8-2009

GENOMIC, PROTEOMIC AND METABOLOMIC APPROACHES TO STUDY DROUGHT RESPONSES IN AQUILEGIA

David Henry
Clemson University, dhenry@clemson.edu

GENOMIC, PROTEOMIC AND METABOLOMIC APPROACHES TO STUDY DROUGHT RESPONSES IN *AQUILEGIA*

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
PLANT AND ENVIRONMENTAL SCIENCES

---

by
DAVID CHAKERIS HENRY
August 2009

Accepted by:
EMERSON R. SHIPE, Committee Chair
HONG LUO
MICHAEL G. SEHORN
HALINA T. KNAP

ABSTRACT

Global population is expected to increase 30% by 2040, which will result in an increased need for crop production to feed the growing population. Combined with projected increased drought conditions worldwide, plant genetic research is necessary to gain a deeper knowledge of the molecular factors involved in plant drought response in order to engineer crop species with improved drought tolerance. *Aquilegia* has been recently developed as a model species for gene exploration based on its ability to thrive in a wide variety of environments including arid locations. An attractive asset of *Aquilegia* is its evolutionary position, equidistant between rice and *Arabidopsis*. Multifaceted molecular biology techniques were employed in these studies to identify biological components associated with *Aquilegia*'s response to drought. Techniques utilized were; suppression subtractive hybridization, using drought-stressed and un-stressed tissue-derived mRNAs to selectively amplify differentially expressed genes, metabalomic profiling as a means to examine the identity and possible function of accumulated metabolites, and exploration of proteomes obtained under different levels of drought stress. Subtractive hybridization yielded numerous sequences such as DREB known to be involved with drought and other genes with no known association to drought response. The metabalomics approach yielded vitexin and vitexin 2'-O-β-D-glucoside, identified via LCMS and HPLC. Vitexin concentration within *Aquilegia* leaf tissue increased 357% over the time course of the drought experiment. Previously, vitexin had not been associated with drought response. Proteome analysis identified various proteins known to play some defensive role in drought stress. This parallel initiative yielded

several possible gene candidates for genetic engineering of crop species. Results from this project confirm *Aquilegia*'s role as an excellent genetic resource for studying responses to a wide range of abiotic stresses, specifically drought.

DEDICATION

This manuscript and the countless hours of work are a testament to the love and support of my family, my friends and my beautiful wife.

ACKNOWLEDGMENTS

I would like to acknowledge and sincerely thank all members of my graduate committee, Dr. Emerson Shipe, my Committee Chair, Dr. Hong Luo, Dr. Halina Knap, and Dr. Michael Sehorn. Without their guidance and motivation, this dissertation would not have been possible. I would also like to acknowledge members of the Clemson University Genomics Institute for all their help; Michael Atkins, Dr. Eric Fang, Lynn Illeck, Emily Jones, Dr. Christopher Saski, Dr. Margaret E. Staton, Dr. Jeff Tomkins, and Jeanice Troutman, Dr. Yonnie Wu.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

List of Figures (continued).....Page

CHAPTER ONE

INTRODUCTION

Drought Effect on Plants

Since plants remain in one location during their entire existence, they have developed numerous ways to deal with environmental stresses such as drought, salt, cold, heat, and extreme temperature change. Drought occurs throughout the world on an annual basis. High temperature combined with water deficits affects crop production (Boyer, 1982). Crop loss due to drought stress most likely exceed losses from all other causes combined (Chaves et al., 2003). More research into how plants respond to drought conditions is needed and will become more important in the future based on climate change predictions of an increase in arid areas (Petit et al., 1999). Understanding plant responses to drought is of great importance in order to make plants more stress tolerant (Reddy et al., 2004). Advances in the understanding of these processes may lead to genetically modified drought tolerant crop plants.

Plants experience drought stress when water supply to roots is decreased and transpiration rate is higher than water uptake. Decreased water conditions lead to many adverse effects on plants such as disruption of water potential gradients, loss of turgor, denaturation of proteins, disruption of membranes, and other detrimental effects (Kozlowski and Pallardy, 2002). Some of these effects may take place in just a few seconds, such as a change in phosphorylation of proteins, while other responses occur over days such as the gene expression profile of a plant. The severity of plant response to drought depends on many factors, such as plant species, the length of time exposed to

decreased water conditions and soil conditions. Some species of plants are able to maintain growth better in drought conditions; these species can be studied to help identify genes and pathways that may be introduced to agricultural crops to increase their resistance to drought. Genetically engineering drought tolerant crop plants has huge economical prospects. In order to successfully modify plants, there is much to be learned about the molecular mechanisms involved in drought tolerance (Shinozaki, 2006).

Arabidopsis has been used as a model species for the dissection of stress responsive pathways in plants. Numerous genes that are responsible for defense against stress have been discovered in *Arabidopsis*. The monocot *Oryza sativa* (rice) is another model species that has particularly large economic impacts worldwide. These two species have been studied comparatively to determine mutual stress responsive pathways. Another newer model species being discussed here is *Aquilegia. Aquilegia* (Columbine) is positioned equal distance between the model species rice and *Arabidopsis* and is a member of the basel most edudicot clade *(Ranunculales).* Most *Aquilegia* are interfertile between species. This asset can be used to determine variations responsible for environmental adaptations and dissect genetic diversity among the clade. Information gained from this study may be applied to other model species and prospective gene candidates transformed into other crop species.

Aquilegia is currently being developed as a new model species for general study of plant adaptation to the environment because the genus has: 1) a wide range and variability in its environment, 2) most *Aquilegia* are interfertile between species, 3) fast evolutionary changes, yielding low sequence variability, and 4) diploid (n=7)

chromosome number with a small genome (1C= 320-400 Mbp) (Hodges and Kramer 2007).

Aquilegia has a natural range from Alaska to Florida (USDA, 2008) and can be found from an elevation of 10,000 feet to sea level. The wide geographic distribution ultimately shows the broad adaptability of *Aquilegia*. *Aquilegia* is found in shaded forest, alpine zones, desert spring areas and rocky outcrops (Munz, 1946). Because of the wide geographic range of Columbine and the extreme environments it thrives in, it is an excellent candidate for gene exploration for the possible transfer of genetic information into crop plants for enhanced resistance to drought.

Plant Defenses to Drought Stress

Plants have developed numerous defenses to deal with drought, which are categorized by three basic mechanisms; drought escape, drought avoidance and drought tolerance (Levitt, 1972). Some plants have evolved to escape drought conditions by completing their life cycle and reproducing while there is still moisture present in the soil, before drought conditions are present (Mooney et al., 1987). Avoiding tissue dehydration due to drought is a challenge met in many different ways, with the ultimate goals being minimization of water loss and maximum water uptake. Some avoidance mechanisms are closing stomata, reducing the surface area of leaves to sunlight exposure (Ehleringer, 1992), and increasing trichome level (Larcher, 2000). Some plants attempt to avoid dehydration by various physical mechanisms such as regulating their transpiration rate. Some decrease their leaf surface area, leaf number, or leaf longevity

which in turn reduces the photosynthesis rate, resulting in a reduction in crop yield. When escape and avoidance to drought are impossible, a plant must become accustomed to reduced water availability and adapt to tolerate the low water availability within the environment. This is done in several ways such as osmotic adjustment (Morgan, 1984), acumulation of osmolytes (Williams and Leopold, 1989), and stomata regulation, (Jones, 1973). The specific response to drought is dependent on the type of plant and conditions present in the environment.

Stomata Regulation

It has long been debated how photosynthesis is down-regulated due to drought and whether it is due to stomata closure or metabolic blockage (Jones, 1973). In an effort to reduce the transpiration rate when exposed to arid conditions, some plants have evolved ways to regulate stomata size. Stomata closure reduces transpiration rate and leads to a lowered level of carbon uptake, resulting in lower level of photosynthesis in leaves (Chaves, 1991). A reduction in the number of leaves reduces water loss due to transpiration and can also be viewed as a way to redistribute nutrients in older leaves and stems before they are shed.

The cascade of events that leads to stomata closure has long been sought. It is known that stomata often close in response to drought before any change in water content in the leaf is detected (Socias et al., 1997). A major player in the signal to close the stomata is abscisic acid (ABA). Abscisic acid is produced in the root where, most likely, the first responses to drought occur (Davies and Zhang, 1987). A direct relationship in

the concentration of ABA in the transport tissue xylem due to drought conditions has been defined only in a few instances (Correia et al., 1995). ABA is not the only factor that induces stomata closure. Other factors include; leaf water potential (Tardieu and Davies, 1992), xylem sap pH (Davies and Wilkinson, 1997), and leaf-to-air vapor pressure deficit regardless of the available water present in the soil (Oren et al., 1999).

The regulation of stomata in relationship to drought, decreased leaf water potential, relative water content, available water, ABA, in addition to other factors, varies widely between different species and genotypes. Due to this fact, some species are better adapted to drought conditions and should be studied to further improve knowledge of the signal cascade leading to drought resistance.

The Role of Abscisic Acid

The hormone ABA helps regulate various stress responses in plants to drought, cold and salinity (Leung and Giraudat, 1998; Davies and Wilkinson, 1997). During drought conditions ABA alters ion transport of guard cells, this in turn promotes stomata closure and prevents stomata opening, allowing the plant to slow the transpiration process. Abscisic acid-activated protein kinase (AAPK) is detected in guard cells but not in roots, leaf epidermal cells, mesophyll cells, flowers, or seeds (Assmann et al., 2000). The AAPK protein is activated by ABA although not by darkness or elevated carbon dioxide level, the conditions that normally trigger stomata closure (Assmann et al., 2000). The increased level of ABA in water stressed plants has been found to stimulate the release of potassium by stomata guard cells, this in turn leads to stomata closure (Hsiao et

al., 1985). It has been noted that moderate starvation of potassium inhibits water stress induced stomata closure and resistance to low water conditions. This may be one of the main causes of plant dehydration in plants growing in low potassium soil (Benlloch-Gonzalez et al., 2007).

Osmolytes

The morphological effects of water deficits are fairly well known and can be easily visualized in a plant, though the causes of these symptoms are not well known on a biochemical and molecular level (Bruce et al., 2002). One of the first responses to water deficit is osmotic adjustment (Leung and Giraudat, 1998). Plants increase their osmotic potential by accumulating osmolytes such as proline, glycine betaine and non-reducing sugars and polyols. Osmolyte accumulation allows additional water to be taken up from the environment buffering the immediate effect of decreased water within the plant. Osmolytes do not interfere with normal cellular biochemical reactions (Williams and Leopold, 1989, Jaleel et al., 2007). Osmolytes increase thermodynamic stability of folded proteins and provide protection against denaturing stresses (Yancey, 2001).

The accumulation of soluble sugars such as sucrose, glucose and fructose is often associated with drought tolerance in plants. It is thought that soluble sugars help protect the plant in times of drought in one of two ways: first, the hydroxyl group of sugars may act as a water molecule to maintain hydrophilic interactions in membranes and proteins by hydrogen bonding with them (Williams and Leopold, 1989). Secondly, these sugars form a highly viscous layer in the cytoplasm of the dehydrated cells reducing the

molecular movement and helping to maintain the structural stability of the macromolecules (Sun and Leopold, 1997). Several genes involved in metabolism of these sugars have been found to increase the ability of different transgenic plants to resist stress (Shinozaki, 2007). Over-expression of drought-inducible galactinol synthetase gene (*AtGolS2)* in *Arabidopsis* yielded increased drought resistance because of the excess accumulation of galactinol and raffinose (Shinozaki, 2007). Also, transgenic plants that express the DREB1A/C- repeat binding factor 3 (CFB3) accumulate more galactinol and raffinose than wild type plants, helping them improve their drought tolerance (Avonce et al., 2004).

In addition to the sugars mentioned above, trehalose is often accumulated in large quantities in response to desiccation. Trehalose is an alpha-linked disaccharide found in nature from a wide variety of sources such as bacteria, fungi, yeast, plants, and invertebrate animals (Elbein, 1974). Trehalose is a factor in tolerance to environmental conditions such as drought, heat, and salt stress (Thevelein, 1996). The high water retention capacity of trehalose helps stabilize proteins and membranes enabling the cells to recover when rehydration conditions are present. Tobacco has been transformed with the trehalose phosphorylase (TP) gene from *Pleurotus sajor-caju* and showed an increased ability to withstand drought. The trehalose synthetase gene *(Tsase)* from *Grifola frondosa* also yielded an increase in the ability to withstand drought in tobacco (Han et al., 2005)

Proline Accumulation

The amino acid proline is thought to play a role in a plant's ability to withstand drought stress by helping regulate osmotic adjustment, stabilizing subcellular structures, scavenging free radicals, and acting as a stress signal. In *Arabidopsis*, over-expression of proline synthetase gene produced plants with an elevated resistance to drought, while suppression of this gene results in plants with a decreased ability to withstand drought (Bartels and Sunkar, 2005).

A strong correlation between the accumulation of proline and drought tolerance has been demonstrated by over-expressing the Δ-pyrroline-5-carboxylate synthase gene (*P5CS*) in *Arabidopsis*. Deficiency of proline from antisense suppression of the *P5CS* gene resulted in decreased drought tolerance in *Arabidopsis.* Gene knockout of proline dehydrogenase (*ProDH*) led to growth inhibition, indicating that excessive proline is detrimental to plant growth (Borsani et al., 2005; Chinnusamy, et al., 2005)

Genes Involved in Drought Resistance

As plants are subjected to increased levels of desiccation, numerous genes are up-regulated in response to surrounding environmental conditions. For example, microarray analyses of *Arabidopsis* using 1300 different cDNAs have revealed 44 different genes that were up-regulated (Seki et al., 2001). Not every gene that is up-regulated is in direct response to drought stress. Some genes that are up-regulated are in response to damage caused by drought induced stress. In general, responsive genes can be grouped into two categories: 1) genes that are turned on immediately, and 2) others that are turned on

within days or weeks of dehydration. It may be that the genes immediately turned on provide initial protection to drought while the genes that are turned on later may be involved in adaptation to the stress (Shinosaki and Shinosaki-Yamaguchi, 2007).

Several attempts to genetically engineer drought tolerant plants have been successful. One such example is the transfer of functional genes that encode for enzymes associated with the synthesis of osmotically active compounds such as transporters, chaperones and reactive oxygen species (Shinozaki, 2006). One recent finding yielded a novel synthetic pathway for glycine betaine. Glycine betaine is an osmoprotectant found in halophilic microorganisms (Reinkkainene, 2000). Transgenic *Arabidopsis*, expressing the betaine synthesis genes, showed a higher level of glycine-betaine than the choline-oxidizing enzyme from the conventional pathway (Takabe, 2005). These plants showed significant dehydration tolerance, possibly due to a protective function of betaine molecules which could stabilize proteins and cellular structures and/or scavenging of free radicals (Chen and Murata, 2002).

It is well known that the plant hormone abscisic acid (ABA) plays a major role in drought tolerance (Elbein, 1974; Shinosaki and Shinosaki-Yamaguchi, 2007). Recent findings have yielded a cytochrome P450 CYP707A family of 8'-hydrolases that play a role in ABA levels in seed imbibitions and dehydration stress conditions (Saito et al., 2004; Shinozaki, 2006). An insertional mutant of CYP707A3 exhibited drought tolerance in conjunction with a reduction in the transpiration rate in *Arabidopsis* (Kushiro et al., 2004).

Photosynthesis Regulation

Photosystems I and II obtain high-energy by capturing electrons as sunlight is absorbed by chlorophyll molecules and the electrons become excited. The two photosystems are linked and transfer electrons from water to NADP+ to form NADPH, generating O_2 as a byproduct. The H+ is pumped into the thylakoid space and the backflow of H+ occurs through an ATP synthase producing ATP in the chloroplast stroma. These products of photosynthesis, ATP and NADPH, are used in all biological pathways. The carbon fixation pathway is one of these; carbon dioxide enters the pathway and combines with ribulose 1, 5 biophosphate. With the help of the catalytic action of rubisco it is converted to triose phosphates which are exported to the cell cytosol to provide a carbon source for cellular reactions (Alberts et al., 1994).

Water is an essential part of life for a plant because plant metabolism requires water to form energy. Drought stress is one of the most important environmental factors inhibiting photosynthesis (Bradford and Hsiao, 1982). If a plant cannot carry out photosynthesis, it cannot produce ATP which is needed for virtually all plant processes. This leads to one theory of how decreased water availability chemically triggers the slowdown in carbon assumption (CO_2) from the environment.

Late Embryogenesis Proteins

Late Embryogenesis Abundant Proteins (LEAs) are known to accumulate during seed maturation, and are also expressed when a plant is exposed to water-stress due to drought or freezing. These LEA proteins vary over a large molecular weight range from

9-200 kD; they are thermostable and contain a high proportion of glycine and lysine residues. LEAs are found in a broad range of vascular plants, mosses, ferns, lichens, and algae. Because LEAs do not catalyze any metabolic reactions, their molecular function is somewhat unclear. There are different groups of LEAs based on chemical properties and motif similarity. Group 1 LEAs contain a 20 amino acid motif. Group 2 LEA proteins are referred to as dehydrins. They are characterized by a 15 amino acid motif that forms an amplipathic α-helix, a tract of continuous serine residues and a conserved sequence, DEYGNP at the N-terminus (Cuming, 1999). It is thought that due to the amphipathic properties of these LEAs they act as membrane stabilizers. These dehydrins contain α-helices which form lipid binding domain that can help stabilize the hydrophobic parts of proteins. The random coil structure yields an exceptional water binding capacity (Campbell and Close, 1997). Group 3 LEA proteins share an 11 amino acid stretch that forms an α-helix. Group 4 LEAs contain a conserved N-terminus that is predicted to form an α-helix and a diverse C-terminal with a random coil structure. Group 5 LEA proteins contain more hydrophobic residues than the other groups, are not soluble after boiling and are likely to adopt a globular structure (Alberts et al., 1994).

The dehydrin gene from *Cicer pinnatifidum* (chickpea), DHN1, can be induced by drought, salt stress, ABA, and expressed during seed development. Over-expression of DHN1 in tobacco plants positively affected the growth of seedlings under artificial water stress. Mature plants, transformed with *DHN1* had larger root mass and were able to recover more readily after re-irrigation than the controls (Beck et al., 2007).

Aquilegia as a Model Species for Adaptation

Aquilegia's ability to thrive in a wide variety of biotic and abiotic environments has allowed it to become a model species in studies dealing with environmental adaptation (Hodges, Kramer, 2007). The fact that speciation has occurred rapidly allows resources created in one species to be applicable in other species because of low sequence variation, this combined with *Aquilegia* relatively small genome of 320-400 Mbp, allows for efficient use of current resources available to *Aquilegia* such as three Bacterial Artificial Chromosome (BAC) Libraries, a genetic map, 90,000 cDNA clones and a physical map. Because *Aquilegia* is positioned equidistant between *Arabidopsis* and Rice, information obtained from *Aquilegia* may be useful in studies of *Arabidopsis* and Rice.

The complexity of drought responses in plants can be explored on many levels such as identification of; genes transcribed, proteins accumulated, and osmolytes accumulated in response to drought. In order to fully understand what is occurring in a drought stressed plant, a multifaceted approach increases the likelihood of learning important information valuable and applicable to other plant species. Here I propose utilization of a genomic, proteomic and metabolomics approach to identify genes, metabolites, and proteins that play an important role in *Aquilegia*'s ability to withstand drought conditions.

CHAPTER TWO

IDENTIFICATION OF DROUGHT RESPONSIVE GENES FROM AQUILEGIA THROUGH SUPPRESSION SUBTRACTIVE HYBRIDIZATION

Abstract

In an effort to ease data analysis and reduce constitutively expressed gene transcripts, a suppression subtractive hybridization library was constructed from both drought stressed *Aquilegia* tissue, termed "tester", and unstressed tissue termed "driver". PolyA mRNA was obtained from stressed and unstressed *Aquilegia* tissue, cDNA was created, and competitive hybridization was utilized followed by adaptor specific PCR amplification to increase number of differential expressed transcripts. These were cloned and sequenced. One hundred and seventy three transcripts from the stressed and one hundred and eight from the control library were identified and classified into various functional categories. One hundred and thirteen sequences were found only in the drought stressed library. Many of these genes show homology to known gene sequences. Just as importantly, 24% of the sequences showed no homology to any known gene sequence in searched databases. The unknown sequences are possible novel gene sequences involved in *Aquilegia'*s ability to withstand drought conditions. Our goal is to identify genes and pathways used in a drought response. Knowledge acquired here will not be limited solely to drought stress responses, but can be applied to other abiotic stress responses that share mutual pathways.

Introduction

Aquilegia (Columbine) has a natural range from Alaska to Florida (USDA, 2008) and can be found from an elevation of 10,000 feet to sea level. Columbines are found in shaded forest, alpine zones, desert spring areas and rocky outcrops (Munz, 1946). *Aquilegia*s ability to thrive in many different environments makes it an ideal candidate for novel gene exploration, with plant transformation of crop species the ultimate goal. Understanding the mechanisms that facilitate regulation of genes is an important issue for genetic improvement of various crop species via plant transformation.

A method to selectively identify genes that are present in one sample but not another is the use of subtractive suppression hybridization (SSH) (Gurskaya et al., 1996). This techniques allows for amplification of differentially expressed transcripts, without the redundant sequencing of the abundant transcripts. There are numerous sequences present from *Aquilegia* at Harvards Gene Index Project (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=aquilegia). These sequences originated from mixed tissue of shoot, floral apical meristems, flower buds, leaves and roots. The advantage of an SSH library and data presented in the manuscript is the fact that the tissue was subjected to moderate drought conditions and the common transcripts were “subtracted” out leaving primarily only genes that are present in drought stressed tissue.

Materials and Methods

Plant Growth

Aquilegia F1, Origami Mix cultivar seeds were obtained from Harris Seed Company in Rochester, NY. The seeds were rinsed and cleaned with distilled water and then allowed to germination in a Petri dish containing moist filter paper at room temperature. After germination, seeds were placed in moist germination soil in a covered growth flat under constant light. Once the seeds had developed a root system, they were moved to larger 18 section flats containing more growth area. Flats were then placed in a growth room for 6 months to increase tissue mass. The temperature was maintained at an average temperature of 20.5°C with 12 hours of artificial light. Plants were watered to saturation every 2 days for approximately 6 months.

Plants were grown to a height of 10 cm. The plants were then subjected to drought conditions. This was performed by watering plants on Day 0. On Day 1 plant tissue from root, stem and leaf was collected and frozen in liquid nitrogen, to be used for the "control or driver RNA". Tissue samples were collected every 12 hours over a 6 day period with no water applied to plants. Samples were taken at 8:00 am and 12 hours later at 8:00 pm. Each day plant tissue was collected as mentioned above and frozen at -80°C until total RNA could be extracted. The morphogical changes to the leaves and stems appeared late on Day three. Stressed plants showed slight signs of wilting of the stems and leaves. The earliest morphogical change that hinted the *Aquilegia* plants were undergoing drought was our point of interest. The early levels of drought were of interest so that identification of transcription factor genes can be examined.

RNA Extraction and PolyA Enrichment

All instruments and supplies were cleaned appropriately to assure there was no RNAase present. When all tissue samples were obtained, they were removed from -80°C and total RNA extraction was performed. The RNA extraction was based on the method by Meisel et al., 2005. Extracted total RNA was checked for quality by gel electrophoresis (1.2% agarose, 80V for 30 minutes). Only high quality RNA with a 28S/18S rRNA ratio of 1.8 – 2 was utilized for polyA isolation. PolyA RNA enrichment was performed on all total RNA samples according to the instruction and materials contained within the MicroPoly(A) Purist™ Kit (Ambion, Hilden, Germany).

Quantitative Real Time PCR

The cDNA was generated from RNA using VERSO™ Reverse Transcriptase in conjunction with Thermo's (Waltham, MA) SYBR® Green 1-Step QRT-PCR Low ROX Kit according to the manufacturer's instructions. To confirm that conditions were right to induce a drought response, Quantitative Real Time PCR (qRT-PCR) was performed as mentioned on a set of genes that are well known to be involved in the drought response, *DREB1*, and *DREB2A,* while *actin* was used as a constitutively expressed transcript to normalize data. Each cDNA sample was amplified in triplicate with each of the primers listed in **Table 2.1.** Primers for qRT-PCR were designed based on sequences obtained from Harvard University's DFCI *Aquilegia* Gene Index in addition to sequences obtained from our research (data not shown).

PCR conditions were; cDNA synthesis at 50°C for 15 minutes, Thermo-Start® activation at 95°C for 15 minutes, denaturation of DNA at 95°C for 15 minutes; 50 cycles of 95°C for 15 seconds, 52°C for 30 seconds, and 60°C for 60 seconds. To determine specificity of the primers a melt curve analysis was performed from 55°C to 99°C at 0.4°C intervals. RNA was extracted in triplicate for each 12 hour time point and those were run in triplicate for each of the 3-primer sets. The threshold cycle (C_t) of each was recorded. The $\Delta\Delta C_t$ method (Livak and Schmittgen, 2001) was performed to determine the relative transcript level for each time point. Data from the replicated experiments was examined to determine the transcript level of genes of interest due to the increase in drought conditions.

The RNA extractions were carried out in duplicate and run on qRT-PCR in triplicate; the average cycle threshold (C_t) was taken to identify the level of expression.

The formula;

$$[\Delta][\Delta]C_t = [\Delta]C_{t,sample} - [\Delta]C_t$$

Where C_t = the number of cycles it takes for the sample to cross an arbitrary line. This method was used to determine the relative amount of transcripts of genes of interest present in the *Aquilegia* RNA samples over the different time points.

Table 2.1 Primer sequences for DREB1, DREB2A and actin to determine steady state transcript level.

DREB2A	5'GTGAGGGAGTATGATATGTCCAAG 3' 5'GTCCACCATATACACCTGAGTGAT 3'
DREB1	5'TTTCATTATGTTCCTCTTCGATCTC 3' 5'GTTTGTATTTTTGGCTTGTCACT 3'
ACTIN	5' GTATAGTAAGCAACTGGGACGACA 3' 5' GGATAGCAACATACATAGCAGGAG 3'

Suppression Subtractive Hybridization (SSH) Library Construction

DNA was synthesized using Super SMART cDNA synthesis kit from CLONETECH according to the manufacturer's instructions. For construction of the suppression subtractive hybridization library, a PCR-select subtraction kit cDNA library kit (CLONETECH) (**FIGURE 2.1**) was used. Root, leaf and stem cDNA from Day 3 and 4 were combined in equal amounts and used as the tester, while Day 1 unstressed control cDNA was utilized as the driver. The final subtracted products were cloned into pGEM T Easy vector. After overnight ligation at 16°C, the ligase was heat denatured at 65°C, placed on nitrocellulose membrane, desalted on 5% PEG and then transformed into ElectroMAX™ DH10B™ Electrocompetent Cells from Invitrogen. These cells were allowed to recover in SOC media for one hour, and then plated onto LB media containing 300ug/mL of carbenicillin with X-gal and IPTG (isopropyl beta-D-thiogalactopyranoside. Both a forward reaction (tester) resulting in a library containing clones associated with drought stress and a reverse library (driver) containing clones present in unstressed plant tissue were created according to manufacturer's instructions.

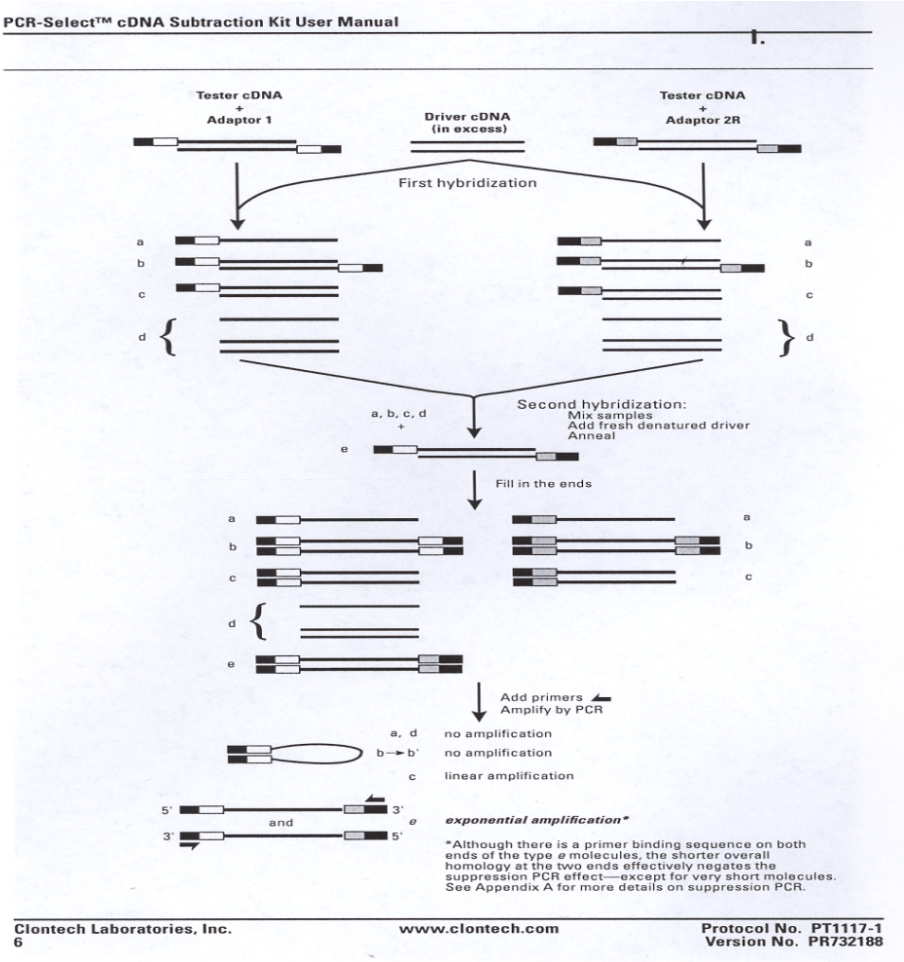


Figure 2.1: Diagrammatic representation of Suppression Subtractive Hybridization and cDNA library construction. Subtractive SubHybridization library creation. Schematic of ligation of adaptors to both subsets of tester cDNA. The first hybridization when combined with driver cDNA, followed by exponential amplification of tester cDNA containing different adaptor sequences on ends.

Positive recombinant clones were selected via blue and white screening to obtain a total of 1920 transformants from the “tester” and a total of 960 of the “driver”. They were picked robotically using the Genetix Q-bot and stored as individual clones in Genetix 384-well microtiter plates as glycerol stocks at -80°C. Plates were decondensed for sequencing into 96-well microtiter plates for use.

Bacterial Mini-Prep

DNA isolation was performed using a modified alkaline lysis protocol. One hundred and fifty µL of 2XYT media containing carbenicillin was inoculated with about 1 µL of culture on a 96 well replicator. These were then allowed to grow for 18 hours in a 37ºC incubator. After 18 hours these cells were then lysed utilizing SDS (10%), and sodium hydroxide (200mM). Cell debris was precipitated by addition of ice cold 7.5M ammonium acetate. In order to separate cell debris from DNA, the solution containing the cell debris and DNA was added to Pall Filters (East Hills, NY) and placed in ABgene® collection plates and centrifuged for 7.0 minutes at 4200 rpms in a Beckman centrifuge (Allegra X-15R) at 22ºC. Centrifugation yielded a clean filtrate that was precipitated with 0.7 volumes of isopropanol, centrifuged for 30 minutes at 3500 rpms, followed by a 70% Ethanol wash, 15 min at 3500rpms. The 70% ethanol was decanted and the DNA pellets were allowed to dry in a laminar flow hood for 1 hour. Pellets were resuspended in 80 µL of DNAse free water (GIBCO, Carlsbad, CA). to remove any RNA. Samples were digested with RNAse for 30 minutes at 37 ºC.

Sequencing

A total of 960 clones from the Day 1 control, unstressed/un-subtracted SSH library and 1920 clones from Days 3 and 4 SSH stressed library were subjected to Sanger style sequencing using ABI PRISM® BigDye™ Terminator v3.1 reaction mix (Applied Biosystems, Foster City, CA). SP6 primer (5'GATTTAGGTGACACTATAG3') was used to sequence from the 3' end of both libraries. PCR conditions for the sequencing reactions were carried out using the following conditions: 96ºC for 5 minutes, and 35

cycles of (96°C for 45 seconds, 50°C for 5 seconds, and 60°C for 4 minutes) in an MJ Research DNA engine cycler. Sequencing results were created and collected from an ABI 3730xl Genetic Analyzer from Applied Biosystems (Foster City, CA).

Analysis of Sequence Data

Sequences were trimmed of vector and adaptor sequences along with any low quality sections of bases with a Phred score less than 20. Sequences that had fewer than 100 high quality bases were not processed. A unigene set was constructed using CAP3 (Huang and Madan, 1999). Gene Ontology (GO) of the Extended Sequence Tags (EST) was performed using Blast2Go (http://www.blast2go.de/). This software was utilized to perform BLASTX similarity searches against GenBanks non-redundant protein database (http://www.ncbi.nlm.nih.gov/blast/blast.cgi). In order to detect simple sequence repeats (SSRs) a version of the Perl script SSRIT was used (Temnykh et al., 2001). The modified version, CUGISSR (Jung et al., 2005) was used to search for SSRs ranging from di-to penta-nucleotide repeats.

Results and Discussion

We constructed a Suppressive Subtracted Hybidization (SSH) library to identify genes and pathways differentially expressed in response to drought. These gene are thought to play some role in the ability of *Aquilegia* to conserve water in its surrounding and hold retain water present within the plant.

The day three and four time-point was chosen for evaluation of differentially expressed transcripts, with the idea being the genes expressed would be some of the earliest responses to desiccation conditions. To assure that most of these early response genes were "captured", Days 3 and 4 RNA were combined equally from root leaf and stem tissue RNA.

Under increasing drought conditions *Aquilegia* displayed very little morphological change when water was withheld until Day 3 (**Figure 2.2**). In this experiment 18 section flats were utilized so available water was significanty less than if larger one gallon containers were utilized as in the other experiments.

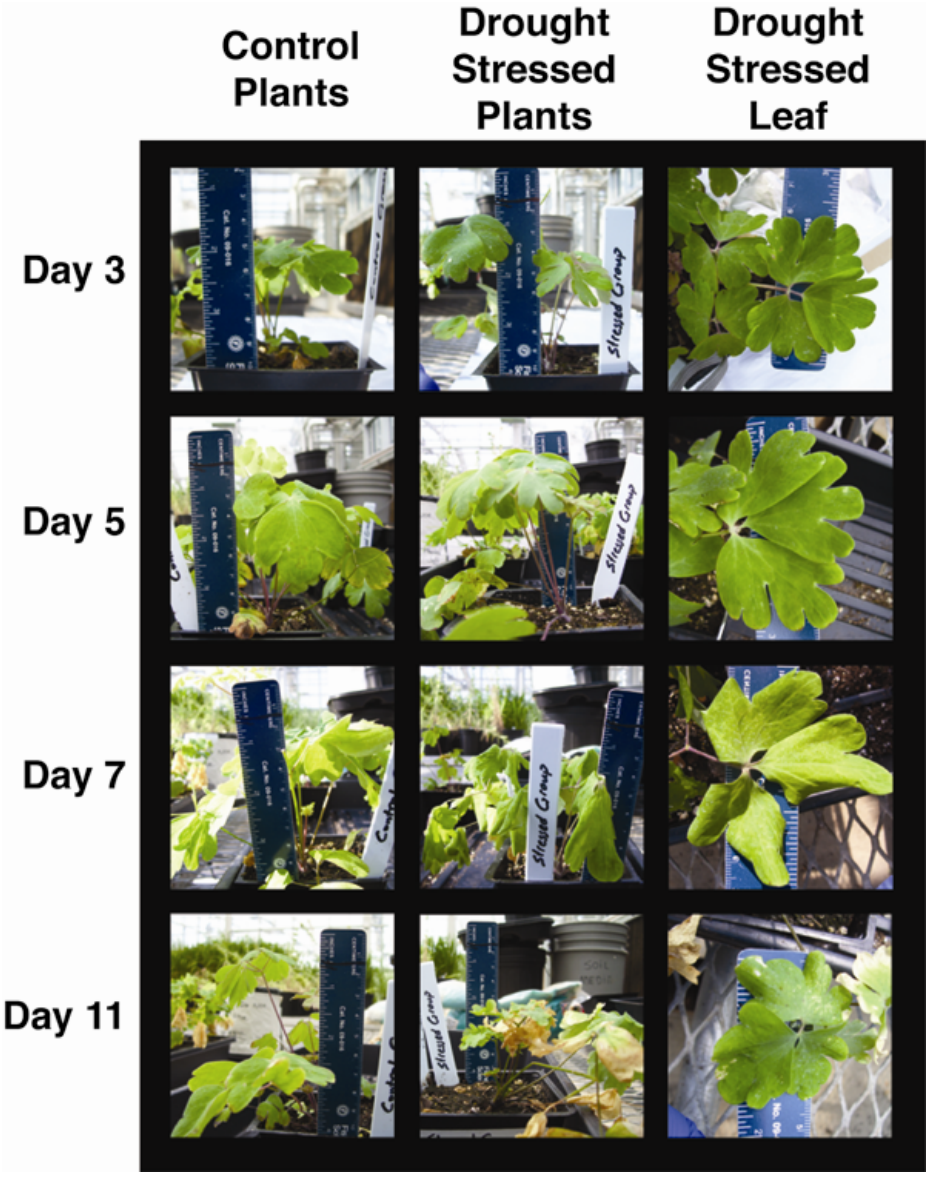


Figure 2.2: Morphological changes in *Aquilegia* plants in drought conditions. Column one: no drought conditions. Column two: overall plant. Column three: close up of leaf. Plants were surveyed for morphologic variation as drought conditions increased over an eleven day experiment.

Quantitative Real Time PCR

Leaf and stem samples were taken between 8:00 and 9:00 AM and again 12 hours later between 8:00 and 9:00 PM. Samples were divided into two different data sets, AM

and PM, because of the variation in metabolic activity with increased exposure to light over the 12 hour period thus increasing the level of genes of interest.

From the data extrapolated from the qRT-PCR results, we found that both DREB1A and DREB2 were up-regulated under the presented drought stress conditions (**Figures 2.3 to 2.4)**.

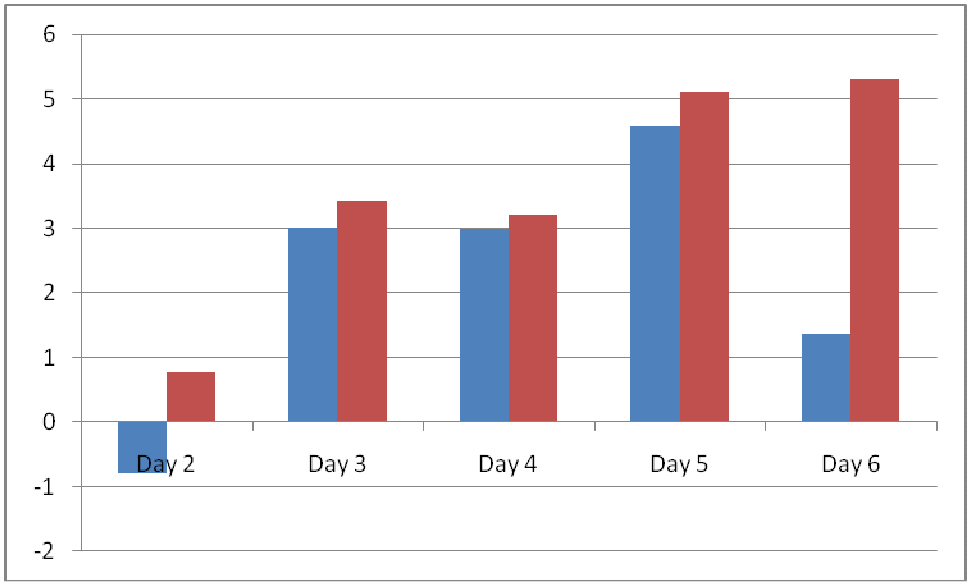


Figure 2.3: Relative transcript *DREB1* Level in morning and evening
RNA was obtained in morning (BLUE) and evening (RED) for this quantitative assay. The $[\Delta][\Delta]C_t$ method was utilized for relative transcription level as *Aquilegia* was undergoing drought.

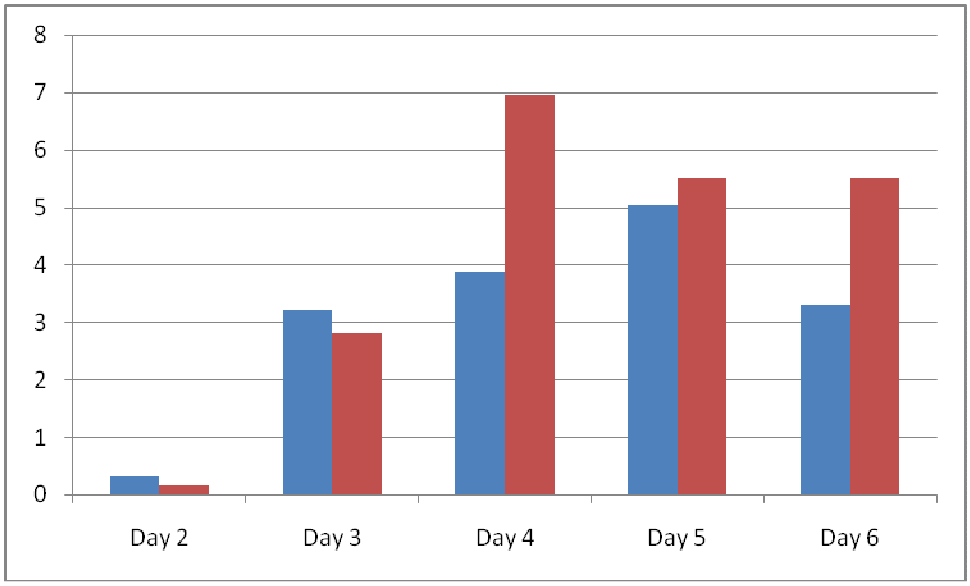


Figure 2.4: Relative transcript *DREB2A* Level in morning and evening

RNA was obtained in morning (BLUE) and evening (RED) for this quantitative assay. The $[\Delta][\Delta]C_t$ method was utilized for relative transcription level as *Aquilegia* was undergoing drought.

Subtractive Hybridization Library Characterization

One major setback in utilizing traditional cDNA libraries for rare transcript genes is a difference in mRNA abundance. More than half of clones produced in these libraries can come from several highly and moderately expressed mRNAs, and the remaining fraction of sequences are from several thousand different mRNAs (Soares et al., 1994). In order to obtain the rarer transcripts present in one sample but not another, the creation of supressive subtractive hybridization library was utilized. This utilizes competitive hybridization to bind transcripts present in driver sample to homologous transcripts present in the tester cDNA population. This is followed by adaptor specific PCR of adaptor ligated tester samples to amplify low abundance an mRNA population. This

method decreases the amount of sequencing needed to find rare transcripts, and decreases the data analysis complexity.

Competitive hybridization was utilized to subtract abundant transcripts present in Day 1 (unstressed) samples from Day 3 (stressed) samples. This technique resulted in exponential amplification of differentially expressed transcripts from Day 3, followed by TA cloning into pGEMgem T Easy vector termed a forward library. A second library was created utilizing the same technique but reversely, resulting in cloning of differentially expressed genes present on Day 1 but not on Day 3, termed a reverse library.

It is assumed that due to the increased drought conditions from well watered Day 1 to the initial stages of drought on Day 3 and 4 there will be a change in the gene profile of *Aquilegia*. It was hypothesized that identification of genes and their tentative function could be utilized to expand current knowledge of the drought signal cascade. Identifying genes up-regulated under drought conditions but not present, or present at extremely reduced levels, under non-stress conditions will further understanding of different mechanisms of drought resistance.

Suppressive subtractive hybridization libraries for Day 1 (AHOTEg, *Aquilegia* Hybrid Origami mix Tomkins Est g) and Day 3 and 4 combined (AHOTEh, *Aquilegia* Hybrid Origami mix Tomkins Est h) *Aquilegia* were constructed by ligation of *Rsa*I digested cDNA into pGEM T Easy vector. From the AHOTEg library 1920 clones were randomly selected via blue white screening in the presence of carbenicillin. AHOTEg inserts were released with *Eco*RI the inserts ranged from 300 base pair (bp) to 2000 bp

with an average insert size of 761 (base pairs) (**Figure 2.5**). This library contained 0% non-recombinant clones. Sequencing success from AHOTEg was 95.29%, and yielded a success rate of 86.18% after trimming of adaptor sequence, poly-A tails and vector contamination.

From the AHOTEh library 960 recombinant clones were randomly selected in the presence of carbenicillin, Xgal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside), and IPTG (isopropyl beta-D-thiogalactopyranoside). AHOTEh yielded an average insert size of 573bp with sizes ranging from 350bp to 1000 bp and contained 3% non-recombinant clones (**Figure 2.6**). Sequencing success from AHOTEh was 97.6 % and yielded a success rate of 88.23% after trimming of adaptor sequence, poly-A tails and vector contamination.

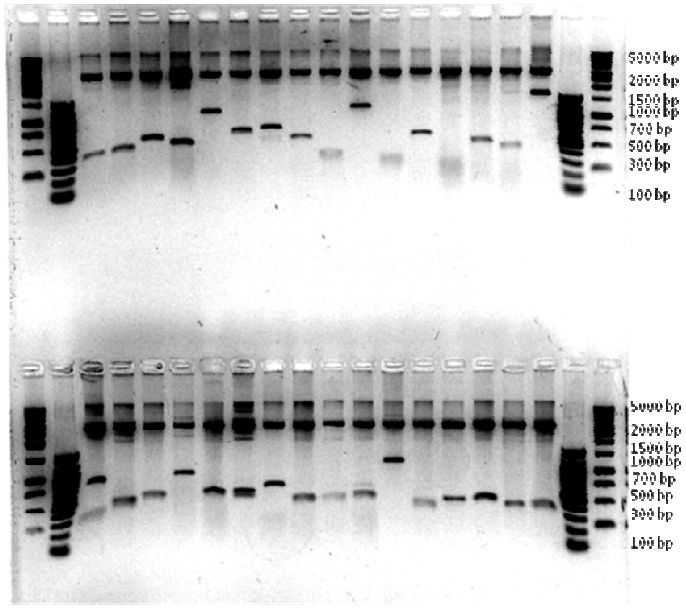


Figure 2.5: Insert size of AHOTEg.
The insert sizes of 32 randomly picked cDNA clones were determined by *Eco*RI digestion. Average insert size was 761 base pairs with 0.0 % non-recombinant clones.

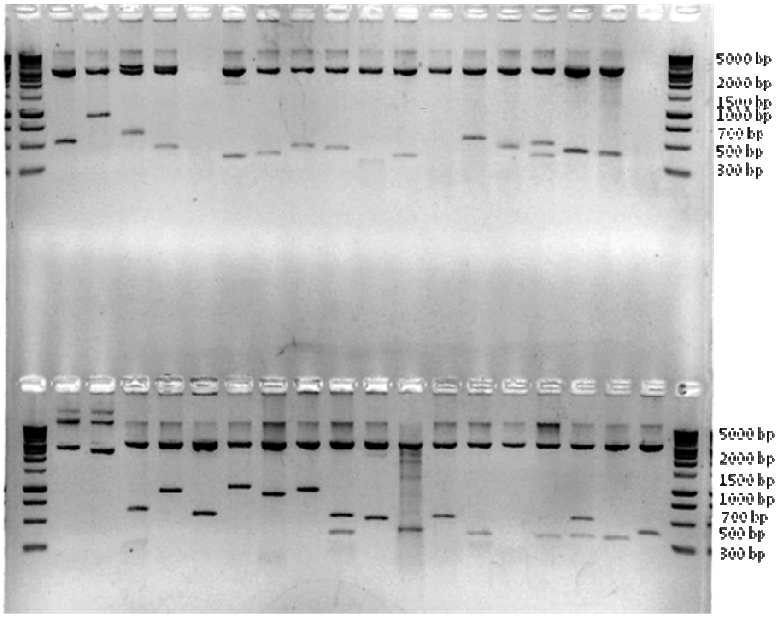


Figure 2.6: Insert size AHOTEh.
The insert sizes of 32 randomly picked cDNA clones were determined by *Eco*RI digestion. Average insert size was 573 base pairs with 3.0 % non-recombinant clones.

Assembly and Annotation of Sequences

Sequences that were trimmed of vector sequence were subjected to a CAP3 assembly program (Huang and Madan, 1999). This program is designed to align homolgous stretches of sequence and assemble them into one contiguous (contig) stretch of sequences to reduce the size of a dataset and also to create a more redundant and longer sequences for further searches.

For the stressed library AHOTEg the CAP3 program yieled 173 contigs and 272 singlets. For the unstressed Day 1 library AHOTEh, the CAP3 program yieled 108 contigs and 324 singlets. These unigene sets were subjected to Gene Ontology based

data mining by Blast2GO (Conesa et al., 2005). These sequences were searched against the non-redundant database of NCBI, using a BLASTX cut off value 1 E-3.

Of the 108 contigs and 324 singlets obtained from the Day 1 library, 55% were annotated using the Blast2GO software, 28% showed no homology to any database, 11% yielded no annotation, and 5% yielded no mapping results (**Figure 2.7)**. Day 3 and 4 library was searched and, 56% were annotated using the Blast2GO software, 24% showed no homology to any to any database searched, 14% yielded no annotation, and 6% yielded no mapping results (**Figure 2.8**).

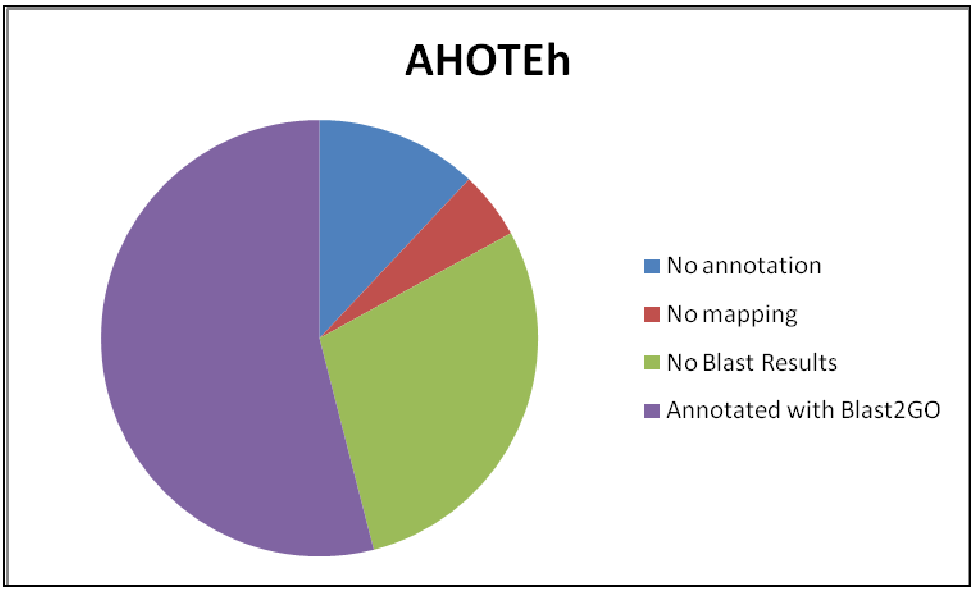


Figure 2.7: Data distribution from Day 1, AHOTEh SSH Library.

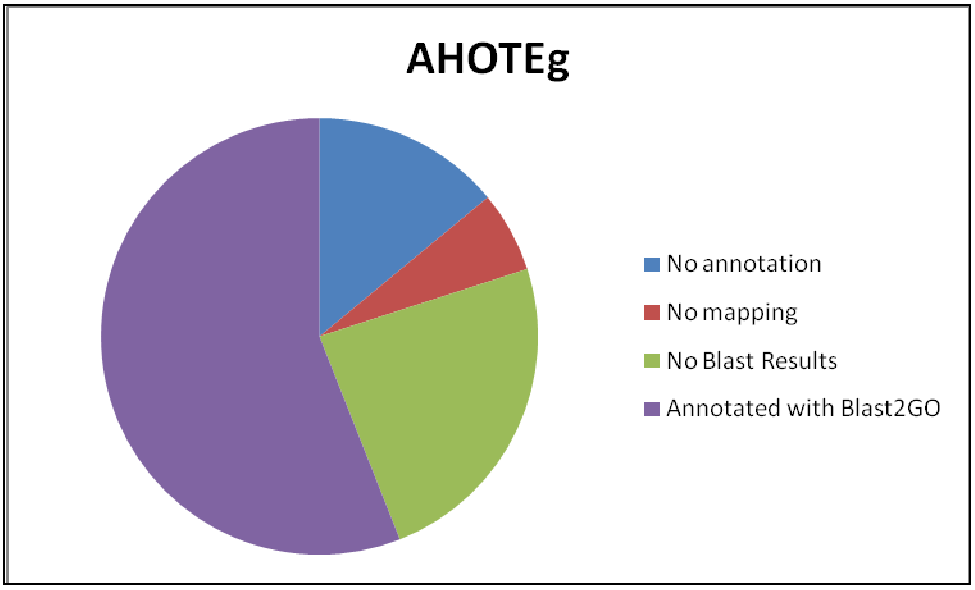


Figure 2.8: Data distribution from AHOTEg, Day 3 /4 SSH Library.

Further analysis of the AHOTEg stress data set yielded the number of individual sequences present within each Contig. Of the 173 contigs, the one that contained the most sequences was a major latex protein 28 (MLP28). The number of clones present in the stressed library gives some idea into the abundance of these sequences in the transcriptome of *Aquilegia* in drought conditions. Most of these genes are known to play some role in defensive mechanisms of plants in helping deal with drought (Figure 2.9).

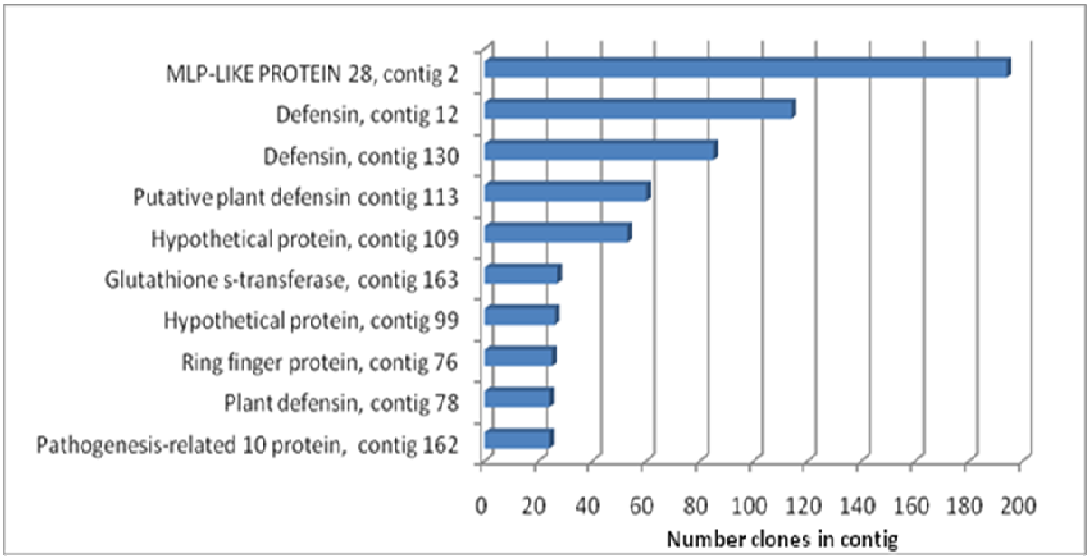


Figure 2.9: AHOTEg clone frequency. Annotation of Stressed libraries contigs. Top ten contigs containing most individual sequences.

Genes with known function were sorted into biological, molecular and cellular functional categories. The three catagories are represented in **Figure 2.10, 2.11, and 2.12,** respectively, for the unstressed library (AHOTEh), and **Figure 2.13, 2.14, and 2.15,** respectively, for the drought stressed library (AHOTEg). Percentages can add up to greater than 100% because individual genes can fall into multiple catagories. Sequences associated with a response to stimulus make up the largest fraction of the biological function subset in the stressed library (14%). This was expected due to the drought conditions present at the time of RNA extraction and gives validity to the subtraction hybridization method used to selectively amplify differentially expressed genes.

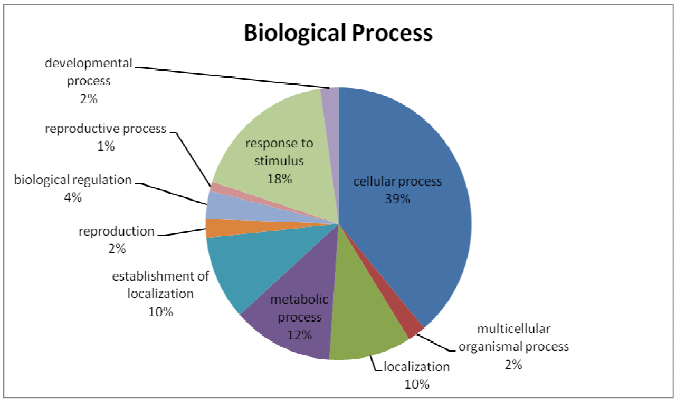


Figure 2.10: GO Mapping of Aquilegia Day 1 (AHOTEh) biological process unstressed library

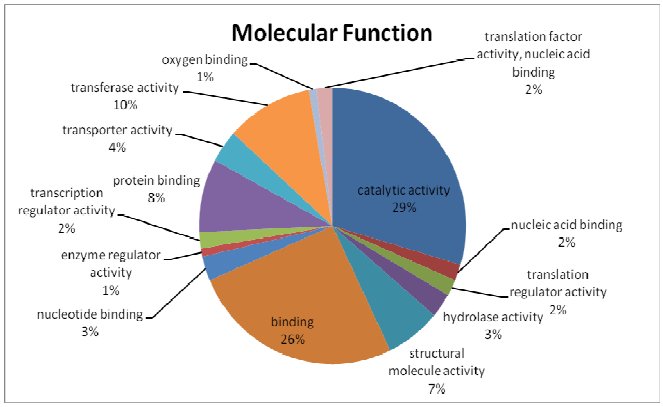


Figure 2.11: GO Mapping of Aquilegia Day 1 (AHOTEh) Molecular Function unstressed library

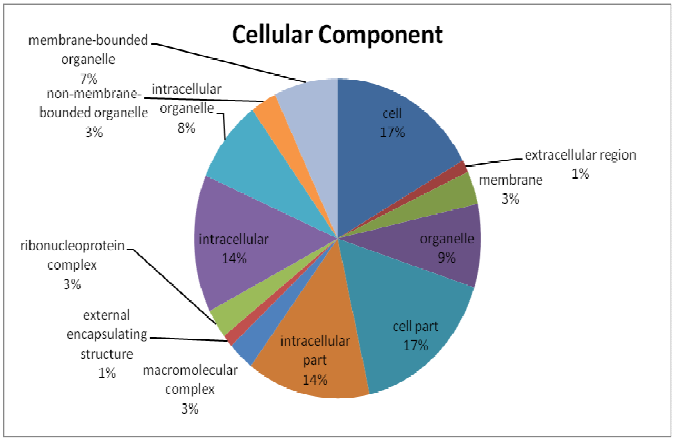


Figure 2.12: GO Mapping of Aquilegia Day1 (AHOTEh) cellular component unstressed library

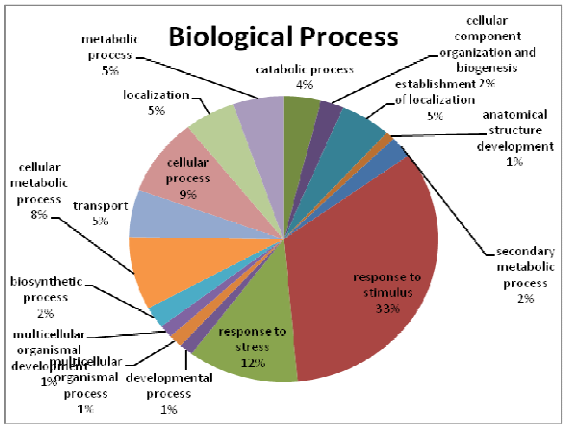


Figure 2.13: GO Mapping of Aquilegia Day 3/4 (AHOTEg) Biological Process stressed library

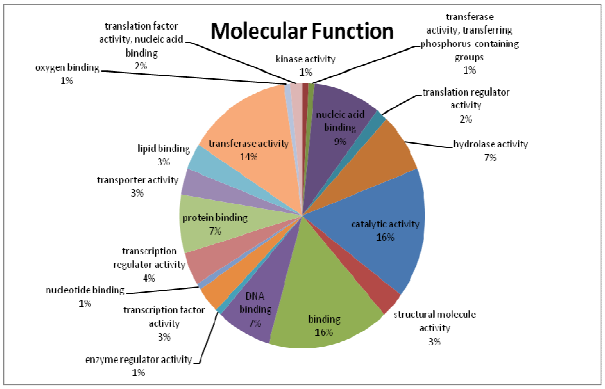


Figure 2.14: GO Mapping of Aquilegia Day 3/4 (AHOTEg) Molecular Function stressed library

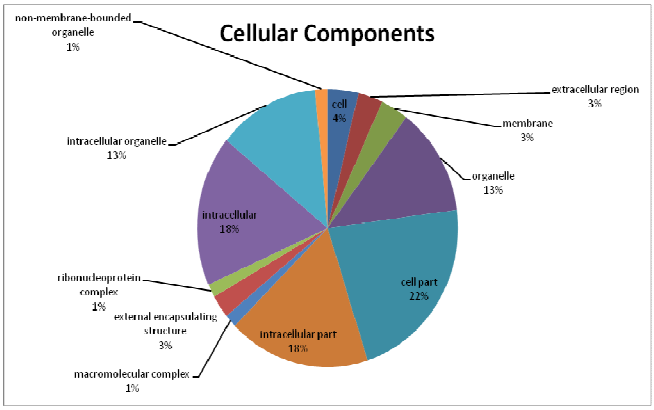


Figure 2.15: GO Mapping of Aquilegia Day3/4 (AHOTEg) Cellular component stressed library

Suppression subtractive hybridization (SSH) amplifies rare mRNA transcripts, and differentially expressed genes present in one sample but not another. The expressed sequences present in only the stressed library can be seen in **Table 2.2.** Sequences found only in the unstressed library can be seen in **Table 2.3**. Of the 173 Tentative Contigs (TCs) present in the stressed library, 113 of these sequences were differentially expressed, meaning they had no homology to sequences that were obtained from the un-stressed library. Many genes known to be involved in drought and various stress responsive pathways were identified from the stressed library. Transcription factors such as *DREB* containing sequences, ethylene-responsive transcriptional co-activator, and Homeodomain leucine zipper protein *HDZ2* were present only in the stressed library. *DREB* is known to have a role in plant response to drought (Shinozaki, 1999). Overexpression of *DREB1A/CBF3* in *Arabidopsis* under control of cauliflower mosaic (CaMV) 35S promoter yielded an increased tolerance to freezing (Jaglo-Ottosen et al., 1998). Overexpression of *DREB1A* yielded an increase in tolerance to drought, high-salt, and freezing stress (Liu et al., 1998; Kasuga et al., 1999). Other proteins found in the stressed library known to play a role in drought response are aquaporons which help regulate water levels within plant cells, peroxidases which help with detoxification of H_2O_2 (Aroca et al., 2005), Glutathione S-transferase which has been shown to increase during drying of the drought-tolerant moss *Tortula ruralis* (Dhindsa, 1991) and low temp and salt responsive protein LT16B. These are only a few of the proteins with known homology to drought responsive genes found in the stressed subtractive hybridization library.

One significant observation was that 20 Tentative Contigs had no known function, or were listed only as hypothetical proteins. The large number of completed plant genome sequences and the ease of availability of this sequence data suggest that the genes represented here represent novel genes or novel pathways. These genes could be interesting for further functional analysis and could possibly lead to a further and more in-depth understanding of known drought defense pathways. It can be anticipated that these sequences will be a focus of further study and will have more light shed on them in the coming years with a draft sequence of *Aquilegia* currently in progress at the Joint Genome Institute (http://www.jgi.doe.gov/sequencing/why/51280.html).

Table 2.2: Annotation of differentially expressed ESTs obtained from drought induced SSH library

Protein of Known Function	Homologous organism	e Value	ID#
BIOLOGICAL PROCESS			
Localization			
Temperature-induced lipocalin	*Lycopersicon esculentum*	1.62E-72	Contig10
Aquaporin PIP1-2	*Gossypium hirsutum*	1.09E-42	Contig89
Acyl-COA binding protein	*Panax ginseng*	1.39E-38	Contig168
Autophagy-related protein 8 precursor	*Lodderomyces elongisporus*	6.61e-33	Contig65
Ferritin-3, chloroplast precursor	*Glycine max*	4.13E-48	Contig15
Response to stimulus			
Norcoclaurine synthase	*Thalictrum Flavum*	9.06E-29	Contig27
Antifungal protein	*Sinapis alba*	3.47E-24	Contig39
Low temp and salt responsive protein LT16B	*Arabidopsis thaliana*	3.39E-12	Contig100
Peroxidase 25	*Ricinus communis*	8.56E-40	Contig46
GAST1 protein precursor	*Ricinus communis*	3.98E-25	Contig32
SAUR family protein	*Populus trichocarpa*	1.21E-30	Contig49
Metabolic process			
Atp-dependent cop protease proteolytic subunit	*Ricinus communis*	4.11E-33	Contig16
Cysteine proteinase	*Ricinus communis*	1.74E-75	Contig129
Xyloglucan endotransglycosylase hydrolase	*Gerbera hybrid cultivar*	4.98E-115	Contig66
Triosephhosphate isomerase	*Coptis japonica*	7.19E-85	Contig150
Beta-cyanoalanine synthase	*Malus x domestica*	1.06E-23	Contig14
Putative arogenate dehydrogenase isoform 2	*Oryza sativa*	2.55E-57	Contig112
Beta-1,3 glucanase	*Vitis vinifera*	1.61E-50	Contig19
Beta-1,3 glucanase basic isoform precursor	*Ricinus communis*	1.01E-57	Contig38
Beta-glucosidase	*Ricinus communis*	5.21E-29	Contig151
Oxidation reduction			
SFerH-3	*Glycine max*	4.13E-48	Contig15
Peroxidase 25	*Ricinus communis*	1.41E-42	Contig46
Canadine synthase	*Thalictrum flavum*	1.79E-13	Contig135
Biological regulation			
Pi starvation-induced protein	*Cicer arietinum*	4.23E-31	Contig159
MOLECULAR FUNCTION			
Catalytic activity			
S-norcoclaurine synthase	*Thalictrum flavum subsp. glaucum*	9.03E-29	Contig27
Actoylglutathione lyase family protein	*Arabidopsis thaliana*	4.78E-54	Contig21
Transferase activity			
3'hydroxy-N-methyl-(S)-coclaurine 4'-OMT	*Thalictrum flavum* subsp. g*laucum*	7.39E-60	Contig3
Glutathione s-transferase	*Ricinus communis*	2.82E-79	Contig35
Glutathione s-transferase GST 22	*Glycine max*	3.14E-77	Contig163
S-scoulerine-9-OMT	*Thalictrum flavum* subsp. *Glaucum*	4.53E-40	Contig165
Pavine-N-methyltransferase	*Thalictrum flavum*	4.97E-80	Contig86

Table 2.2: (Continued) Annotation of differentially expressed ESTs obtained from drought induced SSH library

Protein of Known Function	Homologous organism	e Value	ID#
Oxidoreductase activity			
Putative phosphatase	*Solanum lycopersicum*	2.49E-95	Contig84
UDP-glycosyltransferase 85A8	*Stevia rebaudiana*	1.34E-53	Contig173
Binding			
Fiber annexin	*Gossypium hirsutum*	1.00E-56	Contig158
Phosphatidylethanolamine binding protein	*Ricinus communis*	*2.25*E-69	Contig171
Oxidoreductase Activity			
Cytochrome P450	*Populus trichocarpa*	1.19E-52	Contig41
Cytochrome P450	*Petunia x hybrid*	2.15E-94	Contig33
Ion binding			
Metallothionein-like protein	*Arachis hypogaea*	5.77E-20	Contig147
Nucleic acid binding			
Ethylene-responsive element-DREB	*Nicotiana tabacum*	3.41E-33	Contig1
Ethylene-responsive transcriptional coactivator	*Elaeis guineensis*	1.89E-62	Contig133
Homeodomain leucine zipper protein HDZ2	*Phaseolus vulgaris*	2.80E-69	Contig64
Structural constituent of ribosome			
60S Ribosomal protein L6	*Ricinus communis*	3.11E-73	Contig63
60S Ribosomal Protein L13a	*Ricinus communis*	1.04E-97	Contig98
60S Ribosomal protein L23	*Zea mays*	7.30E-18	Contig172
Transporter activity			
Transparent testa 12 protein	*Zea Mays*	1.35E-37	Contig90
CELLULAR FUNCTION			
Putative root storage protein	*Ranunculus asiaticus*	1.11E-42	Contig36
AWPM-19-like membrane family protein	*Arabidopsis thaliana*	1.27E-51	Contig139
Erwinia induced protein	*Zea mays*	8.23E-41	Contig34
NO KNOWN FUNCTION			
Predicted protein	*Populus trichocarpa*	6.36E-19	Contig91
Predicted protein	*Populus trichocarpa*	1.07E-34	Contig67
Drm3-like protein	*Solanum tuberosum*	7.46E-29	Contig11
Auxin-repressing protein	*Zea mays*	1.59E-9	Contig20
LEA protein group3	*Zea mays*	5.33E-19	Contig22
Zinc-finger protein 1	*Datisca glomerata*	4.63E-24	Contig40

Table 2.2: (Continued) Annotation of differentially expressed ESTs obtained from drought induced SSH library

Protein of Known Function	Homologous organism	e Value	ID#
NO ANNOTATION			
Predicted protein	*Vitis vinifera*	4.9E-11	Contig28
Predicted protein	*Populus trichocarpa*	6.13E-6	Contig42
Hypothetical protein	*Vitis vinifera*	1.58E-37	Contig43
Putative chorismate prephenate dehdratase	*Oryza sativa*	1.30E-68	Contig44
Dehydrin	*Citrus X paradise*	3.99E-8	Contig53
Hypothetical protein	*Vitis vinifera*	2.15E-11	Contig62
Protein disulfide	*Vitis vinifera*	1.09E-8	Contig70
PR-10 norcoclaurine synthase-like protein	*Eschscolzia californica*	8.69E-39	Contig71
Pathogenesis-related protein	*Cucumis sativus*	1.25E-30	Contig75
Transparent testa 12 protein	*Zea mays*	2.71E-9	Contig77
Conserved hypothetical protein	*Ricinus communis*	4.37E-5	Contig81
Ripening-related protein	*Medicago truncatula*	2.58E-100	Contig87
Predicted protein	*Populus trichocarpal*	1.22E-30	Contig94
Unnamed protein product	*Vitis vinifera*	5.35E-31	Contig95
Hypothetical protein	*Vitis vinifera*	1.99E-46	Contig43
Protein	*Vitis vinifera*	2.01E-18	Contig97
Little protein 1	*Arachis hypogaea*	1.47E-29	Contig99
Conserved hypothetical protein	*Ricinus communis*	1.12E-48	Contig104
Conserved hypothetical protein	*Ricinus communis*	1.50E-23	Contig109
Histone h2	*Populus trichocarpal*	3.30E-38	Contig117
Putative staygreen protein	*Nicotiana tabacum*	4.31E-101	Contig118
Antifungal protein	B*rassica napus*	3.19E-5	Contig119
Hypothetical protein isoform 2	*Vitis vinifera*	5.83E-28	Contig120
Hypothetical protein	*Vitis vinifera*	2.97E-7	Contig121
Hypothetical protein	*Vitis vinifera*	7.43E-63	Contig126
Universal stress protein 1	*Gossypium arboreum*	8.62E-56	Contig127
Hypothetical protein	*Vitis vinifera*	4.32E-67	Contig128
Hypothetical protein	*Ricinus communis*	4.63E-35	Contig143
Homeobox-leucine zipper protein athb-12	*Arabidopsis thaliana*	3.32E-32	Contig145
Polyphenol oxidase	*Camellia sinensis*	2.67E-24	Contig153
Hypothetical protein	*Vitis vinifera*	1.98E-54	Contig155
CBL-interacting serine threonine-protein	*Ricinus communis*	2.98E-40	Contig160
Conserved hypothetical protein	*Ricinus communis*	1.17E-4	Contig161
Hypothetical protein	*Vitis vinifera*	2.70E-89	Contig167
Unknown protein	*Arabidopsis thaliana*	5.46E-30	Contig169
Protein	*Populus trichocarpal*	6.95E-8	Contig170
Putative carboxy-layase	*Ricinus communis*	2.26E-44	Contig174

Table 2.3: Annotation of differentially expressed ESTs obtained from non- induced SSH library

Protein of Known Function	Homologous organism	e Value	ID#
BIOLOGICAL PROCESS			
Localization			
TMP-C Plasma membrane intrinsic protein 1;4	*Arabidopsis thaliana*	6.90E-84	Chontig17
Putative H-type thioredoxin	*Citrus* hybrid cultivar	2.34E-49	Chontig24
Thioredoxin H-type, putative	*Ricinus communis*	9.77E-38	Chontig34
Lipid binding protein	*Ricinus communis*	6.61E-42	Chontig38
Protein transport protein SEC61 gamma subunit	*Oryza sativa* Japonica group	1.58E-22	Chontig84
ADP-ribosylation factor	*Elaesis guineensis*	3.93E-99	Chontig87
Response to stimulus			
Nucleoside diphosphate kinase B	*Flaveria bidentis*	7.24E-39	Chontig66
Low molecular weight heat-shock protein	*Papaver somniferum*	1.25E-40	Chontig5
Heat shock protein 90-1	*Glycine max*	9.14E-42	Chontig19
Cationic peroxidase	*Nelumbo nucifera*	3.83E-102	Chontig36
Metabolic process			
Progesterone 5beta reductase-B	*Nicotiana tabacum*	1.29E-51	Chontig92
Cinnamyl alcohol dehydrogenase 1b	*Festuca arundinacea*	2.09E-67	Chontig80
Catechol O-methyltransferase Chontig108	*Thalictrum tuberosum*	4.96E-40	
Cysteine protease	*Vitis vinifera*	3.37E-92	Chontig22
NAD-dependent malate dehydrogenase	*Prunus persica*	2.42E-18	Chontig37
Dimethyladenosine transferase	*Zea mays*	2.09E-55	Chontig78
Phenylalanine ammonia lyase	*Robinia pseudoacacia*	4.21E-55	Chontig74
Putative heat shock protein	*Ricinus communis*	1.08E-42	Chontig19
Triose phosphate isomerase cytosolic isoform	*Solanum chacoensel*	1.95E-23	Chontig4
Aspartate aminotransferase putative	*Ricinus communis*	5.92E-25	Chontig61
Oxidation reduction			
1-aminocyclopropane-1-carboxylate oxidase	*Momordica charantia*	4.95E-11	Chontig75
Nitrogen compound metabolic process			
Isoflavone reductase-like protein 5	*Vitis vinifera*	1.79E-35	Chontig79
MOLECULAR FUNCTION			
Catalytic activity			
S-adenosyl-L-methionine	*Coptis chinensis*	7.18E-71	Chontig47
Structural constituent of ribosome			
Ribosomal protein L30, putative	*Ricinus communis*	8.77E-51	Chontig76
60S ribosomal protein L35	*Ricinus communis*	4.70E-38	Chontig42
Putative L24 ribosomal protein	*Arachis hypogaea*	1.40E-50	Chontig32

Table 2.3: Annotation of differentially expressed ESTs obtained from non- induced SSH library

Protein of Known Function	Homologous organism	e Value	ID#
Structural constituent of ribosome			
60S ribosomal protein L29	*Arabidopsis thaliana*	4.17E-18	chontig14
Binding			
Calmododulin-related protein, putative	*Arabidopsis thaliana*	8.60E-38	chontig98
CELLULAR COMPONENT			
Organelle			
Putative frnE protein-like	*Arabidopsis thaliana*	1.91E-78	chontig70
Chloroplast pohotosystem II light-inducabel protein	*Pachysandra terminalis*	1.52E-40	chontig101
Putative photosystem I reaction center subunit II	*Arabidopsis thaliana*	4.90E-27	chontig12
NO KNOWN FUNCTION			
Hypothetical protein	*Vitis vinifera*	3.14E-48	chontig99
NO ANNOTATION			
Glutathione peroxidase 4	*Populus trichocarpa*	5.32E-26	chontig8
Desacetoxyvindoline 4-hydrolase, putative	*Ricinus communis*	7.34E-52	chontig9
Progesterone 5-beta-reductase	*Digitalis cariensis*	3.30E-34	chontig13
Transparent testa 12 protein	*Zea mays*	2.23E-32	chontig26
Protein	*Vitis vinifera*	2.89E-41	chontig46
TR11	*Cleome spinosa*	3.23E-61	chontig54
Hypothetical protein	*Ricinus communis*	1.00E-21	chontig83
24 kda seed maturation protein	*Elaeis quineensis*	4.06E-13	chontig93
UP-9A	*Nicotiana tabacum*	1.84E-16	chontig100
Tropinone reductase homolog	*Datura stramonium*	1.18E-56	chontig102
Thioredoxin m	*Populus trichocarpa*	1.32E-51	chontig107

Recent studies suggest that simple sequence repeats (SSRs) located within protein coding regions and their untranslated regions (UTRs) can lead to a gain or loss of gene function via frameshift mutation or expanded mRNA which can ultimately lead to disruption of cellular function (Li, 2004). SSRs present within genes show an increased rate of mutation, similar to the increased mutation rate of SSRs in the UTRs of genes (Li, 2004). The elongation/shortening process leads to an increase in biologic complexity. These SSR regions are listed in Appendices A and B. These sequences are possible candidates for use as genetic markers in marker assisted breeding leading to a more in depth understanding of drought resistance and an additional avenue to improve various crop species. The rarest and most specific SSR can be used for identification among different crosses to help dissect molecular traits. Based on the SSR analysis several contigs contain numerous SSRs ranging from one to four nucleotide repeats. The genes containing SSRs, could be compared with their homologs in *Arabidopsis* and other species for comparison to help with the study of function.

From the drought-stressed library (AHOTEg) 113 of the sequences were differentially expressed. SSR analyses of these tentative contigs revealed that 39 of them contained at least one SSR and are composed of di- to tetra- nucleotide repeating units (Appendix A). Nearly 44% of all SSRs are tri-nucleotide repeats. The next most common SSR consists of tetra-nucleotide repeats and accounts for 31% of the SSRs. Di-nucleotide repeats make up 24% (Appendix A)**.**

The un-stressed library resulted in 62 differentially expressed sequences; these tentative contigs were analyzed for SSRs. Of which, 22 TCs were found to contain SSRs.

Of the sequences containing SSRs, 58% of these were found to be tri-nucleotide repeats, 16% tetra nucleotide repeats and 26% di-nucleotide repeats (Appendix B).

Conclusions

Response to drought and the means plants utilize to cope with drought conditions is a highly complex cascade of events. In order to attempt to simplify the comprehension of these various pathways SSH libraries are an excellent resource when Sanger sequencing is to be utilized. This study is the first utilization of SSH to study changes in the transcriptome of *Aquilegia* in response to drought. Sequences obtained from the forward subtracted SSH library yielded numerous gene sequences of proteins with functions known to deal with drought conditions in plants. Genes encoding for proteins such as, Aquaporins, Low temp and salt responsive protein, glutathione S-transferase and ethylene-responsive element-DREB were identified in this data set.

Numerous gene sequences were obtained with no known function. Identification and functional analysis of these genes with no known function will be valuable in furthering understanding of the various pathways involved with drought response. Twenty nine pecent of the sequences obtained from the stressed library produced no homology to any sequences in the public databases. This proves that the SSH approach is a powerful tool for identifying uncharacterized genes, specifically related to increased stress conditions. Additional research is needed to study these novels genes to determine if they are truly involved in a defensive mechanism for *Aquilegia.*

SSR analysis has yielded various possible genetic markers for *Aquilegia* that may be utilized in a breeding program to improve the overall drought resistance of *Aquilegia*, with any information gained applicable to various species because of its evolutionary position. These sequences need to be annotated as to what their homologous genes are and examined in depth the future as to the number of SSRs present in each sequence. This data should be very useful when aligned to homologs present in the NCBI database. This may yield information applicable to a marker assisted breeding program to future dissect *Aquilegia'*s ability to withstand drought conditions.

The information gained from this subtractive hybridization library is an excellent resource for parallel experiments dealing with what particular genes, enzymes, or proteins are instrumental in *Aquilegia's* ability to withstand drought. This type study will contribute to determining the molecular basis of *Aquilegias* elevated resistance to drought. Equally important will be determining new genes involved in drought tolerance to possibly be transferred to crop species for an increased ability to withstand drought and increase yield in water limited climates.

CHAPTER THREE

IDENTIFICATION AND CHARACTERIZATION OF TWO FLAVANOIDS ACCUMULATED IN *AQUILEGIA* WHEN EXPOSED TO DROUGHT

Abstract

Plants possess numerous mechanisms to deal with decreased water conditions such as reduction of leaf size and number, stomata closure and accumulation of osmolytes. In *Aquilegia*, a response to severe drought is reflected by increasing production of osmolyte-like compounds. Methanol soluble, polar compounds were isolated from *Aquilegia* under drought stress and quantitatively analyzed by HPLC in relation to unstressed plants. Two compounds with the highest rate of change were unambiguously identified structures by mass spectrometry when compared to known standard compounds. We baseline purified four fractions corresponding to four different fractions that were significantly increased over the two-week drought induced time span. From the MS and MS/MS fragmentation patterns two of the compounds were identified as apigenin 8-C-glucoside (vitexin) and vitexin 2'-O-β-D-glucoside. The identification of these glucosides and their structural similarity to a known osmolyte, trehalose, support the hypothesis that these compounds play an important role in *Aquilegia's* ability to withstand severe drought conditions

Introduction

The term "-omics" refers to the identification, analysis, characterization and interactions of groups of biological molecules systematically such as metabolites,

proteins, or the transcriptome. This often means analysis of hundreds to thousands of biological molecules present at any given time within a cell, simultaneously. Bridging the genome and proteome to the metabalome is one of the major goals in plant genetic studies. Plant genome sequencing efforts have yielded identification of gene function that lead to exploration into the proteome and subsequently into the metabalome to better understand plant metabolism.

Plants have numerous ways to deal with drought; many drought responsive genes and proteins in plants have been identified by transcriptome analysis and proteomic analysis (Komatsu, 2006). It is observed that the mRNA level may not directly correlate with a protein expression level (Gygi et al., 1999). Similarly, proteins can be present within cells but not in active form (Sumner et al., 2003). In other words, the transcriptome and proteome are only partial representations of the overall metabolic processes that are occurring within tissue at any given time, an area where metabalomics can shed some light in understanding the discrepancy in stress responses at the metabolic level.

This study used metabalomics, for differential profiling of *Aquilegia*'s metabalome under various drought conditions. Metabolites which accumulated in larger quantities when subjected to drought conditions were identified. Metabolites are end products from their respective gene expression pathways and take some part in a variety of different functions within the biological process. Here we report on quantitative analysis of metabolites by reverse phase HPLC of methanol fraction of *Aquilegia* subjected to increased levels of drought conditions over a two-week induction period.

Aquilegia (Columbine) has a natural range from Alaska to Florida (USDA, 2008) and can be found from an elevation of 10,000 feet to sea level. This wide geographic distribution shows the wide adaptability of *Aquilegia*. Columbines are found in shaded forest, alpine zones, desert spring areas and rocky outcrops (Munz, 1946). Because of the wide range of environments Columbine thrives in, it is an excellent candidate for identifying novel genes for drought resistance which becomes increasingly more important as world population increases.

The morphological effects of water deficits are fairly well known and can be easily visualized in a plant, but the cause of these symptoms are not as well known on a biochemical and molecular level (Bruce et al., 2002). One of the first responses to water deficit is osmotic adjustment (Leung and Giraudat, 1998). Plants increase their osmotic potential by accumulating osmolytes such as proline, glycine betaine and non-reducing sugars and polyols. Osmolyte accumulation allows additional water to be taken up from the environment, buffering the immediate effect of decreased water within the organism. Osmolytes increase thermodynamic stability of folded proteins and provide protection against denaturing stresses (Bolen and Wang, 1997).

The accumulation of soluble sugars such as sucrose, glucose and fructose is often associated with drought tolerance in plants. It is thought that soluble sugars help protect the plant in times of drought in one of two ways, first, the hydroxyl group of sugars may take the place of water molecules helping maintain hydrophilic interactions in membranes and proteins by hydrogen bonding with them (Williams and Leopold, 1989). Second, the sugars form a highly viscous layer in the cytoplasm of the dehydrated cells

reducing the molecular movement and help maintain the structural stability of the macromolecules (Sun and Leopold, 1997). The true mechanism is still unknown. Several genes involved in the metabolism of these sugars have been found to increase the ability of different transgenic plants to resist stress (Shinozaki and Yamaguchi-Shinozaki, 2007). Over-expression of drought-inducible galactinol synthetase gene (*AtGolS2)* in *Arabidopsis* yielded increased drought resistance because of the excess accumulation of galactinol and raffinose (Shinozaki and Yamaguchi-Shinozaki, 2007). Additionally, transgenic plants that express the DREB1A/C- repeat binding factor 3 (CFB3) accumulate more galactinol and raffinose than wild type plants, helping them improve their ability to withstand drought (Avonce et al., 2004).

In addition to the sugars mentioned above, trehalose is often accumulated in large quantities in response to desiccation. Trehalose is an α-linked disaccharide found in nature from a wide variety of sources such as bacteria, fungi, yeast, plants and invertebrate animals (Elbein, 1974). Trehalose is also associated with heat and salt stress related responses (Thevelein, 1996). The high water retention capacity of trehalose is thought to help stabilize proteins and membranes enabling the cells to recover when rehydration conditions are present. Tobacco has been transformed with the trehalose phosphorylase (*TP*) gene from *Pleurotus sajor-caju* and showed an increased ability to withstand drought. The trehaolase synthetase gene *(Tsase)* from *Grifola frondosa* also yielded an increase in the ability to withstand drought in tobacco (Han et al., 2005).

Study of accumulated metabolites may shed some light on the process plants use to cope with increasing drought conditions. A systematic and quantitative survey of a

large number of metabolites in a time course study undergoing severe desiccation treatment should provide clues of the inner-working machinery in drought resistance.

Materials and Methods:

Plants

Aquilegia F1 Cv. Origami Mix seeds were obtained from Harris Seed Company in Rochester, NY. Seeds were rinsed with distilled water then germinated in a Petri dish containing moist filter paper at room temperature. After germination, seeds were placed in moist germination soil in a covered growth flat under constant light. When seeds had developed a root system, they were moved to larger 18 section flats with more growth area. Seedlings were then placed in a growth room for 6 months to build tissue mass. Temperature was kept at an average of 20.5°C (21°C max, 20°C min), with 12 hours of artificial light. Plants were watered to saturation at two day intervals. When plants had outgrown their flats they were transplanted into larger one gallon containers. *Aquilegia* plants were then transferred to the greenhouse to adjust and establish root systems and build tissue mass. At around six months the gallon pots containing the *Aquilegia* plants were moved from the greenhouse to a growth room with better controlled environment at 20.5°C (21°C max, 20°C min). The *Aquilegia* plants were kept in the growth room for a week prior to stress treatments.

Before tissue was collected all plants were watered to saturation on Day 0 with tissue samples being taken the following day (day one). Dehydrated *Aquilegia* tissue was pooled from 18 different plants to represent the average drought response. Tissue

samples were taken on days 1, 3, 5, 7, 9, 11, and 14 from stems and leaves separately, and then frozen in liquid nitrogen immediately and stored at -80^{0}C.

Chemicals

All chemicals were purchase from Sigma. Chemical standard (vitexin) was purchased from Sigma-Aldrich (St. Louis, MO). The vitexin standard was solubilized in 80% methanol; a serial dilution in methanol was made to construct a standard curve. Vitexin was also subjected to LC-MS for fragmentation patterns. To do this, ammonium hydroxide was added for negative ion mode acquisition of MS spectra.

Extraction of Methanol-Soluble Polar Compounds

The frozen tissue (0.5g) collected from Day 1, 3, 5, 7, 9, 11, and 14, samples were removed from -80^{0}C and were homogenized using mortar and pestle under liquid nitrogen to powder-like material. Ground leaf tissue was added to 9.8 mL of extraction buffer (80% methanol, 20% water and 100ul of 2% DMSO, and 100 µl of berberine (500 µg/ml)) in a 50mL tube, and then homogenized using a tissue homogenizer (PowerGen 125, Fisher Scientific) for 60 seconds followed by sonication (Bransonic 32, Bransonic Cleaning Equipment, Shelton, CN) for 15 minutes at room temperature. The sample was vacuum filtered through a 0.2 µm filter and the filtrate was transferred into a 1.5 mL micro centrifuge tube and spun at 13,000 rpm for 3 minutes. The supernatant was collected and aliquoted evenly into micro centrifuge tubes (500 µL) and placed in a speed vacuum (Lab Conco, Kansas City, MO) to dry completely. Dried samples were

resuspended in 150 µL 50% methanol in water (each tube) and spun at 13,000 rpm for 3 minutes. The top 100 µL of supernatant was collected and combined before injection into HPLC for separation and quantitative analysis.

HPLC Analysis

Methanol extracts of *Aquilegia* were analyzed by reverse-phase HPLC (717 plus Autosampler, 1525 Binary HPLC Pump, and 2487 Dual wavelength detector, Waters). Separation used a Symmetry 4.6 x 150mm reverse phase column (Waters). Absorbance was measured by duel wavelength at 214nm and 280nm. The LC chromatogram was generated using linear gradient (1 minute at 95% A: 5% B, 5 minutes at 80% A : 20% B, 35 minutes at 50%A : 50%B, and 36 minutes at 5%A: 95%B). Solution A was 0.1M acetic acid/ammonium acetate. Solution B was 90% methanol, 10% acetonitrile. Each sample was run on the HPLC in triplicate with 50 µl injection volume. Data was acquired and processed (Breeze, version 3.30). Fractions were collected automatically (Waters Fraction Collector III) in 30 second intervals in 96 well microtiter plates (ABgene, Epsom, UK) for bioassay. Fractions of interest were collected manually in 1.5 mL centrifuge tubes and dried for identification by LC-MS and MS/MS.

Quantification of Compounds

Berberine was used as an internal control for normalizing variations in sample preparation and injection. Relative levels of metabolites were determined by integrating the peak Area Under the Curve (AUC) of the chromatogram acquired from the Breeze

software. AUCs from three injections were averaged and plotted in Excel spreadsheet for visualization of relative peak levels over the period tested.

Quantification of Vitexin

Vitexin was solubilized in 80% methanol and serial dilutions were made. These were injected in triplicate into HPLC using the same protocol as for the *Aquilegia* metabalome extracts. The points were averaged and the AUC was calculated in the Breeze software (Waters). This information was then placed in Microsoft excel for analysis and creation of a standard curve representing AUC vs. concentration of vitexin present. Based on this standard, all peak 23s from HPLC runs for each time point were measured and quantified.

Crystallization of Fractions

In an attempt to identify the structure of compounds of interest that accumulated in response to drought, fractions containing fractions of interest were collected individually from repeated HPLC runs, pooled and dried in a Speed Vac. They were then re-solubilized in various solvents, placed in glass screw-top vials from Waters, and placed at 4^{o}C with a small opening in caps to allow solvent to evaporate slowly.

Liquid Chromatography-Mass Spectroscopy

Fractions of interest from HPLC were further analyzed by a nanoflow capillary HPLC coupled to a ESI-tandem mass spectrometer (Q-TOF-micro; Waters) Operation of instrumentation used MassLynx software (V4.0, Waters).

All compounds were registered in both positive and negative ion mode to cover broad spectrum. For positive ion mode, formic acid was added to protonate ions (0.1% final). In negative ion mode, ammonium hydroxide was added to deprotonate ions (1% final). Fractions were injected into ion source directly at a flow rate of 1.0µl /minute for both positive and negative ion mode acquisition. The ion source voltages were at ±3000V, sample cone was at 35V, the extraction cone was at 3V, and source temperature was at 100ºC. The quadrupole was scanned from 100 to 800 m/z at 1s with 0.1s inter-scan delay in continuum mode, and the nebulizer gas (nitrogen) flow was at 30L/h For MS/MS product ion acquisition. TOF was scanned from 50-800m/z. To calibrate mass real-time infusion, Sodium formate was used for the accurate mass calibration.

Semi- Quantitative Real Time PCR

For Real Time PCR, tissue samples were collected every 12 hours over a 6 Day period with no water supplied. Samples were taken between 8:00 and 9:00 AM and 12 hours later between 8:00 and 9:00 PM. PolyA RNA enrichment was performed on total RNA sample using manufacture's protocol (MicroPoly A Purist™ Kit, Ambion, Hilden, Germany). cDNA was generated from the polyA RNA using VERSO™ Reverse Transcriptase in conjunction with Thermo's (Walthan, MA) SYBR® Green 1-Step QRT-

PCR Low ROX Kit according to the manufacturer's instructions. Each sample was run in triplicate for each of the primers pairs. Primers for both actin and UDP-Glucose: vitexin 2"-O-β-D-glucosyltransferase, used for RT-PCR were designed based on sequences obtained from a Subtractive Hybridization Library created from stressed tissues isolated from *Aquilegia*. The primer sequences are shown in **Table 3.1**. To determine the specificity of the primers before RTPCR, genomic DNA obtained from *Aquilegia* was subjected to PCR reaction with expression of *actin* as control for variation in sample preparation and concentration. PCR conditions used cDNA synthesis at 50° C for 15 min, Thermo-Start activation at 95° C for 15 min, and denaturation at 95° C for 15 min, 25 cycles of; (95° C for 15 sec, 52° C for 30 sec, and 60 °C for 60 sec). Data from the replicated reactions were examined to determine the transcript level of the particular genes of interest.

Table 3.1: Primer sequences of glucosyltransferase and actin to determine steady state transcripts level. RT PCT performed to determine if there was an increase in transcript level in response to drought in relation to actin for normalization.

vitexin 2"-O-β-D-glucosyltransferase	5'TCAGTAGCTTCTTCTGCTTTTTGTT 3' 5'TGTACTTCTTCTATACGGGGAATTG 3'
Actin	5' GTATAGTAAGCAACTGGGACGACA 3' 5' GGATAGCAACATACATAGCAGGAG 3'

Spore Germination and Appressorium Development

Conidia were harvested from a 12-day-old culture of *Magnaporthe grisea*, growing on V8 agar (10% (v/v) V8, X (w/v) $CaCO_3$, 1.5% agar) at room temperature to make a spore

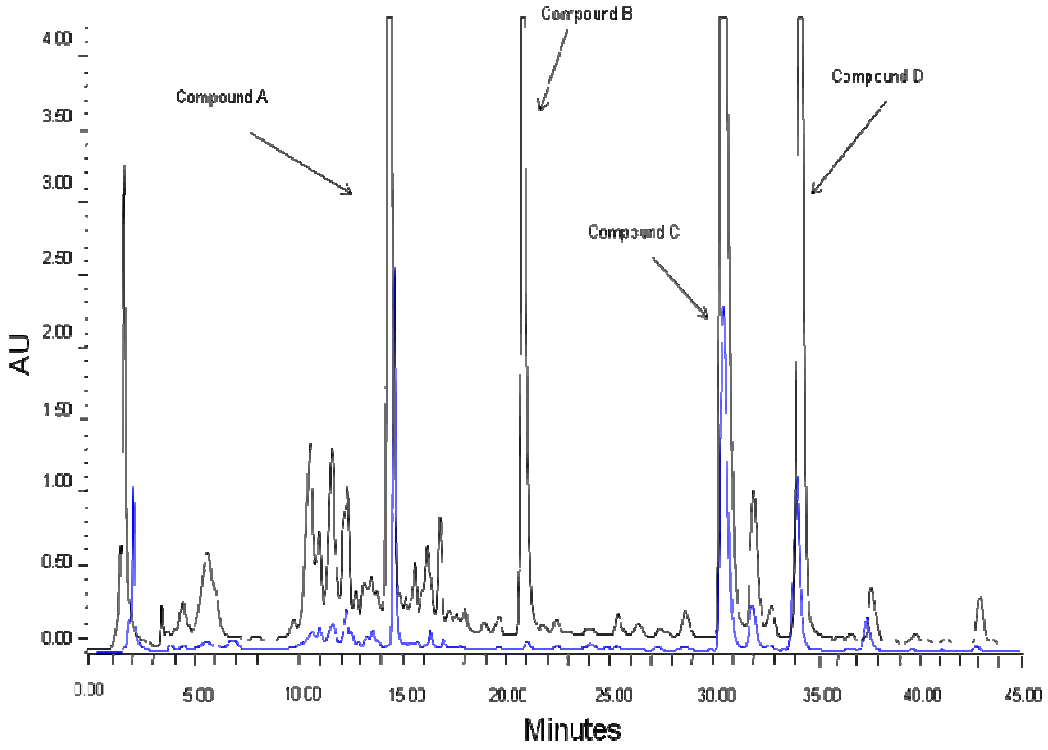
suspension of 2 x 10^6 conidia/ml. Droplets of 20 µl of the spore suspension were placed on glass slides containing 20 µl of vitexin of different concentrations, including 500 µM, 250 µM, 100 µM, 50 µM, 10 µM and distilled water, used as the negative control. All treatments were covered with cover glasses and incubated in a 100% moisture chamber at room temperature for 5 hr and 24 hr. Spore germination and the development of appressorium were recorded from each treatment by an inverted light microscope (Zeiss Axiovert 200 M). The metabolic activity of conidia from each treatment was estimated by MitoTracker Red (MT Red) fluorescence analysis.

Results and Discussion

The methanol extraction and subsequent UV-VIS HPLC analysis are sensitive to detect differentially expressed small molecules. Four metabolites increased significantly over the two-week period when *Aquilegia* plants were subjected to increasing drought conditions, while most metabolites remained stationary. Liquid chromatography procedure not only enabled relative quantification of the difference, but also purified compounds to baseline separation, thus allowing the integration of quantity.

An 80% methanol solution was found to yield the greatest number of compounds in the hydrophilic to moderate hydrophobic range by a linear gradient from aqueous to acetonitrile HPLC. We are mainly interested in hydrophilic metabolites based on their quantity and complexity in the chromatogram. The chromatograms were analyzed so that individual peaks and retention times were recorded and integrated as the Area Under the Curve (AUC). The AUC of 38 selected peaks were recorded and placed in a Microsoft

Excel spreadsheet for further analysis. A more in depth analysis using different extraction solvents, such as chloroform, warm water or methylene chloride, should lead to the extraction of additional compound not studied here. Of the 38 peaks that were given a relative quantification over the time course of our experiment, four peaks were found to be present at higher levels as drought conditions progressed. They were identified tentatively as peaks 13, 23, 30 and 33, having retention time of 13.9 minutes, 23.15 minutes, 32.38 minutes, and 34.79 minutes, respectively (**Figure 3.1)**.



Separation of small metabolytes on reversed-phase HPLC. Solvent A (0.1M ammonium acetate/acetic acid, pH4), Solvent B (90% acetonitrile, 10% methanol), Flow rate 1.0 ml/min, Detection: UV 280nm. Column: Symmetry® C18, 5µm, 4.5x150 mm, Waters, Instruments: Waters 1525 (pump), 717 (autosampler), 2487 (dual UV detector), Breeze 3.2V (software). Day 1 in blue, Day 14 in Black.

Figure 3.1: HPLC profile of Aquilegia using methanol extraction. Obtained from leaf tissue under increased drought conditions. Black trace is drought stressed tissue, Blue trace is control tissue.

The individual fractions were collected on at least three runs from drought-stressed tissues for calculation of standard deviation of the quantity. **Table 3.2** shows the corresponding fraction number, retention time from HPLC chromatogram, monoisotopic mass (m/z), daughter ions from MS/MS scan in positive and negative ion mode, and identification (if known).

Table 3.2: Mass spectroscopy data from four compounds of significant accumulation over course of drought experiment Peaks identified as being up-regulated via HPLC analysis in drought conditions. Compounds were extracted via methanol extraction buffer from leaf tissue. Retention time, mono-isotopic mass, daughter ions in positive and negative ion mode and any identification information is shown.

Peak Identification	**Retention time (min)**	**Mass (Accurate mass)**	**Daughter ions (Positive ion mode)**	**Daughter ions, (Negative ion mode)**	**Name**	**CAS number**
Fraction 13, Compound A	13.9	594.53	595.16, 577.15, 559.14, 529.13, 499.12, 475.12, 445.11, 415.10,397.09, 379.08, 337.07, 313.07, 283.06	593.20, 551.33, 503.29, 473.16, 431.18, 385.16, 281.09, 269.11, 163.06	Vitexin 2”-O-β-D-glucoside	XxX
Fraction 23, Compound B	23.15	432.38	433.11, 415.08, 397.08, 379.04, 367.06, 337.06, 313.04, 283.05	431.09, 341.04, 311.03, 283.03, 269.2, 271.05	Apigenin 8-C-glucoside (Vitexin)	3681-93-4
Fraction 30, Compound C	32.37	463.2849	447.13, 441.30, 429.11, 371.09, 355.06, 304.25, 282.27, 240.15, 210.11	N/A	unknown	unknown
Fraction 33, Compound D	34.80	447.1179	430.12, 411.11, 393.08, 381.10, 351.08, 327.08, 297.07, 285.07 ,	N/A	unknown	unknown

In order to identify the compounds that accumulated under drought conditions, the dried pooled individual fractions were re-solubilized in 80% methanol and placed in individual glass vials, which were then capped. A small hole was pierced in the top to allow minimal air flow in and out, then maintained at 4ºC to allow the solvent to evaporate slowly at 4ºC. Fraction 23 formed small yellow tread-like crystals after 5 days which were allowed to continue to grow until most of solvent had evaporated. The identified compound formed a complex of yellow, needle-like crystals. This crystallization pattern was supported by information obtained from the Merck Index (Budavari, 1989). Compound A corresponding to fraction 13 formed a small brown cube, crystal but both of the resulting crystals were deemed too small for X-ray crystallography.

Crystallization of these molecules was difficult due to the highly hydroscopic nature of the compounds of interest. When exposed to atmospheric conditions the dried crystals would absorb water from the atmosphere disrupting any crystals that had formed. To alleviate this, once the small crystals formed, they were placed directly in the speed vacuum, allowed to dry and when removed, flooded with nitrogen gas and placed in an air tight vial. Even at this level, the structures were unstable and formed thick, supersaturated, viscous liquids when left at room temperature for an elongated amount of time. This was likely due to some residual water contained within the vial. This observation adds to the hypothesis that these molecules are vital in helping plants retain and cope with water loss, due to the high water retention capacity of the compounds.

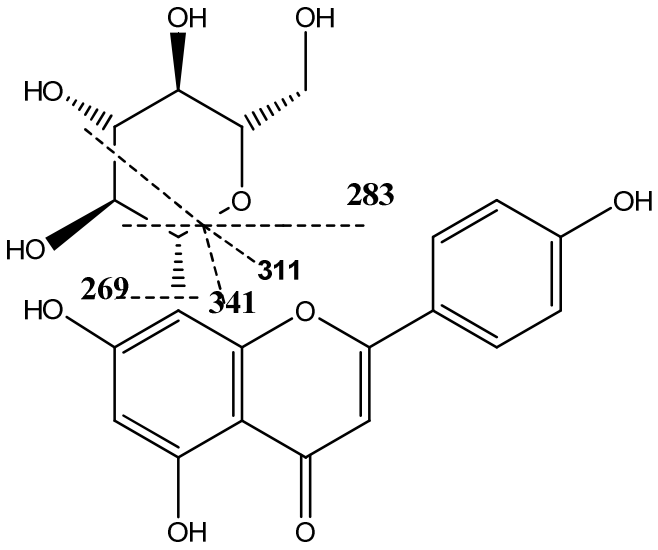


Figure 3.2: Molecular structure and fragmentation breakdown of apigenin 8-C-glucoside (vitexin), m/z 431

Fraction 23, with a retention time of 23.15 minutes, was identified as Apigenin 8-C-glucoside (Vitexin) (**Figure 3.2**) by identification of its accurate mass and MS/MS fragmentation pattern match to published results (March et al., 2006). For further verification, a known standard compound of vitexin (Sigma) was purchased and subjected to LC-MS in both positive and negative ion mode. The fragmentation pattern of the standard vitexin matched the mass and MS/MS fragmentation patterns of the corresponding fraction 23 from the *Aquilegia* extract (**Figure 3.3 and 3.4**). Previous studies have shown a Mass Spectrum for vitexin in negative ion mode which can be seen in **Figure 3.3**. The slight differences in spectrum can be caused by a difference in voltage in the cone, ion source, and extraction cone.

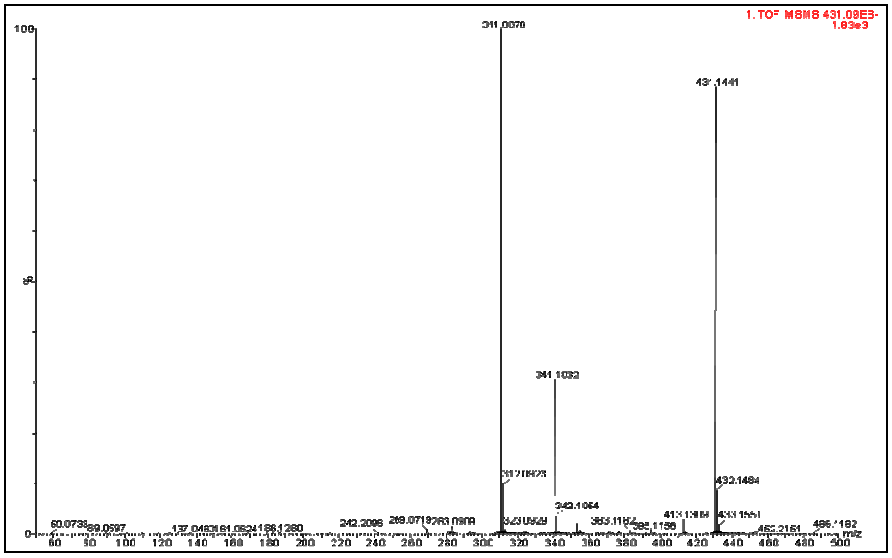


Figure 3.3: Mass spectra of Apigenin 8-C-glucoside, molecular weight 432, in negative ion mode obtained from *Aquilegia.*

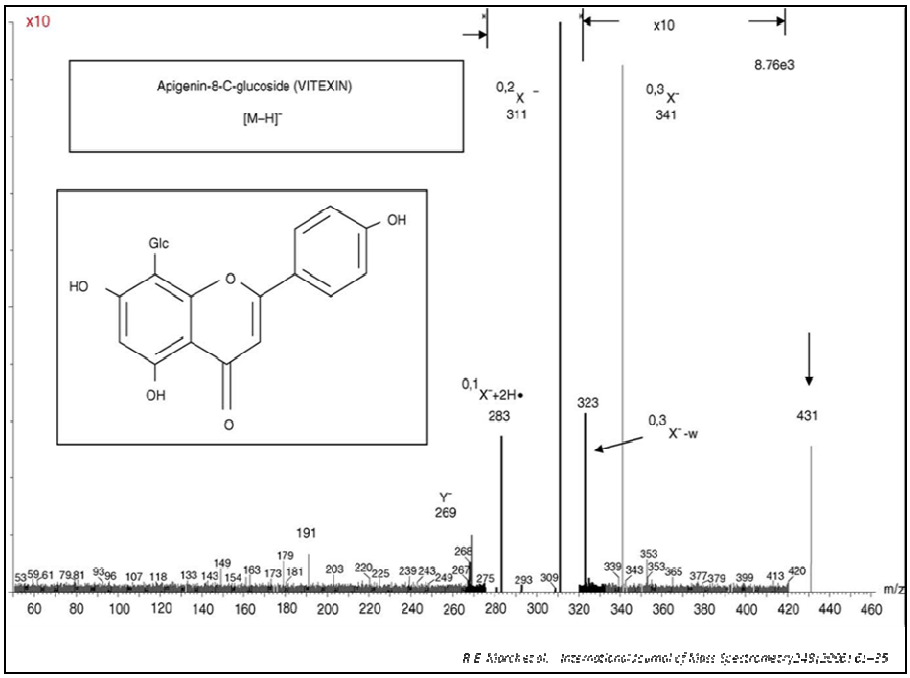


Figure 3.4: Mass spectra of Apigenin 8-C-glucoside, molecular weight 432, in negative ion mode. Spectra copied from (March, et al. 2006).

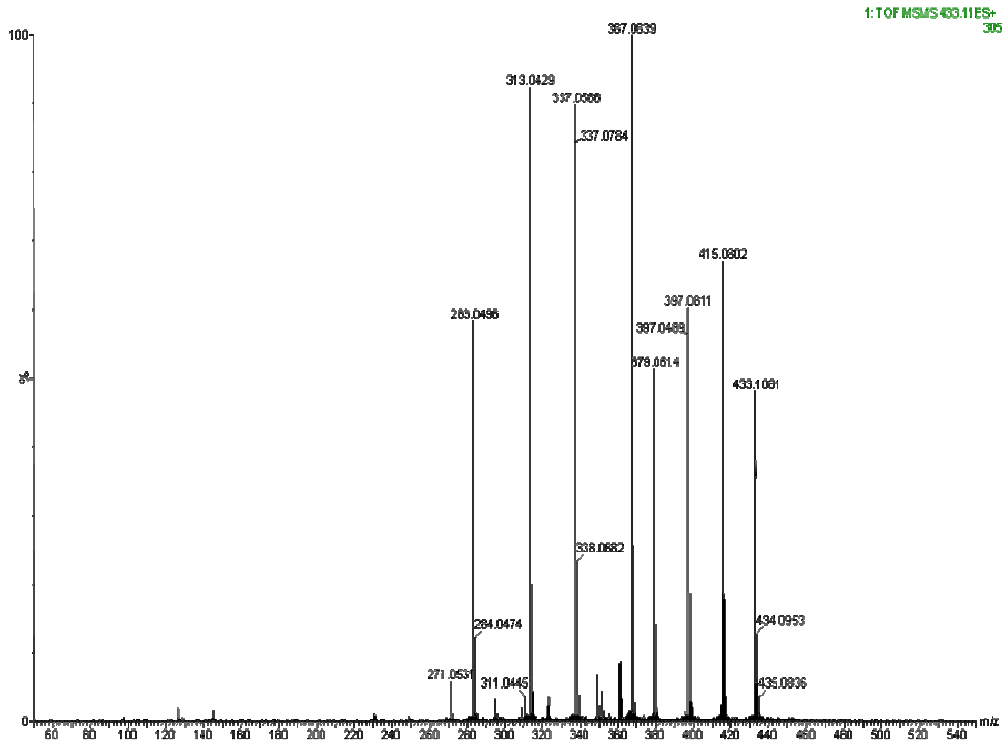


Figure 3.5: Mass spectra of Apigenin 8-C-glucoside, molecular weight 432, in positive ion mode.

A standard curve of vitexin was created by a series of dilutions run and separated by HPLC using the same protocol as for *Aquilegia*. This allowed for accurate quantification of fraction 23 present over the 2 week experimental span. The compound increased 357% from Day 1 to Day 14. At Day 14 the percentage of vitexin was 0.9% of the total weight of the leaf. **Figure 3.6** shows the increasing total percentage of vitexin (in weight) present per gram of leaf tissue on Day 1 through Day 14 as drought stress increased.

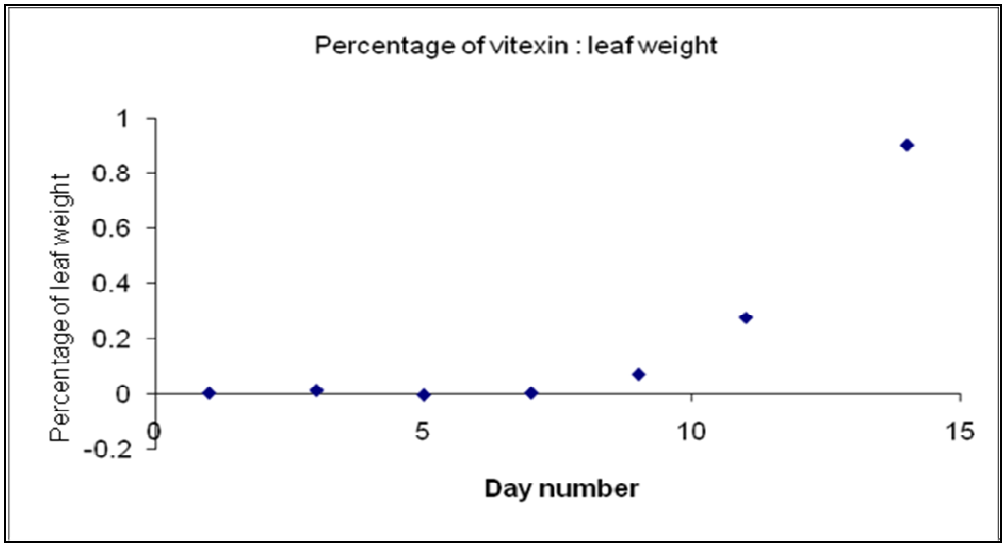


Figure 3.6: Content of vitexin in *Aquilegia* leaf tissue. Percentage of vitexin present in *Aquilegia* leaf tissue weight using a standard curve run with vitexin purchased from Sigma. Vitexin concentration was measured for each time point and quantified in percent of total tissue mass.

In addition to identification of fraction 23, fraction 13 with retention time of 13.9 minutes was identified as vitexin 2-O-β-D-glucoside (**Figure 3.7**). Based on the HPLC data Vitexin 2”-O-β-D-glucoside is present at higher levels throughout the life of the tissue. This is possibly because it is the end product of the vitexin to vitexin 2”-O-β-D-glucoside pathway (Heinsbroek et al., 1980), with the precursor vitexin being amassed in preparation for enzymatic synthesis of final product in response to increasing dehydration conditions. The fragmentation pattern in both negative and positive ion mode can be seen in **Figures 3.8 and 3.9**.

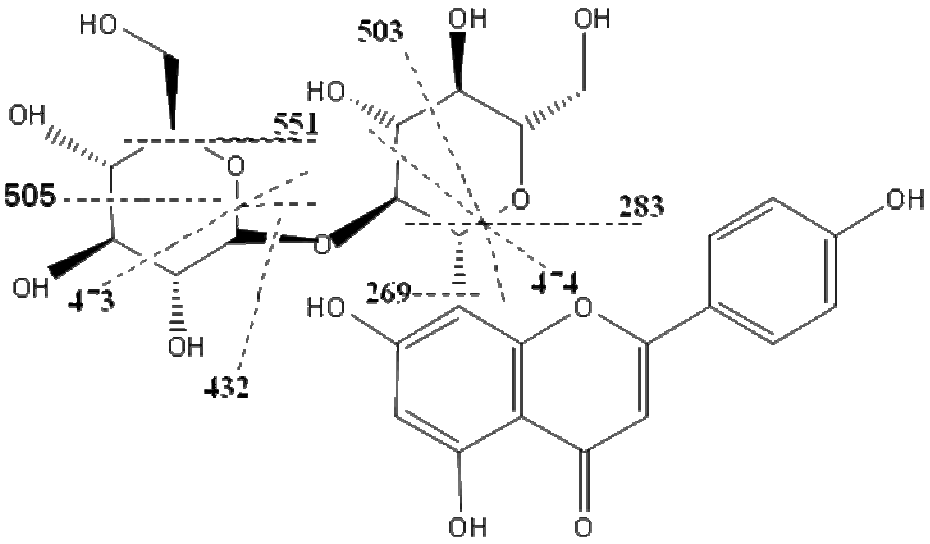


Figure 3.7: Molecular structure and fragmentation scheme [M-H] Vitexin 2”-O-β-D-glucoside, m/z 593

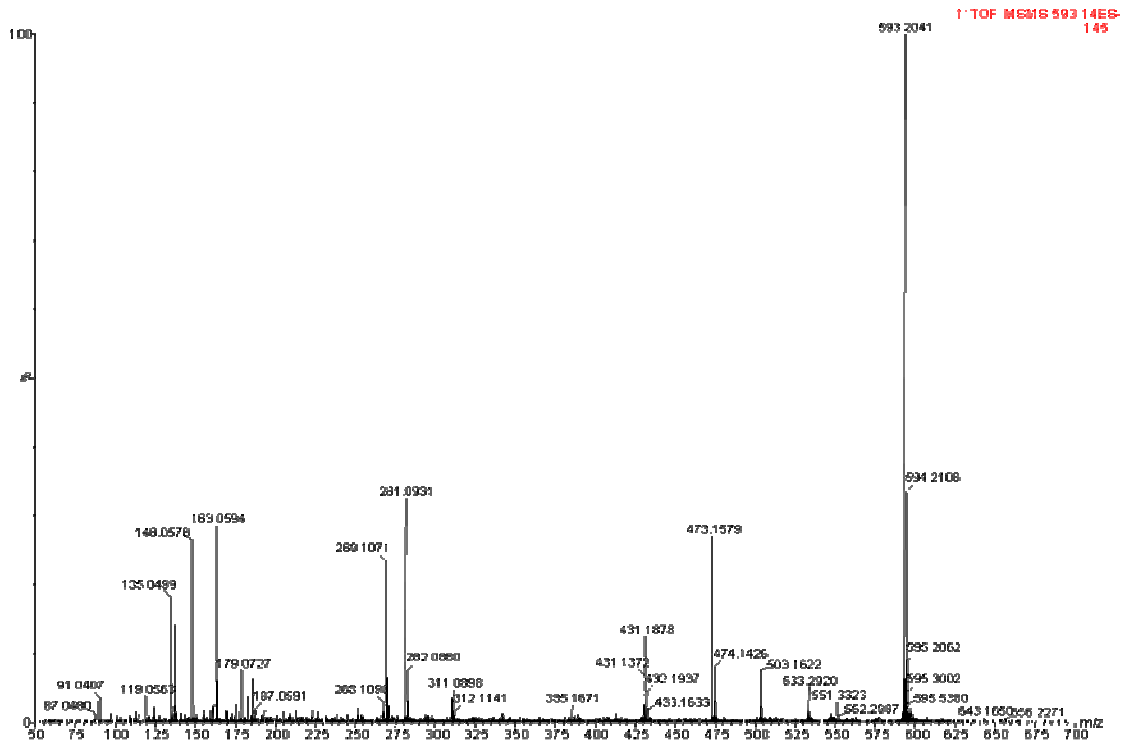


Figure 3.8: Mass spectra of vitexin 2”-O-β-D-glucoside, molecular weight 594, negative ion mode

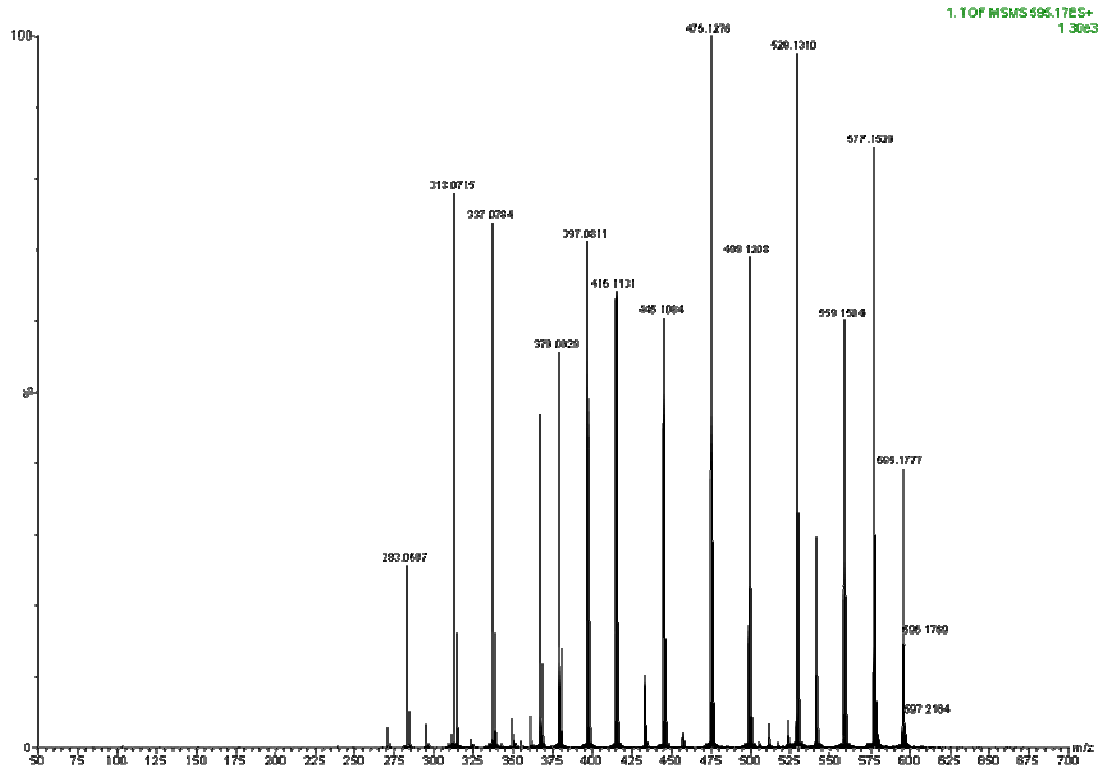


Figure 3.9: Mass spectra of Vitexin 2”-O-β-D-glucoside, m/z 594, in positive ion mode.

Interestingly, both of these metabolites are associated with the same pathway, a vitexin β-glucosyltransferase reaction (**Figure 3.10**). Apigenin, being the precursor to vitexin, is catalyzed by an unknown enzyme. Vitexin being the precursor to Vitexin 2”-O-β-D-glucoside is catalyzed by the enzyme UDP-Glucose: vitexin 2”-O-β-D-glucosyltransferase (Heilsbroek et al., 1980). Real Time PCR of drought stressed mRNA obtained from *Aquilegia* was performed on a vitexin–β-glucosyltransferase homolog, which catalyzes the reaction Apigenin 8-C-glucoside (vitexin) to vitexin 2”-O-β-D-glucoside. The partial gene sequence was obtained from an *Aquilegia* drought stressed subtractive hybridization library as one of the differentially expressed genes in response to drought. This resulted in the obvious upregulation of our particular gene of interest when compared to the control transcript actin (**Figure 3.11**).

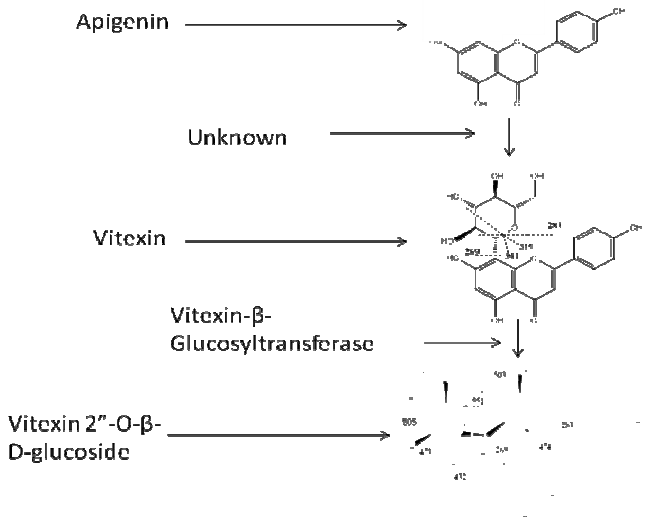


Figure 3.10: Schematic of metabolic pathway of apigenin to vitexin 2'-O-β-D-glucosice.

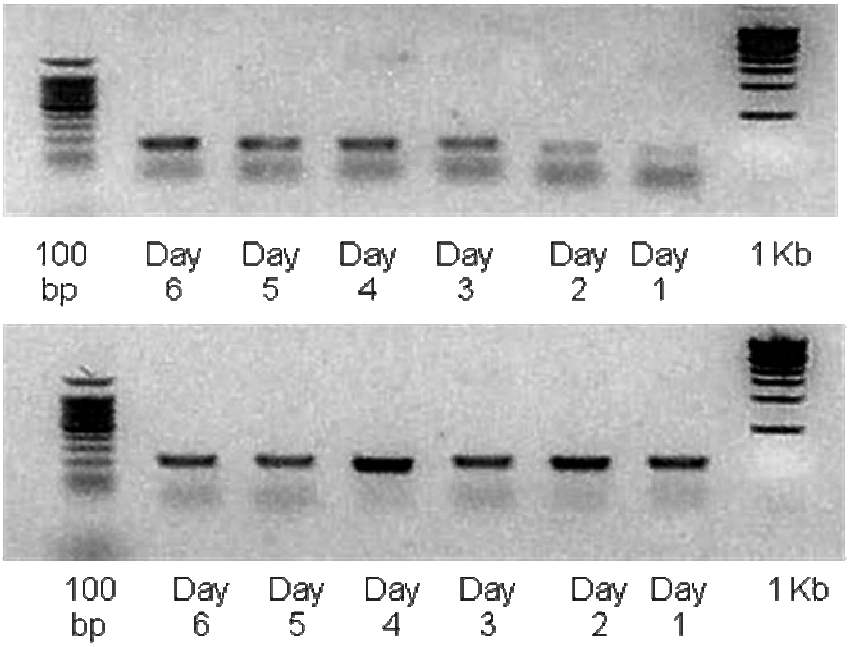


Figure 3.11: Semi-quantitative RTPCR *Aquilegia* UDP-Glucose: vitexin 2"-O-β-D-glucosyltransferase homolog (top). The increase of transcript level shows that there is an increase in gene number as drought increases. Expression of constitutively expressed actin is seen in bottom window.

Based on findings from both our subtractive hybridization libraries (Chapter two) and metabolite study, we conclude that the enzymatic reaction that yields vitexin β glucosyltransferase plays a vital role in *Aquilegia's* drought defense. The accumulation of sugars and osmolytes in plants when subjected to drought-like conditions has been known for some time, but vitexin and Vitexin 2"-O-β-D-glucoside have yet to be associated with drought resistance. Vitexin 2"-O-β-D-glucoside's suger moiety closely resembles trehalose, a known and highly studied osmolyte. Gene pathways associated with Apigenin 8-C-glucoside (Vitexin) and Vitexin 2"-O-β-D-glucoside could be valuable gene overexpression candidates for increasing drought tolerance in plants (Li et al. 2007).

Accumulation of soluble sugars in plant tissue in reponse to drought has been documented as a defensive response to drought-like conditions. For example, pea cultivars with highest sugar content had the highest osmotic adjustment (OA) capacity (Ayerbe et al., 1998). Also tobacco (Pilon-Smits et al., 1995) and sugar beet (*Beta vulgaris*) (Goddijn and van Dun,1999) transformed with *SacB* gene, a levansucrase, from *Bacillus subtillis* yielded increased drought tolerant plants.

Previous studies have shown that vitexin posseses antioxidant activity in cell culture (Pyo et al., 2005). This is important because reactive oxygen species (ROS) are known to damage cellular components. A good source of ROS inhibitors and free radical scavengers are plants. For example, green tea (*Camellia sinensis*) has been examined extensivly for its polyphenols. Metabolites such as epigallocatechin-3-gallate has been

shown to inhibit ultraviolet radiation-induced oxidate stress on human skin (Katiyar et al., 2000).

To test the biological activity of vitexin on fungus, *Magnaporthe grisea* was utilized. *M. grisea* was surveyed for any inhibition or delay of spore germination. As shown in **Figure 3.12 and Figure 3.13,** high concentrations, 500 µM and 250 µM, of Vitexin caused delayed spore germination and further differentiation for appressorium. Reduced number of germinated conidia and short (<10 µm) germination tubes were observed from these two treatments at 5 hr post-treatment. The appressoriums are used by parasitic fungi to enter host organisms by utilization of turgor pressure to penetrate host cells. From the images of MT Red fluorescence analysis for the metabolic activity, there was high activity at the tapered end of germinating conidia where the germ tubes initiated and the tip of the germ tube that was at early stage of differentiation for appressorium; whereas, the activity was absent from the treatments with 500 and 250 µM of vitexin. However, there is no significant suppression in both germination and differentiation after 24 hr post-treatment. The metabolic activity could be observed at conidia end, the knot of the germ tubes, and the mature appressoria from all treatments.

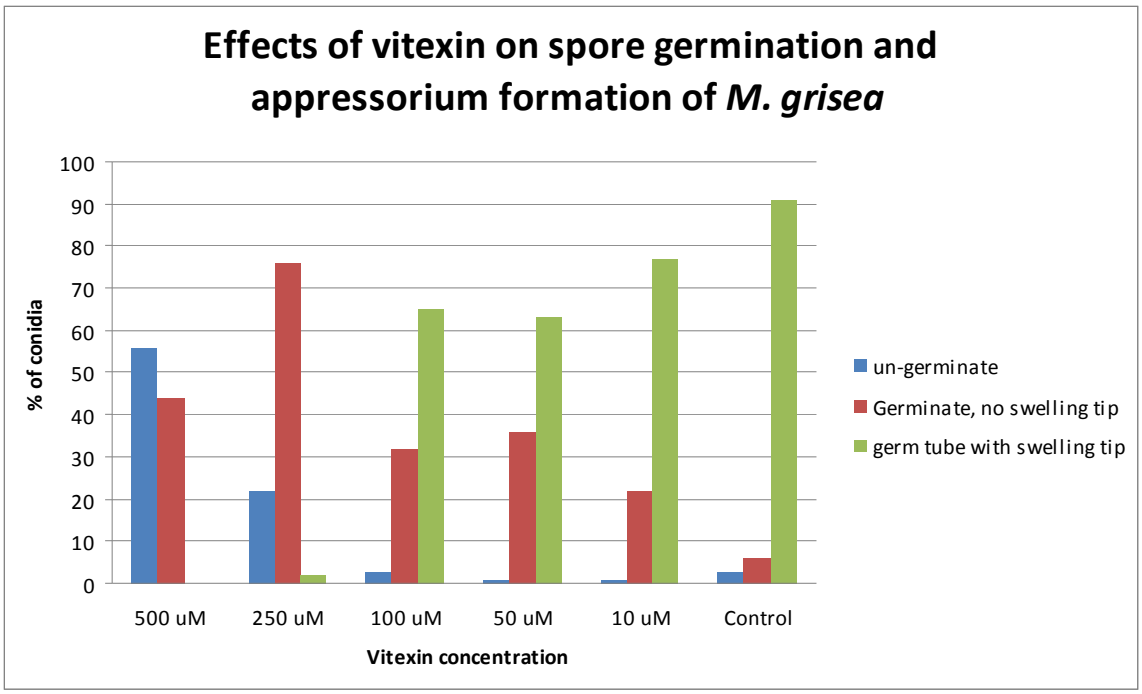


Figure 3.12: High concentration of vitexin suppressed spore germination at 5 hr post-inoculation. Spore suspension was treated with different concentrations of vitexin and incubated at room temperature for 5 hours. Germination rate and further differentiation of the germ tubes were then examined.

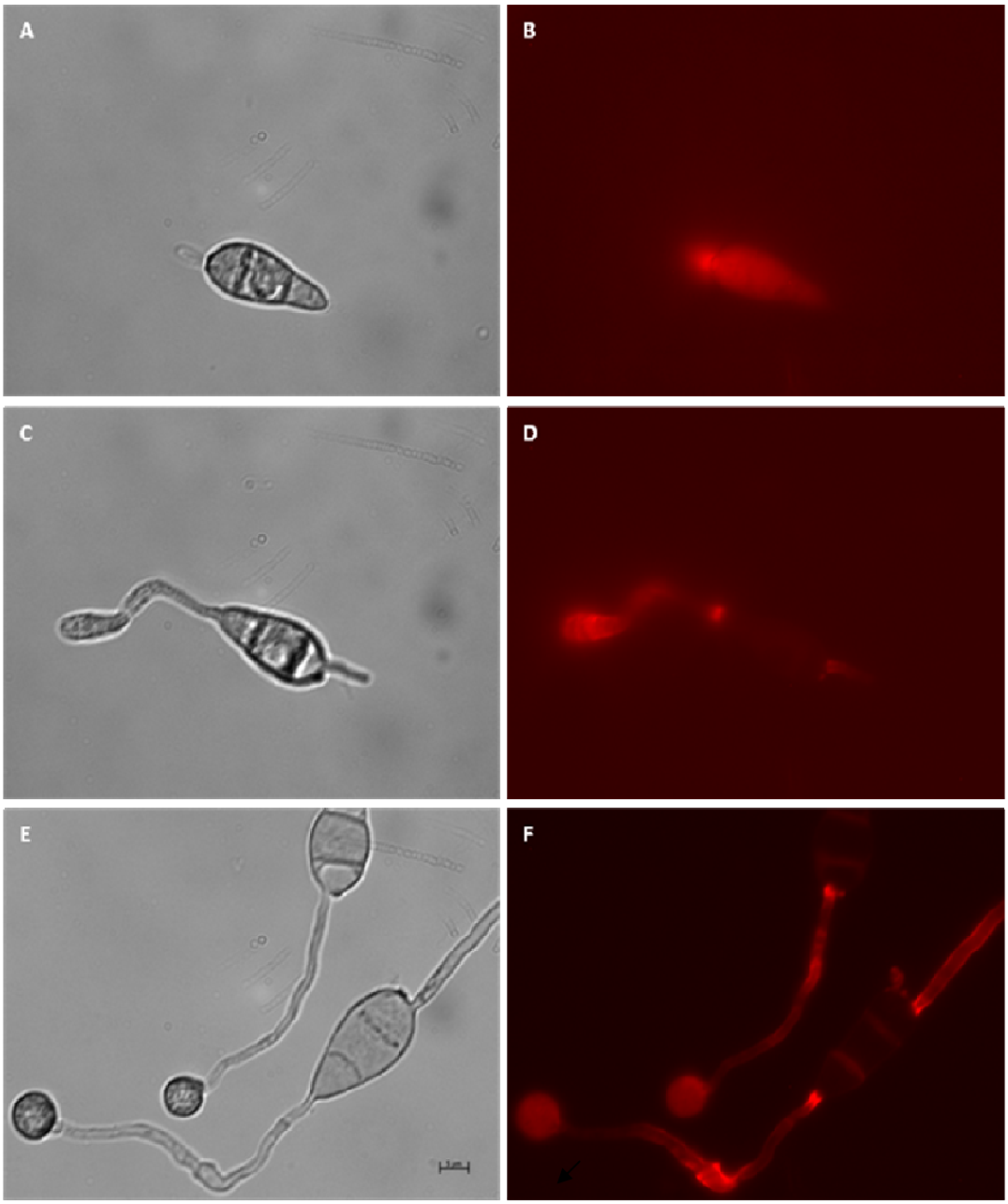


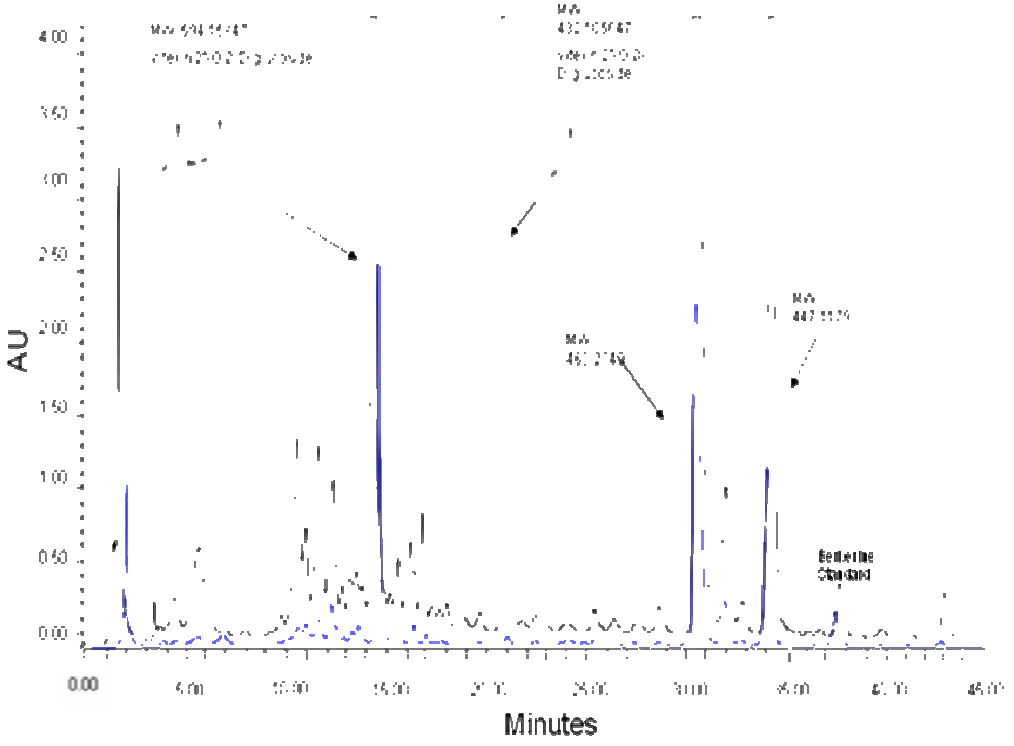
Figure 3.13: Vitexin delayed spore germination and appressorium development of *M. grisea*.
A and **B**, spore treated with 500 µM Vitexin at 5 hr post treatment. Only less than 5 um germ tube was visible. **C** and **D**, spore treated with dH2O at 5 hr. Extended germ tube (~30 um long) began to develop swelling tip. **E** and **F**, spores treated with 500 µM vitexin at 24 hr post treatment. mature melanized appressoria were developed. A, C and E were bright field microscopy and B, D and F were MT Red fluorescence images from A, C and E, respectively.

Conclusion

This study attempted to determine if any metabolites were accumulated in *Aquilegia* leaves under imposed drought conditions. Based on the methanol extraction, we found many different small molecules present in different levels over the time period of the experiment. The four compounds that accumulate the greatest were the focus of this study.

This study yielded the identification of two metabolites never known to be associated with drought. Different plants could be examined in the same manner as leaves were examined in this experiment to possibly identify other compounds associated with drought (**Figure 3.14**). Lastly, more studies need to be performed to determine the enzyme responsible for catalyzing the apigenin to vitexin reaction. Information gained from this study furthers understanding into plant defense to drought with the hope of being able to apply this knowledge to produce crop plants better suited to tolerate drought conditions.

Additionally, the biologic activity needs to be surveyed further. Results of the spore germination data presented show that vitexin possesses some inhibitory properties that could be commercially valuable.



Separation of small metabolytes on reversed-phase HPLC

Figure 3.14: HPLC scan with identified molecules and molecular weight of other accumulated metabolites

CHAPTER FOUR

PROTEOME ANALYSIS OF *AQUILEGIA* WITH INCREASED LEVELS OF DROUGHT

Introduction

Specific research is necessary to gain a deeper knowledge of the molecular factors involved in plant drought response in order to engineer crop cultivars with improved drought tolerance. The use of comparative proteome analysis in determining possible proteins involved in the defensive mechanism of plants has important implications in development of drought tolerant transgenic crops. The goal of this study is to compare proteins present in stressed *Aquilegia* to unstressed *Aquilegia.* Dynamic changes in protein levels, both qualitatively and quantitatively, reflect the mobilization of important players directly involved in drought responses. Proteins of interest will be investigated as related to results from genomic and metabalomics data obtained from *Aquilegia.*

Materials and Methods

Preparation of Plant Material

Aquilegia Cv. F1 Origami Mix seeds were obtained from Harris Seed Company in Rochester, NY. Seeds were rinsed with distilled water then germinated in a Petri dish containing moist filter paper at room temperature. After germination, seeds were placed in moist germination soil in a covered growth flat under constant light. When seedlings had developed a root system, they were transplanted to larger, 18 section flats. Flats were placed in a growth chamber and temperature was kept at an average of 20.5°C (21°C

max, 20°C min), with 12 hours of artificial light. Plants were watered at two day intervals. Plants were then placed in a greenhouse and transferred to one gallon containers to keep roots from becoming root bound in the smaller containers and allowed to grow for six months before harvesting of tissue.

Stress Treatment - Induction of Drought

At around 7 months *Aquilegia* plants were subjected to increased periods of drought. This was performed by watering plants on Day 0. Water was withheld from thereon until the end of the sampling time to provide the drought stress needed to stimulate a response in tissues over the time points from Day1 to Day 6. Tissue collected on Day 1 and Day 6 was weighed to obtain the total fresh weight. The tissue collected at each sampling time was quickly frozen in liquid nitrogen then transferred to -80ºC for storage before usage.

Extraction of Soluble Proteins

Day 1 and Day 6 tissue, frozen at -80ºC, was weighed (10grams) and then ground in mortar and pestle under liquid nitrogen into a fine powder. Ground tissue was then transferred to individual 50 mL conical tubes containing extraction buffer (at a ratio of 1 g tissue / 2 ml buffer) consisting of 1mM $NaHO_4$, 1.5 mM NaH_2PO_4, 10 mM KCl, 0.2 mM EDTA, 0.1mM PMSF, 0.2 mM Thiourea, 1.5% insoluble PVPP, and protease inhibitor tablet (Sigma *FAST*/100mL). Plant powder was homogenized by a polytron tissue homogenizer (OMNI international) to break the cell walls for 30-60 seconds until a

uniform consistence paste was achieved. The sample was then extracted for proteins for one hour on shaker at 4 ºC. The resulting slurry was centrifuged (Beckman Allegra X-15R) for 10 min at 3,000 rpm to remove cell debris. The supernatant was poured into a new 50 mL conical tube and proteins were differentially precipitated using increasing percentages of ammonium sulfate. First, ammonium sulfate was added slowly up to 20%, and the sample was allowed to equilibrate for 15 minutes at 4 ºC. The tubes were then centrifuged at 3,000 rpm for 10 minutes to remove precipitated residue. The supernatant was poured into a new conical tube and ammonium sulfate was gradually added in increments to the final concentration of 80% saturation while the conical tubes were shaking at 4 ºC. The sample was then placed in a centrifuge for 15 min at 4,400 rpm to concentrate remaining proteins. The supernatant was decanted and remaining pellet was resuspended into three mL of distilled water and vortexed until the pellet was fully solubilized. The sample was then dialyzed in 12 – 14 kDalton molecular weight cutoff dialysis tubing, against four liters of 20 mM Tris pH 6.0 for 6 hours at 4 ºC. The buffer was changed to a fresh 4 liters of dialysis buffer and allowed to dialyze overnight. The following day, samples were removed and aliquoted into 1.5 mL microcentrifuge tubes. These were lyophilized (frozen while concentrating) in a Speed Vac (Lab Conco, Kansas City, MO) to a small volume. This was the total soluble protein extract to be used for comparison study among the different time points.

Protein Separation by Gel Electrophoresis

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE), 1 Dimensional (1D) and 2 Dimensional (2D), was utilized to separate proteins based on their molecular weight and charge. For 1D-gel, the protein samples were solubilized in 30 µl of sample loading buffer (100 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 10% 2-mercaptoethanol, 0.002% bromophenol blue) and heated to 95ºC for 4 minutes. A 10 lane, 12% acrylamide separating gel with a 5% stacking gel was used for protein separation. Total protein extracts were loaded in individual wells of the gel and electrophoresis occurred in a Mini-Protein II Cell (Bio-Rad, U.S.A). This electrophoresis was run at 200 volts until dye front reached the bottom of gel (about 1.2 hours.). Protein markers were loaded along with protein extract for size references and to assure gel electrophoresis occurred properly.

For 2D-gels, the procedure used a modified protocol of O'Farrell (1975). For separation in the first dimension (isoelectric focusing), protein pellets were solubilized in 450µl rehydration buffer (8 M urea, 2% CHAPS, 50 mM DTT, 0.2% Bio-lyte ampholytes), vortexed, and centrifuged for 3 minutes at 10,000 rpm to remove any insoluble proteins. Supernatant was loaded onto immobilized pH gradient (IPG strips pH 3-10, BioRad, U.S.A) with gel side facing down in Isoelectric Focusing (IEF) chamber. One mL of mineral oil was used to cover the chamber to prevent evaporation. Strips were allowed to rehydrate the IPG strip for 12 hours at 10 ºC at 50 volts (active rehydration). Following active rehydration, isoelectric focusing was carried out at, 250 V, rapid ramp for 1 hour, increased to 10,000 V 2.0 hours and held at 10,000 V for

16 hours, at a maximum of 50µA for two strips for duration of isoelectric focusing. After isoelectric focusing the IPG strips were reduced and alkylated with equilibration buffer 1, (6 M urea, 0.375 M Tris-HCL, pH 8.8, 2% SDS, 20% glycerol, 2% w/v DTT), for 20 min. This was followed by equilibration with equilibration buffer 2 (6 M urea, 0.375 M Tris-HCL, pH 8.8, 2% SDS, 20% glycerol, 2.5% (w/v) iodoacetamide), for 20 minutes in dark. The IPG strips were rinsed in 1X SDS running buffer to remove any residual equilibrium buffers.

Raised protein bands in the IPG strip, as well as the dye indicator were visually examined to assure the completion of focusing. Proteins in the IPG strips were then subjected to the second dimensional separation based on molecular weight. The IPG strips were placed on top of a slab SDS-PAGE gels (20cm x 20cm x 1mm, BioRad) made of 12% acrylamide separating gel, with a 5% stacking gel. Electrophoresis was performed at a constant 400 V until dye front reached end of the gel. When run was completed, the gel was removed and placed in 200 mL fixing solution, (20% [v/v] ethanol, 8% [w/v] acetic acid, 0.08% [w/v] Coomassie Brilliant Blue G-250) and stained for 2 hours. The gels were destained in distilled water with gentle agitation, with numerous changes of water and addition of paper towel on top to absorb the dye from gel. After destaining, gels were preserved by drying. Drying was carried out by placing gel in a 10% glycerol, 20% ethanol solution for 2 hours followed by placement between two sheets of drying film (BioRad), clamped into a drying rack, and allowed to air dry to completion for storage.

In-Gel Trypsin Digestion

In order to identify proteins present in a time point sample, an in-gel trypsin digestion was performed to digest all protein bands present on SDS/PAGE to complete the shotgun proteomics for comparative studies. These digestions were performed in accordance to reported methods by Rosenberg (2006). Briefly, protein bands covering the entire lane were excised 1.5 mm apart from top to bottom continuously. Each band was cut into 3 to 4 cubes, placed in an Eppendorf tube and destained to remove Coomassie dye in 100mM ammonium biocabonate, 50% acetonitrile. Following dye removal, proteins were reduced with 3 µl of 45mM DTT (20 minutes at 37 ºC), alkylated with 4 µl of 100 mM IAA for 20 minutes in dark. After removing all chemicals, gel pieces were semi-dried at 37ºC for 30 minutes in an incubator until no liquid was present and gel pieces were shrunk. Trypsin digestion was commenced by adding 4µl of 0.04 µg/µL trypsin to gel pieces, where trypsin, smaller than the pore size of gel matrix, entered the gel matrix through rehydration force. Enough 100 mM ammonium bicarbonate was used to cover gel pieces (~30 µl, dependent on volume of gel pieces). Digestions were incubated for 18 hours at 37 ºC. The supernatant recovered from digestion was then transferred individually to clean tubes and the gel pieces containing tryptic digested peptides were extracted twice with 100µl of 60% acetonitrile, 0.1% TFA. These were sonicated for 30 minutes and resulting peptide fractions were dried under vacuum to completion.

LCMS and Identification of Peptides and Shotgun Proteomics

Tryptic peptides were reconstituted in 50 µl of injection solution (50% methanol, 0.1% formic acid) and 4 µl injected into a capillary liquid chromatography (CapLC) system (Waters Corporation, Milford, MA) using an autosampler for high-throughput data acquisition. Peptides were partitioned by high performance liquid chromatography (HPLC) on a C18 reverse phase column (NanoEase C18, Waters) over a 2-40% acetonitrile gradient in 30 minutes. Liquid chromatography was coupled to the ion source of a tandem mass spectrometer (Q-Tof *micro*™, Waters) which enables a real time acquisition of peptide ions (retention time and m/z) and their de novo sequences using the DDA method (data-dependent acquisition). For each run, an external standard peptide (NaFormate; Sigma) was injected for real-time calibration of accurate masses. *De novo* sequencing of peptide ions and protein database search were performed by utilization of Proteinlynx Global Servers software (Version 2.1, Waters) (**Figure 4.1)** The interpretation of a peptide sequence is achieved through a series of statistical analysis, whose authenticity is reflected by the e-values that rank the probability of the event.

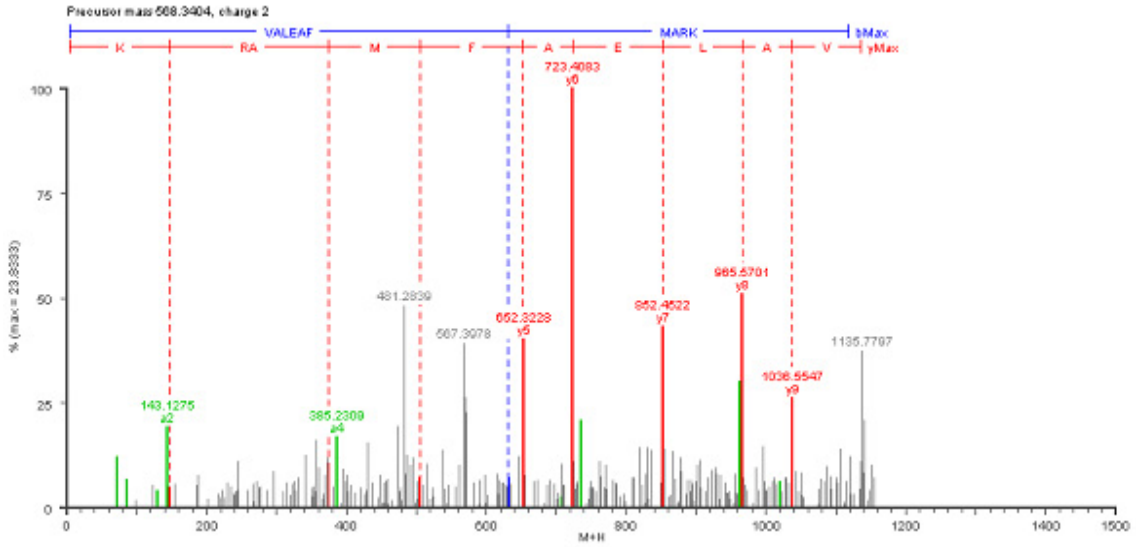


Figure 4.1: ProteinLynx Software
A raw mass spec data file by DDA method is processed by ProteinLynx that produces a peak list file composed of daughter ions, also known as fragment ions from parent ion 568.34^{2+}. Trypsin cleaves at basic residues, which produces more Y ions (red) than b ions (blue). Sequence read from b ions corresponding to N-terminal to C-terminal sequence, which agrees with Y-ions, read from C-terminal to N-terminal.

Verification of Proteomic Results with Genomic Tools

To identify possible matches of peptide sequences obtained from microsequencing of peptides derived from protein gel, protein sequences were first converted into a Fasta file format and assigned an arbitrary number for peptide name. TBLASTN of peptides sequences against AQGI.062608, the DFCI-Aquilgia Gene Index (AqGI), supported by the Dana Farber Cancer Institute, was performed with BLAST filtering turned off and word size set to 2. Peptide sequences produced significant homology to sequences found in the AqGI database. These proteins were manually annotated with the help of Blast2Go to confirm the fidelity of the match.

Results and Discussion

In this proteomic study, an overall glimpse into the proteome of *Aquilegia* under drought stress was examined as an independent verification to the results obtained from the genomic and metabalomic approaches. Total proteins were extracted from unstressed Day 1 tissue and subjected to 1 D-gel SDS-PAGE (**Figure 4.2**) and 2 D-gel SDS PAGE (**Figure 4.3**) separation.

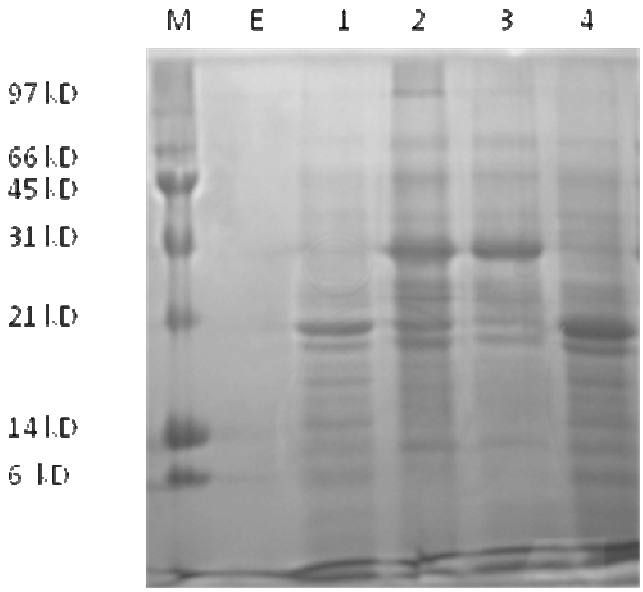


Figure 4.2: 1D gel of proteins from stressed and unstressed tissues. Lane M, Protein marker; Lane E, empty, Lane 3, Day 1 protein extract 40% ammonium sulfate; Lane 2, Day 1 protein extract 80% ammonium sulfate; Lane 3, Day 6 protein extract 80% ammonium sulfate; Lane 4, Day 6 protein extract 40% ammonium sulfate.

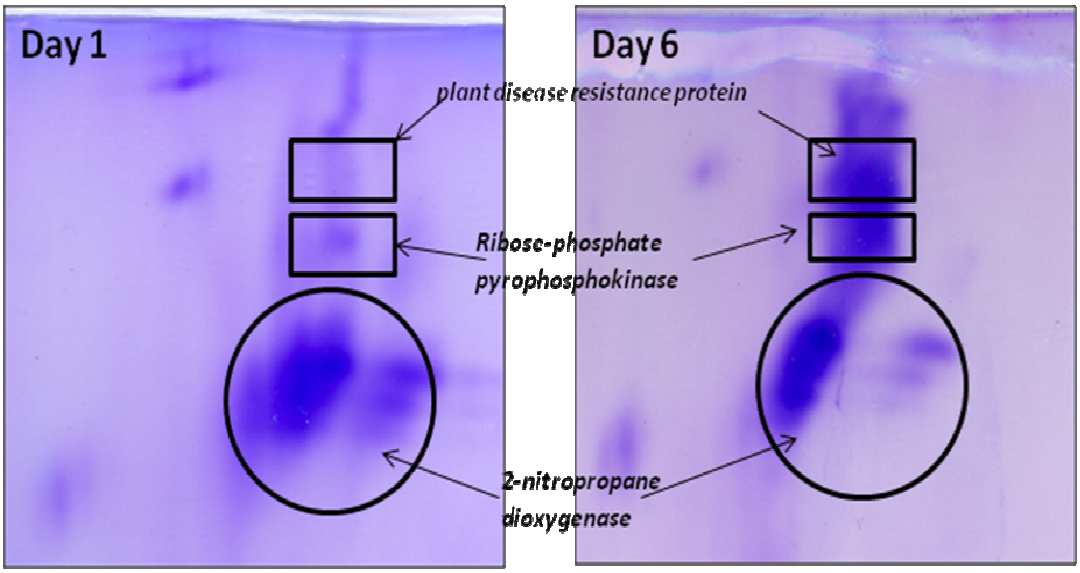


Figure 4.3: 2Dgel-SDS/PAGE highlighting regions with different Day 1 and Day 6 protein expressions

1D-Gel Shotgun Proteome

1D-gel shotgun proteomics is performed to sequence all proteins present in a sample. Proteins digested with trypsin yield peptides with unique masses belonging to specific amino acid sequences. Separation of these peptides is achieved by liquid chromatography. Data analysis using these peptides were searched against protein database, this allows for identification unknowns based on each peptide's unique sequences.

Total soluble leaf protein extracts from the time point samples were run on a one dimensional polyacrylamide gel (1D-gel, 12% acrylamide). Proteins from control and stressed tissues were separated based on their molecular weight and each extract migrated over entire lane evenly. Overall banding pattern were similar between control and treated

tissue, but some were clearly different. For example, at the 30 kD region, the intensity of the bands are visually different between control and treatment sample: At the 20kD region, the relative intensity (lane 1) was higher than other fractions in lanes 2 to 4. It was expected 1D-gel should give numerous bands acrossed the individual lanes that are clearly focused and spread out over entire length of a gel (**Figure 4.2**).

These lanes were subjected to the in-gel trypsin digestion. Since more protein was recovered from 1D-gel, protein sequences obtained from the 1D-gel SDS/PAGE outnumber peptide sequences obtained from the 2D-gel.

2D-Gel Proteome

Two Dimensional-gel SDS/PAGE was performed on both un-stressed and stressed protein extracts. Due to the complexity of recovering soluble proteins from leaf tissue, loss of protein post-extraction, and plant derived hydrophilic macromolecules interferring with isoelectric focusing, a large number of proteins did not resolve to the anticipated level of consistency. Two Dimensional-gels typically require more protein (50~100 fold) than that of 1D-gel (a lane contains 10 ~ 20 μg total proteins across can be seen comfortably by Coomassie staining). Interfering compounds, and/or inhibitory compounds present in the extract interferred with the immobilized pH gradient. However there were isolated areas within the gel where proteins did produce spots (**Figure 4.3**).

Protein Identification

Mass Spectrometry analysis yielded a total of 1974 tentative peptide sequences from the in-gel trypsin digestions from both stressed and unstressed protein extracts with both 1D and 2D gel results combined (**Table 4.1**).

Some sequences did show homology to gene sequences present in DFCI-Aquilgia Gene Index (AqGI). The protein homologs from Day 1 can be seen in **Table 4.1** and Day 6 proteins are seen in **Table 4.2**. These proteins are categorized into one of eight different classes according to their function; enzyme, regulatory, transport, storage, motile, structural, scaffold, or protective proteins.

Table 4.1: Results of BLAST search of *Aquilegia* EST database from Day 1 protein sequences

Name	Accession	*Species*	e-value	peptide	homolgy	seq name	EST
Enzyme	**(#)**			**(#)**	**(%)**		
receptor protein kinase, putative	EEF30210	*Ricinus communis*	3.1	1	32%	Seq_986	DT767316
Glutathione transferase	EEF30504.1	*Ricinus communis*	0.88	3	76%	Seq_1074	DR949706
lactoylglutathione lyase, putative	EEF43857	*Ricinus communis*	1.2	1	100%	seq_1449	TC20087
polyamine oxidase1	NP_001105106	*Zea mays*	0.39	1	83%	seq_1220	TC24685
fructose-1,6-bisphosphatase	NP_190973	*Arabidopsis thaliana*	1.6	1	91%	Seq_1308	TC22551
2-nitropropane dioxygenase	NP_568988	*Arabidopsis thaliana*	1.5		100%	Seq_20	TC24230
NADP isocitrate dehydrogenase	BAC77065	*Lupinus albus*	4.4	1	57%	seq_1162	TC22980
flavonoid 3-hydroxylase	EEF40985	*Ricinus communis*	2.6	1	37%	seq_861	TC25151
beta-D-galactosidase	BAF31234	*Persea americana*	0.75	1	41%	seq_1119	TC25259
sucrose cleavage protein-like	AAM62596	*Arabidopsis thaliana*	5.4	1	76%	seq_1473	TC26550
glutathione-s-transferase theta	EEF30504	*Ricinus communis*	6.3	4	100%	seq_1501	TC26575
cytokinin oxidase	XP_002308930	*Populus trichocarpa*	8.3	1	45%	seq_1502	TC27340
receptor-kinase, putative	EEF43563	*Ricinus communis*	0.47	1	33%	seq_1119	TC27557
polyamine oxidase 1	NP_001105106	*Zea mays*	1	2	72%	seq_928	TC29091
Regulatory							
zinc finger (C3HC4-type RING finger)	NP_001078548	*Arabidopsis thaliana*	1.2	1	33%	Seq_845	EST1113245
KNAT3-like transcription factor	ACJ09316	*Juglans nigra*	9	1	45%	seq_767	TC24805
Chain A, Polyamine Oxidase	1B37_A	*Zea mays*	0.39	1	83%	seq_1220	TC24685
calcium ion binding protein	EEF30929	*Ricinus communis*	3.3	1	39%	seq_1426	TC25390
Transport							
condensin subunit	XP_002304405	*Populus trichocarpa*	9.7	1	57%	Seq_1573	EST1108620
Motile,							
peroxisomal membrane protein 34	EEF49824	*Ricinus communis*	0.43	1	39%	Seq_858	TC22446
Structural							
putative alpha7 proteasome subunit	CAC43323	*Nicotiana tabacum*	2.7	1	57%	seq_1014	TC21745

Table 4.2: BLAST hits to Aquilegia EST database from Day 6 protein sequences

Name	Accession	Species	e-value	peptide	homolgy	seq name	EST
Enzyme	(#)			(#)	(%)		
Ribose-phosphate pyrophosphokinase 4	Q9XGA1	*Spinacia oleracea*	7	1	39%	Seq_437	TC21792
candidate plant disease resistance protein	AAL76166	*Glycine max*	4.9	2	42%	Seq_489	EST1177598
2-nitropropane dioxygenase	NP_568988	*Arabidopsis thaliana*	1.5	4	84%	ID_61	TC24230
Beta-amylase	AAY40266	*Glycine max*	1.4	1	68%	ID_274	DT732275
Fructose-bisphosphate aldolase	EEF30862	*Ricinus communis*	0.099	3	87%	ID_10	TC23892
2-oxoglutarate-dependent dioxygenase	XP_00229812	*Populus trichocarpa*	3.2	1	73%	ID_239	TC20425
class I beta-1,3-glucanase	XP_00227716	*Populus trichocarpa*	1.2	2	100%	ID_146	TC23874

The protein spots that did resolve on the 2D-gel SDS/PAGE from both stressed and unstressed samples were excised and subjected to in-gel trypsin digestion followed by subsequent LC/MS for identification of proteins. Based on the results and their tentative identification, some spots were found to be differentially expressed from Day 1 to Day 6 tissue. For example, 2-nitropropane dioxygenase was one of the abundant protein present in the protein extract in day 6 tissue (**Figure 4.3)**. The circled proteins in **Figure 4.3** all yielded BLAST hits to 2-nitropropane dioxygenase. The spot pattern comparison of the 2-nitropropane dioxygenase suggest that modification occurred to this particular protein leading to a shift in isoelectric point and molecular weight, which produced the various isoforms detected on 2D-gel. Post-transcriptional modifications such as phosphorylation can cause shifts in charge and glycosylation can cause shifts in molecular weight large enough to be detected on 2D gels. Various post transcriptional modificatons are known to affect the activity of enzymes, for instance, activating or inactivating enzymes (Voet and Voet, 2004). The variation of states of this particular enzyme may be indicative of the forms needed in response to drought conditions. In this instance 2-nitropropane dioxygenase is known to be a Reactive Oxygen Species generator. ROSs are known to

be detrimental to cells especially in stress conditions. The variations in isoforms from Day 1 to Day 6 may indicate a modification to control function of enzyme.

Another protein identified from the 2D-gel as being present in the stressed but not in unstressed protein extract is a putative plant disease resistance protein (R protein). This obvious difference in spot pattern can be seen in the highlighted square in **Figure 4.3**. Because this protein is only present in detectable amount in Day 6 (severely drought-stressed plant leaf tissue), it is assumed that this particular protein plays a role in general defense response overlapping dessiccation condition. Plant disease resistance proteins are normally associated with detection of pathogens, acting early on as foreign proteins are detected instigating a signal cascade where, metabolism is altered and defenses are activated (McDowell and Woffenden, 2003). It is known that plants utilize one gene for many different functions depending on conditions and a need for expression of certain products. This may be the case here , where an R-protein is expressed in response to drought, signalling a drought defensive pathway.

Yet another protein that was tentativly identified from the 2D-gel was a ribose-phosphate pyrophosphokinase. This particular protein is the key enzyme in the biosynthesis of nucleic acids and amino acids catalyzing the formation of alpha-D-ribose-5-Phosphate from Phosphoribosyl pyrophosphate.

Many of the proteins that are shown in the representative table (**Table 4.1 and 4.2)** are known to have some role in stress tolerance. For example, glutathione S transferase is known to play a role in both detoxification of metabolites and stress tolerance in plants (Qing-Yin et al., 2005). Polyamine Oxidase tDNA mutants have been

shown to alter their expression in response to drought stress in the flavonoid biosynthesis pathway (Nishimura, 2008). This may be related to our results presented in Chapter 2 highlighting the importance of two flavonoids in *Aquilegia's* response to drought. Another identified peptide showed similarity to a flavonoid 3-hydroxylase, *SbF3'H2,* found in sorghum, which has been suggested to play some role in pathogen-specific 3-deoxyanthocyanidin synthesis (Lo et al., 2006).

From the Day 6 protein extract, a fructose-bisphosphate aldolase had the most BLAST hits. Interestingly, this protein has been associated with drought response by Zhang et al., 2009. An SSH library was constructed from *Sesuvium portulacastrum* (a halophyte species). From this library, a gene (*SpFPA)* with high homology to a sequence from *Mesembryanthemum nodiflorum*, (a succulent plant able to withstand high salt concentrations) was identified. When *SpFPA* was isolated and cloned into *Escherichia coli,* recombinant *E. coli* has shown an increased tolerance to salinity and related abiotic stimulus (Zhang et al., 2009).

Peptides Compared to Subtractive Hybridization Library

Protein hits in **Table 4.2** were compared to cDNA sequences obtained from the subtractive hybridization libraries. Proteins that showed homology to SSH gene sequences were identifed via manual annotation in **Table 4.3**.

Table 4.3: Proteins with homologous sequences to drought stressed suppression sbtractive-hybridization library

flavonoid 3-hydroxylase	**EEF40985**	***Ricinus communis***	**2.6**	**1**	**37%**	**seq_861**	**TC25151**
glutathione-s-transferase theta	**EEF30504**	***Ricinus communis***	**6.3**	**4**	**100%**	**seq_1501**	**TC26575 TC29438**
zinc finger (C3HC4-type RING finger)	**NP_001078548**	***Arabidopsis thaliana***	**1.2**	**1**	**33%**	**Seq_845 Seq_847**	**>EST1113245 DR921706**

Conclusions

This study attempted to determine any quantitative or qualitative differences seen in *Aquilegia*'s proteome over a six day period of increasing drought conditions. Based on the obtained results, some differences were noted. The 1D-SDS PAGE produced differences between samples. Most of the Day 6 proteins that were identified based on their homologous peptide sequences were associated with some form of metabolite synthesis. This strongly corresponds with published research showing the importance of accumulated osmolytes as a defensive mechanism to drought and results presented in Chapter 2 (Yancey, 2001).

In addition to a difference of peptides identified from the 1D gel data, 2D SDS PAGE showed difference in the expression pattern of section of the 2D gel. These proteins were identified as showing homology to a; 2-nitropropane dioxygenase , plant disease resistance protein, and Ribose-phosphate pyrophosphokinase. 2-nitropropane dioxygenase is known to be a reactive oxygen species generator. The change of isoforms from Day 1 to Day 6 suggest that a change in some regulatory element takes place

possibly to down regulate the function of this particular gene. These differences represent a real change in the expression of Aquilegia's proteome due to drought stress.

Information obtained from this study stressed the importance of metabolites and osmolyte accumulation as a means to maintain stability and structure of plants undergoing drought stress. In order to gain a deeper understanding of how the proteome, the transcriptome, and the metabolome are connected, a more indepth study utilizing 2D SDS PAGE consisting of more time points should be performed . This should allow for a more quantitative analysis of proteome differences.

CHAPTER FIVE

CONCLUSIONS

This type of study, using multiple approaches will contribute to determining the molecular basis of *Aquilegia's* enhanced resistance to drought. Equally important will be identification of new genes involved in drought tolerance by testing transgenic crop species for increased ability to withstand drought and increase yield in water limited climates.

Plant response to drought and the genetic mechanisms plants utilize to cope with drought conditions are highly complex. To attempt to dissect these various pathways, SSH libraries are an excellent tool utilizing Sanger sequencing. The availability of SSH library from drought stress tissue is a useful reference for other parallel stress experiments such as examination of proteome and metabalome. The usefulness of a SSH library as a reference point can clearly be seen in this study: when vitexin and vitexin 2"-O-β-D-glucoside were positively identified from the metabalomics work, being two of the four accumulated molecules shown in Chapter 2. It was found that the vitexin to vitexin 2"-O-β-D-glucoside reaction was catalyzed by the enzyme UDP-glycosyltransferase 85A8. A homologous sequence to UDP-glycosyltransferase 85A8 was present in the *Aquilegia* SSH library. This partial gene sequence was subsequently used for the Real Time PCR over increased levels of drought treatment. Results show this particular gene is upregulated under drought conditions. To date, neither vitexin nor

vitexin 2”-O-β-D-glucoside has been reported in literature to be accumulated or associated with drought response in plants.

This is the first study to utilize SSH to detect changes in the transcriptome of *Aquilegia* in response to drought. Sequences obtained from the forward subtracted subhybridization library have yielded gene sequences of proteins with known functions dealing with drought conditions in plants. Genes encoding for proteins such as; aquaporin, low temperature and salt responsive protein, glutathione S-transferase and ethylene-responsive element-DREB were identified that are known to be related to drought response and were present in the SSH library. Other sequences found in the SSH library such as ferritin-3, chloroplast precursor, ribosomal proteins, norcoclaurine synthase and many more have not been implicated in drought response directly. There is a possibility that some of these genes may actually be found to have some role based on the fact they were present in the tester SSH library. For example, until now *UDP-glycosyltransferase* has not been classified as a drought tolerance gene. In Chapter two we provide evidence linking this gene to sugar metabolism that possibly acts in a drought-buffering small molecule synthesis pathway (Chapter Two).

Identification and functional analysis of these unknown genes will be valuable in further understanding of the various pathways involved in drought response in other plant species. Twenty-nine percent of the assembled contigs obtained from the stressed tissue library produced no homology to any sequences in the public databases. Additional research is needed to study these novel genes sequences to determine if they have a role in drought stress mechanism for *Aquilegia*. There is a high likelihood that one of the

genes present in the stressed SSH library and showing no homology to known sequences in the dataset, might be a partial gene sequence of the lacking enzyme responsible for catalyzing the apigenin to vitexin reaction. If positively identified from functional analysis of the gene, the pathway from apigenin to the final amassed product, vitexin 2"-O-β-D-glucoside, could be genetically engineered and transformed into crop species for studies involved in drought resistance.

SSR analysis of assembled contigs has yielded various possible genetic markers for *Aquilegia* that may be utilized in a breeding program to improve the overall drought resistance of *Aquilegia*. Because of its evolutionary position, any information gained may be applicable to a large number of species. These SSR containing contigs need to be annotated as to what their homologous genes are and examined in depth. The complexity of repeats increases the possibility of producing polymorphisms within different *Aquilegia* plants. A comparison of the SSR containing regions should be very useful when aligned to homologs present in the NCBI database from different species such as *Arabidopsis* and rice. This may yield information applicable to a marker assisted breeding program to future dissect *Aquilegia'*s ability to withstand drought conditions.

It has been shown from previous studies that over-expression of trehalose-6-phosphate and trehalose-6-phosphate phosphatase provides transgenic *Arabidopsis thaliana* with an elevated resistance to drought, freezing, salt and heat tolerance (Iturriaga et al., 2007). It is well known that accumulation of sugars and osmolytes yield protection against drought. Vitexin 2"-O-β-D-glucoside's two glucose residues closley resemble trehalose. Combined with the water retention capacity previously discussed, there is

evidence to make the assumption that vitexin and vitexin 2”-O-β-D-glucoside play a critical role in *Aquilegia’s* ability to withstand drought conditions. The UDP-glycosyltransferase 85A8 gene obtained from the subtractive hybridization library is an excellent candidate for gene transformation and functional analysis in *Arabidopsis* and/or rice. Additional research should be pursued to identify the enzyme responsible for the apigenin to vitexin reaction.

Based on the metabolyte data, both vitexin and vitexin 2”-O-β-D-glucoside are accumulated at later stages of drought. Because the RNA was obtained at the earliest morphologic changes of drought, this may explain why there were not more sequences present from the *UDP-glycosyltransferase 85A8* gene. A subtractive hybridization library from a later time point may provide a better understanding of later drought response by showing a different set of gene sequences related to drought. This could produce a partial gene sequence for the enzyme responsible for the apigenin to vitexin reaction among other possibilities.

Analysis of the proteome via SDS-PAGE and LCMS yielded numerous peptides with homology to proteins known to be involved in drought responses. Also, tentative identification was given to regulatory, transport, and structural proteins. The majority of proteins extracted from the day 6 extraction were identified as showing homology to functional enzymes, such as Ribose-phosphate pyrophosphokinase 4, 2-nitropropane dioxygenase, β-amylase, Fructose-bisphosphate aldolase, 2-oxoglutarate-dependent dioxygenase, and class I β-1,3-glucanase. An additional candidate plant disease resistance protein was identified from the day 6 protein sample based on homology to

NCBI database. The majority of these proteins are enzymatic, functional proteins reacting on a specific substrate to create a specific product.

In addition to a difference of peptides identified from the 1D gel data, 2D SDS PAGE provided differences in the expression pattern. These differences were correspond to a; 2-nitropropane dioxygenase , plant disease resistance protein, Ribose-phosphate pyrophosphokinase. 2-nitropropane dioxygenase is known to be a reactive oxygen species generators. The change of isoforms from Day 1 to Day 6 suggest that a change in some regulatory element takes place possibly to down regulate the function of this particular gene.

The spot pattern comparison of the 2-nitropropane dioxygenase suggest that modification occurred to this particular protein leading to a shift in isoelectric point and molecular weight, which produced the various isoforms detected on 2D-gel. Post-transcriptional modifications such as phosphorylation can cause shifts in charge and glycosylation can cause shifts in molecular weight large enough to be detected on 2D gels. Various post transcriptional modifications are known to affect the activity of enzymes, for instance, activating or inactivating enzymes (Voet and Voey, 2004). The variation of states of this particular enzyme may be indicative of the forms needed in response to drought conditions. ROSs are known to be detrimental to cells especially in stress conditions. The variations in isoforms from Day 1 to Day 6 may be a modification to control function of enzyme.

Another protein identified from the 2D-gel as being present in the stressed but not in unstressed protein extract is a putative plant disease resistance protein (R protein).

Because this protein is only present in detectable amount in Day 6 (severely drought-stressed plant leaf tissue), it is assumed that this particular protein plays a role in general defense response overlapping desication conditions. Plant disease resistance proteins are normally associated with detection of pathogens, acting early on as foreign proteins are detected instigating a signal cascade where, metabolism is altered and defenses are activated (McDowell and Woffenden, 2003). It is known that plants utilize one gene for many different functions depending on conditions and a need for expression of certain products. This may be the case here , where an R-protein is expressed in response to drought, signalling a drought defensive pathway.

Information obtained from this study stressed the importance of metabolites and osmolyte accumulation as a means to maintain stability and structure of plants undergoing drought stress. In order to gain a deeper understanding of how the proteome, the transcriptome, and the metabolome are connected, a more indepth study utilizing 2D SDS PAGE consisting of more time points should be performed. This will allow for a more quantitative analysis of proteome differences.

The three approaches, genomic, metabalomic and proteomics, of this project are interconnecting and complementary, with each aspect giving an overall glimpse of biological processes. The subhybridization library yielded homologous sequences for the enzyme responsible for the vitexin to vitexin 2”-O-β-D-glucoside reaction found in the metabolomic work. The proteomic analysis yielded tentative homologous proteins with implications in the formation of different metabolites such as flavonoid 3-hydroxylase, Fructose-bisphosphate aldolase, and β-amylase. The large increase of specific sugar

containing molecules lends value to the proteomic data that showed the presence of many metabolic enzymes in the late stages of drought.

Our knowledge and understanding of drought and stress related pathways have been expanded with the large amount of sequence data, and the various in-depth studies into this stress mechanism. Until now vitexin and vitexin 2”-O-β-D-glucoside have not been associated as a defense mechanism to drought conditions. Functional analysis of genes responsible for the synthesis of these two compounds could yield important results in the engineering of more drought tolerant crop species.

APPENDICES

Appendix A

SSR analysis of AHOTEg,

Simple sequence repeats present differentially expressed contigs from stressed library AHOTEg. Table shows: name description, name, number of repeats in each contig, overall sequence length, SSR location, and number of repeats.

Description	SeqLen	SSR					
		# SSRs	# bp	Motif	# Repeats	Start	Stop
Contig1	741	2	3	aac	5	219	233
Contig1	741	2	3	ctt	7	299	319
Contig11	687	2	2	ca	5	187	196
Contig11	687	2	4	gata	3	261	272
Contig14	731	1	2	ct	11	418	439
Contig28	592	3	3	aca	4	171	182
Contig28	592	3	4	aaga	3	428	439
Contig28	592	3	2	ta	5	507	516
Contig34	679	3	2	tc	11	602	623
Contig34	679	3	4	caca	3	623	634
Contig34	679	3	2	ac	6	624	635
Contig35	806	1	4	acaa	3	118	129
Contig36	764	1	3	aag	4	639	650
Contig39	518	1	4	tttg	3	479	490
Contig40	521	2	3	caa	10	11	40
Contig40	521	2	3	act	4	283	294
Contig43	756	1	2	gt	7	691	704
Contig48	625	2	3	gta	5	467	481
Contig48	625	2	3	ttc	5	508	522

Contig49	582	2	2	at	6	81	92
Contig49	582	2	2	at	6	127	138
Contig53	671	2	3	taa	7	29	49
Contig53	671	2	3	gta	6	245	262
Contig56	762	1	4	tagc	3	593	604
Contig62	582	1	3	ttc	4	324	335
Contig63	824	1	3	tct	5	355	369
Contig64	1157	1	2	ag	5	1122	1131
Contig65	605	2	4	ttct	3	52	63
Contig65	605	2	3	ctt	8	67	90
Contig80	576	1	3	aga	4	451	462
Contig81	550	1	3	caa	4	136	147
Contig91	532	1	2	tc	9	511	528
Contig94	1096	1	4	aatg	8	1046	1077
Contig100	472	1	4	aaca	3	135	146
Contig104	1094	1	3	tcc	4	106	117
Contig106	277	1	4	gtct	3	187	198
Contig117	519	1	3	tct	4	257	268
Contig118	1090	1	3	caa	5	881	895
Contig126	597	1	2	gt	5	222	231
Contig133	569	2	4	caaa	3	20	31
Contig133	569	2	4	aaag	3	43	54
Contig135	409	1	4	ttaa	3	161	172
Contig136	365	1	4	agaa	3	129	140
Contig143	1182	1	2	ta	5	69	78
Contig147	450	1	4	acat	3	58	69
Contig153	568	1	2	at	5	408	417
Contig156	290	2	4	gttg	4	157	172
Contig156	290	2	4	tgtt	3	242	253

Contig159	587	1	3	gca	4	480	491
Contig167	1247	1	3	ttc	4	217	228
Contig169	627	1	3	cag	7	120	140
Contig170	543	3	3	aac	5	1	15
Contig170	543	3	3	tca	4	60	71
Contig170	543	3	3	tct	4	188	199

Frequency of repeat motifs Simple sequence repeats present in all contigs from stressed library AHOTEg. Table shows sequence motif, number of times repeated and percentage.

Motif	Frequency	Frequency (%)	Motif Length
aaag	2	2.739726027	4
aaat	2	2.739726027	4
aac	4	2.054794521	3
aaca	1	2.739726027	4
aag	4	2.054794521	3
aaga	1	2.739726027	4
aagaa	2	3.424657534	5
aatg	1	2.739726027	4
ac	3	1.369863014	2
aca	1	2.054794521	3
acaa	1	2.739726027	4
acat	1	2.739726027	4
acc	1	2.054794521	3
acgt	1	2.739726027	4
act	2	2.054794521	3

ag	2	1.369863014	2
aga	2	2.054794521	3
agaa	1	2.739726027	4
agaag	1	3.424657534	5
agct	1	2.739726027	4
agt	2	2.054794521	3
at	7	1.369863014	2
ata	2	2.054794521	3
atc	2	2.054794521	3
atg	1	2.054794521	3
atga	1	2.739726027	4
att	1	2.054794521	3
atta	1	2.739726027	4
ca	3	1.369863014	2
caa	4	2.054794521	3
caaa	1	2.739726027	4
caca	1	2.739726027	4
cag	1	2.054794521	3
cat	1	2.054794521	3
cca	1	2.054794521	3
cct	1	2.054794521	3
ct	1	1.369863014	2
cta	1	2.054794521	3
ctag	1	2.739726027	4
ctt	4	2.054794521	3
cttt	1	2.739726027	4
ga	1	1.369863014	2
gata	1	2.739726027	4
gca	2	2.054794521	3

gct	1	2.054794521	3
gctt	1	2.739726027	4
gt	3	1.369863014	2
gta	2	2.054794521	3
gtct	1	2.739726027	4
gtg	1	2.054794521	3
gtt	1	2.054794521	3
gttg	1	2.739726027	4
ta	4	1.369863014	2
taa	1	2.054794521	3
taaa	1	2.739726027	4
taat	2	2.739726027	4
tac	1	2.054794521	3
tag	1	2.054794521	3
tagc	1	2.739726027	4
tc	5	1.369863014	2
tca	2	2.054794521	3
tcc	1	2.054794521	3
tct	11	2.054794521	3
tctt	2	2.739726027	4
tg	3	1.369863014	2
tgtt	1	2.739726027	4
ttaa	1	2.739726027	4
ttc	10	2.054794521	3
ttcc	1	2.739726027	4
ttct	1	2.739726027	4
ttctc	1	3.424657534	5
ttg	3	2.054794521	3
ttgt	1	2.739726027	4

ttta	1	2.739726027	4
tttc	2	2.739726027	4
tttg	1	2.739726027	4
tttgc	1	3.424657534	5

SSR Summary Report AHOTEg

Selected Motif Type	Min # repeats	
dinucleotides	5	
trinucleotides	4	
tetranucleotides	3	
pentanucleotides	3	
Number of sequences searched for SSRs 113		
Number of unique sequences with at least one repeat.. 39		
Number of motifs that are unique 40		
Motif Length	**Frequency**	**Frequency (%)**
2 bp	13	24.07407407
3 bp	24	44.44444444
4 bp	17	31.48148148

Appendix B

SSR analysis of AHOTEh

SSR analysis of AHOTEh, Simple sequence repeats present in all contigs from stressed library AHOTEh. Table shows Name description, number of repeats in each Contig, overall sequence length, SSR location.

Description	SeqLen	SSR					
		# SSRs	# bp	Motif	# Repeats	Start	Stop
Contig3	613	1	4	tttc	5	586	605
Contig5	729	2	3	tgt	4	520	531
Contig5	729	2	3	tga	4	694	705
Contig12	449	1	3	gaa	7	288	308
Contig15	341	1	3	aga	5	89	103
Contig24	622	1	2	at	6	10	21
Contig29	142	1	2	gt	5	85	94
Contig31	521	1	2	tc	5	468	477
Contig34	575	1	3	ttc	5	528	542
Contig44	235	2	3	agt	4	188	199
Contig44	235	2	3	cta	5	213	227
Contig46	685	2	3	ttc	6	354	371
Contig46	685	2	3	caa	4	379	390
Contig48	536	1	2	gt	5	258	267
Contig50	1162	1	3	cta	6	272	289
Contig60	487	1	3	tcg	4	381	392
Contig82	580	1	4	aaag	3	257	268
Contig84	530	2	2	ag	5	496	505
Contig84	530	2	4	agaa	3	504	515
Contig85	580	3	3	att	4	459	470

Contig85	580	3	3	ttc	6	490	507
Contig85	580	3	4	aaag	4	537	552
Contig86	616	1	2	ac	7	193	206
Contig87	907	1	2	ct	6	83	94
Contig89	912	1	3	tct	7	827	847
Contig98	683	2	3	tat	4	391	402
Contig98	683	2	4	taag	3	667	678
Contig100	774	1	2	at	5	762	771
Contig103	645	1	4	caaa	4	166	181

Frequency of repeat motifs of un-stressed AHOTEh subtractive hybridization library

Motif	Frequency	Frequency(%)	Motif Length
aaag	2	13.79310345	4
ac	1	6.896551724	2
ag	1	6.896551724	2
aga	1	10.34482759	3
agaa	1	13.79310345	4
agt	1	10.34482759	3
at	2	6.896551724	2
att	1	10.34482759	3
caa	1	10.34482759	3
caaa	1	13.79310345	4
ct	1	6.896551724	2
cta	2	10.34482759	3
gaa	1	10.34482759	3

gt	2	6.896551724	2
taag	1	13.79310345	4
tat	1	10.34482759	3
tc	1	6.896551724	2
tcg	1	10.34482759	3
tct	1	10.34482759	3
tga	1	10.34482759	3
tgt	1	10.34482759	3
ttc	3	10.34482759	3
tttc	1	13.79310345	4

SSR Summary Report AHOTEh

Selected Motif Type	Min # repeats	
dinucleotides	5	
trinucleotides	4	
tetranucleotides	3	
pentanucleotides	3	
Number of sequences searched for SSRs 62		
Number of unique sequences with at least one repeat .. 14		
Number of motifs that are unique 17		
Motif Length	**Frequency**	**Frequency (%)**
2 bp	5	26.31578947
3 bp	11	57.89473684
4 bp	3	15.78947368

LITERATURE CITED

Alberts B, Bray D, Lewis J, Raff M, Roverts K, Watson JD (1994) In: Molecular Biology of the Cell (2nd edition). Garland Publishing, Inc. New York, NY

Aroca R, Amodeo G, Fernandez,-Illescas S, Herman EM, Chaumont F, Chrispeels. (2005) The role of aquaporins and membrane damage in chilling and hydrogen peroxide induced changes in the hydraulic conductance of maize roots. Plant Physiology 137: 341-353.

Assmann SM, Snyder JA, Lee YRJ (2000) ABA-deficient (*aba1*) and ABA-insensitive (*abi1-1*, *abi2-1*) mutants of Arabidopsis have a wild-type stomatal response to humidity. Plant Cell Environ. 23:387-395.

Avonce N, Leyman B, Mascorro-Gallardo JO, Van Dijck P, Thevelein J, Iturriaga G (2004) The Arabidopsis trehalose-6P-synthase *AtTPS1* gene is a regulator of glucose, ABA and stress signaling. Plant Physiol. 136:3649-3659.

Ayerbe Luis, Sanchez F, Manzanares M, Andres E, Tenorio J (1998) Turgor maintenance, osmotic adjustment and soluble sugar and proline accumulation in 49 pea cultivars in response to water stress. Field Crops Res 59: 225-235.

Bartels D, Sunkar R (2005) Drought and salt tolerance in plants. Critical Reviews in Plant Science. 24: 23-58.

Beck EH, Fettig S, Knake C, Hartig K,Bhattarai T (2007) Specific and unspecific responses of plants to cold and drought stress. J Biosci 32:501-510.

Benlloch-Gonzalez M, Octavio Arquero O, Jose Maria Fournier JM,, Diego Barranco D, Manuel Benlloch M (2007) K+ starvation inhibits water-stress induced stomatal closure. Journal of Plant Physiology 165: 623-630.

Bolen DW, Wang A. (1997) A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. Biochemistry 36: 9101-9108.

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005). Endogenous siRNA derived from a pair of natural cis-antisence transcripts regulated salt tolerance in Arabidopsis. Cell 123:1279-1291.

Boyer JS (1982) Plant productivity and environment. Science 218:443-448.

Bradford KJ, Hsiao TC (1982) Physiological responses to moderate water stress. In: Water relations and photosynthetic productivity, Vol 12B. Lange OL Nobel PS, Osmond CB, Ziegler H (eds.), Encyclopedia of Plant Physiology, New Series, Springer-Verlag, Berlin, Heidelberg, New York, pp 263-324.

Bruce W, Edmeades G, Barker T (2002) Molecular and physiological approaches to maize improvement for drought tolerance. Journal of Experimental Botany 53: 13-25.

Budavari, S. (Ed.). (1989) The Merck Index an Encyclopedia of Chemicals, Drugs and Biologicals. Rahway, NJ, USA: Merck & Co., Inc.

Campbell SA, Close TJ (1997) Dehydrins; genes proteins and association with phenotypic traits. New Phytologist 137:61-74.

Chaves MM (1991) Effects of water deficits on carbon assimilation. J Exp Bot 42:1-6.

Chaves MM, Marco JP, Pereira JS (2003) Understanding plant responses to drought-from genes to the whole plant. Funct Plant Biol 30:239-264.

Chen TH, Murata TC (2002) Enhancement of tolerance of abiotic stress by metabolic engineering of betaines and other compatible solutes. Currt Opin Plant Biol 5*:*250-257.

Chinnusamy V, Jagendorf A, Zhu JK (2005) Understanding and improving salt tolerance in plants. Crop Sci 45:437-448.

Conesa A, Gotz S, Garcia Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatic 21:3674-3676.

Correia MJ, Pereira JS, Chaves MM, Rodrigues ML, Pacheco CA (1995) ABA xylem concentrations determine maximum daily leaf conductance of field-grown *Vitis vinifera* L. plants. Plant Cell Environ 18:511-521.

Cuming A (1999) LEA Proteins. Dordrecht, The Netherlands: Kluwer Academic.

Davies WJ, Wilkinson S (1997) Xylem sap pH increase: A drought signal received at the apoplastic face of the guard cell that involves the suppression of saturable abscisic acid uptake by the epidermal symplast. Plant Physiol 113:559-573.

Davies WJ, Zhang J (1987) Increased synthesis of ABA in partially dehydrated root tips and ABA transport from root to leaves. Journal of Experimental Biology 38: 2015-2023.

Dhindsa RS (1991) Drought stress, enzymes of glutathione metabolism, oxidation injury, and protein syntheis in *Tortula ruralis*. Plant Physiol 95:648-651.

Ehleringer Jr, TC (1992) On the role of orientation in reducing photoinhibitory damage in photosynthetic-twig desert shrub plant. Cell & Environment 15: 301-306.

Elbein A(1974) The metabolism of alpha-trehalose. Adv Carbohydr Chem Biochem 30:227-256.

Goddijn OJ, van Dun K (1999) Trehalose metabolism in plants. Trends Plant Sci 4:315-319.

Gurskaya N, Diatchenko L, Chenchik A, et al. (1996) Equilizing cDNA subtraction based on selective suppression of polymerase chain reaction; cloning of jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. *Anal Biochem* 240: 90-97.

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19:1720-1730.

Han S-E, Park S-R, Kwon H-B, Yi B-Y, Lee G-B, Byun M-O (2005) Genetic engineering of drought-resistant tobacco plants by introducing the trehalose phosphorylase (TP) gene from *Pleurotus sajor-caju*. Plant Cell Tissue Organ Cult 82:151-158.

.

Heilsbroek R, Van Brederode J, Van Nigtevecht G, Maas J, Kamsteeg J, Besson E (1980) The 2“-O-glucosylation of vitexin and isovitexin in petals of *Silene alba* is catalysed by two different enzymes. Phytochemistry 19:1935-1937.

Hodges S, Kramer E (2007) Columbines. Current Biology 17: 992-994.

Hsiao TC, Silk WK, Jing Jj (1985) Leaf growth and water deficits: biophysical effects. 27 Cambridge University Press, Cambridge, MA pp 239-266.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868-877.

Iturriaga G, Miranda J, Avonce N, Suarez R, Thevelein J, VanDijck P, (2007) A bifunctional TPS-TPP enzyme from yeast confers tolerance to multiple and extreme abiotic-stress conditions in transgenic *Arabidopsis.* Planta 226:1411-1421.

Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF (1998) Arabidopsis CBF1 overexpression induces cor genes and enhances freezing tolerance. Science (5360):104-106.

Jaleel CA, Manivannan P, Kishorekumar A, Sankar B, Gopi, R, Somasundaram R, Panneerselvam R (2007) Alterations in osmoregulation, antioxidant enzymes and indole alkaloid levels in *Catharanthus roseus* exposed to water deficit. Colloids surfaces. B: Biointerfaces. 59(2):150-157.

Jones HG (1973) Moderate-term water stresses and associated changes in some photosynthetic parameters in cotton. New Phytol 72:1095-1105.

Jung S, Abbott A, Jesudurai C, Tomkins J, Main D (2005) Frequencey, type, distribution, and annotation of simple sequence repeats in Rosaceae ESTs. Funct Integr Genomics 5:136-143.

Karp P, Caspi R, Foerster H, Fulcher C, Kaipa P, Krummenacker M, Latendresse M, Paley K, Rhee S, Shearer A, Tissier C, Walk T, Zhang (2007) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research 36:D623-D631.

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat Biotechnol 17:287-291.

Katiyar SK, Ahmad N, Mukhtar H (2000) Green tea and skin. Arch Dermatol 136: 989-994.

Komatsu (2006) Plant proteomics databases: Their status in 2005. Current Bioinformatics 1:33-36.

Kozlowski TT and Pallardy SG (2002) Acclimation and adaptive responses of woody plants to environmental stresses. The Botanical Review 68:270-334.

Kushiro T, Okamoto M, Nakabayashi K, Yamagishi K, Kitamura S, Asami T, Hirai N, Koshiba T, Kamiya Y, Nambara E (2004) The Arabidopsis cytochrome P450 CYP707A encodes ABA 8'- hydroxylases; key enzymes in ABA catabolism. EMBO J 23:1647-1656.

Larcher W (2000) Temperature stress and survival ability of Mediterranean sclerophyllous plants. Plant Biosyst 134:279-295.

Levitt J (1972) Responses of plant to environmental stresses. New York: Academic Press, New York.

Leung J, Giraudat J (1998) Abscisic acid signal transduction. Annual Review of Plant Physiology and Plant Molecular Biology 49: 199-222.

Li Y, Korol A, Fahima, Nevo E (2004) Microsatellites within genes: structure, function and evolution. Mol Biol Evol 21: 991-1007.

Li XB, Tai FJ, Wang XL, Xu WL (2007) Characterization and expression analysis of two cotton genes encoding putative UDP-Glycosyltransferases. Molecular Biology 42: 50-58.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25:402-408.

Liu Q, Kasuga, M, Sakuma y, Abe H, Miura S, Yamaguchi-Shinozaki K, 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, respectively, in *Arabidopsis*. Plant Cell 10:1391-1406.

Lo C, Chun-Hat Shih, Chu I, Yip W (2006) Differential expression of two flavonoid 3'-hydroxylase cDNAs involved in biosynthesis of anthocyanin pigments and 3-deoxyanthocyanidin phytoalexins in sorghum. Plant and Cell Physiology 10:1412-1419.

March R, Lewars E, Stadey C, Miao X, Zhao X, Metcalfe C (2006). A comparison of flavonoid glycosides by electrospray tandem mass spectrometry. International Journal of Mass Spectrometry 248:61-85.

McDowell J, Woffenden B (2003) Plant disease resistance genes: recent insights and potential applications. Trends in Biotechnology 21:178-183.

Meisel L, Fonseca B, González S, Baeza-Yates R, Cambiazo V, Campos R, Gonźalez M, Orellana A, Retamales J, Silva H (2005) A rapid and efficient method for purifying high quality total RNA from peaches (*Prunus persica*) for functional genomics analyses. Biol Res 38:83-88.

Mooney HA, Pearcy RW, Ehleringer J (1987) Plant physiological ecology today. BioSci 37:18-20.

Morgan JM (1984) Osmoregulation and water stress in higher plants. Ann Rev Plant Physiol 35:299-319.

Munz PA (1946) *Aquilegia* the cultivated and wild columbines. Gentes Herb 7:116.

Nishimura M, Kamada-Nobusada T, hayashi M, Fukazawa M, Sakakibara H (2008) A putative peroxisomal polyamine oxidase, AtPA04, is involved in polyamine catabolism in *Arabidopsis thaliana* Plant and cell physiology 49:1272-1282.

O'Farrell P (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007-4021.

Oren R, Sperry JS, Katul GG, Pataki DE, Ewers BE, Phillips N, Schäfer KVR (1999) Survey and synthesis of intran and interspecific variations in stomatal sensitivity to vapour pressure deficit. Plant Cell Enviro 22:1515-1526.

Petit JR, Jouzel J, Raynaud D, Barkov NI, Barnola J-M, Basile I, Bender m, Chappellaz J, Davis M, Delaygue G, Delmotte M, Kotlyakov VM, Legrand M, Lipenkov VY, Lorius C, Pépin L, Ritz C, Saltzman E, Stievenard M (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core Antarctica. Nature 399:429-436.

Pilon-Smits EAH, Ebskamp M, Paul M, Jeuken M, Weisbeek P, Smeekens SC (1995) Improved performance of transgenic fructan-accumulating tobacco under drought stress. Plant Physiology 107: 125-130.

Pyo HB, Kim JH, Lee BC, Kim JH, Sim GS, Lee DH, Lee KE, Yun YP, (2005) The isolation and antioxidative effects of Vitexin from *Acer palmatum*. Arch Pharm Res 28:195-202.

Qing-Yin, Zeng, Hai Lu, and Xiao-Ru Wang (2005) Molecular characterization of a glutathione transferase from *Pinus tabulaeformis* (Pinaceae). Biochimie 87: 445-455.

Reddy AR, Chaitanya KV, Vivekanandan M (2004) Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. J Plant Physiol 161:1189-1202.

Reinkkainene T (2000) Extreme halophiles synthesize betaine from glycine by methylation. J Biol Chem 275:22196-22201.

Rosenberg, IM (2006) Protein analysis and purification benchtop techniques. 2nd edition. Birkhauser, Boston.

Saito S, Hirai N, Matsumoto C, Ohigashi H, Ohta D, Sakata K, Mizutani M (2004) Arabidopsis CYP707As encode (+)-abscisic acid 8'-hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid. Plant Physiol 134:1439-1449.

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

Shinozaki K (2007) Regulatory metabolic networks in drought stress responses. Curren Opin Plant Biol 10:296-302.

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. J Exp Bot 58:221-227.

Shinozaki K (1999) Biological function of proline in morphogenesis and osmotolerance revealed in antisense transgenic *Arabidopsis thaliana*. Plant J 18:185-193.

Shinozaki K (2006) CYP707A3, a major ABA 8'-hydroxylase involved in dehydration and rehydration response in *Arabidopsis thaliana*. Plant J 46*:*171-182.

Shinozaki K , Umezawa T, Okamoto M, Kushiro T, Nambara E, Oono Y, Seki M, Kobayashi M, Koshiba T, Kamiya Y (2006) CYP707A3, a major ABA 8'-hydroxylase involved in dehydration and rehydration response in *Arabidopsis thaliana*. Plant J 46:71–182.

Smirnoff N, and Bryant J (1999) DREB takes the stress out of growing up. Nature Biotechnology 17: 229-230.

Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, Efstratiadis A (1994) Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci USA 91:9228-9232.

Socias X, Correia MJ, Chaves M, Medrano H (1997) The role of abscisic acid and water relations in drought responses of subterranean clover. J Exp Bot 48:1281-1288.

Sumner ER, Avery AM, Houghton JE, Robins RA, Avery SV (2003) Cell cycle- and age-dependent activation of Sod1p drives the formation of stress resistant cell subpopulations within clonal yeast cultures. Mol Microbiol 50:857-870.

Sun WQ, Leopold AC (1997) Cytoplasmic vitrification and survival of anhydrobiotic organisms. Comp. Biochem Physiol 117A:327–333.

Takabe T (2005) Genes for direct methylation of glycine provide high levels of glycine and abiotic-stress tolerance in *Synechococcus* and *Arabadopsis*. Proc Natl Acad Sci USA 102:1318-1323.

Tardieu F, Davies WJ (1992) Stomatal response to abscisic acid is a function of current plant water status. Plant Physiol 98:540-545.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:141-1452.

Thevelein JM (1996) Regulation of trehalose metabolism and its relevance to cell growth and function. In: The Mycota,. Brambl R, Marzluf GA (eds), Springer, Berlin, pp 395-420.

USDA (2008) Plants Profile for *Aquilegia vulgaris* (European columbine) USDA PLANTS. Retrieved July 15, 2008, from http://plants.usda.gov/java/profile?symbol=AQUIL

Voet D, Voet D (2005) Biochemistry. 3rd edition, John Wiley and Sons, INC Hoboken, NJ.

Yamaguchi-Shinozak K, Fujita M, Fujita Y, Shinozaki K ,Umezawa T (2006) Engineering drought tolerance in plants: Discovering and tailoring genes to unlock the future. Curr Opinions Biotech 17:113-122.

Yancey P, (2001) Water stress, osmolytes and proteins. American Zoologist 41:699-700.

Zhang Y, Fan W, Zhang Z (2009) Cloning and molecular characterization of fructose-1,6-biphosphate aldolase gene regulated by high-salinity and drought in *Sesuvium portulacastrum*. Physiology and Biochemistry 28:975-984.