Use of Bioinformatics to Investigate Abiotic Stress in Arabidopsis and to Design Primers for Pathogen Detection

Shrinivasrao Mane

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Plant Pathology, Physiology, and Weed Science

Ruth Grene, Ph.D., Chairman

Oswald R. Crasta, Ph.D

Glenda Gillaspy, Ph.D.

Lenwood S. Heath, Ph.D.

T. M. Murali, Ph.D.

April 9, 2007
Blacksburg, Virginia

Use of Bioinformatics to Investigate Abiotic Stress in Arabidopsis and to Design Primers for Pathogen Detection

by

Shrinivasrao Mane

Committee Chairman: Ruth Grene, Ph.D.

Plant Pathology, Physiology, and Weed Science

(ABSTRACT)

The focus of the work has been on computational approaches to solving biological problems. First, microarray analysis was used to study the role of PLDα 1 in drought stress in Arabidopsis. Second, a tool for designing and in-silico testing of primers for PCR-based pathogen detection will be discussed. Phospholipase D (PLD) has been implicated in a variety of stresses including osmotic stress and wounding. PLDα 1-derived phosphatidic acid interacts with ABI1 phosphatase 2C and promotes abscisic acid signaling. Plants with abrogated PLDα 1 show insensitivity to ABA and impaired stomatal conductance. My goal is to identify PLDα-mediated downstream events in response to progressive drought stress in Arabidopsis. *Arabidopsis thaliana* (Col-0) and antisense-PLDα 1 (Anti-PLDα) were drought stressed by withholding water. Anti-PLDα experienced severe water stress at the same time period that Col-0 experienced less water stress. Diurnal leaf water potential (LWP) measurements showed that Anti-PLDα had lower LWP than Col-0 under drought stress conditions. qRT-PCR revealed up to 18-fold lower values for Anti-PLDα transcripts in stressed Anti-PLDα plants when compared to stressed Col-0. Microarray expression profiles revealed distinct gene expression patterns in Col-0 and Anti-PLDα. ROP8, PLDδ and lipid transfer proteins were among the differentially expressed genes between the two genotypes. Different microarray analyses methods (TM4 and Expresso) were also compared on two different data sets. The results obtained from Expresso analysis were more accurate when compared with quantitative RT-PCR data.

Rapid diagnosis of disease-causing agents is extremely important since delayed diagnosis can result in disease spread and delayed prophylaxis. It is even more important in an era where disease-causing agents are used as bioterrorism agents. Rapid advances in sequencing technology have resulted in the sequencing of thousands of microorganisms in recent years. Availability of genomic

sequences has made it possible to identify and characterize microorganisms at the molecular level. PCR-based detection is powerful for pathogen diagnostics since it is rapid and sensitive. We have developed a tool, PathPrime, that can design primers, computationally test them against target genes, and potential contaminant sequences, and identify a minimum set of primers that can unambiguously detect a given list of sequences.

To my Family ...

My parents for their love and support

My brother for inspiring and encouraging me to pursue my dream

My wife Veena for her understanding, patience, and love

My son Mehul for bringing all the joy and happiness in my life

ACKNOWLEDGEMENTS

First, I would like to express gratitude to my advisor Dr. Ruth Grene for her valuable guidance, supervision, constructive criticism and constant encouragement in all aspects of my study and investigation.

I sincerely owe my deep sense of gratitude to Dr. Oswald Crasta for providing me guidance and support when I needed it the most. I thank him providing me the opportunity to work at VBI. Working at VBI provided me a lot of opportunities to do and learn new things. I would also like to thank Drs. Glenda Gillaspy, Lenwood Heath and T.M. Murali for their valuable guidance and help.

I thank my lab members Cecilia, Verlyn, Brooke and Julie for their help. I would also like to thank my friends and colleagues Allan, Amrita, Joshua, Nataraj, Mike and people in PATRIC.

I would like to thank my family and friends who have always supported and encouraged me. My parents receive my deepest gratitude and love for their endless love and encouragement. I would not have made it so far without their love and support.

I would like to express special appreciation to my wife, Veena, for her profound love, understanding, sacrifice, and tremendous support which made my education and scientific endeavor possible.

TABLE OF CONTENTS

1 Introduction **1**

1.1 Drought signaling . . . 2

1.1.1 Drought signal perception . . . 3

1.1.2 Drought signal transduction . . . 4

1.1.3 ABA-dependent pathway . . . 5

1.1.4 ABA-independent pathway . . . 7

1.2 Phospholipid signaling . . . 7

1.2.1 Phospholipase C . . . 8

1.2.2 Phospholipase D . . . 8

1.2.3 Phosphatidic Acid . . . 9

1.3 Microarray Technology . . . 12

1.3.1 DNA Microarray fabrication and use . . . 13

1.3.2 Data analysis: Normalization . . . 14

1.3.3 Data analysis: Assessing differential gene expression . . . 15

2 Responses of Wildtype and Antisense Plants to Drought Stress **20**

2.1 Introduction . . . 20

2.2 Materials and Methods . . . 22

2.2.1 Plant Material . . . 22

2.2.2 Progressive drought stress experiment . . . 22

2.2.3 RNA isolation . . . 22

2.2.4 DNase treatment . . . 23

2.2.5 Reverse Transcription . . . 23
2.2.6 Microarray experiment . . . 24
2.2.7 Quantitive-RT-PCR . . . 25
2.3 Results . . . 25
2.3.1 Drought stress experiment: Wildtype *vs* Antisense-GR2 . . . 25
2.3.2 Drought stress experiment: Wildtype *vs* Antisense-PLDα 1 . . . 26
2.3.3 Transcriptome Profiling . . . 26
2.4 Discussion . . . 30
2.4.1 Drought stress experiment: Wildtype *vs* Anti-GR2 . . . 30
2.4.2 Drought stress experiment: Wildtype *vs* Anti-PLDα 1 . . . 30

3 Microarray Analysis 64
3.1 Introduction . . . 64
3.2 Materials and Methods . . . 66
3.2.1 Median and integrated intensity values . . . 66
3.2.2 Microarray data sets . . . 67
3.2.3 Real-time quantitative RT-PCR . . . 68
3.2.4 Expresso analysis . . . 68
3.2.5 TM4 microarray software suite . . . 72
3.2.6 MIDAS data normalization methods and filters . . . 73
3.2.7 TM4 MEV analysis . . . 74
3.3 Results and Discussion . . . 75
3.3.1 Normalization and low intensity filtering in TM4 . . . 75
3.3.2 Choice of intensity signal data . . . 77
3.3.3 Comparison of Statistical Methods . . . 77
3.4 Conclusion . . . 79

4 PathPrime - A tool for designing and testing primers for disease diagnostics 95
4.1 Introduction . . . 95
4.1.1 Diagnostic methods . . . 95
4.1.2 Designing primers for PCR-based pathogen detection . . . 96
4.2 PathPrime workflow . . . 99

4.3 Biological Case Studies . . . 103
4.3.1 Designing primers for Lyssaviruses . . . 103
4.3.2 Validation of results . . . 104
4.3.3 Conclusion and Future Work . . . 104

5 Bibliography **109**

A List of all differentially expressed genes in Anti-PLDα stress vs Col-0 stress **133**

B PathPrime Perl Code **175**
B.1 Perl Code: PathPrime . . . 175
B.2 Perl Code: makePCRtemplate . . . 196

C PathPrime configuration files **199**
C.1 Primer Design parameters for Lyssaviruses . . . 199

LIST OF TABLES

2.1 Significantly regulated genes in Anti-PLDα control vs Col-0 control 36
2.2 Functional categorization of Drought Regulated Genes in Col-0 and Anti-PLDα . . . 47
2.3 Detailed functional characterization of drought regulated genes in Col-0 and Anti-PLDα 48
2.4 Drought-regulated signaling genes 50
2.5 Regulation of Small GTP-related proteins in Col-0 and Anti-PLDα 50
2.6 Comparison of Drought Regulated Genes in Col-0 and Anti-PLDα 51
2.7 Drought Regulated Genes Involved in Hormone Signaling 52
2.8 Drought-regulated lipid signaling proteins 53
2.9 Quantitative RTPCR of selected genes 53
2.10 Genes involved in metabolism 54
2.11 Differentially regulated transcription Factors 57
2.12 Differentially regulated signaling genes 59
2.13 List of primers used in qRT-PCR 63

3.1 Numbers of identified differentially expressed genes in the GP and GOT models after preprocessing by 0 or more MIDAS pipeline steps — IIV, $\mathrm{IIV_T}$, $\mathrm{IIV_{TL}}$, $\mathrm{IIV_{TLS}}$, or $\mathrm{IIV_{TLSF}}$. 86
3.2 Retention counts (RC) and retention percentages (RP) for the differentially expressed gene sets of Table 3.1 87
3.3 Comparison of MIV and IIV input data with respect to the sets of genes ultimately identified as differentially expressed. 88
3.4 Comparison of the number of genes identified as differentially expressed by Expresso analysis and the MEV *t*-test. 89

3.5 Comparison of qRT-PCR results with identified up-expressed and down-expressed genes in Experiment 2 using Expresso and TM4. 90
3.6 Correlation of qRT-PCR results from the Expresso GP model and the MEV t-test. . 91
3.7 Genes Subjected to qRT-PCR 94

4.1 Programs and Perl modules used by PathPrime. 102
4.2 Starting data for the primer design process. 105
4.3 Primer-pair sequences and number of Lyssavirus genomes amplified. 107
4.4 Sequences amplified by each primer pair 108

A.1 Significantly regulated genes in Anti-PLDα stress vs Col-0 stress 133

LIST OF FIGURES

1.1 Formation of phosphatidic acid by PLC and PLD pathways. 10
1.2 The domain structures of the Arabidopsis PLD protein family. 17
1.3 A model for PLD and PA in signaling ABA responses 18
1.4 Result of normalization of microarray data by three different methods. 19

2.1 Drought stress experimental setup in growth chamber 38
2.2 Time-points for leaf water potential and photosynthesis measurements 38
2.3 Leaf water potential in Col-0 and Anti-GR2 39
2.4 Net photosynthesis in Col-0 39
2.5 Net photosynthesis in Anti-GR2 40
2.6 Leaf water potential in Col-0 and Anti-GR2 after rehydration 40
2.7 Leaf water potential in Col-0 and Anti-PLDα 41
2.8 Col-0 and Anti-PLDα after 10 days of water stress 42
2.9 Col-0 and Anti-PLDα recovery at 16 hours after rehydration. The stressed Col-0 and Anti-PLDα 1 were rehydrated after the end of $10^t h$ day. LWP and photosynthesis measurements were taken the following afternoon. 43
2.10 Photosynthesis in Col-0 and Anti-PLDα 44
2.11 Venn diagram showing differential expression of genes in Col-0 and Anti-PLDα . . . 45
2.12 Expression of PLDδ in Col-0 and Anti-PLDα 46

3.1 The overall flow of the statistical analyses of microarray data 80
3.2 RI-plots illustrating specific intensity-dependent dye bias as microarray data is processed. 81

3.3 Comparison of estimated $\log_2$(fold change) and the corresponding p-value estimates for the GP and GOT model results of the WS ecotype values in Table 3.4 82
3.4 Comparison of qRT-PCR results to Expresso and MEV t-test results for first 16 of 50 selected genes of the Col-0 genotype of Experiment 2. 83
3.5 Comparison of qRT-PCR results to Expresso and MEV t-test results for the middle 17 of 50 selected genes of the Col-0 genotype of Experiment 2. 84
3.6 Comparison of qRT-PCR results to Expresso and MEV t-test results for the last 17 of 50 selected genes of the Col-0 genotype of Experiment 2. 85

4.1 The PathPrime workflow. 100
4.2 Locations of the amplicons produced by the 3 primer-pairs 106

Chapter 1

Introduction

Drought is one of the most important constraints of food production in many areas of the world. On a global basis, drought, in conjunction with coincident high temperature and radiation, poses the most important environmental constraint to plant survival and to crop productivity. The effects for drought spread far beyond the physical effects of reduced water availability. Drought impacts economic, environmental and social aspects of life. In fact, drought affects more people than any other natural hazard (Wilhite, 2000).

According to the National Drought Mitigation Center, loss in agriculture in the United States due to drought was estimated to be approximately $11 billion in 2002 (Hayes *et al.*, 2004). With the global shortages of water now emerging, reducing water consumption in crop production has now been generally recognized as an essential strategy for sustainable agriculture. It has also been gradually recognized as an important strategy for crop production, even for areas where water supplies are currently abundant. In addition, reduced levels of irrigation will decrease levels of water contamination and energy consumption, thus producing a significant positive impact for our environmental conservation efforts.

Drought stress affects all stages of plant growth and development. Indeed, physiological drought also occurs during cold and salt stresses, when the main damage caused to the living cell can be related to water deficit (McCue & Hanson, 1990). At a cellular level, among other events, it is

caused by denaturation of proteins caused by loss of water in the cell or movement of solutes that can change the pH and affect protein stability.

Plants have developed different mechanisms to cope with drought stress. These mechanisms include physical adaptations as well as interactive molecular and cellular changes. Molecular responses to stress begin when the stress is perceived, this perception initiates signal transduction pathways, through which physiological changes occur and genes are activated, modifying molecular and cellular processes. Although significant progress has been made towards understanding the underlying molecular mechanisms of drought tolerance much remains to be learned (Cushman & Bohnert, 2000; Hasegawa *et al.*, 2000; Bartels & Salamini, 2001; Zhu, 2002; Seki *et al.*, 2003; Bray, 2004; Zhang *et al.*, 2005; Anna *et al.*, 2005). Current knowledge of the regulatory network governing the drought-stress responses is incomplete, with little information on signal perception. However, signal transduction and the promoter modules of response genes, are starting to be elucidated.

Some of the most recent efforts to understand gene function have used molecular genetics, and these studies have significant implications for crop development. Plant breeding has already provided an enormous improvement in the drought tolerance of crop plants, with selection often allowing desired traits to be transferred from close wild relatives. Root architecture is one of the most important traits for drought tolerance (Malamy, 2005). A dynamic root system is fine-tuned to soil moisture status and is known to regulate the amount of water available to the plant depending on its distribution in the soil (Toorchi *et al.*, 2004; Eapen *et al.*, 2005). Several studies have showed correlation between drought tolerance and root morphology (Pinheiro *et al.*, 2005; Yue *et al.*, 2005). Although several quantitative trait loci (QTLs) related to root morphology have been identified, their molecular basis is often not clearly understood.

1.1 Drought signaling

Plants perceive and respond to drought stress by altering transcriptional, biochemical and physiological processes. The mechanism of molecular responses of plants to drought stress has been extensively studied (Ingram & Bartels, 1996; Rabbani *et al.*, 2003; Rizhsky *et al.*, 2004; Seki *et al.*,

2002; Yamaguchi-Shinozaki & Shinozaki, 2006). Several genes have been studied that respond to drought, salt or cold stress at the transcriptional level. Genes that respond to osmotic stress caused by water deficit and/or high salt stress also respond to cold stress (Zhu *et al.*, 1997; Thomashow, 1998; Zhu, 2002; Zhang *et al.*, 2004a). The products of these stress-inducible genes can been classified into two groups: those that directly respond to and protect against environmental stresses and those that regulate gene expression and signal transduction in the stress response. The first group includes proteins that function by protecting cells from dehydration, such as the enzymes required for the biosynthesis of various osmoprotectants, late embryogenesis abundant proteins, antifreeze proteins, chaperones, and detoxification enzymes (Bartels & Salamini, 2001; Goyal *et al.*, 2005; Rontein *et al.*, 2002; Yamaguchi-Shinozaki & Shinozaki, 2006). The second group of gene products includes transcription factors, protein kinases, phosphatases, G-protein heteromers and monomers, and enzymes involved in phospholipid signaling (Agarwal *et al.*, 2006; Wang *et al.*, 2006; Yamaguchi-Shinozaki & Shinozaki, 2006; Zhang *et al.*, 2004b). Stress-inducible genes have been used to improve the stress tolerance of plants by gene transfer (Hasegawa *et al.*, 2000; Yamaguchi-Shinozaki & Shinozaki, 2006).

1.1.1 Drought signal perception

Despite accumulating vast knowledge about drought signal transduction such as transcription factors, chromatin-remodeling factors, and RNA-processing proteins, second messengers and phosphoprotein cascades, nothing is known about how the physical signals of dehydration stress are perceived by the roots and converted into biochemical signals (Kacperska, 2004; Verslues & Zhu, 2005; Nilson & Assmann, 2007). The most common model of sensing an external stimuli is the sensing of chemical ligand binding to a receptor. However, water does not have a chemical ligand that can sensed by receptors. Other factors such as changes in turgor, membrane pressure, molecular agglomeration are suspected to be the stimuli (Morgan, 1984; Wood, 1999). Several physiological experiments have shown that changes in cell turgor due to stress can be sensed as mechanical change in cell shape and cell volume (Jia *et al.*, 2001). Hsiao (1973) has argued that physical properties of water change little over a range of water potential sensed by plants whereas turgor changes substantially. This may be sensed as the initial stimulus for water deficiency in the cell.

Studies on yeast osmosensing mechanisms have given some molecular clues. Synthetic Lethal

of N-End Rule1 (*SLN1*), a two-component histidine kinase, senses cellular turgor pressure in yeast by an unknown mechanism (Reiser *et al.*, 2003). Another osmosensing protein, SH3-Domain Osmosensor 1 (SHO1) with a very similar function also been found in yeast (Tamas *et al.*, 2000). Two Arabidopsis histidine kinase genes, *Arabidopsis thaliana* Histidine Kinase1 (*ATHK1*) and Cytokinin Response1 (*CRE1*) were shown to complement *sln1* deletion mutants in yeast (Reiser *et al.*, 2003). *CRE1* can also respond to changes in turgor pressure when expressed in yeast (Urao *et al.*, 1999). *ATHK1* is highly expressed in roots especially under both low and high osmolarity conditions. However, these proteins have not yet been shown to function as osmosensors in plants.

1.1.2 Drought signal transduction

Specific signaling events that precede and control ABA accumulation are still lacking in plants. Studies in yeast stress signal transduction provide some clues. Upon perception of stress signals in yeast, SLN1 and SHO1 are shown to activate a specific MAPK (mitogen-activated protein kinase) signaling cascade commonly referred to as HOG1 (Brewster *et al.*, 1993). HOG1 was the first component of the MAPK signaling pathway to be discovered. Genes of all three components of MAPK signaling cascade (MAPK, MAPK kinase and MAPK kinase kinase) have been shown to be upregulated during dehydration stress in Arabidopsis (Mizoguchi *et al.*, 2000; Ichimura *et al.*, 2000; Shinozaki *et al.*, 2003). These genes are also shown to interact to produce a fully functional signaling cascade in yeast. In addition to dehydration stress, other abiotic stresses also activate protein phosphorylation activity of some MAPKs in Arabidopsis (Mizoguchi *et al.*, 1996; Chinnusamy *et al.*, 2004). MPK4 and MPK6 are activated by cold, salt, drought, wounding and touch (Ichimura *et al.*, 2000). Burnett *et al.,* (2000) observed that treatment with MAPKKK inhibitor, 098059, inhibited ABA-regulated stomatal activity as well as the ABA-induced gene expression in the epidermal peels from pea. This suggests the involvement of MAP kinase cascades in ABA signal transduction pathways. Drought stress activates many ABA-dependent and ABA-independent pathways. Both pathways will be reviewed in detail.

1.1.3 ABA-dependent pathway

Abscisic Acid

Abscisic acid was first identified and characterized in 1963 by a group that was studying compounds responsible for the abscission of fruits in cotton. It was called "abscisin II" originally because it was thought to play a major role in abscission of fruits (Addicott *et al.*, 1964). ABA influences most aspects of plant growth and development to some extent. These include stomatal movement, root and shoot growth, synthesis of storage protein in seeds, dormancy, germination of seeds, embryogenesis and response to biotic and abiotic stresses (Davies, 1995; Salisbury & Ross, 1992).

ABA is the most important drought-responsive hormone in plants (Evans & Hetherington, 2001; Hetherington, 2001; Desikan *et al.*, 2004). ABA production increases in response to drought stress. ABA accumulation is controlled by several metabolic processes. These processes include ABA synthesis, catabolism and/or ABA conjugation. Forward genetic analysis has revealed upstream signaling components of ABA. The most important among them being NCED (9-*cis*-epoxycarotinoid dioxygease) gene family. NCED catalyses the cleavage of the C_{25} carotenoids 9-*cis*-neoxanthin or 9-*cis*-violaxanthin to the C_{15} ABA precursor xanthoxin (Schwartz *et al.*, 2003). This reaction is the rate limiting step in ABA biosynthesis (Chernys & Zeevaart, 2000; Iuchi *et al.*, 2001; Thompson *et al.*, 2000). Arabidopsis has 9 NCED family genes among which NCED3 is the most strongly induced gene by dehydration (Iuchi *et al.*, 2001; Tan *et al.*, 2003).

Several functions are thought to be affected by ABA in response to drought. In plants, 90% of water loss occurs mainly due to transpiration via stomatal pores (Pei *et al.*, 1998). ABA reduces water loss by inducing stomatal closure in leaves (Evans & Hetherington, 2001; Hetherington, 2001; Schroeder *et al.*, 2001; Nambara & Marion-Poll, 2005). Open stomata allow for CO_2 influx and water efflux thereby assisting the transpirational process (Hetherington, 2001; Schroeder *et al.*, 2001; Hetherington & Woodward, 2003). During water stress, plants redistribute foliar ABA and upregulate ABA synthesis in roots, which is translocated to the aerial portion of the plant leading to roughly a 30-fold increase in ABA concentration in the of guard cell apoplast (Pei & Kuchitsu, 2005; Sauter *et al.*, 2001). Several cellular components essential for ABA signal transduction have been identified by physiological experiments. Cytosolic Ca^{2+} rapidly increases within minutes of

ABA exposure (McAinsh *et al.*, 2000; Allen *et al.*, 2001; Webb *et al.*, 2001; Young *et al.*, 2006). The level of cytosolic Ca^{2+} is dependent upon phospholipase C activity (Staxen *et al.*, 1999; Hunt *et al.*, 2004), and transient increases in inositol 1,4,5-triphosphate (IP_3) (Lee *et al.*, 1996; Lecourieux *et al.*, 2006). IP_3 levels increase by 90% ten seconds after 10 μM ABA addition (Lee *et al.*, 1996; Lecourieux *et al.*, 2006). The role of IP_3 in stomatal movement has also been reported (Gilroy *et al.*, 1990).

Stomatal closure requires many components including an increase in cytosolic Ca^{2+} (Gilroy *et al.*, 1990; Allan *et al.*, 1994; McAinsh *et al.*, 1995, 2000). ABA triggers stomatal closure by increasing cytosolic Ca^{2+} levels (Schroeder *et al.*, 2001) as ABA induces cytosolic Ca^{2+} oscillations (Gilroy *et al.*, 1991; McAinsh *et al.*, 1992; Grabov & Blatt, 1998; Allen *et al.*, 1999; Schroeder *et al.*, 2001). The sensitivity of stomata to ABA is dependent on the pH of the extracellular environment (Prokic *et al.*, 2006). The culmination of these events is the inhibition of plasma membrane H^+-ATPases and inward rectifying K^+ channels (Schroeder *et al.*, 2001). Activation of outward rectifying K^+ channels allows for ion efflux from the guard cell and a decrease in guard cell turgor. This results in closure of the guard cells (Blatt *et al.*, 1990; Lemtiri-Chlieh *et al.*, 2000; Lemtiri-Chlieh & MacRobbie, 1994; Schroeder *et al.*, 2001; Pei & Kuchitsu, 2005).

Despite such knowledge, no receptors for ABA had been found until recently. Two ABA receptors have been identified (Razem *et al.*, 2006; Shen *et al.*, 2006). FCA is an ABA receptor involved in RNA metabolism and in controlling flowering time (Razem *et al.*, 2006). ABA-binding protein (ABAR) the gene for which encodes the H subunit of Mg-chelatase (CHLH) specifically binds ABA, and mediates ABA signaling as a positive regulator in seed germination, post-germination growth and stomatal movement Shen *et al.* (2006).

Many drought, cold and salt stress inducible genes are also responsive to ABA (Zhu *et al.*, 1997; Shinozaki & Yamaguchi-Shinozaki, 2000). The promoters of these genes contain a motif known as ABA Responsive Element (ABRE). These ABRE motifs are known target binding sites for b-Zip transcription factors. Responsive to desiccation 29B (*RD29B*) gene contains two copies of ABRE in its promoter, and therefore, is ABA dependent and is induced by drought and high salinity. Another ABA-dependent pathway involves genes containing MYC and MYB recognition sequences. Respon-

sive to desiccation 22 (*RD22*) is controlled by MYC and MYB transcription factors (Urao *et al.*, 1993; Abe *et al.*, 1997, 2003). ABA induces expression of several stress responsive genes including *RAB18*, *cor15a* , *kin1* and *kin2* (Lang & Palva, 1992; Wilhelm & Thomashow, 1993; Wang & Cutler, 1995; Kurkela & Borg-Franck, 199).

1.1.4 ABA-independent pathway

The ABA independent pathway involves genes that respond to both drought and cold stress containing strikingly similar elements in their promoter regions, namely *cis*-acting dehydration-responsive (DRE), and the cold-responsive (C repeat) elements. These elements share the core motif CCGAC (Yamaguchi-Shinozaki & Shinozaki, 1994; Haake *et al.*, 2002). Transcription factors that bind to DRE/C repeat (DREB-DRE binding; CBF- C-repeat binding) have been isolated. *RD29A* contains both ABRE and DREB motifs and can respond to drought, cold and salt stresses. The combination of ABRE and DREB motifs in a single promoter provides a mechanism for crosstalk between ABA-dependent and independent pathways. The timing of induction of different transcription factors coupled with diverse combination of regulatory elements in the promoter regions of their target genes lead to a complex network of gene regulation in response to drought and other abiotic stresses (Shinozaki & Yamaguchi-Shinozaki, 2000).

1.2 Phospholipid signaling

Phospholipids are the backbone for membranes and are important components of lipid metabolism. They also serve as rich sources of signaling messengers. Phospholipases not only affect the structure and stability of cell membranes but also regulate many cellular functions. However, unlike structural phospholipids, signaling phospholipids are present in minute amounts, but their levels increase in response to appropriate stimuli. Activation of phospholipases is often the first step in generating lipid-based signal messengers. Phospholipase C (PLC) and phospholipase (PLD) have emerged as components of drought-responsive signal transduction pathways (Hirayama *et al.*, 1995; Sang *et al.*, 2001b).

1.2.1 Phospholipase C

PLC signaling represents the paradigm for phospholipid-based signal transduction. Upon activation, PLC hydrolyses the minor lipid phosphatidylinositol-4-5-bisphosphate (PIP_2) into two second messengers: inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). IP_3 releases Ca^{2+} from intracellular stores whereas DAG activates certain members of the protein kinase C (PKC) superfamily. During osmotic stress, plant cells may increase the production of PIP_2 by upregulating the expression of *PI5K* (Mikami *et al.*, 1998), a gene that encodes a phosphatidylinositol 4-phosphate 5-kinase (PIP5K) functioning in the production of PIP_2. Several genes encoding components of the PLC system are strongly expressed in response to drought and/or salt stress. These include PLC, PIP5K and DGK (Katagiri *et al.*, 1996). Under drying conditions, the plant must prevent water loss via transpiration and must therefore control stomatal aperture. There is evidence that PLC is involved in the closing mechanism. ABA, the phytohormone produced during water stress, also induces stomatal closure and this can be blocked by a PLC-inhibitor.

1.2.2 Phospholipase D

Phospholipase D is a ubiquitous enzyme in plants. It plays diverse roles in plant cells including in programmed cell death, cell growth, cell patterning, vesicular trafficking, cytoskeletal changes, root growth, abiotic stress responses and the oxidative burst (Lee *et al.*, 2003; Li *et al.*, 2004; Ohashi *et al.*, 2003; Potocký *et al.*, 2003; Sang *et al.*, 2001b; Wang, 2005). PLD cleaves the terminal phosphodiester bond of phospholipids to phosphatidic acid (PA) and water soluble free-head groups (Fig 1.1). Arabidopsis contains 12 distinct PLD genes (Eliáš *et al.*, 2002). In comparison, mammals posses two PLDs. Eukaryotic PLDs are characterized by two highly conserved carboxy-terminal (C-terminal) domains and an amino-terminal (N-terminal) lipid-binding region (Fig 1.2). PLDs also posses two highly conserved catalytic HxKxxxxD (HKD) motifs that interact and play an essential role in lipase activity. The plant PLDs can be grouped into two subfamilies: C2 -domain-containing and phox homology (PX)-pleckstrin homology (PH)-domain containing PLDs. C2-subfamily contains 10 Arabidopsis isoforms containing the phospholipid/Ca^{2+}-binding domain called the C2/CalB domain. In general, C2-PLDs hydrolyze different membrane lipids such as PA, phosphatidylethanolamine (PE) and phosphtidylglycerol (PG) *in vitro*. However, different PLDs have differing substrate preferences

(Wang, 2002; Qin *et al.*, 2002). The PLDs can be further classified into six subfamilies. Five C2-PLDs namely, PLDα(-1,-2,-3,-4), PLDβ(-1,-2), PLDγ(-1,-2,-3) and PLD δ. PLDζ(-1,-2) from the PH/PX family.

The various PLDs have different requirements for Ca^{2+}, phosphoinositides and varied substrate preferences (Wang, 2000). In addition, new regulatory motifs that interact with G proteins, calcium and phosphoinositides have also been identified. The molecular and biochemical heterogeneities of plant PLD family play key roles in the regulation of diverse functions of individual PLDs. PLDα 1 directly interacts with Gα, the only canonical α-subunit of the heterotrimeric G protein in Arabidopsis (Zhao & Wang, 2003). PLDα 1-Gα interaction inhibits PLDα activity, stimulating intrinsic GTPase activity of Gα (Zhang *et al.*, 2004c), and regulates stomatal movement (Mishra *et al.*, 2006). ABA-independent activation of PLD by Gγ subunit has also been reported in barley aleurone (Ritchie & Gilroy, 2000).

Depression of PLDα 1 activity by antisense suppression or its ablation by knockout mutation resulted in alteration in several stress-related physiological processes such as reactive oxygen production (Sang *et al.*, 2001a), abscisic acid signaling (Fan *et al.*, 1997), wounding response (Wang *et al.*, 2000), freezing tolerance (Welti *et al.*, 2002) and increased water loss (Sang *et al.*, 2001b). PLDα 1-silenced Arabidopsis plants showed impaired response to ABA treatment and/or water stress (Sang *et al.*, 2001b). ABA is known to induce production of PLDα 1 (Fan *et al.*, 1997; Sang *et al.*, 2001b; Zhang *et al.*, 2004b). The role of PLD-derived PA in drought responses is discussed in further detail below (Section 1.2.3)

1.2.3 Phosphatidic Acid

Phosphatidic acid (PA) is the simplest form of membrane phospholipid. PA molecular species plays an important role in membrane biogenesis and glycerolipid metabolism. It is also a key intermediate in synthesis of membrane lipids and storage lipids. In recent years, PA has emerged as an important second messenger (Jenkins & Frohman, 2005; Testerink & Munnik, 2005). The PA involved in cell signaling is generated via two phospholipase pathways, PLD and PLC (Figure 1.1). PA is formed in response to several biotic and abiotic stresses including salinity, drought, hyperosmotic stress,low

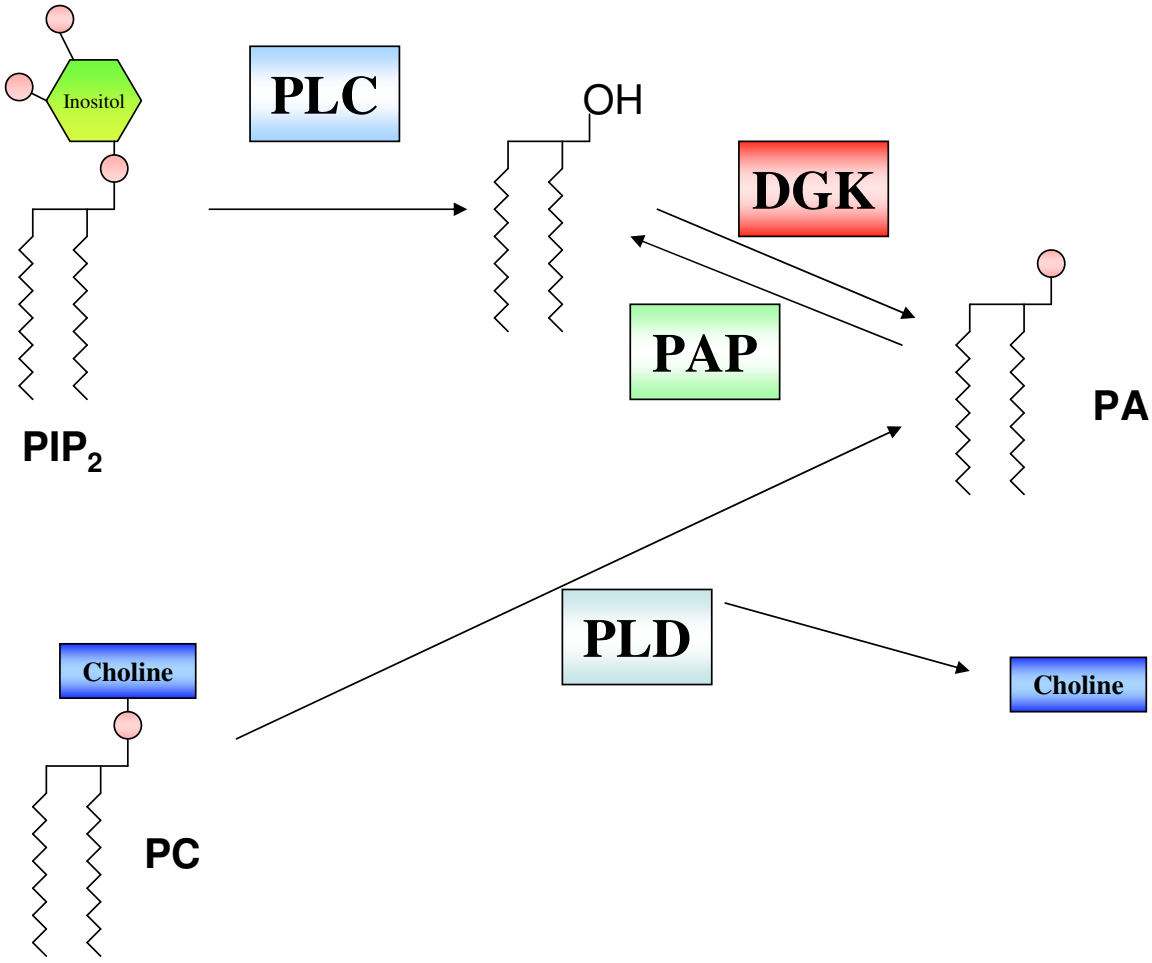


Figure 1.1: Formation of phosphatidic acid by PLC and PLD pathways. Abbreviations: DAG, diacylglycerol; P, phosphate; PAP, PA phosphatase; PC, phosphatidylcholine; PIP_2, phosphatidylinositol-(4,5)-bisphosphate; PLC, phospholipase C; PLD, phospholipase D.

temperature, oxidative stress, and pathogen attack. During stress, ABA induces production of PA via PLDα pathway (Fan *et al.*, 1997; Sang *et al.*, 2001b; Zhang *et al.*, 2004b). PA also plays important roles in growth and development such as seed germination, polarized cell growth, cell patterning, ion flux, cytoskeletal changes, anti-cell death, secretion and vescicular trafficking (Wang *et al.*, 2006).

Arabidopsis plants with antisense or knockout PLDα 1 showed impaired stomatal closure in response to ABA and were more sensitive to drought stress (Sang *et al.*, 2001b). External application of PA induces stomatal closure (Zhang *et al.*, 2004b). PLDα 1-derived PA has been shown to interact with ABI1, a protein phosphatase 2 C (PP2C)-like enzyme (Zhang *et al.*, 2004b). ABI1 is a negative regulator of ABA responses. ABI1 was initially identified by analysis of mutants displaying altered ABA responses (Leung *et al.*, 1994). PA binds to ABI1 and reduces its translocation to the nucleus by tethering it to the plasma membrane, thus promoting ABA responses. In addition, PLDα 1 directly interacts with the Gα-subunit of the heterotrimeric protein in Arabidopsis thereby reciprocally affecting activities of both PLDα 1 and Gα (Figure 1.3). GDP-bound Gα inhibits PLD activity where as PLDα 1 has GTPase-activating protein (GAP) activity which inactivates Gα function. PLDα 1 and PA interact with PP2C to signal ABA promoted stomatal closure, whereas PLDα 1 interacts with the Gα-subunit to mediate ABA inhibition of stomatal opening (Mishra *et al.*, 2006).

Reactive oxygen species (ROS) are produced by cells in response to biotic and abiotic stresses. PLDα 1 and PA have been implicated in increasing NADPH oxidase activity and ROS production (Sang *et al.*, 2001a). PA also plays a role in ROS responses. In Arabidopsis, PLDδ is activated by H_2O_2 and the PA derived from PLDδ leads to a decrease in H_2O_2-induced programmed cell death (Zhang *et al.*, 2003). Antisense PLDα 1 plants are more tolerant to cold stress than wildtype (Welti *et al.*, 2002). It has been suggested that increased PLDα 1 activity converts PC to PA in the membranes thereby destabilizing membrane integrity and increasing membrane leakage. PA is also produced rapidly in response to both hyper- and hypo-osmotic stresses (Testerink & Munnik, 2005). PA increases rapidly in cell suspension cultures in tomato during salt stress and in the dehydrated leaves of resurrection plant *Craterostigma plantagineum* (Munnik *et al.*, 2000).

In mammalian cells, several PA targets have been identified (Ktistakis *et al.*, 2003). These include protein kinases, protein phosphastases, phosphodiesterases, ATPase, small GTPases, sph-

ingosine kinase and protein kinase C (Delon *et al.*, 2004; Ghosh *et al.*, 1996; Jenkins *et al.*, 1994; Lopez-Andreo *et al.*, 2003; Manifava *et al.*, 2001). In plant systems, less is known. Apart from ABI1, PA has also been shown to bind PDK1, a protein kinase which acts upstream of a MAP kinase pathway (Anthony *et al.*, 2004; Rentel *et al.*, 2004).PA binding to PDK1 results in activation of the oxidative stress-response protein kinase OXI1/AGC2-1 (Anthony *et al.*, 2006). PA also binds to MAPK6-related protein, an important mediator in stress and ethylene signaling, in soybean cells (Liu & Zhang, 2004). Several other components of the ABA pathway are thought to be targets of PA. For example, sphingosine kinase, an enzyme that produces sphingosine-1 phosphate, has been recently implicated in ABA signaling in plants (Coursol *et al.*, 2003). In mammalian cells, sphingosine kinase 1 (SK1) binds PA and translocates to PA-enriched membranes (Delon *et al.*, 2004). This is yet to be shown in plants. Other PA binding targets have been identified in a proteome-wide study (Testernik *et al.*, 2004). These include SnRK2 protein kinase, a PP2A phosphatase regulatory subunit (RCN1), two HSP90 isoforms and DRG1, a new class of G proteins (Testernik *et al.*, 2004). However, an overall understanding of any PA-mediated drought responses is still lacking for plants.

1.3 Microarray Technology

Rapid progress has been made in whole genome and large-scale EST sequencing projects in the recent years. Completed genome sequences of dozens of organisms, of which three are plants (Arabidopsis, rice and poplar; Arabidopsis Genome Initiative, 2000; Goff *et al.*, 2002; Yu *et al.*, 2002; Tuskan *et al.*, 2006), have been launched or completed. Although allowing cross-genome comparisons, genomic sequences do not yield much information about respective functions of genes, how they bring about living cells nor how those cells are organized into tissues and whole organisms. Such questions form the basis of functional genomics. The main goal of functional genomics is not only to assign function to each gene in an organism, but also to understand how they work together to make functioning cells and organisms.

Microarrays are one of the most powerful technologies to take advantage of sequence data generated by genome sequencing projects. Microarray technology utilizes nucleic acid hybridization techniques and advanced computational data analysis methods to evaluate the mRNA expression

profile of thousands of genes in a single experiment. DNA arrays can be used for many different purposes including expression profiling (Schena *et al.*, 1995), identification and genotyping of mutant alleles and DNA marker polymorphisms (Wang *et al.*, 1998), DNA resequencing (Hacia, 1999) and DNA-protein interactions, commonly known as Chromatin-immunoprecipitation (Lee *et al.*, 2006). The major types of DNA arrays are spotted arrays and synthesized arrays. Oligonucleotide of 25-mer in length and are directly synthesized on the array by photolithography (Affymetrix) (Lockhart *et al.*, 1996). Spotted arrays consist of cDNA or 70-mer long oligonucleotides spotted on slides. My primary focus will be on discussing long-oligonucleotide arrays.

1.3.1 DNA Microarray fabrication and use

DNA microarrays are generated by physically depositing oligonucleotides to known locations on glass surfaces (typically microscope slides coated with any of a number of substances that foster binding to DNA). Microarrays use the same DNA probe detection principle used by southern blotting (Southern, 1975). A two-probe microarray experiment consists of several general steps: fabrication of microarrays, RNA isolation, probe labeling and hybridization, image analysis, data extraction, and data management and analysis (Duggan *et al.*, 1999). In the first step, 70-mer synthesized oligonucleotides are spotted and immobilized onto a glass slide bearing chemically modified surfaces (usually amino-silane) using a robotic printing device. In a single glass side, typically tens of thousands of oligonucleotides can be spotted. Next, cDNA is obtained by reverse transcribing RNA from two different treatment groups and further labeled separately with one of two fluorescent dyes (typically N,N8-(dipropyl)- tetramethylindocarbocyanine (Cy-3-) or N,N8-(dipropyl)-tetramethylindodicarbocyanine (Cy-5-dUTP); Shalon *et al.,* 1996). Instead of cDNA, cRNA can also be used for hybridization. The cDNA pools are then mixed and hybridized to the array. cDNAs from the mixed pool hybridize to homologous sequences spotted on the array. The unbound cDNA is removed from the slide by washing. The signal intensities are recorded using two separate wavelengths depending upon the dye molecules used. The images resulting from the detection process are further used to quantify signal intensities of each spot.

1.3.2 Data analysis: Normalization

Data normalization is the first step in analyzing microarray data after background subtraction. Normalization is necessary to remove systematic errors introduced at various stages of a microarray experiment. The purpose of normalization is to adjust for effects which arise from variation in the microarray technology. Imbalances between the two (say red and green) fluorescent dyes may arise from differences between the labeling and detection efficiencies, by the use of different scanner settings, non-uniform hybridization or the initial RNA from the two samples used in the assay, in addition to differences in printing tips or uniformity of glass slides (Quackenbush, 2001). Usually the imbalance is more complicated than a simple scaling of one channel relative to the other. The degree of dye bias varies with intensity and position on the slide. Between-array difference can also arise from differences in experimental conditions, scanner settings and print quality. Therefore, between and within array normalization needs to be performed.

There are many approaches to normalizing expression data. According to Dobbin *et al.* (2005) dye bias can be divided into four different types.

Type I Same for all genes in an array.

Type II Dependence on overall spot intensity.

Type III Dependence on subset of genes.

Type IV Dependence on combination of genes and samples.

Type I dye bias is assumed to be the same for all genes on an array, therefore it can be removed by median centering of an array, a variant of global normalization. Thus, the red (Cy5) and green (Cy3) intensities are related by a constant factor k.

$$k = median\frac{R}{G} \tag{1.1}$$

$$R' = R, G' = kG \tag{1.2}$$

$$log_2 = \frac{R'}{G'} = log_2\frac{R}{kG} = log_2\frac{R}{G} - log_2 k \tag{1.3}$$

$$M = log_2\frac{R}{G}, A = \frac{1}{2}log_2(RG) \tag{1.4}$$

where R and G represent the background-corrected red and green intensities for each spot, and R' and G' indicate normalized values of R, G respectively.

This normalization brings only the shift of M values in an 'MA-plot' without changing the shape of scatter plot where M means the log intensity ratio and A the mean log intensity for each spot (Fig. 1.4B). Type II dye bias is thought to be dependent on the overall spot intensity curve. Lowess smoothing, a locally weighted regression, is the most widely used normalization method for intensity dependent bias (Yang *et al.*, 2001). Global lowess normalization can efficiently remove intensity dependent dye bias (Fig. 1.4C). Type III bias is due to specific subset of genes. Reciprocal dye swapping experiment can remove gene-specific bias. Finally, Type IV dye bias depends on both genes and samples. No systematic method is available to eliminate type IV dye bias (Dombkowski *et al.*, 2004). In addition to dye bias described above, spatially dependent dye bias occurs from print-tips. Yang *et al.* (2002) proposed the print-tip lowess normalization method to remove this bias. Figure 1.4D shows that the print-tip lowess removes intensity-dependent bias for each print-tip is easily which was not completely removed by global lowess method.

1.3.3 Data analysis: Assessing differential gene expression

The most common use of expression data is the identification of differentially expressed genes. Earlier studies used a cutoff of two or more fold expression change between two comparisons to identify differentially expressed genes (Schena *et al.*, 1996; Khan *et al.*, 1999). Using such arbitrary thresholds fails to identify biologically important genes that may have very small fold change yet consistent across biological replicates. Many robust statistical methods are available to identify differentially expressed genes. The mixed model ANOVA method reliably detect changes in gene expression (Wolfinger *et al.*, 2001; Chu *et al.*, 2002). In this model, statistical significance is not based on fold-change, but, rather, on how well the data for that gene fits the model and how well the data from repeats fit. This more robust estimation of significance allows identification of differentially expressed genes with confidence, even at relatively small fold changes. Conversely, the mixed model may reject large fold changes for which the variance is very high. The mixed model approach, which is widely used (Byrne *et al.*, 2005; Kirst *et al.*, 2005; Brinker *et al.*, 2004; Tan *et al.*, 2003) has been found to rank highly among other methods for analysis of gene expression data (Reverter *et al.*, 2005;

Allison *et al.*, 2006). Many groups have employed a fold change cutoff value combined with a value of statistical significance to identify differentially regulated genes (Aharoni *et al.*, 2000; Arbeitman *et al.*, 2002; Guo *et al.*, 2006). Another powerful use of microarrays is the development of global expression profiles by clustering (DeRisi *et al.*, 1997; Hughes *et al.*, 2000; Arbeitman *et al.*, 2002). Clustering allows one to identify genes that show similar expression patterns because genes with similar expression behavior are more likely to be functionally related.

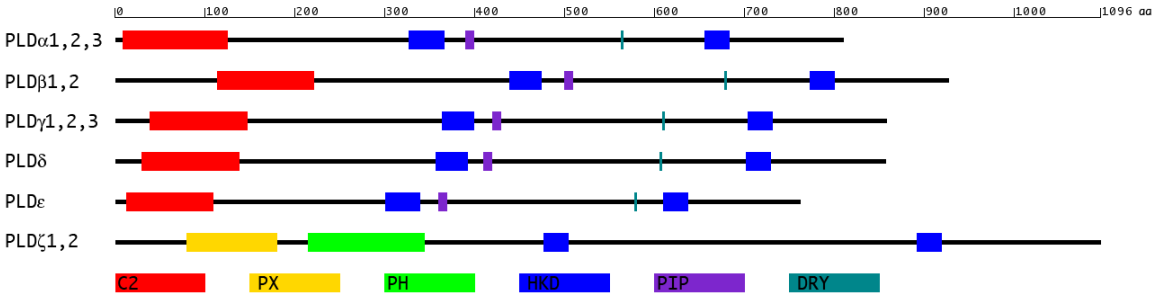


Figure 1.2: The domain structures of the Arabidopsis of PLD family. Plant PLDs consist of two distinctive groups: the C2-PLDs and the PX/PH-PLDs. The 12 Arabidopsis PLDs are classified into six types. C2, Ca^{2+}/phospholipid binding domain; PH, Pleckstrin homology domain; PX, phox homology domain; DRY, DRY motif; PIP, phosphatidylinositol-(4,5)-bisphosphate-binding domain. The duplicated HKD motifs are involved in catalysis.

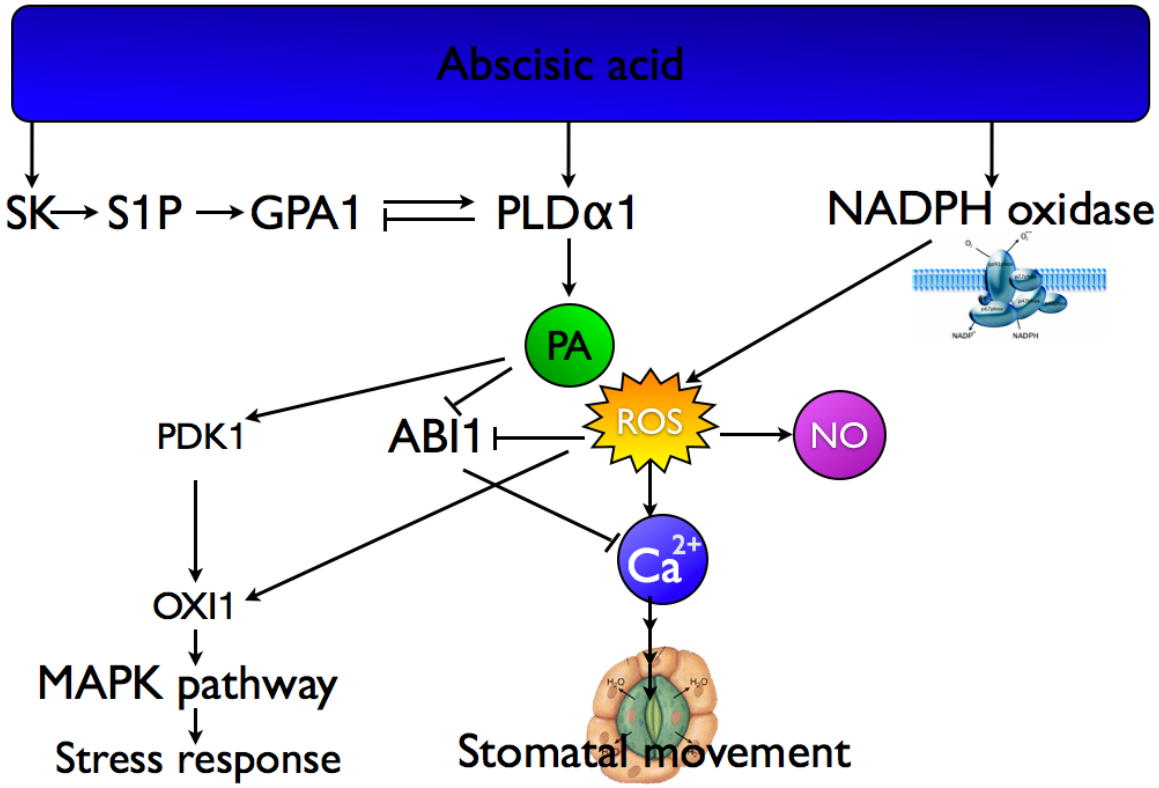


Figure 1.3: A model for PLD and PA in signaling ABA responses. PLDα 1-derived PA binds to ABI1, and tethers ABI1 to the plasma membrane and also decreases its PP2C activity. PLDα-1 binds to the Gα subunit of heterotrimeric G proteins. They reciprocally regulate each other, and thus control the production of PA. PA bind to various targets to control ABA responses. ABA, Abscisic acid; ABI1, ABA insensitive 1; GPA1, G protein α sub-unit 1; MAPK, mitogen-activated protein kinase; NADPH, nicotinamide adenine dinucleotide phosphate; NO, nitric oxide; OXI1, oxidative signal-inducible 1; PA, phosphatidic acid; PDK1, phosphoinositide-dependent protein kinase 1; PLD, Phospholipase D; ROS, reactive oxygen species; S1P, sphingosine-1 phosphate; SK, sphingosine kinase.

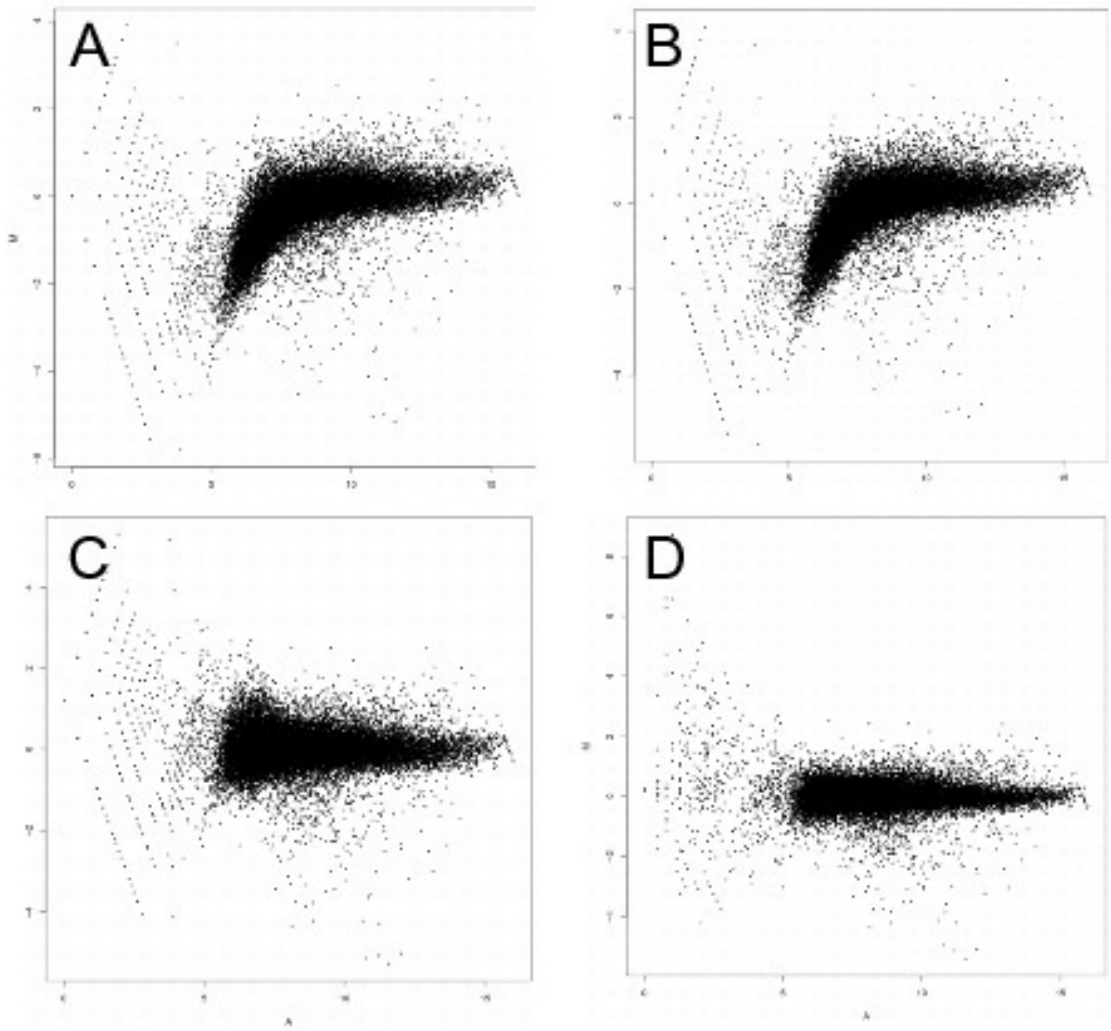


Figure 1.4: Result of normalization of microarray data by three different methods. A. The original raw data. B. Normalized data by median centering. C. Normalized data by global lowess. D. Normalized data by print-tip lowess.

Chapter 2

Responses of Wildtype and Antisense Plants to Drought Stress

2.1 Introduction

The results presented in this chapter have been published (Mane *et al.*, 2007). The copyright is held by Journal of Experimental Botany. The consent of co-authors (C. Vasquez-Robinet, A. Sioson, L. Heath and R. Grene) is appreciated.

Plants perceive and respond to drought stress by altering transcriptional, biochemical and physiological processes. The mechanism of molecular responses of plants to drought stress has been extensively studied (Ingram & Bartels, 1996; Rabbani *et al.*, 2003; Rizhsky *et al.*, 2004; Seki *et al.*, 2002), and phospholipid based signaling, involving phospholipase C (PLC) and phospholipase (PLD) have emerged as components of drought-responsive signal transduction pathways (Hirayama *et al.*, 1995; Sang *et al.*, 2001b).

Phospholipase D genes play diverse roles in plant cells in cell signaling and in metabolism by hydrolyzing phospholipids to phosphatidic acid (PA). PLDs have been shown to be involved in programmed cell death, cell growth, cell patterning, vesicular trafficking, cytoskeletal changes, root growth, abiotic stress responses and oxidative burst (Lee *et al.*, 2003; Li *et al.*, 2004; Ohashi *et al.*,

2003; Potocký *et al.*, 2003; Sang *et al.*, 2001b; Wang, 2005). The various PLDs have different requirements for Ca^{2+}, phosphoinositides and varied substrate preferences. Depression of PLDα 1 activity by antisense suppression or its ablation by knockout mutation resulted in alteration in several stress-related physiological processes such as reactive oxygen production (Sang *et al.*, 2001a), abscisic acid signaling (Fan *et al.*, 1997), wounding response (Wang *et al.*, 2000), freezing tolerance (Welti *et al.*, 2002) and increased water loss (Sang *et al.*, 2001b).

PLDα 1 directly interacts with Gα, the only canonical α-subunit of the heterotrimeric G protein in Arabidopsis (Zhao & Wang, 2003). PLDα 1-Gα interaction inhibits PLDα activity, stimulating intrinsic GTPase activity of Gα (Zhang *et al.*, 2004b), and regulates stomatal movement (Mishra *et al.*, 2006). ABA-independent activation of PLD by Gα subunit has also been reported in barley aleurone (Ritchie & Gilroy, 2000).

In mammalian cells, several PA targets have been identified (Ktistakis *et al.*, 2003). These include protein kinases, protein phosphastases, phosphodiesterases, ATPase, small GTPases, sphingosine kinase and protein kinase C (Delon *et al.*, 2004; Ghosh *et al.*, 1996; Jenkins *et al.*, 1994; Lopez-Andreo *et al.*, 2003; Manifava *et al.*, 2001). In plant systems, less is known. In Arabidopsis, PLDα 1-derived PA has been shown to interact with ABI1, a protein phosphatase 2 C (PP2C) which is a negative regulator of ABA responses (Zhang *et al.*, 2004b). PA binds to ABI1 and reduces its translocation to the nucleus by tethering it to the plasma membrane, thus promoting ABA responses. PLDα 1 and PA interact with PP2C to signal ABA promoted stomatal closure, whereas PLDα 1 and PA interact with the Gα- subunit to mediate ABA inhibition of stomatal opening (Mishra *et al.*, 2006). PA also binds to PDK1, a protein kinase which acts upstream of a MAP kinase pathway (Anthony *et al.*, 2004; Rentel *et al.*, 2004). PA has been shown to bind MAPK6-related protein, an important mediator in stress and ethylene signaling, in soybean cells (Liu & Zhang, 2004). Other PA binding targets have been identified in a proteome-wide study (Testernik *et al.*, 2004). However, an overall understanding of any one PLD-mediated pathway is still lacking for plants.

2.2 Materials and Methods

2.2.1 Plant Material

2.2.2 Progressive drought stress experiment

Arabidopsis thaliana seeds were sown on moist soil (Sunshine Mix 1), covered with plastic wrap and placed at 4 °C for 24 hours. The flats were then transferred to a growth chamber set on 12 hour, 100% fluorescent light at 22 °C, 12 hours, 0% light at 18 °C and a constant relative humidity of 75% (Conviron 4030, Winnipeg, MB. Canada). The flats were watered on alternate days. After four weeks, the plants were transplanted to trays with a spacing of two inches (Figure 2.1). At the time of transplantation, the growth chamber was changed to an 8 h light, 16 h dark photoperiod to avoid flowering. Drought stress was imposed by withholding water just prior to flowering, about 7 weeks after sowing prior to flowering.

Sampling began on the 8^{th} day after imposition of drought stress. Samples were taken five times a day to study the physiological state of plants in response to stress. Relative to dawn, the samples were taken at the following times: -0.5, 1, 2, 4, and 9 hours (Figure 2.2). Leaf water potential (LWP) was measured using a plant water status console (Soil Moisture Equipment Corp., Santa Barbara, CA). Two leaves from each plant were taken for water potential measurements. Photosynthesis measurements were taken using Li-6400 portable Photosynthesis System (Li-Cor Inc. Lincoln, NE) at 1, 2, 4 hours after dawn. All measurements were made in 2 biological replicates with 2-3 leaves perl plant unless otherwise stated. The aerial portion of the plant was extracted from the soil, wrapped in aluminum foil and flash frozen in liquid nitrogen and stored at 70 °C for further studies.

2.2.3 RNA isolation

Leaf tissue was ground to a fine powder under liquid nitrogen using a mortar and pestle. The fine ground powder was then transfered to a 50 ml conical tube containing 1 volume of grinding buffer (0.18M TRIS, 0.09M Licl, 4.5mM EDTA and 1% SDS; pH 8.2), 1 volume of phenol (pH 4.7) $1/10^{th}$ volume of 2M NaOAc pH 4.0 and 10 µl of mercaptoethanol/ml extraction buffer. The sample was throughly mixed on an orbital shaker (300 rpm) for 10 min. The sample was then

centrifuged for 10 min at (9000 x G). The supernatant was transfered new tube and 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) ph 6.6. The sample was then centrifuged for 10 min at (9000 x G). The supernatant was transfered to a new tube and an equal volume of chloroform was added. The sample was then centrifuged for 10 min at (9000 x G). The supernatant was transfered to a new tube and 1/2 volume 10 M LiCl was added. The sample was incubated at -20 °C overnight for RNA precipitation. The sample was then centrifuged for 25 min at (10,000 x G). The supernatant was discarded and the pellet was washed with 80 % ethanol, air dried and resuspended in nuclease free water.

2.2.4 DNase treatment

RNA was purified by DNase treatment using an Ambion DNA-free protocol (Ambion Inc., Austin, TX). To the RNA, 0.1 volume of 10X DNase I buffer and 1 µl DNase I (2 units) were added. The sample was gently mixed and incubated at 37 °C for 30 min. Resuspended DNase Inactivation Reagent (0.1 volume) was added to the sample and incubated for 2 min at room temperature. The sample was centrifuged at 10,000 x g for 2 min to pellet the DNase Inactivation Reagent.

2.2.5 Reverse Transcription

To a 0.6 ml eppendorf tube, 2 µg of RNA and 0.5 µg/µl of 18-mer Oligo-dT were added. Nuclease-free water was added to bring the volume to 20 µl. The sample was heated to 70 °C for 10 min and placed on ice. To the sample, 5 µl of 5x first strand buffer (Invitrogen, Carlsbad, CA), 1.25 µl of 0.1 M DTT, 1.25 µl of 10mM RNase-free dNTPs and 1.25 µl of SuperscriptII reverse transcriptase (Invitrogen, Carlsbad, CA) were added. The sample was incubated at 42 °C for 60 min and then heated to 75 °C for 10 min, and subsequently diluted with nuclease-free water (Sigma-Aldrich, St. Louis, Missouri) to 12.5 ng/µl cDNA.

2.2.6 Microarray experiment

Arrays

Microarray slides including 26,000 elements consisting of 70-mer (Qiagen/Operon, Valencia, CA) gene-specific oligonucleotides were used (see: `http://www.ag.arizona.edu/microarray/`). There are 48 blocks per microarray, 25 rows by 24 columns (600 spots) per block, and 28,800 spots per microarray, including spots for the 25,712 gene-specific 70-mers and 302 control elements. The remaining 2,786 spots are blank.

Hybridization

Total RNA, 100 µg per sample, was labeled using the indirect labeling procedure as described by Hegde *et al.* (2000). RNA was mixed with 0.8 mM of dATP, dCTP, and dGTP, 0.5 mM of dTTP and 0.3 mM of 5-(3-aminoallyl)-2deoxyuridine-5triphosphate (Sigma, St. Louis, MO), 2 µg oligo dT (Invitrogen, Carlsbad, CA) followed by first strand cDNA synthesis (Superscript II RT; Invitrogen, Carlsbad, CA). After incubation for 2h at 42 °C, the cDNA was treated with 2 units of RNaseH (Invitrogen, Carlsbad, CA) for 15 min at 37 °C and purified (Qiagen/Operon, Valencia, CA). After binding the cDNA, columns were washed with phosphate-ethanol buffer (5 mM potassium phosphate, pH 8.0 and % ethanol) and the cDNA was eluted with phosphate buffer (4 mM K-phosphate, pH 8.5).

Purified cDNA was dried, and subjected to a coupling reaction with Cy3-dUTP or Cy5-dUTP dye ester (Amersham-Pharmacia, St Louis, MO) and the labeled cDNA was purified (Qiagen). Glass slides were pre-treated and hybridized as described Kawasaki *et al.* (2001). Hybridizations were done using the two-dye (Cy3 and Cy5) method. In total, four hybridizations were performed with two hybridizations per treatment (dye-swaps with biological replicate).

Data Analyses

Microarray data analyses was carried as described in Sioson *et al.* (2006). The details of microarray analyses are discussed in section 3.2.

2.2.7 Quantitive-RT-PCR

Each SYBR Green I RT-PCR assay included: a standard curve of 7 serial dilution points of larger PCR fragments (440-850 bp) of genes, a no-template control, and 25 ng of each test cDNA. RT-PCR amplification mixtures (25 µl) contained 25 ng template cDNA, 2x SYBR Green I Master Mix buffer (12.5 µl) (Applied Biosystems, Foster City, CA) and 300 nM forward and reverse primer. While designing primers, care was taken to avoid amplification of the antisense region of PLDα1. The list of all primers and their sequences are provided in Table 2.13. The reactions were run on an ABI 7300 Real time PCR System (Applied Biosystems, Foster City, CA). The cycling conditions comprised 10 min denaturation at 95 °C and 35 cycles at 95 °C for 15 sec, 56 °C for 30 sec 72 °C for 30 sec. Chloroplastic GAPDH (AT3G26650) was used a control gene. Each assay was performed in triplicate. All PCR efficiencies were above 95 %. Sequence Detection Software (version 1.2.3) (Applied Biosystems, Foster City, CA) was used for data analysis.

2.3 Results

2.3.1 Drought stress experiment: Wildtype *vs* Antisense-GR2

In this experiment, I compared physiological drought responses of wildtype (Col-0) and Antisense-Glutathione Reductase 2 (Anti-GR2). Under control conditions, there was no observable phenotypic difference in between Col-0 and Anti-GR2 plants. Leaf water potential (LWP) measurements began at 4^{th} day after withholding water. Diurnal LWP measurements were similar in unstressed Col-0 and Anti-GR2 plants (Figure 2.3). Differences in LWP became apparent on day 8 after withholding water. Anti-GR2 experienced severe water stress at the same time period that Col-0 experienced less water stress. Difference in LWP began to appear from day 6 after stress imposition. The antisense plants were unable to recover fully in the dark. Net photosynthesis was also more affected in Anti-GR2 than in Col-0 (Figures 2.4 and 2.5). The Anti-GR2 plants showed severe symptoms of leaf yellowing. They experienced so much stress that they were not able to recover after rehydration at 9 days of stress (Figure 2.6).

2.3.2 Drought stress experiment: Wildtype *vs* Antisense-PLDα 1

In this experiment, physiological drought responses of wildtype (Col-0) and Antisense-PLDα 1 (Anti-PLDα) plants were compared. Under control conditions, there was no observable phenotypic difference between Col-0 and Anti-PLDα plants (Figure 2.8). Leaf water potential (LWP) measurements began at 8^{th} day after withholding water. LWP measurements were taken 30 min before dawn, 1, 2 and 4 hours during the day and 1 hour after dusk. Diurnal LWP measurements were similar in unstressed Col-0 and Anti-PLDα plants (Figure 2.7). Differences in LWP became apparent on day 8 after withholding water. Anti-PLDα experienced severe water stress at the same time period that Col-0 experienced less water stress (-0.31 MPa in Col-0 and -1.28 MPa in Anti-PLDα at peak drought stress) (Figure 2.7). Diurnal LWP measurements showed that Anti-PLDα had significantly lower LWP than Col-0 as the drought stress progressed. However, Anti-PLDα plants were able to recover in the dark. Net photosynthesis was also more affected in Anti-PLDα than in Col-0 (Figure 2.10). Both the untransformed and the Anti-PLDα plants recovered fully (0.11 MPa in Col-0 and 0.15 MPa in Anti-PLDα) upon rehydration at 10 days after the end of stress imposition (Figure 2.7 and 2.9).

2.3.3 Transcriptome Profiling

Plants from day 8, 4^{th} time-point of drought stress were chosen for gene transcription profiling in order to identify drought-mediated transcript changes associated with PLDα function at an early stage of drought stress when physiological differences between stressed Col-0 and Anti-PLDα plant began to appear. RNAs from two biological replicates were hybridized on two microarray slides with dye-swap. To determine drought-regulated genes, a two-step ANOVA was performed using Expresso Sioson *et al.* (2006). To determine the effect of antisense suppression of PLDα1 under control conditions, gene expression patterns of unstressed Col-0 and Anti-PLDα plants were compared. A total of only 39 (0.15 % of the transcriptome) genes were differentially expressed between unstressed Col-0 and Anti-PLDα (Table 2.1).

Genes that were identified as significantly up-/down-regulated under drought stress at $p \leq 0.05$ are shown in Figure 2.11 and appendix A.1. A total of 920 (4.4 % of the transcriptome) genes were

drought regulated in Col-0 with 460 up-regulated and 460 down-regulated genes. In Anti-PLDα, 765 (3.7 of the transcriptome) genes were drought regulated with 431 up-regulated and 334 down-regulated. Only 14.67 % and 16.25 % of the drought up- and down-regulated genes, respectively were common to both Col-0 and Anti-PLDα. Significantly regulated genes were classified according to the Munich Information Center for Protein Sequences (MIPS) categorization. In both Col-0 and Anti-PLDα, the most represented category was transcription followed by metabolism, signaling and subcellular localization (Table 2.2;Table 2.3). In most categories, barring transcription, a greater number of genes were downregulated in Col-0 than in Anti-PLDα.

Transcription

Forty-seven transcription-associated genes were upregulated and 35 were downregulated by drought in Col-0. In Anti-PLDα, 43 genes were upregulated and 24 were downregulated (Table 2.6). Among the drought regulated transcription factors, only ten genes were common to both genotypes. The transcription factors were classified by their domains into 14 groups. Zinc finger proteins were the most regulated followed by MYB and AP2/ERF (Apetela 2 /Ethylene-Responsive Element Binding Factor). One AP2 gene (AT4G27950) was upregulated in both Col-0 and Anti-PLDα while another gene, AT1G36060 was upregulated only in Col-0. Three genes encoding bZIP transcription factors were upregulated in Anti-PLDα, out of which one gene (AT1G53490) was downregulated in Col-0. A CCAAT-box-binding transcription factor and 2 homeodomain/ lipid-binding START domain-containing (HD-START) genes were upregulated in Col-0. Out of the 2 MADS box genes downregulated in Col-0, one gene (AT5G65330) was upregulated in Anti-PLDα. All significantly regulated genes belonging to MYB and NAC were differentially regulated in both Col-0 and Anti-PLDα. Not only were the zinc finger proteins the most regulated group of transcription factors but they were also the most differentially regulated between Col-0 and Anti-PLDα. ut of 16 upregulated genes in Col-0, only two genes were upregulated in Anti-PLDα and out of 13 downregulated genes, only one gene was downregulated in Anti-PLDα plants.

Metabolism

The second most drought-responsive group of genes was that associated with metabolism. Among metabolism genes, 28 genes were upregulated in Col-0 while 34 genes were upregulated in Anti-PLDα with nine genes common to both and 37 genes were downregulated in Col-0 while 19 genes were downregulated in Anti-PLDα with eight genes common to both (Table 2.10). More genes belonging to Carbon (C)-compounds and carbohydrate metabolism were downregulated in wild-type (17 genes) than in Anti-PLDα (7 genes) with three genes common to both. On the other hand, more genes associated with C-compounds and carbohydrate metabolism were upregulated in Anti-PLDα (21 genes) than in Col-0 (12 genes) with five genes common to both.

Signal transduction genes responsive to drought

Calcium plays an important role in early signal transduction events of many stresses. In Col-0, 2 14-3-3 proteins (AT1G34760, AT5G38480) and a calmodulin-binding protein were upregulated and 2 calmodulin binding proteins downregulated while in Anti-PLDα, 14-3-3 and 3 calmodulin proteins were upregulated and one 14-3-3 protein (AT5G10450) were downregulated. Only the 14-3-3 protein GF14 omicron (grf11) (AT1G34760) response was common to both genotypes (Table 2.4). Eight protein kinases including five putative protein kinases, a receptor-like kinase, a serine/threonine kinase and a LRR-transmembrane kinase were upregulated in Col-0 while two putative protein kinases, two receptor-like kinases, two serine/threonin kinases and a LRR-transmembrane kinase were upregulated in Anti-PLDα. Five phosphatases including two PP2Cs, one PP2A were upregulated and 3 phosphatases were downregulated in Col-0 while two PP2Cs (including ABI1) were upregulated in Anti-PLDα. Two response regulators (including a pseudoresponse regulator, APRR5) were upregulated in Col-0. Nine GTPases were downregulated in Col-0, but were either up-regulated (AT5G64990, AT2G44690) or unchanged in Anti-PLDα (AT1G43890, AT5G13650, AT3G05310, AT5G60860). ROP8 was downregulated in Col-0 but upregulated in Anti-PLDα.

Eleven and seven genes involved in phospholipid signaling were regulated by drought (Table 2.8) in Col-0 and Anti-PLDα, respectively. Phosphatidylinositol-4-phosphate 5-kinase (AT1G21920) and lipid transfer protein (LTP; AT2G10940) were upregulated and lipid-binding START domain containing protein was downregulated in both Col-0 and Anti-PLDα. Genes upregulated only in

Col-0 are inositol-1,4,5-trisphosphate 5-phosphatase, scramblase-like and phospholipase-like protein. PLDδ and a LTP were downregulated in Col-0 and upregulated in Anti-PLDα (Table 2.4).

Protein fate

Genes belonging to "protein fate" showed differential expression in Col-0 and Anti-PLDα. These included proteins involved in protein modification, degradation and folding. In Col-0, HSP60 (AT2G33210) was increased and three genes involved in protein degradation were downregulated (AT3G62250, AT5G53300 and AT5G42990). None of these were drought-regulated in the antisense genotype.

Hormone-related genes

Five genes related to ABA signaling were upregulated in Col-0 while two of them were downregulated and one unchanged in Anti-PLDα (Table 2.7). Three cytokinin genes (IPT2 and two UDP-glucosyl transferase genes) were upregulated in Col-0 but down regulated in Anti-PLDα. Several other UDP-glucosyl transferase genes were differentially regulated in Col-0 and Anti-PLDα. Ethylene responsive element binding factor family (EREBP) was down regulated in Anti-PLDα. Gibberellin response modulator (GAI) (RGA2) was downregulated in Col-0.

Quantitative RTPCR (qRTPCR) PLDα 1 transcript levels were 16-fold less in drought-stressed Anti-PLDα than in Col-0. To validate the microarray data, qRTPCR analyses were performed on six genes that responded on the microarrays. qRTPCR results confirmed that these genes are regulated by drought stress (Table 2.9). The qRTPCR data showed highly significant correlation ($\alpha = 0.01$) with the microarray results with a correlation coefficient of 0.96 and 0.95 in Col-0 and Anti-PLDα, respectively. qRTPCR was also performed on samples taken from other time points to determine the kinetics of PLD expression in both genotypes. At the initial time point, PLD was repressed, while at later time points PLD was upregulated in Col-0. PLD was upregulated at all time points in Anti-PLDα (Figure 2.12).

2.4 Discussion

2.4.1 Drought stress experiment: Wildtype *vs* Anti-GR2

Glutathione (GSH; *upgamma*-glutamylcysteinyl glycine) is an important multifunctional metabolite in most living organisms. It has unique structural properties, abundance and broad redox potential. One of the most well known functions of glutathione is as an antioxidant participating in the ascorbate-glutathione cycle (Foyer & Halliwell, 1976; Alscher, 1989). This cycle functions in the reduction of hydrogen peroxide (H_2O_2) to water by ascorbate, glutathione, and other enzymes. Ascorbate is oxidized during suppression of H_2O_2 and forms dehydroascorbate (DHA). Glutathione, in turn, provides the electrons needed to reduce DHA back to ascorbate through the action of dehydroascorbate reductase (DHAR). Glutathione reductase (GR), at the expense NADPH, converts oxidized glutathione (GSSG) back to its active reduced form (GSH) (Foyer & Halliwell, 1976). During abiotic stress, the levels and redox state of glutathione affects both expression and activity of detoxifying enzymes. The levels of glutathione in cells and organelles is tightly controlled by several facters including *de novo* synthesis, recycling by Glutathione reductase (GR) activity, conjugation reactions, transport from the cell, and degradation. The redox state of glutathione depends on the balance of oxidative processes and *in vivo* GR activity. Arabidopsis has two GRs, the organellar GR (GR1) and the cytosolic GR (GR2).

To better understand the role glutathione reductase in abiotic stress antisense GR2 plants were used in this study. The stressed antisense plants showed severe wilting when compared to Col-0. The diurnal measurements of LWP revealed that the stressed antiGR2 plants were not able to recover at night. Photosynthesis was affected more in antisense plants. The leave of AntiGR2 plants turned yellow after 8 days of stress. The plants failed to recover after rehydration. This was in contrast to Anti-PLDα which was able to recover after similar levels of stress suggesting that the loss of PLDα 1 did not affect the plants' ability to recover.

2.4.2 Drought stress experiment: Wildtype *vs* Anti-PLDα 1

Significant work has been done to understand the role of PLD in drought signaling, using short-term "shock" treatments. However, little or no work has been done to study its role in natural progressive drying. Our aim in this study was to identify PLDα1-regulated genes at early stages of progressive

drought. To study the role of PLDα1 in response to drought stress, Arabidopsis wildtype (Col-0) and Anti-PLDα were subjected to progressive drought by withholding water. LWP was measured five times a day to monitor diurnal changes in plant moisture status. LWP was reduced in both Col-0 and Anti-PLDα in response to drought. Anti-PLDα plants are less sensitive to ABA and have impaired stomatal closure during water deficits (Sang *et al.*, 2001b). Antisense suppression of PLDα1 did not show any observable phenotypic differences when compared to wildtype under control conditions. During drought, Anti-PLDα plants were more wilted than Col-0 indicating a greater transpirational loss of water in Anti-PLDα (Figure 1). However, Anti-PLDα plants were able to recover to Col-0 moisture status during the night suggesting that depletion of PLDα1 does not affect nocturnal stomatal closure. Moreover, reduction in PLDα1 did not affect the plants ability to recover fully after rehydration. Photosynthesis was also more affected in the antisense plants as a result of reduced plant moisture content, and it also recovered fully upon rehydration.

In order to understand more about the role of PLDα1 in drought signaling, I chose the 8^{th} day-4^{th} time-point (initial stage of drought) for gene expression analysis in response to progressive drought. This was when Anti-PLDα plants just began to experience higher water loss than Col-0. Our goal was to capture early PLDα-mediated signaling events in response to progressive drought stress. The effect of antisense suppression of PLDα1 on global gene expression was minimal under unstressed conditions. About 0.15% of the transcriptome was differentially regulated between Col-0 and Anti-PLDα. Lipid profiling studies have shown that PLDα1 deletion does not significantly alter basal PA levels in leaves (Devaiah *et al.*, 2006). Monitoring changes in drought-induced gene expression revealed significant differences between drought-stressed Col-0 and Anti-PLDα. The expression of very few (15 %) drought-regulated genes were common to both Col-0 and Anti-PLDα (Figure 2.11; ST 1). This result suggests that changes in plant PA levels via the PLDα pathway significantly alter global gene expression under progressive drought stress.

In wildtype plants, stomata are closed for increasingly longer periods during the day under progressive drought, starting mid-morning (Tenhunen *et al.*, 1987). Due to stomatal closure in response to drought there is reduction in water loss as well as gas exchange leading to reduced carbon assimilation and leading to a near optimization of carbon assimilation in relation to water supply (Cowan, 1981; Jones, 1992). This partly explains why Col-0 has more downregulation of genes involved in

metabolism (C-compound and carbohydrate metabolism) than Anti-PLDα (12 genes up, 17 down and 21 up and seven down in Col-0 and Anti-PLDα, respectively).

Responses of transcription factors to drought stress

The relatively large numbers of up-/down-regulated genes (9 % of the drought regulated genes) involved in transcription suggest that the plants are undergoing significant changes in transcription in response to water loss. The number of upregulated genes encoding transcription factors was higher than downregulated genes suggesting that drought stress leads to mainly activation of transcription factors. (Table 2.11). In general, there was little overlap of responsive genes (7 %) between Col-0 and Anti-PLDα in this category. AP2 transcription factors play important role in cold and drought stress responses. RAP2 (AT1G36060), a part of the CBF regulon and possibly controlling a sub-regulon (Zhang *et al.*, 2004c; Fowler & Thomashow, 2002), was upregulated in Col-0, but not in antisense plants.

Evidence for PLDα-dependent and independent drought-responsive pathways

Several transcription factors involved in plant development were differentially regulated in Col-0 and Anti-PLDα by drought. These include Homeodomain (HD-START), NAC, and MADS. NAC transcription factors have been reported to be involved in abiotic stress signaling (Fujita *et al.*, 2004; Tran *et al.*, 2004). In our experiment, these two groups were downregulated in both Col-0 and Anti-PLDα, suggesting the existence of a drought-responsive pathway that is independent of PLDα. MYB and basic helix-loop-helix (bHLH) transcription factors showed diverse responses. HD-START domain proteins were upregulated in Col-0. These proteins bind to phosphatidylcholine via START domain and modulate the activity and transcription of gene via homeodomain (Schrick *et al.*, 2004). By analogy with animal systems, this provides a mechanism whereby lipid content in the cytosol regulates transport/sequestration of the transcription factor to the nucleus. These data suggest that the HD-START mechanism is on the same pathway as PLDα.

Zinc finger domain-containing transcription factors were by far the largest group of transcription factors regulated by drought (29 in Col-0 and 19 in Anti-PLDα). C2H2 type zinc fingers have

been reported to be involved in stress regulation (Sakamoto *et al.*, 2004). Several C3HC4-type and CHP-type zinc finger proteins (four and three genes, respectively) were upregulated in Anti-PLDα but down-regulated or unchanged in Col-0 (ST-2). Three CCCH-type zinc fingers were upregulated in Col-0, out of which only one was upregulated in Anti-PLDα.

Drought signal transduction

PLDα1 may be involved in upstream activation of 14-3-3 and ankyrin-repeat containing protein genes: Calcium is a critical component in stress signal transduction pathways. Consistent with the central role of calcium, more calcium binding proteins (three Calmodulin-binding proteins and two 14-3-3 proteins) were upregulated in Anti-PLDα (Table 2.4 and 2.12). One of them, rare cold inducible (RCI1: AT5G38480) was upregulated only in Col-0. Interestingly, 14-3-3 protein GF14 λ (grf6/AFT1) was downregulated in Anti-PLDα. Overexpression of the Arabidopsis 14-3-3 protein GF14 λ in cotton leads to a "Stay-Green" phenotype and improves tolerance to moderate drought stress (Yan *et al.*, 2002). 14-3-3 protein GF14 λ interacts with ankyrin repeat-containing protein 2 (AKR2). AKR2 is involved in both disease resistance and antioxidation metabolism (Yan *et al.*, 2002). Four ankyrin-repeat proteins were upregulated in Col-0 out of which one gene (AT3G24210) was upregulated and two (AT4G19660, AT1G62050) genes were downregulated in Anti-PLDα. These results suggest that PLD α1 may be involved in upstream activation of 14-3-3 and ankyrin-repeat containing proteins.

A possible interaction between cytokinin and drought/cold signaling

Phosphorylation and dephosphorylation play an important role in stress signal transduction. The number of protein kinases and phosphatases was the highest (56) out of a total of 99 drought signaling genes (Table 2.4). Fifty percent of the downregulated genes were common to Col-0 and Anti-PLDα. These include LRR-transmembrane kinase, putative protein kinases, receptor and receptor-like protein kinases suggesting that these genes act upstream of the PLD signaling pathway, or are on another pathway altogether. ABI1 was downregulated in Anti-PLDα. ABI1, a PP2C, interacts with PLDα1-derived PA and negatively regulates ABA signaling (Zhang *et al.*, 2003). Two ARR (Arabidopsis Response Regulator) genes were upregulated by PLDα1. ARR5 has been implicated

in circadian rhythms (Fujimori *et al.*, 2005). ARR15 was downregulated in Anti-PLDα. ARR15, a type-A response regulator, acts as a negative regulator in cytokinin-mediated signal transduction in Arabidopsis and is also implicated in ethylene signaling (Kiba *et al.*, 2003). In their study, Lee *et al.* (2005) found that other response regulators (ARR10 and ARR16) are downregulated during cold stress. These genes are also involved in cytokinin-signaling. This suggests an interaction between cytokinin and drought/cold signaling.

Involvement of other phospholipases in drought signaling

Phosphatidylinositol-4-phosphate 5-kinase (PI5K; AT1G21920) was upregulated in both Col-0 and Anti-PLDα. Increased phosphatidylinositol 4,5-bisphosphate (PIP2) production is associated with upregulation of PI5K during osmotic stress (Mikami *et al.*, 1998). Two β-tubulin (AT5G62700, AT2G29550) and three calmodulin genes were upregulated in Anti-PLDα. In animals, β-tubulin binds to PIP2 and enhances PLC activity (Chang *et al.*, 2005). PIP2 also plays a critical role in PLD activation in mammals and plants. PIP2 is required for activities of Arabidopsis PLDβ and PLDγ (Qin *et al.*, 1997). This suggests that plants channel drought signaling through other members of PLD or PLC pathway (possibly signal amplification via IP3-mediated Ca^{2+} release) and calmodulin signaling when PLDα availability is greatly reduced. However, no changes in gene expression of PLC family members were observed suggesting the possibility of increased post-translational activation of PLC and PLD pathway.

PLDδ may act as "damage control" and possibly function in late drought stress signaling

PLDδ, was highly upregulated in Anti-PLDα when compared to Col-0. In earlier studies, Anti-PLDα plants did not show any change in activity or expression of PLDβ, PLDγ (Pappan *et al.*, 1997; Wang *et al.*, 2000) or PLDδ (Sang *et al.*, 2001a) under control conditions. Drought conditions were not used in these experiments. PLDδ is known to be activated by hydrogen peroxide (H_2O_2), and the resulting PA functions to decrease H_2O_2-promoted programmed cell death (Zhang *et al.*, 2003). However, a recent study showed that abrogation of PLDδ does not affect tolerance to drought stress (Li *et al.*, 2005). Studies have revealed that PLDα 1 and PLDδ have different, and sometimes opposite, roles in signaling (Welti *et al.*, 2002; Li *et al.*, 2004). Although PLDα1 and PLDδ

produce PA, they may prefer phospholipid substrates with different fatty acid side-chains. This, in turn, would change the PA species pool, thereby activating distinct intracellular signals. Our results suggest PLDδ acts as "damage control" and has a role in late drought stress signaling.

PLDα negatively regulates small GTPases

Nine GTPases were downregulated in Col-0 but were either upregulated or unchanged in Anti-PLDα (Table 2.5). A Rho-like GTPase (ROP8) was downregulated in Col-0 but upregulated in Anti-PLDα. Nothing is known about the function of ROP8 but other members of ROP (ROP6, ROP10) gene family are involved in negative regulation of ABA responses. Our results are consistent with the hypothesis that PLDα1 may negatively regulate small GTPases.

Hormone Signaling

Much information has accumulated on the role of ABA in stomatal movement and its importance in dehydrating roots and ABA movement in the plants. There is a very limited knowledge about the exact relationship between water deficit and ABA long-distance signaling in field drought conditions. Under rapid dehydration, ABA is synthesized mainly in roots and transported to the shoots. Contrary to most dehydration studies, very few genes were involved in ABA signaling under the conditions of progressive drought stress. There are two possible reasons why there was no induction of ABA biosynthesis genes. First, I did not include root tissue in our microarray experiment. Second, in leaves, ABA biosynthesis is increased only when leaf water potential falls to the point where turgor approaches zero. In this study, I collected leaf samples when the leaf water potential just began to drop. Five ABA responsive genes were up regulated in Col-0 while only one (ABI1) gene was upregulated in Anti-PLDα suggesting a role for PLDα1 in ABA signaling.

Table 2.1: Significantly regulated genes in Anti-PLDα control vs Col-0 control. Changes in transcript profile are shown as positive (+) or negative (-)

Gene	PCvsWC	Annotation
AT1G04940	+	tic20 family protein similar to Tic20
AT1G05010	+	1-aminocyclopropane-1-carboxylate oxidase
AT1G07020	+	Expressed protein
AT1G11060	+	expressed protein
AT1G19100	+	ATP-binding region, ATPase-like domain
AT1G53250	+	expressed protein
AT1G66070	+	translation initiation factor-related
AT1G69490	+	No apical meristem (NAM) protein family
AT1G71030	+	myb family transcription factor
AT1G72900	+	disease resistance protein (TIR-NBS class), putative
AT2G36240	+	putative salt-inducible protein
AT2G37620	+	actin 3
AT2G39280	+	RabGAP/TBC domain-containing protein
AT2G41740	+	putative villin 2
AT3G04840	+	40S ribosomal protein S3A (S phase specific)
AT3G43590	+	zinc knuckle (CCHC-type) family protein
AT3G47340	+	glutamine-dependent asparagine synthetase
AT4G16660	+	HSP like protein
AT5G14800	+	pyrroline-5-carboxylate reductase
AT5G35200	+	epsin N-terminal homology (ENTH) domain
AT5G59890	+	actin depolymerizing factor 4 - like protein
AT1G12160	-	flavin-containing monooxygenase (FMO) family
AT1G27690	-	putative lipase
AT1G28760	-	hypothetical protein
AT1G29200	-	hypothetical protein

Continued on next page

Gene	PCvsWC	Annotation
AT1G37035	-	pseudogene, Mutator-like transposase
AT1G50030	-	Target of Rapamycin (TOR) protein (pTOR)
AT1G65470	-	chromatin assembly factor-1 (FASCIATA1) (FAS1)
AT1G69020	-	putative protease
AT1G76310	-	Cyclin B2;4
AT2G36470	-	expressed protein
AT2G37020	-	translin-like protein
AT3G21965	-	pseudogene At3g21965
AT3G22104	-	phototropic-responsive NPH3 protein-
AT4G15340	-	pentacyclic triterpene synthase (04C11) (ATPEN1)
AT4G27140	-	NWMU1 - 2S albumin 1 precursor
AT5G51930	-	mandelonitrile lyase-like protein
AT5G54820	-	F-box protein family
AT5G58190	-	expressed protein

Figure 2.1: Drought stress experimental setup in growth chamber. The plants were grown under short photoperiod for six weeks before stress was imposed

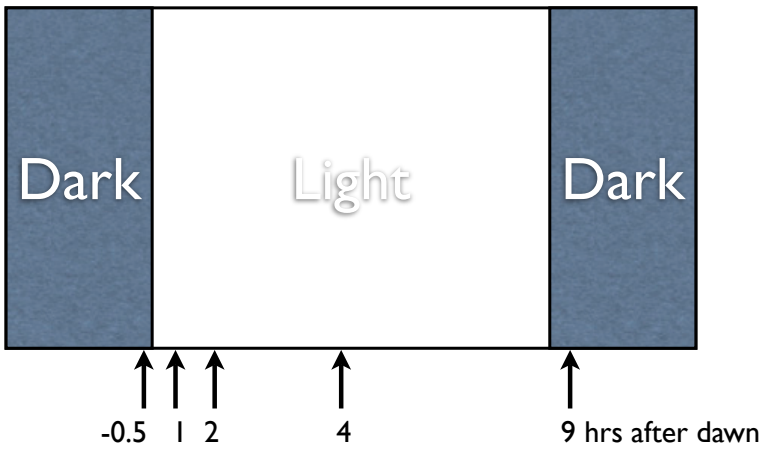


Figure 2.2: Time-points for leaf water potential (LWP) and photosynthesis measurements. LWP measurements were taken five times (-0.5, 1, 2, 4 and 9 hours relative to dawn) and photosynthesis measurements were taken 3 times (1, 2 and 4 hours relative to dawn).

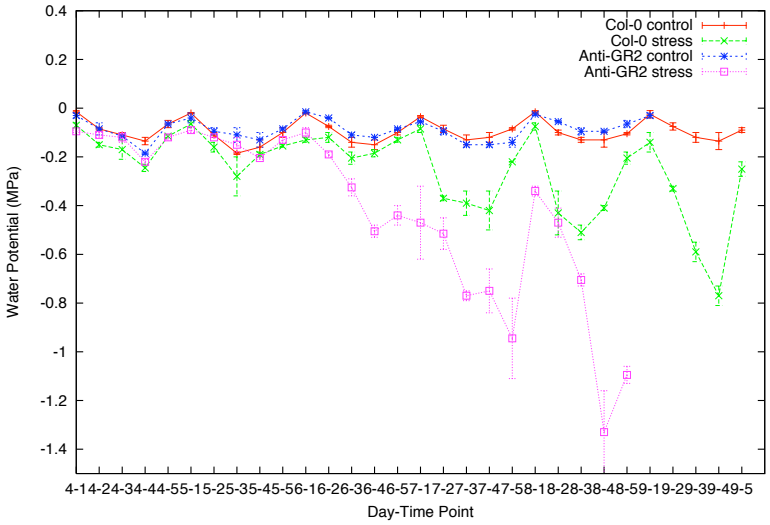


Figure 2.3: Leaf water potential in Col-0 and Anti-GR2. Leaf water potential was measured at five time points using a pressure bomb and is expressed in Mega Pascals. Three leaves per plant were used for measuring LWP. The Anti-GR2 plants turned yellow and died after 8 days of drought stress.

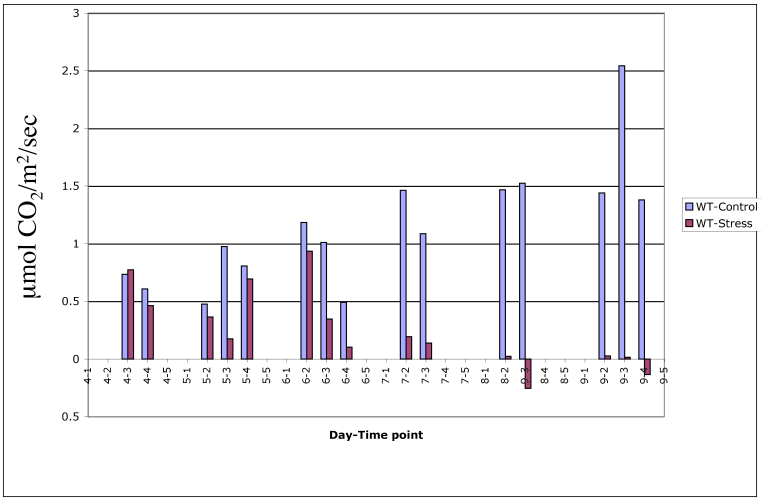


Figure 2.4: Net photosynthesis in Col-0. Col-0 plants showed negative photosynthesis after 8 days of stress imposition.

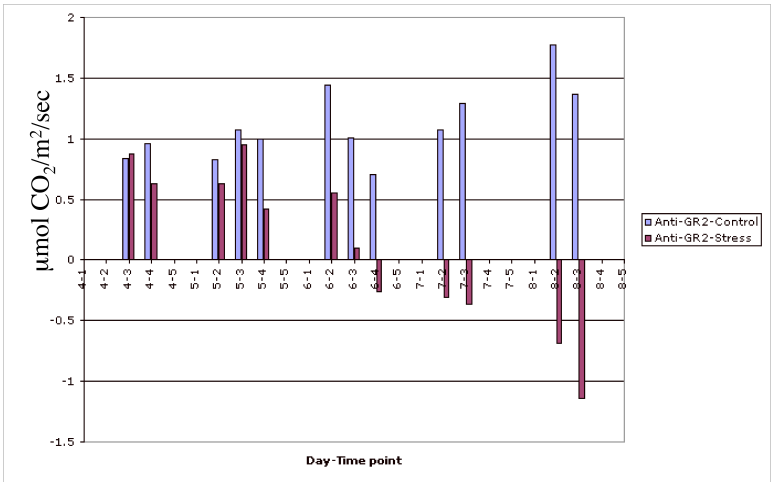


Figure 2.5: Net photosynthesis in Anti-GR2. Anti-GR2 plants showed negative photosynthesis after 6 days of stress imposition.

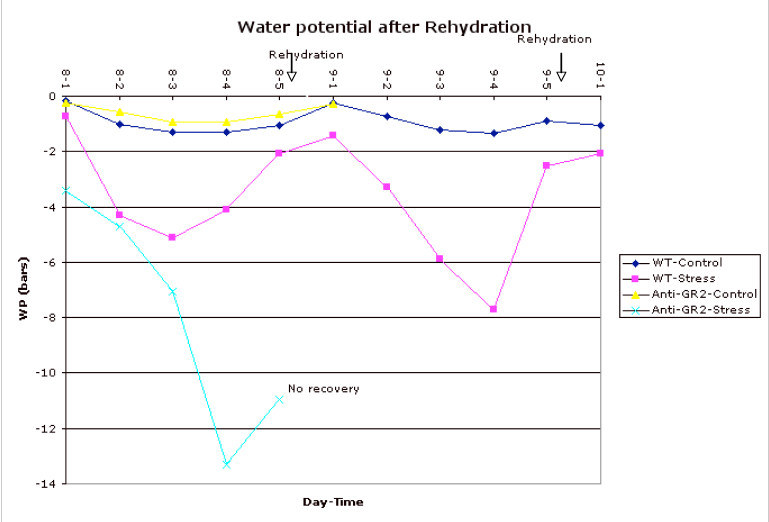


Figure 2.6: Leaf water potential (in MPa) in Col-0 and Anti-GR2 after rehydration. Plants were rehydrated at the end of 8^{th} day. LWP was measured after 16 hours of rehydration

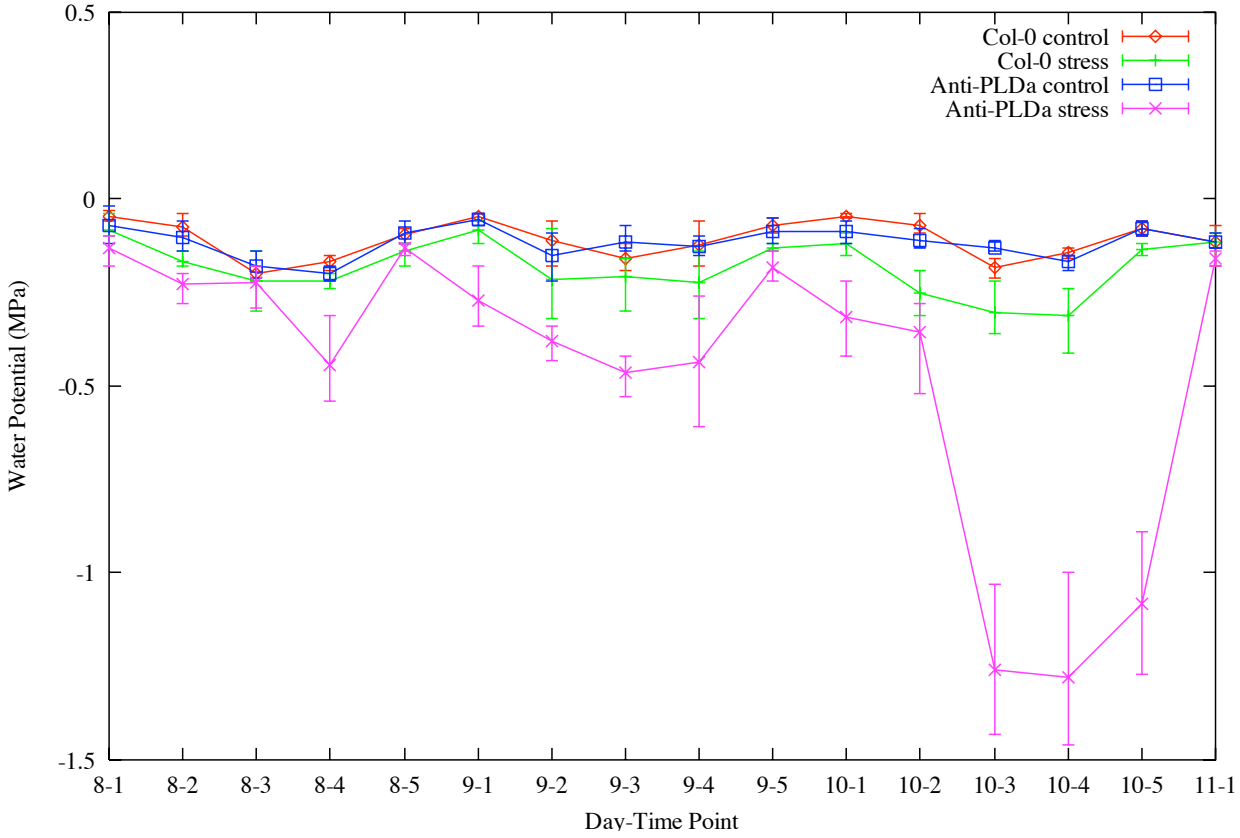


Figure 2.7: Leaf water potential (in MPa) in Col-0 and Anti-PLDα. The LWP of Anti-PLDα stressed plants were similar to stressed wild-type during dark.

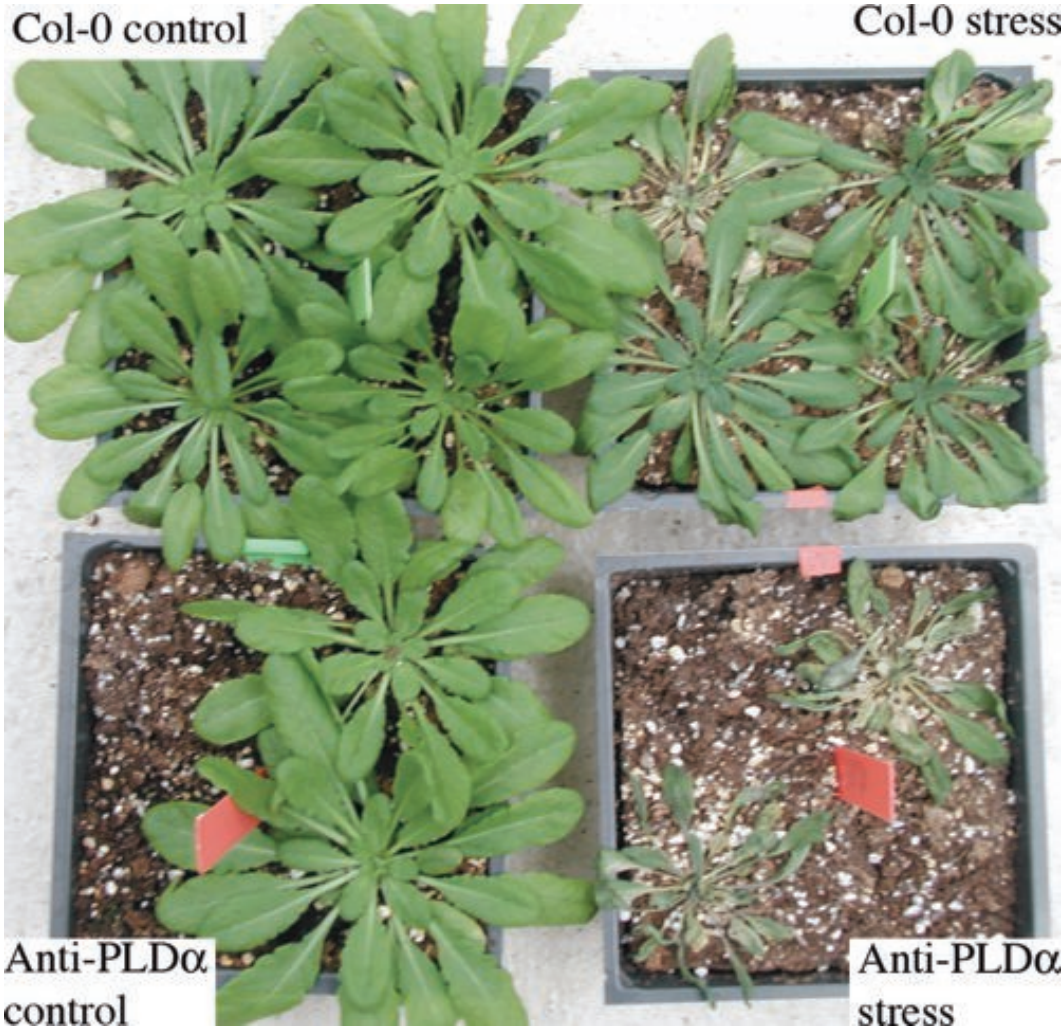


Figure 2.8: Col-0 and Anti-PLDα after 10 days of water stress. The photograph of representative samples were taken at 10-4 time point.

Figure 2.9: Col-0 and Anti-PLDα recovery at 16 hours after rehydration. The stressed Col-0 and Anti-PLDα 1 were rehydrated after the end of 10^th day. LWP and photosynthesis measurements were taken the following afternoon.

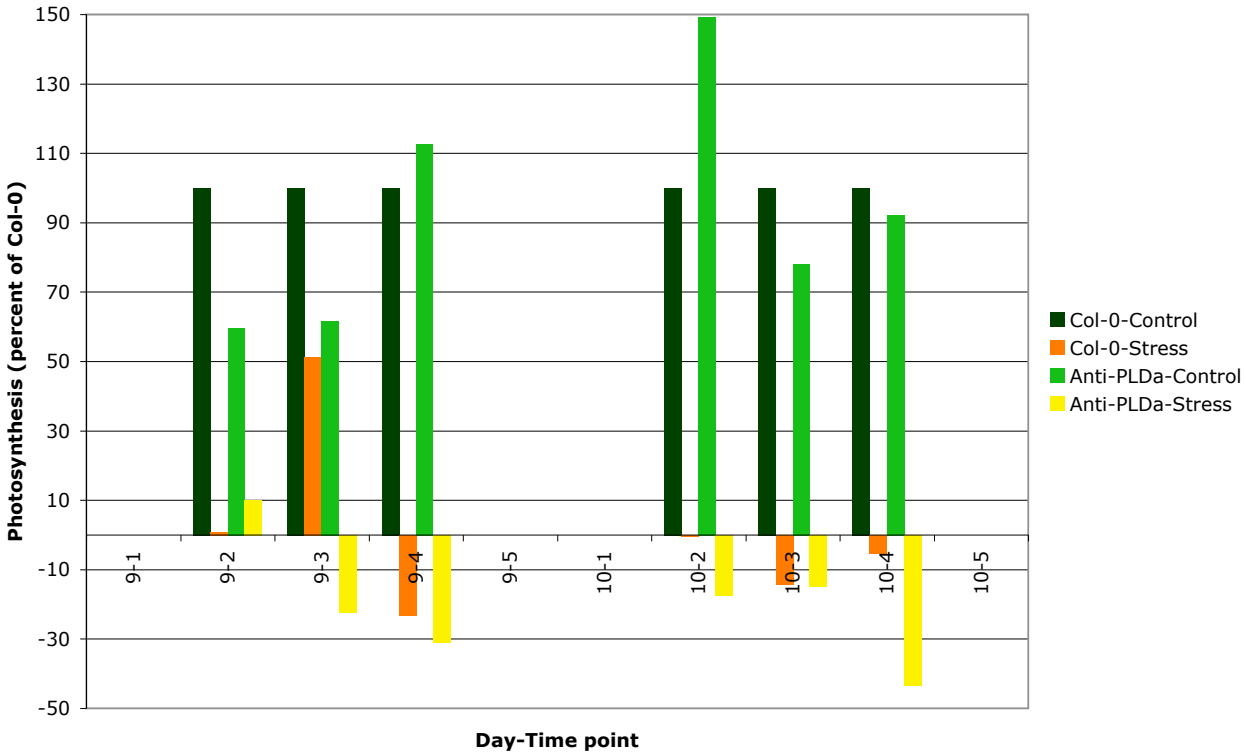


Figure 2.10: Photosynthesis in Col-0 and Anti-PLDα. The photosynthesis measurements re expressed as percent of Col-0 photosynthesis values.

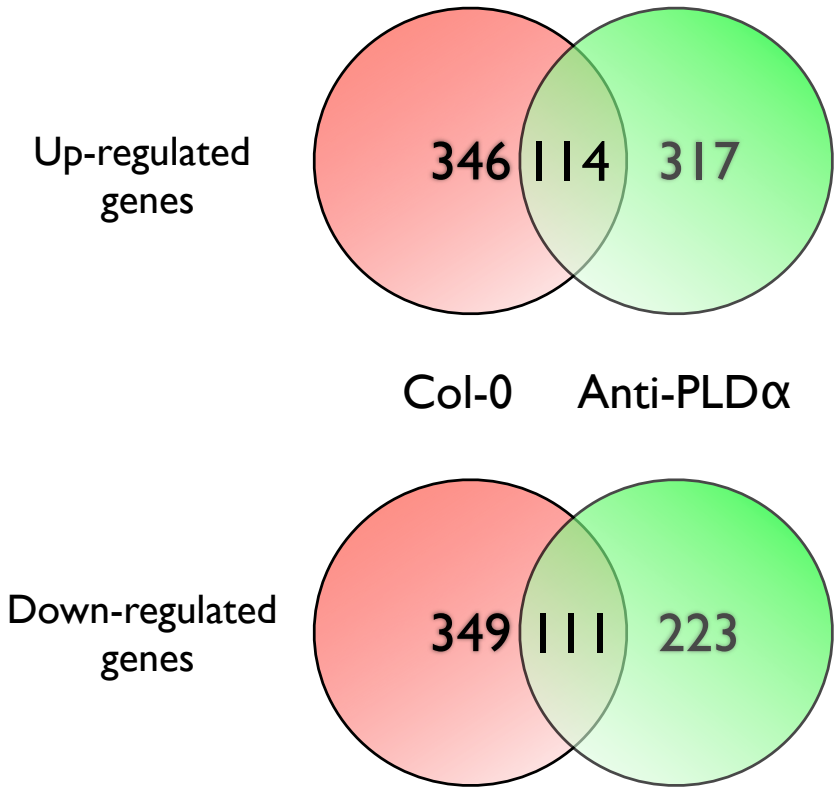


Figure 2.11: Venn diagram showing differential expression of genes in Col-0 and Anti-PLDα. Genes were assessed as up-/down-regulated at α=0.05. More genes were up-/down-regulated in Col-0 than in Anti-PLDα.

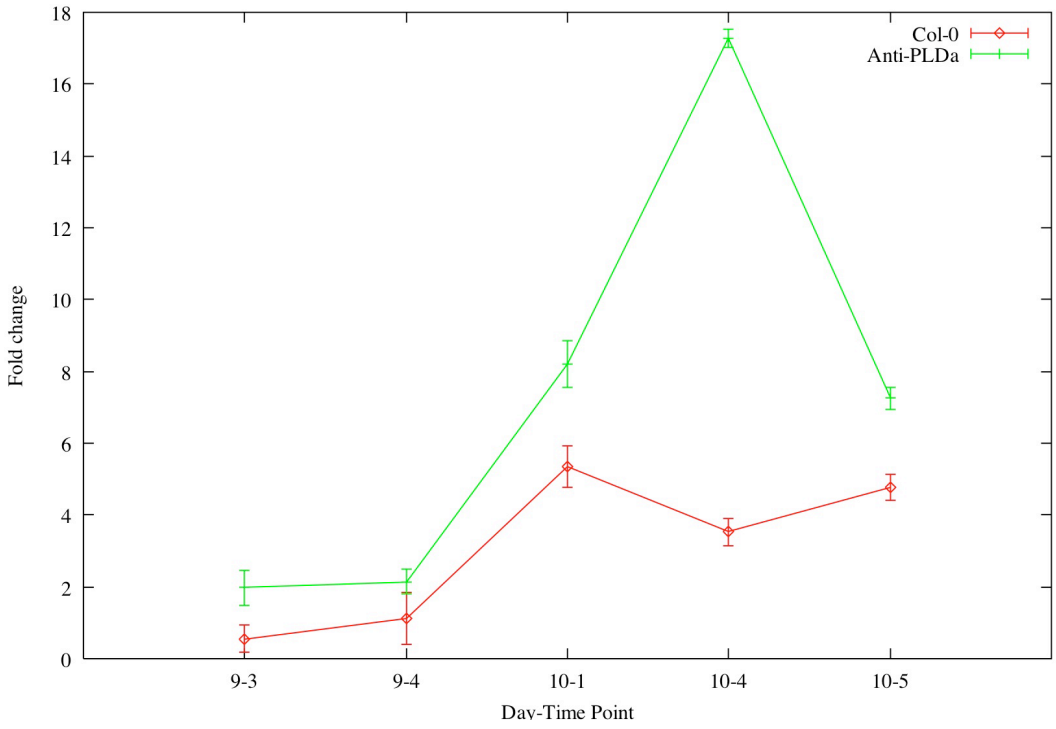


Figure 2.12: Expression of PLDδ in Col-0 and Anti-PLDα. GAPDH was used as control.

Table 2.2: Functional categorization of Drought Regulated Genes in Col-0 and Anti-PLDα. Numbers in parentheses indicate the number of responsive genes common to both Col-0 and Anti-PLDα.

Functional Category	Col-0		Anti-PLDα	
	Up	Down	Up	Down
Metabolism	28(9)	37(8)	34	19
Subcellular localization	14(5)	35(11)	19	20
Signal transduction	27(3)	34(13)	22	28
Cellular transport, transport facilitation and transport routes	10(3)	17(5)	10	15
Protein with binding function or cofactor requirement (structural or catalytic)	8(3)	21(10)	11	13
Transcription	47(8)	35(2)	43	24

Table 2.3: Detailed functional characterization of drought regulated genes in Col-0 and Anti-PLDα. The total number of genes is higher because many genes belong to multiple categories

Functional Category	Col-0		Anti-PLDα	
	Up	Down	Up	Down
Metabolism	28	37	34	19
amino acid metabolism	4	4	5	5
nitrogen and sulfur metabolism			1	
nucleotide metabolism	1	5	3	2
phosphate metabolism		1		
C-compound and carbohydrate metabolism	12	17	21	7
lipid, fatty acid and isoprenoid metabolism	3	3		2
metabolism of vitamins, cofactors, and prosthetic groups	8	1		
secondary metabolism		5	4	2
Energy	4	8	8	2
Storage protein	2			1
Cell cycle and DNA processing	6	6	7	
Transcription	47	35	43	24
Protein synthesis	2	19	2	10
Protein fate (folding, modification, destination)	12	20	9	7
Protein with binding function or cofactor requirement (structural or catalytic)	8	21	11	13
Protein activity regulation	1	5	1	3
Cellular transport, transport facilitation and transport routes	10	17	10	15
Cellular communication/signal transduction mechanism	13	19	12	19
Cell rescue, defense and virulence	11	10	11	10
Interaction with the cellular environment	6	5	3	3
Interaction with the environment (systemic)	2	1	2	
Transposable elements, viral and plasmid proteins	4	2	4	3

Continued on next page

Functional Category	Col-0		Anti-PLDα	
	Up	Down	Up	Down
Cell fate	4	10	5	4
Development (systemic)	4	7	7	3
Biogenesis of cellular components	8	17	15	7
Cell type differentiation	1	2	2	
Tissue differentiation	1	1	3	
Subcellular localization	14	35	19	20
cell wall	1		1	1
cytoplasm	4	12	4	7
cytoskeleton	2	7	4	1
endoplasmic reticulum	1	1		1
eukaryotic plasma membrane / membrane attached		2	1	1
Golgi	1	4		3
intracellular transport vesicles	1			
nucleus	3	5	7	5
mitochondrion	1	2	1	1
endosome	1		1	
plastid	1		1	
peroxisome		1		1
vacuole or lysosome		2		
extracellular / secretion proteins		1		
Cell type localization	1	1	1	
Tissue localization		1	1	
No clear classification/unclassified	355	325	334	238
Total	508	576	515	386

Table 2.4: Drought-regulated signaling genes. Numbers in parentheses indicate the number of responsive genes common to both Col-0 and Anti-PLDα.

Category	Col-0		Anti-PLDα	
	Up	Down	Up	Down
Calcium binding protein	3(1)	2	5	1
Protein phosphatase	5	3	2	2
Protein Kinase	10	17(9)	9	17
Response regulator	2			1
GTP-related protein		8(3)	1	4
Lipid Signaling protein	7(2)	4(1)	5	3

Table 2.5: Regulation of Small GTP-related proteins in Col-0 and Anti-PLDα. Changes in transcript profile are shown as positive (+), negative (-) or unchanged (0)

Gene	Col-0	Anti-PLDα	Annotation
AT1G43890	-	0	Ras-related GTP-binding protein (RAB1Y)
AT3G05310	-	0	GTP-binding protein - related
AT5G60860	-	0	Ras-related GTP-binding protein
AT5G64990	-	0	Ras-related GTP-binding protein
AT1G09180	-	-	putative GTP-binding protein, SAR1B
AT1G49300	-	-	Ras-related GTP-binding protein
AT3G54840	-	-	Rab GTPase (ARA6)
AT2G44690	-	+	Rac-like GTPase (ARAC9)
AT2G27440	0	+	rac GTPase activating protein

Table 2.6: Comparison of Drought Regulated Genes in Col-0 and Anti-PLDα. Numbers in parentheses indicate the number of responsive genes common to both Col-0 and Anti-PLDα.

Category	Col-0		Anti-PLDα	
	Up	Down	Up	Down
AP2/ERF	2(1)	1	1	1
B3				1
bZIP		1	3	
CCAAT	1			
Homeodomain	2			
HSF				1
MADS	1	2	1	
MYB	4	3	4	3
bHLH	2	1	2	1
NAC		1	1	
PHD finger	2	2		1
WRKY		2(1)		1
Zinc	16(2)	13(1)	13	6
NAM	2		2	1
other DNA binding	15(5)	9	16	8

Table 2.7: Drought Regulated Genes Involved in Hormone Signaling. Changes in transcript profile are shown as positive (+), negative (-) or unchanged (0).

Gene	WT	Anti-PLDα	Annotation
ABA			
AT4G08500	+	-	MEKK1/MAP kinase kinase kinase (MAPKKK8)
AT4G18350	+	-	9-cis-epoxycarotenoid dioxygenase (NCED1), putative
AT3G23050	+	0	AUXIN RESISTANT 2 (AXR2)
AT5G49160	-	-	DNA METHYLTRANSFERASE (DMT01)
AT4G26080	0	+	protein phosphatase ABI1
AT3G50980	+	0	dehydrin-like protein
AT3G51250	+	0	senescence/dehydration-associated protein similar to ERD7
Cytokinin			
AT2G27760	+	0	putative tRNA isopentenylpyrophosphate transferase
AT2G30150	+	0	putative glucosyltransferase
AT3G16520	+	0	putative glucosyltransferase
AT2G36760	-	+	putative glucosyl transferase
AT5G17040	-	+	UDP glucose:flavonoid 3-o-glucosyltransferase -like protein
AT1G05680	-	0	putative indole-3-acetate beta-glucosyltransferase
AT3G46720	-	0	glucuronosyl transferase-like protein
AT1G22340	0	+	UDP-glucose glucosyltransferase, putative
AT1G74890	0	-	Two component response regulator 15 (ARR15)
AT3G22250	0	-	glycosyltransferase family
Ethylene			
AT4G08500	+	-	MEKK1/MAP kinase kinase kinase
AT2G27070	+	0	Two component response regulator 13 (ARR13)
AT3G08580	-	0	adenine nucleotide translocator 1 (ANT1)
AT5G09410	0	+	calmodulin-binding protein
AT1G28160	0	-	ethylene responsive element binding factor family (EREBP)
AT1G74890	0	-	Two component response regulator 15 (ARR15)
GA			
AT1G14920	-	0	signal response protein (GAI)

Table 2.8: Drought-regulated lipid signaling proteins. Changes in transcript profile are shown as positive (+), negative (-) or unchanged (0).

Gene	WT	Anti-PLDα	Annotation
AT1G21920	+	+	phosphatidylinositol-4-phosphate 5-kinase, putative
AT2G10940	+	+	lipid transfer protein (LTP) family
AT5G19940	+	-	plastid-lipid associated protein PAP-related
AT2G04940	+	0	scramblase-related
AT2G04940	+	0	scramblase-related
AT3G63240	+	0	inositol-1,4,5-trisphosphate 5-Phosphatase-like protein
AT4G38550	+	0	Phospholipase like protein
AT1G55260	-	+	lipid transfer protein (LTP) family
AT4G35790	-	+	PLDδ
AT3G54800	-	-	pleckstrin homology (PH), lipid-binding START domain-containing protein
AT4G27350	-	0	membrane lipo protein
AT5G45560	0	+	PH, lipid-binding START domain-containing protein

Table 2.9: Quantitative RTPCR of selected genes. Changes in transcript profile are shown as positive (+), negative (-) or unchanged (0). †indicates fold change values.

Gene	Col-0		Anti-PLDα 1		Annotation
	qRT-PCR†	Microarray	qRT-PCR†	Microarray	
AT2G44690	0.4448	-	1.65187	+	ROP8/ARAC9
AT4G35790	1.1189	-	2.14744	+	Phospholipase Dd
AT5G13650	0.4862	-	1.31883	+	GTP-binding protein typA
AT1G62050	3.6429	+	0.19720	-	Ankyrin repeat protein
AT1G21920	1.4654	+	1.51108	+	phosphatidylinositol-4-phosphate 5-kinase
AT5G17220	0.1462	-	6.54938	+	glutathione S-transferase, putative

Table 2.10: Genes involved in metabolism. Changes in transcript profile are shown as positive (+), negative (-) or unchanged (0).

Gene	Col-0	Anti-PLDα	Annotation
AT5G58770	+	+	DEDOL-PP synthase, putative
AT5G23790	+	+	galactinol synthase
AT3G04200	+	+	germin-like protein
AT5G26120	+	+	glycosyl hydrolase family 51
AT1G35190	+	+	hyoscyamine 6-dioxygenase hydroxylase, putative
AT5G26140	+	+	lysine decarboxylase family protein
AT3G42950	+	+	polygalacturonase, putative
AT2G21590	+	+	ADP-glucose pyrophosphorylase large subunit
AT3G53900	+	+	uracil phosphoribosyltransferase-like protein
AT3G61300	+	-	anthranilate phosphoribosyltransferase-like
AT4G37330	+	-	cytochrome p450 family
AT5G47770	+	-	farnesyl pyrophosphate synthetase 1 (FPS1)
AT4G27040	+	-	SNF8 like protein
AT4G27070	+	-	tryptophan synthase beta-subunit (TSB2)
AT4G23920	+	-	UDPglucose 4-epimerase - like protein
AT4G15290	+	0	cellulose synthase like protein
AT3G28740	+	0	cytochrome p450 family
AT4G35420	+	0	dihydroflavonol 4-reductase family
AT4G32540	+	0	dimethylaniline monooxygenase - like protein
AT1G33220	+	0	glycosyl hydrolase family 17 (beta-1,3-glucanase)
AT5G01190	+	0	laccase (diphenol oxidase), putative
AT3G46480	+	0	oxidoreductase, 2OG-Fe(II) oxygenase family
AT4G38550	+	0	Phospholipase like protein
AT3G51480	+	0	putative glutamate receptor
AT3G09300	+	0	putative oxysterol-binding protein
AT1G01090	+	0	pyruvate dehydrogenase E1 alpha subunit

Continued on next page

Gene	Col-0	Anti-PLDα	Annotation
AT3G55650	+	0	pyruvate kinase - like protein
AT5G17310	+	0	UDP-glucose pyrophosphorylase
AT4G33510	-	+	2-dehydro-3-deoxyphosphoheptonate aldolase
AT4G18440	-	+	adenylosuccinate lyase - like protein
AT4G02120	-	+	CTP synthase - like protein
AT1G60470	-	+	galactinol synthase, putative
AT3G28180	-	+	glycosyltransferase family 2
AT5G13520	-	+	leukotriene-A4 hydrolase-like protein
AT4G37160	-	+	pectinesterase (pectin methylesterase) family
AT2G36760	-	+	putative glucosyl transferase
AT4G22880	-	+	putative leucoanthocyanidin dioxygenase
AT3G03080	-	+	putative NADP-dependent oxidoreductase
AT3G04050	-	+	putative pyruvate kinase
AT5G65205	-	+	short-chain dehydrogenase/reductase family
AT5G17040	-	+	UDP glucose:flavonoid 3-o-glucosyltransferase
AT3G02360	-	-	6-phosphogluconate dehydrogenase, putative
AT3G51840	-	-	acyl-coA dehydrogenase
AT3G26180	-	-	cytochrome p450 family
AT5G49160	-	-	DNA (cytosine-5)-methyltransferase
AT4G22100	-	-	glycosyl hydrolase family 1
AT4G11010	-	-	nucleoside diphosphate kinase 3 (ndpk3)
AT3G49160	-	-	pyruvate kinase -like protein
AT3G45080	-	-	sulfotransferase-like protein
AT5G47760	-	0	4-nitrophenylphosphatase-like protein
AT5G17920	-	0	5-methyltetrahydropteroyltriglutamate-
			-homocysteine S-methyltransferase
AT3G08580	-	0	adenylate translocator
AT5G07720	-	0	alpha galactosyltransferase protein
AT5G65010	-	0	asparagine synthetase (gb—AAC72837.1)

Continued on next page

Gene	Col-0	Anti-PLDα	Annotation
AT4G12270	-	0	copper amine oxidase like protein (fragment1)
AT3G46720	-	0	glucuronosyl transferase-like protein
AT1G13440	-	0	glyceraldehyde-3-phosphate dehydrogenase
AT2G13560	-	0	malate oxidoreductase (malic enzyme), putative
AT3G59970	-	0	methylenetetrahydrofolate reductase MTHFR1
AT4G13050	-	0	oleoyl-(acyl-carrier-protein) hydrolase-like
AT1G56190	-	0	phosphoglycerate kinase, putative
AT3G12780	-	0	phosphoglycerate kinase, putative
AT3G61490	-	0	polygalacturonase, putative
AT2G38040	-	0	putative alpha-carboxyltransferase
AT3G16050	-	0	putative ethylene-inducible protein
AT4G23990	0	+	cellulose synthase catalytic subunit-like
AT5G56720	0	+	cytosolic malate dehydrogenase
AT3G60120	0	+	glycosyl hydrolase family 1
AT1G68560	0	+	glycosyl hydrolase family 31 (alpha-xylosidase)
AT3G45060	0	+	high-affinity nitrate transporter - like
AT3G13930	0	+	putative acetyltransferase
AT3G02020	0	+	putative aspartate kinase
AT2G21330	0	+	putative fructose bisphosphate aldolase
AT2G22780	0	+	glyoxysomal malate dehydrogenase precursor
AT3G22740	0	+	putative selenocysteine methyltransferase
AT4G03210	0	+	xyloglucan endotransglycosylase, putative
AT5G48070	0	+	xyloglucan endotransglycosylase, putative
AT3G47340	0	-	glutamine-dependent asparagine synthetase
AT4G31590	0	-	glycosyltransferase family 2
AT5G04360	0	-	pullulanase-like (starch debranching enzyme)
AT5G14800	0	-	pyrroline-5-carboxylate reductase
AT1G79440	0	-	succinic semialdehyde dehydrogenase, putative

Table 2.11: Differentially regulated transcription Factors. Changes in transcript profile are shown as positive (+) or negative (-)

Gene	Col-0	Anti-PLDα	Annotation
AT4G27950	+	+	AP2 domain transcription factor
AT1G36060	+	0	AP2 domain RAP2, A6 sub-family
AT3G25890	-	0	AP2 domain , B6 sub-family
AT5G67000	0	-	AP2 domain , B6 sub-family
AT5G18090	0	-	transcriptional factor B3 family
AT1G53490	-	+	bZIP protein
AT5G10030	0	+	bZIP transcription factor, OBF4
AT5G42910	0	+	bZIP protein
AT1G72440	+	0	CCAAT-box-binding transcription factor
AT1G05230	+	0	homeodomain protein/lipid-binding START domain-containing
AT5G17320	+	0	homeodomain protein/lipid-binding START domain-containing
AT5G16820	0	-	Heat Shock Factor 3 (HSF3)
AT5G26625	+	0	MADS-box protein
AT5G65330	-	+	MADS-box protein
AT2G22540	-	0	MADS-box protein
AT1G42670	+	-	myb family protein
AT3G01530	+	-	myb family transcription factor
AT1G18330	+	-	myb family transcription factor
AT5G01200	+	0	myb family transcription factor
AT4G26930	-	+	myb family transcription factor (MYB97)
AT5G03780	-	0	myb -like protein
AT4G17780	-	0	myb family transcription factor (MYB39)
AT4G37780	0	+	myb DNA-binding protein (AtMYB87)
AT4G04580	0	+	myb family transcription factor
AT3G52250	0	+	myb family transcription factor
AT3G56980	+	-	bHLH protein

Continued on next page

Gene	Col-0	Anti-PLDα	Annotation
AT5G62610	+	0	bHLH protein
AT3G05800	-	+	bHLH protein
AT4G02590	0	+	bHLH protein
AT5G62380	-	+	Vascular related NAC-Domain 6 (VND6)
AT3G61740	+	-	PHD finger family protein (ATX3)
AT3G14740	+	0	PHD finger family protein
AT3G05670	-	0	PHD finger family protein
AT2G01810	-	0	PHD finger family protein
AT2G21900	-	-	WRKY family transcription factor
AT1G55600	0	-	WRKY family transcription factor
AT5G56930	+	+	zinc finger (CCCH-type) family
AT2G43220	+	-	CHP-rich zinc finger protein, putative
AT2G26000	+	-	C3HC4-type RING finger
AT5G42200	+	0	C3HC4-type zinc finger protein family
AT2G28920	+	0	C3HC4-type zinc finger protein family
AT1G61840	+	0	CHP-rich zinc finger protein, putative
AT4G38180	+	0	far-red impaired responsive/SWIM zinc finger
AT2G26940	+	0	putative C2H2-type zinc finger protein
AT1G74410	+	0	putative RING zinc finger protein
AT5G07640	+	0	zinc finger (C3HC4-type RING finger) family
AT1G03790	+	0	zinc finger (CCCH-type) family protein
AT3G55980	+	0	zinc finger (CCCH-type) family protein
AT4G31670	+	0	zinc finger (MYND type) family protein
AT1G29570	+	0	zinc finger protein, putative
AT3G43590	+	0	zinc knuckle (CCHC-type) family protein
AT5G40590	-	+	CHP-rich zinc finger protein, putative
AT4G09690	-	+	CHP-rich zinc finger protein, putative
AT1G44030	-	+	CHP-rich zinc finger protein, putative
AT3G54810	-	+	GATA zinc finger protein

Continued on next page

Gene	Col-0	Anti-PLDα	Annotation
AT5G51195	-	+	similar to putative zinc finger protein
AT3G07200	-	+	zinc finger (C3HC4-type RING finger) family
AT1G04020	-	+	zinc finger (C3HC4-type RING finger) family
AT3G45560	-	-	zinc finger (C3HC4-type RING finger) family
AT5G02750	-	0	C3HC4-type zinc finger protein family
AT5G45290	-	0	C3HC4-type zinc finger protein family
AT3G21175	-	0	GATA zinc finger protein
AT2G28530	-	0	putative RING zinc finger protein
AT5G52010	-	0	zinc finger (C2H2 type) family protein
AT1G57730	0	+	C3HC4-type zinc finger protein family
AT3G09760	0	+	C3HC4-type zinc finger protein family
AT5G03450	0	+	zinc finger (C3HC4-type RING finger) family
AT5G43630	0	+	zinc knuckle (CCHC-type) family protein
AT4G02670	0	-	zinc finger (C2H2 type) family protein
AT1G55040	0	-	zinc finger (Ran-binding) family protein
AT1G32700	0	-	zinc-binding family protein
AT2G46770	+	-	no apical meristem (NAM) family protein
AT2G17040	+	0	no apical meristem (NAM) family protein
AT4G01520	0	+	no apical meristem (NAM) family protein
AT3G04430	0	+	no apical meristem (NAM) family protein

Table 2.12: Differentially regulated signaling genes. Changes in transcript profile are shown as positive (+) or negative (-)

Gene	Col-0	Anti-PLDα	Annotation
Calcium			
AT5G10450	0	-	14-3-3 protein GF14 lambda (grf6/AFT1)
AT5G16050	0	+	14-3-3 protein GF14 upsilon (grf5)
AT1G34760	+	+	14-3-3 protein GF14 omicron (grf11)

Continued on next page

Gene	Col-0	Anti-PLDα	Annotation
AT5G38480	+	0	14-3-3 protein GF14 psi (grf3/RCI1)
AT5G65930	-	+	putative kinesin calmodulin-binding protein
AT2G18750	-	+	calmodulin-binding protein
AT5G09410	0	+	calmodulin-binding protein
AT3G59440	+	0	calmodulin-like protein
Kinase			
AT3G04690	+	-	protein kinase family
AT2G40560	+	-	protein kinase family
AT1G53730	+	0	LRR transmembrane protein kinase 1, putative
AT3G57770	+	0	protein kinase - like protein
AT4G23130	+	0	protein kinase - like protein
AT3G57640	+	0	protein kinase family protein
AT5G01560	+	0	receptor like protein kinase
AT5G45430	+	0	serine/threonine-protein kinase protein
AT1G66830	-	+	LRR transmembrane protein kinase, putative
AT5G35370	-	+	receptor-like protein kinase - like protein
AT4G29450	-	+	leucine-rich repeat protein kinase
AT3G56370	-	-	LRR transmembrane protein kinase, putative
AT5G38210	-	-	protein kinase - like protein
AT5G66790	-	-	protein kinase family
AT4G38470	-	-	protein kinase like protein
AT2G05940	-	-	protein kinase, putative
AT4G27300	-	-	putative receptor protein kinase
AT2G14440	-	-	putative receptor-like protein kinase
AT4G18950	-	0	protein kinase - like protein
AT3G44610	-	0	protein kinase-like protein
AT4G38830	-	0	receptor-like protein kinase - like protein
AT1G43895	-	0	similar to protein kinase - like protein
AT3G12200	0	+	protein kinase, putative

Continued on next page

Gene	Col-0	Anti-PLDα	Annotation
AT2G39110	0	+	protein kinase, putative
AT4G39110	0	+	putative receptor-like protein kinase
AT5G08590	0	+	serine/threonine-protein kinase
AT1G35710	0	-	LRR transmembrane protein kinase, putative
AT1G74360	0	-	LRR transmembrane protein kinase, putative
AT3G46930	0	-	protein kinase 6-like protein
AT2G21480	0	-	protein kinase family
AT1G70460	0	-	protein kinase -related
AT5G02290	0	-	serine/threonine-specific protein kinase APK1A
Phosphatase			
AT1G03960	+	-	protein phosphatase 2A -related
AT1G03590	+	0	protein phosphatase 2C (PP2C)
AT3G12620	+	0	protein phosphatase 2C, putative
AT4G27800	-	0	protein phosphatase homolog (PPH1)
AT5G16480	-	0	tyrosine specific protein phosphatase family
AT2G14270	0	+	protein phosphatase 2C -related
AT4G26080	0	+	protein phosphatase ABI1
Response regulator			
AT1G74890	0	-	response regulator 7, putative
AT2G27070	+	0	Response Regulator: B- Type (ARR13)
AT5G24470	+	0	pseudo-response regulator, APRR5
GTP-related			
AT1G43890	-	0	Ras-related GTP-binding protein (RAB1Y)
AT3G05310	-	0	GTP-binding protein - related
AT5G60860	-	0	Ras-related GTP-binding protein
AT5G64990	-	0	Ras-related GTP-binding protein

Continued on next page

Gene	Col-0	Anti-PLDα	Annotation
AT1G09180	-	-	putative GTP-binding protein, SAR1B
AT1G49300	-	-	Ras-related GTP-binding protein
AT3G54840	-	-	Rab GTPase (ARA6)
AT2G44690	-	+	Rac-like GTPase (ARAC9)
AT2G27440	0	+	rac GTPase activating protein
Phospholipid			
AT1G21920	+	+	phosphatidylinositol-4-phosphate 5-kinase
AT2G10940	+	+	protease inhibitor/lipid transfer protein
AT5G19940	+	-	plastid-lipid associated protein
AT2G04940	+	0	scramblase-related
AT2G04940	+	0	scramblase-related
AT3G63240	+	0	inositol-1,4,5-trisphosphate 5-Phosphatase-like
AT4G38550	+	0	Phospholipase like protein
AT1G55260	-	+	lipid transfer protein (LTP) family
AT4G35790	-	+	PLDδ
AT3G54800	-	-	pleckstrin homology (PH) domain-containing/ lipid-binding START domain-containing
AT4G27350	-	0	membrane lipo protein
AT5G45560	0	+	(PH) domain- / lipid-binding START domain
AT2G18730	0	-	putative diacylglycerol kinase

Gene	AGI code	Forward	Reverse	Comments
PLDα 1	AT3G15730	TGGCAAGATTCTGGCAAATG	CTTCCTGGTTCTCCCAACTC	Phospholipase Dα 1
PLDα 1	AT3G15730	CTGACCACCGAACGATTGAG	AGCCTTGTCTCTGCGAGAAG	602 bp fragment for std. curve
PLDδ	AT4G35790	ACGATCCATGTGTTTGGGTT	GCCTCAGCCTCATCTTCATATT	Phospholipase Dδ
PLDδ	AT4G35790	GCTAGACCCACCGATACCGATG	CCCCAAGCGTCATATCCTTATCTC	572 bp fragment for std. curve
ROP8	AT2G44690	CTTATTTCCTACACCAGCAACAC	GTATCCCAGAGACCCAGATTGAC	ROP8/ARAC9
ROP8	AT2G44690	CATCATCAGCAACAGCTACAACG	CAAACCAAAGCCAGAGACTTGAG	633 bp fragment for std. curve
GR2	AT3G24170	TGTTCTTGCTTTGTCGCTTC	CGCCACCTTATCAATCTCACC	glutathione reductase 2
GR2	AT3G24170	TGTTCTTGCTTTGTCGCTTC	CGCCACCTTATCAATCTCACC	627 bp fragment for std. curve
GST	AT5G17220	TTGGTCGAGGATCTCAAAGTGAAG	GCATGTGCGTCAAATCAGCC	glutathione S-transferase, putative
GST	AT5G17220	TTTGAATCACGAGCCATCGC	CAGTGACCAGCCAGCACCAT	452 bp fragment for std. curve
Ankyrin-repeat	AT1G62050	TGGCGAGAAACCCGATGCTC	GCGTTTGTTGCTCCTCCTCTTC	Ankyrin repeat protein
Ankyrin-repeat	AT1G62050	AGATCAGTGGCTCTACCTTC	ATTGGTATAGCGAATGGGTC	515 bp fragment for std. curve
Tu-GTP	AT5G13650	GTTTGTAGGTTCTGGAGTGG	TTATGTTTGTCGCTGCCTTC	GTP-binding protein typA
Tu-GTP	AT5G13650	TATGCAGGGTGTTGGGTCGG	ACTGCCTGCCTTTCTTTGCC	512 bp fragment for std. curve
PIP5k	AT1G21920	GATTCAGGGGTTCTTGTGAC	ATTCACTGCCTTATTCGCTG	phosphatidylinositol -4-phosphate 5-kinase
PIP5k	AT1G21920	GGTGAGTGGTTTAATGGTCAAAGC	GCCGTAGCAGCCTTATTAGC	443 bp fragment for std. curve
GAPDH	AT3G26650	CGTGATCTAAGGAGAGCAAGA	TTCCCTTGAGGTTAGGGAGC	Glyceraldehyde-3 -phosphate dehydrogenase
GAPDH	AT3G26650	GAGTCACCGTGATCTAAGGAGA	CTTCACCGATATTTTGAGAACC	500 bp fragment for std. curve

Table 2.13: List of primers used in qRT-PCR

Chapter 3

Microarray Analysis

3.1 Introduction

DNA microarrays are a powerful means of monitoring the expression of thousands of genes simultaneously. A variety of computational and statistical methods have been employed to extract information from the large quantity of data generated from microarray experiments. Many methods assume the use of cDNA labeled with one of two fluorescent dyes to differentiate two treatments on a single microarray, implying data from two images to be analyzed. These methods include a number of data normalization techniques to reduce the effects of systematic errors and various kinds of statistical tests to identify differentially expressed genes in comparisons among different experimental conditions. There is as yet no single method that can be recommended under all circumstances for either normalization or identification of differential gene expression.

In recent years, ANOVA methods have gained popularity for identification of differential gene expression. The power of ANOVA methods derives from their flexibility in fitting and comparing different models to a given set of data Churchill (2004). One such method is the two-stage, log-linear ANOVA mixed models technique of Wolfinger *et al.* (2001). Its first stage uses a normalization

model designed to remove global effects across all microarrays. Its second stage uses a gene-specific model to estimate gene-treatment interactions as ratios of gene expression under control and treated conditions, along with a statistical significance. Kerr (2003) notes that the global normalization model employed in this technique is conducive to combining data across genes for realistic and robust models of error, especially when random effects are included. Pan (2002) compares different microarray statistical analysis methods and demonstrates that the log-linear ANOVA mixed model approach performs better than the t-test and regression approaches. The regression approach, although flexible and robust, assumes that the data is drawn from a normal distribution, while the t-test is limited due to very few degrees of freedom. Chu *et al.* (2004) compare two log-linear ANOVA mixed models for probe-level, oligonucleotide array data and found that both types of models capture key measurable sources of variability of oligonucleotide arrays for real and simulated data. Cui & Churchill (2003) review the use of a mixed ANOVA model for analyzing a cDNA microarray experiment and conclude that such models provide a powerful way to obtain information from experiments with multiple factors or sources of variation. Rosa *et al.* (2005) review issues of analyzing cDNA microarrays with mixed linear models and puts such analysis in the larger context of Bayesian analysis procedures and adjustments for multiple testing.

Data normalization is the first step in analyzing microarray data; numerous data normalization methods have been proposed and investigated. While refinements of existing methods continue to appear (e.g., Futschik & Crompton, 2004), naive methods, such as total intensity normalization, are still in use (e.g., Held *et al.*, 2004). Xie *et al.* (2004) did a comparative study of normalization methods and test statistics to analyze the results of a DNA-protein binding microarray experiment. Using performance and bias correction criteria, Bolstad *et al.* (2003) evaluate the cyclic lowess method, the contrast method, the quantile method, and baseline array scaling methods, both linear and non-linear; they demonstrate that normalization methods incorporating data from all microarrays perform better than methods employing a baseline array.

Several software tools that combine data normalization and statistical analysis are currently available. Dudoit *et al.* (2003) review these software tools with an emphasis on the TM4 microarray software suite, Bioconductor in R, and the BioArray Software Environment (BASE) system. Saeed *et al.* (2003) describe the features and capabilities of TM4, while Quackenbush (2002) describes

the normalization and transformation methods implemented in it. Williams *et al.* (2004), Zhu *et al.* (2004), and Khaitovich *et al.* (2004) have used TM4 in microarray data analysis. Another system is Expresso, an experiment management system that serves as a unifying framework to study data driven applications such as microarray experiments (Watkinson *et al.*, 2003; Sioson *et al.*, 2003; Heath *et al.*, 2002). Expresso has adapted the two-stage ANOVA mixed models technique of Wolfinger *et al.* (2001) to the particular needs of individual microarray data sets. Our experience with numerous such data sets has demonstrated that modeling the underlying experiment carefully and completely is essential to obtaining meaningful and defensible results. Use of tools that require experiments to conform to their analysis methods are less than satisfactory.

Here, the Expresso analysis methodology is compared to the approach provided in the TM4 microarray analysis software suite Saeed *et al.* (2003). Each tool is used to identify differentially expressed genes in two experimental data sets, each of which uses an *Arabidopsis thaliana* oligonucleotide array. Along the way, differences between the use of integrated intensity values (IIV) and median intensity values (MIV) as inputs are demonstrated. The interactions between normalization and gene identification methods are discussed. Finally, quantitative reverse-transcriptase PCR (qRT-PCR) results are used to assess the consistency of genes reported by TM4 and Expresso as having significant differential expression.

3.2 Materials and Methods

3.2.1 Median and integrated intensity values

This research considers two ways of measuring spot intensity, one or both of which are reported by typical microarray image processing software. The median intensity value (MIV) of a spot is the median value of all the pixels identified as part of the spot. The integrated intensity value (IIV) of a spot is the total value of all the pixels identified as part of the spot. In this research, both are background-corrected values. If the IIV data is unavailable, but the radius of the bounding circle of the spot and its average intensity value are available, then the IIV data can be estimated as the product of the average intensity and the number of pixels in the circle. This is the estimate used by ExpressConverter (below) when it converts GPR format data to MEV format data.

3.2.2 Microarray data sets

We used data sets from two experiments that employed the Arabidopsis Oligonucleotide Microarray (see: `http://www.ag.arizona.edu/microarray/`), which include 25,712 elements, each a gene-specific 70-mer (Qiagen/Operon, Valencia, CA) for a known or putative open reading frames in *Arabidopsis thaliana*. There are 48 blocks per microarray, 25 rows by 24 columns (600 spots) per block, and 28,800 spots per microarray, including spots for the 25,712 gene-specific 70-mers and 302 control elements. The remaining 2,786 spots are blank.

Experiment 1. Experiment 1 compares the responses of *Arabidopsis thaliana* wild type, ecotype Columbia-0 (henceforth, WT), and of an antisense plant for phospholipase D α (Anti-PLDα) to drought stress (Mane *et al.*, 2007; section 2.3.2,). Plants were harvested at a single time point, and two biological replicate hybridizations were done for each of WT and Anti-PLDα. ScanArray Express (PerkinElmer Life and Analytical Sciences, Inc., Boston, MA USA) was used to quantitate the four microarrays. By default, ScanArray Express performs a global lowess normalization of median intensities per microarray. ScanArray Express output its results to four files in GenePix (GPR) format, which constitute the Experiment 1 data set.

Experiment 2. Li *et al.* (2006) compare the responses to elevated CO_2 of a wild *Arabidopsis thaliana* relative (*Thellungiella halophila*, ecotype Shandong; Th) and of three *Arabidopsis thaliana* ecotypes: Wassilewskija (WS), Columbia (Col-0), and Cape Verde Islands (Cvi-0). Three biological replicate hybridizations were done for each genotype. GenePix (Axon Instruments, Union City, CA USA) was used to quantitate the twelve microarrays. GenePix also performs by default a global lowess normalization of median intensities per microarray. The output of GenePix is twelve GPR files, which constitute the Experiment 2 data set.

3.2.3 Real-time quantitative RT-PCR

For verification of microarray results in Experiment 2, Li *et al.* (2006) performed real-time quantitative reverse-transcriptase PCR (qRT-PCR) for selected genes — 55 in Col-0; 52 in Cvi-0; 59 in WS; 26 in Th. Table 3.7 contains the annotation of the selected genes. In brief, primer pairs were selected to represent unique sequences in the *Arabidopsis thaliana* genome and in the *Thellungiella* sequences deposited in NCBI. *Thellungiella* actin (CX129618) cDNA primers and *Arabidopsis thaliana*. Ubiquitin-10 cDNA primers were used as internal controls in the qRT-PCR analyses. RT-PCR products were detected using the fluorescent dye SYBR-green (Applied Biosystems, Forster City, CA USA) and the ABI PRISM/Taqman 7900 Sequence Detection System (Applied Biosystems, Forster City, CA USA). Dissociation curves were generated for each reaction to ensure specific amplification. Three repeats were done for each gene. The averaged threshold cycle numbers were used to estimate original mRNA levels.

3.2.4 Expresso analysis

Expresso analysis employs a general and flexible method to identify differentially expressed genes that is adapted from the two-stage analysis method of Wolfinger *et al.* (2001). In general, Expresso analysis consists of two log-linear ANOVA mixed models, called the normalization model and gene model. The first estimates and removes the experiment-wise systematic errors, while the second estimates and removes the gene specific errors. The residual that remains is the log-ratio estimate for each gene. In particular, the Tukey-Kramer multiple comparison of treatment effects on each gene is performed to estimate its expression level and the significance of (confidence in) that expression level. Expresso analysis is implemented for and executed on SAS (SAS/STAT version 8.2, SAS Institute Inc., Cary, NC USA).

The original model of Wolfinger *et al.* (2001) includes the treatment and the array as the main effects. In previous Expresso analysis, we have extended that model to experiment-appropriate models that include additional fixed and random effects. Here, the design of the two-dye oligonucleotide microarray used in Experiment 1 and Experiment 2 includes various controls strategically positioned in different blocks of the microarray. This makes it possible to estimate the random block effect in

each microarray. Furthermore, the dye effect is included in the normalization model to estimate and remove the global dye bias.

For this research, two Expresso models were developed, one whose gene model assesses the gene-(plant sample) effect (the GP model) and the other whose gene model assesses the gene-genotype-treatment effect (the GOT model). The GP model is much like previous Expresso models and is applicable to both Experiment 1 and Experiment 2. However, the GOT model is specific to analyzing Experiment 2. In both experiments, we used the GP model to estimate the differences in response of individual genotypes to treatment (drought stress versus control or ozone stress versus ambient ozone). However, the GOT model was used to estimate the effects of treatment (ozone stress), aside from the effect of individual genotypes.

Expresso GP model. The normalization model is

$$y_{spdab} \quad = \quad \mu + P_p + D_d + A_a + (P \times A)_{pa} + B_{ba} + r_{spdab}.$$

Each y_{spdab} value is the $\log_2$-transformed intensity of spot s within the dye d image in block b of array a. (A spot may represent a gene, a control, or a blank.) The global mean of the y_{spdab} values, over all microarrays, is μ. The fixed effects in the model are the plant sample effect P_p, where p indexes the various distinct plant samples from which mRNA was obtained, and the dye effect D_d, where d has two values for the two dyes. The random effects in the model are the array effect A_a, where a indexes the microarrays, the interaction effect $(P \times A)_{pa}$ of plant sample p with microarray a, and the block effect B_{ba}, where b identifies the block within microarray a. The model residual is r_{spdab}. This differs from the normalization model in Wolfinger *et al.* (2001), in that it incorporates dye and block effects. It is a refinement of the Expresso normalization model in Watkinson *et al.* (2003), in that it has no printing pin effect, which is specific to the analysis in the earlier paper, and includes the block effect.

The second stage of the analysis uses the residual values r_{spdab} computed in the first stage to estimate the interaction between an individual gene g and each plant sample p at a significance level $\leq \alpha = 0.05$. Index g is added to the residual values r_{spdab} resulting to r_{gspdab}. The value of g is determined using the mapping of s index values to g index values. The gene model is

$$r_{gspdab} \quad = \quad G_g + (G \times P)_{gp} + (G \times D)_{gd} + (G \times A)_{ga} + \lambda_{gspdab}.$$

Here, g is a spot that represents a gene (not a blank or control) within the dye d image in block b of array a. The value G_g is the mean of residual values for all spots that represent gene g in all images. The interactions $(G \times P)_{gp}$ of gene g with plant sample p and of $(G \times D)_{gd}$ of gene g with dye d are the fixed effects. The interaction $(G \times A)_{ga}$ of gene g with microarray a is a random effect. The λ_{gspdab} values are stochastic errors. This differs from the gene model in Wolfinger *et al.* (2001), in that it incorporates interactions between gene and dye and between gene and array. It refines the Expresso gene model in Watkinson *et al.* (2003) to include the interaction between gene and array.

The estimate of the expression level of each gene in each treatment comparison is done by computing the pair-wise least square mean differences of gene-treatment effects. The Tukey-Kramer multiple comparison of gene-(plant sample) effects on each gene is made to estimate the p values associated with each calculated expression level. If there are ρ plant samples, then there are $\binom{\rho}{2}$ possible pairwise comparisons. If we index the plant samples from 1 to ρ, then the null hypothesis for gene g and comparison i, j, where $1 \leq i < j \leq \rho$, is

$$H_o : (G \times P)_{gi} \quad = \quad (G \times P)_{gj}.$$

The difference $(G \times P)_{gi} - (G \times P)_{gj}$ is the estimate of the $\log_2$(fold change) of gene g in the experimental comparison P_i versus P_j.

The above GP model was used to analyze both Experiment 1 and Experiment 2. In both experiments, there are 48 blocks per array. In Experiment 1, there are four arrays and four plant samples, namely, WT-control, WT-stressed, antiPLD-control, and antiPLD-stressed. In Experiment 2, there are 12 arrays and eight plant samples, namely, Col-0-test, Col-0-control, Cvi-0-test, Cvi-0-control, WS-test, WS-control, Th-test, and Th-control.

Expresso GOT model We wanted to estimate the gene-treatment effects separately from gene-genotype-treatment interaction effects using just one model, To do this, we designed the gene-

genotype-treatment model, an alternative set of log-linear ANOVA mixed models, for the elevated CO_2 experiment where we unfold the genotype information from the plant sample factor in the GP model. This resulted in a normalization model that includes the genotype effect (O_o) with 4 levels (Col-0, Cvi-0, WS, and Th) and basic treatment effect (T_t) with 2 levels (test and control). The random array (A_a) effect however needs to be removed from the model since it confounds the genotype effect.

The normalization model is

$$y_{sotdab} \quad = \quad \mu + O_o + T_t + D_d + (O \times T)_{ot} + B_{ba} + r_{sotdab}.$$

Each y_{sotdab} value is the $\log_2$-transformed intensity of spot s for genotype o and treatment t within the dye d image in block b of array a. We have that μ is as in the GP model. The fixed effects in the model are the genotype effect O_o, where o indexes the genotype (organism), the treatment effect T_t, where t is the treatment, and the dye effect D_d, where d has two values for the two dyes. The random effects in the model are the interaction effect $(O \times T)_{ot}$ of genotype o with microarray a, and the block effect B_{ba}, where b identifies the block within microarray a. The model residual is r_{sotdab}. This differs from the normalization model in Wolfinger *et al.* (2001), in that it incorporates genotype (organism), dye, and block effects. It is a refinement of the Expresso normalization model in Watkinson *et al.* (2003), in that it has no printing pin effect, and includes the genotype and block effects.

The second stage of the analysis uses the residual values r_{sotdab} computed in the first stage to estimate the interaction among an individual gene g, each genotype o, and each treatment t, at a significance level $\leq \alpha = 0.05$. Index g is added to the residual values r_{sotdab} resulting to $r_{gsotdab}$. The value of g is determined using the mapping of s index values to g index values. The gene model is

$$r_{gsotdab} \quad = \quad G_g + (G \times O)_{go} + (G \times T)_{gt} + (G \times O \times T)_{got} + (G \times D)_{gd} + \lambda_{gsotdab}.$$

Here, G_g are as for the GP model. The interactions $(G \times O)_{go}$ of gene g with genotype o, $(G \times T)_{gt}$ of gene g with treatment t, $(G \times O \times T)_{got}$ of gene g with genotype o and treatment t, and $(G \times D)_{gd}$

of gene g with dye d are fixed effects. The $\lambda_{gsotdab}$ values are stochastic errors. This differs from the gene model in Wolfinger *et al.* (2001), in that it incorporates interactions between gene and genotype, between gene and genotype with treatment, and between gene and array. It refines the Expresso gene model in Watkinson *et al.* (2003) to include interactions between gene and genotype and between gene and genotype with treatment.

We do pairwise comparisons as in the GP model, but there are two plausible classes of null hypotheses to test within the GOT gene model. If τ is the number of treatments, then the class 1 null hypothesis for gene g and comparison i, j, where $1 \leq i < j \leq \tau$, is

$$H_1 : (G \times T)_{gi} \quad = \quad (G \times T)_{gj}.$$

The difference $(G \times T)_{gi} - (G \times T)_{gj}$ is the estimate of the $\log_2$(fold change) of gene g in the T_i versus T_j comparison. This particular comparison looks for gene-treatment effects that are independent of genotype.

We can still estimate the expression level of each gene with respect to a specific genotype level by computing the pair-wise least square differences of gene-genotype-treatment interaction effects. The class 1 null hypothesis for gene g and comparison i, j, where $1 \leq i < j \leq \tau$, is

$$H_2 : (G \times O \times T)_{goi} \quad = \quad (G \times O \times T)_{goj}.$$

The difference $(G \times O \times T)_{goi} - (G \times O \times T)_{goj}$ is the estimate of the $\log_2$(fold change) of gene g of a specific genotype o for the T_i versus T_j comparison.

The GOT model applies to Experiment 2 in a straightforward way. There are 12 arrays, two treatments, and four genotypes, namely, Col-0, Cvi-0, WS, and Th.

3.2.5 TM4 microarray software suite

The TM4 microarray software suite consists of several components freely available from the TM4 web site (`http://www.tm4.org/`). This research employed the ExpressConverter, Microarray Data

Analysis Software (MIDAS), and Microarray Experiment Viewer (MEV) components. ExpressConverter converts microarray data from various data formats, such as the GenePix Results (GPR) format, to the MEV format, which is used by MIDAS and MEV. MEV format includes only integrated intensity values (IIV), which is the kind of intensity values expected of all TM4 components.

3.2.6 MIDAS data normalization methods and filters

Low intensity and saturated spots are marked by quantitation programs. These spots are filtered out from the data before doing any further normalization or statistical analysis. Data normalization methods proceed from the assumption that only a relatively small proportion of the genes change significantly in expression level between the two hybridized mRNA samples. This assumption is reasonable for these data sets since the hybridizations and subsequent analysis address nearly all *Arabidopsis thaliana* genes. The MIDAS component of TM4 provides a number of data normalization methods and filters and supports applying them in a pipelined fashion (Quackenbush, 2002; Saeed *et al.*, 2003).

Total intensity normalization. While our assumption implies that the average measured intensities of the two channels of a cDNA or oligonucleotide microarray should be almost the same, these averages are often significantly different, due to differences in the inherent fluorescence of the two dyes. The total intensity normalization step in TM4 is a straightforward means to eliminate this global dye bias. For each spot i, where $1 \leq i \leq n$, let R_i and G_i be the measured intensities of the spot in the two channels. The normalized intensity data for spot i is $G'_i = \kappa G_i$ and $R'_i = R_i$, where κ is the normalization factor $\kappa = (\sum_{i=1}^{n} R_i)/(\sum_{i=1}^{n} G_i)$. Quackenbush Quackenbush (2002) discusses this normalization in the context of several variations that are possible to address differing channel intensities.

Lowess normalization. Beyond the global dye bias, there is dye bias that is dependent on the measured spot intensities (Yang *et al.*, 2002b,a). TM4 constructs a scatter plot, called an RI-plot, of the points (x_i, y_i), where $1 \leq i \leq n$, given by $x_i = \log_{10}(R_i G_i)$ and $y_i = \log_2(R_i/G_i)$. Under our

assumption, the RI-plot should be very nearly symmetric with respect to the line $y = 0$. In lowess normalization, TM4 applies the lowess method of Cleveland (1979) to fit a locally weighted regression curve to the RI-plot; TM4 then adjusts spot intensities to eliminate any systematic intensity-dependent bias. Additional details on correcting intensity-dependent bias is found in Quackenbush (2002).

Standard deviation regularization. After total intensity and lowess normalizations eliminate dye bias on a global (per microarray) scale, TM4 employs standard deviation regularization to ensure that the per-block variances of $\log(R_i/G_i)$ values are the same (Yang *et al.*, 2002b; Huber *et al.*, 2002). Quackenbush (2002) provides the formulas for this normalization step.

Low intensity filtering. Since the relative error in the $\log(R_i/G_i)$ values increases if R_i or G_i is close to background levels, spots with low intensities are filtered out. Quackenbush (2002) provides additional details, which essentially require that both R_i and G_i intensities be above two standard deviations of the respective backgrounds.

The MIDAS pipeline. We applied a MIDAS pipeline consisting of total intensity normalization, lowess normalization, standard deviation regularization, and low intensity filtering to both microarray data sets. MIDAS default parameters were used throughout; the default low intensity filter cut-off is $R_i G_i < 10,000$.

3.2.7 TM4 MEV analysis

The Multi Experiment Viewer (MEV) component of TM4 provides a number of statistical analyses and clustering algorithms to identify differentially expressed genes. We report results from the one-class t-test analysis applied to output of the MIDAS pipeline. This test assumes that the paired distribution of treated and control groups is normally distributed. Since the intensities measured from the same spot are correlated, we can apply the one-class t-test for the two-group comparison.

3.3 Results and Discussion

Here, a portion of the results obtained in the comparison of Expresso analysis and the TM4 pipeline are reported (section 3.2). Figure 3.1 illustrates the overall flow of the statistical analyses of microarray data that were done in this study. We began with microarray data in GPR format from Experiment 1 and Experiment 2. Median intensity values (MIV) from the GPR files can be analyzed by the Expresso GP and GOT models directly. ExpressConverter provides integrated intensity values (IIV) for further Expresso and TM4 analysis. The MIDAS normalization and filtering pipeline executes these steps in order: total intensity normalization (subscript T), lowess normalization (subscript L), standard deviation regularization (subscript S), and low intensity filter (subscript F). MIDAS allows tapping the output of any step in the pipeline; for example, IIV_{TL} signifies an MEV file after total intensity normalization followed by lowess normalization. The identification of genes with significant differential expression was performed on all GPR and MEV files, using the Expresso GP and GOT models and the *t*-test in MEV.

3.3.1 Normalization and low intensity filtering in TM4

Quackenbush (2002) describes the use of ratio-intensity plots (RI-plots) to detect and normalize for any systematic intensity-dependent dye bias using lowess normalization (section 3.2). We evaluated the effect of lowess normalization within the context of the flow in figure 3.1 by creating RI-plots after each step for the second replicate microarray in Experiment 1, WT plant. figure 3.2 contains these RI-plots. The IIV_{TL} is indeed effective, for this data set, in correcting systematic dye bias, suggesting that preprocessing by these two normalization steps in MIDAS may be a good practice in many situations.

The normalization and filtering pipeline affects the number of genes identified as differentially expressed in both the GP and GOT models. See Table 3.1. For example, in Experiment 1, the GP model using IIV input data identifies 567 up-expressed genes in the WT microarrays, while it identifies only 460 WT genes as up-expressed if IIV_{TLSF} (processed by the complete MIDAS pipeline) input data is used.

Small changes in the number of genes identified as up- or down-expressed after successive MIDAS steps may mask larger changes in the composition of sets of up- and down-expressed genes. To obtain a more precise view of the effects of MIDAS changes, we computed retention counts (RC) and retention percentages (RP) between the gene successive sets whose numbers are in Table 3.1. RC is the number of genes in the set before the MIDAS step that remain in the set after the step. RP is the percentage of remaining genes with respect to the number of genes in the set after the MIDAS step. Table 3.2 contains the RC and RP values corresponding to the counts in Table 3.1. For Experiment 1, there is a tremendous drop in retention during the lowess normalization that follows the total intensity normalization. There is not a drop of corresponding magnitude for Experiment 2. For both experiments, normalization has a significant effect on the sets of genes identified as differentially expressed.

In Experiment 1 results, the number of genes commonly assessed by Expresso as significantly expressed when using IIV and IIV_{T} is high. For example, there is 95.45% retention of WT genes (545 total) assessed as up-expressed when using IIV_{T} data in Expresso compared to that when using IIV data. Retention percentage of these genes assessed as expressed however went down after doing lowess normalization. There is only 15.20% retention of WT genes (74 total) assessed as up-expressed in the results when using IIV_{TS} data in Expresso compared to when using IIV_{T}. While we observe increase in the retention percentage in IIV_{TLS} (from IIV_{TL}) and IIV_{TLSF} (from IIV_{TLS}), there's low retention percentage in the results using IIV_{TLSF} from IIV data. This can be traced in the low retention percentage of IIV_{TL} from IIV_{T}. Hence, the normalization method that affects the results in Experiment 1 the most is lowess normalization.

The results of Expresso on Experiment 2 show that low retention percentages happen after application of total intensity normalization (lowest is 59.30%) and after application of low intensity filtering (lowest is 61.19%). The low retention percentages shown in the IIV $\cap$ IIV_{TLSF} column implies that the normalization pipeline also significantly affects the results in Expresso analysis of Experiment 2.

3.3.2 Choice of intensity signal data

The input to statistical analysis of microarray experiments is a set of real numbers that represent the measured intensity signal for each spot in a microarray. Much statistical analysis of microarray data has traditionally used median intensity values (MIV). The alternative used in TM4 is the integrated intensity value (IIV). (section 3.2) Since IIV is intended to integrate the measured intensity across the biological sample printed at a spot, one might expect IIV to be a more accurate assessment of the biological measurement than MIV data. For example, a spot having 100 pixels and a median intensity of 5,000 has the same IIV as a spot having 50 pixels and a median intensity of 10,000.

This study provides the opportunity to observe the difference that choosing MIV or IIV makes on the sets of genes ultimately identified as differentially expressed. We used the GP model to analyze unnormalized MIV and IIV data from Experiment 1, and we used the GOT model to analyze unnormalized MIV and IIV data from Experiment 2. Table 3.3 reports a summary of the results. In Experiment 1, 725 WT genes are assessed as up-expressed and 774 WT genes are down-expressed when MIV data are used in Expresso. These numbers decreased to 567 up-expressed genes and 552 down-expressed genes when IIV data are used instead. A similar trend is observed in Experiment 2 results when using MIV and IIV data. These results suggest that employing IIV input data with Expresso analysis leads to more conservative results than employing MIV input data.

3.3.3 Comparison of Statistical Methods

We compared the performance of the GP model, the GOT model, and the t-test of MEV in identifying differentially expressed genes in Experiment 2. We used the IIV_{TLSF} data of Experiment 2 as input to these methods. We also contrast these results with MIV data analyzed in GP. Table 3.4 reports counts for these analyses.

The plot in figure 3.3a demonstrates that the estimates of $\log_2$(fold change) are the same in GP and GOT. As figure 3.3b shows, the p values by GOT are smaller compared to the p values calculated by GP. Use of the MEV t-test resulted in fewer genes assessed as significantly expressed when compared to the numbers for the GP model. The results obtained when MIV data was used

as input to GP, is closest to the results when using IIV_{TLSF}.

To compare the effectiveness of Expresso and TM4 in identifying gene differential expression, we compared the identified direction of differential expression of a select set of genes per genotype in Experiment 2 with results obtained by qRT-PCR. See Table 3.5. The lowest overall percentage (71.9%) of agreement is between the qRT-PCR results and the MEV t-test results using IIV_{TLSF}. The log(fold change) estimates of the GP model has 77.1% percentage agreement with the qRT-PCR results, which is slightly higher than the percentage for the MEV t-test. The results of the GP model using MIV data demonstrated the greatest agreement, 90.1%, with the qRT-PCR results.

Figures 3.4, 3.5, and 3.6 present the actual assessed $\log_2$(fold change) values for 50 selected Col-0 genes in Experiment 2, along with their qRT-PCR values. These are the 50 Col-0 genes, among the 55 with qRT-PCR values, for which we have expression values for all methods. For each gene, a histogram of the $\log_2$(fold change) estimates of qRT-PCR, the GP model using MIV, the GP model using IIV_{TLSF}, and the MEV t-test is given. The 50 histograms are spread over three figures to enhance readability and are in increasing order by qRT-PCR estimated change. In general, the $\log_2$(fold change) estimates of the GP model and of the MEV t-test, all IIV_{TLSF} input data, are approximately the same, while being slightly different from estimates of the GP model using MIV input data. As might be expected, disagreement between qRT-PCR and microarray results are more prevalent for small estimated $\log_2$(fold change) values. The histograms for genes AT4G09020 (figure 3.4), AT1G35580 (figure 3.5), and AT3G29360 (figure 3.5) show that the direction of log(fold change) estimate of qRT-PCR matches the direction of the GP model using MIV input data, while differing from the direction of the estimates of the GP model and the MEV t-test using IIV_{TLSF} input data.

Table 3.6 summarizes genotype-specific correlation results, which demonstrate that the GP model using MIV input data has the highest correlation with qRT-PCR compared to the GP model and the MEV t-test using IIV_{TLSF} input data. The highest correlation of 0.85 is for the Col-0 results of qRT-PCR versus the GP model using MIV input. The corresponding correlations for Cvi-0, WS, and Th are 0.73, 0.66, and 0.84, respectively, which are all best among the analysis methods.

3.4 Conclusion

Our integration and comparison of Expresso analysis and the capabilities of TM4 has highlighted successes in microarray analysis, some similarities, and some differences. The success of microarray analysis is demonstrated by considerable agreement between qRT-PCR results and the results of all the examined microarray analysis methods. The greatest agreement was found when median intensity value (MIV) inputs were analyzed with the Expresso GP analysis model. We also found that the use of integrated intensity value (IIV) inputs for Expresso analysis consistently resulted in fewer genes identified as differentially expressed when compared to results from MIV inputs. This suggests that the use of IIV inputs is more conservative than the use of MIV inputs, while MIV inputs may give greater agreement to qRT-PCR results than IIV inputs.

Our results demonstrate that the MIDAS normalization and filtering pipeline corrects systematic intensity-dependent dye bias on a per microarray basis. The normalization stage in Expresso analysis removes global effects across all microarrays and complements the per microarray normalization methods of MIDAS. The generally better agreement of Expresso analysis with qRT-PCR results when compared to the MEV t-test suggests that it would be desirable for MEV to have an ANOVA test that has the greater flexibility of the Expresso gene model.

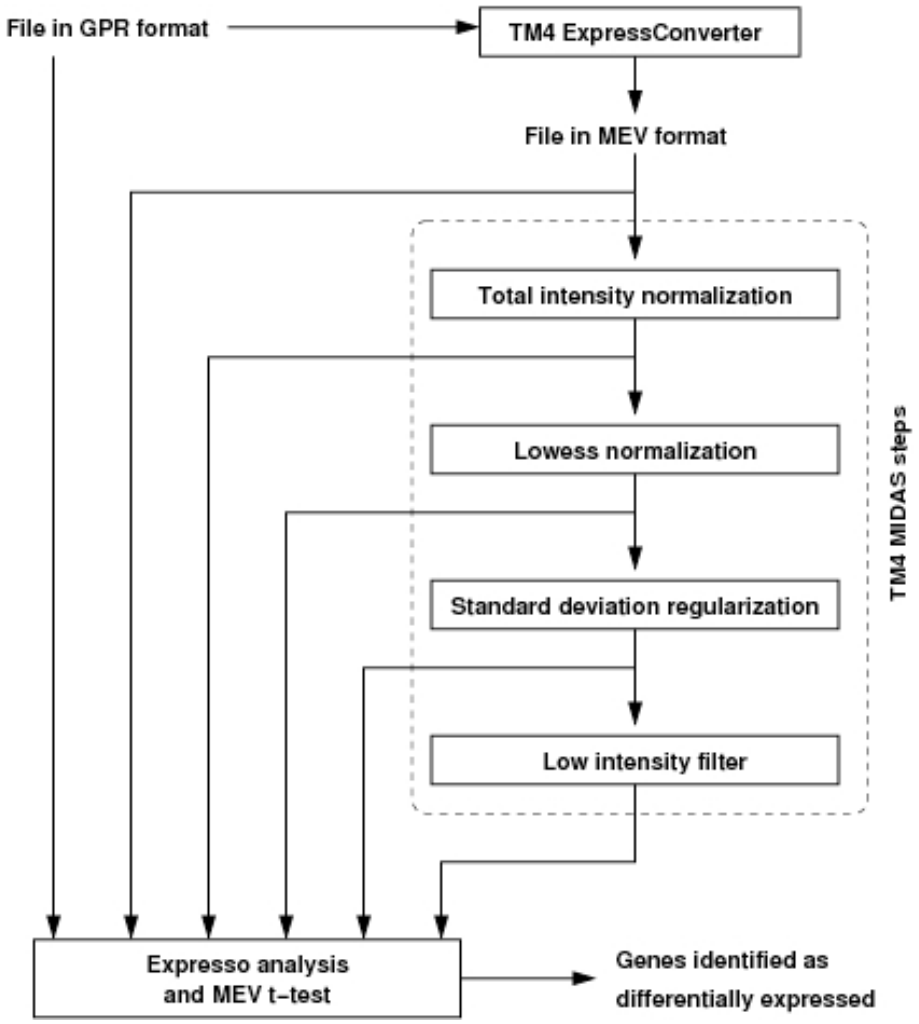
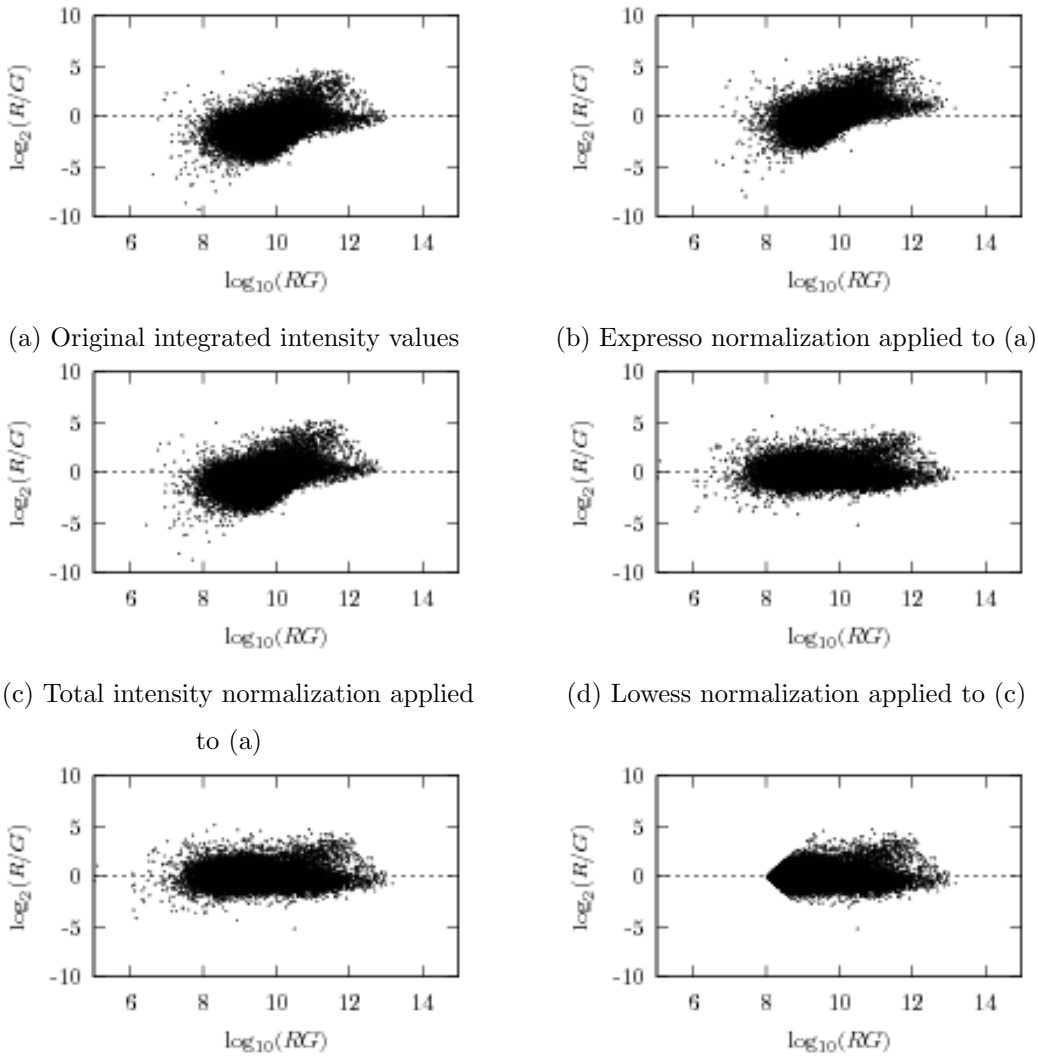


Figure 3.1: The overall flow of the statistical analyses of microarray data. Data input is in GPR format and provides the MIV for each spot. TM4 analysis requires ExpressConverter to generate MEV format containing the IIV for each spot. The normalization steps performed by MIDAS are T (total intensity normalization), L (lowess normalization), S (standard deviation regularization), and F (low intensity filter). Differential gene expression is obtained from the Expresso GP model, the Expresso GOT model (Experiment 2 only), and the *t*-test in MEV.



(a) Original integrated intensity values

(b) Expresso normalization applied to (a)

(c) Total intensity normalization applied to (a)

(d) Lowess normalization applied to (c)

(e) Standard deviation regularization applied to (d)

(f) Low intensity filtering applied to (e)

Figure 3.2: RI-plots illustrating specific intensity-dependent dye bias as microarray data is processed. (a) The original data (IIV) for the Experiment 1, WT plant samples, and second replicate microarray. (b) Expresso normalization ($\mathrm{IIV_E}$, where E signifies Expresso normalization) applied to IIV. (c) MIDAS total intensity normalization ($\mathrm{IIV_T}$) applied to IIV. (d) MIDAS lowess normalization ($\mathrm{IIV_{TL}}$) applied to $\mathrm{IIV_T}$. (e) MIDAS standard deviation regularization ($\mathrm{IIV_{TLS}}$) applied to $\mathrm{IIV_{TL}}$. (f) MIDAS low intensity filtering ($\mathrm{IIV_{TLSF}}$) applied to $\mathrm{IIV_{TLS}}$.

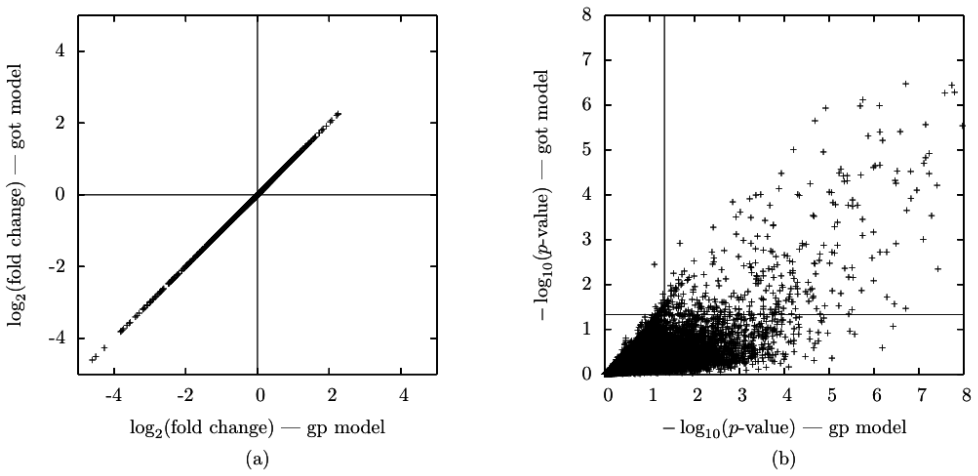


Figure 3.3: Comparison of estimated $\log_2$(fold change) and the corresponding p-value estimates for the GP and GOT model results of the WS ecotype values in Table 3.4 (a) This is a scatter plot of the estimated $\log_2$(fold change) values from the GP and GOT models; these values are essentially identical. (b) This is a scatter plot of the $-\log_{10}$(p-value), again for the GP and GOT models. The dotted lines correspond to $-\log_{10}(0.05)$, as our significance cutoff is 0.05. For the GP model, it is the points to the right of the dotted line that are significant. For the GOT model, it is the points above the dotted line that are significant.

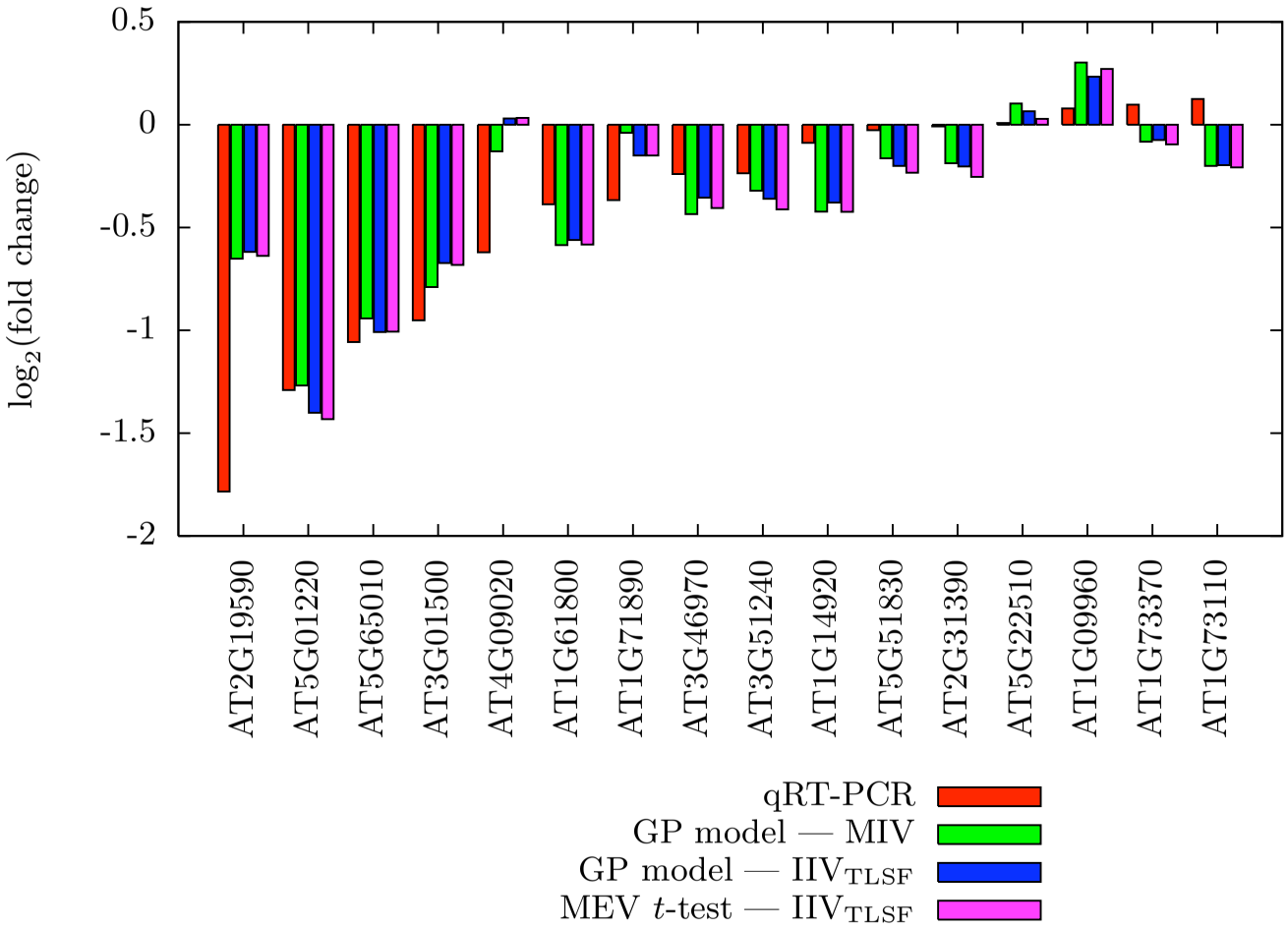


Figure 3.4: Comparison of qRT-PCR results to Expresso and MEV t-test results for first 16 of 50 selected genes of the Col-0 genotype of Experiment 2.

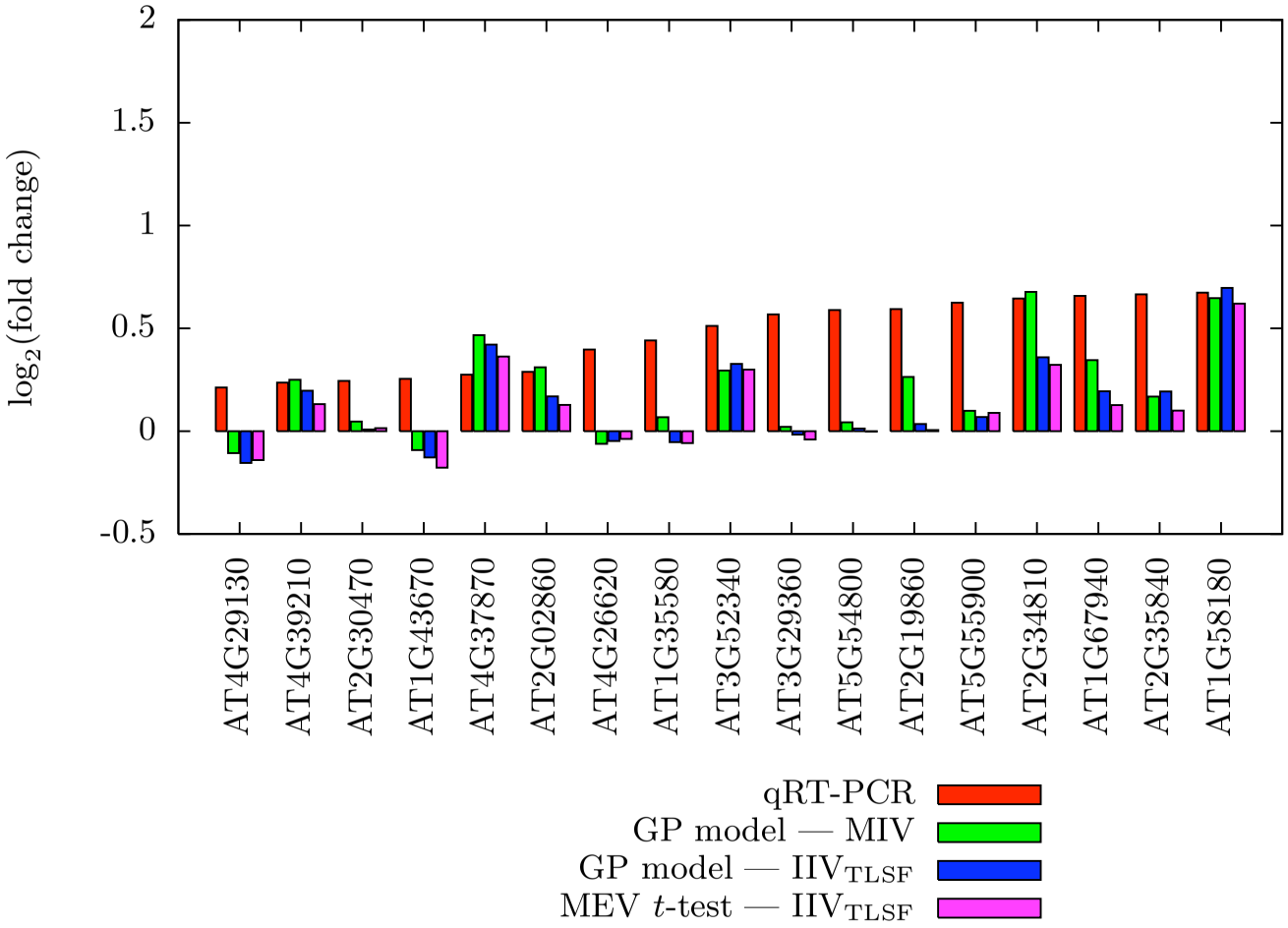


Figure 3.5: Comparison of qRT-PCR results to Expresso and MEV t-test results for the middle 17 of 50 selected genes of the Col-0 genotype of Experiment 2.

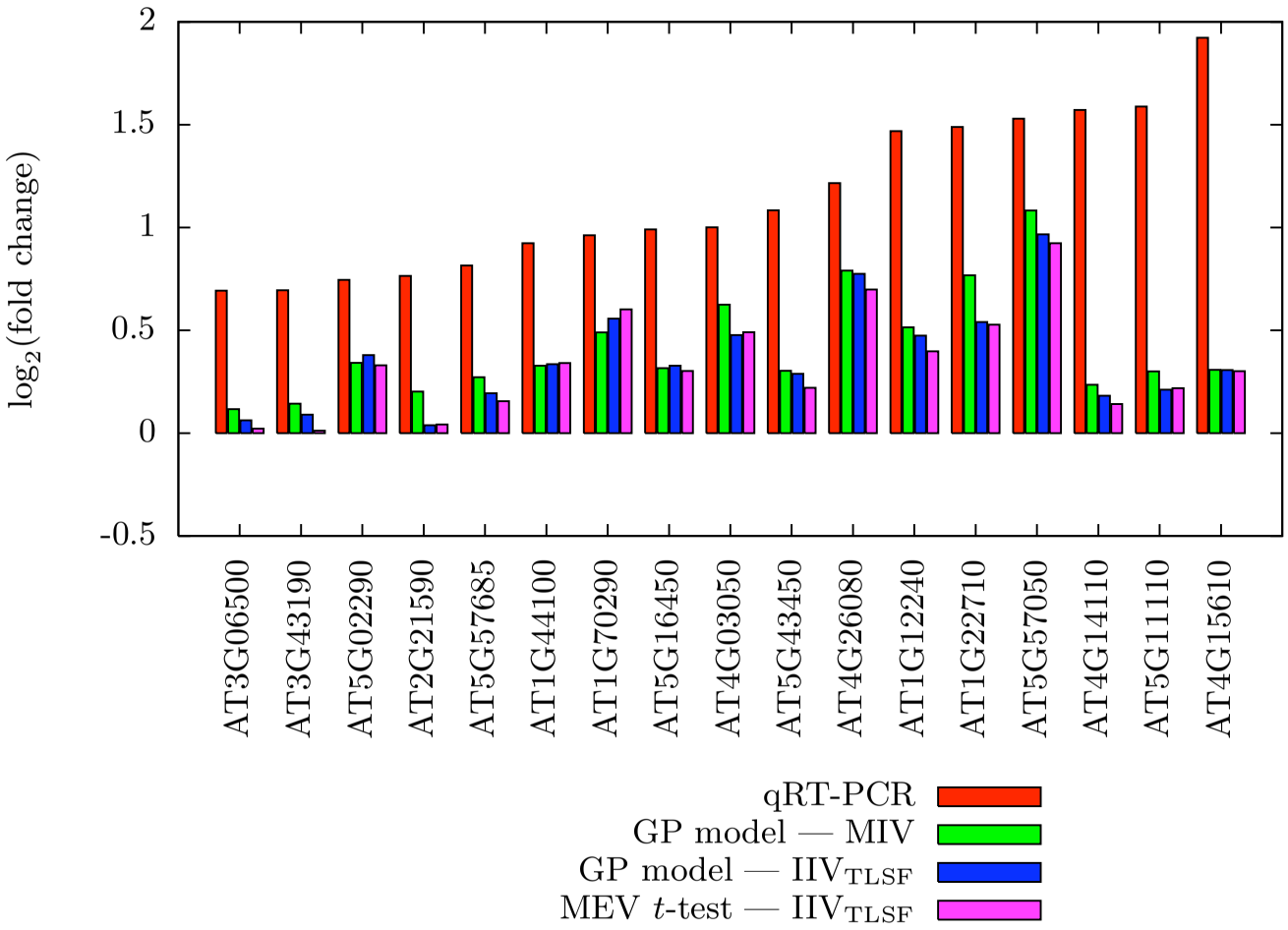


Figure 3.6: Comparison of qRT-PCR results to Expresso and MEV t-test results for the last 17 of 50 selected genes of the Col-0 genotype of Experiment 2.

Table 3.1: Numbers of identified differentially expressed genes in the GP and GOT models after preprocessing by 0 or more MIDAS pipeline steps — IIV, IIV_{T}, IIV_{TL}, IIV_{TLS}, or IIV_{TLSF}. For Experiment 1, up-expression (+) and down-expression (−) numbers are given for both WT and Anti-PLDα. For Experiment 2, + and −1 numbers are given for all 4 genotypes separately. The ALL entries correspond to the H1 hypotheses of the GOT model (see section 3.2).

Genotype		IIV	IIV_{T}	IIV_{TL}	IIV_{TLS}	IIV_{TLSF}
		GP Model — Experiment 1				
WT	+	567	571	487	495	460
WT	−	552	553	499	499	460
Anti-PLDα	+	442	421	440	471	431
Anti-PLDα	−	336	346	381	363	334
		GOT Model — Experiment 2				
Col-0	+	227	309	242	240	270
Col-0	−	122	171	178	180	183
Cvi-0	+	106	108	82	78	71
Cvi-0	−	63	93	74	71	67
WS	+	234	346	279	269	150
WS	−	331	347	340	340	224
Th	+	238	398	350	363	361
Th	−	196	245	241	246	240
ALL	+	298	428	357	354	290
ALL	−	202	261	253	248	195

Table 3.2: Retention counts (RC) and retention percentages (RP) for the differentially expressed gene sets of Table 3.1. RC is the number of genes in a set before the MIDAS step that remain in the set after that step. RC is the number of genes in the set before the MIDAS step that remain in the set after that step. RP is the percentage of remaining genes with respect to the number of genes in the set after the MIDAS step. RC and RP are reported for the intersections $\mathrm{IIV}_{\cap}\mathrm{IIV}_{\mathrm{T}}$, $\mathrm{IIV}_{\mathrm{T}} \cap \mathrm{IIV}_{\mathrm{TL}}$, $\mathrm{IIV}_{\mathrm{TL}} \cap \mathrm{IIV}_{\mathrm{TLS}}$, and $\mathrm{IIV}_{\mathrm{TLS}} \cap \mathrm{IIV}_{\mathrm{TLSF}}$, as well as intersection $\mathrm{IIV}_{\cap} \cap \mathrm{IIV}_{\mathrm{TLSF}}$, which corresponds to the effect of MIDAS steps from the start of the pipeline to the end. The ALL entries correspond to the H1 hypotheses of the GOT model (see section 3.2).

		IIV ∩ $\mathrm{IIV}_{\mathrm{T}}$		$\mathrm{IIV}_{\mathrm{T}}$ ∩ $\mathrm{IIV}_{\mathrm{TL}}$		$\mathrm{IIV}_{\mathrm{TL}}$ ∩ $\mathrm{IIV}_{\mathrm{TLS}}$		$\mathrm{IIV}_{\mathrm{TLS}}$ ∩ $\mathrm{IIV}_{\mathrm{TLSF}}$		IIV ∩ $\mathrm{IIV}_{\mathrm{TLSF}}$	
Genotype		RC	RP	RC	RP	RC	RP	RC	RP	RC	RP
		GP Model — Experiment 1									
WT	+	545	95.45	74	15.20	285	57.58	456	99.13	59	12.83
WT	−	532	96.20	55	11.02	306	61.32	459	99.78	53	11.52
Anti-PLDα	+	403	95.72	58	13.18	278	59.02	430	99.77	46	10.67
Anti-PLDα	−	326	94.22	45	11.81	203	55.92	330	98.80	33	9.88
		GOT Model — Experiment 2									
Col-0	+	224	72.49	221	91.32	219	91.25	172	63.70	159	58.89
Col-0	−	121	70.76	150	84.27	165	91.67	120	65.57	92	50.27
Cvi-0	+	81	75.00	65	79.27	71	91.25	49	69.01	36	50.70
Cvi-0	−	60	64.52	60	81.08	63	88.73	41	61.19	38	56.72
WS	+	223	64.45	230	82.44	248	92.19	108	72.00	101	67.33
WS	−	293	84.44	267	78.53	314	92.35	180	80.36	149	66.52
Th	+	236	59.30	308	88.00	330	90.91	256	70.91	201	55.68
Th	−	183	74.69	210	87.14	225	91.46	172	71.67	144	60.00
ALL	+	282	65.89	307	85.99	333	94.07	200	68.97	177	61.03
ALL	−	196	75.10	196	77.47	230	92.74	133	68.21	103	52.82

Table 3.3: Comparison of MIV and IIV input data with respect to the sets of genes ultimately identified as differentially expressed. The genotype and set labelings are the same as those in Table 3.1. The GP model was used to analyze the unnormalized MIV and IIV data from Experiment 1. The GOT model was used to analyze the unnormalized MIV and IIV data from Experiment 2. Up-expressed and down-expressed counts are reported for both experiments and all genotypes. The count of genes in the intersections is found in column Common. For convenience, percentages of the intersection with respect to the MIV and IIV sets are tabulated in the last two columns. The ALL entries correspond to the H1 hypotheses of the GOT model (see section 3.2).

Genotype		MIV	IIV	Common	% of common in MIV	% of common in IIV
		GP Model — Experiment 1				
WT	+	725	567	501	69.10%	88.36%
WT	−	774	552	515	66.54%	93.30%
Anti-PLDα	+	518	422	357	68.92%	84.60%
Anti-PLDα	−	505	336	313	61.98%	93.15%
		GOT Model — Experiment 2				
Col-0	+	311	227	198	63.67%	87.22%
Col-0	−	190	122	110	57.89%	90.16%
Cvi-0	+	185	106	89	48.11%	83.96%
Cvi-0	−	97	63	59	60.82%	93.65%
WS	+	326	234	181	55.52%	77.35%
WS	−	482	331	273	56.64%	82.48%
Th	+	363	238	219	60.33%	92.02%
Th	−	251	196	166	66.14%	84.69%
ALL	+	448	298	249	55.58%	83.56%
ALL	−	294	202	159	54.08%	78.71%

Table 3.4: Comparison of the number of genes identified as differentially expressed by Expresso analysis and the MEV t-test. Counts for both up-expressed and down-expressed genes, as well as all four genotypes of Experiment 2, are reported. The first three analyses — the GP model, the GOT model, and the MEV t-test — take the IIV_{TLSF} data as input. For point of comparison, the last analysis uses the GP model on MIV input.

	GP Model IIV_{TLSF}		GOT Model IIV_{TLSF}		MEV t-test IIV_{TLSF}		GP Model MIV	
Genotype	+	−	+	−	+	−	+	−
Col-0	1837	1563	270	183	2125	2007	1761	1212
Cvi-0	817	1248	71	67	478	823	669	1403
WS	2606	2002	150	224	535	388	2422	1184
Th	2016	1314	361	240	1144	827	2395	1457
Totals	7276	6127	852	714	4282	4045	7247	5256

Table 3.5: Comparison of qRT-PCR results with identified up-expressed and down-expressed genes in Experiment 2 using Expresso and TM4. Results of qRT-PCR are available for each of the four ecotypes in the numbers n in the second column. Numbers of agreement or non-agreement are shown in the S, D, and F columns. The S (same) column tabulates the number of genes for which the sign of the log(fold change) for statistically significant differential expression matches the direction of change in the corresponding qRT-PCR result. The D (differing) column tabulates the number of genes for which the sign of the log(fold change) for statistically significant differential expression is in the opposite direction of the change in the corresponding qRT-PCR result. The F (filtered) column tabulates the number of genes for which there is a qRT-PCR result, but for which either the gene was filtered by MIDAS low intensity filtering or the analysis method did not assess the change in expression as statistically significant. The MEV t-test (first grouping) results are for the typical TM4 process, which involves IIV input data followed by the four MIDAS steps, which we denote IIV_{TLSF}. The GP model (second grouping) gives the same numbers and uses the same IIV_{TLSF} input data. the GP model (third grouping) results use MIV input data and has no filtered genes.

		GP Model		GP Model			MEV t-test		
	qRT-PCR	MIV		IIV_{TLSF}			IIV_{TLSF}		
Genotype	n	S	D	S	D	F	S	D	F
Col-0	55	50	5	43	8	4	40	9	4
Cvi-0	52	46	6	38	9	5	36	7	9
WS	59	54	5	46	10	3	41	9	9
Th	26	23	3	21	3	2	19	3	4
Total	192	173	19	148	30	14	138	28	26
Percentage		90.1%	9.9%	77.1%	15.6%	7.3%	71.9%	14.6%	13.5%

Table 3.6: Correlation of qRT-PCR results from the Expresso GP model and the MEV t-test. For each comparison, the analytical result for each gene for which there is a qRT-PCR result for a particular genotype is assembled into a result vector. SAS was employed to compute a Pearson correlation of each result vector with the corresponding vector of qRT-PCR results. The computed correlations are as reported above.

	Genotype-Specific Correlation			
Comparison	Col-0	Cvi-0	WS	Th
qRT-PCR versus GP model with MIV input	0.85	0.73	0.66	0.84
qRT-PCR versus GP model with IIV_{TLSF} input	0.83	0.69	0.61	0.78
qRT-PCR versus MEV t-test with IIV_{TLSF} input	0.83	0.65	0.64	0.77

	Amino acid metabolism
AT5GG5010	asparagine synthetase 2 (ASN2)
	Carbon metabolism
AT1G09960	sucrose transporter / sucrose-proton symporter (SUT4)
AT1G12240	beta-fructosidase (BFRUCT4) / beta-fructofuranosidase / invertase, vacuolar
AT1G22710	sucrose transporter / sucrose-proton symporter (SUC2)
AT1G35580	beta-fructofuranosidase / invertase / saccharase / beta-fructosidase
AT1G43670	fructose-1,6-bisphosphatase / D-fructose-1,6-bisphosphate 1-phosphohydrolase
AT1G58180	carbonic anhydrase family protein / carbonate dehydratase family protein
AT1G73110	ribulose bisphosphate carboxylase/oxygenase activase / RuBisCO activase
AT1G73370	sucrose synthase / sucrose-UDP glucosyltransferase
AT2G21590	ADPglucose pyrophosphorylase Large subunit 3, APL3
AT2G31390	pfkB-type carbohydrate kinase family protein
AT2G35840	sucrose-phosphatase 1 (SPP1)
AT3G01500	carbonic anhydrase 1, chloroplast / carbonate dehydratase 1 (CA1)
AT3G06500	beta-fructofuranosidase / invertase / saccharase / beta-fructosidase
AT3G29360	UDP-glucose 6-dehydrogenase
AT3G43190	sucrose synthase / sucrose-UDP glucosyltransferase
AT3G46970	Glucan phosphorylase (cytosolic), PHS2
AT3G52340	sucrose-phosphatase 3b (SPP3b)
AT4G03050	encodes a 2-oxoglutarate-dependent dioxygenase
AT4G09020	isoamylase / starch debranching enzyme
AT4G26620	sucrase-related
AT4G37870	phosphoenolpyruvate carboxykinase (ATP) / PEP carboxykinase / PEPCK

Continued on next page

AT4G39210	ADPglucose pyrophosphorylase Large subunit 4, APL4
AT5G11110	SPS5a, sucrose-phosphate synthase -like protein
AT5G22510	beta-fructofuranosidase / invertase / saccharase
AT5G43450	1-aminocyclopropane-1-carboxylate oxidase
AT5G51830	pfkB-type carbohydrate kinase family protein
AT5G55900	sucrase-related
	Hormones
AT1G14920	gibberellin response modulator (GAI) (RGA2) / gibberellin-responsive modulator
AT2G19590	1-aminocyclopropane-1-carboxylate oxidase / ACC oxidase
	Lipid metabolism
AT5G01220	UDP-sulfoquinovose:DAG sulfoquinovosyltransferase / sulfolipid synthase (SQD2)
	Nucleic acids
AT2G30470	ABI3/VP1 family regulatory protein
	Secondary metabolism
AT3G51240	naringenin 3-dioxygenase / flavanone 3-hydroxylase (F3H)
AT5G16450	dimethylmenaquinone methyltransferase family protein
	Signaling
AT2G19860	hexokinase 2 (HXK2)
AT4G14110	COP9 signalosome subunit / CSN subunit (CSN8), CSN8, FUS7
AT4G26080	ABI1 protein phosphatase ABI1
AT4G29130	hexokinase 1 (HXK1)
AT5G02290	serine/threonine-specific protein kinase NAK
AT5G57050	ABI 2 protein phosphatase 2C
	Stress
AT1G70290	TPS8, trehalose-6-phosphate synthase, putative

Continued on next page

	TRANSPORT
AT1G44100	amino acid permease 5
AT1G61800	glucose-6-phosphate/phosphate translocator
AT1G67940	ABC transporter family protein
AT1G71890	sucrose transporter / sucrose-proton symporter (SUC5)
AT2G02860	sucrose transporter / sucrose-proton symporter (SUC3)
AT5G54800	glucose-6-phosphate/phosphate translocator
	UNCATEGORIZED
AT2G34810	putative berberine bridge enzyme
AT4G15610	integral membrane family protein
AT5G57685	expressed protein

Table 3.7: Genes Subjected to qRT-PCR. Annotation of genes subjected to qRT-PCR. For verification of microarray results in Experiment 2, performed real-time quantitative reverse-transcriptase PCR (qRT-PCR) for selected genes — 55 in Col-0; 52 in Cvi-0; 59 in WS; 26 in Th.

Chapter 4

PathPrime - A tool for designing and testing primers for disease diagnostics

4.1 Introduction

Rapid diagnosis of disease-causing agents is extremely important since delayed diagnosis can result in disease spread and delayed prophylaxis. It is even more important in an era where disease-causing agents are used as bioterrorism and agroterrorism agents (Fauci, 2003). Pathogen diagnostics has become an important part in several areas including biodefense, health care, agriculture and food safety, clinical research and drug discovery. Lack of a proper diagnostic assay can sometimes have serious consequences to the patient and may result in disease spread. A diagnostic assay should be able to provide a very rapid, highly sensitive and specific result.

4.1.1 Diagnostic methods

There are many ways to diagnosing pathogens. One of the oldest ways is visual observation of the microorganism under a microscope and identifying it based on morphological characteristics. Microbes can also be identified by growing them on differential culture media. Another approach

to identifying pathogens is by analyzing microbial macromolecular composition and their metabolic by-products, and use of specific immunologic reagents have created a variety of systems. The gold standard of pathogen identification has been the detection of host immune response (*e.g.,* antibodies) to known organism.

However, these conventional methods have several drawbacks. Visual appearance of microorganisms is nonspecific making it hard to identify the right pathogen. Culture-based methods are insensitive for detecting some bacteria and viruses. Culture-based methods are limited to pathogens with known growth requirements. These methods are not good at discriminating microbes with common behavioral traits. These methods fail to identify infection caused by organisms that do not elicit a detectable host immune response. Immuno-assays require special chemical consumables that add considerably to the logistic burden and costs. These conventional techniques are not well suited for identifying large numbers of environmental or clinical samples and viral diagnostics, and are often time-consuming and laborious.

Rapid advances in sequencing technology has resulted in sequencing of thousands of microorganisms in recent years. Availability of genomic sequences has made it possible to characterize microorganisms at the molecular level. Each pathogen species contains characteristic DNA or RNA sequences that can be used to differentiate from other organisms. The developments in the area of genomics and the efforts to overcome problems like culturing of microorganisms, false positives and low sensitivity in pathogen detection have led to the development of DNA-based diagnostic methods such as polymerase chain reaction (PCR) amplification techniques. PCR-based methods are rapid, highly specific and sensitive, often do not require culturing of pathogens and can be deployed in a high throughput manner with minimum labor.

4.1.2 Designing primers for PCR-based pathogen detection

Polymerase chain reaction involves a pair of primers (forward and reverse) to amplify a region of DNA/RNA. PCR is a sensitive method capable of amplifying even a single DNA molecule. In order to reliably detect pathogen or family of pathogens, the primers should be highly specific. So for PCR-based detection methods, identifying genomic regions that are unique (based on existing sequence data), conserved (across different strains of pathogen) and long enough to match the needs

of the detection technology is the key. For larger group of pathogens and for viruses (especially retroviruses, which have high mutation rates), it is not possible to design a single primer-pair. Several groups have suggested and successfully used degenerate primer-pairs for pathogen identification and cloning of homologous genes (Rose *et al.*, 2003; Briese *et al.*, 2004; Moës *et al.*, 2005). Although use of degenerate primers increases flexibility of PCR, it also increases complexity of primer design. Degeneracy is a critical factor in sensitivity of PCR. It also increases the number of false positives and reduces specificity. There are several tools available for both degenerate and non-degenerate primer design (see `http://www.humgen.nl/primer_design.html` for a comprehensive list of free and commercial primer designing software). However, these are not suitable for designing primers for diagnostics since, among other reasons, these software tools do not check for specificity of the resulting primers. Therefore, it is possible that the primers resulting from the use of these tools may amplify non-target contaminant sequences (*e.g.,* host sequences or sequences of common lab contaminants) or closely related non-pathogenic organisms leading to false positives.

Existing tools for signature detection/primer design

Real public awareness of bioterrorism began in the US after the anthrax attacks in 2001. Numerous problems, including false positives and false negatives, from different types of detection assays exposed the weaknesses of pathogen diagnostics at that time. Since then several groups have developed tools to identify pathogen specific genomic signatures for pathogen detection. Some of these tools are reviewed below.

KPATH. Slezak *et al.,* (2003) developed a computational pipeline called KAPTH. KPATH can take a single or multiple sequences as input, aligns them using DIALIGN (for smaller genomes) or MGA (for larger genomes). It then generates a *consensus gestalt*, which is then used by Vmatch to compare against *all_microbes* database. Any subsequence of at least 14 nucleotides in the *consensus gestalt* that matches *all_microbes* database is masked out. The resulting *uniqueness gestalt* is then mined for potential candidate signatures. The signatures are then tested with Primer3 program and filtered by a "custom software" to yield a final set of signature candidates. This computational pipeline was not available for download at the time of this writing.

Signature Detection Pipeline. This pipeline was developed at Los Alamos Laboratory by Jason Gans' group. This pipeline takes a group of sequences, partitions them into a subsequences and compares them against background sequences using mpiBLAST. Unique sequences are isolated by subtracting from background genomes. A minimal set of candidate signatures (oligomers) are identified by filtering with experimental constraints. This minimal set of sequences are validated once again against all sequences. This computational pipeline was not available for download at the time of this writing.

Insignia. Insignia is a computational pipeline to generate unique DNA signatures for any and all pathogens in the Insignia database. It involves two steps. The first step involves a computationally intensive subtraction of target genome sequences of a group of pathogens against background sequences. A list of subsequences that are unique to the pathogen group is isolated. In the second step, the subsequences from each pathogen within the group is compared to each other in a pair-wise manner using MUMmer. The final output is a list of all unique oligomers of user-specified length specific to a group of pathogens.

Greene SCPrimer. Greene SCPrimer is a web-based primer design program developed by Jabado *et al.* (2006). It takes an alignment of desired sequences as input. Extracts sub-alignments of length appropriate for PCR primer, computes pair-wise identity and constructs a phylogenetic tree. It then scores and filters primers based on common PCR parameters such as melting temperature, GC content, etc. It then identifies a minimum set of primers using a greedy set covering algorithm (Slavik, 1997). The user is then required to match forward and reverse primers for product size and melting temperature.

The above software tools are computationally very expensive and require large computing clusters (KPATH, Signature Detection Pipeline and Insignia) or require user intervention (Greene SCPrimer) at different stages of primer design. Greene SCPrimer also does not screen primers for background/non-target genome amplification. KPATH and Signature Detection Pipeline are not

available for download.

Here I present PathPrime, a tool that can design primers, computationally test them against target genes and potential contaminant sequences, and identify a minimum set of primers that can unambiguously detect a given list of sequences.

4.2 PathPrime workflow

PathPrime involves two broad steps. The first step involves preparation of PCR template files for background or non-target genomes/sequences and the second step involves actual primer design and testing. Preparation of PCR template files is done by using the script `makePCRtemplate`. The script, `makePCRtemplate`, uses `famap` and `fahash`, which are part of NCBIs `e-PCR` tool (Schuler, 1997). `famap` creates a mmapped-file with sequence data in fast random-accessible format and `fahash` extracts a string of consecutive letters (a "word") of specified length from the $3'$ end of each sequence and converts it to a unique integer (a "hash value"). Using the hash value for indexed access to a table of sequences allows potential matches to be evaluated efficiently as the query primers are scanned. This word-based strategy significantly speeds up the search for primer annealing within the sequence.

A schematic representation of PathPrime primer design workflow is depicted in Figure 4.1. The script `PathPrime` takes a single multi-fasta file containing nucleotide sequences as input. Each step in `PathPrime` is described below.

1. **Sequence clustering.** In the first step, the sequences are grouped into sub categories using NCBI's `BLASTCLUST` tool (Altschul *et al.*, 1990). `BLASTCLUST` automatically and systematically clusters DNA sequences based on pairwise matches found using the Mega BLAST algorithm. Mega BLAST search is performed for all the sequences combined against a database created from the same sequences. `BLASTCLUST` finds pairs of sequences that have statistically significant matches and clusters them using single-linkage clustering.

2. **Multiple sequence alignment.** Clusters having two or more sequences are used to construct multiple sequence alignment. `MUSCLE` is used in case of smaller genomes (Edgar, 2004). For larger genomes, `DIALIGN` is used (Morgenstern, 2004). IUPAC consensus sequences are

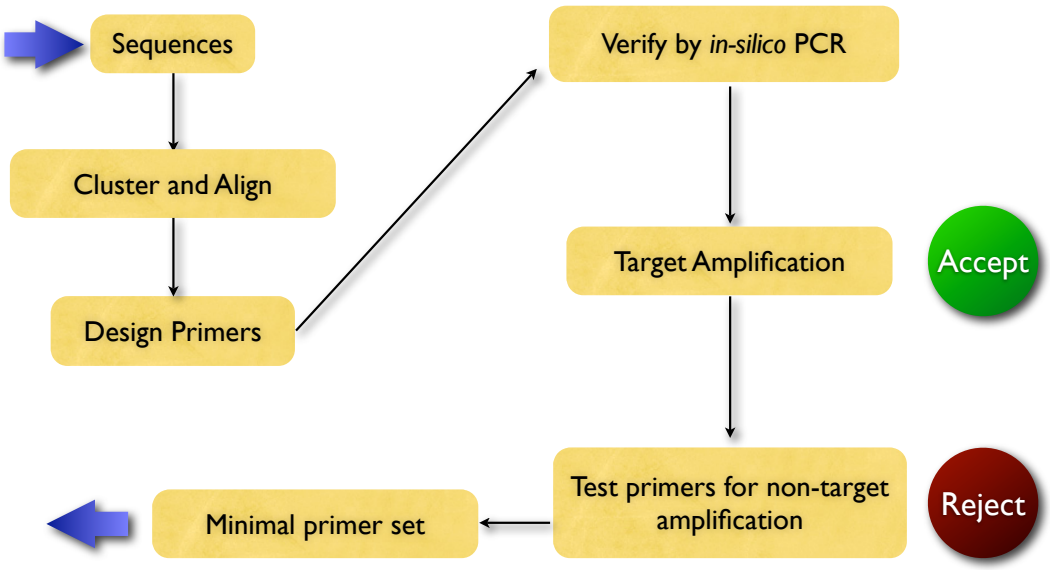


Figure 4.1: The PathPrime workflow. Input consists of a fasta file with one or more genome sequences. The sequences are clustered using BLASTCLUST. Clusters with two or more sequences are aligned to generate consensus sequences. The consensus sequences together with singlet sequences are used to design primers using Primer3. The resulting primer-pairs are tested against target and background sequences. Primer-pairs that amplify background sequences are discarded. The final output is a minimal set of primer-pairs that can amplify all sequences in the input file.

generated from each multiple sequence alignment. Degenerate nucleic acid codes in IUPAC consensus sequences are replaced with 'N's.

3. **Primer design.** Modified consensus sequences and singlet sequences are used to design primers using `Primer3`. Primer design parameters such as melting temperature, GC content, amplicon length, etc can be specified in a separate configuration file. The configuration file can also be used to set/change parameters used by other PathPrime steps as well. A total of four best primer-pairs will be designed for each input sequence.

4. ***In-silico* testing against targets.** The primer-pairs will then be used for *in-silico* testing against target sequences (input sequences) using `re-PCR`. Primer-pairs are ranked by the number of sequences they amplify.

5. ***In-silico* testing against background sequences.**The primer-pairs are then tested against background sequences with relaxed PCR conditions using `re-PCR`. Any primer-pair that amplify a background sequence will be discarded.

6. **Identifying a minimal primer set** Identifying a minimal set of primers is a NP-hard problem so a greedy algorithm was used. The greedy algorithm for set coverage chooses primer-pairs according to one rule: at each stage, choose the primer-pair which can amplify the largest number of un-amplified (by the previous primer-pairs) sequences. It can be shown that this algorithm achieves an approximation ratio of $H(s)$, where s is the size of the largest set and $H(n)$ is the n^{th} harmonic number:

$$H_n = \sum_{k=1}^{n} \frac{1}{k} \leq \ln n + 1$$

The greedy algorithm achieves an approximation ratio of $\log_2(n)/2$.

PathPrime was written in Perl. In order to run the script, Perl v5.8.6 or higher needs to be installed. The PathPrime script uses several perl modules and programs. The list of modules and programs used is provided in Table 4.1.

Table 4.1: Programs and Perl modules used by PathPrime.

Program/Module Name	**Description**	**URL**
`BLASTCLUST`	This program is a part of NCBI local blast package	NCBI
`e-PCR`	e-PCR provides a search mode using primer sequences against a sequence database	NCBI
`MUSCLE`	A tool for clustering nucleic acid and protein sequences	Drive5
`Primer3`	Primer3 is a widely used program for designing PCR primers	Primer3
`BioPerl`	Two packages from BioPerl, Core and Run packages, are needed. Core package provides functions for basic handling and manipulation of sequences. Run packages provide wrappers for executing common bioinformatics applications. The following BioPerl modules are used by our pipeline: `Bio::SeqIO`, `Bio::Tools::Run::Alignment::Muscle`, `Bio::Tools::Run::Primer3`	BioPerl
`Array::Unique`	This module lets you create an array which will allow only one occurrence of any value	CPAN
`Tie::DxHash`	Tie::DxHash keeps insertion order and allows duplicate hash keys	CPAN
`Algorithm::SetCovering`	This module is used to identify a minimum set of primer-pairs described in section 4.2	CPAN

4.3 Biological Case Studies

The PathoSystems Resource Integration Center (PATRIC) is one of eight Bioinformatics Resource Centers (BRCs) at Virginia Bioinformatics Institute (VBI) funded by the National Institute of Allergy and Infection Diseases (NIAID) to create a data and analysis resource for selected NIAID priority pathogens, specifically proteobacteria of the genera Brucella, Rickettsia and Coxiella, and Coronavirus, Calicivirus, Lyssavirus and viruses associated with Hepatitis A and E. The goal of PATRIC is to provide a comprehensive bioinformatics resource for these pathogens, including consistently annotated genome, proteome and metabolic pathway data to facilitate research into countermeasures, including drugs, vaccines and diagnostics. Here I use PathPrime to design primers for Lyssaviruses.

4.3.1 Designing primers for Lyssaviruses

Twenty one complete genome sequences were downloaded from GenBank. The parameters used for primer-design were specified in the configuration file (Appendix C.1). In the first step, PathPrime produced 4 clusters. Cluster 1 contained 16 Rabies viruses, cluster 2 contained 2 Mokola viruses, cluster 3 contained 2 Australian bat lyssaviruses and cluster 4 contained a single sequence from Rabies virus strain HEP-Flury. Multiple sequence alignments from each cluster with more than one sequences were generated using `MUSCLE`. The consensus sequences were processed to replace degenerate nucleotides with 'N's. The sequences were then used to design primers using `Primer3`. A total of 15 primer-pairs were designed. The resulting primer-pairs were used to perform *in-silico* PCR against the input sequences. Stringent amplification conditions were used for target sequences. Primer-pairs producing at least one amplicon were retained. In this case, all primer-pairs produced at least one amplicon.

In-silico testing against non-target (background) genomes that include human, canine and feline, and partial or complete sequences of Coronaviruses (3299 sequences), other Rhabdoviruses (759 sequences) was carried out. Amplification conditions were relaxed, and up to two bp mismatch and/or two gaps/primer was allowed in case of non-target genomes. None of the primers were able to produce an amplicon of at most 2kb from any of the background genomes. This suggests that the primers are highly specific to Lyssaviruses. Using the set-covering algorithm, a minimal set of

3 primer-pairs were obtained. The primer sequences and the number of sequences they amplify are given in Table 4.3. The relative location and length of each amplicon on the Lyssavirus genome is shown in Figure 4.2 and Table 4.4.

4.3.2 Validation of results

In order to assess the validity of the primers, the primers were compared with existing degenerate diagnostic primers currently in use by CDC (Rupprecht Lab). The forward and reverse sequences of the degenerate primer-pairs are ACGCTTAACGAMAAA and GTRCTCCARTTAGCRCACAT, respectively. Nearly 5,000 partial and complete Lyssavirus sequences were downloaded from GenBank. Both sets of primers were compared using *in-silico* PCR. The degenerate primers were able to amplify 6 sequences while the primer-pairs obtained from PathPrime were able to amplify 99 sequences. The primers will also be tested in wet-lab for validation of the tool.

4.3.3 Conclusion and Future Work

We have developed a tool called PathPrime to design and test primers that can specifically amplify target pathogens. *In-silico* testing of primers shows that it is a valuable tool for design primers for pathogen diagnostics. The primer designed for Lyssaviruses will be tested in the lab to confirm their effectiveness in identifying different strains of Lyssaviruses. The future goal is to make a web-based tool where users can directly upload sequences and get the results without have to install it locally on their computers.

Table 4.2: Starting data for the primer design process. The DNA sequences for 21 Lyssavirus complete genome sequences were collected from NCBI as the input data. The GenBank accession numbers are listed in this table.

Accession no.	**Remarks**
Rabies virus	
AB009663.2	Rabies virus genomic RNA
AB044824.1	Rabies virus strain:Nishigahara
AB085828.1	Rabies virus strain:HEP-Flury
AB128149.1	Rabies virus strain:Ni-CE
AF499686.2	Rabies virus strain SRV9
AY956319.1	Rabies virus serotype 1
DD161903.1	Attenuated Rabies Virus
DD161904.1	Attenuated Rabies Virus
DD161905.1	Attenuated Rabies Virus
DD161906.1	Attenuated Rabies Virus
DQ099524.1	Rabies virus strain FluryLEP
DQ099525.1	Rabies virus strain PM1503
DQ875050.1	Rabies virus strain MRV
DQ875051.1	Rabies virus strain DRV
M13215.1	RAVMMGN Rabies virus
M31046.1	RAVCGA Rabies virus (strain SAD B19)
NC_001542.1	Rabies virus, complete genome
Australian bat lyssavirus	
AF081020.2	Australian bat lyssavirus
NC_003243.1	Australian bat lyssavirus
Mokola virus	
NC_006429.1	Mokola virus
Y09762.1	Mokola virus

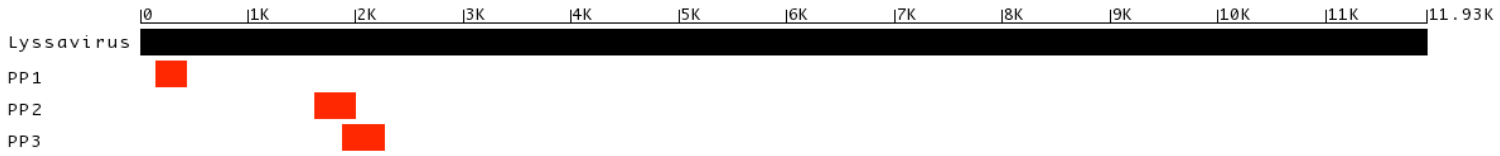


Figure 4.2: Locations of the amplicons produced by the 3 primer-pairs. Primer-pairs were designed and screened as described in section 4.3. The location of the amplicon for each primer pair, in red, is shown relative to the representative Lyssavirus genome, in black.

Table 4.3: Primer-pair sequences and number of Lyssavirus genomes amplified. The three primer-pairs that amplify all of the 21 Lyssavirus genomes are presented in this table. The name of the pair, the sequences of the forward and reverse primers, and the number of nucleoprotein sequences predicted to be amplified by the primer pair is displayed here.

ID	**Forward**	**Reverse**	**NSA***	**Remarks**
PP1	TCAATATGAGTACAAGTACCCTGC	GCCCAATTCCCTTCTACATC	17	Rabies
PP2	GAGCAACCAAGCTCACCTTC	GAGCTAACTGAAGCGGTTGG	2	Australian bat
PP3	AAGTCAACCCAGACGACACC	CTTGGAGTGAGACGATGCAA	2	Mokola

*NSA: Number of sequences amplified

Table 4.4: Sequences amplified by each primer pair. *In-silico* PCR was used to determine which target sequences were amplified by each primer pair. The table shows the name of the primer pair, the accession number for each sequence it amplifies, the corresponding amplicon length as well as the target sequence coordinates for start and end of the amplicon.

Primer-pair	Target Accession	Start position	End Position	Amplicon Length
PP1	NC_001542.1	142	432	291
PP1	DQ875051.1	142	432	291
PP1	DQ875050.1	142	432	291
PP1	DD161906.1	142	432	291
PP1	DD161905.1	142	432	291
PP1	DD161904.1	142	432	291
PP1	DD161903.1	142	432	291
PP1	AB009663.2	142	432	291
PP1	AB044824.1	142	432	291
PP1	DQ099525.1	72	362	291
PP1	DQ099524.1	72	362	291
PP1	AY956319.1	142	432	291
PP1	AF499686.2	142	432	291
PP1	AB128149.1	142	432	291
PP1	AB085828.1	142	432	291
PP1	M13215.1	142	432	291
PP1	M31046.1	142	432	291
PP2	NC_006429.1	1622	1990	369
PP2	Y09762.1	1622	1990	369
PP3	NC_003243.1	1876	2262	387
PP3	AF081020.2	1876	2262	387

Chapter 5

Bibliography

Bibliography

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) functions as transcriptional activators in abscisic acid signaling. *The Plant Cell*, **15**, 63–78.

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Toshisuke, I., Hosokawa, D. & Shinozaki, K. (1997) Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *The Plant Cell*, **9**, 1859–1868.

Addicott, F. T., Carns, H. R., Lyon, J. L., Smith, O. E. & McMeans, J. L. (1964) On the physiology of abscission. *Regulateurs naturels de la croissance vegetale* (ed. J. P. Nitsch), pp. 687–703. Cent. Nat. Rech. Sci., Paris.

Agarwal, P., Agarwal, P., Reddy, M. & Sopory, S. (2006) Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Reports*, **25**, 1263–74.

Aharoni, A., Keizer, L. C. P., Bouwmeester, H. J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H. A., Blaas, J., van Houwelingen, A. M. M. L., De Vos, R. C. H., van der Voet, H., Jansen, R. C., Guis, M., Mol, J., Davis, R. W., Schena, M., van Tunen, A. J. & O'Connell, A. P. (2000) Identification of the SAAT Gene Involved in Strawberry Flavor Biogenesis by Use of DNA Microarrays. *The Plant Cell*, **12**, 647–662.

Allan, A., Fricker, M., Ward, J., Beale, M. & Trewavas, A. (1994) Two transduction pathways mediate rapid effects of abscisic acid in Commelina guard cells. *The Plant Cell*, **6**, 1319–1328.

Allen, G., Chu, S., Harrington, C., Schumacher, K., Hoffmann, T., Tang, Y., Grill, E. & Schroeder, J. (2001) A defined range of guard cell calcium oscillation parameters encodes stomatal movements. *Nature*, **411**, 1053–1057.

Allen, G., Kwak, J., Chu, S., Llopis, J., Tsien, R., Harper, J. & Schroeder, J. (1999) Cameleon calcium indicator reports cytoplasmic calcium dynamics in Arabidopsis guard cells. *The Plant Journal*, **19**, 735–747.

Allison, D., Cui, X., Page, G. & Sabripour, M. (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nature Reviews Genetics*, **7**, 55–65.

Alscher, R. (1989) Biosynthesis and antioxidant function of glutathione in plants. *Physiologia Plantarum*, **77**, 457–464.

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **215**, 403–410.

Anna, A., Bohnert, H. J. & Bressan, R. A. (2005) Abiotic Stress and Plant Genome Evolution. Search for New Models. *Plant Physiology*, **138**, 127–130.

Anthony, R., Henriques, R., Helfer, A., Meszaros, T., Rios, G., Testerink, C., Munnik, T., Deak, M., Koncz, C. & Bogre, L. (2004) A protein kinase target of a PDK1 signalling pathway is involved in root hair growth in Arabidopsis. *The EMBO Journal*, **23**, 572–581.

Anthony, R. G., Khan, S., Costa, J., Pais, M. S. & Bogre, L. (2006) The Arabidopsis Protein Kinase PTI1-2 Is Activated by Convergent Phosphatidic Acid and Oxidative Stress Signaling Pathways Downstream of PDK1 and OXI1. *Journal of Biological Chemistry*, **281**, 37536–37546.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Arbeitman, M. N., Furlong, E. E. M., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W. & White, K. P. (2002) Gene Expression During the Life Cycle of *Drosophila melanogaster*. *Science*, **297**, 2270–2275.

Bartels, D. & Salamini, F. (2001) Desiccation tolerance in the resurrection plant *Craterostigma plantagineum*: A contribution to the study of drought tolerance at the molecular level. *Plant Physiology*, **127**, 1346–1353.

Blatt, M., Thiel, G. & Trentham, D. (1990) Reversible inactivation of K^+ channels of Vicia stomatal guard cells following the photolysis of caged inositol 1,4,5- trisphosphate. *Nature*, **346**, 766–769.

Bolstad, B., Irizarry, R., Astrand, M. & Speed, T. (2003) A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias. *Bioinformatics*, **19**, 185–193.

Bray, E. A. (2004) Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. *Journal of Experimental Botany*, **55**, 2331–2341.

Brewster, J., de Valoir, T., Dwyer, N., Winter, E. & Gustin, M. (1993) An osmosensing signal transduction pathway in yeast. *Science*, **259**, 1760–1763.

Briese, T., Rambaut, A. & Lipkin, W. I. (2004) Analysis of the medium (M) segment sequence of Guaroa virus and its comparison to other orthobunyaviruses. *Journal of General Virology*, **85**, 3071–3077.

Brinker, M., van, L., Zyl, Liu, W., Craig, D., Sederoff, R., Clapham, D. & von, S., Arnold (2004) Microarray Analyses of Gene Expression during Adventitious Root Development in *Pinus contorta*. *Plant Physiology*, **135**, 1526–1539.

Burnett, E. C., Desikan, R., Moser, R. C. & Neill, S. J. (2000) ABA activation of an MBP kinase in *Pisum sativum* epidermal peels correlates with stomatal responses to ABA. *Journal of Experimental Botany*, **51**, 197–205.

Byrne, K., Wang, Y., Lehnert, S., Harper, G., McWilliam, S., Bruce, H. & Reverter, A. (2005) Gene expression profiling of muscle tissue in Brahman steers during nutritional restriction. *Journal of Animal Science*, **83**, 1–12.

Chang, J.-S., Kim, S.-K., Kwon, T.-K., Bae, S., Min, D., Lee, Y., Kim, S.-O., Seo, J.-K., Choi, J. & Suh, P.-G. (2005) Pleckstrin Homology Domains of Phospholipase C-γ 1 Directly Interact with β-Tubulin for Activation of Phospholipase C-γ1 and Reciprocal Modulation of β-Tubulin Function in Microtubule Assembly. *Journal of Biological Chemistry*, **280**, 6897–6905.

Chernys, J. T. & Zeevaart, J. A. D. (2000) Characterization of the 9-Cis-Epoxycarotenoid Dioxygenase Gene Family and the Regulation of Abscisic Acid Biosynthesis in Avocado. *Plant Physiology*, **124**, 343–354.

Chinnusamy, V., Schumaker, K. & Zhu, J.-K. (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *Journal of Experimental Botany*, **55**, 225–236.

Chu, T.-M., Weir, B. & Wolfinger, R. (2002) A systematic statistical linear modeling approach to oligonucleotide array experiments. *Mathematical Biosciences*, **176**, 35–51.

Chu, T.-M., Weir, B. & Wolfinger, R. D. (2004) Comparison of Li-Wong and Loglinear Mixed Models for the Statistical Analysis of Oligonucleotide Arrays. *Bioinformatics*, **20**, 500–506.

Churchill, G. A. (2004) Using ANOVA to Analyze Microarray Data. *BioTechniques*, **37**, 173–177.

Cleveland, W. (1979) Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, **74**, 829–836.

Coursol, S., Fan, L.-M., Stunff, H. L., Spiegel, S., Gilroy, S. & Assmann, S. M. (2003) Sphingolipid signalling in Arabidopsis guard cells involves heterotrimeric G proteins. *Nature*, **423**, 651–654.

Cowan, I. (1981) *Regulation of water use in relation to carbon gain in higher plants*. Berlin: Springer.

Cui, X. & Churchill, G. A. (2003) Statistical Tests for Differential Expression in cDNA Microarray Experiments. *Genome Biology*, **4**.

Cushman, J. C. & Bohnert, H. J. (2000) Genomic approaches to plant stress tolerance. *Current Opinion in Plant Biology*, **3**, 117–124.

Davies, P. (1995) *Plant Hormones: Physiology, Biochemistry and Molecular Biology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Delon, C., Manifava, M., Wood, E., Thompson, D., Krugmann, S., Pyne, S. & Ktistakis, N. (2004) Sphingosine Kinase 1 Is an Intracellular Effector of Phosphatidic Acid. *Journal of Biological Chemistry*, **279**, 44763–44774.

DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997) Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science*, **278**, 680–686.

Desikan, R., Cheung, M.-K., Bright, J., Henson, D., Hancock, J. & Neill, S. (2004) ABA, hydrogen peroxide and nitric oxide signalling in stomatal guard cells. *Journal of Experimental Botany*, **55**, 205–212.

Devaiah, S., Roth, M., Baughman, E., Li, M., Tamura, P., Jeannotte, R., Welti, R. & Wang, X. (2006) Quantitative profiling of polar glycerolipid species from organs of wild-type Arabidopsis and a phospholipase D α 1 knockout mutant. *Phytochemistry*, **67**, 1907–1924.

Dombkowski, A., Thibodeau, B., Starcevic, S. & Novak, R. (2004) Gene-specific dye bias in microarray reference designs. *FEBS Letters*, **560**, 120–124.

Dudoit, S., Gentleman, R. C. & Quackenbush, J. (2003) Open Source Software for the Analysis of Microarray Data. *BioTechniques*, **34**, s45–s51.

Duggan, D., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. (1999) Expression profiling using cDNA microarrays. *Nature Genetics*, **21**, 10–14.

Eapen, D., Barroso, M., Ponce, G., Campos, M. & Cassab, G. (2005) Hydrotropism: root growth responses to water. *Trends in Plant Science*, **10**, 44–50.

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Eliáš, M., Potocký, M., Cvrčková, F. & Žárský, V. (2002) Molecular diversity of phospholipase D in angiosperms. *BMC Genomics*, **3**, 2.

Evans, N. & Hetherington, A. (2001) Plant physiology: the ups and downs of guard cell signalling. *Current Biology*, **11**, R92–R94.

Fan, L., Zheng, S. & Wang, X. (1997) Antisense Suppression of Phospholipase D α Retards Abscisic Acid and Ethylene-Promoted Senescence of Postharvest Arabidopsis Leaves. *The Plant Cell*, **9**, 2183–2196.

Fauci, A. S. (2003) Biodefence on the research agenda. *Nature*, **421**, 787–787.

Fowler, S. & Thomashow, M. F. (2002) Arabidopsis Transcriptome Profiling Indicates That Multiple Regulatory Pathways Are Activated during Cold Acclimation in Addition to the CBF Cold Response Pathway. *The Plant Cell*, **14**, 1675–1690.

Foyer, C. & Halliwell, B. (1976) The presence of glutathione and glutathione reductase in chloroplasts: a proposed role in ascorbic acid metabolism. *Planta*, **133**, 21–25.

Fujimori, T., Sato, E., Yamashino, T. & Mizuno, T. (2005) PRR5 (PSEUDO-RESPONSE REGULATOR 5) plays antagonistic roles to CCA1 (CIRCADIAN CLOCK-ASSOCIATED-1) in *Arabidopsis thaliana*. *Bioscience Biotechnology and Biochemistry*, **69**, 426–430.

Fujita, M., Fujita, Y., Maruyama, K., Seki, M., Hiratsu, K., Ohme-Takagi, M., Tran, L.-S. P., Yamaguchi-Shinozaki, K. & Shinozaki, K. (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *The Plant Journal*, **39**, 863–876.

Futschik, M. & Crompton, T. (2004) Model Selection and Efficiency Testing for Normalization of cDNA Microarray Data. *Genome Biology*, **5**.

Ghosh, S., Strum, J., Sciorra, V., Daniel, L. & Bell, R. (1996) Raf-1 Kinase Possesses Distinct Binding Domains for Phosphatidylserine and Phosphatidic Acid. *Journal of Biological Chemistry*, **271**, 8472–8480.

Gilroy, S., Fricker, M., Read, N. & Trewavas, A. (1991) Role of Calcium in Signal Transduction of Commelina Guard Cells. *The Plant Cell*, **3**, 333–344.

Gilroy, S., Read, N. & Trewavas, A. (1990) Elevation of cytoplasmic calcium by caged calcium or caged inositol triphosphate initiates stomatal closure. *Nature*, **346**, 769– 771.

Goff, S., Ricke, D., Lan, T., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W., Chen, L., Cooper, B., Park, S., Wood, T., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. & Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica) *Science*, **296**, 92–100.

Goyal, K., Walton, L. J. & Tunnacliffe, A. (2005) LEA proteins prevent protein aggregation due to water stress. *Biochemical Journal*, **388**, 151–157.

Grabov, A. & Blatt, M. (1998) Membrane voltage initiates Ca^{2+} waves and potentiates Ca^{2+} increases with abscisic acid in stomatal guard cells. *Proceedings of the National Academy of Sciences, USA*, **95**, 4778–4783.

Guo, M., Rupe, M. A., Yang, X., Crasta, O., Zinselmeier, C., Smith, O. S. & Bowen, B. (2006) Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. *Theoretical and Applied Genetics*, **113**, 831–845.

Haake, V., Cook, D., Riechmann, J., Pineda, O., Thomashow, M. & Zhang, J. (2002) Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. *Plant Physiology*, **130**, 639–648.

Hacia, J. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics*, **21**, 42–47.

Hasegawa, P. M., Bressan, R. A., Zhu, J.-K. & Bohnert, H. J. (2000) Plant Cellular and Molecular Responses to High Salinity. *Annual Review of Plant Physiology and Plant Molecular Biology*, **51**, 463–499.

Hayes, M. J., Svoboda, M. D., Knutson, C. L. & Wilhite, D. A. (2004) Estimating the Economic Impacts of Drought. *The 15th Symposium on Global Change and Climate Variations*.

Heath, L. S., Ramakrishnan, N., Sederoff, R. R., Whetten, R. W., Chevone, B. I., Struble, C. A., Jouenne, V. Y., Chen, D., van Zyl, L. M. & Grene, R. (2002) Studying the Functional Genomics of Stress Responses in Loblolly Pine using the Expresso Microarray Management System. *Comparative and Functional Genomics*, **3**, 226–243.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J. E., Snesrud, E., Lee, N. & Quackenbush, J. (2000) A concise guide to cDNA microarray analysis. *Biotechniques*, **29**, 548–550.

Held, M., Gase, K. & Baldwin, I. T. (2004) Microarray in Ecological Research: A Case Study of a cDNA Microarray for Plant-Herbivore Interactions. *BMC Ecology*, **4**.

Hetherington, A. (2001) Guard cell signaling. *Cell*, **107**, 711–714.

Hetherington, A. & Woodward, F. (2003) The role of stomata in sensing and driving environmental change. *Nature*, **424**, 901–908.

Hirayama, T., Ohto, C., Mizoguchi, T. & Shinozaki, K. (1995) A Gene Encoding a Phosphatidylinositol-Specific Phospholipase C is Induced by Dehydration and Salt Stress in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA*, **92**, 3903–3907.

Hsiao, T. C. (1973) Plant Responses to Water Stress. *Annual Review of Plant Physiology*, **24**, 519–570.

Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. (2002) Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression. *Bioinformatics*, **18**, S96–S104.

Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P. Y., Stepaniants, S. B., Shoemaker, D. D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M. & Friend, S. H. (2000) Functional Discovery via a Compendium of Expression Profiles. *Cell*, **102**, 109–126.

Hunt, L., Otterhag, L., Lee, J., Lasheen, T., Hunt, J., Seki, M., Shinozaki, K., Sommarin, M., Gilmour, D., Pical, C. & Gray, J. (2004) Gene-specific expression and calcium activation of *Arabidopsis thaliana* phospholipase C isoforms. *New Phytologist*, **162**, 643–654.

Ichimura, K., Mizoguchi, T., Yoshida, R., Yuasa, T. & Shinozaki, K. (2000) Various abiotic stresses rapidly activate Arabidopsis MAP kinases ATMPK4 and ATMPK6. *The Plant Journal*, **24**, 655–665.

Ingram, J. & Bartels, D. (1996) The Molecular Basis of Dehydration Tolerance in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, **47**, 377–403.

Iuchi, S., Kobayashi, M., Taji, T., Naramoto, M., Seki, M., Kato, T., Tabata, S., Kakubari, Y., Yamaguchi-Shinozaki, K. & Shinozaki, K. (2001) Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in Arabidopsis. *The Plant Journal*, **27**, 325–333.

Jabado, O. J., Palacios, G., Kapoor, V., Hui, J., Renwick, N., Zhai, J., Briese, T. & Lipkin, W. I. (2006) Greene SCPrimer: a rapid comprehensive tool for designing degenerate primers from multiple sequence alignments. *Nucleic Acids Research*, **34**, 6605–6611.

Jenkins, A., Bootman, M., Berridge, M. & Stone, S. (1994) Differences in intracellular calcium signaling after activation of the thrombin receptor by thrombin and agonist peptide in osteoblast-like cells. *Journal of Biological Chemistry*, **269**, 17104–17110.

Jenkins, G. M. & Frohman, M. A. (2005) Phospholipase D: a lipid centric review. *Cellular and Molecular Life Sciences*, **62**, 2305–2316.

Jia, W., Zhang, J. & Liang, J. (2001) Initiation and regulation of water deficit-induced abscisic acid accumulation in maize leaves and roots: cellular volume and water relations. *Journal of Experimental Botany*, **52**, 295–300.

Jones, H. (1992) *Plants and microclimate*. Cambridge: Cambridge University Press.

Kacperska, A. (2004) Sensor types in signal transduction pathways in plant cells responding to abiotic stressors: do they depend on stress intensity? *Physiologia Plantarum*, **122**, 159–168.

Katagiri, T., Mizoguchi, T. & Shinozaki, K. (1996) Molecular cloning of a cDNA encoding diacylglycerol kinase (DGK) in *Arabidopsis thaliana*. *Plant Molecular Biology*, **30**, 647–53.

Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D. & Bohnert, H. J. (2001) Gene Expression Profiles during the Initial Phase of Salt Stress in Rice. *The Plant Cell*, **13**, 889–906.

Kerr, M. K. (2003) Linear Models for Microarray Data Analysis: Hidden Similarities and Differences. *Journal of Computational Biology*, **10**, 891–901.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W. & Paabo, S. (2004) A Neutral Model of Transcriptome Evolution. *PLoS Biology*, **2**, 682–689.

Khan, J., Bittner, M., Saal, L., Teichmann, U., Azorsa, D., Gooden, G., Pavan, W., Trent, J. & Meltzer, P. (1999) cDNA microarrays detect activation of a myogenic transcription program by the PAX3-FKHR fusion oncogene. *Proceedings of the National Academy of Sciences, USA*, **96**, 13264–13269.

Kiba, T., Yamada, H., Sato, S., Kato, T., Tabata, S., Yamashino, T. & Mizuno, T. (2003) The Type-A Response Regulator, ARR15, Acts as a Negative Regulator in the Cytokinin-Mediated Signal Transduction in *Arabidopsis thaliana*. *Plant and Cell Physiology*, **44**, 868–874.

Kirst, M., Basten, C., Myburg, A., Zeng, Z.-B. & Sederoff, R. (2005) Genetic Architecture of Transcript-Level Variation in Differentiating Xylem of a Eucalyptus Hybrid. *Genetics*, **169**, 2295–2303.

Ktistakis, N., Delon, C., Manifava, M., Wood, E., Ganley, I. & Sugars, J. (2003) Phospholipase D1 and potential targets of its hydrolysis product, phosphatidic acid. *Biochemical Society Transactions*, **31**, 94–97.

Kurkela, S. & Borg-Franck, M. (199) Structure and expression of *Kin2*, one of two cold- and ABA-induced genes of *Arabidopsis thaliana*. *Plant Molecular Biology*, **19**, 689–692.

Lang, V. & Palva, E. (1992) The expression of rab-related gene, RAB18, is induced by abscisic acid during the cold acclimation process of *Arabidopsis thaliana* (L.) Heynh. *Plant Molecular Biology*, **20**, 951–962.

Lecourieux, D., Ranjeva, R. & Pugin, A. (2006) Calcium in plant defence-signalling pathways. *New Phytologist*, **171**, 249–269.

Lee, B.-h., Henderson, D. & J-K., Z. (2005) The Arabidopsis Cold-Responsive Transcriptome and Its Regulation by ICE1. *The Plant Cell*, **17**, 3155–3175.

Lee, S., Park, J. & Lee, Y. (2003) Phosphatidic acid induces actin polymerization by activating protein kinases in soybean cells. *Molecules and Cells*, **15**, 313–319.

Lee, T. I., Johnstone, S. E. & Young, R. A. (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nature Protocols*, **1**, 729–748.

Lee, Y., Coi, Y., Suh S, J., Lee, Assmann, S., Joe, C., Kelleher, J. & Crain, R. (1996) Abscisic acid-induced phosphoinositide turnover in guard cell protoplasts of *Vicia faba*. *Plant Physiology*, **110**, 987–996.

Lemtiri-Chlieh, F. & MacRobbie, E. (1994) Role of calcium in the modulation of Vicia guard cell potassium channels by abscisic acid: a patch-clamp study. *Journal of Membrane Biology*, **137**, 99–107.

Lemtiri-Chlieh, F., MacRobbie, E. & Brearley, C. (2000) Inositol hexakisphosphate is a physiological signal regulating the K^+-inward rectifying conductance in guard cells. *Proceedings of the National Academy of Sciences, USA*, **97**, 8687–8692.

Leung, J., Bouvier-Durand, M., Morris, P., Guerrier, D., Chefdor, F. & Giraudat, J. (1994) Arabidopsis ABA response gene ABI1: features of a calcium-modulated protein phosphatase. *Science*, **264**, 1448–1452.

Li, P., Mane, S. P., Sioson, A. A., Vasquez-Robinet, C., Heath, L. S., Bohnert, H. J. & Grene, R. (2006) Effects of Chronic Ozone Exposure on Gene Expression in *Arabidospsis thaliana* Ecotypes and in *Thellungiella halophila*. *Plant, Cell and Environment*, **29**, 854–868.

Li, W., Li, M., Zhang, W., Welti, R. & Wang, X. (2004) The plasma membrane-bound phospholipase D δ enhances freezing tolerance in *Arabidopsis thaliana*. *Nature Biotechnology*, **22**, 427–433.

Li, W., Welti, R. & Wang, X. (2005) The different roles of phospholipase D α 1 and δ in water-deficit stress signaling in *Arabidopsis thaliana*. *International Conference on Plant Lipid-Mediated Signaling: Building Connections.*, pp. 7–8. Raleigh, NC.

Liu, Y. & Zhang, S. (2004) Phosphorylation of 1-Aminocyclopropane-1-Carboxylic Acid Synthase by MPK6, a Stress-Responsive Mitogen-Activated Protein Kinase, Induces Ethylene Biosynthesis in Arabidopsis. *The Plant Cell*, **16**, 3386–3399.

Lockhart, D., Dong, H., Byrne, M., Follettie, M., Gallo, M., Chee, M., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. & Brown, E. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology*, **14**, 1675–1680.

Lopez-Andreo, M., Gomez-Fernandez, J. & Corbalan-Garcia, S. (2003) The Simultaneous Production of Phosphatidic Acid and Diacylglycerol Is Essential for the Translocation of Protein Kinase C epsilon to the Plasma Membrane in RBL-2H3 Cells. *Molecular Biology of the Cell*, **14**, 4885–4895.

Malamy, J. (2005) Intrinsic and environmental response pathways that regulate root system architecture. *Plant Cell and Environment*, **28**, 67–77.

Mane, S. P., Vasquez-Robinet, C., Sioson, A. A., Heath, L. S. & Grene, R. (2007) Early PLDα-Mediated Events in Response to Progressive Drought Stress in Arabidopsis: a Transcriptome Analysis. *Journal of Experimental Botany*, **58**, 241–252.

Manifava, M., Thuring, J., Lim, Z.-Y., Packman, L., Holmes, A. & Ktistakis, N. (2001) Differential Binding of Traffic-related Proteins to Phosphatidic Acid- or Phosphatidylinositol (4,5)-Bisphosphate-coupled Affinity Reagents. *Journal of Biological Chemistry*, **276**, 8987–8994.

McAinsh, M., Brownlee, C. & Hetherington, A. (1992) Visualizing Changes in Cytosolic-Free Ca^{2+} during the Response of Stomatal Guard Cells to Abscisic Acid. *The Plant Cell*, **4**, 1113–1122.

McAinsh, M., Gray, J., Hetherington, A., Leckie, C. & Ng, C. (2000) Ca^{2+} signalling in stomatal guard cells. *Biochemical Society Transactions*, **28**, 476–481.

McAinsh, M., Webb, A., Taylor, J. & Hetherington, A. (1995) Stimulus-Induced Oscillations in Guard Cell Cytosolic Free Calcium. *The Plant Cell*, **7**, 1207–1219.

McCue, K. & Hanson, A. (1990) Drought and salt tolerance : towards understanding and application. *Trends in Biotechnology*, **8**, 358–362.

Mikami, K., Katagiri, T., Iuchi, S., Yamaguchi-Shinozaki, K. & Shinozaki, K. (1998) A gene encoding phosphatidylinositol-4-phosphate 5-kinase is induced by water stress and abscisic acid in *Arabidopsis thaliana*. *The Plant Journal*, **15**, 563–568.

Mishra, G., Zhang, W., Deng, F., Zhao, J. & Wang, X. (2006) A Bifurcating Pathway Directs Abscisic Acid Effects on Stomatal Closure and Opening in Arabidopsis. *Science*, **312**, 264–266.

Mizoguchi, T., Irie, K., Hirayama, T., Hayashida, N., Yamaguchi-Shinozaki, K., Matsumoto, K. & K., S. (1996) A gene encoding a mitogen-activated protein kinase kinase kinase is induced simultaneously with genes for a mitogen-activated protein kinase and an S6 ribosomal protein kinase by touch, cold, and water stress in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA*, **93**, 765–9.

Mizoguchi, T., Yoshida, R., Yuasa, T. & Shinozaki, K. (2000) Various abiotic stresses rapidly activate Arabidopsis MAP kinases ATMPK4 and ATMPK6. *The Plant Journal*, **24**, 655–665.

Moës, E., Vijgen, L., Keyaerts, E., Zlateva, K., Li, S., Maes, P., Pyrc, K., Berkhout, B., van der Hoek, L. & Ranst, M. V. (2005) A novel pancoronavirus RT-PCR assay: frequent detection of human coronavirus NL63 in children hospitalized with respiratory tract infections in Belgium. *BMC Infectious Diseases*, **5**, 6.

Morgan, J. M. (1984) Osmoregulation and Water Stress in Higher Plants. *Annual Review of Plant Physiology*, **35**, 299–319.

Morgenstern, B. (2004) DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Research*, **32**, W33–36.

Munnik, T., Meijer, H. J. G., ter Riet, B., Hirt, H., Frank, W., Bartels, D. & Musgrave, A. (2000) Hyperosmotic stress stimulates phospholipase D activity and elevates the levels of phosphatidic acid and diacylglycerol pyrophosphate. *The Plant Journal*, **22**, 147–154.

Nambara, E. & Marion-Poll, A. (2005) Abscisic acid biosynthesis and catabolism. *Annual Review of Plant Biology*, **56**, 165–185.

Nilson, S. E. & Assmann, S. M. (2007) The Control of Transpiration. Insights from Arabidopsis. *Plant Physiology*, **143**, 19–27.

Ohashi, Y., Oka, A., Rodrigues-Pousada, R., Possenti, M., Ruberti, I., Morelli, G. & Aoyama, T. (2003) Modulation of Phospholipid Signaling by GLABRA2 in Root-Hair Pattern Formation. *Science*, **300**, 1427–1430.

Pan, W. (2002) A Comparative Review of Statistical Methods for Discovering Differentially Expressed Genes in Replicated Microarray Experiments. *Bioinformatics*, **18**, 546–554.

Pappan, K., Zheng, S. & Wang, X. (1997) Identification and characterization of a novel phospholipase D that requires polyphosphoinositide and submicromolar calcium for activity in Arabidopsis. *Journal of Biological Chemistry*, **272**, 7048–7054.

Pei, Z., Ghassemian, M., Kwak, P., C. McCourt & Schroeder, J. (1998) Role of farnesyltransferase in ABA regulation of guard cell anion channels and plant water loss. *Science*, **282**, 252–253.

Pei, Z. & Kuchitsu, K. (2005) Early ABA Signaling Events in Guard Cells. *Journal of Plant Growth Regulation*, **24**, 296–307.

Pinheiro, H. A., DaMatta, F. M., Chaves, A. R. M., Loureiro, M. E. & Ducatti, C. (2005) Drought Tolerance is Associated with Rooting Depth and Stomatal Control of Water Use in Clones of *Coffea canephora*. *Annals of Botany*, **96**, 101–108.

Potocký, M., Eliáš, M., Profotová, B., Novotná, Z., Valentová, O. & Žárský, V. (2003) Phosphatidic acid produced by phospholipase D is required for tobacco pollen tube growth. *Planta*, **217**, 122–130.

Prokic, L., Jovanovic, Z., McAinsh, M. R., Vucinic, Z. & Stikic, R. (2006) Species-dependent changes in stomatal sensitivity to abscisic acid mediated by external pH. *Journal of Experimental Botany*, **57**, 675–683.

Qin, C., Wang, C. & Wang, X. (2002) Kinetic analysis of Arabidopsis phospholipase D δ: Substrate preference and mechanism of activation by Ca^{2+} and phosphatidylinositol 4, 5-bisphosphate. *Journal of Biological Chemistry*, **277**, 49685–49690.

Qin, W., Pappan, K. & Wang, X. (1997) Molecular Heterogeneity of Phospholipase D (PLD) Cloning of PLDγ and Regulation of Plant PLDγ , -β , and -α by Polyphosphoinositides and Calcium. *Journal of Biological Chemistry*, **272**, 28267–28273.

Quackenbush, J. (2001) Computational Analysis of Microarray Data. *Nature Reviews Genetics*, **2**, 418–427.

Quackenbush, J. (2002) Microarray Data Normalization and Transformation. *Nature Genetics Supplement*, **32**, 496–501.

Rabbani, M., Maruyama, K., Abe, H., Khan, M., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2003) Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses. *Plant Physiology*, **133**, 1755–1767.

Razem, F. A., El-Kereamy, A., Abrams, S. R. & Hill, R. D. (2006) The RNA-binding protein FCA is an abscisic acid receptor. *Nature*, **439**, 290–294.

Reiser, V., Raitt, D. C. & Saito, H. (2003) Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure. *The Journal of Cell Biology*, **161**, 1035–1040.

Rentel, M., Lecourieux, D., Ouaked, F., Usher, S., Petersen, L., Okamoto, H., Knight, H., Peck, S., Grierson, C., Hirt, H. & Knight, M. (2004) OXI1 kinase is necessary for oxidative burst-mediated signalling in Arabidopsis. *Nature*, **427**, 858–861.

Reverter, A., McWilliam, S. M., Barris, W. & Dalrymple, B. P. (2005) A rapid method for computationally inferring transcriptome coverage and microarray sensitivity. *Bioinformatics*, **21**, 80–89.

Ritchie, S. & Gilroy, S. (2000) Abscisic Acid Stimulation of Phospholipase D in the Barley Aleurone Is G-Protein-Mediated and Localized to the Plasma Membrane. *Plant Physiology*, **124**, 693–702.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S. & Mittler, R. (2004) When Defense Pathways Collide. The Response of Arabidopsis to a Combination of Drought and Heat Stress. *Plant Physiology*, **134**, 1683–1696.

Rontein, D., Basset, G. & Hanson, A. (2002) Metabolic Engineering of Osmoprotectant Accumulation in Plants. *Metabolic Engineering*, **4**, 49–56.

Rosa, G. J., Steibel, J. P. & Tempelman, R. J. (2005) Reassessing Design and Analysis of Two-colour Microarray Experiments using Mixed Effects Models. *Comparative and Functional Genomics*, **6**, 123–131.

Rose, T., Henikoff, J. & Henikoff, S. (2003) CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Research*, **31**, 3763–3766.

Saeed, A., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. & Quackenbush, J. (2003) TM4: A Free, Open-Source System for Microarray Data Management and Analysis. *BioTechniques*, **34**, 374–378.

Sakamoto, H., Maruyama, K., Sakuma, Y., Meshi, T., Iwabuchi, M., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2004) Arabidopsis Cys2/His2-Type Zinc-Finger Proteins Function as Transcription Repressors under Drought, Cold, and High-Salinity Stress Conditions. *Plant Physiology*, **136**, 2734–2746.

Salisbury, F. B. & Ross, C. W. (1992) *Plant Physiology*, pp. 357, 531. Wadsworth Pub. Co., Belmont, Calif.

Sang, Y., Cui, D. & Wang, X. (2001a) Phospholipase D and Phosphatidic Acid-Mediated Generation of Superoxide in Arabidopsis. *Plant Physiology*, **126**, 1449–1458.

Sang, Y., Zheng, S., Li, W., Huang, B. & Wang, X. (2001b) Regulation of plant water loss by manipulating the expression of phospholipase D α. *The Plant Journal*, **28**, 135–144.

Sauter, A., Davies, W. & Hartung, W. (2001) The long-distance abscisic acid signal in the droughted plant: the fate of the hormone on its way from root to shoot. *Journal of Experimental Botany*, **52**, 1991–1997.

Schena, M., Shalon, D., Davis, R. & Brown, P. (1995) Quantitative monitoring of gene- expression patterns with a complementary-DNA microarray. *Science*, **270**, 467–470.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. & Davis, R. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences, USA*, **93**, 10614–10619.

Schrick, K., Nguyen, D., Karlowski, W. & Mayer, K. (2004) START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biology*, **5**, R41.

Schroeder, J., Kwak, J. & Allen, G. (2001) Guard cell abscisic acid signalling and engineering drought hardiness in plants. *Nature*, **410**, 327–330.

Schuler, G. D. (1997) Sequence Mapping by Electronic PCR. *Genome Research*, **7**, 541–550.

Schwartz, S. H., Qin, X. & Zeevaart, J. A. D. (2003) Elucidation of the Indirect Pathway of Abscisic Acid Biosynthesis by Mutants, Genes, and Enzymes. *Plant Physiology*, **131**, 1591–1601.

Seki, M., Kamei, A., Yamaguchi-Shinozaki, K. & Shinozaki, K. (2003) Molecular responses to drought, salinity and frost: common and different paths for plant protection. *Current Opinion in Biotechnology*, **14**, 194–199.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. & Shinozaki, K. (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant Journal*, **31**, 279–292.

Shalon, D., Smith, S. & Brown, P. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research*, **6**, 639–645.

Shen, Y.-Y., Wang, X.-F., Wu, F.-Q., Du, S.-Y., Cao, Z., Shang, Y., Wang, X.-L., Peng, C.-C., Yu, X.-C., Zhu, S.-Y., Fan, R.-C., Xu, Y.-H. & Zhang, D.-P. (2006) The Mg-chelatase H subunit is an abscisic acid receptor. *Nature*, **443**, 823–826.

Shinozaki, K. & Yamaguchi-Shinozaki, K. (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Current Opinion in Plant Biology*, **3**, 217–223.

Shinozaki, K., Yamaguchi-Shinozaki, K. & Seki, M. (2003) Regulatory network of gene expression in the drought and cold stress responses. *Current Opinion in Plant Biology*, **6**, 410–7.

Sioson, A. A., Mane, S. P., Li, P., Sha, W., Heath, L. S., Bohnert, H. J. & Grene, R. (2006) The statistics of identifying significantly expressed genes in Expresso and TM4: a comparison. *BMC Bioinformatics*, **7**, 215.

Sioson, A. A., Watkinson, J. I., Vasquez-Robinet, C., Ellis, M., Shukla, M., Kumar, D., Ramakrishnan, N., Heath, L. S., Grene, R., Chevone, B. I., Kadafar, K. & Watson, L. T. (2003) Expresso and Chips: Creating a Next Generation Microarray Experiment Management System. *Proceedings of the Next Generation Software Systems Workshop, 17th International Parallel and Distributed Processing Symposium (IPDPS '03)*, p. 209b. Nice, France.

Slavik, P. (1997) A tight analysis of the greedy algorithm for set cover. *Journal of Algorithms*, **25**.

Slezak, T., Kuczmarski, T., Ott, L., Torres, C., Medeiros, D., Smith, J., Truitt, B., Mulakken, N., Lam, M., Vitalis, E., Zemla, A., Zhou, C. E. & Gardner, S. (2003) Comparative genomics tools applied to bioterrorism defence. *Briefings in Bioinformatics*, **4**, 133–149.

Southern, E. (1975) Detection of specific sequences among DNA fragments separated by gel-electrophoresis. *Journal of Molecular Biology*, **98**, 503–517.

Staxen, I., Pical C, L., Montgomery, Gray, J., Hetherington, A. & McAinsh, M. (1999) Abscisic acid induces oscillations in guard-cell cytosolic free calcium that involve phosphoinositide-specific phospholipase C. *Proceedings of the National Academy of Sciences, USA*, **96**, 1779–1784.

Tamas, M., Rep, M., Thevelein, J. & S., H. (2000) Stimulation of the yeast high osmolarity glycerol (HOG) pathway: evidence for a signal generated by a change in turgor rather than by water stress. *FEBS Letters*, **472**, 159–165.

Tan, B.-C., Joseph, L. M., Deng, W.-T., Liu, L., Li, Q.-B., Cline, K. & McCarty, D. R. (2003) Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. *The Plant Journal*, **35**, 44–56.

Tenhunen, J., Pearcy, R. & Lange, O. (1987) *Diurnal variations in leaf conductance and gas exchange in natural environments*. Stanford: Stanford University Press.

Testerink, C. & Munnik, T. (2005) Phosphatidic acid: a multifunctional stress signaling lipid in plants. *Trends in Plant Science*, **10**, 368–375.

Testernik, C., Dekker, H., Lim, Z.-Y., Johns, M., Holmes, A., Koster, C., Ktistakis, N. & Munnik, T. (2004) Isolation and identification of phosphatidic acid targets from plants. *The Plant Journal*, **39**, 527–536.

Thomashow, M. (1998) Role of cold-responsive genes in plant freezing tolerance. *Plant Physiology*, **118**, 1–7.

Thompson, A. J., Jackson, A. C., Symonds, R. C., Mulholland, B. J., Dadswell, A. R., Blake, P. S., Burbidge, A. & Taylor, I. B. (2000) Ectopic expression of a tomato 9-cis-epoxycarotenoid dioxygenase gene causes over-production of abscisic acid. *The Plant Journal*, **23**, 363–374.

Toorchi, M., Shashidhar, H., Sharma, N. & Hittalmani, S. (2004) Tagging QTLs for maximum root length in rainfed lowland rice by combined selective genotyping and STMs markers. Proceedings for the 4th International Crop Science Congress, Brisbane, Australia, 26 September – 1 October 2004.

Tran, L.-S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2004) Isolation and Functional Analysis of Arabidopsis Stress-Inducible NAC Transcription Factors That Bind to a Drought-Responsive cis-Element in the early responsive to dehydration stress 1 Promoter. *The Plant Cell*, **16**, 2481–2498.

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P. M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Déjardin, A., dePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D. R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V.,

Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., de Peer, Y. V. & Rokhsar, D. (2006) The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray) *Science*, **313**, 1596–1604.

Urao, T., Yakubov, B., Satoh, R., Yamaguchi-Shinozaki, K., Seki, M., Hirayama, T. & Shinozaki, K. (1999) A Transmembrane Hybrid-Type Histidine Kinase in Arabidopsis Functions as an Osmosensor. *The Plant Cell*, **11**, 1743–1754.

Urao, T., Yamaguchi-Shinozaki, K., Urao, S. & Shinozaki, K. (1993) An Arabidopsis MYB homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequences. *The Plant Cell*, **5**, 1529–1539.

Verslues, P. & Zhu, J. (2005) Before and beyond ABA: upstream sensing and internal signals that determine ABA accumulation and response under abiotic stress. *Biochemical Society Transactions*, **33**, 375–379.

Wang, C., Zien, C., Afitlhile, M., Welti, R., Hildebrand, D. & Wang, X. (2000) Involvement of Phospholipase D in Wound-Induced Accumulation of Jasmonic Acid in Arabidopsis. *The Plant Cell*, **12**, 2237–2246.

Wang, D., Fan, J., Siao, C., Berno, A., Young, P., Sapolsky, R., Gandour, G., Perkins, N., Winchester, E. & Lander, E. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280**, 1077–1082.

Wang, H. & Cutler, A. (1995) Promoters from *kin1* and *cor6.6*, two *Arabidopsis thaliana* low-temperature and ABA-inducible genes, direct strong β-glucuronidase expression in guard cells, pollen and young developing seeds. *Plant Molecular Biology*, **28**, 619–634.

Wang, X. (2000) Multiple forms of phospholipase D in plants: the gene family, catalytic and regulatory properties, and cellular functions. *Progress in Lipid Research*, **39**, 109–149.

Wang, X. (2002) Phospholipase D in hormonal and stress signaling. *Current Opinion in Plant Biology*, **5**, 408–414.

Wang, X. (2005) Regulatory Functions of Phospholipase D and Phosphatidic Acid in Plant Growth, Development, and Stress Responses. *Plant Physiology*, **139**, 566–573.

Wang, X., Devaiah, S., Zhang, W. & Welti, R. (2006) Signaling functions of phosphatidic acid. *Progress in Lipid Research*, **45**, 250–278.

Watkinson, J. I., Sioson, A. A., Vasquez-Robinet, C., Shukla, M., Kumar, D., Ellis, M., Heath, L. S., Ramakrishnan, N., Chevone, B., Watson, L. T., van Zyl, L., Egertsdotter, U., Sederoff, R. R. & Grene, R. (2003) Photosynthetic Acclimation is Reflected in Specific Patterns of Gene Expression in Drought-Stressed Loblolly Pine. *Plant Physiology*, **133**, 1702–1716.

Webb, A., Larman, M., Montgomery, L., Taylor, J. & Hetherington, A. (2001) The role of calcium in ABA-induced gene expression and stomatal movements. *The Plant Journal*, **26**, 351–362.

Welti, R., Li, W., Li, M., Sang, Y., Biesiada, H., Zhou, H.-E., Rajashekar, C., Williams, T. & Wang, X. (2002) Profiling Membrane Lipids in Plant Stress Responses. Role of Phospholipase D α in freezing-induced lipid changes in Arabidopsis. *Journal of Biological Chemistry*, **277**, 31994–32002.

Wilhelm, K. & Thomashow, M. (1993) *Arabidopsis thaliana cor l5b*, an apparent homologue of *cor l5a*, is strongly responsive to cold and ABA, but not drought. *Plant Molecular Biology*, **23**, 1073–1077.

Wilhite, D. A. (2000) Drought as a natural hazard: concepts and definitions. *Drought: a global assessment* (ed. D. Wilhite), *Volume 1*, pp. 1–18. Routledge, New York.

Williams, R. D., Hing, S. N., Greer, B. T., Whiteford, C. C., Wei, J. S., Natrajan, R., Kelsey, A., Rogers, S., Campbell, C., Pritchard-Jones, K. & Khan, J. (2004) Prognostic Classification of Relapsing Favorable Histology Wilms Tumor using cDNA Microarray Expression Profiling and Support Vector Machines. *Genes, Chromosomes and Cancer*, **41**, 65–79.

Wolfinger, R., Gibson, G., Wolfinger, E., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C. & Paules, R. (2001) Assessing Gene Significance from cDNA Microarray Expression Data via Mixed Models. *Journal of Computational Biology*, **8**, 625–637.

Wood, J. M. (1999) Osmosensing by Bacteria: Signals and Membrane-Based Sensors. *Microbiology and Molecular Biology Reviews*, **63**, 230–262.

Xie, Y., Jeong, K. S., Pan, W., Khodursky, A. & Carlin, B. P. (2004) A Case Study on Choosing Normalization Methods and Test Statistics for Two-Channel Microarray Data. *Comparative and Functional Genomics*, **5**, 432–444.

Yamaguchi-Shinozaki, K. & Shinozaki, K. (1994) A novel *cis*-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *The Plant Cell*, **6**, 251–264.

Yamaguchi-Shinozaki, K. & Shinozaki, K. (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annual Review of Plant Biology*, **57**, 781–803.

Yan, J., Wang, J. & Zhang, H. (2002) An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *The Plant Journal*, **29**, 193–202.

Yang, I., Chen, E., Hasseman, J., Liang, W., Frank, B., Wang, S., Sharov, V., Saeed, A., White, J., Li, J., Lee, N., Yeatman, T. & Quackenbush, J. (2002a) Within the Fold: Assessing Differential Expression Measures and Reproducibility in Microarray Assays. *Genome Biology*, **3**, 1–12.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J. & Speed, T. P. (2002b) Normalization for cDNA Microarray Data: A Robust Composite Method Addressing Single and Multiple Slide Systematic Variation. *Nucleic Acids Research*, **30**, e15.

Yang, Y. H., Dudoit, S. & Luu, T. P., P.and Speed (2001) Normalization for cDNA microarray data. *Microarrays: Optical Technologies and Informatics* (eds. M. L. Bittner, Y. Chen, A. N. Dorsel & E. R. Dougherty), *Volume 4266*. *Proceedings of SPIE*, pp. 141–152.

Young, J. J., Mehta, S., Israelsson, M., Godoski, J., Grill, E. & Schroeder, J. I. (2006) CO_2 signaling in guard cells: Calcium sensitivity response modulation, a Ca^{2+}-independent phase, and CO_2 insensitivity of the *gca2* mutant. *Proceedings of the National Academy of Sciences, USA*, **103**, 7506 – 7511.

Yu, J., Hu, S., Wang, J., Wong, G., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X.,

Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L. & Yang, H. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica) *Science*, **296**, 79–92.

Yue, B., Xiong, L., Xue, W., Xing, Y., Luo, L. & Xu, C. (2005) Genetic analysis for drought resistance of rice at reproductive stage in field with different types of soil. *Theoretical and Applied Genetics*, **111**, 1127–1136.

Zhang, J., Creelman, R. & Zhu, J. (2004a) From laboratory to field. Using information from Arabidopsis to engineer salt, cold, and drought tolerance in crops. *Plant Physiology*, **135**, 615–621.

Zhang, W., Qin, C., Zhao, J. & Wang, X. (2004b) Phospholipase Dα1-derived phosphatidic acid interacts with ABI1 phosphatase 2C and regulates abscisic acid signaling. *Proceedings of the National Academy of Sciences, USA*, **101**, 9508–9513.

Zhang, W., Ruan, J., Ho, T., You, Y., Yu, T. & Quatrano, R. (2005) Cis-regulatory element based targeted gene finding: genome-wide identification of abscisic acid- and abiotic stress-responsive genes in *Arabidopsis thaliana*. *Bioinformatics*, **21**, 3074–3081.

Zhang, W., Wang, C., Qin, C., Wood, T., Olafsdottir, G., Welti, R. & Wang, X. (2003) The Oleate-Stimulated Phospholipase D, PLD δ, and Phosphatidic Acid Decrease H_2O_2-Induced Cell Death in Arabidopsis. *The Plant Cell*, **15**, 2285–2295.

Zhang, X., Fowler, S., Cheng, H., Lou, Y., Rhee, E., S.Y. abd Stockinger & Thomashow, M. (2004c) Freezing-sensitive tomato has a functional CBF cold response pathway, but a CBF regulon that differs from that of freezing-tolerant Arabidopsis. *The Plant Journal*, **39**, 905–9919.

Zhao, J. & Wang, X. (2003) Arabidopsis phospholipase Da1 interacts with the heterotrimeric G-protein a subunit through a motif analogous to the DRY motif in G-protein-coupled receptors. *Journal of Biological Chemistry*, **279**, 1794–1800.

Zhu, J.-K. (2002) Salt and Drought Stress Signal Transduction in Plants. *Annual Review of Plant Biology*, **53**, 247–273.

Zhu, J.-K., Hasegawa, P. & Bressan, R. (1997) Molecular aspects of osmotic stress in plants. *Critical Reviews in Plant Sciences*, **16**, 253–277.

Zhu, X., Hart, R., Chang, M. S., Kim, J.-W., Lee, S. Y., Cao, Y. A., Mock, D., Ke, E., Saunders, B., Alexander, A., Grossoehme, J., Lin, K.-M., Yan, Z., Hsueh, R., Lee, J., Scheuermann, R. H., Fruman, D. A., Seaman, W., Subramaniam, S., Sternweis, P., Simon, M. I. & Choi, S. (2004) Analysis of the Major Patterns of B Cell Gene Expression Changes in Response to Short-Term Stimulation with 33 Single Ligands. *The Journal of Immunology*, **173**, 7141–7149.

Appendix A

List of all differentially expressed genes in Anti-PLDα stress vs Col-0 stress

Table A.1: Significantly regulated genes in Anti-PLDα stress vs Col-0 stress. Changes in transcript profile are shown as positive (+) or negative (-)

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G01090	0.865	+	0.034	0	pyruvate dehydrogenase E1 alph
AT1G03350	-0.328	-	0.486	+	BSD domain-containing protein
AT1G03590	0.523	+	-0.125	0	protein phosphatase 2C (PP2C)
AT1G03790	0.294	+	-0.029	0	zinc finger (CCCH-type) family
AT1G03960	1.572	+	-0.609	-	protein phosphatase 2A -relate
AT1G04020	-0.307	-	0.156	+	zinc finger (C3HC4-type RING f
AT1G04050	-0.751	-	-0.659	0	SET-domain transcriptional reg
AT1G04240	-0.013	0	1.090	+	putative auxin-induced protein
AT1G04280	0.089	+	0.110	+	expressed protein
AT1G04330	2.127	+	-1.173	-	expressed protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G05010	-0.728	0	-1.495	-	1-aminocyclopropane-1-carboxyl
AT1G05230	0.433	+	0.018	0	homeodomain protein
AT1G05480	0.197	0	-0.667	-	SNF2domain/helicase domain-con
AT1G05520	0.037	0	0.591	+	transport protein, putative
AT1G05620	-0.209	-	0.013	0	inosine-uridine preferring nuc
AT1G05680	-0.582	-	-0.130	0	putative indole-3-acetate beta
AT1G06570	1.854	+	-0.244	0	4-hydroxyphenylpyruvate dioxyg
AT1G06950	0.339	+	0.018	0	chloroplast inner envelope pro
AT1G06960	0.051	0	-0.619	-	small nuclear ribonucleoprotei
AT1G07070	-0.586	-	-0.401	-	ribosomal protein, putative
AT1G07540	-0.270	0	-0.616	-	DNA-binding protein, putative
AT1G07705	0.207	+	0.079	0	NOT2/NOT3/NOT5 family protein
AT1G08080	0.100	0	0.542	+	storage protein, putative
AT1G08170	0.575	+	0.465	+	histone H2B, putative
AT1G08200	-0.551	-	0.687	+	dihydroflavonol reductase -rel
AT1G08360	-1.377	-	-0.803	0	60S ribosomal protein L10A, pu
AT1G09060	1.316	+	-0.169	0	transcription factor jumonji (
AT1G09130	1.629	+	-0.253	0	ATP-dependent Clp protease pro
AT1G09140	0.303	+	0.279	+	arginine/serine-rich splicing
AT1G09180	-0.293	-	-0.890	-	putative GTP-binding protein,
AT1G09330	0.046	+	0.139	+	expressed protein
AT1G09790	-0.096	0	0.808	+	putative phytochelatin synthet
AT1G10220	-0.061	0	-0.566	-	hypothetical protein
AT1G10630	-1.074	-	-0.979	-	ADP-ribosylation factor, putat
AT1G10990	1.078	+	-0.405	-	hypothetical protein
AT1G11210	0.127	0	0.909	+	expressed protein
AT1G11460	-0.658	-	0.677	+	nodulin MtN21 family protein
AT1G11520	1.480	+	0.255	0	hypthetical protein
AT1G12970	-0.415	-	-0.237	-	leucine rich repeat protein-re

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G13180	1.663	+	-0.834	0	actin-like protein
AT1G13440	-1.004	-	0.118	0	glyceraldehyde-3-phosphate deh
AT1G13800	-0.528	-	-0.295	-	pentatricopeptide (PPR) repeat
AT1G14140	0.446	+	-0.711	-	mitochondrial carrier protein
AT1G14920	-0.438	-	0.120	0	signal response protein (GAI)
AT1G14930	-0.904	-	-0.309	0	major latex protein (MLP)-rela
AT1G14950	-0.645	-	-0.073	0	major latex protein (MLP)-rela
AT1G14970	0.826	+	-0.024	0	auxin-independent growth promo
AT1G15010	0.898	+	0.250	0	expressed protein
AT1G15030	-0.034	0	0.566	+	expressed protein
AT1G15050	0.617	+	-0.003	0	putative glycerol kinase
AT1G15730	1.644	+	0.164	0	PRLI-interacting factor L, put
AT1G15930	-0.569	-	-0.916	-	40S ribosomal protein S12, put
AT1G16340	-0.239	0	0.463	+	putative 3-deoxy-D-manno-2-oct
AT1G16400	0.351	+	0.653	+	cytochrome p450 family
AT1G16510	0.501	+	-1.192	-	putative auxin-induced protein
AT1G16630	-0.590	0	1.584	+	expressed protein
AT1G16860	-0.297	-	-0.115	-	merozoite surface protein-rela
AT1G17170	1.047	+	-0.210	-	glutathione transferase, putat
AT1G17500	0.874	+	-0.275	-	P-type ATPase, putative
AT1G17745	0.560	+	-0.230	0	D-3-phosphoglycerate dehydroge
AT1G18180	0.285	0	-0.736	-	hypothetical protein
AT1G18260	-0.338	-	0.061	0	suppressor of lin-12-like prot
AT1G18270	-0.443	-	-0.559	-	ketose-bisphosphate aldolase c
AT1G18330	1.025	+	-0.766	-	myb family transcription facto
AT1G18630	0.403	+	0.050	0	glycine-rich RNA-binding prote
AT1G18660	-0.685	-	0.018	0	RING finger protein family
AT1G18690	-0.204	-	0.299	+	alpha galactosyltransferase, p
AT1G19570	-1.438	-	0.259	0	dehydroascorbate reductase, pu

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G19710	-0.003	0	-0.436	-	glycosyltransferase family 1
AT1G20060	2.252	+	-1.182	-	kinesin-related protein
AT1G20340	-0.722	-	0.069	0	plastocyanin
AT1G20370	-0.111	-	-0.103	-	tRNA pseudouridine synthase fa
AT1G20540	-0.708	-	-0.163	0	transducin / WD-40 repeat prot
AT1G20860	1.713	+	-0.659	-	putative inorganic phosphate t
AT1G21360	-0.107	-	0.240	+	unknown protein
AT1G21630	1.499	+	-0.162	-	calcium-binding EF-hand family
AT1G21920	0.576	+	0.240	+	phosphatidylinositol-4-phospha
AT1G22260	-0.453	-	0.252	0	expressed protein
AT1G22340	0.074	0	0.770	+	UDP-glucose glucosyltransferas
AT1G22720	0.436	0	-0.853	-	WAK-like kinase (WLK)
AT1G22750	-0.049	0	0.562	+	expressed protein
AT1G22990	-0.160	0	0.641	+	copper chaperone (CCH)-related
AT1G23390	-0.371	-	-0.326	-	Kelch repeat containing F-box
AT1G23490	-0.311	-	-0.688	-	putative ADP-ribosylation fact
AT1G23890	0.853	+	-0.075	0	NHL repeat-containing protein
AT1G24040	1.582	+	-0.942	-	GCN5-related N-acetyltransfera
AT1G25350	0.696	0	2.071	+	tRNA-glutamine synthetase, put
AT1G25430	-0.260	0	-0.294	-	reverse transcriptase (RNA-dep
AT1G26120	-0.647	-	-0.814	-	esterase-related contains simi
AT1G26350	1.163	+	-0.030	0	hypothetical protein
AT1G26370	-0.672	-	-0.045	0	RNA helicase, putative
AT1G26500	-0.097	0	-0.169	-	pentatricopeptide (PPR) repeat
AT1G27530	-0.527	-	-0.129	0	expressed protein
AT1G28160	0.206	0	-0.533	-	ethylene responsive element bi
AT1G28500	-0.780	-	0.861	+	hypothetical protein
AT1G28630	0.133	0	0.259	+	hypothetical protein
AT1G28760	-0.781	-	0.366	0	hypothetical protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G29070	0.659	+	0.052	0	plastid ribosomal protein L34
AT1G29170	0.079	0	0.392	+	hypothetical protein
AT1G29220	0.259	+	-0.263	-	transcriptional regulator fami
AT1G29330	-0.111	-	-0.348	-	ER lumen protein retaining rec
AT1G29570	2.061	+	0.108	0	zinc finger protein, putative
AT1G30050	0.316	+	-0.191	0	hypothetical protein
AT1G30090	0.064	+	0.550	+	Kelch repeat containing F-box
AT1G30150	0.807	+	-0.636	-	En/Spm-like transposon protein
AT1G30540	-0.317	-	0.557	+	ATPase, BadF/BadG/BcrA/BcrD-ty
AT1G30900	-0.235	-	-0.743	-	spot 3 protein and vacuolar so
AT1G31220	-0.240	0	-0.742	-	putative phosphoribosylglycina
AT1G31500	0.771	+	-0.735	-	endonuclease/exonuclease/phosp
AT1G31580	-1.050	-	-0.933	-	ORF1, putative
AT1G31600	2.003	+	-0.654	0	RRM-containing protein
AT1G31780	0.116	0	-0.465	-	conserved oligomeric Golgi com
AT1G32390	0.015	0	0.431	+	hypothetical protein
AT1G32700	0.026	0	-0.297	-	zinc-binding family protein si
AT1G33220	0.946	+	-0.281	0	glycosyl hydrolase family 17 (
AT1G33240	0.721	+	0.232	+	DNA-binding factor, putative
AT1G33290	-0.433	-	0.089	+	sporulation protein-related is
AT1G33490	0.394	+	0.088	+	expressed protein
AT1G33920	0.330	+	-0.215	0	lectin-related
AT1G34070	0.177	+	0.075	+	hypothetical protein
AT1G34520	-0.678	-	0.585	+	long-chain-alcohol O-fatty-acy
AT1G34740	-0.198	-	0.217	+	mutator-like transposase, puta
AT1G34760	0.397	+	0.570	+	14-3-3 protein GF14 omicron (g
AT1G35190	0.110	+	0.123	+	hyoscyamine 6-dioxygenase hydr
AT1G35420	-0.236	-	-0.145	0	dienelactone hydrolase family
AT1G35710	-0.580	0	-1.196	-	leucine-rich repeat transmembr

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G36000	-0.368	-	-0.207	0	lateral organ boundaries (LOB)
AT1G36060	2.353	+	-0.811	0	AP2 domain transcription facto
AT1G36910	-0.870	-	0.572	+	plant transposase (Ptta/En/Spm
AT1G36970	0.924	+	-0.098	0	hypothetical protein
AT1G36990	0.604	+	0.177	0	expressed protein
AT1G42040	0.926	+	0.175	+	hypothetical protein
AT1G42200	0.640	+	-0.546	0	transposon protein, putative
AT1G42670	0.903	+	-0.865	-	myb family protein
AT1G43560	-0.418	-	-0.395	-	thioredoxin, putative
AT1G43650	0.348	0	-0.786	-	nodulin-like protein
AT1G43750	0.253	0	0.410	+	hypothetical protein
AT1G43763	-0.054	0	-0.620	-	F28H19.2; pseudogene, putative
AT1G43775	1.549	+	0.205	0	pseudogene, similar to polypro
AT1G43830	-0.608	-	-0.094	0	pseudogene, similar to Athila
AT1G43890	-0.623	-	-0.282	0	GTP-binding protein(RAB1Y), pu
AT1G43895	-1.106	-	0.156	0	similar to protein kinase - li
AT1G44030	-0.485	-	0.608	+	CHP-rich zinc finger protein,
AT1G44100	-0.139	-	0.141	+	amino acid permease, putative
AT1G44890	0.390	0	0.642	+	expressed protein
AT1G44940	-0.604	-	0.132	0	F27F5.29; pseudogene, putative
AT1G45688	-0.279	-	-0.364	-	expressed protein
AT1G47540	0.216	+	-0.124	-	trypsin inhibitor 2, putative
AT1G47560	-0.014	0	0.745	+	hypothetical protein
AT1G47620	-0.228	-	0.226	+	cytochrome P450, putative
AT1G48370	1.152	+	-0.246	-	oligopeptide transporter OPT f
AT1G49300	-0.070	-	-0.313	-	small GTP-binding protein, put
AT1G49340	-0.842	-	-0.540	0	phosphatidylinositol 3- and 4-
AT1G50010	-1.100	-	-0.564	0	tubulin alpha-2/alpha-4 chain,
AT1G50360	-0.549	-	0.180	0	myosin, putative

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G50520	-0.708	-	-0.528	-	cytochrome p450 family
AT1G50750	-0.512	-	0.130	+	expressed protein
AT1G51115	0.354	+	0.677	+	unknown protein
AT1G51360	0.534	+	0.120	0	unknown protein
AT1G51490	0.440	+	-0.485	-	glycosyl hydrolase family 1
AT1G51550	1.322	+	-0.046	0	F-box protein ZEITLUPE/FKF/LKP
AT1G52030	-1.219	-	0.157	0	myrosinase binding protein, pu
AT1G52080	0.198	0	0.675	+	actin binding protein family c
AT1G52680	-0.268	-	0.083	0	late embryogenesis abundant (L
AT1G53035	-0.148	0	0.334	+	expressed protein
AT1G53250	0.509	+	0.119	0	expressed protein
AT1G53400	-0.485	-	-0.100	0	pinoresinol-lariciresinol redu
AT1G53490	-1.265	-	0.905	+	bZIP protein
AT1G53550	0.686	+	0.780	+	F-box protein family
AT1G53730	1.215	+	-0.213	0	leucine-rich repeat transmembr
AT1G54020	0.880	+	-0.339	0	myrosinase-associated protein,
AT1G54430	1.034	+	-0.390	0	hypothetical protein
AT1G54540	-0.694	-	0.309	0	hypothetical protein
AT1G55040	0.274	0	-1.533	-	zinc finger (Ran-binding) fami
AT1G55150	0.227	0	-0.316	-	ethylene-responsive DEAD/DEAH
AT1G55260	-0.271	-	0.342	+	protease inhibitor/seed storag
AT1G55600	0.032	0	-1.013	-	WRKY family transcription fact
AT1G55840	-0.607	-	0.121	0	polyphosphoinositide binding p
AT1G55890	0.700	+	0.044	0	pentatricopeptide (PPR) repeat
AT1G55970	0.753	+	-0.299	0	CREB-binding protein, putative
AT1G56190	-0.859	-	-0.038	0	phosphoglycerate kinase, putat
AT1G56230	1.098	+	0.255	+	unknown protein
AT1G57640	0.583	+	0.048	0	polyprotein, putative
AT1G57720	-0.834	-	-0.813	-	elongation factor 1B-gamma, pu

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G57730	-0.074	0	0.192	+	C3HC4-type zinc finger protein
AT1G58025	0.494	+	-0.284	0	DNA-binding protein family
AT1G58260	-0.683	-	0.436	+	cytochrome p450 family
AT1G59730	0.227	+	-0.588	-	thioredoxin, putative
AT1G60470	-0.057	-	0.285	+	galactinol synthase, putative
AT1G60650	0.694	+	-0.141	0	glycine-rich RNA-binding prote
AT1G60740	-0.085	0	-0.681	-	peroxiredoxin, putative
AT1G60800	1.161	+	-0.163	-	receptor-like kinase, putative
AT1G61220	0.020	0	0.402	+	hypothetical protein
AT1G61255	1.339	+	0.128	0	hypothetical protein
AT1G61290	0.902	+	-0.433	0	syntaxin SYP124
AT1G61390	-0.715	-	0.032	0	receptor kinase, putative
AT1G61410	0.939	+	0.003	0	tolA protein-related contains
AT1G61520	-1.107	-	0.157	0	light-harvesting chlorophyll a
AT1G61840	0.468	+	-0.258	0	CHP-rich zinc finger protein,
AT1G61920	-0.238	-	0.122	+	hypothetical protein
AT1G62050	2.112	+	-0.454	-	ankyrin repeat protein-related
AT1G62310	0.188	0	0.696	+	transcription factor jumonji (
AT1G62510	0.361	0	1.439	+	similar to 14KD proline-rich p
AT1G62880	-0.555	-	-0.129	0	cornichon family protein conta
AT1G63340	-0.324	-	0.660	+	flavin-containing monooxygenas
AT1G63520	-0.291	-	-0.088	0	hypothetical protein
AT1G64070	-0.454	-	0.047	0	disease resistance protein (TI
AT1G64355	0.860	+	0.233	0	expressed protein
AT1G64700	1.075	+	-0.040	-	hypothetical protein
AT1G64860	1.532	+	0.248	0	plastid RNA polymerase sigma-s
AT1G65110	-1.266	-	0.141	0	ubiquitin carboxyl-terminal hy
AT1G65230	0.338	+	0.672	+	expressed protein
AT1G65470	-0.141	0	0.423	+	chromatin assembly factor-1 (F

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G65630	-0.308	-	0.221	+	DegP protease
AT1G65730	0.432	+	-0.597	-	oligopeptide transporter OPT f
AT1G65770	0.625	+	-0.004	0	F-box protein family
AT1G66220	1.454	+	-0.226	0	subtilisin-like serine proteas
AT1G66830	-0.946	-	0.508	+	leucine-rich repeat transmembr
AT1G67270	-0.207	0	-0.538	-	hypothetical protein
AT1G68380	1.163	+	-0.276	-	hypothetical protein
AT1G68560	0.005	0	0.254	+	glycosyl hydrolase family 31 (
AT1G68580	-0.005	0	0.275	+	agenet domain-containing prote
AT1G68990	0.120	+	0.122	+	DNA-directed RNA polymerase
AT1G69020	-0.294	-	0.392	+	putative protease
AT1G69680	-0.741	-	-0.042	0	expressed protein
AT1G69710	2.275	+	-0.326	0	putative regulator of chromoso
AT1G69850	-0.255	0	-0.893	-	nitrate transporter (NTL1)
AT1G70170	0.240	+	-0.419	-	matrix metalloproteinase, puta
AT1G70320	0.435	+	0.173	0	ubiquitin-protein ligase 2 (UP
AT1G70350	0.848	+	0.800	+	expressed protein
AT1G70460	0.023	0	-0.631	-	protein kinase -related
AT1G70850	0.096	0	0.204	+	Csf-2-related
AT1G70880	0.179	0	0.639	+	Csf-2-related
AT1G71150	0.097	0	0.469	+	unknown protein
AT1G71340	0.713	+	0.041	0	glycerophosphoryl diester phos
AT1G71680	0.476	+	1.443	+	amino acid permease, putative
AT1G71760	-0.127	-	0.019	0	hypothetical protein
AT1G71860	0.555	+	-0.324	0	protein tyrosine phosphatase
AT1G72060	1.331	+	0.077	0	expressed protein
AT1G72160	0.844	+	0.657	+	cytosolic factor, putative
AT1G72190	-0.493	-	-0.692	-	phosphoglycerate dehydrogenase
AT1G72270	-0.011	0	0.807	+	hypothetical protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G72440	0.256	+	0.046	0	CCAAT-box-binding transcriptio
AT1G72500	-0.356	0	-0.817	-	inter-alpha-trypsin inhibitor
AT1G72840	0.671	+	0.134	0	disease resistance protein (TI
AT1G72900	-0.973	-	-1.157	-	disease resistance protein (TI
AT1G74020	0.339	0	-1.044	-	strictosidine synthase family
AT1G74040	0.606	+	-0.770	-	putative 2-isopropylmalate syn
AT1G74360	0.960	0	-1.479	-	leucine-rich repeat transmembr
AT1G74410	0.768	+	-0.110	0	putative RING zinc finger prot
AT1G74470	-0.260	-	0.096	+	geranylgeranyl reductase
AT1G74890	0.159	0	-0.238	-	response regulator 7, putative
AT1G75020	0.001	0	-0.132	-	putative acyl-CoA:1-acylglycer
AT1G75200	-0.400	-	-0.404	-	flavodoxin family
AT1G75520	0.122	+	0.084	+	lateral root primordium (LRP)
AT1G75560	2.423	+	-0.663	0	DNA-binding protein
AT1G75590	-0.208	-	0.326	+	auxin-induced (indole-3-acetic
AT1G75760	1.241	+	-0.134	0	ER lumen protein retaining rec
AT1G76310	-0.455	0	0.853	+	Cyclin B2;4
AT1G76580	-0.756	-	-0.061	-	SPL1-Related3 protein (SPL1R3)
AT1G76940	0.040	0	0.171	+	RNA recognition motif (RRM)-co
AT1G77310	-0.585	-	0.497	+	wound-responsive protein, puta
AT1G77650	-0.418	-	0.160	0	F-box protein family
AT1G78790	0.484	+	0.118	0	hypothetical protein
AT1G79130	1.259	+	0.234	0	auxin-induced (indole-3-acetic
AT1G79440	-0.012	0	-0.366	-	succinic semialdehyde dehydrog
AT1G79800	0.658	+	-0.366	-	plastocyanin-like domain conta
AT1G79860	-0.123	0	-0.578	-	hypothetical protein
AT1G79990	0.486	+	0.196	0	putative coatomer protein comp
AT1G80110	-0.363	-	0.322	+	expressed protein
AT1G80450	0.735	+	0.501	+	VQ motif-containing protein co

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT1G80460	0.539	+	-0.329	-	putative glycerol kinase
AT1G80540	-0.124	0	0.772	+	hypothetical protein
AT1G80670	0.199	0	-0.701	-	mRNA export protein, putative
AT1G80820	-0.345	-	0.069	+	cinnamoyl CoA reductase, putat
AT2G01310	-0.428	-	-0.127	-	unknown protein
AT2G01490	-0.241	0	-1.462	-	phytanoyl-CoA dioxygenase (Phy
AT2G01670	-0.577	0	-0.745	-	MutT/nudix family protein
AT2G01810	-0.758	-	0.329	0	PHD finger family protein cont
AT2G02670	-0.299	-	-0.301	-	hypothetical protein
AT2G03020	0.712	+	-0.235	0	predicted by genefinder
AT2G03140	-0.236	-	-0.129	-	similar to late embryogenesis
AT2G03760	1.038	+	-0.085	0	putative steroid sulfotransfer
AT2G03850	1.038	+	0.189	+	putative cold-regulated protei
AT2G04020	-0.143	-	-0.951	-	putative GDSL-motif lipase/hyd
AT2G04600	0.118	+	0.797	+	unknown protein
AT2G04920	0.361	+	0.058	0	F-box protein family, AtFBX9
AT2G04940	0.350	+	-0.117	0	scramblase-related weak simila
AT2G05000	-0.825	-	-0.073	0	hypothetical protein
AT2G05420	-0.736	-	-0.058	0	meprin and TRAF homology domai
AT2G05580	-0.023	0	0.710	+	glycine-rich protein
AT2G05940	-0.501	-	-0.294	-	protein kinase, putative
AT2G05950	-0.222	-	0.232	+	En/Spm-like transposon protein
AT2G06990	0.811	+	-0.349	-	DEAD/DEAH box RNA helicase (HU
AT2G10440	-0.317	-	0.494	+	hypothetical protein
AT2G10940	1.256	+	0.182	+	protease inhibitor/seed storag
AT2G11240	-0.545	-	0.454	0	putative non-LTR retroelement
AT2G12920	0.994	+	0.444	+	putative retroelement pol poly
AT2G13510	0.456	+	-0.246	0	putative Ta11-like non-LTR ret
AT2G13560	-0.690	-	0.010	0	malate oxidoreductase (malic e

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT2G13720	1.267	+	-0.117	0	hypothetical protein
AT2G13780	-0.358	0	-0.414	-	unknown protein
AT2G14100	0.781	+	0.242	0	cytochrome p450 family
AT2G14240	-0.378	0	1.298	+	hypothetical protein
AT2G14270	0.167	0	0.453	+	protein phosphatase 2C -relate
AT2G14440	-0.299	-	-0.172	-	putative receptor-like protein
AT2G14570	-0.312	0	1.494	+	Mutator-like transposase
AT2G14610	-0.117	0	-2.203	-	pathogenesis-related PR-1-like
AT2G14690	-0.639	-	0.079	0	glycosyl hydrolase family 10
AT2G14700	0.732	+	-0.114	0	hypothetical protein
AT2G14720	-0.364	-	0.382	+	spot 3 protein and vacuolar so
AT2G15080	-0.429	-	-0.692	-	disease resistance protein fam
AT2G15100	0.829	+	0.333	+	putative retroelement pol poly
AT2G15880	-0.742	-	-0.142	0	leucine-rich repeat extensin f
AT2G16030	-0.284	-	0.495	+	expressed protein
AT2G16640	-0.700	-	-0.073	0	putative chloroplast outer mem
AT2G16710	-0.600	-	-0.189	0	putative HesB-like protein
AT2G16820	-1.172	-	0.220	0	Mutator-like transposase
AT2G17040	1.039	+	0.668	0	NAM (no apical meristem)-like
AT2G17190	0.442	0	-0.679	-	putative ubiquitin-like protei
AT2G17250	-0.523	-	-0.230	0	hypothetical protein
AT2G17300	0.437	+	0.785	+	expressed protein
AT2G17510	-0.343	-	-0.150	-	putative mitotic control prote
AT2G18730	-0.338	0	-0.882	-	putative diacylglycerol kinase
AT2G18750	-0.524	-	0.341	+	calmodulin-binding protein
AT2G19385	0.501	+	-0.563	-	Expressed protein
AT2G19560	-0.145	-	-0.265	-	proteasome protein-related wea
AT2G19890	-0.699	-	0.026	0	hypothetical protein
AT2G20370	0.875	+	-0.083	0	exostosin family protein conta

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT2G20490	-0.467	0	-1.068	-	nucleolar RNA-binding Nop10p f
AT2G20700	1.276	+	-0.541	0	expressed protein
AT2G20960	-0.734	-	-0.350	0	pEARLI 4 protein
AT2G21130	-1.093	-	-0.498	0	cyclophilin (CYP2)
AT2G21220	-0.449	-	0.406	+	putative auxin-regulated prote
AT2G21330	-0.247	0	0.608	+	putative fructose bisphosphate
AT2G21350	0.792	+	-0.134	-	unknown protein
AT2G21370	0.471	+	0.414	+	putative xylulose kinase
AT2G21480	0.294	0	-0.666	-	protein kinase family
AT2G21580	-1.176	-	-0.266	-	40S ribosomal protein S25
AT2G21590	0.726	+	0.515	+	putative ADP-glucose pyrophosp
AT2G21900	-0.518	-	-0.332	-	WRKY family transcription fact
AT2G21960	0.682	+	0.146	0	expressed protein
AT2G22100	0.343	+	0.185	+	RRM-containing RNA-binding pro
AT2G22360	0.955	+	-0.168	-	putative DnaJ protein
AT2G22420	0.112	0	0.439	+	putative peroxidase
AT2G22540	-0.589	-	-0.061	0	MADS-box protein
AT2G22780	0.237	0	0.384	+	putative glyoxysomal malate de
AT2G23510	-0.293	0	0.392	+	putative anthranilate N-hydrox
AT2G24010	0.845	+	-0.121	0	putative serine carboxypeptida
AT2G24200	-1.307	-	0.036	0	putative leucine aminopeptidas
AT2G24550	-0.704	-	0.119	0	expressed protein
AT2G25020	0.573	+	-0.190	0	hypothetical protein
AT2G25580	1.311	+	-0.165	0	putative selenium-binding prot
AT2G25590	0.942	+	-0.266	0	agenet domain-containing prote
AT2G25690	-0.118	0	0.349	+	senescence-associated protein
AT2G25700	-0.059	0	0.125	+	E3 ubiquitin ligase SCF comple
AT2G25710	-0.656	-	0.471	+	biotin holocarboxylase synthet
AT2G25720	1.108	+	-0.236	0	expressed protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT2G25980	-0.559	0	0.880	+	similar to jasmonate-inducible
AT2G26000	0.762	+	-0.575	-	zinc finger (C3HC4-type RING f
AT2G26050	1.099	+	-0.353	0	hypothetical protein
AT2G26340	-0.295	-	0.551	+	expressed protein
AT2G26380	0.801	+	-0.446	-	disease resistance protien-rel
AT2G26520	-0.065	0	-0.713	-	expressed protein
AT2G26680	0.334	0	0.585	+	expressed protein
AT2G26830	0.432	+	-0.371	-	putative choline kinase
AT2G26940	0.379	+	-0.080	0	putative C2H2-type zinc finger
AT2G27010	0.577	+	-0.194	0	cytochrome p450, putative
AT2G27070	0.955	+	0.188	0	putative two-component respons
AT2G27120	-0.250	-	-0.062	-	putative DNA polymerase epsilo
AT2G27320	-0.791	-	0.514	0	hypothetical protein
AT2G27440	-0.378	0	0.488	+	putative rac GTPase activating
AT2G27750	0.066	0	0.517	+	nucleolar matrix protein-relat
AT2G27760	0.902	+	-0.393	0	putative tRNA isopentenylpyrop
AT2G28190	0.601	+	0.456	+	copper/zinc superoxide dismuta
AT2G28240	0.360	+	0.140	0	hydroxyproline-rich glycoprote
AT2G28530	-0.674	-	0.167	0	putative RING zinc finger tran
AT2G28600	-0.646	0	-0.668	-	putative ATP-dependent RNA hel
AT2G28690	0.541	+	-0.286	-	hypothetical protein
AT2G28920	1.863	+	-0.275	0	C3HC4-type zinc finger protein
AT2G29140	0.774	+	0.153	0	pumilio-family RNA-binding pro
AT2G29190	-0.317	-	0.464	+	pumilio-family RNA-binding pro
AT2G29450	-0.806	-	0.328	0	glutathione transferase (103-1
AT2G29550	-0.009	0	0.452	+	tubulin beta-7 chain
AT2G29790	0.830	+	0.068	0	hypothetical protein
AT2G30150	0.554	+	-0.200	0	putative glucosyltransferase
AT2G30170	0.194	+	-0.486	-	expressed protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT2G30300	-0.084	0	0.461	+	nodulin-related weak similarit
AT2G31020	-0.919	-	-0.088	0	putative oxysterol-binding pro
AT2G31410	-0.176	-	0.195	+	expressed protein
AT2G31670	0.783	+	0.125	0	expressed protein
AT2G31730	0.094	0	0.382	+	ethylene-responsive protein, p
AT2G31865	-0.672	0	-0.767	-	similar to putative poly(ADP-r
AT2G31945	0.883	+	-0.137	-	expressed protein
AT2G31970	-0.490	-	-0.115	-	putative RAD50 DNA repair prot
AT2G32160	1.134	+	-0.421	0	hypothetical protein
AT2G32250	-0.452	-	0.024	0	Mutator-like transposase
AT2G32300	1.236	+	-0.690	0	putative uclacyanin I
AT2G32440	0.547	+	0.282	+	ent-kaurenoic acid oxidase (cy
AT2G33210	1.313	+	-0.511	0	mitochondrial chaperonin (HSP6
AT2G33550	-0.683	-	0.787	+	gt-2-related weak similarity t
AT2G33670	0.832	+	-0.480	-	similar to Mlo proteins from H
AT2G34110	0.108	0	0.673	+	hypothetical protein
AT2G34320	0.326	+	0.434	+	putative non-LTR retroelement
AT2G34490	1.287	+	0.236	0	cytochrome p450 family
AT2G35250	-0.073	0	-0.892	-	hypothetical protein
AT2G35570	0.785	+	-0.117	0	putative serpin
AT2G35610	-0.628	-	-0.122	0	expressed protein
AT2G35640	-1.451	-	1.105	+	proline-rich protein family
AT2G35950	-1.228	-	-0.001	0	expressed protein
AT2G36060	-0.270	0	-0.581	-	E2, ubiquitin-conjugating enzy
AT2G36120	-0.421	-	-0.275	-	glycine-rich protein
AT2G36220	-0.176	-	-0.375	-	expressed protein
AT2G36330	-0.466	-	0.553	+	integral membrane protein, put
AT2G36340	0.039	0	0.391	+	DNA-binding storekeeper protei
AT2G36760	-0.737	-	1.051	+	putative glucosyl transferase

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT2G37140	-0.363	-	0.565	+	terpene synthase/cyclase-relat
AT2G37190	-0.561	-	-0.570	-	60S ribosomal protein L12
AT2G37480	-0.329	-	-0.489	-	expressed protein
AT2G37620	-0.827	-	-0.229	0	actin 3
AT2G37690	-0.354	-	-0.205	-	putative phosphoribosylaminoim
AT2G38040	-0.853	-	-0.336	0	putative alpha-carboxyltransfe
AT2G38180	-0.133	-	0.283	+	GDSL-motif lipase/hydrolase pr
AT2G38250	0.394	+	-0.415	-	putative GT-1-like transcripti
AT2G38580	-0.622	-	-0.080	0	hypothetical protein
AT2G38610	-0.177	-	-0.210	-	KH domain protein
AT2G39110	-0.135	0	0.352	+	protein kinase, putative
AT2G39310	0.426	+	0.730	+	putative myrosinase-binding pr
AT2G39470	-0.196	0	-0.430	-	oxygen-evolving complex 25.6 k
AT2G40060	0.467	+	0.357	+	expressed protein
AT2G40090	0.433	+	-0.061	0	putative ABC1 protein
AT2G40560	1.671	+	-1.009	-	protein kinase family
AT2G40790	-0.750	-	-0.120	0	putative thioredoxin H
AT2G40960	-0.465	-	-0.196	0	expressed protein
AT2G41960	0.008	0	0.313	+	unknown protein
AT2G42250	0.792	+	0.319	0	cytochrome p450, putative
AT2G42690	-1.044	-	0.223	0	putative lipase
AT2G42980	-0.159	0	0.489	+	putative chloroplast nucleoid
AT2G42990	-0.790	-	-0.597	-	putative APG isolog protein
AT2G43220	0.602	+	-0.091	-	CHP-rich zinc finger protein,
AT2G43460	-1.438	-	-0.540	0	60S ribosomal protein L38
AT2G43920	-0.076	0	0.444	+	thiol methyltransferase
AT2G44690	-0.505	-	0.338	+	rac family GTP-binding protein
AT2G45470	-1.516	-	0.114	+	fasciclin-like arabinogalactan
AT2G45550	0.735	+	0.652	+	cytochrome p450 family

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT2G45560	-0.572	-	0.753	+	putative cytochrome P450
AT2G45580	-0.020	0	1.088	+	cytochrome p450 family
AT2G45780	0.171	+	0.225	+	hypothetical protein
AT2G45830	1.815	+	-0.202	0	expressed protein
AT2G45990	-0.782	-	0.100	0	expressed protein
AT2G46380	-0.221	0	-0.644	-	hypothetical protein
AT2G46450	-0.706	-	-0.510	-	cyclic nucleotide-regulated io
AT2G46720	0.012	0	-0.094	-	putative beta-ketoacyl-CoA syn
AT2G46770	1.013	+	-0.541	-	NAM (no apical meristem)-like
AT2G47140	0.414	+	-0.234	0	putative alcohol dehydrogenase
AT2G47990	-0.250	-	-0.237	-	transducin / WD-40 repeat prot
AT2G48020	-0.687	-	0.246	0	putative sugar transporter
AT2G48080	1.127	+	-0.322	0	oxidoreductase, 2OG-Fe(II) oxy
AT3G01120	-1.021	-	-0.307	0	putative cystathionine gamma-s
AT3G01130	-0.605	-	-0.536	-	expressed protein
AT3G01530	0.777	+	-0.406	-	myb family transcription facto
AT3G01890	-0.034	-	0.402	+	SWIB complex BAF60b domain-con
AT3G01920	0.105	+	0.014	0	yrdC family protein contains P
AT3G01980	0.168	0	0.843	+	putative dehydrogenase
AT3G02020	0.075	0	1.115	+	putative aspartate kinase
AT3G02080	-0.917	-	-0.752	0	putative 40S ribosomal protein
AT3G02360	-0.459	-	-0.527	-	6-phosphogluconate dehydrogena
AT3G02400	-0.317	-	-0.366	-	forkhead-associated domain-con
AT3G02560	-0.551	-	-0.878	-	putative 40S ribosomal protein
AT3G02740	0.137	0	-0.567	-	putative aspartyl protease
AT3G02875	0.204	+	0.728	+	IAA-amino acid hydrolase (ILR1
AT3G03060	-0.239	-	0.891	+	putative 26S proteosome regula
AT3G03080	-0.600	-	0.591	+	putative NADP-dependent oxidor
AT3G03300	-0.193	0	-0.817	-	DEAD/DEAH box helicase carpel

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G03640	-0.116	0	-1.112	-	glycosyl hydrolase family 1
AT3G03710	0.828	+	-0.143	0	polyribonucleotide nucleotidyl
AT3G03740	0.256	+	0.025	0	speckle-type POZ protein-relat
AT3G04050	-0.066	-	0.193	+	putative pyruvate kinase
AT3G04200	0.581	+	0.753	+	germin-like protein
AT3G04330	0.366	+	-0.538	-	putative trypsin inhibitor
AT3G04430	0.077	0	0.322	+	No apical meristem (NAM) prote
AT3G04690	0.042	+	-0.580	-	protein kinase family
AT3G04730	-0.642	-	0.433	+	auxin-induced protein
AT3G04900	0.875	+	-0.375	-	heavy-metal-associated domain-
AT3G05060	-0.359	-	0.483	+	putative SAR DNA-binding prote
AT3G05280	-0.386	-	-0.490	-	integral membrane Yip1 family
AT3G05310	-0.702	-	0.102	0	GTP-binding protein - related
AT3G05530	0.122	0	0.781	+	26S proteasome AAA-ATPase subu
AT3G05600	2.521	+	-0.346	0	putative epoxide hydrolase
AT3G05670	-0.805	-	-0.092	0	PHD finger family protein cont
AT3G05800	-0.154	-	0.822	+	bHLH protein
AT3G05850	-0.291	-	0.405	+	unknown protein
AT3G06000	0.057	0	-0.228	-	RAN GTPase activating protein-
AT3G06090	0.813	+	0.149	0	hypothetical protein
AT3G07040	0.541	+	-0.741	-	disease resistance protein RPM
AT3G07200	-0.358	-	0.709	+	putative RING zinc finger prot
AT3G07260	0.890	+	-0.408	-	forkhead-associated domain-con
AT3G07680	-0.672	-	-0.419	0	putative coated vesicle membra
AT3G07700	0.465	+	0.449	+	ABC1 family protein contains P
AT3G07800	1.181	+	0.209	0	putative thymidine kinase
AT3G08580	-1.430	-	-0.687	0	adenylate translocator
AT3G08943	-0.111	0	-0.768	-	pseudogene, importin beta subu
AT3G09300	1.030	+	0.066	0	putative oxysterol-binding pro

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G09360	0.660	+	-0.612	-	putative transcription factor
AT3G09730	1.150	+	-0.330	-	hypothetical protein
AT3G09740	-0.832	-	-0.585	-	syntaxin of plants 71 (SYP71)
AT3G09760	-0.056	0	0.489	+	C3HC4-type zinc finger protein
AT3G09805	-0.547	-	-0.170	0	putative isocitrate dehydrogen
AT3G10260	-0.099	-	0.180	+	reticulon family protein weak
AT3G10420	1.183	+	-0.217	0	expressed protein
AT3G11350	-0.816	-	0.515	0	hypothetical protein
AT3G11670	-0.702	-	0.607	+	digalactosyldiacylglycerol syn
AT3G12010	0.344	+	-0.293	-	expressed protein
AT3G12110	0.965	+	-0.049	0	actin 11 (ACT11)
AT3G12200	-0.179	0	0.529	+	protein kinase, putative
AT3G12540	0.493	0	-0.932	-	hypothetical protein
AT3G12620	0.516	+	-0.107	0	protein phosphatase 2C, putati
AT3G12780	-0.800	-	-0.343	0	phosphoglycerate kinase, putat
AT3G12900	0.347	+	-0.089	0	oxidoreductase, 2OG-Fe(II) oxy
AT3G13235	-0.806	-	-1.086	-	ubiquitin family
AT3G13350	-0.377	-	0.494	+	high mobility group (HMG1/2) f
AT3G13740	0.317	+	0.078	+	URF 4-related similar to URF 4
AT3G13800	0.474	+	0.366	+	hydrolase, putative
AT3G13930	0.343	0	0.596	+	putative acetyltransferase
AT3G14280	0.489	+	0.179	0	expressed protein
AT3G14500	-0.303	-	0.244	+	hypothetical protein
AT3G14720	0.444	0	0.633	+	putative MAP kinase
AT3G14740	1.289	+	0.067	0	zinc finger protein-related
AT3G14910	0.249	+	0.420	+	hypothetical protein
AT3G15230	0.200	+	-0.228	-	unknown protein
AT3G15270	-0.535	-	-0.086	0	squamosa promoter binding prot
AT3G15280	-0.101	-	0.842	+	expressed protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G15760	-0.281	-	-0.198	-	expressed protein
AT3G15790	0.313	+	0.540	+	methyl-CpG-binding domain-cont
AT3G15810	0.427	+	-0.938	-	expressed protein
AT3G16050	-0.836	-	-0.120	0	putative ethylene-inducible pr
AT3G16110	1.241	+	0.310	0	protein disulfide isomerase fa
AT3G16180	-0.596	-	0.869	+	putative transport protein
AT3G16400	-1.290	-	1.152	+	putative lectin
AT3G16520	0.807	+	0.110	0	putative glucosyltransferase
AT3G17030	-0.086	0	0.836	+	hypothetical protein
AT3G17110	0.188	+	-0.153	-	glycine-rich protein
AT3G17130	0.850	+	0.128	0	invertase/pectin methylesteras
AT3G17290	-0.057	0	0.321	+	hypothetical protein
AT3G17530	-0.357	-	0.363	+	F-box family protein contains
AT3G18000	-0.747	-	0.083	0	methyltransferase, putative
AT3G18060	0.025	0	-0.267	-	putative WD-repeat protein (WD
AT3G18330	0.643	+	0.560	+	F-box family protein contains
AT3G18780	0.882	+	-0.468	-	actin 2
AT3G19550	0.829	+	0.136	0	expressed protein
AT3G19690	0.864	+	-0.179	0	PR-1 protein, putative
AT3G19930	0.705	+	-0.315	-	monosaccharide transport prote
AT3G20120	1.051	+	-0.465	0	cytochrome p450 family
AT3G20150	-0.734	-	0.134	+	kinesin-like protein
AT3G20250	0.097	0	0.599	+	pumilio-family RNA-binding pro
AT3G20490	-0.369	-	-0.692	-	hypothetical protein
AT3G20870	0.206	+	0.476	+	metal transporter family
AT3G21175	-0.364	-	-0.277	0	GATA zinc finger protein
AT3G21820	-0.420	0	0.998	+	SET-domain transcriptional reg
AT3G22104	-0.500	-	0.172	+	phototropic-responsive NPH3 pr
AT3G22250	-0.283	0	-0.543	-	glycosyltransferase family

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G22350	0.820	+	0.165	0	F-box protein family
AT3G22360	1.719	+	-0.280	0	alternative oxidase 1b precurs
AT3G22740	0.099	0	0.824	+	putative selenocysteine methyl
AT3G22790	0.718	+	0.450	0	kinase interacting family prot
AT3G23050	1.186	+	-0.392	0	indoleacetic acid (IAA)-induci
AT3G23070	0.107	0	0.437	+	expressed protein
AT3G23960	-0.230	-	0.000	0	F-box protein family
AT3G24030	-0.412	-	0.321	+	hydoxyethylthiazole kinase, pu
AT3G24150	0.393	+	0.415	+	expressed protein
AT3G24200	-0.438	-	-0.078	0	flavin-dependent hydroxylase V
AT3G24210	0.391	+	0.423	+	ankyrin repeat family protein
AT3G24430	0.117	+	0.183	+	mrp protein, putative
AT3G24830	-1.038	-	-0.665	-	60S ribosomal protein, putativ
AT3G25500	0.527	+	0.350	0	formin homology 2 (FH2) domain
AT3G25700	-0.304	-	-0.052	0	chloroplast nucleoid DNA-bindi
AT3G25855	-0.012	0	0.630	+	Expressed protein
AT3G25890	-0.947	-	-0.024	0	AP2 domain transcription facto
AT3G25940	0.780	+	0.019	0	transcription factor S-II (TFI
AT3G26180	-0.458	-	-0.257	-	cytochrome p450 family
AT3G26330	0.283	+	0.779	+	cytochrome p450 family
AT3G27280	-0.420	-	0.125	+	prohibitin, putative
AT3G27290	-0.162	0	0.548	+	F-box family protein-related c
AT3G27360	-1.201	-	0.387	0	histone H3, putative
AT3G27670	-0.119	-	-0.383	-	expressed protein
AT3G27840	1.135	+	0.175	0	50S ribosomal protein L12-B
AT3G27940	0.035	0	-0.272	-	lateral organ boundaries (LOB)
AT3G28180	-0.520	-	0.448	+	glycosyltransferase family 2
AT3G28740	0.747	+	-0.105	0	cytochrome p450 family
AT3G28880	0.241	0	-0.526	-	ankyrin repeat family protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G28990	-0.427	-	0.302	+	hypothetical protein
AT3G29060	-0.378	0	0.972	+	EXS family protein / ERD1/XPR1
AT3G29380	0.360	+	-0.317	-	transcription initiation facto
AT3G29410	0.818	+	0.057	0	terpene synthase/cyclase famil
AT3G29595	0.473	+	-0.250	0	pseudogene, hypothetical prote
AT3G29612	-0.176	0	0.442	+	pseudogene, hypothetical prote
AT3G29635	0.626	+	0.337	+	transferase family
AT3G29670	0.351	+	-0.393	-	anthocyanin 5-aromatic acyltra
AT3G30396	0.051	0	-0.532	-	En/Spm transposon protein, put
AT3G30480	0.905	+	-0.533	-	hypothetical protein
AT3G31470	0.959	+	-0.224	0	pseudogene, putative reverse t
AT3G31930	-0.341	-	0.351	+	hypothetical protein
AT3G32930	0.284	+	0.297	+	expressed protein
AT3G42210	0.798	+	-0.232	-	putative protein
AT3G42320	0.117	+	-0.646	-	putative protein
AT3G42360	1.578	+	0.369	0	pseudogene, putative protein
AT3G42460	0.964	+	0.759	+	putative protein
AT3G42650	0.285	+	0.392	+	plant transposase (Ptta/En/Spm
AT3G42670	0.276	+	-0.113	0	SNF2domain/helicase domain-con
AT3G42790	0.743	+	-0.220	-	nucleic acid binding protein-l
AT3G42910	0.823	+	-0.440	-	hypothetical protein
AT3G42950	0.486	+	0.190	+	polygalacturonase, putative
AT3G43060	-1.344	-	-0.136	0	putative protein
AT3G43090	-0.179	-	-0.356	-	putative protein
AT3G43530	0.057	0	0.811	+	putative protein
AT3G43580	0.508	0	-0.685	-	hypothetical protein
AT3G43590	0.890	+	-0.698	0	zinc knuckle (CCHC-type) famil
AT3G43950	-0.607	-	-0.019	0	putative protein
AT3G44380	-0.792	-	0.021	0	putative protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G44610	-0.431	-	0.038	0	protein kinase-like protein
AT3G44830	0.136	+	0.243	+	lecithin:cholesterol acyltrans
AT3G45060	0.010	0	0.379	+	high-affinity nitrate transpor
AT3G45080	-0.198	-	-0.234	-	sulfotransferase-like protein
AT3G45270	0.660	+	0.726	+	hypothetical protein
AT3G45290	-0.708	-	0.210	0	seven transmembrane MLO protei
AT3G45370	1.619	+	-0.056	0	hypothetical protein
AT3G45560	-0.170	-	-0.108	-	zinc finger (C3HC4-type RING f
AT3G45620	-0.535	-	-0.967	-	transducin / WD-40 repeat prot
AT3G46000	-0.385	-	0.494	+	actin depolymerizing factor 2
AT3G46220	-0.118	0	0.389	+	expressed protein
AT3G46240	-0.515	-	-0.157	0	glycine-rich protein
AT3G46360	-0.510	-	0.182	+	hypothetical protein
AT3G46440	-0.133	0	0.256	+	dTDP-glucose 4-6-dehydratases-
AT3G46450	1.558	+	0.093	+	SEC14 cytosolic factor, putati
AT3G46480	1.048	+	0.268	0	oxidoreductase, 2OG-Fe(II) oxy
AT3G46520	-0.833	-	0.415	0	A.t. DNA chromosome 3, BAC clo
AT3G46580	0.747	+	0.175	0	methyl-CpG-binding domain-cont
AT3G46720	-0.680	-	0.084	0	glucuronosyl transferase-like
AT3G46930	-0.015	0	-0.546	-	protein kinase 6-like protein
AT3G47030	-0.531	-	0.136	0	F-box protein family
AT3G47060	0.017	0	0.241	+	FtsH protease, putative
AT3G47340	-0.063	0	-1.372	-	glutamine-dependent asparagine
AT3G47470	-0.605	-	0.356	+	light-harvesting chlorophyll a
AT3G47610	0.468	+	0.197	+	putative protein
AT3G48650	0.133	0	-0.732	-	hypothetical protein
AT3G49010	-1.255	-	-0.784	0	60S ribosomal protein L13, BBC
AT3G49070	0.665	+	0.285	0	putative protein
AT3G49130	-0.224	-	0.268	+	putative protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G49160	-0.280	-	-0.388	-	pyruvate kinase -like protein
AT3G49240	-1.085	-	0.479	0	pentatricopeptide (PPR) repeat
AT3G49390	0.578	+	-0.411	-	RNA-binding protein, putative
AT3G49730	-0.447	-	-0.260	-	pentatricopeptide (PPR) repeat
AT3G49740	0.711	+	0.148	0	pentatricopeptide (PPR) repeat
AT3G49950	-0.294	-	0.245	+	scarecrow transcription factor
AT3G50430	0.252	+	0.316	+	expressed protein
AT3G50860	-0.079	0	-0.390	-	putative clathrin coat assembl
AT3G50980	1.296	+	0.024	0	dehydrin-like protein
AT3G51070	0.582	+	0.464	+	dehydration-induced protein-re
AT3G51190	-0.232	-	0.719	+	ribosomal protein L8 homolog
AT3G51250	0.415	+	-0.150	0	senescence/dehydration-associa
AT3G51400	0.989	+	0.235	0	expressed protein
AT3G51480	1.416	+	-0.410	0	putative glutamate receptor
AT3G51840	-0.266	-	-0.373	-	acyl-coA dehydrogenase
AT3G52070	1.279	+	-0.634	0	expressed protein
AT3G52080	1.319	+	-0.107	0	cation/hydrogen exchanger, put
AT3G52155	0.249	+	0.327	+	expressed protein
AT3G52250	-0.048	0	0.544	+	myb family transcription facto
AT3G52330	-0.253	-	0.202	+	F-box protein-related
AT3G53780	-0.800	-	0.376	0	membrane protein, Rhomboid fam
AT3G53890	-1.903	-	-0.745	0	40S ribosomal protein S21 homo
AT3G53900	0.133	+	0.244	+	uracil phosphoribosyltransfera
AT3G54310	-0.078	-	0.730	+	putative protein
AT3G54410	-0.238	-	-0.525	-	hypothetical protein
AT3G54500	-0.233	-	0.314	+	putative protein
AT3G54680	0.046	0	-0.472	-	b/t At3g54680.1 proteophosphog
AT3G54800	-0.929	-	-0.373	-	pleckstrin homology (PH) domai
AT3G54810	-0.361	-	0.422	+	GATA zinc finger protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G54840	-0.270	-	-0.162	-	Rab family GTP-binding protein
AT3G54970	-0.867	-	0.129	0	expressed protein
AT3G55360	-0.329	-	-0.252	-	synaptic glycoprotein SC2-like
AT3G55620	1.122	+	-0.220	0	eukaryotic translation initiat
AT3G55650	0.698	+	0.227	0	pyruvate kinase - like protein
AT3G55980	0.875	+	-0.044	0	zinc finger (CCCH-type) family
AT3G56090	0.713	+	-0.087	0	ferritin, putative similar to
AT3G56110	1.028	+	0.096	0	prenylated rab acceptor (PRA1)
AT3G56170	1.265	+	-0.371	0	Ca(2+)-dependent nuclease
AT3G56210	-0.256	-	0.155	+	expressed protein
AT3G56290	0.788	0	0.939	+	expressed protein
AT3G56360	-0.330	0	-0.862	-	expressed protein
AT3G56370	-0.567	-	-0.484	-	leucine-rich repeat transmembr
AT3G56630	0.225	+	0.472	+	cytochrome P450, putative
AT3G56890	-0.545	0	0.769	+	putative protein
AT3G56910	-0.030	0	-0.197	-	ribosomal protein, chloroplast
AT3G56980	1.190	+	-0.190	-	bHLH protein
AT3G57430	0.448	+	0.066	0	pentatricopeptide (PPR) repeat
AT3G57640	0.523	+	-0.034	0	protein kinase family protein
AT3G57680	-1.188	-	0.688	0	D1 CtpA arboxy-terminal protea
AT3G57770	1.082	+	-0.488	0	protein kinase - like protein
AT3G58670	1.138	+	-0.250	0	putative protein
AT3G58730	-1.057	-	0.216	0	v-ATPase subunit D (vATPD)
AT3G59310	-0.320	-	0.074	0	(Anm1) anthocyanin-related mem
AT3G59440	0.855	+	0.122	0	calmodulin-like protein
AT3G59970	-0.812	-	0.363	0	methylenetetrahydrofolate redu
AT3G59980	-0.420	-	-0.054	0	tRNA-binding region domain-con
AT3G60120	0.252	0	0.867	+	glycosyl hydrolase family 1
AT3G60490	0.999	+	0.413	+	transcription factor - like pr

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT3G60520	-0.341	-	0.084	0	expressed protein
AT3G60780	0.651	+	-0.003	0	expressed protein
AT3G60810	0.922	+	0.677	+	expressed protein
AT3G60850	0.027	0	-0.297	-	putative protein
AT3G61050	-0.236	0	0.537	+	CaLB protein
AT3G61300	0.099	+	-0.151	-	anthranilate phosphoribosyltra
AT3G61380	2.036	+	0.720	+	putative protein
AT3G61490	-0.724	-	0.333	0	polygalacturonase, putative
AT3G61640	0.530	+	-0.013	0	arabinogalactan-protein (AGP20
AT3G61740	0.599	+	-0.026	-	PHD finger family protein (ATX
AT3G61970	-1.076	-	0.681	+	RAV-like protein
AT3G62250	-1.574	-	-0.230	0	ubiquitin extension protein (U
AT3G62270	0.312	0	0.564	+	anion exchange protein family
AT3G62500	0.524	+	0.918	+	putative protein
AT3G62890	1.056	+	-0.252	0	pentatricopeptide (PPR) repeat
AT3G62950	1.572	+	-0.743	0	glutaredoxin -like protein
AT3G63010	0.493	+	0.039	0	putative protein
AT3G63020	-0.651	-	-0.034	0	putative protein
AT3G63170	1.265	+	-0.105	-	expressed protein
AT3G63240	0.386	+	0.050	0	inositol-1,4,5-trisphosphate 5
AT3G63360	-0.988	-	0.602	+	defensin-related contains weak
AT3G63400	0.488	+	0.134	+	cyclophylin -like protein
AT4G00355	-0.687	-	-0.852	-	Expressed protein
AT4G00660	-0.741	-	-0.493	-	DEAD/DEAH box RNA helicase, pu
AT4G00780	-0.368	-	0.112	+	coded for by A. thaliana cDNA
AT4G00880	-0.374	-	0.402	+	coded for by A. thaliana cDNA
AT4G01440	-1.018	-	0.284	0	nodulin N21 family protein
AT4G01450	-0.231	-	0.065	0	nodulin MtN21 family protein
AT4G01520	-0.265	0	0.939	+	putative NAM-like protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT4G01650	-0.287	0	0.586	+	expressed protein
AT4G01892	-0.454	-	0.203	+	similar to NIMIN-1b protein A
AT4G02100	-0.244	-	-0.109	0	DnaJ domain-containing protein
AT4G02120	-0.918	-	0.212	+	CTP synthase - like protein
AT4G02140	-0.392	-	0.611	+	expressed protein
AT4G02590	-0.840	0	1.274	+	bHLH protein
AT4G02670	-0.034	0	-0.373	-	putative zinc finger protein
AT4G02710	-0.461	0	0.939	+	kinase interacting family prot
AT4G02880	0.234	+	-0.103	-	hypothetical protein
AT4G03160	0.268	+	0.641	+	hypothetical protein
AT4G03210	-0.351	0	0.586	+	xyloglucan endotransglycosylas
AT4G03410	-0.402	-	-0.181	-	peroxisomal membrane protein-r
AT4G03490	0.791	+	0.049	0	ankyrin repeat family protein
AT4G03520	-1.072	-	-0.121	-	putative M-type thioredoxin
AT4G03540	-0.352	-	0.347	+	hypothetical integral membrane
AT4G03610	0.612	+	0.296	0	putative hydrolase
AT4G03640	-0.674	-	0.646	+	hypothetical protein
AT4G03880	-0.093	-	0.116	+	putative transposon protein
AT4G03916	-0.096	0	0.630	+	similar to putative helicase
AT4G04410	0.133	+	0.725	+	pseudogene, similar to polypro
AT4G04580	-0.179	0	0.567	+	myb family transcription facto
AT4G04640	0.012	0	0.424	+	coded for by A. thaliana cDNA
AT4G05590	-0.698	-	-0.610	0	putative protein
AT4G07380	-0.036	0	1.194	+	hypothetical protein
AT4G07460	1.234	+	0.118	+	contains similarity to an Arab
AT4G07700	-0.438	-	0.433	+	putative athila transposon pro
AT4G07720	0.043	+	-0.287	-	hypothetical protein
AT4G07730	1.078	+	-0.702	-	putative athila transposon pro
AT4G08500	1.025	+	-0.621	-	MEKK1/MAP kinase kinase kinase

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT4G08980	-0.341	-	0.383	+	F-box protein family, AtFBW2
AT4G09070	0.502	0	-0.716	-	putative protein
AT4G09420	1.122	+	-0.094	0	disease resistance protein (TI
AT4G09540	-0.125	0	-0.765	-	putative protein
AT4G09690	-0.633	-	0.201	+	CHP-rich zinc finger protein,
AT4G09800	-1.252	-	-0.692	-	S18.A ribosomal protein
AT4G09980	-0.400	-	0.086	0	methyltransferase MT-A70 famil
AT4G10570	0.363	+	0.109	+	ubiquitin carboxyl-terminal hy
AT4G11010	-1.652	-	-0.560	-	nucleoside diphosphate kinase
AT4G11380	-0.350	-	-0.054	0	beta-adaptin - like protein
AT4G11460	-0.730	-	-0.809	-	serine/threonine kinase-like p
AT4G11590	-0.323	-	0.365	+	F-box protein family
AT4G11970	-0.139	-	0.585	+	YT521-B-like family protein co
AT4G12270	-0.777	-	-0.262	0	copper amine oxidase like prot
AT4G12560	1.718	+	0.001	0	F-box protein family
AT4G12920	-0.317	-	0.351	+	aspartyl protease family prote
AT4G13050	-0.482	-	0.423	0	oleoyl-(acyl-carrier-protein)
AT4G13390	-0.728	-	-0.339	-	extensin-like protein
AT4G13690	-0.172	-	0.892	+	hypothetical protein
AT4G13800	0.044	0	0.924	+	permease-related contains 9 pr
AT4G13820	-0.036	0	-0.209	-	leucine rich repeat protein fa
AT4G13830	-0.159	-	0.136	0	DnaJ-like protein
AT4G13960	-0.298	-	0.436	+	F-box protein family
AT4G14020	-0.389	-	-0.343	-	rapid alkalinization factor (R
AT4G14290	-0.298	-	-0.910	-	hypothetical protein
AT4G14320	-0.447	-	-1.159	-	ribosomal protein
AT4G15290	1.083	+	-0.119	0	cellulose synthase like protei
AT4G15545	-0.771	-	0.044	0	expressed protein
AT4G15590	0.606	+	-0.090	0	reverse transcriptase like pro

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT4G15740	0.750	+	0.030	0	C2 domain-containing protein
AT4G15750	-0.264	-	0.195	0	invertase/pectin methylesteras
AT4G15790	0.016	0	0.056	+	expressed protein
AT4G16190	-0.230	0	0.673	+	cysteine proteinase
AT4G16370	-0.017	0	0.095	+	isp4 like protein
AT4G16660	-0.200	0	-0.501	-	HSP like protein
AT4G17020	0.345	+	-0.704	-	transcription factor like prot
AT4G17070	-0.449	0	-0.963	-	expressed protein
AT4G17300	-0.254	-	0.411	+	asparagine–tRNA ligase
AT4G17780	-0.344	-	-0.141	0	similar to putative myb-relate
AT4G17960	-0.118	0	0.806	+	putative protein
AT4G18350	0.462	+	-0.727	-	9-cis-epoxycarotenoid dioxygen
AT4G18440	-0.479	-	0.639	+	adenylosuccinate lyase - like
AT4G18500	-0.124	0	-0.639	-	hypothetical protein
AT4G18600	-0.211	-	0.273	+	gene 11-1 protein - like
AT4G18950	-1.061	-	-0.098	0	protein kinase - like protein
AT4G19360	-0.232	-	0.159	+	putative protein
AT4G19370	-0.327	-	-0.383	-	hypothetical protein
AT4G19400	-0.081	-	-0.300	-	expressed protein
AT4G19570	1.051	+	-0.498	0	DnaJ protein family
AT4G19600	-0.338	-	0.125	+	cyclin family
AT4G19610	0.416	+	-0.135	-	RRM-containing protein
AT4G19660	0.147	+	-0.131	-	ankyrin repeat family protein
AT4G19740	0.428	+	-0.363	-	glycosyl hydrolase family 18
AT4G19960	0.388	+	0.167	+	potassium transporter-like pro
AT4G20410	0.085	0	0.954	+	gamma-soluble NSF attachment p
AT4G20810	0.587	+	0.063	0	transcription initiation facto
AT4G20850	0.390	+	0.326	+	subtilase family protein conta
AT4G20920	-0.384	-	0.304	0	double-stranded RNA-binding do

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT4G21020	0.677	+	-0.096	0	late embryogenesis abundant (L
AT4G21930	-0.388	-	0.131	+	putative protein
AT4G22100	-0.331	-	-0.835	-	glycosyl hydrolase family 1
AT4G22150	-0.497	-	-0.284	-	UBX domain-containing protein
AT4G22880	-0.484	-	0.252	+	putative leucoanthocyanidin di
AT4G22990	0.296	+	0.276	+	SPX (SYG1/Pho81/XPR1) domain-c
AT4G23130	1.403	+	0.403	0	protein kinase - like protein
AT4G23460	1.121	+	-0.205	0	beta adaptin - like protein
AT4G23620	-0.572	-	0.291	+	50S ribosomal protein-related
AT4G23920	0.926	+	-0.411	-	UDPglucose 4-epimerase - like
AT4G23990	-0.125	0	0.900	+	cellulose synthase catalytic s
AT4G24140	-0.564	-	-0.122	0	hydrolase, alpha/beta fold fam
AT4G24175	-0.225	0	-0.739	-	expressed protein
AT4G24310	-0.399	-	0.244	+	expressed protein
AT4G24670	0.344	0	1.036	+	putative alliin lyase
AT4G25200	0.067	0	-0.336	-	mitochondrion-localized small
AT4G25650	-0.609	-	0.145	0	Rieske (2Fe-2S) domain-contain
AT4G25730	-0.328	0	-0.435	-	FtsJ-like methyltransferase fa
AT4G25750	1.266	+	-0.210	0	ABC transporter family protein
AT4G25920	-0.229	-	0.200	+	expressed protein
AT4G26080	-0.218	0	0.863	+	protein phosphatase ABI1
AT4G26090	-0.028	0	0.234	+	disease resistance protein RPS
AT4G26500	-0.236	-	-0.017	0	BolA-like family protein / Fe-
AT4G26650	-0.468	0	-0.940	-	heterogeneous nuclear ribonucl
AT4G26670	0.384	+	0.658	+	mitochondrial import inner mem
AT4G26930	-1.250	-	0.374	+	myb family transcription facto
AT4G27040	0.201	+	-0.328	-	SNF8 like protein
AT4G27070	0.387	+	-0.298	-	tryptophan synthase beta-subun
AT4G27300	-0.319	-	-0.519	-	putative receptor protein kina

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT4G27340	0.095	0	-0.236	-	Met-10+ like family protein ;
AT4G27350	-0.351	-	-0.155	0	membrane lipo protein lipid at
AT4G27800	-0.905	-	0.045	0	protein phosphatase homolog (P
AT4G27810	-0.855	-	0.045	0	hypothetical protein
AT4G27950	1.383	+	0.178	+	AP2 domain transcription facto
AT4G28085	0.022	0	0.348	+	expressed protein
AT4G28170	0.672	+	-0.823	-	hypothetical protein
AT4G28670	-0.307	-	0.371	+	serine/threonine kinase-like p
AT4G28840	0.145	0	0.389	+	expressed protein
AT4G28900	-0.341	0	-0.707	-	expressed protein
AT4G28940	1.547	+	-0.286	0	nucleosidase-related contains
AT4G28980	0.643	+	0.178	0	Cyclin-dependent kinase F;1
AT4G29140	1.189	+	-0.039	0	MATE efflux protein - related
AT4G29450	-0.121	-	0.439	+	serine/threonine-specific rece
AT4G29520	2.203	+	-0.616	0	expressed protein
AT4G29760	-0.360	-	0.474	+	putative protein
AT4G30910	-1.052	-	-0.336	0	leucyl aminopeptidase - like p
AT4G31200	0.735	+	-0.005	0	SWAP (Suppressor-of-White-APri
AT4G31260	1.035	+	-0.044	0	hypothetical protein
AT4G31320	-0.072	0	0.514	+	auxin induced protein-related
AT4G31550	1.217	+	-0.179	0	putative putaive DNA-binding p
AT4G31590	-0.318	0	-0.899	-	glycosyltransferase family 2
AT4G31670	0.371	+	-0.033	0	ubiquitin carboxyl-terminal hy
AT4G32030	-0.018	0	-0.530	-	putative protein
AT4G32375	1.282	+	-0.435	0	similar to polygalacturonase,
AT4G32530	0.478	+	-0.253	-	H+-transporting ATPase - like
AT4G32540	1.253	+	-0.691	0	dimethylaniline monooxygenase
AT4G32610	1.219	+	-0.340	0	mitochondrial glycoprotein fam
AT4G32620	-0.035	-	0.173	+	putative protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT4G32780	0.036	+	-0.600	-	putative protein
AT4G32850	-0.174	-	-0.005	0	putative poly(A) polymerase
AT4G32860	0.169	+	0.083	0	putative protein
AT4G33170	-0.285	-	-0.430	-	pentatricopeptide (PPR) repeat
AT4G33350	0.382	0	-1.029	-	Tic22 -like protein
AT4G33510	-0.397	-	0.331	+	2-dehydro-3-deoxyphosphohepton
AT4G33640	-1.078	-	-0.048	0	putative protein
AT4G33830	-0.556	-	-0.759	-	glycosyl hydrolase family 10
AT4G33890	0.255	+	0.738	+	hypothetical protein
AT4G33905	0.264	+	-0.116	0	peroxisomal membrane protein,
AT4G33970	0.741	+	-0.218	-	extensin-like protein
AT4G34450	-0.729	-	-0.063	0	Nonclathrin coat protein gamma
AT4G34730	0.312	+	-0.081	0	ribosome-binding factor A fami
AT4G34760	0.357	+	0.216	0	Putative auxin-regulated prote
AT4G35090	1.255	+	0.415	+	catalase 2
AT4G35180	0.029	0	-0.243	-	amino acid permease - like pro
AT4G35220	-0.002	0	-0.355	-	cyclase family protein contain
AT4G35300	0.169	+	0.342	+	putative sugar transporter pro
AT4G35420	0.926	+	-0.007	0	dihydroflavonol 4-reductase (d
AT4G35620	1.556	+	-0.080	0	Cyclin B2;2
AT4G35790	-0.583	-	0.457	+	phospholipase D -related
AT4G37160	-0.142	-	0.044	+	pectinesterase (pectin methyle
AT4G37290	-0.072	0	-0.311	-	expressed protein
AT4G37300	2.080	+	-0.143	0	putative protein
AT4G37330	0.396	+	-0.445	-	cytochrome p450 family
AT4G37400	0.020	0	-0.656	-	cytochrome p450 family
AT4G37430	0.293	+	-0.615	-	cytochrome p450 family
AT4G37500	-0.039	0	-0.340	-	xanthine dehydrogenase family
AT4G37680	-0.474	-	0.350	0	expressed protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT4G37740	0.231	+	0.124	+	transcription activator (GRL2)
AT4G37780	-0.546	0	0.708	+	myb DNA-binding protein (AtMYB
AT4G37910	-0.916	-	-0.127	0	glycine hydroxymethyltransfera
AT4G38130	0.940	+	0.518	+	Histone deacetylase
AT4G38180	0.656	+	0.264	0	far-red impaired responsive pr
AT4G38470	-0.435	-	-0.242	-	protein kinase like protein
AT4G38550	1.071	+	-0.537	0	Phospholipase like protein
AT4G38830	-0.613	-	-0.286	0	receptor-like protein kinase -
AT4G38920	-0.396	-	0.013	0	H+-transporting ATPase 16K cha
AT4G39110	0.220	0	0.535	+	putative receptor-like protein
AT4G39580	-0.643	-	-0.024	0	Kelch repeat containing F-box
AT4G39740	-0.537	-	-0.574	-	electron transport SCO1/SenC f
AT5G01100	1.307	+	0.029	0	expressed protein
AT5G01190	0.402	+	0.089	0	laccase (diphenol oxidase), pu
AT5G01200	0.586	+	0.352	0	myb family transcription facto
AT5G01350	-0.476	-	0.700	+	expressed protein
AT5G01500	0.278	+	0.348	+	mitochondrial carrier protein
AT5G01560	0.657	+	-0.026	0	receptor like protein kinase
AT5G01650	-0.450	-	-0.289	-	Macrophage migration inhibitor
AT5G01840	0.203	+	-0.074	0	ovate family protein 59% simil
AT5G02170	-0.530	-	0.089	0	amino acid transporter family
AT5G02260	-0.010	0	0.584	+	expansin, putative
AT5G02290	0.022	0	-0.252	-	serine/threonine-specific prot
AT5G02410	-0.099	-	-0.088	-	DIE2/ALG10 family contains Pfa
AT5G02750	-0.594	-	0.057	0	C3HC4-type zinc finger protein
AT5G02780	0.165	+	-0.137	-	In2-1 protein, putative
AT5G02970	1.510	+	-0.226	0	esterase/lipase/thioesterase f
AT5G03130	1.051	+	-0.336	0	putative protein
AT5G03390	-0.517	-	0.280	+	expressed protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G03400	-0.550	-	-0.076	0	hypothetical protein
AT5G03450	-0.099	0	1.041	+	zinc finger (C3HC4-type RING f
AT5G03680	0.016	0	-0.413	-	GT2 -like protein
AT5G03780	-0.576	-	-0.065	0	myb -like protein
AT5G03850	-1.216	-	-0.538	0	RIBOSOMAL PROTEIN S28- like
AT5G03890	0.755	+	0.250	+	putative protein
AT5G04000	0.002	0	0.277	+	putative protein
AT5G04010	0.794	+	0.438	+	putative protein
AT5G04360	0.745	0	-1.100	-	pullulanase-like protein (star
AT5G04570	0.692	+	0.131	0	putative protein
AT5G04830	0.204	0	-0.803	-	expressed protein
AT5G05340	-0.263	0	-0.839	-	peroxidase, putative
AT5G05400	0.336	+	0.303	+	disease resistance protein (CC
AT5G05560	0.550	+	-0.621	-	E3 ubiquitin ligase APC1, puta
AT5G06390	-0.075	0	0.956	+	beta-Ig-H3 domain-containing p
AT5G06520	0.567	+	0.753	+	SWAP (Suppressor-of-White-APri
AT5G06550	-0.601	0	0.842	+	transcription factor jumonji (
AT5G06620	0.248	+	0.073	0	hypothetical protein
AT5G06780	-0.041	0	-0.299	-	emsy N terminus domain-contain
AT5G07060	0.094	0	0.945	+	RRM-containing RNA-binding pro
AT5G07300	-0.458	-	-0.171	0	copine-like protein
AT5G07520	1.342	+	-0.313	0	glycine-rich protein GRP18
AT5G07640	1.482	+	-0.011	0	zinc finger (C3HC4-type RING f
AT5G07720	-0.650	-	0.520	0	alpha galactosyltransferase pr
AT5G08060	1.117	+	-1.063	-	expressed protein
AT5G08230	0.087	0	0.288	+	PWWP domain protein
AT5G08310	0.400	+	0.356	+	pentatricopeptide (PPR) repeat
AT5G08330	0.616	+	-0.480	0	TCP family transcription facto
AT5G08400	0.605	+	-0.432	-	expressed protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G08560	0.549	+	-0.314	0	WD-repeat protein-like
AT5G08590	0.010	0	0.625	+	serine/threonine-protein kinas
AT5G09410	0.388	0	0.613	+	calmodulin-binding protein
AT5G09530	-0.024	0	-0.084	-	surface protein PspC-related
AT5G09570	-0.745	-	0.556	+	expressed protein
AT5G09810	-1.495	-	-0.303	0	ACTIN 2/7 (sp—P53492)
AT5G09860	0.464	+	0.735	+	nuclear matrix protein-related
AT5G09930	1.041	+	-0.231	-	ABC transporter family protein
AT5G10030	0.105	0	0.454	+	bZIP transcription factor, OBF
AT5G10300	-0.400	-	-0.097	0	alpha-hydroxynitrile lyase-lik
AT5G10450	-0.499	0	-0.802	-	14-3-3 protein GF14 lambda (gr
AT5G11180	0.846	+	-0.065	-	glutamate receptor family (GLR
AT5G11560	-0.245	-	-0.518	-	PQQ enzyme repeat-containing p
AT5G11790	-0.071	0	-0.328	-	Ndr family
AT5G12910	-0.553	-	0.072	0	histone H3 -like protein
AT5G12920	-0.185	0	0.600	+	hypothetical protein
AT5G13370	0.716	+	0.162	+	auxin reponsive - like protein
AT5G13520	-0.440	-	0.360	+	leukotriene-A4 hydrolase-like
AT5G13590	1.792	+	-0.179	0	expressed protein
AT5G13620	-0.352	-	0.849	+	putative protein
AT5G13650	-0.667	-	0.433	+	GTP-binding protein typA (tyro
AT5G13780	-0.541	-	0.256	+	N-acetyltransferase, putative
AT5G14530	-0.628	-	-0.076	0	transducin / WD-40 repeat prot
AT5G14680	-0.243	-	-0.428	-	universal stress protein (USP)
AT5G14800	-0.367	0	-0.546	-	pyrroline-5-carboxylate reduct
AT5G14830	0.214	+	0.201	+	myosin heavy chain-related
AT5G14880	1.511	+	-0.441	-	putative cation transport prot
AT5G14970	0.028	0	0.623	+	seed maturation-like protein
AT5G15470	-0.331	-	0.297	+	glycosyltransferase family 8

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G15520	-0.898	-	-0.148	0	40S RIBOSOMAL PROTEIN S19 - li
AT5G15640	-0.511	-	0.304	0	mitochondrial carrier protein
AT5G15680	1.259	+	0.078	+	expressed protein
AT5G15870	0.556	+	0.212	0	glycosyl hydrolase family 81
AT5G16050	-0.080	0	0.169	+	14-3-3 protein GF14 upsilon (g
AT5G16310	-0.131	0	0.377	+	ubiquitin C-terminal hydrolase
AT5G16380	0.115	0	-0.256	-	expressed protein
AT5G16460	0.162	0	0.226	+	putative protein
AT5G16480	-0.932	-	-0.010	0	tyrosine specific protein phos
AT5G16660	-0.637	-	0.062	0	protein; similar to unknown pr
AT5G16780	-0.528	0	-0.804	-	SART-1 family protein contains
AT5G16820	-0.331	0	-0.635	-	Heat Shock Factor 3
AT5G16920	0.254	0	0.420	+	expressed protein
AT5G17040	-0.702	-	0.634	+	UDP glucose:flavonoid 3-o-gluc
AT5G17120	-0.049	0	0.331	+	putative protein
AT5G17220	-0.718	-	1.534	+	glutathione transferase, putat
AT5G17310	0.903	+	-0.143	0	UDP-glucose pyrophosphorylase
AT5G17320	2.066	+	0.027	0	homeodomain protein
AT5G17360	1.670	+	-0.712	0	putative protein
AT5G17720	-0.570	-	-0.104	0	hydrolase, alpha/beta fold fam
AT5G17920	-1.345	-	-0.032	0	5-methyltetrahydropteroyltrigl
AT5G18090	-0.219	0	-0.463	-	transcriptional factor B3 fami
AT5G18230	0.432	+	0.143	+	transcription regulator NOT2/N
AT5G18810	0.611	+	0.102	0	arginine/serine-rich protein S
AT5G19260	0.585	+	0.034	0	expressed protein
AT5G19520	-0.053	0	0.721	+	mechanosensitive ion channel d
AT5G19940	0.539	+	-0.449	-	plastid-lipid associated prote
AT5G20000	0.882	+	-0.186	0	26S proteasome AAA-ATPase subu
AT5G21050	0.910	+	0.170	0	expressed protein

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G21090	-0.324	-	0.023	0	leucine-rich repeat protein
AT5G22030	-0.330	0	1.038	+	ubiquitin-specific protease 8
AT5G22710	-0.369	-	-0.153	-	unknown protein
AT5G23390	-0.219	-	-0.149	-	putative protein
AT5G23440	-0.127	-	0.400	+	ferredoxin-thioredoxin reducta
AT5G23640	0.889	+	0.214	+	unknown protein
AT5G23790	0.971	+	1.461	+	galactinol synthase
AT5G24370	0.444	+	-0.084	0	invertase/pectin methylesteras
AT5G24470	1.421	+	-0.023	0	pseudo-response regulator, APR
AT5G25100	0.366	+	0.025	+	putative multispanning membran
AT5G26120	0.314	+	1.178	+	glycosyl hydrolase family 51
AT5G26140	0.366	+	0.776	+	lysine decarboxylase family pr
AT5G26360	-0.155	-	-0.030	-	chaperonin gamma chain - like
AT5G26570	0.503	+	0.605	+	glycoside hydrolase starch-bin
AT5G26625	1.288	+	-0.025	0	MADS-box protein
AT5G26734	0.194	0	-0.570	-	putative protein
AT5G26980	-0.123	0	-0.331	-	similar to S-locus protein 4-l
AT5G27060	-0.022	0	-0.922	-	disease resistance protein fam
AT5G27210	-0.788	-	-0.440	-	expressed protein
AT5G27470	-0.257	0	-0.377	-	seryl-tRNA synthetase
AT5G28510	-1.143	-	1.070	+	glycosyl hydrolase family 1
AT5G28776	0.523	+	-0.657	-	similar to GB:AAC64917 gag-pol
AT5G35200	-0.299	-	-0.493	-	epsin N-terminal homology (ENT
AT5G35370	-0.094	-	0.929	+	receptor-like protein kinase -
AT5G35700	0.335	+	-0.100	-	putative fimbrin protein
AT5G35720	0.477	+	-0.167	-	similar to putative reverse tr
AT5G35740	0.868	+	-0.452	-	glycosyl hydrolase family 17
AT5G35910	-1.192	-	0.350	0	nucleolar protein-like
AT5G36110	0.035	0	0.202	+	cytochrome p450 family

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G36200	0.247	+	0.318	+	putative protein
AT5G36210	0.147	+	-0.114	-	acyl-peptide hydrolase-like
AT5G36270	-1.766	-	0.393	+	dehydroascorbate reductase, pu
AT5G37930	-0.134	0	-0.569	-	seven in absentia (sina) prote
AT5G38210	-0.474	-	-0.437	-	protein kinase - like protein
AT5G38340	-0.639	-	0.570	0	disease resistance protein (TI
AT5G38400	0.166	0	-1.352	-	unknown protein
AT5G38480	1.849	+	0.629	0	14-3-3 protein GF14 psi (grf3/
AT5G38610	-0.522	0	1.072	+	invertase/pectin methylesteras
AT5G39090	-0.618	-	-0.179	0	acyltransferase -like protein
AT5G39400	-0.396	-	-0.284	-	PTEN -like protein
AT5G39450	-0.350	-	-0.201	-	F-box protein family
AT5G39980	-0.812	-	0.209	0	pentatricopeptide (PPR) repeat
AT5G40290	0.023	0	-0.535	-	metal-dependent phosphohydrola
AT5G40590	-0.638	-	0.157	+	CHP-rich zinc finger protein,
AT5G40740	0.000	0	0.752	+	putative protein
AT5G40790	0.819	+	0.155	+	putative protein
AT5G41270	-0.233	-	0.147	0	expressed protein
AT5G41850	0.157	0	0.354	+	unknown protein
AT5G41960	-1.076	-	0.747	0	expressed protein
AT5G41970	-0.072	0	0.784	+	GAMM1 protein-like
AT5G42200	0.527	+	0.017	0	C3HC4-type zinc finger protein
AT5G42210	-0.336	0	0.634	+	putative protein
AT5G42500	0.141	0	0.773	+	disease resistance response pr
AT5G42770	0.092	0	-0.552	-	Maf family protein contains Pf
AT5G42910	0.199	0	0.436	+	bZIP protein
AT5G42980	-1.119	-	0.044	0	thioredoxin (clone GIF1) (pir—
AT5G42990	-0.130	-	0.012	0	ubiquitin-conjugating enzyme-l
AT5G43080	-0.325	-	0.207	+	Cyclin A3;1

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G43630	-0.089	0	0.423	+	zinc knuckle (CCHC-type) famil
AT5G43700	0.967	+	-0.071	0	auxin-induced protein AUX2-11
AT5G43790	-0.326	-	-0.351	-	pentatricopeptide (PPR) repeat
AT5G44130	0.613	+	-0.100	0	fasciclin-like arabinogalactan
AT5G44760	0.602	+	0.475	+	C2 domain-containing protein-l
AT5G45020	0.236	0	0.686	+	putative protein
AT5G45120	1.398	+	0.332	+	chloroplast nucleoid DNA-bindi
AT5G45250	-0.130	0	-0.666	-	disease resistance protein RPS
AT5G45260	-0.371	-	0.243	+	disease resistance protein (TI
AT5G45290	-0.575	-	0.250	0	C3HC4-type zinc finger protein
AT5G45330	-0.573	-	0.260	0	hypothetical protein
AT5G45430	0.591	+	-0.012	0	serine/threonine-protein kinas
AT5G45530	1.665	+	-0.145	0	hypothetical protein
AT5G45560	0.086	0	0.157	+	pleckstrin homology (PH) domai
AT5G45870	0.121	0	0.589	+	Bet v I allergen family
AT5G45880	0.176	0	0.851	+	Ole e I (main olive allergen)-
AT5G45900	-0.372	-	-0.377	-	ubiquitin activating enzyme E1
AT5G46430	-0.891	-	-0.640	-	ribosomal protein L32
AT5G46950	0.453	+	-0.135	0	invertase/pectin methylesteras
AT5G47180	-0.578	-	0.392	+	VAMP (vesicle-associated membr
AT5G47240	-1.102	-	0.455	+	mutT domain protein-like
AT5G47660	-0.741	-	-0.353	0	DNA-binding protein-related si
AT5G47710	0.198	0	0.786	+	C2 domain-containing protein
AT5G47760	-0.265	-	0.077	0	4-nitrophenylphosphatase-like
AT5G47770	0.576	+	-0.375	-	farnesyl pyrophosphate synthet
AT5G48070	-0.295	0	0.874	+	xyloglucan endotransglycosylas
AT5G48080	-0.238	0	0.626	+	unknown protein
AT5G48150	-0.355	-	0.214	0	SCARECROW gene regulator-like
AT5G48370	-0.001	0	-0.759	-	thioesterase family

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G48980	-0.665	-	-0.034	0	Kelch repeat containing F-box
AT5G49160	-0.083	-	-0.463	-	DNA (cytosine-5)-methyltransfe
AT5G49170	-0.207	-	0.225	+	putative protein
AT5G49380	-0.898	-	0.360	+	putative protein
AT5G49910	-0.523	-	0.228	0	heat shock protein 70 (gb—AAF2
AT5G50090	0.899	+	0.141	0	unknown protein
AT5G50570	-0.699	-	-0.477	0	squamosa promoter-binding prot
AT5G51020	-0.117	0	0.312	+	expressed protein
AT5G51170	0.451	+	0.108	0	expressed protein
AT5G51195	-0.727	-	0.946	+	similar to putative zinc finge
AT5G51500	0.427	0	1.430	+	pectinesterase family
AT5G51920	-0.162	0	0.986	+	hypothetical protein
AT5G51930	0.121	0	1.514	+	mandelonitrile lyase-like prot
AT5G52010	-0.479	-	-0.192	0	zinc finger (C2H2 type) family
AT5G52240	-0.537	-	-0.405	-	progesterone-binding protein-l
AT5G52460	-0.176	0	0.237	+	F-box protein family
AT5G53110	-0.037	-	0.325	+	expressed protein
AT5G53270	0.387	+	-0.222	0	seed maturation family protein
AT5G53300	-0.991	-	-0.136	0	ubiquitin-conjugating enzyme
AT5G53470	0.313	+	0.055	0	acyl-CoA binding protein (ACBP
AT5G53520	-0.118	0	0.572	+	isp4 protein
AT5G53920	-0.250	-	-0.633	-	ribosomal protein L11 methyltr
AT5G54130	-0.823	-	1.632	+	calcium-binding EF-hand family
AT5G54190	-0.727	-	0.427	0	protochlorophyllide oxidoreduc
AT5G54210	-0.160	-	0.325	+	NLI interacting factor (NIF) f
AT5G54420	-0.068	0	0.332	+	putative protein
AT5G54440	-0.022	0	-0.406	-	expressed protein
AT5G54560	-0.469	-	0.312	0	hypothetical protein
AT5G54820	-0.763	-	0.125	0	F-box protein family

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G54980	-0.132	-	0.423	+	integral membrane family prote
AT5G55140	-0.727	-	0.011	0	ribosomal protein L30p family
AT5G55280	-0.625	-	0.170	+	putative cell division protein
AT5G55930	-0.035	0	0.753	+	sexual differentiation process
AT5G56220	-0.562	-	0.707	+	expressed protein
AT5G56550	-0.543	-	-0.024	0	putative protein
AT5G56720	0.107	0	0.543	+	cytosolic malate dehydrogenase
AT5G56800	0.424	+	1.053	+	putative protein
AT5G56810	-0.150	0	0.899	+	F-box protein
AT5G56930	0.111	+	0.188	+	zinc finger (CCCH-type) family
AT5G57120	-0.742	-	0.309	0	putative protein
AT5G57200	0.689	+	-0.425	0	epsin N-terminal homology (ENT
AT5G57655	-0.466	-	0.192	0	xylose isomerase family protei
AT5G57790	0.955	+	-0.176	-	expressed protein
AT5G58440	-0.382	0	-0.521	-	phox (PX) domain-containing pr
AT5G58470	1.212	+	-0.233	0	RNA/ssDNA-binding protein - li
AT5G58530	0.030	0	0.745	+	glutaredoxin family protein
AT5G58770	0.502	+	0.587	+	dehydrodolichyl diphosphate sy
AT5G59500	-0.271	-	0.413	+	putative protein
AT5G59830	-0.309	-	0.049	0	putative protein
AT5G59890	-0.454	-	-0.250	-	actin depolymerizing factor 4
AT5G60560	-0.165	0	0.340	+	F-box family protein contains
AT5G60700	0.392	+	0.124	0	glycosyltransferase family 2
AT5G60790	-1.093	-	-0.233	0	ABC transporter family protein
AT5G60860	-0.970	-	-0.368	0	GTP-binding protein, putative
AT5G61120	0.486	+	-0.036	0	expressed protein
AT5G61320	0.953	+	0.152	0	cytochrome p450 family
AT5G61630	-0.593	-	0.065	0	putative protein
AT5G61840	0.296	+	0.764	+	(AtGUT1) NpGUT1 homolog

Continued on next page

Gene	WT-FC	WT-exp	Anti-fc	Anti-exp	Annotation
AT5G62080	0.111	+	-0.125	-	A9 protein precursor - like
AT5G62280	0.571	+	0.269	+	putative protein
AT5G62380	-0.138	-	0.250	+	NAC - like protein
AT5G62610	0.318	+	0.161	0	bHLH protein family
AT5G62620	-0.428	-	0.289	+	galactosyltransferase family
AT5G62700	-1.625	-	0.727	+	tubulin beta-2/beta-3 chain (s
AT5G62780	0.699	+	-0.197	0	DNAJ heat shock N-terminal dom
AT5G63340	-0.666	-	-0.104	0	unknown protein
AT5G63470	0.290	+	0.007	0	transcription factor Hap5a-lik
AT5G64120	0.459	+	-0.366	-	peroxidase, putative
AT5G64990	-0.329	-	-0.061	0	GTP-binding protein, putative
AT5G65010	-0.551	-	0.127	0	asparagine synthetase (gb—AAC7
AT5G65205	-0.665	-	0.564	+	short-chain dehydrogenase/redu
AT5G65330	-0.437	-	0.250	+	MADS-box protein
AT5G65560	0.255	0	1.178	+	pentatricopeptide (PPR) repeat
AT5G65930	-1.010	-	0.885	+	putative kinesin calmodulin-bi
AT5G65980	-0.272	-	-0.625	-	auxin efflux carrier protein f
AT5G66040	0.266	+	0.415	+	senescence-associated protein
AT5G66150	2.352	+	-1.159	0	glycosyl hydrolase family 38
AT5G66790	-0.153	-	-0.205	-	protein kinase family
AT5G66800	-0.354	0	0.957	+	expressed protein
AT5G66860	0.336	0	0.682	+	expressed protein
AT5G66870	0.269	+	1.203	+	lateral organ boundaries (LOB)
AT5G67000	0.012	0	-0.540	-	AP2 domain transcription facto
AT5G67020	0.862	+	-0.220	-	hypothetical protein
AT5G67090	0.591	+	0.432	+	subtilisin-like serine proteas
AT5G67190	0.277	0	0.597	+	TINY-like protein
AT5G67630	1.175	+	-0.240	-	RuvB DNA helicase-like protein

Appendix B

PathPrime Perl Code

B.1 Perl Code: PathPrime

```
#!/usr/bin/perl -w
##############################################################################
# Program name: PathPrime
# Author: Shrinivasrao P. Mane
# Date (Begin):12/14/06
# Description:
#
##############################################################################



##############################################################################
# DISCLAIMER:
# This software comes with no guarantee of usefulness.
# Use at your own risk. Check any solutions you obtain.
# S. P. Mane assumes no responsibility for the use of this software.
##############################################################################


=pod

=head1 NAME

PathPrime - A tool for designing and testing primers for disease diagnostics

=head1 VERSION
```

```
Version 1.00, 30 December, 2006



=head1 USAGE

PathPrime (-vh) [-n BackgroundList.txt] [-i InFile.fasta] [-o OutFile.txt]
[-c config.conf]

-i File with sequences in fasta format 'InputFile.fasta'.

-o File with primers and their sequences 'Output.txt'.

-n Non-Target Contaminant Organisms' List 'BackgroundList.txt'
(optional).

-c Configuration file

-h Help: print this help message.

-v Turn on verbose output.

NOTE: Type perldoc $0 for documentation.


=head1 DESCRIPTION

PathPrime can design, test and identify a minimum set of primers to amplify
a given list of sequences.

=cut

use strict;
use Getopt::Std;
getopts('i:n:o:c:hv');
our($opt_i, $opt_n, $opt_o, $opt_h, $opt_v,$opt_c);
my $file=$opt_i;
my $Bkd_ePCRFileList=$opt_n;
my $outFile=$opt_o;
my $configFileName=$opt_c;
my $doBkgdTest="yes"; # To test for non-target primer annealing
my %config;

#Print detailed help message
HelpMsg() if defined $opt_h;
```

```
#Print usage
if(!defined $opt_i or !defined $opt_o)
{
die Usage();

}

my $verbose=0;
$verbose=1 if defined $opt_v;
$doBkgdTest="no" if !defined $opt_n;

## check if all necessary binaries are installed
#  Else warn and exit
checkPrograms();

## Print message (Brag!)
printMessage();

## Load config file if used
#
readConfigFile($configFileName) if defined $opt_c;

################################
# Declare global vairables  #
################################

# BLASTCLUST VARIABLES
my $cl_output;
my $cl2cnt=0; # counts # of genes per cluster
my @cl2files=""; # holds list of clusters with at least 2 sequences
my $consen_file="consensus.fasta"; # do not change this

# Run NCBI's blastclust
run_blastclust();

# open the cluster file get the sequences for each gene in the cluster and
# save each cluster in a separate file
getClusterSeq();

# build alignments for all clusters that have at least 2 sequences and save
# the consensus sequences
makeAlignment() if $cl2cnt!=0;

# Design primers using primer3
runPrimer3($consen_file);
```

```
my $HashFileName=makePCRtemplate($file);

# Run ePCR on target genes
my $IfTargetSeq="yes"; # Make sure that it runs ePCR on target genes
run_ePCR($HashFileName,$IfTargetSeq);

# Run ePCR on background/non-target genes
if($doBkgdTest eq "yes"){
$IfTargetSeq="no";
run_ePCR($Bkd_ePCRFileList,$IfTargetSeq);
}

# Find minimal primer set
findMinimumPrimerSet($outFile);

exit;




########################################
#       SUBROUTINES       #
########################################


# Print Usage
sub Usage{
return "Usage: $0 [-n BackgroundList.txt] [-i InFile.fasta] " .
   "[-o OutFile.txt]\n$0 -h for help message\n";
}




# Print detailed help message
sub HelpMsg{
print STDERR Usage();
print STDERR <<EOM;
-i File with sequences in fasta format 'Sequences.fasta'.
-o File with primers and their sequences 'output.txt'.
-n Non-Target Contaminant Organisms' List 'backgroundlist.txt'
             (optional).
-c Configuration file
-h Help: print this help message.
```

```
-v Turn on verbose output.

NOTE: Type perldoc $0 for documentation.
EOM
exit(0);

}

## Reads parameters from the config file
#
sub readConfigFile{
my $ConfigFile=shift;
print STDERR "Reading config file            ... ";
open(CONF,$ConfigFile) or die "Configuration file ". $ConfigFile .
" not found\n $!";
while(<CONF>){
next if m/^#/;
next if m/^$/;
chomp;
m/(.+)=(.+)/;
my $k = $1;
my $v = $2;
$k =~s/\s+//g;
$v =~s/\\/\//g;
$v =~s/\s+//g;
$config{$k}=$v;
#print "$k\t$v\n";
}
print STDERR "done.\n";
}


=pod

=head1 SUBROUTINES

=head2 run_blastclust

This subroutine takes the fasta file and clusters them using NCBI bastclust.
The resulting output file is stored as 'blastclust_95.txt'.

=cut

## Run NCBI's blastclust
#
sub run_blastclust{
```

```
if($verbose==1){
print STDERR "\n##################################################\n";
print STDERR "Clustering sequences using BLASTCLUST\n";
print STDERR "##################################################\n";
print STDERR "Clustering sequences (please have patience, \n".
"it may take a while) ... ";
}else{
print STDERR "Clustering sequences           ... ";
}
my ($ncpu,$similarity,$length,$both);
my $blastclust="blastclust"; # program name
$ncpu=$config{"NumberOfCPU"} or $ncpu="1"; # you can change the # of
# processors here (useful if you have a multiprocessor computer)
$similarity=$config{"ScoreCoverageThreshold"} or $similarity="98";
$length =$config{"LengthCoverageThreshold"} or $length ="0.9";
$both=$config{"BothNeighbours"} or $both="T";
$cl_output="blastcluster_"
. $similarity . ".txt";  # name of the cluster output file


my $blastclustx="\"$blastclust\" -i \"$file\" -S \"$similarity\" " .
"-L \"$length\" -b \"$both\" > \"$cl_output\"";
my $blastclustOut=`$blastclustx 2>&1`;
print STDERR "done.\n";
print STDERR "The cluster file is saved in $cl_output\n" if $verbose==1;
}

=pod

=head2 getClusterSeq

This subroutine reads output from run_blastclust (i.e., 'blastclust_95.txt')
and fetches sequence for each gene in the cluster. It stores each cluster in
a separate file. (e.g., Cluster_001.txt).  This subroutine uses getFastaSeq()
and IUPAC2gestalt().

=cut

## open the cluster file get the sequences for each gene in the cluster and #
## save each cluster in a separate file    #
#
sub getClusterSeq{
open (CLUST,$cl_output) or die "Could not open $cl_output . Where did it go?!\n";
chomp(my @clusters=<CLUST>);
close(CLUST);
my $clustcnt=1; # counts the # of clusters
```

```
my $clname="clust_"; # prefix for fasta files of each cluster
my $sno=""; # holds printf formatted number
my $clfname=""; # holds formatted cluster filename
my $fastaseq=""; # holds fasta sequence


foreach my $cluster (@clusters){
# skip the header line
next if ($cluster=~/(Start clustering|\[NULL)/);
my @genes= split(" ", $cluster);
my $numGenes=@genes; # count the # of genes in each cluster
$sno= sprintf("%03d",$clustcnt);
$clfname=$clname.$sno. ".txt";
if ($numGenes>=2){
$cl2files[$cl2cnt]=$clfname; # store clusters with >=2 genes
$cl2cnt++; # count clusters with >=2 genes
}
if($verbose==1){
print STDERR "Cluster_$sno ($numGenes genes)\n" if $numGenes>1;
print STDERR "Cluster_$sno ($numGenes gene)\n" if $numGenes==1;
}

foreach my $gene (@genes){
# get Fasta sequences and write to the cluster files
getFastaSeqBio($gene,$file,$clfname);
if ($numGenes<2){
getFastaSeq($gene,$file,$consen_file)
}
}
#print "\n";
$clustcnt++;
}
print STDERR "There are a total of ", $clustcnt-1," clusters and ".
$cl2cnt ." clusters have at least 2 genes\n" if $verbose==1;
}

=pod

=head2 getFastaSeq

This subroutine fetches the sequence of a gene.

=cut


# GET FASTA SEQUENCES
```

```
## This is my version
#
sub getFastaSeq{
my ($gene, $file, $out)=@_;
open( FILE, "< $file" ) or die "Can't open $file : $!";
open( OUT, ">> $out" ) or die "Can't open $out : $!";
my $keyword =$gene;
my $match=0;
while( <FILE> ) {
if (/>.*$keyword/i){
chomp;
print OUT "\n$_\n";  # print the gene info/header line
$match=1;  # turn flag ON
next  # go to next line
}
# Turn flag OFF if the beginning of a new sequence is encountered
$match=0 if />/;
if ($match==1){
chomp;
s/[a-z]/n/g;
s/[BDEFHIJKLMNOPQRSUVWXYZ]/N/g;
print OUT "$_";
}
}
close FILE;
close OUT;
}

## this is Bioperl version
#
sub getFastaSeqBio{
my ($id,$infile,$outfile)=@_;
use Bio::SeqIO;
my $in = Bio::SeqIO->new('-file' => "$infile",'-format' => 'fasta') ;
my $out = Bio::SeqIO->new('-file' => ">>$outfile",'-format' => 'fasta');
while(my $seq_obj = $in->next_seq){
if ($seq_obj->id()=~m/$id/){
$out->write_seq($seq_obj);
}
}

}

=pod
```

```
=head2 makeAlignment

This subroutine takes all clusters that have at least 2 sequences and builds
alignments. The resulting consensus sequence is processed to replace all
degenerate nucleotide notations with 'N' and stored in 'consensus.fasta'.
Clusters with single gene are stored in 'consensus.fasta' without further
processing. The subroutine uses runMuscle().

=cut


## Build alignments for all clusters and save the consensus sequences
#
sub makeAlignment {
open(CONSENSUS,">>$consen_file") or die "Cannot write to $consen_file\n";
if($verbose==1){
print STDERR "\n##################################################\n";
print STDERR "Aligning sequences in each cluster using MUSCLE\n";
print STDERR "##################################################\n";
}else{
print STDERR "Aligning sequences             ... "
}

print CONSENSUS "\n";
foreach my $cl2file (@cl2files){
my $cl2aln=$cl2file;
$cl2aln=~s/txt/aln/i;
my $clustName=$cl2file;
$clustName=~s/\.txt//;
$clustName=~s/clust_/Cluster /;
print "Building alignment for $clustName ..." if $verbose==1;
# write output to files and get consensus sequences
my $consensus= runMuscle($cl2file,$cl2aln);
print " done.\n" if $verbose==1;
# convert all consensus notations to "N"
my $gestalt=IUPAC2gestalt($consensus);
my $fasta_header=$cl2file;
$fasta_header=~s/\.txt//i;
print CONSENSUS "\>$fasta_header\n$gestalt\n";
}
close(CONSENSUS);
print STDERR "done.\n" if $verbose==0;
}

=pod
```

```
=head2 runMuscle

This subroutine builds multiple sequence alignments using MUSCLE.

=cut

## runMuscle takes input file, aligns and saves the output file
## it returns consensus sequence
#
sub runMuscle{
my ($infile, $outfile)=@_;
use Bio::Tools::Run::Alignment::Muscle;
use Bio::AlignIO;
my $alignout = new Bio::AlignIO('-file' => ">$outfile",'-format' => 'clustalw');
my $muscle = new Bio::Tools::Run::Alignment::Muscle(quiet => 1,
LOGA=>'muscle.log');
my $aln = $muscle->align($infile);
my $iupac_consensus = $aln->consensus_iupac();
$alignout->write_aln($aln);
return $iupac_consensus;
}

=pod

=head2 IUPAC2gestalt

This subroutine replaces all non-ATGC character with 'N's in a DNA sequence.

=cut


#Replaces all non-ATGC characters with N
sub IUPAC2gestalt{
my ($seq)=@_;
$seq=~s/[a-z]/n/g;  # convert partially conserved nucleotides to "n"
$seq=~s/[BDEFHIJKLMNOPQRSUVWXYZ]/N/g;  # convert all non-ATGC notations to "n"
return $seq;
}

=pod

=head2 runPrimer3

This subroutine designs five primer-pairs for each (consensus) sequence in
'consensus.fasta' according to the primer design parameters specified by the
user. The results are stored in 'primer_list_epcr.txt'. The details of primer
```

```
design are logged in 'primer3.log'.

=cut

## Takes a list of sequences and design 5 primers per sequence
#
sub runPrimer3{
my ($ConsenusFile)=@_;
my $totalPrimerCnt=0;
if($verbose==1){
print STDERR "\n################################################\n";
print STDERR "        Designing primers using Primer3\n";
print STDERR "################################################\n";
}else{
print STDERR "Designing primers              ... ";
}

use Bio::Tools::Run::Primer3;
use Bio::SeqIO;
my $seqio = Bio::SeqIO->new('-file'=>"$ConsenusFile");
my $cntSeq=0;
my ($primerMinTm, $primerMaxTm, $PrimerOptSize,$Primer3EndGC,
$AmpLenRange, $AmplOptLen);
$primerMinTm=$config{"PrimerMinTemp"}        or $primerMinTm=60;
$primerMaxTm=$config{"PrimerMaxTemp"}        or $primerMaxTm=70;
$PrimerOptSize=$config{"PrimerOptSize"}      or $PrimerOptSize=20;
$Primer3EndGC=$config{"Primer3EndGC"}        or $Primer3EndGC=0;
$AmpLenRange=$config{"AmpliconLengthRange"}  or $AmpLenRange="100-400";
$AmplOptLen=$config{"AmpliconLengthOptimum"} or $AmplOptLen=300;
while (my $seq = $seqio->next_seq){#;
$cntSeq++;
my $primer3 = Bio::Tools::Run::Primer3->new(-seq => $seq,
-outfile => ">primer3.log");
$primer3->program_name('primer3_core');
unless ($primer3->executable) {
print STDERR "primer3 can not be found. Is it installed?\n";
exit(-1);
}
 my $args = $primer3->arguments;
 $primer3->add_targets(
'PRIMER_MIN_TM'             => $primerMinTm,
'PRIMER_MAX_TM'             => $primerMaxTm,
'PRIMER_PRODUCT_OPT_SIZE'   => $AmplOptLen,
'PRIMER_PRODUCT_SIZE_RANGE' => $AmpLenRange,
'PRIMER_GC_CLAMP'           => $Primer3EndGC,
'PRIMER_OPT_SIZE'           => $PrimerOptSize
```

```
);
my $results = $primer3->run;
    # get specific results
    my ($numres)=$results->number_of_results;
    my $all_results=$results->all_results;
    #get sequence ID of the gene
my ($seq_id)=${$all_results}{'PRIMER_SEQUENCE_ID'};
#print results in a file
open (EPCR,'>>primer_list_epcr.txt')  or
die 'could not write to file\n';
for(my $i=0;$i<$numres;$i++){
my $tab_res=$results->primer_results($i);
print EPCR $seq_id, "_", $i+1, "\t",
 ${$tab_res}{'PRIMER_LEFT_SEQUENCE'},"\t" ,
 ${$tab_res}{'PRIMER_RIGHT_SEQUENCE'},"\t100-800\n";
}
close(EPCR);
if($verbose==1){
print STDERR "Could not find primers for $seq_id\n"
 if $results->number_of_results==0;
}

 $totalPrimerCnt=$totalPrimerCnt+$results->number_of_results;
}
if($totalPrimerCnt==0){ # Exit if no primers are found
print "\nCould not find any primers.\n" .
  "Change some parameters and re-run $0\n\n";
exit;
}
print STDERR "In all $totalPrimerCnt primer pairs were designed on " .
 $cntSeq ." sequences\n" if $verbose==1;
print STDERR "done.\n"
}

=pod

=head2 makePCRtemplate

This subroutine builds PCR template files for target genes: <filename>.famap
and <filename>.hash. These files are required to run e-PCR.

=cut

## make template files for e-PCR of target genes
#
sub makePCRtemplate{
```

```
my $file=shift;
use Cwd;
use File::Path;
use File::Basename;
my $basename = basename($file);
    my $dirname  = dirname($file);
my $currWorkDir = &Cwd::cwd();
my $map_file=$basename;
$map_file=~s/fa.*$/famap/;
my $hash_file=$basename;
$hash_file=~s/fa.*$/hash/;
my $newDir="target_epcr";
my $fullPath=$currWorkDir."/".$newDir;
&File::Path::mkpath($fullPath);
my $fullFastaFileName=$currWorkDir .
   "/" . $dirname . "/". $basename;
my $fullMapName=$fullPath . "/" . $map_file;
my $fullHashName=$fullPath . "/" . $hash_file;

#The programs (famap and fahash) need fullpath for the filenames
print STDERR "Preparing E-PCR template       ... ";
my $famap="famap -tN -b $fullMapName $fullFastaFileName";
#capture a command's STDERR but discard its STDOUT
my $famapOut=`$famap 2>/dev/null`;
print STDERR "$famapOut" if $?!=0;
my $fahash="fahash -b $fullHashName -w 12 -f3 $fullMapName";
my $fahashOut=`$fahash 2>/dev/null`;
print STDERR "$fahashOut" if $?!=0;
print STDERR "done.\n";
return $fullHashName;
}

=pod

=head2 run_ePCR

This subroutine takes 'primer_list_epcr.txt' from runPrimer3() and performs
in-silico/electronic PCR to verify the primers against the target sequences.
The results are stored in 'target_result.epcr'. The input file
('primer_list_epcr.txt') can also be used to test against non-target/conta-
minant organism sequences. In that case, the results are stored in
'OrganismName_bkgd.epcr'.

=cut

## Run e-PCR on target and non-targets
```

```
#
sub run_ePCR{
  my($fullname,$IfTargetSequence)=@_;

  if($IfTargetSequence eq "yes"){
   if($verbose==1){
   print STDERR "\n#################################################\n";
   print STDERR "      Running In-silico PCR on Target genes\n";
   print STDERR "#################################################\n";
   system("re-PCR -S $fullname -n0 -g0 -o target_result.epcr
primer_list_epcr.txt");
   print "Done.\n";
   }else{
   print STDERR "Performing in-silico PCR       ... ";
   my $repcr=`re-PCR -S $fullname -n0 -g0 -o target_result.epcr
 primer_list_epcr.txt 2>/dev/null`;
   print STDERR "done.\n";

   }



  }else{
   print STDERR "\n#################################################\n";
   print STDERR "   Running In-silico PCR on non-target genes\n";
   print STDERR "#################################################\n";
   open(BKGFILE,$fullname) or die "Cannot open file: $fullname $!";
   my ($mismatch,$gap);
   $mismatch=$config{"Mismatch"} or $mismatch=2;
   $gap=$config{"Gap"} or $gap=2;
   while(<BKGFILE>){
   next if /\#/;
   chomp;
   my($OrganismName,$FileName)=split /\t/;
   print STDERR "Running e-PCR on background: $OrganismName\n";
   $OrganismName=~s/\s+/\_/g;
   $OrganismName=$OrganismName."_bkgd.epcr";
   system("re-PCR -S $FileName -n $mismatch -g $gap -o
$OrganismName primer_list_epcr.txt");
   print STDERR "Done.\n";
   }
   close BKGFILE;
  }
}

=pod
```

```
=head2 findMinimumPrimerSet

This subroutine takes 'target_result.epcr' and counts the number of target
sequences amplified by each primer-pair. It then reads all
'OrganismName*_bkgd.epcr' files and removes any primer-pair that can
potentially amplify non-target genes. The resulting primer list is used to
find the minimum set of primers that can amplify all sequences. The details
of findMinimumPrimerSet() are logged in 'minimal_set.log'. The results are
saved in user-specifed output filename.

=cut

sub findMinimumPrimerSet{
my ($outFile)=@_;
my $infile="target_result.epcr";
my $primerSeqInfoFile="primer_list_epcr.txt";
use lib '/home/smane/.perlmod';
use Array::Unique; # Holds unique elements
use Tie::DxHash; # Allows duplicate keys and preserves their order
use Algorithm::SetCovering; # to identify a minimum set of primers
tie my @p, 'Array::Unique'; # holds primer names (unique list)
tie my @g, 'Array::Unique'; # holds gene names (unique list)
my(%pg);  # this holds the primers and name of each amplified gene
tie %pg, 'Tie::DxHash' ; # hash variable hold duplicate primer names
my ($alg);
my (%p_cnt); # this holds the primers and the no. of amplified genes
my (%pr_seq); # holds primer name, forward and reverse sequences
my @row;
my @to_be_opened;
my %bkgdAmplifyingPrimers; # primers that amplify "non-target" genes
my @targetAmplifyingPrimers; # primers that amplify "target" genes
my $min_setfinderLogFile="minimal_set.log";
if($verbose==1){
print STDERR "################################################\n";
print STDERR "         Finding Minimal Primer Set\n";
print STDERR "################################################\n";
}else{
print STDERR "Finding minimal set of primers ... ";
}
# read results from non-taget e-PCR
my @bkgEPCRresultFiles=glob("*_bkgd.epcr");
foreach my $bkgdGenes(@bkgEPCRresultFiles){
open (NTFILE, $bkgdGenes) or die "can't open file\n";
while(<NTFILE>){
chomp;
```

```
if ($_!~/^\#/){
my ($primer)=split /\t/;
$bkgdAmplifyingPrimers{$primer}=$primer;
}
}
}
open (FILE, $infile) or die "Could not open file: $infile\n";
while(<FILE>){
chomp;
next if /^#/;
push @targetAmplifyingPrimers,$_;
}
close(FILE);
@targetAmplifyingPrimers=sort  @targetAmplifyingPrimers;
my $prev="";
my $cnt=0;
my $printprev="";

foreach(@targetAmplifyingPrimers){
chomp;
my   ($pr,$ge)=split /\t/;
next if exists $bkgdAmplifyingPrimers{$pr};
push (@p,$pr);
push (@g,$ge);
$pg{$pr}=$ge;
if($prev eq $pr){
$cnt++;
$printprev= "$prev\t$cnt\n";
# store primer and no. of amplified genes
$p_cnt{$prev}=$cnt if $cnt>0;
}else{
# store primer and no. of amplified genes
$p_cnt{$prev}=$cnt if $cnt>0;
$cnt=1;
$prev = $pr;
}

}

my $nump=@p;
# To store detailed result incl. intermediate results (for debugging)
open(LOG,">$min_setfinderLogFile") or
 die "Cannot write to log file $! \n";
# Stores the final output (summary)
open(FINALOUT,">$outFile") or die "Cannot write to $outFile $! \n";
my ($sec,$min,$hour,$mday,$mon,$year,$wday,$yday,$isdst)=localtime(time);
```

```
print FINALOUT "PathPrime\tRun Date: ";
printf FINALOUT "%4d-%02d-%02d %02d:%02d:%02d\n\n",
$year+1900,$mon+1,$mday,$hour,$min,$sec;
print STDERR "Total number of primers designed :\t$nump\n"
if $verbose==1;
print LOG "Total number of primers designed :\t$nump\n";
print FINALOUT "Total number of primers designed :\t$nump\n";

my @sortedGeneList=sort @g;
my $len=@sortedGeneList;
print STDERR "Number of target genes amplified :\t$len\n\n"
if $verbose==1;
print LOG "Number of target genes amplified :\t$len\n\n";
print FINALOUT "Number of target genes amplified :\t$len\n\n";
$alg = Algorithm::SetCovering->new(columns => $len, mode  => "greedy");
# another option for mode is "brute_force"
# Use it for really small matrix: O(2^n-2)
my ($checker,$row2add);
for (my $i=0;$i<$len;$i++){
if($i==$len-1){
$checker=$checker. "$sortedGeneList[$i]";
}else{
$checker=$checker. "$sortedGeneList[$i]\t";
}
}
print LOG "Primer:\t$checker\n";
@row=split("\t",$checker);
@to_be_opened = ((@row));
my $g_cnt=0;
my $pg_cnt=0;
foreach my $primer (@p){
for(my $g_cnt=0;$g_cnt<=$len;$g_cnt++){
my $amplifiedGene=$pg{$primer};
#print "$amplifiedGene\n";
if($g_cnt<$len && $amplifiedGene eq $sortedGeneList[$g_cnt]){
if($g_cnt==$len-1){
$row2add=$row2add."1";
}else{
$row2add=$row2add."1\t";
}
}else{
while($g_cnt<$len){
if($amplifiedGene ne $sortedGeneList[$g_cnt]){
if($g_cnt==$len-1){
$row2add=$row2add."0";
}else{
```

```
$row2add=$row2add."0\t";
}
$g_cnt++;
}
if($g_cnt<$len &&
$amplifiedGene eq $sortedGeneList[$g_cnt]){
if($g_cnt==$len-1){
$row2add=$row2add."1";
}else{
$row2add=$row2add."1\t";
}
$g_cnt++;
$amplifiedGene=$pg{$primer};
#print "$amplifiedGene\n";
}
}
}

}

print LOG "$primer:\t$row2add\n";
@row=split("\t",$row2add);
$alg->add_row(@row);
#Flush the variables
$row2add="";
@row="";
}

# Store forward and reverse primer seq. for each primer-pair in a Hash
open(PRSEQ,$primerSeqInfoFile) or
 die "cannot find primer sequence file $!\n";
while(<PRSEQ>){
chomp;
next if /^\#/;
my ($pri,$ForwardSeq,$ReverseSeq)=split /\t/;
$pr_seq{$pri}="$ForwardSeq\t$ReverseSeq";
}
close(PRSEQ);

my @minimumPrimerSet = $alg->min_row_set(@to_be_opened);
print STDERR "done.\n" if $verbose==0;
print STDERR "To amplify all genes, we need ", scalar @minimumPrimerSet,
 " primer(s):\n" if $verbose==1; # primers needed to amplify all genes
print LOG "To amplify $len genes, we need ", scalar @minimumPrimerSet,
" primer(s):\n"; # no.of primers needed to amplify all genes
print FINALOUT "To amplify $len genes, we need ", scalar
```

```
 @minimumPrimerSet, " primer(s):\n";# primers needed to amplify all genes
# print the min.primer set with seq. and the no. of genes they amplify
print STDERR "Primer ID\tForward Primer\tReverse Primer\t" .
"No. of sequences amplified\n" if $verbose==1;
print FINALOUT "Primer ID\tForward Primer\tReverse Primer\t" .
"No. of sequences amplified\n";
print LOG "Primer ID\tTracking ID\tForward Primer\tReverse Primer\t" .
"No. of sequences amplified\n";
my $ppCnt=0;
for (@minimumPrimerSet) {
my $prim=$p[$_];
$ppCnt++;
print STDERR "PP$ppCnt\t$pr_seq{$prim}\t$p_cnt{$prim}\n"
if $verbose==1; # print primer list
print FINALOUT "PP$ppCnt\t$pr_seq{$prim}\t$p_cnt{$prim}\n";
print LOG "PP$ppCnt\t$prim\t$pr_seq{$prim}\t$p_cnt{$prim}\n";
}
close LOG;
print STDERR "\nThe results are stored in \"". $outFile."\"\n";

}

## Check if all necessary binaries are installed and/or are in system path
#
sub checkPrograms{
my $os=$^O;
my @progs;
if ($os=~/MSWin32/i){
@progs =
qw/muscle.exe primer3.exe
blastclust.exe famap.exe fahash.exe re-PCR.exe/;

}else{
@progs=qw/muscle primer3_core blastclust famap fahash re-PCR/;
}
my @prog_not_installed;
my $cnt=0;
if ($os=~/MSWin32/i){
foreach my $prog(@progs){
my $out=`$prog`;
if ($out=~m/internal or external command$/i){
$cnt++ ;
push @prog_not_installed,uc $prog
}
#print "$prog:\t$cnt\t$out\n"
}
```

```
}else{

foreach my $prog(@progs){
my $out=`which $prog`;
if ($out=~m/^no /){
$cnt++ ;
push @prog_not_installed,uc $prog
}
}
}
if ($cnt>0){
print STDERR "Cannot proceed! ".
"The following programs are not installed:\n";
print STDERR join "\n", @prog_not_installed,"\n";
print STDERR "Please install them before you proceed." .
 "Or add them to the system path\n";
exit;
}

}

sub printMessage{
print STDERR
"\n**************************************************************************
***********\n";
print STDERR "Program: $0\n";
print STDERR "Version: 1.0 -- Jan 2007\n";
print STDERR
"Purpose: Design and In-silico Testing of Primers for Pathogen Diagnostics\n";
print STDERR "Author : Shrinivasrao P. Mane\n";
print STDERR "Contact: smane\@vbi.vt.edu\n";
print STDERR
"Adrress: Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA, USA\n";
print STDERR
"**************************************************************************
***********\n\n";

}

=pod

=head1 REQUIREMENTS

PathPrime has been tested on Mac OS X and should work on any Unix-like
operating system. The following software tools and modules are required to
install and run PathPrime.
```

```
=over 4

=item * BLASTCLUST: This program is a part of NCBI local blast package.

=item * e-PCR: e-PCR provides a search mode using primer sequences against a
sequence database.

=item * MUSCLE: A tool for clustering nucleic acid and protein sequences.

=item * Primer3: Primer3 is a widely used program for designing PCR.

=item * Perl 5.8.6 or higher.

=back

The following Perl modules are needed (available from CPAN):

=over 4

=item * C<Algorithm::SetCovering>: This module is used to identify a minimum
set of primer-pairs.

=item * C<Array::Unique>: This module lets you create an array which will
allow only one occurrence of any value.

=item * BioPerl: Two packages from BioPerl, Core and Run packages, are
needed. Core package provides functions for basic handling and manipulation
of sequences. Run packages provide wrappers for executing common
bioinformatics  applications. The following BioPerl modules are used by
PathPrime:  C<Bio::SeqIO>, C<Bio::Tools::Run::Alignment::Muscle>,
C<Bio::Tools::Run::Primer3>

=item * C<Tie::DxHash>: C<Tie::DxHash> keeps insertion order and allows
duplicate hash keys.

=back

=head1 AUTHOR

Shrinivasrao P. Mane, E<lt>smane@vbi.vt.eduE<gt>

L<https://patric.vbi.vt.edu/>
```

B.2 Perl Code: makePCRtemplate

```
#!/usr/bin/perl -w

=pod

=head1 NAME

makePCRtemplate.pl is a part of PathPrime

PathPrime - A tool for designing and testing primers for disease diagnostics

=head1 VERSION

Version 1.00, December, 2006


=head1 USAGE

Usage: makePCRtemplate.pl  [-i InFile.fasta] [-d Output directory]

-i File with sequences in fasta format 'Sequences.fasta'.

-d Full directory name where output files should be saved.

-h Help: print this help message.

-v Turn on verbose output.

NOTE: Type perldoc makePCRtemplate.pl for documentation.


=head1 DESCRIPTION

This subroutine builds PCR template files for Non-target genes:
<filename>.famap and <filename>.hash.

These files have to made if you want to run e-PCR on non-target genes.

=head1 AUTHOR

Shrinivasrao P. Mane, E<lt>smane@vbi.vt.eduE<gt>

L<https://patric.vbi.vt.edu/>

=cut
```

```
use Cwd;
#use File::Path;
use Getopt::Std; # for switch and argument processing
getopts('i:d:hv');
our ($opt_h, $opt_i, $opt_d,$opt_v);
my $file=$opt_i;
my $outDir=$opt_d;

#Print detailed help message
HelpMsg() if defined $opt_h;

if(!defined $opt_i or !defined $opt_d){
die Usage();
}
my $verbose=0;
$verbose=1 if defined $opt_v;


my $currWorkDir = &Cwd::cwd();
my $map_file=$file;
$map_file=~s/fa.*$/famap/ or die "File name end in .fa or .fasta";
my $hash_file=$file;
$hash_file=~s/fa.*$/hash/;
my $fullFastaFileName=$currWorkDir . "/" . $file;
my $fullMapName=$outDir . "/" . $map_file;
my $fullHashName=$outDir . "/" . $hash_file;


print "Preparing E-PCR Map  template ... ";
my $famap="famap -tN -b $fullMapName $fullFastaFileName";
my $famapOut=`$famap 2>/dev/null`;
print $famapOut if $verbose==1;
print "Done.\n";

print "Preparing E-PCR Hash template ... ";
my $fahash="fahash -b $fullHashName -w 12 -f3 $fullMapName";
my $fahashOut=`$fahash 2>/dev/null`; #capture a command's STDERR but discard its STDOUT 2>/dev/nu
print $fahashOut if $verbose==1;
print "Done.\n";

print "The template hash file is stored in:\n$fullHashName\n\n";

# Print Usage
```

```
sub Usage{
return "Usage: $0  [-i InFile.fasta] [-d Output directory]\n$0 -h for help message\n";
}

# Print detailed help message
sub HelpMsg{
print Usage();
print <<EOM;
-i File with sequences in fasta format 'Sequences.fasta'.
-d Full directory name where output files should be saved.
-h Help: print this help message.
-v Turn on verbose output.

NOTE: Type perldoc $0 for documentation.
EOM
exit(0);

}
```

Appendix C

PathPrime configuration files

C.1 Primer Design parameters for Lyssaviruses

```
## This is the configuration file for PathPrime
# NOTE: All the default values used here built into the program by default.
# You don't have to use this file unless you want to change the default
# parameters used below



################################################################################
##                      PathPrime related options                             ##
################################################################################


# File where consensus sequences are stored

#ConsensusFastaFile = consensus.fasta

################################################################################
##                      NCBI Blastclust related options                       ##
################################################################################

# NOTE: For details of the options used below, see the NCBI BLAST manual
# Not all options have been used

# Specify the number of CPUs to use on a multiprocessor machine
# if you have a Dual core processor, you can set it to 2
# option -a in blastclust
# % $ncpu="1"

NumberOfCPU = 2
```

```
# Score coverage threshold (bit score / length if < 3.0, percentage of
# identities otherwise). Please refer to the BLAST manual to change this one
# minimum identity between nucleotides to form a cluster
# Reduce this number if you want to get a smaller list of primers. However,
# you may not be able to get any primers for a cluster if has little
# consensus.
# option -S in blastclust
# % $similarity="98";

ScoreCoverageThreshold = 85

# Length coverage threshold
# option -L
# % $length ="0.9";

LengthCoverageThreshold = 0.75

# Require coverage on both neighbours? [T/F]. If true, the program requires
# both sequences to pass the coverage criteria set by option -L before they
# are called neighbours and clustered together
# % $both="T"

BothNeighbours = T

##############################################################################
##                          Primer 3 related options                        ##
##############################################################################

# NOTE: For details of the options used below, see the Primer3 manual
# Not all options have been used

# set the minimum primer melting temperature (Tm)

PrimerMinTemp = 50

# set the maximum primer melting temperature (Tm)

PrimerMaxTemp = 70

# set the optimum primer melting temperature (Tm)
# % 'PRIMER_OPT_TM'=>' 60

PrimerOptTemp = 60

# Optimal primer size. Primers will be close to this value in length
```

```
# % 'PRIMER_OPT_SIZE'=>22

PrimerOptSize = 20

# Number of Gs and Cs at the 3 prime end
# % 'PRIMER_GC_CLAMP'=>0

Primer3EndGC = 0

# set the size range of PCR product
# % 'PRIMER_PRODUCT_SIZE_RANGE'=>'100-300'

AmpliconLengthRange = 100-900

# set the most desired size of PCR product
# % 'PRIMER_PRODUCT_OPT_SIZE'=>300

AmpliconLengthOptimum = 500


##############################################################################
##                        NCBI e-PCR related options                        ##
##############################################################################

# NOTE: For details of the options used below, see the NCBI e-PCR manual
# Not all options have been used

# Maximum distance between to look for between a primer pair both in target
# and non-target sequences.

AmpliconLength = 100-800

# No of mismatches to be allowed between primer and the template in an
# in-silico PCR reaction against non-target sequences.
# Note: No mismatches are allowed in target sequences

Mismatch = 2

# No of gaps to be allowed between primer and the template in an
# in-silico PCR reaction against non-target sequences.
# Note: No gaps are allowed in target sequences

Gap = 2
##############################################################################
```