and analysis of the expression of the corresponding genes. Plant Cell Physiol. 37: 1073-1082

Kang J, Choi H, Im M and Kim SY (2002) Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell 14: 343-357.

Katageri T, Takahashi S and Shinozaki K (2001) Involvement of a novel Arabidopsis phospholipase D AtPLD, in dehydration-inducible accumulation of phosphatidic acid in stress signalling. Plant J. 26: 595-605.

Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D and Bohnert HJ (2001) Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13: 889-906.

Kerr MK and Churchill GA (2001) Statistical design and the analysis of gene expression microarray data. Genet. Res. 77: 123-128.

Koyuturk M, Grama AY, and Ramakrishnan, N (2002) Algebraic techniques for analysis of large discrete-valued datasets, in Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'02).

Lee GJ and Vierling E (2000) A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. Plant Physiol. 122: 189-198.

Lev-Yadun S and Sederoff R (2000) Pines as model gymnosperms to study evolution, wood formation and perennial growth. J. Plant Growth Reg. 19: 290-305.

Li ZY and Chen SY (2000a) Isolation and characterization of a salt- and drought-inducible gene for S-adenosylmethionine decarboxylase from wheat (Triticum aestivum L.). J. Plant Physiol. 156: 386-393.

Li ZY and Chen SY (2000b) Differential accumulation of the S-adenosylmethionine decarboxylase transcript in rice seedlings in response to salt and drought stresses. Theor. Appl. Genet. 100: 782-788.

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K and Shinozaki K (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, in arabidopsis. Plant Cell 10: 1391-1406.

Muggleton S (1999) Scientific knowledge discovery using inductive logic programming, Communications of the ACM, 42:42-46.

Muggleton S and Feng C (1990) Efficient induction of logic programs. In Proceedings of the First International Conference on Algorithmic Learning Theory, Arikawa S., Goto S., Ohsuga, S., and Yokomori, T (eds.), Japanese Society for Artificial Intelligence, Tokyo, 368-381.

Ozturk ZN, Talame V, Deyholos M, Michalowski CB, Galbraith DW, Gozukirmizi N, Tuberosa R and Bohnert HJ (2002) Monitoring large-scale changes in transcript abundance in drough- and salt-stressed barley. Plant Molec. Biol. 48: 551-573.

Perez-Amador MA, Lidder P, Johnson MA, Landgraf F, Wisman E and Green PJ. 2001. New molecular phenotypes in the dst mutants of arabidopsis revealed by DNA microarray analysis. Plant Cell 13: 2703-2717.

Rajasekaran LR and Blake TJ (1999) New plant growth regulators protect photosynthesis and enhance growth under drought of jack pine seedlings. J. Plant Growth Regul. 18: 175-181.

Ristic Z, Yang GP, Martin B and Fullerton S (1998) Evidence of association between specific heat shock proteins and the drought and heat tolerance phenotype in maize. J. Plant Physiol. 153: 497-505.

Ruan Y, Gilmore J and Conner T (1998) Towards arabidopsis genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. Plant J. 15:821-833.

Sang Y, Zheng S, Li W, Huang B and Wang X. 2001. Regulation of plant water loss by manipulating the expression of phospholipase D. Plant J. 28: 135-144.

Scheideler M, Schlaich NL, Fellenberg K, Beissbarth T, Hauser NC, Vingron M, Slusarenko AJ and Hoheisel JD (2001) Monitoring the switch from housekeeping to defense metabolism in Arabidopsis thaliana using cDNA microarrays. J. Biol. Chem. 277: 10555-10561.

Schultz RP (1999) Loblolly - the pine for the twenty-first century. New Forests 17: 71-88.

Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y and Shinozaki K (2001) Monitoring the expression patterns of 1300 arabidopsis genes under drought and cold streses by using a full-length cDNA microarray. Plant Cell 13:61-72.

Sioson A, Watkinson JI, Vasquez-Robinet C, Ellis M, Shukla M, Kumar D, Ramakrishnan N, Heath LS, Grene R, Chevone, BI, Kafadar, K and Watson LT (2003) Expresso and Chips: Creating a Next Generation Microarray Experiment Management System, in *Proceedings of the Next Generation Software Workshop, 17th International Parallel and Distributed Processing Symposium*, Nice, France, 8 pages, IEEE Computer Society.

Somerville C and Somerville S (1999) Plant functional genomics. Science 285: 380-383.

Stasolla C, van Zyl L, Egertsdotter U, Craig D, Liu W, and Sederoff RR (2003) The effects of polyethylene glycol on gene expression of developing white spruce somatic embryos. Plant Physiol. 131: 49-60.

Taylor TE and Whitelaw GA (2001) Signals in abcission. New Phytol. 151: 323-339.

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr. Opin. Plant Biol. 5: 218-223.

Xu D,. Duan X, Wang B, Hong B, Ho THD and Wu R (1996) Expression of a late embryogenesis abundant protein gene, HVA1, from barley confers tolerance to water deficit and salt stress in transgenic rice. Plant Physiol. 110: 249-156.

Zhang L, Ohta A, Takagi M, and Imai R (2000) Expression of plant group 2 and group 3 lea genes in Saccharomyces cerevisiae revealed functional divergence among LEA proteins. J. Biochem. 127: 611-616.

Table I. Effect of mild or severe drought stress on net photosynthesis in one year old loblolly pine rooted cuttings.

Net Photosynthesis (μmol CO_2 m^{-2} s^{-1})[a]			
Condition	**Cycle**	**Control**	**Stressed**
Mild	**1**	**4.28**	**2.48**
	2	**3.54**	**3.82**
	3	**4.75**	**3.28**
Severe	**1**	**3.67**	**0.88**
	2	**3.00**	**0.19**
	3	**2.90**	**0.77**

[a]Measurements were made on the first fully mature fasicle in each case, using the Li-Cor 6400. Three or four repeated measurements were made in each case.

Table II. Effect of mild or severe drought stress on number of clones showing differential regulation

Significant[a] Gene Expression			
Condition	**Cycle**	**# of clones (+) expressed (T:C)**	**# of clones (-) expressed (T:C)**
Mild	**1**	**133**	**94**
	2	**213**	**159**
	3	**62**	**90**
Severe	**1**	**145**	**144**
	2	**162**	**156**
	3	**135**	**53**

[a] Significant changes in gene expression were identified using a linear model to normalize our data and a linear mixed model to find the significance of the interaction between gene and treatment. (Wolfinger et al. 2001). Only those changes with a 0.05 significance or lower were considered. (T = treatment, C= control.) 2,178 clones on array.

Table III. Number of Genes out of 2178 where the interactions *GT*, *GA*, and *GD* are significant at $\alpha = 0.05$ in mild vs. control cycles and severe vs. control cycles.

	Mild Treatment vs. Control			Severe Treatment vs. Control		
	Cycle 1	Cycle 2	Cycle 3	Cycle 1	Cycle 2	Cycle 3
GT	1382	1620	1292	1535	1448	1340
GA	1764	1427	1481	1406	1496	1381
GD	1736	1776	1835	1693	1912	1635

Table IV. Real Time PCR results of selected genes shown by microarray analysis to have altered transcript levels during either mild or severe drought stress.

Clone ID	Mild Drought Stress		Severe Drought Stress	
	Cycle 2	Cycle 3	Cycle 2	Cycle 3
07 H08	0.899[a] ± 0.326	3.66 ± 0.469	0.993 ± 0.249	2.06 ± 0.371
NXSI_063_D01	0.407 ± 0.182	2.1 ± 1.73	1.28 ± 1.5	1.59 ± 0.369
PC14G04	0.845 ± 0.836	0.62 ± 0.332	2.54 ± 1.01	1.74 ± 0.421
NXCI_002_G10	0.79 ± 0.122	2.13 ± 0.562	1.45 ± 0.327	1.75 ± 0.486
ST40F04	0.51 ± 0.506	2.93 ± 0.909	1.00 ± 1.41	1.29 ± 1.53

[a] Indicates fold change (where 1 is equal to no change, numbers greater than 1 represent a positive change in gene expression and numbers less than 1 represent negative changes in gene expression) plus or minus one standard deviation in RNA level of target gene in treated samples as compared to control samples.

Figure 1. Graphical representation of experimental design for mild and severe drought stress treatments (does not reflect actual time period)

Water was withheld from treated trees until a desired water potential was reached (-1MP, -10 bars for mild stress, and -1.5 Mpa, -15 bars for severe stress) as determined by pre-dawn, water potential measurements. At maximum drought stress, photosynthesis was measured and plants were re-watered. Needles and stems were harvested 24 h after re-watering.

Figure 2 Gene significance results for the first cycle of milddrought stress vs. controls on rooted cuttings of loblolly pine. The result of analysis of the first cycle data of Mild vs Control experiment is shown in Figure 1. The vertical axis represents $-\log_{10}(p\text{-}value)$ while the horizontal axis represents lg(*fold change*). Each spot at the right of the vertical axis and above the threshold value $-\log_{10}(0.05) \approx 1.3$ are positively expressed clones. Similarly, those at the left and above the threshold line are negatively expressed.

Figure 3. Venn diagrams representing number of unique clones responding either positively (A) or negatively (B) to mild or severe stress.

Genes showing significant changes in transcript level were compared between mild and severe stress. The results are presented in the form of Venn diagrams here.

Figure 4. Schematic diagram of hierarchy of gene functional categories.

Clones were assigned to functional categories based on information in the gene annotation and information from the litereature. The categories are based on known plant

stress responses and the process that are thought to be protected by those stresses. By virtue of being placed in a category, a clone automatically becomes a member of all parent categories. In addition, a clone can be placed in multiple functional categories that are not direct ancestors of each other (see chaperones above).

Figure 5. Signature patterns of changes in gene expression profiles for trees grown under mild drought stress. The data generated through microarray analysis were redescribed using ILP such that expression patterns were related to functional categories. The rules generated by ILP are represented as colored ovals in the diagram. Subcategories are represented as ovals within larger ovals. The green oval represents a subcategory that responded and is shared between two main categories.

Figure 6. Signature patterns of changes in gene expression profiles for trees grown under severe drought stress. Diagrams were developed as in Figure 4.

Figure 7. Response of genes encoding chaperones to mild and severe drought stress as discovered through microarrays and ILP.
Genes encoding chaperones are grouped according to the chaperone class in which they fall. Changes in transcript profiles are shown as positive (+), negative (-) or unchanged (0)

Figure 8. Transcript profiles of genes in the category of carbon metabolism during three cycles of mild or severe drought stress.

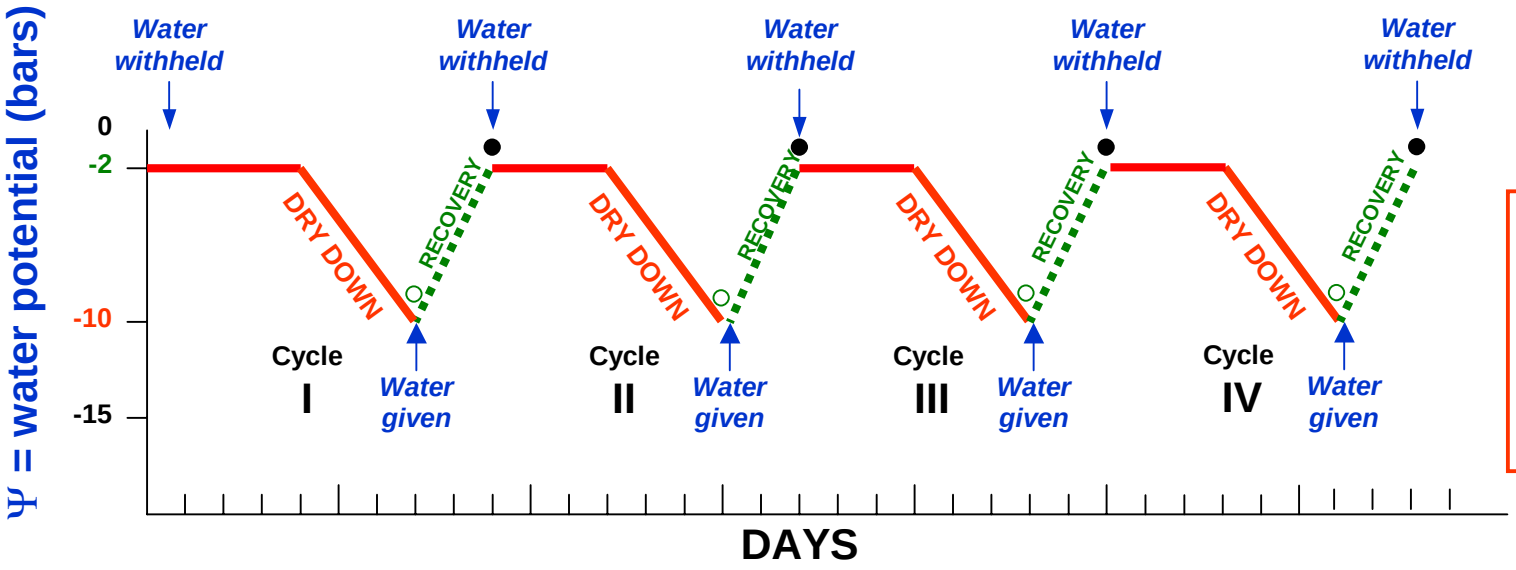
Genes within the categories of 'carbon metabolism' were identified by ILP as having increased transcript levels in trees grown under mild stress. The genes within these categories that are present on the array are represented here . Expression is indicated as positive (+), negative (-) or unchanged (0). Blue squares indicate genes that were not represented on the array.

Figure 9. Transcript profiles of genes in the categories of 'phenylpropanoid metabolism' and 'aromatic amino acids' during three cycles of either mild or severe drought stress. Genes within the categories of 'phenylpropanoid metabolism' and 'aromatic amino acids' were identified by ILP as having increased transcript levels in trees grown under mild stress. The genes within these categories (identified by annotation) that are present in a pathway are represented here. Expression is indicated as positive (+), negative (-) or unchanged (0). Blue squares indicate genes that were not represented on the array.

Figure 10. Phylogenetic tree of different groups of LEA proteins from Pinus taeda with those from other species.

The tree was generated using pine contigs (that contain the ESTs with differential expression). BLASTX was used to obtain homologous sequences in Genbank. Alignment of translated pine sequences (translated using the SIXFRAME tool in workbench http://workbench.sdsc.edu) and homologous sequences was performed using the CLUSTALW tool of workbench. The parameters used were the default parameters from workbench. The njplot tool was used to visualize the resulting tree and to generate a postscript file. Color and numbering was added to the postscript file. Pine clones are identified by Contig number and clone ID, other plant proteins are identified by GenBank

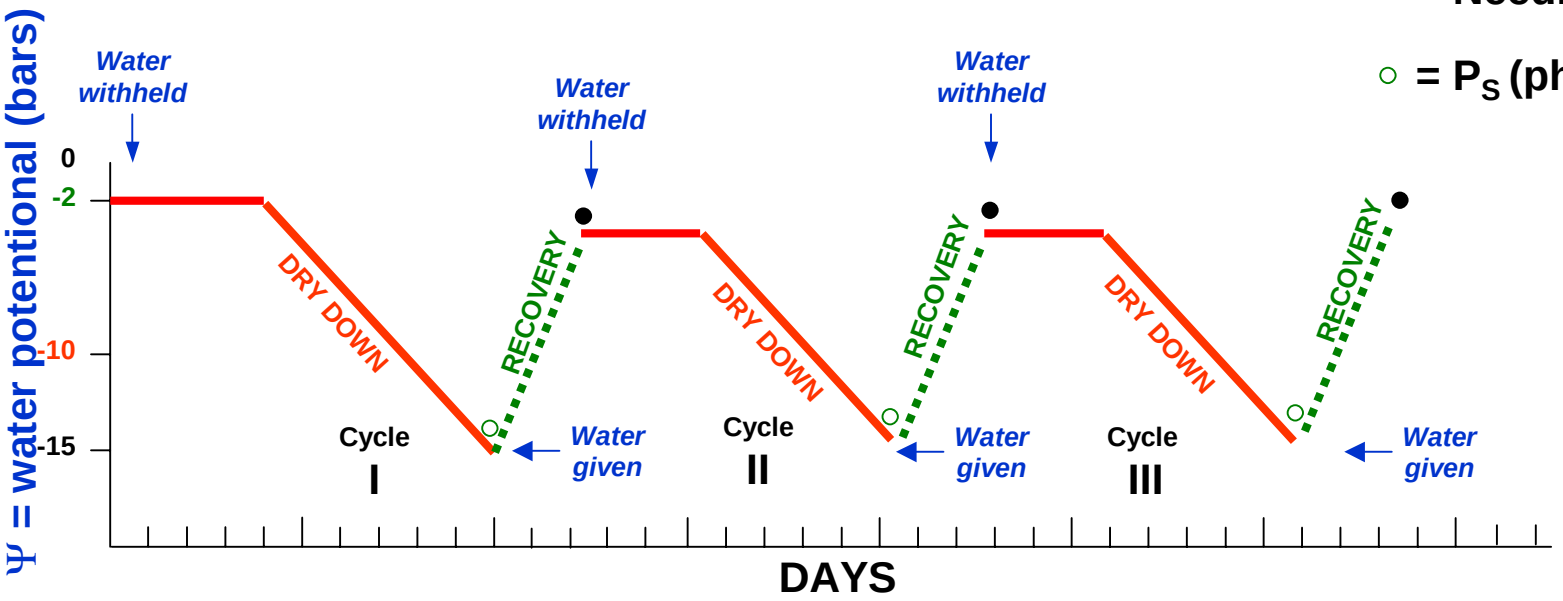
numbers. Group 4 (G4) LEAs make up the first main group in black. Group 2 (G2) LEAs are colored in blue, Group 3 (G3) LEAs are colored in burgundy, and Group 5 LEAs are colored green. The Group 1 (G1) LEAs are colored black and lie between Group 2 and Group 3.



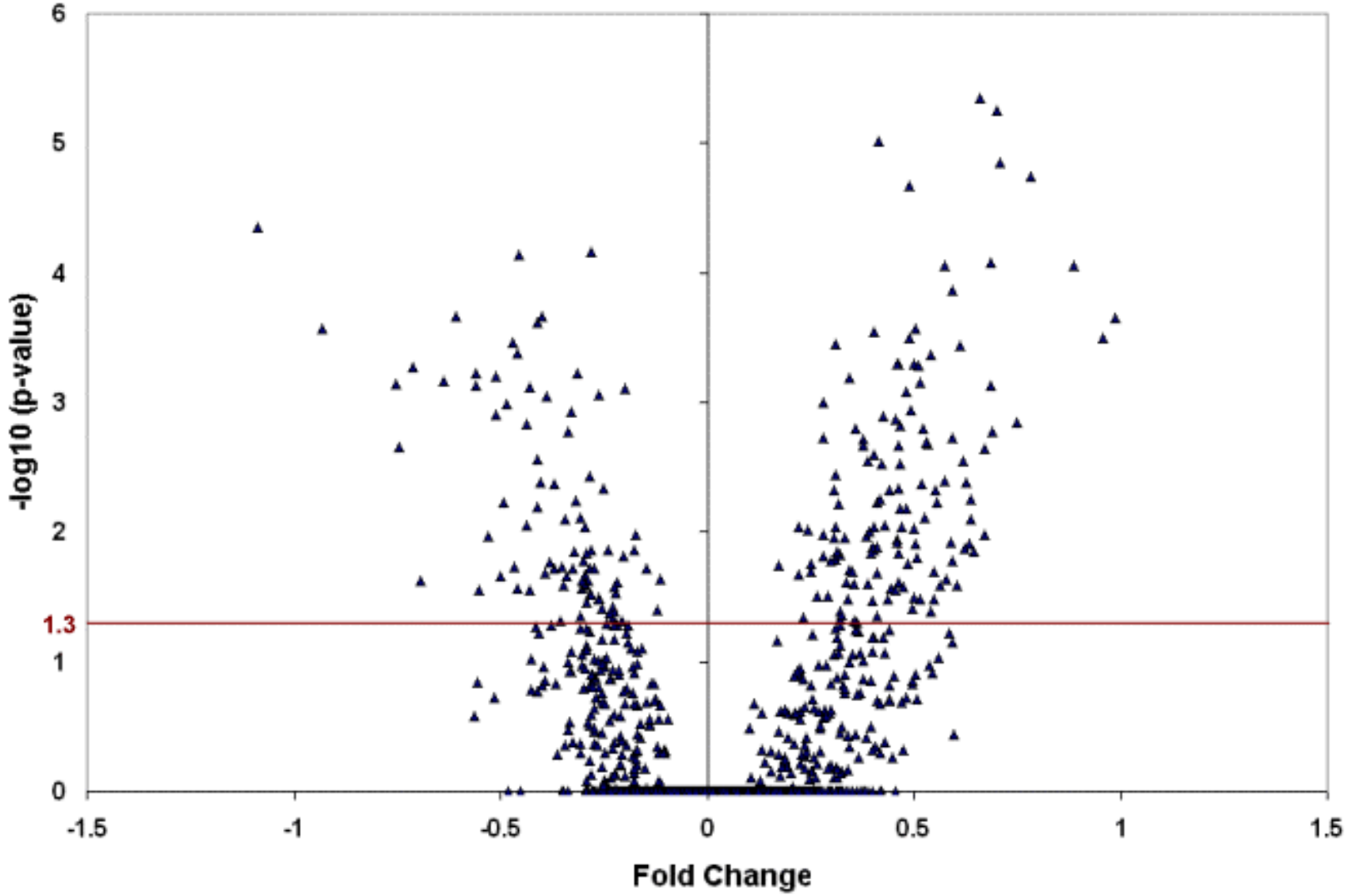
Experiment 1: Cycles of Mild Drought Stress

● = Needles harvest

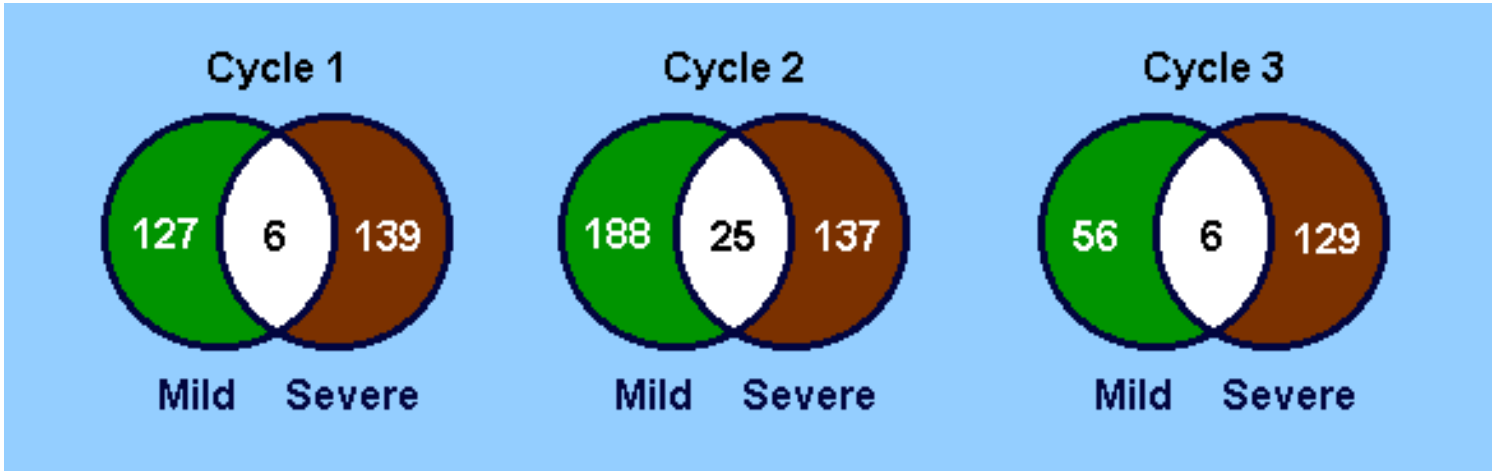
○ = P_S (photosynthesis)



Experiment2: Cycles of Severe Drought Stress

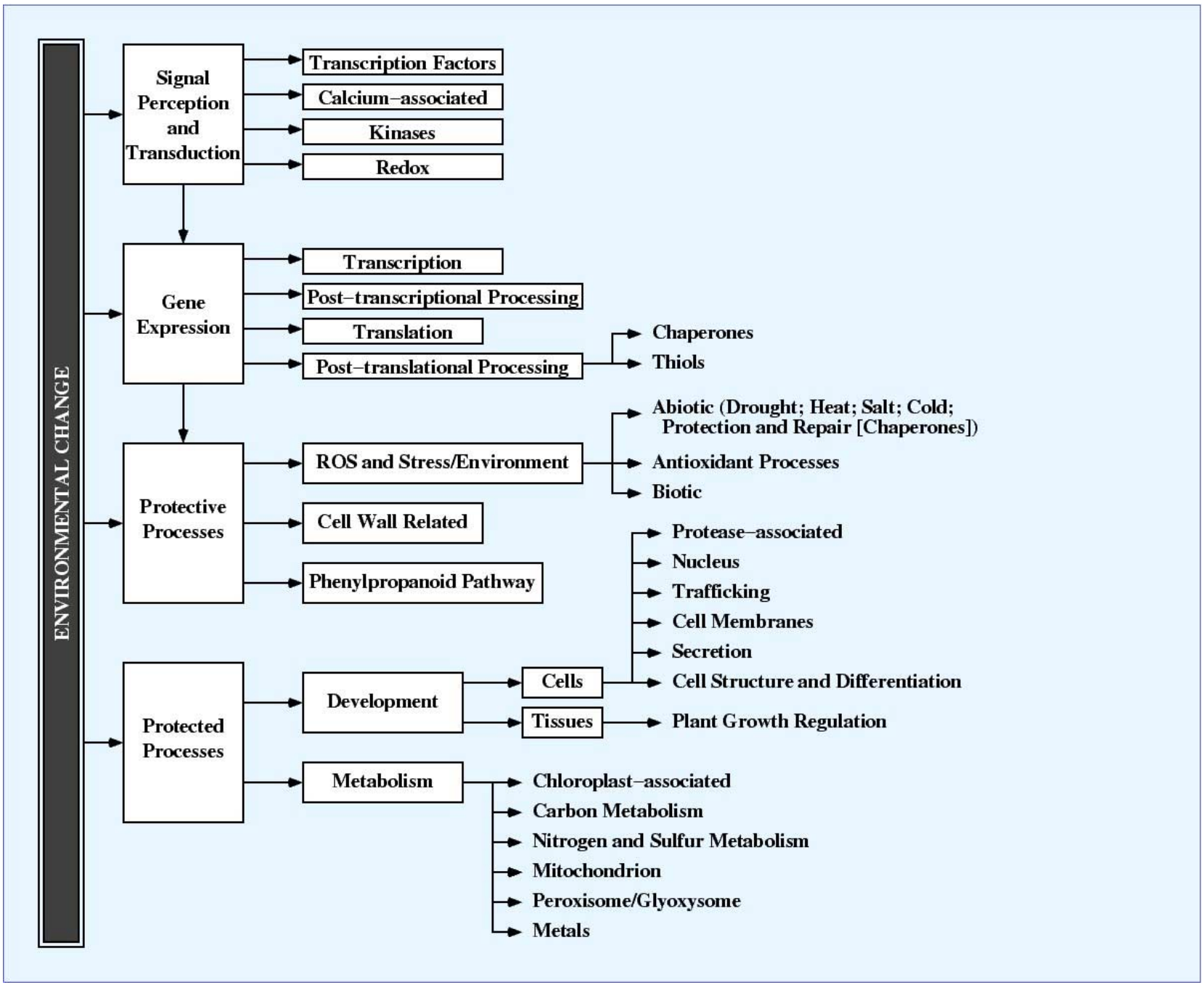

6
5
4
3
2
1.3
1
0
-log10 (p-value)
-1.5
-1
-0.5
0
0.5
1
1.5
Fold Change

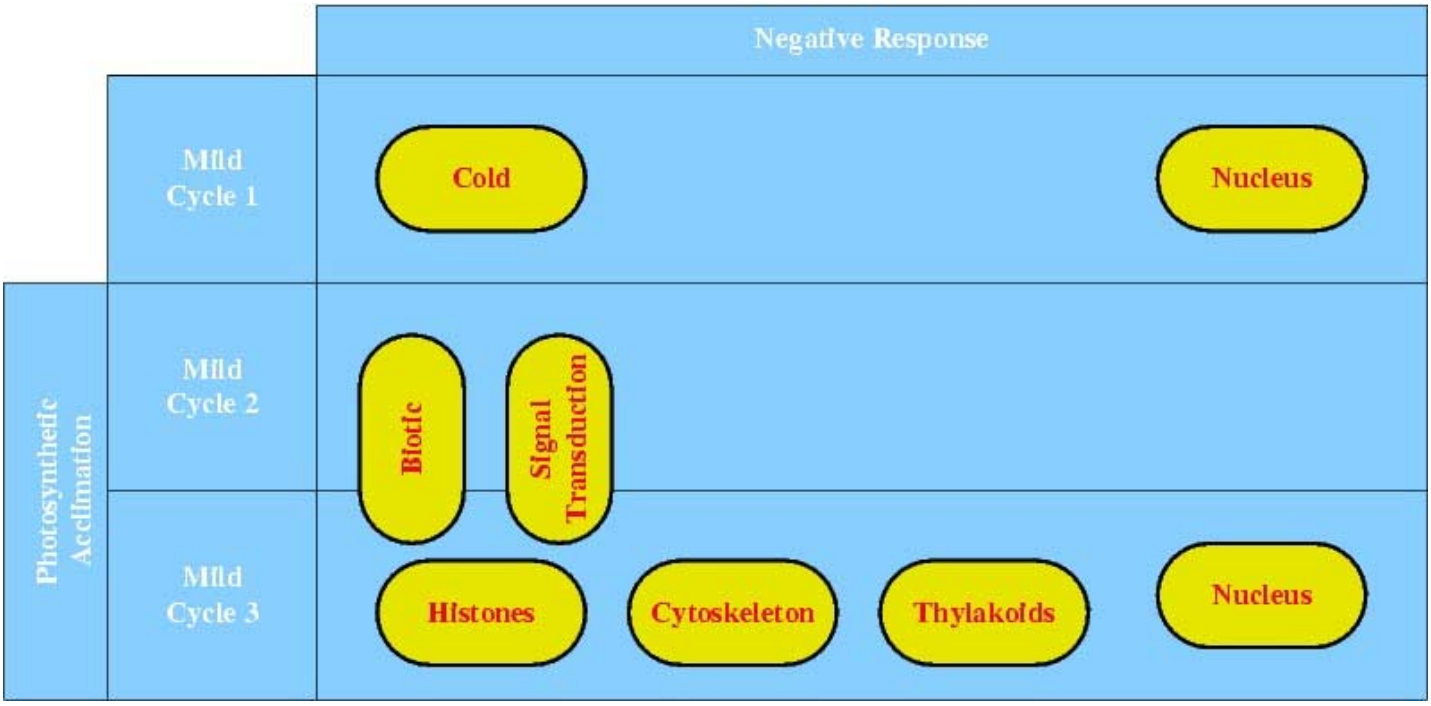
A.

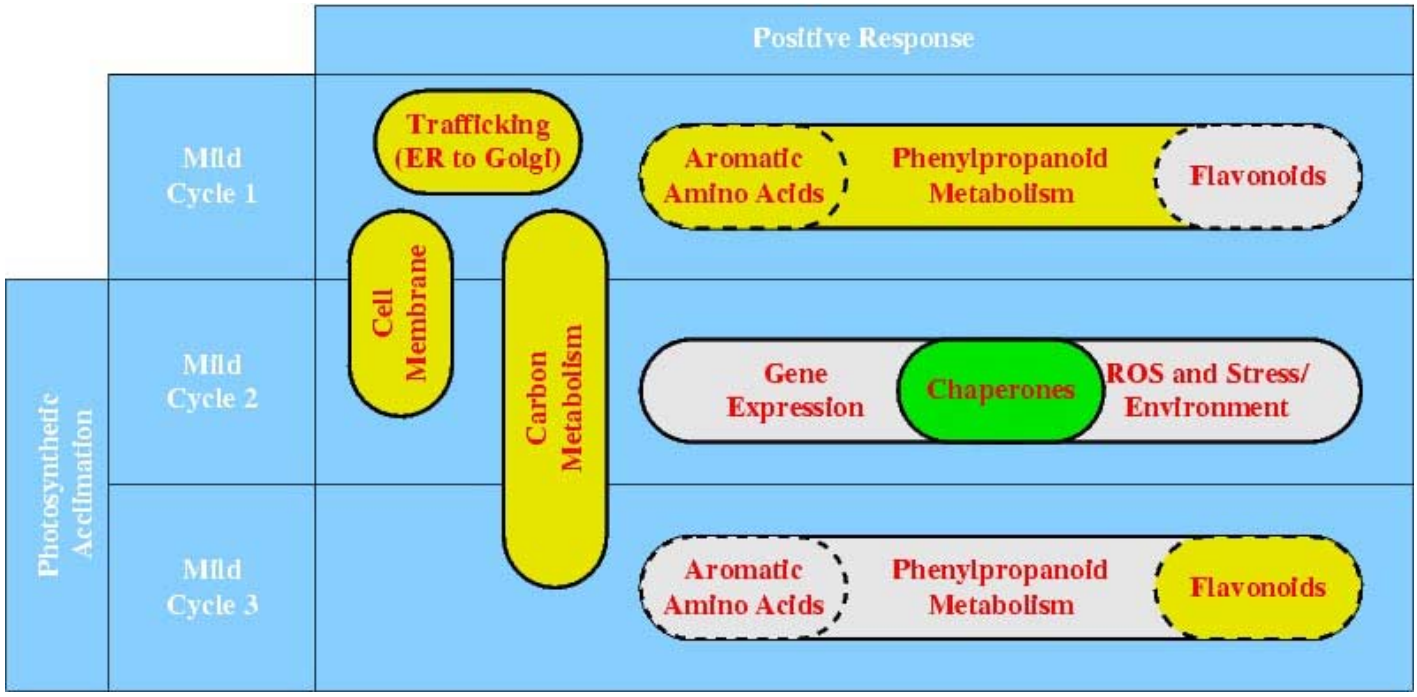

Cycle 1
127
6
139
Mild
Severe
Cycle 2
188
25
137
Mild
Severe
Cycle 3
56
6
129
Mild
Severe

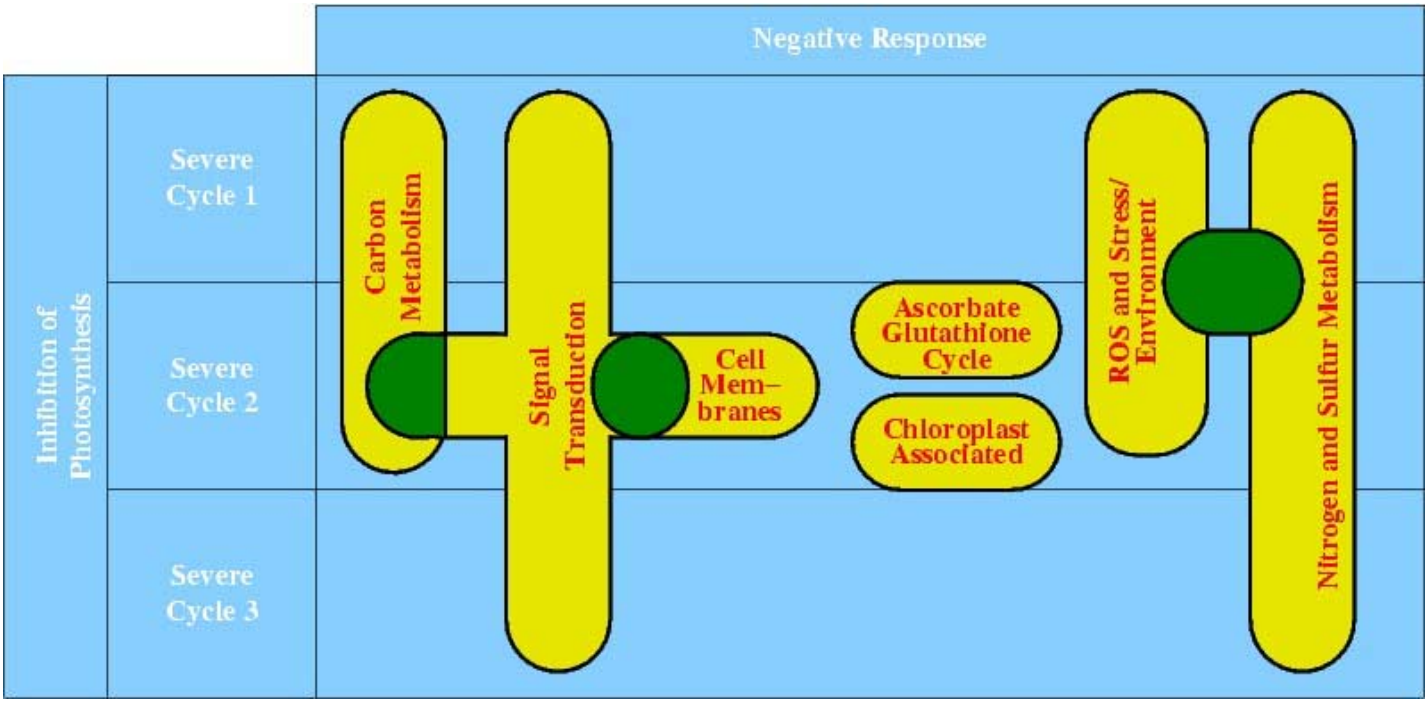
B.

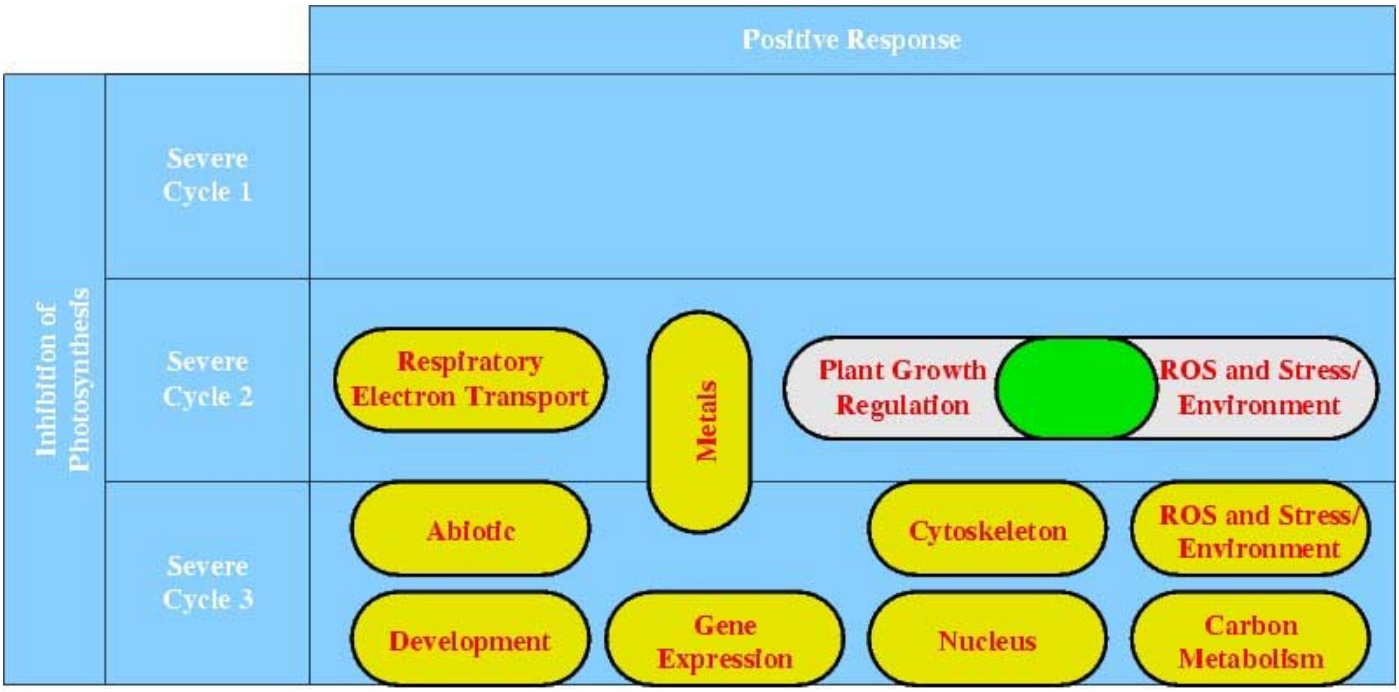

Cycle 1
82
12
132
Mild
Severe
Cycle 2
146
13
143
Mild
Severe
Cycle 3
89
1
52
Mild
Severe


ENVIRONMENTAL CHANGE
Signal Perception and Transduction
Transcription Factors
Calcium–associated
Kinases
Redox
Gene Expression
Transcription
Post–transcriptional Processing
Translation
Post–translational Processing
Chaperones
Thiols
Protective Processes
ROS and Stress/Environment
Abiotic (Drought; Heat; Salt; Cold; Protection and Repair [Chaperones])
Antioxidant Processes
Biotic
Cell Wall Related
Phenylpropanoid Pathway
Protected Processes
Development
Cells
Protease–associated
Nucleus
Trafficking
Cell Membranes
Secretion
Cell Structure and Differentiation
Tissues
Plant Growth Regulation
Metabolism
Chloroplast–associated
Carbon Metabolism
Nitrogen and Sulfur Metabolism
Mitochondrion
Peroxisome/Glyoxysome
Metals


Negative Response
Photosynthetic Acclimation
Mild Cycle 1
Mild Cycle 2
Mild Cycle 3
Cold
Nucleus
Biotic
Signal Transduction
Histones
Cytoskeleton
Thylakoids
Nucleus


Positive Response
Photosynthetic Acclimation
Mild Cycle 1
Mild Cycle 2
Mild Cycle 3
Trafficking (ER to Golgi)
Cell Membrane
Carbon Metabolism
Aromatic Amino Acids
Phenylpropanoid Metabolism
Flavonoids
Gene Expression
Chaperones
ROS and Stress/ Environment
Aromatic Amino Acids
Phenylpropanoid Metabolism
Flavonoids

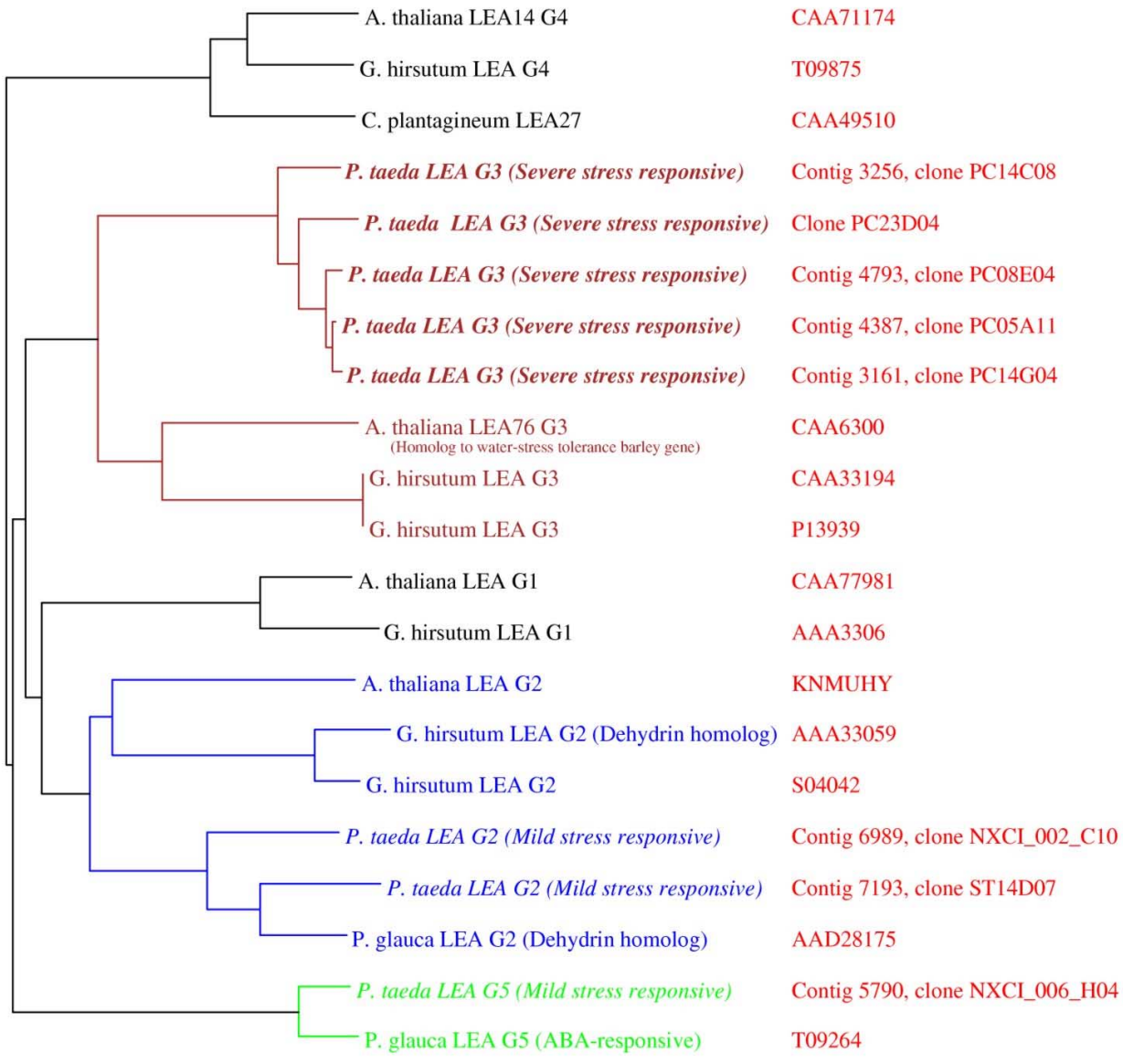

Negative Response
Inhibition of Photosynthesis
Severe Cycle 1
Severe Cycle 2
Severe Cycle 3
Carbon Metabolism
Signal Transduction
Cell Mem–branes
Ascorbate Glutathione Cycle
Chloroplast Associated
ROS and Stress/ Environment
Nitrogen and Sulfur Metabolism


Positive Response
Inhibition of Photosynthesis
Severe Cycle 1
Severe Cycle 2
Severe Cycle 3
Respiratory Electron Transport
Metals
Plant Growth Regulation
ROS and Stress/ Environment
Abiotic
Development
Gene Expression
Cytoskeleton
Nucleus
ROS and Stress/ Environment
Carbon Metabolism

Clone ID	Annotation	Class	Mild			Severe		
			1	2	3	1	2	3
14D07	Putative Dehydrin	LEA Group 2	0	+	0	0	0	0
NXCI_002_C10	Putative Dehydrin	LEA Group 2	0	0	+	0	0	0
NXCI_006_H04	Embyonic abunndant protein (white spruce)	LEA Group ?	0	0	+	0	0	0
PC23D04	LEA76 homolog	LEA Group 3	0	0	0	0	+	0
ST01E01	Putative LEA	LEA Group 3	0	0	0	0	+	+
PC14C08	LEA76 homolog	LEA Group 3	0	0	0	0	0	+
PC14G04	LEA76 homolog	LEA Group 3	-	0	-	+	+	+
PC05A11	LEA76 homolog	LEA Group 3	0	0	-	+	+	0
PC08E04	LEA76 homolog	LEA Group 3	0	0	-	+	0	+
39A03	DNA J-like protein	DNA J	0	+	0	-	0	0
40F04	Low molecular weight heat shock protein	sHSP	+	0	-	0	0	0
NXNV_132_E06	DNA J homolog	DNA J	+	0	0	0	0	0
NXSI_116_B04	Heat Shock Protein 82	HSP 90	0	+	0	0	0	0
NXSI_117_C08	DNA K type molecular chaperone hsc 70	HSP 70	0	+	0	-	0	0
NXNV_149_E10	Putative Heat Shock Protein 90	HSP 90	0	0	0	0	+	0
NXCI_022_G01	Heat Shock 70 Kda protein, mitochonrial	HSP 70	0	0	0	+	+	0
NXNV_	Heat Shock Protein, 82K, precursor	HSP 90	0	0	0	+	0	0
13C12	Putative peptidyl-prolyl cis-trans isomerase	Ppiase	0	0	0	+	0	0
38H03	Peptidyl proline isomerase	Ppiase	0	+	0	0	0	0
39H08	Peptidyl prolyl cis-trans isomerase	Ppiase	0	+	0	0	0	0

	Clone ID	Annotation	Mild			Severe		
			1	2	3	1	2	3
Glycolysis	NXCI_007_H12	Sucrose synthase					-	
	NXCI_032_F09	Sucrose synthase					+	
	NXCI_106_C10	Sucrose synthase		-		-	-	
	NXSI_116_F02	Hexokinase		+				
	NXNV_079_G08	Fructokinase		-				
	NXCI_157_B10	Fructokinase				-		
	NXSI_021_D06	Glucose-6-P isomerase		-		-		
	NXCI_034_B04	PPi dep. Phosphofructokinase						
	NXNV_144_H09	Aldolase		-				
	07 E05	Aldolase		+				
	NXNV_124_C02	Triose-P isomerase		-		-		
	NXSI_034_D06	Triose-P isomerase	+		-			
	NXSI_064_G04	Glyceraldehyde-3-P Dehydrog						
	NXSI_134_C01	Glyceraldehyde-3-P Dehydrog						
	NXSI_031_H06	Phosphoglycerate Kinase						
		Phosphoglycerate mutase						
	NXCI_122_A09	Enolase		+				
	NXSI_143_H06	Pyruvate kinase						+
	NXCI_126_D02	Pyruvate kinase	+	+	+			
	NXNV_066_D03	Pyruvate decarboxylase						
	20 A11	Alcohol dehydrogenase	+	+				
	NXSI_100_H03	Alcohol dehydrogenase					-	
TCA Cycle	NXNV_074_H11	Pyruvate dehydrogenase	+		+			
	NXCI_150_E08	Pyruvate dehydrogenase			+		-	
	NXCI_094_G11	Pyruvate dehydrogenase	+					
	29 A09	Citrate synthase						+
		Aconitate hydratase						
		Isocitrate dehydrogenase						
	NXSI_066_A02	2 oxoglutarate dehydrogenase						
	NXSI_039_A11	Succinly CoA synthase						
		Succinate dehydrogenase						
	NXCI_106_D10	Fumarase						
	NXSI_048_D06	Malate dehydrogenase					-	
OPPP		Glucose-6-P dehydrogenase						
		Phosphoglucono lactonase						
	NXCI_018_D09	Phosphogluconate dehydrog						
	NXCI_153_D09	Ribose-5-P isomerase		-				
	NXNV_075_A12	Ribose-5-P isomerase			+		-	-
	NXCI_146_H08	Transketolase			+		-	
	NXSI_145_D04	Transaldolase	+	+				+
RPPP	39 F01	RUBISCO					+	+
	NXCI_115_A02	Phosphoglycerate Kinase					-	
	NXSI_134_C01	Glyceraldehyde-3-P Dehydrog						
	NXCI_144_H09	Glyceraldehyde-3-P Dehydrog			+			+
	NXCI_071_F03	Glyceraldehyde-3-P Dehydrog						+
		Triose-P isomerase						
	NXNV_144_H09	Aldolase		-				
	07 E05	Aldolase		+				
		Fructose 1,6 bisphosphatase						
	NXCI_146_H08	Transketolase			+		-	
	NXSI_145_D04	Transaldolase	+	+				+
		Sedohept 1,7 bisphosphatase						
	NXCI_153_D09	Ribose-5-P isomerase		-				
	NXNV_075_A12	Ribose-5-P isomerase			+		-	-
		Ribose-5-P epimerase						
	18 H10	Phosphoribulokinase		+				

	Clone ID	Annotation	Mild 1	Mild 2	Mild 3	Severe 1	Severe 2	Severe 3
Aromatic Amino Acid	NXCI_047_C05	DAHP synthase	+			-	-	
	NXCI_071_C01	3-dehydroquinate synthase	+		+			-
	NXCI_117_D08	3-dehydroquinate dehydratase						
	NXNV_185_H02	Shikimate dehydrogenase						
	NXCI_034_B01	Shikimate kinase						
		EPSP Synthase						
	NXCI_163_G07	Chorismate synthase	+				-	+
	NXSI_051_F10	Chorismate synthase						
	NXCI_016_F11	Chorismate mutase		+				
		Prephenate aminotransferase						
		Arogenate dehydratase						
		Arogenate dehydrogenase						
Phenyl-propanoid	NXCI_093_H05	PAL					-	
	NXSI_118_A03	Cinnimate 4 hydroxylase						
	NXCI_087_F07	Cinnimate 4 hydroxylase						
	NXCI_045_B07	Cinnimate 4 hydroxylase	+		+			
Lignin	12 E05	Caffeoyl O methyl transferase						+
	NXSI_055_H08	Caffeoyl O methyl transferase						
	NXSI_130_F05	Caffeoyl O methyl transferase						
	02 B03	Cinnamyl alcohol dehydrogenase	-					
	NXNV_162_F07	Cinnamyl alcohol dehydrogenase			-			
	NXCI_165_H04	Cinnamoyl CoA reductase				-		
	34 F04	Cinnamoyl CoA reductase						
	NXNV_044_G05	Laccase						-
	NXSI_127_C02	Laccase	+			-		
	NXNV_136_F10	Laccase		-		+	+	
	NXCI_005_C10	Laccase	-					
	NXCI_018_F10	Pinoresinol reductase						
Flavonoids		Chalcone synthase						
	NXCI_098_F10	Chalcone/Flavone isomerase	+		+			
	07 H08	Chalcone/Flavone isomerase	+		+			+
	NXNV_127_E04	Isoflavone reductase						
	NXNV_127_F01	Isoflavone reductase						
	NXCI_002_E07	Isoflavone reductase		+	+	-		
	NXSI_063_D01	Naringenin-2-oxo dioxygenase	+		+	+		
	28 B11	Naringenin-2-oxo dioxygenase		+				+
	13 H06	Leucoanthocyanidin reductase	+					


A. thaliana LEA14 G4
CAA71174
G. hirsutum LEA G4
T09875
C. plantagineum LEA27
CAA49510
P. taeda LEA G3 (Severe stress responsive)
Contig 3256, clone PC14C08
P. taeda LEA G3 (Severe stress responsive)
Clone PC23D04
P. taeda LEA G3 (Severe stress responsive)
Contig 4793, clone PC08E04
P. taeda LEA G3 (Severe stress responsive)
Contig 4387, clone PC05A11
P. taeda LEA G3 (Severe stress responsive)
Contig 3161, clone PC14G04
A. thaliana LEA76 G3
(Homolog to water-stress tolerance barley gene)
CAA6300
G. hirsutum LEA G3
CAA33194
G. hirsutum LEA G3
P13939
A. thaliana LEA G1
CAA77981
G. hirsutum LEA G1
AAA3306
A. thaliana LEA G2
KNMUHY
G. hirsutum LEA G2 (Dehydrin homolog)
AAA33059
G. hirsutum LEA G2
S04042
P. taeda LEA G2 (Mild stress responsive)
Contig 6989, clone NXCI_002_C10
P. taeda LEA G2 (Mild stress responsive)
Contig 7193, clone ST14D07
P. glauca LEA G2 (Dehydrin homolog)
AAD28175
P. taeda LEA G5 (Mild stress responsive)
Contig 5790, clone NXCI_006_H04
P. glauca LEA G5 (ABA-responsive)
T09264