
Electronic Thesis and Dissertation Repository

10-10-2017 2:30 PM

The Mystery of Nuclear Localization of AROGENATE DEHYDRATASE5 from Arabidopsis thaliana

Sara Abolhassani Rad
The University of Western Ontario

Supervisor
Susanne Kohalmi
The University of Western Ontario

Graduate Program in Biology
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy
© Sara Abolhassani Rad 2017

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

 Part of the [Molecular Biology Commons](#)

Recommended Citation

Abolhassani Rad, Sara, "The Mystery of Nuclear Localization of AROGENATE DEHYDRATASE5 from Arabidopsis thaliana" (2017). *Electronic Thesis and Dissertation Repository*. 4941.
<https://ir.lib.uwo.ca/etd/4941>

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

Arogenate dehydratases (ADTs) have been identified to catalyze the last step of phenylalanine (Phe) biosynthesis in plants. All ADTs have a transit peptide sequence that targets them into the chloroplasts where the biosynthesis of Phe happens. Subcellular localization studies using fluorescently tagged *Arabidopsis thaliana* ADTs demonstrated that all six ADTs localize to chloroplast stromules (stroma filled tubules). However, one member of this family, ADT5, was also detected in the nucleus. As dual targeting of proteins to different cell compartments is an indication of multifunctionality, ADT5 nuclear localization suggests that this member of the ADT protein family is a moonlighting protein with a non-enzymatic role in the nucleus.

In this study, first the nuclear localization of the ADT5 was confirmed by expression of ADT5-CFP under the regulation of ADT5 native promoter. Using confocal microscopy and Western blot analysis it was shown that ADT5 localizes into the nucleus. Next, different possible mechanisms that could result to the nuclear localization of ADT5 were studied. It was tested if ADT5 can move directly from chloroplast stroma to the nucleus through stromules or if ADT5 enter the nucleus from cytoplasm using the nuclear import system. Data presented are consistent with a translocation from cytoplasm. A combination of an AQEH motif and a single amino acid Asn28 present in the N-terminus of ADT5 ACT domain were identified as potential protein interaction sites required for nuclear targeting of ADT5. Y2H screenings and protein-protein interaction analyses suggest that nuclear localization of ADT5 occurs through the interaction of the cytosolic portion of an ER membrane bound protein, PHOSPHOLIPID DIACYLGLYCEROL ACYLTRANSFERASE1 (PDAT1). A nuclear targeting sequence was identified in the N-terminal cytosolic portion of PDAT1. Hence, it is possible that PDAT1 piggybacks ADT5 into the nucleus through the nuclear import system. This study introduces ADT5 as a moonlighting protein and identifies a possible mechanism for ADT5 nuclear translocation.

Keywords

ADT5, *Arabidopsis thaliana*, chloroplast, moonlighting proteins, NLS, nuclear targeting, PDAT1, phenylalanine, protein-protein interactions, subcellular localization.

Co-Authorship Statement

This thesis contains information that has been published and the permission of use can be found in Appendix 4.

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

My contribution to this paper included cloning the *ADT5* native promoter into pCB5 and analyzing ADT5 localization under the control of its native promoter. I also performed all the Western blots for ADTs. All the pCB constructs were previously generated by Crystal Bross and the subcellular localization analysis were performed by Travis Howes. Travis Howes and Ornela Kljakic performed stromule inhibition analyses and the statistic analysis.

Only portions of the publication which have been performed by myself are included in this thesis.

Acknowledgments

I would like to express my sincere appreciation to all those who made this Ph.D thesis possible. First and foremost, I am thankful to my supervisor, Dr. Susanne Kohalmi, for her immense support, excellent guidance and constant encouragement. I am very grateful for her inspiring advice and thoughtful help during my time in her lab. Many thanks also go to my advisors Dr. Denis Maxwell and Dr. Yuhai Cui for their valuable consultations and suggestions throughout every step of my degree.

I wish to express my gratitude to Dr. Rima Menassa for providing me with the opportunity to perform part of my research project in her lab and patiently reading and commenting on my thesis. I would like to extend my acknowledgements to all of the Menassa lab members for their help and friendship. Especially, I would like to thank Dr. Igor Kolotilin for training and helping me to perform the chloroplast genome transformations and always inspiring me to push myself a little bit further. I am also thankful to Dr. Reza Sabrianfar and Angelo Kaldis who never hesitated to answer my endless questions and they helped me a lot.

I was fortunate to work with wonderful people past and present in the Kohalmi group who were always there to support me and made my graduate time more enjoyable. I would particularly like to thank Megan Smith-Uffen for generating the chimeric constructs and Colette van Dyk for spending a whole summer by my side, helping me perform many transformations and plasmid preps. A special thank goes to Emily Clayton for her kindness and her generosity of taking time to proofread my thesis. I also wish to thank all my colleagues and friends in the department of biology at Western University and Agriculture and Agri-food Canada for their companionship and all the good memories we made together.

To my parents, Mahnaz and Hamid, and my only brother Sina for their constant love and encouragement, despite the long distance, that always helped me be confident and be the best I can be. I love you and I am always grateful to have you in my life.

Last, but not least, I would like to thank my husband, Ali, who traveled this journey with me from the beginning with his unflagging support. You felt every obstacle and success as deeply as I did during the past five years. This thesis is dedicated to you.

Table of Contents

Abstract	ii
Co-Authorship Statement.....	iv
Acknowledgments.....	v
Table of Contents	vi
List of Tables	x
List of Figures	xi
List of Appendices	xv
List of Abbreviations	xvi
1 Introduction	1
1.1 Importance of phenylalanine.....	1
1.2 Phe biosynthesis and role of arogenate dehydratases	2
1.3 AROGENATE DEHYDRATASE family	5
1.4 Subcellular localization of ADTs.....	8
1.5 Moonlighting: the process of multitasking	11
1.6 Dual targeting of a protein in the cell	14
1.7 Plastid to nucleus retrograde signaling	15
1.8 Possible role of stromules in molecule trafficking	17
1.9 Nuclear pore complex and nuclear import system.....	18
1.10 Protein-protein interactions and nuclear targeting	20
1.11 Research goals and objectives.....	21
2 Material and Methods	22
2.1 Media, solutions and media additives.....	22
2.1.1 Media	22

2.1.2	Solutions and buffers	22
2.1.3	Media additives	24
2.2	Organisms and growth conditions	25
2.2.1	Bacterial strains and growth conditions	25
2.2.2	Plant material and growth conditions	25
2.2.3	Yeast strains and growth conditions	26
2.3	Cloning procedures and construct design	26
2.3.1	Primer design	26
2.3.2	Polymerase chain reaction conditions	27
2.3.3	Plasmid DNA isolation and sequencing	27
2.3.4	Cloning systems and plasmid strains	27
2.4	Bacterial and yeast transformations	28
2.5	Plant transformations	29
2.5.1	Transient transformation of <i>N. benthamiana</i>	29
2.5.2	Chloroplast genome transformation of <i>N. tabacum</i>	29
2.5.3	Stable transformations of <i>A. thaliana</i>	30
2.6	Plant genomic DNA isolation	30
2.7	Bimolecular Fluorescence Complementation Assay	31
2.8	Confocal microscopy	31
2.9	Y2H assay and cDNA library screening	31
2.10	Protoplast preparation	32
2.11	Protein isolation and Western blotting	33
2.12	Histochemical GUS staining	33
2.13	Sequence analysis	34
3	Results	35

3.1	Nuclear localization of ADT5.....	35
3.2	Nuclear localization of ADT5 using the native promoter.....	40
3.3	Stromules, tunnels from chloroplasts to the nucleus?.....	50
3.4	Does ADT5 have a nuclear localization sequence.....	71
3.4.1	<i>In silico</i> analysis to identify a putative NLS in ADT5.....	71
3.4.2	Generating truncated fusion proteins to find a NLS	76
3.4.3	Generating chimeric ADT5-ADT2 proteins	86
3.4.4	N- terminal ACT domain <i>in silico</i> analysis to find the NLS	93
3.5	Protein-protein interaction and nuclear localization	93
3.5.1	<i>In silico</i> analysis of potential ADT5 interactors	93
3.5.2	<i>In planta</i> interaction of ADT5 and IMPA6.....	100
3.6	Identifying novel ADT5 interactors.....	107
3.7	PDAT1 <i>In silico</i> analysis	110
3.8	<i>In planta</i> identification of ADT5 and PDAT1 interaction.....	124
4	Discussion	131
4.1	ADT5 localizes to the stromules of chloroplast and into the nucleus.....	131
4.2	Functional relevance of stromule localization of ADTs	132
4.3	Close positioning of chloroplasts and nucleus.....	132
4.4	The importance of the ACT domain	133
4.5	Involvement of the ACT domain in protein conformation	136
4.6	Protein-protein interaction and nuclear localization	137
4.7	PDAT1 and its similarity to lipid synthesis enzymes	138
4.8	ER membrane bound transcription factors	140
4.9	ADT5: a moonlighting protein.....	141
4.10	Concluding remarks and future directions	142

5	References	145
	Appendices.....	157
	Curriculum Vitae	163

List of Tables

Table 1 Predicted ADT5 interators in BAR database.....	96
Table 2 Putative ADT5 interactors identified by Y2H screening.....	108

List of Figures

Figure 1 The final two steps of Phe biosynthesis.	3
Figure 2 Amino acid sequence alignment of <i>A. thaliana</i> ADTs.....	6
Figure 3 Stromule structure.	9
Figure 4 ADT2 and ADT5 localization patterns.....	12
Figure 5 Expression pattern observed in control transformations in <i>N. benthamiana</i>	36
Figure 6 Subcellular localization of ADT-CFPs in <i>N. benthamiana</i> leaves.	38
Figure 7 Protein expression of ADT-CFPs in <i>N. benthamiana</i>	41
Figure 8 Expression pattern of GFP and GUS in <i>N. benthamiana</i> and <i>A. thaliana</i> using ADT5 native promoter.....	43
Figure 9 Subcellular localization of the ADT5-CFP with <i>proNat5</i> and control transformations in <i>N. benthamiana</i>	46
Figure 10 Detection of ADT5-CFP expressed with <i>proNat5</i> in transiently transformed <i>N. benthamiana</i> leaves.	48
Figure 11 Close proximity of stromules and nucleus in <i>N. benthamiana</i> leaf samples expressing ADT5-CFP.	51
Figure 12 Schematic representation of full-length and deletion constructs used for chloroplast genome transformation in <i>N. tabacum</i>	54
Figure 13 Genotyping of transplastomic calli.....	56
Figure 14 Controls in chloroplast transformation experiments.	58
Figure 15 Localization pattern of ADT-FL-GFP in transplastomic plantlets.....	60

Figure 16 Localization pattern of ADT-IS-GFP in transplastomic plantlets.	62
Figure 17 Localization pattern of ADT-S-GFP in transplastomic plantlets.	65
Figure 18 Phenotype of transplastomic GFP control, ADT5-IS and ADT4-IS plants.	67
Figure 19 Leaf phenotype and localization pattern of ADT-GFP fusion proteins in transplastomic <i>N. tabacum</i>	69
Figure 20 Protoplast images exhibiting patterns of ADT5-IS-GFP in young, intermediate and old leaves of transplastomic plants.	72
Figure 21 Protein expression of ADT-GFPs in <i>N. tabacum</i> transplastomic plants.	74
Figure 22 Schematic representations of ADT5 and ADT4 full-length and deletion constructs.	77
Figure 23 Subcellular localization of ADT5-CFPs in <i>N. benthamiana</i> leaves.	80
Figure 24 Subcellular localization of ADT4-CFP in <i>N. benthamiana</i> leaves.	82
Figure 25 Protein expression of the full length and truncated ADT5 and ADT4 in <i>N. benthamiana</i>	84
Figure 26 Schematic representation of full-length and chimeric ADT2/ADT5 constructs.	87
Figure 27 Subcellular localization of the ADT2-CFP and ADT5-CFP in <i>N. benthamiana</i> leaves.	89
Figure 28 Subcellular localization of the C1-C7 chimeric proteins in <i>N. benthamiana</i> leaves.	91
Figure 29 Sequence alignment of ACT domain of six <i>Arabidopsis</i> ADTs.	94

Figure 30 Control transformations and interactions involving ADT5 and IMPA6 assayed by Y2H in Yeast AH109.....	98
Figure 31 Control transformations ADT5 and IMPA6 fusion constructs in <i>N. benthamiana</i>	101
Figure 32 Interactions involving ADT5 and IMPA6 assayed by BiFC in <i>N. benthamiana</i> leaves.	103
Figure 33 Interactions involving IMPA6 and ADT5 assayed by BiFC in <i>N. benthamiana</i> leaves.	105
Figure 34 Control transformation for ADT5 interactors alone and with DB vector in Yeast AH109.	111
Figure 35 Interaction of ADT5 interactors with ADT1, ADT2 and ADT3 in Yeast AH109.	113
Figure 36 Interaction of ADT5 interactors with ADT4, ADT5 and ADT6 in Yeast AH109.	115
Figure 37 Interactions of PDAT1 and ADTs in Yeast AH109.....	117
Figure 38 Phylogenetic analysis of <i>A. thaliana</i> PDAT/LCATs.....	119
Figure 39 Alignment of PDAT1 and LCAT2 protein sequences from <i>A. thaliana</i>	122
Figure 40 Phylogenetic analysis of <i>A. thaliana</i> PDAT/LCAT sequences together with model organisms (landmark) PDAT/LCATs.	125
Figure 41 Control transformations and interactions involving ADT5 and PDAT1 assayed by BiFC in <i>N. benthamiana</i> leaves.....	127

Figure 42 Interactions involving PDAT1 and ADTs assayed by BiFC in <i>N. benthamiana</i> leaves.	129
---	-----

List of Appendices

Appendix 1 Primers.	157
Appendix 2 List of vectors.....	160
Appendix 3 Viability of yeast populations.	161
Appendix 4 Journal of Experimental Botany authorship permission.	162

List of Abbreviations

NOTE: SI units are not listed.

ACT	Aspartokinase-chorismate mutase-TyrA
Asn	Asparagine
Asp	Aspartic acid
BAP	Benzyl amino purine
BAR	The bio-analytic resource for plant biology
BiFC	Biomolecular fluorescence complementation assay
BME	Beta mercapto ethanol
BSA	Bovin serum albumin
bZIP	Basic domain leucine zipper protein
CaMV	Cauliflower mosaic virus
CAT	Catalytic domain
CFP	Cyan fluorescent protein
ChIP	Chromatin immunoprecipitation
co-IP	Co-immunoprecipitation
CTAB	Hexadecyl trimethyl ammonium bromide
DB	DNA binding
DMSO	Dimethyl sulfoxide
dpi	Days after infiltration
EDTA	Ethylene diamine tetracetic acid
ER	Endoplasmic reticulum
FL	Full length
GFP	Green fluorescent protein
His	Histidine
IS	I-region and short length
ISIS	Interaction sites identified from sequence
LB	Lysogen broth
Leu	Leucine
MS	Murashige and Skoog

OD	Optimal density
MES	Morpholino ethanesulfonic acid
MTF	Membrane bound transcription factors
NAA	1-Naphthalene acetic acid
NCBI	National Center for Biotechnology Information
NE	Nuclear envelope
NLS	Nuclear localization signal
NPC	Nuclear pore complexes
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PEB	Protein extraction buffer
PCR	Polymerase chain reaction
Phe	Phenylalanine
PMSF	Phenylmethanesulfonylfluoride
PPVP	Polyvinylpolypyrrolidone
PPI	Protein-protein interactions
RMOP	Revised medium for organogenesis
ROS	Reactive oxygen species
SCOP	Structural classification of proteins
SD	Synthetic dextrose
SDS	Sodium dodecyl sulfate
TA	Transcription activation
TAIR	The <i>Arabidopsis</i> information resource
TF	Transcription factor
TP	Transit peptide
Trp	Tryptophan
TSP	Total soluble protein
Tyr	Tyrosine
UAS	Upstream activating sequence
X-Gluc	5-bromo-4-chloro-3-indolyl-glucuronide
YC	C-terminal half of YFP

YEB	Yeast extract and beef
YFP	Yellow fluorescent protein
YN	N-terminal half of YFP
YPDA	Yeast peptone dextrose adenine
Y2H	Yeast two hybrid

1 Introduction

This study focuses on the characterization of AROGENATE DEHYDRATASE5 (ADT5) *in planta*. ADT5 is a member of the ADT protein family that catalyzes the last step of phenylalanine (Phe) biosynthesis in *Arabidopsis thaliana*. This protein is found in two different subcellular compartments, the chloroplast and the nucleus, suggesting the possibility that ADT5 could have a second non-enzymatic role in the nucleus.

1.1 Importance of phenylalanine

Phenylalanine (Phe) is one of twenty amino acids that serve as building blocks in protein synthesis in all organisms. Only plants and microorganisms are able to synthesize Phe *de novo* while animals need to obtain this amino acid through their diet, making Phe an important dietary component for animals (Herrmann and Weaver, 1999). In addition to protein synthesis, in plants, Phe is a major precursor for the biosynthesis of secondary metabolites such as lignin, flavonoids and isoflavonoids (Herrmann and Weaver, 1999; Vogt, 2010).

Lignin is one of the most abundant biopolymers on earth and more than 30% of the photosynthetic fixed carbon in vascular plants is directed into lignin biosynthesis (Maeda and Dudareva, 2012). It has an important structural role in strengthening cell walls and mechanical stability of plant cells (Vanholme et al., 2010). Furthermore, lignin solidifies the cell wall and defends plants against pathogens, fungi and bacteria (Frei, 2013). More recently, lignin has become of great importance for the biofuel industry due to its role in inhibition of microbial fermentation in cell walls, which decreases the efficiency of converting plant material into energy and causes revenue losses (Frei, 2013).

Flavonoids and isoflavonoids are abundant in plants and have a wide variety of functions. They are responsible for the marvelous colors and aroma of flowers and fruits which are not only pleasant to humans but more importantly help to attract pollinators and consequently are involved in plants reproductive success (Samanta et al., 2011). The antioxidant characteristics of flavonoids protect plants against biotic and abiotic stresses and serve as a unique UV-filter. As stress and high light conditions cause the formation of highly reactive chemical species that damage the plant cell, flavonoids are known as components of detoxification and receive the electrons from these highly reactive

molecules and act as antioxidants (Zhou et al., 2016). Moreover, the antioxidant properties of flavonoids have an important effect on human health, hence flavonoids are suggested to be used in prevention and treatment of cardiovascular diseases, cancer and aging (Yao et al., 2004).

The importance of Phe and its downstream products to plant and animal life explains the need to increase our knowledge of the enzymes and pathways involved in its synthesis.

1.2 Phe biosynthesis and role of arogenate dehydratases

Phe biosynthesis begins with a series of enzymatic reactions in the shikimate pathway which connect the metabolism of carbohydrates to the synthesis of aromatic amino acids (Herrmann and Weaver, 1999). Chorismate is the final product of the shikimate pathway and is the last common precursor for the biosynthesis of the three amino acids tyrosine (Tyr), tryptophan (Trp) and Phe (Herrmann and Weaver, 1999; Tzin and Galili, 2010). In Phe biosynthesis chorismate initially is converted to prephenate by chorismate mutase (CM) and then prephenate is converted to Phe via one of the two alternative routes (Figure 1; Maeda and Dudareva, 2012).

In microorganisms, prephenate is first decarboxylated/dehydrated to phenylpyruvate by a prephenate dehydratase (PDT) and then transaminated to Phe by a phenylpyruvate aminotransferase (PPAT; Bentley, 1990; Tzin and Galili, 2010). In plants, these enzymatic steps occur in reverse order. Prephenate first undergoes transamination to arogenate by prephenate aminotransferase (PAT) and is then decarboxylated/dehydrated to Phe by an enzyme called arogenate dehydratase (ADT; Cho et al., 2007; Maeda et al., 2010).

Biosynthesis of Phe through the ADT pathway was shown first in chloroplasts of *Nicotiana glauca* and *Spinacia oleracea* where no PDT was detected (Jung et al., 1986). Since then, ADT activity has been observed in chloroplasts of wide a range of plants and been predominantly accepted (Cho et al., 2007; Yamada et al., 2008; Chen et al., 2016). However, PDT activity is still observed in *Arabidopsis* and *Petunia* which has lead to the suggestion that PDT pathway exists in plants (Bross et al., 2011; Maeda and Dudareva, 2012; Yoo et al., 2013).

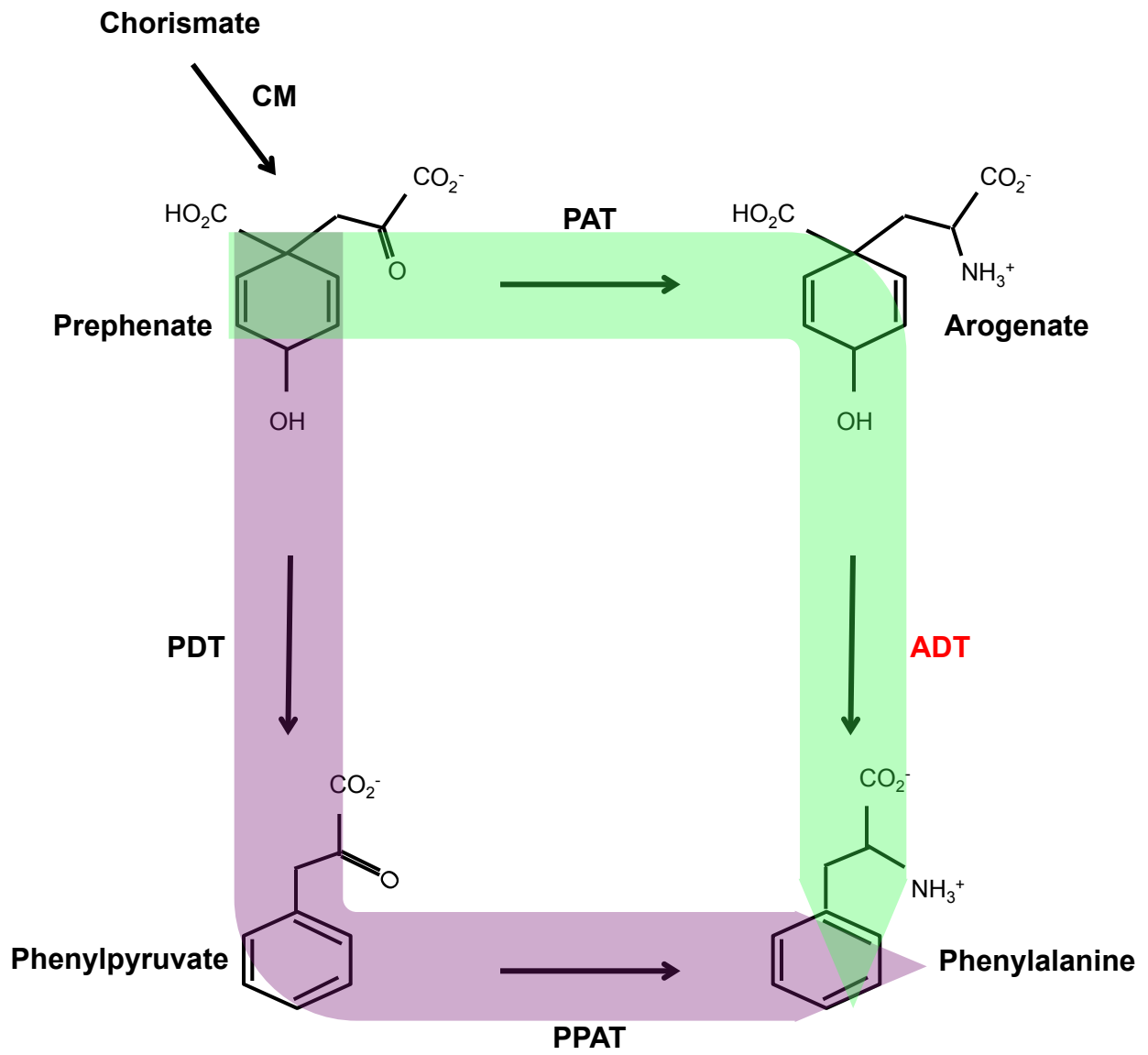
Figure 1 The final two steps of Phe biosynthesis.

Chorismate, the final product of the shikimate pathway is converted to prephenate by chorismate mutase. Prephenate can then be used for Phe biosynthesis via two pathways. In the prephenate pathway, prephenate is first dehydrated/decarboxylated by a prephenate dehydratase to phenylpyruvate, and then phenylpyruvate is transaminated to Phe by a phenylpyruvate aminotransferase. In the arogenate pathway prephenate is first transaminated to arogenate by a prephenate aminotransferase and then is dehydrated/decarboxylated by an arogenate dehydratase to Phe.

Purple arrow: PDT pathway, green arrow: ADT pathway, red: ADT enzyme.

ADT: arogenate dehydratase, CM: chorismate mutase, PPAT: phenylpyruvate aminotransferase, PAT: prephenate aminotransferase, PDT: prephenate dehydratase.

Modified from Tzin and Galili (2010).



1.3 AROGENATE DEHYDRATASE family

A family of six ADTs were identified for the first time in *A. thaliana* based on their high degree of sequence similarities to bacterial PDTs (Cho et al., 2007; Ehrling et al., 2005). Further analysis demonstrated that these enzymes predominantly produce Phe through the aroenate pathway (Cho et al., 2007; Bross et al., 2011) and were named ADTs (ADT1: At1g11790, ADT2: At3g07630, ADT3: At2g27820, ADT4: At3g44720, ADT5: Atg22630 and ADT6: At1g08250). Today families of ADTs have been identified in many different plant species, including five ADTs in *Oryza sativa*, three in *Petunia hybrida* and nine in *Pinus taeda* (Maeda et al., 2010; Yamada et al., 2008; El-Azaz et al., 2016). The presence of multiple isoforms of these enzymes suggests that ADTs might have evolved different properties or they might have faced different transcriptional and/or post-translational regulatory processes to gain distinct functional roles. For instance, aside from aroenate, two members of *A. thaliana* ADTs (ADT1 and ADT2) as well as two members of ADT family in *Petunia* are able to use prephenate as a substrate meaning that they can act as ADTs and PDTs (Cho et al., 2007; Bross et al., 2011; Maeda et al., 2010). Furthermore, genetic analysis demonstrated that ADTs are expressed ubiquitously but in a tissue-specific manner and they preferentially contribute to biosynthesis of specific downstream secondary metabolites (Corea et al., 2012b). For instance a study of *A. thaliana* mutants demonstrated that ADT4 and ADT5 play a dominant role in lignin biosynthesis in stem tissues (Corea et al., 2012a) while ADT1 and ADT3 have a high contribution to anthocyanin biosynthesis in leaves (Chen et al., 2016).

Plant ADTs have three domains: an N-terminal transit peptide (TP) responsible for targeting the proteins into the chloroplasts, an internal catalytic domain (CAT) catalyzing the decarboxylation/dehydration reaction and a C-terminal ACT domain (aspartokinase-chorismate mutase-TyrA) (Cho et al., 2007). The ACT domain is involved in allosteric regulation mediated by ligand binding (Vivan et al., 2008). Both catalytic and ACT domains are conserved in ADTs and PDTs across plants, fungi and bacteria while TP sequences are unique to plant ADTs (Cho et al., 2007; Bross et al., 2011). Alignments of six full-length *A. thaliana* ADTs show high sequence similarity in CAT domain and ACT domain (62-98% and 61-92% similarity, respectively) but the TP sequences are far more diverse (Figure 2).

Figure 2 Amino acid sequence alignment of *A. thaliana* ADTs.

Amino acid alignments of *A. thaliana* ADTs, reveal a high degree of sequence similarity. (ADT1: At1g11790, ADT2: At3g07630, ADT3: At2g27820, ADT4: At3g44720, ADT5: Atg22630 and ADT6: At1g08250). The most conserved sequences are found in the catalytic domain (62-98%) and ACT domain (61-92%), while the transit peptide sequence are less conserved. It is not known whether the I-region is part of the transit peptide or the catalytic domain.

Green: ACT domain, light blue: I-region, pink: catalytic domain, yellow: transit peptide. Shading within the sequences indicates amino acid conservation; Black: 100%, medium grey: 75%, light gray: 50%, white <50%, numbers: amino acid position.

Adapted from Styranko (2011).

ADT1MALRCFPIWVCQTTHHRSPLMGLAEFDAKRRRFCLE/CSSSASQIRAVTAIEGELPF:SELKKSSIDEILGL	71
ADT2MAMHTVRLSPATQLHGCISSNLSPPNRKPNNSIVRYGCGSSKRFIRIVTVLASLIRENDANGRDNS.....	64
ADT3MRTLL.P..SHTPATVTTAARRRHVIHCAGKRSDSFSI.N..S.SSDWQSSCAILSSKVN:SQEQ:SESLS.SN:NGSSSYIHVSAVINGHNNGAG.VSIDNLVPE	94
ADT4	...MQAATSCDLKFRSTDPTSRNK.CFSHAIPKRVAVTCG.YRSESFSPNGVSVSRDWQSSCAILSSKVAS/ENT/GGIA.DKI..AA...V...INGHTNGS..V.INLB/LVA/	97
ADT5	MQTISPAFSCDLKSVIQPNLTAKKARYSHVNGKRVSVRCS.YRSESFSPNGVGSRAWDQSSCAILASKV:SAENSS.....SV...AV...V...INGHSNGS..V.IDLSILVPS	97
ADT6MKA.L.S..SSSP..ILGASQP..AT..A...T.ALIA.R..S.GRSEWQSSCAILT:SKVI:SQEESISLPVPPV:SGGV).IHL...NGHNS/AAARV/GMNLVPI	81
Consensus	1	
ADT1	TQETQSLSFHRDLSMLPKPLTANSIYSSDGDLSKVRISFQCTEGAYSETAALKAEPNCETVPOCEFAAFCAVELIWL/DKAVLP:ENS:GGSIHHRNYDLLLRHLHIVDEVHLP	185
ADT2	VRAVEVKKIFEDSPLLPKPLSSNCLTESVSNCSRVRVAYQGVGAYSEAAACKAYPNCEAVPCEFFDPAFEAVERWL/DRAVLP:ENS:GGSIHHRNYDLLLRHLHIVGEVKLA	178
ADT3	NNNQS...I.QSK....KPLSISDLSAPAMHGSNLRVAYQGVGAYSEAAACKAYPNQQAIPDCQFEVAFCAVELIWIADRAVLP:ENS:GGSIHHRNYDLLLRHLHIVGEVQLP	200
ADT4	.ES.T.....NGKLAPAQPLTITDLSAPALHGSILRVAYQGVGAYSEAAACKAYPNQDAIPDCQFDVAFCAVELIWIADRAVLP:ENS:GGSIHHRNYDLLLRHLHIVGEVQIP	204
ADT5	.KSQH....NGKPGLIQPLTITDLSAPASHCSTLRVAYQGVGAYSEAAACKAYPNSEAIPDCQFDVAFCAVELIWIADRAVLP:ENS:GGSIHHRNYDLLLRHLHIVGEVQIP	205
ADT6	EKSDSNPLVPQHRHNPLKPLSMTDLSAPAMHGSNLRVAYQGVGAYSEAAACKAYPNQQAIPDCQFEVAFCAVELIWIADRAVLP:ENS:GGSIHHRNYDLLLRHLHIVGEVQLP	195
Consensus	pl l s r qg gayse aa ka pn pc f af ave w d avlp ens ggsihrnydlllrh lhiv ev	
ADT1	VNHCLLGVPGVKKEDIKCMISHPOALDQCVNSLNNLG..IQRLSAKDTATAAQIVSSSGKILVCAIASVRAANTYGLDILAEINIQDLVNNVTREILAREPIIPRTIRPEKTSI	297
ADT2	VRHCLLANHGWNIEDLRVLSHPQALACCENTITKLG..LVREAVDDTAGAAQIAFENLNDAWASEKAAKIYCNIVAKIQDDCIDNVTRFIMLAREPIIPSTIRLEKTSI	290
ADT3	VHHCILALPGVRKEFLTRVLSHPQGLACCEHTITKLGILNVAREAVDDTAGAAEFIAANNIRLTAAMASAFAAEITYGLEIIEIDGIQDDASNVTRFVMLAREPIIPRTIRPEKTSI	314
ADT4	VHHCILALPGVRTDCVSRVLSHPQALACTEHSIDVLTPHAAREAFHDATAAAEYISANDLEDATAVASAFAAEYLYICILADGIQDDPGNVTRFIMLAREPIIPRTIRPEKTSI	318
ADT5	VHHCILALPGVRTDCITRVLSHPQALACTEGSINKLTPKAAIEAFHDATAAAEYIAANNLHDTAAVASAFAAEYLYICILADGIQDDAGNVTRFIMLARDPIIPRTIRPEKTSI	319
ADT6	VHHCILALPGVRKEFLTRVLSHPQGLACCEHTITKLGILNVAREAVDDTAGAAEFIASNNLRDTAAAMASAFAAEITYGLEIIEIDGIQDDVSNVTREVMMLAREPIIPRTIRPEKTSI	309
Consensus	shcl gv v shpq l q l l dta aa d a as aa y l i iqdd nvtrf lar p ip t r ktsi	
ADT1	VFSLEE..GPCVLFKALAVFAIRSNLSKIESRPQRRRPLRVVGDGSNN:SAKFEIYIFYIDFEASMADEFAQALGHCEFA:SFIRILG:YPMILV:R.....	392
ADT2	VFSLEE..GPCVLFKALAVFAIRCNILSKIESRPLRKHPFASGG....LKGFEDYIFYVDFEASMADEVAQNALRHLEEFATIFIRVLGSYPMDITML.....	381
ADT3	VFA.HE.KCTOVLFKVL:SAFAFRNISL:KIESRPNHNVP:IRLVDEANVTAKFEFFYIFYVDFEASMAESFAQNALSEVCEFTSFLIRVLGSYPMDTSWSP:SS:SSSTFSL..	424
ADT4	VFAAQEHKGT:SVLFKVL:SAFAFRDISL:KIESRPHHNRP:IRLVVDGSGF:GTSKNFEFFYIFYVDFEASMAEPHAQNALAEVCEYTSFLIRVLGSYPMDTPW:MTSTNEA.....	424
ADT5	VFAAQEHKGT:SVLFKVL:SAFAFRNISL:KIESRPHQNC:IPRVVGDENVGT:SKFEFFYIFYVDFEASMAEFAQNALAEVCEYTSFLIRVLGSYPMDTPWSTLP:SEOV.....	425
ADT6	VFA.HE.KGTSVLFKVL:SAFAFRDISL:KIESRPNHNRP:IRLVDDANVTAKFEFFYIFYVDFEASMAEFAQNALAEVCEFTSFLIRVLGSYPMDTPWSTPST:SS.....	413
Consensus	svf e g vlfk l fa r i l kiesrp p r k f y fy dfeasma aq al e f r lg yp d	

In addition, there is a short sequence of 17 amino acids immediately upstream of the CAT domain of ADTs that is more conserved than TP sequences and does not exist in PDTs. This region is called the intermediate region (I region) as it is not clear if it actually belongs to the CAT domain of plant ADTs or to the TP and its importance is still unknown (Figure 2).

1.4 Subcellular localization of ADTs

Fluorescently tagged fusion proteins and studies of their subcellular localization have revolutionized the approaches toward identifying the functional roles of proteins in cell (Sparkes and Brandizzi, 2012). Using these tags, new and unexpected roles have been identified for members of protein families. For example, dual localization of a member of *A. thaliana* hexokinase family in chloroplasts and mitochondria led to identifying two roles for this protein in glucose metabolism and glucose signaling (Karve et al., 2008).

According to prediction programs like ChloroP (Emanuelsson et al., 1999) the TP targets ADTs to chloroplasts which is consistent with the enzymatic role of ADTs in Phe biosynthesis in chloroplasts (Jung et al., 1986; Cho et al., 2007). First experimental observations of *A. thaliana* ADTs identified they localize as punctuated patterns in chloroplasts (Rippert et al., 2009). Later, it was demonstrated that most *A. thaliana* ADTs are targeted to the stromules of chloroplasts (Figure 3A, Bross et al., 2017). Stromules are stroma-filled projections of the chloroplast membrane and can form in different shapes like elongated tubes or short protrusions (Gray et al., 2001; Gunning, 2005; Köhler and Hanson, 2000) (Figure 3B). In confocal microscopy, chloroplast autofluorescence can be seen as red ovoid structures due to the presence of chlorophyll II in the thylakoid membrane. Since stromules do not contain thylakoids they do not autofluoresce, however they can be visualized in the presence of a stroma targeted, fluorescently tagged protein (Gray et al., 2001).

In addition to stromule, two of the *A. thaliana* ADTs have additional subcellular localization patterns suggesting novel and non-enzymatic roles for these proteins (Bross et al., 2017).

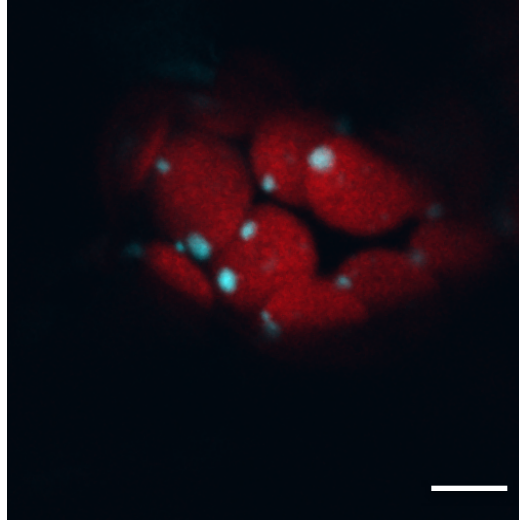
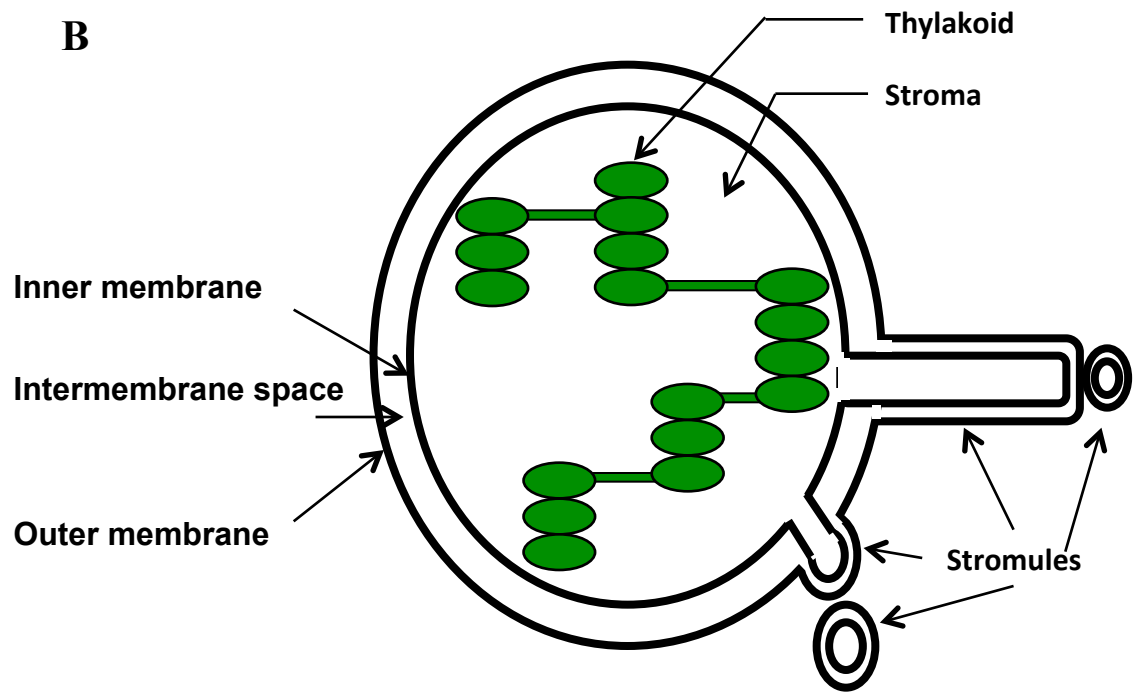
Figure 3 Stromule structure.

Two membranes surround chloroplasts, the inner and outer membranes that are separated by the intermembrane space. The stroma is a protein-rich fluid within the chloroplast. Thylakoids are another membrane system located within the stroma of chloroplasts. Thylakoids have the photosynthetic components and chlorophyll, which is a pigment and causes chloroplast auto fluorescence in confocal microscopy. Stroma-filled protrusions formed by the chloroplast membranes are named stromules and since they do not contain thylakoids, they do not autofluoresce and hence are not visible in confocal microscopy unless a fluorescent fusion protein accumulates there. Stromules are variable in size and shape and they can form as long tubes or short protrusions. They can bud off and form vesicles.

A) A laser scanning confocal microscopy picture presenting the stromule localization of ADT4-CFP in *N. benthamiana*. Chloroplasts autofluoresce as red round structures and ADT4-CFP localizes in short protrusions stromules.

B) Schematic showing the chloroplast and stromules structure.

Adapted from Bross et al. (2017).

A**ADT4-CFP/Chlorophyll****B**

ADT2 localizes at the equatorial plane of chloroplasts as a ring structure (Figure 4A left). Co-localization of ADT2 with a known/well described chloroplast division protein FtsZ (Vitha et al., 2001) showed that both proteins formed a ring at the same time in the same place (Figure 4B, Abolhassani Rad et al. Manuscript submitted). ADT2's role in chloroplast division is further substantiated as chloroplast morphology was affected in *A. thaliana* *adt2* mutant lines and proper expression of FtsZ requires the presence of ADT2 (Bross et al., 2017). Furthermore, transmission electron microscopy of *adt2* mutant lines demonstrated that *adt2* lines have long and misshapen chloroplasts (Figure 4C left) compared to the ovoid and uniform size of wild type chloroplasts (Figure 4C right, Abolhassani Rad et al. Manuscript submitted). Taken together this evidence suggests that ADT2 is one of as yet unidentified components required for chloroplast division.

Moreover, ADT5 was observed in the nucleus in *N. benthamiana* and *A. thaliana* (Bross et al., 2017). The nuclear localization of ADT5 was confirmed by co-expression of ADT5-CFP and a nuclear membrane marker (NUCLEOPORIN-1) NUP1-YFP in *N. benthamiana* (Figure 4A right). Chloroplast targeting of ADT5 by TP and the enzymatic role of this protein is clear, however how this protein localizes into the nucleus and what function ADT5 has in the nucleus is yet unknown.

1.5 Moonlighting: the process of multitasking

The unique and different localization patterns of ADT2 and ADT5 in the ADT family suggest that these two ADTs are moonlighting proteins. Moonlighting refers to proteins that perform multiple autonomous and often unrelated roles (Jeffery, 1999). Different roles of moonlighting proteins are not due to different RNA splice variants or to gene fusions and are not separated in different protein domains (Huberts and van der Klei, 2010; Espinosa-Cantú et al., 2015). The existence of moonlighting proteins was first reported when vertebrates crystallins were found to have an enzymatic role in low levels in many tissues, in addition to a structural function in the eye lenses in high levels (Piatigorsky, 2003). Since then, an increasing number of moonlighting proteins have been identified (Mani et al., 2015) and among them enzymes very commonly have a second non-enzymatic function including functions as structural components and as gene

Figure 4 ADT2 and ADT5 localization patterns.

ADT2 and ADT5 are the only members of the ADT family in *A. thaliana* that have dual localization pattern.

A) Confocal images showing the localization patterns of ADT2-CFP and ADT5-CFP transiently expressed in *N. benthamiana* leaves.

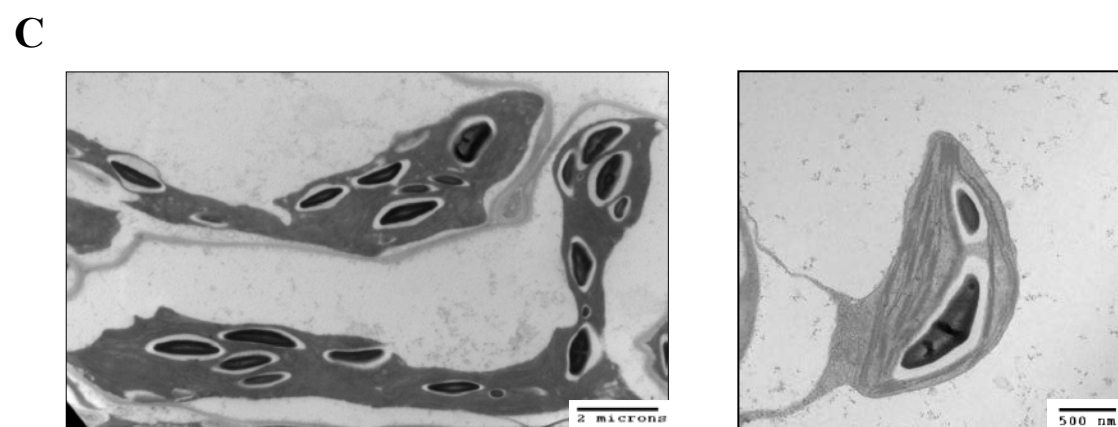
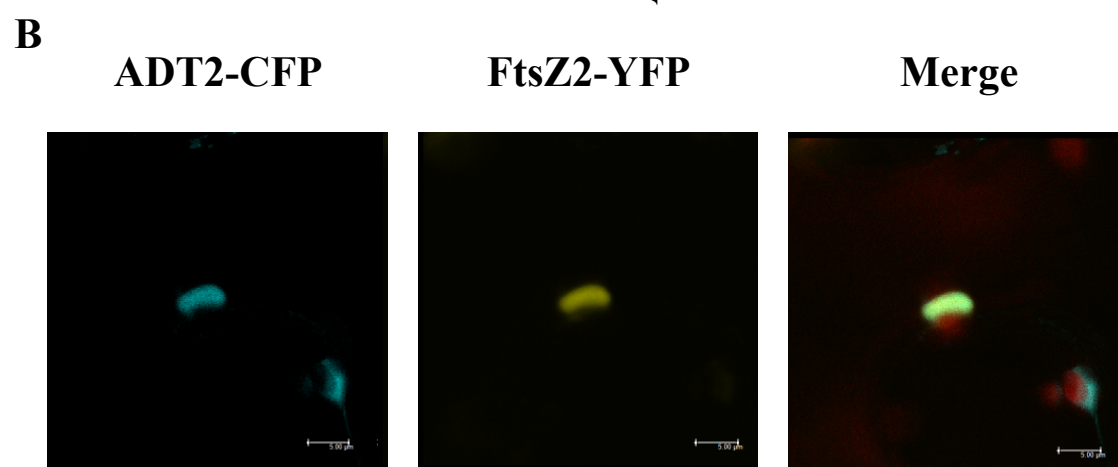
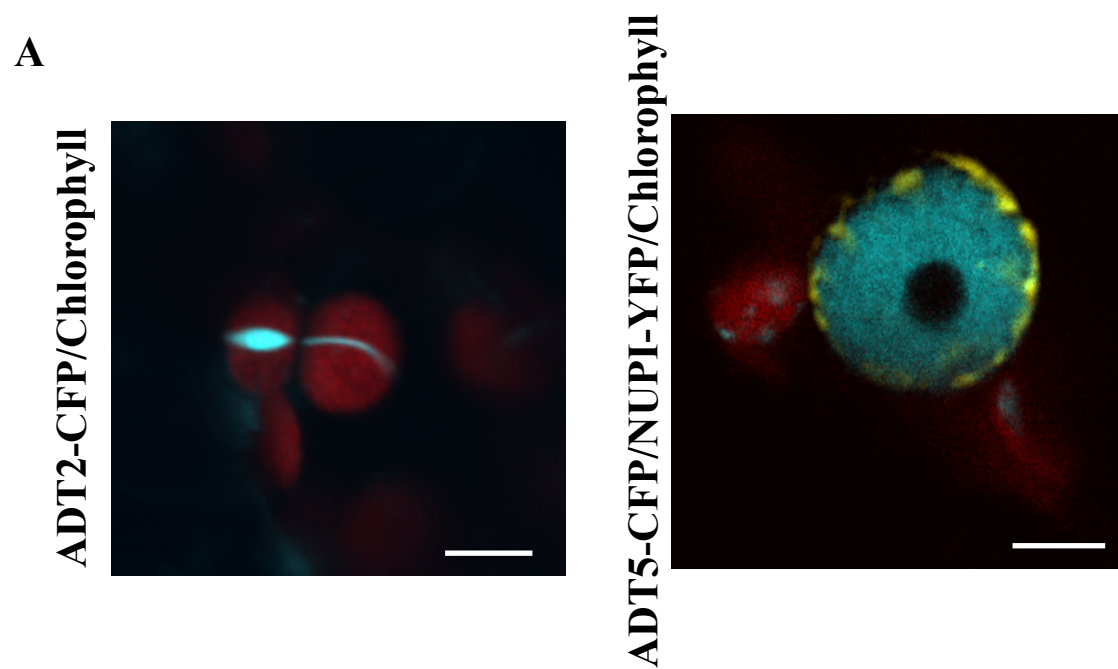
Left: ADT2-CFP localizes as a ring at the equatorial plane of chloroplasts.

Right: ADT5- CFP localizes into the nucleus and the nuclear localization was confirmed by co-expression of the nuclear membrane marker NUP1-YFP with ADT5-CFP.

B) ADT2-CFP (left) and FtsZ-YFP (middle) co-expressed and both fusion proteins were detected as a chloroplast ring (right).

C) Chloroplasts in *Arabidopsis* *adt2* mutant lines as observed by transmission electron microscopy. Long and misshapen chloroplasts were observed containing several starch granules (left) and wild type chloroplast has almond shape with one or two starch granules surrounded with thylakoids and stroma.

Scale bars are 5 μ m for confocal images (A and B), 2 μ for C (left) and 500 nm for C (right).



transcription or translation regulators. For instance, a citrate synthase in *Tetrahymena* has an enzymatic activity in mitochondria while in the cytosol it can polymerize to form 14 nm filaments which acts as a cytoskeleton protein (Kojima et al., 1997). Another example is HEXOKINASE1 (HXK1) in *A. thaliana* an enzyme in glucose metabolism in mitochondria, which in the nucleus is part of a transcription factor protein complex that is involved in transcription of glucose signaling genes (Cho et al., 2006).

Different mechanisms have been proposed for multifunctioning of moonlighting proteins such as different subcellular localizations, cell types, homodimerization of proteins or heterodimerization and interaction with other proteins, the availability of the substrates and cofactors or in many cases a combination of these parameters (Jeffery, 2016). SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION3 (STAT3) is an example of a moonlighting protein in *Mus musculus* that has two distinct functions in two different cell compartments (Wegrzyn et al., 2009). It acts as a transcription factor in the nucleus and has a second function in the electron transport chain during cell respiration in mitochondria. However, one important question about moonlighting in different cell compartments is how these proteins localize to different membrane enclosed organelles in the cell. In the following sections different possible routes for dual localization of proteins in the cell are described. The main focus will be on the dual localization of proteins in the chloroplast and nucleus.

1.6 Dual targeting of a protein in the cell

Having a single gene product dually targeted to different destinations is one mechanism to diversify protein function without increasing genome size. In dual targeting, either exactly the same protein or a shorter version of the protein can localize into different organelles. However to be able to enter to membrane-enclosed organelles in the cell specific targeting mechanism are necessary to make the dual targeting possible (Karniely and Pines, 2005). Some proteins such as dual targeted chloroplast and mitochondrial proteins have an ambiguous targeting signal that can be recognized by the import system of both organelles (Silva-Filho, 2003). One good example of these proteins is pea glutathione reductase (GR). GR is dually targeted to mitochondria and chloroplast by an N-terminal 60 amino acid signal peptide that can be recognized by both stromal

processing peptidase in the chloroplast import system and mitochondrial processing peptidase in the mitochondrial import system (Rudhe et al., 2004). Other proteins have two targeting signals and which signal is used depends on the cell physiology, the availability of receptors or the affinity of each signal for its target (Karniely and Pines, 2005). For example the type II NAD(P)H dehydrogenase (ND) protein family in *A. thaliana* has seven members and typically are located in mitochondria by an N-terminal mitochondrial targeting signal (Elhafez et al., 2006), however three members of this family have a C-terminal peroxisomal targeting signal that target them into the peroxisome (Carrie et al., 2008).

Dual targeting of proteins with either an ambiguous signal or two targeting signals is dependent on the accessibility of the signals. There is evidence that process like protein folding or interactions with other proteins can mask the targeting signals. Similarly, modifications of the protein sequence can alter the affinity of the targeting signal for its target and change the population of the protein in different compartments (Yogev and Pines, 2011). For example, Cytochrome P4501A in mammals has a N-terminal bipartite signal including ER targeting sequence followed by a mitochondrial targeting sequence. The mitochondrial localization of this protein is dependent on cleavage of the ER targeting signals by cytosolic endoprotease that activates the cryptic mitochondrial targeting signal (Addya et al., 1997).

Retrograde movement of one protein from one compartment to another in the cell is a method that has been purposed for dual localization of proteins (Yogev and Pines, 2011). This relocalization of proteins is a mechanism that allows response to changes within the cell or in environment. Although, the mechanism of retrograde movement is not fully understood, there is evidence that mitochondrial or chloroplast proteins relocalize to the nucleus suggesting a communication system between these organelles in the cell, which will be discussed more specifically for chloroplast and nucleus in the following section.

1.7 Plastid to nucleus retrograde signaling

Chloroplasts are photosynthetic organelles evolved from cyanobacteria that were engulfed by a eukaryotic cell through endosymbiosis (Dyall et al., 2004). It is estimated that the chloroplast genome of *A. thaliana* has about 100 genes, while more than 3000

proteins are recognized for chloroplast functioning, this means the majority of chloroplast proteins are encoded in the nucleus, translated in the cytoplasm and post translationally targeted to the chloroplast (McFadden, 2001; Leister, 2003). Therefore, chloroplasts and nucleus have a complex signaling system to ensure proper expression and targeting of proteins in plant cells (Bobik and Burch-Smith, 2015). For example, in response to environmental stresses, pathogen defense or chloroplast biogenesis, retrograde signals from the chloroplast communicate with the nucleus to regulate the expression of specific genes (Barajas-López et al., 2013). The first evidence of retrograde signaling and regulation of nuclear genes is based on experiments performed with the *albostrians* barley mutant (Bradbeer et al., 1979). It was shown that chloroplast-synthesized RNAs acted as signals to regulate the expression of nuclear genes encoding the photosynthetic components phosphoribulokinase and D-glyceraldehyde-3-phosphate NADP⁺ oxidoreductase (Bradbeer et al., 1979). Since then, several chloroplast metabolites have been proposed to act as retrograde signals like reactive oxygen species (ROS) (Galvez-Valdivieso and Mullineaux, 2010) and redox signals (Baier and Dietz, 2005). Recently, proteins with dual location in the chloroplast and nucleus have been suggested as retrograde signals. One such protein is PTM (PHD type transcription factor with transmembrane domain) in *A. thaliana*. PTM is associated with the outer membrane of chloroplasts and a truncated form of this protein has been detected in the nucleus (Sun et al., 2011). Several domains have been identified in the PTM protein sequence including a DNA-binding domain, a transcription activation domain and a plant homeodomain at its N-terminus and four transmembrane domains at the C-terminus (Sun et al., 2011). In response to the photooxidative stress, PTM was released to the cytosol from the chloroplast membrane through proteolytic cleavage and the N-terminal cleaved fragments containing the transcription activation domain accumulated in the nucleus (Sun et al., 2011). However, the molecular mechanism of transferring the protein from the cytosol to the nucleus is unknown. Another example is WHIRLY1 from *A. thaliana*. This protein contributes to plastid genome stability in chloroplasts (Maréchal et al., 2009) and in the nucleus acts as a transcription factor for pathogen response genes (Isemer et al., 2012). Expression of WHIRLY1 tagged with a GFP in the chloroplast genome and detection of the full-length fusion protein in the nucleus of the same cell suggested direct

translocation of the proteins from the chloroplast into the nucleus (Isemer et al., 2012). However, the mechanism of this translocation remains to be investigated.

1.8 Possible role of stromules in molecule trafficking

Stromules are highly dynamic and they can extend from the surface of the plastids like chloroplasts, amyloplasts, leucoplasts and chromoplasts (Natesan et al., 2005). They are morphologically extremely variable from tissue to tissue and at different stages of plant development (Köhler and Hanson, 2000). The primary function of stromules has not been clearly elucidated, however some possible functions have been proposed. Stromules increase the plastid surface area without increasing the plastid volume, potentially increasing the plastid capacity to translocate metabolites and macromolecules to other areas of the cell (Natesan et al., 2005). Consistent with this role, transfer of fluorescently labeled proteins between plastids have been shown to occur through stromules (Kwok and Hanson, 2004a). In addition, stromules have been reported to position closely to other organelles like mitochondria (Gunning, 2005), endoplasmic reticulum (Schattat et al., 2011), and nuclei (Kwok and Hanson, 2004b). The close association of the stromules and nuclei suggests that stromules can facilitate the transport of macromolecules and metabolites between these two organelles. Although the translocation mechanism of macromolecules through the stromules is not well understood, there are some suggestions based on experimental observations. For instance, stromules can form vesicles and this process is referred to as shedding (Gunning, 2005). It has been suggested that these vesicles can transport stromal proteins to another plastid. Alternatively, they might dispose of plastid proteins by sending them into vacuoles (Hanson and Sattarzadeh, 2011). Recent studies suggest the possibility of physical interaction between chloroplasts and nuclei specifically in response to pathogens in *A. thaliana* and *N. benthamiana* (Caplan et al., 2015). Stromule formation in these plants was induced and close connection of the stromule and nucleus was observed which correlated with accumulation of the chloroplast-localized N RECEPTOR INTERACTING PROTEIN1 (NIRP1) defense protein in the nucleus (Caplan et al., 2015). These results suggest a direct translocation of the NIRP1 from the chloroplast to the nucleus through the stromules, however the exact mechanism of nuclear entry remains unsolved.

1.9 Nuclear pore complex and nuclear import system

To enter the nucleus, proteins must cross the inner and outer membranes of the nuclear envelope (NE). The inner membrane is adjacent to the nuclear lamina and carries only proteins specific to the nucleus. The outer membrane is continuous with the endoplasmic reticulum (ER) membrane and is studded with ribosomes on its cytoplasmic side. The space between the two membranes is connected to the ER lumen (Raikhel, 1992; Boruc et al., 2012). Despite the complexity of the NE, genetic material is not totally isolated from the rest of the cell as the nuclear pore complexes (NPC) provide a gateway for exchanging mRNA and proteins between nucleus and cytoplasm (Raikhel, 1992). Water, sugars, ions and proteins smaller than 40 kDa diffuse freely into the nucleus from the cytoplasm; however, passage of macromolecules is restricted and the NPC prevents the passing of non-specific large proteins into the nucleus (Wente and Rout, 2010; Parry, 2015). Therefore, many proteins active in the nucleus have at least one targeting signal, a nuclear localization signal (NLS), which consists of one or two clusters of basic residues (Rout and Aitchison, 2001). Classical NLSs (cNLSs) are composed of basic amino acids lysine (K) and arginine (R) that are recognized by nuclear receptors, known as importins, and are well conserved in eukaryotes (Kosugi et al., 2009a). Classical NLSs are divided in two major groups: Monopartite and bipartite NLSs. Monopartite NLSs have only a single cluster of basic amino acids and can be further subdivided to class one which is a continuous sequence of at least 4 basic residues (**PKKRKKV**) and class two which is represented by one or two basic amino acids with any other amino acids in between (**PAAKRVKLD**) (Raikhel, 1992). Bipartite NLS contains two clusters of basic residues separated by a 10-12 amino acid linker for example **SPPKAVKRPAATKKAGQAKKKKLDKEDES**, the first cluster has at least two basic residues and the second cluster has at least 3 basic residues and both clusters cooperate in binding to importin proteins, with the linker probably facilitating the interaction (Raikhel, 1992). Although the putative consensus sequences of classical NLSs have been defined, there are three classes of noncanonical NLSs identified experimentally that do not match the classical NLSs (Kosugi et al., 2009a). Class three has **KRXWFXXAF** and class four has (P/R)**XXKR(K/R)** core sequence where X can be any amino acid. Class five of the noncanonical NLSs **LGKR(K/R)W** is specific to plants

and has been identified in *A. thaliana* where as the other classes are functional in mammals, yeast and plants (Kosugi et al., 2009a).

In the nuclear protein import system, importin proteins and shuttling factors including GTP-binding protein Ran mediate the translocation of cargo proteins (proteins that need to get transported into the nucleus) from the cytoplasm into the nucleus (Wente and Rout, 2010). Importin is a type of karyopherins that transports cargo proteins into the nucleus by binding to the NLS and has two subunits: importin α and importin β . Different isoforms of importin proteins have been identified in different plant species; for example, there are 8 predicted isoforms of importin α and 17 of importin β in *A. thaliana* (Merkle, 2011). Ran proteins are small soluble proteins that are known as molecular switches and are important for nuclear import and export. There is a high RanGDP concentration in the cytoplasm and a high RanGTP concentration in the nucleus which establish a gradient inside and outside of the nucleus and determines the direction of protein transport (Cook et al., 2007; Wente and Rout, 2010). Importin β has the main shuttling role in protein translocation. This protein has a cargo-binding domain that binds to the NLS of a cargo protein as well as a NPC binding domain that interacts with the NPC. Importin β also has a binding domain to bind a Ran GTP to complete the transporting complex (Cook et al., 2007). In the cytoplasm RanGDP binds importin β -cargo and this protein complex passes through the NPC. In the nucleus RanGDP disassociates importin β , which releases the cargo in the nucleus. Then RanGTP binds Importin β in the nucleus and the complex is exported to the cytoplasm through the NPC for further shuttling (Wente and Rout, 2010).

Although, importin β proteins can bind directly to cargo proteins and translocate cargo protein to the nucleus, most of the time the NLS of the cargo protein is recognized by an adaptor protein, importin α , which then binds to importin β (Riddick and Macara, 2007). The cargo-importin α -importin β complex translocate through the NPC into the nucleus by association of RanGDP to importin β . Importin α in the nucleus is exported to the cytosol by a specific protein called CAS, which also promotes the release of cargo protein from the adaptor (Riddick and Macara, 2007; Merkle, 2011). Adaptor proteins are suggested to provide more control on cargo accumulation in the nucleus under different cellular conditions and help the organism respond to environmental stimuli (Riddick and Macara, 2007).

1.10 Protein-protein interactions and nuclear targeting

Many molecular processes in the cell are mediated by molecular complexes that are composed of different numbers of proteins established by protein-protein interactions (PPI). Some protein assemblies are stable in the cell due to of their continuous role as macromolecular protein complexes, like ATP synthase. However, there are protein interactions in cells that are transient and they are only found in certain cellular contexts like specific times, specific tissues or external factors. For example activation of gene expression by interaction with transcription factors activates a gene promoter in pathogen defense (de Las Rivas and Fontanillo, 2010). In many cases protein interaction is a mechanism of gaining multifunctionality and study of PPIs has lead to reveal unknown functions of many proteins. For example and as mentioned before, a monomer of citrate synthase in *Tetrahymena* has an enzymatic activity in mitochondria, however by forming homo-oligomers it acts as a cytoskeleton protein in the cytosol (Kojima et al., 1997). In addition, PPI has been identified as a fascinating import mechanism for cargo proteins without targeting signals into different compartments of the cell, termed a “piggyback” mechanism (Thoms, 2015; Genoud et al., 2008). For example the nuclear localization of light-activated phytochrome A photoreceptor proteins (PHYA) in *A. thaliana* has been identified to depend on interaction with FAR-RED ELONGATED HYPOCOTYL1 (FHY1) in the cytoplasm (Genoud et al., 2008). FHY1 has a NLS and piggybacks PHYA into the nucleus through the NPC.

Protein interactions can be determined by small or large-scale screens with different technologies. Binary methods study the physical interactions between protein pairs and co-complex methods determine physical interactions among group of proteins (Yu et al., 2008). Yeast Two Hybrid (Y2H) and Biomolecular Fluorescence Complementation assay (BiFC) are two common binary assays that are used to study PPIs of plant proteins (de Las Rivas and Fontanillo, 2010; Kerppola, 2006). These methods were used to identify binary interactions between ADTs in *A. thaliana* and it was found that all six *At*ADTs are able to form hetero- and homodimers (Styranko, 2011). *In planta* dimerization of ADTs was detected in the chloroplast stromules, however, ADT5 homo-dimers and ADT5-containing hetero-dimers had nuclear localization (Styranko, 2011).

1.11 Research goals and objectives

ADT5 is a member of the ADT protein family in *A. thaliana* and has an enzymatic role in the decarboxylation/dehydration of arogenate to Phe in Phe biosynthesis in chloroplasts. Aside from the chloroplast and more specifically stroma localization, ADT5 was detected in the nucleus. The nuclear localization of ADT5 is intriguing and suggests a moonlighting role. However, not much is known about how ADT5 localizes to the nucleus or its potential role in the nucleus. These questions form the basis of my thesis. The specific objectives of my project are:

1. To confirm the nuclear localization of ADT5 and identify the localization of ADT5 under the control of the native ADT5 promoter.
2. To identify how ADT5 translocates into the nucleus. Three different possibilities were examined:
 - a) Direct translocation from chloroplast to the nucleus through stromules.
 - b) Cytosolic translocation of ADT5 using the nuclear import system.
 - c) Interaction with other proteins to piggyback into the nucleus.
3. To identify the possible sequence within ADT5 that makes it distinct from other ADTs and enables the nuclear localization of this protein.

2 Material and Methods

2.1 Media, solutions and media additives

2.1.1 Media

For preparing solid media 15 g agar was added to 1 liter of liquid media except for Murashige and Skoog (MS) medium and Revised Medium for Organogenesis of *Nicotiana plumbaginifolia* (RMOP) medium that 7 g agar was used.

Lysogen broth (LB)

For 1 L: 10 g tryptone, 5 g yeast extract, 10 g NaCl.

Murashige and Skoog (MS) medium

For 1 L: 4.43 g Murashige and Skoog nutrient mix (Sigma M 5519), 40 g sucrose, pH 5.8.

Regeneration medium

For 1 L: RMOP, 250 µl BAP, 26 µl NAA, 100 µl thiamine, 2 ml myo-inositol.

RMOP medium

For 1 L: 4.43 g Murashige and Skoog nutrient mix pH 5.8, 30 g sucrose, pH 5.8.

Rooting medium

For 1 L: RMOP, 5 ml spectinomycin (50 mg/ml).

Synthetic Dextrose (SD) medium

For 1 L: 20 g glucose, 6.7 g yeast nitrogen base, 1.5 g appropriate drop out powder.

Yeast Extract and Beef (YEB) medium

For 1 L: 5 g beef extract, 1 g yeast extract, 5 g peptone, 5 g sucrose, 0.49 g MgSO₄.

Yeast Peptone Dextrose Adenine (YPDA) medium

For 1 L: 20 g glucose, 10 g yeast extract, 20 g peptone, 40 mg adenine hemisulfate.

2.1.2 Solutions and buffers

Cellulose solution

For 50 ml: 0.75 g cellulose R10 (Yakult Pharmaceutical, 216016), 0.2 g macerozyme[®] R10 (Yakult Pharmaceutical, 202051), 3.65 M mannitol, 1 ml 1 M KCl, 1

ml 1 M morpholino ethanesulfonic acid (MES) (Sigma M8250) pH 5.7, 0.5 ml 1 M CaCl_2 , 17.5 μl β -mercapto ethanol (BME) (Sigma M6250) and 0.5 ml bovin serum albumin (BSA) (Sigma A6793).

DNA coating mixture

For 5 shots: 50 μl 0.6 μm gold particles (Bio-Rad 165-2262), 50 μl 2.5 M CaCl_2 , 20 μl 0.1 M spermidine (Sigma S2626).

Fertilizer solution

For 1 L: 0.25 g water-soluble fertilizer N: P: K=20: 8: 20 (Plant Products, Brampton, ON, Canada).

Floral dip solution

For 500 ml: 25 g sucrose, 100 μl Silwet L-77 (Plant Media 30630216-2).

Flotation medium

For 50 ml: 10 g sucrose, 0.5 ml 1 M MES (Sigma M8250) pH 5.7 and 1 ml 1 M KCl.

Gamborg's solution

For 1 L: 3.2 g Gamborg B5 medium and vitamins (RPI 20200), 20 g sucrose, 10 ml 1 M MES (Sigma M8250) and 1 ml 200 μM 4'-hydroxy 3', 5'-dimethoxyacetophenone (acetosyringone) (Sigma D134406).

Hexadecyl Trimethyl Ammonium Bromide (CTAB) extraction buffer

For 500 ml: 10 g CTAB (Sigma 855820), 20 ml 0.5 M Ethylene Diamine Tetracetic Acid (EDTA) pH 8, 140 ml 5 M NaCl, 50 ml 1 M Tris-HCl pH 8.

Phosphate Buffered Saline (PBS)

For 1 L: 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4 , 0.24 g KH_2PO_4 , pH 7.4.

Protein Extraction Buffer (PEB)

For 100 mL: 96 ml PBS, 100 μl Tween-20, 2 g poly vinyl poly pyrrolidone (PVPP), 200 μl 0.5 mM EDTA pH 8, 1 ml 100 mM phenylmethanesulfonylfluoride (PMSF), 100 μl 1 mg/ml leupeptin, 2 g 100 mM L-ascorbate.

Seed sterilization solution

For 100 ml: 20 ml 5% bleach, 100 μl 2% sodium dodecyl sulfate (SDS).

Washing solution

For 50 ml: 18.2 g mannitol, 0.8 ml 1M MES (pH 5.7), 4 ml 1 M KCl.

X-Gluc solution

For 100 ml: 1 ml 100 mM 5-bromo-4-chloro-3-indolyl-glucuronide (X-Gluc) (Thermo Scientific™ B1691), 20 ml methanol, pH 7.0.

2.1.3 Media additives

Acetosyringone

For 200 mM stock: 196.2 mg acetosyringone (Sigma D134406) dissolved in 5 ml dimethyl sulfoxide (DMSO). Final concentration of 200 μ M was used.

Amino acid dropout powder

For 25 L: 1 g adenine hemisulphate, 0.5 g L-arginine-HCL, 2.5 g L-aspartic acid, 2.5 g L-glutamic acid, 0.75 g L-isoleucine, 1.5 g L-leucine, 0.75 g L-lysine-HCL, 0.5 g L-methionine, 1.25 g phenylalanine, 9.35 g L-serine, 5 g L-threonine, 1 g L-tryptophan, 0.75 g L-tyrosine, 0.5 g uracil, 3.72 g L-valine, 2 g L-histidine HCL.

Antibiotics

Stock solution of ampicillin (100 mg/ml), gentamycin (50 mg/ml), kanamycin (60 mg/ml), spectinomycin (50 mg/ml), streptomycin (50 mg/ml) was prepared by dissolving each in MiliQ water. Aureobasidin A (500 μ g/ml) and rifampicin (25 mg/ml) dissolved in absolute ethanol and dissolved in dimethyl sulfoxide (DMSO), respectively. Stocks were added to the media after autoclaving to a final concentration of 100 μ g/ml, 200 ng/ml, 15 μ g/ml, 60 μ g/ml, 10 μ g/ml, 50 μ g /ml and 50 μ g /ml for ampicillin, aureobasidin A, gentamycin, kanamycin, rifampicin, spectinomycin and streptomycin, respectively.

Benzyl Amino Purine (BAP)

For 4 g/L stock: 0.2 g BAP (Sigma B3408) dissolved in 1 ml 1 N NaOH and mix with 49 ml MiliQ water. Final concentration of 1 mg/L was used.

Myo- inositol

For 50 g/L stock: 2.5 g myo-inositol (Sigma I5125) dissolved in 50 ml MiliQ water. Final concentration of 0.1 g/L was used.

1-Naphthalene Acetic Acid (NAA)

For 4 g/L stock: 0.2 g NAA (Sigma 317918) dissolved in 1 ml 1 N NaOH and mix with 49 ml MiliQ water. Final concentration of 0.1 mg/L was used.

Thiamine

For 10 mg/L stock: 0.5 g thiamine hydrochloride (Sigma 47858) dissolved in 50 ml MiliQ water. Final concentration of 1 mg/L was used.

X- α -Gal

For 20 mg/ml stock: 100 mg X- α -Gal (Clontech 630407) dissolved in 5 ml dimethyl formamide. Final concentration of 20 μ g/ml was used.

2.2 Organisms and growth conditions

2.2.1 Bacterial strains and growth conditions

Escherichia coli DH5 α (Invitrogen 11319019) was used for maintenance of plasmids. For liquid cultures cells were grown in LB media at 250 rpm at 37°C for 16-18 hours. Appropriate antibiotic was added to media for selection of cells harboring plasmids.

Agrobacterium tumefaciens LBA4404 (NCCB accession PC2760) and GV3101 (kindly provided by Dr. Grbic, Western University, London, ON; Hellens and Mullineaux, 2000) were used for plant transformations. *A. tumefaciens* liquid cultures were grown at 250 rpm at 28°C for 16-48 hours in YEB or LB medium supplemented with appropriate antibiotics.

2.2.2 Plant material and growth conditions

All plants were grown in growth chambers at 22°C with a 16/8 hours light/dark photoperiod at a light density of 110 μ mol⁻²s⁻¹ unless specified otherwise.

A. thaliana wild type (Columbia-0; TAIR accession number CS907) seeds were used in this study. *A. thaliana* seeds were first germinated on MS media and then transferred onto soil. Prior to this seeds were sterilized using sterilization solution and shaken gently for 10 minutes followed with 5 washes with MiliQ water to remove the sterilization solution. Then seeds were transferred onto MS medium and vernalized at 4°C in the dark for 3 days and then transferred into a growth chamber for 2 weeks. If there was a need of selection, appropriate selectable components were added to the media. After 2 weeks germinated seedlings were transferred to soil and grown in the growth chamber. To maintain the humidity pots with seedlings were covered with clear

covers in the growth chamber for the first 2 days and after that they exposed to the environment.

Nicotiana benthamiana and *Nicotiana tabacum* (Male sterile 81V9) seeds were kindly provided by Dr. Menassa, Agriculture and AgriFood Canada, London, ON.

N. benthamiana seeds were sown on soil and grown in a growth chamber for 4-6 weeks. Plants were watered with water-soluble fertilizer every other day.

N. tabacum seeds were first sterilized with 70% ethanol and rinsed twice with MiliQ water. Then they were grown on MS medium in sterile Magenta boxes. Plants were grown for 4 weeks in a growth chamber and used for chloroplast genome bombardment. Transplastomic tobacco plants were potted and transferred to the green house with a 16/8 hours light/dark photoperiod. Green house temperature was 26°C at day and 18°C at night.

2.2.3 Yeast strains and growth conditions

Saccharomyces cerevisiae AH109 (Clontech 630444; *MAT α* , *trp1-901*, *leu2-3 112*, *ura3-52*, *his3-200*, *gal4 Δ* , *gal80 Δ* , *LYS2::GAL1-UAS-GAL1_{TATA}-HIS*, *GAL2-UAS-GAL2_{TATA}-ADE2*, *URA3::MEL1-UAS-MEL1_{TATA}-lacZ*, *MEL1*) and Y187 (Clontech 630457; *MAT α* , *ura3-52*, *his3-200*, *ade2-101*, *trp1-901*, *leu2-3 112*, *gal4 Δ* , *gal80 Δ* , *MEL1*, *URA3::GAL1_{UAS}-gal1_{TATA}-lacZ*) were used. Yeast cells were grown on appropriate YPD or supplemented SD medium at 30°C and liquid cultures were shaken at 220 rpm unless specified otherwise.

2.3 Cloning procedures and construct design

2.3.1 Primer design

Gateway primers were designed (Appendix 1) to amplify full length and truncated *ADT5* (At5g22630), *ADT4* (At3g44720) and *ADT2* (At3g07630) sequences as well as full length *IMPA6* (At1g02690) and *PDAT* (At5g13640) based on the coding sequences available on TAIR. All forward primers were designed with 5' *att* sites and the reverse primers were designed with 3' *att* sites. For cloning of the native promoter into the pCB5, the forward primer was designed with 5' *MauBI* and reverse primer was designed with 3' *XhoI* restriction site using the genomic sequence of the *ADT5* available on TAIR. All primers for generating the transplastomic plants were designed to amplify full length and

truncated *ADT5*, *ADT4*, *ADT2* sequences and added 5' and 3' *NheI* restriction sites. In addition, some internal gene specific primers were designed for sequencing purposes. All primers were tested for melting temperature, balanced GC content, self- complementary and primer pair complementary using DNAMAN (Lynnon BioSoft, Version 6).

2.3.2 Polymerase chain reaction conditions

A standard PCR condition (denaturation of the template 1 minute at 95°C, annealing of the primers at 55°C, extension 3 minutes at 72°C, 25-30 cycles of amplification) was used. All PCR reactions were performed using *Taq* polymerase (New England BioLabs®_{Inc} M0273) to establish the conditions prior to using Phusion®_{Inc} high-fidelity DNA polymerase (New England BioLabs®_{Inc} M0530S). For both polymerases, reactions were set up according to the recommended company protocols. Previously cloned *ADT* sequences (Cho et al., 2007) were used for cloning *ADT* sequences and an *A. thaliana* cDNA library from 14 days old seedlings (kindly provided by Dr. Cui) was used as template for cloning of *IMPA6* and *PDAT* sequences.

2.3.3 Plasmid DNA isolation and sequencing

Plasmid DNA was isolated from *E. coli* using the QIAprep Spin Miniprep Kit (QIAGEN 12123) following the manufacture's protocol. The same kit was used for isolating plasmid DNA from yeast cells with a small modification. At the very first step 425-600 micron glass beads, acid washed (Sigma G8772) were added to the samples to break down the cell wall and release plasmid DNA. For sequence confirmation of the constructs, all the samples were sent to the sequencing laboratory at London Regional Genomic Center at the Robarts Research Institute (Western University). Sequence results were analyzed using the DNAMAN (Lynnon BioSoft, Version 6).

2.3.4 Cloning systems and plasmid strains

Two cloning systems were used in this project: Gateway® and ligation cloning. All the plasmids used in this study are listed in Appendix 2. In the Gateway® system (Hartley et al., 2000) PCR fragments were introduced into pDonor221 vector (Invitrogen 1236017) using a BP Clonase (Invitrogen 11789020) and BP recombination reaction. Resulting entry vectors were sequenced to confirm that no PCR errors were introduced

and sequences were in frame. The confirmed fragments were recombined into a destination vectors using LR Clonase (Invitrogen 11791020) to generate expression vectors.

Different Gateway[®] destination vectors were used in this project based on the purpose of the experiments (Appendix 2): To determine subcellular localization plant expression vectors pEarlyGate 101 and pEarlyGate 102 (Earley et al., 2006) and for promoter analysis pKGWFS7 (Karimi et al., 2002) were used. The Gateway[®] modified compatible expression vectors pGBKT7-DEST (GAL4-DB) and pGADT7-DEST (GAL4-TA) were used for Y2H experiments (Clontech 630489 and 630490 modified by Lu et al. 2010). The Gateway[®] modified compatible expression vectors pEarleyGate201-YN, pEarleyGate 202-YC (Earley et al., 2006; Lu et al., 2010) were used for BiFC assays. Restriction enzyme digestion reactions were used for confirmation of proper insertion.

For ligation-dependent cloning, PCR amplified DNA fragments and desired vectors were first digested with appropriate restriction enzymes and fragments separated on 0.7 % (w/v) agarose gel. Then digested products were visualized and the desired digested PCR fragments and digested vectors were removed and gel purified using QIAquick Gel Extraction Kit (QIAGEN 28704). Digested PCR fragments were cloned into vectors using T4 DNA ligase (Thermo Scientific[™] EL0014) and sequenced to confirm that no PCR errors were introduced and sequences were in frame. The pCEC4 vector (Kolotilin et al., 2013) was used for chloroplast genome bombardment and pCB-ADT5 (Bross et al., 2017) was used for expression of *ADT5* with its native promoter (Appendix 2).

2.4 Bacterial and yeast transformations

E. coli chemical competent cells were prepared as described (Renzette, 2011) and transformed using the Chemically Competent *E. coli* transformation protocol (Invitrogen C4040-03 Manual). *A. tumefaciens* electroporation competent cells were prepared and transformed using a Gene Pulser II system (Bio-Rad) as described (Wise et al., 2006) with the following settings: 2 kV, 400 Ω , 25 μ F. Immediately after transformation *E. coli* and *A. tumefaciens* cells recovered at 37°C and 30°C, respectively, for 1 hour in non-selective medium before transferring on selective medium.

Yeast competent cells were prepared using the Fast Yeast Transformation™ kit (G-Biosciences GZ-1) and used for transformation according to the manufacture's instructions. To increase the efficiency of transformation, cells were incubated at 30°C for 3 hours before plating on selective medium.

2.5 Plant transformations

2.5.1 Transient transformation of *N. benthamiana*

Agrobacterium cells harboring desired constructs were grown to an optical density of 0.8 (OD₆₀₀). Cells were collected by centrifugation at 1000 g for 30 minutes and resuspended in Gamborg's solution to the final OD₆₀₀ of 1 and incubated with gentle shaking for 1 hour at 21°C. To increase the expression of the transgenes, *Agrobacterium* cells harboring p19 at OD₆₀₀ of 1 was used in all transient expression experiments. The p19 vector encodes a 19 kDa protein from *Cymbidium ringspot* virus (CymRSV) that suppresses post transcriptional gene silencing (Silhavy et al., 2002; Voinnet et al., 2003). Therefore, *Agrobacterium* cells harboring the desired construct and *Agrobacterium* cells harboring p19 in a 1:1 ratio suspension were used for infiltration of the abaxial leaf epidermis of 4 weeks old *N. benthamiana* plants using a 1 ml syringe (Kapila et al., 1997). The fluorescence signals were detected by confocal microscopy 4-5 days after infiltration (dpi).

2.5.2 Chloroplast genome transformation of *N. tabacum*

A biolistic method was used to generate transplastomic plants (Verma et al., 2008). Briefly, 5 µg of desired plasmid DNA was gold-coated using the coating mixture at room temperature. The coated plasmids were washed twice with 70% ethanol and following a wash with 100% ethanol. 10 µl of the coated DNA was used per transformation. 8-10 week old plants were used for bombardment with Bio-Rad PDS-1000/ HeTM Biolistic Delivery system (Bio-Rad 165-2257) under sterile conditions. For each constructs, 10 leaves abaxial side up were bombarded.

Bombarded leaves were kept in the dark for 2 days at 22°C on RMOP plates. Then the leaves were cut into 5 mm² pieces and placed on RMOP medium containing spectinomycin for selection of cells containing transformed chloroplasts. Leaves were incubated in a growth chamber for 6 weeks and were inspected weekly for the appearance

of green calli. Positive transformed calli were identified by PCR and regenerated for 2 generations on regeneration medium. The regenerated plantlets were rooted on rooting medium and after 4 weeks were potted and transferred to the green house. Young plantlets were covered with clear covers for 2 days after transferring to the green house to help them to adjust.

2.5.3 Stable transformations of *A. thaliana*

Agrobacterium cells harboring desired constructs were transferred to the wild type *A. thaliana* plants at flowering stage using the floral dip method (Zhang et al., 2006). Briefly, *Agrobacterium* cells were grown to an OD₆₀₀ of 1.5-2 and were collected by centrifugation at 4000 g for 10 minutes at room temperature. Cells were gently resuspended in dipping solution. Then the aerial parts of the *A. thaliana* plants were dipped into the cell suspension for 10 seconds. Transformed plants were covered with clear plastic bags for 1 day in the room temperature in the dark and then transferred into the growth chamber for one month until dry seeds were collected. Seeds were placed on MS media containing kanamycin and transgenic plants were selected after two weeks. Germinated transformed seedlings were transferred to soil and allowed to grow to maturity and seed collection.

2.6 Plant genomic DNA isolation

To extract the genomic DNA from *A. thaliana* or *N. tabacum* leaves the Fast Plant Genomic DNA isolation method was used (Murray and Thompson, 1980). Briefly, 30 mg leaf material were placed in a 2 ml Eppendorf tube and dipped in liquid nitrogen. Frozen tissue samples were grounded using pestles and mixed with 300 µl of CTAB extraction buffer by vortexing. Then 300 µl chloroform was added to the tube and mixed well by inverting several times. Samples were centrifuged at 10.000 g for 10 minutes and the upper aqueous phase was transferred to a fresh tube. 300 µl isopropanol was added to the solution and mixed well by inverting the tube several times and centrifuged at 12.000 g for 10 minutes. The supernatant was removed and the pellet was washed with 300 µl 70% room temperature ethanol. Pellet was dissolved with 50 µl MiliQ water.

2.7 Bimolecular Fluorescence Complementation Assay

The Bimolecular Fluorescence Complementation assay (BiFC) was used to test for protein-protein interactions *in planta* (Kerppola, 2006). Desired genes were recombined into modified pEarlyGate 201-nYFP and pEarlyGate 202-cYFP (section 2.3.4) where each has a half N-terminal and C-terminal fragments of YFP (Lu et al., 2010). Constructs were transformed into *Agrobacterium* cells and were transiently co-expressed in 4 weeks old *N. benthamiana* leaves. Infiltrated leaf samples were inspected 3 dpi by confocal microscopy for detection of the YFP signal. Interaction of the fused proteins would bring the two half of the yellow fluorescence protein within proximity and emit its fluorescence signal (Kerppola, 2006).

2.8 Confocal microscopy

To visualize the fluorescent fusion proteins, the abaxial epidermal cells of the leaf disc (9 mm) were imaged using Lecia SP2 and Olympus FV3000 confocal laser scanning microscope equipped with 63x water immersion objective lens. Chlorophyll and CFP were imaged with a 405 blue diode laser and emission was detected from 440 nm to 485 nm and from 630 nm to 690 nm, respectively. For imaging GFP a 488 nm argon laser was used and emission was detected from 500 nm to 525 nm and YFP was excited with a 514 nm argon laser and its emission was collected from 540 nm to 550 nm. For colocalization experiments, CFP and YFP emissions were detected sequentially to avoid crosstalk of fluorophore pair (Shaner et al., 2005). Chlorophyll II, YFP, CFP and GFP fluorescence was colored red, yellow, cyan and green, respectively.

2.9 Y2H assay and cDNA library screening

Y2H assays (Chien et al., 1991) were used to detect and analyze protein-protein interactions *in vivo*. Desired genes were cloned into pGADT7-DEST (has an N-terminal GAL4-AD domain) and pDGBKT7-DEST (has an N-terminal GAL4-BD) (Lu et al., 2010) and transformed pairwise into yeast AH109 cells using Fast Yeast Transformation™ kit (G-Biosciences GZ-1). Transformed cells were grown to a cell density of 5×10^6 cells/ml and then spotted on appropriately supplemented SD medium. Results were recorded 3-7 days after spotting. Three reporter genes, *ADE2*, *HIS3* and

MEL1, at three different chromosomal locations in yeast cells AH109 were used. *ADE2* and *HIS3* are genes required in yeast for adenine and histidine biosynthesis and *MEL1* encodes for an enzyme called α -galactosidase that if expressed in the presence of X- α -Gal produces a color change in the colonies. Growth of white cells in the absence of histidine and adenine in the medium and production of a blue pigment in the presence of X- α -gal indicates activation of the reporter genes and interaction.

To identify potential interactors for ADT5 a commercial *A. thaliana* Y2H cDNA library (Mate and PlateTM Library- Universal *Arabidopsis*, Clontech 630487) was used following the recommended protocol (Clontech, 2010). This commercial library has been normalized, hence the representation of highly expressed genes like housekeeping genes has been reduced which increase the possibility of interactions with proteins with low expression level (Clontech, 2010).

Briefly, AH109 cells harboring pGBKT7- ADT5 constructs were grown about 20 hours to an OD₆₀₀ of 0.8. Then, cells were combined with Mate and PlateTM Library yeast Y187 cells (Clontech 630487) in a 2 L flask with 45 ml 2x YPDA and incubated for 24 hours slowly shaking at 30 rpm to induce mating. Cells were evaluated for the presence of diploid zygotes (three-lobed zygotes) using hemocytometer and a light microscope. Once zygotes were detected, cells were centrifuged at 1000 rpm for 10 minutes and the pellet was re-suspended in 2xYPDA and plated on appropriate SD medium. Plates were incubated at room temperature for 5 days. The three reporter genes (*ADE2*, *HIS3* and *MEL1*) in addition of antibiotic resistance reporter gene *AURI* recommended by the manufacture for screening of Y2H libraries were used to identify colonies expressing a putative interactor. *AURI* encodes an enzyme called inositol phosphorylceramide synthase and enables the yeast cells to grow in presence of antibiotic aureobasidin A.

2.10 Protoplast preparation

To take better images of chloroplasts, protoplasts were isolated as described (Sheen, 2011) and analyzed by microscopy. Briefly, 1 g of leaf tissue was cut into 0.5-1 mm strips and submerged in 5 ml cellulose solution; vacuum infiltrated for 30 minutes and incubated overnight in the dark at room temperature to digest the cell wall. Protoplasts were released by shaking at 80 rpm for 5 minutes. The solution with protoplasts was

filtered through 100 μm mesh filters and then centrifuged for 10 minutes at 620 rpm at 4°C. The supernatant was removed and protoplasts were mixed with an equal volume of ice cold washing solution. The suspension was centrifuged for 10 minutes at 620 rpm at 4°C and the supernatant was removed protoplasts were resuspended in 2.5 ml washing solution and transferred gently to the top of flotation medium and centrifuged for another 10 minutes at 620 rpm at 4°C. Protoplasts floated on the top layer of the flotation medium and were transferred gently to ice cold washing solution. 300 μl of the suspension loaded on a cavity well microscopy slide (VWR 10118-600) and was visualized by confocal microscopy.

2.11 Protein isolation and Western blotting

For protein isolation three leaf discs (9 mm) were collected. Total soluble protein was extracted and quantified as described (Conley et al., 2009). Briefly, tissue samples were flash frozen with liquid nitrogen and homogenized with TissueLyser (Qiagen). Then 200 μl PEB was added to the samples. Samples were centrifuged at 16,000 rpm 4°C for 10 minutes and the supernatant was transferred to a fresh tube. The total soluble protein (TSP) concentration was quantified with a Bradford assay (Bradford, 1976). 10 μg of TSP from each sample was size separated using Mini-Protean® TGXTM Precast 4-20% (w/v) polyacrylamide gradient gel (Bio-Rad 456-1091).

The fusion proteins were probed with a primary anti-GFP antibody (1:5000 dilution; Clontech 632380) that designed to recognize GFP and GFP derivatives including CFP. Then fusion proteins probed subsequently with a secondary goat anti-mouse IgG (H+L) horseradish peroxidase conjugated (1:3000 dilution; Bio-Rad 170-6516). Proteins were visualized with the enhanced chemiluminescence (ECL) detection kit according to the manufacture instruction (GE Healthcare, Mississauga, ON, Canada).

2.12 Histochemical GUS staining

To stain the plant tissues for GUS activity, GUS staining was performed as described (Crone et al., 2001). Seedlings or sections of plant tissues were immersed in 1 ml of fresh GUS solution and vacuum infiltrated for 15 minutes. Samples were incubated at 37°C overnight. Next day the staining solution was removed and samples were washed with 75% ethanol and incubated at 37°C for 1 hour several times until tissues cleared and all

pigments removed. GUS stained tissues were imaged by microscope Nikon Ecolipse E 800 (Champigny sur Marne, France).

2.13 Sequence analysis

Sequences for *A. thaliana* and different organisms were obtained from TAIR and NCBI, respectively. Multiple sequence alignments were performed with DNAMAN using an optimal sequence alignment and a BLOSUM protein weight matrix (Henikoff and Henikoff, 1992). The phylogenetic trees were generated using the Neighbor Joining method (DNAMAN Version 6; Saitou and Nei, 1987).

3 Results

This study focuses on discovering the routes that account for ADT5 dual localization *in planta*.

3.1 Nuclear localization of ADT5

Subcellular localization studies of ADT-CFPs *in planta* demonstrated a nuclear localization for ADT5 in addition to the chloroplast stroma localization typically seen for all ADTs (Bross, 2011). However, some reports demonstrated that fluorescent tags, such as GFP or CFP, can get separated from fusion recombinant proteins in localization studies (Francin-Allami et al., 2011; Kaldis et al., 2013). Since the cleaved fluorescent products are small (25-27 kDa), they can easily diffuse into the nucleus (Wente and Rout, 2010). To test if the fluorescence detected in the nucleus was due to the localization of ADT5-CFP or a CFP cleavage product, further experiments were necessary. Therefore, *Agrobacterium* cells harboring ADT-CFP constructs (pCB-ADTs) (Bross, 2011) were transiently expressed in *N. benthamiana* leaves. In each of the pCB-ADT plasmids, CFP is fused to the C-terminus of one of the six ADTs and the CaMV 35S promoter controls the expression of the fusion constructs. Since all transformations are performed in the presence of an *Agrobacterium* strain containing p19 (section 2.5.1), leaves also were transformed with this strain alone as a negative control. Furthermore, an *Agrobacterium* strain carrying a GFP plasmid was used to see the expression pattern of this fluorescent protein alone.

Leaf samples were visualized at 4 dpi using confocal microscopy. In control infiltrated leaves expressing p19 no fluorescence was detected (Figure 5A) and GFP alone was present in the cytosol and nucleus (Figure 5B). The subcellular localization pattern of ADTs was as expected based on previous observations (Figure 6). With the exception of ADT6 that had a cytosolic localization pattern (Figure 6F) all ADTs were detected in the chloroplast and more specifically the stroma of chloroplasts (Figure 6A-E). ADT5 was the only ADT observed in the nucleus (Figure 6E). Next, to determine that the CFP signal in the nucleus is due to the localization of ADT5-CFP and not the cleaved CFP, proteins were isolated from the same infiltrated leaf tissues used for confocal microscopy to perform a Western blot.

Figure 5 Expression pattern observed in control transformations in *N. benthamiana*.

N. benthamiana leaves were infiltrated either with *Agrobacterium* cells harboring p19 or harboring GFP plasmid alone to ensure that p19 does not produce fluorescent signals and to see the expression pattern of the fluorescent protein, respectively.

Chlorophyll fluorescence and fluorescent proteins are shown separately in the left and middle column, respectively, and a merged image is shown in the right column. Summary of the observations are shown on the right side of the panel.

White arrows: Nucleus.

C: Cytoplasm, N: Nucleus.

Scale bars are 5 μm .

(A) No signal was detected in p19 infiltrated leaves.

(B) GFP alone was detected in the cytosol and nucleus.

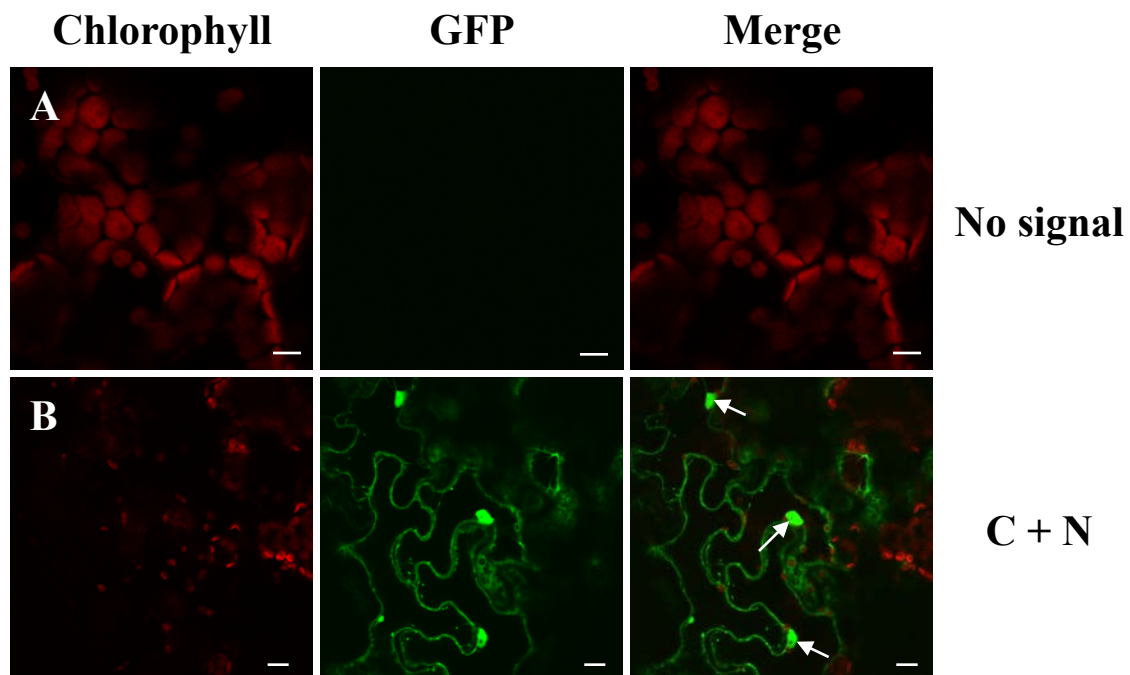


Figure 6 Subcellular localization of ADT-CFPs in *N. benthamiana* leaves.

To ensure the nuclear localization pattern is specific to ADT5, all six ADTs were transiently expressed in *N. benthamiana* leaves and the localization of the fusion proteins was visualized by confocal microscopy at 4 dpi.

Chlorophyll fluorescence and CFP fluorescence are shown separately in the left and middle column, respectively, and the merged image is shown in the right column. A summary of the observations is shown on the right side of the panel.

White arrow: Nucleus.

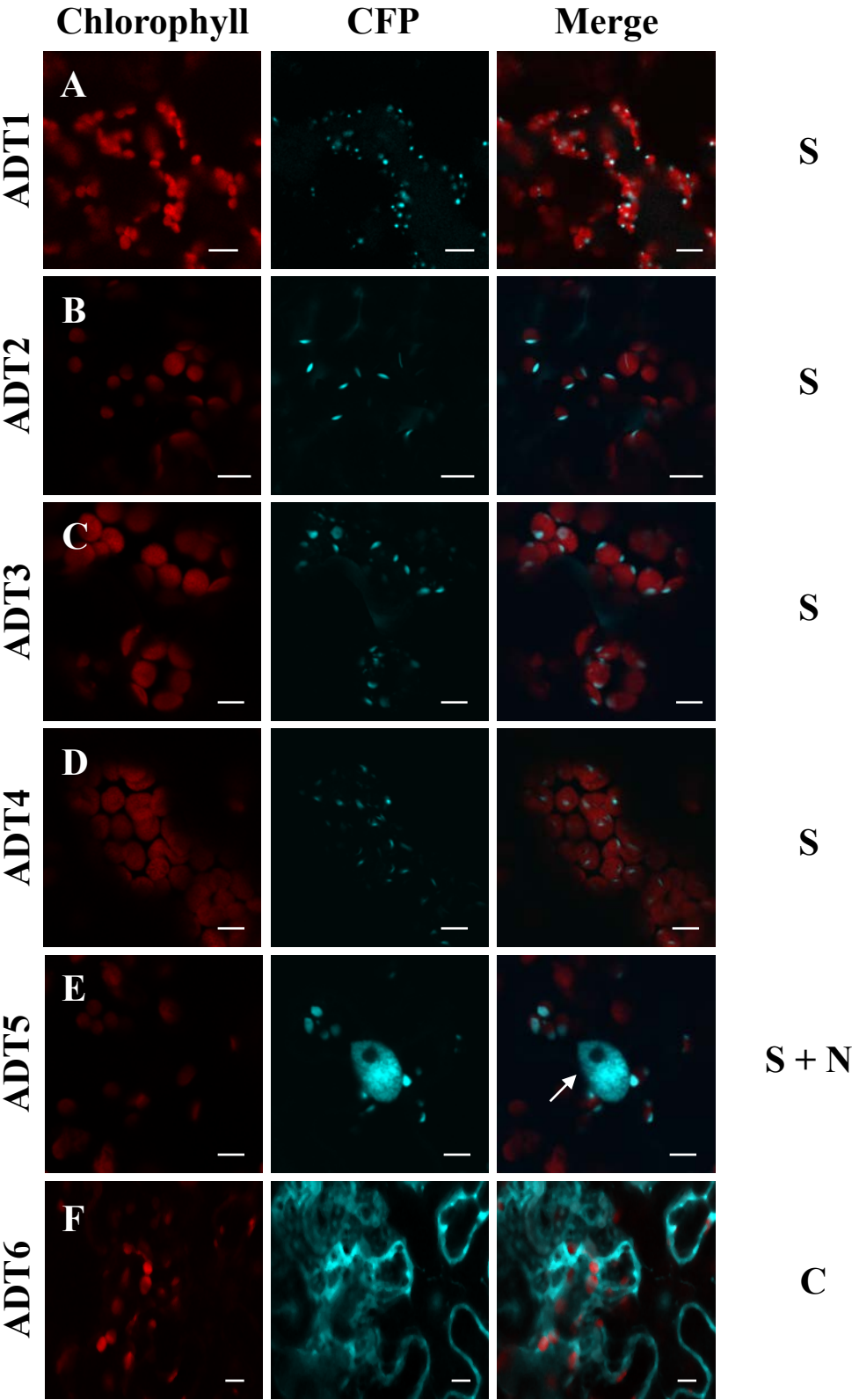
C: Cytoplasm, N: Nucleus, S: Stromule.

Scale bars are 5 μm .

(A-D) ADT1, ADT2, ADT3 and ADT4 were detected in stromules.

(E) ADT5 was detected in stromules and nucleus.

(F) ADT6 was found mostly in the cytosol.



Using an anti-GFP antibody no band was detected in protein samples from *N. benthamiana* leaves expressing p19 only and a single band of 25 kDa in GFP samples as expected for the GFP (Figure 7). For all the ADT-CFP fusion proteins a distinct band of predicted size of approximately 75 kDa was detected (see figure legend for details). In addition to these bands, additional bands of higher molecular weight can be seen for all fusion proteins. These bands are consistent in size for ADT-CFP dimers (150 kDa), tetramers (300 kDa) and higher multimers. A faint band at the size of GFP (25 kDa) was detected for some of the ADTs consistent with the size of a cleaved CFP. However, no nuclear localization was observed for any of the ADTs except ADT5. Since CFP alone is small in molecular size and can diffuse to the nucleus, these results indicate that the level of cleaved CFP products for ADTs is not sufficient to accumulate in the nucleus and account for the signal detected by confocal microscopy. Hence, the observed CFP signal in the nucleus is indeed due to the localization of the ADT5-CFP.

3.2 Nuclear localization of ADT5 using the native promoter

Up to this point, all the *ADT-CFP* constructs were expressed using the CaMV 35S promoter. Therefore, it was suggested that the nuclear localization of ADT5 might have been due to overexpression of the recombinant protein by the constitutive CaMV 35S promoter (So et al., 2005; Kay et al., 1987). To ensure that ADT5 nuclear localization is not an overexpression artifact, it was expressed under the control of its native promoter (*proNat5*). Sequences 1 kb upstream of the *ADT5* start codon were PCR-amplified using *A. thaliana* genomic DNA as a template. This fragment was recombined into a pKGWFS7 plant destination vector using the Gateway system (section 2.3.4) and the resulting construct was named *proNat5:GFP:GUS*. In this vector, the *proNat5* sequences regulate the expression of the two reporter genes *GFP* and *GUS*.

The *proNat5:GFP:GUS* construct was introduced into *Agrobacterium* GV3101 by electroporation. *Agrobacterium* cells containing *proNat5:GFP:GUS* were infiltrated into *N. benthamiana* leaves. The activity of *proNat5* was identified by detecting a cytosolic GFP expression pattern in the infiltrated leaf samples by confocal microscopy at 5 dpi (Figure 8A).

Figure 7 Protein expression of ADT-CFPs in *N. benthamiana*.

Western blot showing the recombinant ADT-CFP proteins expressed in *N. benthamiana* leaves. Ten µg of total soluble protein was size-separated on a SDS-PAGE, blotted with nitrocellulose membrane and detected with a GFP antibody. Predicted protein size for fusions ADT-CFP are ADT1: 71.5 kDa, ADT2: 70.0 kDa, ADT3: 74 kDa, ADT4: 73.8 kDa, ADT5: 73.8 kDa, ADT6: 72.7 kDa. Larger bands represent ADT dimers (150 kDa), tetramers (300 kDa) and higher multimers. As negative and positive controls, protein samples from P19 and GFP (25 kDa) infiltrated leaves were isolated and run on the gel next to the ADTs.

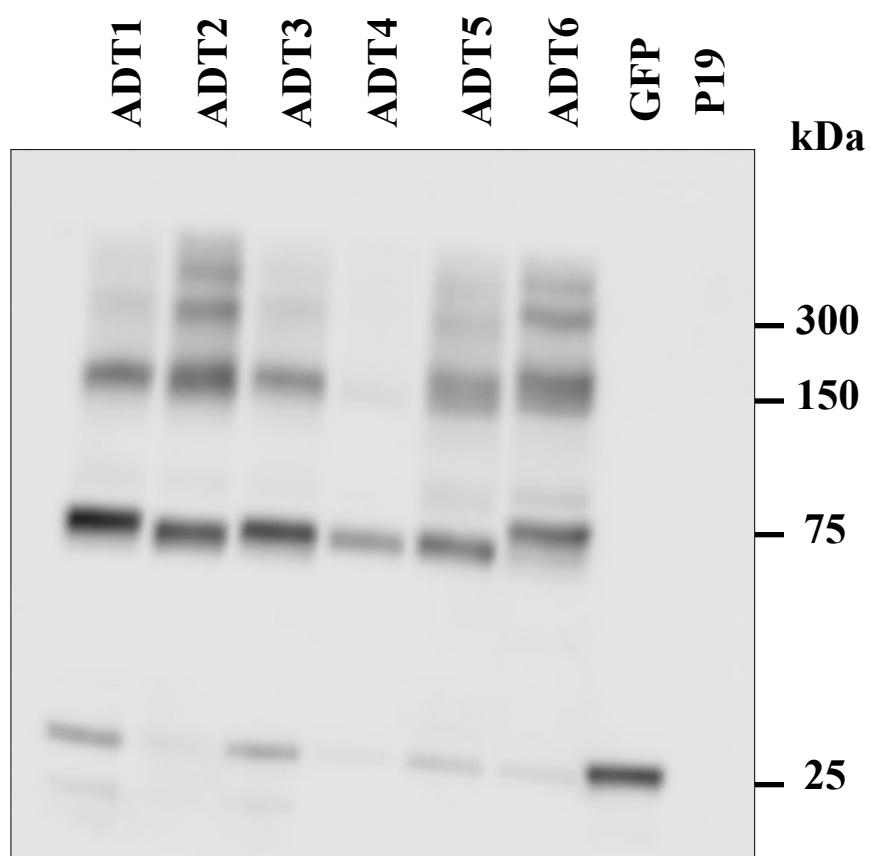


Figure 8 Expression pattern of GFP and GUS in *N. benthamiana* and *A. thaliana* using ADT5 native promoter.

To ensure the *proNat5* is active transcriptionally, *proNat5*:GUS:GFP was transiently expressed in *N. benthamiana* and stably transformed into *A. thaliana*.

(A) Activity of the *proNat5* in regulating the GFP expression was examined in *N. benthamiana* leaves and visualized by confocal microscopy. At 5 dpi, the GFP signal was detected in the cytosol.

Chlorophyll fluorescence and fluorescent proteins are shown separately in the left and middle columns, respectively, and the merged image is shown in the right column.

Scale bars are 5 μ m

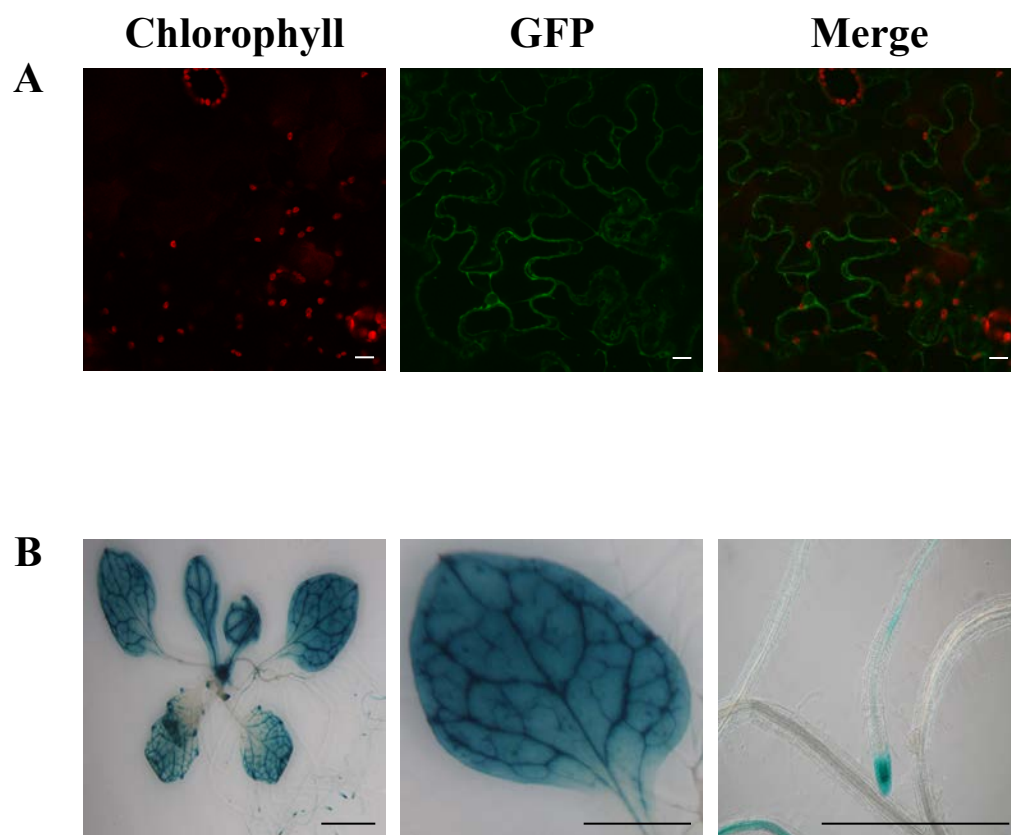
(B) Two week-old transgenic *A. thaliana* seedlings were stained with GUS solution and GUS expression was detected using light microscopy.

Left: Gus expression was detected in the leaves and roots of the seedlings.

Middle: At higher magnification, GUS expression was mainly observed at the leaf veins.

Right: At higher magnification, GUS expression was detected at the root tips.

Scale bars 1 mm.



The *proNat5* activity was also examined in different tissues of stably transformed *A. thaliana* plants by GUS assay (section 2.12). The left image on Figure 8 shows the GUS expression pattern in 2 weeks old *A. thaliana* seedlings and GUS expression was detected mostly in the leaves and roots. Higher magnification images of a leaf and a root demonstrated that the GUS expression was mostly in the leaf veins (Figure 8B middle panel) and the root tip (Figure 8B right panel).

Since the transcriptional activity of the *proNat5* promoter was confirmed with reporter genes, the same sequence was cloned upstream of ADT5 to determine if ADT5 can be detected in the nucleus. To do this, *proNat5* was PCR-amplified with appropriate primers (Appendix 1) to introduce restriction enzymes 3' *Mau*BI and 5' *Xho*I cut sites. As the CaMV 35S promoter in the pCB-ADT5 also had 3' *Mau*BI and 5' *Xho*I, the two promoter sequences were exchanged (section 2.3.4). The new construct was named proNat5::ADT5:CFP and was sequenced to ensure that no errors were introduced by PCR. Then, proNat5::ADT5:CFP was transformed into *Agrobacterium* cells GV3101 by electroporation. *Agrobacterium* cells harboring proNat5::ADT5:CFP were co-infiltrated with p19 cells into *N. benthamiana* leaves (section 2.5.1). *Agrobacterium* containing p19 and *Agrobacterium* carrying GFP vector were used as negative and positive controls, respectively. Expression of the recombinant proteins was visualized at 5 dpi by confocal microscopy. The localization of ADT5-CFP was detected in the chloroplast stromules as short protrusions and long projection structures (Figure 9A). In addition, it was detected in the nucleus (Figure 9B) and this demonstrates that the nuclear localization of ADT5 happens under the regulation of its native promoter. As expected no signal was observed in the samples expressing p19 (Figure 9C) and a cytosolic and nuclear expression pattern was observed for GFP (Figure 9D). To ensure that the observed CFP signal in the nucleus was not due to cleaved CFP products, TSP was isolated from the same infiltrated leaf tissues used for confocal microscopy to perform a Western blot.

No recombinant protein was detected in samples isolated from p19-infiltrated leaves while one band for GFP (25 kDa) and a band of degraded product of GFP (20 kDa) were detected (Figure 10). A distinct band at the expected size of ADT-CFP (73.8 kDa) was detected for samples from proNat5::ADT5:CFP-infiltrated leaves. In addition, one band at the size of the ADT5-CFP dimers (150 kDa) and a very faint band

Figure 9 Subcellular localization of the ADT5-CFP with *proNat5* and control transformations in *N. benthamiana*.

Agrobacterium cells harboring *proNat5::ADT5:CFP* were co-infiltrated with *Agrobacterium* cells with p19 into the *N. benthamiana* leaves to determine the localization of ADT5-CFP. *Agrobacterium* cells with p19 and *Agrobacterium* cells with GFP alone were infiltrated individually as controls. Fluorescent signals were visualized by confocal microscopy at 5 dpi.

Chlorophyll fluorescence and fluorescent proteins are shown separately in the left and middle columns, respectively, and the merged image is shown in the right column.

Scale bars are 5 μm .

White arrows: Nucleus.

(A) ADT5-CFP is detected in the chloroplast stromules as short protrusions and long protrusions.

(B) ADT5-CFP is detected in the nucleus.

(C) No signal is detected in leaf samples infiltrated with p19.

(D) GFP is detected in the cytosol and nucleus.

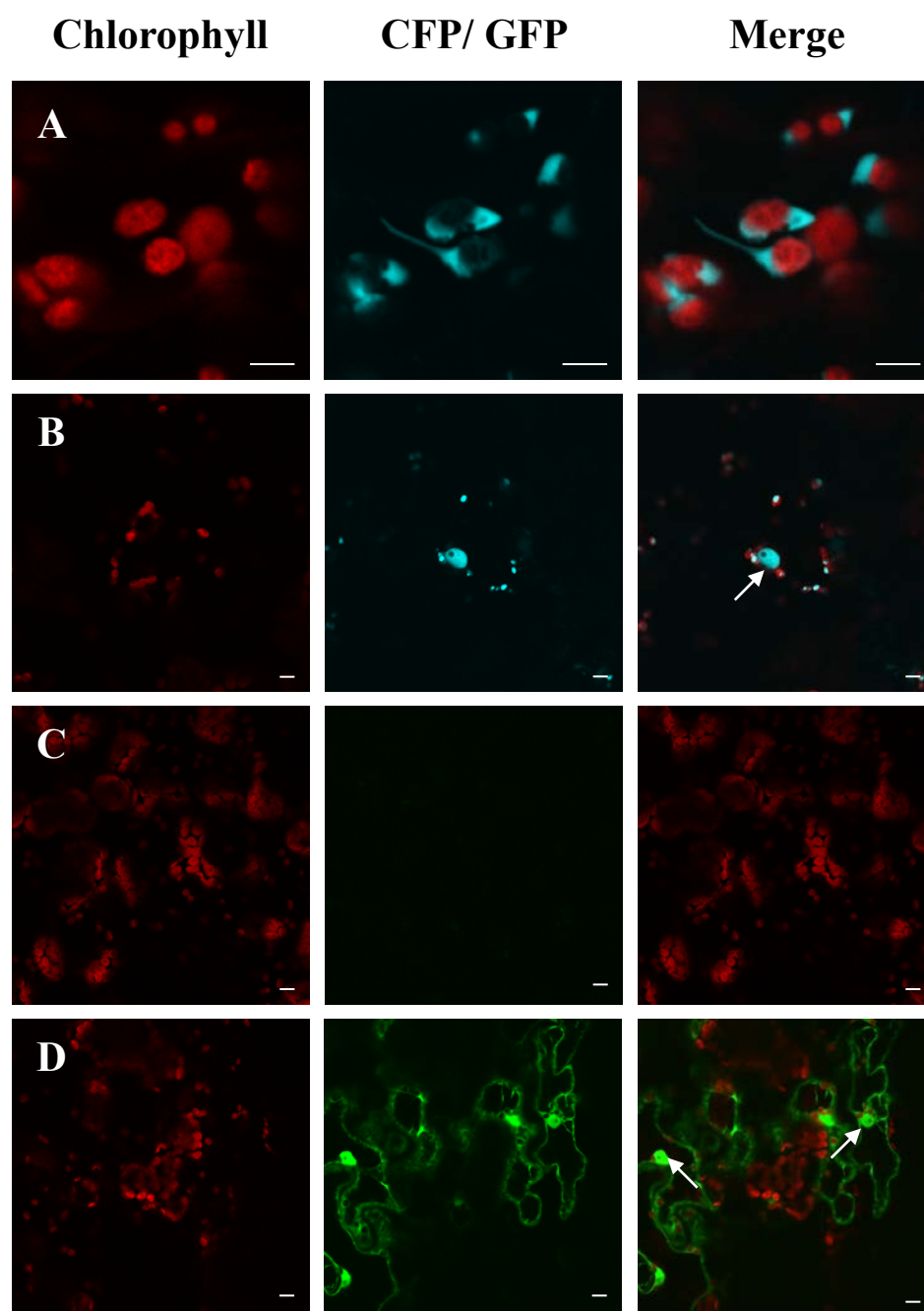
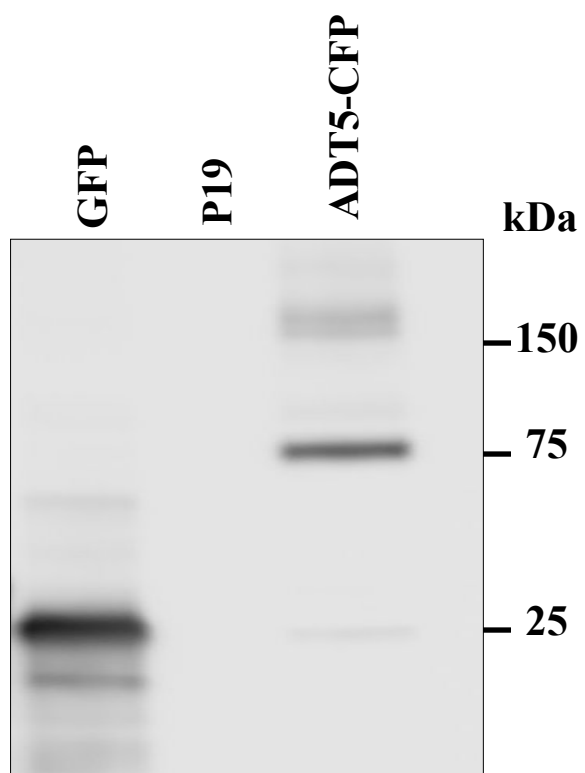


Figure 10 Detection of ADT5-CFP expressed with *proNat5* in transiently transformed *N. benthamiana* leaves.

Ten µg of TSP was size separated by SDS-PAGE, blotted with nitrocellulose membrane and detected with a GFP antibody. Expected sizes for ADT5-CFP and GFP are 73.8 kDa and 25 kDa, respectively. One band of the expected size for monomeric ADT5-CFP was detected and a second band was detected at 150 kDa, a size consistent with ADT5-CFP dimers. No band was detected for the p19 negative control. One band at the size of 25 kDa and a second band at approximately 20 kDa possibly a GFP degraded product were observed for the GFP positive control.



at size of GFP (25 kDa) were detected. Since it was found that amount of cleaved products were not sufficient to accumulate in the nucleus and account for the signal detected by confocal microscopy (section 3.1), it was concluded that nuclear localization of ADT5 was *bona fide*.

3.3 Stromules, tunnels from chloroplasts to the nucleus?

Stromules are projections from chloroplasts that increase the chloroplast surface area and potentially increase the trafficking of signals and metabolites from plastids to other plastid and organelles in the cell (Köhler and Hanson, 2000; Kwok and Hanson, 2004c; Hanson and Sattarzadeh, 2013). Accordingly, close positioning of plastids and nuclei in various tissues of *N. tabacum* and *A. thaliana* has been reported where stromules were observed to reside in grooves and in-folds of the nuclear envelope (Kwok and Hanson, 2004b). These close contacts between stromules and nuclei suggest that a physical connection may exist that increases the communications between chloroplasts and nucleus.

Transient expression of the ADT5-CFP proteins in *N. benthamiana* leaves displayed various stromule patterns (“dot” pattern, short protrusions and longer projections) from chloroplasts close to the nucleus (Figure 11). These stromule patterns and position close to the nucleus suggests possible roles for stromules in translocating ADT5 from chloroplasts to the nucleus. To test if stromules interact with the nucleus and transfer ADT5 from chloroplasts to the nucleus, ADT5-GFP fusion constructs were transformed into the chloroplast genome of *N. tabacum* and hence they were expressed in the chloroplast stroma instead of the cytosol. Several constructs were prepared: full-length *ADT5* and N-terminal deletion constructs: *IS5* (deletion of the TP) and *S5* (deletion of the TP and I region). Appropriate sequences were PCR-amplified and 5’ and 3’ restriction enzyme *NheI* cut sites were introduced. PCR products were cloned into the chloroplast transformation vector pCEC4 by ligation (section 2.3.4). Deletion constructs were generated based on reports indicating that cleavage of transit peptides might be required for activation of proteins in the chloroplast (McFadden, 2001; Bobik and Burch-Smith, 2015). Since it was not clear if the “I” region was part of the TP or not, two deletion

Figure 11 Close proximity of stromules and nucleus in *N. benthamiana* leaf samples expressing ADT5-CFP.

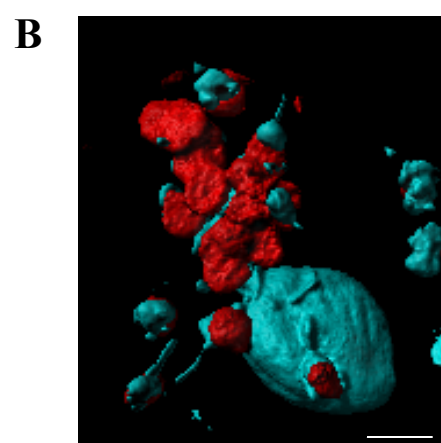
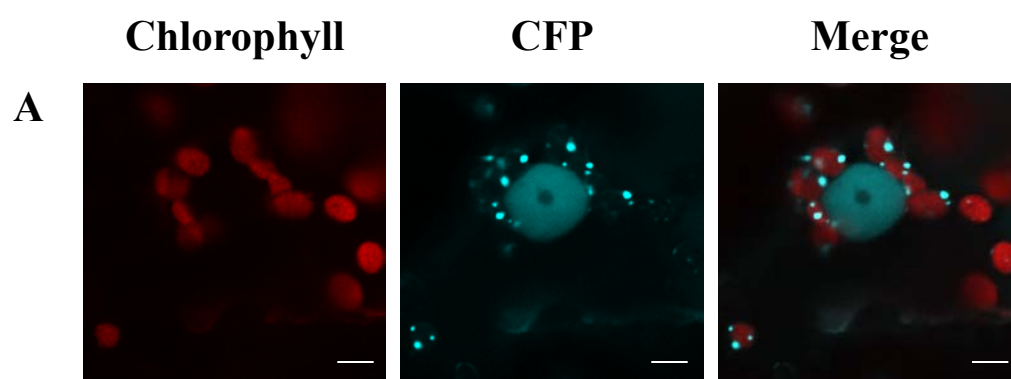
Transient expression of ADT5-CFP in *N. benthamiana* leaves was detected in the stromules with various patterns and nucleus. Confocal images were used to generate a 3D picture to show the close positioning of chloroplasts and stromules to the nucleus.

Chlorophyll fluorescence and fluorescent proteins are shown separately in the left and middle column, respectively, and a merged image is shown in the right column.

Scale bars are 5 μm (A) and 7 μm (B).

(A) ADT5-CFP was often detected at the nucleus and the stromules of chloroplast surrounded the nucleus.

(B) 3D image showing the stromules pattern and the close positioning of stromules and nucleus.



constructs were generated with and without the “I” region. The resulting plasmids were named ADT5-FL, ADT5-IS and ADT5-S (Figure 12). In these plasmids the ADT5 coding sequences were fused to *GFP* and the expression of the fusion proteins was regulated by a tobacco plastid promoter *PpsbA* and *TrbcL* terminator (Kolotilin et al., 2013). Similarly, constructs for *ADT2* and *ADT4* were generated as controls to compare the localization patterns of the expressed fusion proteins. It was expected that only ADT5-GFP would localize into the nucleus, as ADT4 and ADT2 have never been detected in the nucleus. In addition, the pCEC4-GFP (kindly provided by Dr. Menassa, Agriculture and AgriFood Canada, London, ON) was used as a control for stromal GFP expression, and transformation efficiency (GFP control).

Plasmids were transformed into the chloroplast genome of *N. tabacum* by gene bombardment (section 2.5.2). After six weeks of incubation on antibiotic selective medium some green calli appeared on the bombarded leaf tissue. Positively transformed calli were identified by PCR using *PpsbA* forward and *GFP* reverse primers (Figure 13). Three positive calli for each *ADT* plasmid and two for *GFP* were transferred on appropriate media (section 2.5.2) for regeneration of transplastomic plantlets. Finally, the expression pattern of the fusion proteins in regenerated plantlets was analyzed by confocal microscopy. For comparison, an untransformed leaf sample from wild type *N. tabacum* was used as a negative control and a leaf of the transplastomic plantlets expressing GFP was used as a positive control. No GFP signal was detected in wild type leaf samples (Figure 14A) and the GFP signal was detected throughout the chloroplast stroma in GFP transplastomic plantlets (Figure 14B). Typical localization patterns for full-length (FL) ADT5, ADT4 and ADT2 are shown in Figure 15. ADT5-FL-GFP was accumulated at a low level and was detected as small dots around chloroplasts like stromules (Figure 15A) while no signal was detected for ADT4-FL-GFP (Figure 15B). In contrast, ADT2-FL displayed accumulation at higher levels and was clearly observed in all the chloroplasts as green patches like stromules (Figure 15C).

Typical localization patterns for IS constructs are shown in Figure 16. ADT5-IS-GFP was observed as a circular, distinct pattern surrounding most of the chloroplasts (Figure 16A). Similar to ADT4-FL-GFP, ADT4-IS-GFP did not accumulate and no signal was observed.

Figure 12 Schematic representation of full-length and deletion constructs used for chloroplast genome transformation in *N. tabacum*.

To determine if ADT5 can translocate from chloroplast directly into the nucleus through stromules, full-length ADT5 and two deletion constructs were generated. The same construct for ADT4 and ADT2 were generated as controls. Schematic not drawn to scale.

Full-length constructs include TP, CAT and ACT domain. IS constructs include I region, CAT and ACT domain. S constructs include CAT and ACT domain.

ACT: ACT regulatory domain; CAT: catalytic domain; GFP: Green Fluorescence Protein tag; I: intermediate region; *PpsbA*: tobacco plastid promoter; TP: transit peptide, *TrbcL*: transcription terminator.

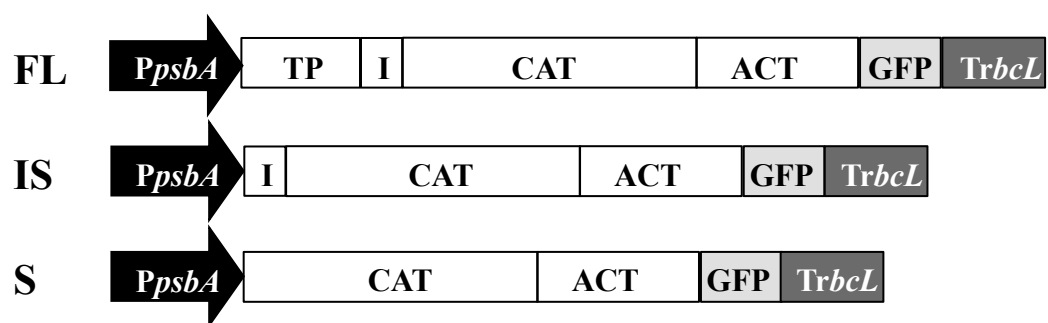


Figure 13 Genotyping of transplastomic calli.

DNA was extracted from calli and used as template for PCR amplification using *PpsbA* forward primer and *GFP* reverse primer to identify positive transplastomic calli. All PCR products were size-separated in 1% agarose gels. Predicted sizes for the amplified ADT-GFP fragments are as follow: ADT5-FL: 1447 bp, ADT5-IS: 1117 bp, ADT5-S: 1067 bp, ADT4-FL: 1442 bp, ADT4-IS: 1118 bp, ADT4-S: 1067 bp, ADT2-FL: 1313 bp, ADT2-IS: 1073 bp and ADT2-S: 1019 bp, GFP control 200 bp.

White arrows: positive calli used for further analysis.

L: 1 kb or 1 kb plus ladders.

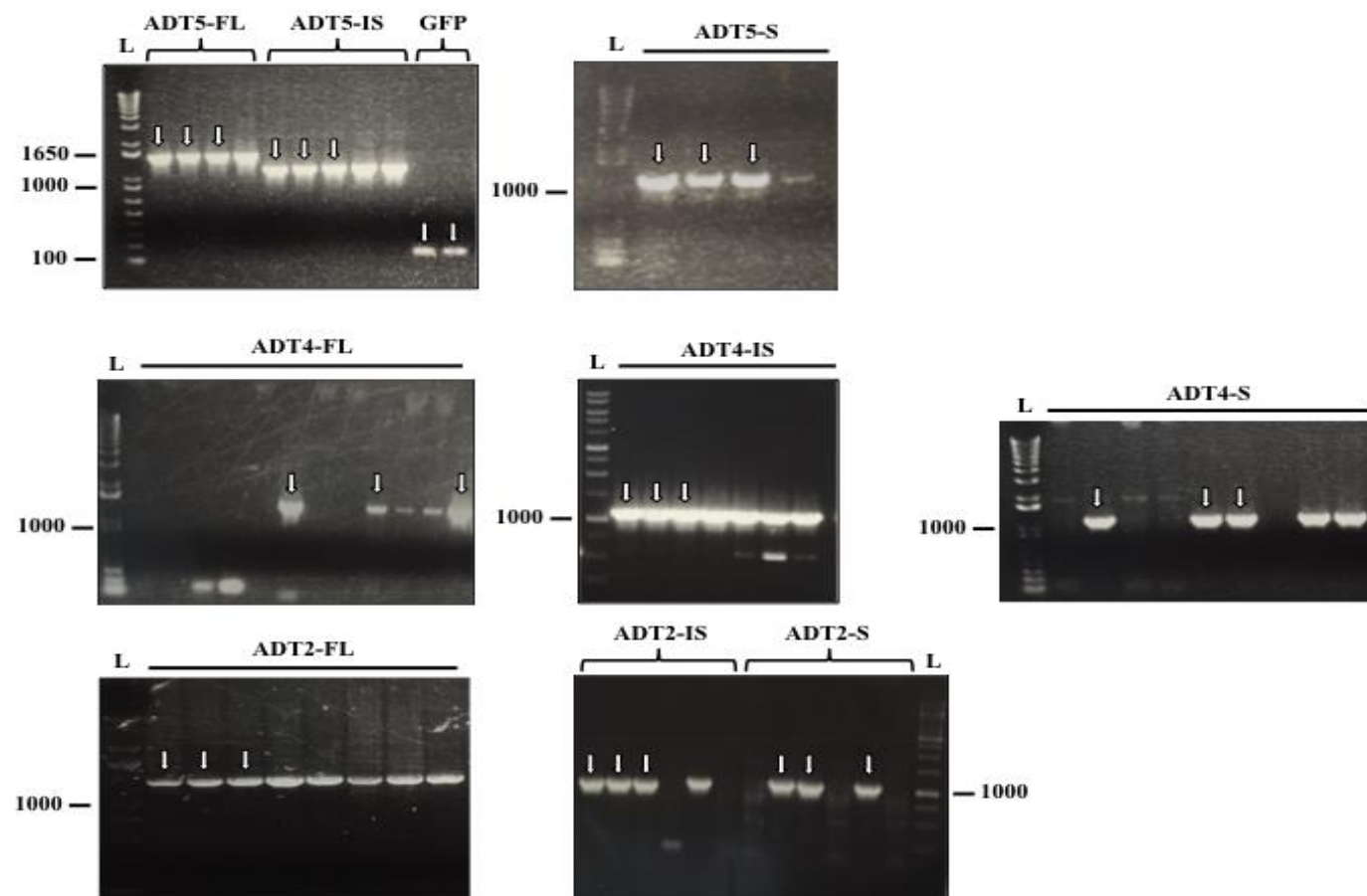


Figure 14 Controls in chloroplast transformation experiments.

A leaf sample of wild type *N. tabacum* and GFP control transplastomic plantlets (4 weeks old) were analyzed by confocal microscopy.

Chlorophyll fluorescence and GFP fluorescence are shown separately in the left and middle columns, respectively, and the merged image is shown in the right column.

Scale bars are 5 μm .

(A) No GFP signal is detected in wild type leaf samples.

(B) Strong GFP fluorescence is observed in the stroma of chloroplasts in GFP expressing transplastomic plantlets leaves.

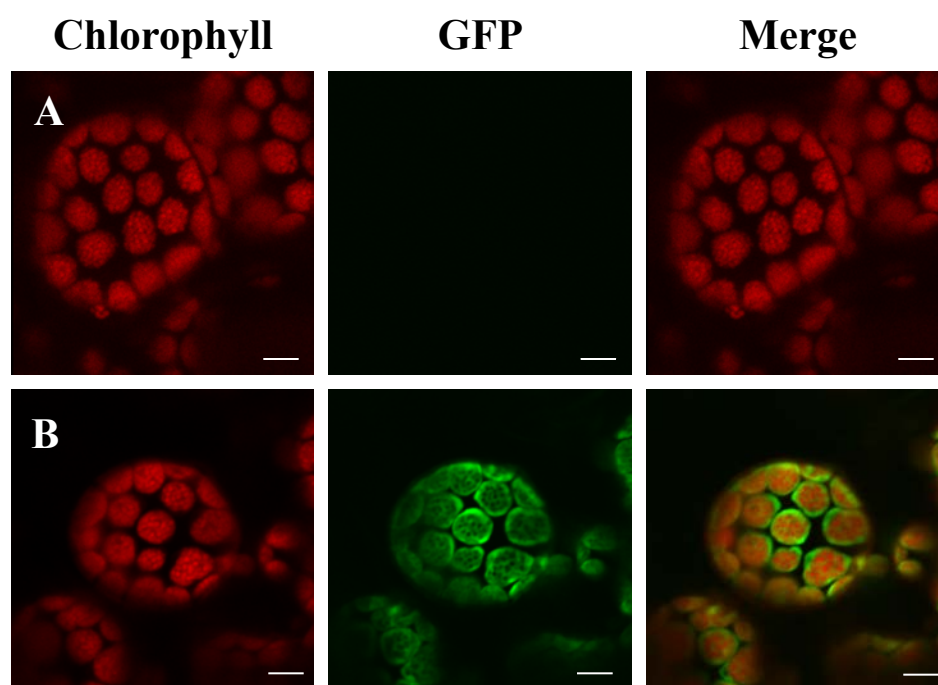


Figure 15 Localization pattern of ADT-FL-GFP in transplastomic plantlets.

To identify the localization pattern of ADT-FL-GFP fusion proteins, mesophyll cells of *N. tabacum* transplastomic plantlets (4 weeks old) were analyzed by confocal microscopy.

Chlorophyll fluorescence and GFP fluorescence are shown separately in the left and middle columns, respectively, and the merged image is shown in the right column. Scale bars are 5 μm .

(A) ADT5-FL-GFP is detected in stromule around the chloroplasts.

(B) No signal is detected for ADT4-FL-GFP.

(C) ADT2-FL-GFP is observed in green patches stromules around chloroplasts.

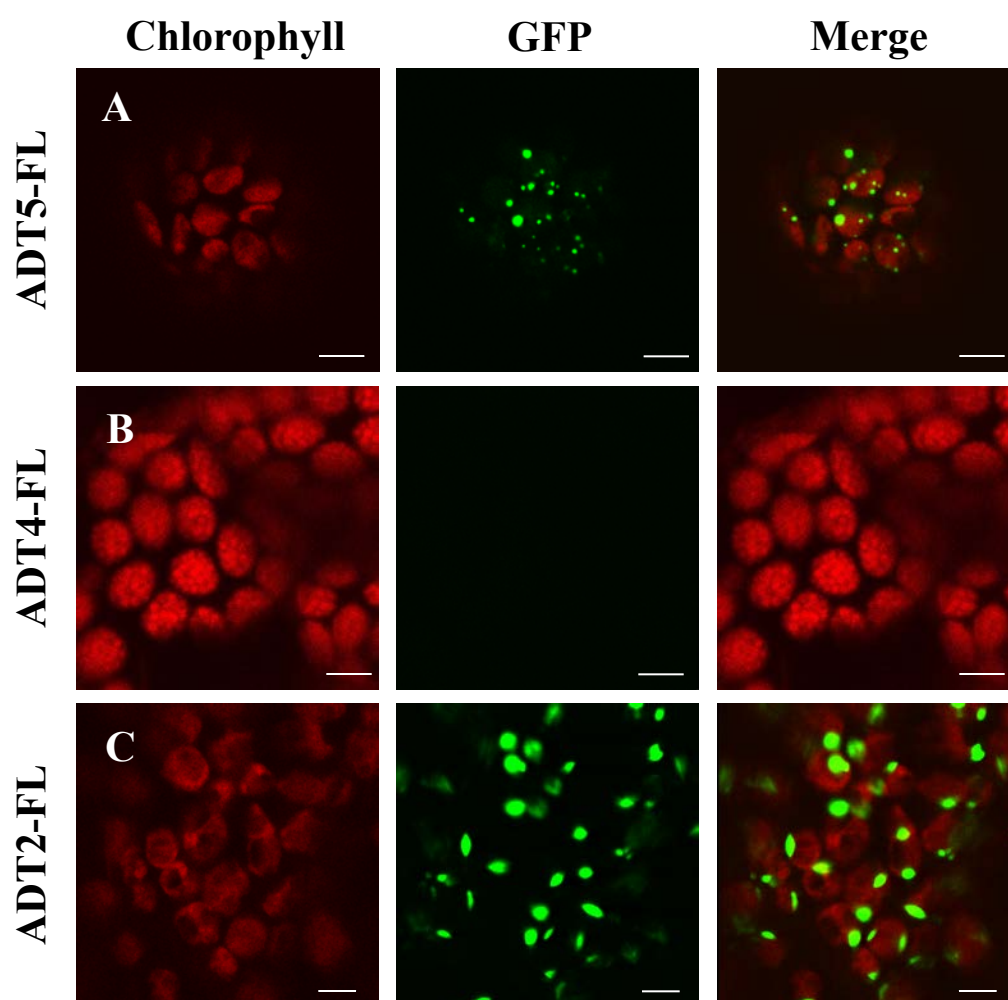


Figure 16 Localization pattern of ADT-IS-GFP in transplastomic plantlets.

To identify the localization pattern of ADT-IS-GFP fusion proteins, mesophyll cells of *N. tabacum* transplastomic plantlets (4 weeks old) were visualized by confocal microscopy.

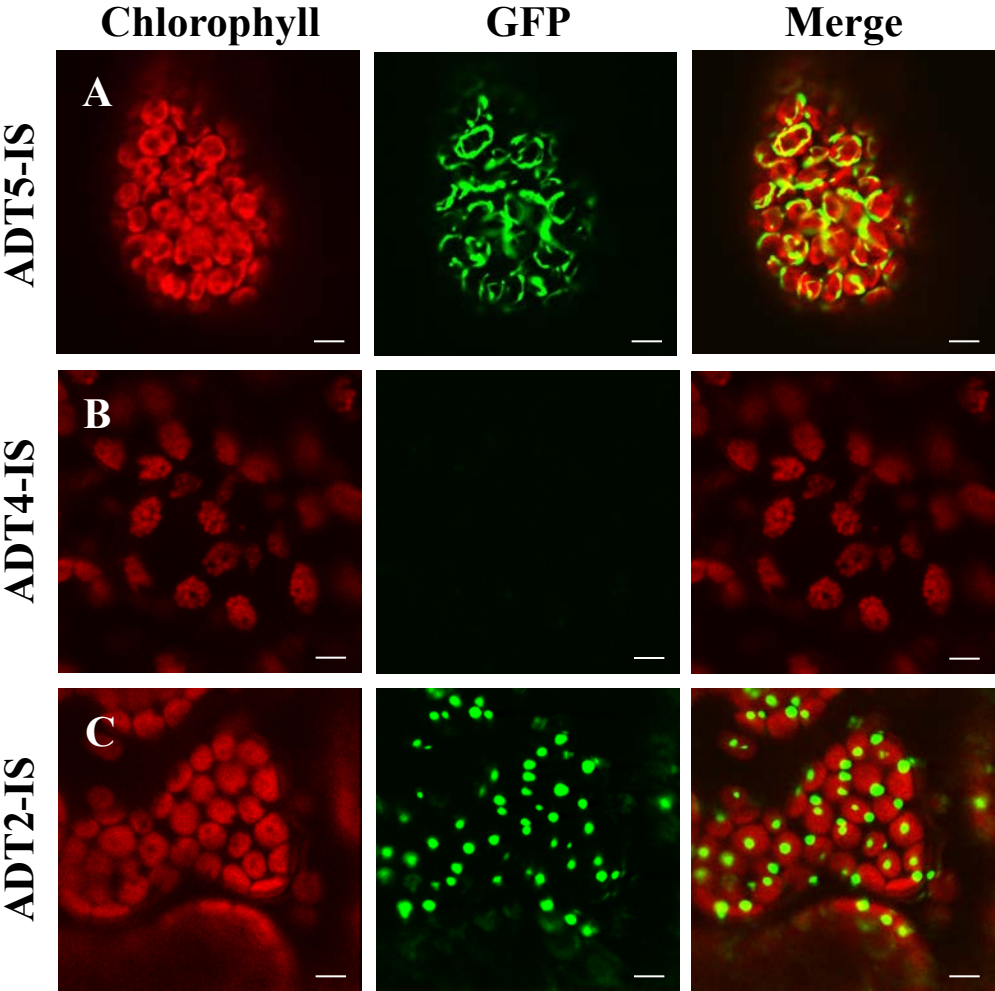
Chlorophyll fluorescence and GFP fluorescence are shown separately in the left and middle column, respectively, and the merged image is shown in the right column.

Scale bars are 5 μm .

(A) ADT5-IS-GFP is detected as circular patterns surrounding the most of the chloroplasts.

(B) No signal is observed for ADT4-IS-GFP and the autofluorescence of chloroplasts do not seem to distribute evenly.

(C) ADT2-IS-GFP is localized as big protrusions stromules.



However the chloroplast shapes were different compared to wild type oval chloroplasts (Figure 16B and Figure 14A). Similarly to ADT2-FL-GFP, the signal for ADT2-IS was strong and detected in big globular protrusions from the chloroplasts like stromules (Figure 16C).

The expression patterns for short ADTs are shown in Figure 17. ADT5-S-GFP was not detected (Figure 17A) while ADT4-S-GFP appeared as small dot pattern stromules (Figure 17B). The expression of the ADT2-S-GFP generated a strong signal, and appeared as small disc and rod shapes mainly situated on the sides of chloroplasts (Figure 17C).

Although the accumulation of the fusion proteins were different based on the ADT and length of the sequences, no nuclear localization of any of the fusion proteins was detected. To see if the expression pattern of the fusion proteins would change in mature plants, small plantlets were transferred to soil and moved to the greenhouse. In the mean time, the phenotype of the plants was observed to see if the expression of the ADT-GFP fusion proteins in chloroplasts causes any changes in phenotype. The ADT4-FL plantlets did not survive in the greenhouse and the phenotype of transplastomic ADT5-FL, ADT2-FL, ADT2-IS, ADT5-S, ADT4-S and ADT2-S plants were similar to GFP control plants. In contrast, ADT5-IS and ADT4-IS had striking phenotypes. ADT5-IS plants had variegated leaves but otherwise the plants looked comparable to other transplastomic plants and they were able to produce seeds. ADT4-IS plants were dwarf, the leaves were pale in appearance, and plants grew slowly and did not produce flowers or seeds (Figure 18).

Figure 19A shows a summary of leaf phenotype of the transplastomic plants. The mature transplastomic plant leaves were investigated by confocal microscopy for the presence of ADT-GFP within the chloroplast and/or nucleus. Several leaves were checked representing different developmental stages (from top of the plant to the bottom: young, intermediate and old leaves). While no signal was detected for ADT5-S-GFP and ADT4-IS-GFP, the localization pattern for ADT5-FL-GFP, ADT2-FL-GFP, ADT2-IS-GFP, ADT4-S-GFP and ADT2-S-GFP were the same as what had been observed in the plantlet stage.

Figure 17 Localization pattern of ADT-S-GFP in transplastomic plantlets.

To identify the localization pattern of ADT-S-GFP fusion proteins, mesophyll cells of *N. tabacum* transplastomic plantlets were visualized at 4 weeks old stage by confocal microscopy.

Chlorophyll fluorescence and GFP fluorescence are shown separately in the left and middle column, respectively, and the merged image is shown in the right column.

Scale bars are 5 μm .

(A) No signal is detected for ADT5-S-GFP.

(B) ADT4-S-GFP is observed in small dot stromules.

(C) The expression of the ADT2-S-GFP is as small disc shapes from chloroplast.

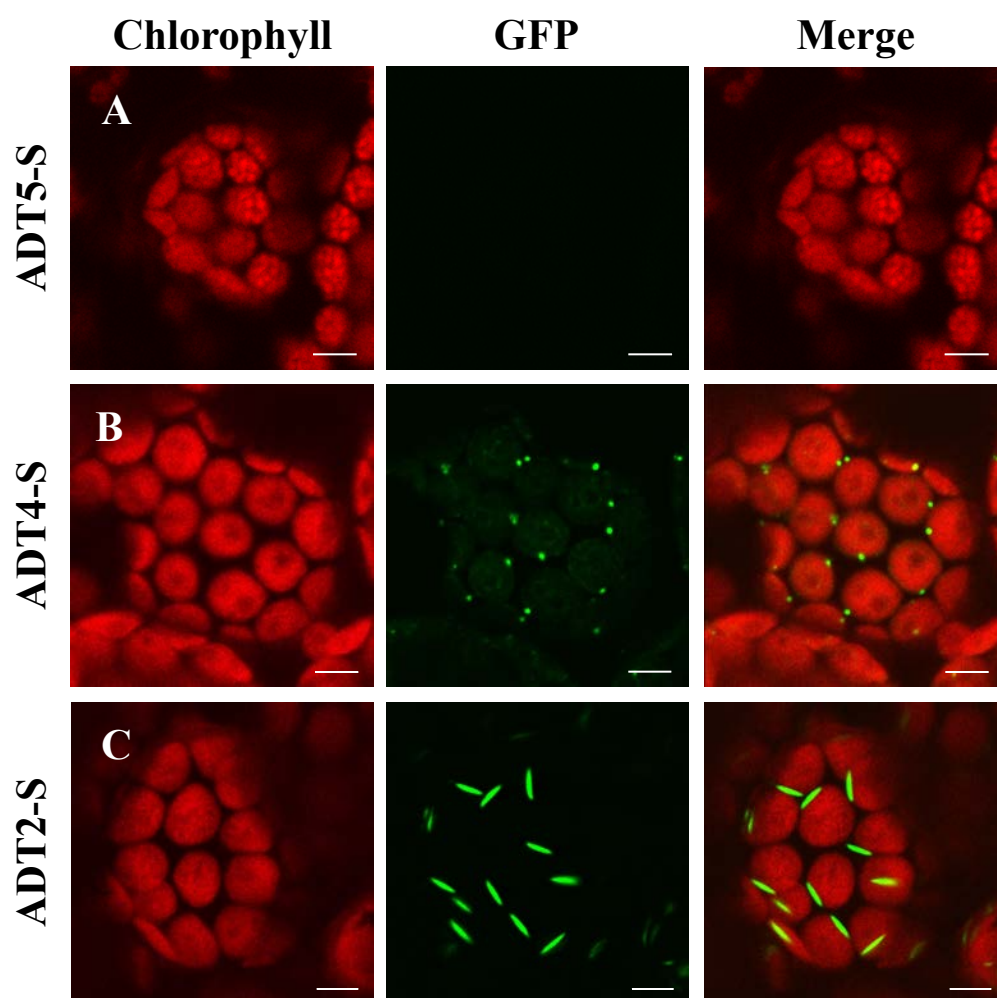


Figure 18 Phenotype of transplastomic GFP control, ADT5-IS and ADT4-IS plants.

Mature transplastomic GFP control, ADT5-IS and ADT4-IS at 4 months of age.

Scale bar 1 m.

(A) GFP-control plants grew tall with green leaves.

(B) ADT5-IS plants had variegated leaves but grew the same as GFP control plants.

(C) ADT4-IS plants grew very slowly compared to the GFP-control plants. Plants were dwarf and had pale leaves.

GFP



ADT5-IS



ADT4-IS



Figure 19 Leaf phenotype and localization pattern of ADT-GFP fusion proteins in transplastomic *N. tabacum*.

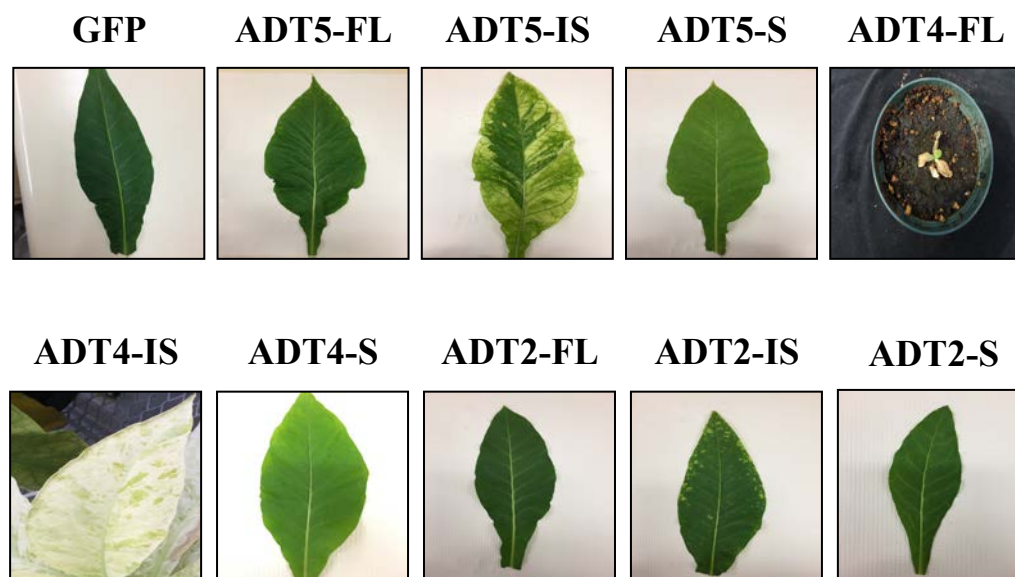
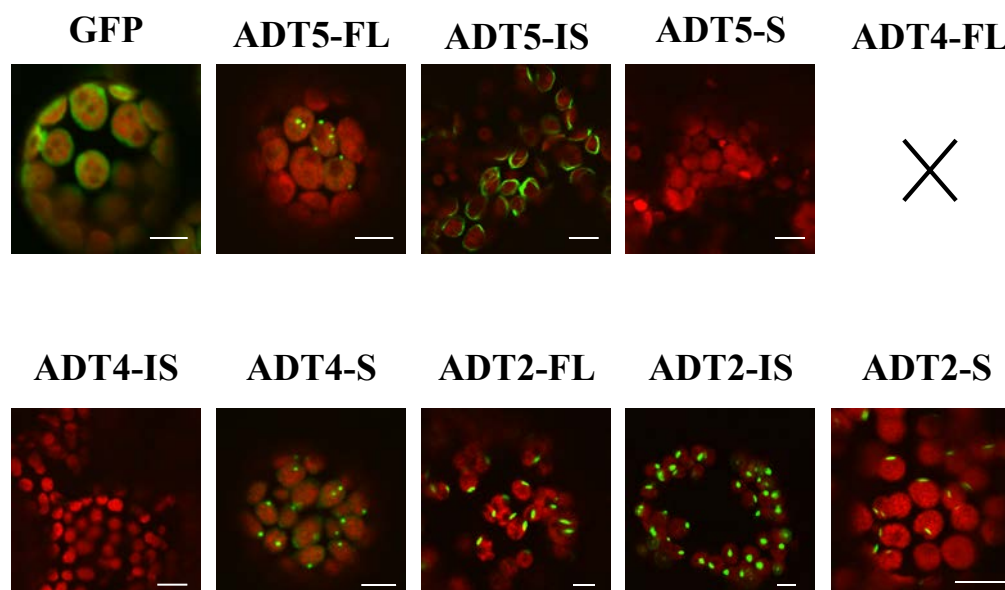
Phenotype of mature transplastomic plants and the localization pattern of the ADT-GFP fusion proteins were studied to see if there are any changes in phenotype of the plants or localization of the ADT-GFPs.

(A) One intermediate leaf represents the leaf phenotype of transplastomic plants. No phenotype change was observed for most of the transplastomic plants (GFP, ADT5-FL, ADT5-S, ADT4-S, ADT2-FL, ADT2-IS AND ADT2-S). However, while ADT4-FL transplastomic plantlets did not survive in greenhouse, ADT5-IS had variegated leaves and ADT4-IS had pale leaves.

(B) The ADT-GFP localization of the intermediate leaves is shown. GFP signal is detected in chloroplast stroma of the GFP control plants. ADT5-FL-GFP, ADT4-S-GFP, ADT2-FL-GFP and ADT2-IS-GFP show stromule localization. ADT5-IS-GFP is detected as circular, distinct pattern surrounding the chloroplasts and ADT2-S-GFP was detected as small disc and rod shapes mainly situated on the sides of chloroplasts.

Only merged images are shown.

Scale bars are 5 μm .

A Leaf phenotype of transplastomic plants**B Localization pattern of ADT-GFP fusion proteins**

While no signal was detected for ADT5-S-GFP and ADT4-IS-GFP, the localization pattern for ADT5-FL-GFP, ADT2-FL-GFP, ADT2-IS-GFP, ADT4-S-GFP and ADT2-S-GFP were the same as what had been observed in the plantlet stage. Figure 19B shows a summary of the localization patterns observed in an intermediate leaf of transplastomic plants. As a striking localization pattern was detected for ADT5-IS-GFP, protoplasts were extracted from young, intermediate and old leaf samples of ADT5-IS plants to obtain more informative confocal images (Figure 20). As control, protoplasts were isolated from an intermediate leaf of GFP transplastomic plants. In the GFP control, GFP was detected in the stroma (Figure 20A). In young ADT5-IS-GFP leaf sample the GFP was observed as a ring around chloroplasts (Figure 20B) while long projection pattern stromules were detected in intermediate leaves. Some of the stromules seem to form bridges from one chloroplast to another (Figure 20C). In older leaves, the fusion proteins were observed in the stroma (Figure 20D). However, despite the long stromules observed in the intermediate leaves of ADT5-IS transplastomic plants, no extensions between chloroplasts and nuclei were observed and no signal was detected in the nucleus. As no GFP signal was detected for ADT5-S-GFP and ADT4-IS-GFP, total soluble proteins were isolated from leaves of ADT5 and ADT4 transplastomic plants to perform a Western blot and check for the presence of the ADT-GFP proteins (Figure 21). A band at the expected size of the fusion proteins was detected for ADT5-FL-GFP, ADT5-IS-GFP and ADT4-S-GFP. However, no band was detected for ADT5-S-GFP and ADT4-IS-GFP, consistent with the absence of a fluorescent signal.

3.4 Does ADT5 have a nuclear localization sequence

3.4.1 *In silico* analysis to identify a putative NLS in ADT5

Small proteins (40 kDa and less) enter into the nucleus from the cytoplasm by passive diffusion and large proteins use the active importin-mediated system (Wente and Rout, 2010). ADT5-CFP (73.8 KDa) is too large to diffuse into the nucleus. Hence, it needs to be actively transported across the nuclear envelope, which often requires the presence of specific targeting sequences within the protein (Rout and Aitchison, 2001). Therefore, an *in silico* analysis was used to identify potential NLS in ADT5.

Figure 20 Protoplast images exhibiting patterns of ADT5-IS-GFP in young, intermediate and old leaves of transplastomic plants.

Protoplasts from leaf samples of ADT5-IS transplastomic plant at 6 months of age were extracted and the localization pattern of the ADT5-IS-GFP was detected by confocal microscopy. Protoplasts of the GFP control plants were visualized as control.

Chlorophyll fluorescence and fluorescence proteins are shown separately in the left and middle column, respectively, and the merged image is shown in the right column. Scale bars are 5 μm .

(A) GFP signal is detected in all the stroma for the GFP control protoplasts.

(B) In young leaves the ADT5-IS-GFP localization pattern is observed as long tubules all around the chloroplasts.

(C) In intermediate leaves ADT5-IS-GFP is detected in the chloroplast stroma as well as long stromules from chloroplasts. Some of the stromules seem to form bridges from one chloroplast to another.

(D) In old leaves the ADT5-IS-GFP signal is in the stroma and also it is observed as dot pattern stromules.

White arrows: Long stromules.

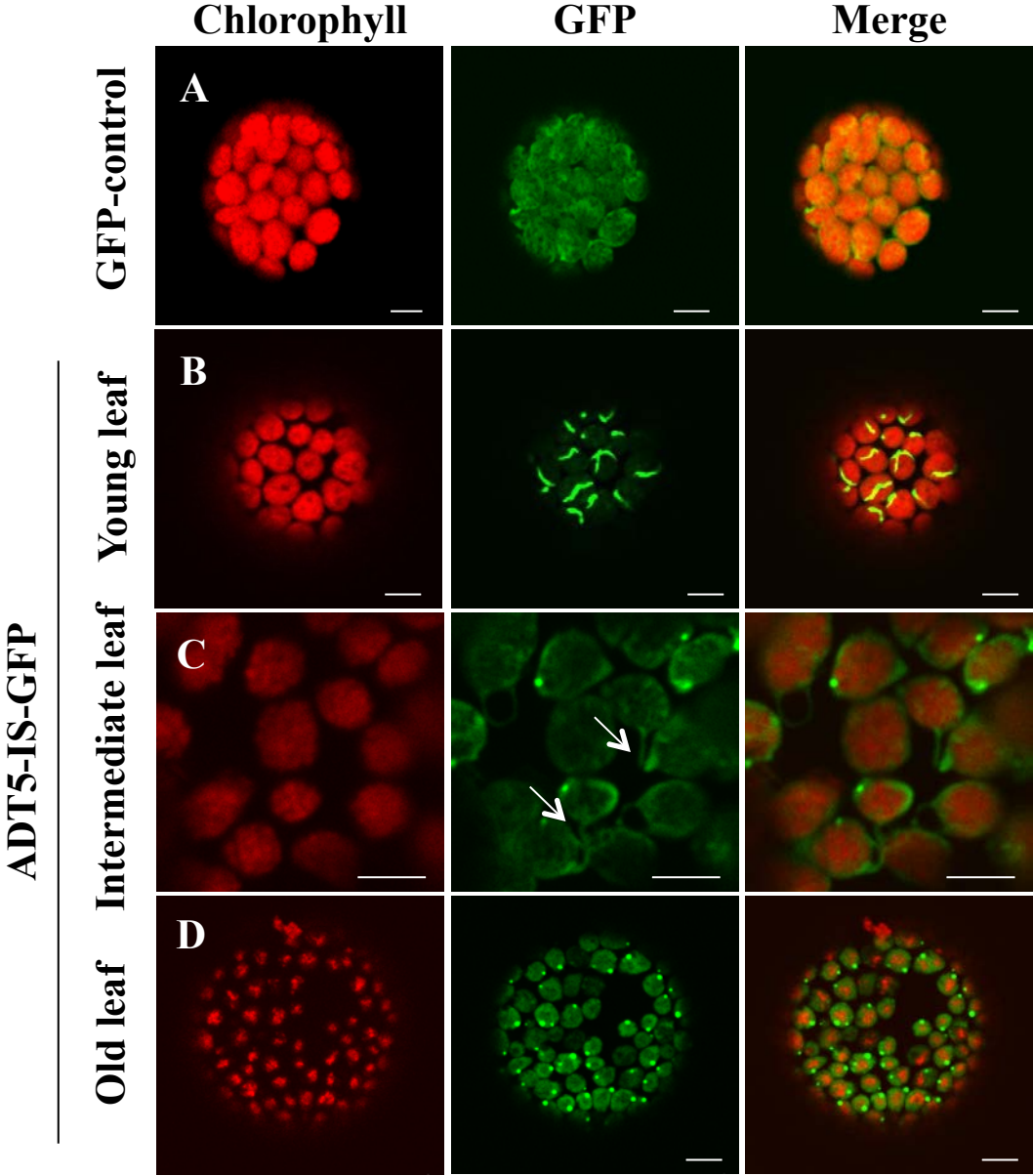
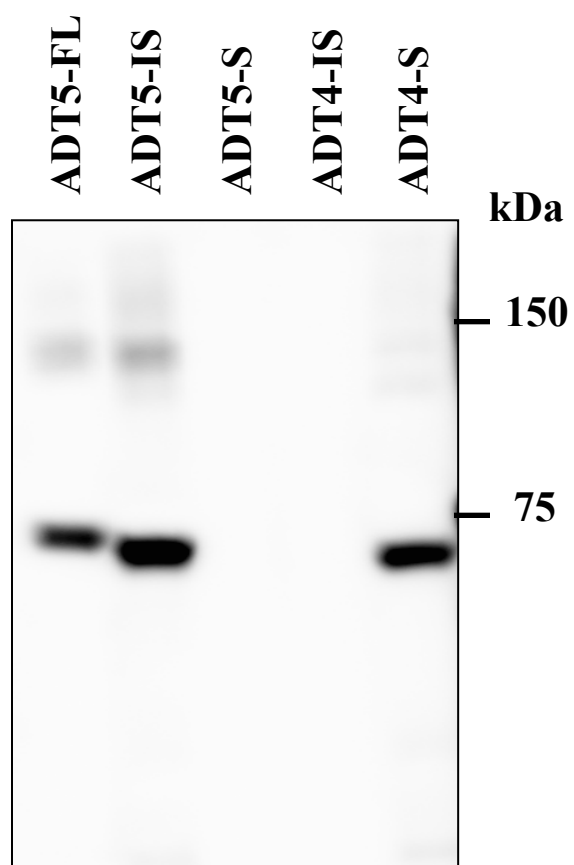


Figure 21 Protein expression of ADT-GFPs in *N. tabacum* transplastomic plants.

Western blot showing the recombinant ADT-GFP proteins expressed in *N. tabacum* chloroplasts. Ten µg of total soluble protein was size separated on a SDS-PAGE gel, blotted with nitrocellulose membrane and detected with a GFP antibody. Predicted protein size for fusions are ADT5-FL-GFP: 73.8 kDa, ADT5-IS-GFP: 67.8 kDa, ADT5-S-GFP: 66.2 kDa, ADT4-IS-GFP: 67.8 kDa, ADT4-S-GFP: 66.3 kDa.



The full-length ADT5 sequence was submitted to two programs: Predict Protein (LockTree2; Goldberg et al., 2012) and NLS Mapper (Kosugi et al., 2009b). The LockTree2 searches for potential organelle targeting sequences in protein sequences and ranks the likelihood of a subcellular localization by a confidence score ranging from 0 (unreliable) to 100 (reliable). Using this program only a chloroplast targeting sequence (TP) at the N-terminus of ADT5 was detected with a confidence score of 100. The NLS Mapper is able to predict monopartite NLS or bipartite NLS connected by a linker of 15-20 amino acids within proteins. The program searches for NLS in proteins and predict the nuclear localizations by a confidence score ranging from 1 to 10. Higher confidence score indicate a stronger NLS. Score confidence of 8, 9 and 10 indicates exclusive nuclear localization, 7 and 8 partial nuclear localization, 3, 4 and 5 both in the nuclear and cytoplasm localization and score 1 and 2 cytoplasm localization. Using this program no NLS was identified for ADT5.

3.4.2 Generating truncated fusion proteins to find a NLS

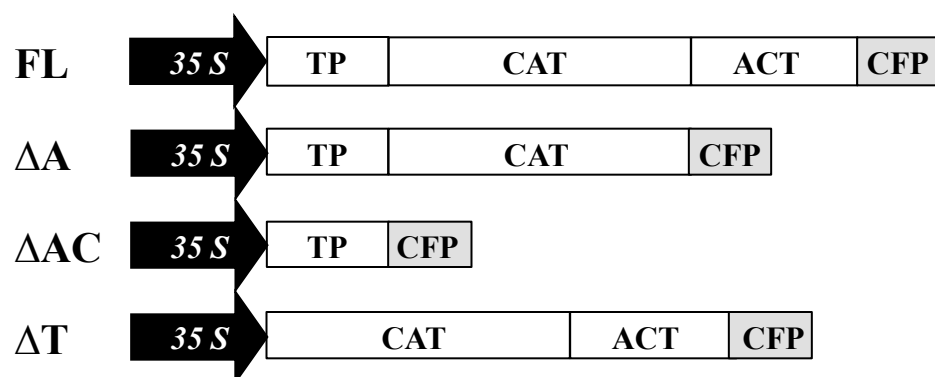
Computer programs were not successful to predict a NLS for ADT5. However prediction programs have been shown to miss the identification of NLSs (Krebs et al., 2010). Hence, an experimental approach was used to identify any potential nuclear localization sequences within ADT5. To identify a region within ADT5 that may contain a NLS, deletion constructs were generated as follows: lacking the ACT regulatory domain (ΔA), lacking ACT and CAT domains (ΔAC) and lacking the transit peptide domain and I region (ΔT). Full-length ADT5 and deletion fragments were PCR amplified using pCB5 vector (Bross, 2011) as template and appropriate gateway primers (Appendix 1). Amplified fragments were recombined into the Gateway plant destination vector pEarlyGate102 (section 2.3.4). In this vector, fragments are cloned with *CFP* and their expression is regulated by the CaMV 35S promoter (Earley et al., 2006). Expression constructs were checked by restriction enzyme digestion reactions and verified constructs were named ADT5-FL, ADT5- ΔA , ADT5- ΔAC , ADT5- ΔT (Figure 22). Analogous constructs were generated for *ADT4* to be used as controls since ADT4 is the most similar in sequence to ADT5 but has never been observed in the nucleus.

Figure 22 Schematic representations of ADT5 and ADT4 full-length and deletion constructs.

To identify a NLS within ADT5, constructs were generated for full-length and deletion sequences. For comparison, analogous constructs were generated for ADT4. Schematic not drawn to scale.

Full-length constructs include: TP, CAT and ACT domains. ΔA : deletion ACT domain, ΔAC : deletion ACT and CAT domains, ΔT : deletion TP. The predicted protein size for fusion proteins are: ADT5-FL: 73.8 kDa, ADT5- ΔA : 61 kDa, ADT5- ΔAC : 40 kDa, ADT5- ΔT : 58.6 kDa, ADT4-FL: 73.8 kDa, ADT4- ΔA : 60.9 kDa, ADT4- ΔAC : 40 kDa and ADT4- ΔT : 59.4 kDa.

ACT: ACT regulatory domain; CAT: catalytic domain; CFP: Cyan Fluorescence Protein; 35 S: CaMV 35S promoter; TP: transit peptide



To avoid diffusion into the nucleus due to the small size of the fusion proteins, all truncated fusion proteins were at least 40 kDa (see figure legend for details). All the constructs were introduced into *Agrobacterium* LBA4404 by electroporation and were co-infiltrated with p19 carrying cells into *N. benthamiana* leaves (section 2.5.1). The localization of the fluorescent fusion proteins was determined with confocal microscopy at 4 dpi. ADT5-FL fusion proteins were found in stromules and nucleus (Figure 23A). No signal was detected for ADT5- Δ A (Figure 23B), ADT5- Δ AC was seen in stromules, cytosol and nucleus and ADT5- Δ T was observed in cytosol and nucleus (Figure 23C). While ADT4-FL only localized to the stromules (Figure 24A) the observations for truncated ADT4 fusion proteins were similar to ADT5. No signal was detected for ADT4- Δ A (Figure 24B), ADT4- Δ AC was seen in stromules, cytosol and nucleus (Figure 24C) and ADT4- Δ T were found in cytosol and nucleus.

Δ AC proteins were close to size exclusion limit (ADT5- Δ AC: 40 kDa and ADT4- Δ AC: 40 kDa) and probably could diffuse into the nucleus. However, Δ T fusion proteins were too large to diffuse into the nucleus (ADT5- Δ T: 58.6 kDa and ADT4- Δ T: 59.4 kDa), therefore the nuclear CFP signal could be as result of the fusion proteins or cleaved products. In addition, ADT5- Δ A and ADT4- Δ A were not detected by confocal microscopy and it was not clear if those fusion proteins did not express or if their accumulation was too low to be detected. Therefore, TSPs were isolated from the same infiltrated leaf tissues used for microscopy and a Western blot was performed for full length and truncated Δ A and Δ T proteins (Figure 25). As controls, isolated p19 and GFP proteins (section 3.2) were used. No band for p19 protein samples and one 25 kDa band for GFP protein samples were detected. For ADT5-FL and ADT4-FL samples only one band at the expected size at approximately 73.8 kDa was detected. One faint band was detected for ADT5- Δ A but not at the expected size of the fusion protein (61 kDa), indicating a cleaved product but it was not detected by confocal microscopy. No band was detected for ADT4- Δ A consistent with no signal in microscopy images. N terminally truncated proteins ADT5- Δ T and ADT4- Δ T both had a band at the expected size of the fusion protein and some cleaved products. The cleaved band in ADT4- Δ T protein samples was more clear than the ADT5- Δ T protein samples however it was still not clear if the nuclear localization of these proteins was due to the accumulation of fusion protein

Figure 23 Subcellular localization of ADT5-CFPs in *N. benthamiana* leaves.

To determine the localization pattern of fusion proteins, full-length and deletion ADT5 constructs were transiently expressed in *N. benthamiana* leaves and fusion proteins were visualized at 4 dpi.

Chlorophyll fluorescence and CFP fluorescence are shown separately in the left and middle columns, respectively, and the merged image is shown in the right column. Construct name is mentioned on the left side and a summary of the confocal observations is shown on the right side.

White arrows: Nucleus.

C: Cytoplasm, N: Nucleus, S: Stromule.

Scale bars are 5 μm .

(A) ADT5-FL is localized in the stromules and nucleus.

(B) No CFP signal is detected for ADT5- ΔA .

(C) ADT5- ΔAC is observed in the cytosol, chloroplasts and nucleus.

(D) ADT5- ΔT is found in the cytosol and nucleus.

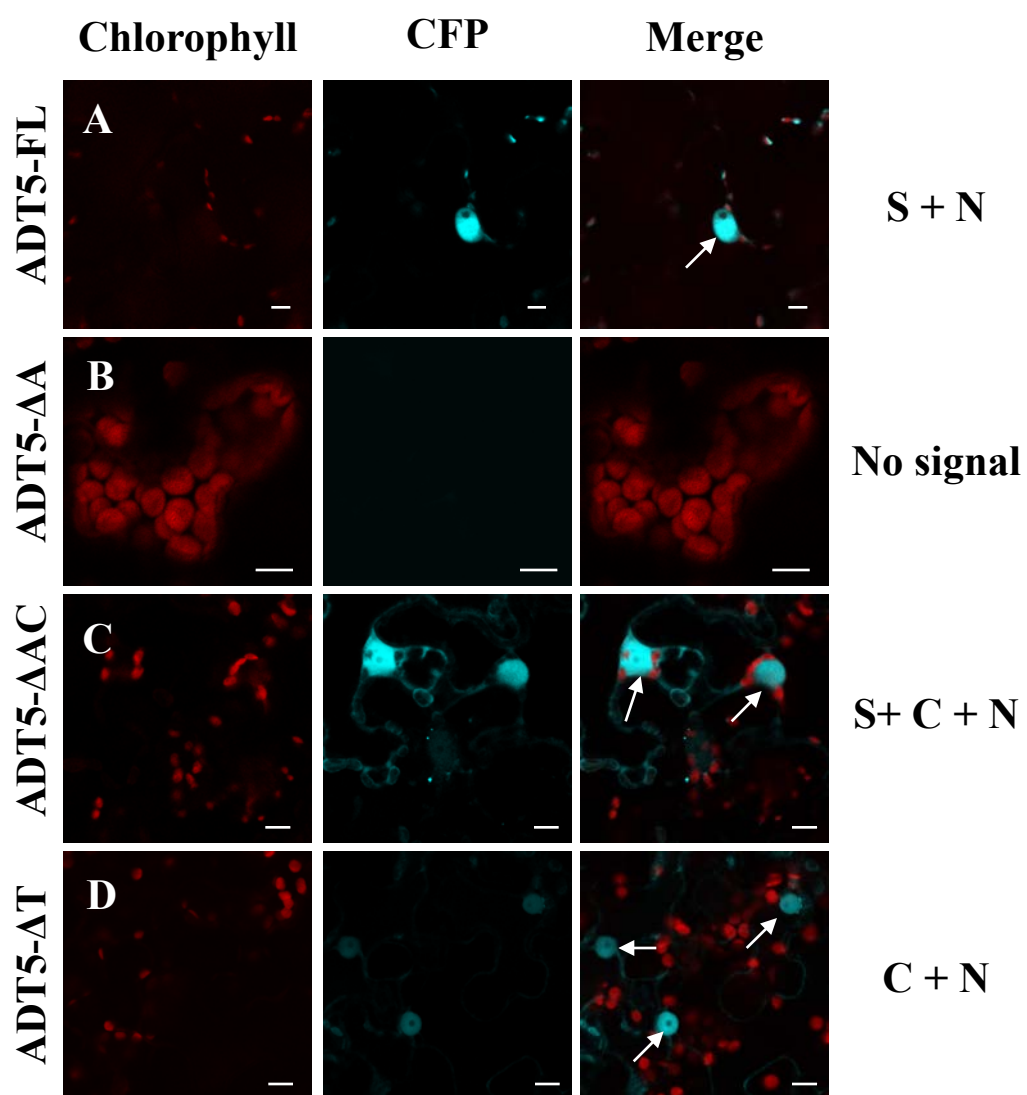


Figure 24 Subcellular localization of ADT4-CFP in *N. benthamiana* leaves.

Full length and deletion constructs for ADT4 were transiently expressed in *N. benthamiana* leaves and localization pattern of ADT4-CFP proteins were analyzed at 4 dpi.

Chlorophyll fluorescence and CFP fluorescence are shown separately in the left and middle columns, respectively, and the merged image is shown in the right column. Construct name is mentioned on the left side and a summary of the confocal observations is shown on the right side.

White arrows: Nucleus.

C: Cytoplasm, N: Nucleus, S: Stromule.

Scale bars are 5 μm .

(A) ADT4-FL is detected in the chloroplast stromules.

(B) No CFP signal is detected for ADT4- ΔA .

(C) ADT4- ΔAC is found in stromules and nucleus. It is also observed in the cytosol.

(D) ADT4- ΔT is observed mostly in the cytosol and it is also detected in the nucleus.

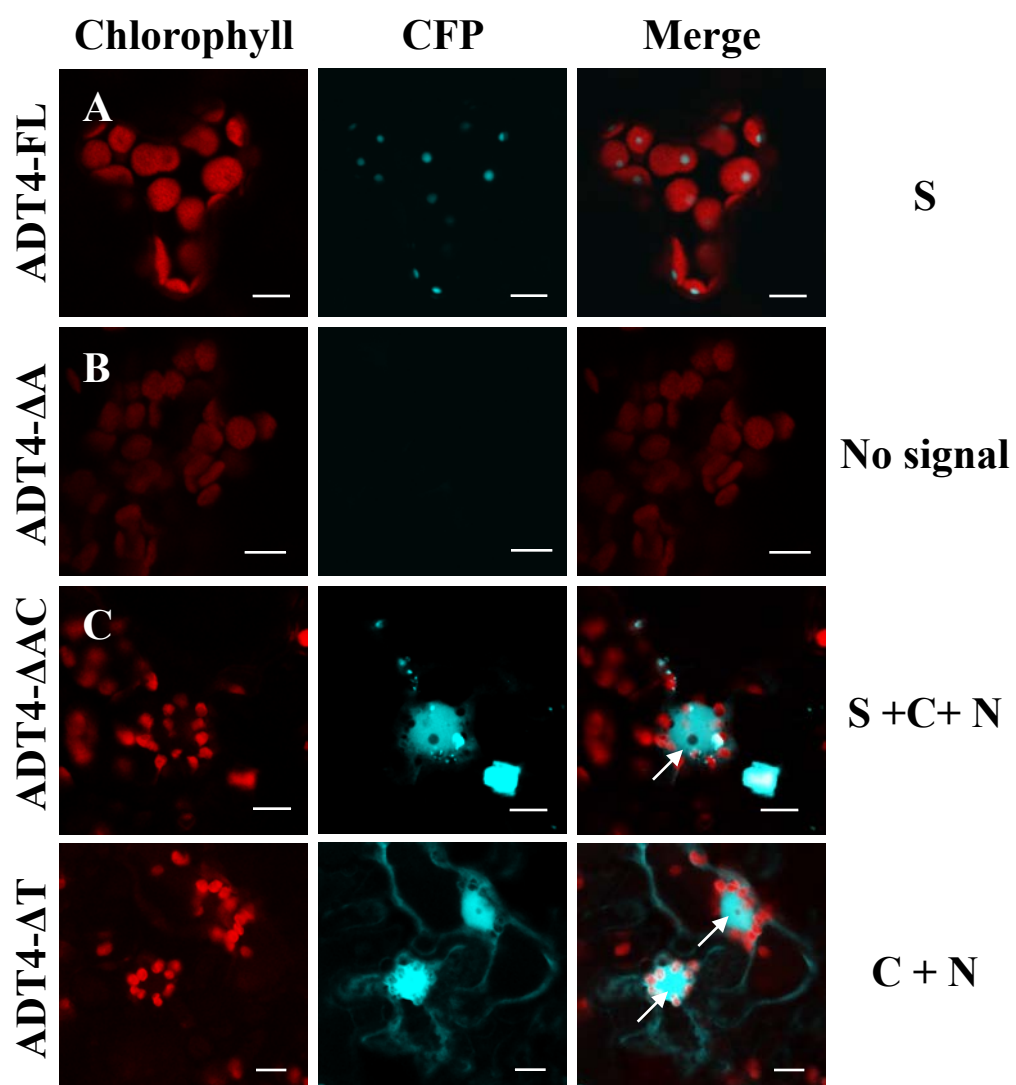
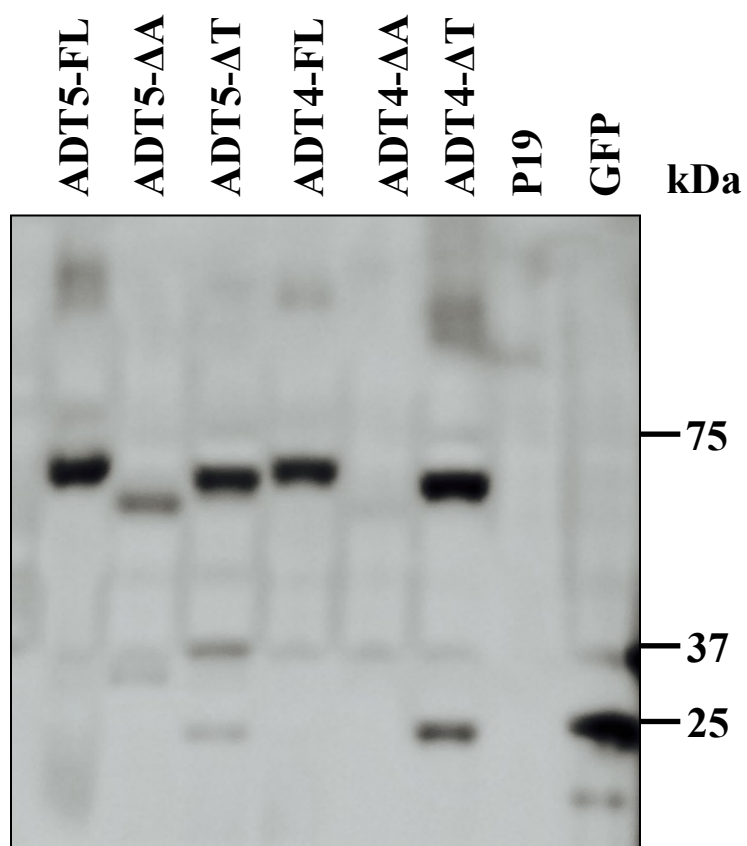


Figure 25 Protein expression of the full length and truncated ADT5 and ADT4 in *N. benthamiana*.

Total soluble proteins were isolated from *N. benthamiana* leaves infiltrated with ADT5-FL, ADT5-ΔA, ADT5-ΔT, ADT4-FL, ADT4-ΔA, and ADT4-ΔT. As controls, isolated proteins samples from GFP and p19 were used. Ten μg of TSP was size-separated by SDS-PAGE, blotted with nitrocellulose membrane and detected with a GFP antibody. The predicted protein sizes for fusion proteins are ADT5-FL: 73.8 kDa, ADT5-ΔA: 65.3 kDa, ADT5-ΔT: 66.2 kDa, ADT4-FL: 73.8 kDa, ADT4-ΔA: 65.3 kDa, ADT4-ΔT 66.3 kDa and GFP: 25 kDa.



in the nucleus or cleaved fluorescent products.

In summary, the deletion constructs did not directly identify a NLS in ADT5. However, the observations on confocal images and Western blot data suggest that the ACT domain might be important to ensure stable expression of ADT4 and ADT5 fusion proteins.

3.4.3 Generating chimeric ADT5-ADT2 proteins

As the deletion constructs did not clarify the location of a potential NLS within ADT5, an alternative approach was chosen. Domains were swapped between ADT5 (nuclear localization) and ADT2 (no nuclear localization) to generate chimeric ADT5-ADT2 fusion proteins. The advantage of this approach is the expressed chimeric proteins are all full length, above 70 kDa, which will avoid the possibility of diffusion or no or low accumulation due to lacking a portion of the protein. Since ADT2 was never observed in the nucleus, it was expected that only fusion proteins with an ADT5 domain containing a NLS would be able to enter into the nucleus. ADT5-ADT2 chimeric fragments, previously generated by Smith-Uffen (2014), were amplified and introduced into the plant destination vector pEarlyGate102. Resulting chimeric constructs were named “C” constructs (C1-C7) and are shown in Figure 26. C constructs were introduced into *Agrobacterium* GV3101 and then were co-infiltrated with p19-carrying *Agrobacterium* cells in *N. benthamiana* leaves. Localization of the fusion proteins was visualized at 4 dpi by confocal microscopy. ADT5-FL and ADT2-FL (Bross, 2011) were used as controls. As observed before (Figure 6) ADT2 was only present in stromules (Figure 27A) while ADT5 was observed in the stromules and nucleus (Figure 27B). All chimeric fusion proteins were expressed and observed in the chloroplasts and specifically stromules (Figure 28) meaning that domain swapping did not interfere with the expression or accumulation of the proteins. Only three chimeric proteins (C1, C5 and C6) were detected in the nucleus (Figure 28A, Figure 28E and Figure 28F, respectively). These three chimeric proteins had only the N-terminal portion of the ADT5 ACT domain in common. Therefore, it was concluded that the ADT5 N-terminal ACT domain had sequences that target the fusion proteins into the nucleus.

Figure 26 Schematic representation of full-length and chimeric ADT2/ADT5 constructs.

To identify the domain that potentially had the NLS, different domains were swapped between ADT5 and ADT2. Schematic not drawn to scale.

ACT: ACT regulatory domain; CAT: catalytic domain; CFP: Cyan Fluorescence Protein; 35 S: CaMV 35S promoter; TP: transit peptide.

Dashed pattern: ADT2 domains, solid pattern: ADT5 domains.

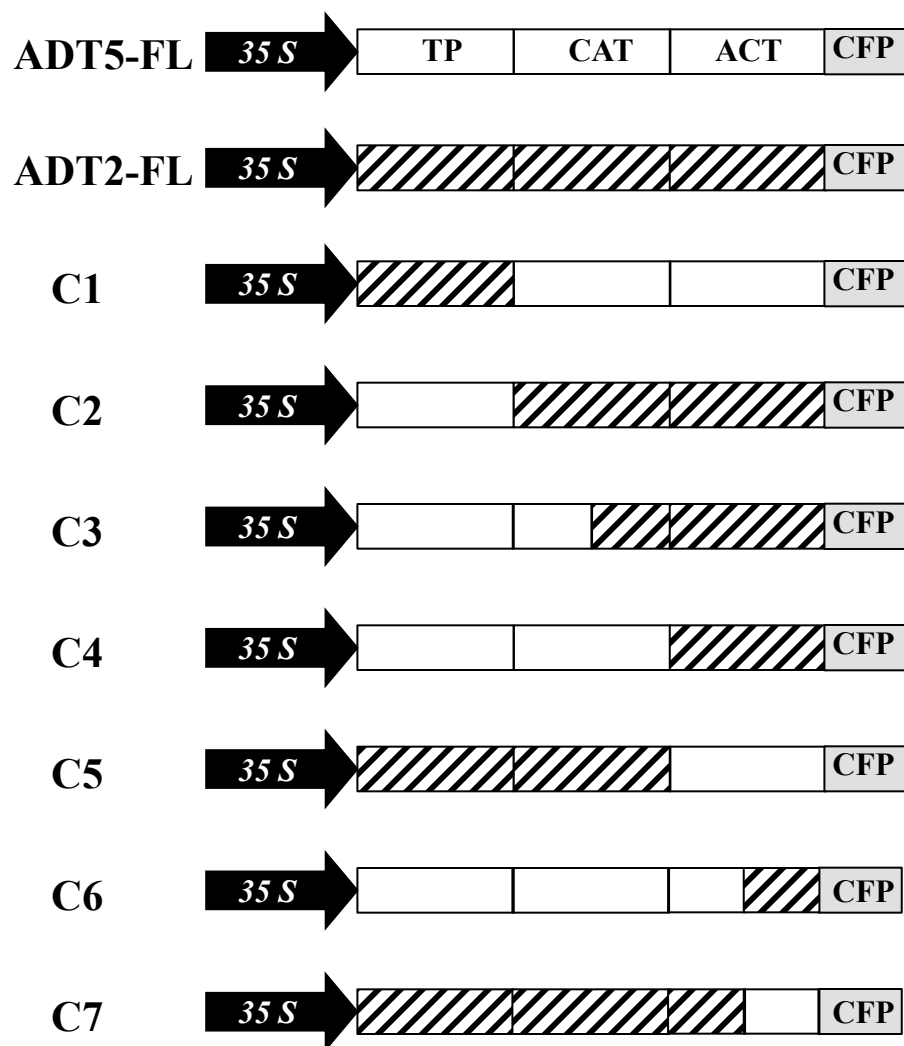


Figure 27 Subcellular localization of the ADT2-CFP and ADT5-CFP in *N. benthamiana* leaves.

To compare the expression pattern of ADT2-CFP and ADT5-CFP with the C constructs, *N. benthamiana* leaves were either infiltrated with *Agrobacterium* cells harboring ADT2-CFP or ADT5-CFP constructs. The localization pattern was visualized 4 dpi by confocal microscopy.

Chlorophyll fluorescence and CFP fluorescence are shown separately in the left and middle column, respectively, and a merged image is shown in the right column. Construct name is shown on the left side and a summary of the confocal observations is shown on the right side.

Scale bars are 5 μm .

(A) ADT2-CFP is detected in the chloroplast stromules.

(B) ADT5-CFP is detected in the stromules and nucleus.

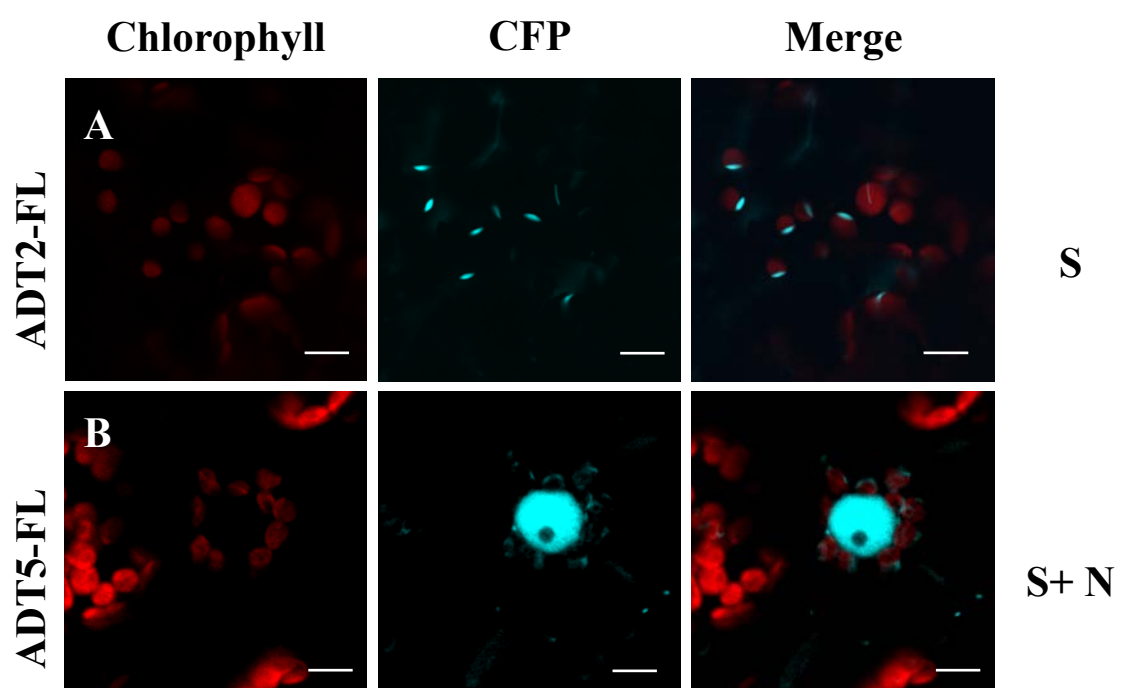


Figure 28 Subcellular localization of the C1-C7 chimeric proteins in *N. benthamiana* leaves.

Agrobacterium cells harboring C constructs were transiently expressed in *N. benthamiana* leaves and the subcellular localization of the chimeric proteins was visualized at 4 dpi.

Chlorophyll fluorescence and CFP fluorescence are shown separately in the left and middle column, respectively, and the merged image is shown in the right column. Construct name and schematic is shown on the left side and a summary of the confocal observations is shown on the right side.

C: Cytoplasm, N: Nucleus, S: Stromule.

Scale bars are 5 μm .

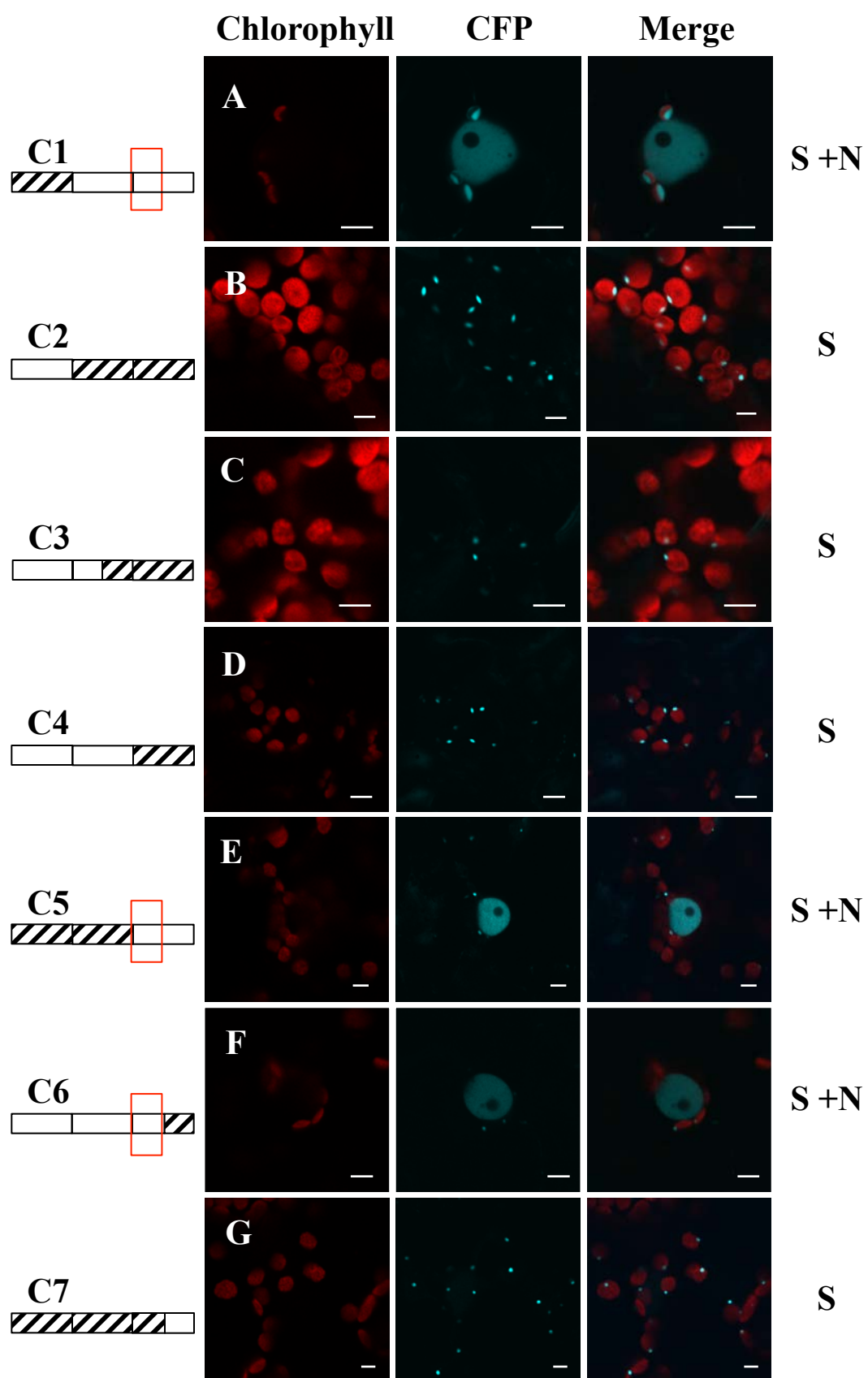
(A) C1 is observed in stromules and nucleus.

(B), (C) and (D) C2, C3 and C4 are visualized only in stromules.

(E) and (F) C5 and C6 are observed in stromules and nucleus.

(G) C7 is found in stromules.

Red box: N-terminal ADT5 ACT domain.



3.4.4 N- terminal ACT domain *in silico* analysis to find the NLS

To identify the amino acid sequence unique to ADT5 and potentially responsible for targeting the fusion proteins into the nucleus, the ACT domain sequence of all six ADTs were aligned (Figure 29). ADTs are highly similar in their N-terminal ACT domain and there are more variations in the C-terminus of the ACT domain. A motif of 4 amino acids (AQEH) at position 8 to 11 of ADT5 ACT domain was found unique to ADT5 and ADT4. The only difference between ADT4 and ADT5 is an asparagine (Asn) at position 28 N-terminus of ADT5 ACT domain which is substituted with an aspartic acid (Asp) in ADT4. However, an asparagine is present at the same position in ADT3, while neither ADT4 nor ADT3 has been observed in the nucleus. Therefore, it was concluded that presence of the AQEH motif and Asn28 in N-terminus of ADT5 ACT domain are potentially important for nuclear targeting of ADT5. In addition, the ACT domain sequence of ADT5 was submitted to the ISIS program (Ofra and Rost, 2007). This program predicts the protein-protein interaction sites. Interestingly, the histidine (His) in the AQEH motif and Asn28 were predicted as protein interaction regions in N-terminal ACT domain. Hence, it was concluded that ADT5 nuclear localization could happen through protein-protein interactions.

3.5 Protein-protein interaction and nuclear localization

Protein complexes and protein-protein interactions mediate many molecular processes in the cell. For plants, protein interactions have been identified that allow proteins to piggyback into the nucleus (Genoud et al., 2008). Hence, ADT5 nuclear targeting may happen via interactions with other proteins as part of a piggyback system. To explore this possibility, protein interaction partners of ADT5 were identified.

3.5.1 *In silico* analysis of potential ADT5 interactors


Initially, to identify potential ADT5 interactors, the Bio-Analytic Resource for Plant Biology (BAR) (Waese and Provart, 2016) was used. The interaction viewer tool in this database shows for any given *A. thaliana* protein predicted or confirmed interacting proteins. Based on this website ADT5 has the following predicted interactors: ALPHA CRYSTALLINE DOMAIN 31.2 (ACD 31.2), ADT4, IMPORTIN ALPHA ISOFORM 6

Figure 29 Sequence alignment of ACT domain of six *Arabidopsis* ADTs.

An alignment of the ACT domain of six *Arabidopsis* ADTs was performed to find an amino acid sequence unique to ADT5 N-terminal ACT domain that potentially can target ADT5 into the nucleus. A motif of 4 amino acids (AQEH) at position 8 to 11 from ACT domains was found unique to ADT5 and ADT4. At position 28, Asn in ADT5 is substituted with Asp in ADT4.

Light blue: N-terminal ACT domain, dark blue: C-terminal ACT domain, orange boxes: the identified motif unique to ADT4 and ADT5, yellow box: the identified position that one amino acid is different between ADT4 and ADT5.

Numbers on the left are count acids from the beginning of the ACT domain.



ADT1 KTSTVESL..EEGFGVLEKALAVEAIPSTINISKIESFEQFFRFLRWGCSNNGSAKYETLYLFYILFEASMAITRAQHAIGHIQEEASHIRIICQYEMDIVF..... 99
 ADT2 KTSTVESL..EEGFGVLEKALAVEAIPQINLTIESFRLRKHPLFASCG.....IKYETLYLFYILFEASVALEVAQNALRHLEEFATHIRVIGSYEMITMI..... 95
 ADT3 KTSTVEAH..EKGTQVLEKVLSAFAERNISITKIESRFENENVEIRIVDEANVCTAKHEEYMFYILFEASVAESRAQNALSEVQEFTSHIRVIGSYEMMTISWSESSSSSSSTFS 113
 ADT4 KTSTVEAAQEEHKCTSVLEKVLSAFAERDISITKIESRFENENRFLRWGCGSECTSKNEEYMFYILFEASVAEFRAQNALAEVQEYTSIRVIGSYEMMTIFWSTIESTEEA..... 110
 ADT5 KTSTVEAAQEEHKCTSVLEKVLSAFAERDISITKIESRFHÇNCFRVVGLENVCTSKHEEYTFYILFEASVAEFRAQNALAEVQEYTSIRVIGSYEMMTIFWSTIESTELV..... 110
 ADT6 KTSTVEAH..EKGTQVLEKVLSAFAERDISITKIESRFENENRFLRWLLANVCTAKHEEYMFYILFEASVAEFRAQNALAEVQEFTSHIRVIGSYEMMTIFWSTISSTISS..... 108

Table 1 Predicted ADT5 interactors in BAR database.

Description	Abbr.	AGI NO.	Function	Localization	Reference
ALPHA- CRYSTALIN DOMAIN 32.1	ACD 32.1	At1g06460	Heat response	Peroxisomes ^a	(Eubel et al., 2008)
AROGENATE DEHYDRATASE 4	ADT4	At3g44720	Phe biosynthesis	Chloroplasts ^a	(Bross et al., 2017)
IMPORTIN ALPHA ISOFORM 6	IMPA6	At1g02690	Protein import into nucleus	Cytoplasm ^a , nucleus ^a	(Marín et al., 2012)
ZIM 17 TYPE ZINC FINGER 3	ZR3	At3g54826	Protein folding, protein import into mitochondrial matrix	Mitochondria ^a	(Kluth et al., 2012)
REDUCED LATERAL ROOT FORMATION	RLF	At5g09680	Lateral root formation	Cytoplasm ^a , nucleus ^b	(Ikeyama et al., 2010)

^aExperimentally determined, ^bPredicted.

(IMPA6), ZIM 17 TYPE ZINC FINGER 17 and REDUCED LATERAL ROOT FORMATION (RLF; Table 1). Among these predicted interactors, IMPA6 was a prime candidate for guiding proteins into the nucleus. IMPA6 is a member of the importin α family in *A. thaliana* that acts as a cytosolic receptor and facilitates translocation of a wide range of proteins across the NPC (Smith et al., 1997). Therefore, it is possible that IMPA6 can interact with ADT5 and assist in the transfer into the nucleus through the nuclear envelope. As protein interactions listed on the BAR are based on large-scale Y2H screens or biochemical methods (Geisler-Lee et al., 2007) it is important to exclude the possibility of false positives. For this reason two protein-protein interaction assays were used: the Y2H assay that tests a binary interaction between the two proteins and BiFC assay that tests the protein interaction *in planta*.

3.5.1.1 *In vivo* protein interaction of ADT5 and IMPA6

The Y2H assay was used to confirm the interaction between ADT5 and IMPA6. The full-length coding sequences of *IMPA6* and *ADT5* were PCR amplified using primers, which incorporate *att* sites for Gateway cloning (Appendix 1). The PCR-amplified fragments were recombined into Y2H gateway vectors pGBKT7-DEST (GAL4-DB) and pGADT7-DEST (GAL4-TA) (section 2.3.4). Each construct was transferred individually into the yeast AH109 strain and tested for self-interaction. Each strain was also co-transformed with a corresponding empty DB or TA vector, or a vector expressing an unrelated *CRUCIFERINA* (*CRA1*; accession At5g44120) DB- or TA-fusion constructs. *CRA1* is a seed storage protein and is often used as a negative control in protein-protein interaction assays for ADTs (Kohalmi et al., 1998). Activation of the three reporter genes *HIS3*, *ADE2* and *MEL1* was evaluated and it was found that none of the ADT5 and IMPA6 fusion proteins (DB and TA-fusions) activated the reporters alone (Figure 30). As well, no interaction was observed between DB-ADT5, TA-ADT5 and TA-IMPA6 with either the corresponding empty vector or the *CRA1* fusion protein (Figure 30B, Figure 30D and Figure 30E). However, DB-IMPA6 was able to activate all the three reporter genes in the absence of an interacting protein, and also activated the *HIS3* when expressing with TA-*CRA1* (Figure 30B and Figure 30C). Therefore, *DB-IMPA6* was not used for any further experiments and only the interaction ability of the

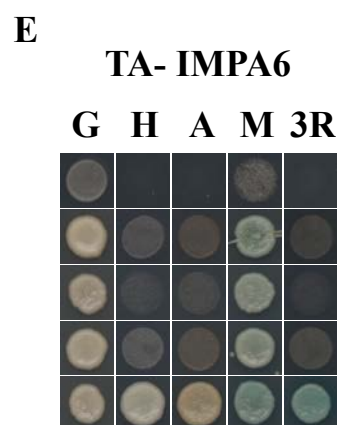
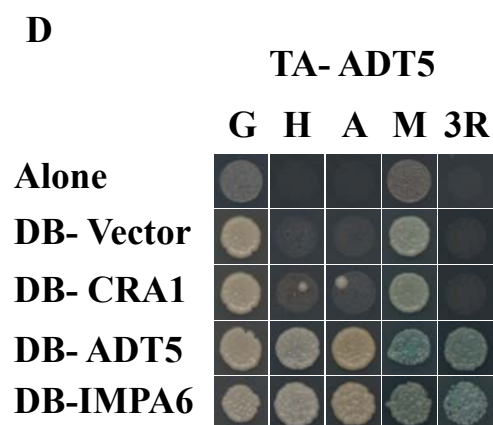
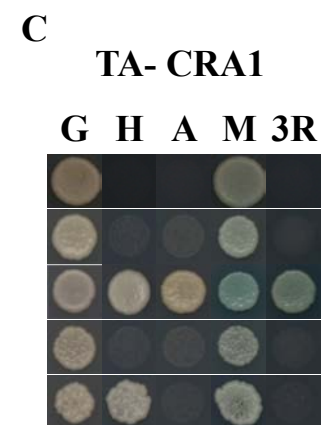
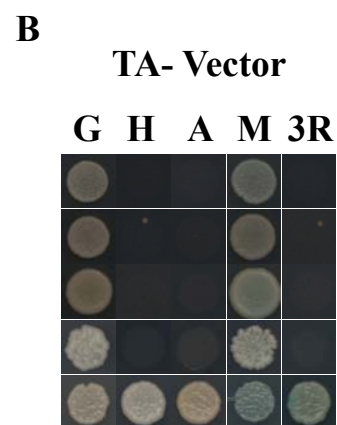
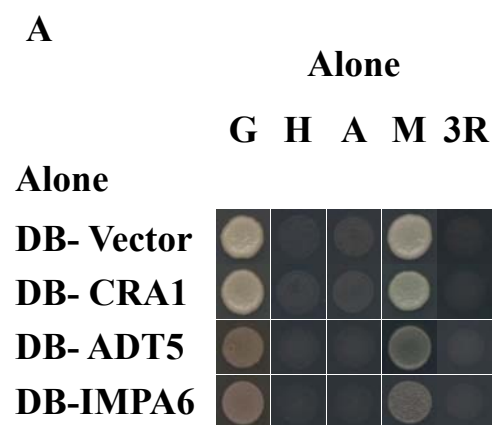
Figure 30 Control transformations and interactions involving ADT5 and IMPA6 assayed by Y2H in Yeast AH109.

Yeast transformants were assayed for growth, histidine prototrophy (H), adenine prototrophy (A), α -galactosidase activity (M) as well as the activation of the all three reports simultaneously (3R). Growth of white cells in the absence of histidine, growth of white cells in the absence of adenine, and production of a blue pigment in the presence of X- α -gal indicates activation of the reporter genes and interaction. Growth of blue colonies is expected when all 3 reporters are activated simultaneously.

Activation of the reporter genes was tested for each construct alone, in combination with the corresponding empty construct and CRA1 constructs as controls as well as homo- and heterodimerization of ADT5 and IMPA6.

- (A) None of the DB and TA constructs were able to activate the reporter genes alone.
- (B) Co-transformation of TA-vector with corresponding DB vectors resulted no interactions except for DB-IMPA6.
- (C) No interaction was observed for TA-CRA1 corresponding DB vectors resulted no interactions except for DB-IMPA6. CRA1 was able to form homodimer.
- (D) No interaction was detected for TA-ADT5 and DB vector or DB-CRA1. ADT5 was able to form homodimer.
- (E) TA-IMPA6 did not interact with corresponding DB vectors. Homodimerization of IMPA6 was detected.
- (F) Table shows a summary of the interactions.

A: adenine prototrophy; DB: GAL4 DNA binding domain; G: growth; H: histidine prototrophy; M: *MEL1* α -galactosidase; TA: GAL4 transactivation domain; 3R: 3 reporter genes simultaneously.



F

	Alone	TA- vector	TA- CRA1	TA- ADT5	TA- IMPA6
Alone		-	-	-	-
DB- vector	-	-	-	-	-
DB- CRA1	-	-	+	-	-
DB- ADT5	-	-	-	+	-
DB- IMPA6	-	+	+		+

fusion proteins DB-ADT5 and TA-IMPA6 was evaluated. It was found that none of the reporter genes were activated and hence it was concluded that ADT5 and IMPA6 are not able to interact in the Y2H system (Figure 30E).

3.5.2 *In planta* interaction of ADT5 and IMPA6

As an alternative to Y2H system, the interaction of IMPA6 and ADT5 was examined with the *in planta* BiFC assay. With this assay the interaction of the proteins can be detected in a plant environment and the subcellular localization of the interactions can be visualized. *IMPA6* and *ADT5* were PCR amplified using primers, which incorporate the *att* sites for Gateway cloning system (Appendix 1). *IMPA6* and *ADT5* were recombined into the BiFC vectors, pEarlyGate 201-nYFP and pEarlyGate 202-cYFP (section 2.3). Verified constructs were transformed into *Agrobacterium* cells and were transiently expressed in *N. benthamiana* leaves. Prior to analysis of ADT5 and IMPA6 interaction, the constructs were infiltrated alone to ensure they cannot fluorescence in the absence of a corresponding protein. As well, *YN-CRA1* and *YC-CRA1* were co-infiltrated with corresponding *ADT5* and *IMPA6* constructs to detect if random collisions that were not specific protein interactions lead to fluorescent signal. None of the ADT5 and IMPA6 fluoresced on their own and no fluorescence from random collisions was detected (Figure 31). Next dimerization of ADT5 and IMPA6 was tested and it was observed that both ADT5 and IMPA6 were able to form homodimers and heterodimers. The ADT5 homodimers were detected in the chloroplast stromules and nucleus (Figure 32A) and the IMPA6 homodimers were detected in the nucleus (Figure 32B). The heterodimers of ADT5 and IMPA6 were observed only in the nucleus (Figure 32C and Figure 32D). These data confirm that ADT5 and IMPA6 interact *in planta*.

As only ADT5 was detected in the nucleus (Bross, 2011) it was tested if the interaction with IMPA6 was specific to ADT5. To do that, IMPA6 was co-infiltrated with other ADTs. All ADTs were able to interact with IMPA6 and the interaction was detected in the nucleus (Figure 33). On the basis of these observations it was concluded that ADT5 and IMPA6 interact *in planta* however this interaction is not unique to ADT5.

Figure 31 Control transformations ADT5 and IMPA6 fusion constructs in *N. benthamiana*.

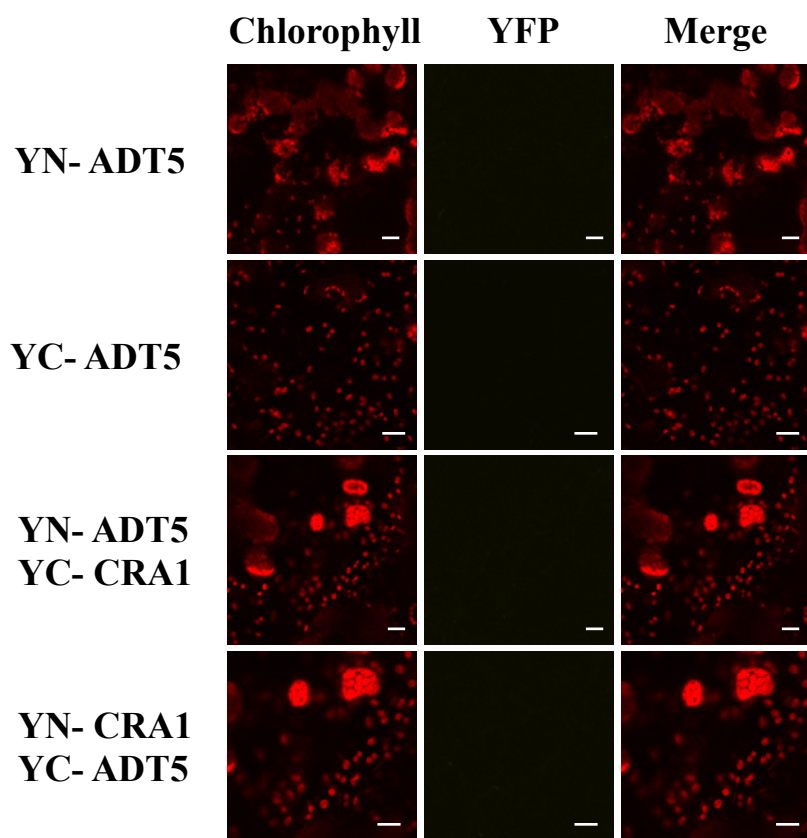
ADT5 and IMPA6 *YN*- and *YC*- constructs were transiently expressed in *N. benthamiana* leaves alone or with corresponding *YC-CRA1* and *YN-CRA1* constructs and analyzed at 3 dpi. Expression combinations are identified to the left.

Chlorophyll fluorescence and YFP fluorescence are shown separately in the left and middle column, respectively, and the merged image is shown in the right column.

Scale bars are 5 μm .

No YFP signal was detected in leaf samples infiltrated with fusion constructs alone or in combination with corresponding CRA1 constructs.

A



B

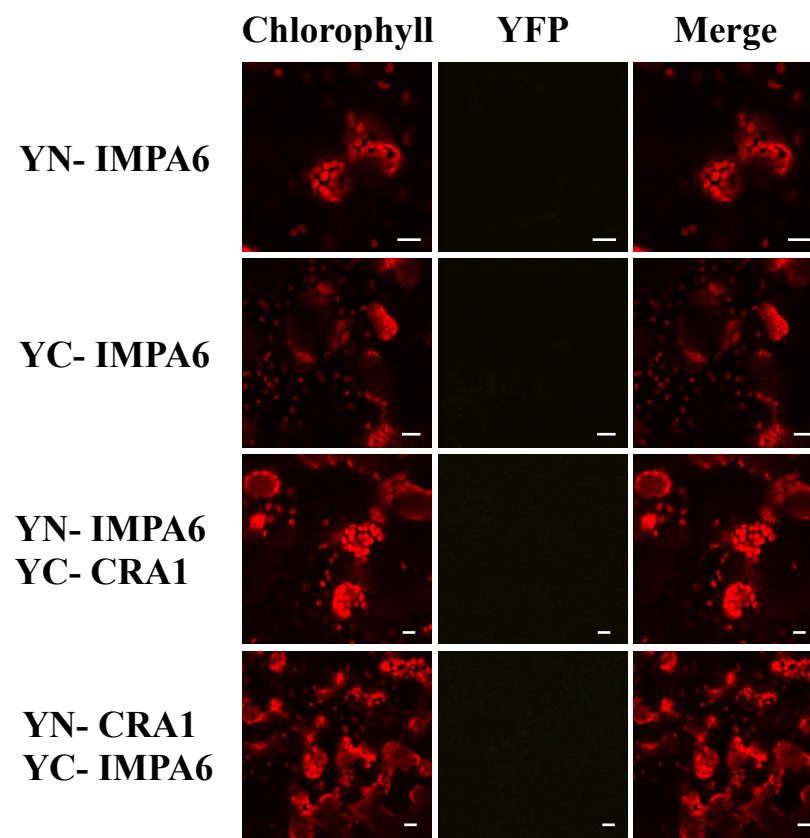


Figure 32 Interactions involving ADT5 and IMPA6 assayed by BiFC in *N. benthamiana* leaves.

To identify homodimerization and heterodimerization between ADT5 and IMPA6 fusion proteins, corresponding constructs were co-infiltrated in *N. benthamiana* leaves. The proximity of two fusion proteins results in a functioning fluorophore. Yellow fluorescence indicates dimer formation.

Chlorophyll fluorescence and YFP fluorescence are shown separately in the left and middle column, respectively, and the merged image is shown in the right column. Summary of the observations are shown on the right side of the panel.

C: Cytoplasm, N: Nucleus, S: Stromule.

Scale bars are 5 μm .

- (A) ADT5 forms homodimers, which are detected in the stromules and nucleus.
- (B) IMPA6 forms homodimers, which are localized in the nucleus.
- (C) Heterodimerization of YN-ADT5 and YC-IMPA6 is observed in the nucleus.
- (D) Heterodimerization of YC-ADT5 and YN-IMPA6 is detected in the nucleus.

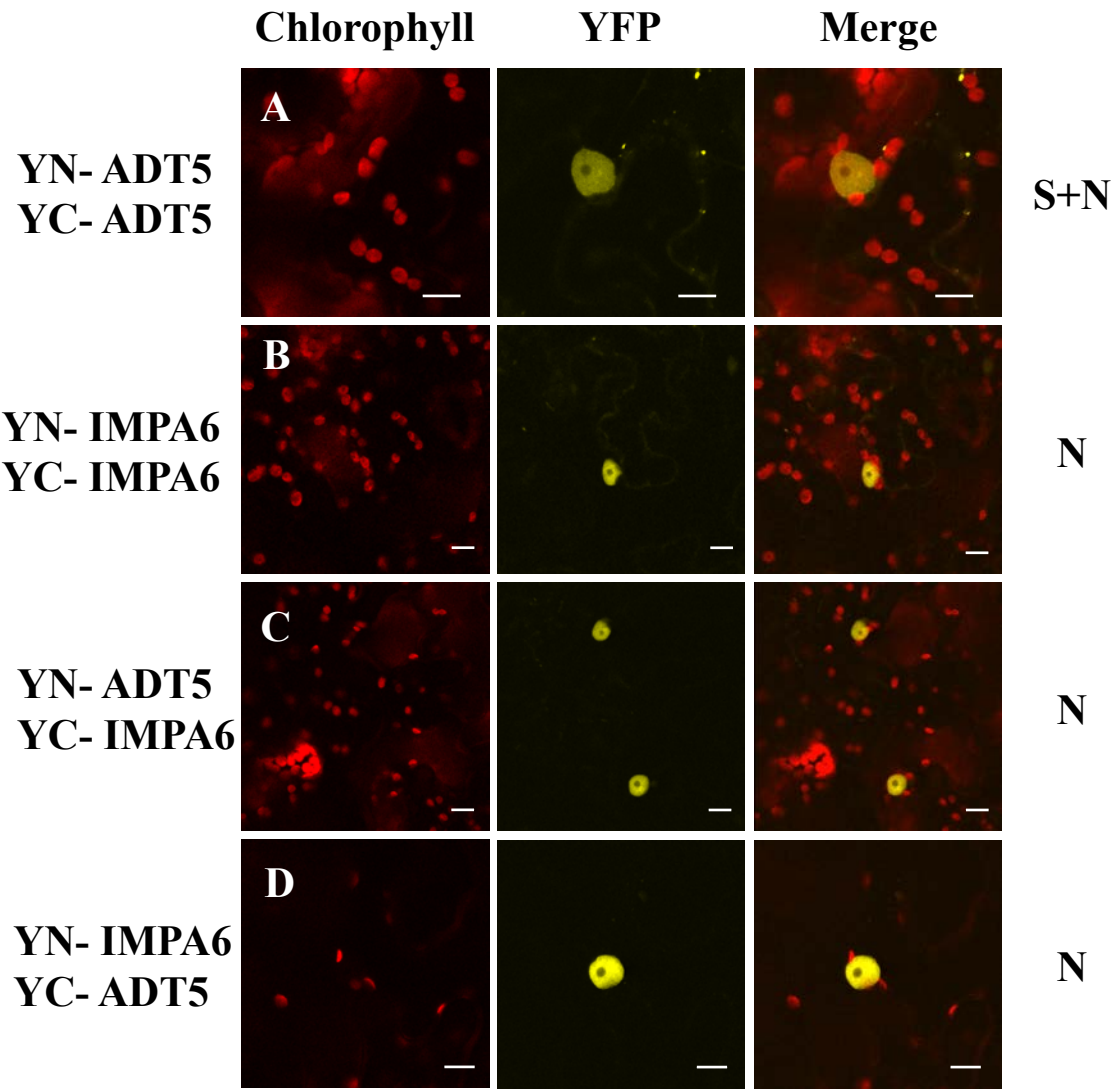


Figure 33 Interactions involving IMPA6 and ADT5 assayed by BiFC in *N. benthamiana* leaves.

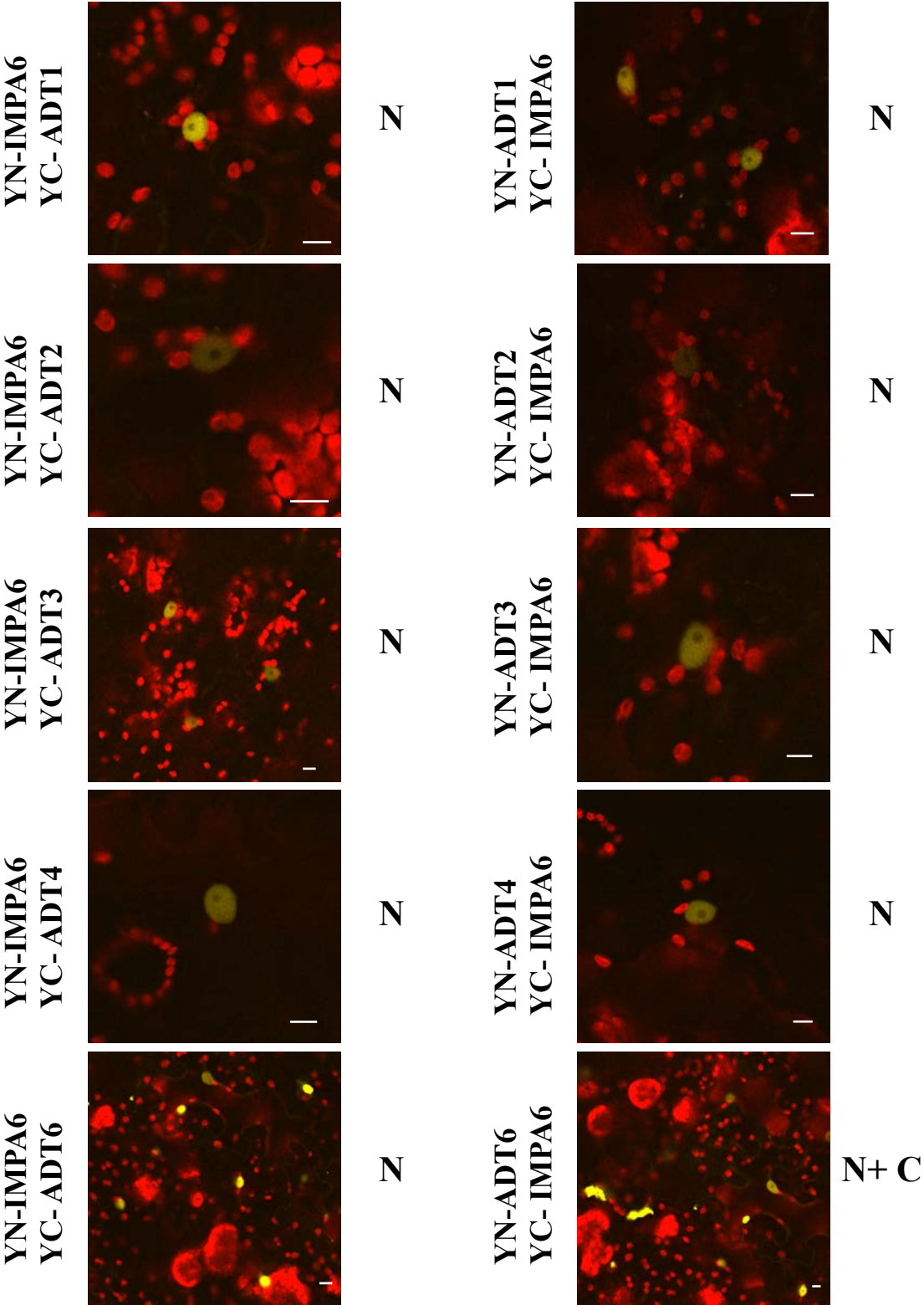
Interactions between IMPA6 and ADTs were assayed by co-expression of corresponding IMPA6 and ADT constructs in *N. benthamiana* leaves using BiFC assay. The proximity of two fusion proteins results in a functioning fluorophore. Yellow fluorescence indicates dimer formation.

Only merged images are shown. Summary of the observations is shown on the right side of the panel.

C: Cytoplasm, N: Nucleus.

Scale bars are 5 μm

All the six ADTs are able to form heterodimers with IMPA6 and the interactions are observed in the nucleus.



3.6 Identifying novel ADT5 interactors

Since ADT5 is the only ADT that localizes into the nucleus, another mechanism must be in play for its import into the nucleus. It is possible that it interacts with other proteins and uses the piggyback system for transport in the nucleus. To gain a better understanding of ADT5 putative interactors, a large scale *A. thaliana* Y2H cDNA library screen was performed using yeast AH109 cells expressing *DB-ADT5* (section 3.5.1.1) as bait and commercial Mate and Plate Library in yeast Y187 cells as prey (section 2.9). Mated cells were plated on two types of media: (A) selecting for activation of the *HIS* and *ADE2* reporter genes and (B) selecting for activation of the *AURI* reporter. All plates were incubated at room temperature for 5 days. Viability of haploid AH109 harboring bait plasmid and Y187 cells harboring prey plasmids as well as diploids indicated that a mating efficiency of 3.3% was achieved (Appendix 3). Optimal mating efficiency using this library has a range of 2%-5% (Clontech, 2010).

In the primary screening, 72 colonies were found on media A and 438 colonies on media B. Then, all the 72 colonies from group A and 26 colonies from group B were tested for activation of the three reporter genes *HIS*, *ADE2* and *AURI* individually and simultaneously. None of the colonies from media B were able to activate the *HIS* or *ADE2* reporter genes, therefore the rest of colonies in this group were not analyzed further. Sixty-seven colonies from media A were able to activate all three reporters. Plasmid DNA was isolated from these positive colonies and the insert of the prey plasmids were sequenced. Sequencing results were submitted to the TAIR website (The *Arabidopsis* Information Resource) to identify the genes of putative ADT5 interactors in *A. thaliana* genome. The screen identified 21 proteins as putative interactors including 3 ADTs, 2 unknown proteins and 16 new interactors. All the interactors were classified in different groups based on their subcellular localization in the cell (Table 2). Seven proteins were classified as nuclear localized proteins.

The ADTs were not processed further as ADTs hetero- and homodimerization were studied previously in our lab (Styranko, 2011). Some interactors were recognized more than once; in that case only one full-length sequence representative was used for further analysis. To evaluate the interaction between ADT5 and newly identified interactors, several control experiments were performed. For all analyses, activation of

Table 2 Putative ADT5 interactors identified by Y2H screening.

Interactor Name	Abbr.	Identification	Interactions Recorded	Function	Reference
Proteins with nuclear localization					
SNF1 KINASEK HOMOLOG 10	KIN10	At3g01090	6	Protein binding, lipid biosynthesis	(Zhai et al., 2017)
NUCLEAR TRANSPORT FACTOR 2	NTF2	At5g43960	1	Nucleo-cytoplasmic transport	(Reichel et al., 2016)
POLYPEPTIDE RNA-BINDING KH DOMAIN	PEP	At4g26000	1	RNA processing	(Lee et al., 2011)
NUCLEOSOME/CHROMATIN ASSAMBELY FACTOR D	HMG	At3g28730	4	DNA repair, DNA replication,	(Launholt et al., 2007)
EMBRYO DEFECTIVE 1507	BRR2	At1g20960	2	RNA splicing, embryo development	(Golisz et al., 2013)
EIN3-BINDING F BOX PROTEIN 1	EBF1	At2g25490	1	Ethylene activated signaling pathway	(Gagne et al., 2004)
AROGENATE DEHYDRATASE 5	ADT5	At5g22630	1	Phe biosynthesis	(Bross et al., 2017)
Proteins with ER localization					
PHOSPHOLIPID DISCYLGLYCEROL ACYLTRSNDERSASE	PDAT	At5g13640	1	Glycerol metabolic process, Lipid metabolic process	(Stahl et al., 2004)
TRANSLOCON-ASSOCIATED PROTEIN	TRAP	At2g21160	1	Regulate the retention of ER proteins	(Wang et al., 2008)
Proteins with chloroplast localization					
AROGENATE DEHYDRATASE 6	ADT6	At1g08250	34	Phe biosynthesis	(Bross et al., 2017)

Interactor Name	Abbr.	Identification	Interactions Recoded	Function	Reference
AROGENATE DEHYDRATASE 2	ADT2	At3g07630	2	Phe biosynthesis	(Bross et al., 2017)
TRAF LIKE FAMILY PROTEIN	TRAF	At5g26280	1	Unknown	(Bao et al., 2014)
DNAJ HEAT SHOCK FAMILY PROTEIN	DJA6	At2g22360	1	Response to heat	(Zhichang et al., 2010)
Proteins with mitochondria localization					
MITOCHONDRIAL PYRUVATE CARRIERS1	MPC1	At5G20090	1	Pyruvate transmembrane transporter	(Li et al., 2014)
2-OXOGLUTARATE DEHYDROGENASE	2OXO*	At5g65750	1	Response to cadmium ion	(Millar et al., 1999)
Proteins with golgi localization					
MANNAN SYNTHESIS RELATED 1	MSR1	At3g21190	1	Transferring glycosyl groups	(Wang et al., 2013)
PHOSPHATIDYLINOSIT OL 4-OH KINASE BETA1	PI4KBE	At5g64070	1	Phosphatidylinositol phosphorylation	(Janda et al., 2014)
Proteins with vacuole localizations					
SAPOSIN B DOMAIN- CONTAINING PROTEIN	SaBP*	At3g51730	1	Lipid metabolic process	(Carter, 2004)
Proteins with cytosolic localizations					
PROTEIN PHOSPHATASE 2C FAMILY PROTEIN	PP2C*	At1g47380	1	Protein dephosphorylation	(Xue et al., 2008)
Unknown proteins					
-	Un1*	At1g11125	5	-	-
-	Un2*	At2g47010	1	-	-

* Random abbreviations were made for analysis.

four reporter genes *HIS3*, *ADE2*, *MEL1* and *AURI* was recorded. When testing for self-activation ability it was found none of them were able to activate the reporters alone (Figure 34). Then, the interactors were co-transformed with DB empty vector to see if they were able to activate the transcription of the reporter genes in the absence of ADT5. Some of the interactors including PEP, TRAF, DJA6, MSR1 and PI4K were able activate the reporter genes. These proteins were excluded from further analysis. The remaining ADT5 interactors were co-transformed with all six ADTs to determine if they were only interacting with ADT5 (Figure 35 and Figure 36). All the interactors were able to interact with at least one of the ADTs other than ADT5, except for one protein that was identified to specifically interact with ADT5 (Figure 37). The ADT5 specific interactor was PHOSPHOLIPID DIACYLGLYCEROL ACYLTRANSFERASE1 (PDAT1) which has a known enzymatic role in glycerol biosynthesis (Dahlqvist et al., 2000). This protein has been described as an integral protein present in the endoplasmic reticulum (ER) but it is also predicted to have a nuclear localization. Therefore, PDAT1 was a piggyback protein candidate and was used for further analysis.

3.7 PDAT1 *In silico* analysis

AtPDAT1 (At5g13640) was identified as an interactor specific to ADT5 in *A. thaliana*. This protein is 671 amino acids in length and has a calculated molecular weight of 74.15 kDa (The *Arabidopsis* Information Resource). PDAT was first characterized in three different oil seeds: sunflower, castor bean and *Crepis palaestina* and later on in yeast (*ScPDAT*), as an enzyme that catalyzes an acyl-CoA-independent formation of triacylglycerol as a homolog of human lecithin: cholesterol acyltransferase (*HsLCAT*) (Dahlqvist et al., 2000). In addition to PDAT1, five gene sequences with sequence similarity to *ScPDAT* and *HsLCAT* have been identified in the *Arabidopsis* genome database (*AtLCAT1*: At1g27480, *AtLCAT2*: At3g44830, *AtLCAT3*: At3g03310, *AtPSAT1*: At1g04010 and *AtPSAT2*: At4g19860) (Stahl et al., 2004). To identify the extent of sequence similarity between these *A. thaliana* sequences, an amino acid alignment was performed using the full-length sequences and a phylogenetic tree was generated (Figure 38A). It was found that *A. thaliana* sequence similarity values range

Figure 34 Control transformation for ADT5 interactors alone and with DB vector in Yeast AH109.

Each construct was transformed individually and co-transformed with DB vector in yeast AH109 and was assayed for growth, histidine prototrophy, adenine prototrophy, α -galactosidase activity and *AURAI* activity as well as the activation of all four reporters simultaneously. Growth of white colonies in the absence of histidine, growth of white colonies in the absence of adenine, production of a blue pigment in the presence of X- α -gal and growth of white colonies in the presence of ABA indicate an interaction. Growth of blue colonies is expected when all 4 reporters are activated simultaneously.

None of the interactors were able to activate the reporter genes alone. However, some of them were able to interact with the empty DB-GAL4.

Red boxes: Interactors that were able to activate the reporter genes in the absence of a second protein.

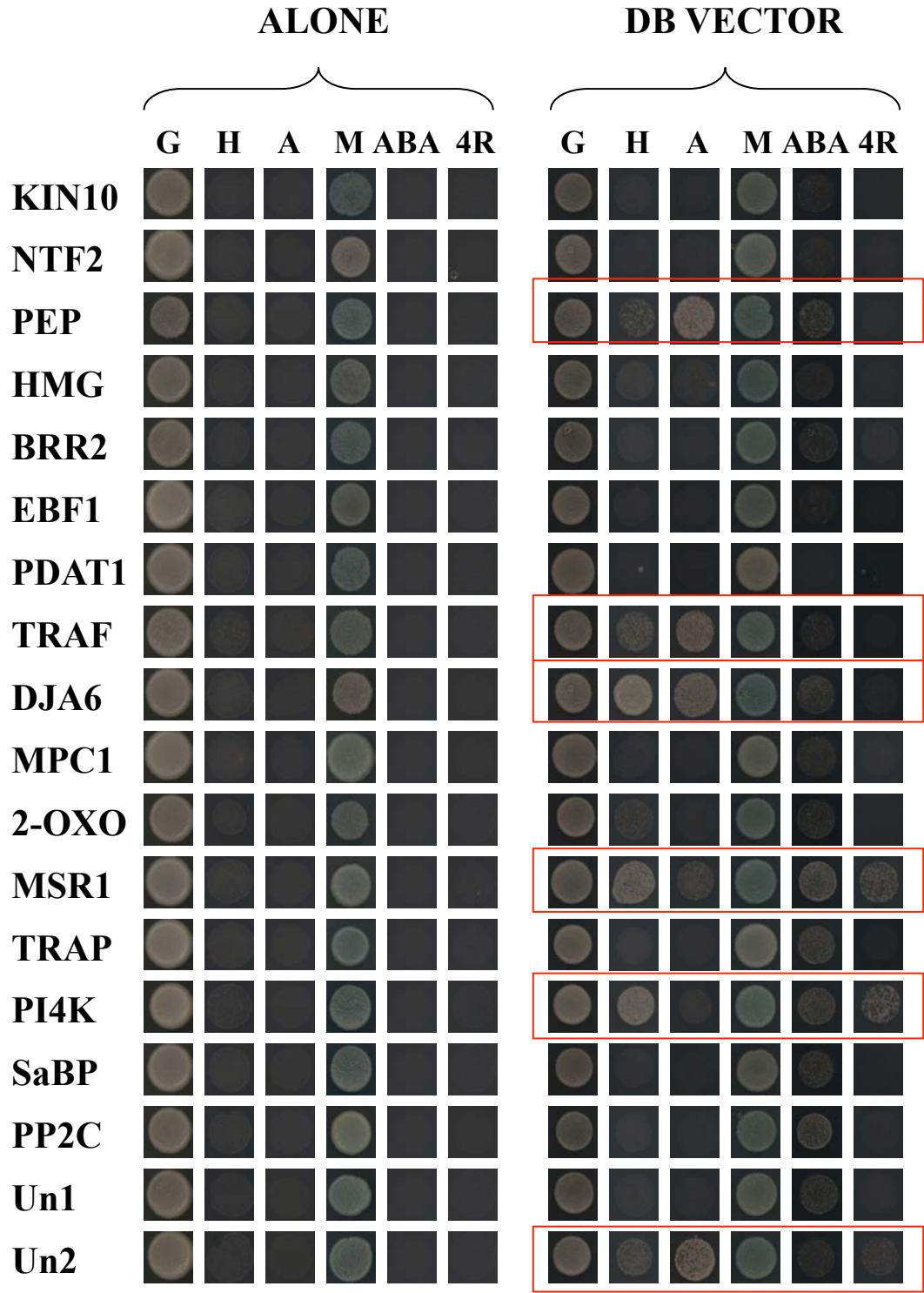


Figure 35 Interaction of ADT5 interactors with ADT1, ADT2 and ADT3 in Yeast AH109.

Each construct was co-transformed with DB-ADT1, DB-ADT2 and DB-ADT3 vectors in yeast AH109 and was assayed for growth, histidine prototrophy, adenine prototrophy, α -galactosidase activity and *AURA1* activity as well as the activation of all four reporters simultaneously. Growth of white colonies in the absence of histidine, growth of white colonies in the absence of adenine, production of a blue pigment in the presence of X- α -gal and growth of white colonies in the presence of ABA indicate an interaction. Growth of blue colonies is expected when all 4 reporters are activated simultaneously.

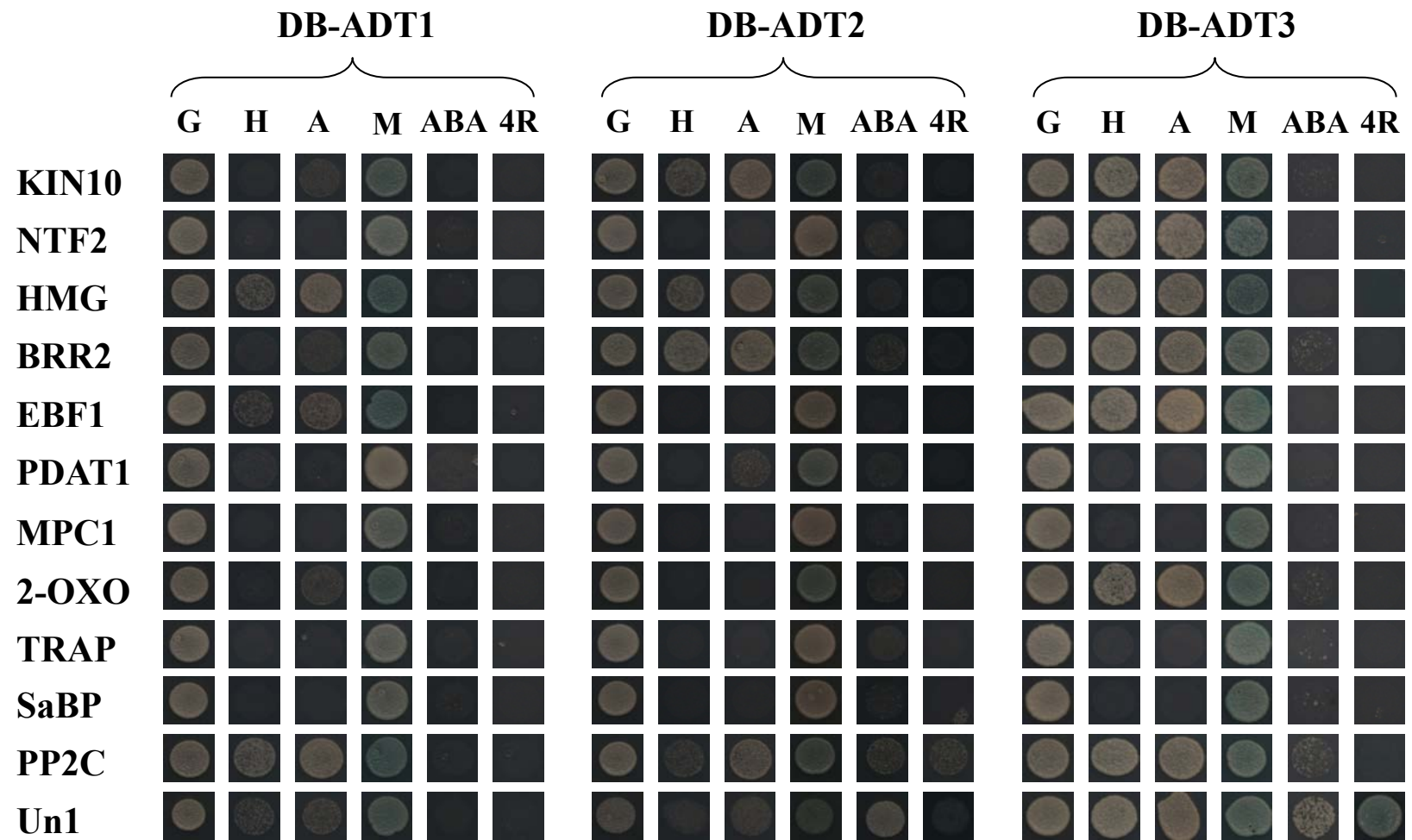


Figure 36 Interaction of ADT5 interactors with ADT4, ADT5 and ADT6 in Yeast AH109.

Each construct was co-transformed with DB-ADT4, DB-ADT5 and DB-ADT6 vectors in yeast AH109 and was assayed for growth, histidine prototrophy, adenine prototrophy, α -galactosidase activity and *AURAI* activity as well as activation of the all four reporters simultaneously. Growth of white colonies in the absence of histidine, growth of white colonies in the absence of adenine, production of a blue pigment in the presence of X- α -gal and growth of white colonies in the presence of ABA indicate an interaction. Growth of blue colonies is expected when all 4 reporters are activated simultaneously.

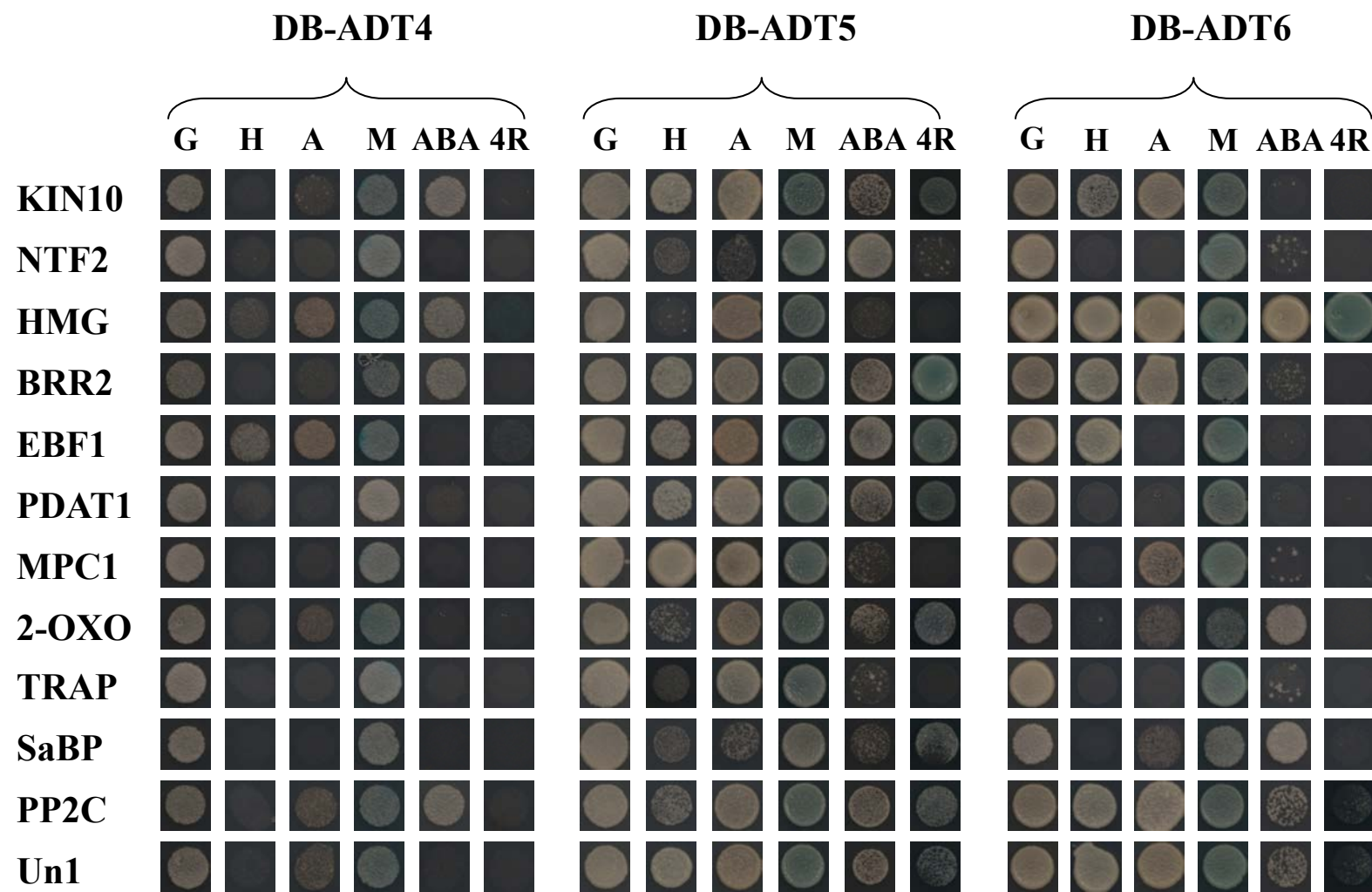


Figure 37 Interactions of PDAT1 and ADTs in Yeast AH109.

PDAT1 was co-transformed with ADTs in yeast AH109 and was assayed for growth, histidine prototrophy, adenine prototrophy, α -galactosidase activity and *AURAI* activity as well as the activation of the all four reporters simultaneously. Growth of white colonies in the absence of histidine, growth of white colonies in the absence of adenine, production of a blue pigment in the presence of X- α -gal and growth of white colonies in the presence of ABA indicate an interaction. Growth of blue colonies is expected when all 4 reporters are activated simultaneously.

PDAT1 was only interacted with ADT5 and as a result all the four reporter genes were activated.

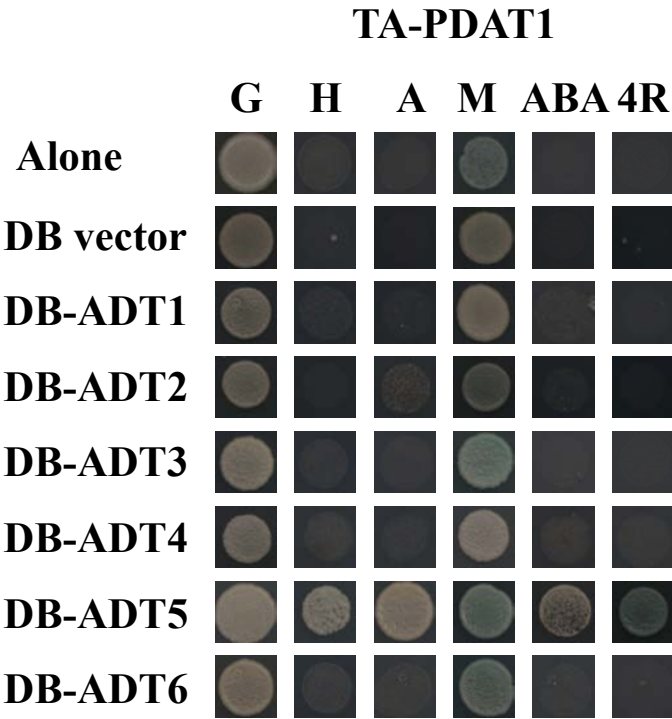
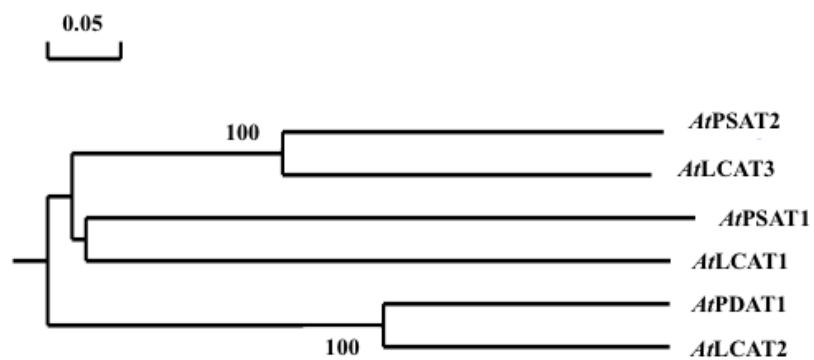


Figure 38 Phylogenetic analysis of *A. thaliana* PDAT/LCATs.

A. Protein sequences were compared using DNAMAN and a rooted tree was generated with a bootstrap of 1000. Numbers at the branch points give the bootstrapping values and the horizontal scale indicates sequence divergence. *Arabidopsis thaliana* sequences: *At*PSAT1: phospholipid sterol acyltransferase 1 (*At*1g04010), *At*PSAT2: phospholipid sterol acyltransferase2 (*At*4g19860), *At*LCAT3: lecithin cholesterol acyltransferase3 (*At*3g03310), *At*LCAT1: lecithin cholesterol acyltransferase1 (*At*1g27480), *At*PDAT1: phospholipid diacylglycerol acyltransferase1 (*At*5g13640), *At*LCAT2: lecithin cholesterol acyltransferase 2 (*At*3g44830).

B. Homology matrix of amino acid sequences of six *A. thaliana* PDAT1 homologs. Sequence similarity values range from 11.1% (*At*PSAT1 versus *At*LCAT2) to 59.7% (*At*PDAT1 versus *At*LCAT2).

A**B**

	<i>AtPSAT2</i>	<i>AtLCAT3</i>	<i>AtPDAT1</i>	<i>AtLCAT2</i>	<i>AtLCAT1</i>	<i>AtPSAT1</i>
<i>AtPSAT2</i>	100					
<i>AtLCAT3</i>	48.6	100				
<i>AtPDAT1</i>	14.8	15.4	100			
<i>AtLCAT2</i>	13.7	15.7	59.7	100		
<i>AtLCAT1</i>	19	20.3	13.1	13.1	100	
<i>AtPSAT1</i>	14.6	14.7	11.8	11.1	16.5	100

from 11.1% (*At*PSAT1 versus *At*LCAT2) to 59.7% (*At*PDAT1 versus *At*LCAT2; Figure 38B). *At*LCAT2 was identified as the most similar sequence in *A. thaliana* to *At*PDAT1.

*At*PDAT1 and *At*LCAT2 have a N-terminal cytoplasmic tail, a transmembrane-spanning region that anchors the proteins to the membrane (Sonnhammer et al., 1998) and the C terminus of the protein as well as an aromatic amino acid rich ER retrieval motif are localized in the lumen of the ER (Figure 39, Stahl et al., 2004; McCartney et al., 2004). There are conserved residues in *Hs*LCAT and *A. thaliana* sequences that have been shown to be required for phospholipase A activity (Stahl et al., 2004).

When the full-length sequences of *At*PDAT1 and *At*LCAT2 were analyzed with the cNLS Mapper (Kosugi et al., 2009b) both *At*PDAT1 and *At*LCAT2 were found to have predicted NLSs. *At*PDAT1 has a predicted bipartite NLS at positions 5 and 131 with a score of 6.4 and 7, respectively and *At*LCAT2 has two predicted monopartite NLS at positions 29 and 116 with the scores of 6.8 and 8, respectively (Figure 39). In addition, the full-length sequences of *At*PDAT1 and *At*LCAT2 were submitted to SomeNA. This software predicts the DNA binding residues based on the amino acid sequence of the proteins (Hönigschmid, 2012). SomeNA identified a His at position 5 located at the N-terminal cytosolic tail of *At*PDAT1 as a DNA binding site (Figure 39).

Next, a genome-wide search was performed using the PDAT1 amino acid sequence as query to perform a protein-protein BLAST against model organism sequences listed in the NCBI (National Center for Biotechnology Information database). Candidate proteins with higher sequence similarities were found in *Capsella rubella*, *Glycine max* (Soybean), yeast (*Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*) humans (*Homo sapiens*), mice (*Mus musculus*) and bacteria (*Microcystis aeruginosa* and *Mycobacterium tuberculosis*). Multiple hits were identified in *Arabidopsis* and soybean and only one hit was identified for other organisms. In total 20 sequences were identified and 35% were annotated as PDAT proteins, 20% LCAT, 15% PSAT and 20% were other proteins with functions in lipid biosynthesis like phospholipase A. A phylogenetic analysis of the full-length sequences of 20 candidates and *At*PDAT1 was carried out (Figure 40). The PDAT1 and LCAT2 sequences in *Arabidopsis* and PDATs from soybean and yeast grouped together (group I) whereas PSATs (except *At*PSAT1), LSAT and GLYMA sequences grouped together (group II).

Figure 39 Alignment of PDAT1 and LCAT2 protein sequences from *A. thaliana*.

Identical amino acids in both sequences are shown with lines. Numbers on the right side of the sequences count the amino acids.

Black asterisk: The active phospholipase A site similar to *HsLSAT*, Blue box: ER retrieval motif, Green triangle: DNA binding position, Orange box: Membrane spanning regions, Red box: NLS.

*At*PSAT1 and *Gm*PSATXI grouped together and microbial sequences had a separate branch. All the group I sequences were submitted to the cNLS Mapper (Kosugi et al., 2009b) to identify potential NLS sequences and nuclear localizations. It was found that all the sequences in group I except *Gm*PDAT3, have predicted NLS sequences.

3.8 *In planta* identification of ADT5 and PDAT1 interaction

For further assessment of the interaction of ADT5 with PDAT1, BiFC assays were performed to determine if the interaction could be detected *in planta*. The full-length coding sequence of *PDAT1* was recombined into BiFC destination vectors (section 2.3). Prior to co-infiltration of ADT5 and PDAT1, each construct was infiltrated individually to ensure no fluorescence was detected in the presence of a single vector only (Figure 41A and Figure 41B). By co-infiltration of *YN-PDAT1* and *YC-PDAT1* it was demonstrated that PDAT1 was able to form homodimers and the YFP signal was detected in the ER as indicated by outlining the cell and the nucleus, while excluded from the nucleus (Figure 41C). Reciprocal co-infiltration *YN-* and *YC-* *ADT5* and *PDAT1* fusions also showed that interactions were detected in the nucleus (Figure 41D and Figure 41E).

Furthermore, to ensure the specificity of PDAT1 interaction with ADT5, PDAT fusion constructs were co-infiltrated with all other ADT constructs. The YFP signal was not detected for any of the other ADTs (Figure 42). Therefore, it was concluded that the PDAT1 interaction was specific to ADT5 and the detection of the YFP signal in the nucleus was due to the presence of ADT5 since PDAT1 homodimers were only detected in the ER.

Figure 40 Phylogenetic analysis of *A. thaliana* PDAT/LCAT sequences together with model organisms (landmark) PDAT/LCATs.

Protein sequences were aligned using DNAMAN and a rooted tree was generated with a bootstrap of 1000. Numbers at the branch points give the bootstrapping values and the horizontal scale indicates sequence divergence. *Arabidopsis thaliana* sequences: *AtPSAT1*: phospholipid sterol acyltransferase1 (*At1g04010*), *AtPSAT2*: phospholipid sterol acyltransferase2 (*At4g19860*), *AtLCAT3*: lecithin cholesterol acyltransferase 3 (*At3g03310*), *AtLCAT1*: lecithin cholesterol acyltransferase1 (*At1g27480*), *AtPDAT1*: phospholipid diacylglycerol acyltransferase (*At5g13640*), *AtLCAT2*: lecithin cholesterol acyltransferase2 (*At3g44830*). *Capsella rubella* sequence: *CrCARUB* (XP_006286627). *Glycine max* sequences: *GmPSAT XI*: phospholipid sterol O-acyltransferase-like isoform XI (XP_003553502), *GmLCAT4*: lecithin cholesterol acyltransferase-like4 (XP_003529428), *GmLCAT3*: phospholipase A (XP_003552461), *GmGLYMA*: hypothetical protein (04G047000), *GmPDAT1*: phospholipid diacylglycerol acyltransferase1 (XP_003541296), *GmPDAT2*: phospholipid diacylglycerol acyltransferase2 (XP_006592298), *GmPDAT3*: phospholipid diacylglycerol acyltransferase3 (XP_014619101), *GmPDAT4*: phospholipid diacylglycerol acyltransferase4 (XP_003528441). *Microcystis aeruginosa* sequence: *MaLpqU*: lipoprotein Lpqu (NP_215538). *Mycobacterium tuberculosis* sequence: *MtRib HII*: ribonuclease HII (WP_002798654). *Mus musculus* sequence: *MmPLA2*: Group XV phospholipase A2 precursor (NP_598553). *Homo sapiens* sequence: phosphatidylcholine-sterol acyltransferase precursor (NP_000220). *Saccharomyces cerevisiae* sequence: phospholipid diacylglycerol acyltransferase (NP_014405). *Schizosaccharomyces pombe* sequence: phospholipid diacylglycerol acyltransferase p1h1 (NP_596330).

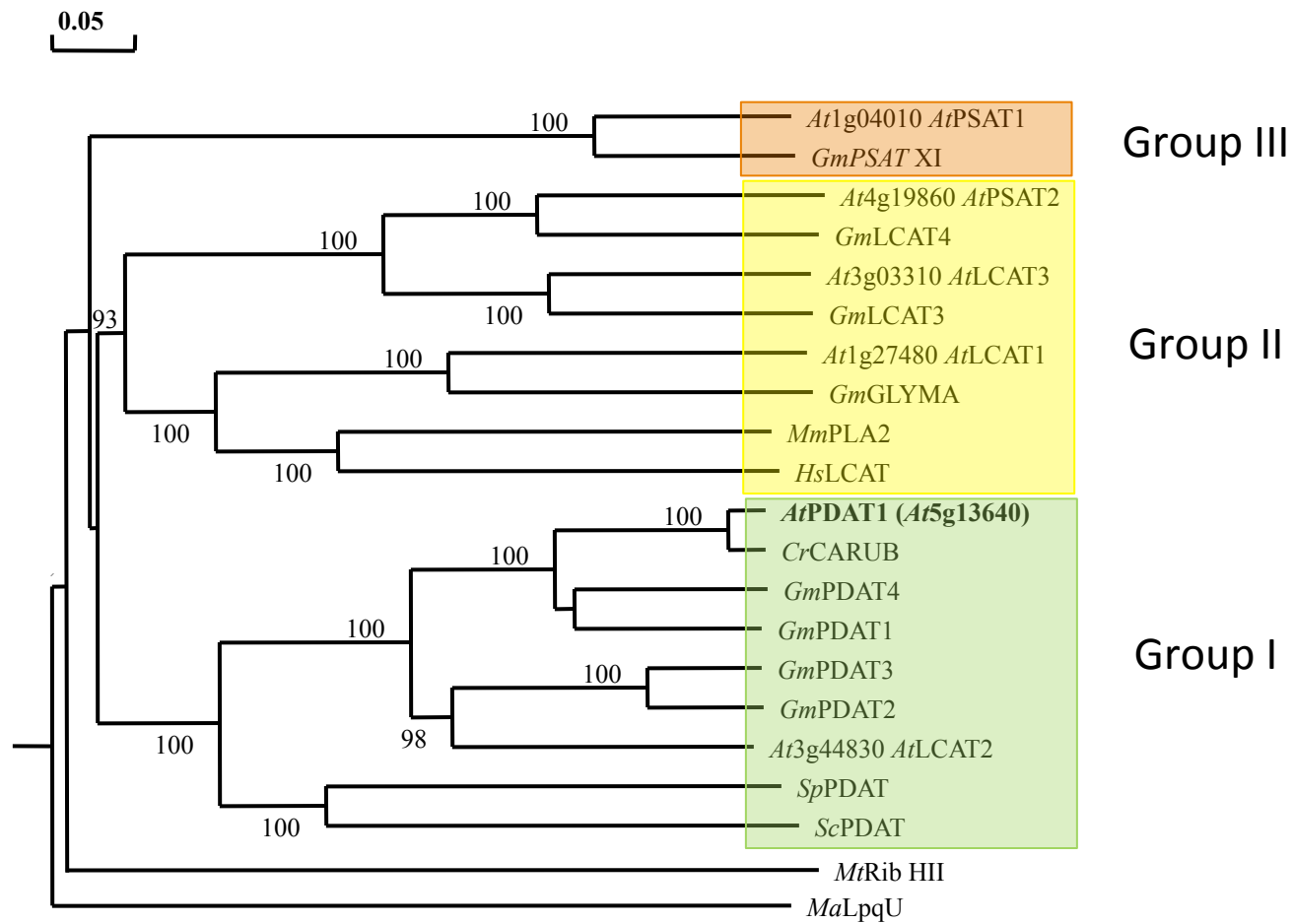


Figure 41 Control transformations and interactions involving ADT5 and PDAT1 assayed by BiFC in *N. benthamiana* leaves.

YN- and *YC-PDAT1* constructs were transiently expressed in *N. benthamiana* leaves alone, together and in combination with corresponding ADT5 constructs and analyzed at 3 dpi by confocal microscopy. Proximity of two fusion proteins results in functioning fluorophore. Accumulation of yellow fluorescence indicates dimer formation. Infiltration combinations are identified on the left.

Chlorophyll fluorescence and YFP fluorescence are shown separately in the left and middle column, respectively, and the merged image is shown in the right column. Summary of the observations is shown on the right side of the panel.

E: ER, N: Nucleus.

Scale bars are 5 μ m.

(A) and **(B)** No YFP signal is detected in leaf samples infiltrated with only *YC-PDAT1* or *YN-PDAT1*.

(C) YFP signal is detected in leaf samples co-infiltrated with *YN-PDAT1* and *YC-PDAT1* in the ER.

(D) and **(F)** Interaction of YN-PDAT1 and YC-ADT5 as well as YC-PDAT1 and YN-ADT5 is observed in the nucleus.

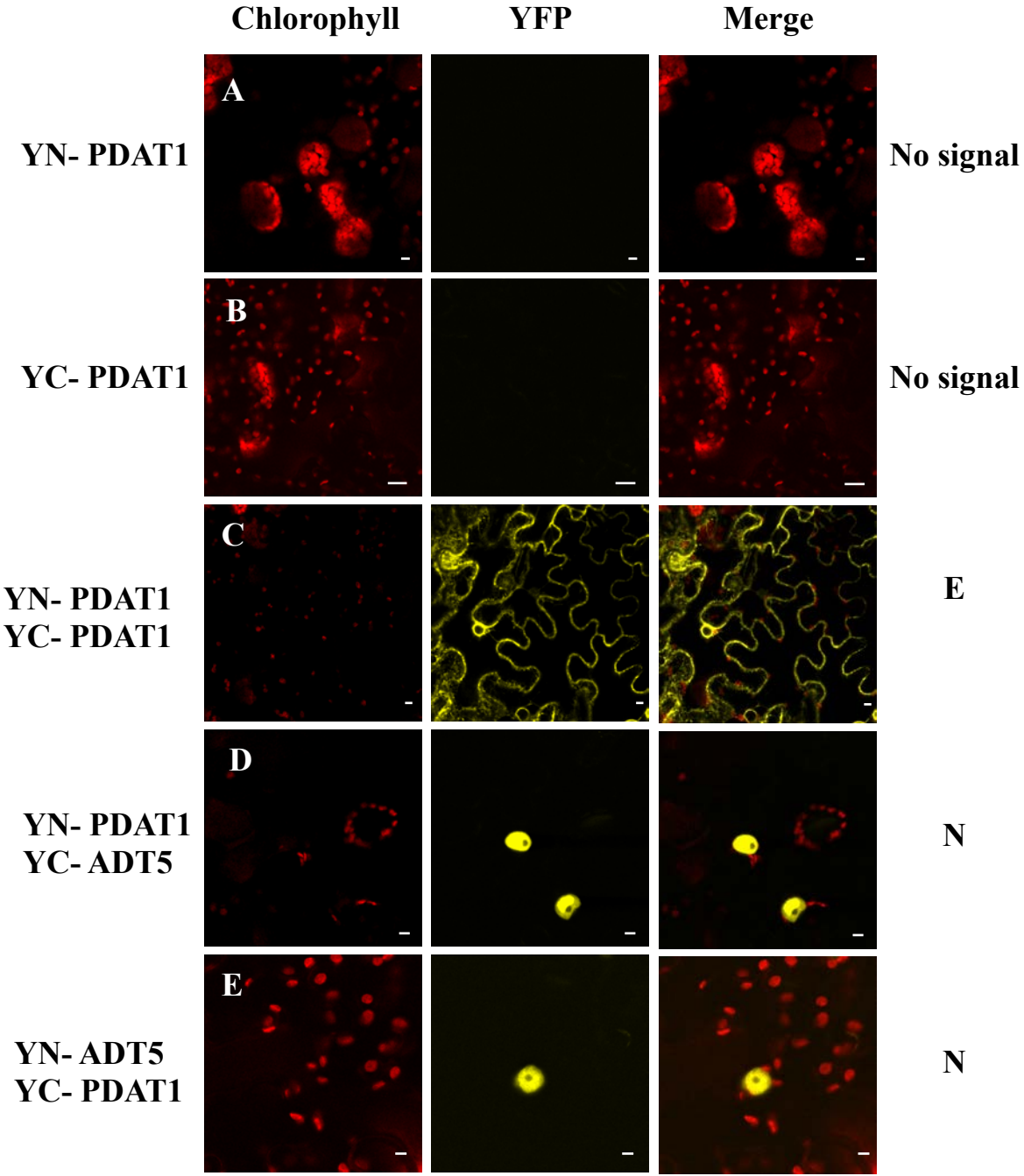


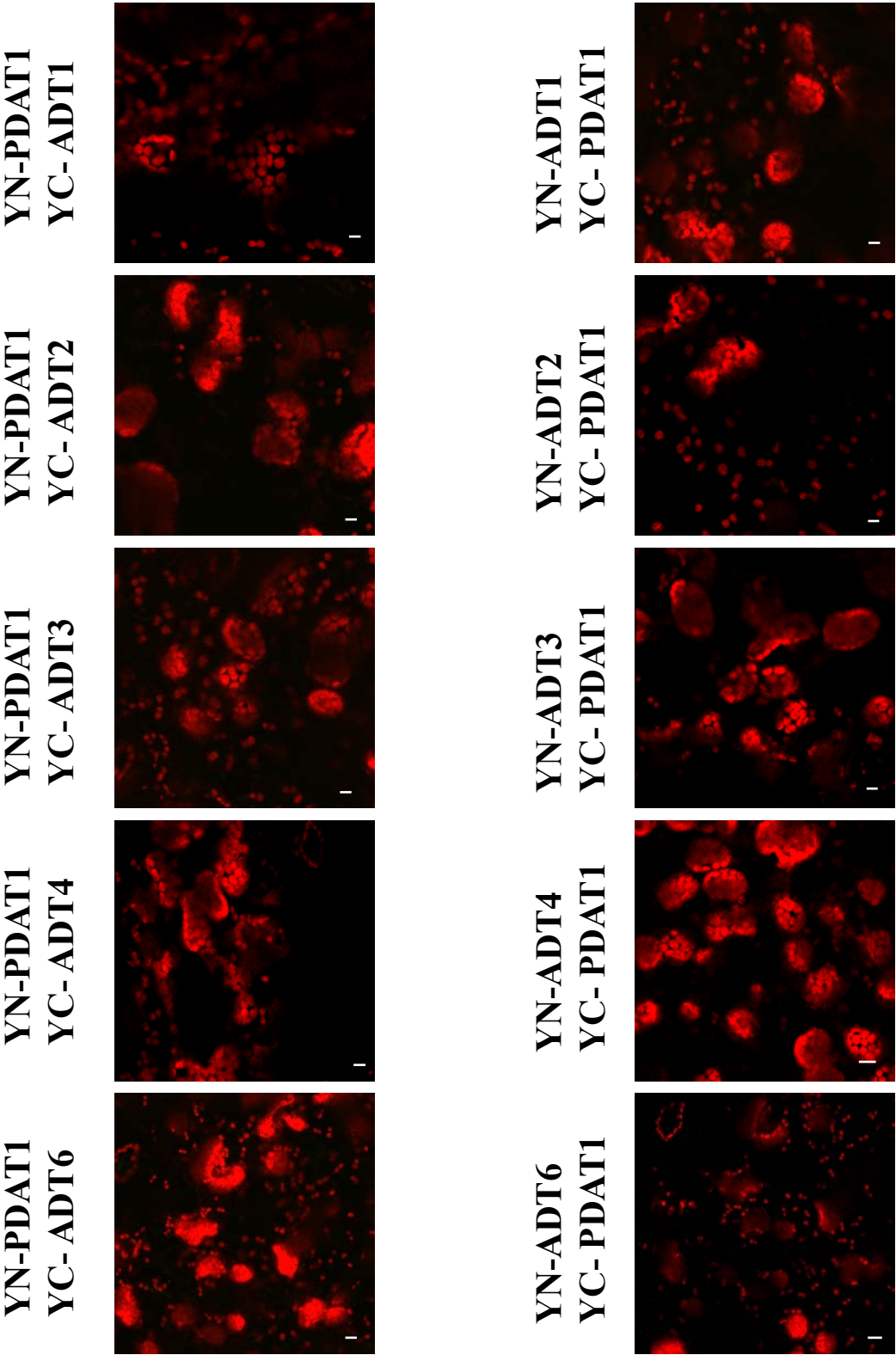
Figure 42 Interactions involving PDAT1 and ADTs assayed by BiFC in *N. benthamiana* leaves.

Interactions between PDAT1 and ADTs were assayed by co-expression of *YN-PDAT1* or *YAC-PDAT1* with corresponding *YN-* or *YC-*ADTs in *N. benthamiana* leaves and BiFC assay. Proximity of two fusion proteins results in functioning fluorophore. Accumulation of yellow fluorescence indicates dimer formation.

Chlorophyll fluorescence and YFP fluorescence are shown separately in the left and middle column, respectively, and the merged image is shown in the right column.

Scale bars are 5 μm .

No interaction is detected.



4 Discussion

The dual localization of ADT5 in the chloroplast and nucleus suggests a novel non-enzymatic role for this protein in *A. thaliana*. This study focused on the possible routes for nuclear localization of ADT5 including direct translocation of the protein from chloroplast to the nucleus through stromules and nuclear targeting by interacting with other proteins. Using an *A. thaliana* large-scale cDNA library screening, I was able to identify a specific intractor of ADT5 named PHOSPHOLIPID DIACYLGLYCEROL ACYLTRANSFERASE1 (PDAT1). PDAT1 is an enzyme that catalyzes the biosynthesis of triacylglycerol in the ER. *In silico* analysis predicted a bipartite NLS as well as a DNA binding site at the N-terminus of PDAT1. Here it will discuss how this protein might mediate ADT5 translocation into the nucleus. My results suggest that ADT5 is a new moonlighting protein with non-enzymatic roles in the regulation of gene expression in the nucleus.

4.1 ADT5 localizes to the stromules of chloroplast and into the nucleus

Dual localization of ADT5 in the stromules and nucleus was observed in *N. benthamiana* and *A. thaliana* (Bross, 2011; Howes, 2013). To confirm the nuclear localization of ADT5-CFP, all six ADT-CFP proteins were transiently expressed in *N. benthamiana* leaves. Fusion proteins were visualized by confocal microscopy and verified by Western blotting. Since there was no ADT-specific antibody available the ADT-CFP fusion proteins were detected by an anti-GFP antibody. This antibody has been shown to recognize GFP as well as GFP variants including CFP (Clontech 632380). The Western blot analyzing the ADT-CFP samples (Figure 7) showed a clear band at the expected size of the fusion proteins for all the ADTs including ADT5 as well as larger sizes consistent with ADT dimers or multimers (Figure 5). In addition to ADT bands, a faint band at 25 kDa corresponding to the size of CFP was detected. However, this band was detected more abundantly in samples for ADT1 and ADT3 and these proteins were never observed in the nucleus by confocal microscopy. Hence, it was concluded that ADT5 nuclear localization is not due to the CFP cleavage product. Moreover, expressing ADT5-CFP with its native promoter showed the same nuclear localization by confocal

microscopy and no cleaved products were detected by Western blotting (Figure 9 and Figure 10). Hence, ADT5-CFP indeed dually localizes into the chloroplasts and nucleus.

4.2 Functional relevance of stromule localization of ADTs

Stromules are dynamic structures that vary in size and shape, from small dots or projections to long elaborate tubules (Köhler and Hanson, 2000; Gunning, 2005). Although the formation of stromules has been the subject of debate, the importance of stromules in translocation of molecules and some proteins synthesized within the plastids to other cell compartments is generally accepted (Hanson and Sattarzadeh, 2013; Schattat et al., 2011; Caplan et al., 2015). Expression of ADT5, ADT4 and ADT2 fusion proteins in the chloroplast stroma resulted in accumulation of the proteins in stromules, similar to our previous observation when ADTs were transiently expressed in *N. benthamiana* leaves. Hence, localization of ADTs in stromules is consistent with the proposed role of stromules, as the end product of ADT enzymatic activity Phe, is required for protein synthesis and phenylpropanoid (for example lignin and flavonoids) biosynthesis in the cytosol (Boulter, 1970; Vogt, 2010; Dastmalchi et al., 2016). Therefore, it is feasible that high concentrations of Phe in stromules as a result of ADT activity can potentially facilitate Phe export into the cytosol through the large surface area of stromules. There are other stromule-localized enzymes with enzymatic products that are used in the cytosol. For example, geranylgeranyl diphosphate synthase (GGPS) that synthesizes geranylgeranyl diphosphate (GGPP) from isopentenyl diphosphate and dimethylallyl diphosphate in *Catharanthus roseus* accumulates in the stromules to facilitate the export of the GGPP to the cytosol for production of geranylgeranylated proteins (Thabet et al., 2011).

4.3 Close positioning of chloroplasts and nucleus

When transiently expressing ADT5, it was often observed that ADT5-CFP was localized to stromules in chloroplasts surrounding the nucleus at close proximity, and appears to make contact with the nucleus (Figure 11). This close proximity of stromules to the nucleus was recently explained as a result of high abundance of actin filaments around the nucleus (Erickson et al., 2017). Stromules are closely associated with actin filaments (Kwok and Hanson, 2003; Holzinger et al., 2007) and destabilization of actin

polymers in the cell decreases the stromules appearance (Kwok and Hanson, 2003; Howes, 2013). Therefore, more actin around the nucleus can increase stromule abundance in that area of the cell (Erickson et al., 2017). Frequent observations of stromules surrounding and anchoring to the nucleus provides a possibility of signal and macromolecule transfer directly from chloroplasts into the nucleus (Caplan et al., 2008, 2015; Leister, 2012; Bobik and Burch-Smith, 2015). Furthermore, previous studies in our lab demonstrated that the ADT5-CFP nuclear localization was decreased by inhibition of stromule formation in *N. benthamiana* leaves (Howes, 2013). Therefore, it is possible that ADT5 translocate into the nucleus via stromules. To test this possibility, ADT5-GFP fusion constructs were generated (Figure 12) and transformed into the chloroplast genome of *N. tabacum*, where they are expressed and translated in the chloroplast stroma instead of the cytosol to see if the fusion proteins can directly transfer to the nucleus via stromules. Analogous constructs were generated for ADT2 and ADT4 as controls as they have never been detected in the nucleus. ADT5-IS-GFP was detected in long tubular stromules in transplastomic plants, and sometimes they appeared to elongate between chloroplasts. These type of stromule formations have been reported before and it was proposed that small proteins can transfer between the chloroplasts through the stromules (Hanson and Sattarzadeh, 2013). However no fluorescent signal was detected in the nucleus for any of the fusion proteins. Considering most reports on stromule-mediated protein transfers suggest that translocation from chloroplasts to the nucleus via stromules occurs under stress conditions, like abscisic acid or pathogen attack (Gray et al., 2012; Caplan et al., 2015) it is possible that no translocation of ADT5 to the nucleus was observed as transplastomic plants were grown under standard growth conditions.

4.4 The importance of the ACT domain

As ADT5 translocation to the nucleus through stromules was not detected, it raised the possibility that ADT5 translocation happens from the cytosol, as ADT5 is a nuclear encoded protein that is translated in the cytosol. For ADT5 import from the cytosol into the nucleus a NLS is needed. Using an *in silico* approach it was not possible to identify a NLS in the ADT5 sequence (section 3.4.1). However, it has been shown that computer programs can fail to identify NLSs. For example, a bipartite NLS with a

17 amino acid linker was missed by computer programs for members of the DOF transcription factor family in *A. thaliana* (Krebs et al., 2010). Similarly, predicting a NLS for ZIC3, a nuclear-localized protein with a zinc finger domain, was not possible by using the available programs as formation of the NLS for this protein is dependent on proper folding of the protein (Hatayama et al., 2008). Hence an experimental approach was used to identify potential NLSs in the ADT5 sequence.

Two alternative approaches were used to identify if there is a functional NLS in ADT5. First, deletion constructs were generated and truncated fusion ADT5 and ADT4 (control) proteins were transiently expressed in *N. benthamiana* leaves (section 3.4.2). As small proteins less than 40 kDa can diffuse into the nucleus (Wente and Rout, 2010; Parry, 2015), all the constructs were designed to express fusion-truncated proteins of at least 40 kDa. It was found that full-length ADT4 only localized in the stromules and full-length ADT5 was detected in both stromules and nucleus. Surprisingly, all the truncated proteins for ADT4 and ADT5 were localized into the nucleus with the exception of ADT4- Δ A and ADT5- Δ A (lacking the ACT domain) that were not detected (Figure 23 and Figure 24). As ADT4- Δ AC and ADT5- Δ AC were transit peptides tagged with CFP were small enough for diffusion (40 kDa), it was possible that these proteins diffused into the nucleus. However, localization of Δ T fusion proteins in the nucleus and no detection for Δ A proteins was interesting. Western blots showed bands consistent in size with cleaved CFP products especially for ADT4- Δ T and might cause the fluorescence observed in the nucleus (Figure 25). No band at the expected cleavage product size of Δ A fusion proteins was detected on a Western blot. This is consistent with no detected CFP signal in confocal images, suggesting that the deletion of the ACT domain in Δ A fusion proteins interferes with the expression or accumulation of these fusion proteins. In summary, these data suggested an important role of the ACT domain in expression or accumulation of ADTs, but did not identify a NLS in ADT5. Therefore, a domain swapping approach was used between ADT5 and ADT2. All chimeric ADT2-ADT5 fusion proteins were close in size to the full-length ADT proteins and too large for diffusion. Chimeric proteins were transiently expressed in *N. benthamiana* leaves and they were all detected in stromules (Figure 28). Interestingly, only 3 of the chimeric proteins (C1, C5 and C6) were detected in the nucleus. The only sequence common to

C1, C5 and C6 was the N-terminus of the ADT5 ACT domain (Figure 28). Therefore, the results of this experiment suggest that ACT domain is involved, in some capacity, in the nuclear localization of ADT5.

The ACT (aspartokinase, chorismate mutase, TyrA; Aravind and Koonin, 1999) domain is a ligand-binding domain and can be found in a wide variety of proteins that are involved in amino acid metabolism (Liberles et al., 2005). These include both metabolic enzymes and transcription factors. In fact, the ACT domain-containing transcription factors regulate the expression of other ACT domain containing proteins (Grant, 2006). For example, Lrp-like (leucine-responsive regulatory protein) transcription factor is a protein with an ACT domain and regulates expression of proteins involved in degradation or biosynthesis of amino acids (Leonard et al., 2001). The DNA binding capability of this protein depends on interactions between ACT domains and the formation of homodimer. Therefore, the ACT domain is generally referred to as “the regulatory domain in amino acid biosynthesis” in the SCOP (structural classification of proteins) database (Grant, 2006). The ACT domain-containing enzymes are allosterically regulated by ligand binding of end products in amino acid biosynthesis (Liberles et al., 2005). In ADTs, binding of Phe to the ACT domain allosterically inhibits their enzymatic activity similar to what has been described for PDTs in microorganisms (Tzin and Galili, 2010). A highly conserved motif “ESRP” located at the N-terminus of the ACT domain has been identified to be essential for Phe binding and feedback inhibition (Huang et al., 2010). Mutations in this motif in *A. thaliana* ADT2 results in sensitivity changes for binding Phe and mutant plants show a severe phenotype including leaf morphology and dwarf and sterile plants (Huang et al., 2010). However, other evidence indicates that this motif in ADT4 and ADT5 is not sufficient to regulate their enzymatic activity (Chen et al., 2016). Plants overexpressing ADT4 and ADT5 were found to be insensitive to Phe and no feedback inhibition occurred. Phe concentration in overexpressed plants was high and consequently resulted in high anthocyanin levels (Chen et al., 2016). Transgenic plants overexpressing ADT4 and ADT5 had yellow/white leaves and some of the ADT4 transgenic plants were dwarf and sterile, similar to the morphological phenotype of ADT2 mutants (Chen et al., 2016). Overall, these observations are similar to the morphological phenotypes of transplastomic ADT5-IS and ADT4-IS, which had pale

leaves, and were dwarf and sterile (Figure 18). Although the level of Phe or any secondary metabolites were not quantified in these transplastomic plants (as it was not the main focus of this thesis), the observed phenotypes are consistent with high accumulation of Phe and/or secondary metabolites. However, ADT2 transplastomic plants did not have any noticeable phenotype probably due to the feedback control of Phe. Chen et al. (2016) suggested that some specific amino acids are only present in ADT4 and ADT5 ACT domains that might cause the insensitivity to Phe.

It is not unexpected that sequence changes have occurred in ACT domains over time. As the ACT domains are found in proteins from bacteria and archaea and they have an ACT or ACT like domain, this suggests that this domain appeared very early on (Grant, 2006). There are examples that this domain evolved to bind other ligands and molecules and perhaps participating in other cellular processes. For example, NikR transcription regulator binds nickel and has a role in the regulation of intracellular nickel levels (Schreiter et al., 2003). Similarly, YKoF protein that bind thiamine and are involved in thiamine transport (Goldschmidt et al., 2007). Although binding to amino acids was the original function of the ACT domain, it seems the role of the ACT domain as a regulatory domain is expanding.

An alignment of the ACT domains of all six ADTs indicates that the AQEH motif in the N-terminus is unique to ADT4 and ADT5 (Figure 29). However, as ADT4 was never detected in the nucleus, this motif is not sufficient for nuclear targeting. The only difference between the ADT4 and ADT5 N-terminal ACT domain is an asparagine (Asn) of ADT5 ACT domain, which is substituted with an aspartic acid (Asp) in ADT4. Therefore, I propose that the combination of AQEH and Asn28 is important for targeting ADT5 into the nucleus. A protein prediction interaction program (Ofra and Rost, 2007) supports this idea as it predicted both the histidine (His) in the AQEH motif and Asn28, are required for protein interaction in the ADT5 ACT domain.

4.5 Involvement of the ACT domain in protein conformation

There is no crystal structure available for plant ADTs, and all the predictions for ADT structures are based on PDT crystal structures generated for these bacterial PDTs (Tan et al., 2008; Vivan et al., 2008). Crystal structures for the PDTs indicate that CAT

domain and ACT domain are connected through a small linker that allows for the movement of ACT domain. In addition, it is shown that PDTs form tetramers (dimers of dimers) but a single PDT dimer is believed to be the basic catalytic unit. In a PDT homodimer, two catalytic domains form the catalytic cleft, which forms the active site (Tan et al., 2008). When Phe is not bound to the ligand-binding site, located at the N-terminus of the ACT domain, PDT dimers exist in an open conformation and prephenate has access to the catalytic cleft (Grant, 2006). However, when two Phe molecules simultaneously bind to the ACT domain of dimers, the ACT domains shift into a closed position, which decreases the accessibility of prephenate to the active site. This movement affects the entire protein and changes the overall conformation of the protein and the dimers. As the protein interaction sites important for nuclear targeting has been identified in the N-terminus of the ADT5 ACT domain, it is very possible that the overall conformation of the protein is important in the accessibility of the interaction site to the interaction partners.

Six ADTs of *A. thaliana* have been shown to form homodimers and every combination of heterodimers (Styranko, 2011). This is consistent with data presented in this study where Western blots show bands larger than the expected size of ADT monomers but consistent in size with the formation of ADT dimers and multimers (Figure 7). However, it is still unknown if dimerization is required for the enzymatic activity of ADTs.

4.6 Protein-protein interaction and nuclear localization

The BAR database predicts 5 proteins that are able to interact with ADT5 (Table 1). Among these, IMPA6 was the best candidate for being involved in nuclear targeting of ADT5. IMPA6 is a member of the *A. thaliana* importin family that transfer proteins into the nucleus through the nuclear import system and nuclear pores (Smith et al., 1997; Goldfarb et al., 2004). To confirm the interaction of ADT5 and IMPA6, a Y2H assay was used. Control transformations showed that DB-IMPA6 activated the reporters in the absence of an interacting protein (Figure 30B). Such self-activation has been reported for proteins that contain motifs that are capable of recruiting the transcription complex in Y2H assays (Piskacek et al., 2007). Therefore, only TA-IMPA6 was used to test for

possible interaction with ADT5, however no interaction was detected (Figure 30B). In addition to Y2H, the ADT5 and IMPA6 interaction was also analyzed using the BiFC assay *in planta*. The BiFC results indicate that IMPA6 and ADT5 are able to interact and form heterodimers and that interaction occurs in the nucleus (Figure 32C and Figure 32D). However, this interaction was not specific to ADT5 as IMPA6 was able to interact with all ADTs (Figure 33). This type of assay cannot distinguish if ADTs other than ADT5 are directly interacting with IMPA6 or if an ADT5 homolog in *N. benthamiana* mediates the interaction and hence all translocate into the nucleus. This scenario is feasible as it has been shown that all heterodimers with ADT5 are detected in the nucleus (Styranko, 2011).

Next, to identify interactors specific to ADT5, an *Arabidopsis* Y2H cDNA library screen was performed. Initially 21 putative ADT5 interactors were identified and based on their subcellular localizations 7 proteins were classified as nuclear proteins (Table 2). Control transformations showed that some of the interactors were able to activate the reporter genes in the absence of ADT5 and they were not used for further analysis (Figure 34). Among the remaining interactors, proteins with a role in lipid biosynthesis were observed several times (KIN10, SaBP and PDAT1) and interestingly PDAT1 was identified to specifically interact with ADT5 (Figure 35, Figure 36 and Figure 37).

Additionally, a BiFC assay was performed to confirm the interaction of ADT5 and PDAT1 *in planta*. BiFC results supported the Y2H data and it was found that PDAT1 only interacts with ADT5 and the ADT5 and PDAT1 heterodimers were detected in the nucleus (Figure 41D and Figure 41E). Although interaction of ADT5 with an ER transmembrane protein was not expected, interaction of ADTs (ADT6) with an ER-anchored isoflavone synthase (IFS2) in soybean has been reported recently (Dastmalchi et al., 2016).

4.7 PDAT1 and its similarity to lipid synthesis enzymes

Triacylglycerol (TAG) is a well-known lipid-based energy source. For a long time it was believed that DIACYLGLYCEROL ACYLTRANSFERASE (DAG) is the only enzyme that catalyzes the last step of TAG biosynthesis, a reaction that transfers the acyl from acyl-CoA to diacylglycerol in the ER (Ohlrogge and Browse, 1995). In 2000,

an enzyme called PHOSPHOLIPID DIACYLGLYCEROL ACYLTRANSFERASE (PDAT) was characterized as an enzyme that is also involved in the synthesis of TAG in yeast and plants but this enzyme acts independently of acyl-CoA (Dahlqvist et al., 2000). In *Saccharomyces cerevisiae*, *LRO1* has been identified to encode a PDAT and was shown to have sequence similarity to animal lecithin:cholesterol acyltransferases (LCAT; Dahlqvist et al., 2000). Six gene sequences with sequence similarity to *ScPDAT* and *HsLCAT* were identified in the *Arabidopsis* genome database (*AtLCAT1*: At1g27480, *AtLCAT2*: At3g44830, *AtLCAT3*: At3g03310, *AtPSAT1*: At1g04010 and *AtPSAT2*: At4g19860) (Stahl et al., 2004) with *AtPDAT1* being most similar to *ScPDAT* (28% amino acid identity). To identify the extent of sequence similarity between the *A. thaliana* sequences, an amino acid alignment was performed using full-length sequences (Figure 38A). Comparing the sequences, it was found that the *A. thaliana* sequences are diverse and the *AtPDAT1* and *AtLCAT2* shared the highest sequence similarity (59.7%; Figure 38B). Due to their sequence similarity, *AtLCAT2* has been named *AtPDAT2* in some reports (Pan et al., 2015).

AtPDAT1 and *AtLCAT2* have a short N-terminal cytoplasmic tail, a transmembrane-spanning region which anchors the proteins into the ER membrane and a large C-terminus ending in an aromatic amino acid-rich ER retrieval motif which is located in the lumen of the ER (Figure 39). Both *AtPDAT1* and *AtLCAT2* are predicted to have NLS using the cNLS mapper software (Kosugi et al., 2009a) and the SomeNA protein-DNA binding prediction software (Hönigschmid, 2012) identified His5 as a DNA binding residue in the *AtPDAT1* sequence (Figure 39). Both, the DNA binding position and the NLS in the *AtPDAT1* sequence are located in the N-terminus of the protein suggesting that this portion of PDAT1 may contain sequences allowing this protein to be targeted to the nucleus.

An amino acid sequence alignment was performed with sequences most closely resembling PDAT1 in other model organisms in NCBI (National Center for Biotechnology Information). The sequences fall into three groups and *AtPDAT1* and *AtLCAT2* were found to be the most similar to the PDAT sequences from yeast and soybean in group I (Figure 40). No nuclear localization has been reported for any of these proteins from group I, however the cNLS Mapper program (Kosugi et al., 2009b) found

that all the sequences in group I except *GmPDAT3*, have predicted NLS sequences located in the N-terminal sequences of these proteins. This N-terminal sequence is not present in *HsLCAT* (Stahl et al., 2004), therefore it is apparently specific to PDATs and raises the possibility that these proteins are able to localize to the nucleus.

4.8 ER membrane bound transcription factors

Although, PDAT1 is an ER localized protein, it is possible that it can localize into the nucleus as it has a NLS sequence (Figure 39). In the nucleus, it might act as transcription factor as it has a predicted DNA binding site.

TFs are translated in the cytoplasm and often they stay inactive until stimulation, upon which they translocate into the nucleus using the NLS nuclear transport system (Seo et al., 2008). Hence, activation of dormant transcription factors is an important molecular feature and several processes in the cell including posttranslational modifications and interaction with other proteins have been seen to regulate this activity (Espenshade and Hughes, 2007). One mechanism for the activation of TFs is the proteolytic cleavage of membrane-bound transcription factors (MTFs) which has been identified as an adaptive strategy to respond to environmental changes (Vik and Rine, 2000). Different TFs have been reported to be activated in this way, such as the yeast transcription factors SPT23 and MGA2 (Chellappa et al., 2001). These proteins are associated with the ER membrane and when unsaturated fatty acids are needed, the transcription factor domain of these proteins is released and activate the transcription of *OLE1*, which encodes an enzyme involved in unsaturated fatty acid biosynthesis (Hoppe et al., 2005). Another example is the STEROL REGULATORY ELEMENT BINDING DOMAIN PROTEIN (SREBP), a mammalian transcription factor controlling lipid biosynthesis and an integral ER membrane protein (Magaña et al., 2000). Upon the cleavage of the N-terminal cytosolic portion of this protein that contains the transcription factor domain, the cleavage product localizes into the nucleus where it regulates the expression of genes controlling lipid biosynthesis (Brown et al., 2000). Furthermore, recent studies identified several plant TFs that are membrane proteins. Often these are ER membrane proteins that are linked to ER stress responses. For example, three members of the BASIC DOMAIN LEUCINE ZIPPER PROTEIN (bZIP) transcription factor family are ER membrane proteins (Iwata

and Koizumi, 2005; Liu et al., 2007). A truncated form of these proteins without the transmembrane domain has been shown to translocate into the nucleus. These examples are reminiscent of PDAT1 and the position of the NLS and DNA binding residues at the cytoplasmic N-terminus of PDAT is consistent with structures found in other ER membrane-bound TFs. It remains to be determined if this cytosolic domain is cleaved in PDAT1.

4.9 ADT5: a moonlighting protein

It is fascinating that in protein families as described here, members share a high degree of sequence similarity with a well-established catalytic role and sometimes one or a few members are moonlighting proteins (Mani et al., 2015). This phenomenon can be found in many organisms including plants. For example, MYO-INOSITOL PHOSPHATE SYNTHASE (MIPS) protein family in *A. thaliana* has 3 highly similar enzymes that catalyze the myo-inositol biosynthesis in the cytoplasm (Donahue et al., 2010). However, MIPS1 is found to interact with histone methyltransferases ATXR5 and ATXR6 in the nucleus to control its own expression via chromatin changes in response to pathogens (Latrasse et al., 2013). Gaining a new function sometimes is possible through only a few changes in the amino acid sequence of proteins. For example, delta1 and delta2 crystalins in birds are 94% identical and differ only in 27 amino acid residues. However, these differences cause a change in conformational structure of the protein which results in delta2 acquiring an additional function (Piatigorsky and Horwitz, 1996; Sampaleanu et al., 2001). Therefore, if only a few changes are needed to gain a function, those changes could occur over a relatively short evolutionary time and give rise to multifunctional proteins. Following the initial functional change, further sequence alternations can then affect either one of the functions or both functions of the protein (Scott and Pillus, 2010; Sluchanko and Gusev, 2017). Sometimes even a single amino acid substitution is sufficient for a protein to gain multifunctionality. For example, the mitochondrial dihydrolipoamide dehydrogenase (DLD) is catalytically active as a homodimer and converts dihydrolipoic acid to lipoic acid (Carothers et al., 1989). It is found that a single amino acid substitution destabilizes the homodimer interface and

enables this enzyme to function as a protease via a hidden protease constitutive site (Babady et al., 2007).

ADTs in *A. thaliana* share a high degree of sequence similarity and ADT4 has been identified to share the highest sequence similarity with ADT5 in CAT and ACT domains (84%; Bross et al., 2011). It is possible that only a few changes can define the nuclear localization of ADT5 and/or a new associated function. The AQEH motif in combination with the Asn28 could be such sequences that allow ADT5 to form new interactions, leading to new subcellular localization and/or new function in the nucleus without disrupting the enzymatic function of the protein.

There are a growing number of dually targeted proteins to plastids and nucleus either with defined roles in both organelles or with a known function only in one of the organelles (Krause et al., 2012). One such protein is phosphate-iso-penthytransferase 3 an enzyme required for cytokinin biosynthesis in plastids. This enzyme is also found in the nucleus but the nuclear role of this protein is still not understood (Galichet et al., 2008). Currently, the role of ADT5 in the nucleus is unknown. However, previous studies indicate that ADTs differentially contribute to the synthesis of downstream products of phenylpropanoid pathway (Corea et al., 2012a; Chen et al., 2016) and ADT5 has been identified to be important in biosynthesis of lignin (the Phe-derived cell structural components; Corea et al., 2012b). Therefore, it is possible that ADT5 regulates the expression of *ADT* genes or genes acting within the phenylpropanoid pathway to ensure that the gene products are available in an orchestrated fashion.

4.10 Concluding remarks and future directions

ADTs are important for Phe biosynthesis; however, ADT5 has been shown in this study to have a nuclear localization in addition to its stromule localization. Expression of truncated ADTs as well as chimeric ADT fusion proteins suggested that the ACT domain is important for stability of ADT4 and ADT5. In addition, presence of the AQEH motif in combination with Asn28 in the N-terminal region of the ACT domain was identified to be important for protein interaction. To ensure that these positions are involved in nuclear targeting of ADT5, Asp in ADT4 should be substituted to Asn to determine if this targets ADT4 into the nucleus. The reciprocal change in ADT5

(replacing Asn with Asp) should be resulted in no nuclear targeting. However, there might be still more residues involved in overall conformation of ADT5 that affect the interaction ability and the moonlighting function of ADT5. Unfortunately, the lack of structural information on ADTs, has greatly influenced our understanding and interpretation of our observations. So far, all the predictions are based on structural information from PDTs (that do not have a TP) consequently they may not accurately reflect the conformation of the full length ADT proteins. Therefore, future studies need to concentrate on determining the crystal structure of ADTs, especially ADT5.

In addition, it has been shown in this study as well as in previous studies of the Kohalmi lab that ADTs are able to form dimers (Figure 7 and Styranko, 2011). There are many proteins whose functional activity relies on their ability to form dimers or multimers. For example PDTs are enzymatically active as dimers (Vivan et al., 2008). Therefore, it can be anticipated that the ADT5 enzymatic activity or its moonlighting function is dependent on dimerization. It is still not experimentally determined if dimerization is necessary for the active functionality of plants ADT. Therefore, crystal structure information can also greatly assist to better understand ADT5 function.

The Y2H screen identified PDAT1 as a specific interactor of ADT5. Mutation of the identified putative interaction sites in ADT5 ACT domain can be used to validate them as PDAT1 interaction sites by Y2H. Moreover, the *in silico* analysis indicates that PDAT1 has a NLS in the N-terminal cytosolic portion. Although these data suggest that PDAT1 is able to localize into the nucleus, it should still be experimentally confirmed. It remains to be investigated if the full length PDAT1 or only the cytosolic portion of the protein can localize into the nucleus. Comparing transient expression of PDAT1 fused to N-terminal and C-terminal fluorescence tag proteins (for example YFP) can help to analyze the localization of PDAT1 by confocal microscopy. Western blot analysis further can demonstrate if a full length or truncated version of the protein localizes into the nucleus.

The use of *N. benthamiana* as a host for ease and efficiency of transient expression of the *A. thaliana* proteins has been shown to achieve representative results (Howes, 2013; Bross et al., 2017). However, for more confirmation of ADT5 and PDAT1 interaction it is important to determine if the same localization and interactions happen in

the natural environment of these enzymes, *A. thaliana*, ideally under the regulation of the native promoters. To this end, stably transformed *A. thaliana* plants expressing ADT5-CFP and YFP-PDAT1 can be generated. These transgenic plants can then be used for several analyses. First the localization of PDAT1 and ADT5 in the nucleus can be investigated in stable transformed plants. This will extend our observations of these proteins localization from leaves to other tissues. In addition co-immunoprecipitation (co-IP) can be performed with the total protein extracts from these *A. thaliana* lines to confirm the interaction of ADT5 and PDAT1. The co-IP can also lead us to identify new binding partners (if any) and potentially increase our understanding of ADT5 roles. Currently, we are limited by the lack of ADT5 and PDAT1 antibodies. However, as ADT5-CFP fusion proteins include either Flag or HA epitope tags (Earley et al., 2006) and YFP-PDAT1 has the YFP tag, commercially available antibodies can be used to perform the co-IP experiments using transgenic plants. Alternatively, efforts should be taken to develop antibodies against ADT5 and PDAT1 that will allow a more direct detection.

Although the predicted residue in PDAT1 as a putative DNA binding site is interesting, it seems unusual that a single amino acid is sufficient to bind DNA. Possibly there are still more residues required that are not predicted by computer programs. Hence, it is worthwhile to determine if PDAT1 can bind DNA through chromatin immunoprecipitation (ChIP), and if it can the sequences it binds should be determined by sequencing approaches. It is also suggested that ChIP analysis is performed for ADT5. As it is possible that PDAT1 only piggybacks ADT5 into the nucleus and the ADT5 function in the nucleus is not related to its interaction with PDAT1.

Understanding proteins as a whole by combining structural information, computational and experimental analyses will open a door to unknown aspects of protein functions. Perhaps alternative functions can be anticipated for proteins with dual localizations in the cell. This study provides evidence that introduces ADT5 as a potential multifunctional protein and it adds a new twist to moonlighting plant enzymes.

5 References

- Abolhassani Rad, S., Clayton, E.J., Cornelius, E.J., Howes, T.R., and Kohalimi, S.E.** Moonlighting protein: putting the spotlight on enzymes. *Plant Signal. Behav.* Manuscript submitted.
- Addya, S., Anandatheerthavarada, H.K., Biswas, G., Bhagwat, S. V., Mullick, J., and Avadhani, N.G.** (1997). Targeting of NH₂-terminal-processed microsomal protein to mitochondria: A novel pathway for the biogenesis of hepatic mitochondrial P450MT2. *J. Cell Biol.* **139**: 589–599.
- Aravind, L. and Koonin, E. V** (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.* **287**: 1023–1040.
- Babady, N.E., Pang, Y.P., Elpeleg, O., and Isaya, G.** (2007). Cryptic proteolytic activity of dihydrolipoamide dehydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 6158–63.
- Baier, M. and Dietz, K.J.** (2005). Chloroplasts as source and target of cellular redox regulation: A discussion on chloroplast redox signals in the context of plant physiology. *J. Exp. Bot.* **56**: 1449–1462.
- Bao, Y., Wang, C., Jiang, C., Pan, J., Zhang, G., Liu, H., and Zhang, H.** (2014). The tumor necrosis factor receptor-associated factor (TRAF)-like family protein SEVEN IN ABSENTIA 2 (SINA2) promotes drought tolerance in an ABA-dependent manner in *Arabidopsis*. *New Phytol.* **202**: 174–187.
- Barajas-López, J. de D., Blanco, N.E., and Strand, Å.** (2013). Plastid-to-nucleus communication, signals controlling the running of the plant cell. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**: 425–437.
- Bentley, R.** (1990). The Shikimate Pathway: A metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* **25**: 307–384.
- Bobik, K. and Burch-Smith, T.M.** (2015). Chloroplast signaling within, between and beyond cells. *Front. Plant Sci.* **6**: 781.
- Boruc, J., Zhou, X., and Meier, I.** (2012). Dynamics of the plant nuclear envelope and nuclear pore. *Plant Physiol.* **158**: 78–86.
- Boulter, D.** (1970). Protein synthesis in plants. *Annu. Rev. Plant Biol.* **21**: 91–114.
- Bradbeer, J.W., Atkinson, Y.E., Börner, T., and Hagemann, R.** (1979). Cytoplasmic synthesis of plastid polypeptides may be controlled by plastid-synthesised RNA. *Nature* **279**: 816–817.
- Bradford, M.M.** (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248–254.
- Bross, C.D.** (2011). Characterizing substrate specificity and subcellular localization of arogenate dehydratases from *Arabidopsis thaliana*. Biology, M.Sc., Western University, London, Ontario.
- Bross, C.D., Corea, O.R.A., Kaldis, A., Menassa, R., Bernards, M.A., and Kohalimi, S.E.** (2011). Complementation of the pha2 yeast mutant suggests functional differences for arogenate dehydratases from *Arabidopsis thaliana*. *Plant Physiol. Biochem.* **49**: 882–890.

- Bross, C.D., Howes, T.R., Rad, S.A., Kljakic, O., and Kohalmi, S.E.** (2017). Subcellular localization of *Arabidopsis* aroenate dehydratases suggests novel and non-enzymatic roles. *J. Exp. Bot.* **68**: 1425–1440.
- Brown, M.S., Ye, J., Rawson, R.B., and Goldstein, J.L.** (2000). Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* **100**: 391–398.
- Caplan, J.L., Kumar, A.S., Park, E., Padmanabhan, M.S., Hoban, K., Modla, S., Czymmek, K., and Dinesh-Kumar, S.P.** (2015). Chloroplast stromules function during innate immunity. *Dev. Cell* **34**: 45–57.
- Caplan, J.L., Mamillapalli, P., Burch-Smith, T.M., Czymmek, K., and Dinesh-Kumar, S.P.** (2008). Chloroplastic protein NRIP1 mediates innate immune receptor recognition of a viral effector Jeffrey. *Cell* **132**: 449–462.
- Carothers, D.J., Pons, G., and Patel, M.S.** (1989). Dihydrolipoamide dehydrogenase: Functional similarities and divergent evolution of the pyridine nucleotide-disulfide oxidoreductases. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **268**: 409–425.
- Carrie, C., Murcha, M.W., Kuehn, K., Duncan, O., Barthet, M., Smith, P.M., Eubel, H., Meyer, E., Day, D.A., Millar, A.H., and Whelan, J.** (2008). Type II NAD(P)H dehydrogenases are targeted to mitochondria and chloroplasts or peroxisomes in *Arabidopsis thaliana*. *FEBS Lett.* **582**: 3073–3079.
- Carter, C.** (2004). The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. *Plant Cell Online* **16**: 3285–3303.
- Chellappa, R., Kandasamy, P., Oh, C.S., Jiang, Y., Vemula, M., and Martin, C.E.** (2001). The membrane proteins, Spt23p and Mga2p, play distinct roles in the activation of *Saccharomyces cerevisiae* OLE1 gene expression: Fatty acid-mediated regulation of Mga2p activity is independent of its proteolytic processing into a soluble transcription act. *J. Biol. Chem.* **276**: 43548–43556.
- Chen, Q., Man, C., Li, D., Tan, H., Xie, Y., and Huang, J.** (2016). Aroenate dehydratase isoforms differentially regulate anthocyanin biosynthesis in *Arabidopsis thaliana*. *Mol. Plant* **9**: 1609–1619.
- Chien, C.T., Bartel, P.L., Sternglanz, R., and Fields, S.** (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci. U. S. A.* **88**: 9578–9582.
- Cho, M.H. et al.** (2007). Phenylalanine biosynthesis in *Arabidopsis thaliana*: Identification and characterization of aroenate dehydratases. *J. Biol. Chem.* **282**: 30827–30835.
- Cho, Y.H., Yoo, S.D., and Sheen, J.** (2006). Regulatory functions of nuclear HEXOKINASE1 complex in glucose signaling. *Cell* **127**: 579–589.
- Clontech** (2010). Matchmaker™ Gold Yeast Two-Hybrid System User Manual. **1**: 1–41.
- Conley, A.J., Joensuu, J.J., Jevnikar, A.M., Menassa, R., and Brandle, J.E.** (2009). Optimization of elastin-like polypeptide fusions for expression and purification of recombinant proteins in plants. *Biotechnol. Bioeng.* **103**: 562–573.
- Cook, A., Bono, F., Jinek, M., and Conti, E.** (2007). Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem.* **76**: 647–671.

- Corea, O.R.A., Bedgar, D.L., Davin, L.B., and Lewis, N.G.** (2012a). The aroenate dehydratase gene family: Towards understanding differential regulation of carbon flux through phenylalanine into primary versus secondary metabolic pathways. *Phytochemistry* **82**: 22–37.
- Corea, O.R.A., Ki, C., Cardenas, C.L., Kim, S.J., Brewer, S.E., Patten, A.M., Davin, L.B., and Lewis, N.G.** (2012b). Aroenate dehydratase isoenzymes profoundly and differentially modulate carbon flux into lignins. *J. Biol. Chem.* **287**: 11446–11459.
- Crone, D., Rueda, J., Martin, K.L., Hamilton, D., and Mascarenhas, J.P.** (2001). The differential expression of a heat shock promoter in floral and reproductive tissues. *Plant Cell Environ.* **24**: 869–874.
- Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., Ronne, H., and Stymne, S.** (2000). Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc. Natl. Acad. Sci.* **97**: 6487–6492.
- Dastmalchi, M., Bernards, M.A., and 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 J.* **85**: 689–706.
- Donahue, J.L., Alford, S.R., Torabinejad, J., Kerwin, R.E., Nourbakhsh, A., Ray, W.K., Hernick, M., Huang, X., Lyons, B.M., Hein, P.P., and Gillasp, G.E.** (2010). The *Arabidopsis thaliana* *myo-inositol 1-phosphate synthase1* gene is required for Myo-inositol synthesis and suppression of cell death. *Plant Cell* **22**: 888–903.
- Dyall, S.D., Brown, M.T., and Johanson, P.J.** (2004). Ancient Invasions: From endosymbionts to organelles. *Science*. **304**: 253–257.
- Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S.** (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**: 616–629.
- Ehlting, J. et al.** (2005). Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *Plant J.* **42**: 618–640.
- El-Azaz, J., de la Torre, F., Avila, C., and Caovas, F.M.** (2016). Identification of a small protein domain present in all plant lineages that confers high prephenate dehydratase activity. *Plant J.* **87**: 215–229.
- Elhafez, D., Murcha, M.W., Clifton, R., Soole, K.L., Day, D.A., and Whelan, J.** (2006). Characterization of mitochondrial alternative NAD(P)H dehydrogenases in *Arabidopsis*: Intraorganelle location and expression. *Plant Cell Physiol.* **47**: 43–54.
- Emanuelsson, O., Nielsen, H., and von Heijne, G.** (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8**: 978–984.
- Erickson, J.L., Kantek, M., and Schattat, M.H.** (2017). Plastid-nucleus distance alters the behavior of stromules. *Front. Plant Sci.* **8**: 1135–1151.
- Espenshade, P.J. and Hughes, A.L.** (2007). Regulation of sterol synthesis in eukaryotes. *Annu. Rev. Genet.* **41**: 401–427.
- Espinosa-Cantú, A., Ascencio, D., Barona-Gómez, F., and De Luna, A.** (2015). Gene duplication and the evolution of moonlighting proteins. *Front. Genet.* **6**: 1–7.

- Eubel, H., Meyer, E.H., Taylor, N.L., Bussell, J.D., O'Toole, N., Heazlewood, J.L., Castleden, I., Small, I.D., Smith, S.M., and Millar, A.H.** (2008). Novel proteins, putative membrane transporters, and an integrated metabolic network are revealed by quantitative proteomic analysis of arabidopsis cell culture peroxisomes. *Plant Physiol.* **148**: 1809–1829.
- Francin-Allami, M., Saumonneau, A., Lavenant, L., Boudier, A., Sparkes, I., Hawes, C., and Popineau, Y.** (2011). Dynamic trafficking of wheat g-gliadin and of its structural domains in tobacco cells, studied with fluorescent protein fusions. *J. Exp. Bot.* **62**: 4507–4520.
- Frei, M.** (2013). Lignin: Characterization of a multifaceted crop component. *Sci. World J.* **2013**.
- Gagne, J.M., Smalle, J., Gingerich, D.J., Walker, J.M., Yoo, S.-D., Yanagisawa, S., and Vierstra, R.D.** (2004). *Arabidopsis* EIN3-binding F-box 1 and 2 form ubiquitin-protein ligases that repress ethylene action and promote growth by directing EIN3 degradation. *Proc. Natl. Acad. Sci.* **101**: 6803–6808.
- Galichet, A., Hoyerová, K., Kamínek, M., and Gruissem, W.** (2008). Farnesylation directs AtIPT3 subcellular localization and modulates cytokinin biosynthesis in *Arabidopsis*. *Plant Physiol.* **146**: 1155–1164.
- Galvez-Valdivieso, G. and Mullineaux, P.M.** (2010). The role of reactive oxygen species in signalling from chloroplasts to the nucleus. *Physiol. Plant.* **138**: 430–439.
- Geisler-Lee, J., O'Toole, N., Ammar, R., Provart, N.J., Millar, A.H., and Geisler, M.** (2007). A predicted interactome for *Arabidopsis*. *Plant Physiol.* **145**: 317–329.
- Genoud, T., Schweizer, F., Tscheuschler, A., Debrieux, D., Casal, J.J., Schäfer, E., Hiltbrunner, A., and Fankhauser, C.** (2008). FHY1 mediates nuclear import of the light-activated phytochrome A photoreceptor. *PLoS Genet.* **4**.
- Goldberg, T., Hamp, T., and Rost, B.** (2012). LocTree2 predicts localization for all domains of life. *Bioinformatics* **28**: 458–465.
- Goldfarb, D.S., Corbett, A.H., Mason, D.A., Harreman, M.T., and Adam, S.A.** (2004). Importin alpha: A multipurpose nuclear-transport receptor. *Trends Cell Biol.* **14**: 505–514.
- Goldschmidt, L., Cooper, D.R., Derewenda, Z.S., and Eisenberg, D.** (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci.* **16**: 1569–1576.
- Golis, A., Sikorski, P.J., Kruszka, K., and Kufel, J.** (2013). *Arabidopsis thaliana* LSM proteins function in mRNA splicing and degradation. *Nucleic Acids Res.* **41**: 6232–6249.
- Grant, G.A.** (2006). The ACT domain: A small molecule binding domain and its role as a common regulatory element. *J. Biol. Chem.* **281**: 33825–33829.
- Gray, J.C., Hansen, M.R., Shaw, D.J., Graham, K., Dale, R., Smallman, P., Natesan, S.K.A., and Newell, C.A.** (2012). Plastid stromules are induced by stress treatments acting through abscisic acid. *Plant J.* **69**: 387–398.
- Gray, J.C., Sullivan, J.A., Hibberd, J.M., and Hansen, M.R.** (2001). Stromules: Mobile protrusions and interconnections between plastids. *Plant Biol.* **3**: 223–233.
- Gunning, B.E.S.** (2005). Plastid stromules: Video microscopy of their outgrowth, retraction, tensioning, anchoring, branching, bridging, and tip-shedding. *Protoplasma* **225**: 33–42.

- Hanson, M.R. and Sattarzadeh, A.** (2011). Stromules: recent insights into a long neglected feature of plastid morphology and function. *Plant Physiol* **155**: 1486–1492.
- Hanson, M.R. and Sattarzadeh, A.** (2013). Trafficking of proteins through plastid stromules. *Plant Cell* **25**: 2774–2782.
- Hartley, J.L., Temple, G.F., and Brasch, M.A.** (2000). DNA cloning using *in vitro* site-specific recombination. *Genome Res.* **10**: 1788–1795.
- Hatayama, M., Tomizawa, T., Sakai-Kato, K., Bouvagnet, P., Kose, S., Imamoto, N., Yokoyama, S., Utsunomiya-Tate, N., Mikoshiba, K., Kigawa, T., and Aruga, J.** (2008). Functional and structural basis of the nuclear localization signal in the ZIC3 zinc finger domain. *Hum. Mol. Genet.* **17**: 3459–3473.
- Hellens, R. and Mullineaux, P.** (2000). Guide to *Agrobacterium* binary vector. *Trends Plant Sci.* **5**: 446–451.
- Henikoff, S. and Henikoff, J.G.** (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89**: 10915–10919.
- Herrmann, K.M. and Weaver, L.M.** (1999). The shikimate pathway. *Annu Rev Plant Physiol Plant Mol Biol* **50**: 473–503.
- Holzinger, A., Wasteneys, G.O., and Lütz, C.** (2007). Investigating cytoskeletal function in chloroplast protrusion formation in the arctic-alpine plant *Oxyria digyna*. *Plant Biol.* **9**: 400–410.
- Hönigschmid, P.** (2012). Improvement of DNA- and RNA protein binding prediction. Bioinformatic Department, Ph.D. thesis, Technische Universität München, München, Germany.
- Hoppe, T., Matuschewski, K., Rape, M., Schlenker, S., Ulrich, H.D., and Jentsch, S.** (2005). Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing. *Trends Biochem. Sci.* **30**: 577–586.
- Howes, T.R.** (2013). Analysis of subcellular localization patterns suggest non-enzymatic roles for select Arogenate dehydratases. Biology, M.Sc., Western University, London, Ontario.
- Huang, T., Tohge, T., Lytovchenko, A., Fernie, A.R., and Jander, G.** (2010). Pleiotropic physiological consequences of feedback-insensitive phenylalanine biosynthesis in *Arabidopsis thaliana*. *Plant J.* **63**: 823–835.
- Huberts, D.H.E.W. and van der Klei, I.J.** (2010). Moonlighting proteins: An intriguing mode of multitasking. *Biochim. Biophys. Acta - Mol. Cell Res.* **1803**: 520–525.
- Ikeyama, Y., Tasaka, M., and Fukaki, H.** (2010). RLF, a cytochrome b5-like heme/steroid binding domain protein, controls lateral root formation independently of ARF7/19-mediated auxin signaling in *Arabidopsis thaliana*. *Plant J.* **62**: 865–875.
- Isemer, R., Mulisch, M., Schäfer, A., Kirchner, S., Koop, H.U., and Krupinska, K.** (2012). Recombinant Whirly1 translocates from transplastomic chloroplasts to the nucleus. *FEBS Lett.* **586**: 85–88.
- Iwata, Y. and Koizumi, N.** (2005). An *Arabidopsis* transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 5280–5285.
- Janda, M., Šašek, V., and Ruelland, E.** (2014). The *Arabidopsis* pi4kIIIβ1β2 double mutant is salicylic acid-overaccumulating: a new example of salicylic acid influence on plant stature. *Plant Signal. Behav.* **9**: 805–816.

- Jeffery, C.J.** (1999). Moonlighting Proteins. *Trends Biochem. Sci.* **24**: 8–11.
- Jeffery, C.J.** (2016). Protein species and moonlighting proteins: Very small changes in a protein's covalent structure can change its biochemical function. *J. Proteomics* **134**: 19–24.
- Jung, E., Zamir, L.O., and Jensen, R.** (1986). Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate. *Proc. Natl. Acad. Sci. U. S. A.* **83**: 7231–5.
- Kaldis, A., Ahmad, A., Reid, A., Mcgarvey, B., Brandle, J., Ma, S., Jevnikar, A., Kohalmi, S.E., and Menassa, R.** (2013). High-level production of human interleukin-10 fusions in tobacco cell suspension cultures. *Plant Biotechnol. J.* **11**: 535–545.
- Kapila, J., De Rycke, R., Van Montagu, M., and Angenon, G.** (1997). An *Agrobacterium*-mediated transient gene expression system for intact leaves. *Plant Sci.* **122**: 101–108.
- Karimi, M., Inze, D., and Depicker, A.** (2002). GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**: 193–195.
- Karniely, S. and Pines, O.** (2005). Single translation- dual destination: mechanisms of dual protein targeting in eukaryotes. *EMBO Rep* **6**: 420–425.
- Karve, A., Rauh, B.L., Xia, X., Kandasamy, M., Meagher, R.B., Sheen, J., and Moore, B. d.** (2008). Expression and evolutionary features of the hexokinase gene family in *Arabidopsis*. *Planta* **228**: 411–425.
- Kay, R., Chan, A., Daly, M., and Mcpherson, J.** (1987). Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes Robert. *Science* **236**: 8–11.
- Kerppola, T.K.** (2006). Visualization of molecular interactions by fluorescence complementation the Bimolecular Fluorescence Complementation assay. *Nat Rev Mol Cell Biol* **7**: 449–456.
- Kluth, J., Schmidt, A., März, M., Krupinska, K., and Lorbiecke, R.** (2012). *Arabidopsis* ZINC RIBBON3 is the ortholog of yeast mitochondrial HSP70 escort protein HEP1 and belongs to an ancient protein family in mitochondria and plastids. *FEBS Lett.* **586**: 3071–3076.
- Kohalmi, S.E., Reader, L.J. V., Alon, S., Nowak, J., Haughn, G.W., and Crosby, W.L.** (1998). Identification and characterization of protein interactions using the yeast 2-hybrid system. *In: Plant Molecular Biology Manual*, Kluwer Academic Publishers, Netherlands.
- Köhler, R.H. and Hanson, M.R.** (2000). Plastid tubules of higher plants are tissue-specific and developmentally regulated. *J. Cell Sci.* **113**: 81–89.
- Kojima, H., Watanabe, Y., and Numata, O.** (1997). The dual functions of Tetrahymena citrate synthase are due to the polymorphism of its isoforms. *J. Biol. Chem.* **122**: 998–1003.
- Kolotilin, I., Kaldis, A., Pereira, E.O., Laberge, S., and Menassa, R.** (2013). Optimization of transplastomic production of hemicellulases in tobacco: effects of expression cassette configuration and tobacco cultivar used as production platform on recombinant protein yields. *Biotechnol. Biofuels* **6**: 65–80.
- Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H.** (2009a). Six classes of nuclear localization signals specific to different binding grooves of importin. *J. Biol. Chem.* **284**: 478–485.

- Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H.** (2009b). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 10171–10176.
- Krause, K., Oetke, S., and Krupinska, K.** (2012). Dual targeting and retrograde translocation: Regulators of plant nuclear gene expression can be sequestered by plastids. *Int. J. Mol. Sci.* **13**: 11085–11101.
- Krebs, J., Mueller-Roeber, B., and Ruzicic, S.** (2010). A novel bipartite nuclear localization signal with an atypically long linker in DOF transcription factors. *J. Plant Physiol.* **167**: 583–586.
- Kwok, E.Y. and Hanson, M.R.** (2004a). GFP-labelled Rubisco and aspartate aminotransferase are present in plastid stromules and traffic between plastids. *J. Exp. Bot.* **55**: 595–604.
- Kwok, E.Y. and Hanson, M.R.** (2003). Microfilaments and microtubules control the morphology and movement of non-green plastids and stromules in *Nicotiana tabacum*. *Plant J.* **35**: 16–26.
- Kwok, E.Y. and Hanson, M.R.** (2004b). Plastids and stromules interact with the nucleus and cell membrane in vascular plants. *Plant Cell Rep.* **23**: 188–195.
- Kwok, E.Y. and Hanson, M.R.** (2004c). Stromules and the dynamic nature of plastid morphology. *J. Microsc.* **214**: 124–137.
- de Las Rivas, J. and Fontanillo, C.** (2010). Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.* **6**: 1–8.
- Latrasse, D., Jégu, T., Meng, P.H., Mazubert, C., Hudik, E., Delarue, M., Charon, C., Crespi, M., Hirt, H., Raynaud, C., Bergounioux, C., and Benhamed, M.** (2013). Dual function of MIPS1 as a metabolic enzyme and transcriptional regulator. *Nucleic Acids Res.* **41**: 2907–2917.
- Launholt, D., Grønlund, J.T., Nielsen, H.K., and Grasser, K.D.** (2007). Overlapping expression patterns among the genes encoding Arabidopsis chromosomal high mobility group (HMG) proteins. *FEBS Lett.* **581**: 1114–1118.
- Lee, S.C., Hwang, I.S., and Hwang, B.K.** (2011). Overexpression of the pepper antimicrobial protein CaAMP1 gene regulates the oxidative stress- and disease-related proteome in *Arabidopsis*. *Planta* **234**: 1111–1125.
- Leister, D.** (2003). Chloroplast research in the genomic age. *Trends Genet.* **19**: 47–56.
- Leister, D.** (2012). Retrograde signaling in plants: from simple to complex scenarios. *Front. Plant Sci.* **3**: 1–9.
- Leonard, P.M., Smits, S.H.J., Sedelnikova, S.E., Brinkman, A.B., Vos, W.M. De, Oost, J. Van Der, Rice, D.W., and Rafferty, J.B.** (2001). Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. *EMBO J.* **20**: 990–997.
- Li, C.L., Wang, M., Ma, X.Y., and Zhang, W.** (2014). NRG1, a putative mitochondrial pyruvate carrier, mediates ABA regulation of guard cell ion channels and drought stress responses in *Arabidopsis*. *Mol. Plant* **7**: 1508–1521.
- Liberles, J.S., Thórólfsson, M., and Martínez, A.** (2005). Allosteric mechanisms in ACT domain containing enzymes involved in amino acid metabolism. *Amino Acids* **28**: 1–12.

- Liu, J.X., Srivastava, R., Che, P., and Howell, S.H.** (2007). An endoplasmic reticulum stress response in *Arabidopsis* is mediated by proteolytic processing and nuclear relocation of a membrane-associated transcription factor, bZIP28. *Plant Cell Online* **19**: 4111–4119.
- Lu, Q., Tang, X., Tian, G., Wang, F., Liu, K., Nguyen, V., Kohalmi, S.E., Keller, W.A., Tsang, E.W.T., Harada, J.J., Rothstein, S.J., and Cui, Y.** (2010). *Arabidopsis* homolog of the yeast TREX-2 mRNA export complex: Components and anchoring nucleoporin. *Plant J.* **61**: 259–270.
- Maeda, H. and Dudareva, N.** (2012). The shikimate pathway and aromatic amino Acid biosynthesis in plants. *Annu. Rev. Plant Biol.* **63**: 73–105.
- Maeda, H., Shasany, A.K., Schnepf, J., Orlova, I., Taguchi, G., Cooper, B.R., Rhodes, D., Pichersky, E., and Dudareva, N.** (2010). RNAi suppression of Arogenate Dehydratase1 reveals that phenylalanine is synthesized predominantly via the arogenate pathway in petunia petals. *Plant Cell* **22**: 832–849.
- Magaña, M.M., Koo, S., Towle, C., and Osborne, T.F.** (2000). Different sterol regulatory distinct co-regulatory factors to activate the promoter for fatty acid synthase different sterol regulatory element-binding protein-1 isoforms utilize distinct co-regulatory factors to activate the promoter for fatty acid synthase. *J. Biol. Chem.* **275**: 4726–4733.
- Mani, M., Chen, C., Amblee, V., Liu, H., Mathur, T., Zwicke, G., Zabad, S., Patel, B., Thakkar, J., and Jeffery, C.J.** (2015). MoonProt: A database for proteins that are known to moonlight. *Nucleic Acids Res.* **43**: 277–282.
- Maréchal, A., Parent, J.S., Véronneau-Lafortune, F., Joyeux, A., Lang, B.F., and Brisson, N.** (2009). Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 14693–8.
- Marín, M., Thallmair, V., and Ott, T.** (2012). The intrinsically disordered N-terminal region of AtREM1.3 remorin protein mediates protein-protein interactions. *J. Biol. Chem.* **287**: 39982–39991.
- McCartney, A.W., Dyer, J.M., Dhanoa, P.K., Kim, P.K., Andrews, D.W., McNew, J.A., and Mullen, R.T.** (2004). Membrane-bound fatty acid desaturases are inserted co-translationally into the ER and contain different ER retrieval motifs at their carboxy termini. *Plant J.* **37**: 156–173.
- McFadden, G.I.** (2001). Chloroplast origin and integration. *Plant Physiol.* **125**: 50–53.
- Merkle, T.** (2011). Nucleo-cytoplasmic transport of proteins and RNA in plants. *Plant Cell Rep.* **30**: 153–176.
- Millar, A.H., Hill, S.A., and Leaver, C.J.** (1999). Plant mitochondrial 2-oxoglutarate dehydrogenase complex: purification and characterization in potato. *Biochem. J.* **343**: 327–334.
- Murray, M.G. and Thompson, W.F.** (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Res.* **6**: 715–732.
- Natesan, S.K.A., Sullivan, J.A., and Gray, J.C.** (2005). Stromules: A characteristic cell-specific feature of plastid morphology. *J. Exp. Bot.* **56**: 787–797.
- Ofran, Y. and Rost, B.** (2007). ISIS: Interaction sites identified from sequence. *Bioinformatics* **23**: 13–16.
- Ohlrogge, J. and Browse, J.** (1995). Lipid biosynthesis. *Plant Cell* **7**: 957–70.

- Pan, X., Peng, F.Y., and Weselake, R.J.** (2015). Genome-wide analysis of PHOSPHOLIPID : DIACYLGLYCEROL ACYLTRANSFERASE (PDAT) genes in plants reveals the eudicot-wide PDAT gene expansion and altered selective pressures acting on the core eudicot PDAT paralogs. *Plant Physiol.* **167**: 887–904.
- Parry, G.** (2015). The plant nuclear envelope and regulation of gene expression. *J. Exp. Bot.* **66**: 1673–1685.
- Piatigorsky, J.** (2003). Gene sharing, lens crystallins and speculations on an eye/ear evolutionary relationship. *Integr. Comp. Biol.* **43**: 492–499.
- Piatigorsky, J. and Horwitz, J.** (1996). Characterization and enzyme activity of argininosuccinate lyase delta-crystallin of the embryonic duck lens. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1295**: 158–164.
- Piskacek, S., Gregor, M., Nemethova, M., Grabner, M., Kovarik, P., and Piskacek, M.** (2007). Nine-amino-acid transactivation domain: Establishment and prediction utilities. *Genomics* **89**: 756–768.
- Raikhel, N.** (1992). Nuclear targeting in plants. *Plant Physiol.* **100**: 1627–1632.
- Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A.-M., Horos, R., Hentze, M.W., Preiss, T., and Millar, A.A.** (2016). In planta determination of the mRNA-binding proteome of *Arabidopsis* etiolated seedlings. *Plant Cell* **28**: 2435–2452.
- Renzette, N.** (2011). Generation of transformation competent *E. coli*. *Curr. Protoc. Microbiol.*
- Riddick, G. and Macara, I.G.** (2007). The adapter importin-alpha provides flexible control of nuclear import at the expense of efficiency. *Mol. Syst. Biol.* **3**: 118.
- Rippert, P., Puyaubert, J., Grisollet, D., Derrier, L., and Matringe, M.** (2009). Tyrosine and phenylalanine are synthesized within the plastids in *Arabidopsis*. *Plant Physiol.* **149**: 1251–1260.
- Rout, M.P. and Aitchison, J.D.** (2001). The nuclear pore complex as a transport machine. *J. Biol. Chem.* **276**: 16593–16596.
- Rudhe, C., Clifton, R., Chew, O., Zemam, K., Richter, S., Lamppa, G., Whelan, J., and Glaser, E.** (2004). Processing of the dual targeted precursor protein of glutathione reductase in mitochondria and chloroplasts. *J. Mol. Biol.* **343**: 639–647.
- Saitou, N. and Nei, M.** (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- Samanta, A., Gouranga, D., and Kumar Sanjoy, D.** (2011). Roles of flavonoids in Plants. *Int J Pharm Sci Tech* **6**: 12–35.
- Sampaleanu, L.M., Vallée, F., Slingsby, C., and Howell, P.L.** (2001). Structural studies of duck delta1 and delta2 crystallin suggest conformational changes occur during catalysis. *Biochemistry* **40**: 2732–2742.
- Schattat, M., Barton, K., Baudisch, B., Klosgen, R.B., and Mathur, J.** (2011). Plastid stromule branching coincides with contiguous endoplasmic reticulum dynamics. *Plant Physiol* **155**: 1667–1677.
- Schreiter, E.R., Sintchak, M.D., Guo, Y., Chivers, P.T., Sauer, R.T., and Drennan, C.L.** (2003). Crystal structure of the nickel-responsive transcription factor NikR. *Nat. Struct. Biol.* **10**: 794–799.

- Scott, E.M. and Pillus, L.** (2010). Homocitrate synthase connects amino acid metabolism to chromatin functions through Esa1 and DNA damage. *Genes Dev.* **24**: 1903–1913.
- Seo, P.J., Kim, S.G., and Park, C.M.** (2008). Membrane-bound transcription factors in plants. *Trends Plant Sci.* **13**: 550–556.
- Shaner, N., Steinbach, P., and Tsien, R.** (2005). Full-Text. *Nat. Methods* **2**: 905–909.
- Sheen, J.** (2011). Transient expression in *Arabidopsis* mesophyll protoplasts. *Nat. Protoc.* **2**: 660–664.
- Silhavy, D., Molnár, A., Lucioli, A., Szittya, G., Hornyik, C., Tavazza, M., and Burgyán, J.** (2002). A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. *EMBO J.* **21**: 3070–3080.
- Silva-Filho, M.C.** (2003). One ticket for multiple destinations: Dual targeting of proteins to distinct subcellular locations. *Curr. Opin. Plant Biol.* **6**: 589–595.
- Sluchanko, N.N. and Gusev, N.B.** (2017). Moonlighting chaperone-like activity of the universal regulatory 14-3-3 proteins. *FEBS J.* **284**: 1279–1295.
- Smith-Uffen, M.E.S.** (2014). Changing the substrate specificity of arogenate dehydratases (ADTs) from *Arabidopsis thaliana*. Biology, M.Sc., Western University, London, Ontario.
- Smith, H.M., Hicks, G.R., and Raikhel, N. V** (1997). Importin alpha from *Arabidopsis thaliana* is a nuclear import receptor that recognizes three classes of import signals. *Plant Physiol.* **114**: 411–417.
- So, Y.Y., Bomblies, K., Seung, K.Y., Jung, W.Y., Mi, S.C., Jong, S.L., Weigel, D., and Ji, H.A.** (2005). The 35S promoter used in a selectable marker gene of a plant transformation vector affects the expression of the transgene. *Planta* **221**: 523–530.
- Sonnhammer, E.L., von Heijne, G., and Krogh, A.** (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings* **6**: 175–182.
- Sparkes, I. and Brandizzi, F.** (2012). Fluorescent protein-based technologies: Shedding new light on the plant endomembrane system. *Plant J.* **70**: 96–107.
- Stahl, U., Carlsson, A., Lenman, M., Dahlqvist, A., Huang, B., Banás, W., Banás, A., and Stymne, S.** (2004). Cloning and functional characterization of a phospholipid : diacylglycerol acyltransferase from *Arabidopsis*. *Am. Soc. Plant Biol.* **135**: 1324–1335.
- Styranko, D.M.** (2011). Profiling *Arabidopsis* arogenate dehydratases: dimerization and subcellular localization pattern. Biology, M.Sc. thesis, Western University, London, Ontario.
- Sun, X., Feng, P., Xu, X., Guo, H., Ma, J., Chi, W., Lin, R., Lu, C., and Zhang, L.** (2011). A chloroplast envelope-bound PHD transcription factor mediates chloroplast signals to the nucleus. *Nat. Commun.* **2**: 477.
- Tan, K., Li, H., Zhang, R., Gu, M., Clancy, S.T., and Joachimiak, A.** (2008). Structures of open (R) and close (T) states of prephenate dehydratase (PDT)- Implication of allosteric regulation by l-phenylalanine. *J. Struct. Biol.* **162**: 94–107.
- Thabet, I., Guirimand, G., Guihur, A., Lanoue, A., Courdavault, V., Papon, N., Bouzid, S., Giglioli-Guivarc’h, N., Simkin, A.J., and Clastre, M.** (2011). Characterization and subcellular localization of geranylgeranyl diphosphate synthase from *Catharanthus roseus*. *Mol. Biol. Rep.* **39**: 3235–3243.

- Thoms, S.** (2015). Import of proteins into peroxisomes: piggybacking to a new home away from home. *Open Biol.* **5**: 150148.
- Tzin, V. and Galili, G.** (2010). New Insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Mol. Plant* **3**: 956–972.
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W.** (2010). Lignin biosynthesis and structure. *Plant Physiol.* **153**: 895–905.
- Verma, D., Samson, N.P., Koya, V., and Daniell, H.** (2008). A protocol for expression of foreign genes in chloroplasts. *Nat. Protoc.* **3**: 739–758.
- Vik, A. and Rine, J.** (2000). Membrane biology: Membrane-regulated transcription. *Curr. Biol.* **10**: 869–871.
- Vitha, S., McAndrew, R.S., and Osteryoung, K.W.** (2001). FtsZ ring formation at the chloroplast division site in plants. *J. Cell Biol.* **153**: 111–119.
- Vivan, A.L., Caceres, R.A., Abrego, J.R.B., Borges, J.C., Neto, J.R., Ramos, C.H.I., De Azevedo, W.F., Basso, L.A., and Santos, D.S.** (2008). Structural studies of prephenate dehydratase from *Mycobacterium tuberculosis* H37Rv by SAXS, ultracentrifugation, and computational analysis. *Proteins Struct. Funct. Genet.* **72**: 1352–1362.
- Vogt, T.** (2010). Phenylpropanoid biosynthesis. *Mol. Plant* **3**: 2–20.
- Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D.** (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**: 949–956.
- Waese, J. and Provart, N.J.** (2016). The bio-analytic resource: data visualization and analytic tools for multiple levels of plant biology. *Curr. Plant Biol.* **7**: 2–5.
- Wang, Y., Mortimer, J.C., Davis, J., Dupree, P., and Keegstra, K.** (2013). Identification of an additional protein involved in mannan biosynthesis. *Plant J.* **73**: 105–117.
- Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.J., Su, Z., and Wu, W.H.** (2008). Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiol.* **148**: 1201–1211.
- Wegrzyn, J. et al.** (2009). Function of mitochondrial STAT3 in cellular respiration. *Science* **323**: 793–7.
- Wente, S.R. and Rout, M.P.** (2010). The nuclear pore complex and nuclear transport. *cold spring Harb. Perspect. Biol.* **2**: 1–20.
- Wise, A., Liu, Z., and Binns, A.N.** (2006). Three methods for the introduction of foreign DNA into *Agrobacterium*. *Meth. Mol. Biol.* **343**: 43–53.
- Xue, T., Wang, D., Zhang, S., Ehrling, J., Ni, F., Jakab, S., Zheng, C., and Zhong, Y.** (2008). Genome-wide and expression analysis of protein phosphatase 2C in rice and *Arabidopsis*. *BMC Genomics* **9**: 550.
- Yamada, T., Matsuda, F., Kasai, K., Fukuoka, S., Kitamura, K., Tozawa, Y., Miyagawa, H., and Wakasa, K.** (2008). Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of phenylalanine and tryptophan. *Plant Cell* **20**: 1316–29.
- Yao, L.H., Jiang, Y.M., Shi, J., Tomas-Barberan, F. a, Datta, N., Singanusong, R., and Chen, S.S.** (2004). Flavonoids in food and their health benefits. *Plant Foods Hum. Nutr.* **59**: 113–122.

- Yogev, O. and Pines, O.** (2011). Dual targeting of mitochondrial proteins: Mechanism, regulation and function. *Biochim. Biophys. Acta - Biomembr.* **1808**: 1012–1020.
- Yoo, H., Widhalm, J.R., Qian, Y., Maeda, H., Cooper, B.R., Jannasch, A.S., Gonda, I., Lewinsohn, E., Rhodes, D., and Dudareva, N.** (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine:phenylpyruvate aminotransferase. *Nat. Commun.* **4**: 2833.
- Yu, H. et al.** (2008). High-quality binary protein interaction map of the yeast interactome network. *Science* **322**: 104–110.
- Zhai, Z., Liu, H., and Shanklin, J.** (2017). Phosphorylation of WRINKLED1 by KIN10 Results in its Proteasomal Degradation, Providing a Link Between Energy Homeostasis and Lipid Biosynthesis. *Plant Cell* **29**: 871–889.
- Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H.** (2006). *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* **1**: 641–646.
- Zhichang, Z., Wanrong, Z., Jinping, Y., Jianjun, Z., Zhen, L., Xufeng, L., and Yang, Y.** (2010). Over-expression of *Arabidopsis* DnaJ (Hsp40) contributes to NaCl-stress tolerance. *African J. Biotechnol.* **9**: 972–978.
- Zhou, R., Su, W.H., Zhang, G.F., Zhang, Y.N., and Guo, X.R.** (2016). Relationship between flavonoids and photoprotection in shade-developed *Erigeron breviscapus* transferred to sunlight. *Photosynthetica* **54**: 201–209.

Appendices

Appendix 1 Primers.

Primer name ^a	Sequence (5' to 3') ^b	Direction	Recognition sequence	Annealing temperature (C)
PRM-ADT5-F	<u>AATCCGCGCGCGGGAGCGATATGGATTGGACCATGTTC</u>	Forward	<i>MauBI</i>	62
PRM-ADT5-R	<u>TTCAC TCGAGGATTTGAGTCGAAAAATGAAAAAATTGCT</u> <u>GTTTATTTATTTTTCTGGTAAGAGGAAACAAAAGAGTAA</u>	Reverse	<i>XhoI</i>	62
ADT5-FL-F	<u>TACATTGCTAGCCAAACCATTTGCCTGCGTTCTCG</u>	Forward	<i>NheI</i>	65
ADT5-IS-F	<u>TACATTGCTAGCCCGTTAACGATAACAGATCTATCTCCCG</u> <u>C</u>	Forward	<i>NheI</i>	63
ADT5-S-F	<u>TACATTGCTAGCCCGTGTGCGGTATCAAGGAGTTCCG</u>	Forward	<i>NheI</i>	63
ADT5-R	<u>ATTGTAGCTAGCTACGTCTTCGCTAGGTAACGTGGACCAT</u> <u>G</u>	Reverse	<i>NheI</i>	64
ADT4-FL-F	<u>TACATTACTAGTCAAGCCGCAACGTCGTGTGATCT</u>	Forward	<i>SpeI</i>	63
ADT4-IS-F	<u>TACATTACTAGTCCGTTGACTATTACTGATCTATCTCCGG</u> <u>CAC</u>	Forward	<i>SpeI</i>	64
ADT4-S-F	<u>TACATTACTAGTCGTGTAGCTTACCAAGGCGTTCCC</u>	Forward	<i>SpeI</i>	62
ADT4-R	<u>ATGTAACTAGTTGCTTCTTCTGTGGATGTCATGGACC</u>	Reverse	<i>SpeI</i>	62
ADT2-FL-F	<u>TACATTACTAGTGCAATGCACACTGTTTCGATTGTCG</u>	Forward	<i>SpeI</i>	61
ADT2-IS-F	<u>TACATTACTAGTAAGCCGTTATCATCGAACCAGCTCAC</u>	Forward	<i>SpeI</i>	61
ADT2-S-F	<u>TACATTACTAGTCGTGTTGCGTATCAGGGAGTACGAG</u>	Forward	<i>SpeI</i>	61
ADT2-R	<u>ATGTAACTAGTCTGCTTGTTGTCGAGCATTGTAGTGTCCA</u> <u>CTGGGTAGCTTC</u>	Reverse	<i>SpeI</i>	62
pEG102-ADT5-FL-F	<u>GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCAAACC</u> <u>ATTTTCGCCTGCG</u>	Forward	<i>attB1</i>	67
pEG102-ADT5-FL-R	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTTTACGTCTTC</u> <u>GCTAGGTAACG</u>	Reverse	<i>attB2</i>	68

Primer name ^a	Sequence (5' to 3') ^b	Direction	Recognition sequence	Annealing temperature (C)
pEG102-ADT5-ΔA-R	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTAATAGACTA</u> <u>GAACCATGCAACGG</u>	Reverse	<i>attB2</i>	67
pEG102-ADT5-ΔAC-R	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTTAAGAGTAG</u> <u>ATCCATGAGACGG</u>	Reverse	<i>attB2</i>	68
pEG102-ADT5-ΔT-F	<u>GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCGTGTC</u> <u>GCGTATCAAGGAGTTCCG</u>	Forward	<i>attB1</i>	69
pEG102-ADT4-FL-F	<u>GGGGACAAGTTTGTACAAAAAAGCAGGCTCAAGCCGCA</u> <u>ACGTCGTG</u>	Forward	<i>attB1</i>	68
pEG102-ADT4-ΔA-R	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTACGCTAGGG</u> <u>CTTGCGGGTGAGAG</u>	Reverse	<i>attB2</i>	68
pEG102-ADT4-ΔAC-R	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTATAGACTAG</u> <u>AACCATGCAACGG</u>	Reverse	<i>attB2</i>	68
pEG102-ADT4-ΔT-F	<u>GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCGTGTA</u> <u>GCTTACCAAGGCGTTCCC</u>	Forward	<i>attB1</i>	68
IMPA6-FL-F	<u>GGGGACAAGTTTGTACAAAAAAGCAGGCTCTTCTTACA</u> <u>AACCAAGCGCG</u>	Forward	<i>attB1</i>	83
IMPA6-FL-RY	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTGTCAACCAA</u> <u>AGTTGAATCCACC</u>	Reverse	<i>attB2</i>	85
IMPA6-FL-RBi	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTGACCAAAGT</u> <u>TGAATCCACC</u>	Reverse	<i>attB2</i>	84
Plip-FL-F	<u>GGGGACAAGTTTGTACAAAAAAGCAGGCTCTATGCCCC</u> <u>TTATTCATCGG</u>	Forward	<i>attB1</i>	82
Plip-FL-RBi	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTGCAGATTAV</u> <u>GGTCAATACG</u>	Reverse	<i>attB2</i>	84
Plip-FL-R	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACTTCA</u> <u>GGTCAATACG</u>	Reverse	<i>attB2</i>	83
Plip-FL-R	<u>GGGGACCACTTTGTACAAGAAAGCTGGGTTTCACTTCA</u> <u>GGTCAATACG</u>	Reverse	<i>attB2</i>	83

Primer name ^a	Sequence (5' to 3') ^b	Direction	Recognition sequence	Annealing temperature (C)
Plip-FL-R	GGGGACCACTTTGTACAAGAAAGCTGGGTT <u>CAGCTTC</u> <u>AGGTCAATACG</u>	Reverse	<i>attB2</i>	83
Seq PRM-ADT5	TGCAAGAGTACACGTCATTCCTCAG	Forward		65
GFP-seq-R	CCCGGTGGTGCAGATGAAC	Reverse		61
Seq ADT5-FL	ACGCCATGGTCCACGTTACC	Forward		60
PpsbA-F	CCAAGATTTTACCT	Forward		60

^aFL: full length, IS: I region and short form, S: short form, A: ACT domain, AC:ACT and CAT domain, T: transit peptide, pEG102: pEarlyGate 102, PRM: Native promoter, Δ: deletion primer, RY: Yeast-S-Hybrid primer, Bi: Bi-molecular Fluorescence. Seq: Sequencing primers.

^bBold: nucleotides added to provide a docking for the restriction enzymes or promote efficient recombination; Underline: sequence complementary to template, Italic: nucleotide added to adjust frame.

Appendix 2 List of vectors.

Plasmid name	Promoter ^a	Fusion Domains ^b		Selectable Marker(s) ^c	Host(s) ^d
		N	C		
pDONOR221	-	-	-	<i>kan^r</i>	<i>E. coli</i>
pEarlyGate101	35S35S	-	YFP	<i>kan^r, ccdB, BAR</i>	<i>E. coli, Agro, Plant</i>
pEarlyGate102	35S35S	-	CFP	<i>kan^r, ccdB, BAR</i>	<i>E. coli, Agro, Plant</i>
pEarlyGate201-YN	35S35S	HA Tag	YFP-N	<i>kan^r, ccdB, BAR</i>	<i>E. coli, Agro, Plant</i>
pEarlyGate202-YC	35S35S	Flag Tag	YFP-C	<i>kan^r, ccdB, BAR</i>	<i>E. coli, Agro, Plant</i>
pKGWFS7	-	-	GFP, GUS	<i>sp^r, ccdB, kan^r</i>	<i>E. coli, Agro, Plant</i>
pCB5	35S35S	-	CFP	<i>gen^r, BAR</i>	<i>E. coli, Agro, Plant</i>
pCEC4	<i>PpsbA</i>	-	GFP	<i>amp^r, sp^r</i>	<i>E.coli, Plant</i>
pGBKT7-DEST	<i>ADH1</i>	cMyc Tag, GAL4-BD	-	<i>kan^r, TRP1</i>	<i>E. coli, Yeast</i>
pGADT7-DEST	<i>ADH1</i>	HA Tag, GAL4-AD	-	<i>amp^r, LEU2</i>	<i>E. coli, Yeast</i>
P19	35S35S	-	-	<i>kan^r</i>	<i>Agro</i>
GFP	35S35S	-	-	<i>kan^r</i>	<i>Agro</i>
GFP	<i>PpsbA</i>	-	-	<i>amp^r, sp^r</i>	<i>E.coli, Plant</i>

^a35S35S: Cauliflower Mosaic Virus 35S, *PpsbA*: Photosystem II protein D1, *ADH1*: *ALCOHOL DEHYDROGENASE 1*,

^bA tag or additional domain fused at the N- or C- terminus of the protein. YFP- N or YFP- C has N or C terminal half of YFP.

^c*kan^r*: kanamycin resistance, *ccdB*: bacterial negative selectable marker, *BAR*: Basta resistance, *sp^r*: spectinomycin resistance, *amp^r*: ampicillin resistance, *gen^r*: gentamycin resistance, *TRP1*: tryptophan prototrophy, *LEU2*: leucine prototrophy.

^d*Agro*: *A. tumefaciens*.

Appendix 3 Viability of yeast populations.

Yeast Populations	Cells/mL
DB-ADT5 in haploid AH109	2×10^7
TA-cDNA in haploid Y187 yeast	3×10^6
Diploids	1×10^5

Appendix 4 Journal of Experimental Botany authorship permission.



RightsLink®



Title: Subcellular localization of Arabidopsis arogenate dehydratases suggests novel and non-enzymatic roles

Author: Bross, Crystal D.; Howes, Travis R.

Publication: Journal of Experimental Botany

Publisher: Oxford University Press

Date: 2017-03-10

Copyright © 2017, Oxford University Press

Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

Curriculum Vitae

Sara Abolhassani Rad

EDUCATION

2012- 2017	Ph.D. Biology (Cell and Molecular Biology) Western University, London, Ontario, Canada
2006-2009	M.Sc. Biotechnology Engineering, Islamic Azad University, Science & Research Branch, Tehran, Iran
2001-2004	BSc, Agricultural Engineering, Shahid Bahonar University, Kerman, Iran

MAJOR AWARDS & DISTINCTIONS

2012-2016	Western graduate research scholarship (WGRS)
2015	Duff travel bursary award, Society of Canadian Plant Biologists
2015	Graduate students travel award, Biology department
2014	Duff travel bursary award, Society of Canadian Plant Biologists
2012-2014	Dean's doctoral scholarship, Western University

WORK EXPERIENCE

2012-2017	Teaching Assistant, Department of Biology, Western University
2012-2017	Research Assistant, Department of Biology, Western University
2014-2017	Doctoral Researcher, Agriculture and Agri-Food, London, Canada

REFEREED PUBLICATIONS

Sara Abolhassani Rad, Emily J. Clayton, Emily J. Cornelius, Travis R. Howes, Susanne E. Kohalmi. Moonlighting protein: putting the spotlight on enzymes.
Manuscript submitted in Plant Signaling and Behavior

Sara Abolhassani Rad, Igor Kolotilin, Rima Menassa, Susanne E. Kohalmi. The diverse destiny of closely related fusion proteins when expressed in chloroplasts.
Manuscript in preparation

Sara Abolhassani Rad, Susanne E. Kohalmi. PDAT1 plays an adapter role in nuclear localization of ADT5 in *Arabidopsis thaliana*. Manuscript in preparation

Crystal D. Bross, Travis R. Howes, **Sara Abolhassani Rad**, Susanne E. Kohalmi. Subcellular localization of *Arabidopsis* arogenate dehydratase suggests novel and differential roles. *Journal of Experimental Botany*, 68:1425-1440, 2017

Sara Abolhassani Rad, Hossein Shahbani Zahiri, Kambiz Akbari Noghabi, Sarah Rajai, Reza Heidari, and Leila Mojalali. Type 2 IDI performs better than type 1 for improving lycopene production in metabolically engineered *E. coli* strains. *World Journal of Microbiology and Biotechnology*, 28. 1: 313-321, 2012

Hossein Shahbani Zahiri, Kambiz Akbari Noghabi, Mojtaba Samoodi, Negar Omid Yeganeh, **Sara Abolhassani Rad**, Azam Safari, Fatemeh Hosseini, and Reza Hajhosseini. Effect of concomitant lycopene biosynthesis on CoQ10 accumulation in transformed *Escherichia coli* strains. *Iranian Journal of Biotechnology*, 7: 224-232, 2009

NON-REFEREED CONTRIBUTIONS

Sara Abolhassani Rad, Susanne E. Kohalmi. The mystery of ADT5 nuclear localization. Canadian Society of Plant Biologists Eastern Regional Meeting, Hamilton, Canada, Nov. 2016

Sara Abolhassani Rad, Susanne E. Kohalmi. Transcriptional and translational expression patterns of ADT5 in transgenic *Arabidopsis*. Plant Biotech Conference, Kingston, Canada, June 2016

Sara Abolhassani Rad, Susanne E. Kohalmi. Stromules, Tunnels from chloroplasts to the nucleus? Fallona Family showcase, London, Canada, Dec. 2015

Sara Abolhassani Rad, Susanne E. Kohalmi. Piggy- back transport of Arogenate dehydratase 5 into the nucleus in *Arabidopsis thaliana* with Importin α isoform 6. Canadian Society of Plant Biologists Eastern Regional Meeting, Toronto, Canada, Nov. 2015

Sara Abolhassani Rad, Susanne E. Kohalmi. Analyzing the non-enzymatic role of Arogenate dehydratase 5 in *Arabidopsis thaliana*. Botany Conference, Edmonton, Canada, Jul. 2013

Sara Abolhassani Rad, Travis R. Howes, Susanne E. Kohalmi. Identification of sequences required for nuclear targeting of ADT5. Canadian Society of Plant Biologists Conference, Quebec City, Canada, Jun. 2013

Sara Abolhassani Rad and Hossein Shahbani Zahiri. Microbial production of lycopene by using the metabolic engineering tools. 3rd International Congress of Biochemistry and Molecular Biology, Tehran, Iran, Nov. 2009

Azam Safari, Hossein Shahbani Zahiri, **Sara Abolhassani Rad**, Fatemeh Hosseini. Thermostable alkaline amylase production by a native bacterium isolated from Tehran soil. 10th National Congress of Biochemistry, Tehran, Iran, Nov. 2009

Sara Abolhassani Rad, Hossein Shahbani Zahiri, Azam Safari. Evaluation of chimeric mevalonate pathway to enhance lycopene production in *E. coli*. 14th European Congress on Biotechnology, Barcelona, Spain, Sep. 2009

Sara Abolhassani Rad, Azam Safari, Fatemeh Hosseini, Hossein Shahbani Zahiri. Lycopene production in *E. coli* by metabolic engineering. 6th National Biotechnology Congress, Tehran, Iran, Jul. 2009