Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

11-3-2017 10:00 AM

Characterization of six Arabidopsis AROGENATE DEHYDRATASE promoters

Emily J. Cornelius, *The University of Western Ontario*

Supervisor: Dr. Susanne E. Kohalmi, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology
© Emily J. Cornelius 2017

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biology Commons

Recommended Citation
Cornelius, Emily J., "Characterization of six Arabidopsis AROGENATE DEHYDRATASE promoters" (2017). *Electronic Thesis and Dissertation Repository*. 5081.
https://ir.lib.uwo.ca/etd/5081

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

Phenylalanine is an important aromatic amino acid synthesized by higher plants, and is a major component of numerous specialized metabolites including structural components, pigments, and defense compounds. The last step in the synthesis of phenylalanine is catalyzed by an enzyme called AROGENATE DEHYDRATASE, of which there are six different isoenzymes encoded by the *Arabidopsis* genome. All six have specialized roles within the plant, and are differentially expressed during development and under stressful conditions. To deduce the potential specialized role of each ADT, unique patterns of regulatory motifs were identified for all six *ADT* promoters, as well as corresponding transcription factors with similar expression profiles to each enzyme. Seven stable transgenic *Arabidopsis* lines were also generated using *ADT* promoter-eGFP/GUS constructs to test expression in all tissues during development, and under stressful conditions.

Keywords

Phenylalanine, specialized metabolism, arogenate dehydratase, cis regulatory motif, co-expression, transcriptional regulation

Acknowledgments

Firstly, I would like to thank my supervisor, Dr. Susanne Kohalmi, for her undying guidance, support, and mentorship throughout this thesis. Thank you for giving me the chance to learn, and for teaching me countless lessons about science, research, and life. I will be forever grateful for the opportunities you have given me along this entire journey. I would also like to thank my advisors, Dr. Norm Huner and Dr. Ryan Austin, for their much appreciated support and contributions throughout the process.

Second, I would like to thank Emily Clayton, Sara Rad, and Caryn Dooner, not only for being the most wonderful lab mates, but also great friends. Your tireless support and encouragement means the world to me. Also thanks to Jessica Belcastro, Daniel Bour and Michelle Millones. Daily life in the lab would not have been the same without all of you.

Third, I would like to thank my friends inside and outside the department, especially Tian Wu. Thank you for helping me realize my potential, and for keeping a smile on my face throughout every challenge. Thank you also to the rest of my friends for all the memories that I am so grateful to take with me from the last two years. I will truly miss watching ever major television event together.

Finally, I would like to thank my family. Your unconditional love and support gives me the courage to persevere when times are at their most difficult.

Table of Contents

Abstract i
Acknowledgments ii
List of Tables vii
List of Figures vii
List of Appendices viii
List of Abbreviations ix

1 INTRODUCTION **1**
1.1 Secondary Metabolism and the Phenylpropanoid Pathway 1
1.2 Phenylalanine Biosynthesis 3
1.3 Arogenate Dehydratases 4
1.3.1 Alternative Roles and Differential Expression 7
1.4 Transcriptional Regulation in Eukaryotes 8
1.5 Promoter Organization 14
1.6 Duplications and Gene Family Evolution 18
1.7 A Data Mining Approach to Gene Family Analysis 21
1.8 Research Questions, Experimental Design and Objectives 23

2 MATERIALS AND METHODS **24**
2.1 Media, Solutions and Buffers 24
2.1.1 Media 24
2.1.2 Solutions 24
2.1.3 Buffers 25
2.2 Bacterial Strains and Plasmids 25
2.2.1 Bacterial Strains and Growth Conditions 25
2.2.2 Plasmids 26
2.3 Plant Material and Standard Growth Conditions 26
2.4 DNA Isolation 28
2.4.1 Plasmid DNA Isolation from Bacteria 28
2.4.2 Plant Genomic DNA Isolation 28
2.5 PCR Amplification and Purification of *ADT* Promoter Regions 28

2.5.1 Primer Design and Sequence Amplification 30
2.5.2 Gel Electrophoresis of PCR Products and DNA Extraction.................. 30
2.6 Gateway Cloning Procedure 30
2.7 Transformations 31
2.7.1 *E. coli* Transformations... 31
2.7.2 *A. tumefaciens* Transformations.. 31
2.7.3 *N. benthamiana* Transformations 34
2.7.4 *A. thaliana* Transformations 34
2.8 Seed Collection, Sterilization and Storage 35
2.9 Histochemical Detection of GUS 35
2.10 Confocal Microscopy 35
2.11 *In Silico* Methods 36
2.11.1 Sequence Analyses.. 36
2.11.2 Motif Pattern Analysis... 36
2.11.3 Co-expression Analysis... 37

3 RESULTS... 38
3.1 *In Silico* Results... 38
3.1.1 Sequence Analyses.. 38
3.1.2 Regulatory Motif Categories.. 41
3.1.3 Proportions of Motif Categories in Each Promoter 45
3.1.4 Motifs Common to all Promoters 45
3.1.5 Unique Motifs 51
3.1.6 Significantly Enriched Motifs 58
3.1.7 Co-expression Results 67
3.2 Cloning and Expression of *ADT* Promoter-Reporter Constructs. 76
3.2.1 Promoter Sequence Amplification from gDNA 79
3.2.2 Construct Confirmation... 80
3.2.3 Transient Expression by *ADT* Promoter Sequences............................ 80
3.2.4 Stable Transformations of *Arabidopsis* 89

4 DISCUSSION 92
4.1 *ADT* Regulation and Alternative Roles... 92

4.1.1 Do all ADTs have a role in development? 93
4.1.2 A role for ADT1 in flavonoid biosynthesis.. 94
4.1.3 ADT2: mediator of chloroplast homeostasis?...................................... 95
4.1.4 A supportive role in stress and development for ADT3 96
4.1.5 ADT4 and ADT5: leading role in structure and stress response? 97
4.1.6 Does ADT6 have a role in cold acclimatization? 98
4.2 Candidate Transcription Factor Gene Families Identified by *in silico* Analyses 99
4.3 Role of Promoter Organization and Introns in Gene Regulation 103
4.4 The Role of Promoters in Gene Family Evolution 104
4.6 Limitations and Future Work 106
4.7 Conclusions 108

5 REFERENCES 109

6 APPENDICES 124

7 CIRRICULUM VITAE 137

List of Tables

Table 1: Plasmids used in this Study 27

Table 2. Primers used in this Study.. 29

Table 3. Primary Data Used to Generate Pie Charts. 48

Table 4. Motifs Present in all Six *ADT* Promoters 49

Table 5. Motifs Present in Only One *ADT* Promoter. 52

Table 6. Significantly Enriched Motifs Present in Each *ADT* Promoter. 59

Table 7. Select Co-expressed Genes. 68

List of Figures

Figure 1. Last Two Steps of Phenylalanine Biosynthesis.. 5
Figure 2. *ADT* RNA Expression Under Temperature Stress 9
Figure 3. Heat Maps of Leaf-specific *ADT* Expression Over Time. 11
Figure 4. Schematic of Eukaryotic Gene Promoter Region. 15
Figure 5. The Fate of Duplicated Genes. 19
Figure 6. Gateway Cloning Procedure. 32
Figure 7. Comparisons of ADT Nucleotide and Promoter Sequences. 39
Figure 8. General Distribution of Motifs Present in *ADT* Promoters. 42
Figure 9. Categories of Motifs Identified in Each *ADT* Promoter. 46
Figure 10. Promoter Region Map Showing Sequences Amplified from gDNA...... 77
Figure 11. Confirmation of Promoter Sequences in Expression Vector. 81
Figure 12. Negative Controls for Transient Expression in Tobacco Leaves. 83
Figure 13. Fluorescence Images of eGFP Expression Driven by *ADT* 85
Figure 14. eGFP Fluorescence Driven by Intron-Containing Constructs. 87
Figure 15. Stable Expression of GUS in *Arabidopsis* Leaves. 90
Figure 16. Summary of Predicted Roles for Each ADT 100

List of Appendices

Appendix A. List of all *cis* Regulatory Motifs: Abiotic Stress[1,2] 125
Appendix B. Co-expressed Gene Totals for Each Category[1,2]............................ 136

List of Abbreviations

A	adenine
ABA	abscisic acid
ABRC	*Arabidopsis* Biological Resource Centre
AS	Abiotic Stress
att	attachment
BAR	Bio-analytic Resource
bp	base pairs
BS	Biotic Stress
C	cytosine
CDS	coding sequence
Col-0	Columbia-0
D	Development
ddH_2O	double distilled water
DNA	deoxyribonucleic acid
EDTA	ethylene diaminetetraacetic acid
GA	gibberellic acid
gDNA	genomic deoxyribonucleic acid
G	guanine
GR	General Response
GUS	β-glucuronidase
eGFP	enhanced green fluorescent protein
Kb	kilobase
LB	lysogeny broth
L	Light Response
MES	2-(N-morpholino) ethanosulfonic acid
mRNA	messenger RNA
MS	Murashige and Skoog
OD_{600}	optical density measured at 600 nm
PBS	phosphate buffered saline

PCR	polymerase chain reaction
Phe	phenylalanine
PLACE	Plant *cis*-acting Regulatory DNA Elements database
PPM	Plant Preservative Medium
PR	Phenylpropanoid-Related
PSSM	position-specific scoring matrix
RNA	ribonucleic acid
RPM	revolutions per minute
SD	Stress and Development
SDS	sodium dodecylsulphate
SOC	Super Optimal Growth with Catabolite repression
T	thymine
TAE	Tris base, glacial acetic acid, EDTA
TAIR	The *Arabidopsis* Information Resource
T-DNA	transfer deoxyribonucleic acid
Ti-plasmid	tumour-inducing plasmid
T_m	melting temperature
TSS	transcriptional start site
TrSS	translational start site
UTR	untranslated region
UV	ultraviolet
WT	wild type
w/v	weight per volume
X-Gluc	5-bromo-4-chloro-3-indoxyl-beta-D-glucuronide cyclohexylammonium salt

*SI units not included

1 INTRODUCTION

This study focused on the analysis of the promoter sequences and transcriptional regulation of *AROGENATE DEHYDRATASE (ADT)* genes in *Arabidopsis thaliana* using two approaches. An *in silico* approach was used to characterize promoter sequences and identify regulatory motifs and transcription factors that may contribute to differential *ADT* regulation. Promoters were subsequently isolated and cloned, and used to generate stable transgenic *Arabidopsis* plants that will allow testing of differential expression *in vivo.*

1.1 Secondary Metabolism and the Phenylpropanoid Pathway

Plants are both dependent upon and hindered by the abiotic and biotic factors that shape their environment. On one hand, they depend heavily on light energy, adequate water, seed dispersers, and pollinators (Winkel-Shirley, 2001) to survive and reproduce. On the other hand, they can be threatened by high or low light intensity or temperature, flooding or drought, and pathogens or herbivores. Unless each of these variables is perfectly controlled, plants must find ways to adjust to their changing environments. To facilitate the positive interactions, and to reduce the effects of negative interactions, plants have evolved the ability to produce secondary metabolites. Secondary metabolites are considered to be anything not directly involved in protein or nucleotide metabolism, and can include protective structural components, antimicrobials, coloured pigments, and scent compounds (Vogt, 2010). The term “secondary” was coined by scientists when it was still unclear what roles these metabolites play in plants, and can be misleading because it implies that these compounds are not essential for plant survival. As a result, secondary metabolites are now more appropriately referred to as specialized metabolites (Pichersky and Lewinsohn, 2011)

One of the most important specialized metabolic pathways in plants is the phenylpropanoid pathway. Depending on the plant species, and whether it is woody or herbaceous, up to 50% of the carbon assimilated from photosynthesis can be incorporated into phenylalanine (Phe)-derived metabolites (Corea et al., 2012; van Heerden et al., 1996). Among the most important of these metabolites

are flavonoids and lignins. To better illustrate the importance of the phenylpropanoid pathway, these two examples of Phe-derived metabolites will be discussed in more detail.

Flavonoids are the chemical compounds seen as red, blue and purple pigments in plants, and their major role is attracting pollinators and seed dispersers through bright floral and fruit displays and scents (Winkel-Shirley, 2001). To date, more than 10 000 different types of flavonoids have been identified across higher plants (Zhao et al., 2017b), allowing the possibility for a much greater variety of needs for these compounds across plant species. Flavonoids also have roles in symbiotic signaling, and play a particularly important role in attracting rhizobia and other symbiotic soil microbes to roots of legumes under low nitrogen conditions (Liu and Murray, 2016). Some flavonoids are also cytotoxic, and can deter insect feeding and oviposition by preventing digestion or influencing skeletal muscle contractions, or reducing hatching success if eggs are laid on leaves or in surrounding soil (Mierziak et al., 2014). Furthermore, flavonoids are thought to act as plant “sunscreens”, and are synthesized in response to increased UV-B radiation. They accumulate in leaf epidermal cells, and reduce the effects of harmful free radicals and oxidative DNA damage (Landry et al., 1995; Winkel-Shirley, 2001). Aside from these essential roles in plants, flavonoids also have benefits to humans. The antioxidant properties of deeply red and purple pigmented berries and grapes are thanks to flavonoids. Foods rich in flavonoids are thought to reduce inflammation, risk of cancer, and cardiovascular diseases in humans (Skrovankova et al., 2015). These brightly coloured pigments are also highly desirable features of ornamental plants, and these properties are often engineered to generate new varieties with different colours or patterns (Nishihara and Nakatsuka, 2011). In turn, flavonoids are major contributors to the almost $1.3 billion in flower plant sales annually in Canada (Statistics Canada, 2015).

Much like flavonoids, lignins are very important specialized metabolites that play multiple roles in plants. Lignin is one of the most abundant polymers on earth (Boerjan et al., 2003) and is thought to have played an essential role in the evolution of land plants. The transition from water to land posed new threats to

plants, including wind and mechanical damage, desiccation, and higher light intensity. A recent study by Renault et al. (2017) uncovered an essential developmental role in mosses for P450 oxygenase, which delineates the starting point of lignin metabolism in angiosperms. The first part of the lignin biosynthesis pathway is essential in forming a cuticle layer in moss that controls structural growth and water exchange (Renault et al., 2017). Therefore, it is thought that this early phenol-rich lignin precursor in mosses was the first example of lignin, and gave rise to its roles in many other processes including fungal and pathogen protection (Bhuiyan et al., 2009). It has been shown that lignin biosynthesis increases and accumulates around sites of fungal or insect penetration, reducing susceptibility to mechanical damage and infection (Bhuiyan et al., 2009; Thakur and Sohal, 2013). Like flavonoids, lignins are also relevant to humans. As lignins cannot be completely digested by humans, they are a component of dietary fibre (Slavin, 2013). Lignin also provides the rigidity and strength of hardwoods used in structural building foundations, furniture, and flooring. Additionally, lignin is a major by-product of paper production. Studies in renewable energy have suggested the use of this leftover lignin for biofuel production, making our understanding of lignin biosynthesis even more important (Slavin, 2013).

1.2 Phenylalanine Biosynthesis

The major precursor of the phenylpropanoid pathway is the aromatic amino acid Phe. Since plants, fungi, bacteria are the only organisms that can synthesize Phe (Herrmann and Weaver, 1999), it is essential for animals and must be obtained through their diet. In humans, this amino acid is required for protein synthesis, but also has many other roles, including in the generation of neurotransmitters and melanin (Fernstrom and Fernstrom, 2007). In addition, Phe deficiency is associated with several human diseases, such as vitiligo (loss of skin pigment in blotches), for which it can be used as a treatment (Cohen et al., 2015). Since it is a molecule of exceptional importance to plants, animals and humans, Phe biosynthesis is a topic of great interest across all areas of science. Hence, understanding the regulation of Phe biosynthesis is essential.

All aromatic amino acids are synthesized *de novo* through the shikimate pathway (Herrmann and Weaver, 1999). In higher plants, the last two steps of Phe synthesis can occur via two ways. In the prephenate pathway, prephenate is decarboxylated and dehydrated to form phenylpyruvate, which is then transaminated to form Phe (Figure 1). The prephenate pathway is more commonly used by bacteria and fungi, but has also been described in plants (Tzin and Galili, 2010). In the arogenate pathway, prephenate is first transaminated to form arogenate, which is then decarboxylated and dehydrated by an enzyme called arogenate dehydratase (ADT) to form Phe (Figure 1)(Cho et al., 2007; Ehlting et al., 2005; Jung et al., 1986). The arogenate pathway is almost exclusively described in higher plants. However, our lab has previously shown that two *Arabidopsis* ADTs (ADT1 and ADT2) have retained the ability to accept prephenate as a substrate, and may act as ADT/PDTs under certain conditions (Bross et al., 2011). PDT activity was also found for some ADTs in petunia (Maeda et al., 2010) and rice (Yamada et al., 2008). As they catalyze the crucial last step in Phe synthesis, initiating the phenylpropanoid pathway, ADTs are the enzyme of interest in this study.

1.3 Arogenate Dehydratases

ADTs have been identified in every higher plant analyzed to date, and most encode several versions of ADT isoenzymes. For instance, petunia (*Petunia hybrida*) encodes three *ADTs* (Maeda et al., 2010), pine (*Pinus taeda*) encodes nine (El-Azaz et al., 2016) and *Arabidopsis* encodes six (Cho et al., 2007). All ADT proteins are localized within or around chloroplasts where Phe is synthesized (Bross et al., 2017; Jung et al., 1986; Rippert et al., 2009). It is still, as of yet, not fully understood why multiple ADTs are necessary, as most bacteria and yeast only encode one (Bross et al., 2011).

In *Arabidopsis*, all 6 *ADT*s are similar in protein structure and amino acid sequence (Cho et al., 2007). Each protein has an N-terminal transit peptide

Figure 1. Last Two Steps of Phenylalanine Biosynthesis

The last two steps of Phe biosynthesis can occur via two pathways.

Top: The prephenate pathway. Prephenate is first decarboxylated and dehydrated by a prephenate dehydratase (PDT) to form phenylpyruvate. Phenylpyruvate is then transaminated by a phenylpyruvate aminotransferase (PPAT) to form Phe. This pathway is mainly described for bacteria and fungi.

Bottom: The arogenate pathway. Prephenate is first transaminated by a prephenate aminotransferase (PAT) to form arogenate. Arogenate is then decarboxylated and dehydrated by an arogenate dehydratase (ADT). This pathway is predominantly used by higher plants.

ADT: arogenate dehydratase, PAT: prephenate aminotransferase, PDT: prephenate dehydratase, PPAT: phenylpyruvate aminotransferase

Prephenate Pathway

PDT

Phenylpyruvate

PPAT

Prephenate

PAT

Arogenate

ADT

Phenylalanine

Arogenate Pathway

domain, a catalytic domain, and a C-terminal ACT domain. Transit peptide sequences are quite variable, but the catalytic and ACT domains are conserved with approximately 80% similarity in amino acid sequence (Cho et al., 2007). Although *Arabidopsis* ADTs are all similar in sequence, and all six can decarboxylate and dehydrate arogenate, there is evidence that each ADT might have a more specialized role in *Arabidopsis*.

1.3.1 Alternative Roles and Differential Expression

In *Arabidopsis,* it has been shown that the activity of certain ADTs is more important for certain branches of the phenylpropanoid pathway. For instance, it was demonstrated that ADT5 and possibly ADT4 play a more pronounced role channeling Phe into lignin biosynthesis than the other ADTs (Corea et al., 2012). In this study, *Arabidopsis adt4/5* and *adt5* knockout lines showed severely low lignin (wilted) phenotypes compared to other single and double *adt* knockouts. Research also indicates more specific roles for *ADT1* and *ADT3* in regulating biosynthesis of anthocyanins, which are flavonoid-derived pigments (Chen et al., 2016b), as mutants with a non-functional *adt1* or *adt3* only produced around half of the wild type levels of anthocyanins. There is also evidence that ADT3 plays a role in reactive oxygen species (ROS) homeostasis through synthesis of photo-protective flavonoid compounds and epicuticular waxes (Para et al., 2016).

ADT2 and ADT5 might also have non-enzymatic roles within the cell in addition to their enzymatic roles (Bross et al., 2017). ADT2 is thought to have a role in chloroplast division, and forms a ring around chloroplast equatorial planes similar to the well-known chloroplast division protein FtsZ, a structural homolog of tubulin (Vitha et al., 2001). ADT5 is localized to the nucleus as well as chloroplasts, and is thought to have a second role as a transcription factor (Bross et al., 2017). Since they both have an enzymatic role in chloroplasts, and appear to have a second non-enzymatic function, both ADT2 and ADT5 are considered moonlighting proteins. Moonlighting proteins are defined as having multiple functions that are not a result of splicing, gene fusion or dimerization, or due to pleiotropic effects (Jeffery, 2015). Furthermore, *ADT*s are differentially expressed

in different tissues, at different developmental stages, and under different environmental conditions (Figure 2; Corea et al., 2012). Each *ADT* has a unique expression pattern under different conditions (Figure 2). Higher expression levels are seen for *ADT1, ADT2, ADT4* and *ADT6* during heat shock (Figure 2A), while *ADT3, ADT4* and *ADT6* are more highly expressed during cold shock (Figure 2B). Other data from previous students in our lab also indicate different expression patterns in various tissues (Hood, 2008), and the Bio-analytic Resource for Plant Biology (BAR) database also contains expression data for developmental stages (Austin et al., 2016). As an example, the heat map in Figure 3 shows the different levels of expression in leaves at each developmental stage from week one through twelve. Since all ADTs catalyze the same reaction in the same area of the cell, but are differentially expressed, there is a strong indication that they are regulated at the level of transcription.

1.4 Transcriptional Regulation in Eukaryotes

At any given time, DNA that is not expressed is wrapped tightly around histone proteins to prevent transcription, degradation of DNA, or excessive energy expenditure (Kornberg, 2007; Larch et al., 1987). This is the case for most genes that are only expressed under certain conditions, such as stress response. To initiate transcription of a gene for any purpose, histone proteins need to be shifted from the promoter sequence to unfold the DNA from the nucleosome and make it accessible to other proteins (Boeger et al., 2003). For this reason, the promoter sequence is essential for transcription to occur.

Promoters are non-coding DNA sequences usually found immediately upstream of a gene (Danino et al., 2015; Novina and Roy, 1996; Roy and Singer, 2015) and their role is essentially to initiate and regulate transcription (Kadonaga, 2012). Proximal promoters are considered to be the region approximately 500 to 1000 bp upstream from the transcriptional start site (TSS), and distal promoters are found much further upstream, and include enhancer regions (Korku et al., 2014; Kristiansson et al., 2009). The transcription factors responsible for initiating basal levels of transcription recognize what is known as the core promoter, often

Figure 2. *ADT* RNA Expression Under Temperature Stress

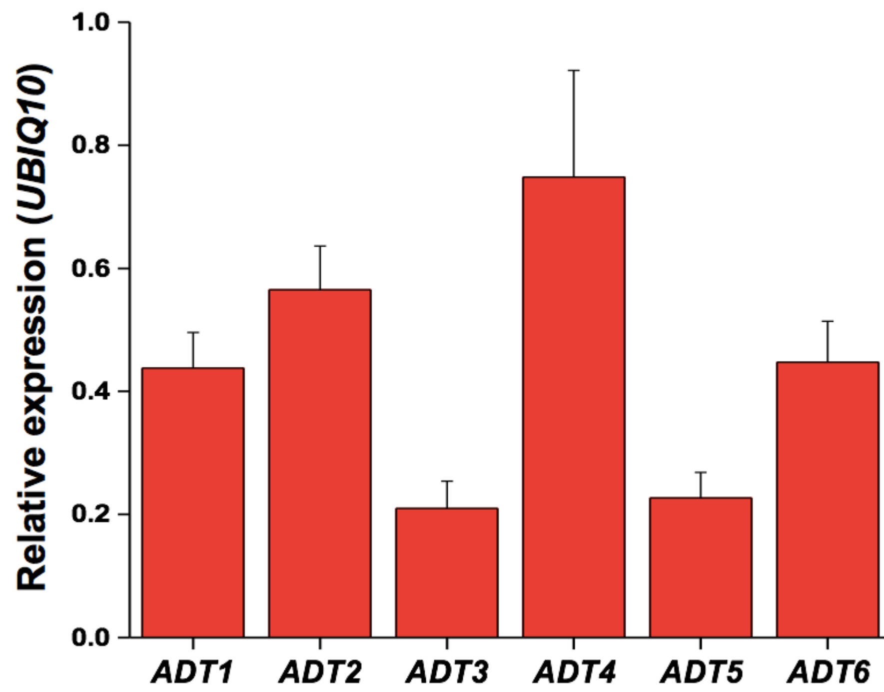
RT-PCR data showing relative expression of *ADTs* to the *UBIQUITIN10* gene as an internal control under standard conditions. Template RNA was isolated from plants that were exposed to their respective conditions under 16 h light and 8 h dark. Each *ADT* is differentially expressed under these conditions, suggesting that ADTs have varying roles during temperature stress.

A. Heat shock (38°C) for 24 h

B. Cold shock (6°C) for 24 h

Adapted from Hood, 2008.

A



B

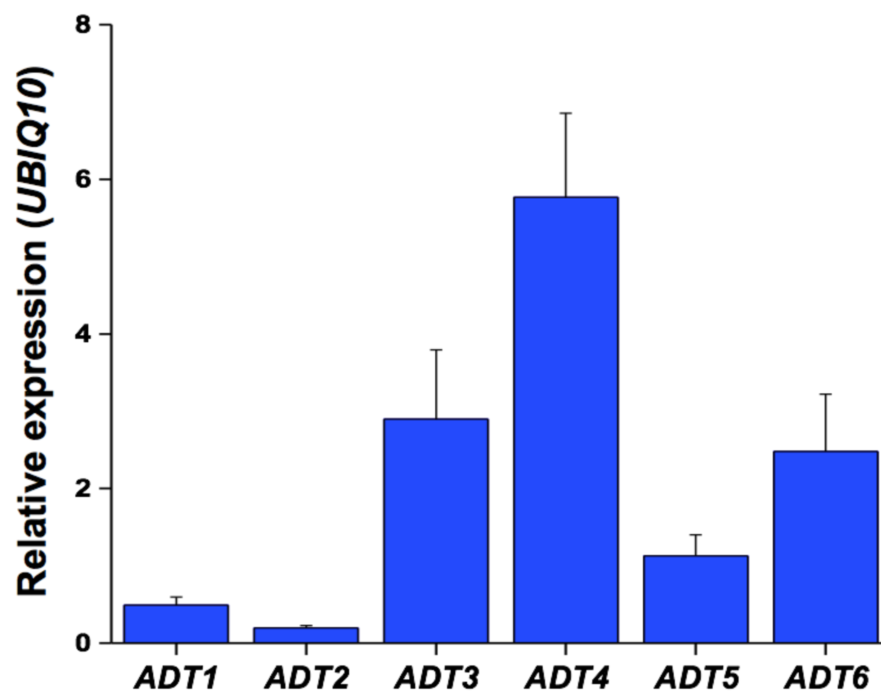
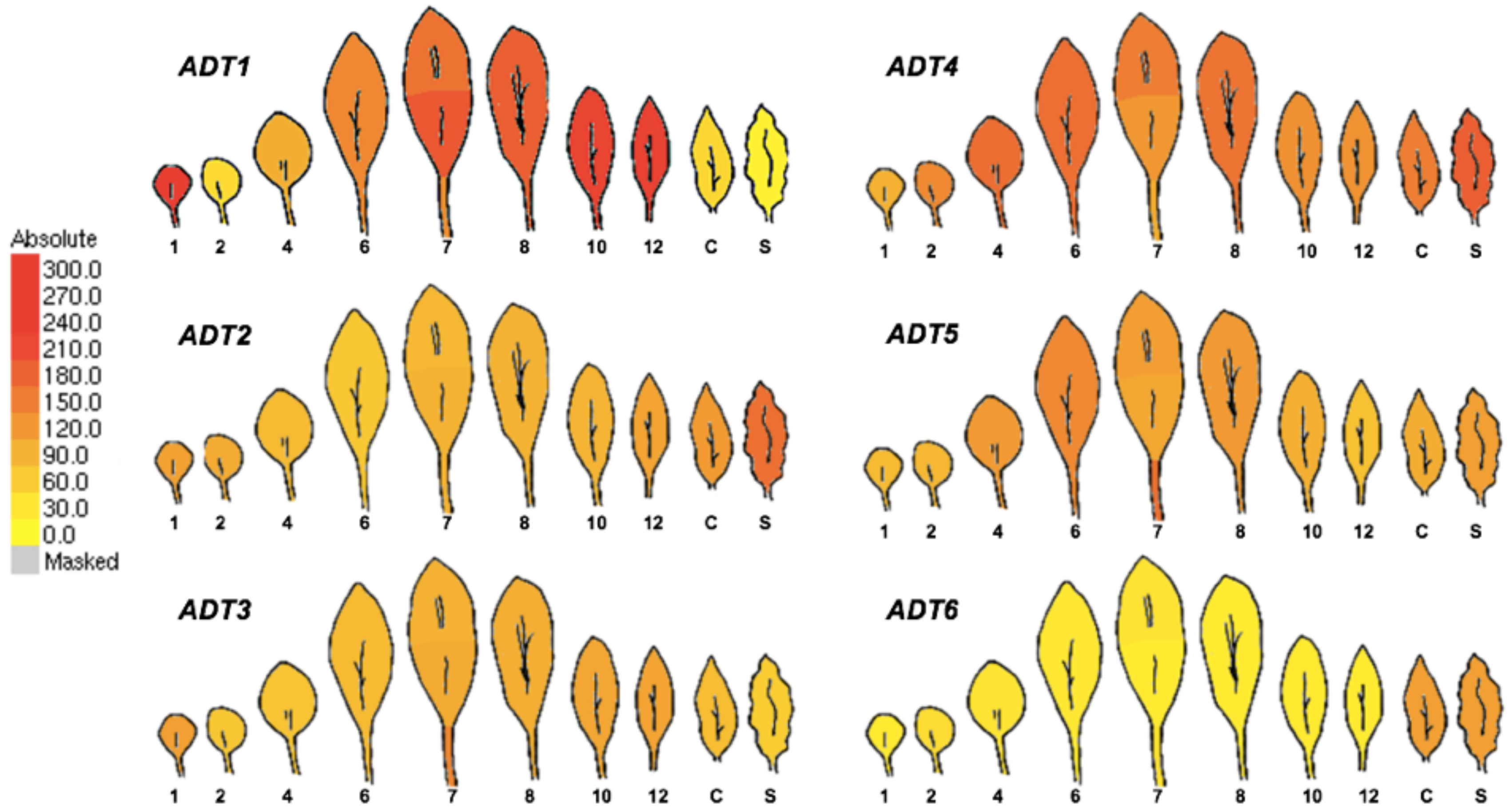


Figure 3. Heat Maps of Leaf-specific *ADT* Expression Over Time.

Heat map representations of ADT RNA levels in *Arabidopsis* leaves over time from week one of development to senescence. Heat maps were generated using the BAR database *Arabidopsis* eFP Browser (Winter et al., 2007) and are based on published microarray data. The scale represents the absolute expression level of a given *ADT.*

Continuous scale from yellow (low expression) to red (high expression). Number below each leaf represents the number of weeks since germination.

C: cauline, S: senescent


ADT1
ADT2
ADT3
ADT4
ADT5
ADT6
Absolute
300.0
270.0
240.0
210.0
180.0
150.0
120.0
90.0
60.0
30.0
0.0
Masked
1 2 4 6 7 8 10 12 C S

located within the first 50 bp upstream from the TSS (Molina and Grotewold, 2005; Novina and Roy, 1996). The core promoter contains very specific sequence motifs, of which one of the most well-characterized is the TATA box. Although not all core promoters contain actual TATA boxes, the core promoter sequence is usually very A/T-rich, creating a hydrophobic surface for the TATA-binding protein (TBP) to recognize and bind to (Sainsbury et al., 2015). There are 6 general transcription factors that form a complex required for RNA Polymerase II recruitment and initiation of transcription: Transcription Factor RNA Polymerase II A (TFIIA), TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Kadonaga, 2012; Kornberg, 2007; Sainsbury et al., 2015). The entire complex is referred to as the pre-initiation complex (PIC), and this is widely accepted to be the most basic machinery of transcription initiation, or the "on/off switch" (Kadonaga, 2012).

In mammalian promoters, there are two major core sequence element types: the TATA box and the CpG island. TATA boxes are often present in environmentally-responsive genes, and CpG islands are more common in housekeeping genes (Molina and Grotewold, 2005; Yamamoto et al., 2009, 2011). In plants, core promoter types are more variable, and do not contain known CpG islands. Instead, the main core types in plants are TATA boxes, Y patches (or pyrimidine patches), GA elements, and the less common CA elements (Yamamoto et al., 2009). Promoters that do not contain any known elements are referred to as coreless promoters, although they may have some sequence characteristics similar to one or more core types.

TATA boxes are seen as highly responsive, high expression specialists in plants and vertebrates (Yamamoto et al., 2009, 2011). TATA box-containing promoters tend to be longer than other promoters. This suggests that promoters controlling gene expression through environmentally-responsive signaling pathways are longer to accommodate more regulatory motifs, and to ensure the complicated network of internal and external cues can precisely control transcription (Kristiansson et al., 2009; Yamamoto et al., 2011). This is in contrast with housekeeping genes having a short promoter sequence, as their expression is not quite as dependent on intricate networks of responses, and sharp peaking

levels of expression are not characteristic of ubiquitously present proteins such as tubulin.

Though CpG islands are thought to control housekeeping gene expression in mammals, there are no known core types in plants that are specifically thought to regulate housekeeping gene expression. In fact, the other core types in plants, aside from TATA boxes, do not have any significant associations, other than coreless promoters having generally lower expression (Yamamoto et al., 2011). Although Y-patches are fairly common and sometimes thought to be the CpG equivalent in plants, they are still poorly understood in terms of function. This suggests that aside from the TATA box, plants and vertebrates differ considerably in core promoter type (Gagniuc and Ionescu-Tirgoviste, 2012; Yamamoto et al., 2009, 2011).

1.5 Promoter Organization

Core promoters only make up one part of eukaryotic promoters (Figure 4). The core promoter, usually located within 50 bp upstream of the TSS, is the region containing motifs recognized by the PIC and RNA Polymerase II (the on/off switch) (Molina and Grotewold, 2005; Novina and Roy, 1996). The proximal promoter then extends up to 1000 bp upstream from the TSS (Korku et al., 2014; Kristiansson et al., 2009). This 1000 bp stretch contains a number of short nucleotide sequences, usually anywhere between 5 and 49 nucleotides long, called *cis* regulatory elements (CREs), or regulatory motifs. These motifs can be recognized by transcription factors, and in turn, regulate transcription (Figure 4) (Korku et al., 2014; Kornberg, 2007). Since a promoter sequence controls transcription of both the forward and reverse DNA strands, motifs can be found on either strand. However, the same motif does not necessarily need to be in the same area on both strands, as many transcription factors' jobs are ultimately to assist in the unwinding of promoter DNA (Kornberg, 2007; Larch et al., 1987).

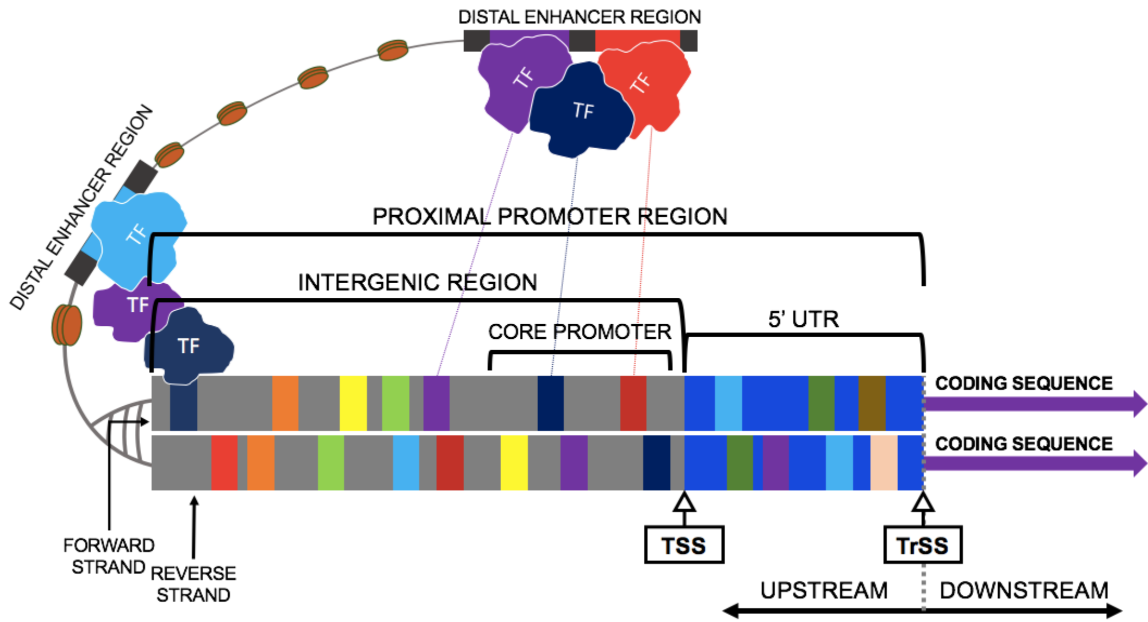
In a eukaryote, cells require different genes to be activated or turned off at different times, and as all genes cannot be expressed all at once, there is cell type-specific transcriptional regulation. Aside from the core promoter, there are

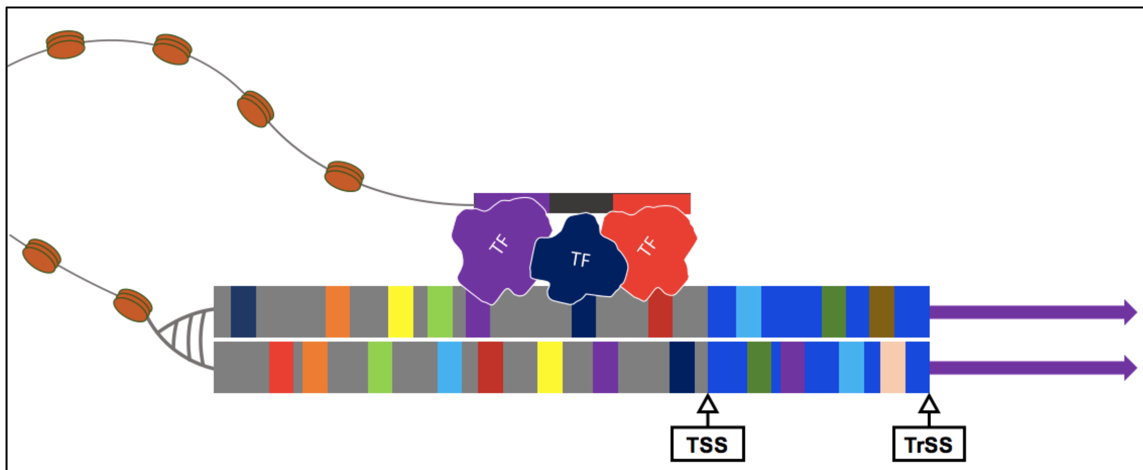
Figure 4. Schematic of Eukaryotic Gene Promoter Region.

Top: The proximal gene promoter region is immediately upstream of a given coding sequence (purple arrows) and can extend up to 1 Kb from the transcriptional start site (TSS), spanning any untranslated regions (UTRs) and intergenic regions between one gene and the next upstream gene. Aside from the core promoter, there are multiple short sequence motifs, usually 5 to 49 bp long that are scattered on both strands in this 1 Kb region. Motifs are shown as small boxes in various colours, where each colour represents a binding site for a specific corresponding transcription factor (TF) and therefore a different function. There are also distal enhancer regions that can be several Kb away from the TSS.

Bottom: Enhancer regions contain regulatory motifs that transcription factors bind to and cause folding of the DNA to reach the promoter. Transcription levels are then influenced by these transcription factors.

TrSS: translational start site


DISTAL ENHANCER REGION
TF
TF
TF
DISTAL ENHANCER REGION
TF
TF
TF
PROXIMAL PROMOTER REGION
INTERGENIC REGION
CORE PROMOTER
5' UTR
CODING SEQUENCE
CODING SEQUENCE
FORWARD
STRAND
REVERSE
STRAND
TSS
TrSS
UPSTREAM
DOWNSTREAM


TF
TF
TF
TSS
TrSS

cis-acting motifs throughout the promoter region and in distal enhancer regions that are recognized by cell-specific transcription factors, which result in either increased or decreased transcription (Korku et al., 2014). If the core promoter is the on/off switch, proximal regulatory motifs and cell type-specific or environmentally-responsive transcription factors can be considered the volume control of gene expression. There are two steps in this process: 1) a motif is recognized by a transcription factor, and binding of the transcription factor can inhibit or enhance transcription of that gene by 2) causing alterations to the DNA structure, or recruiting other complexes of transcription factors (D’haeseleer, 2006; Zhu et al., 2015). Through a variety of signaling pathways, transcription factors integrate environmental cues and internal signals to provide an organism with the appropriate tools to facilitate a proper interaction with its environment (Babbitt et al., 2017). They therefore govern every aspect of survival, from biochemical pathways, to DNA repair, to stress responses to ensure survival.

On the 5’ and 3’ ends of transcribed RNA sequences are untranslated regions (UTRs) (Figure 4). UTRs are transcribed but are not part of the translated protein, but the DNA encoding these UTRs contains regulatory motifs, meaning they can also have roles in transcriptional regulation (Baxter et al., 2012; van der Velden and Thomas, 1999). Additionally, regulatory motifs can be found in introns of genes, and even several thousand base pairs away in enhancer regions (Figure 4). It has been shown in multiple studies that the presence of introns in coding sequences has a positive effect on gene expression (Gallegos and Rose, 2015; Rose et al., 2016). This phenomenon is referred to as intron-mediated enhancement (IME), and is still poorly understood, but is thought to be due to the regulatory motifs present within these intron sequences. Enhancer regions are located far away from a given TSS (Figure 4), and are especially important for cell type-specific regulation (Hnisz et al., 2016). Enhancers often control gene expression by interacting with transcription factors, and then bringing them into the proximity of the promoter sequence through DNA folding and looping (Figure 4; Hnisz et al., 2016; Sainsbury et al., 2015). The interactions between the promoter

sequence and transcription factors in the loop can contribute to the further unwinding of promoter DNA to initiate transcription at that site.

1.6 Duplications and Gene Family Evolution

Gene families are groups of related genes, usually with similar functions, that arise most often from duplication events, and are observed in almost all eukaryotes, especially plants. In fact, no other group of organisms has a greater incidence of polyploidy or duplicated DNA than plants (Wendel et al., 2009). It is thought that whole genome duplications (WGDs) were a driving force of the rapid diversification and world dominance of angiosperms (Airoldi and Davies, 2012). Once duplicated, a gene can have one of three fates: neo-functionalization, sub-functionalization, or loss (Figure 5). The most common fate is the loss of a gene, as it can be energetically costly to synthesize multiple redundant copies of a gene that all perform the same function. However, if a WGD occurs, the multiple copies of genes that are involved in complex networks, including transcription factors or signal transduction components, are often retained. It is believed that this happens because their function relies more on the balance of components of the complex relative to each other rather than the numbers of each one. So, if everything is duplicated and the balance is preserved, this is not an issue, but if one or two components are lost, it can be detrimental to the cell (Edger and Pires, 2009). Additionally, if all copies are retained, as long as one of them can still perform the original network function, there is flexibility for the other copies to adopt a new function (neo-functionalization) through mutation or interaction, where they either only perform the new function (Figure 5B), or they keep the old function and the new function (Figure 5C; Airoldi and Davies, 2012). Sometimes the role of the duplicated gene can be partitioned between two copies so that both copies must be present for the complete function (Figure 5D). These copies can also be regulated differently during certain conditions, and ultimately provide more complexity for the cell (Airoldi and Davies, 2012; Lynch and Conery, 2000).

Figure 5. The Fate of Duplicated Genes.

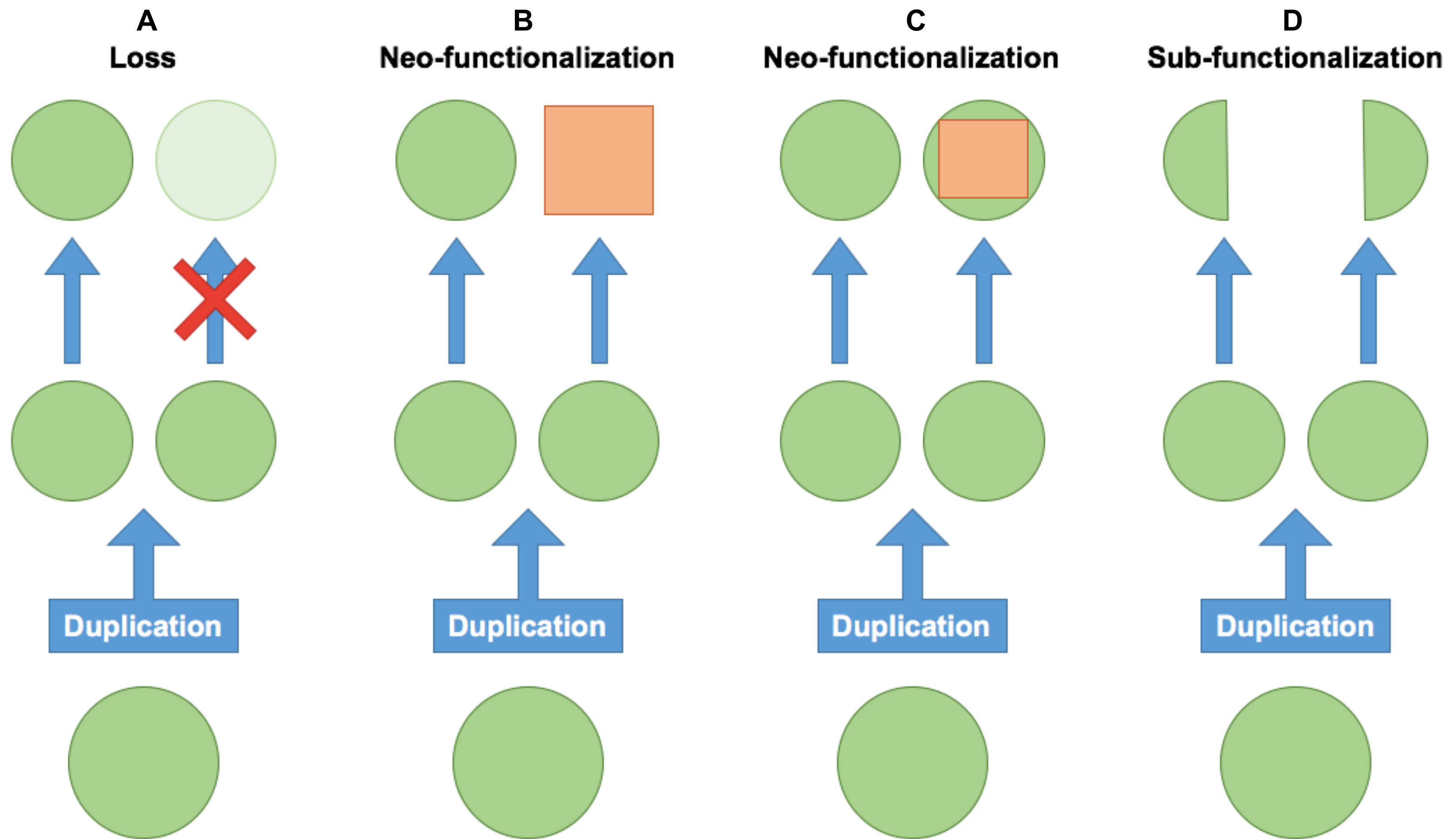
Following a gene duplication event, the extra copy of a gene may be lost, adopt a new function, or have its function partitioned into two parts.

A. In most cases, extra copies become dysfunctional and are lost.

There are two options for neo-functionalization:

B. The new copy can adopt a new function and lose the original function
C. The new copy can retain its original function and also adopt a new function, so it is able to perform two functions (middle right).

D. A duplicated gene may also be partitioned into two parts (sub-functionalization). Since there are two copies, it might be more efficient for a plant to encode two parts of the protein so it can differentially regulate both parts, providing the plant with more functional options.


A
Loss
B
Neo-functionalization
C
Neo-functionalization
D
Sub-functionalization
Duplication
Duplication
Duplication
Duplication

Since plants can sustain WGD better than any other organism (Airoldi and Davies, 2012), gene family diversification is well-described in plants. It is common for a plant to encode hundreds of members of a gene family that may have some overlapping functions, but can still carry out distinct roles in the cell (Airoldi and Davies, 2012). For example, as a result of numerous WGD events, the MADS family of transcription factors in *Arabidopsis* contains more than 100 members with various roles in floral development and embryogenesis (Becker and Theissen, 2003). For instance, AGL15-like MADS box proteins are all involved in development. However, AGL15 is expressed in embryos of developing seeds, but not in endosperm, whereas AGL18 is expressed in endosperm but not in any stage of embryo development (Becker and Theissen, 2003). Given the preference for plants to retain proteins such as transcription factors after a WGD event, it is not surprising that the MADS family has grown so large over time. As *ADTs* are also involved in a complex network of metabolic signaling and activity, they may have been preferentially retained rather than lost by higher plants, which allowed neo-functionalization. One important point to keep in mind is that WGDs duplicate not only the coding sequences themselves, but also all non-coding DNA (Lynch and Conery, 2000; Wendel et al., 2009). The combination of having multiple copies of a gene, each with their own regulatory sequence that can sustain any number of substitutions, presents a good opportunity for neo-functionalization. Promoter sequences can sustain much higher variation than coding sequences can (Vedel and Scotti, 2011), and this variation can affect the level of expression of each gene copy at any given time. Therefore, promoter sequences can harbour a wealth of knowledge about the neo-functionalization of different gene family members, including *ADTs.*

1.7 A Data Mining Approach to Gene Family Analysis

Since promoters control transcription, and transcription is responsive to internal and external environmental cues via signaling pathways, it is possible to make predictions about gene function based on experimentally determined promoter motifs. The PLACE (Plant *cis*-Acting Regulatory DNA Elements)

database (Higo et al., 1999) contains information, based on primary literature, on numerous *cis* regulatory motifs and their functions *in planta.* This database is one of the largest and most comprehensive of its kind, and is therefore an excellent tool for determining the putative motif composition of a given plant gene promoter, ultimately providing an idea as to which pathways a gene may be involved in. An analysis of these elements can not only help make predictions about gene function, but also identify candidate transcription factors for further study.

Though the basic functions of ADTs are established (Bross et al., 2011, 2017; Cho et al., 2007), their roles in preferential supply of Phe to the phenylpropanoid pathway are still poorly understood (Bross et al., 2017; Corea et al., 2012; Para et al., 2016). Previous studies have used the PLACE and other motif databases to gain relevant insight about specific gene function and regulation in newly discovered and poorly characterized gene families. However, these studies either characterize one gene promoter *in silico* accompanied by an *in planta* analysis (Kumar et al., 2015; Sohrabi et al., 2015; Srivastava et al., 2014) or characterize multiple promoters *in silico* without an *in planta* analysis (Song et al., 2011a). For instance, Srivastava et al. (2011) characterized the promoter of the *SIEVE ELEMENT OCCLUSION (SEOF1)* gene from *Pisum sativum* using the PLACE database, and found a number of stress-related motifs. Transient expression of the *GUS* reporter gene and quantitative analysis of expression changes under different stressful conditions then supported the *in silico* prediction. Though this analysis is an important step in understanding the multiple roles of *PsSEOF1* in stress response, it is only one promoter and does not differentiate between promoters of a gene family. There are also multiple *in silico* studies that characterize promoter structure across an entire genome (Molina and Grotewold, 2005; Yamamoto et al., 2009, 2011; Zhu et al., 2015), but this information does not specifically pertain to a given gene and often only focuses on core promoters rather than specific regulatory motifs. There are only few studies where a comparative analysis of all promoters of a gene family are analyzed both *in silico* and *in vivo.*

1.8 Research Questions, Experimental Design and Objectives

All six ADT enzymes in *Arabidopsis* are capable of catalyzing the same reaction in Phe biosynthesis, but they are differentially expressed, suggesting that they can channel Phe into different branches of the phenylpropanoid pathway depending on the needs of the plant. It is hypothesized that each *ADT* promoter sequence has a unique motif pattern that is recognized by transcription factors which are tailored to that ADT's role in *Arabidopsis.* There are two main objectives of this thesis.

The first objective is an *in silico* analysis of each *ADT* promoter region to identify putative regulatory motifs that might contribute to differential expression of each *ADT.* The motif sequence analysis will be complimented by a meta-analysis of existing expression data to identify any transcription factors that can validate the motifs identified. If the same transcription factor is identified in both analyses, it is a candidate gene for further study where transcription factor expression or motif presence can be modified.

The second objective is an *in planta* analysis, where promoter sequences of each *ADT* will be cloned 5' to reporter genes, whose expression can be monitored as a proxy for *ADT* gene expression. Promoter-reporter constructs will then be used for two types of plant transformations. Transient expression analyses in *N. benthamiana* will be performed first to check that promoters are able to generate reporter expression *in planta.* Stable *Arabidopsis* transformants will also be generated so expression can be viewed in all tissues at different developmental stages, and under different stressful conditions. Subcellular localization patterns can also be determined under these varying conditions.

2 MATERIALS AND METHODS

2.1 Media, Solutions and Buffers

2.1.1 Media

All media were autoclaved to sterilize unless stated otherwise.

Gamborg's Solution

For 100 mL: 0.32 g Gamborg's solution powder with B5 + vitamins, 2.0 g of 20 g/L sucrose, 1.0 mL of 1 M stock MES (pH 5.6), 100 μL of 200 mM acetosyringone.

Lysogeny Broth (LB)

For 1 L: 950 mL ddH_2O, 10 g tryptone, 10 g NaCl, 5 g yeast extract. For solid medium add 15 g agar per L before autoclaving.

Murashige & Skoog (MS) Medium

For 1 L: 4.3 g Murashige & Skoog salts, 10 g sucrose, 0.5 g MES. For solid medium add 8 g agar per L. If plating seeds, add 1 mL of 100 mg/L^{-1} carbenicillin and 5 mL Plant Preservative Mixture (PPM) after autoclaving and cooling.

Super Optimal Broth with Catabolite Repression (SOC)

For 1 L: 970 mL ddH_2O, 0.5 g NaCl, 0.186 g KCl, 20 g tryptone, 5 g yeast extract. After autoclaving and cooling, add 10 mL of 1 M $MgCl_2$ and 3.6 g glucose.

2.1.2 Solutions

0.7% Agarose Gel

0.21 g agarose, 30 mL 1X TAE buffer, heat 1 min to dissolve.

Antibiotic Solutions

100 mg/mL ampicillin, 100 mg/mL carbenicillin, 50 mg/mL gentamycin, 60 mg/mL kanamycin, 100 mg/mL spectinomycin were dissolved in ddH_2O and filter sterilized to prepare stock solutions.

GUS Staining Solution

For 1 mL: 830 μL ddH_2O, 100 μL of 1 M $NaPO_4$ (pH 7.0), 20 μL of 0.5 M EDTA (pH 8.0), 10 μL of 10% Triton X-100, 20 μL of 50 mM $K_3Fe(CN)_6$, 20 μL of 0.1 M X-Gluc (50 mg/mL) dissolved in dimethylformamide.

Mini Prep Solution I (A)

For 100 mL: 5 mL of 50 mM glucose, 2.5 mL of 25 mM Tris (pH 8.0), 0.2 mL of 10 mM EDTA (pH 8.0), 92.3 mL ddH_2O

Mini Prep Solution II

For 10 mL: 1 mL of 1% (w:v) SDS, 0.4 mL of 0.2 NaOH, 8.6 mL ddH_2O

Mini Prep Solution III (B)

For 100 mL: 60 mL of 3 M K-acetate, 11.5 mL glacial acetic acid, 28.5 mL ddH_2O

Seed Sterilization Solution

For 1 mL: 500 μL undiluted bleach, 500 μL ddH_2O, 50 μL PBS Tween

2.1.3 Buffers

50X TAE Buffer

For 1 L: 242 g Tris, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA (pH 8.0).

PBS Tween

For 1 L: 800 mL ddH_2O, 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4, 2 mL Tween-20. Adjust volume to 1 L and pH to 7.2.

Quick DNA Extraction Buffer

For 100 mL: 20 mL 1M Tris (pH 7.5), 1.46 g NaCl, 5 mL 0.5 M EDTA, 5 mL 10% SDS.

2.2 Bacterial Strains and Plasmids

2.2.1 Bacterial Strains and Growth Conditions

The *Escherichia coli* bacterial strain used for plasmid maintenance and gateway cloning reactions was DH5α. All cells were grown in liquid LB or SOC media in a shaker incubator at 220 RPM at 37°C.

The *Agrobacterium tumefaciens* bacterial strain used for all floral dip transformations was GV3101. This strain carries the helper plasmid pMP90. All cells were grown in liquid LB or SOC media at 220 RPM at 30°C.

Bacteria containing plasmids were grown in media supplemented with appropriate antibiotics, and all stocks were stored at -80°C in 25% glycerol.

2.2.2 Plasmids

The pDONR™221 (Invitrogen) vector (see Table 1 for details of all plasmids) was used as the donor vector for all Gateway® reactions. This vector carries a kanamycin resistance selectable marker for *E. coli.* The pKGWFS7 (Invitrogen) vector was used as the destination vector for all Gateway™ reactions. This vector carries two selectable markers conferring resistance to spectinomycin in bacteria, and kanamycin for plants (Table 1). This vector also contains the coding sequences for eGFP and GUS being expressed as fusion reporter proteins to allow determination of expression patterns generated by cloned promoter sequences in plants. Both the donor vector and the expression vector contain the *ccdb* gene between *att* sites, which is lethal to DH5α *E. coli*. Presence of this gene prevents growth of unsuccessful recombinants.

The pMP90 helper plasmid carries a selectable marker that confers gentamycin resistance for selection in *A. tumefaciens* (Hellens et al., 2000).

The p19 vector encodes a 19 kDa tomato bushy stunt virus protein, which is a suppressor of post-translational gene silencing in plants (Silhavy et al., 2002). It was used in this study to prevent silencing of reporter genes in transient *in planta* expression experiments. This vector carries a selectable marker that confers resistance to kanamycin.

The 1.2 kb *ADT4* promoter sequence was ordered and is integrated in the pUC57 vector (Bio Basic Inc. J508021-0001) which carries an ampicillin resistance gene for selection in *E. coli.*

2.3 Plant Material and Standard Growth Conditions

Arabidopsis Columbia-0 (Col-0) wild type seeds (stock number CS1092) were obtained from the *Arabidopsis* Biological Resource Centre (ABRC). All *Arabidopsis* seeds were planted in water-saturated soil, covered with plastic wrap, and placed at 4°C to vernalize. After 3 days of vernalization, pots were moved to a Conviron growth chamber and incubated at 22°C with 16 h light and 8 h dark (long

Table 1: Plasmids used in this Study

Plasmid	Selectable Marker		Reference
	Bacteria	**Plants**	
pDONR™221	Kanamycin	--	(Invitrogen)
pKGWFS7	Spectinomycin	Kanamycin	(Invitrogen)
pMP90	Gentamycin	--	(Hellens et al., 2000)
p19	Kanamycin	--	(Silhavy et al., 2002)
pUC57	Ampicillin	--	(Bio Basic Inc.)

day) photoperiod, and 120 μmol•m^2s^1 light. These conditions were also used for screening of primary transformants on MS selective media plates.

Nicotiana benthamiana wild type seeds were generously provided by Dr. Rima Menassa and Hong Zhu (Agriculture and Agri-Food Canada, London, Ontario). All *N. benthamiana* seeds were planted in water-saturated soil, covered with plastic wrap and grown in the same conditions as described for *Arabidopsis.*

Plastic wrap was used to cover all pots and maintain high humidity while seeds germinated, and removed after approximately one week once small seedlings appeared.

2.4 DNA Isolation

2.4.1 Plasmid DNA Isolation from Bacteria

Plasmid DNA for PCR amplification, sequencing and cloning was isolated using a Geneaid Presto™ Mini Plasmid Kit (FroggaBio PDH300) following the manufacturer's instructions.

2.4.2 Plant Genomic DNA Isolation

One to two *Arabidopsis* leaves were placed in a 1.5 mL Eppendorf tube with 200 μL of prepared quick DNA extraction buffer and mashed using a pestle. Another 200 μL extraction buffer was added and the solution was vortexed to mix before centrifugation at 14 000 RPM for 5 min. 300 μL of the supernatant was transferred to a new Eppendorf tube and 300 μL of room temperature isopropanol was added before vortexing and centrifugation at 14 000 RPM for 10 min. Supernatant was removed and pellet was air-dried for 5 min before being dissolved in 100 μL of 10 mM Tris. Isolated DNA was stored at -20°C.

2.5 PCR Amplification and Purification of *ADT* Promoter Regions

Wild type *Arabidopsis* gDNA was used to amplify *ADT* promoter sequences. All primers (Table 2) used for amplification of promoters were designed with *att* sites so that the final PCR products had the necessary *att* sites for Gateway® cloning at 5' and 3' ends. Primer pairs were also designed to have a T_m of no more

Table 2. Primers used in this Study

Primer Name	**Sequence (5' to 3')**	**T_m (°C)**	**Length (bp)**
proattB1ADT1F	GGGGACAAGTTTGTACAAAAAAGCAGGCT**TACCTTTAGAACATATGG**	64.8	47
proattB2ADT1R	GGGGACCACTTTGTACAAGAAAGCTGGGT**CCATAGCAAAGCAGGGAG**	69.6	47
proattB2ADT1Rint	GGGGACCACTTTGTACAAGAAAGCTGGGT**GACTGTTTGCAGTTAGCGG**	69.6	48
proattB1ADT2F	GGGGACAAGTTTGTACAAAAAAGCAGGCT**ACCTTTTCGATTCTAATTCC**	65.5	49
proattB2ADT2R	GGGGACCACTTTGTACAAGAAAGCTGGGT**TGATGTTGTTTTGACGGC**	68.6	47
proattB2ADT2Rint	GGGGACCACTTTGTACAAGAAAGCTGGGT**GGTTCGATGATAACGGC**	69.0	46
ADT3P—F_925	GGGGACAAGTTTGTACAAAAAAGCAGGCT**GGTCTGACAGTGAGACTGC**	68.8	48
ADT3P—R_35	GGGGACCACTTTGTACAAGAAAGCTGGGT**GTTGCCGGAGTATGGGAAGG**	70.8	49
ADT4P—F_1179	GGGGACAAGTTTGTACAAAAAAGCAGGCT**GCCAGCTGATGTGTCAGAGC**	67.2	49
ADT4P—R_1	GGGGACCACTTTGTACAAGAAAGCTGGGT**GGTTTGGTAATGATGGTAAG**	67.7	49
proattB1ADT6F	GGGGACAAGTTTGTACAAAAAAGCAGGCT**TTTGCGGCGATTATAAATTACG**	66.6	51
proattB2ADT6R	GGGGACCACTTTGTACAAGAAAGCTGGGT**GTTTTAGCAATGGCGTC**	68.6	46

[1]F indicates forward, R indicates reverse, int indicates intron.
[2]The four guanines at the beginning of each sequence are recommended by the Gateway™ manual. Underlined sequences are *att* sites, bolded sequences are *ADT* promoter primer sequences.

than 5°C apart from each other, and were analyzed using the Integrated DNA Technologies Oligo Analyzer 3.1 tool to check for the possibility of primer dimers. Sequences the primers were designed to be complimentary to were searched in NCBI to ensure they were unique.

2.5.1 Primer Design and Sequence Amplification

For all promoter sequences except *ADT4,* the PCR protocol used was: 1 cycle of 30 s at 95°C, 35 cycles of 20 s at 95°C, 45 s at 58°C, 1 min at 68°C, and a final extension of 72°C for 5 min before a final hold at 4°C.

Due to the high AT content, amplification of the *ADT4* promoter sequence was unsuccessful with several different sets of primers. Therefore, a vector containing the sequence of the 1.2 kb region upstream of the *ADT4* TrSS was obtained from Bio Basic Inc. The original primers (Table 2) were then used to amplify the region with a modified PCR protocol consisting of: 1 cycle of 2 min at 94°C, 30 cycles of 20 s at 94°C, 10 s at 52°C, 10 s at 48°C, 8 min at 65°C, and a final hold at 4°C.

2.5.2 Gel Electrophoresis of PCR Products and DNA Extraction

PCR fragments were size-separated on a 0.7% agarose gel in 1 X TAE buffer. For visualization under UV light, RedSafe™ Nucleic Acid Staining Solution (FroggaBio Cat. No. 21141) was used to stain the DNA fragments. PCR fragments of the predicted size were excised from the gel and purified using the GenepHlow™ Gel/PCR Kit (FroggaBio Cat. No. DFH300). Purified DNA was resuspended in 30 μL ddH_2O and the concentration was determined using a Nanodrop™ 1000 Spectrophotometer. If the concentration was at least 20 ng/μL and the 260/280 value was at least 1.7, the DNA quality was considered acceptable.

2.6 Gateway Cloning Procedure

Purified PCR fragments were first recombined into the pDONR221™ vector (Invitrogen) using Gateway® BP Clonase™ II Enzyme mix (Invitrogen 11789020)

(Figure 6). Insert DNA was sequenced after cloning to ensure that a correct known *ADT* promoter sequence was present. Inserts were then recombined into the compatible destination vector pKGWFS7 (Invitrogen) using Gateway® LR Clonase™ II Enzyme mix (Invitrogen 11791020). All Gateway™ reactions and vectors were performed and maintained in *E.coli* (Invitrogen, 2003). Figure 6 outlines this procedure in detail.

2.7 Transformations

2.7.1 *E. coli* Transformations

A rubidium chloride procedure (Renzette, 2011) was used to prepare chemically competent DH5α *E. coli* cells. These cells were transformed with plasmid DNA using a heat shock method (Sambrook and Russell, 2001) and all liquid cultures were grown in non-selective media for 1-1.5 h immediately after. Cells were then plated on solid LB media containing the appropriate antibiotics for selection of transformants and grown overnight. The next day, cells were picked with a toothpick and grown overnight in 5 mL liquid LB with appropriate antibiotics to generate liquid cultures for storage.

2.7.2 *A. tumefaciens* Transformations

Electro-competent GV3101 *A. tumefaciens* cells were prepared using the protocol outlined by Weise (2013). These cells were transformed with plasmid DNA using an electroporation method (Weise, 2013), and all liquid cultures were grown without selection in SOC media at for 1.5 h immediately after electroporation to recover. Cells were then plated on solid LB media containing appropriate antibiotics for selection of the PKGWFS7,0 vector and pMP90 helper plasmid. Plates were placed at 30°C for 48 h, and successful colonies were picked and grown in 5 mL LB liquid selective media overnight to generate liquid cultures for storage, and used to generate starter cultures for plant transformations.

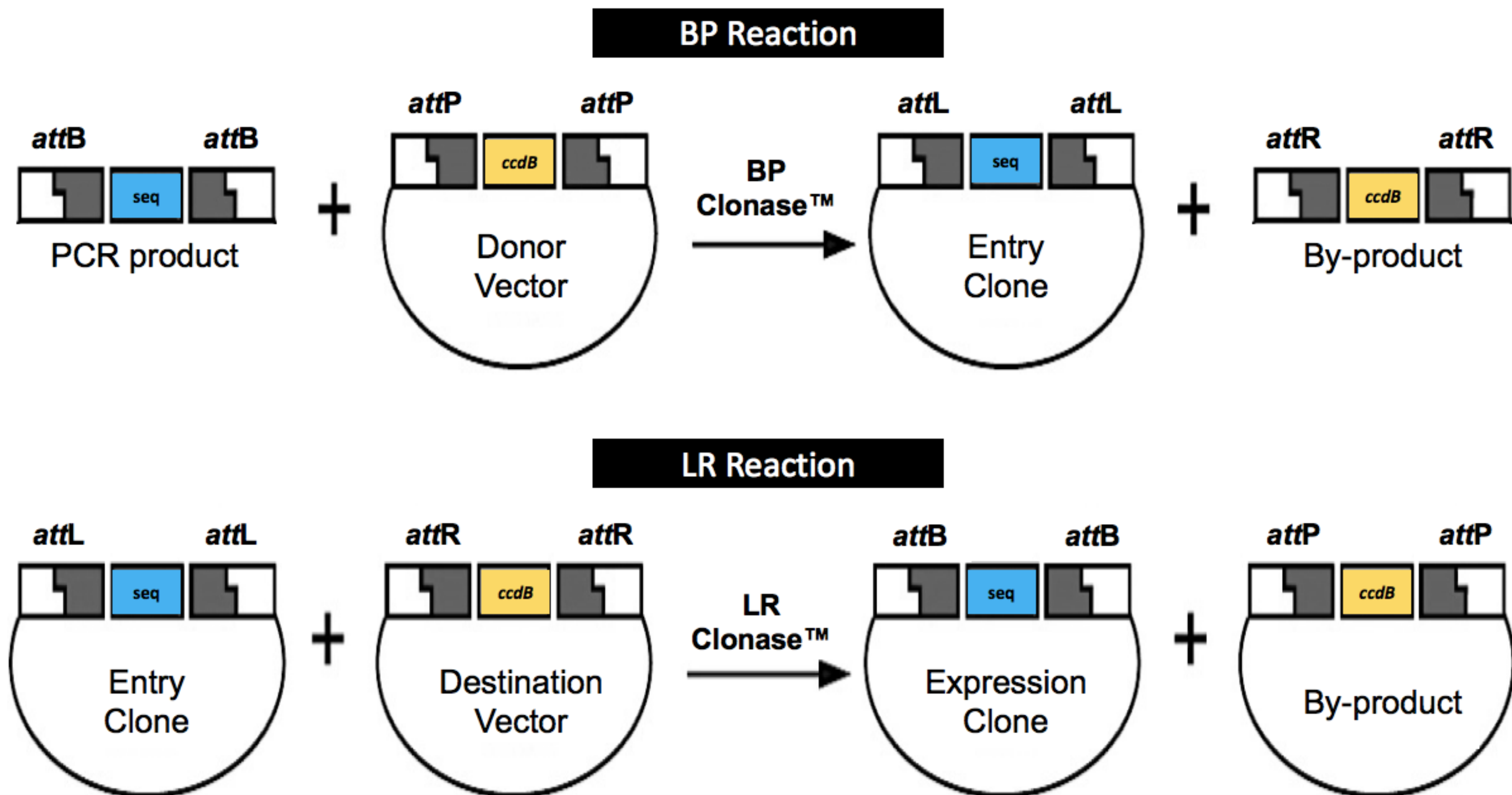
Figure 6. Gateway Cloning Procedure.

After PCR amplification and addition of *att* sites, the sequence is recombined into the donor vector through a BP reaction using BP Clonase™. This reaction results in the displacement of the *ccdB* gene, and replacement with the sequence of interest. The final products of the BP reaction are an entry vector containing the sequence of interest, and the *ccdB* gene fragment as a by-product. The *attL* sites flanking the sequence in the donor vector are then recognizable by the *attR* sites flanking the *ccdB* gene in the destination vector.

In the LR reaction, the sequence of interest is recombined into the destination vector, again replacing the *ccdB* gene using LR Clonase™. The final products of the LR reaction are an expression vector containing the sequence of interest, and the donor vector containing the *ccdB* gene as a by-product.

Blue: sequence of interest
Yellow: *ccdB* gene

Adapted from Invitrogen Gateway Manual (2003).


BP Reaction
attB
attB
seq
PCR product
+
attP
attP
ccdB
Donor
Vector
BP
Clonase™
attL
attL
seq
Entry
Clone
+
attR
attR
ccdB
By-product
LR Reaction
attL
attL
seq
Entry
Clone
+
attR
attR
ccdB
Destination
Vector
LR
Clonase™
attB
attB
seq
Expression
Clone
+
attP
attP
ccdB
By-product

2.7.3 *N. benthamiana* Transformations

Agroinfiltration (Yang et al. 2000) was used for all *N. benthamiana* transformations. Five days prior to infiltration, freezer stocks of *A. tumefaciens* containing the *ADT* promoter constructs were streaked out on LB plates with appropriate antibiotics and grown for 2 days. A single colony from each plate was used to inoculate a 3 mL feeder culture of liquid LB media with appropriate antibiotics, and grown overnight. The next day, 10 μL of each feeder culture was used to inoculate 10 mL of LB media containing appropriate antibiotics. This 10 mL culture was grown until cells were in log phase (OD_{600}= 0.7-0.9). The cultures were then centrifuged at 3000 RPM for 30 minutes and the pelleted cells were resuspended in 10 mL of Gamborg's solution at an OD_{600}= 1.0, and incubated at room temperature for 1 hour at 220 RPM. Undersides of the leaves of 6-week-old *N. benthamiana* plants were inoculated with these final cultures using a blunt-ended syringe. Inoculated plants were placed back into standard growth conditions, and reporter gene expression was analyzed after 4-5 d.

2.7.4 *A. thaliana* Transformations

A floral dip method was used for all stable transformations of *A. thaliana* (Zhang et al., 2006). Feeder cultures of *A. tumefaciens* were prepared as in section 2.7.3. The entire feeder culture was used to inoculate a 500 mL liquid LB culture with appropriate antibiotics which was grown for 16-24 h. Cells were then collected by centrifugation at 3000 RPM for 30 min, and then resuspended in 500 mL of freshly made 5% (wt/vol) sucrose solution in a 1 L beaker. 100 μL of Silwet L-77 was added to the 500 mL solution and swirled to mix. Inflorescences of potted wild type plants were dipped into the *A. tumefaciens* cell suspension for 10 s with gentle agitation, and drained for 3-5 s so that a visible film of the solution could be seen coating the plants. Three plants were dipped for each construct to generate three separate lines of independent stable transformants. All dipped plants were placed laying on their sides, still potted, in separate clear plastic bags and an elastic band was used to close the opening below the pot and maintain high humidity. Plants inside plastic bags were placed

laying down in the dark for 16-24 h. They were then removed from the plastic and placed back into standard growth conditions for one month.

2.8 Seed Collection, Sterilization and Storage

Once siliques of transformed plants started to turn brown, watering was withheld and *Arabidopsis* plants were allowed to dry out. After one month, when completely dry, plants were removed from the growth chamber and bolts were trimmed below the lowest siliques. The bundle of bolts was laid flat on a piece of cheese cloth over a fresh sheet of white paper. The cheese cloth was folded tightly around the bundle and seeds were sloughed off onto the white paper to be sorted into labelled 1.5 mL microcentrifuge tubes. Seeds were sterilized using the bleach method (Zhang et al., 2006), where they were first soaked in 70% ethanol for 1 minute, and then in seed sterilization solution with vigorous vortexing for 2-5 min. Seeds were then rinsed three times with ddH_2O before being plated. All seeds were stored in the dark at room temperature.

2.9 Histochemical Detection of GUS

N. benthamiana Leaves that had been inoculated 3 days prior, or tissues of stably transformed *A. thaliana* were removed from the plants and cut into 1 cm^2 pieces to fit into 12-well tissue culture plates. Leaf pieces were immersed in 1 mL fresh GUS staining solution, vacuum infiltrated for 15 minutes and then incubated overnight at 37°C in the dark. GUS staining solution was then removed and leaves were rinsed with ddH_2O and repeated washes of 90% ethanol until tissue turned clear and select cells were visibly blue. An un-infiltrated or WT leaf was used as a negative control for each assay.

2.10 Confocal Microscopy

Confocal images were generated using an Olympus Fluoview FV1200 confocal laser scanning microscope at Agriculture and Agri-food Canada (London, ON). Slides were prepared by putting a single drop of water per sample in the middle of the slide, and lining the perimeter with Vaseline. Leaf samples

approximately 3 mm^2 in size were cut out of leaves using a razor and placed on the water droplet. A cover slide was placed on top and pressed down lightly to create a tight seal around the leaf. A 63X water immersion objective lens was used for all imaging. A 488 nm laser was used to excite eGFP and emission was collected at 509 nm. For chloroplast autofluorescence, a 559 nm laser was used to excite dsRED2 and emission was collected between 640 and 700 nm.

2.11 *In Silico* Methods

2.11.1 Sequence Analyses

To find sequence similarities and differences, the program Geneious® 8.0.5 was used to compare all six ADTs at the nucleotide level. Global sequence alignments were performed using the default settings under the built-in Geneious Alignment algorithm. Coding and promoter sequences were compared using free end gaps.

2.11.2 Motif Pattern Analysis

The “Cistome” feature on the BAR Database (Austin et al., 2016) was used to analyze the 1 kb region preceding each *ADT* translational start site (TrSS). Cistome detects the presence of regulatory motifs in promoter sequences using published microarray data. Motifs were found from the list of All PLACE Elements based on position-specific scoring matrices (PSSMs). The suggested functional depth cut-off of 0.7 was used for all analyses in this research to ensure consistency and accuracy.

To determine whether the motifs present in each promoter were unique, or less commonly seen in high numbers or at all in other *Arabidopsis* promoters, the default settings (Ze cutoff of 3.0 and expected proportion of 0.5) were used for all analyses when conducting the PSSM motif analysis as described. Motifs yielded from this search are considered “significantly enriched”.

2.11.3 Co-expression Analysis

The “Expression Angler” feature on the BAR Database (Austin et al., 2016) was used to identify genes with the most similar expression patterns compared to *ADTs*. The “top 25 most similar” expression profiles option was chosen to find similar patterns during development, chemical stress, abiotic stress and for root expression, as these treatments were described in detail and biotic stress was not being tested. For each condition and each *ADT*, genes with one or more of the following criteria were selected to be included in this analysis to narrow down the total number of genes and reduce irrelevant information: it is a transcription factor, it is relative to *ADT* function or intracellular communication, it is involved in metabolism or the phenylpropanoid pathway, it has a correlation coefficient of at least 0.9.

3 RESULTS

3.1 *In Silico* Results

Cis regulatory motifs and the transcription factors that bind to them provide specificity and precise control of transcription. This is particularly important in gene families, where nucleotide sequences and protein functions are similar, but expression levels or secondary roles can vary. Since the *ADT* family members are very similar in protein sequence, their promoters are thought to be the reason for their varying levels of expression and specific involvement in distinct biological processes (Bross et al., 2017; Corea et al., 2012; Para et al., 2016). To further understand this differential regulation, the promoter sequences of all six *ADT*s were analyzed using Cistome (Austin et al., 2016) to identify regulatory motif patterns based on those documented in the PLACE database. A co-expression analysis was also conducted to determine whether any transcription factors that recognized those motifs had similar expression patterns when compared to *ADTs*. Co-expression data were analyzed using Expression Angler (Austin et al., 2016) and compared to motif data to identify patterns and potential networks of motifs and other genes. Together, these data will provide a better understanding of the role of promoters and transcriptional regulatory networks in the differential regulation of *ADT*s*.*

3.1.1 Sequence Analyses

Since there are six members of the *ADT* family, nucleotide coding sequences were first compared, and can be viewed as a phylogenetic tree (Figure 7A). Using MEGA7 software, a Maximum Likelihood tree diagram was assembled using *ADT* nucleotide coding sequences. As expected, the nucleotide alignment data were consistent with previous amino acid sequence alignments (Bross et al., 2011). All six nucleotide coding sequences are at least 50% similar to each other (Figure 7B). The most similar sequences are *ADT3* and *ADT6* (74.4%), and *ADT4* and *ADT5* (72.9%), and all pairwise comparisons except those involving *ADT1* and *ADT2* were above 65% similarity. All comparisons between *ADT1* or *ADT2* and another *ADT* were around 55% similarity.

Figure 7. Comparisons of ADT Nucleotide and Promoter Sequences.

A. *ADT* nucleotide coding sequences were used to generate a Maximum Likelihood rooted phylogenetic tree with a bootstrap of 1000 using MEGA7 software. The numbers on branches indicate bootstrapping values, and the horizontal scale represents sequence divergence.

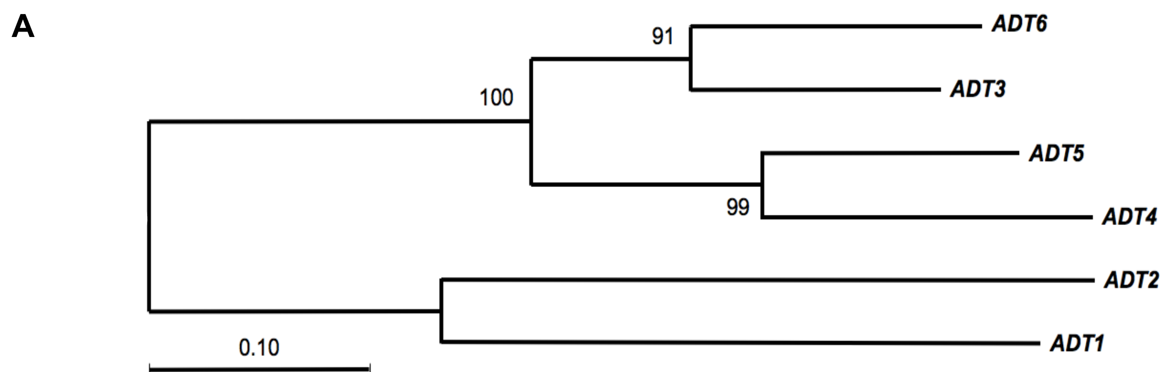
B+C: Pairwise sequence alignments were performed using Geneious® 8.0.5 software. Similarities based on the default settings under the built-in Geneious Alignment algorithm were used for all alignments. Numbers represent the % of identical sites based on the shorter of the two sequences being compared.

B: Complete nucleotide coding sequences.

Accession Numbers for Nucleotide Coding Sequences: AY081528 (*ADT1*), AY113967 (*ADT2*), BT025989 (*ADT3*), BT008862 (*ADT4*), AY090235 (*ADT5*), AY056290 (*ADT6*).

C: Promoter sequences are the 1 kb region preceding each *ADT* TrSS from chromosome sequences in the TAIR database (*ADT1*: At1g11790, *ADT2*: At3g07630, *ADT3*: At2g27820, *ADT4*: At3g44720, *ADT5*: At5g22630, *ADT6*: At1g08250).

Blue colours are a continuous scale from dark (higher % similarity) to light (lower % similarity) used to show similarity patterns.



B: Nucleotide Coding Sequences

	ADT1	*ADT2*	*ADT3*	*ADT4*	*ADT5*	*ADT6*
ADT1		56.1	54.6	54.5	53.4	55.5
ADT2			54.8	53.6	54.2	56.1
ADT3				65.2	66.3	74.4
ADT4					72.9	66.1
ADT5						65.6
ADT6						

C: Promoter Sequences

	ADT1	*ADT2*	*ADT3*	*ADT4*	*ADT5*	*ADT6*
ADT1		36.6	36.2	41.2	38.4	38.9
ADT2			39.7	39.0	36.4	33.8
ADT3				36.0	36.8	35.0
ADT4					45.9	36.4
ADT5						37.8
ADT6						

To determine how similar *ADT* promoters are, all six promoter sequences were aligned (Figure 7C). Sequences aligned were the 1 kb region preceding the TrSS, including any UTRs, and these were used for all *in silico* analyses in this study. All six sequences were at least 33% similar to each other. The most similar promoters were *ADT4* and *ADT5* (45.9%) followed by *ADT1* and *ADT4* (41.2%). All other pairs shared less than 40% similarity.

These sequence similarities indicate that *ADT* promoter sequences show different patterns of sequence similarity than their respective nucleotide coding sequences, and sequence conservation is lower, allowing room for distinct promoter elements. To better understand how these differences contribute to differential *ADT* expression, an analysis of regulatory motifs for each sequence was performed.

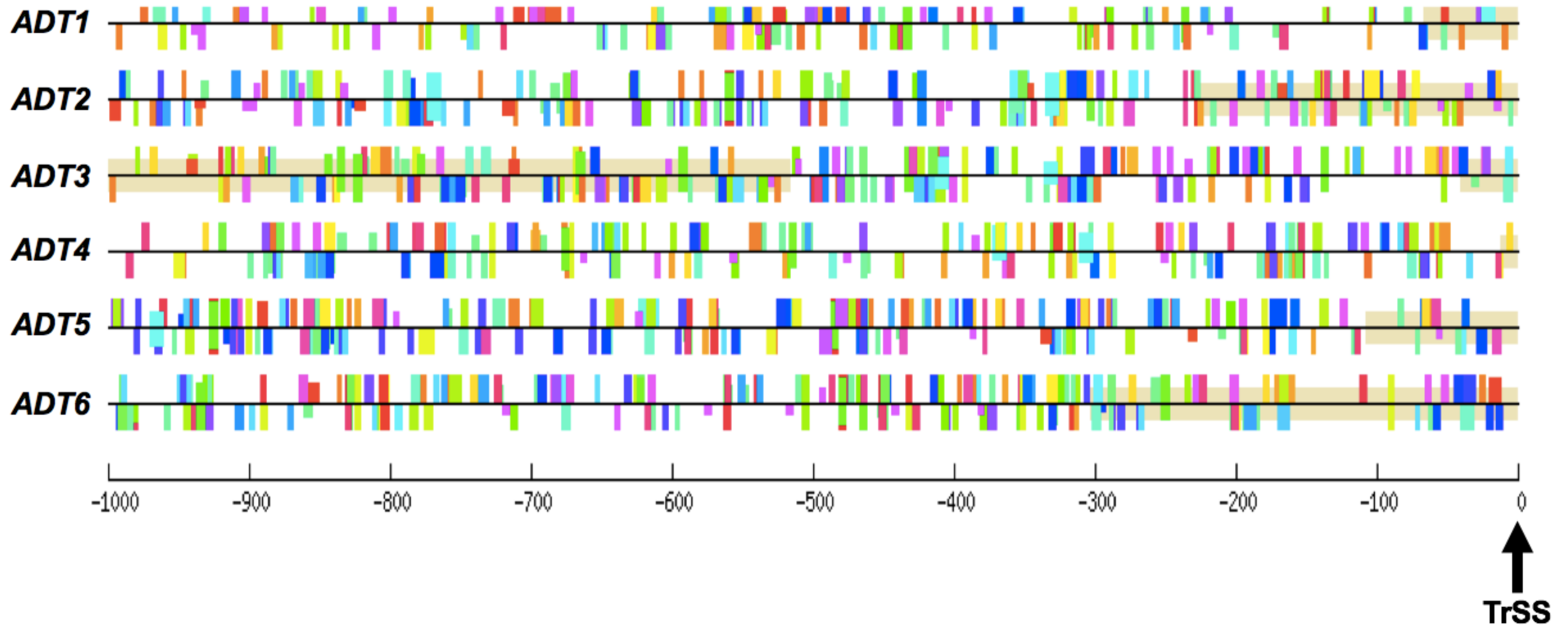
3.1.2 Regulatory Motif Categories

To identify transcription factors that bind to *ADT* promoters, and ultimately make predictions about the functions of individual ADTs, a data mining approach was used to search among known, previously documented motifs in the PLACE database (Higo et al., 1999). The Cistome feature within the BAR database (Austin et al., 2016) was used to identify a total of 121 motifs, each one present in at least one *ADT* promoter. Figure 8 shows the distribution of these 121 motifs (each motif was given a color by the program, hence colors used in this figure are different from the colors used to identify functional categories in all later images). The figure shows that the motifs are found throughout the 1 kb region, with no discernible pattern for any given *ADT* promoter.

As a first step to characterizing the motifs further, motifs were placed into different categories based on their function in *Arabidopsis* (Appendix A). The function of a motif was determined using information found in primary literature. Seven categories were used for sorting: Abiotic Stress (AS), Biotic Stress (BS), Development (D), Stress and Development (SD), Light Response (L), Phenylpropanoid-Related (PR), and General Responses (GR). The AS category

Figure 8. General Distribution of Motifs Present in *ADT* Promoters.

Between 45 and 65 unique motifs were identified in each *ADT* promoter for a total of 121 motifs. This image was generated by Cistome after inputting all 6 *ADT*s as query. The region shown represents the 1 kb region preceding the TrSS. Beige rectangles represent 5' UTRs of *ADT*s and the 3' UTR of the gene upstream from *ADT3.* Small coloured rectangles represent the locations of motifs identified from the PLACE database, where each different colour represents a different motif, and the recurrence of one colour indicates the recurrence of that motif. These colours have been randomly assigned by Cistome and are independent of the colours used to differentiate each category in motif analyses in this study.


ADT1
ADT2
ADT3
ADT4
ADT5
ADT6
-1000
-900
-800
-700
-600
-500
-400
-300
-200
-100
0
TrSS

includes motifs which have been associated with stress conditions such as high or low temperature, salinity, flooding or drought, restricted nutrient availability, or those involved in general abiotic stress responses. Motifs were placed into the BS category if they were involved in pathogen or herbivore responses, wounding (although this type of stress can be caused by abiotic factors such as wind, in this study this type of damage was most often a result of herbivory), or general biotic stress responses. Motifs involved in hormone signalling were either placed into the AS or BS category, depending on the cause/initiation of hormone signalling. In the D category are motifs with specific roles in cell division or differentiation, organ or tissue development, reproduction, and growth. The SD category contains motifs that have multiple roles, either in abiotic and biotic stress responses, or in one or more stress responses in addition to a role in development. The L category contains motifs that mediate response to light, including responses to changes in light intensity or quality (regardless if the changes were stressful or not), circadian rhythms or changes in seasonal cues, and phytochrome signaling. The PR category contains motifs that are only found in the promoters of phenylpropanoid-related genes. To make this category more stringent, motifs were not placed here if they interact with transcription factors that have other roles in stresses or development. The GR category contains core promoter elements, and motifs with general roles in transcriptional regulation, mRNA editing, and tissue-specific expression.

Every motif was only assigned to one category which was not always straightforward, especially if motifs were associated with more than one function. The SD category is somewhat an exception, as it requires that a motif was involved either in biotic and abiotic stress, or in a stress response in combination with a developmental role. In each case, a conservative approach was taken when assigning a motif to a category. For example, if a motif was described to be required for pollen-specific expression, it was not assumed that this motif was involved in pollen development. Hence, it was placed into the GR category rather than the D category.

3.1.3 Proportions of Motif Categories in Each Promoter

Each promoter had motifs from at least 6 of the 7 categories in a unique proportion (Figure 9, Table 3). The only two promoters that contained motifs from all categories were *ADT1* and *ADT4*. The other 4 promoters (*ADT2, ADT3, ADT5, ADT6*) did not contain any motifs from the PR category. It is not surprising that the *ADT1* and *ADT4* promoters have similar proportions, as they had the highest similarity in the pairwise sequence alignment (Figure 6C). However, it is somewhat surprising that all 6 promoters did not contain PR motifs, as all ADTs provide Phe to the phenylpropanoid pathway.

ADT2, ADT3 and *ADT5* promoters were most similar to each other, as over 50% of their total motifs belonged to only 3 categories: GR, AS, and BS (Figure 9, Table 3). The *ADT6* promoter was unique as it contained the highest number of L motifs. A complete list of all motifs and category assignments can be found in Appendix A.

After defining the composition of motifs by broad categories, motifs were analyzed in more detail by defining which ones were found in all *ADT* promoters, which ones were unique to one promoter, and which ones were significantly enriched in a promoter.

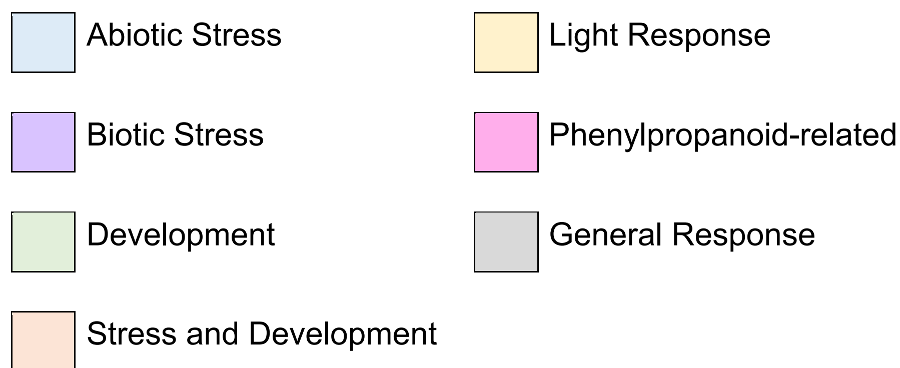
3.1.4 Motifs Common to all Promoters

A total of 12 motifs were present in all six *ADT* promoters (Table 4). The largest number of these motifs belonged to the SD category (5 out of 12). Two motifs in this category, MYCCONSENSUSAT and EBOXBNNAPA, have the same consensus sequence (CATATG), but are recognized by two different members of the bHLH transcription factor family (Toledo-Ortiz et al., 2003). The other SD motifs are involved in multiple abiotic and biotic signaling pathways, including pathogen response, touch and wound response, cold and drought response, or senescence.

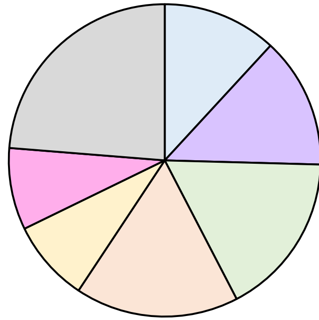
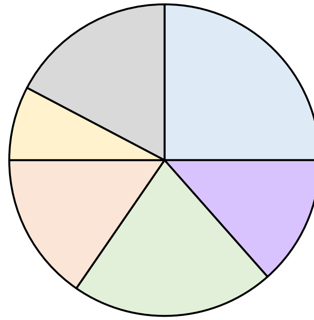
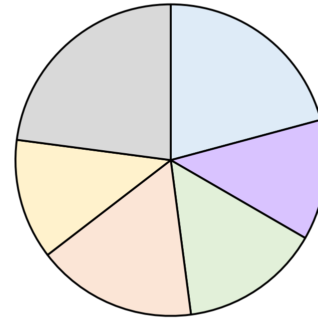
The second most prominent category was GR (3 out of 12) and all 3 were involved in tissue-specific expression. One of these motifs is specifically involved

Figure 9. Categories of Motifs Identified in Each *ADT* Promoter.

Pie charts showing how the total number of motifs are distributed among the 7 categories. Primary data used for generating the pie charts are located in Table 1.



ADT1 ***ADT2*** ***ADT3***

ADT4 ***ADT5*** ***ADT6***

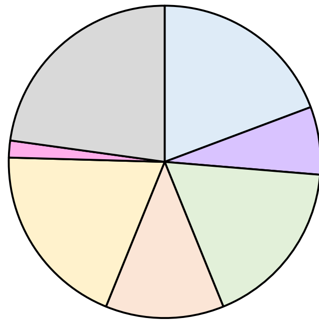
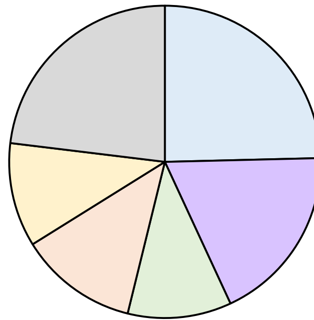
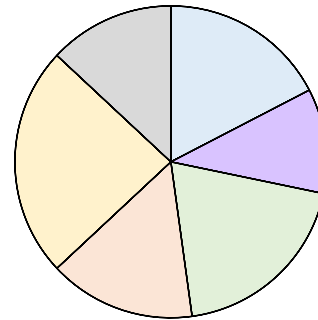




Table 3. Primary Data Used to Generate Pie Charts.

ADT	AB	BS	D	SD	L	PR	GR	Total
1	7 (12%)	8 (14%)	10 (17%)	10 (17%)	5 (8%)	5 (8%)	14 (24%)	59
2	13 (25%)	6 (12%)	11 (22%)	8 (16%)	4 (8%)	--	9 (18%)	51
3	10 (21%)	6 (13%)	7 (15%)	8 (17%)	6 (13%)	--	11 (23%)	48
4	11 (19%)	4 (7%)	10 (18%)	7 (12%)	11 (19%)	1 (2%)	13 (23%)	57
5	16 (25%)	12 (18%)	7 (11%)	8 (12%)	7 (11%)	--	15 (23%)	65
6	8 (17%)	5 (11%)	9 (20%)	7 (15%)	11 (24%)	--	6 (13%)	46

[1]AB= abiotic stress, BS= biotic stress, D= development, SD= stress and development, L= light response, PR= phenylpropanoid-related, GR= general response.
[2]Percentages represent the proportion of a specific category of motifs out of the total number of all motifs.
Data are shown as: number (% of total).

Table 4. Motifs Present in all Six *ADT* Promoters

Motif ID[1,2]	Consensus	TF/Name	Predicted Function	Ref.[3]
ARR1AT	NGATT	ARR1 (Arabidopsis response regulator 1)	Bacterial response	(Sakai et al., 2000)
POLLEN1LELAT52	AGAAA	WRKY34*	Pollen development	(Guan et al., 2014)
DOFCOREZM	AAAG	DOF (DNA binding with one finger)	Positive or negative regulator in numerous signaling pathways	(Yanagisawa, 2004)
RAV1AAT	CAACA	RAV1-A (also ERF4; ethylene response DNA binding factor 4)	Drought/cold stress, touch response, senescence	(Kagaya et al., 1999)
MYCCONSENSUSAT	CATATG	bHLH (basic helix-loop-helix family)	JA-induced wound response	(Abe et al., 2003)
EBOXBNNAPA	CATATG	bHLH (basic helix-loop-helix family)	Light responsive and tissue-specific activation of phenylpropanoid genes*	(Yadav et al., 2005)
GT1GMSCAM4	GAAAAA	GT-3b (GT-1-like transcription factor)	Pathogen and salt stress response	(Park et al., 2004)
GATABOX	GATA	GATA (Type IV zinc finger family)	Light-regulated expression	(Reyes et al., 2004)

Table 4. Continued.

Motif ID[1,2]	Consensus	TF/Name	Predicted Function	Ref.[3]
GT1CONSENSUS	GRAAAW	GT-1 (trihelix DNA binding protein similar to GATA)	Light-regulated expression	(Nagata et al., 2010)
CACTFTPPCA1	TACT	MEM1 (mesophyll expression module 1)	Mesophyll-specific expression	(Gowik et al., 2017)
GTGANTG10	GTGA	Pollen-specific element	Pollen-specific transcription	(Rogers et al., 2001)
CAATBOX1	CAAT	Sequence for tissue-specific expression	Tissue-specific expression	(Fauteux and Strömvik, 2009)

[1]Colours of boxes indicate the predicted function of that motif (purple: biotic stress, green: development, orange: stress and development, yellow: light response, grey: general response).
[2]Motif IDs are the names of each motif from the PLACE database.
[3]Functions are predicted based on the indicated reference.
* indicates that transcription factor or its function is predicted in literature and has not been confirmed experimentally.

in mesophyll expression, which is not surprising, as all six ADTs are localized to chloroplasts (Bross et al., 2017), most of which are found in mesophyll cells. A second motif for tissue specificity is involved in regulating pollen-specific transcription. This compliments the only D motif common to all 6 promoters (POLLEN1LELAT52), which is predicted to be a recognition site for WRKY34, and also required for pollen development (Guan et al., 2014). It is also interesting that two L motifs were present in all 6 promoters, and were either a GATA or GATA-like motif. GATA motifs are commonly found in eukaryotic promoters for regulation of various types of genes, however, in plants they regulate photosynthetic processes through light cues (Reyes et al., 2004). This is not surprising, as *ADTs* are metabolic enzymes, and plants are more metabolically active during daylight. The remaining category is BS, from which only one motif was present in all 6 promoters. This motif is involved in pathogenic bacterial response, which is again not surprising, as some products of the phenylpropanoid pathway in plants are antimicrobials (Vogt, 2010).

Overall, 7 out of 12 total motifs belonged to one of the BS, D, or SD categories, indicating all six ADTs are involved to some degree in development and stress response.

3.1.5 Unique Motifs

One of the goals of this study was to identify regulatory motifs that are unique to only one *ADT* promoter (Table 5), as these might help to predict specialized roles for ADTs.

The *ADT1* promoter had the highest total number of unique motifs (11), and was the only promoter with unique PR motifs. Two of these PR motifs are PAL boxes, which are tripartite sequences found upstream of phenylpropanoid-related genes (Huang et al., 2010; Olsen et al., 2008). There were also two MYB recognition sites, one of which is specifically involved in flavonoid biosynthesis.

There were 4 unique motifs identified in the *ADT2* promoter, and 2 of these belong to the AS category, both of which are involved in drought stress response.

Table 5. Motifs Present in Only One *ADT* Promoter.

ADT	Motif ID[1,2]	Consensus	TF/Name	Predicted Function	Ref.[3]
1	ELRECOREPCRP1	TTGACC	WRKY	Elicitor response	(Schluttenhofer and Yuan, 2015)
	SEF3MOTIFGM	AACCCA	SEF3 (soybean embryo factor 3)	Embryo development	(Fauteux and Strömvik, 2009)
	TGACGTVMAMY	TGACGT	Sequence required for alpha-amylase expression	Seed development	(Yamauchi, 2001)
	GCN4OSGLUB1	TGAGTCA	bZIP	Endosperm expression, environmental response	(Jakoby et al., 2002)
	GLMHVCHORD	ATGAGTCAT	bZIP	Endosperm expression, Environmental response	(Jakoby et al., 2002)
	TATCCACHVAL21	TATCCAC	Part of GA response complex (GARC)	Development in response to environmental cues	(Isabel-LaMoneda et al., 2003)
	MYBPLANT	CACCAACC	MYB	Phenylpropanoid-related gene regulation	(Liu et al., 2015a)
	MYBPZM	CCAACC	MYB	Phenylpropanoid-related gene regulation	(Liu et al., 2015a)
	PALBOXLPC	TCTCACCAACC	Box- L (one of 3 *cis* elements)	Elicitor and light-responsive regulation	(Olsen et al., 2008)

Table 5. Continued.

ADT	Motif ID[1,2]	Consensus	TF/Name	Predicted Function	Ref.[3]
1	PALBOXPPC	TTCTCACCAACCCC	Box- P (one of 3 *cis* elements)	Elicitor and light-responsive regulation	(Olsen et al., 2008)
	-300MOTIFZMZEIN	ATGAGTCAT	Enhancer element	Endosperm expression	(Thomas and Flavell, 1990)
2	ABRELATERD1	ACGTG	ABRE-like sequence	Drought stress response	(Nakashima et al., 2009)
	MYB2AT	TAACTG	MYB2	Drought stress response	(Abe et al., 2003)
	T/GBOXATPIN2	AACGTG	MYC (bHLH family)	Jasmonate and wound response	(Appel et al., 2014)
	SBOXATRBCS	CACCTCCA	S-box; ABI4 (ABA intensive 4)	Light- and sugar-responsive ABA signaling	(Baxter et al., 2012)
3	CARGNCAT	CCWWWWWWWWGG	AGL15 (bZIP family)	Gibberellin response and metabolism	(Hill et al., 2008)
	MRNASTA2CRPSBD	TGAGTTG	mRNA stability determinant	mRNA stabilty and processing	(Nickelsen, 2000)
4	P1BS	GCATATTC	MYB	Phosphate starvation response	(Dubos et al., 2010)

Table 5. Continued

ADT	Motif ID[1,2]	Consensus	TF/Name	Predicted Function	Ref.[3]
4	HDZIP2ATATHB2	TAATAATTA	ATHB-2 (*Arabidopsis thaliana* homeobox protein)	Auxin response and cell proliferation	(Ohgishi et al., 2001)
	RHERPATEXPA7	TCACGT	Root hair-specific element	Root hair distribution	(Zhiming et al., 2011)
	REBETALGLHCB21	CGGATA	GATA-like sequence	Phytochrome signaling	(Kawoosa and Gahlan, 2014)
	SORLREP3AT	TGTATATAT	Sequence found upstream of light-induced genes	PhyA/ light response	(Teakle et al., 2002)
	TATABOX2	TATAAAT	TATA-like motif	Transcription initiation	(Yukawa et al., 2000)
	TATAPVTRNALEU	TTTATATA	TATA-like motif	Transcription initiation	(Yukawa et al. 2000)
5	ANAERO2CONSENSUS	AGCAGC	In promoters of anaerobically-induced genes	Hypoxia response in submerged plants	(Mohanty et al., 2005)
	DRE1COREZMRAB17	ACCGAGA	DRE1 core found upstream of ABA responsive genes	ABA responsive signaling	(Busk et al., 1997)
	PREATPRODH	ACTCAT	Hypo-osmolarity response element	Hypo-osmolarity response	(Satoh et al., 2002)

Table 5. Continued.

ADT	Motif ID[1,2]	Consensus	TF/Name	Predicted Function	Ref.[3]
5	ACGTTBOX	AACGTT	bZIP	Stimulus-dependent gene activation	(Alves et al., 2013)
	AMYBOX2	TATCCAT	Amylase box	Sugar starvation response	(Loreti et al., 2003)
	SP8BFIBSP8BIB	TACTATT	SPF1 (WRKY family)	Sucrose- or polygalacturonic-acid-induced expression	(Ishiguro and Nakamura, 1994)
	RYREPEATVFLEB4	CATGCATG	FUS3 (RY- repeat motif)	Embryogenesis	(Wang et al., 2014)
	MARTBOX	TTTTTTTTTT	T-box found upstream of light-regulated genes	Light-response	(Yukawa et al., 2000)
	SORLIP1AT	GCCAC	Sequence found upstream of light-induced genes	PhyA/ light response	(Kawoosa and Gahlan, 2014)
	NAPINMOTIFBN	TACACAT	Sequence identified upstream of *Napin* genes	*Seed-specific expression	(Ericson et al., 1991)
6	LTRE1HVBLT49	CCGAAA	LTRE (low temperature response element)	Low temperature response	(Catalá et al., 2011)

Table 5. Continued.

ADT	Motif ID[1,2]	Consensus	TF/Name	Predicted Function	Ref.[3]
6	LTRECOREATCOR15	CCGAC	LTRE (low temperature response element)	Low temperature response	(Catalá et al., 2011)
	SEBFCONSSTPR10A	TTGTCTC	SEBF (silencing element binding factor)	Silences PR-10a pathogen defense gene	(Boyle and Brisson, 2001)
	E2FCONSENSUS	ATTCCCCC	E2F	Cell cycle regulation	(Ramirez-Parra et al., 2003)
	ARFAT	TGTCTC	ARF1 (auxin response factor 1)	Auxin signaling response	(Ulmasov, 1997)
	IBOX	GATAAG	MYB	Light-regulated activation	(Hartmann et al., 2005)
	PRECONSCRHSP70A	CCGATTATGAC-ACTCCACCAAGAG	PRE (plastid response element)	Acts as a light-responsive enhancer	(von Gromoff et al., 2006)
	TBOXATGAPB	ACTTTG	GAP box	Light-regulated expression	(Chan et al., 2001)

[1]Colours of boxes indicate the predicted function of that motif (blue: abiotic stress, purple: biotic stress, green: development, orange: stress and development, yellow: light response, pink: phenylpropanoid-related, grey: general response).
[2]Motif IDs are the names of each motif from the PLACE database.
[3]Functions are predicted based on the indicated reference.
ABA: abscisic acid, GA: gibberellic acid, TF: transcription factor

There was also 1 BS and 1 SD motif specific to this promoter, and both have changes in activity in response to different plant hormones.

The *ADT3* promoter had only 2 unique motifs, which is the lowest number out of all 6 promoters. One of these motifs is recognized by the AGAMOUS-LIKE 15 (AGL15) transcription factor, which is part of the MADS family, and is highly expressed during embryogenesis and floral organ development (Becker and Theissen, 2003). AGL15 can act as a transcriptional activator or repressor, and its activity is thought to be controlled by binding to other proteins to form regulatory complexes (Hill et al., 2008). The other is an mRNA stability determinant, and is involved in mRNA processing and steady-state cycling (Nickelsen, 2000).

For *ADT4*, there were 7 unique motifs, including 2 different types of TATA boxes. There were 5 total types of TATA boxes identified in *ADT* promoters (Appendix A), and each promoter had at least one TATA box, but *ADT4* was the only one that had any that were unique to it. *ADT4* was also the only promoter that contained a root hair-specific motif for root hair morphogenesis that is recognized by EXPANSIN A7 (EXPA7) (Kim et al., 2006; Zhiming et al., 2011).

The *ADT5* promoter had 10 unique motifs, with 3 in each of the AS and BS categories. An unexpected motif found in the *ADT5* promoter was ANAERO2CONSENSUS, which is typically found upstream of genes involved in response to very low levels or absence of oxygen. This commonly occurs when soil is waterlogged and the diffusion of oxygen into roots is too slow to meet the needs of respiration during dark, wet conditions (Mohanty et al., 2005).

Figure 9 showed that *ADT6* has the highest proportion of light-response motifs, so it is not surprising that it also contains the highest number of unique light-response motifs. The light response motifs are not specifically involved in light stress responses, but have roles in light-regulated activation and expression (circadian or daylight changes). It also contains 2 low temperature response elements, which are the only temperature-specific motifs from Table 5.

There were three clear patterns from the unique motif analysis. The first is that the *ADT1* promoter contained the highest number of unique motifs, with a large portion of them having roles in environmental interactions, including elicitor

response and flavonoid biosynthesis. The second is that the *ADT5* promoter contained the highest number of stress response motifs from both the AS and BS categories. The third is that the *ADT6* promoter contained the highest number of L category motifs. Although this gives some indication as to what specialized roles these three ADTs might have, the roles of the remaining three are still unclear.

3.1.6 Significantly Enriched Motifs

This study also aimed to determine which of the motifs in *ADT* promoters are not commonly seen in other *Arabidopsis* promoters, as these might set *ADT* promoters apart. An analysis of significantly enriched motifs (motifs that are not usually found in high numbers or at all in other *Arabidopsis* promoters) was therefore conducted (Table 6). Since all 6 ADTs are predicted to be involved to some extent in development and stress response, it was expected that each one might have a select few motifs from the AS, BS, D or SD categories that were significantly enriched.

There were no motifs that were significantly enriched in all 6 promoters. However, the CARGCW8GAT motif, which is recognized by the AGL15 (AGAMOUS-LIKE15) transcription factor (Section 3.1.5), was significantly enriched in 5 out of 6 promoters (all but *ADT5*, in which it was absent)(Appendix A).

Another interesting result is that the ROOTMOTIFTAPOX1 motif occurred 25 times in the *ADT4* promoter. This motif is involved in root elongation and vascular tissue-specific expression (Pastore et al., 2011). Its consensus sequence is ATATT, and all 25 occurrences are evenly spread throughout the intergenic region, are found in roughly the same number on the forward and reverse strands, but are not found in the 5' UTR. Three of these motifs overlap with instances of SEF1MOTIF, a motif involved in flowering time (March-Diaz et al., 2007) with the consensus sequence ATATTTAWW. There are also 5 TATABOX2 motifs present in the *ADT4* promoter, and although all *ADT* promoters contain at least one type of TATA box (Appendix A), *ADT4* is the only one that has the TATABOX2 motif, or any that are significantly enriched (Section 3.1.5). The next highest occurrence after ROOTMOTIFTAPOX1 was the MYCCONSENSUSAT/EBOXBNNAPA motif

Table 6. Significantly Enriched Motifs Present in Each *ADT* Promoter.

ADT	Motif ID[1,2]	Strand[3]	Consensus[4]	Match[5]	Start[6]	End	Instances[7]
1	CARGCW8GAT	>	CTATWTATWG	CTATTTATTG	335	326	3
		<	CTATWTATWG	CTATATATAG	765	774	3
		>	CTATWTATWG	CTATATATAG	774	765	3
	BOXLCOREDCPAL	>	ACCWWCC	ACCAACC	105	99	2
		>	ACCWWCC	ACCTTCC	171	165	2
	CGCGBOXAT	<	ACGCGT		344	349	2
		>	ACGCGT		349	344	2
	INRNTPSADB	>	TTCARTYC	TTCAGTCC	33	26	1
	MYBPLANT	>	CACCAACC		106	99	1
	PALBOXLPC	>	TCTCACCAACC		109	99	1
	ABRERATCAL	>	CACGCGT		350	344	1
	TRANSINITDICOTS	>	AATATGGC		387	380	1
	-300MOTIFZMZEIN	>	ATGAGTCAT		509	501	1
	GLMHVCHORD	>	ATGAGTCAT		509	501	1
2	SITEIIATCYTC	<	TGGGCC		64	69	4
		>	TGGGCC		120	115	4
		<	TGGGCC		171	176	4
		>	TGGGCC		178	173	4

Table 6. Continued.

ADT	Motif ID[1,2]	Strand[3]	Consensus[4]	Match[5]	Start[6]	End	Instances[7]
2	SORLIP2AT	<	GGGCC		65	69	4
		>	GGGCC		119	115	4
		<	GGGCC		172	176	4
		>	GGGCC		177	173	4
	TATCCAOSAMY	<	TATCCA		153	158	3
		<	TATCCA		638	643	3
		>	TATCCA		881	876	3
	CARGCW8GAT	>	CWAAWWAAAG	CAAAAAAAAG	311	302	2
		>	CWAAWWAAAG	CTAATTAAAG	373	364	2
	SBOXATRBCS	>	CACCTCCA		56	49	1
	TRANSINITDICOTS	>	AAAATGGC		545	538	1
	DPBFCOREDCDC3	>	ACACTGG		700	694	1
3	EBOXBNNAPA/ MYCCONSENSUSAT	>	CANWTG	CAAATG	140	135	9
		>	CANWTG	CAAATG	435	430	9
		>	CANWTG	CAAATG	595	590	9
		<	CANWTG	CAATTG	663	668	9
		>	CANWTG	CAATTG	668	663	9
		>	CANWTG	CAGATG	781	776	9

Table 6. Continued.

ADT	Motif ID[1,2]	Strand[3]	Consensus[4]	Match[5]	Start[6]	End	Instances[7]
3	EBOXBNNAPA/ MYCCONSENSUSAT	<	CANWTG	CAAATG	821	826	9
		>	CANWTG	CAAATG	838	833	9
		>	CANWTG	CAGATG	847	842	9
	CARGCW8GAT	>	CWWTTATWTG	CTTTTATATG	336	327	2
		>	CWWTTATWTG	CAATTATTTG	414	405	2
	EECCRCAH1	>	GAGTTGC		276	270	1
	TRANSINITDICOTS	>	ACAATGGC		807	800	1
	CGCGBOXAT	>	ACGCGG		943	938	1
	ABRERATCAL	>	AACGCGG		944	938	1
4	ROOTMOTIFTAPOX1	>	ATATT		298	294	25
		<	ATATT		302	306	25
		<	ATATT		326	330	25
		>	ATATT		402	398	25
		<	ATATT		463	467	25
		>	ATATT		466	462	25
		<	ATATT		594	598	25
		>	ATATT		595	591	25
		<	ATATT		655	659	25
		>	ATATT		658	654	25

Table 6. Continued.

ADT	Motif ID[1,2]	Strand[3]	Consensus[4]	Match[5]	Start[6]	End	Instances[7]
4	ROOTMOTIFTAPOX1	<	ATATT		707	711	25
		>	ATATT		727	723	25
		<	ATATT		736	740	25
		>	ATATT		737	733	25
		<	ATATT		784	788	25
		<	ATATT		804	808	25
		>	ATATT		807	803	25
		<	ATATT		832	836	25
		>	ATATT		877	873	25
		<	ATATT		904	908	25
		<	ATATT		910	914	25
		>	ATATT		913	909	25
		<	ATATT		963	967	25
		<	ATATT		978	982	25
		>	ATATT		981	977	25
	POLASIG3	<	AATAAT		273	278	9
		>	AATAAT		475	470	9
		>	AATAAT		618	613	9

Table 6. Continued.

ADT	Motif ID[1,2]	Strand[3]	Consensus[4]	Match[5]	Start[6]	End	Instances[7]
4	POLASIG3	>	AATAAT		647	642	9
		>	AATAAT		697	692	9
		>	AATAAT		700	695	9
		>	AATAAT		750	745	9
		>	AATAAT		839	834	9
		>	AATAAT		997	992	9
	TATABOX2	>	TATAAAT		62	56	5
		<	TATAAAT		729	735	5
		>	TATAAAT		812	806	5
		<	TATAAAT		858	864	5
		>	TATAAAT		861	855	5
	S1FBOXSORPS1L21	<	ATGGTA		13	18	3
		>	ATGGTA		356	351	3
		<	ATGGTA		434	439	3
	SEF1MOTIF	>	ATATTTAWW	ATATTTAAT	727	719	3
		>	ATATTTAWW	ATATTTATA	737	729	3
		<	ATATTTAWW	ATATTTATA	804	812	3
	P1BS	<	GAATATTC		461	468	2

Table 6. Continued.

ADT	Motif ID[1,2]	Strand[3]	Consensus[4]	Match[5]	Start[6]	End	Instances[7]
4	P1BS	>	GAATATTC		468	461	2
	SORLREP3AT	>	TGTATATAT		484	476	2
		<	TGTATATAT		487	495	2
	CIACADIANLELHC	>	CAACAATATC		310	301	1
	BOXLCOREDCPAL	>	ACCATCC		438	432	1
	DPBFCOREDCDC3	>	ACACAGG		641	635	1
	CARGCW8GAT	>	CTTAATATAG		970	961	1
5	SREATMSD	>	TTATCC		453	448	3
		<	TTATCC		559	564	3
		>	TTATCC		832	827	3
	AMYBOX2	>	TATCCAT		452	446	2
		<	TATCCAT		988	994	2
	NAPINMOTIFBN	>	TACACAT		627	621	2
		<	TACACAT		770	776	2
	PROLAMINBOXOSGLUB1	>	TGCAAAG		478	472	2
		<	TGCAAAG		980	986	2
	RYREPEATGMGY2	<	CATGCAT		801	807	2
		>	CATGCAT		808	802	2

Table 6. Continued.

ADT	Motif ID[1,2]	Strand[3]	Consensus[4]	Match[5]	Start[6]	End	Instances[7]
5	RYREPEATLEGUMINBOX	<	CATGCAT		801	807	2
		>	CATGCAT		808	802	2
	RYREPEATVFLEB4	<	CATGCATG		801	808	2
		>	CATGCATG		808	801	2
	TATCCAYMOTIFOSRAMY3D	>	TATCCAT		452	446	2
		<	TATCCAT		988	994	2
	-300ELEMENT	>	TGAAAAGG		933	926	1
	DRE1COREZMRAB17	>	ACCGAGA		351	345	1
	MARTBOX	>	TTATATTATT		951	942	1
6	LTRECOREATCOR15	>	CCGAC		318	314	3
		<	CCGAC		837	841	3
		<	CCGAC		877	881	3
	INRNTPSADB	>	CTCATTTT		484	477	2
		>	CTCATTTT		713	706	2
	CARGCW8GAT	>	CAATTTTAAG		26	17	1
	CANBNNAPA	>	CGAACAC		418	412	1
	CIACADIANLELHC	>	CAACCAAATC		506	497	1
	E2FANTRNR	>	TTTCCCGC		528	521	1

Table 6. Continued.

ADT	Motif ID[1,2]	Strand[3]	Consensus[4]	Match[5]	Start[6]	End	Instances[7]
6	E2FCONSENSUS	>	TTTCCCGC		528	521	1

[1]Colours of boxes indicate the predicted function of that motif (purple: biotic stress, green: growth and development, orange: stress and development, yellow: light response, grey: environmentally-independent). See appendix for details of motif function.
[2]Motif IDs are the names of each motif from the PLACE database.
[3]> indicates forward strand, < indicates reverse strannd.
[4]W= A or T, R= purine, Y= pyrimidine, N= any nucleotide.
[5]Match represents the actual sequence from a promoter for comparison to the consensus if consensus has ambiguous nucleotides.
[6]Start and end sites indicate the position upstream from the TrSS.
[7]Instances represents the total number of that motif identified in that promoter sequence.

(9 times in the *ADT3* promoter)(Section 3.1.4), and the POLASIG3 motif (9 times in the *ADT4* promoter). The POLASIG3 motif is a plant polyadenylation signal (Ashraf et al., 2014) that is found in the *ADT1, ADT2, ADT4* and *ADT5* promoters, but only significantly enriched in the *ADT4* promoter. All occurrences of this motif were also evenly distributed in the *ADT4* promoter. Another observation is that *ADT6* is the only promoter significantly enriched in the LTRECOREATCOR15 motif, which is a low temperature response element that regulates cold-induced genes involved in anthocyanin biosynthesis (Catalá et al., 2011).

All six promoters contained significantly enriched motifs from the D category, indicating a role for all ADTs in development. All promoters also contained some combination of AS, BS or SD motifs, reiterating the idea that ADTs may be involved in various stress responses. However, there were no clear patterns pointing to one single role for each ADT, suggesting that they may all be involved in developmental and stress response networks to some extent. There were also no correlations in terms of the area of a promoter that a specific motif was enriched in, and each occurrence in a different promoter always occurred in a different spot. Overall, it is clear from this analysis that the *ADT4* promoter is very AT-rich, and few significantly enriched motifs are found in 5' UTRs of any promoters.

3.1.7 Co-expression Results

The co-expression analysis was conducted to identify any genes of interest (as described in section 2.11.3) that have similar expression patterns to *ADT*s (Table 7), as co-expression can be an indication of co-regulation. If any of the co-expressed genes had functions related to ADTs, or were transcription factors that complimented regulatory motifs identified, they are excellent candidates for further investigation.

The Expression Angler feature within the BAR database was used for the co-expression analysis because it contains published microarray data for co-expressed genes under various growth and stress conditions. Since the program

Table 7. Select Co-expressed Genes.

r-value[1]	*ADT*	Name[2]	Location	Function[3]	Reference
Abiotic Stress					
0.871	5	*PHD finger*	Nucleus	Binds to histones during epigenetic regulation (GR)	(Ascencio-Ibanez et al., 2008)
0.869	5	*TFIIS*	Nucleus	Transcription Factor IIS, initiation of transcription (GR)	(Roy and Singer, 2015)
0.864	5	**WRKY23*	Nucleus	Embryo development, auxin response, flavonol biosynthesis (SD)	(Schluttenhofer and Yuan, 2015)
0.860	5	*BGLU15*	Multiple	Beta glucosidase 15; Abiotic stress response and recovery (AS)	(Roepke and Bozzo, 2015)
0.860	5	*BGLU46*	Extracellular	Beta glucosidase 46; Lignin biosynthesis (PR)	(Escamilla-Trevin et al., 2006)
0.790	1	*CYP97A3*	Chloroplast	Carotenoid and xanthophyll biosynthesis (L, SD)	(Kim et al., 2006)
0.769	1	*UGT78D1*	Chloroplast	UDP-glucosyl transferase 78D1; flavonoid biosynthesis (PR)	(Yin et al., 2012)
0.768	4	*WRKY15*	Nucleus	Stress response in mitochondria and chloroplasts (SD)	(Van Aken et al., 2016)
0.766	1	**PDH1*	Chloroplast	Prephenate dehydrogenase; tyrosine biosynthesis (PR)	(Schenck et al., 2014)
0.763	1	**FTSZ2-2*	Chloroplast	Plastidial division protein; chloroplast organization (D)	(Johnson et al., 2015)
0.763	1	**TIC21*	Chloroplast	Translocon at inner membrane of chloroplasts 21, homeostasis and protein import into chloroplasts (GR)	(Teng et al., 2006)
Chemical Stress					
0.948	2	**NRP1*	Cytosol/ Nucleus	Histone chaperone; nucleosome assembly and cell division/differentiation; root formation (D)	(Zhu et al., 2017)
0.934	4	*MYB51*	Nucleus	Pathogen defense; stress response; phenylpropanoid-related genes (SD)	(Gigolashvili et al., 2007)
0.911	4	**WRKY33*	Nucleus	Diverse stress responses (SD)	(Lai et al., 2011)

Table 7. Continued.

r-value[1]	*ADT*	Name[2]	Location	Function[3]	Reference
Chemical Stress					
0.910	4	**CYP81F2*	Cell membrane	Hypoxia and bacterial stress response (SD)	(Xu et al., 2016)
0.753	3	**4CL2*	Nucleus	Phenylpropanoid pathway regulation; stress response (SD, PR)	(Li et al., 2015)
0.725	3	*PAL1*	Cytosol	Phenylalanine ammonia lyase 1; cinnamic acid biosynthesis, stress response and development (SD, PR)	(Bilgin et al., 2010)
0.718	6	**4CL1*	Cytosol/ Nucleus	Fungal and wounding response; phenylpropanoid pathway regulation (SD, PR)	(Li et al., 2015)
Development					
0.902	1	**ATPREP2*	Chloroplast/ Mitochondria	Signal peptide degrading enzyme targeted to mitochondria and chloroplasts (GR)	(Bhushan et al., 2005)
0.877	3	*HUA1*	Cytosol/ Nucleus	Enhancer of AG-4 1; cell fate determination, flower development (D)	(Xu et al., 2016)
0.861	1	*LPA66*	Chloropalst	Low PSII Accumulation 66; RNA modification of chloroplast genes (GR)	(Cai et al., 2009)
0.854	2	*GAS41*	Nucleus/ Chloroplast	Histone acetylation; flowering and development in response to light cues (D)	(Su et al., 2017)
0.851	1	*ATECB2*	Chloroplast	Early chloroplast biogenesis, growth and response to light stimulus (D)	(Yu et al., 2009)
0.851	3	*ARID/ BRIGHT*	Nucleus	Regulates cell cycle and development, glucosinolate metabolism (D)	(Webb et al., 2011)

Table 7. Continued.

Development

r-value[1]	*ADT*	Name[2]	Location	Function [3]	Reference
0.844	3	*FH6*	Nucleus/ Spindle	Actin cytoskeleton, cell component organization; pathogen response (SD)	(Favery et al., 2004)
0.844	3	*DDB1*	Nucleus	Damaged DNA binding; embryo development; protein ubiquitination and stress-induced germination (SD)	(Fernando and Schroeder, 2016)
0.843	2	**GUN1*	Chloroplast/ Nucleus	Nuclear gene regulation and chloroplast- & mitochondrial-nuclear signaling (GR)	(Colombo et al., 2016)
0.843	3	**DYNLL1*	Cytosol	Microtubule motor activity; microtubule-based processes (GR)	(Filatov et al., 2006)
0.842	3	*PGM*	Stromule	Phosphoglucomutase involved in carbohydrate metabolism, detection of gravity and cold response (SD)	(Hoermiller et al., 2017)
0.819	2	**SSN2*	Nucleus	Bacterial response and homologous recombination (SD)	(Song et al., 2011b)
0.807	2	**RSW7*	Kinesin complex	Microtubule-based movement, cytokinesis, mitotic spindle organization, involved in cell wall patterning (D)	(Gillmor et al., 2016)
0.802	2	**ATORC2*	Nucleus	Target of E2F, involved in origin recognition for DNA replication (GR)	(Ascencio-Ibanez et al., 2008)
0.794	2	**ROXY2*	Cytosol, Nucleus	Transcriptional repressor, controls anther development; redox homeostasis (D)	(Wang et al., 2012)
0.750	4	**MYB15*	Nucleus	Cell differentiation, hormone response, multiple stress responses (SD)	(Kim et al., 2017)
0.730	6	*JAR1*	Cytosol, Nucleus	Jasmonate and hormone signaling; phytochrome A signaling and regulation of ROS, pathogen and wound response (SD)	(Nie et al., 2017)
0.729	5	*RAP2.4*	Nucleus	ERF/AP2 transcription factor; light and ethylene signaling(SD)	(Iwase et al., 2017)
0.723	6	*C4H*	Membrane;	Cinnamate-4-hydroxylase: phenylpropanoid metabolism, development, stress response (PR, SD)	(Zhou et al., 2017)

Table 7. Continued.

Development

r-value[1]	*ADT*	Name[2]	Location	Function[3]	Reference
0.721	5	**ATSZF2*	Nucleus	Salt-inducible zinc finger; fungal defense; salt and cold stress response (SD)	(Sun et al., 2007)
0.720	5	*BGLU46*	Extracellular	Beta glucosidase 46; Lignin biosynthesis (PR)	(Escamilla-Trevin et al., 2006)
0.720	5	*ORA47*	Nucleus	DREB subfamily; cell division, wounding and insect defense (SD)	(Chen et al., 2016a)
0.709	4	*ADT5*	Chloroplast	Arogenate dehydratase 5 (PR)	
0.709	5	**ADT4*	Chloroplast	Arogenate dehydratase 4 (PR)	
0.706	5	**RRFT1*	Nucleus	Redox-responsive, member of ERF family, fungal defense, cell division, ethylene signaling (SD)	(Matsuo et al., 2015)
0.703	4	**JAZ5*	Nucleus	Regulates jasmonic acid signaling via protein binding as a co-repressor, pathogen defense, wound response (BS)	(de Torres Zabala et al., 2016)

Root

r-value[1]	*ADT*	Name[2]	Location	Function[3]	Reference
0.896	5	**4CL5*	Cytosol/ Nucleus	Lignin and flavonoid biosynthesis (PR)	(Li et al., 2015)
0.883	4	*C4H*	Cell membrane	Cinnamate-4-hydroxylase: phenylpropanoid metabolism, development, stress response (PR, SD)	(Zhou et al., 2017)
0.881	3	**LysoPL2*	Chloroplast, Plasma Membrane	Lysophospholipase 2; lignin biosynthesis, ROS response (PR)	(Gao et al., 2010)
0.881	5	*CAD5*	Cytosol	Lignin biosynthesis and redox processes (PR)	(Tronchet et al., 2010)

Table 7. Continued.

Root

r-value[1]	*ADT*	Name[2]	Location	Function[3]	Reference
0.880	3	*UGT89C1*	Nucleus, Membrane	Flavonol biosynthesis (PR)	(Kuhn et al., 2016)
0.871	3	*PAT1*	Nucleus, Cytosol	Phytochrome A signaling, photomorphogenesis (SD)	(Torres-Galea et al., 2006)
0.871	3	**NF-YA3*	Nucleus	CCAAT-binding complex, embryogenesis (D)	(Zhao et al., 2017a)
0.860	5	**NAC062*	Nucleus	Cold, defense and chitin response (SD)	(Seo and Park, 2010)
0.850	5	**MYB15*	Nucleus	Cell differentiation, hormone response, multiple stress responses (SD)	(Kim et al., 2017)
0.841	4	**OMT1*	Cytosol, Nucleus, Membrane	Lignin and flavonol biosynthesis (PR)	(Byeon et al., 2014)
0.833	2	*RACK1*	Cytosol, Nucleus, Membrane	Germination, protein complex scaffold activity, hormone signaling (SD)	(Su et al., 2015)
0.819	4	**ZF3*	Nucleus	Cold and salt stress and chitin response (SD)	(Ding et al., 2013)
0.810	4	**4CL5*	Cytosol, Nucleus	Lignin and flavonoid biosynthesis (PR)	(Li et al., 2015)
0.809	4	**MYB63*	Nucleus	Lignin biosynthesis, cell differentiation, hormone signaling (SD)	(Ehlting et al., 2005)
0.809	4	**TOM1*	Golgi stack	Protein transporter activity (GR)	(Yamanaka et al., 2000)
0.808	4	**F6'H1*	Cytosol	Coumarin and phenylpropanoid biosynthesis, redox reactions (PR)	(Schmid et al., 2014)
0.793	6	**LSH9*	Nucleus	Light response and floral development (L, D)	(Schmid et al., 2014)

Table 7. Continued

Root					
r-value[1]	***ADT***	**Name[2]**	**Location**	**Function[3]**	**Reference**
0.768	6	**ADT4*	Chloroplast	Arogenate dehydratase 4 (PR)	

[1]Correlation coefficients above 0.9 are highlighted in yellow.
[2]A * indicates that the CARGCW8GAT motif is significantly enriched in that promoter.
[3]Locations and functions are predicted using literature from the TAIR or NCBI database. Transcription factors are highlighted in blue, and genes involved in the phenylpropanoid pathway or related processes are highlighted in purple. AS: abiotic stress, BS: biotic stress, D: development, SD: stress and development, L: light response, PR: phenylpropanoid-related, GR: general response

was limited to only four conditions, these were the groups of co-expressed genes. The groups for the co-expression analysis are: Abiotic Stress, Chemical Stress, Development, and Root. The Abiotic Stress group contains data from plants that were analyzed under various abiotic stress conditions, including cold, heat, drought and salt stress. In the Chemical Stress group are genes co-expressed with *ADTs* in plants that were treated with various chemicals to inhibit or up-regulate various hormones or signaling mechanisms. The Development group contains data from all stages of development from 1 through 12 weeks (senescence). The Root group contains co-expression data from the various root cell types. Since the correlation coefficients are representative of all treatments combined, and to ensure the co-expression groups were somewhat consistent with the motif categories, changes in *ADT* expression under the specific treatments for each category are not shown, but are located in Appendix B for reference. Since groups varied, the colour scheme for the co-expression analysis is independent from the motif analysis. See Appendix C for co-expressed genes sorted into the categories from Section 3.1.2.

In the Abiotic Stress co-expression group there were 11 genes in total: five with *ADT5,* five with *ADT1,* and one with *ADT4.* The 5 genes with the most highly correlated expression patterns in this group all belonged to *ADT5*. Two WRKY transcription factors were also identified, one co-expressed with *ADT4* and one with *ADT5. WRKY23* (co-expressed with *ADT5*) is involved in development and flavonoid biosynthesis, and *WRKY15* (co-expressed with *ADT4*) is involved in stress response in mitochondria and chloroplasts, which is not surprising again given the chloroplast localization of *ADT*s. There were also two genes involved in flavonoid or lignin biosynthesis, which are two downstream phenylpropanoid-related processes, correlated with *ADT5* and *ADT1*. The most intriguing gene in this category was *FTSZ2-2*, which is a component of chloroplast division machinery. It is interesting that it was highly co-expressed with *ADT1*, as previous research has only identified *ADT2* as possibly having a role in chloroplast division (Section 1.4.1; Bross et al., 2017).

Four out of the five highest overall correlation coefficients belonged to genes in the Chemical Stress group. Three out of those four were with *ADT4*, suggesting a possible role in hormone or chemical signaling networks for stress response in *Arabidopsis.* Two out of the seven total Chemical Stress genes were *MYB51* (defense and stress response through phenylpropanoid-related genes) and *WRKY33* (general stress response). The remaining co-expressed genes in the Chemical Stress group were *PAL1*, and two 4-coumarate:CoA ligases, which are involved in the very first step of the general phenylpropanoid pathway and the very last step, respectively (Huang et al., 2010; Li et al., 2015; Olsen et al., 2008).

The 5th highest overall correlation coefficient was between *ADT1* and *AtPreP2* during Development. *AtPreP2* is a signal peptide-degrading enzyme targeted to mitochondria and chloroplasts (Bhushan et al., 2005). Since all six *ADTs* are localized to chloroplasts, it is interesting that only *ADT1* is highly correlated with *AtPrep2.* Aside from containing one of the 5 highest correlation coefficients, the Development group was the largest with 26 members. There were few surprising results in this group, as most of the genes are involved in cell division, DNA replication, or other growth and developmental processes. One interesting result was *GUN1* (correlated with *ADT2*), a protein involved in chloroplast- and mitochondria-nucleus retrograde signaling. Also, in the Development group, *C4H* is co-expressed with *ADT6,* and in the Root group *C4H* is co-expressed with *ADT4,* suggesting a more important role for ADT4 in the phenylpropanoid pathway in roots. In the Development group *ADT4* and *ADT5* are also co-expressed with each other. This is not surprising considering the similarities between *ADT4* and *ADT5* in terms of sequence and function (Corea et al., 2012; Sections 1.4 and 3.1.1).

There were several interesting results in the Root co-expression group. Firstly, another coumarate-CoA ligase (*4CL5*) was highly co-expressed with *ADT4* and also *ADT5.* Additionally, *ADT4* and *ADT6* expression is correlated in roots. It is also interesting that half (9 out of 18) members of the Root co-expression group have roles in phenylpropanoid metabolism.

The promoters of genes with correlation coefficients of at least 0.9 were further analyzed using Cistome to determine whether they shared any common motifs with *ADT* promoters. The most notable result of this analysis was that 4 out of the 5 promoters of the most highly co-expressed genes were found to contain the CARGCW8GAT (MADS box) motif, and it was significantly enriched in all 4. After this finding, promoters of the other co-expressed genes were further analyzed using Cistome to check for the presence of this motif. Of all 62 genes listed in the co-expression analysis, 34 of them contained this motif (55%), and it was significantly enriched in all 34 (Table 6). The CARGCW8GAT motif was identified before as being recognized by the AGL15 transcription factor (Section 3.1.5).

Overall, there is ample evidence from the *in silico* analysis that each *ADT* might have a more specified role in the cell, whether it be an alternate function (Section 1.4.1) or a particular end use for Phe (ie. flavonoid biosynthesis for UV protection versus antifungal compound). It is also possible that ADTs might have overlapping secondary roles, and that they work together in networks or in dimers to respond to environmental cues. There was also overlap between some transcription factor families in the co-expression analysis with certain motifs identified, making them excellent candidates for future study.

3.2 Cloning and Expression of *ADT* Promoter-Reporter Constructs.

The other goal of this study was to generate stably transformed *Arabidopsis* plants in which *ADT* promoters control eGFP/GUS reporter gene expression. Although transient transformations are fast and show expression in leaves, they cannot be used for multiple experiments over time and cannot show expression in different tissues. Though generation of stable transformants takes much longer, seeds can be grown under different environmental conditions, and expression in all tissues can be analyzed at any given time.

Figure 10. Promoter Region Map Showing Sequences Amplified from gDNA and Cloned

The region amplified included the 5' UTR of an *ADT* and the 5' intergenic region, as well as the 3' UTR of the upstream gene if it existed. The *ADT* is indicated by the label on the left, orange arrows represent the primers used for amplification, and blue arrows represent the upstream genes. If the blue arrow points to the left, that *ADT* promoter is a putative bidirectional promoter. Dark purple blocks represent exons/coding sequences, and thick light purple lines represent introns. The translational start site (TrSS) is indicated by the vertical dotted line.

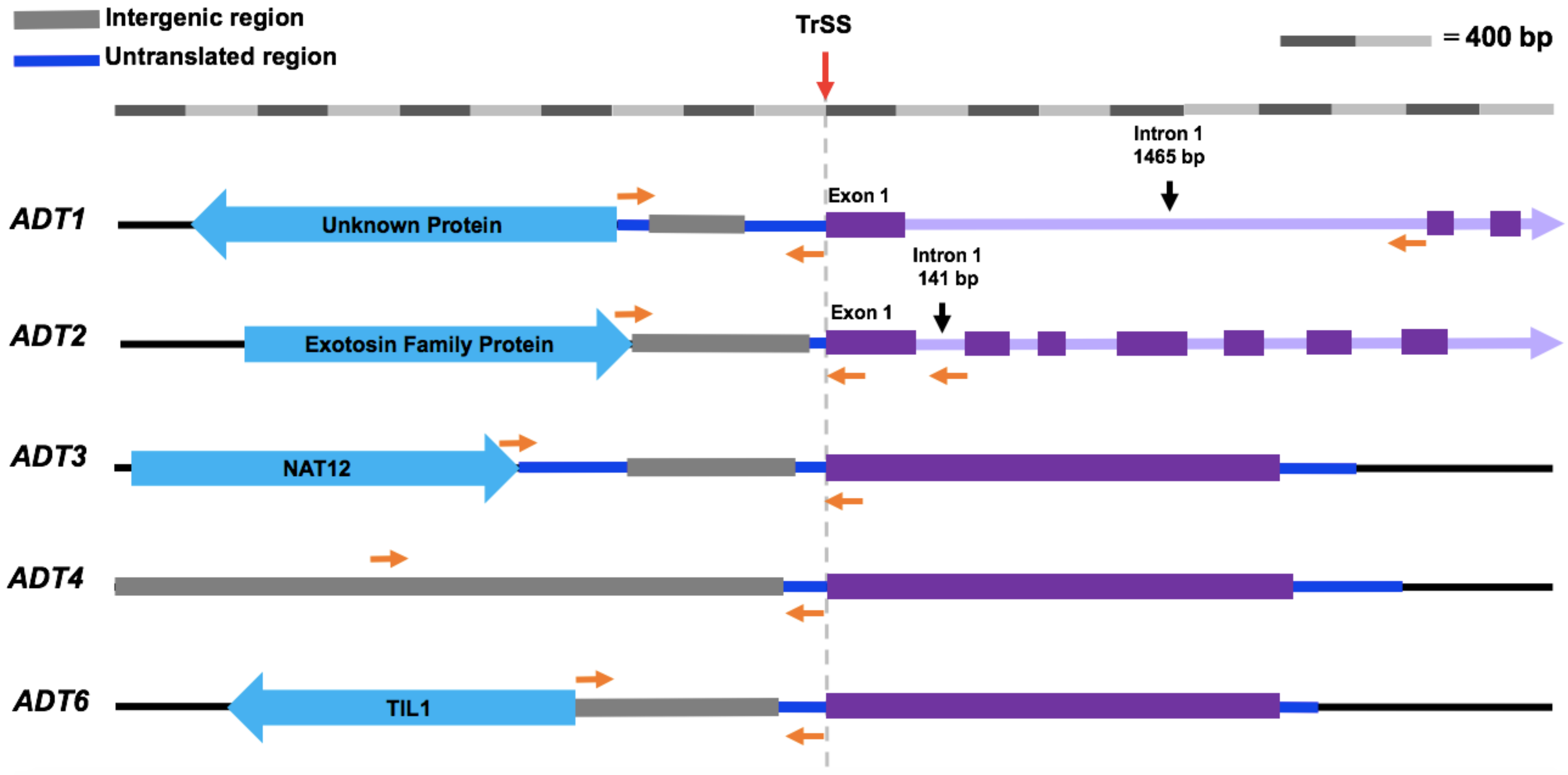
Since *ADT1* and *ADT2* are the only *ADTs* that have introns, the first introns were amplified as well since they could contain regulatory elements. For these two *ADTs*, a second reverse primer was designed to amplify the first intron.

ADT2 and *ADT3* sequences had a high AT content, so primers were designed to be complimentary to sites inside the coding sequences of the upstream gene and the *ADT* itself to ensure the whole promoter region was properly amplified. *ADT4* also had a high AT content, so primers were designed to be outside the 1 kb region as the upstream gene was much further away.

The *ADT5* promoter is not included in this part of the study because it has already been characterized *in planta* by another student in the lab.

Sequence lengths: *ADT1* 600 bp, *ADT1i* 2.2 kb, *ADT2* 550 bp, *ADT2i* 1 kb, *ADT3* 990 bp, *ADT4* 1.2 kb, *ADT6* 750 bp

TAIR Loci: *ADT1*: At1g11790, *ADT2*: At3g07630, *ADT3*: At2g27820, *ADT4*: At3g44720, *ADT5*: At5g22630, *ADT6*: At1g08250.


Intergenic region
Untranslated region
TrSS
= 400 bp
Intron 1
1465 bp
ADT1
Unknown Protein
Exon 1
Intron 1
141 bp
ADT2
Exotosin Family Protein
Exon 1
ADT3
NAT12
ADT4
ADT6
TIL1

3.2.1 Promoter Sequence Amplification from gDNA

The first step in generating stable transformants is the amplification of each promoter sequence from gDNA. Since all *in silico* analyses were completed using the 1 kb region upstream of the TrSS, the promoter sequences that were amplified for use *in planta* were kept as close to this length as possible for consistency. However, since *ADT* promoter sequences are extremely AT rich, and the upstream genes often extend into the 1 kb region, adjustments were made (Figure 10). For *ADT1, ADT2, ADT3* and also *ADT6*, the upstream gene coding sequence was within the 1 kb region upstream of the TrSS, the promoter sequences that were amplified for use *in planta* were kept as close to this length as possible for consistency. However, since *ADT* promoter sequences are extremely AT rich, and the upstream genes often extend into the 1 kb region, adjustments were made (Figure 10). For *ADT1, ADT2, ADT3* and also *ADT6*, the upstream gene coding sequence was within the 1 kb *ADT* region, so these sequences are shorter than 1 kb to ensure only the intergenic regions and UTRs were cloned. For *ADT4*, the next upstream gene was around 5.4 kb away. Since the sequence is highly AT-rich, primers were designed to amplify a sequence approximately 1.2 kb in length to avoid amplification issues.

ADT1 and *ADT2* are the only *ADT*s with introns (Figure 10), and introns have been shown to play a role in gene regulation through regulatory motifs (Section 1.5). Since this study aims to determine the roles of regulatory motifs in differential expression of *ADT*s, the first intron of *ADT1* and *ADT2* coding sequences were also amplified to see whether this has an effect on expression.

All promoter sequences were PCR amplified from WT *Arabidopsis* Col-0 gDNA as template. To ensure the sequences amplified were the right size, gel electrophoresis was used to compare band sizes to the expected promoter sequence sizes from Figure 10. Fragments of the correct size (data not shown) were excised, purified, and recombined into the Gateway® donor vector.

3.2.2 Construct Confirmation

After recombination into the donor vector, inserts were sequenced using M13 primers to ensure there were no errors made during PCR amplification. Plasmids carrying the correct *ADT* promoter sequence were recombined into the destination vector through an LR reaction. To ensure the entire sequence was present and correct, sequences were PCR amplified from isolated destination vector DNA and compared to the same predicted lengths (Compare Figure 10 and Figure 11). The same DNA was also sequenced (data not shown) to double check that there were no errors before transforming the destination vector into *A. tumefaciens.*

3.2.3 Transient Expression by *ADT* Promoter Sequences

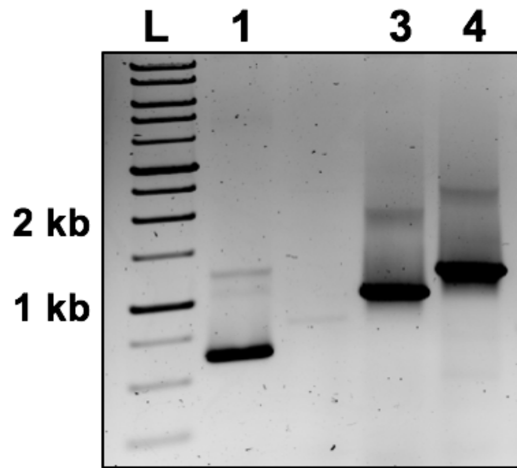
To check whether the amplified promoter sequences were sufficient to drive expression of eGFP/GUS, transformed *Agrobacterium* was used to inoculate leaves of *N. benthamiana.* Three negative controls were used for transient transformations. The controls were an un-infiltrated WT leaf, a leaf inoculated with the empty PKGWFS7 vector only, and a leaf inoculated with the p19 vector only (Figure 12A). It is clear that no eGFP expression is visible for any of the negative controls. Figure 12B shows the characteristic puzzle piece shape of WT leaf epidermal cells from *N. benthamiana* for reference. A pattern of eGFP signal in this shape indicates cytosolic expression.

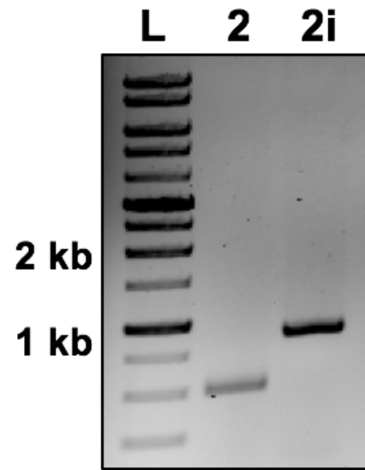
Following transient transformations with *Agrobacterium* strains carrying *ADT* promoter constructs, eGFP reporter gene expression *in planta* was visualized using confocal microscopy. All eGFP expression caused by *ADT* promoter sequences was detected in the cytosol and nuclei of leaf epidermal cells, as the patterns in Figure 13 show the same puzzle piece shape as in Figure 11B. Higher magnification images show eGFP expression is localized around the edges of chloroplasts, but not within (far right columns of Figure 13). These results were as expected. For *ADT1i* and *ADT2i*, the first exon between the promoter region and the first intron (Figure 10) encodes the transit peptide domain of the active *ADT*

Figure 11. Confirmation of Promoter Sequences in Expression Vector.

Once promoter sequences were recombined into the PKGWFS7 expression vector, they were PCR amplified again from isolated plasmid DNA using the same primers from initial promoter amplification to ensure the full sequence of known size was properly cloned. ADT5 is not included in this part of the study because it has already been characterized by a previous student.

The ADT is indicated by the number at the top of its respective lane. The ladder, L, and the 500 bp, 1 and 2 Kb ladder markers are labelled. An i beside a number indicates the presence of an intron in that construct.


L
1
3
4
2 kb
1 kb


L
2
2i
2 kb
1 kb

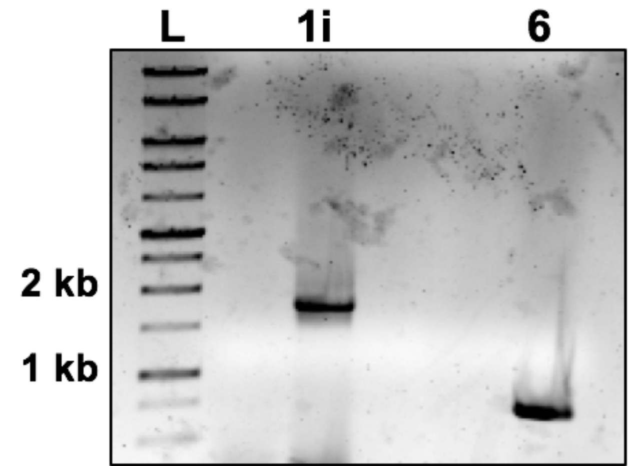
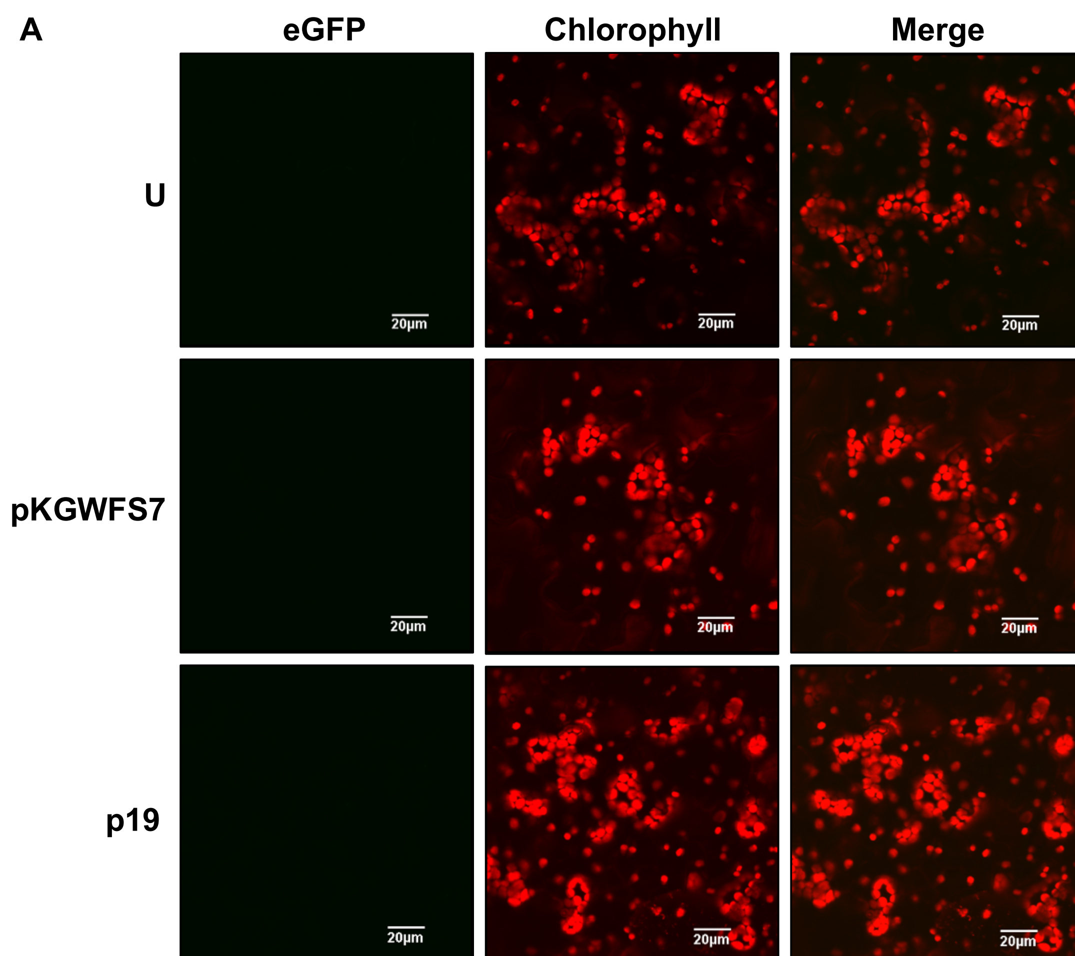

L
1i
6
2 kb
1 kb

Figure 12. Negative Controls for Transient Expression in Tobacco Leaves.

Undersides of 6 week-old *N. benthamiana* leaves were either un-infiltrated, or inoculated with the p19 vector only, or inoculated with an empty pKGWFS7 expression vector as negative controls for all transient transformations. No eGFP expression is visible for any of the control infiltrations. Images are taken 4 dpi with an Olympus Fluoview FV1200 confocal laser scanning microscope. eGFP emission was collected from 470-520 nm (left column), and chlorophyll autofluorescence was collected as dsRed2 from 640-700 nm (middle column). A merge image is shown in the right column.

A. U: Un-infiltrated tobacco leaf, PKGWFS7: tobacco leaf inoculated with empty pKGWFS7 expression vector, p19: Tobacco leaf inoculated with p19 vector only.

B. Light microscope images of *N. benthamiana* epidermal cells. The black box highlights the presence of guard cells in the epidermis.


A
eGFP
Chlorophyll
Merge
U
pKGWFS7
p19
20µm

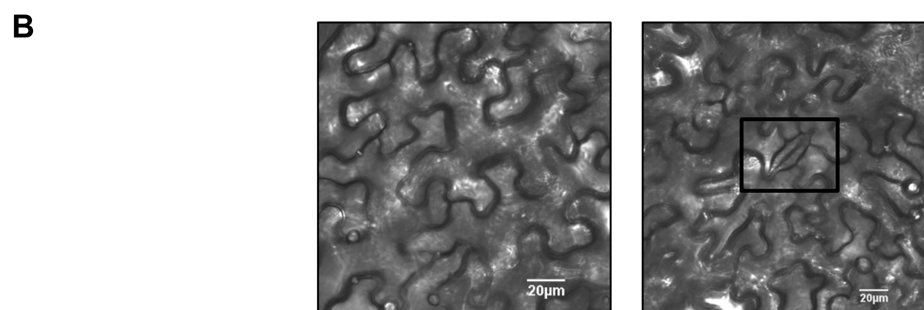

B
20µm

Figure 13. Fluorescence Images of eGFP Expression Driven by *ADT* Promoters.

The undersides of 6-week-old *N. benthamiana* leaves were inoculated with *A. tumefaciens* carrying the expression vectors for each *ADT.* Expression was viewed with an Olympus Fluoview FV1200 confocal laser scanning microscope 4 dpi for all *ADTs* except *ADT3* (3 dpi). eGFP emission was collected from 470-520 nm (far left), and chlorophyll autofluorescence was collected as dsRed2 from 640-700 nm (middle left). For all images, the area in the Zoom Merge column (far right) is the area indicated by the white arrow in the original Merge image (middle right). Expression for all *ADT* promoter constructs is visible in cytosol and nuclei. Green: eGFP, red: chlorophyll autofluorescence.

A. *ADT1*
B. *ADT2*
C. *ADT3*
D. *ADT4*
E. *ADT6*

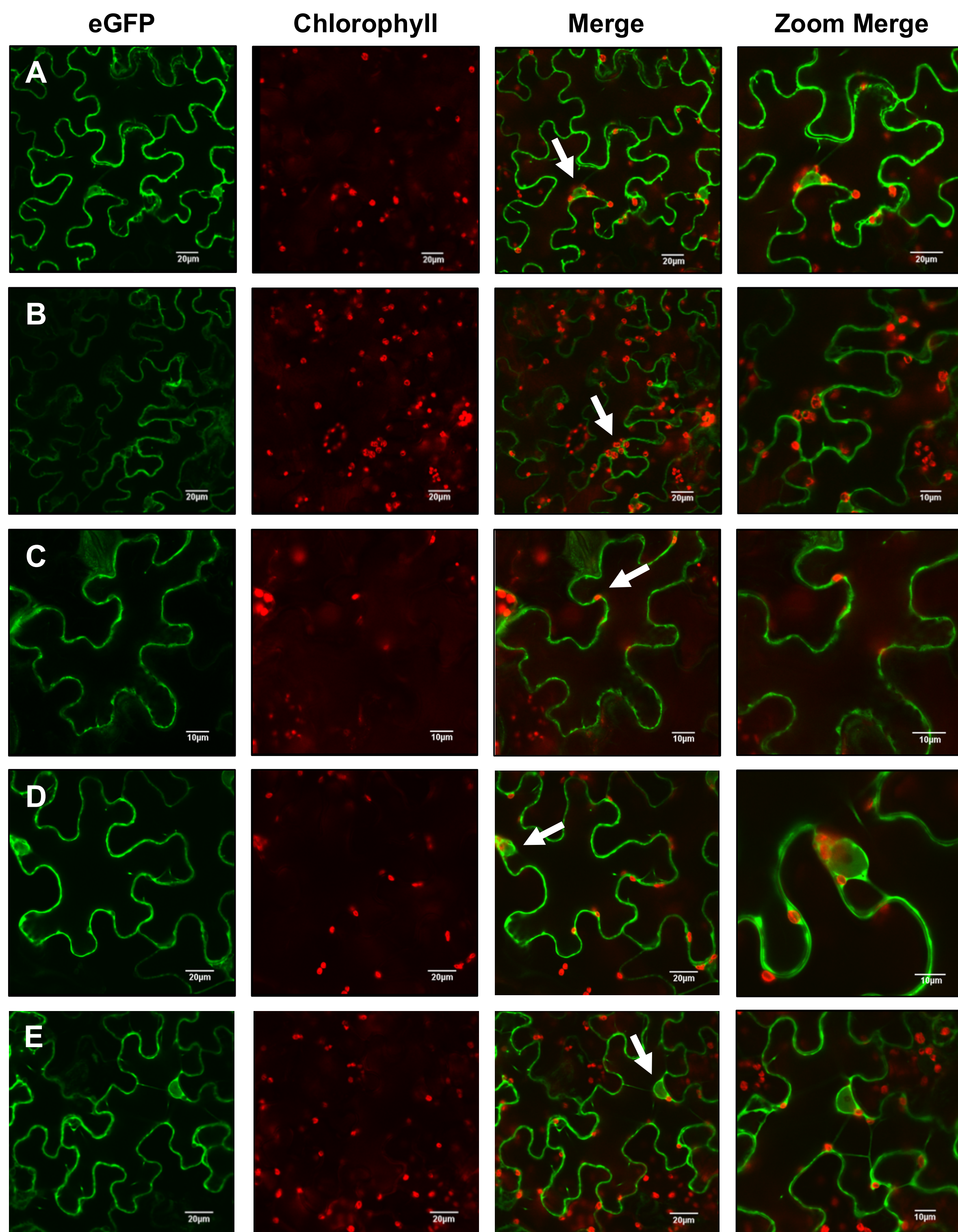
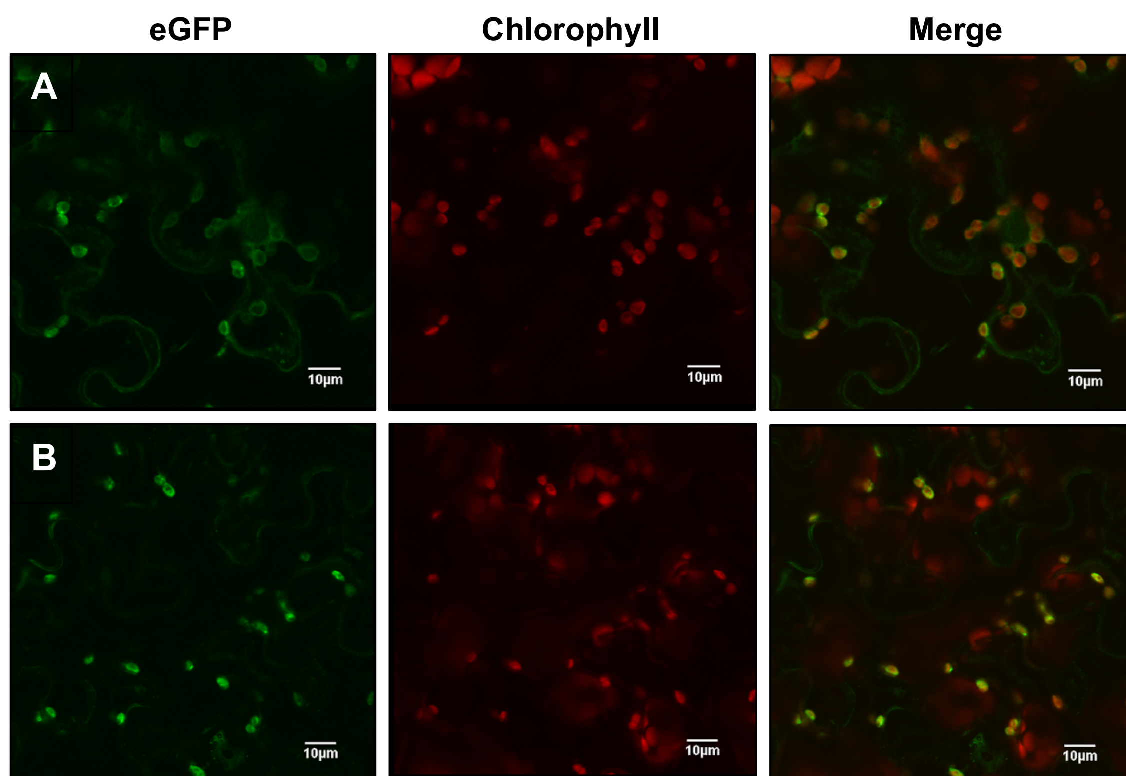

eGFP
Chlorophyll
Merge
Zoom Merge
A
B
C
D
E
20μm
10μm

Figure 14. eGFP Fluorescence Driven by Intron-Containing Constructs.

The undersides of 6-week-old *N. benthamiana* leaves were inoculated with *A. tumefaciens* carrying the expression vectors for each *ADTi.* Expression was viewed with an Olympus Fluoview FV1200 confocal laser scanning microscope 4 dpi. eGFP emission was collected from 470-520 nm, and chlorophyll autofluorescence was collected as dsRed2 from 640-700 nm. Green: eGFP, red: chlorophyll autofluorescence, yellow: overlap.

A: *ADT1i*

B: *ADT2i*


eGFP
Chlorophyll
Merge
A
B
10µm

protein, and is responsible for localization in chloroplasts. This exon was included in the *ADTi* sequences, leading to expression of the eGFP protein containing the transit peptide. As a result, localization of eGFP was almost exclusively in chloroplasts (Figure 14), seen as yellow in the Merge images. It is evident that the transit peptide sequences are necessary and sufficient for localization of eGFP to chloroplasts.

Compared to the other *ADT* constructs, expression of eGFP by the *ADT3* promoter was relatively low. All other constructs were viewed 4 dpi, whereas an image for *ADT3* could only be generated after 3 dpi, indicating a faster turnover rate. Nonetheless, all promoter constructs amplified in this study are sufficient to drive eGFP expression *in planta.*

3.2.4 Stable Transformations of *Arabidopsis*

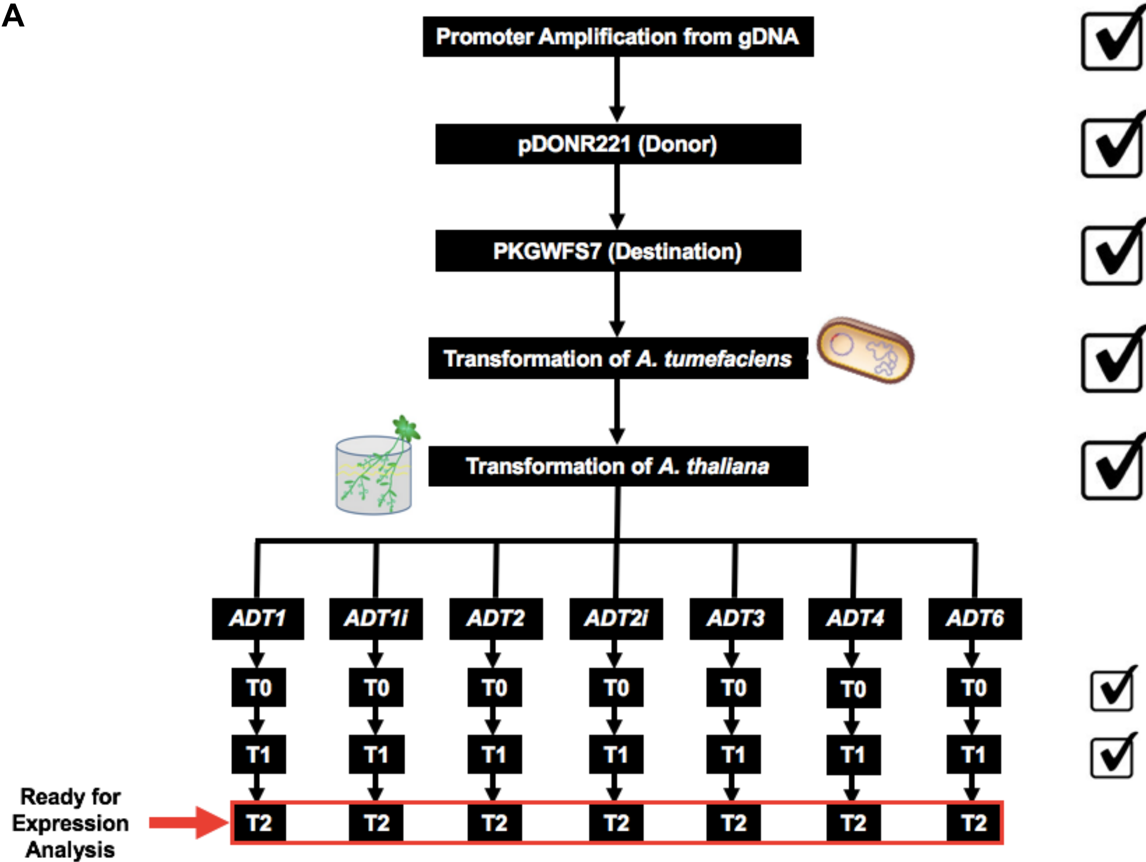
Seven transgenic *Arabidopsis* lines were generated by stable transformation (*ADT1, ADT1i*, *ADT2, ADT2i*, *ADT3, ADT4* and *ADT6*) and seeds were collected and stored. Three independent transformations were performed for each *ADT* to generate three independent lines. This is necessary to ensure expression changes are not due to position effects. Histochemical GUS assays were performed on one leaf from a transformant to view GUS reporter gene expression and confirm the presence of the transgene (Figure 15). Figure 15B represents a stable transformant. GUS reporter gene expression is visible as blue, and shows a specific pattern where it is more highly expressed in leaf veins and in the distal (far with respect to stem) end of the leaf rather than near the petiole. Due to time constraints, only one stable line (*ADT2i*) was successfully selected after 2 generations, and is ready for use in expression analyses.

Figure 15. Stable Expression of GUS in *Arabidopsis* Leaves.

To use them for expression analyses, stable transformants need to be at least 2 generations old (T2) to ensure expression of eGFP and GUS is high enough to make accurate conclusions.

A. Flow chart illustrating the process of stable transformation of *Arabidopsis.* Transformants generated are all still only one generation old. However, each stable line exists and seeds can be harvested for future studies.
B. One stable replicate line of *ADT2i* is in its 2nd generation (T2) and one leaf was stained with GUS staining solution to determine that is was a stable transformant. The seeds from this plant can be used immediately for expression analysis.

A



B

ADT2i -T2

4 DISCUSSION

This study is the first to characterize promoter sequences of all six members of the *Arabidopsis ADT* gene family. To accomplish this, two approaches were used. The first approach was an *in silico* analysis to determine putative regulatory motif patterns for each *ADT* promoter. All motifs identified are known and described in the PLACE database, so it was possible to determine which transcription factors might control *ADT* expression. To compliment the motif analysis, a co-expression analysis was conducted to determine whether any of the transcription factors identified have similar expression patterns to *ADT*s, as this might indicate co-regulation. The second approach was the isolation and cloning of each promoter sequence, and generation of stable *Arabidopsis* transgenic plants using *ADT* promoter-eGFP/GUS reporter constructs. Overall, multiple candidate motifs and transcription factors were identified, and seven stable transgenic *Arabidopsis* lines were generated.

4.1 *ADT* Regulation and Alternative Roles

Transcription is the most common stage of gene regulation, not only because it is the first step in synthesizing a protein, but also because it allows modification of expression levels in response to internal and external environmental cues (Novina and Roy, 1996; Vedel and Scotti, 2011). The ability to alter expression of a gene under any circumstance is essential to every living organism, whether it be for a certain stage of development, or response to a certain stressor. Transcriptional regulation is especially important for plants because they are sessile, and cannot move to avoid stresses or threats, and are constantly competing with their neighbours for resources (Gundel et al., 2014). If a plant lacked proper response mechanisms to everyday fluctuating conditions, survival would be difficult in virtually all environments.

Sequence analyses showed that each *ADT* nucleotide coding sequence is relatively similar (all comparisons above 50% similarity), but promoter sequences are far more distinct (all comparisons below 50% similarity; Figure 7). In all six *ADT* promoters, a TATA box was identified (Appendix A). However, only two TATA

boxes were present in the expected location of 50 to 100 bp upstream: one in the *ADT1* promoter and one in the *ADT4* promoter (data not shown). All other TATA boxes were much further upstream from the TSS (data not shown). This might be explained by the idea of focused versus dispersed transcription (Kadonaga, 2012), where some genes either have multiple weaker recognition sites for transcription initiation (dispersed), one strong site (focused), or some combination of both. However, since the TATA boxes identified in the other *ADT* promoters are further than 500 bp upstream, they may be false hits. Since so little is known about the functions of the other core promoter types (Y-patch, GA, CA, coreless)(Yamamoto et al., 2009, 2011), it is difficult to connect the core promoter type present in a given *ADT* promoter with that *ADT*'s specific role in channelling Phe to a specific pathway.

4.1.1 Do all ADTs have a role in development?

Although the core promoter type can be indicative for some genes or organisms, the remaining regulatory motifs that are present in a promoter can suggest more specific functions for a protein. Based on both the motif analysis and the co-expression analysis, where they were expressed embryonically, it is likely that each *ADT* has some role in development. For instance, the POLLEN1LELAT52 motif is found in all 6 *ADT* promoters (Table 4), and is involved in pollen development (Guan et al., 2014). The GTGANTG10 motif is involved in pollen-specific expression (Rogers et al., 2001), and is also found in all 6 promoters (Table 4). This leads to the prediction that each ADT might be important for pollen development. This would not be surprising, as flavonoid-deficient pollen fails to produce a proper pollen tube for its transport to the ovule, rendering it unable to fertilize any eggs (Cheynier et al., 2013). Also, at least one type of D motif involved in at least one stage of development was significantly enriched in each promoter, including CARGCW8GAT, which was significantly enriched in all promoters but *ADT5* (Table 6). The CARGCW8GAT motif is classified by its CARG sequence, which is recognized by MADS box transcription factors, in this case AGL15 (Airoldi and Davies, 2012). Members of the MADS family of transcription

factors have diverse roles in organogenesis and flowering time, and some MADS protein mutants have late flowering times or abnormal flower morphology (Ng and Yanofsky, 2001; Pastore et al., 2011). For instance, the *apatela1/cauliflower* (MADS) double mutant is responsible for the “cauliflower” phenotype of *Arabidopsis,* where flowers resemble tiny heads of cauliflower (Ng and Yanofsky, 2001). The CARGCW8GAT motif was also significantly enriched in over half of the co-expressed gene promoters (Table 7). The common occurrence of this motif suggests that ADTs may be part of a complex signaling network that regulates multiple developmental decisions.

Overall, as they are expressed in multiple stages of development, it is likely that all six ADTs are important for multiple stages of development, in particular for organogenesis and pollen tube formation, as well as floral organ development and flowering time. More specific roles for ADTs will now be discussed.

4.1.2 A role for ADT1 in flavonoid biosynthesis

It has been shown that *ADT1* plays a role in the synthesis and accumulation of anthocyanins, which are flavonoid-derived pigments (Chen et al., 2016b). The *ADT1* promoter contains the highest number of phenylpropanoid-specific motifs (Appendix A, Figure 9, Tables 3, 5 and 6). These sequences are all variations of PAL boxes, and involved in elicitor and light response, and in flavonoid biosynthesis (Olsen et al., 2008). Compounds derived from flavonoid metabolism are involved in both positive and negative environmental responses (Cheynier et al., 2013; Liu and Murray, 2016; Mierziak et al., 2014; Ng and Yanofsky, 2001; Winkel-Shirley, 2001), which suggests that *ADT1* may be up-regulated when the need for flavonoid biosynthesis increases, for instance if the plant is under pathogen attack, is recruiting symbionts, or if light intensity is too high. In combination with the phenylpropanoid-specific motifs, the *ADT1* promoter also contains a number of unique elicitor response, light response, and developmental motifs (Table 5). As such, *ADT1* may be involved in regulating developmental processes that are sensitive to the environment, such as flowering time, or the changing of leaf colour, which requires Phe synthesis. Accumulation of

anthocyanins is responsible for the change in leaf colour from green to red during the fall (Cheynier et al., 2013). The build-up of these pigments is thought to alleviate the effects of ROS when temperatures drop but there is still ample light, and energy supply is too high for the plant to use efficiently. Leaf colour is also thought to act as a warning sign to herbivores that leaf quality is low, which has potentially facilitated the coevolution of migrating insects (Cheynier et al., 2013). Since migrating insects prefer to feed on green leaves, the change from green to red signifies that it is time to relocate and find new green trees. This also somewhat coincides with *ADT1* having three different motifs involved in elicitor response. Additionally, the heat map in Figure 3 indicates higher expression of *ADT1* than all the other *ADTs* in leaves, especially during the last two weeks before senescence. Accordingly, I predict that *ADT1* plays a role in flavonoid biosynthesis in a manner that is also indicative of the health status of a plant as it approaches senescence, or if conditions are poor.

4.1.3 ADT2: mediator of chloroplast homeostasis?

ADT2 is slightly different from the other ADTs in that it possibly acts as a moonlighting protein, and might play a supplementary role in chloroplast division (Section 1.3.1, Bross et al., 2017) in addition to its role in Phe synthesis. The *ADT2* promoter was significantly enriched in the SORLIP2 light-response motif. This motif is involved in circadian regulation of gene expression (Kawoosa and Gahlan, 2014), which is essential for regulating photosynthetic activity. *ADT2* is also co-expressed with a number of genes that are localized to the nucleus and chloroplast, in particular *GUN1*. GUN1 has been described as a “jack of all trades” in chloroplast homeostasis and signaling (Colombo et al., 2016), as it is involved in processes ranging from chloroplast gene translation, to protein import into chloroplasts, to mediating redox processes. *ADT2* is also co-expressed with *ATORC7,* a part of the kinesin complex, involved in microtubule-based movement away from the nucleus. Together, these data provide a potential indication that *ADT2* may be involved in chloroplast-nucleus communication, possibly mediated by GUN1. This signaling might be in response to abiotic stress (the *ADT2* promoter

does have 2 motifs involved in drought stress response), or in responding to light cues in chloroplasts, and translating photoperiod or redox information between the nucleus and chloroplasts. It is possible that ADT2 is also a “jack of all trades” protein, and aside from chloroplast division, it may be responsible for acting upon received signals from the nucleus to adjust photosynthetic activities under some conditions.

4.1.4 A supportive role in stress and development for ADT3

It has been shown that ADT3 plays a role in ROS homeostasis and cotyledon development (Para et al., 2016), and also in anthocyanin biosynthesis (Chen et al., 2016b). Although no specific motifs involved in any of these processes were identified in the *ADT3* promoter, *ADT3* was co-expressed with *LYSOPHOSPHOLIPASE 2* (*LysoPL2*) during development, which is involved in ROS response (Gao et al., 2010). Furthermore, *ADT3* was also co-expressed with *DAMAGED DNA BINDING PROTEIN 1* (*DDB1*) (involved in stress-induced germination), and *PHYTOCHROME A SIGNAL TRANSDUCTION 1* (*PAT1*) (involved in phytochrome A signaling and photomorphogenesis). Together, these data agree with a role for *ADT3* in ROS response, and might suggest involvement in developmental processes that are regulated by environmental cues. However, this is only a speculation, and there is no strong indication that *ADT3* is specifically involved in any one process.

Some sequences are not recognized by specialized transcription factors, but can be involved in mRNA stability or polyadenylation, and may be important for ribosomal or microRNA attachment, or mRNA editing in different tissues (D’haeseleer, 2006). The MRNASTA2CRPSBD motif, which was specific to the *ADT3* promoter, is involved in mRNA stability and processing (Table 5). The *ADT3* promoter generated lower eGFP expression compared to all other *ADT* promoters (Section 3.2.3), and an image could only be taken 3 dpi rather than 4 dpi (Figure 13), before the signal faded. This suggests that the *ADT3* transcript has a higher turnover rate or is less stable than those of the other *ADTs.* In a study of the MRNASTA2CRPSBD motif in *Chlamydomonas*, it was found to play a role in

allowing a gene transcript to accumulate rather than be degraded (Nickelsen et al., 1999), suggesting that *ADT3* may be less stable. It is possible that instead of channeling Phe into a specific use by the phenylpropanoid pathway, ADT3 plays a supportive role in maintaining ROS homeostasis or anthocyanin biosynthesis. The activity of many proteins is controlled by dimerization with another protein to enhance or reduce activity (Marianayagam et al., 2004). Since *ADT3* has been shown to play a role in anthocyanin biosynthesis, it is possible that it forms a dimer with ADT1 to enhance anthocyanin accumulation. Overall, it is still difficult to speculate about a more specific role for ADT3, but due to its low expression, it is likely that ADT3 is involved in supportive roles in developmental processes and stress responses.

4.1.5 ADT4 and ADT5: leading role in structure and stress response?

It has been shown that *ADT4* and *ADT5* play a role in lignin biosynthesis (Corea et al., 2012), and motifs identified in both promoters supported this finding. Firstly, the *ADT4* promoter was significantly enriched in the ROOTMOTIFTAPOX1 motif. This motif is involved in vascular development and root elongation (Pastore et al., 2011). The *ADT4* promoter was also the only promoter aside from *ADT1* that had PR (phenylpropanoid-related) category motifs, and it has been shown that multiple phenylpropanoid-related genes are highly expressed in roots. This is supported by the fact that roots lignify, and that other metabolites from this pathway are required in roots for defense against soil pathogens, drought stress, nutrient signaling, and recruitment of symbiotic organisms (Cheynier et al., 2013; Hemm et al., 2004). It is possible that *ADT4* has a role in the formation and lignification of roots. The *ADT4* promoter also contained a number of hormone-responsive, AS, and BS motifs. Therefore, one can predict that *ADT4* expression would be highest in roots during any kind of below-ground stress, or hormone signaling to notify the rest of the plant that roots are under stress.

Over half of the motifs unique to the *ADT5* promoter were either AS (abiotic stress) or BS (biotic stress) category motifs, and the five highest correlation coefficients under Abiotic Stress were with *ADT5,* indicating a potential role for

ADT5 in stress response. It has been proposed that ADT5 may be a moonlighting protein that has a role as a transcription factor in the nucleus (Bross et al., 2017). It is possible that abiotic or biotic stress can increase ADT5 expression, and the secondary function can be initiated in the nucleus to bring forward a complete stress response. Additionally, *ADT5* was co-expressed with *β-GLUCOSIDASE 46* (*BLGU46*) in two different groups- Abiotic Stress and Development. BGLU46 is a stem-specific enzyme involved in lignin biosynthesis (Escamilla-trevin et al., 2006). Unpublished data from our lab indicates higher expression of *ADT5* in stems than other *ADTs* (Rad, 2017)(data not shown). Expression data on the BAR database eFP browser (data not shown) also indicates higher *ADT5* expression in stems under standard conditions. This is usually the part of the plant with the highest lignin content (Zhao, 2016), so one can expect *ADT5* to be more highly expressed in stems. Furthermore, lignins play a role in response to mechanical damage and pathogen attack (Bhuiyan et al., 2009). Therefore, it is easy to speculate about the involvement of ADT5 in stress-induced lignin biosynthesis.

Overall, I predict that both ADT4 and ADT5 are involved in lignin biosynthesis that is modulated by both abiotic and biotic stress response.

4.1.6 Does ADT6 have a role in cold acclimatization?

The *ADT6* promoter had the highest number of light response motifs (Table 5). All four of these motifs are involved in regulating expression in response to light cues, but not necessarily light stress. The *ADT6* promoter also contained the only low temperature response elements (LTREs). Temperate plants, including *Arabidopsis,* generally experience 4 seasons in a given year. In turn, they have adopted low-temperature or freezing tolerance mechanisms to survive adverse low temperatures. It has been shown that light is essential to the regulation of proper cold tolerance mechanisms in temperate plants (Catalá et al., 2011). Specifically, light enhances the expression of transcription factors and hormones involved in cold stress response signaling, and also the expression of ROS scavenging proteins (Soitamo et al., 2008). Without light, low temperature would not be as stressful since both the energy source and consumption of energy are decreased.

However, when energy is being provided but metabolism is too slow to use it all, the balance is lost and the plant becomes more stressed (Flügge et al., 2016). Considering the *ADT6* promoter has the most L category motifs, and also the only LTRE motifs, it is possible that ADT6 is involved in cold acclimatization and gene expression during seasonal temperature and daylight changes.

Figure 16 is a summary of the predictions of specialized roles of each ADT. Although the *in silico* analysis did not uncover many obvious connections between an ADT and a specific need for Phe, it was still possible to get a better idea of the reasons for differential *ADT* expression, and make predictions about ADT activity for future experiments.

4.2 Candidate Transcription Factor Gene Families Identified by *in silico* Analyses

Throughout the *in silico* analyses, several members of the same transcription factor families were repeatedly identified, including bHLH, WRKY, bZIP and MYB. The bHLH (basic helix-loop-helix) family consists of at least 147 members with diverse roles in *Arabidopsis* (Toledo-Ortiz et al., 2003), including all stages of development, phytochrome signaling, and some phenylpropanoid-related processes (Toledo-Ortiz et al., 2003). The WRKY family of transcription factors is specific to plants, and there are likely up to 100 members in *Arabidopsis* (Eulgem et al., 2000). This family is known to be involved in a number of defense responses, and has recently been recognized as a regulator of secondary metabolic processes and development (Eulgem et al., 2000). The bZIP (basic leucine zipper) family is made up of 75 distinct members in *Arabidopsis* that regulate diverse processes including pathogen defense, light response and development (Jakoby et al., 2002). The MYB family is present in all eukaryotes, and is made up of over 250 members in *Arabidopsis* (Ambawat et al., 2013; Dubos et al., 2010) with diverse functions from stress response, to cell cycle control, to phenylpropanoid gene regulation in plants (Ambawat et al., 2013).

Figure 16. Summary of Predicted Roles for Each ADT

Functions for each ADT were predicted based on the motif and co-expression analyses. All ADTs are predicted to have a role in development.

ADT1: Biotic Interactions and Flavonoid Biosynthesis

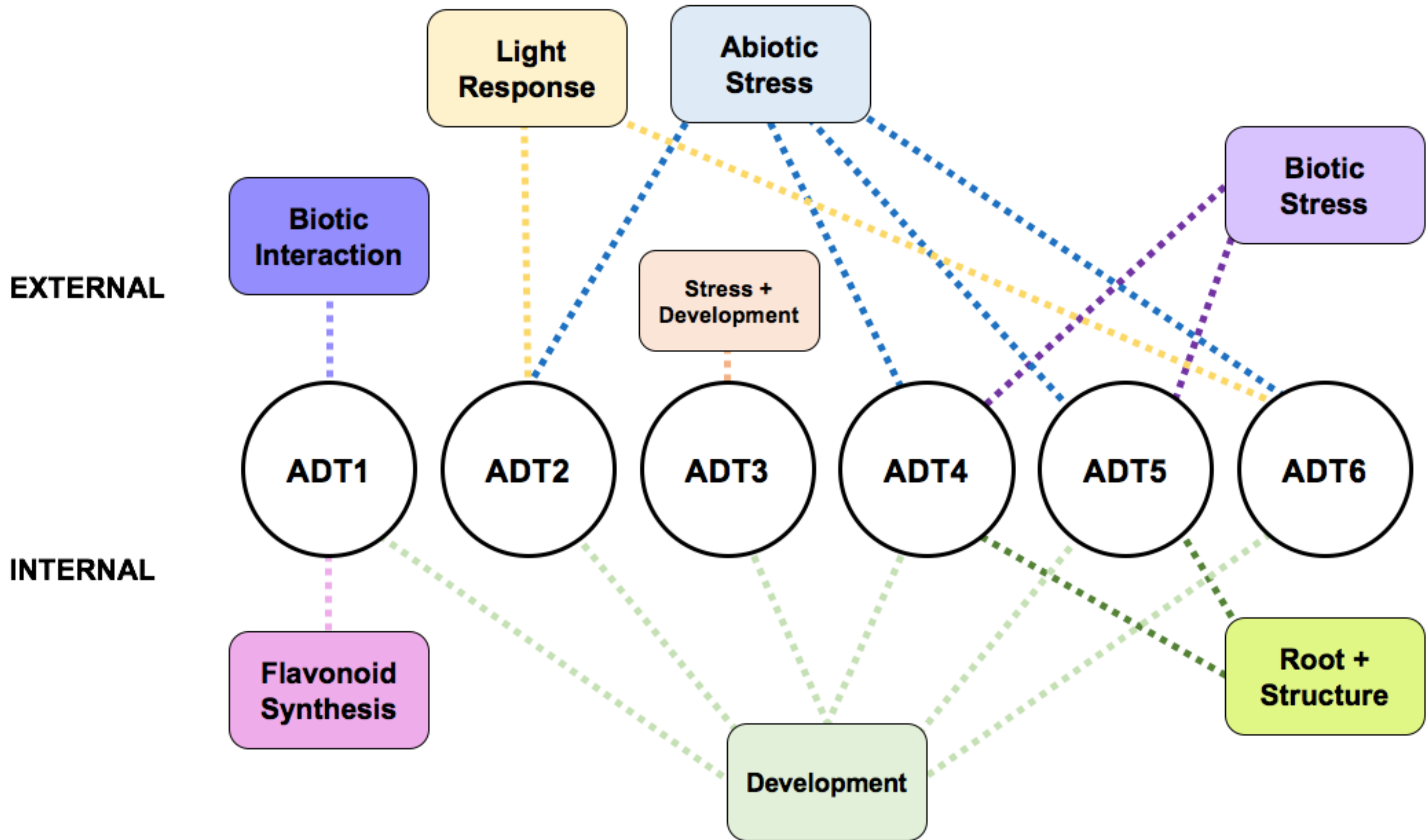
ADT2: Light Response (in particular chloroplast processes), Abiotic Stress

ADT3: General Stress and Development

ADT4: Abiotic and Biotic Stress, Root Expression and Lignin Synthesis

ADT5: Abiotic and Biotic Stress, Root Expression and Lignin Synthesis

ADT6: Light Response and Abiotic Stress



Light Response
Abiotic Stress
Biotic Stress
Biotic Interaction
EXTERNAL
Stress + Development
ADT1
ADT2
ADT3
ADT4
ADT5
ADT6
INTERNAL
Flavonoid Synthesis
Root + Structure
Development

It is not surprising that these four families (bHLH, WRKY, bZIP and MYB) were identified multiple times in the *in silico* analyses. Each one has been shown to play a role in plant specialized metabolism by interacting with each other in complex regulatory networks (Chezem and Clay, 2016; Liu et al., 2015b; Matus, 2016; Schluttenhofer and Yuan, 2015). Phylogenetic analyses suggest that the oldest functions of some bHLH and MYB proteins were in anthocyanin and proanthocyanin production (Chezem and Clay, 2016). In maize, the MYB protein COLOURLESS (C1*)* interacts with the bHLH protein RED (R). This interaction is necessary and sufficient for initiating anthocyanin production through binding to promoters of genes specifically involved in anythocyanin biosynthesis (Chezem and Clay, 2016). The amino acid sequences in the interaction sites of these two proteins, and of other proteins in both families, are also conserved in other species (Zimmermann et al., 2004), suggesting these interactions occurred in an early ancestor of land plants. There are two bHLH motifs present in all six *ADT* promoters, and two MYB binding sites unique to the *ADT1* promoter. It is possible that MYB-bHLH complexes regulate *ADT* expression. Specifically, since *ADT1* is predicted to have a role in flavonoid (and anthocyanin) biosynthesis, MYB-bHLH complexes similar to the C1-R complex might regulate *ADT1* expression.

MYB transcription factors have also been shown to interact with members of the bZIP family. It has been shown in *Arabidopsis* that MYB7 is a negative regulator of the bZIP gene *AB15* (Kim et al., 2015). The AB15 transcription factor is essential for seed germination, and is negatively regulated by MYB7 during unfavourable conditions (Kim et al., 2015). It has also been shown that promoters of WRKY transcription factor genes contain recognition sites for MYB transcription factors (Yang et al., 2013). This suggests a hierarchy for transcriptional control, where MYB and bHLH transcription factors are master regulators of other genes with roles in specialized metabolism (Chezem and Clay, 2016).

The drawback to identifying these large, well-known transcription factor families is that they are likely to be involved in regulating a multitude of specialized metabolic processes and stress responses. This could mean that regulating *ADT* expression is only a side job for some of these transcription factors, so studying

their roles in *ADT* regulation could yield few results. That being said, they are still well-characterized and are a good stepping stone for beginning to understand differential *ADT* regulation.

4.3 Role of Promoter Organization and Introns in Gene Regulation

Motif patterns identified in this study are putative, and their biological relevance is not determined. Therefore, it was important that the patterns seen in *ADT* promoters were representative of biologically relevant motif patterns from other genes or organisms. Non-uniformity in motif distribution is a good indicator that a motif pattern in a given promoter or family of gene promoters is biologically relevant (Casimiro et al., 2008). There was no easily identified distribution pattern in *ADT* promoters, as common motifs were not usually found in the same location, nor the same number in different promoters. This lack of uniformity in distribution and number is an indication that the putative motif patterns found in this study are likely to be representative of a biologically relevant promoter sequence.

Identifying promoter sequences is becoming easier with new sequencing and annotation technologies, and can be accomplished in model organisms by looking at the sequences immediately upstream of a TrSS or TSS (Vedel and Scotti, 2011). Since the *Arabidopsis* genome is sequenced and annotated, and the PLACE, TAIR and NCBI (among others) databases contain information regarding functions of specific genes and motifs, the proximal promoter region is the easiest to characterize. In turn, characterizing this region is an obvious first step towards understanding *ADT* gene family regulation. Although it was possible with this approach to get a better idea of the known signaling pathways ADTs can respond to, the proximal promoter alone does not encompass the entire picture of gene regulation.

There is evidence in literature that introns have a positive effect on gene expression. Removing intron sequences can completely eliminate gene expression, and adding them to an intron-less gene can increase expression (Gallegos and Rose, 2015; Rose et al., 2016). In the case of *ADT1* and *ADT2*, promoter-driven eGFP expression was high enough to visualize, but since

expression is localized to chloroplasts because of the transit peptide, it is difficult to tell whether it is actually higher with the intron than without. The method of intron-enhanced transcription is poorly understood, but believed to involve enhancer elements found within the intron sequence (Rose et al., 2016). It is not clear why *ADT1* and *ADT2* are the only *ADT*s with introns, but considering the effects of introns on gene expression, it is possible that the presence of introns in these gene sequences are not random.

Promoter length can also contribute to the level of expression of a gene and its ability to respond to the environment (Kristiansson et al., 2009). Specifically, longer promoters are thought to be more capable of responding to signaling cascades involved in stress response, as they have a greater range of motifs for transcription factors to recognize. The length of a promoter is also under heavy selection, as a smaller genome can be advantageous when it comes to the speed of replication, and an intergenic region may better sustain variation than a coding sequence (Kristiansson et al., 2009). Generally, the regions between *ADT1* and *ADT2* and their respective upstream gene are shorter than the other *ADTs* (Figure 10). It is possible that the introns in these sequences are advantageous because there is less selection on intron length than there is on promoter length, so the intron sequences might make up for some of the lost promoter length. This might make ADT1 and ADT2 more responsive to signaling pathways involved in environmental or stress response than they otherwise would be.

4.4 The Role of Promoters in Gene Family Evolution

Gene families are a common occurrence in complex organisms, most often arising through single gene or WGD events (Wendel et al., 2009). When multiple copies of a gene are present, functions of these genes become redundant, and through neutral or beneficial mutations or interactions, new functions may arise. It has been shown that genes involved upstream of branched metabolic pathways incur fewer substitutions (Rausher, 2012) which may explain, to some extent, the nucleotide sequence similarities observed between *ADT*s (Figure 6), and the retention of the original dehydratase activity in all six members. This suggests that

the rate of *ADT* evolution might be relatively slow. This also suggests that mutations in coding sequences of genes which control the flux of a substrate into a metabolic pathway are under heavy selection (Rausher, 2012).

It is possible that differential *ADT* expression and channeling of Phe into different end uses of the phenylpropanoid pathway is due to mutations in non-coding regions, specifically promoters, enhancers or introns. If WGD was the reason for the diversification of the *ADT* family, promoter sequences would have been duplicated as well. It was clear that *ADT* promoters are all different despite being similar in coding sequence (Figure 7), suggesting that the rate of mutation in promoters is higher than the rate in coding sequences. It has been shown that only a few changes to regulatory regions can affect gene regulation so much that phenotypes can be modified in few (under 100) generations (Vedel and Scotti, 2011). For example, in *Drosophila biarmipes*, the *YELLOW (Y)* gene controls wing pigmentation and is expressed throughout the wing at low levels. A mutation in the *cis* regulatory region around 1 kb upstream from the *yellow* TSS was shown to cause over-expression of the gene in one area of the wing, creating a spot of dark pigmentation on wings of adult males (Gompel et al., 2005). This trait is passed on to male offspring, and plays a role in sexual selection. The *y* phenotype is therefore an example of how mutations in *cis* regulatory regions can accelerate neo-functionalization over few generations, in this case affecting sexual selection (Gompel et al., 2005).

Promoters are thought to be drivers of rapid adaptation (as in the Y gene) because they can harbour more variation than coding sequences without any negative effects. This is because increasing the number of instances of a motif passed a certain point often has no effect, as long as there are still enough functional copies of that motif (Vedel and Scotti, 2011). Additionally, some transcription factors can still recognize a motif if it contains some variability in nucleotides. This variability in motif sequences can also lead to the creation of new recognition sites for different transcription factors, causing changes in the expression pattern of that gene under different conditions. In turn, promoters are a reservoir of neutral variation, or seemingly unnoticeable changes in tissue-specific

or condition-specific expression, until selection pressure is applied and expression effects become advantageous (Vedel and Scotti, 2011). Therefore, it is possible that promoter variation has been a driving force in allowing differential expression and neo-functionalization of *ADTs*.

4.5 Cloning and Expression Analysis Results

The generation of one stable transgenic line of *Arabidopsis* plants takes several months to complete. To avoid wasting time, it was necessary to determine whether the promoter sequences amplified were sufficient to generate *eGFP* expression *in planta* before following through with stable transformations. It was shown that all seven *ADT* promoter-reporter constructs were sufficient to drive *eGFP* expression *in N. benthamiana* leaves (Figures 13 and 14). Since the amplified promoter sequences were sufficient to generate a reporter signal in tobacco leaves, they are expected to do the same in all tissues of stable *Arabidopsis* transformants. The advantage of having stable transgenic *Arabidopsis* plants is that they can be grown under varying environmental conditions, and expression levels can be determined for all tissues. This is opposed to transient transformations, where reporter gene expression can only be determined for leaves. These stable transgenic plants can be used for a multitude of future studies determining the effects of certain parameters on *ADT* expression in every tissue.

4.6 Limitations and Future Work

An issue that comes with *in silico* analysis of motifs is the inability to determine which individual motifs are biologically relevant. Matches are determined theoretically based on PSSMs. This means only the consensus sequence is considered and there is no experimental evidence that the motif is not a false hit. It is also possible that a motif is recognized by more than one transcription factor, or that its sequence overlaps with another motif. This is especially important for short motifs (4 to 6 bp long) as they may often fall within other motif consensus sequences by chance, or can be ambiguous. The program

used in this analysis also does not consider functional relevance of location within the sequence, nor does it incorporate information from any of the motifs identified together as a group. However, now that promoter sequences have been characterized *in silico,* future studies *in planta* using the transgenic plants generated may provide concrete evidence for some motifs. Since it is more likely that there is an over-representation of motifs present in *ADT* promoters rather than an under-representation, there is now enough information available to begin to decipher functional trends, and to facilitate future studies.

Although promoter sequences are imperative for proper gene expression, as mentioned above, enhancer and intron sequences also contain regulatory motifs. Due to limitations of the program, only the region 1 kb upstream of the translational start site could be analyzed for motifs, so any motifs located in introns have yet to be characterized. Since enhancers can occur thousands of base pairs upstream from a given transcriptional start site, it would be near impossible to determine a definite location for an enhancer specific to an *ADT* with an *in silico* analysis alone.

Limitations aside, this study is the first stepping stone to a number of future studies. Stable transformants that were generated should be grown under varying environmental conditions, and expression checked using confocal microscopy or GUS assay to determine the changes in *ADT* expression. This should first be done to determine any differences in expression between promoter-only and intron-containing constructs. Since there were multiple occurrences of temperature, drought, light-related and pathogen-response motifs, the stress conditions in future studies should reflect these. In particular, since there is no co-expression data available for root cells in response to soil pathogens, *ADT4* and *ADT5* expression should be measured in the roots of stable transformants under these conditions.

To compliment the *in planta* research, it would be helpful to test interactions of motifs and transcription factors using a yeast one-hybrid assay or a chromatin IP (ChIP) assay. This could identify specific transcription factors for further studies that, in the future, could be inhibited or over-expressed, and enhance the plant's response to a particular threat. A good start for this type of study would be to

determine whether any of the MYB, WRKY, bZIP or bHLH motifs identified are recognized by any of the same families of transcription factors identified in the co-expression analysis. These are well-known protein families, and there would likely be a multitude of resources to continue with overexpression or knockdown studies of those transcription factors. The concepts learned here could be applied in the future to crop biotechnology, where synthetic promoters or overexpression of transcription factors could enhance crop survivability in Northern climates, or generally unfavourable environments.

4.7 Conclusions

To date, it is unknown why *Arabidopsis*, and many other plants (Cho et al., 2007; El-Azaz et al., 2016; Maeda et al., 2010) require more than one ADT. Although it has been shown that all are differentially expressed (Cho et al., 2007) and may have neo-functionalized roles (Bross et al., 2017), the ADT family in *Arabidopsis* is still poorly understood. This research is the first in-depth analysis of *Arabidopsis ADT* promoter sequences. Each *ADT* promoter sequence was successfully isolated, and reporter genes were expressed transiently in *N. benthamiana*, and stably in *A. thaliana.* Seeds of stable *A. thaliana* transgenic lines can be grown in future studies under standard and stressful conditions to determine the changes in *ADT* expression *in planta*. Paired with the motif pattern and co-expression data, these *in planta* experiments can provide a well-rounded approach to understanding more about *ADT*s and gene family regulation in plants. This study is also an example of how a data mining approach can be used to analyze existing motif data to compliment an *in planta* analysis, and provide a more complete understanding of how promoters control gene expression.

5 REFERENCES

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

Airoldi, C.A., and Davies, B. (2012). Gene duplication and the evolution of plant MADS-box transcription factors. J. Genet. Genomics *39*, 157–165.

Van Aken, O., De Clercq, I., Ivanova, A., Law, S.R., Van Breusegem, F., Millar, A.H., and Whelan, J. (2016). Mitochondrial and chloroplast stress responses are modulated in distinct touch and chemical inhibition phases. Plant Physiol. *171*, 2150–2165.

Alves, M.S., Dadalto, S.P., Gonzalves, A.B., De Souza, G.B., Barros, V.A., and Fietto, L.G. (2013). Plant bZIP transcription factors responsive to pathogens: A review. Int. J. Mol. Sci. *14*, 7815–7828.

Ambawat, S., Sharma, P., Yadav, N.R., and Yadav, R.C. (2013). MYB transcription factor genes as regulators for plant responses: An overview. Physiol. Mol. Biol. Plants *19*, 307–321.

Appel, H.M., Maqbool, S.B., Raina, S., Jagadeeswaran, G., Acharya, B.R., Hanley, J.C., Miller, K.P., Hearnes, L., Jones, A.D., Raina, R., and Schultz, J.C. (2014). Transcriptional and metabolic signatures of *Arabidopsis* responses to chewing damage by an insect herbivore and bacterial infection and the consequences of their interaction. Front. Plant Sci. *5*, 1-12.

Ascencio-Ibanez, J.T., Sozzani, R., Lee, T.-J., Chu, T.-M., Wolfinger, R.D., Cella, R., and Hanley-Bowdoin, L. (2008). Global analysis of *Arabidopsis* gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol. *148*, 436–454.

Ashraf, M.A., Shahid, A.A., Rao, A.Q., Bajwa, K.S., and Husnain, T. (2014). Functional characterization of a bidirectional plant promoter from cotton leaf curl burewala virus using an *Agrobacterium*-mediated transient assay. Viruses *6*, 223–242.

Austin, R.S., Hiu, S., Waese, J., Ierullo, M., Pasha, A., Wang, T.T., Fan, J., Foong, C., Breit, R., Desveaux, D., Moses, A., and Provart, N.J. (2016). New BAR tools for mining expression data and exploring *cis*-elements in *Arabidopsis thaliana*. Plant J. *88*, 490–504.

Babbitt, C.C., Haygood, R., Nielsen, W.J., and Wray, G.A. (2017). Gene expression and adaptive noncoding changes during human evolution. BMC Genomics *18*, 1–11.

Banerjee, J., Sahoo, D.K., Dey, N., Houtz, R.L., and Maiti, I.B. (2013). An intergenic region shared by At4g35985 and At4g35987 in Arabidopsis thaliana is a tissue specific and stress inducible bidirectional promoter analyzed in transgenic Arabidopsis and tobacco plants. PLoS One *8*, 1-17.

Baranowskij, N., Frohberg, C., Prat, S., and Willmitzer, L. (1994). A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. EMBO J. *13*, 5383–5392.

Baumann, K. (1999). The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and Auxin-regulated expression of the rolB oncogene in plants. Plant Cell *11*, 323–334.

Baxter, L., Jironkin, A., Hickman, R., Moore, J., Barrington, C., Krusche, P., Dyer, N.P., Buchanan-wollaston, V., Tiskin, A., Beynon, J., and Denby, K. (2012). Conserved Noncoding Sequences Highlight Shared Components of Regulatory Networks in Dicotyledonous Plants. Plant Cell *24*, 3949–3965.

Becker, A., and Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol. Phylogenet. Evol. *29*, 464–489.

Bhuiyan, N.H., Selvaraj, G., Wei, Y., and King, J. (2009). Role of lignification in plant defense. Plant Signal. Behav. *4*, 158–159.

Bhushan, S., Ståhl, A., Nilsson, S., Lefebvre, B., Seki, M., Roth, C., McWilliam, D., Wright, S.J., Liberles, D.A., Shinozaki, K., Bruce, B.D., Boutry, M., and Glaser, E. (2005). Catalysis, subcellular localization, expression and evolution of the targeting peptides degrading protease, AtPreP2. Plant Cell Physiol. *46*, 985–996.

Bilgin, D.D., Zavala, J.A., Zhu, J., Clough, S.J., Ort, D.R., and Delucia, E.H. (2010). Biotic stress globally downregulates photosynthesis genes. Plant, Cell Environ. *33*, 1597–1613.

Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. Mol. Cell. *11*, 1587–1598.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. Ann. Rev. Plant Biol. *54*, 519–546.

Boyle, B., and Brisson, N. (2001). Repression of the defense gene PR-10a by the single-stranded DNA binding protein SEBF. Plant Cell *13*, 2525–2537.

Bross, C.D., Corea, O.R.A., Kaldis, A., Menassa, R., Bernards, M.A., and Kohalmi, S.E. (2011). Complementation of the pha2 yeast mutant suggests functional differences for arogenate dehydratases from *Arabidopsis thaliana*. Plant Physiol. Biochem. *49*, 882–890.

Bross, C.D., Howes, T.R., Abolhassani Rad, S., Kljakic, O., and Kohalmi, S.E. (2017). Subcellular localization of *Arabidopsis* arogenate dehydratases suggests novel and non-enzymatic roles. J. Exp. Bot. *68*, 1425–1440.

Busk, P.K., Jensen, A.B., and Pages, M. (1997). Regulatory elements in vivo in the promoter of the abscisic acid responsive gene rab17 from maize. Plant J. *11*, 1285–1295.

Byeon, Y., Lee, H.Y., Lee, K., and Back, K. (2014). Caffeic acid O-methyltransferase is involved in the synthesis of melatonin by methylating N-acetylserotonin in *Arabidopsis*. J. Pineal Res. *57*, 219–227.

Cai, W., Ji, D., Peng, L., Guo, J., Ma, J., Zou, M., Lu, C., and Zhang, L. (2009). LPA66 is required for editing psbF chloroplast transcripts in *Arabidopsis*. Plant Physiol. *150*, 1260–1271.

Casimiro, A.C., Vinga, S., Freitas, A.T., and Oliveira, A.L. (2008). An analysis of the positional distribution of DNA motifs in promoter regions and its biological relevance. BMC Bioinformatics *9*, 1-13.

Catalá, R., Medina, J., and Salinas, J. (2011). Integration of low temperature and light signaling during cold acclimation response in Arabidopsis. PNAS *108*, 16475–16480.

Cercós, M., Gómez-Cadenas, A., and Ho, T.H.D. (1999). Hormonal regulation of a cysteine proteinase gene, EPB-1, in barley aleurone layers: *cis*- and *trans*-acting elements involved in the co-ordinated gene expression regulated by gibberellins and abscisic acid. Plant J. *19*, 107–118.
Chan, C.S., Guo, L., and Shih, M.C. (2001). Promoter analysis of the nuclear gene encoding the chloroplast glyceraldehyde-3-phosphate dehydrogenase B subunit of Arabidopsis thaliana. Plant Mol. Biol. *46*, 131–141.
Chen, H.Y., Hsieh, E.J., Cheng, M.C., Chen, C.Y., Hwang, S.Y., and Lin, T.P. (2016). ORA47 (octadecanoid-responsive AP2/ERF-domain transcription factor 47) regulates jasmonic acid and abscisic acid biosynthesis and signaling through binding to a novel cis-element. New Phytol. *211*, 599–613.
Cheynier, V., Comte, G., Davies, K.M., Lattanzio, V., and Martens, S. (2013). Plant phenolics : Recent advances on their biosynthesis, genetics, and ecophysiology. Plant Physiol. Biochem. *72*, 1–20.
Chezem, W.R., and Clay, N.K. (2016). Regulation of plant secondary metabolism and associated specialized cell development by MYBs and bHLHs. Phytochemistry *131*, 26–43.
Cho, M.H., Corea, O.R.A., Yang, H., Bedgar, D.L., Laskar, D.D., Anterola, A.M., Moog-Anterola, F.A., Hood, R.L., Kohalmi, S.E., Bernards, M.A., Kang, CL., Davin, L.B., and Lewis, N.G. (2007). Phenylalanine biosynthesis in *Arabidopsis thaliana*: identification and characterization of arogenate dehydratases. J. Biol. Chem. *282*, 30827–30835.
Cohen, B.E., Elbuluk, N., Mu, E.W., and Orlow, S.J. (2015). Alternative systemic treatments for vitiligo: a review. Am. J. Clin. Dermatol. *16*, 463–474.
Colombo, M., Tadini, L., Peracchio, C., Ferrari, R., and Pesaresi, P. (2016). GUN1, a Jack-Of-All-Trades in Chloroplast Protein Homeostasis and Signaling. Front. Plant Sci. *7*, 1–14.
C**orea, O.R.A., Ki, C., Cardenas, C.L., Kim, S.J., Brewer, S.E., Patten, A.M., Davin, L.B., and Lewis, N.G.** (2012). Arogenate dehydratase isoenzymes profoundly and differentially modulate carbon flux into lignins. J. Biol. Chem. *287*, 11446–11459.
D'haeseleer, P. (2006). What are DNA sequence motifs? Nat. Biotech. *24*, 423–425.
Danino, Y.M., Even, D., Ideses, D., and Juven-gershon, T. (2015). The core promoter : at the heart of gene expression. BBA Gene Reg. Mech. *1849*, 1116–1131.
Ding, Y., Liu, N., Virlouvet, L., Riethoven, J.-J., Fromm, M., and Avramova, Z. (2013). Four distinct types of dehydration stress memory genes in Arabidopsis thaliana. BMC Plant Biol. *13*, 229.
Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. Trends Plant Sci. *15*, 573–581.
Edger, P.P., and Pires, J.C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosom. Res. *17*, 699–717.

Ehlting, J., Mattheus, N., Aeschliman, D.S., Li, E., Hamberger, B., Cullis, I.F., Zhuang, J., Mansfield, S.D., Samuels, L., Ritland, K., Ellis, B.E., and Douglas, C.J. (2005). Global transcript profiling of primary stems from Arabidopsis thaliana identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. Plant J. *42*, 618–640.
El-Azaz, J., Torre, F. De, and Francisco, M.C. (2016). Identification of a small protein domain present in all plant lineages that confers high prephenate dehydratase activity. Plant J. *87*, 215–229.
Ellerström, M., Stalberg, K., Ezcurra, I., and Rask, L. (1996). Functional dissection of a napin gene promoter: identification of promoter elements requried for embryo and endosperm-specific transcription. Plant Mol. Biol. *32*, 1019–1027.
Ericson, M.L., Muren, E., Gustavsson, H., Josefsson, L., and Rask, L. (1991). Analysis of the promoter region of napin genes from Brassica napus demonstrates binding of nuclear protein in vitro to a conserved sequence motif. Eur. J. Biochem. *197*, 741–746.
Escamilla-trevin, L.L., Chen, W., Card, M.L., Shih, M., Cheng, C., and Poulton, J.E. (2006). Arabidopsis thaliana b-glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. Phytochemistry *67*, 1651–1660.
Eulgem, T., Rushton, P.J., Robatzek, S., and Somssich, I.E. (2000). The WRKY superfamily of plant transcription factors. Trends Plant Sci. *5*, 199–206.
Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K., and Rask, L. (1999). Interaction between composite elements in the napA promoter: Both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. Plant Mol. Biol. *40*, 699–709.
Fauteux, F., and Strömvik, M.V. (2009). Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. BMC Plant Biol. *9*, 1–11.
Favery, B., Chelysheva, L.A., Lebris, M., Jammes, F., Marmagne, A., Almeida-Engler, J. de, Lecomte, P., Vaury, C., Arkowitz, R.A., and Abada, P. (2004). Arabidopsis Formin AtFH6 Is a Plasma Membrane-Associated Protein Upregulated in Giant Cells Induced by Parasitic Nematodes. Plant Cell Online *16*, 2529–2540.
Fernstrom, J.D., and Fernstrom, M.H. (2007). Tyrosine, phenylalanine, and catecholamine synthesis and function in the brain. J. Nutr. *137*, 1539–1547.
Filatov, V., Dowdle, J., Smirnoff, N., Ford-Lloyd, B., Newbury, H.J., and MacNair, M.R. (2006). Comparison of gene expression in segregating families identifies genes and genomic regions involved in a novel adaptation, zinc hyperaccumulation. Mol. Ecol. *15*, 3045–3059.
Flügge, U., Westhoff, P., and Leister, D. (2016). Recent advances in understanding photosynthesis. F1000 Res. *5*, 1–10.
Gagniuc, P., and Ionescu-Tirgoviste, C. (2012). Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. BMC Genomics *13*, 1.
Gallegos, J.E., and Rose, A.B. (2015). The enduring mystery of intron-mediated enhancement. Plant Sci. *237*, 8–15.
Gao, W., Li, H., Xiao, S., Chye, M., Road, P., and Kong, H. (2010). Acyl-CoA-binding protein 2 binds lysophospholipase 2 and lysoPC to promote tolerance to

cadmium-induced oxidative stress in transgenic Arabidopsis. Plant J. *62*, 989–1003.

Gigolashvili, T., Berger, B., Mock, H.P., Müller, C., Weisshaar, B., and Flügge, U.I. (2007). The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in Arabidopsis thaliana. Plant J. *50*, 886–901.

Gillmor, C.S., Roeder, A.H.K., Sieber, P., Somerville, C., and Lukowitz, W. (2016). A genetic screen for mutations affecting cell division in the Arabidopsis thaliana embryo identifies seven loci required for cytokinesis. PLoS One *11*, 1–21.

Gompel, N., Prud, B., Wittkopp, P.J., Kassner, V.A., and Carroll, S.B. (2005). Chance caught on the wing: cis -regulatory evolution and the origin of pigment patterns in Drosophila. Nature *433*, 481–487.

Gowik, U., Schulze, S., Saladié, M., Rolland, V., Tanz, S.K., Westhoff, P., and Ludwig, M. (2017). A MEM1-like motif directs mesophyll cell-specific expression of the gene encoding the C4 carbonic anhydrase in Flaveria. J. Exp. Bot. *68*, 311–320.

von Gromoff, E.D., Schroda, M., Oster, U., and Beck, C.F. (2006). Identification of a plastid response element that acts as an enhancer within the Chlamydomonas HSP70A promoter. Nucleic Acids Res. *34*, 4767–4779.

Guan, Y., Meng, X., Khanna, R., Lamontagne, E., Liu, Y., and Zhang, S. (2014). Phosphorylation of a WRKY Transcription Factor by MAPKs Is Required for Pollen Development and Function in Arabidopsis. PLOS Genet. *10*, 1–12.

Gubler, F., Raventos, D., Keys, M., Watts, R., Mundy, J., and Jacobsen, J. V. (1999). Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. Plant J. *17*, 1–9.

Gundel, P.E., Pierik, R., Mommer, L., and Ballaré, C.L. (2014). Competing neighbors : light perception and root function. Oecologia *176*, 1–10.

Hartmann, U., Sagasser, M., Mehrtens, F., Stracke, R., and Weisshaar, B. (2005). Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol. Biol. *57*, 155–171.

van Heerden, P.S., Towers, G.H.N., and Lewis, N.G. (1996). Nitrogen Metabolism in Lignifying Pinus taeda Cell Cultures. J. Biol. Chem. *271*, 12350–12355.

Hellens*, R., Mullineaux*, P., and Klee, H. (2000). A guide to Agrobacterium binary Ti vectors. Trends Plant Sci. *5*, 1360-1385.

Hemm, M.R., Rider, S.D., Ogas, J., Murry, D.J., and Chapple, C. (2004). Light induces phenylpropanoid metabolism in Arabidopsis roots. Plant J. *38*, 765–778.

Herrmann, K.M., and Weaver, L.M. (1999). The shikimate pathway. Annu Rev Plant Physiol Plant Mol Biol *50*, 473–503.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res. *27*, 297–300.

Hill, K., Wang, H., and Perry, S.E. (2008). A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components. Plant J. *53*, 172–185.

Hnisz, D., Day, D.S., and Young, R.A. (2016). Insulated Neighborhoods :

Structural and Functional Units of Mammalian Gene Control. Cell *167*, 1188–1200.

Hoermiller, I.I., Naegele, T., Augustin, H., Stutz, S., Weckwerth, W., and Heyer, A.G. (2017). Subcellular reprogramming of metabolism during cold acclimation in Arabidopsis thaliana. Plant, Cell Environ. *40*, 602–610.

Huang, J., Gu, M., Lai, Z., Fan, B., Shi, K., Zhou, Y.-H., Yu, J.-Q., and Chen, Z. (2010). Functional analysis of the Arabidopsis PAL gene family in plant growth, development, and response to environmental stress. Plant Physiol. *153*, 1526–1538.

Invitrogen (2003). Gateway ® Technology: A universal technology to clone DNA sequences for functional analysis and expression in multiple systems. User Manual 1–60.

Isabel-LaMoneda, I., Diaz, I., Martinez, M., Mena, M., and Carbonero, P. (2003). SAD: A new DOF protein from barley that activates transcription of a cathepsin B-like thiol protease gene in the aleurone of germinating seeds. Plant J. *33*, 329–340.

Ishiguro, S., and Nakamura, K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5′ upstream regions of genes coding for sporamin and β-amylase from sweet potato. Mol. Gen. Genet. *244*, 563–571.

Iwase, A., Harashima, H., Ikeuchi, M., Rymen, B., Ohnuma, M., Komaki, S., Morohashi, K., Kurata, T., Nakata, M., Ohme-Takagi, M., Grotewold, E., and Sugimoto, K. (2017). WIND1 Promotes Shoot Regeneration through Transcriptional Activation of *ENHANCER OF SHOOT REGENERATION1* in Arabidopsis. Plant Cell *29*, 54–69.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., and Parcy, F. (2002). bZIP transcription factors in Arabidopsis. Trends Plant Sci. *7*, 106–111.

Jeffery, C.J. (2015). Why study moonlighting proteins? Front. in Genetics. *6*, 1-5.

Johnson, C.B., Shaik, R., Abdallah, R., Vitha, S., and Holzenburg, A. (2015). FtsZ1/FtsZ2 turnover in chloroplasts and the role of ARC3. Microsc. Microanal. *21*, 313–323.

J**ung, E., Zamir, L.O., and Jensen, R.** (1986). Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate. Proc. Natl. Acad. Sci. U.S.A. *83*, 7231–7235.

Kadonaga, J.T. (2012). Perspectives on the RNA polymerase II core promoter. WIREs Dev. Biol. *1*, 40–51.

Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. Nucleic Acids Res. *27*, 470–478.

Kamioka, M., Takao, S., Suzuki, T., Taki, K., Higashiyama, T., Kinoshita, T., and Nakamichi, N. (2016). Direct repression of evening genes by CIRCADIAN CLOCK-ASSOCIATED1 in the *Arabidopsis* circadian clock. Plant Cell *28*, 696–711.

Kaplan, B., Davydov, O., Knight, H., Galon, Y., Knight, M.R., Fluhr, R., and Fromm, H. (2006). Rapid transcriptome changes induced by cytosolic Ca2+ transients reveal ABRE-related sequences as Ca2+-responsive cis elements in

Arabidopsis. Plant Cell *18*, 2733–2748.
Kawoosa, T., and Gahlan, P. (2014). The GATA and SORLIP motifs in the 3-hydroxy- 3-methylglutaryl-CoA reductase promoter of Picrorhiza kurrooa for the control of light-mediated expression. Funct. Integr. Genomics *14*, 191–203.
Kim, J., and DellaPenna, D. (2006). Defining the primary route for lutein synthesis in plants: the role of *Arabidopsis* carotenoid beta-ring hydroxylase CYP97A3. Proc. Natl. Acad. Sci. *103*, 3474–3479.
Kim, D.W., Lee, S.H., Choi, S., Won, S., Heo, Y., Cho, M., Park, Y., and Cho, H. (2006). Functional conservation of a root hair cell-specific *cis*-element in angiosperms with different root hair distribution patterns. Plant Cell *18*, 2958–2970.
Kim, J.H., Hyun, W.Y., Nguyen, H.N., Jeong, C.Y., Xiong, L., Hong, S., and Lee, H. (2015). AtMyb7, a subgroup 4 R2R3 Myb, negatively regulates ABA-induced inhibition of seed germination by blocking the expression of the bZIP transcription factor ABI5. Plant, Cell Environ. *38*, 559–571.
Kim, S.H., Kim, H.S., Bahk, S., An, J., Yoo, Y., Kim, J., and Chung, W.S. (2017). Phosphorylation of the transcriptional repressor MYB15 by mitogen-activated protein kinase 6 is required for freezing tolerance in Arabidopsis. Nucleic Acids Res. *45*, 6613–6627.
Korku, P., Schippers, J.H.M., and Walther, D. (2014). Characterization and identification of *cis*-regulatory elements in *Arabidopsis* based on single-nucleotide polymorphism information. Plant Physiol. *164*, 181–200.
Kornberg, R.D. (2007). The molecular basis of eukaryotic transcription. PNAS *104*, 12955–12961.
Kristiansson, E., Thorsen, M., Tama, M.J., and Nerman, O. (2009). Evolutionary forces act on promoter length: identification of enriched *cis*-regulatory elements. Mol. Biol. Evol. *26*, 1299–1307.
Kuhn, B.M., Errafi, S., Bucher, R., Dobrev, P., Geisler, M., Bigler, L., Zazímalová, E., and Ringli, C. (2016). 7-Rhamnosylated Flavonols modulate homeostasis of the plant hormone auxin and affect plant development. J. Biol. Chem. *291*, 5385–5395.
Kumar, S., Asif, M.H., Chakrabarty, D., Tripathi, R.D., and Dubey, R.S. (2015). Comprehensive analysis of regulatory elements of the promoters of rice sulfate transporter gene family and functional characterization of OsSul1 ;1 promoter under different metal stress. Plant Signal. Behav. *10*, 1–6.
Lai, Z., Li, Y., Wang, F., Cheng, Y., Fan, B., Yu, J.-Q., and Chen, Z. (2011). *Arabidopsis* sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. Plant Cell *23*, 3824–3841.
Landry, L.C., Clint, C., Chapple, S., and Last, R. (1995). *Arabidopsis* mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. Plant Physiol. *109*, 1159-1166.
Larch, Y., Lapointe, J.W., and Kornberg, R.D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *49*, 203–210.
Li, Y., Liu, Z.-B., Shi, X., Hagen, G., and Guilfoyle, T.J. (1994). An auxin-inducible element in soybean SAUR promoters. Plant Physiol. *106*, 37–43.
Li, Y., Kim, J.I., Pysh, L., and Chapple, C. (2015). Four isoforms of *Arabidopsis*

thaliana 4-coumarate: CoA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism. Plant Physiol. *169*, 2409-2421.
Liu, C., and Murray, J.D. (2016). The role of flavonoids in nodulation host-range specificity: an update. Plants *5*, 1–13.
Liu, J., Osbourn, A., and Ma, P. (2015). MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. Mol. Plant *8*, 689–708.
Loreti, E., Yamaguchi, J., Alpi, A., and Perata, P. (2003). Sugar modulation of alpha-amylase genes under anoxia. Ann. Bot. *91*, 143–148.
Luo, H., Song, F., Goodman, R.M., and Zheng, Z. (2005). Up-regulation of OsBIHD1, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses. Plant Biol. *7*, 459–468.
Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. Science. *290*, 1151–1156.
Maeda, H., Shasany, A.K., Schnepp, J., Orlova, I., Taguchi, G., Cooper, B.R., Rhodes, D., Pichersky, E., and Dudareva, N. (2010). RNAi suppression of Arogenate Dehydratase1 reveals that phenylalanine is synthesized predominantly via the arogenate pathway in petunia petals. Plant Cell *22*, 832–849.
March-Diaz, R., Garcia-Dominguez, M., Florencio, F.J., and Reyes, J.C. (2007). SEF, a new protein required for flowering repression in *Arabidopsis*, interacts with PIE1 and ARP6 1. Plant Physiol. *143*, 893–901.
Marianayagam, N.J., Sunde, M., and Matthews, J.M. (2004). The power of two: protein dimerization in biology. Trends Biochem. Sci. *29*, 618–625.
Maruyama-Nakashita, A., Nakamura, Y., Watanabe-Takahashi, A., Inoue, E., Yamaya, T., and Takahashi, H. (2005). Identification of a novel *cis*-acting element conferring sulfur deficiency response in *Arabidopsis* roots. Plant J. *42*, 305–314.
Matsuo, M., Johnson, J.M., Hieno, A., Tokizawa, M., Nomoto, M., Tada, Y., Godfrey, R., Obokata, J., Sherameti, I., Yamamoto, Y.Y., Bohmer, F.D., and Oelmuller, R. (2015). High REDOX RESPONSIVE TRANSCRIPTION FACTOR1 Levels Result in Accumulation of Reactive Oxygen Species in Arabidopsis thaliana Shoots and Roots. Mol. Plant *8*, 1253–1273.
Matus, J.T. (2016). Transcriptomic and metabolomic networks in the grape berry illustrate that it takes more than flavonoids to fight against ultraviolet radiation. Front. Plant Sci. *7*, 1–8.
Mierziak, J., Kostyn, K., and Kulma, A. (2014). Flavonoids as important molecules of plant interactions with the environment. Molecules 16240–16265.
Mikami, K., Tabata, T., Kawata, T., Nakayama, T., and Iwabuchi, M. (1987). Nuclear protein(s) binding to the conserved DNA hexameric sequence postulated to regulate transcription of wheat histone genes. FEBS Lett. *223*, 273–278.
Mohanty, B., Krishnan, S.P.T., Swarup, S., and Bajic, V.B. (2005). Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. Ann. Bot. *96*, 669–681.
Molina, C., and Grotewold, E. (2005). Genome wide analysis of *Arabidopsis* core promoters. BMC Genomics *6*, 25.
Nagata, T., Niyada, E., Fujimoto, N., Nagasaki, Y., Noto, K., Miyanoiri, Y., Murata, J., Hiratsuka, K., and Katahira, M. (2010). Solution structures of the trihelix DNA-binding domains of the wild-type and a phosphomimetic mutant of

Arabidopsis GT-1: Mechanism for an increase in DNA-binding affinity through phosphorylation. Proteins Struct. Funct. Bioinf.. *78*, 3033–3047.

Nakashima, K., Ito, Y., and Yamaguchi-Shinozaki, K. (2009). Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. Plant Physiol. *149*, 88–95.

Ng, M., and Yanofsky, M.F. (2001). Function and evolution of the plant MADS-box gene family. Nat. Rev. Genet. *2*, 186–195.

Nickelsen, J. (2000). Mutations at three different nuclear loci of Chlamydomonas suppress a defect in chloroplast psbD mRNA accumulation. Curr. Genet. *37*, 136–142.

Nickelsen, J., Fleischmann, M., Boudreau, E., Rahire, M., and Rochaix, J. (1999). Identification of *cis*-acting RNA leader elements required for chloroplast psbD gene expression in Chlamydomonas. Plant Cell *11*, 957–970.

Nie, P., Li, X., Wang, S., Guo, J., Zhao, H., and Niu, D. (2017). Induced systemic resistance against *Botrytis cinerea* by *Bacillus cereus* AR156 through a JA/ET- and NPR1-dependent signaling pathway and activates PAMP-triggered immunity in *Arabidopsis*. Front. Plant Sci. *8*, 1–12.

Nishihara, M., and Nakatsuka, T. (2011). Genetic engineering of flavonoid pigments to modify flower color in floricultural plants. Biotech. Lett. *33*, 433–441.

Nishiuchi, T., Shinshi, H., and Suzuki, K. (2004). Rapid and transient activation of transcription of the ERF3 gene by wounding in tobacco leaves: Possible involvement of NtWRKYs and autorepression. J. Biol. Chem. *279*, 55355–55361.

Novina, C.D., and Roy, A.L. (1996). Core promoters and transcriptional control. Trends Genet. *12*, 351–355.

Ohgishi, M., Oka, A., Morelli, G., Ruberti, I., and Aoyama, T. (2001). Negative autoregulation of the *Arabidopsis* homeobox gene ATHB-2. Plant J. *25*, 389–398.

Olsen, K.M., Lea, U.S., Slimestad, R., Verheul, M., and Lillo, C. (2008). Differential expression of four *Arabidopsis* PAL genes; PAL1 and PAL2 have functional specialization in abiotic environmental-triggered flavonoid synthesis. J. Plant Physiol. *165*, 1491–1499.

Para, A., Muhammad, D., Orozco-Nunnelly, D.A., Memishi, R., Alvarez, S., Naldrett, M.J., and Warpeha, K. (2016). The dehydratase ADT3 affects ROS homeostasis and cotyledon development. Plant Physiol. *172*, 1045–1060.

Park, H.C., Kim, M.L., Kang, Y.H., Jeon, J.M., Yoo, J.H., Kim, M.C., Park, C.Y., Jeong, J.C., Moon, B.C., Lee, J.H., Yoon, H.W., Lee, S.H., Chung, W.S., Lim, C.O., Lee, S.Y., Hong, J.C. and Cho, M.J. (2004). Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. Plant Physiol. *135*, 2150–2161.

Pastore, J.J., Limpuangthip, A., Yamaguchi, N., Wu, M.-F., Sang, Y., Han, S.-K., Malaspina, L., Chavdaroff, N., Yamaguchi, A., and Wagner, D. (2011). LATE MERISTEM IDENTITY2 acts together with LEAFY to activate APETALA1. Development *138*, 3189–3198.

Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. Ann. Rev. Plant Biol. *62*, 549–566.

Piechulla, B., Merforth, N., and Rudolph, B. (1998). Identification of tomato Lhc promoter regions necessary for circadian expression. Plant Mol. Biol. *38*, 655–-

662.
Planchais, S., Perennes, C., Glab, N., Mironov, V., Inzé, D., and Bergounioux, C. (2002). Characterization of *cis*-acting element involved in cell cycle phase-independent activation of Arath;CycB1;1 transcription and identification of putative regulatory proteins. Plant Mol. Biol. *50*, 111–127.
Plesch, G., Ehrhardt, T., and Mueller-Roeber, B. (2001). Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression. Plant J. *28*, 455–464.
Quinn, J.M., Barraco, P., Eriksson, M., and Merchant, S. (2000). Coordinate Copper- and Oxygen-responsive Cyc6 and Cpx1 expression in Chlamydomonas is mediated by the same element. J. Biol. Chem. *275*, 6080–6089.
Ramirez-Parra, E., Fründt, C., and Gutierrez, C. (2003). A genome-wide identification of E2F-regulated genes in Arabidopsis. Plant J. *33*, 801–811.
Rausher, M.D. (2012). The evolution of genes in branched metabolic pathways. Evolution *67*, 34–48.
Renault, H., Alber, A., Horst, N.A., Lopes, A.B., Fich, E.A., Kriegshauser, L., Wiedemann, G., Ullmann, P., Herrgott, L., Erhardt, M., Pineau, E., Ehlting, J., Schmitt, M., Rose, J.K.C., Reski, R., and Werck-Reichhart, D. (2017). A phenol-enriched cuticle is anestral to lignin evolution in land plants. Nat. Commun. *8*, 1–8.
Renzette, N. (2011). Generation of transformation competent E. coli. Curr. Protoc. Microbiol. Appendix 1.
Reyes, J.C., Muro-Pastor, M.I., and Florencio, F.J. (2004). The GATA family of transcription factors in *Arabidopsis* and rice. Plant Physiol. *134*, 1718–1732.
Rippert, P., Puyaubert, J., Grisollet, D., Derrier, L., and Matringe, M. (2009). Tyrosine and Phenylalanine Are Synthesized within the Plastids in Arabidopsis. Plant Physiol. *149*, 1251–1260.
Roepke, J., and Bozzo, G.G. (2015). Arabidopsis thaliana beta-glucosidase BGLU15 attacks flavonol 3-O-beta-glucoside-7-O-alpha-rhamnosides. Phytochemistry *109*, 14–24.
Rogers, H.J., Bate, N., Combe, J., Sullivan, J., Sweetman, J., Swan, C., Lonsdale, D.M., and Twell, D. (2001). Functional analysis of *cis*-regulatory elements within the promoter of the tobacco late pollen gene g10. Plant Mol. Biol. *45*, 577–585.
Rose, A.B., Carter, A., Korf, I., and Kojima, N. (2016). Intron sequences that stimulate gene expression in Arabidopsis. Plant Mol. Biol. *92*, 337–346.
Roy, S. (2015). Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. Plant Signal. Behav. *2324*, 165-171.
Roy, A.L., and Singer, D.S. (2015). Core promoters in transcription: old problem, new insights. Trends Biochem. Sci. *40*, 165–171.
Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 129–143.
Sakai, H., Aoyama, T., and Oka, A. (2000). *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. Plant J. *24*, 703–711.
Sambrook J., and Russell, D. (2001) Molecular Cloning: a Laboratory Manual. ed 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sandal, N.N., Bojsen, K., and Marcker, K.A. (1987). A small family of nodule specific genes from soybean. Nucleic Acids Res. *15*, 1507–1519.

Satoh, R., Nakashima, K., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). ACTCAT, a novel *cis*-acting element for proline- and hypoosmolarity-responsive expression of the ProDH gene encoding proline dehydrogenase in Arabidopsis. Plant Physiol. *130*, 709–719.

Schenck, C.A., Chen, S., Siehl, D.L., and Maeda, H.A. (2014). Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. Nat. Chem. Biol. *11*, 52–57.

Schluttenhofer, C., and Yuan, L. (2015). Regulation of specialized metabolism by WRKY transcription factors. Plant Physiol. *167*, 295–306.

Schmid, N.B., Giehl, R.F.H., Doll, S., Mock, H.-P., Strehmel, N., Scheel, D., Kong, X., Hider, R.C., and von Wiren, N. (2014). Feruloyl-CoA 6'-Hydroxylase1-dependent coumarins mediate iron acquisition from alkaline substrates in *Arabidopsis*. Plant Physiol. *164*, 160–172.

Seo, P.J., and Park, C.M. (2010). A membrane-bound NAC transcription factor as an integrator of biotic and abiotic stress signals. Plant Signal. Behav. *5*, 481–483.

S**ilhavy, D., Molnár, A., Lucioli, A., Szittya, G., Hornyik, C., Tavazza, M., and Burgyán, J.** (2002). A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. EMBO J. *21*, 3070–3080.

Simpson, S.D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Two different novel *cis* -acting elements of *erd1* , a *clpA* homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. Plant J. *33*, 259–270.

Skrovankova, S., Sumczynski, D., Mlcek, J., Jurikova, T., and Sochor, J. (2015). Bioactive Compounds and Antioxidant Activity in Different Types of Berries. Int. J. Mol. Sci. *16*, 24673–24706.

Slavin, J. (2013). Fiber and prebiotics: mechanisms and health benefits. Nutrients *5*, 1417–1435.

Sohrabi, M., Zebarjadi, A., Najaphy, A., and Kahrizi, D. (2015). Isolation and sequence analysis of napin seed specific promoter from Iranian Rapeseed (*Brassica napus* L.). Gene *563*, 160–164.

Soitamo, A.J., Piippo, M., Allahverdiyeva, Y., Battchikova, N., and Aro, E. (2008). Light has a specific role in modulating *Arabidopsis* gene expression at low temperature. BMC Plant Biol. *8*, 1–20.

Song, J., Durrant, W.E., Wang, S., Yan, S., Tan, E.H., and Dong, X. (2011a). DNA repair proteins are directly involved in regulation of gene expression during plant immune response. Cell Host Microbe *9*, 115–124.

Srivastava, V.K., Raikwar, S., and Tuteja, N. (2014). Cloning and functional characterization of the promoter of PsSEOF1 gene from Pisum sativum under different stress conditions using Agrobacterium-mediated transient assay. Plant Signal. Behav. *9*, 1-10.

Su, J., Xu, J., and Zhang, S. (2015). RACK1, scaffolding a heterotrimeric G protein and a MAPK cascade. Trends Plant Sci. *20*, 405–407.

Su, Y., Wang, S., Zhang, F., Zheng, H., Liu, Y., Huang, T., and Ding, Y. (2017). Phosphorylation of Histone H2A at Serine 95: A plant-specific mark involved in

flowering time regulation and H2A.Z deposition. Plant Cell *29*, 1–37.
Sun, C., Palmqvist, S., Olsson, H., Borén, M., Ahlandsberg, S., and Jansson, C. (2003). A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. Plant Cell Online *15*, 2076–2092.
Sun, J., Jiang, H., Xu, Y., Li, H., Wu, X., Xie, Q., and Li, C. (2007). The CCCH-type zinc finger proteins AtSZF1 and AtSZF2 regulate salt stress responses in Arabidopsis. Plant Cell Physiol. *48*, 1148–1158.
Teakle, G.R., Manfield, I.W., Graham, J.F., and Gilmartin, P.M. (2002). Arabidopsis thaliana GATA factors: organisation, expression and DNA-binding characteristics. Plant Mol. Biol. *50*, 43–57.
Teng, Y.-S., Su, Y.-S., Chen, L.-J., Lee, Y., Hwang, I., and Li, H.-M. (2006). Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. Plant Cell Online *18*, 2247–2257.
Thakur, M., and Sohal, B.S. (2013). Role of elicitors in inducing resistance in plants against pathogen infection: A review. ISRN Biochem. *2013,* 1-10.
Thomas, M.S., and Flavell, R.B. (1990). Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell *2*, 1171–1180.
Thum, K.E., Kim, M., Morishige, D.T., Eibl, C., Koop, H.U., and Mullet, J.E. (2001). Analysis of barley chloroplast psbD light-responsive promoter elements in transplastomic tobacco. Plant Mol. Biol. *47*, 353–366.
Toledo-Ortiz, G., Huq, E., and Quail, P.H. (2003). The *Arabidopsis* Basic / Helix-Loop-Helix transcription factor family. Plant Cell *15*, 1749–1770.
Torres-Galea, P., Huang, L.F., Chua, N.H., and Bolle, C. (2006). The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses. Mol. Genet. Genomics *276*, 13–30.
de Torres Zabala, M., Zhai, B., Jayaraman, S., Eleftheriadou, G., Winsbury, R., Yang, R., Truman, W., Tang, S., Smirnoff, N., and Grant, M. (2016). Novel JAZ co-operativity and unexpected JA dynamics underpin *Arabidopsis* defence responses to Pseudomonas syringae infection. New Phytol. *209*, 1120–1134.
Toyofuku, K., Umemura, T., and Yamaguchi, J. (1998). Promoter elements required for sugar-repression of the RAmy3D gene for alpha-amylase in rice. FEBS Lett. *428*, 275–280.
Tronchet, M., BalaguÉ, C., Kroj, T., Jouanin, L., and Roby, D. (2010). Cinnamyl alcohol dehydrogenases-C and D, key enzymes in lignin biosynthesis, play an essential role in disease resistance in *Arabidopsis*. Mol. Plant Pathol. *11*, 83–92.
Tzin, V., and Galili, G. (2010). The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. Arab. B. *8*, 1–18.
Ulmasov, T. (1997). ARF1, a Transcription Factor That Binds to Auxin Response Elements. Science. *276*, 1865–1868.
Fernando V.C.D. and Schroeder, D.F. (2016). Arabidopsis DDB1-CUL4 E3 ligase complexes in det1 salt/osmotic stress resistant germination. Plant Signal. Behav. *11*, 225–231.
Vedel, V., and Scotti, I. (2011). Promoting the promoter. Plant Sci. *180*, 182–189.

van der Velden, A.W., and Thomas, A.A.M. (1999). The role of the 5' untranslated region of an mRNA in translation regulation during development. Int. J. Biochem. Cell Biol. *31*, 87–106.

Vieweg, M.F., Frühling, M., Quandt, H.-J., Heim, U., Bäumlein, H., Pühler, A., Küster, H., and Perlick, A.M. (2004). The Promoter of the *Vicia faba* L. Leghemoglobin gene *VfLb29* is specifically activated in the infected cells of root nodules and in the arbuscule-containing cells of mycorrhizal roots from different legume and nonlegume plants. Mol. Plant-Microbe Interact. *17*, 62–69.

Vitha, S., Mcandrew, R.S., and Osteryoung, K.W. (2001). FtsZ ring formation at the chloroplast division site in plants. J. Cell Biol. *153*, 111–119.

Vogt, T. (2010). Phenylpropanoid biosynthesis. Mol. Plant *3*, 2–20.

Wang, X.T., Yuan, C., Yuan, T.T., and Cui, S.J. (2012). The arabidopsis LFR gene is required for the formation of anther cell layers and normal expression of key regulatory genes. Mol. Plant *5*, 993–1000.

Wang, Z., Chen, M., Chen, T., Xuan, L., Li, Z., Du, X., Zhou, L., Zhang, G., and Jiang, L. (2014). TRANSPARENT TESTA2 regulates embryonic fatty acid biosynthesis by targeting FUSCA3 during the early developmental stage of Arabidopsis seeds. Plant J. *77*, 757–769.

Webb, C.F., Bryant, J., Popowski, M., Allred, L., Kim, D., Harriss, J., Schmidt, C., Miner, C.A., Rose, K., Cheng, H.-L., Griffin, C., and Tucker, P.W. (2011). The ARID Family Transcription Factor Bright Is Required for both Hematopoietic Stem Cell and B Lineage Development. Mol. Cell. Biol. *31*, 1041–1053.

Weise, S. (2013). *Agrobacterium* transformation and competent cell preparation. 28–30.

Welchen, E., and Gonzalez, D.H. (2005). Differential expression of the Arabidopsis cytochrome c genes Cytc-1 and Cytc-2. Evidence for the involvement of TCP-domain protein-binding elements in anther- and meristem-specific expression of the Cytc-1 gene. Plant Physiol. *139*, 88–100.

Wendel, J.F., Flagel, L.E., and Wendel, J.F. (2009). Gene duplication and evolutionary novelty in plants. New Phytol. *183*, 557–564.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., and Coupland, G. (2006). CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. Plant Cell Online *18*, 2971–2984.

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. *126*, 485–493.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V, and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One 1–12.

Xu, J., Meng, J., Meng, X., Zhao, Y., Liu, J., Sun, T., Liu, Y., Wang, Q., and Zhang, S. (2016). Pathogen-responsive MPK3 and MPK6 reprogram the biosynthesis of indole glucosinolates and their derivatives in *Arabidopsis* immunity. Plant Cell *28*, 1144–1162.

Yadav, V., Mallappa, C., Gangappa, S.N., Bhatia, S., and Chattopadhyay, S. (2005). A basic helix-loop-helix transcription factor in Arabidopsis, MYC2, acts as a repressor of blue light-mediated photomorphogenic growth. Plant Cell *17*, 1953–

1966.
Yamada, T., Matsuda, F., Kasai, K., Fukuoka, S., Kitamura, K., Tozawa, Y., Miyagawa, H., and Wakasa, K. (2008). Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of Phenylalanine and Tryptophan. Plant Cell *20*, 1316–1329.
Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K., and Obokata, J. (2009). Heterogeneity of *Arabidopsis* core promoters revealed by high-density TSS analysis. Plant J. *60*, 350–362.
Yamamoto, Y.Y., Yoshioka, Y., Hyakumachi, M., and Obokata, J. (2011). Characteristics of core promoter types with respect to gene structure and expression in *Arabidopsis thaliana*. DNA Res 333–342.
Yamanaka, T., Ohta, T., Takahashi, M., Meshi, T., Schmidt, R., Dean, C., Naito, S., and Ishikawa, M. (2000). TOM1, an *Arabidopsis* gene required for efficient multiplication of a tobamovirus, encodes a putative transmembrane protein. Proc. Natl. Acad. Sci. U.S.A. *97*, 10107–10112.
Yamauchi, D. (2001). A TGACGT motif in the 5'-upstream region of alpha-amylase gene from Vigna mungo is a cis-element for expression in cotyledons of germinated seeds. Plant Cell Physiol. *42*, 635–641.
Yanagisawa, S. (2004). Dof Domain Proteins: plant-specific transcription Factors associated with diverse processes. Plant Cell Physiol. *45*, 386–391.
Yang, T., and Poovaiah, B.W. (2002). A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. J. Biol. Chem. *277*, 45049–45058.
Yang, Y., Li, R., and Qi, M. (2000). *In vivo* analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. Plant J. *22*, 543–551.
Yang, Z., Patra, B., Li, R., Pattanaik, S., and Yuan, L. (2013). Promoter analysis reveals *cis*‐regulatory motifs associated with the expression of the WRKY transcription factor CrWRKY1 in *Catharanthus roseus*. Planta *238*, 1039–1049.
Yin, R., Messner, B., Faus-Kessler, T., Hoffmann, T., Schwab, W., Hajirezaei, M.R., Von Saint Paul, V., Heller, W., and Schäffner, A.R. (2012). Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-O-glycosylation. J. Exp. Bot. *63*, 2465–2478.
Yoshioka, S. (2004). The novel Myb transcription factor LCR1 regulates the CO2-responsive gene Cah1, encoding a periplasmic carbonic anhydrase in *Chlamydomonas reinhardtii*. Plant Cell Online *16*, 1466–1477.
Yu, D., Chen, C., and Chen, Z. (2001). Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. Plant Cell *13*, 1527–1539.
Yu, Q.B., Jiang, Y., Chong, K., and Yang, Z.N. (2009). AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript accd rna editing and early chloroplast biogenesis in arabidopsis thaliana. Plant J. *59*, 1011–1023.
Yukawa, Y., Sugita, M., Choisne, N., Small, I., and Sugiura, M. (2000). The TATA motif, the CAA motif and the poly(T) transcription termination motif are all important for transcription re-initiation on plant tRNA genes. Plant J. *22*, 439–447.

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H. (2006). *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. Nat. Protoc. *1*, 641–646.

Zhao, Q. (2016). Lignification: flexibility, biosynthesis and regulation. Trends Plant Sci. *21*, 713–721.

Zhao, H., Wu, D., Kong, F., Lin, K., Zhang, H., and Li, G. (2017a). The *Arabidopsis thaliana* nuclear factor Y transcription factors. Front. Plant Sci. *7*, 1–11.

Zhao, L., Feng, C., Wu, K., Chen, W., Chen, Y., Hao, X., and Wu, Y. (2017b). Advances and prospects in biogenic substances against plant virus: A review. Pestic. Biochem. Physiol. *135*, 15–26.

Zhiming, Y., Bo, K., Xiaowei, H., Shaolei, L., Youhuang, B., Wona, D., Ming, C., Hyung-taeg, C., and Ping, W. (2011). Root hair-specific expansins modulate root hair elongation in rice. Plant J. *66*, 725–734.

Zhou, M., Zhang, K., Sun, Z., Yan, M., Chen, C., Zhang, X., Tang, Y., and Wu, Y. (2017). LNK1 and LNK2 corepressors interact with the MYB3 transcription factor in phenylpropanoid biosynthesis. Plant Physiol. *174*, 1348–1358.

Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-wide prediction and validation of intergenic enhancers in *Arabidopsis* using open chromatin signatures. Plant Cell *27*, 2415-2426.

Zhu, Q., Ordiz, M.I., Dabi, T., Beachy, R.N., and Lamb, C. (2002). Rice TATA binding protein interacts functionally with transcription factor IIB and the RF2a bZIP transcriptional activator in an enhanced plant in vitro transcription system. Plant Cell *14*, 795–803.

Zhu, Y., Rong, L., Luo, Q., Wang, B., Zhou, N., Yang, Y., Zhang, C., Feng, H., Zheng, L., Shen, W.-H., Ma, J., and Dong, A. (2017). The histone chaperone NRP1 interacts with WEREWOLF to activate GLABRA2 in *Arabidopsis* root hair development. Plant Cell *29*, 260–276.

Zimmermann, I.M., Heim, M.A., Weisshaar, B., and Uhrig, J.F. (2004). Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R / B-like BHLH proteins. Plant J. *40*, 22–34.

6 APPENDICES

Appendix A. List of all *cis* Regulatory Motifs: Abiotic Stress[1,2]

MOTIF ID	CONSENSUS	*ADT* 1	2	3	4	5	6	TF/FUNCTION
ABRELATERD1	ACGTG		+					ABRE-like sequence; dehydration stress (Nakashima et al., 2009)
ABRERATCAL	CACGCGT	+		+				ABRE-related sequence; responsive to Ca2+ (Kaplan et al., 2006)
ACGTABOX	TACGTA	+		+				A-box recognized by bZIP transcription factors (Jakoby et al., 2002)
ACGTATERD1	ACGT	+	+	+	+	+		Etiolation response (Simpson et al., 2003)
ACGTTBOX	AACGTT					+		T-box; bZIP; stimulus-dependent activation (Alves et al., 2013)
ANAERO1CONSENSUS	AAACAAA	+		+	+		+	Hypoxia/anaerobically-induced expression (Mohanty et al. 2005)
ANAERO2CONSENSUS	AGCAGC					+		Hypoxia/anaerobically-induced expression (Mohanty et al., 2005)
CCAATBOX1	CCAAT		+	+	+	+	+	Found in promoters of heat shock proteins (Wenkel et al., 2006)
CURECORECR	GTAC		+		+	+		CuRE (copper response element) (Quinn et al., 2000)
DRE1COREZMRAB17	ACCGAGA					+		DRE1 core; ABA responsive (Busk et al., 1997)
EECCRCAH1	GAATTAC	+		+	+	+		EEC consensus; MYB (Yoshioka, 2004)
GAREAT	TAACAAA				+	+	+	Recognized by MYB97, MYB101 and MYB120 (Roy, 2015)
LTRE1HVBLT49	CCGAAA						+	Low temperature response element (Catalá et al., 2011)
LTRECOREATCOR15	CCGAC						+	Low-temperature response element (Catalá et al., 2011)

Appendix A. Continued: Abiotic Stress[1,2]

MOTIF ID	CONSENSUS	*ADT*						TF/FUNCTION
		1	2	3	4	5	6	
MYB1AT	AAACCA		+	+	+	+	+	MYB; ABA-inducible expression/drought stress (Abe et al. 2003)
MYB2AT	TAACTG		+					AtMYB2; dehydration stress response (Abe et al., 2003)
MYBCORE	CTGTTG	+	+	+			+	MYB; drought stress; flavonoid biosynthesis (Liu et al., 2015a)
MYBCOREATCYCB1	AACGG		+	+				MYB; cell cycle phase-independent activation of transcription (Planchais et al., 2002)
MYBGAHV	TAACAAA				+	+		MYB; GA response complex (GARC) (Gubler et al., 1999)
MYBST1	GGATA		+		+	+		MYBSt1 binding site (Baranowskij et al., 1994)
MYCATRD22	CACATG		+			+		Recognized by MYC rd22BP1; dehydration and ABA response (Simpson et al., 2003)
P1BS	GCATATTC				+			MYB; phosphate starvation response (Dubos et al., 2010)
PREATPRODH	ACTCAT					+		Hypoosmolarity responsive element (Satoh et al., 2002)
RYREPEATBNNAPA	CATGCA		+			+		RY repeat; ABA response; seed-specific expression (Ezcurra et al., 1999)
SREATMSD	TTATCC				+	+		SRE (sugar responsive element); response to main stem decapitation (Yadav et al., 2005)
SURECOREATSULTR11	GAGAC	+	+	+		+	+	Sulfur response element (SURE); sulfate deficient response in roots (Maruyama-Nakashita et al., 2005)

Appendix A. Continued: Biotic Stress[1,2]

MOTIF ID	CONSENSUS	*ADT* 1	2	3	4	5	6	TF/FUNCTION
ARR1AT	NGATT	+	+	+	+	+	+	ARR1; bacterial response (Sakai et al., 2000)
AMYBOX1	TAACAGA				+	+		Amylase box; sugar starvation response (Loreti et al., 2003)
AMYBOX2	TATCCAT					+		Amylase box, sugar starvation response (Loreti et al., 2003)
ELRECOREPCRP1	TTGACC	+						EIRE; elicitor response element ; WRKY (Schluttenhofer and Yuan, 2015)
NODCON1GM	AAAGAT		+	+		+		One of two nodulin consensus sequences (Sandal et al., 1987)
NODCON2GM	CTCTT	+	+	+	+	+		One of two nodulin consensus sequences (Sandal et al., 1987)
OSE1ROOTNODULE	AAAGAT		+	+		+		Organ specific element (OSE); root infection (Vieweg et al., 2004)
OSE2ROOTNODULE	CTCTT	+	+	+	+	+		Organ specific element (OSE); root infection (Vieweg et al., 2004)
SEBFCONSSTPR10A	TTGTCTC						+	SEBF; repressor of pathogen response (Boyle and Brisson, 2001)
SP8BFIBSP8BIB	TACTATT					+		SPF1 (WRKY); polygalacturonic acid-induced expression (Ishiguro and Nakamura, 1994)
T/GBOXATPIN2	AACGTG		+					Recognized by MYC (bHLH) for jasmonate and wound response (Appel et al., 2014)
WBOXATNPR1	TTGAC	+				+		W-box; WRKY; SA signaling (Yu et al., 2001)
WBOXHVISO1	TGACT	+				+	+	W-box; SUSIBA2 (WRKY) (Sun et al., 2003)
WBOXNTERF3	TGACY	+				+	+	W-box found in promoter of ERF3 (Nishiuchi et al., 2004)
WRKY71OS	TGAC	+	+	+		+	+	Core of TGAC-containing W-box, WRKY71

Appendix A. Continued: Development[1,2]

MOTIF ID	CONSENSUS	*ADT* 1	2	3	4	5	6	TF/FUNCTION
BIHD1OS	TGTCA	+	+	+	+			Recognized by BELL homeodomain transcription factor in rice (Luo et al., 2005)
CANBNNAPA	CTAACAC	+			+		+	Core of (CA)n element in storage protein gene promoters (Ellerström et al., 1996)
CARGCW8GAT	CTATWTAT WG	+	+	+	+		+	Variant of CArG box with longer A/T rich core; recognized by AGL15 transcription factor (Airoldi and Davies, 2012)
CARGNCAT	CCWWWW WWWWGG			+				Response to AGL15; involved in gibberellin metabolism (Hill et al., 2008)
E2FANTRNR	TTTCCCGC						+	E2Fb recognition site; G1 to S phase transition in cell cycle progression (Ramirez-Parra et al., 2003)
E2FCONSENSUS	ATTCCCCC						+	E2F recognition site; cell cycle regulation (Ramirez-Parra et al., 2003)
HDZIP2ATATHB2	TAATAATTA				+			Recognized by homeobox gene ATHB-2; auxin response and cell proliferation (Ohgishi et al. 2001)
POLLEN1LELAT52	AGAAA	+	+	+	+	+	+	1 of 2 co-dependent regulatory elements responsible for pollen-specific activation (Guan et al., 2014)
PYRIMIDINEBOXHV EPB1	TTTTTTCC		+		+	+		Pyrimidine box; coordinates gene expression in response to gibberellins and ABA (Cercós et al., 1999)
PYRIMIDINEBOXOSR AMY1A	CCTTTT	+	+					Pyrimidine box; upstream of sugar-repressed alpha amylase gene; gibberellin reponse (Cercós et al., 1999)
ROOTMOTIFTAPOX1	ATATT	+	+		+	+	+	Root hair growth and elongation (Pastore et al., 2011)
RHERPATEXPA7	TCACGT				+			Root-hair-specific cis element for root distribution (Zhiming et al., 2011)
RYREPEATGMGY2	CATGCAT		+			+		RY repeat; regulates expression of glycinin genes (Fauteux and Strömvik, 2009)

Appendix A. Continued: Development[1,2]

MOTIF ID	CONSENSUS	*ADT* 1	2	3	4	5	6	TF/FUNCTION
RYREPEATLEGUMINBOX	CATGCAT		+			+		RY repeat or legumin box found in seed-storage protein genes in legumes (Ezcurra et al., 1999)
RYREPEATVFLEB4	CATGCATG					+		RY-repeat motif; recognized by FUS3 transcription factor; embryogenesis (Ezcurra et al., 1999)
SEF1MOTIF	ATATTTATA				+		+	Soybean embryo factor; embryo development (Fauteux and Strömvik, 2009)
SEF3MOTIFGM	AACCCA	+						Soybean embryo factor; embryo development (Fauteux and Strömvik, 2009)
SEF4MOTIFGM7S	RTTTTTG	+		+	+		+	Soybean embryo factor; embryo development (Fauteux and Strömvik, 2009)
SITEIIATCYTC	TGGGCC		+	+			+	Site II element; upstream of genes for oxidative phosphorylation (Welchen and Gonzalez, 2005)
TATCCAOSAMY	TATCCA	+	+	+		+		Recognized by MYB family; hormone regulation of alpha-amylase gene expression (Yamauchi, 2001)
TGACGTVMAMY	TGACGT	+						Required for high expression of alpha-amylase; seed development (Yamauchi, 2001)

Stress and Development

MOTIF ID	CONSENSUS	1	2	3	4	5	6	TF/FUNCTION
ASF1MOTIFCAMV	TGACG	+				+		Recognized by ASF-1; auxin/SA responsive transcription (Banerjee et al., 2013)
CGCGBOXAT	ACGCGT	+		+			+	CGCG box; recognized by SR1-6 signal response element; calmodulin binding site; signal transduction (Yang and Poovaiah, 2002)

Appendix A. Continued: Stress and Development[1,2]

MOTIF ID	CONSENSUS	*ADT* 1	2	3	4	5	6	TF/FUNCTION
DOFCOREZM	AAAG	+	+	+	+	+	+	Required for Dof binding (Yanagisawa, 2002)
DPBFCOREDCDC3	ACACTGG		+		+	+		Recognized by bZIP transcription factors involved in ABA-response and embryo-specification (Kamioka et al., 2016)
EBOXBNNAPA	CATATG	+	+	+	+	+	+	bHLH recognition site (Yadav et al., 2005)
GCN4OSGLUB1	TGAGTCA	+						Recognized by bZIP transcription factors; endosperm development (Jakoby et al., 2002)
GLMHVCHORD	ATGAGTCAT	+						Recognized by bZIP transcription factors; nitrogen response and endosperm expression (Jakoby et al., 2002)
GT1GMSCAM4	GAAAAA	+	+	+	+	+	+	GT-1 motif found in promoter of soy CaM isoform (Park et al. 2004)
MYCCONSENSUSAT	CATATG	+	+	+	+	+	+	bHLH recognition site (Abe et al. 2003)
NTBBF1ARROLB	ACTTTA			+		+		Tobacco Dof protein binding site found in Agrobacterium rhizogenes (Baumann, 1999)
RAV1AAT	CAACA	+	+	+	+	+	+	Recognized by RAV1-A transcription factor; development and drought/salinity stress (Kagaya et al. 1999)
SBOXATRBCS	CACCTCCA		+					S-box; recognized by ABI4 transcription factor; connects light- and sugar-responsive signalling pathways (Baxter et al., 2012)
TAAAGSTKST1	TAAAG		+	+	+		+	Found in promoter of KST1 gene; recognized by Dof transcription factors (Plesch et al., 2001)
TATCCACHVAL21	TATCCAC	+						GARC response complex; one of three sequence motifs responsible for GA response (Isabel- LaMoneda et al. 2003)

Appendix A. Continued: Light Response[1,2]

MOTIF ID	CONSENSUS	*ADT*						TF/FUNCTION
		1	2	3	4	5	6	
-10PEHVPSBD	TATTCT	+		+	+			- 10 promoter element found in promoters controlling genes in response to light (Thum et al., 2001)
ARFAT	TGTCTC						+	Recognized by ARF1 (auxin response factor 1) (Ulmasov et al. 1999)
CATATGGMSAUR	CATATG	+			+	+		Sequence found in NDE element in soy SAUR 15A promoter; auxin response (Li et al., 1994)
CCA1ATLHCB1	AAAAATCT	+			+		+	CCA1 recognition site (MYB-related); regulated by phytochrome (Yoshioka, 2004)
CIACADIANLELHC	CAACTTTATC				+		+	Region necessary for circadian expression (Piechulla et al., 1998)
GATABOX	GATA	+	+	+	+	+	+	GATA box important for light-regulated expression (Reyes et al., 2004)
GT1CONSENSUS	GRAAAW	+	+	+	+	+	+	GT-1 binding site found in light regulated genes (Nagata et al., 2010)
IBOX	GATAAG						+	Upstream of light-regulated genes; recognized by MYB transcription factors (Hartmann et al., 2005)
IBOXCORE	GATAA		+	+	+	+	+	Upstream of light-regulated genes; recognized by MYB transcription factors (Hartmann et al., 2005)
MARTBOX	TTTTTTTTTT					+		T-box; upstream of light-regulated genes (Yukawa et al., 2000)
PRECONSCRHSP70A	CCGATTATGAC ACTCCACCAAG AG						+	Plastid response element; acts as a light-responsive enhancer (von Gromoff et al., 2006)

Appendix A. Continued: Light Response[1,2]

MOTIF ID	CONSENSUS	*ADT* 1	2	3	4	5	6	TF/FUNCTION
REBETALGLHCB21	CGGATA				+			GATA-like sequence; phytochrome signaling (Nagata et al., 2010)
S1FBOXSORPS1L21	ATGGTA			+	+			S1F box; involved in regulation of plastid-related genes (Simpson et al., 2003)
SORLIP1AT	GCCAC					+		"Sequences over-represented in light-induced promoters"; PhyA-regulated (Kawoosa and Gahlan, 2014)
SORLIP2AT	GGGCC		+				+	"Sequences over-represented in light-induced promoters"; PhyA-regulated (Kawoosa and Gahlan, 2014)
SORLREP3AT	TGTATATAT				+			"Sequences over-represented in light-induced promoters"; PhyA-regulated (Kawoosa and Gahlan, 2014)
TBOXATGAPB	ACTTTG						+	T-box; light-activated transcription (Chan et al., 2001)

Phenylpropanoid-Related

MOTIF ID	CONSENSUS	1	2	3	4	5	6	TF/FUNCTION
BOXLCOREDCPAL	ACCWWCC	+			+			Core of BoxL in PAL1 promoter; recognized by MYB1 in response to environmental cues (Maeda et al. 2005)
MYBPLANT	CACCAACC	+						MYB recognition site; in promoters of phenylpropanoid-related genes (Liu et al., 2015a)
MYBPZM	CCAACC	+						MYB homolog recognition site; flavonoid biosynthesis (Liu et al., 2015a)
PALBOXLPC	TCTCACCA ACC	+						Box L; 1/3 cis elements; necessary but not sufficient for elicitor- or light-mediated activation (Olsen et al., 2008)

Appendix A. Continued: Phenylpropanoid-Related[1,2]

MOTIF ID	CONSENSUS	*ADT* 1	2	3	4	5	6	TF/FUNCTION
PALBOXPPC	TTCTCACCAA CCCC	+						Box L; one of 3 cis elements; necessary but not sufficient for elicitor- or light-mediated activation (Olsen et al., 2008)
General Response								
-300ELEMENT	TGMAAARK	+	+	+		+		Endosperm-specific expression- ZEIN gene promoter (Thomas and Flavell, 1990)
-300MOTIFZMZEIN	ATGAGTCAT	+						Enhancer element for endosperm expression (Thomas and Flavell, 1990)
AACACOREOSGLUB1	AACAAAC	+			+		+	Motifs in rice glutenin gene promoter; minimal elements required for endosperm expression (Thomas and Flavell, 1990)
BOXIINTPATPB	ATAGAA		+		+	+		Box II; found in tobacco plastid atpB promoter (Reyes et al., 2004)
CAATBOX1	CAAT	+	+	+	+	+	+	CAAT promoter sequence for tissue-specific expression (Fauteux and Strömvik, 2009)
CACTFTPPCA1	TACT	+	+	+	+	+	+	MEM1 (mesophyll expression module 1) motif for mesophyll-specific expression (Gowik et al., 2017)
GTGANTG10	GTGA	+	+	+	+	+	+	Pollen specific element for pollen-specific expression (Rogers et al. 2001)
HEXMOTIFTAH3H4	ACGTCA				+	+		Hexamer motif found in wheat promoter of histone H3 and H4 genes (Mikami et al., 1987)
INRNTPSADB	TTCARTYC	+		+	+	+	+	Initiator element found in tobacco psaDb gene promoter without a TATA box (Novina and Roy, 1996)

Appendix A. Continued: General Response[1,2]

MOTIF ID	CONSENSUS	*ADT*						TF/FUNCTION
		1	2	3	4	5	6	
MARABOX1	AATAAATAAA			+		+		A-box found in SAR (scaffold arrangement region) (Gasser et al. 1989)
MRNASTA2CRPSBD	TGAGTTG			+				mRNA stability determinant (Nickelsen, 2000)
NAPINMOTIFBN	TACACAT					+		Sequene found in 5' upstream region of napin gene; seed-specific expression (Ericson et al., 1991)
POLASIG1	AATAAA	+	+	+	+	+		Cis-acting Poly-A signal mediating polyadenylation (Ashraf et al., 2014)
POLASIG3	AATAAT	+	+		+	+		PolyA signal in rice alpha amylase (Ashraf et al., 2014)
PROLAMINBOXOSGLUB1	TGCAAAG	+				+		Prolamin box found in rice GluB-1 gene promoter (Wu et al. 2000)
TATABOX2	TATAAAT				+			TATA- like sequence for transcription initiation (Yukawa et al. 2000)
TATABOX4	TATATAA	+		+	+	+		TATA- like sequence for transcription initiation (Yukawa et al. 2000)
TATABOX5	TTATTT	+		+		+		TATA- like sequence for transcription intiation (Yukawa et al. 2000)
TATABOXOSPAL	TATTTAA		+		+		+	TATA binding protein binding site (Zhu et al., 2002)
TATAPVTRNALEU	TTTATATA				+			TATA- like sequence for transcription initiation (Yukawa et al. 2000)
TATCCAYMOTIFOSRAMY3D	TATCCAC	+				+		Required for sugar-repression of RAmy3D in rice (Toyofuku et al., 1998)
TRANSINITDICOTS	AATATGGC	+	+	+				Translation initiation context sequence from dicots (Novina and Roy, 1996)

[1]A + sign indicates the presence of that motif in that promoter.
[2]Yellow=significantly enriched, green= common to at least 5 out of 6 promoters, pink= unique to one promoter

Appendix B. Co-expressed Gene Totals for Each Category[1,2]

ADT	AB	BS	D	SD	L	PR	GR
1	--	--	2	1	1	2	3
2	--	--	4	2	--	--	2
3	--	--	3	5	1	4	1
4	--	1	--	8	--	5	1
5	1	--	--	7	--	5	2
6	--	--	1	3	1	3	--

[1]AB= abiotic stress, BS= biotic stress, D= development, SD= stress and development, L= light response, PR= phenylpropanoid-related, GR= general response.
[2]Data are the total number of genes from the co-expression analysis belonging to each category described in Section 3.1.2.

7 CIRRICULUM VITAE

Emily Cornelius, BSc

EDUCATION

MSc Candidate: Cell and Molecular Biology
The University of Western Ontario, London, Ontario 2015-2017
BSc: Honors Specialization in Biology, Minor in Genetics
The University of Western Ontario, London, Ontario 2011-2015

WORK EXPERIENCE

MSc Candidate and Researcher
Western Science Centre, Molecular Genetics Unit, Western University
Lead Teaching Assistant
Biology 2601: Organismal Physiology, Fall Term 2016
Teaching Assistant
Biology 2601A: Organismal Physiology, Fall Term 2015
Biology 2581B: Genetics, Winter Terms 2016, 2017

VOLUNTEER EXPERIENCE

Peer Reviewer
Western Undergraduate Research Journal, 2017
Judge
Thames Valley Science and Engineering Fair, 2017
Instructor
Let's Talk Science- Outreach at Western University, 2016-2017
Coach
Port Weller Soccer League, St. Catharines, ON, 2013-2015
Peer Mentor
Stand By Me Program, Thames Valley District School Board, 2013-2014

AWARDS

-Graduate Student Teaching Award Nominee (Western University), 2016, 2017
-George H. Duff Travel Award (Canadian Society of Plant Biologists), 2016
-Western Graduate Research Scholarship (Western University), 2015-2017

RESEARCH PRESENTATIONS

Radio Interview
CHRW Radio 94.7 FM, Gradcast #97: Getting Promoted with Emily Cornelius, February 8, 2017, London, ON. www.gradcastradio.ca

Oral Presentation
Cornelius, E.J. and Kohalmi, S.E. Promoter sequence diversity and its roles in differential expression of six *Arabidopsis AROGENATE DEHYDRATASE* genes. Canadian Society of Plant Biologists Eastern Regional Meeting, November 19, 2016, Burlington, ON

Oral Presentation
Cornelius, E.J. and Kohalmi, S.E. Sequence-function relationships of *AROGENATE DEHYDRATASE* promoters from *Arabidopsis thaliana.* Biology Graduate Research Forum, October 13-14, 2016, London, ON

Poster Presentation
Cornelius, E.J. and Kohalmi, S.E. Promoter analysis of *AROGENATE DEHYDRATASE* genes from *Arabidopsis thaliana*. Canadian Society of Plant Biologists Annual General Meeting, June 19-21, 2016, Kingston, ON

PUBLICATIONS

Abolhassani Rad, S., Clayton, E.J., Cornelius, E.J., Howes, T.R. and Kohalmi, S.E. (2017) Moonlighting proteins: putting the spotlight on enzymes. Plant Signaling and Behaviour. Manuscript Submitted.