
Electronic Thesis and Dissertation Repository

8-17-2021 9:00 AM

Functional Characterization of Arogenate Dehydratase Isoforms in Soybean

Ramtin Sirjani, *The University of Western Ontario*

Supervisor: Dhaubhadel, Sangeeta, *Agriculture and Agri-Food Canada*

Co-Supervisor: Bernards, Mark, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

© Ramtin Sirjani 2021

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



Part of the [Biology Commons](#), [Biotechnology Commons](#), [Genetics Commons](#), and the [Molecular Genetics Commons](#)

Recommended Citation

Sirjani, Ramtin, "Functional Characterization of Arogenate Dehydratase Isoforms in Soybean" (2021). *Electronic Thesis and Dissertation Repository*. 8017.
<https://ir.lib.uwo.ca/etd/8017>

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 wlsadmin@uwo.ca.

Abstract

Phenylalanine flux is partitioned between phenylpropanoid and protein synthesis. The mechanisms behind the metabolic channeling of phenylalanine are largely unknown. Arogenate dehydratase (ADT) enzymes, which catalyze the last and rate-limiting step in the synthesis of phenylalanine in plants, have been shown to interact with the isoflavonoid metabolon in the cytosol. Cytosolic phenylalanine, however, can only be synthesized through prephenate dehydratase (PDT) activity. In this study, putative soybean ADTs (GmADTs) were characterized for their ADT and PDT activity. This was done using complementation assays with two different knockout yeast strains, *aro8aro9* and *pha2*, which lack prephenate aminotransferase and PDT activity, respectively. Additionally, *GmADTs* with alternate transcripts that exclude the transit peptide were identified through qRT-PCR. It was determined that, of 8 putative GmADTs, GmADT11B had the most ADT and PDT activity. GmADT12B and GmADT12C were found to have some ADT activity but to a lesser degree. The remaining 5 GmADTs had the least ADT activity, if any. Some PDT activity was detected in GmADT12A and GmADT13A, while none was detected in the remaining 5 GmADTs. Furthermore, it was determined that *GmADT12B* and *GmADT11A* contain alternate transcripts that exclude the sequence for the transit peptide. If these GmADTs have a cytosolic isoform, they are likely involved in directing phenylalanine flux to phenylpropanoid synthesis. These findings provide insight into possible mechanisms of regulation controlling specialized metabolite synthesis in plants.

Keywords:

arogenate dehydratase, prephenate dehydratase, isoflavonoids, phenylpropanoid, cytosolic, regulation

Summary for Lay Audience

Plants make molecules called specialized metabolites that they use in their own protection from external stresses, like extreme weather conditions, diseases, and pests. Soybean is a legume, a family of plants that make unique specialized metabolites called isoflavonoids. In addition to protecting plants from external threats, isoflavonoids play a role in allowing soybean to form symbiotic relationships with species of bacteria called rhizobia to obtain nitrogen from the air. Nitrogen fertilizers are unsustainable and cause environmental destruction. Climate change creates harsher weather conditions and may allow for more/new diseases to attack crops where they were previously not a problem. For this reason, understanding the intricate machinery behind how, when, and why isoflavonoids are made will likely prove to be useful for creating crops that require less nitrogen fertilizer and have increased disease protection. To this extent, I explored when, and how isoflavonoids are made by studying an enzyme called aroclate dehydratase (ADT). ADT makes phenylalanine, the starting molecule used to make isoflavonoids. Phenylalanine is also a building block for proteins. How plants divide available phenylalanine between these two different outputs is unknown, but critical to our understanding of plant function. Furthermore, ADTs are found mostly in the chloroplast, but some may also be found in the cytosol. It has been shown that a cytosolic bi-functional ADT directly interacts with the machinery that makes isoflavonoids. Thus, it is likely that cytosolic bi-functional ADTs direct phenylalanine to isoflavonoid production, while the others make phenylalanine for proteins. In my project I confirmed which *GmADT* genes make functional ADT proteins. I then identified which ADT genes could possibly make cytosolic versions of the protein. I concluded that one ADT, GmADT12A, likely directs phenylalanine to isoflavonoid production in soybean. This knowledge serves as an important step towards building a sustainable future.

Acknowledgements

I would like to thank my supervisor, Dr. Sangeeta Dhaubhadel, for her support in my completion of this thesis. She provided a lab environment where we respected the value of all our kits and reagents, while still having availability to the resources I needed to complete my experiments. Additionally, Dr. Dhaubhadel made excellent suggestions regarding my experimental design. Furthermore, her expertise and guidance in scientific communication nurtured a great improvement in my presentation and writing skills. If I felt anxious or uneasy, Dr. Dhaubhadel was patient and did not give up. I still, however, have a long way to go, and on that journey I will be sure not forget all the positive habits she tried to instill in me.

I would also like to thank our laboratory technicians Ling Chen and Kufrom Kufu. When I first began my work in Dr. Dhaubhadel's lab, Ling walked me through all the standard protocols. She was always willing to take time out of her day to answer any questions I may have had. Near the end of my research, space was limited in our lab due to COVID-19 restrictions, so lab members were split between either the morning or afternoon shift. I had the pleasure of working the same hours as Kufrom. He and I would brainstorm explanations for any result that may not have made sense to me initially. Without his thorough technical understanding, acquiring informative results would have been more difficult. I am also grateful to Angelo Kaldis, who was willing to help me think of ways to troubleshoot protocols that were not behaving as expected.

My labmates, Nishat, Arun, Kelsey, Praveen, Petar, Jasmine, Lexie, and Jordan. Thank you for your support to improve my science, and your support as friends. I enjoyed all the meals we ate together, all the conversations we had, and all the struggles we shared. I hope you all find success.

I would like to extend gratitude to Dr. Chris Garnham for serving on my advisory committee. Thank you for your guidance, and that you for always taking the time for me during our committee meetings.

Thank you to my co-supervisor Dr. Mark Bernards for hearing me out whenever I was struggling. I felt that you always guided me to trust the scientific method. I am also grateful for your edits of my writing, as I believe they help me to be clearer and more persuasive.

Finally, I would like to extend a special thank you to Dr. Susanne Kohalmi and her PhD student Emily Clayton. They were kind enough to allow me to work in their lab on Western campus when COVID-19 restrictions prevented me from accessing the facilities at AAFC. Furthermore, they provided me with both mutant yeast strains for the ADT and PDT assays. Susanne and Emily are both yeast and ADT experts and their knowledge was indispensable for this project.

List of Abbreviations

ADT	arogenate dehydratase
AtADT	<i>Arabidopsis thaliana</i> arogenate dehydratase
CM	chorismate mutase
ER	endoplasmic reticulum
FPKM	fragments per kilobase per million mapped reads
Gal	galactose
Glu	glucose
GmADT	<i>Glycine max</i> arogenate dehydratase
His	histidine
IFS	isoflavone synthase
Leu	leucine
LB	lysogeny broth
PAGE	polyacrylamide gel electrophoresis
PAL	phenylalanine ammonia lyase
PCR	polymerase chain reaction
PDT	prephenate dehydratase
PhADT	<i>Petunia hybrida</i> arogenate dehydratase
Phe	phenylalanine
PPA-AT	prephenate aminotransferase

PPY-AT	phenylpyruvate aminotransferase
PVDF	polyvinylidene difluoride
qRT-PCR	quantitative real-time polymerase chain reaction
RACE	rapid amplification of cDNA ends
Raf	raffinose
SD	synthetic defined
SOC	super optimal broth with catabolite repression
TAE	tris-acetate-EDTA
TBE	tris-borate EDTA
TBS	tris buffer saline solution
TBST	tris buffer saline solution with tween-20

Table of Contents

Abstract	i
Summary for Lay Audience	ii
Acknowledgements	iii
List of Abbreviations	iv
List of Tables	ix
List of Figures	x
List of Appendices	xi
Chapter 1: Introduction	1
1.1 Soybean as a Staple Crop.....	1
1.2 Specialized Metabolites and Their Precursors	1
1.3 Phenylalanine, A Key Point of Divergence	4
1.4 ADTs as Part of the Isoflavonoid Metabolon.....	7
1.5 Localization of the Plant PDT Pathway	9
1.6 The Soybean Genome, and Candidate <i>GmADTs</i>	10
1.7 ADT Architecture and Conserved Motifs	11
1.8 Hypothesis.....	14
1.9 Objectives.....	14
Chapter 2: Materials and Methods	15
2.1 Plant Growth Conditions.....	15

2.2	Bacterial and Yeast Strains.....	15
2.3	Bacterial Media	15
2.4	Agarose Gel Electrophoresis	16
2.5	RNA Extraction, Reverse Transcription, and Quantitative Real-Time PCR (qRT-PCR)	16
2.6	Cloning and Transformation	22
2.7	ADT assay	25
2.8	PDT assay.....	25
2.9	Protein Extraction.....	26
2.10	SDS-PAGE and Western Blotting.....	26
Chapter 3: Results.....		28
3.1	Cloning and Transformation of <i>GmADTs</i>	28
3.2	ADT Assay.....	32
3.3	PDT Assay.....	34
3.4	Western Blot Analysis.....	36
3.5	Identification of <i>GmADT</i> Transcript Variants.....	39
Chapter 4: Discussion		42
4.1	GmADTs Behave like Characterized ADTs in Other Species.....	43
4.2	Amino Acid Substitutions in GmADT Clones May Effect Function	47
4.3	Yeast Complementation Assay – Characterizing ADT Activity of Candidate GmADTs	47

4.4	Yeast Complementation Assay – Characterizing PDT Activity in Candidate GmADTs	51
4.5	Yeast Complementation Assay – Western Blot	52
4.6	Identification of GmADT Alternate Transcripts	54
4.7	Future Direction: The Next Steps.....	55
Chapter 5: Conclusion		56
References		58
Appendices.....		63
Curriculum Vitae		66

List of Tables

Table 1.1. Characteristics of candidate <i>GmADTs</i>	8
Table 2.1. Comprehensive primer list	18
Table 3.1. List of candidate <i>GmADTs</i> in this study.....	30

List of Figures

Figure 1.1. The flavonoid branch of the phenylpropanoid biosynthetic pathway	3
Figure 1.2. Biosynthesis of phenylalanine from chorismate.....	6
Figure 1.3. Phylogenetic relationship between GmADTs and plant and bacterial ADT/PDT sequences.....	13
Figure 2.1. Cloning strategy of <i>GmADT</i> genes	23
Figure 3.1. Yeast colony PCR	31
Figure 3.2. Yeast complementation assay testing ADT activity	33
Figure 3.3. Yeast complementation assay testing PDT activity	35
Figure 3.4. Western blot to detect expression of GmADTs in <i>S. cerevisiae aro8aro9</i> PAT ..	37
Figure 3.5. Western blot to detect expression of GmADTs in <i>S. cerevisiae pha2</i>.....	38
Figure 3.6. RT-qPCR to identify <i>GmADT</i> alternate transcripts	41

List of Appendices

Appendix A. High identity pairs of GmADTs	65
Appendix B. Protein alignment of all candidate GmADTs	66
Appendix C. Raw photos of <i>S. cerevisiae</i> aro8aro9 yeast growth in the ADT assay.....	67

Chapter 1: Introduction

1.1 Soybean as a Staple Crop

Agriculture is one of the key drivers of the Canadian economy. Soybean is considered a staple crop, grown mainly in Ontario, Quebec, and Manitoba. Soybean alone accounted for 2.58 billion dollars of Canada's 2020 agricultural exports (soycanada.ca). Soybean is processed into two main products: meal (80%), and oil (20%). Soybean meal is a high protein food source, but mostly is used as a supplement in animal feed. Soybean oil is generally used in food processing and cooking but can also be used as biodiesel or have other industrial application (ncsoy.org). Soybean is therefore a valuable commodity to Canada, and it is in our best interest to maintain its yield.

Due to its unique traits as a legume, cultivating soybean requires a lower cost in fertilizer than other non-legume crop plants (gov.mb.ca). This is because legumes fix nitrogen through their symbiotic interactions with rhizobia. Legumes like soybean also synthesize unique defensive phytochemicals that protect the plant from disease (Boydston *et al.*, 1983).

Challenges such as pests, and climate change may necessitate the engineering of improved crops (Kopittke *et al.*, 2019). To do so, a comprehensive understanding of the molecular context from which defensive traits are derived is required.

1.2 Specialized Metabolites and Their Precursors

Plant specialized metabolites are biosynthetic products with roles separate from growth and reproduction. These are promising research targets when trying to understand how to improve crops, as they are implicated in biotic defense (Ahuja *et al.*, 2012), abiotic regulation (Cheynier *et al.*, 2013; Dudareva *et al.*, 2013), and symbiotic communication (Subramanian *et al.*, 2006). Isoflavonoids are a class of specialized metabolites specific to legumes (Vogt, 2010; Křížová *et al.*, 2019). Isoflavonoids are phenolic compounds, similar in structure to estrogen, and as such are

colloquially referred to as phytoestrogens. When ingested by vertebrates isoflavonoids have been shown to mimic estrogen (Dixon, 2004), but in plants they function mainly as phytoalexins (Dakora and Phillips, 1996). Phytoalexins are synthesized as part of the basal immune response of a plant, accumulating during times of stress. Additionally, RNA interference silencing of isoflavone synthase (IFS) in soybean resulted in severe reduction of root nodule formation, indicating that isoflavonoids are also important for symbiosis with rhizobia (Subramanian *et al.*, 2006). Isoflavonoids are biosynthetic products of the phenylpropanoid pathway (Figure 1.1), that utilizes the amino acid phenylalanine (Phe) as a precursor (Maeda and Dudareva, 2012; Dudareva *et al.*, 2013).

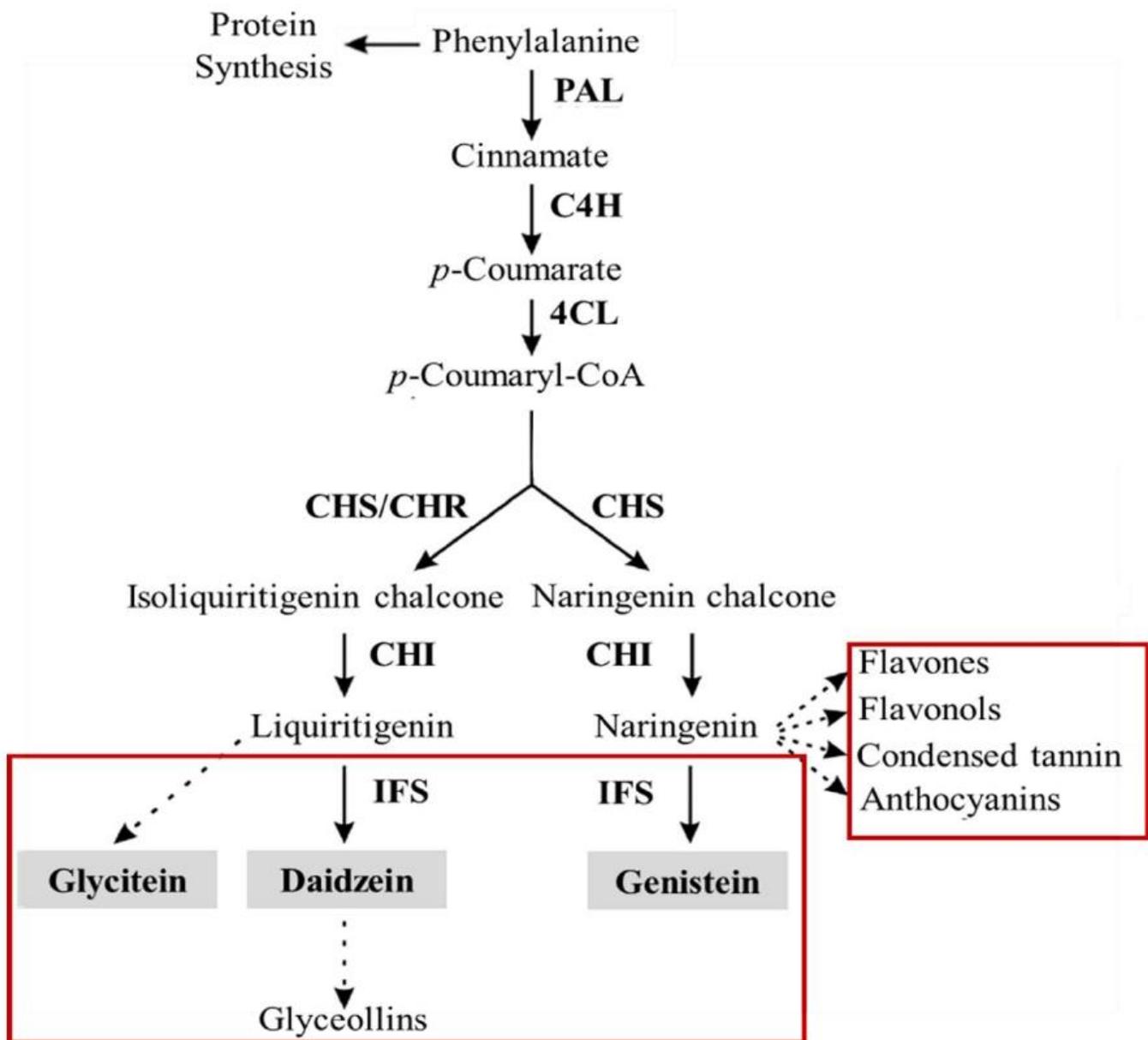


Figure 1.1. The flavonoid branch of the phenylpropanoid biosynthetic pathway

The phenylpropanoid pathway is a phenylalanine sink resulting in the synthesis of a variety of specialized metabolites (shown in red boxes). In the grey boxes are isoflavonoid metabolites. Dashed arrows indicate multiple steps involved. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CHS, chalcone synthase; CHR, chalcone reductase; CHI, chalcone isomerase; IFS, isoflavone synthase.

Modified from Dastmalchi and Dhaubhadel, (2015)

1.3 Phenylalanine, A Key Point of Divergence

Plants and microbes can synthesize Phe, while animals must acquire it through their diets (Maeda and Dudareva, 2012). The shikimate pathway produces chorismate, a precursor to Phe, tyrosine, and tryptophan. Phe and tyrosine synthesis both begin with the conversion of chorismate to prephenate by chorismate mutase (CM). From prephenate, there are two alternative pathways to phe: the prephenate dehydratase (PDT), and arogenate dehydratase (ADT) pathways (Maeda and Dudareva, 2012). As shown in Figure 1.2, the PDT pathway begins with a dehydration/decarboxylation reaction catalyzed by the PDT enzyme, converting prephenate to phenylpyruvate. Phenylpyruvate aminotransferase (PPY-AT) then catalyzes a transamination reaction, converting phenylpyruvate to Phe. The ADT pathway involves the same reactions as the PDT pathway, performed in inverse order. First, prephenate aminotransferase (PPA-AT) (Patel *et al.*, 1977) converts prephenate to arogenate, which subsequently is converted to Phe by the ADT enzyme (Fischer and Jensen, 1987). Most microbes can only synthesize Phe through the PDT pathway, and as they have no organelles, all phe production is cytosolic (Maeda and Dudareva, 2012). In plants most Phe is synthesized through the ADT pathway (Maeda and Dudareva, 2012), but growing evidence suggests that some plant ADTs also have PDT activity. ADTs with PDT activity have been characterized in *Arabidopsis thaliana* (Cho *et al.*, 2007), *Petunia hybrida* (Maeda *et al.*, 2010), and *Pinus pinaster* (El-Azaz *et al.*, 2016). Prephenate and arogenate also act as substrates to synthesize tyrosine, but a greater proportion is directed towards phenylalanine synthesis (Maeda and Dudareva, 2012).

Phe is a precursor to phenylpropanoid metabolites, but it is also a building block for protein synthesis. Regulation of Phe metabolic channeling is therefore necessary to ensure Phe derived compounds are being synthesized in correct proportions. It is believed metabolic channeling through feedback inhibition of phenylpropanoid pathway enzymes is the primary mechanism of regulation (Bubna *et al.*, 2011). Phenylalanine ammonia lyase (PAL) catalyzes the synthesis of cinnamate from Phe. Because all phenylpropanoids are downstream of cinnamate, PAL efficiency

determines the rate of Phe flux into specialized metabolite synthesis (Vogt, 2010; Barros, 2020). Upstream of PAL, ADT/PDT activity is inhibited by the accumulation of Phe (Siehl and Conn, 1988; Yamada *et al.*, 2008). It is possible that more mechanisms controlling Phe flux exist but have yet to be discovered. For example, enzymes in the same phenylpropanoid biosynthesis pathway often aggregate in organizational units called metabolons, allowing for independent regulation of each (Vogt, 2010).

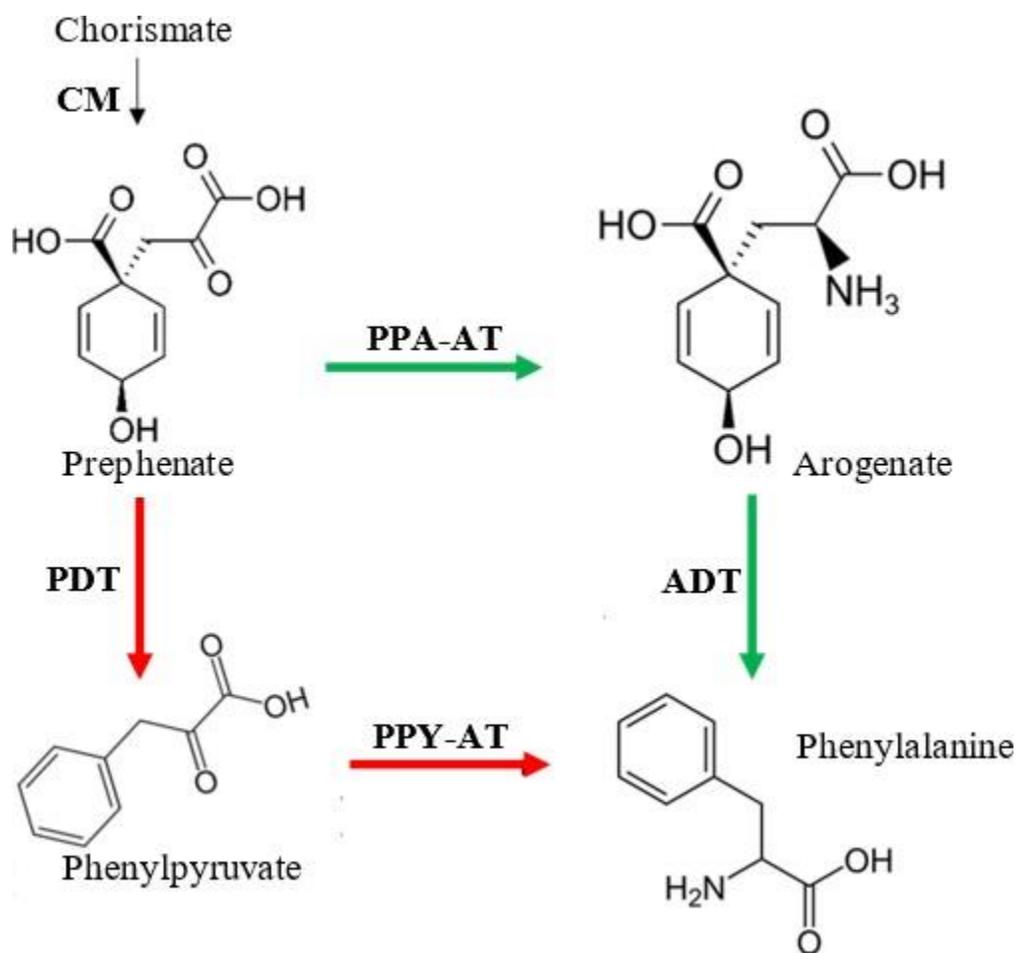


Figure 1.2. Biosynthesis of phenylalanine from chorismate

Green arrows show the ADT pathway, while red shows the PDT pathway. CM, chorismate mutase; PPA-AT, prephenate aminotransferase; ADT, arogenate dehydratase; PDT, prephenate dehydratase; PPY-AT, phenylpyruvate aminotransferase.

1.4 ADTs as Part of the Isoflavonoid Metabolon

A metabolon is the smallest unit of sub-cellular compartmentalization (Ralston, 2006). It consists of enzymes, regulatory elements, and structural elements involved in the same pathway. These pathway components are organized in a way that allows the products of one reaction to be efficiently channeled in as the substrates of the next (Ralston *et al.*, 2005).

Phenylpropanoid metabolons are anchored to the endoplasmic reticulum (ER), via cytochrome P450s (Winkel, 2004). Cytochrome P450s act as catalytic members of the metabolon, with their catalytic side facing the cytosol (Neve and Ingelman-Sundberg, 2008).

Members of the isoflavonoid metabolon are anchored to the ER by the two cytochrome P450s isoflavone synthase (IFS) and cinnamate 4-hydroxylase (Dastmalchi *et al.*, 2016). Two predicted soybean ADTs (GmADTs), Glyma.12G181800 and Glyma.13G319000, were identified as IFS-interacting partners in the metabolon (Table 1.1). As ADTs generally localize to the chloroplast, it is not clear how are they able to interact with the isoflavonoid metabolon in the cytosol. Other members of the metabolon include: chalcone isomerase, chalcone reductase, and chalcone synthase (Dastmalchi *et al.*, 2016). These findings suggest that metabolic channeling of Phe may be partially regulated by differential localization of GmADTs (Pannunzio, 2018). GmADTs in the cytosol may be involved in channeling Phe directly into specialized metabolite synthesis, while those in the chloroplast are not. ADT candidates in soybean, however, have yet to be characterized for their function, so it is not yet known which have ADT, PDT, or bifunctional activity.

Table 1.1. Characteristics of candidate *GmADTs*

Gene Name	Locus Name*	Coding Sequence Length (bp)	Predicted Mol. Weight (kDa)**	Tissue of Max Expression	FPKM***
<i>GmADT9A</i>	Glyma.09G004200	1227	44.5	Root	0.33
<i>GmADT11A</i>	Glyma.11G189100	1287	47.7	Flower	2.856
<i>GmADT12A</i>	Glyma.12G181800	1278	47.4	Leaves	33.836
<i>GmADT12B</i>	Glyma.12G085500	1287	47.7	Flower	28.768
<i>GmADT12C</i>	Glyma.12G193000	1155	43.6	Leaves	53.033
<i>GmADT12D</i>	Glyma.12G072500	933	34.6	Seed	12.215
<i>GmADT13A</i>	Glyma.13G319000	1275	46.9	Stem	24.732
<i>GmADT13B</i>	Glyma.13G309300	963	28.5	Leaves	2.494
<i>GmADT17A</i>	Glyma.17G012600	1209	44.8	Root	5.499
<i>GmADT19A</i>	Glyma.19G053000	251	9.3	-	-
GmADTU4	Glyma.U021400	1158	43.6	Seed	29.686

*Locus name is respective to the Glycine max Wm82.a2.v1 gene model.

**Molecular weight with transit peptide.

*** FPKM: fragments per kilobase per million mapped reads, acquired from RNA Seq data available on [Phytozome 12](#).

Modified from Pannunzio, (2018).

1.5 Localization of the Plant PDT Pathway

To synthesize Phe in the cytosol a functional GmADT alone is insufficient. Transient expression of *P. hybrida* PPY-AT-green fluorescent protein fusions in *Arabidopsis* protoplasts confirmed the localization of PhPPY-AT in the cytosol (Yoo *et al.*, 2013). *In vitro* enzyme assays from cytosolic and plastidial fractions prepared from wild-type petunia petals confirmed that the enzyme was functional, and its most efficient amino donor was tyrosine. In addition, PhPPY-AT activity was enriched in the cytosolic fraction while PPA-AT activity was enriched in the plastid. These results suggest that in addition to plastidial ADT-driven Phe synthesis, some plants also have a microbial-like PDT pathway in the cytosol. To complete this pathway three possible alternatives are postulated: phenylpyruvate is transported from the plastid, prephenate is transported from the plastid and there is PDT activity in the cytosol, or chorismate is transported from the plastid and there are functional isoforms of CM and PDT in the cytosol (Yoo *et al.*, 2013; Qian *et al.*, 2019).

Evidence supports the presence of both CM and PDT activity in the cytosol, suggesting the existence of a complete cytosolic PDT pathway. In plants, a plastidial chorismate mutase (CM1) is predominantly responsible for prephenate synthesis, although a cytosolic chorismate mutase isoform (CM2) has also been characterized. CM2 is less efficient than CM1, but an advantage the former has over the latter is its insensitivity to allosteric feedback inhibition (Goers and Jensen, 1984; Mobley *et al.*, 1999; Neve and Ingelman-Sundberg, 2008). *P. hybrida* ADT3 (PhADT3) is a bifunctional ADT with PDT activity (Maeda *et al.*, 2010). Experimental data from 5'-RACE, and qRT-PCR identified a *PhADT3* transcript variant. This variant lacked the chloroplast targeting sequence and had an alternate transcription start-site located just before sequence encoding the catalytic region (Qian *et al.*, 2019). The presence of multiple transcription start sites in genes is often a contingency used by plants to fine tune stress responses or tissue specific expression (Ayoubi and VanDeVen, 1996). It is possible that increased demand for phenylpropanoid compounds in response to plant biotic and abiotic stress is correlated with the expression and

activity of CM2 and PhADT3.

A. thaliana ADT2 (AtADT2) is another example of an ADT with PDT activity (Cho *et al.*, 2007). Phe produced from AtADT2 is implicated in proper seed development, and as such, *adt2* heterozygous mutants suffered seed development arrest. Since AtADT2 is a bi-functional enzyme, knockdown plants were transformed with either an exclusive ADT (*AtADT3*) or an exclusive PDT (*PHA2*) fused to the *AtADT2* promoter to determine if either would complement AtADT2 function. Both transformants rescued the mutant phenotype, but only when targeted to the chloroplast (El-Azaz *et al.*, 2018). This serves as powerful evidence that cytosolic Phe is channeled directly into phenylpropanoid synthesis, as it is metabolically unavailable for other tasks.

1.6 The Soybean Genome, and Candidate *GmADTs*

The soybean genome has been completely sequenced and annotated, and can be accessed through [Phytozome](#). Sequencing results were derived from the soybean Williams 82 cultivar. Putative transcripts and their expression patterns were determined through bioinformatic analysis of RNA-seq data. Algorithms such as PFAM, KEGG, Gene Ontology, PANTHER, and KOG make predictions about genes/proteins based on sequence similarity to other genes/proteins that have been experimentally characterized. Using the publicly available tools mentioned above, 11 different candidate *Glycine max* ADT genes (*GmADTs*) were identified (Pannunzio, 2018)(Table 1.1). Out of 11 candidate *GmADT* genes, two encode truncated proteins (*GmADT13B* and *GmADT19A*), and as such may not be functional.

Soybean is a paleopolyploid, meaning its genome has undergone at least two full duplication events (Walling *et al.*, 2006). This results in large gene families with varying divergence times. Genes duplicated more recently will likely have higher identity to one another. Additionally, gene identity will remain high if conservation of duplicated genes results in increased fitness, but if the duplicated genes are redundant there is room for one of them to accumulate mutations and change (Wu *et al.*, 2020). In the case of *GmADT* genes, the highest percentage

identity between two family members is 96%, and the lowest is 57% (Pannunzio, 2018). Each candidate *GmADT*, except *GmADT17A*, *GmADT19A*, and *GmADT9A*, has one other family member with a nucleotide identity above 94%, but identity with members outside of this pair is never higher than 84.51%.

A subcellular localization study confirmed that *GmADTs* localize to the chloroplast (Pannunzio, 2018). If this is indeed the case, how they can interact with *GmIFS* on the cytosolic side of the endoplasmic reticulum must be determined. It is possible that, like *PhADT3* (Qian *et al.*, 2019), some *GmADTs* may have an alternate transcription start site that produces a cytosolic *GmADT* isoform.

PhADT3 has both ADT and PDT activity (Maeda *et al.*, 2010). Bi-functional ADTs are a necessary constituent in cytosolic PDT driven Phe synthesis. To elucidate the purpose of *GmADT* interaction with the isoflavonoid metabolon, and whether it is related to the metabolic channeling of Phe, candidate *GmADT* functionality must first be confirmed.

1.7 ADT Architecture and Conserved Motifs

ADTs are composed of three domains: An N-terminal transit peptide, a central catalytic domain, and a C-terminal ACT regulatory domain. All characterized ADTs have transit peptide signals targeting the chloroplast. Studies in *A. thaliana* have demonstrated that ADTs remain functional in yeast with or without their transit peptide domain (Bross *et al.*, 2011). ACT domains are named after the three enzymes in which they were originally discovered: aspartokinase, chorismate mutase, and TyrA. ADTs dimerize through their ACT domains. ADTs are subject to allosteric feedback inhibition through the binding of Phe or tyrosine to ACT domain dimers (Aravind and Koonin, 1999). Among *AtADTs*, the catalytic (62.0-97.8% similarity) and ACT (61.5-91.7% similarity) domains are highly conserved, whereas the transit peptide region is more variable (Crawley 2004). It has been demonstrated that a single residue substitution (F341L) in the ACT domain of *AtADT5* was sufficient to confer it PDT activity (Smith-Uffen, 2014).

ADTs can be assigned to one of three subgroups based on their sequence similarity (Figure 1.3) (Pannunzio, 2018). GmADT19A does not belong to any of the subgroups as it is highly truncated. Subgroup 1 (GmADT12C, GmADT13B, GmADT17A, and GmADT9A) consists of ADTs that share a clade with AtADT1. AtADT1 has been shown to localize to the chloroplast (Bross *et al.*, 2017), have ADT activity, and have a small amount of PDT activity (Cho *et al.*, 2007). As these GmADTs are plastidial, they are less likely to be involved in the metabolic channeling of Phe in the cytosol. GmADTU4 and GmADT12D belong to subgroup 2. ADTs in this subgroup are in the same clade as AtADT2, an ADT which has been thoroughly characterized. AtADT2 is localized to the chloroplast, it is a bi-functional ADT/PDT, and it is required for proper seed development (Bross *et al.*, 2011; Abolhassani Rad, 2017; Bross *et al.*, 2017; El-Azaz *et al.*, 2018). AtADT2 has not however been detected in the cytosol, therefore GmADTs in the same subgroup are likely non-cytosolic as well, and for this reason they are less likely to channel Phe directly to phenylpropanoid synthesis. In subgroup 3 are GmADT11A, GmADT12A, GmADT12B, and GmADT13A. These GmADTs share a clade with PhADT3. There is a higher likelihood that one of these GmADTs may take part in soybean cytosolic PDT-driven Phe synthesis. Additionally, there is experimental evidence that GmADT12A and GmADT13A interact with the isoflavonoid metabolon in the cytosol, which is consistent with the above phylogenetic predictions (Dastmalchi *et al.*, 2016).

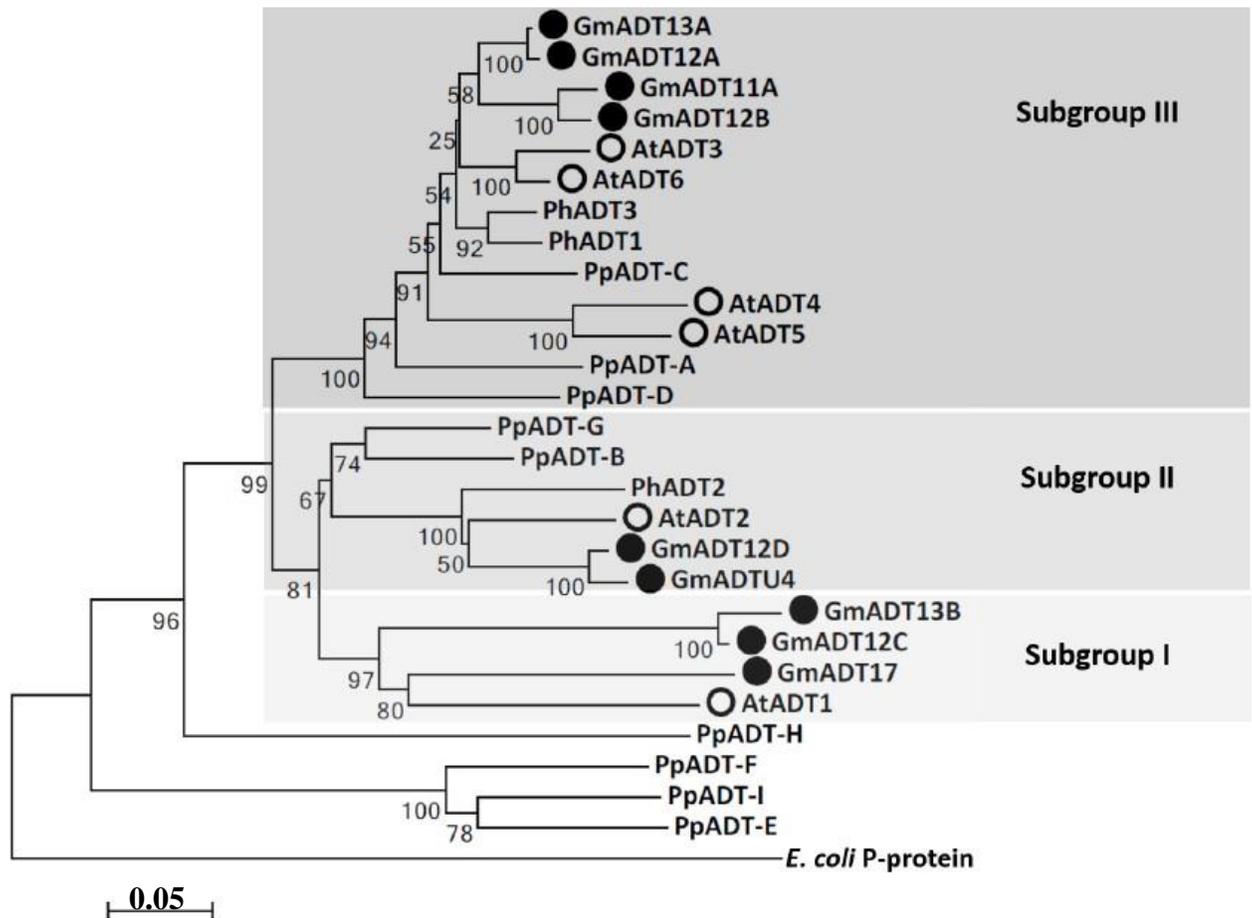


Figure 1.3. Phylogenetic relationship between GmADTs and plant and bacterial ADT/PDT sequences

A neighbour-joining tree constructed from the mature protein sequences of GmADTs, AtADTs, PpADTs, and PhADTs. Most ADTs fall within one of 3 phylogenetic clades, labelled as subgroup 1-3. (Note: GmADTU4 is now renamed GmADT11B; see section 3.1). Proteins preceded by black circles are all GmADTs, and those preceded by hollow circles are AtADTs.

Adapted from Pannunzio, (2018)

1.8 Hypothesis

I hypothesize that candidate GmADTs possess ADT activity and that some are bi-functional with PDT activity. Additionally, some *GmADTs* express two transcripts: one encoding a transit peptide, and one without.

1.9 Objectives

The objectives of my research are the following:

1. To characterize ADT and PDT activity of candidate GmADTs. This entails cloning candidate GmADTs for functional characterization and performing both ADT and PDT assays.
2. To identify candidate *GmADT* transcript variants that exclude sequence encoding the transit peptide.

Chapter 2: Materials and Methods

2.1 Plant Growth Conditions

Soybean (*Glycine max* L. Merr.) cultivar Williams 82 seeds were planted in sterile pots containing PRO-MIX® BX MYCORRHIZAE soil. The growth chamber was maintained at 60% humidity, 24°C, and a light intensity of 78 $\mu\text{mol m}^{-2} \text{s}^{-1}$ with a 16-hour light cycle followed by 8 hours of darkness. After an initial watering, the pots were covered with plastic bags for two days to create a high humidity environment promoting germination. The plants were watered with a fertilizer solution [nitrogen-phosphorus-potassium (20-8-20) dissolved in water at 0.5 g/L] twice, first immediately after planting, and again when seedlings were visible. The plants were grown until they flowered, then harvested for stem, leaf, root, and flower tissue.

2.2 Bacterial and Yeast Strains

Escherichia coli DH5 α was used to store and replicate entry and destination vectors.

To determine the ADT activity of candidate GmADTs *Saccharomyces cerevisiae aro8aro9* was used. It is a double mutant strain, formed by mating *aro8* and *aro9* single mutants (Giaever *et al.*, 2002), that cannot synthesize Phe through the PDT pathway, as its phenylpyruvate aminotransferase activity has been knocked out. Additionally, *aro8aro9* is transformed with a second plasmid (pAG425GAL) containing an *Arabidopsis thaliana* prephenate aminotransferase (AtPAT), to allow for synthesis of Phe through the ADT pathway.

To determine the PDT activity of candidate GmADTs, *S. cerevisiae pha2* strain was used. It is a PDT knockout mutant (Giaever *et al.*, 2002; Bross *et al.*, 2011).

2.3 Bacterial Media

Lysogeny broth (LB) media was used for all bacterial applications. It was made by dissolving tryptone, yeast extract, and NaCl in deionized water, in concentrations of 1%, 0.5%, and 0.5% w/v, respectively. Where solid medium was required, a 1% concentration of agar was used. The LB

medium was then autoclaved for 15-minutes. Filter (0.2 μm) sterilized antibiotics were added to the medium after it had cooled, but before solidification. Super Optimal Broth with Catabolite Repression (SOC) was used to resuspend *E. coli* after transformation by electroporation. It was composed of 1% tryptone, 0.5% yeast extract w/v, 10mM NaCl, 2.5 mM KCl, 10 mM MgCl_2 , and 20 mM glucose. The SOC medium was sterilized in the autoclave for 15 minutes, then, after cooling, divided into 200 μL aliquots. The aliquots were stored at -20 $^\circ\text{C}$.

2.4 Agarose Gel Electrophoresis

Agarose gels were used to separate nucleic acid fragments based on their size. A 1% gel was made when analyzing fragments 1000 bp and larger, while a 2% gel was used for fragments smaller than 1000 bp. Gels were made with 0.5X Tris-borate EDTA (TBE) buffer when running DNA, and 1X Tris-acetate-EDTA (TAE) for RNA. RedSafe stain (iNtROn Biotechnology) was added to gels at 0.01% v/v before they solidified. Bluejuice gel loading buffer 10x (Invitrogen 10816015) was used for DNA. Electrophoresis tanks were filled with either 0.5X TBE (DNA), or 1X TAE (RNA). Gels were run at 100kV for 25 minutes and visualized using the Bio-Rad Gel DocTM.

2.5 RNA Extraction, Reverse Transcription, and Quantitative Real-Time PCR (qRT-PCR)

Total RNA was extracted from soybean root, stem, leaf, and flower tissue (50-70 mg) using the RNeasy plant Mini kit (Qiagen). Before eluting RNA, DNase I (Promega) was used to perform on-column DNA digestion. Purity and quantity of RNA was assessed using a NanoDrop 1000 spectrophotometer (Thermofisher). RNA quality was assessed by running 500 ng of total RNA on a TAE gel. cDNA was synthesized from 1.0 μg of total RNA using oligo dT primers and the SuperScript IV First Strand Synthesis System (Thermofisher) following manufacturer's instructions.

For qRT-PCR, oligo dT primers were used to synthesize cDNA through the

ThermoScript™ RT-PCR System (Invitrogen), from 1 µg of total RNA. *GmADT* gene-specific primer pairs were used to amplify cDNA with the SsoFast EvaGreen Supermix (Bio-Rad) (Table 2.1). Transcript abundance was normalized to the soybean reference gene *CONS4* (Libault et al., 2008). qRT-PCR was performed in a Bio-Rad C1000 Thermal Cycler with the CFX96™ Real-Time PCR system. Each sample was run in triplicate as technical replicates. Quantification cycle (C_q) values were recorded at an amplification threshold of 220 relative fluorescence units. Data were analyzed in CFX Maestro (Bio-Rad).

Table 2.1. Comprehensive primer list

Gateway cloning primers show *attB* sequences in regular font, sequence encoding 6 histidine in bold, and *GmADT* coding sequence underlined.

Purpose	Gene name	Primer name	Sequence 5' to 3'	Amplicon size (bp)
UTR amplification	<i>GmADT11A</i>	GmADT11AUTF	GAGAGAGAGAGTGATGGGGGTAG	1284
		GmADT11AUTRR	CTCAACATGGCGAACAGCAAAG	
	<i>GmADT11B</i>	GmADTU4F	AAACCCAAACACTGTCTCCGTCTGG	1204
		GmADTU4R	CGGGAGTCCACTTTTGAAGATTATCG	
	<i>GmADT12A</i>	GmADT12AF	CTCTCAAACCATAATATGCAGACTCTTTCGCCGCCT	1347
		GmADT12AR	TTTAAATTTATCTCCCCGGGAGGAAGGTGTCCA	
	<i>GmADT12B</i>	GmADT12BF	CAACTAAATTCCCCTTTCCAACC	1370
		GmADT12BR	ATGAAAGAAATGGAGGTGGATG	
	<i>GmADT12C</i>	GmADT12CF	TTGAGAACCGTTGACCTCC	1279
		GmADT12CR	TATTTGGACATGAAGGTAGCTGC	
	<i>GmADT12D</i>	GmADT12DF	GGGCAAAGAATAGGAAGTCTT	1001
		GmADT12DR	GTTGTTGATTCATTGTATGC	
	<i>GmADT13A</i>	GmADT13AF	ATTCCTCTGTCAAGCCACTCG	1322
		GmADT13AR	CAAGAGGGGAAAAAGACGATGC	
	<i>GmADT17A</i>	GmADT17AF	TTCATTTTGATGGCTCTTAAGGCTG	1247
		GmADT17AR	CAGCAAATGAACAGCATGACT	
	<i>GmADT9A</i>	GmADT9F	GGTGAGAACTTCTTTCTTCTTTGAC	1267
		GmADT9R	CGAGAAGATTCATCCCAAGTTGATC	

Table 2.1 con't

Gateway Cloning	<i>GmADT11A</i>	GmGTADT11AF	GGGGACAAGTTTGTAC AAA AAA GCA GGC Ttc <u>ATGCAGACCCTCAATCAA</u>	1366
		GmADT11AR-GWY2	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTc CTA ATG ATG ATG ATG GTG ATG <u>ATTTGGCGCGGACA</u>	
	<i>GmADT11B</i>	GmGTADTU4F	GGGGACAAGTTTGTAC AAA AAA GCA GGC Ttc <u>ATGGCGGCATCGCGAATCGTG</u>	1267
		GmADTU4R-GWY2	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTc CTA ATG ATG ATG ATG GTG ATG <u>CGTCAAGCTAGTGTC</u>	
	<i>GmADT12A</i>	GmGTADT12AF	GGGGACAAGTTTGTAC AAA AAA GCA GGC Ttc <u>ATGCAGACTCTTTCGCC</u>	1366
		GmADT12AR-GWY2	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTc CTA ATG ATG ATG ATG GTG ATG <u>TTTAAATTTATCTCC</u>	
	<i>GmADT12B</i>	GmGTADT12BF	GGGGACAAGTTTGTAC AAA AAA GCA GGC Ttc <u>ATGCAGACCCTCACCC</u>	1366
		GmADT12BR-GWY2	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTc CTA ATG ATG ATG ATG GTG ATG <u>ATTTGGCGCGGAGA</u>	
	<i>GmADT12C</i>	GmGTADT12CF	GGGGACAAGTTTGTAC AAA AAA GCA GGC Ttc <u>ATGGCTGTGACATCACCTCTTG</u>	1234
		GmADT12CR-GWY2	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTc CTA ATG ATG ATG ATG GTG ATG <u>TATGGTTGTATCTAT</u>	
	<i>GmADT12D</i>	GmGTADT12DF	GGGGACAAGTTTGTAC AAA AAA GCA GGC Ttc <u>ATGGCTGCGTCGCGAATC</u>	1012
		GmADT12DR-GWY2	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTc CTA ATG ATG ATG ATG GTG ATG <u>AAGTACCTTTGTAAG</u>	
	<i>GmADT13A</i>	GmGTADT13AF	GGGGACAAGTTTGTAC AAA AAA GCA GGC Ttc <u>ATGCAGAGTCTTTCACCACC</u>	1354
		GmADT13AR-GWY2	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTc CTA ATG ATG ATG ATG GTG ATG <u>GTCTCCCCGGGAGGA</u>	
	<i>GmADT17A</i>	GmGTADT17AF	GGGGACAAGTTTGTAC AAA AAA GCA GGC Ttc <u>ATGGCTCTTAAGGCTGTATC</u>	1288
		GmADT17AR-GWY2	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTc CTA ATG ATG ATG ATG GTG ATG <u>GTTAAGACACTGAAC</u>	
<i>AtPAT</i>	AtPATGTF	GGG GAC AAG TTT GTA CAA AAA AGC AGG CT <u>ATGTCATCTAGAATCTGCGCTATGG</u>	1428	
	AtPATGTR	GGG GAC CAC TTT GTA CAA GAA AGC TGG GT <u>AAACGGAGACAGTGGCACG</u>		

Table 2.1 con't

qPCR Primers	<i>GmADT11A</i>	GmADT11ALqPCRf	CCTGAGACATATGATACCTGACAA	195
		GmADT11ALqPCRR	TTCAATCCCACACCTACCCC	
	<i>GmADT11A</i>	GmADT11AqPCRf	TTAGGGCACAGAACGCACTC	96
		GmADT11AqPCRR	AGGGAGTCCAGGGTGACATA	
	<i>GmADT12A</i>	GmADT12ALqPCRf	ACCCATTTCTCTGCCTGTC	135
		GmADT12ALqPCRR	TTGGAGAAGGGAACCTCAAGC	
	<i>GmADT12A</i>	GmADT12AqPCRf	ACCCATTTCTCTGCCTGTC	93
		GmADT12AqPCRR	GGTATCGCGGAGGTTATTCGT	
	<i>GmADT12B</i>	GmADT12BLqPCRf	AACTATCTCACCCGGGCGC	207
		GmADT12BLqPCRR	GTTGTCGTGGTCGTCGCCGT	
	<i>GmADT12B</i>	GmADT12BqPCRf	GAAACTCGGCCTCACCGTGA	152
		GmADT12BqPCRR	GATTCCGTCGGCCAGGATCTGG	
	<i>GmADT13A</i>	GmADT13ALqPCRf	ACAGCAAGCAATCGTTACCC	131
		GmADT13ALqPCRR	CGCGTTGGAAGTAGGTGGT	
	<i>GmADT13A</i>	GmADT13AqPCRf	CGCCTCACATAGTTGGCGA	84
		GmADT13AqPCRR	CGTGAGGAACTCCTTTCGGA	

Table 2.1 con't

Internal Sequencing Primers	<i>GmADT11A</i>	GmADT11AintF	GCGGGGAAAGCGTATCCG	N/A
		GmADT11AintR	GATTCCGTCGGCC	
	<i>GmADT11B</i>	GmADTU4intF	AGCACAGAAGGCCTACCC	N/A
		GmADTU4intR	ATCATAGGTTCTCGTGCTAAC	
	<i>GmADT12A</i> and <i>GmADT13A</i>	GmADT12AintF	GCACAGGCCAACAACAA	N/A
		GmADT12AintR	GCTCGGATCGTCTTGG	
	<i>GmADT12B</i>	GmADT12BintF	GCTGCGTTTCAAGCGGTG	N/A
		GmADT12BintR	CTCTTCAAAGTGCTTTCCGCG	
	<i>GmADT12C</i>	GmADT12CintF	GGTGGGCTGACAAAGTTAT	N/A
		GmADT12CintR	GCACGAGATAACGACTG	
	<i>GmADT12D</i>	GmADT12DintF	AAGCTGTGCCTTGTGAACAA	N/A
		GmADT12DintR	AGAACTAGCAACTGCTCCT	
	<i>GmADT17A</i>	GmADT17AintF	CCAAGAGATGGATCAAAGGTG	N/A
		GmADT17AintR	CTCCCTTGCAAGGACCAAG	

2.6 Cloning and Transformation

To ensure amplicon specificity, *GmADT* coding sequences were amplified by a two-step nested PCR. The first PCR amplification was performed using cDNA template and *GmADT* gene-specific UTR primers (Table 2.1). The amplicons obtained from the first PCR were used as templates for the second PCR, where primers designed for gateway cloning (GWY2) flanked the *GmADT* coding region with attB sequences. The reverse primer included a sequence encoding six histidine (Table 2.1). *GmADTU4* (Note: *GmADTU4* is now renamed *GmADT11B*; see section 3.1), *GmADT12A*, *GmADT12B*, and *GmADT17A* UTR amplification primers, all *GmADT* gateway forward primers, and all internal sequencing primers were acquired from Pannunzio 2018. Each amplification was catalyzed by Platinum® High Fidelity *Taq* DNA Polymerase. The thermocycler was set to denature at 94°C for 15 seconds, anneal at a primer-specific T_m for 30 seconds, and extend at 68°C for 1.5 minutes. Cycles were repeated 40 times, then completed with a final extension at 68°C for 5 minutes. Amplicons were run on agarose gels, then purified using the EZ- 10 Spin column DNA Gel Extraction Kit (Bio Basic Inc). Purified DNA was quantified using the NanoDrop 1000 spectrophotometer (ThermoScientific).

BP clonase® II Enzyme mix (Invitrogen) was used to recombine *GmADT* GWY2 amplicons into the pDONRzeo entry vector (Invitrogen) (Figure 2.1).

BP reaction (1 μ L) products were transformed into electro-competent *E. coli* DH5 α cells (50 μ L) by electroporation. Electroporation was performed in a 50 μ L Gene Pulser® Cuvette (Bio-Rad Laboratories, Inc.), inserted into a MicroPulser™ electroporator (Bio-Rad Laboratories, Inc.), and electroporated at 1.8 kV. Cells were immediately resuspended in 1 mL of SOC media, transferred to a 2 mL Eppendorf, and incubated for 1 hour at 37°C and 200 rpm in a shaker incubator. Transformation reaction (100 μ L) was plated on LB medium supplemented with zeocin (50 μ g/mL). Plates were incubated overnight at 37°C.

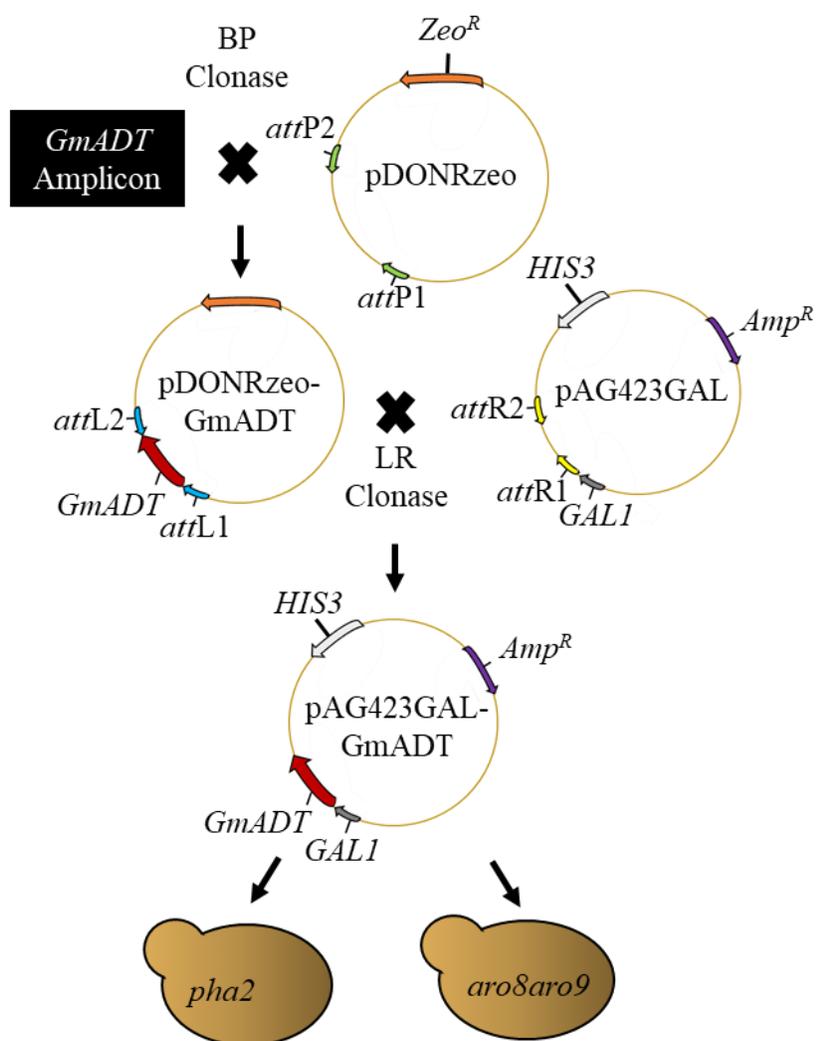


Figure 2.1. Cloning strategy of *GmADT* genes

GmADTs were recombined into the entry vector (pDONRzeo) by BP clonase. The entry clone was propagated in *E. coli* DH5 α . Positive transformants were identified using zeocin resistance as a selectable marker. Finally, pDONRzeo-*GmADTs* were recombined into the destination vector (pAG423GAL) by LR clonase, propagated in *E. coli*, and identified using ampicillin as a selectable marker. The destination clone was transformed into *aro8aro9* and *pha2* for the ADT and PDT assay, respectively.

To confirm transformants, 8-16 overnight grown *E. coli* DH5 α colonies were picked for colony PCR using entry vector specific M13R and M13F primers. The amplicons were run on a 1% agarose gel to check the specific fragment size for *GmADTs*. Positive colonies were re-suspended in LB zeocin (50 $\mu\text{g}/\text{mL}$) liquid medium and grown overnight at 37°C and 200rpm for plasmid DNA isolation. The pDONRzeo-GmADT plasmid DNA was extracted using the EZ-10 Spin Column Plasmid DNA kit (Bio Basic Inc.) following manufacturer's instructions and quantified using the NanoDrop 1000 spectrophotometer (ThermoScientific). Plasmid DNA isolated from transformed colonies (8 μL) and a sequencing primer (4 μL) in pre-labelled tubes were sent to Eurofins Scientific for sequencing. Sequencing results were analyzed in SeqMan Pro 17 (DNASTAR). Following sequence confirmation of *GmADT* genes in the entry vector, the LR clonase® II Enzyme mix (Invitrogen) was used to recombine pDONRzeo-GmADTs into the destination vector pAG423GAL. LR reaction products were transformed into electro-competent *E. coli* DH5 α and plated on LB ampicillin (50 $\mu\text{g}/\text{mL}$). Colony PCR was done using gene-specific GWY2 primers (Table 2.1). Positive colonies were grown overnight using LB ampicillin liquid medium (50 $\mu\text{g}/\text{mL}$) at 37°C and 200 rpm. pAG423GAL-GmADT plasmid DNA was extracted and quantitated as described above.

Competent yeast cells were prepared, and transformation was done as described by the Frozen-EZ Yeast Transformation IITM kit (Zymo Research). Yeast were grown in yeast extract peptone dextrose (YPD) broth to mid-log phase (OD₆₀₀ 0.8-1.0). To make competent cells, 1 mL of mid-log phase culture was centrifuged at 500 x g for 4 minutes in a 10 mL Falcon tube. The pellet was washed twice using 1 mL EZ 1 solution, then resuspended in 100 μL EZ 2 solution. Transformation was performed by adding 1 μg of pAG423GAL-GmADT DNA to 50 μL competent yeast and 500 μL EZ 3 solution, then incubating at 30 °C for 45 minutes, vortexing vigorously 2-3 times during the incubation. pAG423GAL-GmADT constructs were transformed into *S. cerevisiae* *aro8aro9*, and *pha2* for the ADT and PDT assay, respectively. The transformation mixture (100 μL) was plated on either synthetic defined (SD) dropout minus

histidine (-His) (*pha2*), or SD dropout -His minus leucine (-Leu) (*aro8aro9*) media. Positive colonies were confirmed using yeast colony PCR. For this, colonies were picked from plates and transferred to PCR tubes with 20 μ L 0.02 M NaOH to degrade yeast cell walls. The suspended yeast (1 μ L) was used as a template for PCR amplification with gene-specific GWY2 primers. Primers used in the study are listed in Table 2.1.

2.7 ADT assay

Assay media were prepared from minimal synthetic defined (SD) base (3.35 g) with amino acid dropout supplement (0.32 g), the appropriate sugar (10 g), and agar (10 g) if required, in 500 mL H₂O with the pH adjusted to 6.0. Assay media were made using every possible dropout combination of -His, -Leu, and -Phe. Assay plates were made using either galactose (Gal) or glucose (Glu). Raffinose (Raf) -Leu, Raf -His, and Raf -Leu/-His were used for yeast growth pre-assay, because Raf neither induces nor represses the expression of *GmADTs* transformed in yeast.

Overnight cultures of *aro8aro9* transformants were prepared, and a hemocytometer was used to determine concentration of cells/ μ L in each. All cultures were subsequently diluted with water to a final concentration of 1000 cells/10 μ L. Assay plates were spotted separately with 10 μ L of each diluted culture, then incubated at 28°C, and photos were taken after thirteen days.

2.8 PDT assay

pha2 transformants contain pAG423GALGmADTs. The method for making PDT assay plates was the same as that for ADT, except only combinations of -Phe and -His dropouts were required. These yeasts were also grown in overnight cultures using Raf -His medium, and subsequently diluted

with water to 1000 cells/10 μ L. Assay plates were spotted separately with 10 μ L of each diluted culture and incubated at 28°C for thirteen days.

2.9 Protein Extraction

aro8aro9 and *pha2* transformants were cultured in minimal SD induction medium (1% Raf, 2% Gal) for 72 hours. Cell suspensions were centrifuged at 1500 x g for 5 minutes at 4°C. The pellet was washed with ddH₂O, resuspended in 0.1 M NaOH, and incubated at room temperature for 5 minutes. Yeast cells were spun down at 1500 x g for 5 minutes and resuspended in 50 μ L protein extraction buffer containing 10 mM Tris (pH 6.8), 10 mM NaCl, and 2% sodium dodecyl sulfate (SDS). Prior to use, dithiothreitol (100 mM) and 100 x protease inhibitor cocktail (Sigma) were added to the protein extraction buffer. Samples were boiled for 3 minutes at 95°C to denature proteins and separated using SDS-polyacrylamide gel electrophoresis (PAGE).

2.10 SDS-PAGE and Western Blotting

Total protein extracted from each yeast culture were separated by size on a polyacrylamide gel (Bio-Rad). Gels were handcast in mini size (10.0 x 8.0 cm) using 12% resolving and 6% stacking components in a 1.0 mm gel cassette. Solidified gels were mounted in pairs in a vertical electrophoresis chamber, and the chamber was filled with PAGE running buffer. Equal volumes of protein extract (10 μ L) were loaded on 2 gels (one for staining and another for Western blot analysis) and run at 140 v for 1 hour and 15 minutes. A protein ladder (Froggabio) was used to identify approximate size of proteins.

Gel staining was performed using coomassie brilliant blue staining solution for 1 hour. It was subsequently destained using H₂O, methanol, and acetic acid in a ratio of 50/40/10 (v/v/v).

For Western blot analysis, PVDF membrane (Bio-Rad) was cut to the appropriate size, wet in 100% methanol for 1 hour, and moved to transfer buffer for 15 minutes prior to use. Transfers were done using the Trans-Blot® SD Semi-Dry Transfer Cell (Bio-Rad). The transfer

was performed at 12 V for 30 minutes. Once transfer was completed the membrane was washed in Tris buffer saline solution (TBS) for 15 minutes with gentle agitation. It was then blocked overnight in TBS containing 5% non-fat milk powder and 0.1% bovine serum albumin (BSA) at room temperature. After blocking, the membrane was washed in TBS containing 0.05% tween 20 (TBST) for 15 minutes. The membrane was then incubated with mouse anti-His antibody (Sigma) (1:1000 dilution) in TBST at 4°C overnight, followed by 3 washes with TBST, each for 5 minutes at room temperature. Goat anti-mouse antibody with conjugate horse-radish peroxidase (EMD Millipore) at a 1:2000 dilution in TBST was used to probe the membrane for 2 hours, followed by 3 final wash cycles in TBST, each for 5 minutes. Membranes were treated with enhanced chemiluminescence substrate (Bio-Rad) for 5 minutes to induce signal. Signal was detected using the ChemiDoc Imaging System (Bio-Rad).

Chapter 3: Results

3.1 Cloning and Transformation of *GmADTs*

Using the *Glycine max* full genome annotation on [Phytozome 12](#) (Wm82.a2.v1), 11 candidate *GmADT* genes had been identified (Table 1.1) (Pannunzio, 2018). Recently, however, the whole genome sequence data have been re-annotated (Wm82.a4.v1) and the updated version can be accessed through [Phytozome 13](#). Thus, the new assembly was investigated to assess if there had

been changes to previously identified *GmADTs* by Pannunzio (2018). I discovered that *GmADT19A* was no longer annotated as a putative ADT. Furthermore, the previously unscaffolded *GmADTU4* was assigned to chromosome 11, and for this reason was renamed *GmADT11B* (Table 3.1). The remaining *GmADTs* were consistent between the two annotations. *GmADT13B* produces a truncated protein, is likely non-functional, and was excluded from functional characterization. Therefore, in accordance with the new assembly, the soybean genome contained 9 candidate *GmADTs*. All candidate *GmADTs* were cloned in a yeast expression vector, and verified by sequencing, except *GmADT9A*, which could not be amplified despite several attempts. Of the 8 cloned *GmADTs*, three (*GmADT12A*, *GmADT12C* and *GmADT12D*) have sequence polymorphisms which resulted in amino acid substitutions (Table 3.1). Most of the sequence polymorphisms resulting in amino acid substitutions occur in the transit peptide region. *GmADT12D* has an amino acid substitution (I269M) which occurs in a functional region of the protein. Clones were sequenced in both the entry and destination vectors to ensure that replication in *E. coli* DH5 α did not introduce any additional polymorphisms, and to confirm the presence of a sequence encoding 6 histidine residues after the *GmADT* coding sequence. There were no additional polymorphisms and a sequence encoding 6 histidine was present in all *GmADT* sequences except *GmADT11B*. *GmADT* clones were transformed into *aro8aro9* PAT and *pha2* yeast strains.

Presence of all clones in yeast was confirmed by colony PCR using gene-specific primers. As shown in Figure 3.1, *pha2* transformants contain only *GmADT* constructs, while *aro8aro9* PAT transformants contain both *AtPAT* and *GmADT* constructs. Expected sizes of *GmADT* genes are shown in Table 3.1.

Table 3.1. List of candidate *GmADTs* in this study.

Gene Name	Locus Name*	Coding Sequence Length (bp)	Predicted Mol. Weight (kDa)**	N-Terminal Tags	Protein Substitutions
<i>GmADT11A</i>	Glyma.11G189100	1287	47.7	His Tag	N/A
<i>GmADT11B</i>	Glyma.11G151288	1158	43.57	-	N/A
<i>GmADT12A</i>	Glyma.12G181800	1278	47.42	His Tag	P7S
<i>GmADT12B</i>	Glyma.12G085500	1287	47.66	His Tag	N/A
<i>GmADT12C</i>	Glyma.12G193000	1155	43.64	His Tag	G11V H17R H62R
<i>GmADT12D</i>	Glyma.12G072500	933	34.63	His Tag	L55P I269M
<i>GmADT13A</i>	Glyma.13G319000	1275	46.93	His Tag	N/A
<i>GmADT17A</i>	Glyma.17G012600	1209	44.8	His Tag	N/A

*Locus name is respective to the Glycine max Wm82.a4.v1 gene model.

**Molecular weight with transit peptide.

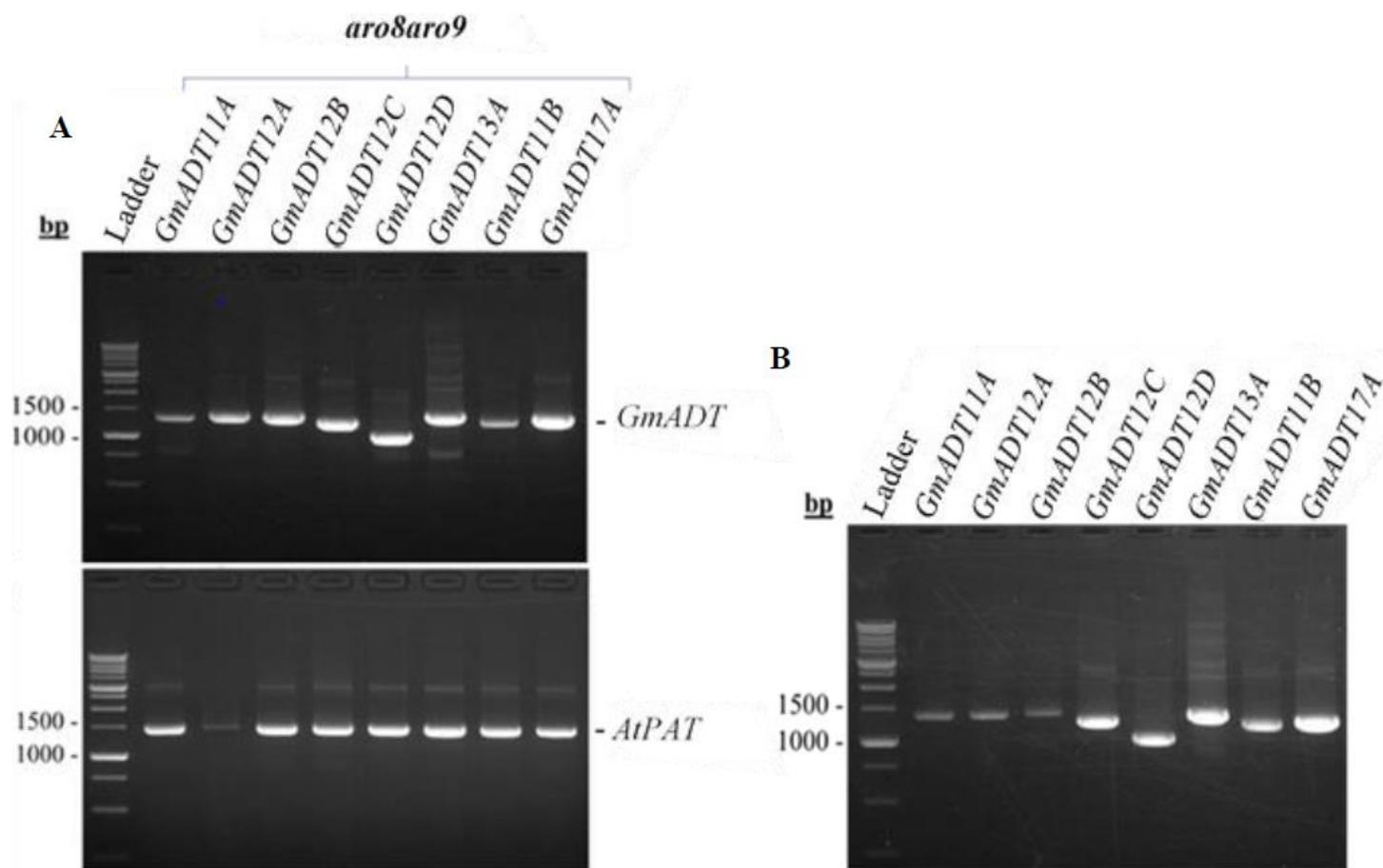


Figure 3.1. Yeast colony PCR

Presence of candidate *GmADTs* and *AtPAT* in *S. cerevisiae aro8aro9* (A) and *GmADTs* in *S. cerevisiae pha2* (B) was confirmed by colony PCR using gene-specific primers.

3.2 ADT Assay

To determine the ADT activity of GmADTs, complementation of PPY-AT knockout *aro8aro9* was performed. *aro8aro9* transformants tested for ADT activity contain two exogenous plasmids, one containing a sequence encoding a functional PAT from *Arabidopsis* and the other a candidate GmADT enzyme. The first plasmid (pAG425GAL) confers leucine synthesis, while the second (pAG423GAL) histidine. These plasmids are necessary and sufficient for *aro8aro9* to grow on media lacking Leu and His, respectively. The ADT enzyme assay showed expected results for the positive and negative controls (Figure 3.2). Transformants with only pAG423GAL were able to grow on SD medium lacking His. Those with only pAG425GAL were able to grow on plates lacking Leu. When both vectors were present, transformants were able to grow on media lacking both His and Leu. *aro8aro9* without the vectors did not grow at all (Figure 3.2).

Both AtPAT and GmADT expressions were under the control of a GAL1 promoter. Presence of galactose induces protein expression, while glucose represses it. Thus, yeast containing GmADTs that have ADT activity were able to grow on Gal medium lacking Phe. *In vitro* enzyme assays have demonstrated that AtADT3 has the most efficient ADT activity (Cho *et al.*, 2007), so it was used as a positive control. There were small visible colonies for all transformants containing both vectors on His, Leu, and Phe triple knockout (TKO) Glu medium (Figure 3.2). The positive, GmADT11B, GmADT12B, and GmADT12C were the only transformants with visible colonies on Gal TKO medium. The GmADT11B colonies on Glu TKO medium were smaller than the ones on Gal TKO medium. This is strong evidence for GmADT11B ADT activity. The remaining transformants with visible colonies on Gal TKO medium were approximately the same size as their Glu counterparts. On Gal TKO plates, only the positive, GmADT11B, GmADT12B, and GmADT12C grew consistently across all replicates (Appendix. C), while on Glu TKO, every transformant carrying both vectors consistently grew. The other two GmADTs, and the positive control also have ADT activity, but they are less efficient than GmADT11B.

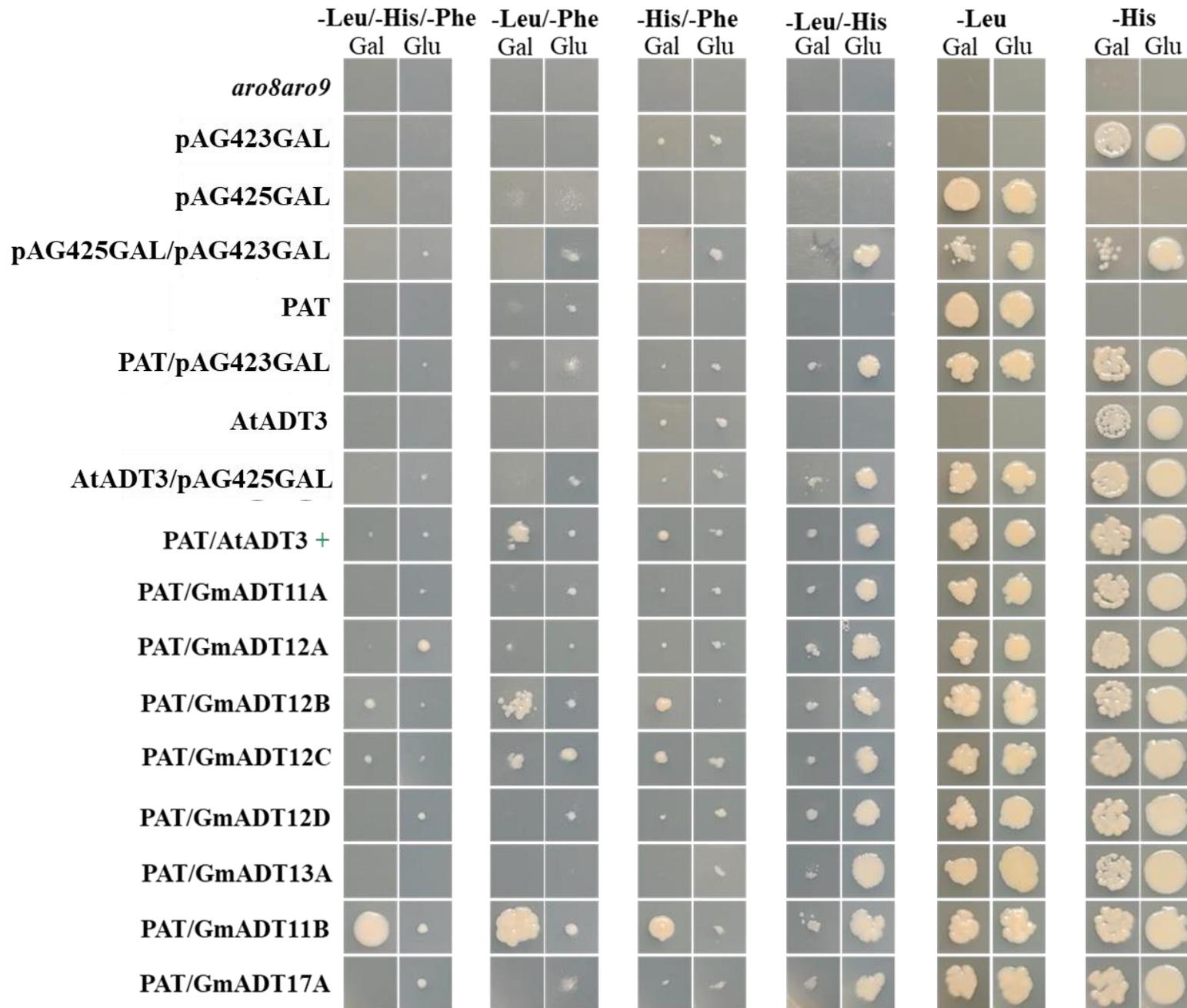


Figure 3.2. Yeast complementation assay testing ADT activity

Yeast cells were spotted (1000 cells in 10 μ L) on SD media containing glucose or galactose, and + or - phe. All plates were incubated at 30 $^{\circ}$ C for 13 days. GmADT11A - GmADT17A: *aro8aro9* containing a pAG423GAL-GmADT vector; *aro8aro9*: untransformed strain; pAG423GAL: *aro8aro9* containing an empty pAG423GAL vector; pAG425GAL: *aro8aro9* containing an empty pAG425GAL vector; PAT: *aro8aro9*.

3.3 PDT Assay

The PDT activity of candidate GmADTs was examined through complementation of the PDT knockout *S. cerevisiae pha2*. This assay required *pha2* to be transformed with only a single vector (pAG423GAL) containing the candidate *GmADT*. As shown in Figure 3.3, all transformants carrying the vector were able to grow on -His SD media with either glucose or galactose. *pha2* without the vector did not grow. AtADT2 was used as a positive control, as it was already confirmed to have PDT activity through *in vitro* enzyme assays (Cho *et al.*, 2007). Colonies were seen for the positive control, GmADT12A, GmADT13A, and GmADT11B on His and Phe double knockout (DKO) Gal medium. No growth was seen on DKO Glu medium. These results suggest that GmADTs with visible colonies on Gal medium have PDT activity to varying degrees. The GmADT11B colony on DKO Gal medium grew the largest in 13 days, so it likely has the highest amount of PDT activity. The next largest colonies were those of GmADT12A, followed by GmADT13A. GmADT12A and GmADT13A have PDT activity but are less efficient than GmADT11B. All plates and expected growth patterns of the yeast transformants can be visualized in Figure 3.3 A and B, for the ADT and PDT assays, respectively.

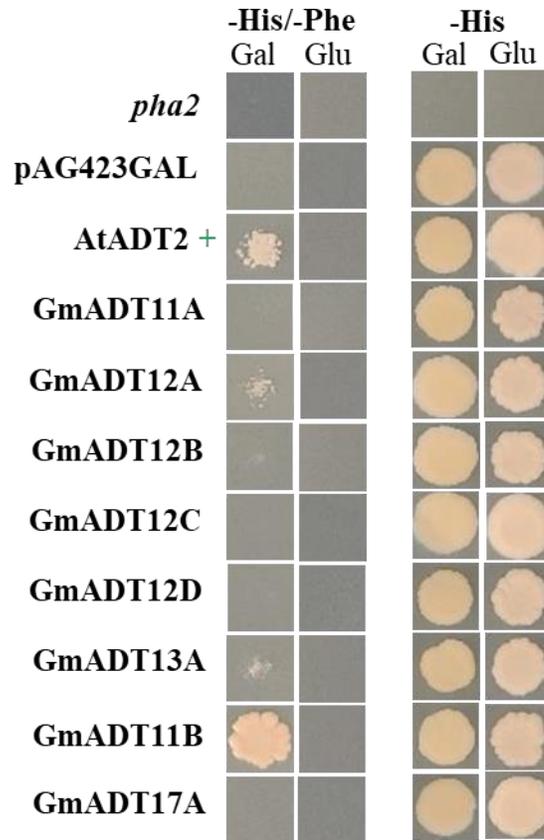


Figure 3.3. Yeast complementation assay testing PDT activity

Yeast cells were spotted (1000 cells in 10 μ L) on SD media containing glucose or galactose, and + or - phe. All plates were incubated at 30 °C for 13 days. GmADT11A - GmADT17A: *pha2* containing a pAG423GAL-GmADT vector; *pha2*: untransformed strain; pAG423GAL: *pha2* containing an empty pAG423GAL vector; AtADT2: *pha2* containing a pAG423GAL-AtADT2 vector. The positive control is marked with a plus.

3.4 Western Blot Analysis

To verify that lack of growth in the ADT and PDT assay was due to GmADTs' inability to complement the mutant phenotype and not because of the absence of the protein, western blot analysis was performed. As described previously in section 2.6, each candidate *GmADT* gene, except for *GmADT11B*, was translationally fused with 6 His residues at the C terminal end of the protein. Total protein was extracted from ADT and PDT assay yeast transformants, and anti-His antibodies were used to detect presence of candidate GmADTs. Purified GmCHR14-His was used as a positive control. The negative control was total protein extracted from GmADT11B transformed yeast, as it lacked the His tag. As shown in Figure 3.4, the signal from the positive control (37.86 kDa) was detected as a band between 35 and 48 kDa. No signal was detected for GmADTs in *aro8aro9* apart from what that was also present in the negative control. This suggests that GmADTs may not have been expressed in *aro8aro9* GmADT transformants that did not grow on -Phe GAL SD dropout medium. Coomassie gels show protein extraction was successful.

The Western blot analysis for presence of GmADTs in *S. cerevisiae pha2* showed a signal for all GmADTs except for GmADT17A (Figure 3.5). This is evidence for lack of PDT activity in GmADTs that failed to complement the *pha2* mutant phenotype. The *pha2* GmADT17A transformants may have failed to complement because of lack of expression rather than lack of PDT activity.

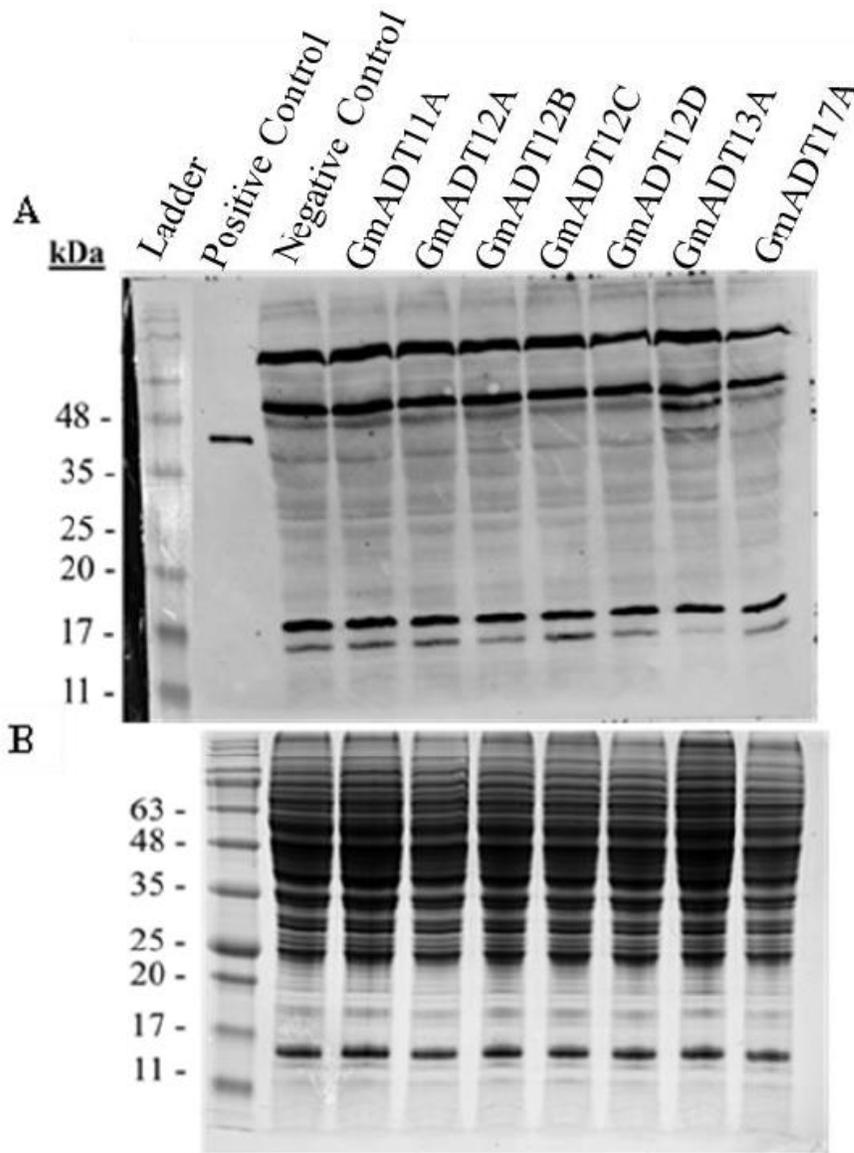


Figure 3.4. Western blot to detect expression of GmADTs in *S. cerevisiae aro8aro9* PAT

A. Total protein extracted from *aro8aro9* transformants grown for 72 hours in 1% raffinose 2% galactose -His -Leu SD dropout media were separated using 13% SDS-PAGE, then transferred to PVDF membrane, and sequentially incubated in 1:1000 mouse anti-his antibody and 1:2000 goat anti-mouse antibody followed by chemiluminescence. No signal is seen for GmADTs. Positive control: Purified GmCHR14-His; negative control: GmADT11B **B.** Samples from the same protein extraction were run in SDS-PAGE and stained with Coomassie Brilliant Blue dye.

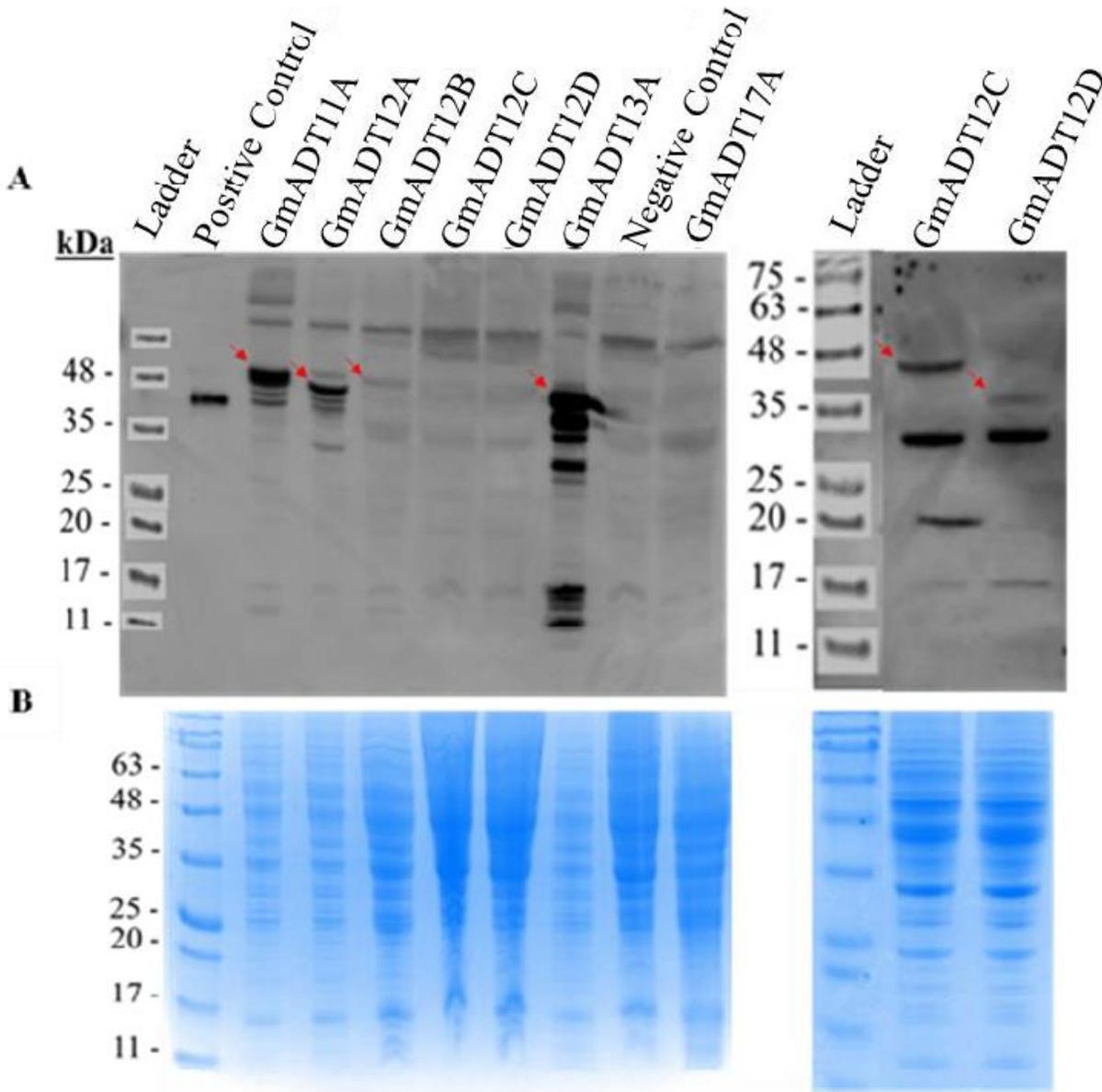


Figure 3.5. Western blot to detect expression of GmADTs in *S. cerevisiae pha2*

A. Total protein extracted from *pha2* transformants grown for 72 hours in 1% raffinose 2% galactose -His SD dropout media were separated using 13% SDS-PAGE, then transferred to PVDF membrane, and sequentially incubated in 1:1000 mouse anti-his antibody and 1:2000 goat anti-mouse antibody followed by chemiluminescence. Red arrows indicate GmADT signal. Positive control: Purified GmCHR14-His; negative control: GmADT11B **B.** Samples from the same protein extraction were run in SDS-PAGE and stained with Coomassie Brilliant Blue dye.

3.5 Identification of *GmADT* Transcript Variants

To identify if any candidate *GmADTs* express transcript variants that lack a sequence encoding a transit peptide, their amino acid sequences were analyzed for presence of an alternate translational start preceding the functional region of the protein. The analysis identified four *GmADTs* (*GmADT11A*, *GmADT12A*, *GmADT12B*, and *GmADT13A*) with a sequence encoding a methionine residue at the end of the predicted transit peptide, just before the catalytic region. This coding sequence topography enables these four *GmADTs* to have an alternate translational start (Figure 3.6).

Further, qRT-PCR was used to assess if these 4 *GmADTs* have alternate transcripts that express a cytosolic version of the protein, some with the transit peptide and others without. Two primer pairs for qRT-PCR were designed for each *GmADT* (Figure 3.6A). Expression of transcripts that encode the transit peptide (only long transcripts) was captured by the first primer pair (L-primers), and expression of transcripts that encode the catalytic region (both long and short transcripts) was captured by the second primer pair (B-primers) (Figure 3.6A). The difference between B-primer and L-primer expression reflects the expression of *GmADT* transcripts without the transit peptide. *GmADTs* have high nucleotide identity, and so most L-primers were designed to amplify the 5' UTR, as it is the most variable region of each transcript. The annotated transcript sequence found on the online database phytozome, however, predicts the 5'-UTR region of each *GmADT* except for *GmADT12B*. Furthermore, the sizes of each predicted 5'-UTR varied greatly, as *GmADT11A*, *GmADT12A*, and *GmADT13A* had 742, 473, and 238 nucleotides, respectively. For this reason, L-primers for *GmADT12B* were designed to amplify the sequence in the transit peptide-encoding region, and those for the other 3 *GmADTs* were designed within 200 nucleotides of the translation start site (Figure 3.6A).

RNA extracted from Soybean Williams 82 flower tissue was used to synthesize qRT-PCR templates for *GmADT11A* and *GmADT12B*, while leaf tissue was used for *GmADT12A* and

GmADT13A. Expression values were normalized against the soybean reference gene *CONS4*. To determine if normalized expression using B-primers was greater than with L-primers, a one-tail two- sample *t*-test assuming unequal variances was used. Normalized expression of B-primers was accepted as greater than that of L-primers if $P < 0.05$. I found that *GmADT11A* ($P = 0.018$) and *GmADT12B* ($P = 0.010$) have alternate transcripts that lack the transit peptide, while *GmADT12A* ($P = 0.107$) and *GmADT13A* ($P = 0.357$) do not (Figure 3.6B).

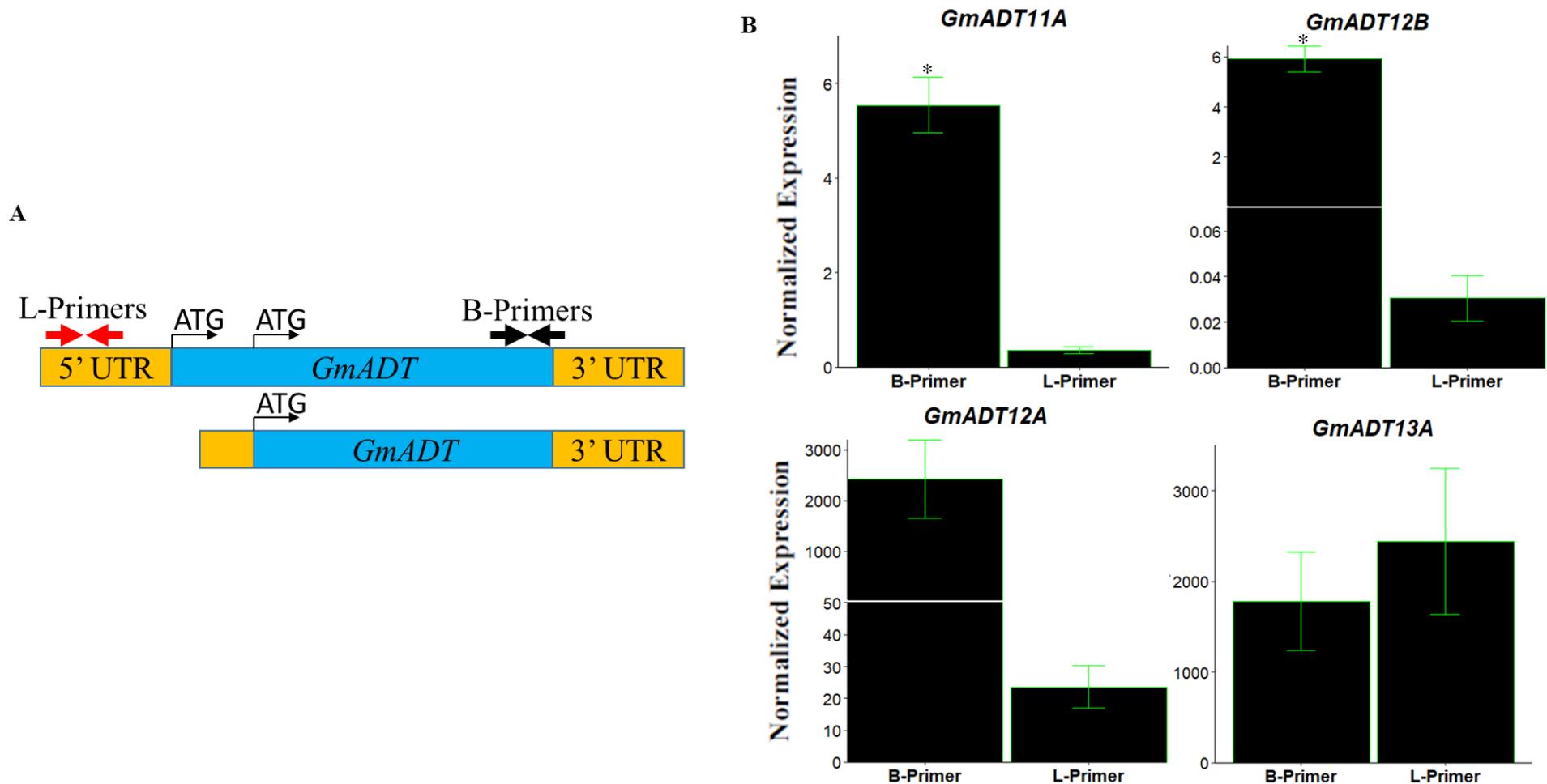


Figure 3.6. RT-qPCR to identify *GmADT* alternate transcripts

A. Schematic diagram of primer design for qRT-PCR. L-primers are intended to amplify only transcripts that encode transit peptides, while B-primers amplify all transcripts that encode *GmADT* functional domains. The upper schematic represents the long transcript that contains transit peptide sequence, and the lower schematic represents the short transcript that excludes transit peptide sequence. **B.** *GmADT* expression normalized to the soybean reference *CONS4*. A one-tailed two-sample t-test assuming unequal variances ($P < 0.05$) was used to determine significance, denoted by the asterisk (*). Error bars indicate SEM of three biological replicates.

Chapter 4: Discussion

Isoflavonoids are a class of legume-specific specialized metabolites (Křížová *et al.*, 2019). They function as phytoalexins, which are a class of molecules toxic to pathogens such as *Phytophthora sojae* (Graham *et al.*, 2007). Generally, phytoalexins are accumulated in response to pathogen attack at the site of the invasion (Ahuja *et al.*, 2012). Isoflavonoids, however, can also be found in apparently healthy seeds, roots, and shoots of legumes (Ingham, 1983). Additionally, isoflavonoids have been shown to stimulate mutualistic root nodule formation by inducing the transcription of nodulation genes in the nitrogen-fixing bacterium *Bradyrhizobium japonicum* (Subramanian *et al.*, 2006). Investigation of isoflavonoid synthesis, therefore, is pertinent to understanding both mutualistic and defensive interactions in legumes. Isoflavonoids, among other specialized metabolites, are products of the phenylpropanoid pathway (Figure 1.1). They are synthesized by a metabolon anchored to the cytosolic side of the endoplasmic reticulum by IFS and C4H. Two GmADTs, GmADT12A and GmADT13A, are found to interact with this metabolon (Dastmalchi *et al.*, 2016). This is noteworthy, as ADTs catalyze the last and rate limiting step in the synthesis of Phe, the precursor to all phenylpropanoids. Additionally, in plants ADTs typically localize to the chloroplast (Jung *et al.*, 1986; Bross *et al.*, 2017), making their presence in the cytosol unexplained. Phenylalanine is also necessary for protein synthesis. The demand for Phe flux into the phenylpropanoid pathway is increased in response to stress. Metabolic channeling of Phe must therefore be dynamically regulated to ensure that flux is sufficient for both Phe sinks. Apart from the feedback regulation of pathway enzymes by their products, the mechanisms of Phe metabolic channeling are largely unknown. I postulate that GmADT interaction with the isoflavonoid metabolon in the cytosol is an element of regulation that directs Phe flux to phenylpropanoid synthesis.

While ADT-driven synthesis is the primary mode of Phe production in plants (Maeda and Dudareva, 2012), microbes use an alternate PDT-driven route (Maeda and Dudareva, 2012). Bi-

functional ADTs with PDT activity have been discovered in *A. thaliana* (Cho et al., 2007), *Petunia hybrida* (Maeda et al., 2010), and *Pinus pinaster* (El-Azaz et al., 2016). Although the plastidial ADT pathway is the chief driver of Phe synthesis in plants, there is evidence supporting the existence of an additional cytosolic microbial-like PDT pathway as well (Yoo et al., 2013; Qian et al., 2019). To complete this pathway, it was shown that the *PhADT3* gene in petunia, which encodes a bi-functional ADT, expresses a transcript variant encoding a cytosolic PhADT3 isoform. Normally, PhADT3 localizes to the chloroplast, but the cytosolic isoform does not have a transit peptide and as a result accumulates in the cytosol. The observation that truncated PhADT3 can accumulate in the cytosol supports cytosolic ADTs as regulatory elements in the channeling of Phe directly to phenylpropanoid synthesis and is further supported by the observation that expression of alternate transcripts is often in response to stress (Ayoubi and VanDeVen, 1996).

Because GmADTs have been found to interact with the isoflavonoid metabolon, the existence of a cytosolic bi-functional GmADT isoform would help to solidify the role of ADTs in the stress-induced upregulation of phenylpropanoids. Additionally, these results could help to distinguish the roles of ADT family members from one another. As such, this study seeks to characterize candidate GmADTs for their ADT and PDT activities and determine if any *GmADTs* express transcript variants that encode cytosolic GmADT isoforms.

4.1 GmADTs Behave like Characterized ADTs in Other Species

ADTs have three distinct domains, the N-terminal transit peptide, followed by the catalytic region, and then the ACT domain. The catalytic region and ACT domains are highly conserved across all ADTs, while the transit peptide is more variable. Furthermore, ADTs are organized into three

different subgroups. Predictions can be made about uncharacterized ADTs based on experimental data from another ADT in the same subgroup.

Of 11 total candidate *GmADT* genes, 2 were truncated, and 1 could not be amplified from cDNA. The first truncated gene was *GmADT13B*, missing a large portion of the sequence that would encode the catalytic and ACT domains, and as a result, is likely non-functional. Reciprocally, *GmADT19A*, the second truncated gene, encodes only the ACT domain (Appendix B). It is possible that a large-scale genomic rearrangement may have divided a once complete *GmADT* into these two truncated pseudo-genes. If indeed this was the case, the aforementioned ‘complete *GmADT*’ composed of *GmADT13B* and *GmADT19A* fused together likely did not produce a protein that increased the fitness of soybean’s ancestor, as the gene for its production was not maintained in the population. It may have been lost randomly, or perhaps it was evolutionarily disadvantageous and selected against. Nevertheless, both truncated genes are likely non-functional, and for this reason, were eliminated from the list of candidate *GmADTs*. *GmADT13B* and *GmADT12C* are a high identity pair, and both are in subgroup 1 with *GmADT17A* and *GmADT9A*. *GmADT17A* and *GmADT9A* have identities of no higher than 70% with any *GmADT*, including each other. *GmADT17A* and *GmADT12C* were successfully amplified from cDNA and cloned, though all attempts to clone *GmADT9A* were unavailing. Interestingly, the high identity pair, *GmADT12C* and *GmADT13B*, are both mostly expressed in flowers, while the other two are in roots (Table 1.1). Furthermore, the genes that were successfully amplified from cDNA, *GmADT12C* and *GmADT17A*, were expressed at higher levels than their subgroup 1 counterparts. These observations further suggest that duplicates of subgroup 1 *GmADTs* are selected against, making them either non-functional or divergent. Although *GmADT9A* amplification was unsuccessful, the annotated reference gene on [phytozome](#) appears to encode a full-length ADT. It is possible that *GmADT9A* is only expressed in response to specific external stimuli. *GmADT17A* expression is highest in root nodules during symbiosis with rhizobia,

suggesting it may have an important role in the process. All characterized subgroup 1 ADTs localize to the chloroplast.

GmADT12D and *GmADT11B* are both members of subgroup 2. These *GmADTs* share 95.93% identity. The tissue of highest expression for both is seed, with FPKM values of 12.2 and 29.7, respectively. Additionally, *GmADT12D* and *GmADT11B* are in the same subgroup as *AtADT2*, which has been thoroughly characterized. *AtADT2* is a bi-function ADT/PDT, and it is essential for seed development. It is likely that soybean's subgroup 2 ADTs also have these function, or that they are split among the two. All characterized subgroup 2 ADTs localize to the chloroplast.

The remaining *GmADTs*, *GmADT11A*, *GmADT12A*, *GmADT13A*, and *GmADT12B*, are in subgroup 3. *GmADT11A* and *GmADT12B* are a high identity (94.52%) pair, and so are *GmADT12A* and *GmADT13A* (95.06%). Interestingly, *A. thaliana* also has 4 ADTs in subgroup 3, *AtADT4* and *AtADT5* (92.2% identity), and *AtADT3* and *AtADT6* (97.7% identity) (Cho *et al.*, 2007). There is evidence that the *A. thaliana* genome has also undergone multiple duplication events (Simillion *et al.*, 2002). Therefore, it is possible that copies of subgroup 3 *AtADTs* were a result of those duplications. If this were the case, why are there four subgroup 3 *AtADTs*, but not the others?

Some subgroup 1 ADTs in soybean are likely non-functional pseudo-genes or have unknown function, suggesting that having more than a single subgroup 1 ADT doesnot improve plant fitness. There is evidence for a high frequency of gene loss post genomeduplication in *Arabidopsis*, thus copies of *ADT1* were likely among those (Simillion *et al.*, 2002).

Soybean has 2 subgroup 2 ADTs, while *A. thaliana* has only 1. Both subgroup 2 ADTs in soybean have sequences encoding full-length proteins, are expressed, were successfully amplified from DNA, and share high identity to one another. This suggests they are both functional to some extent, as unlike soybean subgroup 1 ADTs, sequences for both were conserved over evolutionarytime. However, if this were the case, we would expect to see 4 subgroup 2 ADTs in soybean rather than 2, as soybean has undergone at least two full genome duplications. It is

possible, then, that a single subgroup 2 ADT was first added to the soybean genome at a time between the two duplication events. I postulate that subgroup 2 ADT genes originated in plastids and were later integrated into the plant nuclear genome (Shahmuradov *et al.*, 2003; Ponce-Toledo *et al.*, 2019). In the case of *A. thaliana*, this integration would need to have happened after all genome duplication events, as AtADT2 is its only subgroup 2 ADT. Soybean has two subgroup 2 ADTs, so integration would likely have happened after the first duplication, but before the second. This conjecture works harmoniously with the assumption that *A. thaliana*'s genome duplication events happened longer ago than soybean's.

All full-length subgroup 3 ADTs have been found to localize to the chloroplast, although *PhADT3*, a subgroup 3 ADT, was shown to express a transcript variant that encodes a cytosolic isoform of the protein. Three petunia ADTs were identified from petunia petal-specific EST databases, *PhADT1*, *PhADT2*, and *PhADT3* (Maeda *et al.*, 2010). All three transcripts have sequences encoding an alternate translation start site that would exclude the transit peptide and produce a cytosolic ADT isoform. Using qPCR and 5'-RACE, it was determined that *PhADT3* does produce a cytosolic isoform, while the other two do not. The petunia genome has not been fully sequenced, and no other evidence provides a concrete description of the genomic context from which the *PhADT1*, *PhADT2*, and *PhADT3* transcripts arise. In soybean, only subgroup 3 ADTs have the potential to encode a cytosolic isoform, as none of the other *GmADTs* have a translational start close enough upstream of the catalytic region. Thus, presence of *GmADT11A*, *GmADT12A*, *GmADT12B*, and *GmADT13A* alternate transcripts was explored. Furthermore, as

PhADT3 is believed to function in the cytosol through PDT activity, ADT/PDT activity of all GmADTs was assayed.

4.2 Amino Acid Substitutions in GmADT Clones May Affect Function

For some proteins, a single amino acid substitution is sufficient to impair, or even negate their function. In this study, *GmADT12A*, *GmADT12C* and *GmADT12D* clones had nucleotide polymorphisms that cause amino acid substitutions in their encoded protein (Table 3.1). Most of these, however, occur in the transit peptide region, which would have been removed in the mature protein. It is, therefore, unlikely that amino acid substitutions there will have any effect on catalytic activity. Only GmADT12D has an amino acid substitution (I269M) in a functional domain. Because both isoleucine and methionine have hydrophobic side chains, this is a conservative replacement, meaning it is less likely to alter protein function. It is still however, possible that this amino acid change renders GmADT12D non-functional. This may be why yeasts transformed with *GmADT12D* in the ADT and PDT assay did not show any enzymatic activity (Figure 3.2; Figure 3.3).

4.3 Yeast Complementation Assay – Characterizing ADT Activity of Candidate GmADTs

Normally, enzyme kinetics are determined through *in vitro* assays, where the efficiency of substrate binding and subsequent conversion to product can be measured. *In vitro* assays for ADT activity are difficult to perform, as aroenate is an extremely unstable substrate and it spontaneously converts to phenylalanine *in vitro*. Therefore, it cannot be purchased from a vendor. As yeast complementation assays have been used to characterize PDT activity, a similar assay may also be used to characterize ADT activity. To accomplish this, the mutant yeast *aro8aro9* was developed in Dr. Susanne Kohalmi's laboratory. *aro8aro9* is a double PPY-AT knockout. The

purpose of removing PPY-AT activity from yeast was to prevent synthesis of Phe through the PDT pathway. The double knockout mutant was made by mating *aro8* and *aro9* single knockout mutants, then screening for double knockouts (Wach *et al.*, 1994; Winzeler *et al.*, 1999; Giaever *et al.*, 2002). Yeast cells were screened by growing them on plates without Phe, as those missing both PPY-AT would grow much more slowly. Furthermore, *AtPAT*, a PPA-AT gene, was transformed into *aro8aro9* to synthesize aroenate from prephenate.

For the ADT and PDT assay, it is assumed that transformants that grew larger colonies have more ADT or PDT activity (Figures 3.2 and 3.3). This is because the rate of colony growth decreases with reduced enzymatic activity.

Results generated from candidate GmADT transformants in the ADT assay were unexpected (Figure 3.2). The vector pAG423GAL-GmADT contains *HIS3*, which confers *aro8aro9* the ability to grow on media lacking His, while the vector pAG425GAL-*AtPAT* contains *LEU2*, allowing for growth on media without Leu (Figure 2.3A). To detect the presence of each vector in the transformed yeast, yeast cells were spotted on media lacking His, Leu, and Phe and their combinations. As expected, only yeast carrying pAG423GAL grew on -His plates, only those with pAG425GAL grew on -Leu plates, and those with both grew on any combination of -Leu and -His. Results on -Phe plates, however, were more complicated (Figure 3.2). Both *GmADT* and *AtPAT* expression were controlled by GAL1 promoters, so it is expected that on Gal medium they will be expressed, and if the GmADT has ADT activity, *aro8aro9* mutant phenotype will be rescued and the cells were expected to grow on -Phe plates. Glucose represses the GAL1 promoter, so yeast on Glu media is expected to fail to complement and as a result grow more slowly (West *et al.*, 1984; Giniger *et al.*, 1985). In my ADT assay, most *aro8aro9* PAT GmADT transformants grew more quickly on Glu media than on Gal media (Figure 3.2). This may be in part due to yeast's preference for glucose as a sugar source compared to galactose (Lagunas, 1993). This preference is apparent on the -Leu/-His plates, where colonies on Glu media grew larger than they did on Gal. Alternatively, if *S. cerevisiae*'s endogenous PDT has ADT activity,

even a small amount of AtPAT leakage on Glu media may be sufficient to allow the synthesis of enough Phe to grow. ADT activity, after all, is the rate-limiting step in the arogenate pathway, with PPA-AT activity being at least three orders of magnitude higher (Maeda *et al.*, 2010; Maeda *et al.*, 2011). Results, however, are incongruous with this explanation, as vector-only controls, which cannot express any PAT, also grew on Glu -Phe media. Additionally, there is experimental evidence that PHA2 is a monofunctional PDT, so no growth can be conferred by PHA2 converting arogenate to phenylalanine (El-Azaz *et al.*, 2018). *aro8aro9* must therefore still be able to synthesize a small amount of phenylalanine on its own.

Aminotransferase enzymes, while highly stereoselective, can usually catalyze transfer of amino groups between several amino acid/keto acid pairs (Bommer and Ward, 2013). For example, experimental evidence shows that histidinol-phosphate aminotransferase (HIS5) can use phenylalanine as an amino donor and α -ketoglutarate as the acceptor to produce phenylpyruvate and L-glutamate with a specific activity of about 0.12 $\mu\text{mol}/\text{min}/\text{mg}$ enzyme (Bommer and Ward, 2013). Furthermore, aminotransferase reactions are freely reversible, so HIS5 must also be able to lower the energy barrier for the opposite reaction. *In vitro* enzyme assays show that PhPAT transaminates prephenate to arogenate, and indeed can also catalyze the reverse reaction with an approximate 10-fold lower efficiency (Maeda *et al.*, 2011). It cannot, however, transaminate phenylpyruvate to Phe, which is ideal for the assay to work as intended. Although the ADT assay yeast mutant is a knockout for ARO8 and ARO9, presence of HIS5 may still allow for phenylalanine synthesis through the PDT pathway. If this were the case, it could explain the basal level of growth for vector only controls on -Phe plates (Figure 3.2). If HIS5 does convert phenylpyruvate to phenylalanine, it is possible that GmADTs transformed into *aro8aro9* could be synthesizing phenylalanine through PDT activity. However, there are no characterized plant ADTs with only PDT activity, or PDT activity that is stronger than their ADT activity. Furthermore, AtPAT is likely more efficient at transaminating prephenate to arogenate, than HIS5 is at transaminating phenylpyruvate to phenylalanine. Therefore, it is most probable that

any growth seen in the ADT assay on top of the basal growth from phenylalanine produced in *aro8aro9* can be attributed to ADT activity from transformed GmADTs.

From the ADT assay, it can be concluded that GmADT11B has ADT activity, as may also, to a lesser degree, GmADT12B and GmADT12C (Figure 3.2). AtADT3 was used as a positive control, as it is the most efficient monofunctional ADT ($K_{cat}/K_m=1140 \text{ M}^{-1}\text{s}^{-1}$) in *A. thaliana* (Cho et al., 2007). GmADT11B however is in subgroup 2, the same clade in the phylogenetic tree in which AtADT2 is found (Pannunzio, 2018). AtADT2 has the most efficient ADT activity of all *A. thaliana* ADTs ($K_{cat}/K_m=7650 \text{ M}^{-1}\text{s}^{-1}$), which is reflected here in the growth of GmADT11B-complemented *aro8aro9* yeast. No growth was seen for the other subgroup 2 ADT in soybean, GmADT12D. This is surprising, as subgroup 2 ADTs usually have higher activity than other subgroups, (Cho et al., 2007; Maeda et al., 2010; El-Azaz et al., 2016), and GmADT11B and GmADT12D have 97.27% identity at the amino acid level (Appendix. A). GmADT12D is missing the end of the ACT domain that is present in GmADT11B and other functional ADTs (Appendix B). Taken together, these results suggest that GmADT12D is truncated, and as a result may not function as an ADT. GmADT12B is in subgroup 3, therefore, of the ADTs that grew on triple knockout plates, it is most like AtADT3. This stands to reason, as growth patterns of the two were similar in the assay. This suggests GmADT12B likely does have ADT activity, and it is the highest level of ADT activity from a subgroup 3 ADT in soybean. As for the other subgroup 3 ADTs, GmADT11A, GmADT12A, and GmADT13A, although no growth is seen beyond basal level growth of *aro8aro9*, it is possible that they also have ADT activity, but that the ADT assay is not sensitive enough to detect it. This conclusion is supported by the fact that, all *A. thaliana* subgroup 3 ADTs are functional to some degree (Cho et al., 2007). AtADT3 and AtADT6 have 97.7% identity and have similar amounts of ADT activity (Cho et al., 2007). If GmADT12B is an orthologue of AtADT3, we would expect to see growth of *aro8aro9* complemented with GmADT11A (which has 96.63% identity with GmADT12B and is the likely orthologue of AtADT6). This is not, however, the case. It is possible that because there

are four subgroup 3 ADTs in both species, different ones evolved to fill different niches (Zhao *et al.*, 2018). Finally, GmADT12C is in subgroup 1 along with GmADT17A. AtADT1 is the only subgroup 1 ADT in

A. thaliana, and it has a similar propensity to catalyze the ADT reaction as AtADT3. As GmADT12C transformants grew colonies of similar size to those of AtADT3, it likely does have ADT activity and is a paralogue of AtADT1. No growth is seen for GmADT17A, the other subgroup 1 ADT. As mentioned earlier, genomic data suggests organisms tend to have only a single functional subgroup 1 ADT. From my results, it appears GmADT17A is likely non-functional or has diverged in function.

4.4 Yeast Complementation Assay – Characterizing PDT Activity in Candidate GmADTs

GmADTs were characterized for their PDT activity through another yeast complementation assay. Here, *pha2*, a PDT knockout was transformed by pAG423GAL plasmids carrying GmADTs and transformants were spotted on yeast growth media without Phe.

The GmADT transformant that had the most growth was GmADT11B. GmADT11B was the only GmADT not fused to 6 histidine residues at the C-terminal end of the protein, thus it could be interpreted that GmADTs containing the 6 histidines had their enzymatic activity impaired. This is, however, unlikely, as AtADT2-6xHis and AtADT1-6xHis fusions in the PDT assay were shown to have PDT activity (Bross *et al.*, 2011). It is unknown why subgroup 2 ADTs have PDT activity, as they are all localized to the chloroplast. It has been shown that PHA2, a monofunctional PDT, targeted to the chloroplast, and under the control of the AtADT2 promoter, is able to rescue the seed development arrest phenotype in *A. thaliana* ADT2 knockouts (El-Azaz *et al.*, 2018). It was, however, shown that AtADT3, a monofunctional ADT, targeted to the chloroplast, and under the control of the AtADT2 promoter, also rescues this phenotype. This suggests plastidial phenylalanine synthesis can occur via either the ADT or PDT pathway. GmADT11B, then, can support both modes of phenylalanine synthesis. GmADT12D

transformants did not grow, and therefore do not likely have PDT activity. As all characterized subgroup 2 GmADTs have PDT activity, this further supports the argument that GmADT12D is non-functional.

Small colonies observed for the subgroup 3 GmADT transformants (GmADT13A and GmADT12A) suggest they may have PDT activity. In *Arabidopsis* only a single subgroup 3 ADT, AtADT6, was shown to have PDT activity. If GmADT13A and GmADT12A do indeed have PDT activity, it is minimal compared to GmADT11B and AtADT2. This is consistent with the literature, as AtADT6 has about one fifteenth of AtADT2's efficiency at catalyzing the PDT reaction. GmADT13A and GmADT12A are the same two GmADTs that were found to interact with GmIFS in the cytosolic side of the ER lumen (Dastmalchi *et al.*, 2016). Furthermore, it is predicted that cytosolic Phe synthesis is PDT-driven, and both GmADTs have PDT activity. Therefore, if these GmADTs have a cytosolic isoform, they would be able to complete the PDT pathway in soybean. No growth was seen from either subgroup 1 GmADTs, GmADT12C and GmADT17A, indicating that neither one of them has PDT activity. This contrasts *A. thaliana*'s subgroup 1 ADT, AtADT1, as it has more than twice as much PDT activity than AtADT6 (Cho *et al.*, 2007). GmADT12C and GmADT17A express most highly in root tissue. It is, however, likely that GmADT17A is non-functional, as it has neither PDT nor ADT activity. Additionally, most species have only a single functional subgroup 1 ADT, and soybean's other subgroup 1 ADTs have already diverged or are non-functional.

4.5 Yeast Complementation Assay – Western Blot

To confirm that the lack of growth seen for *aro8aro9* and *pha2* yeast cells on ADT and PDT assay Gal -Phe media was not due to missing recombinant protein expression, presence of GmADT proteins in yeast transformants was verified using Western blot analysis. Results for the PDT assay confirmed the presence of all GmADTs except GmADT17A (Figure 3.5). In contrast, none of the candidate GmADTs were detected in yeast cells used for the ADT assay (Figure 3.4).

As the positive control worked for all western blots, missing signal is not caused by the experimental procedures. Yeast colony PCR results showed that all constructs were present in the transformed yeasts (Figure 3.1). Furthermore, there is evidence that GmADT12B and GmADT12C are expressed, as growth above basal levels was seen for their yeasts in the ADT assay (Figure 3.2). They both, however, failed to be detected in Western blot. As such, it is possible that other GmADTs also were expressed but failed to be detected in Western blot. It is unlikely that non-specific binding of the primary antibody is interfering with visualization of GmADT proteins expressed in ADT assay yeast cells, as non-specific binding to this degree is not seen for yeast cells from the PDT assay. ADT and PDT assay yeast cells were grown under similar conditions, with the distinction that *pha2* transformants were induced in media lacking His, while *aro8aro9* transformants were induced in media lacking His and Leu. It may be notable that *aro8aro9* cells always grew slower and had smaller and fewer cells than their *pha2* counterparts.

4.6 Identification of GmADT Alternate Transcripts

It has been shown that *PhADT3*, a subgroup 3 ADT, can be expressed as an alternate transcript encoding a cytosolic isoform of the protein (Qian *et al.*, 2019). Expression of this cytosolic isoform may be implicated in the regulation of Phe flux towards phenylpropanoid synthesis. Subgroup 3 *GmADTs* (*GmADT11A*, *GmADT12A*, *GmADT12B*, and *GmADT13A*) all have potential alternate translation start sites just before the sequence encoding the ADT catalytic region. Therefore, they were all candidates to express alternate transcripts that may encode cytosolic ADT isoforms in soybean. qRT-PCR analysis confirmed that *GmADT11A* and *GmADT12B* have alternate transcripts that exclude the transit peptide, while *GmADT12A* and *GmADT13A* do not (Figure 3.6B). This is surprising as *GmADT12A* and *GmADT13A* were found to interact with GmIFS in the cytosol (Dastmalchi *et al.*, 2016) and also showed PDT activity by rescuing the *pha2* mutant phenotype (Figure 3.3). Transcript abundance was measured using cDNA synthesized from soybean tissue under normal conditions. As alternate transcript expression is often increased as a response to external stresses like disease (Ayoubi and VanDeVen, 1996), it is possible that alternate transcripts for *GmADT12A* and *GmADT13A* may be detected in cDNA synthesized from soybean tissue under stressed conditions. PDT activity is essential to cytosolic Phe synthesis, as such *GmADT12A* and *GmADT13A* would be the ideal candidates to express cytosolic isoforms. On the other hand, it is also possible that *GmADT11A* and *GmADT12B* have trace amounts of PDT activity that the PDT assay used in my study is not sensitive enough to detect. Small amounts of PDT activity may be sufficient for these ADTs to function in the IFS metabolon, as the organization of enzymes allows for products of one reaction to be efficiently fluxed in as the substrates of the next. For *GmADT12A*, the qPCR data trends in a direction which suggests that it may also express an alternate transcript, and the only reason a significant difference was not seen, is variation of expression between biological replicates was too large. As more replicates are added to the data, variation may decrease. Thus, it is likely that *GmADT12A* expresses a cytosolic isoform that increases Phe flux to

phenylpropanoid synthesis.

4.7 Future Direction: The Next Steps

GmADTs must still be characterized for their ADT activity. Results from the ADT assay were a good start, but they are inconclusive, as GmADT11A, GmADT12A, and GmADT13A may have ADT activity that was not detected. Furthermore, no expression of GmADTs was seen in Western Blot analysis. Additional Western blot analysis should be explored to ensure expression of GmADTs. Moving forward, GmADT11B should be fused to 6x histidine residue to be assayed in Western Blot, as GmADT11B did show ADT and PDT activity, so it must have been expressed. This will help to elucidate if the other GmADT proteins that were not seen in Western Blot are truly not expressed, or if there is another explanation for why no signal was seen. Sensitivity of the ADT assay may be increased by making an *aro8aro9his5* knockout. Perhaps this triple knockout would not be able to synthesize Phe endogenously, and ADT activity in transformants would be clearer. Alternatively, GmADTs can be characterized through *in vitro* enzyme assays. Despite the laborious nature of *in vitro* ADT enzyme assays, they have high precision and sensitivity.

The presence of alternate *GmADT* transcripts should be verified using 5'-RACE and sequence confirmation. If transcripts encoding cytosolic isoforms of ADTs with PDT activity are substantiated, a plethora of new studies can be addressed. It should be determined what external stimuli trigger the expression of these alternate transcripts. Whether or not enzymes upstream of ADT, like PPY-AT and CM, interact with ADT as members of the GmIFS metabolon. As shikimate pathway enzymes are plastidial, at what point(s) is something transported out of the chloroplast to allow for cytosolic Phe synthesis, and what transporter proteins are involved? Does increased phenylpropanoid synthesis improve plant health, and what trade-offs are there?

Chapter 5: Conclusion

The constituents are present for a stress-activated cytosolic PDT pathway in plants (Qian *et al.*, 2019), possibly driving Phe flux towards synthesis of phenylpropanoids. Soybean, as a legume, synthesizes unique defensive phenylpropanoid compounds, making the characterization of this pathway in soybean more pertinent. To this extent, characterization of functional and bi-functional GmADTs was necessary to know which enzymes can participate in Phe synthesis in the cytosol. Furthermore, determining which GmADTs have a cytosolic isoform is necessary, as all characterized full-length ADTs localize to the chloroplast.

Results reported here demonstrate that GmADTs function similarly to other ADTs found in the same subgroup. Candidate GmADTs in subgroup 1 have largely diverged or become non-functional, allowing GmADT12C to be the sole functioning subgroup 1 ADT in soybean. Furthermore, both subgroup 1 ADTs are unlikely to have cytosolic isoforms. GmADT11B has strong ADT and a PDT activity, while GmADT12D appears to be non-functional, making GmADT11B soybean's only functional subgroup 2 ADT. These also are unlikely to function in the cytosol. The remaining 4 GmADTs are in subgroup 3. The only GmADT that showed some ADT activity in the yeast assay was GmADT12B. Additionally, it was learned that GmADT12A and GmADT13A demonstrate small amounts of PDT activity. Furthermore, Subgroup 3 ADTs are most likely to have cytosolic isoforms, as they have sequences that allow for alternate translational start sites that encode functional proteins without transit peptides.

It was determined that *GmADT11A* and *GmADT12B* express alternate transcripts that encode cytosolic ADT isoforms. These proteins, however, were not found to have PDT activity, and for this reason should not be able to function in the cytosol. Although *GmADT13A* does not express a cytosolic transcript variant, the data suggests *GmADT12A* might. A cytosolic isoform of GmADT12A would be able to complete the cytosolic PDT pathway in soybean and interact with the GmIFS metabolon to flux Phe directly to isoflavonoid synthesis. As more mechanisms

controlling plant response to external stimuli are discovered, we acquire a more profound understanding of how complex beneficial traits arise. Therefore, the eventual application of this knowledge to the engineering of better, more sustainable crops, is inevitable.

References

- Abolhassani Rad S.** 2017. The Mystery of Nuclear Localization of Arogenate Dehydratases from *Arabidopsis thaliana*. PhD. *Electronic Thesis and Dissertation Repository*. 4941. <https://ir.lib.uwo.ca/etd/4941>
- Ahuja I, Kissen R, Bones AM.** 2012. Phytoalexins in Defense Against Pathogens. *Trends in Plant Science* **17**, 73-90.
- Aravind L, Koonin EV.** 1999. Gleaning Non-Trivial Structural, Functional and Evolutionary Information about Proteins by Iterative Database Searches. *Journal of Molecular Biology* **287**, 1023-1040.
- Ayoubi TAY, VanDeVen WJM.** 1996. Regulation of Gene Expression by Alternative Promoters. *Faseb Journal* **10**, 453-460.
- Barros J, Dixon, R.A.** 2020. Plant Phenylalanine/Tyrosine Ammonia-lyases. *Trends in Plant Science* **25**, 66-79.
- Bommer M, Ward JM.** 2013. A 1-step Microplate Method for Assessing the Substrate Range of L-Alpha-Amino Acid Aminotransferase. *Enzyme and Microbial Technology* **52**, 218-225.
- Boydston R, Paxton JD, Koeppe DE.** 1983. Glyceollin: A Site-Specific Inhibitor of Electron Transport in Isolated Soybean Mitochondria. *Plant Physiology* **72**, 151-155.
- Bross CD, Corea ORA, Kaldis A, Menassa R, Bernardis MA, Kohalmi SE.** 2011. Complementation of the *pha2* Yeast Mutant Suggests Functional Differences for Arogenate Dehydratases from *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* **49**, 882-890.
- Bross CD, Howes TR, Rad SA, Kljakic O, Kohalmi SE.** 2017. Subcellular Localization of *Arabidopsis* Arogenate Dehydratases Suggests Novel and Non-Enzymatic Roles. *Journal of Experimental Botany* **68**, 1425-1440.
- Bubna GA, Lima RB, Zanardo DY, dos Santos WD, Ferrarese MdLL, Ferrarese-Filho O.** 2011. Exogenous Caffeic Acid Inhibits the Growth and Enhances the Lignification of the Roots of Soybean (*Glycine max*). *Journal of Plant Physiology* **168**, 1627-1633.
- Cheynier V, Comte G, Davies KM, Lattanzio V, Martens S.** 2013. Plant Phenolics: Recent Advances on Their Biosynthesis, Genetics, and Ecophysiology. *Plant Physiology and Biochemistry* **72**, 1-20.
- Cho MH, Corea ORA, Yang H, Bedgar DL, Laskar DD, Anterola AM, Moog-Anterola FA, Hood RL, Kohalmi SE, Bernardis MA, Kang C, Davin LB, Lewis NG.** 2007. Phenylalanine Biosynthesis in *Arabidopsis thaliana*: Identification and Characterization of Arogenate Dehydratases. *Journal of Biological Chemistry* **282**, 30827-30835.

Dakora FD, Phillips DA. 1996. Diverse Functions of Isoflavonoids in Legumes Transcend Anti-Microbial Definitions of Phytoalexins. *Physiological and Molecular Plant Pathology* **49**, 1-20.

Dastmalchi M, Bernardis MA, Dhaubhadel S. 2016. Twin Anchors of the Soybean Isoflavonoid Metabolon: Evidence for Tethering of the Complex to the Endoplasmic Reticulum by IFS and C4H. *Plant Journal* **85**, 689-706.

Dastmalchi M, Dhaubhadel S. 2015. Proteomic Insights Into Synthesis of Isoflavonoids in Soybean Seeds. *Proteomics* **15**, 1646-1657.

Dixon RA. 2004. Phytoestrogens. *Annual Review of Plant Biology* **55**, 225-261.

Dudareva N, Klempien A, Muhlemann JK, Kaplan I. 2013. Biosynthesis, Function and Metabolic Engineering of Plant Volatile Organic Compounds. *New Phytologist* **198**, 16-32.

El-Azaz J, Canovas FM, Avila C, de la Torre F. 2018. The Arogenate Dehydratase ADT2 is Essential for Seed Development in *Arabidopsis*. *Plant and Cell Physiology* **59**, 2409-2420.

El-Azaz J, de la Torre F, Ávila C, Cánovas FM. 2016. Identification of a Small Protein Domain Present in All Plant Lineages that Confers High Prephenate Dehydratase Activity. *The Plant Journal: for Cell and Molecular Biology* **87**, 215-229.

Fischer R, Jensen R. 1987. Arogenate Dehydratase. *Methods in Enzymology* **142**, 495-502.

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo CY, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang YH, Yen G, Youngman E, Yu KX, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M. 2002. Functional Profiling of the *Saccharomyces cerevisiae* Genome. *Nature* **418**, 387-391.

Giniger E, Varnum SM, Ptashne M. 1985. Specific DNA-Binding of GAL4, a Positive Regulatory Protein of Yeast. *Cell* **40**, 767-774.

Goers SK, Jensen RA. 1984. The Differential Allosteric Regulation of 2 Chorismate Mutase Isoenzymes of *Nicotiana glauca*. *Planta* **162**, 117-124.

Graham TL, Graham MY, Subramanian S, Yu O. 2007. RNAi Silencing of Genes for Elicitation or Biosynthesis of 5-Deoxyisoflavonoids Suppresses Race-Specific Resistance and Hypersensitive Cell Death in *Pytophthora sojae* Infected Tissues. *Plant Physiology* **144**, 728-740.

- Ingham JL.** 1983. Naturally Occurring Isoflavonoids. *Fortschritte der Chemie Organischer Naturstoffe* **Vol. 43**, 1-266.
- Jung E, Zamir LO, Jensen RA.** 1986. Chloroplasts of Higher Plants Synthesize L-Phenylalanine via L-Arogenate. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 7231-7235.
- Kopittke PM, Menzies NW, Wang P, McKenna BA, Lombi E.** 2019. Soil and the Intensification of Agriculture for Global Food Security. *Environment International* **132**, 8.
- Křížová L, Dadáková K, Kašparovská J, Kašparovský T.** 2019. Isoflavones. *Molecules* **24**, 1076.
- Lagunas R.** 1993. Sugar Transport in *Saccharomyces cerevisiae*. *FEMS Microbiology Letters* **104**, 229-242.
- Maeda H, Dudareva N.** 2012. The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. *Annual Review of Plant Biology* **63**, 73-105.
- Maeda H, Shasany AK, Schnepf J, Orlova I, Taguchi G, Cooper BR, Rhodes D, Pichersky E, Dudareva N.** 2010. RNAi Suppression of Arogenate Dehydratase 1 Reveals that Phenylalanine is Synthesized Predominantly via the Arogenate Pathway in Petunia Petals. *Plant Cell* **22**, 832-849.
- Maeda H, Yoo HJ, Dudareva N.** 2011. Prephenate Aminotransferase Directs Plant Phenylalanine Biosynthesis via Arogenate. *Nature Chemical Biology* **7**, 19-21.
- Mobley EM, Kunkel BN, Keith B.** 1999. Identification, Characterization and Comparative Analysis of a Novel Chorismate Mutase Gene in *Arabidopsis thaliana*. *Gene* **240**, 115-123.
- Neve EPA, Ingelman-Sundberg M.** 2008. Intracellular Transport and Localization of Microsomal Cytochrome P450. *Analytical and Bioanalytical Chemistry* **392**, 1075-1084.
- Pannunzio K.** 2018. Uncovering a Mystery of the Isoflavonoid Metabolon: Identification and Characterization of the Arogenate Dehydratase Gene Family in Soybean. MSc. *Electronic Thesis and Dissertation Repository*. 5994. <https://ir.lib.uwo.ca/etd/5994>
- Patel N, Pierson DL, Jensen RA.** 1977. Dual Enzymatic Routes to L-Tyrosine and L-Phenylalanine via Pretyrosine in *Pseudomonas aeruginosa*. *Journal of Biological Chemistry* **252**, 5839-5846.
- Ponce-Toledo RI, Lopez-Garcia P, Moreira D.** 2019. Horizontal and Endosymbiotic Gene Transfer in Early Plastid Evolution. *New Phytologist* **224**, 618-624.
- Qian Y, Lynch JH, Guo L, Rhodes D, Morgan JA, Dudareva N.** 2019. Completion of the Cytosolic Post-Chorismate Phenylalanine Biosynthetic Pathway in Plants. *Nature Communications* **10**, 15.

Ralston L, Subramanian S, Matsuno M, Yu O. 2005. Partial Reconstruction of Flavonoid and Isoflavonoid Biosynthesis in Yeast using Soybean Type I and Type II Chalcone Isomerases. *Plant Physiology* **137**, 1375-1388.

Ralston L, Yu, O. 2006. Metabolons Involving Plant Cytochrome P450s. *Phytochemistry Reviews* **5**, 459-472.

Shahmuradov IA, Akbarova YY, Solovyev VV, Aliyev JA. 2003. Abundance of Plastid DNA Insertions in Nuclear Genomes of Rice and *Arabidopsis*. *Plant Molecular Biology* **52**, 923-934.

Siehl DL, Conn EE. 1988. Kinetic and Regulatory Properties of Arogenate Dehydratase in Seedlings of *Sorghum bicolor* (L.) Moench. *Archives of Biochemistry and Biophysics* **260**, 822-829.

Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y. 2002. The Hidden Duplication Past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 13627-13632.

Smith-Uffen ME. 2014. Changing the Substrate Specificity of Arogenate Dehydratases (ADTs) From *Arabidopsis thaliana*. MSc. *Electronic Thesis and Dissertation Repository*. 2551. <https://ir.lib.uwo.ca/etd/2551>

Subramanian S, Stacey G, Yu O. 2006. Endogenous Isoflavones are Essential for the Establishment of Symbiosis Between Soybean and *Bradyrhizobium japonicum*. *Plant Journal* **48**, 261-273.

Vogt T. 2010. Phenylpropanoid Biosynthesis. *Molecular Plant* **3**, 2-20.

Wach A, Brachat A, Pohlmann R, Philippsen P. 1994. New Heterologous Modules for Classical or PCR-Based Gene Disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793-1808.

Walling JG, Shoemaker R, Young N, Mudge J, Jackson S. 2006. Chromosome-level Homeology in Paleopolyploid Soybean (*Glycine max*) Revealed Through Integration of Genetic and Chromosome Maps. *Genetics* **172**, 1893-1900.

West RW, Yocum RR, Ptashne M. 1984. *Saccharomyces cerevisiae* GAL1-GAL10 Divergent Promoter Region - Location and Function of the Upstream Activating Sequence UASG. *Molecular and Cellular Biology* **4**, 2467-2478.

Winkel BSJ. 2004. Metabolic Channeling in Plants. *Annual Review of Plant Biology* **55**, 85-107.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M,

Volckaert G, Ward TR, Wysocki R, Yen GS, Yu KX, Zimmermann K, Philippsen P, Johnston M, Davis RW. 1999. Functional Characterization of the *S. cerevisiae* Genome by Gene Deletion and Parallel Analysis. *Science* **285**, 901-906.

Wu SD, Han BC, Jiao YN. 2020. Genetic Contribution of Paleopolyploidy to Adaptive Evolution in Angiosperms. *Molecular Plant* **13**, 59-71.

Yamada T, Matsuda F, Kasai K, Fukuoka S, Kitamura K, Tozawa Y, Miyagawa H, 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.

Yoo H, Widhalm JR, Qian Y, Maeda H, Cooper BR, Jannasch AS, Gonda I, Lewinsohn E, Rhodes D, Dudareva N. 2013. An Alternative Pathway Contributes to Phenylalanine Biosynthesis in Plants via a Cytosolic Tyrosine:Phenylpyruvate Aminotransferase. *Nature Communications* **4**.

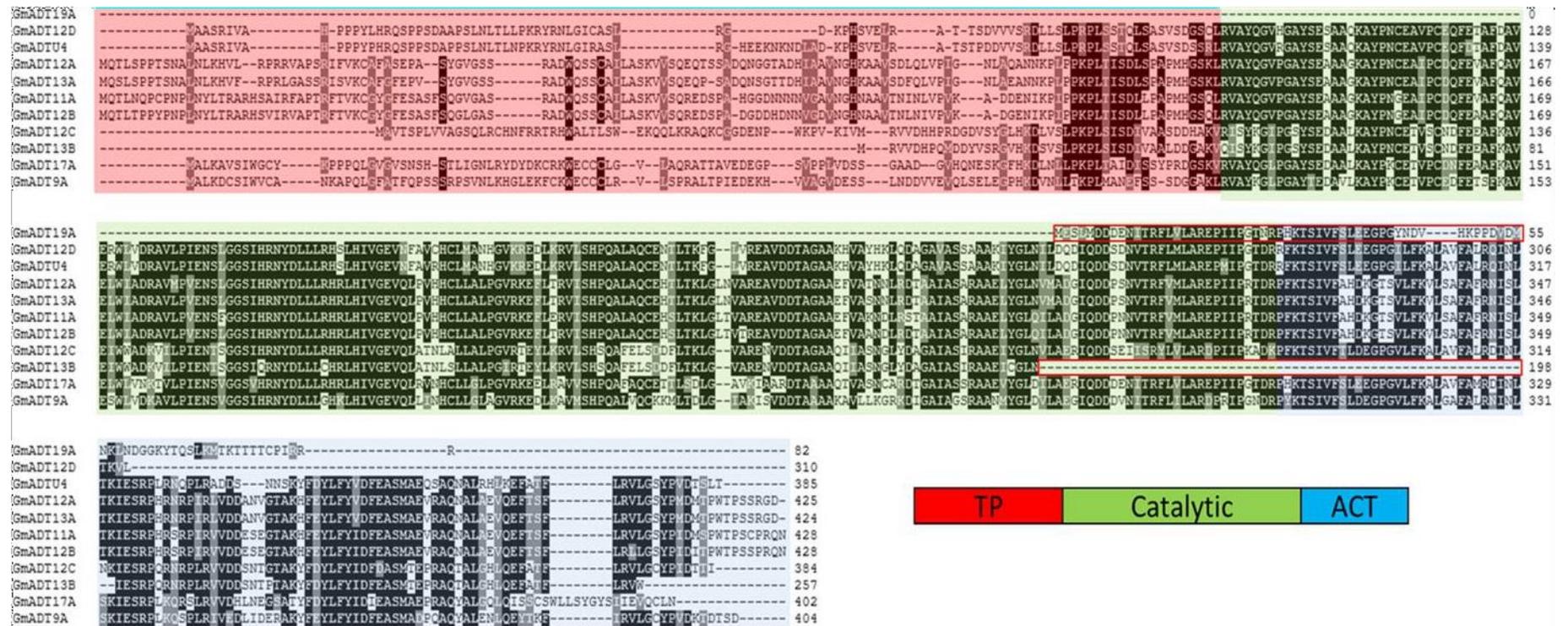
Zhao M, Chen P, Wang W, Yuan F, Zhu D, Wang Z, Ying X. 2018. Molecular Evolution and Expression Divergence of HMT Gene Family in Plants. *International Journal of Molecular Sciences* **19**.

Appendices

	GmADT12B	GmADT13A	GmADT13B	GmADTU4	GmADT9A
GmADT11A	96.63				
GmADT12A		99.32			
GmADT12C			96.50		
GmADT12D				97.27	
GmADT17A					65.1

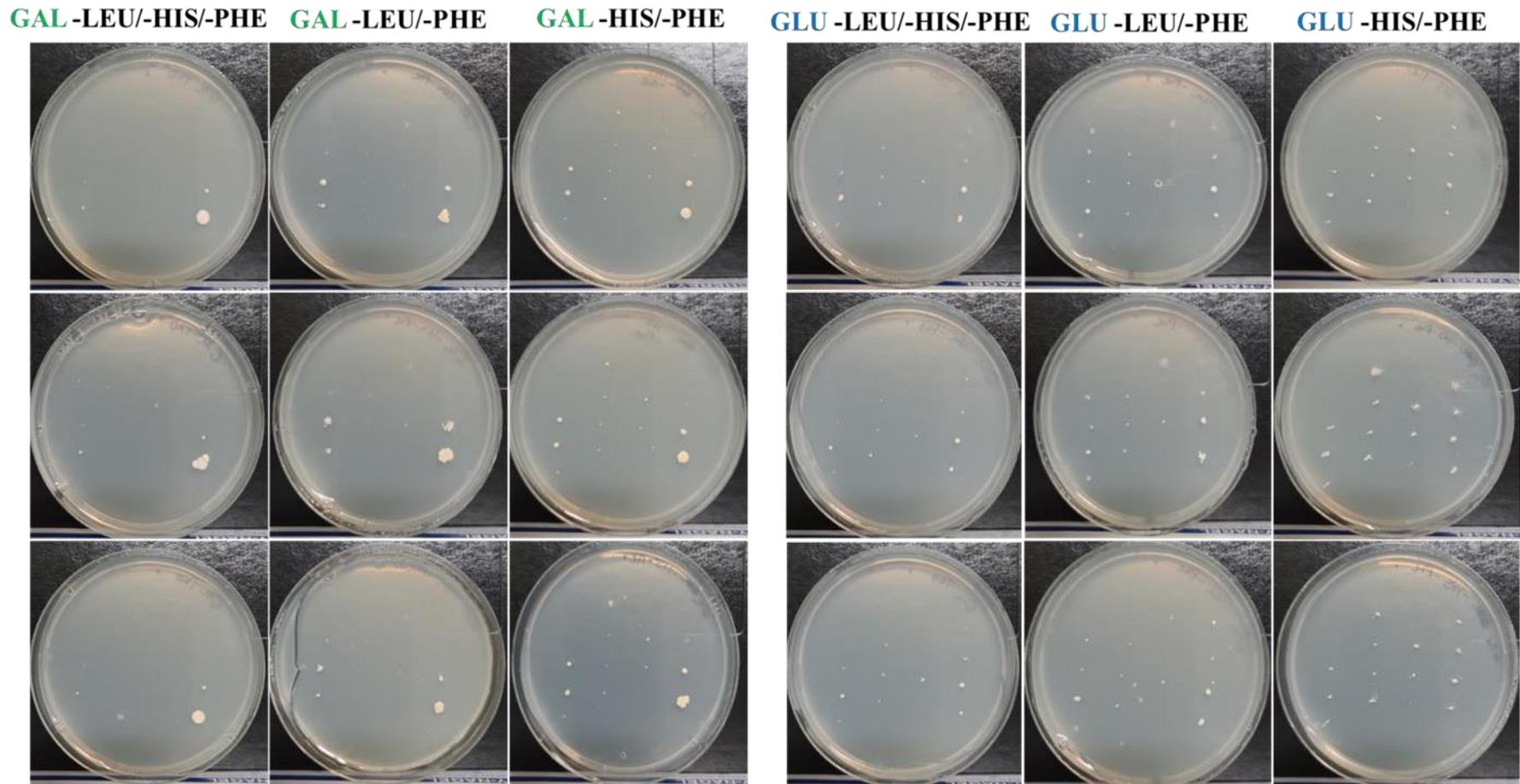
Appendix A. High identity pairs of GmADTs.

Percentage protein identity between each GmADT and its high identity pair



Appendix B. Protein alignment of all candidate GmADTs.

Highlighted in red is the transit peptide, green is the catalytic domain, and blue is the ACT regulatory domain. The red boxes show a large portion of the catalytic and ACT domains missing from GmADT13B, and that these same regions are all that is present for GmADT19A.



Appendix C. Raw photos of *S. cerevisiae* aro8aro9 yeast growth in the ADT assay.

On the right is the template for spotting yeasts.

<i>aro8</i>	423	425	423/ 425
PAT	PAT/ 423	ADT3	ADT3 /425
PAT/ ADT3	PAT/ 11A	PAT/ 12A	PAT/ 12B
PAT/ 12C	PAT/ 12D	PAT/ 13A	PAT/ U4
PAT/ 17A			

Ramtin Sirjani
Curriculum Vitae – June 2021
Department of Biology, Western University

PERSONAL INFORMATION

Date of Birth: October 18, 1994

Citizenships: Canada, Iran

EDUCATION

B.Sc. (Honours), Specialization in Genetics, Department of Biology, Western University, Ontario, Canada. Completed: April 2018

M.Sc. Cell and Molecular Biology, Department of Biology, Western University, Ontario, Canada. Completed: August 2021

RESEARCH INTERESTS

molecular genetics; gene interaction; gene regulation; evolutionary genetics; plant specialized metabolism; bioinformatics

EXPERIENCE**Graduate Research Assistant/ Research Affiliate**

Agriculture and Agri-Foods Canada and Western University; 2019-2021

Teaching Assistant

Biology 2217B: Economic Botany; Winter 2019

Biology 2290F/G: Scientific Methods in biology; Fall 2020 – Winter 2021

AWARDS

Western Graduate Research Scholarship (Western University, London, ON); 2019-2021

RESEARCH PRESENTATIONS

R. Sirjani; K. Pannunzio; S. Kohalmi; S. Dhaubhadel. Isoflavonoid metabolon and arogenate dehydratases in soybean (*Glycine max*): Identification and Functional Characterization. Plant Canada, July 7th – 10th, 2019, Guelph, ON. (*Poster Presentation*)