

**Characterisation of the unknown gene and
the corresponding protein At3g29075 in
*Arabidopsis thaliana***

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch–Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Selvakumar Sukumaran

aus Coimbatore, India

Bonn, 2020

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen
Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Dorothea Bartels
2. Gutachter: Prof. Dr. Peter Dörmann

Tag der Promotion: 23.03.2020

Erscheinungsjahr: 2020

Contents

Abbreviations	1
List of figures	3
List of tables	4
List of supplementary figures	4
1. Introduction	6
1.1. The phospholipase group of enzymes	7
1.1.1. The structural characteristics of phospholipase D and isoforms	7
1.2. Phosphatidic acid	10
1.3. Proteins under the control of phospholipase D	12
1.4. The unknown protein At3g29075 from <i>A. thaliana</i>	13
1.4.1. The protein-coding sequence of At3g29075	13
1.5. The objective of the study	14
2. Materials and methods	16
2.1. Materials	16
2.1.1. Chemicals	16
2.1.2. Equipment	16
2.1.3. Computer programs and databases	17
2.1.4. Enzymes and markers	19
2.1.5. Primer	19
2.1.6. Vectors	21
2.1.7. Kits	22
2.1.8. DNA-sequencing	22
2.1.9. Quantification of proteins and RNA	22
2.2. Plant material	22
2.2.1. Sterilisation of seeds	23
2.3. Growth conditions	23
2.3.1. Growing on soil	23
2.3.2. Growing on MS-plates	23
2.3.3. Germination on Blotting paper	24
2.4. Stress conditions	24
2.5. Microorganisms	25
2.5.1. Bacterial strains	25
2.5.2. Media for microorganisms	26
2.5.3. Media supplements	26

2.5.4. Generation of rubidium chloride–competent cells.....	26
2.5.5. Generation of electro-competent <i>A. tumefaciens</i>	27
2.6. Glycerol Stocks.....	27
2.7. Cloning methods.....	27
2.7.1 Electrophoresis of nucleic acids (Adkins & Burmeister, 1996).....	28
2.7.2. Isolation and purification of plasmid DNA (Sambrook et al., 1989)	28
2.7.3. Purification of DNA.....	29
2.7.4. Restriction digestion	29
2.7.5. Ligation (Sambrook et al., 1989).....	29
2.7.6. Transformation of rubidium chloride-competent <i>E. coli</i> (adapted from Hanahan, 1983)	30
2.7.7. Transformation of electro-competent <i>A. tumefaciens</i> (adapted from Tung & Chow, 1995)	30
2.7.8. Isolation of genomic DNA.....	31
2.8. Amplification of DNA fragments by PCR (Mullis & Faloona, 1987).....	32
2.8.1. Genotyping of overexpression mutants	32
2.8.2. Colony–PCR (Sambrook et al., 1989).....	33
2.9. Extraction of RNA from plant tissue.....	33
2.9.1. RNA extraction with urea (adapted from Missihoun et al., 2011)	33
2.9.2. Phenolic RNA-extraction method (adapted from Valenzuela-Avendaño et al., 2005)	33
2.10. Reverse transcription-polymerase chain reaction.....	34
2.10.1. DNase treatment (adapted from Innis et al., 2012).....	34
2.10.2. Synthesis of cDNA (adapted from Innis et al., 2012)	34
2.11. Extraction of proteins.....	35
2.11.1. Extraction of total proteins (Röhrig et al., 2008).....	35
2.11.2. Direct–method (Laemmli, 1970)	35
2.11.3. Protease inhibitor Assay.....	35
2.12. Quantification of nucleic acids and proteins.....	36
2.12.1. Estimation of nucleic acid concentrations	36
2.12.2. Estimation of protein concentrations (adapted from Bradford, 1976)	36
2.13. Over-expression and isolation of recombinant proteins.....	37
2.13.1. Cobalt column gravity-flow purification	38
2.14. Electrophoresis of proteins.....	39
2.14.1. Isoelectric focusing (first dimension).....	39
2.14.2. SDS-PAGE (second dimension) (adapted from Laemmli, 1970).....	40
2.14.3. Staining of polyacrylamide gels.....	41
2.15. Protein blot (adapted from Towbin et al., 1979).....	42
2.16. Overview of the different antibodies used in this thesis	43

2.17. Protein-lipid interactions	43
2.17.1. Protein-lipid-overlay assay (adapted from Deak et al., 1999).....	43
2.17.2. Liposome-binding assay (adapted from Zhang et al., 2004)	44
2.18. Transient and stable transformation.....	44
2.18.1. Transient transformation via particle gun bombardment	44
2.18.2. <i>A. tumefaciens</i> -mediated stable transformation of <i>A. thaliana</i> (adapted from Clough & Bent, 1998)	45
3. Results.....	46
3.1. Characteristics of At3g29075	46
3.2. Protein-protein interactions	47
3.3. Gene co-expression network of At3g29075.....	48
3.4. Post-translational modifications of At3g29075.....	49
3.5. Detection of At3g29075 protein	50
3.6. Expression analysis of At3g29075	51
3.6.1. Tissue-specific expression of At3g29075	51
3.6.2. Developmental stage-specific expression of At3g29075	52
3.6.3. Expression of At3g29075 during seed germination	53
3.7. Dehydration stress	56
3.7.1. Memory response patterns of At3g29075 upon dehydration stress	57
3.8. Expression of At3g29075 upon salt stress.....	59
3.9. Analysis of promoter region	60
3.9.1. ABA –Treatment.....	61
3.9.2. Expression of At3g29075 upon continuous light and dark condition	62
3.10. Production of recombinant proteins	63
3.10.1. Amplification and cloning of At3g29075 into pQLinkHD	64
3.10.2. Expression and isolation of At3g29075	65
3.11. Cloning of Nt-At3g29075 into pQLinkHD	67
3.11.1. Purification of Nt-At3g29075 recombinant protein	68
3.12. Cloning of Ct-At3g29075 into pET28a	69
3.13. Lipid-binding.....	69
3.13.1. Protein-lipid interactions on nitrocellulose membranes	69
3.13.2. Liposome-binding assay.....	70
3.14. The generation of the At3g29075 over-expression line	72
3.14.1. Selection of Hygromycin-resistant transformants	73
3.15. Phenotypic analysis of mutant plants	74
3.15.1. F ₁ -Generation	74

3.15.2. F ₂ -Generation	74
3.15.3. F ₃ -Generation	75
3.16. At3g29075 protein profiling in wild type and mutants.....	76
3.17. Intracellular localization of At3g29075 - Transient transformation.....	77
4. Discussion.....	78
4.1. Gene analysis of At3g29075.....	78
4.2. At3g29075 is intrinsically disordered.....	78
4.3. Phosphorylation of At3g29075.....	79
4.4. Regulation of At3g29075.....	80
4.5. Expression of At3g29075 during different developmental stages correlated to phosphatidic acid .	80
4.6. Tissue-specific expression of At3g29075 via phosphatidic acid regulation.....	81
4.7. Regulation of At3g29075 under stress condition	82
4.8. The Cloning and production of recombinant At3g29075 protein.....	84
4.9. Interaction of At3g29075 with phosphatidic acid	85
4.10. At3g29075 gene involvement in the flowering mechanism.....	86
4.11. Plant Lysine-rich proteins: A new classification for uncommon and beneficial lysine-rich motif .	87
4.11.1. Classification based on the clustering of amino acids for protein secondary structure prediction	88
4.11.2. Classification based on the functional diversity of the plant lysine-rich protein superfamily .	89
5. Summary	92
6. Outlook	93
7. Supplementary data	95
Vector maps.....	97
8. Reference	114
Acknowledgements	119

Abbreviations

A	adenine	miRNA	micro RNA
AA	amino acid	minute(s)	minute(s)
ABA	abscisic acid	mRNA	messenger RNA
bp	base pair	NO	nitric oxide
BSA	bovine serum albumin	nt	nucleotides
C	cytosine	OD	optical density
Col-0	Columbia-0	PA	phosphatidic acid
cDNA	complementary DNA	PAGE	polyacrylamide gel electrophoresis
CDS	coding sequence	PC	phosphatidylcholine
CRISPR	clustered regularly interspaced short palindromic repeats	PE	phosphatidylethanolamine
EF1α	Elongation factor 1-alpha	PCR	polymerase chain reaction
d	days	pH	pondus Hydrogenii
Da	Dalton	PIP2	phosphatidylinositol 4,5-bisphosphate
DEPC	diethylpyrocarbonate pI isoelectric point	PLD	phospholipase D
DGK	diacylglycerol kinase	PM	plasma membrane
d.H2O	distilled "milli-Q" water	PPI	phosphoinositides
DNA	deoxyribonucleic acid	rpm	rounds per minute
dNTPs	deoxyribonucleoside triphosphates	RNA	ribonucleic acid
e-value	expectation value	ROS	reactive oxygen species
et al.	et alii	RT	room temperature
G	guanine	RT-PCR	Reverse transcriptase-polymerase-chain reaction
g	gram	RuBisCO	Ribulose-1,5-bisphosphate carboxylase oxygenase,
g	gravity acceleration (9.81)	T	thymine
G-	protein guanine nucleotide-binding protein	Taq	Thermus aquaticus
h	hour(s)	T-DNA	transfer DNA
His-tag	histidine-affinity tag	TOF	time of flight
HsP	heat-shock protein	UV	ultraviolet
IDP	intrinsically disordered protein	V	volume
LPA	lysophosphatidyl acyltransferases	[v/v]	volume/volume
M	molar	[w/v]	weight/volume
MALDI	matrix-assisted-laserdesorption/ionization		
MAPK	mitogen-activated protein kinases		

Chemicals and buffers

3-AT	3-amino-1,2,4-triazole	PBS	Phosphate-buffered salt solution
APS	Ammonium persulfate	NaCl	Sodium chloride
CHAPS	3-[3(cholamidopropyl)dimethylammonio]-1-propanesulfonate	NADPH	nicotinamide adenine dinucleotide phosphate
DOC	Sodium deoxycholate	NaOH	Sodium hydroxide
EDTA	Ethylene diaminetetraacetic acid	NTA	Nitrilotriacetic acid
EtBr	Ethidium bromide	PIPES	Piperazine-N, N'-bis (2-ethanesulfonic acid)
HCl	Hydrochloric acid	SDS	Sodium dodecylsulfate
IPTG	Isopropyl- β -D-thiogalactopyranoside	SOC	Super Optimal Broth
KOH	Potassium hydroxide	TAE	Tris-Acetate-EDTA
LB	“Lysogeny broth” or “Luria broth.”	TBS	Tris-buffered salt solution
LiCl	Lithium chloride	TCA	Trichloroacetic acid
MgCl₂	Magnesium chloride	TEMED	Tetramethylethylenediamine
MOPS	3-(N-morpholino) propanesulfonic acid	Tris	Tris(hydroxymethyl)-aminomethane
MS	Murashige-Skoog-Medium	TWEEN	Polyoxyethylene (20) sorbitan Monolaurate

Institutes and databases

BAR	Bio-Analytic Resource for Plant Biology (Toronto, Canada)
EMBL	European Molecular Biology Laboratory (Heidelberg, Germany)
ExpASy	ExpASy Bioinformatic Resources Portal
GABI	German plant genomics research program.
NASC	Nottingham Arabidopsis Stock Centre (Nottingham, Great Britain)
NCBI	National Center for Biotechnology Information (Bethesda USA)
<i>RaptorX property</i>	Toyota Technological Institute at Chicago
SALK	Salk-Institute (La Jolla, USA)
STRING	Known and Predicted Protein-Protein Interactions
TAIR	The Arabidopsis Information Resource (Stanford, USA)
Toxinpred	Protein toxicity prediction based on Swissport and TrEMBL, (New Delhi, India)
UniProt	Uniprot Protein database

List of figures

Figure 1: Schematic representation of the Phospholipase cleavage sites.....	7
Figure 2: Schematic representation of <i>Arabidopsis thaliana</i> PLD genes.....	8
Figure 3: The role of different PLD isoforms throughout hyperosmotic stress such as salt stress and water deficiency.....	9
Figure 4: Schematic representation of the electrostatic/hydrogen bond switch model.....	10
Figure 5: Gene model of At3g29075.	13
Figure 6: BSA calibration line	36
Figure 7: Protein sequence of the At3g29075.	46
Figure 8: Nucleotide and protein sequence of the At3g29075.....	46
Figure 9: The toxicity analysis of the At3g29075 protein residues.	47
Figure 10: Predicted functional partners of the At3g29075 protein.....	48
Figure 11: Gene co-expression network of the At3g29075 gene.....	49
Figure 12: The different antibodies were tested (1 minute (Dilution: 1:2000)) on <i>Arabidopsis thaliana</i> leaves, root tissues, and the recombinant protein At3g29075.....	50
Figure 13: Expression of the At3g29075 gene in different plant tissues of wild type plants.	51
Figure 14: Expression of At3g29075 at the protein level in different tissues of wild type plants.	52
Figure 16: Expression of At3g29075 at the protein level in different developmental stages.	53
Figure 17: Schematic presentation of the transition from dry seed to a seedling at 72 h after the start of imbibition (HAI).....	54
Figure 18: Protein expression in different stages of the seed germination after the start of imbibition (HAI) in <i>Arabidopsis thaliana</i>	54
Figure 19: Two-dimensional analysis of total protein extracts from <i>Arabidopsis thaliana</i>	55
Figure 20: Expression of the At3g29075 gene in response to dehydration.....	56
Figure 21: Expression of the At3g29075 protein in response to dehydration.....	57
Figure 22: Expression of the At3g29075 gene in different stages of dehydration stress condition.	58
Figure 23: Expression of the At3g29075 protein in different stages of dehydration stress in <i>Arabidopsis thaliana</i> plants.	58
Figure 24: Expression of the At3g29075 gene upon salt stress condition from <i>Arabidopsis thaliana</i> plants.....	59
Figure 25: Expression of the protein At3g29075 upon salt stress condition from <i>Arabidopsis thaliana</i> plants.....	59
Figure 26: Promoter sequence of the At3g29075 (last) 1000 bp in <i>Arabidopsis thaliana</i>	60
Figure 27: Expression of the At3g29075 gene in control and ABA-treated conditions..	61
Figure 28: Expression of the At3g29075 protein in control and ABA-treated conditions.....	61
Figure 29: Expression of the At3g29075 gene upon continuous, light, and dark conditions for 24 h, 48 h & 72 h.....	62
Figure 30: Expression of the At3g29075 protein upon continuous, light, and dark conditions for 24 h, 48 h & 72 h.	63
Figure 31: Expression vector pQLinkHD - bearing the His-tagged fragment At3g29075.....	65
Figure 32: Induction of the full-size protein At3g29075.....	66
Figure 33: Induction of the full-size protein At3g29075 upon different time intervals (6-8 h).....	67
Figure 34: His-tag affinity chromatography of At3g29075-full-size of soluble protein fractions.....	67
Figure 35: Cloning strategy of the N-terminal At3g29075 gene in pQLinkHD expression system..	68
Figure 36: His-tag affinity chromatography of the Nt-At3g29075 of soluble protein fractions.....	69
Figure 37: Detection of the At3g29075 protein fragments in protein-lipid-overlay assays.	70
Figure 38: Liposome-binding assay with the At3g29075 protein.....	71

Figure 39: Construct for the generation of the At3g29075 over-expression line.....	72
Figure 40: Rapid selection of the pH2GW7-AT3G29075 overexpression line by Hygromycine B.	73
Figure 41: Phenotype analysis of the AT3G29075 over expression F1 generation line.....	74
Figure 42: Phenotype analysis of the AT3G29075 over expression F2 generation line.....	75
Figure 43: Phenotype analysis of the AT3G29075 over expression F3 generation line.....	75
Figure 44: Expression of the At3g29075 protein level in wild type and mutants.	76
Figure 45: Microscopic analysis of the intracellular localisation of the At3g29075 gene after particle bombardment.	77
Figure 46: Model for the regulation of expression of At3g29075 in response to dehydration in <i>Arabidopsis thaliana</i>	83
Figure 47: The lysine-rich repeats of the At3g29075 protein.	88

List of tables

Table 1: List of primers.....	19
Table 2: List of Antibodies.....	43
Table 3: Prediction of kinase-specific phosphorylation sites in At3g29075.....	49
Table 4: List of vectors and bacterial strains used for the production of recombinant protein At3g29075	63
Table 5: Triplet sate of lysine-rich amino acid clustering.....	88
Table 6: The quintuplet lysine-rich amino acid clustering.	89
Table 7: A. thaliana proteins with possible lysine-rich repeats.....	89

List of supplementary figures

Supplementary figure 1: Tissue-specific expression analysis of At3g29075 based on microarray data in <i>Arabidopsis thaliana</i> Data and pictures obtained from the BAR ePlant.	95
Supplementary figure 2: Immunodetection of At3g29075 in dry seed to a seedling at 72 h after the start of imbibition (HAI) in <i>Arabidopsis thaliana</i>	96
Supplementary figure 3: Post-translation modification sites of At3g29075 protein (left) and the quantitative value of phosphorylation values of the At3g29075 PTM sites (right). Source: FAT-PTM.	96
Supplementary figure 4: Schematic diagram of the pJET1.2 entry vector map.....	97
Supplementary figure 5: pET28a an expression vector map with 6x-his-tag.....	97
Supplementary figure 6: pET43.b expression vector map with his-tag.	98
Supplementary figure 7: pGEX4T-2 expression map with GST tag.....	98
Supplementary figure 8: pTwin1 expression vector map without a tag system.....	99
Supplementary figure 9: pGJ280 expression vector map with GFP tag for localisation.	99
Supplementary figure 10: pDONR tm 201 gateway entry vector map.	100
Supplementary figure 11: pQLinkHD gateway expression vector with His-tag.....	100
Supplementary figure 12: pH2GW7 gateway expression vector map with 35S promoter.	101
Supplementary figure 13: pDONR tm 207 gateway entry vector map.	101
Supplementary figure 14: The pEarlyGate103 gateway expression vector map consists of GFP-tag.	102
Supplementary figure 15: Tandem repeats in At3g29075 are marked in Green.....	102
Supplementary figure 16: Sequence alignment of At3g29075 cDNA with pET28a-At3g29075 insert shows 204 bp shorter.	103
Supplementary figure 17: Immunodetection of pET28a-At3g29075 recombinant protein with C-terminal At3g29075 antibody.....	104
Supplementary figure 18: Sequence alignment of At3g29075 cDNA with pET28a-At3g29075 insert shows 282 bp shorter.	104

Supplementary figure 19: Sequence alignment of At3g29075 cDNA with pET43b-At3g29075 insert shows 199 bp shorter.	105
Supplementary figure 20: Sequence alignment of At3g29075 cDNA with pGEX4T2-At3g29075 insert shows 149 bp shorter.	106
Supplementary figure 21: Construct of At3g29075-GFP in pEarlyGate103 gateway expression system.	107
Supplementary figure 22: Rapid selection of pEarlygate103-A3g29075-GFP overexpression line via Basta.	107
Supplementary figure 23: Sequence alignment of At3g29075 cDNA with pQLinkHD -At3g29075 full-size gene.	108
Supplementary figure 24: The Ct-At3g29075 recombinant protein purification.	109
Supplementary figure 25: Alignment of PLDrp1 and At3g29075 protein.	109
Supplementary figure 26: Amino acid composition of At3g29075 protein.....	110
Supplementary figure 27: Disorder structural prediction of At3g29075 protein by using Raptor X property.	110
Supplementary table 28: List of genes involved in Gene co-expression network of At3g29075	111
Supplementary figure 30: Protein Blast for C terminal part of the At3g29075 showed 88% identical ortholog protein <i>AXX17_At3g31880</i> from <i>Arabidopsis thaliana</i> Ecotype <i>Landsberg erecta</i>	112
Supplementary figure 31: Electrostatic surface model of At3g29075 protein.	113
Supplementary figure 32: Lysine clusters of At3g29075 protein.....	113

1. Introduction

Arabidopsis thaliana has lately become the organism of choice for a wide range of studies in plant sciences (Zachgo et al., 1996). *Arabidopsis* was initially chosen as a model genetic organism by Laibach in Europe and later studied in detail by Rédei in the United States (Meinke et al., 1998). *A. thaliana* is a member of the mustard family (*Brassicaceae*) with broad natural distribution throughout Europe, Asia, and North America (Zachgo et al., 1996) including in areas that are prone to drought at different times of the year. Many different ecotypes (accessions) have been selected from natural populations and are accessible for experimental study. The Columbia and Landsberg ecotypes are the recognised standards for genetic and molecular studies (Meinke et al., 1998). The entire life cycle, including seed germination, the formation of a rosette plant, bolting of the main stem, flowering, and maturation of the first seeds, is completed in 6 weeks; thus, it is the most prominent ideal model in plant molecular biology (Meinke et al., 1998). Another reason is *A. thaliana* is a very smaller plant. As a result, the plants can be grown in Petri plates or maintained in pots located either in a greenhouse or under fluorescent lights in the laboratory (Meinke et al., 1998; Koornneef and Meinke, 2010). *A. thaliana* genome is completely sequenced in the year of 2000. The characterisation of an unknown protein is significant for the understanding of plant growth, development, and stress-adaptation mechanisms. To raise and develop crop production, scientists researched plant stress-adaptation processes and regulatory pathways (Gustavsson et al., 2011).

Nevertheless, the regulatory networks of plant adaptation processes towards environmental stresses are complex and intersect with many different physiological pathways (Liu et al., 2019). A few signalling pathways were identified as vital for plant responses to environmental stresses. Besides, the phytohormone-mediated signalling pathways such as abscisic acid (ABA), dehydration, and salinity, the role of lipid signalling in plant adaptation are grasping scientists' attention. Phospholipases are one of the vital regulators of lipid-mediated stress pathways that involved in the accumulation of lipid signals such as phosphatidic acid (PA) and abiotic stress-induced production (Bargmann and Munnik, 2006).

1.1. The phospholipase group of enzymes

Phospholipases are a group of enzymes that exist in almost all living organisms, and they catalyse the hydrolysis of phospholipids at different cleavage sites. The phospholipase is characterised based on the position of the ester bonds they cleave (Figure 1). The plant phospholipases that are categorised into four groups; which, are phospholipase A (**PLA**), phospholipase B (**PLB**), phospholipase C (**PLC**) and phospholipase D (**PLD**) (Qin and Wang, 2002). Additionally, the phospholipase A enzyme is classified into two subgroups, phospholipase A₁ (**PLA₁**) and phospholipase A₂ (**PLA₂**), as per the position of the hydrolysis of the acyl-ester bond. However, PLA_{1,2} are discharging the membrane lipids into fatty acids. The first and second phosphodiester bonds are cleaved by PLC and PLD,

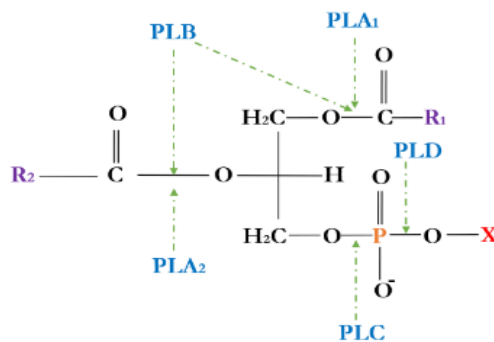


Figure 1: Schematic representation of the Phospholipase cleavage sites.

respectively. Various cofactor requirements, substrate specificity, and reaction conditions are regulating the mode of action of the phospholipases in different physiological and biological processes. Plant phospholipases have been set forth to have numerous roles in cellular regulations, lipid degradation, and membrane lipid remodelling. The products of their activity could also be involved in lipid biosynthesis (Singh et al., 2015). One of the main objectives of this study endeavour on *A. thaliana* phospholipase D and its enzymatic cleavage product phosphatidic acid (PA). The PA regulates the expression of several proteins that control physiological events in the cells of *A. thaliana* (Phospholipase D and phosphatidic acid signalling in plant response to drought and salinity. - PubMed - NCBI).

1.1.1. The structural characteristics of phospholipase D and isoforms

Phospholipase D (PLD) genes are identified in all species such as viruses, bacteria, plants and animals. In *A. thaliana*, twelve isoforms of PLDs have been identified and studied in the past years (Bargmann and Munnik, 2006). The isoforms can be classified into six major categories (α , β , γ , δ , ϵ , and ζ) based on their protein domain structure and biochemical

properties (Li et al., 2009). The PLD is associated with the hydrolysis of the phosphodiester bond of the glycerolipid phosphatidylcholine (PC) to form phosphatidic acid (PA) and choline.

PLD isoforms have two different conserved domains. The first domain HxKxxxxD_xGSxN (HKD) is found in both N and C terminal parts of the gene; as a result, HKD catalysis the hydrolysis of phospholipids. The second domain is sub-classified into two different classes based on the Calcium ion (Ca²⁺) dependency. The one depending on Ca²⁺ binding in the case of most plant PLDs have the C2 domains (Figure 2.1). On the other hand, the Ca²⁺ independent such as PLD ζ contains a PH-PX domain (Figure 2.2) in the N- terminus for PIP₂ binding x (Aloulou et al., 2012).

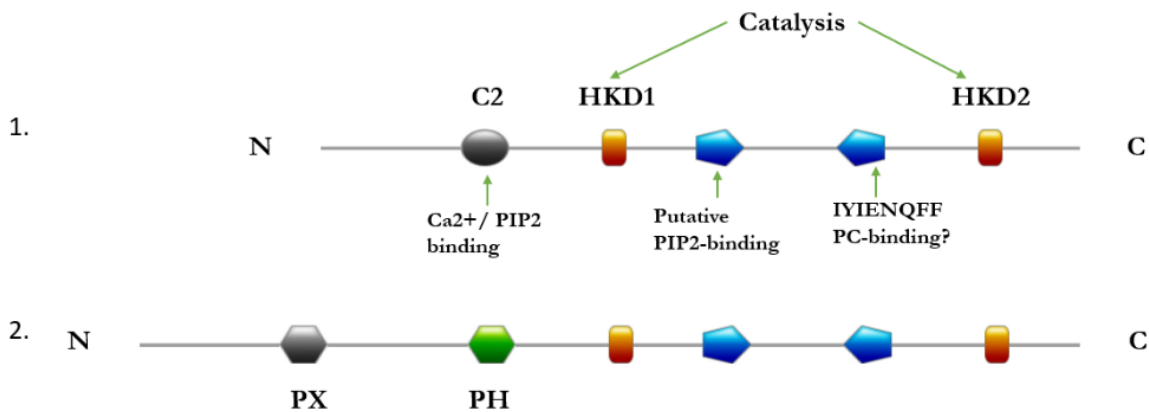


Figure 2: Schematic representation of *Arabidopsis thaliana* PLD genes (Acquired and modified from Qin & Wang, 2002).

PLD isoforms follow diverse patterns for its activation. PLD α is active in a millimolar concentration PLD isoforms are activated via different patterns. The lack of the full complement of acidic amino-acid residues in the calcium-binding C2 domain of PLD α requires the high concentration of Ca²⁺ (20 – 100 mM) for its activation. However, it can hydrolyse phosphatidylcholine (PC) mixed lipid vesicles containing phosphatidylethanolamine (PE) at moderately acidic pH (Wang et al., 2012).

In contrast, all the other PLD isoforms require a micromolar concentration of Ca²⁺ (50 – 100 μM), except PLD ζ. PLD β & γ need hydrolysing phospholipase (PPI) and PE to function on a wide spectrum of a substrate. Their C2 domain contains the full complement of calcium-binding residues, along with Lys/Arg rich PIP₂ binding motif. Even though they could be distinguished from other classes, it is not possible to separate them under *in-vitro* conditions. The rest of the PLD isoforms (Except PLD α isoform) are using N-

acylphosphatidyl ethanolamine (N-APE) to generate PA. However, N-acetyethanolamine is a potent inhibitor of PLD α (Wang et al., 2012).

PLD δ requires 0.5 mM oleic acid for full activity over PE as it prefers it to PC (Wang and Wang, 2001). Due to PLD ϵ unique biochemical and sequence properties, it was positioned in a different class even though its shared architecture with PLD α . It is active under the same conditions as PLD α , PLD δ , and PLD β & γ (Guo and Wang, 2012). Yet, it was not able to survive in a calcium-free condition of PLD ζ (Qin and Wang, 2002).

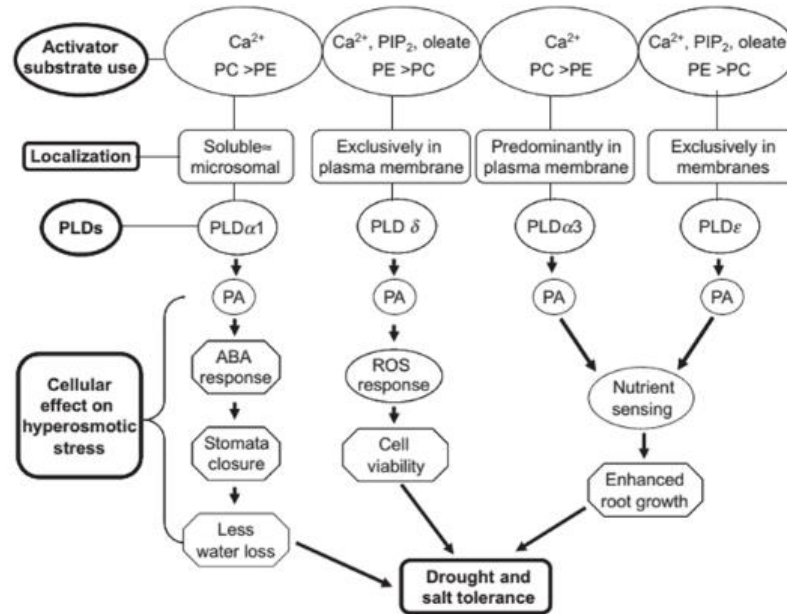


Figure 3: The role of different PLD isoforms throughout hyperosmotic stress such as salt stress and water deficiency (adapted from Hong et al., 2010).

PLD ζ is a calcium-independent class of PLDs, being closer to animal PLDs as it contains a phospho-homology (PH) pleckstrin homology (PX) domain and no C2 motif. It only requires phosphatidylinositol (PIP) for activation (Qin and Wang, 2002).

The most abundant plant PLD isoform is PLD α 1. It is triggered by ABA to produce PA (Figure 3) as *pld α 1* T-DNA insertion plants exhibited increased rates of water loss compared to the wildtype because the stomata were not able to respond to ABA. This phenotype could be compensated by applying external PA (Hong et al., 2010; Liu et al., 2010). Previous studies revealed that it has a negative role during freezing stress as it correlates with phospholipid hydrolysis during freezing and post-freezing phases (Li et al., 2004).

Attributable to high salinity and rapid dehydration stress, H₂O₂ is produced in leaves and cotyledons, which activates and increases the expression of PLD δ , thus demonstrating the role of PLD δ in the vegetative parts of the plant (Katagiri et al., 2001). It is also involved

in the pathogen response during the infection with *Pseudomonas syringae* (Johansson et al., 2014) and *Powdery mildew* (Pinosa et al., 2013).

PLD ϵ protein is localised in the plasma membrane. The PLD ϵ knock-out showed decreased root growth and biomass accumulation, whereas the PLD ϵ over-expression line illustrates the opposite effect, furthermore increased levels of PA under salt stress condition, that possibly aid the plants to uptake and utilise water more efficiently during hyperosmotic stress (Hong et al., 2009).

1.2. Phosphatidic acid

Phosphatidic acid (PA) accounts for less than 1% in-plant total proteins, yet it plays an essential role in lipid metabolism and signalling (Hong et al., 2010). The recent studies emphasis the PA as second messenger in plants; which elaborate signal mediation and responses to stress, such as pathogens (Bah et al., 2015), freezing (Zhang et al., 2013), dehydration (Hong et al., 2008), salinity (Munnik et al., 2000, Yu et al., 2010), and nutrient starvation (Yamaryo et al., 2008). PA is involved in several physiological processes like plant growth, development, and adaptation to environmental stresses (Hong et al., 2010). Through targeting and regulating a wide range of proteins, such as phosphatases, kinases,

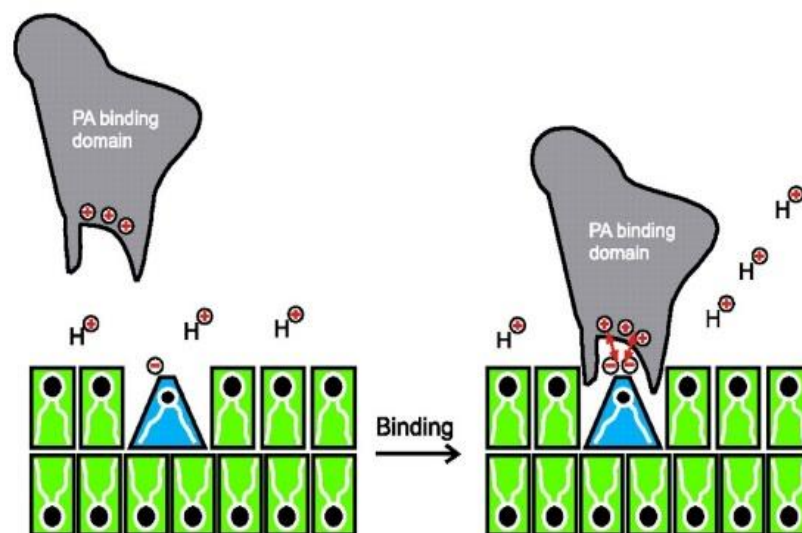


Figure 4: Schematic representation of the electrostatic/hydrogen bond switch model. The molecular shape of PA (a cone shaped lipid) is shown schematically as is the location of the head group of PA deep into the lipid head group region of a PC bilayer. The electrostatic/hydrogen bond switch is incorporated in this model (adapted from Kooijman et al. 2007).

and oxidases, PA affects membrane trafficking, cytoskeletal organisation, root hair growth and elongation, ABA, and ROS response, stomatal movement (Hong et al., 2010).

Furthermore, PA also plays an intermediate role in lipid biosynthesis (Ohlrogge and Browse, 1995). Lysophosphatidyl acyltransferases (LPA) are involved in the generation of plant PA as a precursor for structural phospho-galactolipids (Testerink and Munnik, 2011). The hydrolysis of phospholipids induces the production of PA; nevertheless, PA is also constructed via a two-stage process by the sequential action of PLC and DGK (Singh et al., 2015). The molecular species of PA produced via both pathways differ in the number of carbons and double bonds, which form the two fatty acid chains. The PLC/DGK pathway is mainly involved in the generation of PA 16:0/18:2 and 16:0/18:3 PA species. However, PLD generates other PA species like PA 18:3/18:2 and PA 18:2/18:2 (Vergnolle et al., 2016) (Guo and Wang, 2012) hypothesised that these different molecular forms of PA exhibit diverse affinities to proteins.

Even though the molecular action of PA on target proteins remains indistinguishable, yet binding of PA induces distinct downstream responses (Testerink and Munnik, 2011). The negative charge of PA's phosphate head group may contribute vital information for protein binding. Kooijman et al. (2007) described the “electrostatic/hydrogen-bond switch” model, which proposes interaction of the negatively-charged PA head group with positively charged residues (lysine or arginine) of target proteins. Construction of hydrogen bonds leads to additional deprotonation of the head group, resulting in a positive feedback of further electrostatic and hydrogen bonds between the PA and the target protein (Figure 4). The model explains the absence of a distinct PA-binding domain and denotes to the necessities of domains comprising elementary amino acids like lysine and arginine as potential interacting sides.

However, the model cannot clarify the reason behind diverse affinities shared by various molecular forms of PA towards proteins and diverge in their effect on protein targets (Guo and Wang, 2012). PA-profiling researches distinguished eight molecular PA classes in *A. thaliana* (Wang et al., 2006). This molecular diversity might explain the distinct regulations of multiple target proteins and the involvement of PA in numerous physiological processes.

A second factor for PA's role in many regulatory processes (Hong et al., 2008) might be its cooperation with other cellular signals (Testerink & Munnik, 2011). For example, in response to drought stress, PA induces the formation of nitric oxide (NO) and reactive oxygen species (ROS), which are involved in the ABA-mediated response of stomata

closure (Prelich, 2012). Regulation of PLD activity by intracellular Ca^{2+} concentrations controls PA production upstream of target proteins (Hong et al., 2009); also, the catalyst is required for PA binding networks in some instances (Domínguez-González et al., 2007). A novel PA biosensor was presented by Potocký et al. (2014) to monitor PA dynamics in plant cells, but its inability to discriminate between molecular PA species demands for additional methods and techniques to unravel distinct PA-binding protein networks.

1.3. Proteins under the control of phospholipase D

Arabidopsis-Genome initiative managed to sequence the genome of the plant in 2000. Out of 27,000 genes and the 35,000 proteins they encode, most of their functions and characteristics are not assigned (EMBL). Therefore, several proteins have been illustrated in protein databases as “unknown proteins” since neither function nor additional information is known, even though the protein has been detected. One of those proteins is At5g39570.1 (gene: *At5g39570.1*). The gene encoding for this 40 kDa protein consists of two exons that are separated by a 660 bp long intron towards the 3'-end. However, Ufer (2017) termed this unknown protein as PLD regulated protein1 (PLDrp1).

The total proteins and phosphoproteins of wildtype and double knock-outs (dko) *pld α 1/pld δ* using 2D gel analysed by Anke Kuhn (2009). There was one missing protein spot observed in *pld α 1/pld δ* - double knock-out, which was assigned to the absence of PLD α 1. MALDI-TOF MS predicted the protein as PLDrp1, a protein of unknown function. It is a phosphoprotein, as shown from 2D gel analysis- with 381 aa. There are ten imperfect tandem repeats cited in the C-terminal part of the protein, which is abundant in acidic amino acids that provide a disorganised protein structure. Putative nucleus localization signal found in the C- terminal end of the protein; however, GFP localization analysis identified the protein to be located in cytoplasm and nucleus (Ufer, 2015).

Further studies illustrated that PLD α 1 regulates the expression of PLDrp1. Comparison between the total RNA and polysomal RNA levels of the PLDrp1 gene in the WT and *pld α 1* mutant showed that the total RNA level in both plants are highly similar, while on the polysomal level there is a drastic decrease in the *pld α 1*-mutant plant, suggesting a posttranscriptional modification controlling the amount of protein produced according to the plant developmental state. Likewise, the PLDrp1 protein is up-regulated in the *pld α 1*-

mutant during drought and salt stress. Lipid binding assay exempted that PLDrp1 protein binds to PA (Ufer et al., 2017).

1.4. The unknown protein At3g29075 from *A. thaliana*

Using a BLASTn algorithm to search for similar sequences for PLDrp1 in the NCBI databases, the At3g29075 gene, which contains 1206 bp, was found. It is translated into a glycine-rich protein, and it shares a 49% similarity with the N-terminal amino acids of PLDrp1 protein (Supplementary figure 25), while the C-terminal end of both genes is unique. The identification of the At3g29075 and its putative redundant function led to the necessary of characterisation of At3g29075 protein and the generation of knockdown mutant lines. Work on At3g29075.1 led to a new side project that was followed by a Master student (Nasr, 2015).

At3g29075 protein is documented as one of the unknown proteins in the Tair and NCBI database. The gene encoding for this 34.5 kDa protein consists of two exons, which are separated by 100 bp intron towards 3'-end (Figure 5).

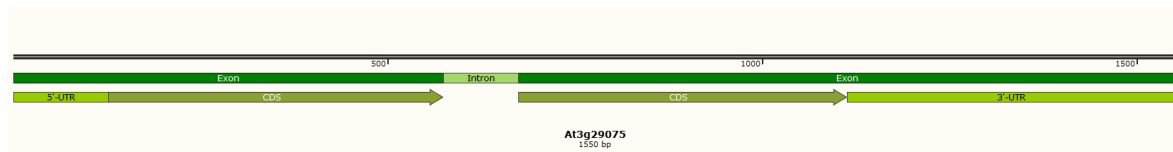


Figure 5: Gene model of *At3g29075*.

1.4.1. The protein-coding sequence of At3g29075

The protein At3g29075 also shows similar protein features of PLDrp1, such as tandem repeats, an acidic pI, and an unusual amino acid composition (Supplementary figure 26). Significantly high contents of charged amino acids such as aspartic acid 19.3%, lysine 21.1% were found in the At3g29075 protein.

Even though lysine is the second most limiting amino acid, it has a primary function of protein deposition in the animal body (Watson and Foglesong, 2007). Lysine aids the absorption of calcium; maintain healthy blood vessels; produce antibodies, enzymes, collagen; and repairs of tissues; it also synthesis carnitine which supports to convert fatty acids into energy and maintain cholesterol in the blood (Ovie and Eze, 2013). The majority of the protein contains 7.2% of lysine; therefore, Lysine is one of the essential amino acids of protein (Miles and Chapman, 2008). Lysines rich-ingredients are the most expensive due to their limiting resources used for feed (Ovie and Eze, 2013). The expert committee of

the world health organisation recently recommended daily uptake of up to 30 mg lysine per kilogram of body weight for humans (Joint, 2007).

By consuming common food proteins with sufficient lysine concentrations, most of the adults fulfil their lysine requirement, infants are more susceptible to lysine-deficiency when fed a diet of proprietary formula (Young and Pellett, 1994). As lysine is much lower concentration in most plant-food proteins than in animals and represents the limiting amino acid in most cereal grains and rice (Young and Pellett, 1994), it is likely people with insufficient or unbalanced diets and some vegans who do not have additional lysine sources like pulses, have the need for additional lysine up-take. The deficiency of lysine has also been reported for professional athletes (Young and Pellett, 1994). The animal's feed is supplemented with industrially produced lysine to overcome the lower lysine concentration in cereals (Izumi et al., 1978).

However, several lysine proteins were studied, and recombinant high lysine-rich proteins were synthesised in cereals to artificially increase the lysine content (Jia et al., 2013; Hejgaard and Boisen, 1980, 2009). Recombinant high lysine-rich proteins contained up to 17% lysine in stable-transformed barley (Hejgaard and Boisen, 2009). In contrast, the native At3g29075 protein from *A. thaliana* obtains 21.1% lysine. The At3g29075 protein contains a very high amount of the essential amino acid lysine; therefore, it is considered as a new protein with high value to study further.

1.5. The objective of the study

Previous proteomic studies of PLDrp1 (Ufer et al., 2017) confirmed the PA binding, which works as a secondary messenger in plant stress-signalling. However, PLDrp1 is a homolog of At3g29075 protein, which makes it an exciting candidate for the study of phospholipase-controlled pathways. This work aims at a detailed investigation of At3g29075 based on the results of the Doctoral thesis of Ufer (2015) and Master's thesis of Nasr (2015).

The characterisation of the unknown At3g29075 protein is the main objective of this study. The study was divided into four different projects.

First, a general characterisation of At3g29075 was pursued via numerous bioinformatics prediction tools, and text-mining algorithms were used to illustrate the first image of this enigmatic protein. The extraordinary, unequal composition of the protein's N- and C-

terminal regions were emphasised *in silico* analysis. Detection of At3g29075 protein with Ct-At3g29075 polyclonal antibody.

Second, the expression of At3g29075 protein during the different developmental stages and also in different tissues of *A. thaliana* was studied. The protein expression pattern upon various environmental stresses was examined due to their role in inducing and mediating signals in response to abiotic stresses.

Third, a preparatory point for the molecular functional characterisation of the unknown protein should be designated. At3g29075 consider as one of the candidate proteins downstream of the PLD-regulated signalling network; to gain further insights into the complex regulatory networks of phospholipases, specifically, their role in PA binding was studied via lipid binding and liposome binding assay.

Finally, a starting point for the identification of the At3g29075 protein's biological process, initial hints for the protein's functions were gathered via *in silico* data analysis and text-mining approaches. Later on, AT3G29075 overexpression lines were generated; the mutant plants were phenotypically monitored with wildtype plants. The cellular localisation of the protein was studied through the transient transformation method.

2. Materials and methods

2.1. Materials

2.1.1. Chemicals

The ensuing list shows an impression of developers, producers, and suppliers from chemicals, equipment, databases, and programs used in this thesis. The next section states the chemicals listed below and will only offer the manufacturer's name.

- AppliChem GmbH (Darmstadt, Germany)
- Apollo Scientific (Ltd Bledsbury, Czech Republic)
- Avanti Polar Lipids (Alabaster, USA)
- Biomol (Hamburg, Germany)
- Bio-Rad (Munich, Germany)
- Carl Roth GmbH (Karlsruhe, Germany)
- Dushefa Biochemie B.V. (Haarlem, Netherlands)
- Fermentas (St. Leon–Rot, Germany)
- GE Healthcare (Freiburg, Germany)
- Invitrogen (Karlsruhe, Germany)
- Labomedic (Bonn, Germany)
- LMS Consult (Brigachtal, Germany)
- Merck AG (Darmstadt, Germany)
- Sigma-Aldrich Chemie GmbH (Munich, Germany)
- Stratagene (Heidelberg, Germany)
- Thermo Scientific (Waltham, MA USA)
- ZVE (Bonn, Germany)

2.1.2. Equipment

- Azure biosystem c300: Chemiluminescent Western Blot Imaging System (Dublin, USA)
- Binocular microscope: SMZ-800 (Nikon, Düsseldorf, Germany)
- Blotting chamber for proteins: “Criterion Blotter” (Biorad, Munich, Germany)
- Chemiluminescence detector: Intelligent Dark Box II, (Fujifilm, Tokyo, Japan)
- Confocal Laser Scanning Microscope: ZE2000 (Nikon, Düsseldorf, Germany)
- Consumables: Pipette tips and centrifugal tubes (Sarstedt AG, Nümbrecht, Germany)
- Desalting columns: “PD–10” (GE Healthcare, Freiburg, Germany)

- Electroporation system GenepulserII Electroporator (Bio-Rad, Hercules, USA)
- Lyophilisator: “LDC–2” (Christ, Osterode am Harz, Germany)
- Gel electrophoresis chambers:
 - “Minigel” (Biometra, Göttingen, Germany)
 - “EasyCast” (Owl, Portsmouth, USA)
- Luminescent Image Analyzer LAS 1000 (Fujifilm Life Science, Stamford, USA)
- Nanodrop: Biospec-Nano (Shimadzu Biotech, Japan)
- PCR–cycler: “T3 Thermocycler” (Biometra, Göttingen, Germany)
- pH–meter (SCHOTT GLAS, Mainz, Germany)
- Rotator: “neoLab-Rotator 2–1175” (neoLab, Heidelberg, Germany)
- Spectrophotometer: “SmartSpec 3000” (Biorad, Hercules, USA)
- Scanner:
 - Typhoon 9200 (Amersham, Piscataway, USA)
 - Image scanner (Amersham, Buckinghamshire, Great Britain)
- Sonification water bath: “Sonorex Super RK102P” (Bandelin electronics, Berlin, Germany)
- Thermocycler: “Eppendorf Mastercycler gradient” (Wesseling-Berzdorf, Germany)
- Centrifuges:
 - Centrifuges: “5415D”; “5417R”, “5810R”; Vacuum centrifuge: “Concentrator 5301” (Eppendorf, Hamburg, Germany)
 - Sorvall centrifuge: “RC50” (DuPont, Hamm-Uentrop, Germany)
 - Ultracentrifuge: “L8-70M” (Beckman Coulter, Brea, USA)

2.1.3. Computer programs and databases

Various bioinformatics tools were used to predict the characteristics of genes and proteins. For, *in silico* analyses, the “ExPASy Bioinformatics Resource Porta” (<http://expasy.org/tools/>) provides a basic set of prediction and analysis tools. The additional computer programs and databases used for detailed *in silico* analysis in this thesis are the following.

2.1.3.1. Computer programs and tools

- APE-A Plasmid Editor v. 1.7
- Blastp-protein blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)
- ClustalW2 (www.ebi.ac.uk/clustalw/)
- cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_y.cgi)
- Cell eFP browser (http://bar.utoronto.ca/cell_efp)
- DisEMBL (<http://dis.embl.de/>)
- GeneMANIA (<http://www.genemania.org/>)
- GlobPlot (<http://globplot.embl.de/>)

- Interpro (<http://www.ebi.ac.uk/interpro/>)
- Lig-Input: http://www.insilico.uni-duesseldorf.de/Lig_Input.html
- Medor v. 1.4 (<http://www.vazymolo.org/MeDor/>)
- Mega X-Molecular evolutionary genetics analysis
- Microsoft Office 2016 (Microsoft, Redmond, USA)
- NetPhos 3.0 (<http://www.cbs.dtu.dk/services/NetPhos/>)
- NetPhosK (<http://www.cbs.dtu.dk/services/NetPhosK/>)
- Kinase Phos 2.0 (<http://kinasephos2.mbc.nctu.edu.tw/>)
- NEBcutter2 (<http://tools.neb.com/NEBcutter2>)
- NucPred (<http://www.sbc.su.se/~maccallr/nucpred/>)
- Photoshop CS and Paint.net
- Primer3 (<http://frodo.wi.mit.edu/primer3/>)
- Protscale (<http://web.expasy.org/cgi-bin/protscale/protscale.pl?1>)
- PONDR-FIT (<http://www.disprot.org/pondr-fit.php>)
- Reverse Complement (www.bioinformatics.org)
- SAM_T08 (http://compbio.soe.ucsc.edu/SAM_T08/)
- SNAPgene
- Subnuclear2 (<http://array.bioengr.uic.edu/cgi-bin/subnuclear/subnuclear2.pl>)
- TRANSFAC (<http://www.gene-regulation.com/pub/databases.html>)
- TRUST “Tracking Repeats Using Significance and Transitivity” (<http://www.ibi.vu.nl/programs/trustwww/>)
- TSSP (<http://www.softberry.com/berry.phtml>)
- Vector NTI (Invitrogen, USA)

2.1.3.2. Databases

- Arabidopsis 1001 genome project (<http://1001genomes.org/>)
- BAR-Database, Arabidopsis eFP Browser (<http://www.bar.utoronto.ca>)
- European Molecular Biology Laboratory (<http://www.embl.de>)
- German plant genomics research program (<http://www.gabi-kat.de>)
- IuPred (<http://iupred.enzim.hu/>)
- Nottingham Arabidopsis Stock Centre (<http://arabidopsis.info>)
- National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>)
- Salk-Institute (<http://www.salk.edu>)
- STRING (<http://www.string-db.org>)
- TAIR (www.arabidopsis.org)
- T-DNA Express (<http://signal.salk.edu/cgi-bin/tdnaexpress>)
- Toxinpred (http://webs.iitd.edu.in/raghava/toxinpred/prot_submitfreq_disp.php?ran=58106)

2.1.4. Enzymes and markers

- UniProt (<http://www.uniprot.org/>)
- DNA-marker (New England Biolabs, Ipswich, USA)
- Restriction enzymes and Pfu-DNA-polymerase (Fermentas, St. Leon–Rot, Germany)
- Taq DNA–polymerase (isolated via F G Pluthero method by Tobias Dieckmann)

2.1.5. Primer

The primer pairs were designed via the primer3 program, a primer with a guanine/cytosine (G/C) content of 45-55% was considered most stable with a higher ratio of GC in terminal nucleotides. Primer pairs were selected based on similar melting temperatures (TM) in the range of 55°C-65°C and low self-complementarity of the sequences. For mutagenesis primers, point mutations avoided when possible. Primers were obtained from Eurofins Genomics. All primers were stored at -20°C and 100 mM concentrations. The following list provides all the primers used in this thesis.

Table 1: List of primers

Name	Sequence (5' → 3')	Restriction Site	Source of clone
Mutagenesis Primers:			
pET28a Clone			
at3gfulLP	CCAAAAACCATGGCGTATTACACCAACG	<i>NcoI</i>	cDNA
at3grp2	GCTTTTTTACTCGAGTCCCTTGTGGTGC	<i>XhoI</i>	
At3gfw37	GAACGATAGCCATGGCGTATTACACC	<i>NcoI</i>	pGJ280-At3g29075
At3grw37	CTTTACCCTCGAGGTGCTTATCCTTC	<i>XhoI</i>	
pET43b.1 Clone			
at3gfw38	CATATGGCGTATTACACCAA	<i>NdeI</i>	pGJ280-At3g29075
at3grp2	GCTTTTTTACTCGAGTCCCTTGTGGTGC	<i>XhoI</i>	
pGEX-4T2 Clone			
at3gfw39	GAACGAATTCATGGCGTATTACACC	<i>EcoRI</i>	pGJ280-At3g29075
at3grp2	GCTTTTTTACTCGAGTCCCTTGTGGTGC	<i>XhoI</i>	
at3gggstfw39a	GAACTCGAGCCATGGCGTATTACACC	<i>XhoI</i>	pGJ280-At3g29075
at3grp2	GCTTTTTTACTCGAGTCCCTTGTGGTGC	<i>XhoI</i>	

pQLinkHD Clone

adaptive primers

attB1-adapter GGGGACAAGTTTGTACAAAAAAGCAGGCT

attB2-adapter GGGGACCACTTTGTACAAGAAAGCTGGGT

Gene specific primer for full size clone

At3g_attB2 AGAAAGCTGGGTATTAGTCATGTCCCTTGTCTGTGCT cDNA

At3g_attB1 AAAAAGCAGGCTTAACCATGCCGTATTACACCAAC

N-terminal part Clone

At3gN_attB2 AGAAAGCTGGGTATTACTCAATTCTTCCACCATA cDNA

At3g_attB1 AAAAAGCAGGCTTAACCATGCCGTATTACACCAAC

pTWIN1 Clone

at3g_Ptwin1_F GGTGGTCCATGGCGTATTACACCAACGACGAC pGJ280-
At3g29075

at3g_Ptwin1_R GGTGGTCTGCAGTTATTAGTCATGTCCCTTGTGGTGCTTA

Primer for the generation of At3g29075.1 overexpression line (pH2GW7)

OVERAT3GFW GAACGATAGCCATGGCGTATTACACC cDNA

OVERAT3GRW TCCCTTTACGGATCCGGTGCTTATCCT

Vector specific primer:

pEarlyGate103 primers

pGate103fw T*TCGCAAGACCCT*TCCTCTA

pgj280 rev TGTGCCCATTAACATCACCA

pJET primers

pJET1.2fwd CGACTCACTATAGGGAGAGCGGC

pJET 1.2 rev AAGAACATCGAT*TTCCATGGCAG

pDONR201/207 primers

SeqLA TCGCGTTAACGCTAGCATGGATCTC

SeqLB GTAACATCAGAGAT*TTGAGACAC

pH2GW7 primers

PH2gw7fw GCGGCCGCACTAGTGATA

PH2gw7rw ACTGGTGAT*TTT*GCGGACT

pQLinkHD primers

newpQEFW CTCGAGCTTAATTAACAACACCAT

pQE276 GGCAACCGAGCGTTCTGAAC

T₇-Promoter TAATACGACTCACTATAGGG

T₇-Terminator GCTAGTTATTGCTCAGCGG

Actin-and poly(A)-primers

ATHACTIN2_FWD ATGGCTGAGGCTGATGATATTCAAC

ATH-ACTIN2_REV GAAACATTTTCTGTGAACGATTCCT

Oligo(dT)18 Primer TTTTTTTTTTTTTTTTTT

2.1.6. Vectors

1. pJET1.2/blunt (Fermentas):

This plasmid is used for blunt-end cloning of PCR-fragments.

2. pET28a (Novagen):

This plasmid encodes for N/C-terminal histidine-tags (His-tag) and possesses an inducible promoter. The vector is used for the expression of His-tagged proteins.

3. pET43b:

The pET-43. b vector designed for cloning and high-level expression of peptide sequences fused with C-term His tag protein.

4. pGEX-4T2:

The pGEX-4T2 vector encodes the GST -tag, and it is used to express the proteins.

5. pGJ280:

This plasmid contains a dual CaMV35S promoter and encodes for the Green Fluorescent Protein (GFP). It is used for over-expression and localization studies. The vector constructed by Dr. G. Jach (Max-Planck-Institute, Cologne, Germany).

6. pDONR201/207 (Invitrogen):

The gateway donor vector used for plasmid blunt-end cloning.

7. pQLinkHD (addgene):

This gateway cassette plasmid encodes for N/C – terminal histidine -tags (His-tag) and contains an inducible promoter.

8. pEarleyGate103:

Gateway compatible plant transformation vector with GFP and 6x his C-terminal tags was used to overexpress and localisation of the protein in plants.

9. pH2GW7:

Gateway vector for Agrobacterium-mediated plant transformation to produce overexpression lines.

10. pTWIN1:

pTWIN1 is an *E. coli* expression vector designed for protein purification with an N-terminal cysteine and/or a C-terminal thioester. The presence of the chitin-binding domain from *Bacillus circulans* facilitates purification.

All plasmid vector maps can be found in the supplementary data.

2.1.7. Kits

- GeneJET Plasmid Miniprep Kit, Fermentas (St. Leon–Rot, Germany)
- CloneJET PCR Cloning Kit, Fermentas (St. Leon–Rot, Germany)
- NucleoSpin Extract II, Macherey-Nagel (Düren, Germany)
- RevertAid First Strand cDNA Synthesis Kit. Fermentas (St. Leon–Rot, Germany)
- BP-clonase, LP-clonase kit (Thermofisher, USA)

Kits were used according to the manufacturer's instructions.

2.1.8. DNA-sequencing

DNA-sequencing of plasmid DNA (100 ng/ μ L) and PCR–DNA (50 ng/ μ L) was performed by GATC (Germany) and Eurofins Genomics (Elbersberg, Germany), respectively.

2.1.9. Quantification of proteins and RNA

Quantification of protein and RNA signals were calculated as described by Dr. Daniel Kraus with the Image J program (http://home.arcor.de/d-kraus/lab/ImageJ_Western_blot.html). Initially, the Bands of interest were selected, marked, and the signal area was quantified automatically using the tracing tools. The same was repeated for the corresponding housekeeping gene (EF1 α - Elongation factor 1-alpha) or protein (RubisCO -Ribulose-1,5-bisphosphate carboxylase oxygenase), respectively. The data transferred to Microsoft Excel, and relative values (%) were calculated based on the reference value (wild-type value set to 100%). The quotient of gene/protein of interest and its housekeeping gene/protein was determined, and the median of three repetitions and associated standard deviations calculated.

2.2. Plant material

The mainstream of the study was done based on the *A. thaliana* (Ecotype Columbia-0). However, the following transgenic lines also used in this thesis,

- at3g29075 knock-down (Ufer, 2015)
- AT3G29075::35S Overexpression line (See section 3.14)

Two different sterilisation methods prevented the contamination of seedlings grown on MS-plates.

2.2.1. Sterilisation of seeds

2.2.1.1. Sterilisation with sodium hypochlorite

Sterilisation solution

- 12 (w/v) Sodium hypochlorite (NaOCl)
- 0.1% (w/v) Sodium dodecylsulfate (SDS)

Seeds were sterilised for two minutes with 70% (v/v) ethanol and subsequently incubated in sterilisation solution for 25 min. Repetitive washing steps in d.H₂O cleaned seeds.

2.2.1.2. Sterilisation with ethanol

- 70% (v/v) EtOH
- **Sterilization solution:** 0.01% (w/v) Polyoxyethylen (20)–sorbitan–monolaurate (TWEEN)

Seeds were sterilised for one minute with 100% EtOH and subsequently incubated in sterilisation solution for 15 min. Repetitive washing steps in d.H₂O cleaned seeds.

2.3. Growth conditions

Sterilised seeds were grown on MS-agar plates supplemented with the appropriate antibiotic.

2.3.1. Growing on soil

Seeds sowed on wet, Lizetan® (Bayer, Leverkusen, Germany)-treated soil and vernalized for two days at 4°C. Seedlings were grown under short-day conditions: Photoperiodic cycle of eight hours of light at 22°C and 16 hours of darkness at 20°C. After two weeks of germination, up to 4 seedlings were transferred into a separate pot. Plants were harvested after 2-6 weeks, depending on the experimental set-up and stored at -70°C. For the production of seeds, six-week-old plants were grown under long-day conditions (16 hours of light at 22°C and 8 hours of darkness at 20°C).

2.3.2. Growing on MS-plates

Vitamin solution

2 mg/L Glycine

0.5 mg/L Niacin

MS–Medium (Murashige & Skoog, 1962)

4.6 g/L MS–salts

20 g/L Saccharose

0.5 mg/L Pyridoxine–HCl	1 mL/L Vitamin solution
0.1 mg/L Thiamine	pH 5.8 (KOH)
in d.H ₂ O sterile filtrated	8 g/L Select Agar
	Media autoclaved for 20 min (121°C, 1.2 bar)

Seeds were sterilised and stratified (2 d, 4°C) earlier to sowing and growing on MS-media. Single seeds applied to MS-media under sterile conditions. After two weeks, seedlings were transplanted on the soil.

2.3.3. Germination on Blotting paper

Seeds of the *A. thaliana* (Col-0) were used for all assays. The timing of testa and endosperm rupture and seedling greening of fully after-ripened seeds was determined as described (Joosen et al., 2010). To sum up, two layers of blotter paper (Macherey-Nagel GmbH & Co.KG, Düren, Germany) were equilibrated with 48 mL of demineralised water in a Petri dish. For germination assays, 50–150 seeds were spread on wetted papers in the Petri dish. The experiment was carried out in a 22°C incubator under a short-day room. Dry seeds were imbibed in three independent biological replicates. Seeds and seedlings were harvested at each physiological state during the seed to seedling transition and frozen in liquid nitrogen. The material was stored at -80°C until further analysed.

2.4. Stress conditions

Before stress treatments, soil-grown seedlings were transferred in pairs of four into separate pots. The pots were placed in Petri dishes and watered equally with 50 mL d.H₂O per week. For osmotic stress treatments, the water replaced by the appropriate solutions of sodium chloride (NaCl) (concentrations from 100–300 mM). Abscisic-acid (ABA) (Final conc. = 100 µM). Drought treatment was performed by withholding the water until the desired relative water content reached. The relative water content (RWC) of plants was calculated as described below:

$$\text{RWC (\%)} = [(W-DW) / (TW-DW)] \times 100$$

RWC: Relative water content (%). W (Fresh weight); TW (Turgescent weight): Weight after the rehydration of leaves for 24 h in H₂O. DW (drought weight): Drought weight of leaves after 24 h at 80°C.

2.5. Microorganisms

2.5.1. Bacterial strains

- ***Escherichia coli* DH10B (Lorow & Jessee, 1990):**

Genotype: μ F⁻mrcA Δ (mrr-hsdRMS-mcrBC) ϕ 80d lacZ Δ ϵ 15 Δ lacX74 endA1 recA1 deoR Δ (ara. leu) 76 λ 7 araDD1 γ galU galK nup6 rps δ \square ⁻

This *E. coli* strain was used for the cloning process.

- ***Escherichia coli* BL21 (DE3) (Pharmacia, Freiburg, Germany):**

Genotype: F⁻. ompT. hsdS (r⁻B. m⁻B). gal. dcm. / \square DE3 (lacI. lacUV5-T7 gene 1. ind1. sam7. nin5).

This *E. coli* strain was used for the over-expression of proteins.

- ***Escherichia coli* C43 (derived from DE3):**

Genotype: F⁻ ompT hsdSB (rB - mB) gal dcm (DE3)

C43(DE3) strains are superior to the parental BL21(DE3) in the transformation and expression of toxic proteins.

- **T7 Express *E. coli*:**

Genotype: fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ (mcrC-mrr)114::IS10

Chemically competent *E. coli* cells suitable for high efficiency transformation and protein expression.

- ***E. coli* RIPL strains BL21-CodonPlus (DE3):**

Genotype: *E. coli* B F⁻ ompT hsdS(rB - mB -) dcm+ Tetr gal λ (DE3) endA Hte [argU proL Camr] [argU ileY leuW Strep/Specr]

Availability of tRNAs in RIPL allows high-level expression of many heterologous recombinant genes that are poorly expressed in conventional BL21 strains.

2.5.2. Media for microorganisms

SOC-media: 2% (w/v) Tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 10 mM MgSO₄, 10 mM MgCl₂.

LB-media: 1 g/L Tryptone, 10 g/L NaCl, 5 g/L yeast extract, pH 7.0.

LB-agar: 15 g/L Select–Agar was added to LB-media.

YEB-media: 5 g Beef extract, 5 g peptone, 5 g sucrose, 1 g yeast extract, pH 7.0. After autoclaving filter-sterilized MgCl₂ solution (final concentration 2 mM) was added.

2.5.3. Media supplements

- Ampicillin stock solution: 100 mg/mL in d.H₂O. Dilution: 1:1000
- Kanamycin stock solution: 50 mg/mL in d.H₂O. Dilution: 1:1000
- Gentamycin stock solution: 25 mg/mL in d.H₂O. Dilution 1:1000
- Spectinomycin stock solution: 50 mg/mL in d.H₂O. Dilution 1:1000
- Rifampicin stock solution: 50 mg/mL in DMSO (dimethyl sulfoxide). Dilution: 1:500
- Hygromycin B: 15 µg /mL
- BASTA (glufosinate ammonium): 50 µM/mL

2.5.4. Generation of rubidium chloride–competent cells

The generation of rubidium chloride-competent cells was generated as described by Stiti *et al.*, (2007) with few modifications. A pre-culture (3 mL) of *E. coli* was inoculated and incubated overnight at 37°C on a shaker (200 rpm). The next day, the primary culture (100 mL) was inoculated and grown to an OD₆₀₀ of 0.35–0.45. Bacterial cells were centrifuged (4000 g, 10 min, 4°C) and resuspended in 15 mL ice-cold TFBI solution, incubated on ice (10 min) and again centrifuged (4000 g, 10 min, 4°C). The pellet was resuspended in 25 mL ice-cold TFBI solution, incubated on ice (5 min) and re-centrifuged. In the end, the pellet was resuspended in 2 mL TFBII and stored at -80°C.

TFBI	TFBII
30 mM Potassium acetate	10 mM MOPS
100 mM Rubidium chloride	75 mM Calcium chloride
10 mM Calcium chloride	10 mM Rubidium chloride
50 mM Manganese chloride	15% (v/v) Glycerol
15% (v/v) Glycerol	pH 6.5 (KOH)
pH 5.8 (Acetic acid)	

2.5.5. Generation of electro-competent *A. tumefaciens*

A pre-culture 5 mL of *A. tumefaciens* grown in YEB medium supplemented with rifampicin overnight at 28°C (250 rpm). The next day a subculture of 50 mL YEB medium with rifampicin was inoculated and grown to an OD600 of 0.5 and subsequently placed on ice for 30 min. Cells harvested by centrifugation (10 min, 5000 rpm, 4°C), and the recovered pellet was resuspended in 25 mL ice-cold sterile water. The cells were centrifuged (10 min, 5000 rpm, 4°C) and washed as follows:

- 1.25 mL 1 mM Hepes pH 7.5
- 12.5 mL 1 mM Hepes pH 7.5
- 10 mL 10% (v/v) Glycerol, 1 mM Hepes pH 7.5
- 5 mL 10% (v/v) Glycerol, 1 mM Hepes pH 7.5
- 2 mL 10% (v/v) Glycerol

2.6. Glycerol Stocks

The cells were finally resuspended in 1 mL 10% (v/v) glycerol, frozen in liquid nitrogen, and stored at -80 °C.

2.7. Cloning methods

A single colony of the desired clone was transferred to appropriate media, supplemented with the appendant antibiotic and incubated overnight. Equal volumes of the culture and sterile glycerol were mixed in a cryotube, frozen in liquid nitrogen, and stored at -70°C.

2.7.1 Electrophoresis of nucleic acids (Adkins & Burmeister, 1996)

50 x TAE–Buffer	10 x Loading buffer
2 M Tris	2.5 mg/mL Bromphenol blue
50 mM EDTA	2.5 mg/mL Xylenxanol
pH 8.0 (Acetic acid)	30% (v/v) Glycerol
	2% (v/v) 50 x TAE–buffer

Electrophoresis was used to separate the nucleic acids on agarose gels with concentrations ranging from 0.8% to 1.5% (w/v) based on their molecular size. Agarose was added to 1x TAE buffer, boiled, and after cooling down supplemented with ethidium bromide (final concentration 10 µg/mL). A 1 kb GeneRuler used as a molecular marker. Electrophoresis performed in 1x TAE–buffer at 90 V and gels were analysed under UV-light.

2.7.2. Isolation and purification of plasmid DNA (Sambrook et al., 1989)

P1–buffer	P2–buffer	P3–buffer
50 mM Tris	200 mM NaOH	3 M Potassium acetate
11 mM EDTA	1% (w/v) SDS	pH 5.5 (Acetic acid)
RNase A (100 µg/µL)		
pH 8.0 (HCl)		

Isolation of plasmid DNA was done as described by Sambrook *et al.* (1989) in a modified version. The overnight culture from a single clone was incubated (200 rpm, 37°C) in 2 mL LB-media and supplemented with antibiotics. The overnight culture centrifuged (16,000 g), and the resulting pellet was resuspended in 100 µL buffer P1. After the addition of 100 µL buffer P2, the solution was mixed carefully and supplemented with 150 µL P3-buffer. The mixture centrifuged (16,000 g, 10 min, RT) and 700 µL of the supernatant was transferred to an equal volume of phenol-chloroform (1:1 v/v) and centrifuged. Subsequently, the upper phase was precipitated with equal amounts of isopropanol on ice for 15 min. Plasmid DNA retained by centrifugation (16000 g, 30 min, RT), washed with 70% (v/v) EtOH, and stored at -20°C. For the extraction of high amounts of pure plasmid,

the “Plasmid extraction Kit” (Fermentas) used according to the manufacturer’s instructions.

2.7.3. Purification of DNA

The DNA samples (100 μ L) were purified by adding an equal volume of phenol and shortly vortexed and centrifuged. The same quantity of chloroform added to the supernatant, and the mixture was supplemented and mixed with two volumes ice-cold 90% (v/v) ethanol, 0.1 volumes 3 M sodium acetate (pH 4.5) and 20-40 μ g glycogen. After centrifugation (20 min, 16,000 g, 4°C), the precipitated DNA was washed twice with 70% (v/v) ethanol, air-dried, and finally dissolved in 20 μ L sterile d.H₂O. Extraction and purification of plasmid DNA from agarose gels done with the NucleoSpin Extract II Kit (Macherey-Nagel) according to the manufacturer’s instructions.

2.7.4. Restriction digestion

Restriction digests were carried out as described by Sambrook *et al.* (1989) at 37°C for 1-3 h or overnight. Alkaline phosphatases were used for dephosphorylation of single digested plasmids to avoid spontaneous re-ligations. Inactivation of the phosphatases was achieved by heating the samples for 15 min at 65°C.

The reaction mixture of Restriction and Digestion:

10 x buffer	2 μ L
DNA / plasmid 100 ng	1 μ g
Restriction enzymes	Each 5 units
d.H ₂ O	Rest
Total:	20 μ L

2.7.5. Ligation (Sambrook et al., 1989)

Before cloning DNA amplicons into plasmids, a restriction digest was carried out. The molar insert concentration was at least 3-fold higher than the plasmids. The ligation calculator Lig Input (University of Düsseldorf), was used for the calculation of molar ratios. Digested plasmids and inserts were ligated at 23°C for 3h (or at 16°C overnight) by

the T4-DNA-ligase. An additional ligase reaction without the insert was used as a negative control. The ligation mix was subsequently used for transformation or stored at -20°C.

A typical ligase reaction:

10 x ligase buffer	1 μ L
T4-DNA-ligase	1 μ L
Plasmid	x ng/ μ L
Insert	3*x ng/ μ L
d.H ₂ O	4 μ L
Total	10 μ L

2.7.6. Transformation of rubidium chloride-competent *E. coli* (adapted from Hanahan, 1983)

Rubidium chloride-competent cells (50 μ L) were supplemented with 2 μ L ligation mix (1 pg – 100 ng) and incubated on ice for 30 min. Bacterial cells were transformed at 42°C for the 30s and consequently cooled on ice for 5 min. After the addition of 950 μ L SOC-medium (sterile filtrated), cells were incubated for one hour at 37°C (200 rpm). Regenerated cells were streaked out on LB-plates with the appropriate antibiotic and incubated overnight (37°C). Transformation without the ligation mix used as a negative control.

2.7.7. Transformation of electro-competent *A. tumefaciens* (adapted from Tung & Chow, 1995)

Plasmid DNA (10-100 ng/ μ L) was added to electrocompetent cells (50 μ L) and mixed carefully. The mixture was transferred to a pre-cooled electro cuvette and placed into the electroporation device (Bio-Rad). An electric pulse was applied for 9 seconds (25 μ F, 2.5 kV, 400 Ω). YEB medium (1 mL) was subsequently added, and cells were regenerated at 38°C (2 h, 250 rpm). Regenerated cells were placed on YEB-plates containing appropriate antibiotics and incubated at 28°C for two days.

2.7.8. Isolation of genomic DNA

Plant material was harvested, frozen in liquid nitrogen and ground with mortar. Depending on the experiment to be carried out, two different DNA-extraction methods used.

2.7.8.1. UREA-extraction method (adapted from Sambrook et al., 1989)

2 x Lysis-buffer:

0.6 M NaCl
0.1 M Tris-HCL, pH 8.0
40 mM EDTA, pH 8.0
4% (w/v) Sarkosyl
1% (w/v) SDS

Ground plant material (50–200 mg) was resuspended in 375 μ L 2x Lyses-buffer and 375 μ L 2 M UREA. After the addition of one volume phenol-chloroform-isoamyl alcohol (25:24:1), the suspension was centrifuged (14,000 rpm, 10 min, RT), and the resulting supernatant precipitated with 0.7 volumes of isopropanol. The DNA-pellet obtained by centrifugation (14,000 rpm, 15 min, 4°C) and subsequently washed twice with 1 mL 70% (v/v) EtOH. The pellet was air-dried and resuspended in 25 μ L resuspension buffer — genomic DNA stored at -20°C.

2.7.8.2. Quick-extraction method (adapted from Edwards et al., 1991)

The “quick-extraction method” provides genomic DNA in a quality that is sufficient for genotyping PCRs. Plant material from *A. thaliana* was transferred to small tubes containing 250 μ L DNA-extraction buffer and subsequently triturated with a small mortar. After the addition of 100 μ L chloroform-isoamyl alcohol (24:1), the suspension was mixed and centrifuged (16,000 g, 10 min, RT). The supernatant was transferred and DNA precipitated by the addition of isopropanol (5 min, RT). After centrifugation (16,000 g, 10 min, RT), genomic DNA has purified in two subsequent washing steps with 70% (v/v) ethanol. The air-dried pellet was finally resuspended in d.H₂O and stored at -20°C.

DNA-extraction buffer:

100 mM Tris-HCL, pH 8.0
100 mM NaCl
10 mM EDTA
1% (w/v) SDS

2.8. Amplification of DNA fragments by PCR (Mullis & Faloona, 1987)

The polymerase-chain-reaction (PCR) was used to amplify specific nucleic sequences by a primer pair and a thermostable DNA-polymerase.

A typical PCR-mixture:

10 x PCR-Buffer	2.0 μ L
Template-DNA (1 μ g/ μ L)	1.0 μ L
dNTPs (10 mM)	1.0 μ L
Fwd.-primer (10 mM)	0.5 μ L
Rev.-primer (10 mM)	0.5 μ L
Taq-polymerase	0.2 μ L
d.H ₂ O	14.8 μ L
Total	20.0 μ L

A standard PCR-program:

Initial denaturation	95 °C	5 min
Denaturation	95 °C	30 s
Annealing	T _a	30 s
Elongation	72 °C	30 s/500 bp
Final elongation	72 °C	10 min
Storage	4 °C	∞

2.8.1. Genotyping of overexpression mutants

The overexpression lines were genotyped before the experiments to reassure the gene overexpression by a gene-specific primer and a vector-specific primer pair. Wild-type DNA was used as a control for gene-specific PCRs.

2.8.2. Colony-PCR (Sambrook et al., 1989)

A single bacterial colony was inoculated in 50 μ L d.H₂O and 2 μ L of this suspension was deployed in a colony-PCR. Gene and vector-specific primers for the construct of interest were used to test the successful transformation of the construct.

2.9. Extraction of RNA from plant tissue

2.9.1. RNA extraction with urea (adapted from Missihoun et al., 2011)

RNA extraction buffer:

6 M Urea
3 M Lithium chloride (LiCl)
10 mM Tris-HCl, pH 8.0
20 mM EDTA, pH 8.0

The extraction buffer was autoclaved prior to use (20 min, 121°C, 1.2 bar). Frozen and pulverised plant material (100 mg) was resuspended and homogenised in 500 μ L extraction buffer. After the addition of 500 μ L phenol-chloroform-isoamyl alcohol (25:24:1), the suspension was centrifuged twice (5 min, 10,000 g, RT) and the upper phase subsequently transferred to a tube containing an equal volume of chloroform-isoamyl alcohol (24:1). The mixture centrifuged (5 min, 10,000 g, 4°C), and the supernatant precipitated with one volume ice-cold isopropanol and 0.1 volume 3 M sodium acetate, pH 5.2. The suspension was incubated on ice for five minutes, centrifuged (10 min, 14,000 rpm, 4°C), and the resulting pellet was washed twice with 70% (v/v) ethanol. The air-dried pellet finally dissolved in 25 μ L d.H₂O, or diethylpyrocarbonate- (DEPC, final concentration 0.1%) treated water and stored at -80°C.

2.9.2. Phenolic RNA-extraction method (adapted from Valenzuela-Avenidaño et al., 2005)

Frozen plant material (100 mg) was subjected to a 1.5 mL RNA extraction buffer. The suspension was vortexed, centrifuged (10,000 g, 10 min, RT), and the supernatant mixed with volumes of chloroform-isoamyl alcohol (24:1). After centrifugation (10,000 g, 10 min, 4°C), the upper phase was precipitated with equal volumes of ice-cold isopropanol and 0.8 M sodium citrate/ 1 M sodium chloride solution. The mixture was incubated (10 min, RT) and subsequently centrifuged (12,000 g, 10 min, 4°C). The RNA was air-dried and resuspended in 25 μ L d.H₂O or DEPC-treated water.

RNA extraction buffer:

38% (v/v) Phenol
0.8 M Guanidine thiocyanate
0.4 M Ammonium thiocyanate
0.1 M Sodium acetate, pH 5.0

The extraction buffer was autoclaved before use (20 min, 121°C, 1.2 bar).

2.10. Reverse transcription-polymerase chain reaction**2.10.1. DNase treatment (adapted from Innis et al., 2012)**

The “RevertAid H_{einus} First Strand cDNA Synthesis Kit” (Fermentas) used according to the manufacturer’s instructions for DNase treatment of RNA.

A typical DNase treatment reaction:

10 x reaction buffer (Fermentas)	1 µL
DNase I, RNase-free (Fermentas)	1 µL
RNA (500 ng/µL)	2 µL
DEPC-treated d.H ₂ O	6 µL
Total	10 µL

The reaction incubated for 30 min at 37°C. Subsequently, the DNase was inactivated by the addition of 1 µL 50 mM EDTA for 10 min at 65°C.

2.10.2. Synthesis of cDNA (adapted from Innis et al., 2012)

The “RevertAid H_{einus} First Strand cDNA Synthesis Kit” (Fermentas) used according to the manufacturer’s instructions for the generation of cDNA.

A typical reaction mix:

5 x reaction buffer (Fermentas)	4 µL
dNTPs (10 mM)	2 µL
Template RNA (1000 ng/µL)	1 µL
Oligo(dT)18 primer (10 mM)	0.5 µL
Transcriptase (Fermentas)	0.5 µL
DEPC-treated d.H ₂ O	rest
Total	20 µL

2.11. Extraction of proteins

The reaction was incubated for 60 min at 42°C and finally terminated by heating for 10 min at 70°C. The reverse transcription reaction product stored at -20°C or directly used in PCR.

2.11.1. Extraction of total proteins (Röhrig et al., 2008)

The quality of extracted proteins was sufficient for 2D- gel analysis.

Dense SDS:

30% (w/v) Sucrose
2% (v/v) SDS
0.1 M Tris-HCl, pH 8.0
5% (v/v) 2-Mercaptoethanol

Pulverised plant material (3 mL) was transferred to a 15 mL reaction tube containing 10 mL ice-cold acetone. The suspension was vortexed, centrifuged (4,000 g, 5 min, 4°C) and washed again with acetone. The resulting pellet was resuspended in 10 mL 10% (w/v) TCA solution and sonicated in an ice-cold water bath (10 min). After centrifugation (4,000 g, 5 min, 4°C), the resulting pellet was successively washed with 10% (w/v) TCA in acetone (3 times), 10% (w/v) TCA and twice with 80% (w/v) acetone. The acetone-wet pellet was thoroughly resuspended in 5 mL Dense SDS and 5 mL phenol, centrifuged, and the upper phase was transferred to new reaction tubes. Proteins were precipitated by the addition of 0.1 M ammonium acetate in methanol for 30 min at -20°C. The protein pellets were washed twice with ice-cold 0.1 M ammonium acetate and once with ice-cold 80% acetone (w/v). Proteins were stored at -20°C before using.

2.11.2. Direct-method (Laemmli, 1970)

Pulverized plant material (200 mg) was affiliated in 100 µL 1 x SDS-sample buffer (2% (w/v) SDS, 10% (w/v) glycerol, 60 mM Tris-HCl, pH 6.8, 0.01% (w/v) bromphenol blue, 0.1 M DTT), boiled at 95°C for 10 min and centrifuged (2,000 g, 10 s, RT). Samples were directly used for gel-electrophoresis (SDS-PAGE) or stored at -20°C.

2.11.3. Protease inhibitor Assay

The proteasome inhibitor cocktail P9599 (Sigma) was used according to the manufacturer's instructions to prevent protein degradation. The treatment was conducted as described by

(Speranza *et al.*, 2001). The inhibitor solution was added to protein extracts (final concentration of 0.2% (w/v)), while control samples were not treated with the inhibitor cocktail.

2.12. Quantification of nucleic acids and proteins

2.12.1. Estimation of nucleic acid concentrations

The concentration of DNA and RNA was estimated with a spectrophotometer (Biospec–Nano, Shimadzu Biotech). The concentration (c) can be calculated with the optical density at 260 nm (OD₂₆₀) in combination with the dilution factor (V) and a DNA/RNA– specific multiplication as follows:

$$\text{Double-stranded DNA: } c \text{ [}\mu\text{g/mL]} = \text{OD}_{260} \times V \times 50$$

$$\text{RNA: } c \text{ [}\mu\text{g/mL]} = \text{OD}_{260} \times V \times 40$$

The OD₂₆₀/OD₂₈₀ quotient describes the purity of the solution. A value between 1.8 and 2.0 indicates pure nucleic acids.

2.12.2. Estimation of protein concentrations (adapted from Bradford, 1976)

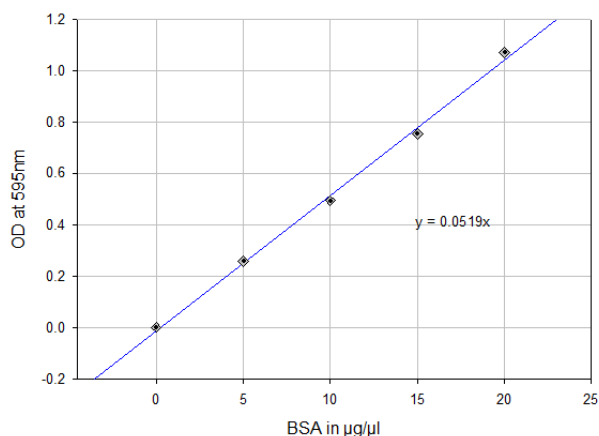


Figure 6: BSA calibration line.

Proteins (100 μL) were dissolved overnight in 7 M urea/2 M thiourea. 20 μL of the protein solution was transferred to a reaction tube containing 0.8 mL d.H₂O and 0.2 mL Bradford reagent (section 2.10.2). The mixture was incubated at RT for 5 min, and the absorbance at 595 nm was measured. The protein concentration was estimated with a calibration line from BSA (Figure 6).

2.13. Over-expression and isolation of recombinant proteins

Buffer A

50 mM Sodium dihydrogenphosphate

0.3 M NaCl

5 mM Imidazole

10% (v/v) Glycerol

0.1% (v/v) Triton X-100

pH 8.0 (NaOH)

Buffer C

50 mM Sodium dihydrogenphosphate

0.3 M NaCl

0.25 M Imidazole

10% (v/v) Glycerol

0.1 % (v/v) Triton X-100

pH 8.0 (NaOH)

Buffer B

50 mM Sodium dihydrogenphosphate

0.3 M NaCl

20 mM Imidazole

10% (v/v) Glycerol

0.1% (v/v) Triton X-100

pH 8.0 (NaOH)

Regeneration buffer

0.1 M EDTA

0.5 M NaCl

20 mM Tris-HCl, pH 7.9

Charge buffer

50 mM Nickel sulfate

A pre-culture of the bacteria of interest was grown in 2 mL LB-medium supplemented with the appropriate antibiotic at 37°C overnight. The next day the main culture (100 mL) was inoculated and grown (37°C, 200 rpm) till an OD600 of 0.5. The culture was subsequently incubated at 26°C for 15 min in the dark, and a sample of the not-induced bacteria was collected (sample 0 h).

Expression of proteins was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration 1mM), and bacteria were grown for 3-12 h at 26°C. Bacterial samples (1 mL) were taken every hour and stored at -20°C. Primary cultures (100 mL) were centrifuged (4,000 g, 20 min, 4°C), and bacterial pellets were stored at -20°C.

Over-expression of recombinant proteins often results in the precipitation of insoluble proteins, so-called “inclusion bodies” the formation of the inclusion body was tested via Palmer & Wingfield, 2001 method with 1 mL samples. The samples were centrifuged (4,000

g, 5 min, 4°C) and resuspended in 200 µL ice-cold PBS. The mixture was sonicated (4 x 20 s) after the addition of volumes of 10% (v/v) Triton X-100.

Bacterial cells were centrifuged (12,000 g, 10 min, 4°C), and pellets and supernatants were separated and resuspended in SDS-sample buffer before electrophoretic analysis. The mixture was incubated on ice for 30 min and subsequently sonicated (6x 20 s). Inclusion bodies were obtained by centrifugation for 30 min (14,000 rpm, 4°C) and stored at -20°C. The supernatant was purified through a filter (0.45 µm), and a control sample (1 mL) was taken before loading the culture on a Ni-NTA column. Before the protein solution was loaded, the column was sequentially washed with d.H₂O (3 mL), charge buffer (5 mL) and buffer A. After 2 mL of the protein solution ran through the column, 20 µL of the flow-through (sample Ft) was collected and frozen at -20°C. The Ni-NTA-bound proteins were washed by the addition of buffer A (10 mL) and buffer B (8 mL) to exclude unspecific binding of proteins. His-tagged proteins were eluted by the sequential addition of buffer C (six times 500 µL). The eluted protein fractions were collected in reaction tubes (F1-F6) and stored at -20°C. The Ni-NTA column was regenerated by the addition of d.H₂O (3 mL), regeneration buffer (3 mL), and 20% (v/v) ethanol (3 mL). The regenerated column was stored in 20% (v/v) ethanol at 4°C.

Recombinant proteins were loaded onto PD10–desalting columns (GE Healthcare) to remove remaining salts from the solution. The desalting columns were used according to the manufacturer's instructions, and the buffer was exchanged to 100 mM ammonium bicarbonate (NH₄HCO₃). Ammonium bicarbonate is a volatile organic compound that is commonly used in protein lyophilisation. Proteins were lyophilised for 2-3 days in a lyophilizer (LDC-2, Christ) and stored at -20°C.

2.13.1. Cobalt column gravity-flow purification

The Cobalt column gravity-flow purification method was used to purify the proteins which are low expressed.

MES Buffer: 20 mM 2-(N-morpholine)-ethanesulfonic acid, 0.1 M sodium chloride;

pH 5.0.

Native conditions buffers:

Equilibration/Wash Buffer: 50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole; pH 7.4.

Elution Buffer: 50 mM sodium phosphate, 300 mM sodium chloride, 150 mM imidazole; pH 7.4.

Denaturing conditions buffers:

Equilibration/Wash Buffer: 50 mM sodium phosphate, 300 mM sodium chloride, 6 M guanidine HCl, 10 mM imidazole; pH 7.4.

Elution Buffer: 50 mM sodium phosphate, 300 mM sodium chloride, 6 M guanidine HCl, 150 mM imidazole.

The HisPur cobalt spin columns also used for gravity-flow purifications. Purifications performed at room temperature. Columns equilibrated at working temperature. Samples prepared by mixing the protein extract with equilibration/wash buffer. Column Equilibrated with two resin-bed volumes of equilibration/wash buffer. The protein extracts add to the column and let it flow through the column via gravity. The flow-through collected in a falcon tube. The step repeated to get more protein for low expressed protein. His-tagged proteins eluted from the resin by adding one resin-bed volume of elution buffer max 3 mL in total. Each fraction collected in separate tubes.

To remove excess imidazole for downstream applications, the PD10 column used with 100 mM NH_4HCO_3 buffer and the PD10 buffer exchange protocol (GE Healthcare) followed. The final product was freeze-dried and stored at -80°C .

2.14. Electrophoresis of proteins

2.14.1. Isoelectric focusing (first dimension)

IPG strips: Immobiline DryStrip, pH 3–10, 7 cm (GE Healthcare)

Protein samples (20-40 μg) were completely dissolved in 130 μL rehydration solution (>1 h, RT). Insoluble particles were removed by centrifugation (12,000 g, 10 min, RT), and 125 μL of the protein solution was then transferred to an IPG strip in a focusing tray. To prevent evaporation during the rehydration process (12-16 h, 20°C), the IPG strip was overlaid with mineral oil. Proteins were separated according to their isoelectric point in the IPGphor II isoelectric focusing system (GE Healthcare).

Typical focusing conditions for the first dimension used in this work:

14 h Rehydration

30 min 500 V

30 min 1000 V

100 min 5000 V

Before separating the proteins in the second dimension, IPG strips were equilibrated in SDS-containing buffers to reduce and alkylate cysteine residues, which might interfere with the staining of two-dimensional gels. IPG strips were incubated for 15 min in 5 mL equilibration buffer, supplemented with 50 mg DTT. Subsequently, IPG strips transferred to a 5 mL equilibration buffer with 125 mg iodoacetamide. Finally, IPG strips were put on SDS-PAGE gels and sealed with IEF agarose.

Equilibration buffer	Rehydration solution	IEF-agarose
50 mM Tris-HCl, pH 6.8	7 M Urea	0.5% (w/v) Agarose
2% (w/v) SDS	2 M Thiourea	0.002% (w/v) Bromphenol blue
6 M Urea	2% (w/v) CHAPS	in 1x running buffer
0.002% (w/v) Bromphenol blue	0.002% (w/v) Bromphenol blue	melted at 100°C, stored at 60°C
30% (v/v) Glycerol	Prior to usage add: 0.2% (w/v) DTT 0.5% (v/v) IPG buffer pH 3-10 (GE Healthcare)	

2.14.2. SDS-PAGE (second dimension) (adapted from Laemmli, 1970)

SDS-sample buffer (2x)	Polyacrylamide gels	Separating gel (12%)	Stacking gel (4%)
4% (w/v) SDS	d.H ₂ O	2.88 mL	2.16 mL
20% (w/v) Glycerol	1.5 M Tris-HCl, pH 8.8	2.34 mL	-
120 mM Tris-HCl, pH 6.8	1 M Tris-HCl, pH 6.8	-	375 µL
Bromphenol blue	Rotiphorese gel 30	3.60 mL	405 µL
0.2 M DTT (added freshly)	10% (v/v) SDS	90 µL	30 µL
	10% (w/v) APS	90 µL	30 µL
	TEMED	3.6 µL	3 µL

1x Running buffer

25 mM Tris

192 mM Glycine

0.1% (w/v) SDS.

Protein pellets (1-fold) and solutions (2-fold) were dissolved in SDS-sample buffer and boiled at 95°C for 10 min before loading on SDS-PAGEs. For the separation of proteins in the second dimension, IPG strips were directly transferred to SDS-PAGEs. Electrophoresis was performed in 1 x running buffer at 20 mA for 2 h.

2.14.3. Staining of polyacrylamide gels

2.14.3.1. Coomassie Staining (adapted from Zehr et al., 1989)

Polyacrylamide gels were fixed in a 100 mL fixation solution for one hour or overnight. Subsequently, the gel was washed three times for 10 min with d.H₂O. Finally, the gels were incubated in a 50 mL staining solution on a shaker overnight. Gels were washed several times with d.H₂O for destaining. The sensitivity limit of the Coomassie stain is about 10-50 ng protein per band.

Fixation solution	Staining stock solution	Staining solution
10% (v/v) Acetic acid	10% (w/v) Ammonium sulfate	80% (v/v) Staining stock solution
40% (v/v) Methanol	1% (v/v) Phosphoric acid	20% (v/v) Methanol
	0.1 % (w/v) Coomassie G250	

2.14.3.2. Silver staining (adapted from Ejvind Mortz et al., 2001)

Polyacrylamide gels are incubated in Fixer (40% ethanol, 10% acetic acid, 50% H₂O) for 1 h and continued with the washing of the gel in H₂O for at least 30 min. Water was changed several times in between to remove all acetic acid and to reduce background staining and increase sensitivity. The gel was sensitised in 0.02% sodium thiosulfate (0.04 g Na₂ S₂O₃, 200 mL H₂O) for only 1 min. Later on, the gel was washed in H₂O for 3 x 20 s and continued with incubation of gel for 20 min in 4°C cold 0.1% silver nitrate solution (0.2 g

AgNO₃, 200 mL H₂O, 0.02% formaldehyde. After that, the gel was washed in H₂O for 3 x 20 s. In a new staining tray, the gel was placed and washed in H₂O for 1 min. The gel was developed in 3% sodium carbonate (7.5 g Na₂CO₃ in 250 mL H₂O), 0.05% formaldehyde. The developer solution immediately changed when it turned yellow. When the staining was sufficient, the gel was washed in H₂O for 20 s and followed by termination of staining in 5% acetic acid for 5 min. The gel was stored at 4 °C in 1% acetic acid. Before MS analysis, the gel was washed in water for 3 x 10 min to ensure complete removal of acetic acid.

2.15. Protein blot (adapted from Towbin et al., 1979)

The protein blot enables the transfer of proteins from polyacrylamide gels to membranes and subsequent immunological detection of specific proteins.

Towbin–buffer	TBS	TBST
25 mM Tris	20 mM Tris, pH 7.5	0.1% (v/v) Tween–20 in TBS
0.2 M Glycine	0.15 M NaCl	
20% (v/v) Methanol		

Proteins were transferred at 70 V for 1-3 h (4°C), as described by Towbin et al. (1979). The successful transfer of proteins to nitrocellulose membrane (Whatman) was confirmed by staining with Ponceau red for 10 min. The membrane was destained with d.H₂O.

The membrane was blocked in 4% (w/v) low-fat milk powder in TBST for 1 hour (RT) or overnight (4°C) to minimise unspecific detection of the antibody. Afterwards, the membrane was incubated with the first antibody (concentration 1:1000 – 1:5000 in blocking solution, depending on the antibody used) for 1 h (RT) or overnight (4°C).

Ponceau Red	Blocking solution
0.2% (w/v) Ponceau S	4% (w/v) low-fat milk powder in TBST
3% (w/v) TCA	

The membrane was washed with TBST for a total of 30 min (1x 15 min; 3x 5 min) and subsequently incubated at RT for 45 min with the second antibody (1:5000 in blocking solution). The membrane was again washed, as described above. The “ECL Western Blotting Detection Reagent” kit (GE Healthcare) was used for the detection of signals according to the manufacturer’s instructions. The secondary antibody is coupled with horseradish peroxidase and binds the primary antibody. By the addition of its substrate,

the enzyme generates chemiluminescence signals, which were detected with the “Intelligent Dark Box II” (Fujifilm Corporation).

Alternative detection was done with photo-sensitive membranes. The blot was incubated with detection reagents as described above and subsequently transferred onto Whatman papers. The set-up was placed into photo-incubation cassettes and covered with photo-sensitive membranes (Fujifilm Corporation). Photo-membranes were incubated in the dark for 2 to 20 minutes and subsequently incubated in a fixation solution. The development of the screens was done according to the manufacturer’s instructions. The protein blots were stored at 4°C.

2.16. Overview of the different antibodies used in this thesis

Table 2: List of Antibodies.

Antibody	Antibody	Antibody	Secondary antibody
Anti-PLDrp1.1	Anti-At3g29075	Anti-His	Anti-IGG Peroxidase:
Rabbit antiserum	Rabbit antiserum	Rabbit antiserum	Goat antiserum
Animal 1115	Animal 15071	GE-Healthcare	Sigma–Aldrich Chemie GmbH
Final bleeding (10.02.2011)	Final bleeding (21.01.2016)	Dilution: 1:1000	Dilution: 1:5000
Dilution: 1:5000	Dilution: 1:5000		

2.17. Protein-lipid interactions

2.17.1. Protein-lipid-overlay assay (adapted from Deak et al., 1999)

Immobilisation of lipids on nitrocellulose filters in the protein-lipid-overlay assay is the method of choice to identify whether a protein can bind to a lipid, but it does not take into account the cellular environment where the lipids occur. The immobilised lipids (each 5 µg) on nitrocellulose membranes were kindly provided by Barbara Kalisch (IMBIO, University of Bonn, Germany). After blocking the membranes one hour in 5% (w/v), BSA in TBS-T at 4°C and then the membrane was incubated overnight with 1 µg/mL of protein dissolved in 5% (w/v) BSA in TBS-T at 4°C. The membranes were washed four times

with TBS-T (5 min), the primary antibody (1 h, RT). Washing steps were repeated, and the membrane was incubated for 45 min with the secondary antibody before analysis.

2.17.2. Liposome-binding assay (adapted from Zhang et al., 2004)

Phosphatidic acid (PA) and Phosphatidylcholine (PC), were dissolved in chloroform/methanol (2:1) Solution (final concentration 4 $\mu\text{g}/\mu\text{L}$) and stored separately at -20°C . For each assay, 250 μg lipids (150 μg PC & 100 μg of the tested lipid) transferred to fresh glass tubes. Solvents were evaporated under the fume hood and resuspended in liposome-binding buffer A or B (0.5 μL $\mu\text{g}/\mu\text{L}$). The mixture incubated at 37°C on a shaker (1 h, 250 rpm), and the resulting liposomes were vortexed (5 min) and subsequently centrifuged (10 min, 20,000 g, 4°C). Liposomes stored at -20°C for no longer than two days or directly used for further analysis.

Prepared liposomes resuspended in liposome-binding buffer A or B containing 0.1 $\mu\text{g}/\mu\text{L}$ of the protein of interest. The mixture incubated at 30°C for 30 min and subsequently spun down (10 min, 10,000 g, 4°C). Supernatants and pellets analysed on polyacrylamide gels.

Liposome-binding buffer A

20 mM MES

30 mM Tris-HCl (pH 7)

100 mM NaCl 0.5 mM NaCl

1 mM DTT

pH 5–9

Liposome-binding buffer B

20 mM MES

30 mM Tris-HCl (pH 7)

2 M Urea

0.5% CHAPS

1 mM DTT

2.18. Transient and stable transformation

2.18.1. Transient transformation via particle gun bombardment

Microcarriers (30 mg gold particles, \varnothing 1.6 μm) were washed thoroughly for 5 min with 1 mL ethanol followed by three washing steps with 1 mL sterile d.H₂O. Gold particles were resuspended in 500 μL sterile 50% (v/v) glycerol, distributed to fresh reaction tubes, and stored at 4°C for further procedure. For coating of microcarriers 25 μg plasmid DNA, 50 μL 2.5 M CaCl₂ and 20 μL 100 mM spermidine were added in the indicated order to the gold suspension and mixed thoroughly by vortexing for 5 min. The microcarriers were centrifuged shortly and subsequently washed with 140 μL 70% (v/v) ethanol and 100%

ethanol. A macro carrier disk was loaded with 25 μ L of the prepared DNA-coated microcarrier suspension and placed into the metal holder. A metal grid was placed on top, and the complete set-up was placed into the assembly unit of the particle gun. A 1/2 MS-media plate was loaded with fresh leaves from *A. thaliana* and placed in the centre of the particle gun. By the application of vacuum (27 mm Hg, 3.6 MPa), the pressure was rising till the desired value (1150 psi) and was then released to shoot the DNA-coated gold particles with high velocity into the leaves. The leaves were incubated in the 1/2 MS plate for 12-48 h and analysed under a confocal laser scanning microscope.

2.18.2. *A. tumefaciens*-mediated stable transformation of *A. thaliana* (adapted from Clough & Bent, 1998)

A definite positive clone of *A. tumefaciens* carrying the construct of interest incubated in 20 mL YEB-medium supplemented with appropriate antibiotics (50 μ g/mL kanamycin and 50 μ g/mL rifampicin) at 28°C (24 h, 250 rpm). This subculture was used to inoculate a 250 mL culture of YEB medium with antibiotics and grown until an OD600 of \sim 0.8 was reached before the transformation, siliques removed from flowering plants. The *Agrobacterium* culture transferred to a 500 mL beaker, and 0.05% (v/v) of the surfactant Silwet Gold L-77 added. The flowering plants were carefully immersed in the medium and dipped with gentle agitation 30 s. After *Agrobacterium*-infiltration, plants were covered with perforated plastic bags and grown under long-day conditions. After one week of recovery, the infiltration was repeated to increase transformation efficiency. Putatively transformed seeds were collected and sown on MS-media supplemented with the appropriate antibiotics.

3. Results

3.1. Characteristics of At3g29075

The structure, expression, and function of the gene *At3g29075.1* are yet unknown. The protein sequence consists of motifs rich in glycine residues such as “GGX,” “GXGX,” and “GXGX/GGX” (Figure 7). Therefore, the protein is classified as a glycine-rich protein (Sachetto-Martins et al., 2000).

```

MPYYTNDNDVDDDFTEYDMPYSGGYDITVITYGRSIPPSDETCYPLSSLSGDAFE
YQRPNFSSNHSSAYDDQALKTEYSSYARPGPVGGSDFGGRKPNISGYGGRTEVEY
GRKTESEHGGSGYGRRIESDYVKPSYGGHEDDGGDDGHHKHSKDYDDGDEKSKKKE
KEKKKDKKKDGNSEDEDFKKKKKKEQYKEHHDDDDYDEKKKKKKDYNDDEKKK
KKHYNDDEKDKKKHNYNDDEKDKKKKKEYHDDDEDKKKKKHYDNDDEKDKKKKH
RDDDDEKDKKKDKHHKGGH
    
```

Figure 7: Protein sequence of the *At3g29075*. Sequences highlighted in green colour shows the glycine-rich motifs.

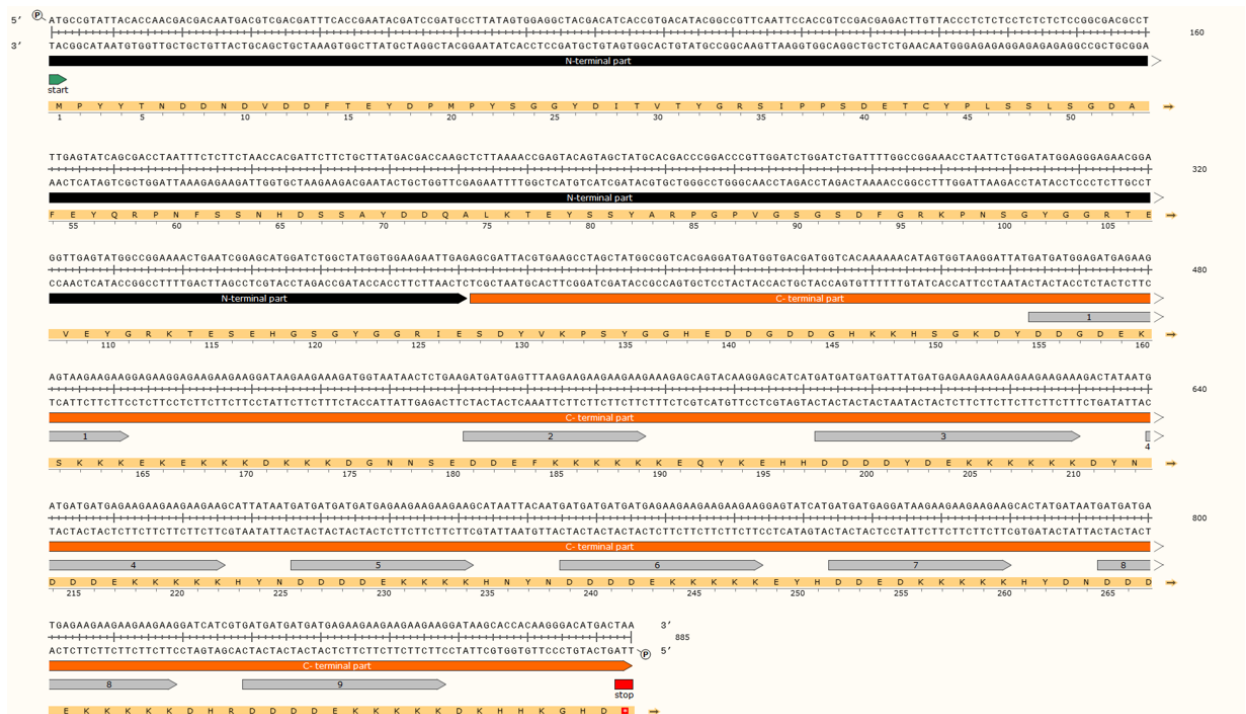


Figure 8: Nucleotide and protein sequence of the *At3g29075*. Green: Start codon; Red: Stop codon; Black: N-terminal part; Orange: C-terminal part; Gray: 9 imperfect repeats (DDDEKSKKK).

The *At3g29075.1* gene encodes a protein consisting of 294 amino acids, with a theoretical molecular weight of 34.4 kDa and a pI of 5.73. The protein *At3g29075* displays high content of the charged amino acids, such as aspartic acid (19.3%) and lysine (21.1%). It

also shows prominent imperfect tandem repeats in their C-terminal region (Figure 8). The repetitive sequences overlap with the least conserved amino acids.

MPYYTNDNDVDDFTEYDMPYSGGYDITVTYGRSIPPSDETCYPLSSLSGDAFEYQRPN
FSSNHDSAYDDQALKTEYSSYARPGVPVGSDFGRKPNNGYGGRTVEVEYGRKTESEHGS
GYGGRIESDYVKPSYGGHEDDGDDGHKKHSGKDYDDGDEKSKKKEKSKKDKKKDGNNNS
EDDEFKSKKKEQYKEHHDDDYDEKSKKSKDYNDDEKSKKSKHYNDDDEKSKKSKHNY
NDDDEKSKKSKKEYHDEDEKSKKSKHYDNDDEKSKSKDHRDDDEKSKSKDKKHHKGH

Figure 9: The toxicity analysis of the At3g29075 protein residues. Residues in red colour represent the starting residue of toxic stretches for selected window length and those in blue colour represent the trailing residues falling in the toxic stretch. Source: Toxinpred.

The protein toxicity was analysed by using Toxinpred website based on dipeptide (swiss-port) quantitative matrix method (QM). E-value cut-off of the motif-based technique was set for 10. The results show that the At3g29075 consist of several potential toxin residues (red colour) especial the tandem repeats of C-terminal parts consist of more toxin residues (Figure 9).

3.2. Protein-protein interactions

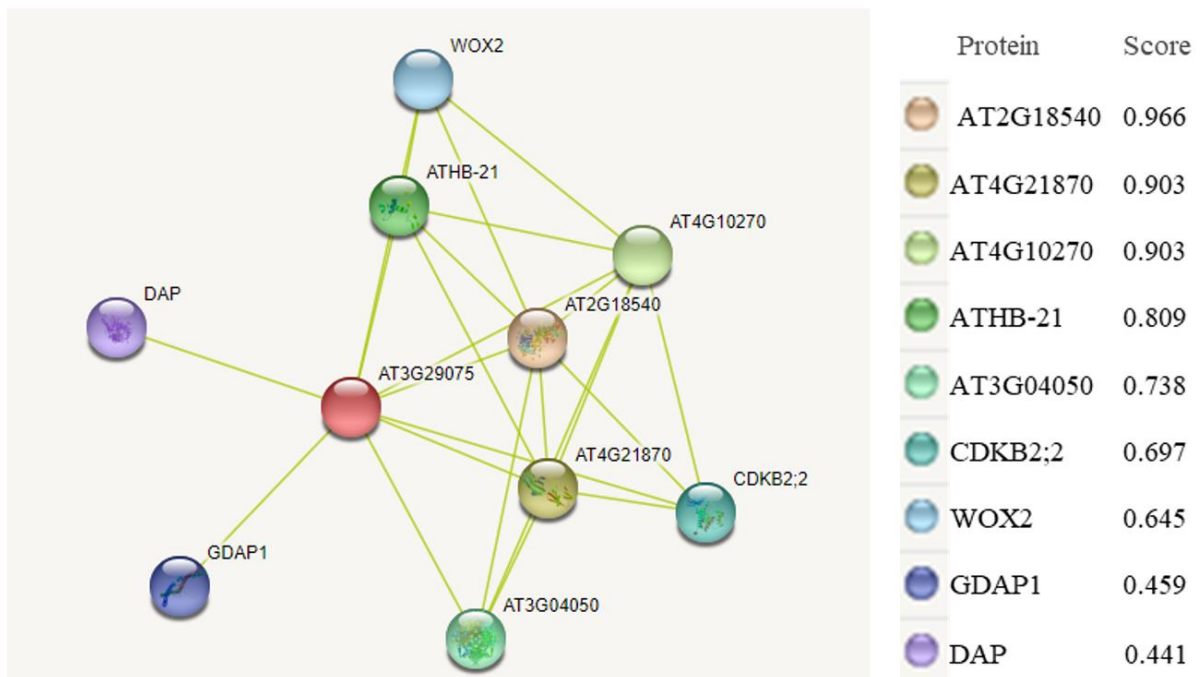


Figure 10: Predicted functional partners of the *At3g29075* protein. The protein-interaction network shows predicted protein interactions based on text mining analyses. A score value of 1 represents a confirmed interaction. *String11* was used to create the interaction network.

Protein-protein interactions (PPIs) are vital for the sustenance of life in living organisms (Brando et al., 2009). A database and web interface for searching and constructing interaction networks based on publicly available protein-protein interaction datasets were used in this study. Based on the prediction, the top three candidates that were found are seed storage protein (*At2g18540*), wound responsive family protein (*At4g10270*), and heat shock protein Hsp20 (*At4g21870*). The cyclin-dependent kinase is known to be expressed in flowers (*CDKB2:2*) (Figure 10). The prediction illustrated two other interacting partners; diaminopimelic acid (DAP) required for lysine biosynthesis, and Ganglioside-induced differentiation-associated protein 1 (GDAP1) encoded a protein that functions as a Golgi apparatus structural component.

3.3. Gene co-expression network of *At3g29075*

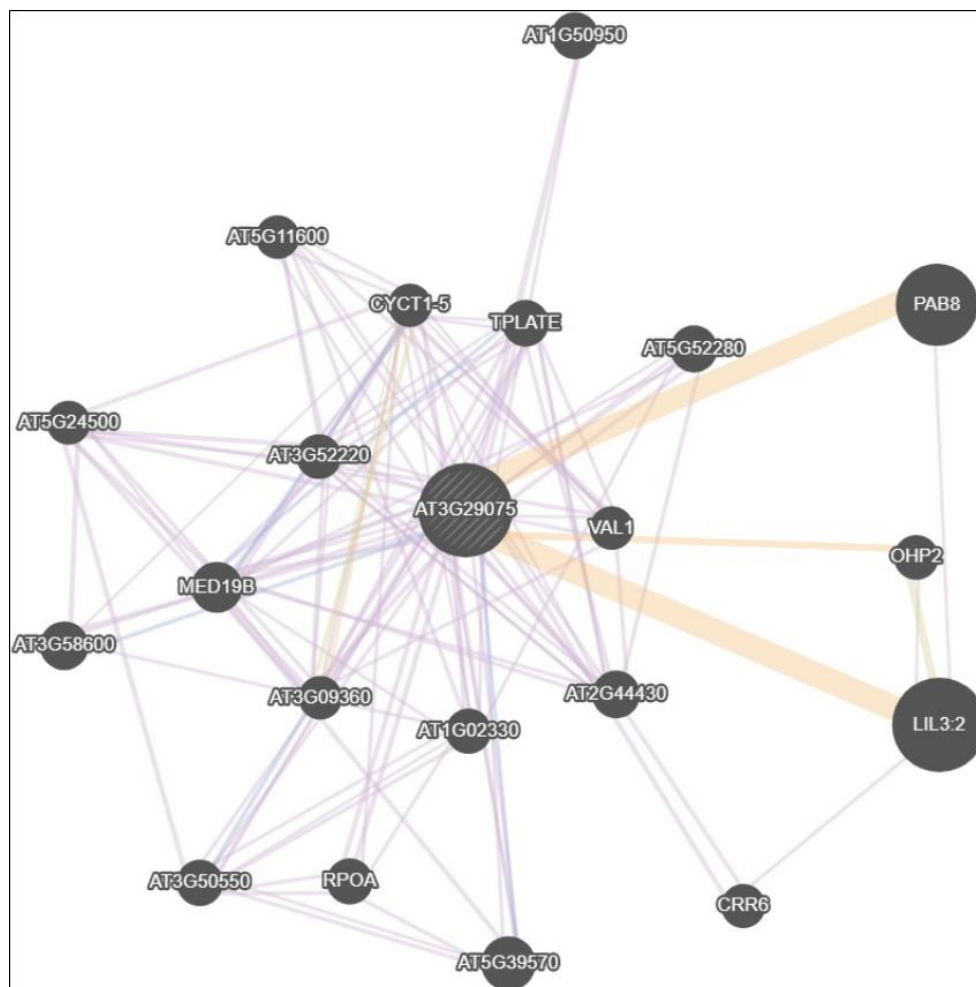

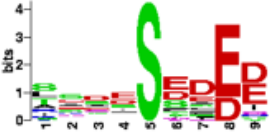



Figure 11: Gene co-expression network of the *At3g29075* gene. Purple line: Predicted co-expression. Blue line: Predicted co-localization. Gene co-expression network generated by geneMANIA.

The interaction partners of protein of interest were identified by an association of gene networks and *in-silico*-localization data. The data from the *in-silico* analysis showed PLDrp1 as one of the top three candidates for the co-expression network, which also shares a 49% gene sequence similarity in the N-terminal part of the *At3g29075* protein (Figure 11). LIL3.2, PAB8 are two of the top candidate genes containing the domains chlorophyll A-B binding family protein, Polyadenylate-binding protein 8 respectively showed high co-expression with the *At3g29075* gene. The *At3g52220* containing a Kinase phosphorylation domain displayed a similar co-expression profile to *At3g29075* (Supplementary table 28).

3.4. Post-translational modifications of *At3g29075*

Table 3: Prediction of kinase-specific phosphorylation sites in *At3g29075*.

AA position	KinasePhos 3.0	NetPhos 3.1	NetPhos 2.0	Kinase	
116	-0.9	0.934	0.935	CKII	
150	-3.0	0.998	0.998	CKII	
179	2.9	0.995	0.995	CKII	

Values in the range of 0.000-1.000 represent possibilities of phosphorylation sites (KinasePhos 3.0 and NetPhos 3.1) and kinase-specific phosphorylation sites (NetPhos2.0). For detailed information on the program, see section 2.1.3.1.

Evaluating posttranslational modification (PTM) patterns within protein molecules and analysing their functional implications provide significant challenges for plant biology.

Various PTMs required for assembly, localisation, function, and degradation in the context of the lifetime of a protein molecule (Miller and Dinesh-Kumar, 2019). Post-translational modifications embrace the addition of functional groups such as phosphate or acetate, carbohydrates, or even lipids. The interaction of phosphorylation and dephosphorylation of a protein is identified as a control for protein activity and function (Reinders and Sickmann, 2007). Three putative phosphorylation sites in At3g29075 were identified using different prediction tools and a support vector machine algorithm (Table 3). The prediction revealed that all the sites are phosphorylated by a casein kinase (CKII). No additional post-translational modification than the three PTM sites with high scores (>90%) was identified (see section 2.1.3). Based on the analysis of FAT-PTM database, AA-S₁₇₉ comes under hypothetical enzyme classes with individual PTM networks (Supplementary figure 3).

3.5. Detection of At3g29075 protein

The N-terminal part of the PLDrp1 protein has a 49% sequence similarity with the At3g29075 protein. The monoclonal N-terminal antibody of PLDrp1 protein acquired from Ufer, 2015. However, polyclonal antibody “ct-At3g29075” produced against At3g29075 protein by using the C-terminal part of At3g29075 (Nasr, 2015).

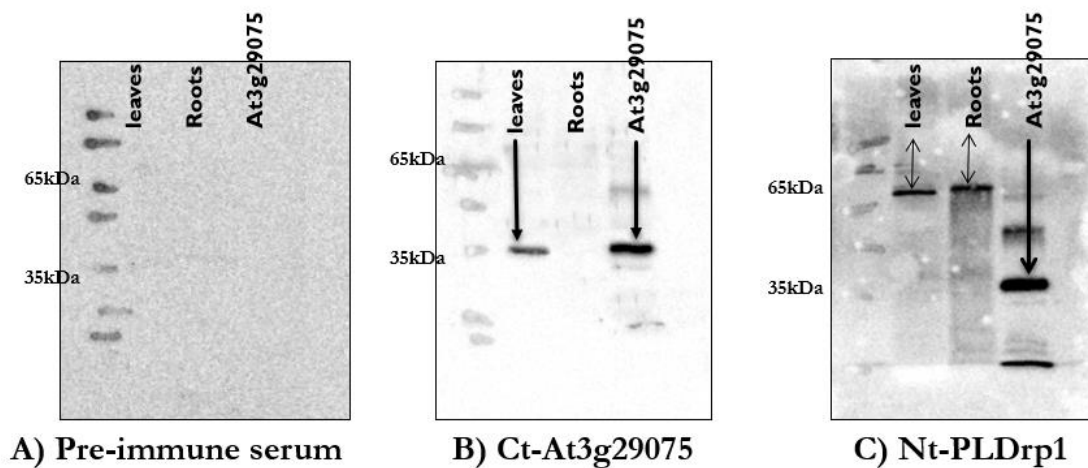


Figure 12: The different antibodies were tested (1 minute (Dilution: 1:2000)) on *Arabidopsis thaliana* leaves, root tissues, and the recombinant protein At3g29075. A) Immuno-detection of Pre-immune serum of Rabbit. B) Immuno-detection of Ct-At3g29075. C) Immuno-detection of Nt- PLDrp1.1. Single line arrow indicates the At3g29075 protein, and a double line arrow indicates the PLDrp1 protein.

The binding specificity of antibodies of Ct-At3g29075 and Nt-PLDrp1 were tested with *A. thaliana* leaves, roots, and the recombinant protein At3g29075.1 from *E. coli* (BL21). Protein blot analysis revealed the antibody Ct-At3g29075 as the specific antibody. It firmly bound the recombinant protein (~33kDa) and also in *A. thaliana* leaves (~33kDa). However, expression of the At3g29075 protein was not detected in the roots of *A. thaliana* (Figure 12B). In contrast, the Nt-PLDrp1 antibody can bind to both At3g29075 and PLDrp1 proteins (Figure 12C). Because it was raised against the N-terminal, part of the PLDrp1 protein and N-terminal region of both proteins sequence was identical, and the C-terminal part varies for both proteins sequence.

The binding of the Ct-At3g29075 polyclonal antibody was more specific to the plant, and recombinant At3g29075 proteins and very few non-specific bands observed; thus, the Ct-At3g29075 polyclonal antibody was used for further experiments.

3.6. Expression analysis of At3g29075

The regulatory network and expression patterns are analysed to characterise the gene and protein. Based on the stress treatments, expression of a gene and protein patterns may vary in different tissues during the various developmental stages. Microarray data provides a primary level of hits for the expression patterns of the unknown gene. The expression of At3g29075 transcripts and protein levels were analysed in different tissues, at different developmental stages at specific stress treatments in *A. thaliana* plants.

3.6.1. Tissue-specific expression of At3g29075

The tissue-specific expression of At3g29075 was evaluated by qRT-PCR in fully developed *A. thaliana* (Col-0).

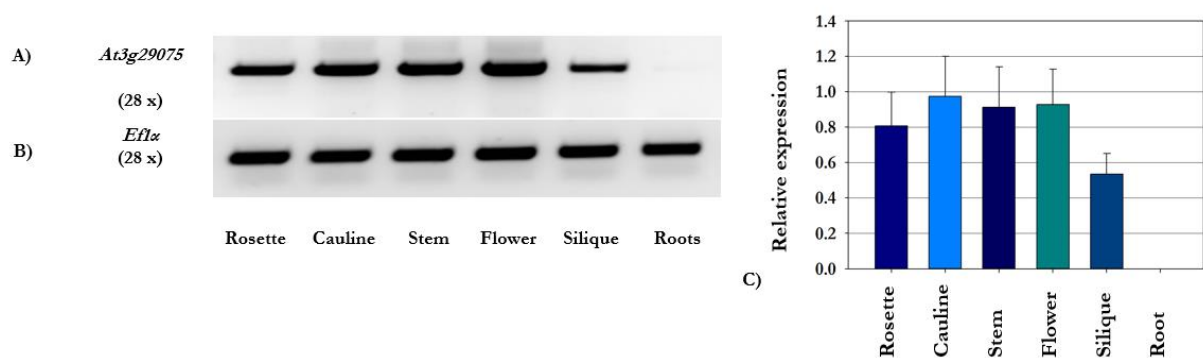


Figure 13: Expression of the At3g29075 gene in different plant tissues of wild type plants. A) 2 µg total RNA extracted from *Arabidopsis thaliana* was used to study the transcript expression level via RT-PCR. B) EF1α used as a loading control.

C) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using the ImageJ program.

Gene expression analysis in various plant tissue samples revealed different patterns of *At3g29075* expression. The gene expression in rosette leaves is lower than in cauline leaves. The highest *At3g29075* expression level observed in plant Cauline leaves. In stem and flower, the *At3g29075* gene expression was higher than in siliques. The *At3g29075* gene is not expressed in roots (Figure 13), *in silico* analysis (eFP browser 2.0) revealed similar results (Supplementary figure 1).

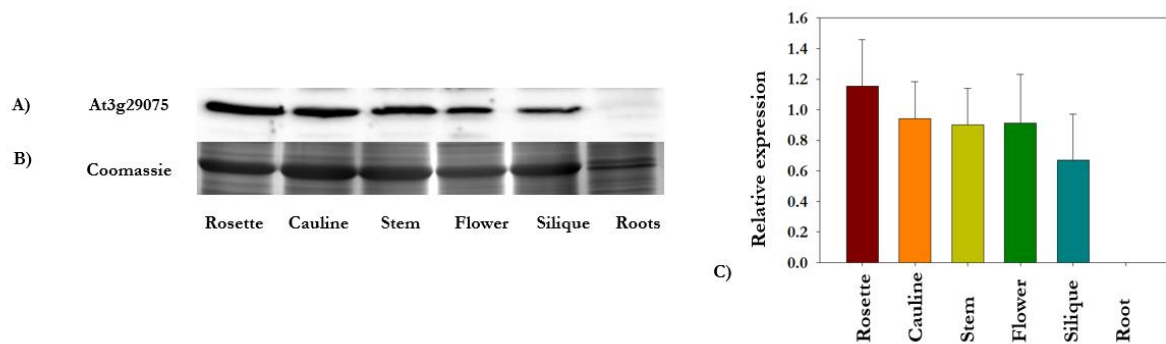


Figure 14: Expression of *At3g29075* at the protein level in different tissues of wild type plants. A) Immuno-detection of protein *At3g29075*. B) Coomassie stain of Rubisco protein used as a loading control. C) Mean values of relative expression levels of the *At3g29075* protein in three technical repetitions and quantified by using the ImageJ program.

The protein blots showed similar expression patterns as the transcript level in different plant tissues, the highest *At3g29075* protein expression observed in the rosette (Figure 14). There was no protein expression detected in the root tissue.

3.6.2. Developmental stage-specific expression of *At3g29075*

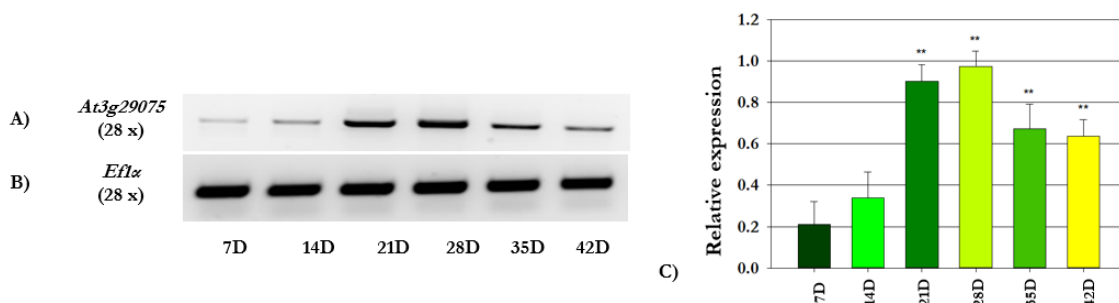


Figure 15: Expression of the *At3g29075* gene in different developmental stages. A) 2- μ g total RNA extracted from *Arabidopsis thaliana* plants used to analyse the transcript level of *At3g29075* via RT-PCR. B) *Efl α* transcripts used as a loading control. The star indicates the levels of significance in comparison to the untreated sample (one-way ANOVA,

Holm-Sidak method): $**p < 0.01$. C) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using the ImageJ program.

In correspondence to microarray studies, the protein level of *At3g29075* increases in senescent leaves during developmental stages. The highest level of transcript expression of *At3g29075* found on the 28th -day (Figure 15).

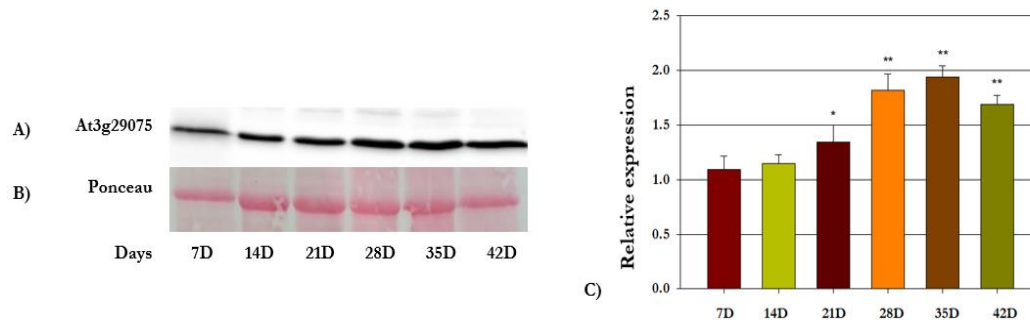


Figure 16: Expression of *At3g29075* at the protein level in different developmental stages. A) Immuno-detection of protein *At3g29075* in various stages of plant development. B) Ponceau stain of Rubisco used as a loading control. C) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using the ImageJ program. The star indicates the levels of significance in comparison to the untreated sample (one-way ANOVA, Holm-Sidak method): $*p < 0.05$; $**p < 0.01$.

The protein blots revealed different expression patterns of *At3g29075* protein during developmental stages of 7, 14, 28, 35, and 42 days old *A. thaliana* leaves (Figure 16). The protein content of *At3g29075* gradually increased from 7-days old seedling to 35 days mature plant. Although the highest expression detected in 35-day-old plants, the expression decreased in 42- days, mature senescent leaves.

3.6.3. Expression of *At3g29075* during seed germination

The translational regulation during the seed germination process was studied by analysing total protein changes via western-blot in five consecutive stages of *A. thaliana* seed germination. Dry seeds imbibed, as mentioned earlier (Section 2.3.3), using three

independent biological replicates. Seeds and seedlings were harvested at each physiological state during the seed to seedling transition up to 7 days (Joosen et al., 2010) (Figure 17).

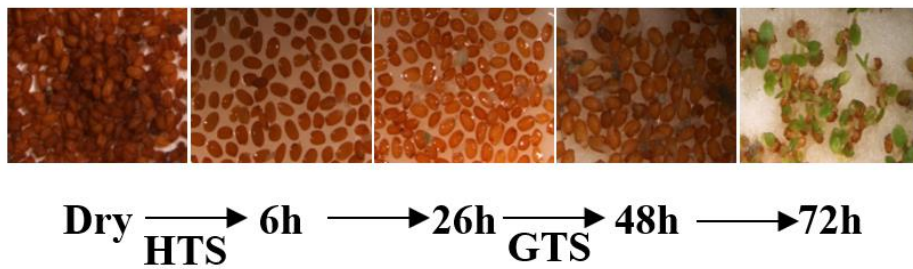


Figure 17: Schematic presentation of the transition from dry seed to a seedling at 72 h after the start of imbibition (HAI). The two phases where large sets of mRNAs are under translational control, the hydration translational shift (HTS) and germination translational shift (GTS) are indicated.

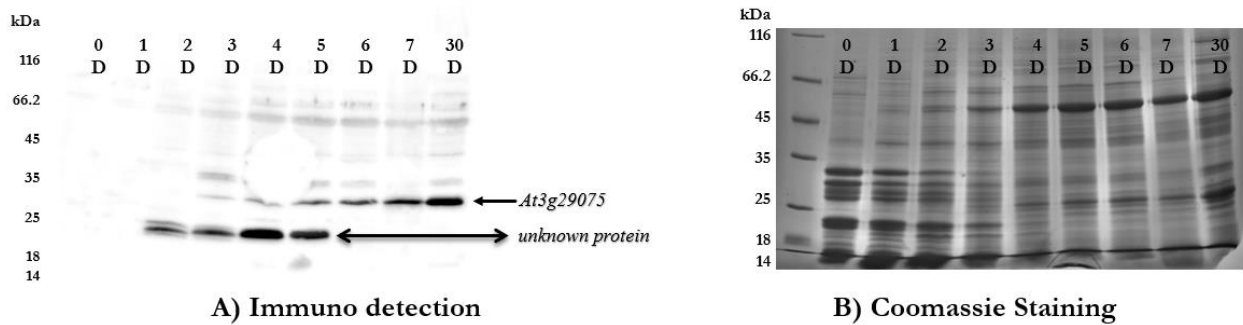


Figure 18: Protein expression in different stages of the seed germination after the start of imbibition (HAI) in *Arabidopsis thaliana*. A) Represents immune detection of the protein *At3g29075*, indicated in single line arrow during seed germination. Double line arrow indicates unknown protein; 30 days(D) sample used as a positive control. B) Coomassie Staining of total protein samples.

The Coomassie stain of the protein gel demonstrates the equal (5 μg) loading of the total protein on each sample (Figure 18B). There was no protein level expression detected at dry seeds (0-day) and on day one imbibition stage (6 h and 26 h) (Supplementary figure 2). In contrast, on the second day of imbibition, the unknown protein expression detected at ~ 20 kDa, and protein expression level increased until 5th day, then the protein expression disappeared from the sixth day of imbibition (Figure 18A). The *At3g29075* (~ 33 kDa)

protein expression detected from the third day of the germination process, and the protein expression gradually increased the 30th day leaves sample used as a positive control.

3.6.3.1. Identification of unknown protein

Two-dimensional gels electrophoresis was performed (Shen et al., 2008) to identify the unknown protein detected in the seed germination stage by the Ct-At3g29075 antibody. Total protein extracted from dry seeds (0-day), 3rd -day germinated sample, and 30th -day old plant leaves sample by following the method of Röhrig et al., (2008). Forty μ g of total protein used for Coomassie-stained gels and 20 μ g total protein samples used for immunoblottings.

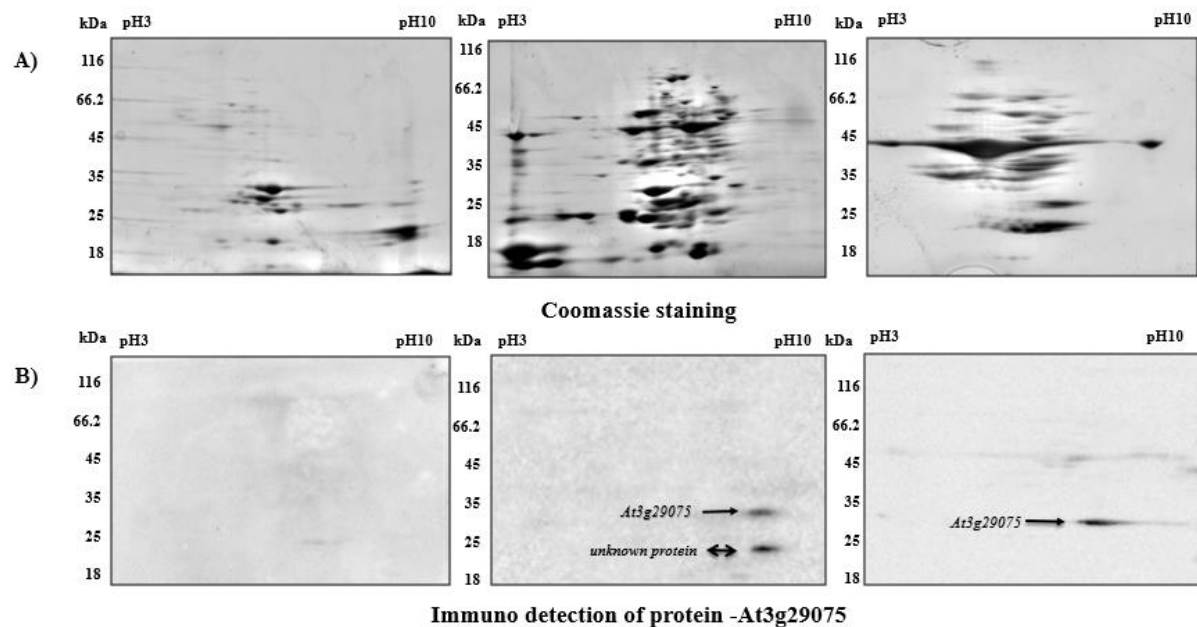


Figure 19: Two-dimensional analysis of total protein extracts from *Arabidopsis thaliana*. Total proteomes of wild type dry seed as a negative control, 3rd -day old seedling, and 30 -day mature plant sample used as a positive control are stained by Coomassie (A) and immune detected (B) by the Ct-At3g29075 polyclonal antibody. Single line arrow represents At3g29075 protein, and the double line arrow shows the unknown protein detected by Ct- At3g29075 antibody.

In dry seeds, the Ct-At3g29075 antibody did not detect the At3g29075 protein; therefore, it is used as a negative control. The total protein from 3rd - day germinated samples under immunodetection revealed two different spots, the At3g29075 protein identified at (~33 kDa) marked by a single line arrow the lower band around 20 kDa detected by the Ct-At3g29075 antibody is unknown protein marked by the double line arrow (Figure 19B).

Total protein from the 30th-day sample shows only At3g29075 protein on immunoblot marked by the single line arrow as a positive control.

However, the 3rd-day sample displayed weak protein signals of unknown protein on Coomassie Staining (Figure 19A). It was not possible to identify unknown protein *via* MS-analysis because of the insufficient yield of the unknown protein.

3.7. Dehydration stress

Drought stress is one of the most prevalent environmental factors limiting crop productivity (Bray, 1997, 2004). Plants have evolved to endure dehydration conditions using an array of morpho-physiological and biochemical adaptations through activation of a cascade of molecular networks (Basu et al., 2016). Under field conditions, crop plants subjected to short-term water deficits of several days to weeks, and some plant species have evolved to quickly limit the cellular damage and continue to grow in a stressful environment (Nakashima et al., 2014). In this study, four weeks old *A. thaliana* plants used, which grown under water limiting conditions for two weeks. The leaf samples collected at different time intervals, and their relative water content measured. *A. thaliana* plants were grown under short-day room condition supplemented with 50 mL of water (control). Dehydration stress was imposed by stopping the supplementation of 50 mL of water for two weeks. Relative water content measured in leaves both control and dehydration stress treated plants.

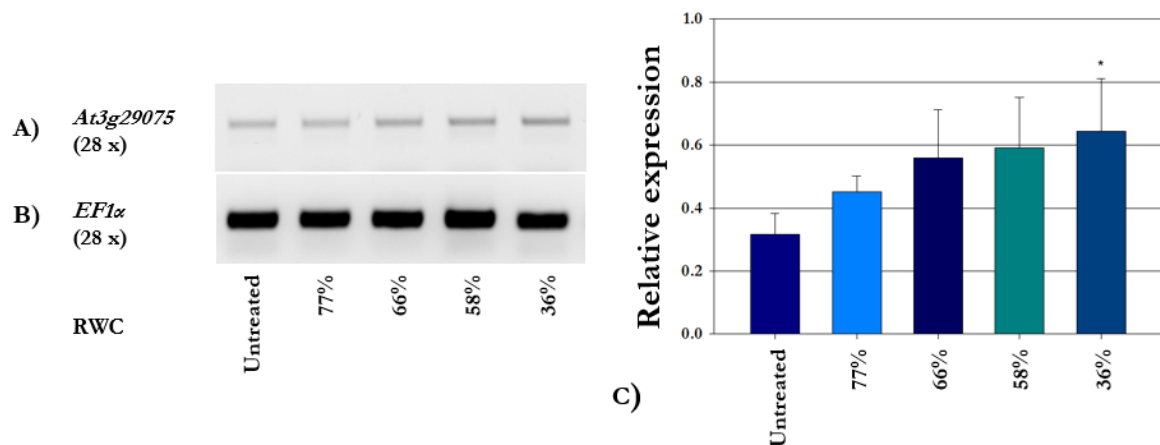


Figure 20: Expression of the *At3g29075* gene in response to dehydration. A) RT-PCR analysis of full-size *At3g29075.1* transcript using 2 μ g total RNA extracted from WT untreated and dehydrated plants. B) *EF1α* transcript served as a loading control (lower panel). C) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using the ImageJ program. RWC: Relative water content. The star indicates the levels of significance in comparison to the untreated sample (one-way ANOVA, Holm-Sidak method): * $p < 0.05$.

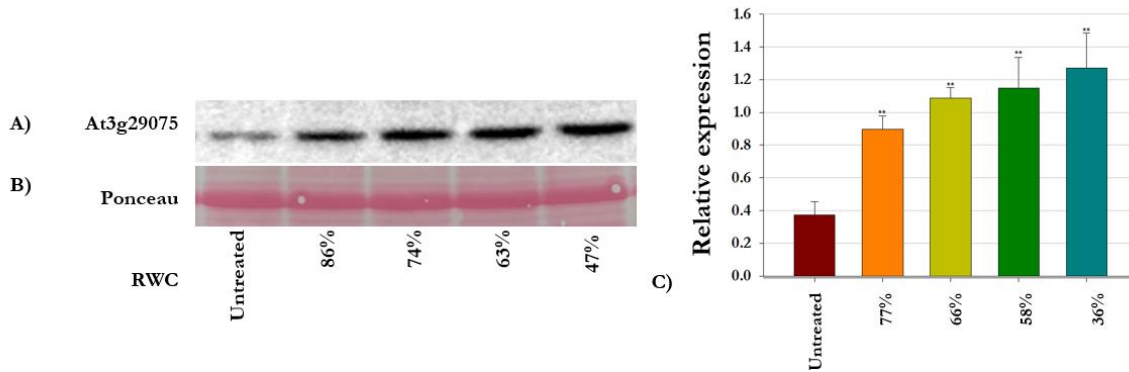


Figure 21: Expression of the *At3g29075* protein in response to dehydration. A) Immuno-detection of protein *At3g29075*, B) Ponceau stain of Rubisco used as a loading control. RWC: Relative water content. C) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using the ImageJ program. The star indicates the levels of significance in comparison to the untreated sample (one-way ANOVA, Holm-Sidak method): ** $p < 0.01$.

To analyse the *At3g29075* gene transcript expression level upon water limiting condition. RT-PCR analysis using dehydration-stressed *A. thaliana* plants showed an increase in transcript levels of *At3g29075* at wilting stage compared to untreated plants, stating that the dehydration stress in *A. thaliana* increased the expression of *At3g29075* gene. The significant level of increased expression found at 36% of the RWC sample compared to the untreated sample (Figure 20).

The protein blot embeds a similar expression pattern level as a transcript expression state of the *At3g29075* gene. The significant level of the protein expression pattern increased during dehydration stress (Figure 21). The highest level of expression was found in 47% RWC sample.

3.7.1. Memory response patterns of *At3g29075* upon dehydration stress

A. thaliana (Col-0) plants grown in potting soil in the short-day growth rooms. Repeated dehydration stresses were performed by air-drying for 2 h followed by a 22 h period of full re-hydration recovery as described. RT-PCR analyses showed on rosette leaves from pre-stressed (W) plants, from plants exposed to the first dehydration treatment (S_1), and during third stress (S_3) following two stress/recovery cycles. Plants from three independent biological samples used.

RT-PCR results showed that the expression of the *At3g29075* gene at the S_1 stage increased significantly than the pre-stressed W condition. However, the expression pattern at the S_3

stage after (two recovery cycles) dropped than the S_1 level (Figure 22). The results observed were similar to microarray data (Ding et al., 2013).

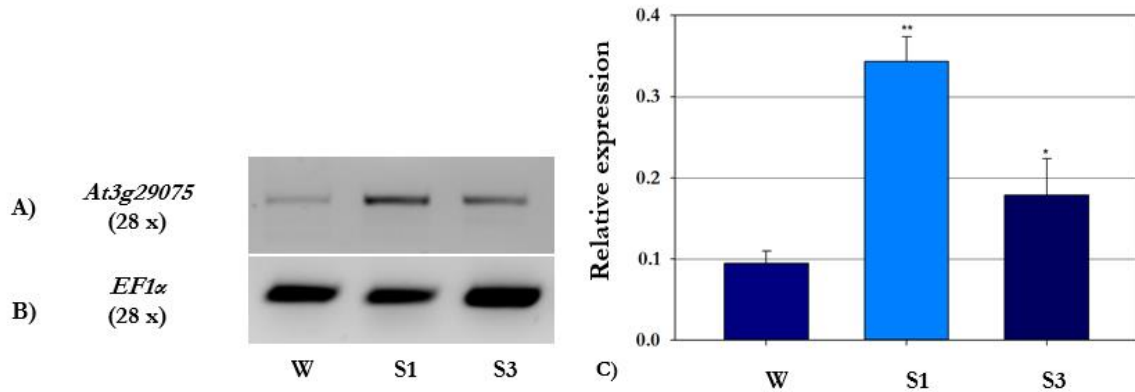


Figure 22: Expression of the *At3g29075* gene in different stages of dehydration stress condition. A) 2 μ g total RNA extracted from *Arabidopsis thaliana* was used to study the *At3g29075* expression at the transcript level by RT-PCR. B) *EF1α* transcripts used as a loading control. C) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using ImageJ program; pre-stressed (W) plants, from plants, exposed to the first dehydration treatment (S_1), and during third stress (S_3) following two stress recovery cycles. The star indicates the levels of significance in comparison to the untreated sample (one-way ANOVA, Holm-Sidak method): * $p < 0.05$; ** $p < 0.01$.

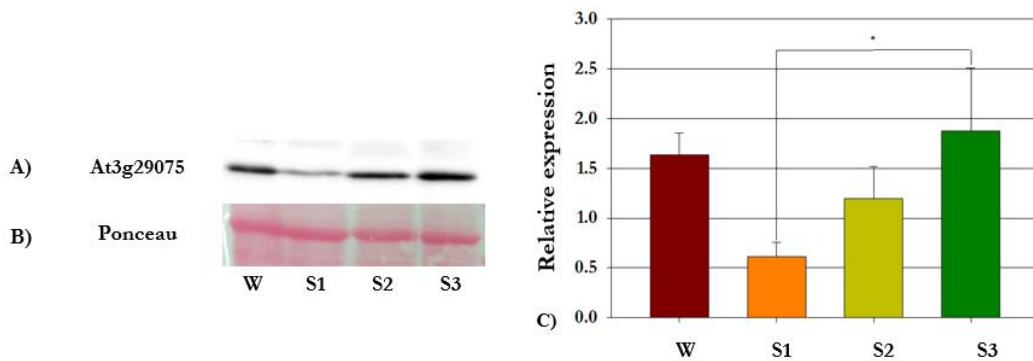


Figure 23: Expression of the *At3g29075* protein in different stages of dehydration stress in *Arabidopsis thaliana* plants. A) Immuno-detection of *At3g29075* protein. B) Ponceau staining of rubisco protein as a loading control. C) Mean values of relative expression levels of protein *At3g29075* in three technical repetitions quantified by using ImageJ program; pre-stressed (W) plants, from plants, exposed to the first dehydration treatment S_1 , S_2 second stress and third stress S_3 following two stress/recovery cycles. The star indicates the levels of significance in comparison to the untreated sample (one-way ANOVA, Holm-Sidak method): * $p < 0.05$.

The memory test upon dehydration stress in protein level showed different expression patterns than the transcript level of the *At3g29075* gene. The protein *At3g29075* expression pattern diminished in the S₁ stage when compared to untreated plants. Later on, the expression pattern increased for the S₂ and S₃ stage (Figure 23). The S₃ stage showed a significant level of increased protein expression than the S₁ stage.

3.8. Expression of *At3g29075* upon salt stress

To determine whether *At3g29075* expression is regulated in response to salt stress, *A. thaliana* plants treated with different concentrations of NaCl (100 mM, 150 mM, 300 mM). The samples were collected during different time laps (2 h, 8 h, 24 h) and later stored at -80°C. Gene expression analysis showed no significant differences in the expression of *At3g29075* under different salt stress conditions. As shown in Figure 24, salt stress has no substantial effect on the expression of the gene *At3g29075* in *A. thaliana*.

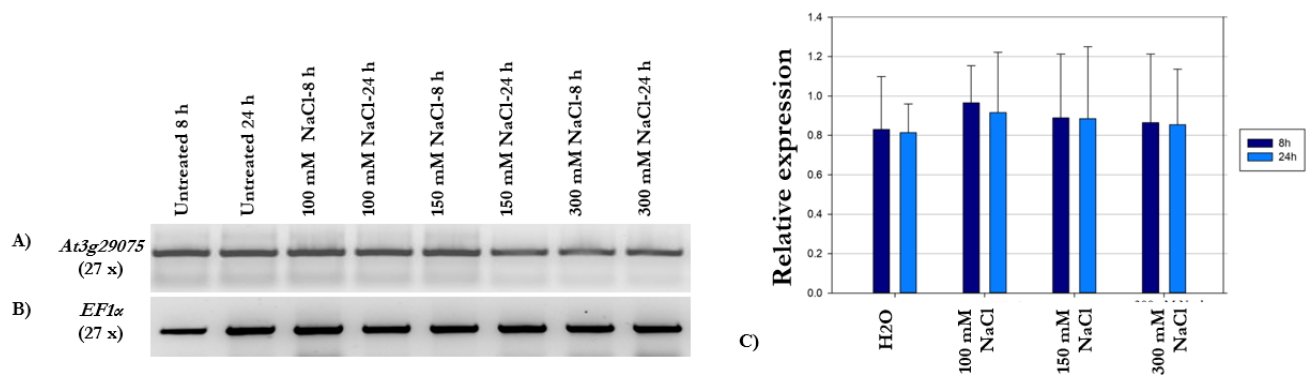


Figure 24: Expression of the *At3g29075* gene upon salt stress condition from *Arabidopsis thaliana* plants. A) RT-PCR of *At3g29075* upon salt stress condition. B) *EF1α* used as control. C) Mean values of relative expression levels of gene *At3g29075* in three technical repetitions and quantified by using the ImageJ program.

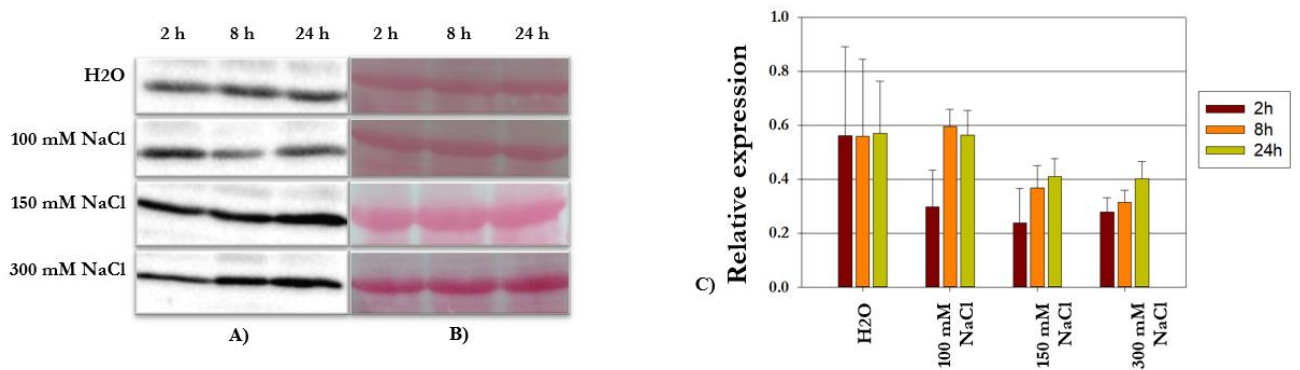


Figure 25: Expression of the protein *At3g29075* upon salt stress condition from *Arabidopsis thaliana* plants. A) Immunodetection of *At3g29075* protein upon salt-stress. B) Ponceau of rubisco protein used as a loading control. C) Mean

values of relative expression levels of protein At3g29075 in three technical repetitions and quantified by using the ImageJ program.

The protein blots revealed no significant level of At3g29075 protein expression in response to salt stress (Figure 25). The transcript and protein expression under salt stress illustrated opposite effects.

3.9. Analysis of promoter region

The phytohormone abscisic acid (ABA) is critical for plant growth and development and plays an essential role in integrating numerous stress signals and regulating downstream stress responses. Plants have to monitor ABA levels continuously in response to changing physiological and environmental conditions (Tuteja, 2007). Light and temperature are the vital abiotic modulators of plant gene expression (Soitamo et al., 2008). The ABA and light stress-responsive motifs observed in the promoter sequence of At3g29075 (Figure 26). In the present work, the plants treated with continuous light and dark conditions and in the presence of ABA stress treatment to examine the At3g29075 expression at the gene and protein level.

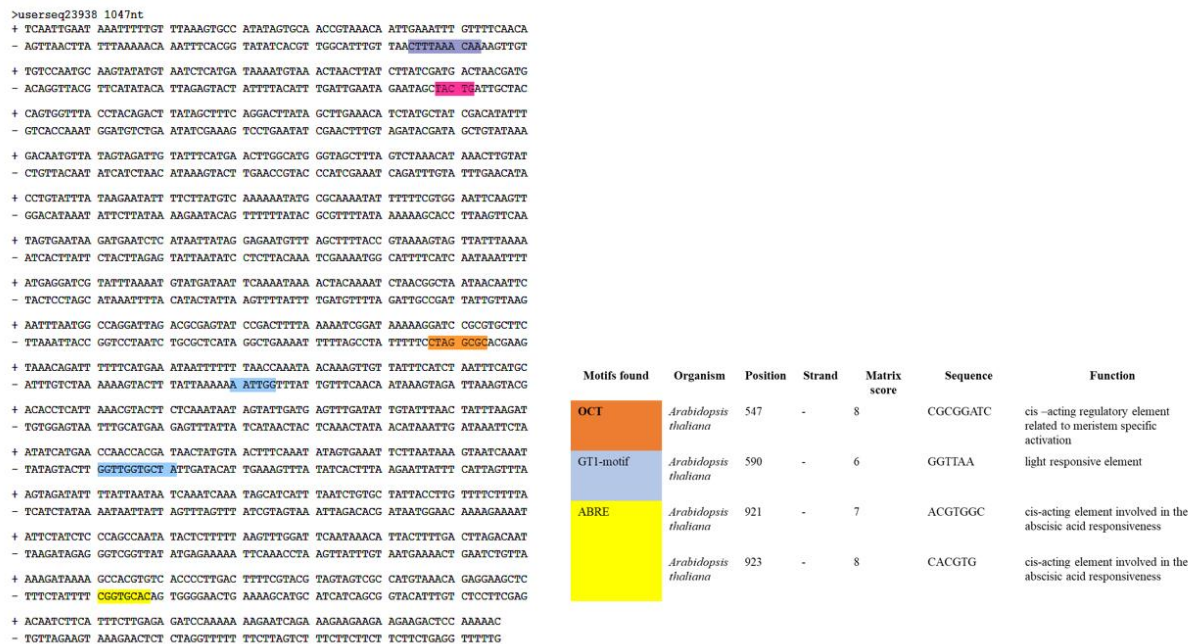


Figure 26: Promoter sequence of the At3g29075 (last) 1000 bp in *Arabidopsis thaliana*. The motifs found are colour-coded. **GT1-motif** - light-responsive elements and **ABRE** - cis-acting element involved in the ABA responsiveness.

3.9.1. ABA –Treatment

To investigate the effect of ABA application, four weeks old *A. thaliana* leaves transferred to the water (control). For the stress treatment, abscisic-acid (ABA) added to the distilled water (Final conc. = 100 μ M) incubated for 24 h & 48 h at a short daylight condition. After 24 h & 48 h, samples were frozen using liquid nitrogen and stored at -70°C . Later, the samples used to analyse the expression of At3g29075. There was a slight increase in the transcript level of the At3g29075 (Figure 27) under ABA stress condition. The At3g29075 protein expression was decreasing in response to ABA treatment which was observed from the protein blot (Figure 28).

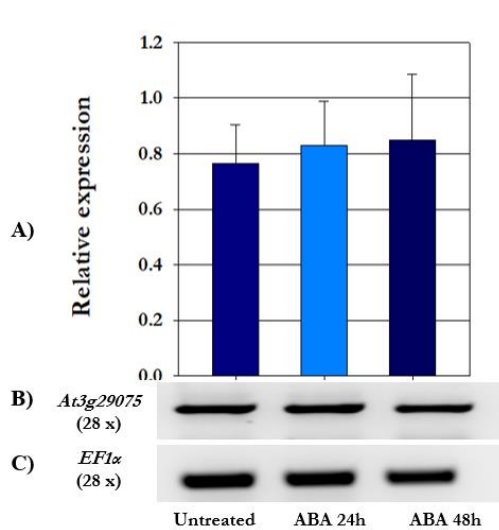


Figure 27: Expression of the At3g29075 gene in control and ABA-treated conditions. A) Mean values of relative expression levels of the At3g29075 gene in three technical repetitions and quantified by using the ImageJ program. B) 2 μ g total RNA extracted from *A. thaliana* plants used to analyse the expression of At3g29075 via RT-PCR. C) Expression of EF1 α transcripts used as a loading control.

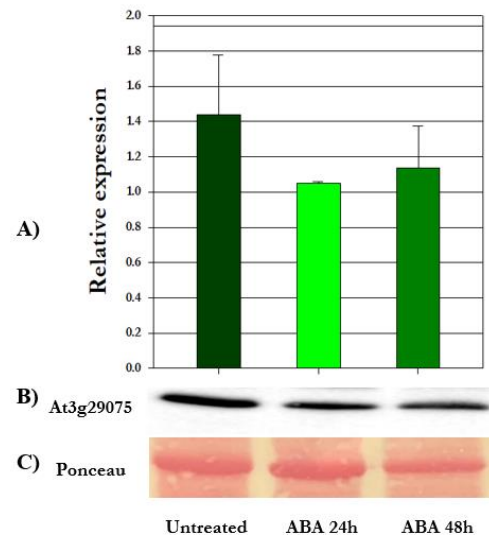


Figure 28: Expression of the At3g29075 protein in control and ABA-treated conditions. A) Mean values of relative expression levels of the At3g29075 gene in three technical repetitions and quantified by using the ImageJ program. B) Immuno-detection of protein At3g29075. C) Ponceau stain of Rubisco used as a loading control.

3.9.2. Expression of At3g29075 upon continuous light and dark condition

Four weeks old, soil-grown WT plants were used to determine the expression of At3g29075 under light-limiting conditions. While the plants kept at RT with continuous lights for 24-72 h, for the dark stress, plants held at the dark chamber at RT for 24-72 h for light treatment, the plants were kept in the dark chamber at RT for 24-72 h for imposing dark treatment. Plants were harvested after 24-48-72 h of stress treatment and stored at -70°C. Gene-specific primers and *EF1a* housekeeping primers used in RT-PCR to test the expression level in 27 cycles.

Light and dark stress conditions revealed different patterns of *At3g29075* gene expression in RT-PCR. However, the gene expression upon light and dark stress increased at 24 h than untreated plants. Subsequently, the gene expression level decreased at 48 h and 72 h upon the stress conditions (Figure 29).

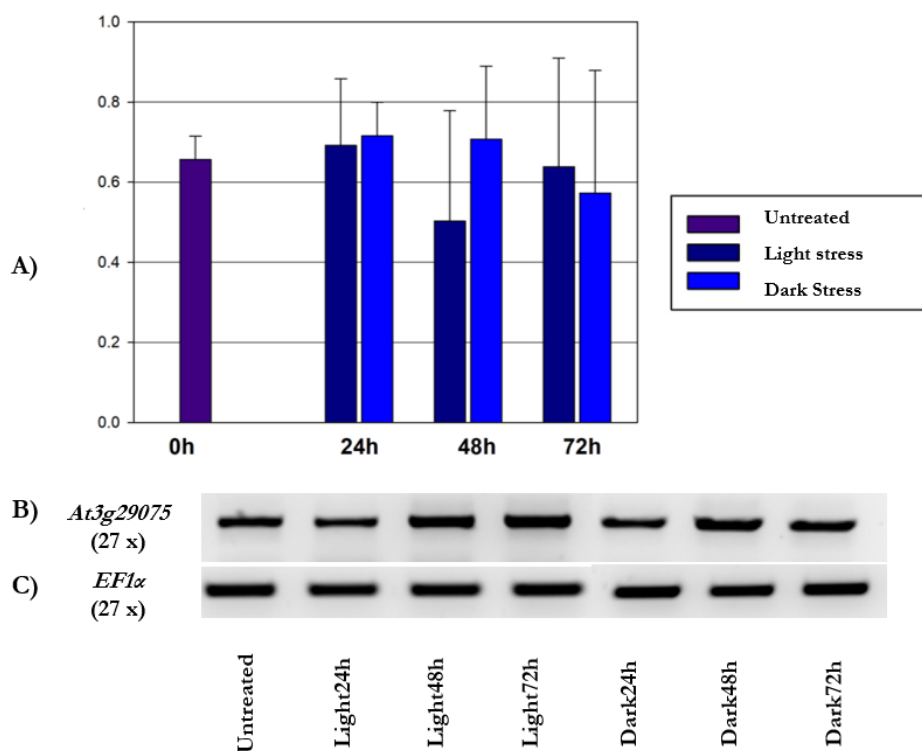


Figure 29: Expression of the *At3g29075* gene upon continuous, light, and dark conditions for 24 h, 48 h & 72 h. A) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using the ImageJ program. B) 2 μ g total RNA extracted from *A. thaliana* used to study the expression at the transcript level by RT-PCR. C) *EF1a* transcripts used as a loading control.

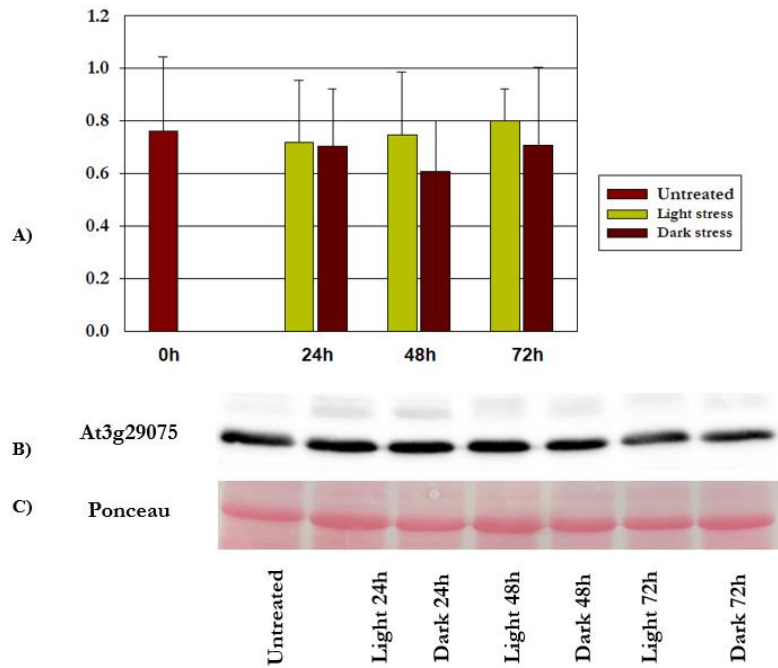


Figure 30: Expression of the *At3g29075* protein upon continuous, light, and dark conditions for 24 h, 48 h & 72 h. A) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using the ImageJ program. B) Immuno-detection of the protein *At3g29075*. C) Ponceau stain of Rubisco used as a loading control.

The protein blot illustrated that *At3g29075* protein expression was upregulated upon the continuous light. Whereas in dark stress, the expression pattern of *At3g29075* protein slightly downregulated when compared to the untreated and light stress samples (Figure 30).

3.10. Production of recombinant proteins

Several expression vectors and bacterial strains used in the process of making the recombinant *At3g29075* protein (Table 4). In most cases, the tandem repeat of the C-terminal part of the *At3g29075* protein was spliced out because of its toxicity (Figure 9, Supplementary figure-16,17,18,19,20). In a combination of Gateway, the cloning method with BL21 DE3 strain was the first successful clone with the full size of the *At3g29075* gene was achieved. However, the expression level of the recombinant protein was lower.

Table 4: List of vectors and bacterial strains used for the production of recombinant protein *At3g29075*; source of clone; different tags expression vector system contains, and the result was documented in this table.

VECTOR	SOURCE	TAG	BACTERIAL STRAIN	RESULT
pET28a	From cDNA	His-6x	DH10B	237 bp shorter
	From pGJ280+At3g29075	His-6x	DH10B	159 bp shorter
pET43b	From pGJ280+At3g29075	His-6x	DH10B	232 bp shorter
pGEX-4T-2	From pGJ280+At3g29075	GST	DH10B	109 bp from the pGJ280 inserted in the gene
pQLinkHD	cDNA+At3g29075	His-6x	DH10B	110 bp shorter
	cDNA+At3g29075	His-6x	BL21 (D.E3)	885 bp full size. Protein expression was low
	cDNA+At3g29075	His-6x	T7 express cells	Low expression
	cDNA+At3g29075	His-6x	RIPL	
	cDNA+At3g29075	His-6x	C43	
pTwin1	From pGJ280+At3g29075	No-tag	DH10B	Not possible to clone
	From pGJ280+At3g29075	No-tag	BL21	
	From pGJ280+At3g29075	No-tag	T7 express cells	
	From pGJ280+At3g29075	No-tag	RIPL	
pEarlygate103	cDNA+At3g29075	GFP	DH10B/ <i>Agrobacterium</i>	NO GFP was expressed

The recombinant protein At3g29075 from the successful pQLinkHD expression system in the BL21 DE3 strain was used for lipid-binding assay and liposome binding assay and other experiments. The rest of the cloning results are documented under supplementary data.

3.10.1. Amplification and cloning of At3g29075 into pQLinkHD

In a first step, an amplicon of 885 bp, containing the complete coding sequence of *At3g29075*, was amplified by RT-PCR. By using the gateway adoptive primer At3g_attB1 (Forward primer) and At3g_attB2 (Reverse primer), the attB sites were introduced into the transcript. The attB-PCR product cloned into the entry gateway system pDONR™201 (Supplementary figure 10) with BP CLONASE™ enzyme mix. The gene At3g29075 from the entry clone into destination vector pQLinkHD (Supplementary figure 11) with LR

CLONASE™ enzyme (Figure 31). Successful cloning confirmed by sequencing with new pQEFW and pQE 276 primers. No other changes observed in the At3g29075 sequence (supplementary figure 23).

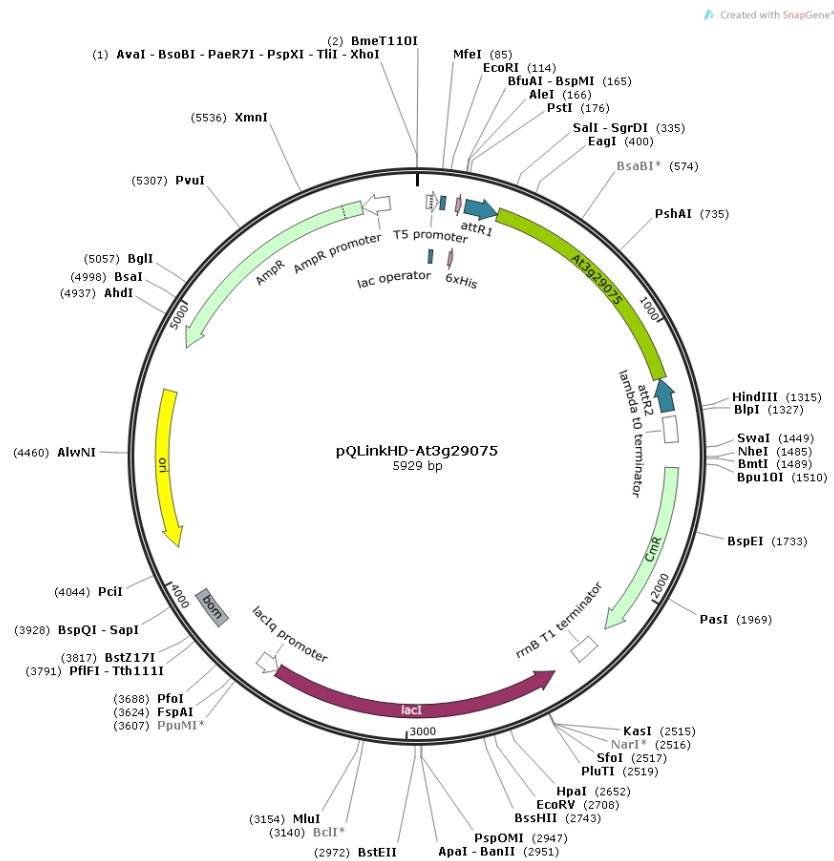


Figure 31: Expression vector pQLinkHD - bearing the His-tagged fragment At3g29075. The green arrow represents the full-size At3g29075 -fragment (885 bp) that fused on the 5'-end to six consecutive histidine residues. The resulting fusion protein consists of 347 amino acids.

3.10.2. Expression and isolation of At3g29075

The sequenced constructs pQLinkHD - At3g29075 transformed into *E. coli* BL21 cells. The T5 promoter induced by the application of IPTG (final concentration 1 mM). The induced At3g29075 protein level was significantly lower than the standard expressing proteins (Dumon-Seignovert et al., 2004). Therefore, pET-28 with Cp LEA-like 11-24 protein in BL21 strain obtained from Prof. D. Bartel's lab was used to test the induction process and by using it as a positive control. Expression of Cp LEA-like 11-24 protein marked with a single line arrow and the expression of At3g29075 marked in green arrow (Figure 32A). For a detailed analysis of full-size At3g29075, small-scale samples (1 mL)

taken and processed, as described in section 2.1. Soluble protein fractions (S0, S1, and S2 & S3) and insoluble proteins (P0, P1, and P2 & P3) separated on SDS-PAGEs (Figure 32B). There was no high level of protein expression found in soluble and insoluble fractions. Therefore, the induction time points increased to 6 h, 8 h, 12 h, and 24 h for At3g29075 protein. Protein samples from pQLinkHD Empty and pQLinkHD-At3g29075 collected and loaded to the gel and analysed by Coomassie Staining (Figure 33A). Immunodetection revealed that 8 h protein samples have small-scale induction of protein At3g29075 (Figure 33B). After the 8 hours of induction, the protein started to denature. Hence, 12 h and 24 h induction samples were not used further for protein purification.

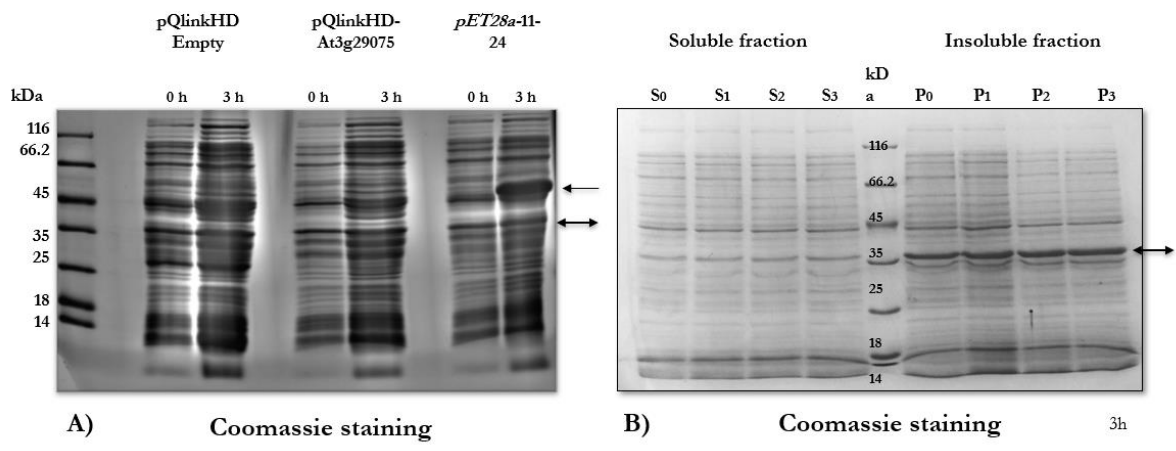


Figure 32: Induction of the full-size protein At3g29075. A) Single line arrow: 11-24 as a control to show the standard level of protein induction; double line arrow: At3g29075 protein. Induction of bacteria for 3 h with empty pQLinkHD plasmid and pQLinkHD with At3g29075 and pET28a with Cp LEA-like 11-24 protein. Not-induced (0 h) and induced (3 h) fractions. B) Detailed analysis of protein At3g29075 synthesis upon induction. S0-S3: Soluble protein fractions before and after induction. P0-P3: Insoluble protein fractions before and after induction. Double line arrow: At3g29075 protein.

The primary culture of the soluble protein fraction prepared as per section 2.13.1. As shown in Figure 34, the soluble protein fraction F0 contains total proteins. After that, a mild washing step yields Ft fraction containing unbound protein fraction of the column. The At3g29075 recombinant protein fraction eluted from the column in six different fractions F1-F6 loaded on the gel. The silver staining indicates the 33kDa protein Figure 34A, and the immunoblot confirms the presence of At3g29075 protein in the fraction F1-F6 (Figure 34B). Later on, the samples desalted by using a PD10 column (GE Healthcare)

by following the manufacturer's instruction. The protein samples were freeze-dried and stored at -20°C for further experiments.

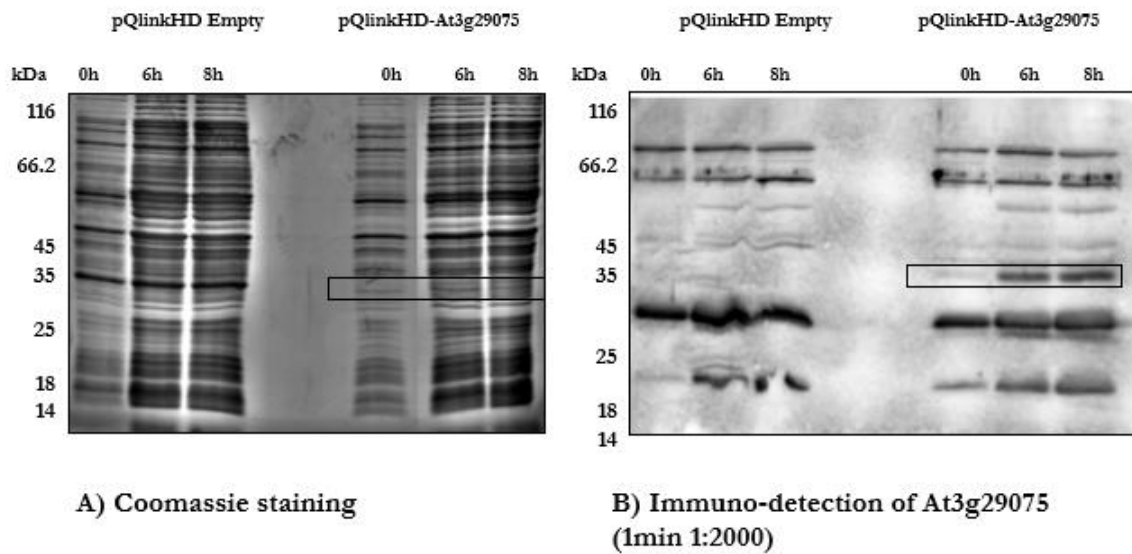


Figure 33: Induction of the full-size protein *At3g29075* upon different time intervals (6-8 h). A) Protein samples from *pQLinkHD Empty* and *pQLinkHD-At3g29075* on Coomassie Staining. B) Immuno-detection of *pQLinkHD Empty* and *pQLinkHD-At3g29075* shows levels of protein induction marked in the red box.

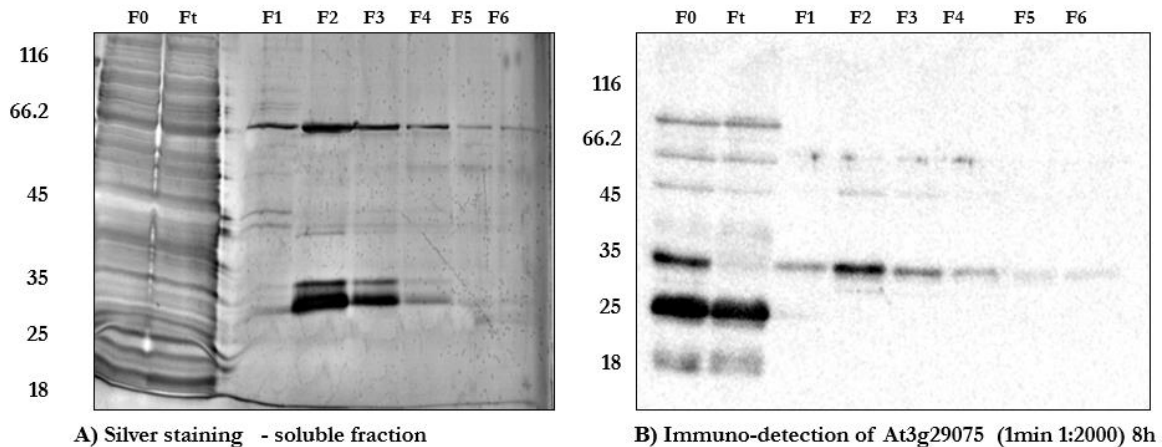


Figure 34: His-tag affinity chromatography of *At3g29075*-full-size of soluble protein fractions. A) Silver staining. B) Immuno-detection. F0: Total proteins before loading onto the column; Ft: Flow-through; F1–F6: Eluted His-tag fractions.

3.11. Cloning of Nt-*At3g29075* into *pQLinkHD*

The Nt-*At3g29075* (381 bp) was amplified from cDNA by RT-PCR with attB sites with *At3g_attB1* and *At3gN_attB2* primers. It was cloned into *pDONRTM201* entry gateway vector with the *BP CLONASETM* enzyme mix. The gene Nt-*At3g29075* sub-cloned from

the entry clone into destination vector pQLinkHD with LR CLONASE™ enzyme (Figure 35). Successful cloning confirmed by sequencing it with newpQEFW and pQE 276 primers.

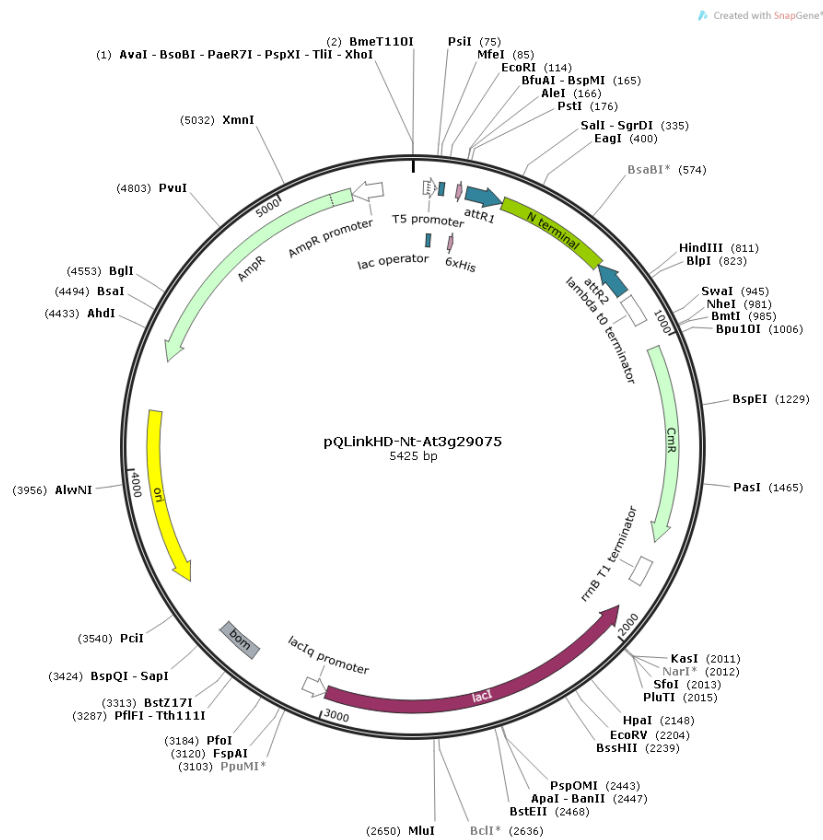


Figure 35: Cloning strategy of the N-terminal *At3g29075* gene in pQLinkHD expression system. Conceptual graphic of Gateway Expression vector pQLinkHD with Nt-*At3g29075* protein.

3.11.1. Purification of Nt-At3g29075 recombinant protein

The pQLinkHD bearing Nt-At3g29075 fragment transformed into *BL21* strains. The Nt-At3g29075 induced for 3 h, as per section 2.13. The primary culture soluble fraction of Nt-At3g29075 used for protein purification as per section 2.13.1.

The Silver staining gel (Figure 36A) picture exhibits the F0 total protein fraction. Ft- the eluted fraction from mild washing of the column, has unbound proteins. F1 to F6 fraction represents the eluted Nt-At3g29075 protein fractions. The immunoblot shows the confirmation of Nt-At3g29075 purified fraction F1-F6 (Figure 36B). The eluted protein fractions F1-F6 pooled together, and then the PD10 column (GE Healthcare) was used to change the buffer. The Nt-At3g29075 protein sample was freeze-dried and stored at -20°C.

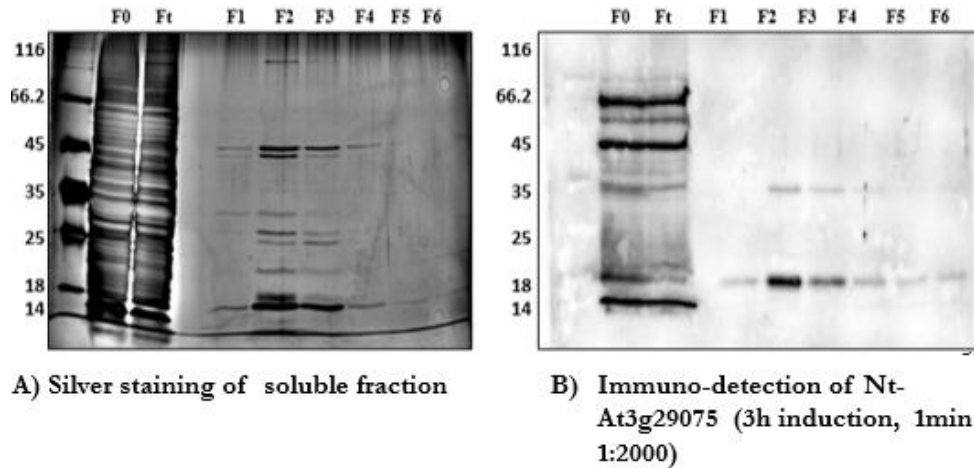


Figure 36: His-tag affinity chromatography of the Nt-At3g29075 of soluble protein fractions. A) Silver staining; B) Immuno-detection of the Nt-At3g29075 protein with Nt- PLDrp1.1 antibody. F0: Total proteins before loading onto the column; Ft: Flow-through; F1–F6: Eluted His-tag fractions.

3.12. Cloning of Ct-At3g29075 into pET28a

This construct containing Ct-At3g29075 (504 bp) cloned into pET28a (Nasr, 2015) for the antibody production, the same construct used for Ct-At3g29075 protein purification and used for lipid binding assay (Supplementary figure 24).

3.13. Lipid-binding

3.13.1. Protein-lipid interactions on nitrocellulose membranes

The protein-lipid-overlay assay is the more rapid method to validate the binding of a target protein to a range of various lipids. In addition to the direct interaction between PLD α 1 and PLDrp1.1 suggests (Sachetto-Martins et al., 2000; Ufer et al., 2017), an interaction of both proteins via phosphatidic acid (PA).

Interestingly a highly similar protein (At3g29075) that shares up to 49% sequence similarity within the N-terminal part of the protein was identified in *A. thaliana*. In this work, the binding of At3g29075 to phosphatidylcholine (PC), phosphatidic acid (PA) tested. The lipid-protein interaction was visualised using Nt-PLDrp1 antibodies directed against the N-terminal part target At3g29075 proteins and with the At3g29075 - specific antibody was

carried out to identify lipid binding of the full-size protein and the C-terminal fragment of At3g29075.

A strong signal for the full-size- At3g29075 protein binding to PA was detected (Figure 37A). While the N-terminal fragment of PLDrp1.1 (Figure 37B) displays slightly decreased signal strength for PA binding, the C-terminal fragment (Figure 37C) reveals a very weak affinity to PA. The antibody raised against the C-terminal part of At3g29075. As a result, it cannot identify the N-terminal protein part. Thus, the Nt-PLDrp1 antibody used to detect the lipid-binding of the N-terminal fragment of At3g2907.

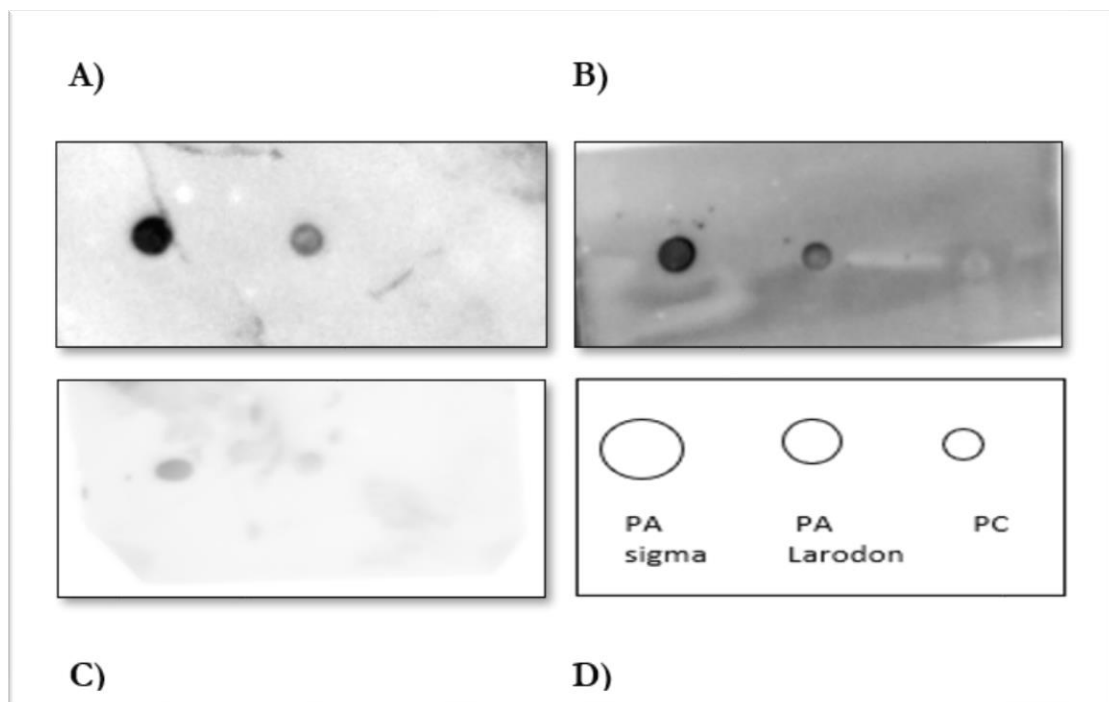


Figure 37: Detection of the At3g29075 protein fragments in protein-lipid-overlay assays. The Ct-At3g29075 antibody (A & C) and the Nt-PLDrp1 antibody (B) used, and blots detected for 1 min. A) Full-size At3g29075 protein; B) N-terminal fragment of the At3g29075 protein; C) C-terminal fragment of the At3g29075 protein; D) Loading scheme of lipids used for all proteins.

To summarise, the At3g29075 protein can bind to PA. A similar activity was observed for PLDrp1 protein with PA interaction (Ufer et al., 2017).

3.13.2. Liposome-binding assay

In contrast to ex-situ protein-lipid-overlay assays on nitrocellulose membrane, the cellular state of lipids as liposomes were taken into account for liposome-binding assays. Therefore, it imitates a more natural environment of lipids. Lipid binding depends on a

variety of factors, such as protein conformation and the pH of the environment (Ruano et al., 1998). Liposome-binding assays performed, as illustrated in section 2.17.2. However, unbound proteins spotted in the supernatant, the liposome-bound protein complexes held in the pellet fractions.

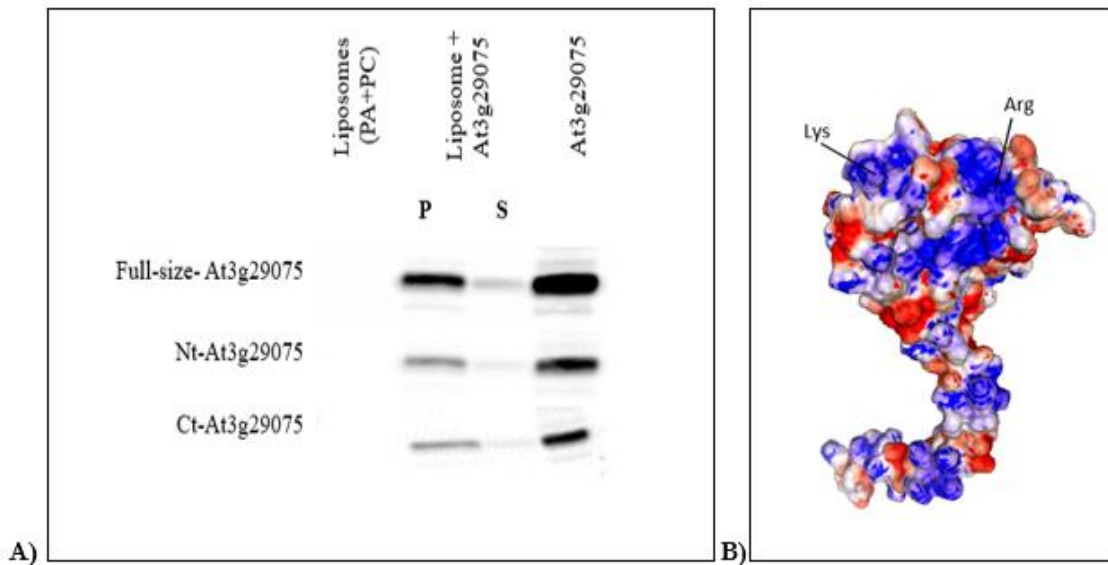


Figure 38: Liposome-binding assay with the At3g29075 protein. A) Binding of PA to At3g29075, Nt & Ct- At3g29075 tested. Pure protein and pure liposomes used as binding controls. S: Supernatant, P: Pellet. B) Electrostatic surface model of the At3g29075 protein. Arginine and lysine clustered in the glycine-rich domain of the protein hypo a positive charge. Source: Charmm.

To verify the earlier witnessed PA binding to At3g29075 protein and the N-terminal and C-terminal region of At3g29075 protein, further PA-binding assays carried out via liposome binding assay. The purified recombinant protein samples dissolved and retained in the supernatant, indicating that the proteins dissolved. Protein-specific antibodies in the western blot cannot detect the liposomes; therefore, liposomes applied as a negative control. Immunoblot revealed strong binding of PA to the full-size At3g29075 protein. However, decreased binding to the N-terminal At3g29075 fragment observed. In addition to that, only faint binding found for the C-terminal At3g29075 protein. Besides, the C-terminal At3g29075 protein showed a weak signal in the pellet fraction (Figure 38A), indicating a lack of solubilisation in the protein fraction. Complete solubilisation in the C-terminal protein was not possible. There are 7 out of 8 arginines and 57 out of 66 lysines are located at the positively charged electrostatic surface model of the At3g29075 protein (Supplementary figure 31 C). There are -20 lysine clusters (Supplementary figure 31 B, 32) were formed by 57 lysine amino acids (Clusters are calculated based on Zheng, 2004) and

therefore it draws the negatively charged liposomes and formulates the liposome-binding (Figure 38B).

3.14. The generation of the At3g29075 over-expression line

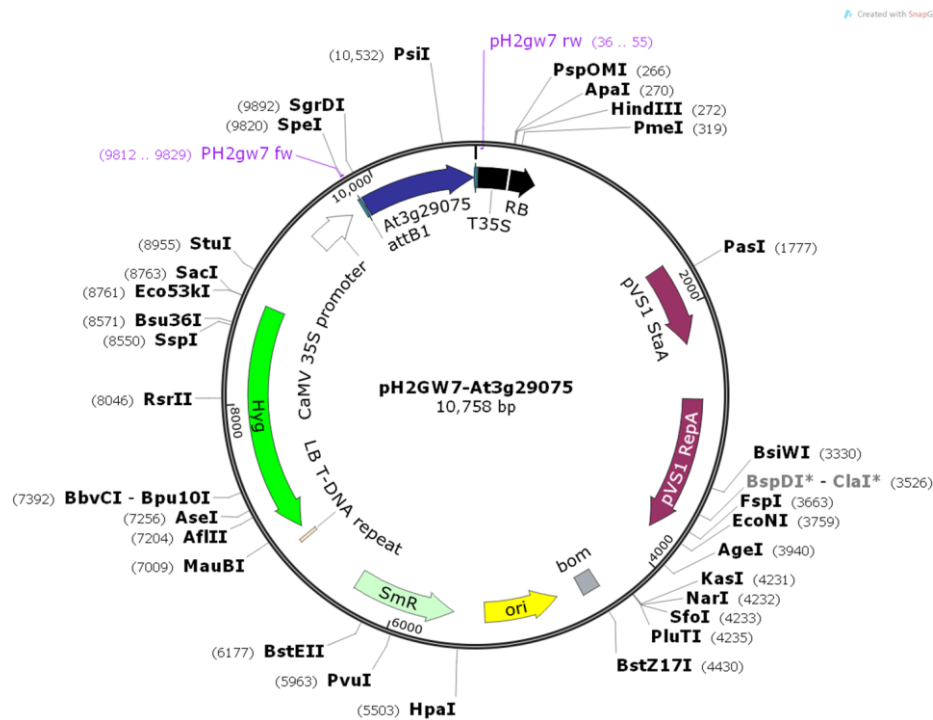


Figure 39: Construct for the generation of the At3g29075 over-expression line.

One of the factors that can lead to severe defects in the growth and development of plants is over-expressing the gene. Thus, it can help to reveal the function of a gene or protein.

The cloning process initiated by using cDNA from *A. thaliana* as a template for mutagenesis PCRs, thereby generating a full-size fragment of At3g29075 with At3g_aatB1 and At3g-aatB1 primers. One μ L of the PCR product (885 bp) was cloned into the pDONR201 entry vector with BP reaction according to the instruction of the Thermo scientific Gateway cloning Kit. Further, sub-cloning of the At3g29075 gene into binary. Gateway expression system pH2GW7 (Figure 39) containing 35S-promoter through LR reaction mix (Thermo scientific Gateway cloning Kit). The final construct was checked by DNA sequencing with At3g_aatB1 and pH2gw7rw primers. Later, the construct transformed into *Agrobacterium tumefaciens*. Subsequently, the construct in the *Agrobacterium tumefaciens* was transformed into *A. thaliana* via the floral dip method.

3.14.1. Selection of Hygromycin-resistant transformants

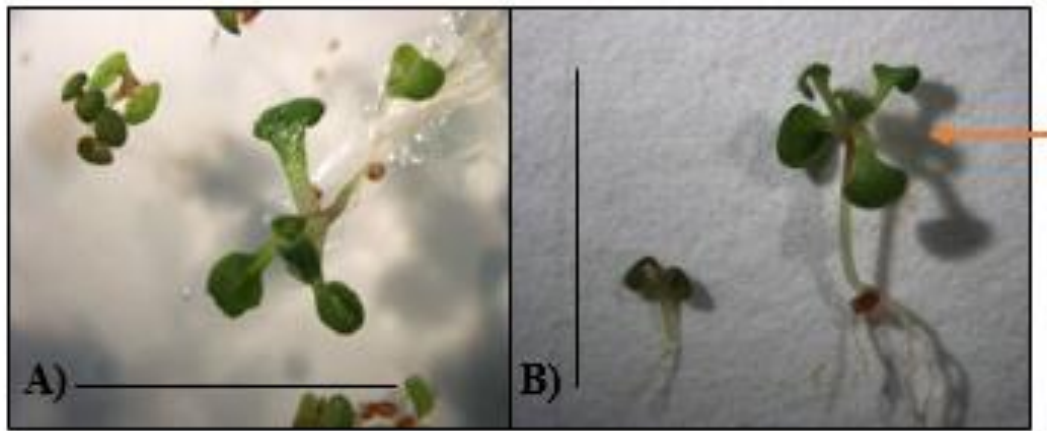


Figure 40: Rapid selection of the pH2GW7-AT3G29075 overexpression line by Hygromycin B. Col-0 after rapid selection procedure (left); Hygromycin B-resistant Arabidopsis seedling after rapid selection procedure (on the right side marked with an arrow). Size bars 1 cm.

In the presence of Hygromycin-B, Hygromycin resistant seedlings showed a different morphology to those grown without any resistance (Harrison et al., 2006). After the second day in the dark on medium containing hygromycin B ($15 \mu\text{g mL}^{-1}$), hygromycin-resistant transformants showed longer hypocotyls of approximately 0.8–1.0 cm, whereas non-transformants had short hypocotyls (0.2–0.4 cm) (Figure 40). However, the selection of hygromycin-resistant seedlings from non-transformants easily achieved by selecting those with elongated hypocotyls. To validate, the phenotypes of seedlings grown in the presence of hygromycin B, wild-type controls grown alongside.

The majority of characterised hygromycin-resistant seedlings (pH2GW7-AT3G29075) had long hypocotyls following rapid selection. Additionally, PCR was done to confirm the presence of the transgene in 24 seedlings with longer hypocotyls. No hypocotyl elongation observed in non-transformed Col-0 controls in the presence of Hygromycin. Positive plants (F₁, F₂ and F₃ generations) phenotypically monitored.

3.15. Phenotypic analysis of mutant plants

3.15.1. F₁–Generation



Figure 41: Phenotype analysis of the *AT3G29075* over expression F₁ generation line. A) Wild type and B) *AT3G29075* overexpression line of the same age showing early flowering.

The *A. thaliana* over-expression lines with Hygromycin resistance grown on the soil in a short-day condition. Only three out of 24 screened *AT3G29075* overexpression lines survived. All the plants monitored on a daily base. The overexpression lines of *AT3G29075* showed early flowering (Figure 41B). The seeds collected from these lines.

3.15.2. F₂–Generation

The F₁ generation seeds of *AT3G29075* overexpression lines screened with the rapid selection method of hygromycin. Selected *AT3G29075* overexpression lines (F₂ generation), wild type and knockdown mutant *at3g29075* (F₂) were grown on soil in short day room conditions. The overexpression lines of *AT3G29075* showed the first visible early flowers (Figure 42A, B). The *AT3G29075* overexpression line shows magnificent development in flowering (Figure 42C) while comparing with wild type and knockdown *at3g29075* mutants. This study conducted for 50 individual plants per overexpression line and all of them showed early flowering.

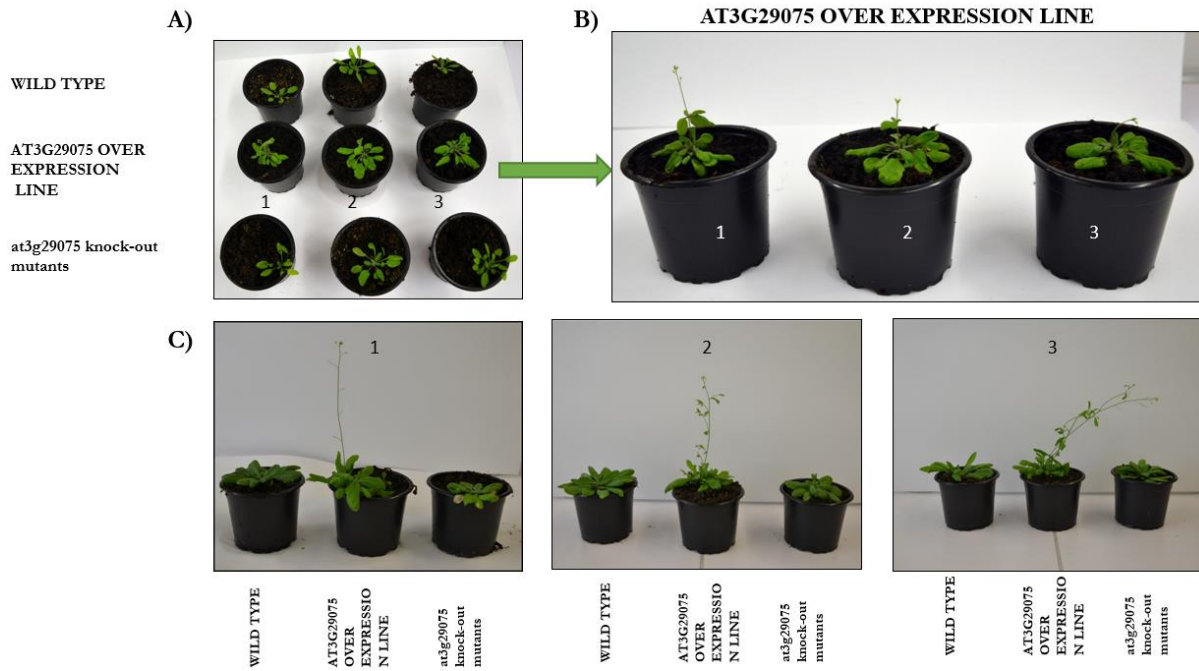


Figure 42: Phenotype analysis of the *AT3G29075* over expression F₂ generation line. Wild type, *AT3G29075* overexpression line, and knockdown *at3g29075* grown on soil in short-day condition. Figure A&B shows the early flowering in the *AT3G29075* overexpression line. C) Shows the development in the flowering of the *AT3G29075* overexpression line.

3.15.3. F₃–Generation

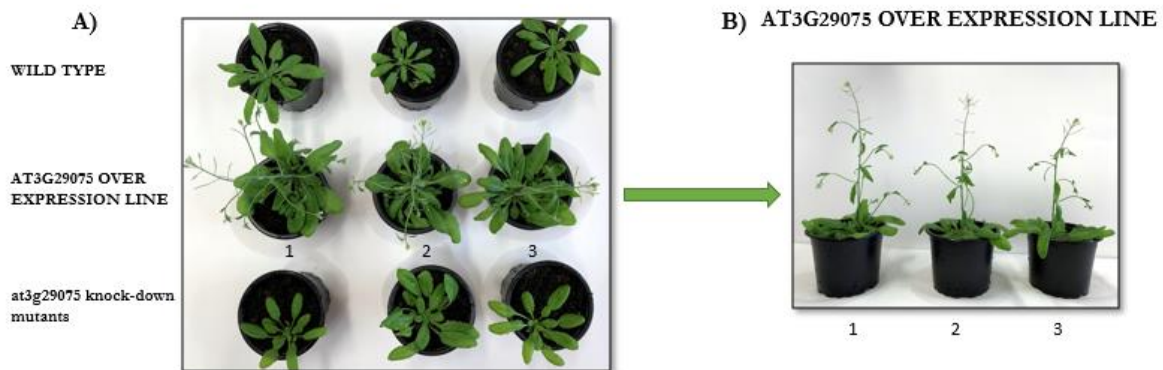


Figure 43: Phenotype analysis of the *AT3G29075* over expression F₃ generation line. Wild type, *AT3G29075* overexpression line, and knockdown *at3g29075* were grown on soil in short-day conditions. A) Shows the early flowering in the *AT3G29075* overexpression line. B) Indicates the development in the flowering of the *AT3G29075* overexpression line.

The hygromycin rapid selection method was used to screen seeds from F₂ *AT3G29075* Overexpression lines. The WT, F₃ generation of knockdown *at3g29075* mutant and F₃ generation Overexpression *AT3G29075* line were grown in the soil in short-day conditions. The *AT3G29075* Overexpression lines (F₃) exhibited the early flowering

(Figure 43B); however, the WT and knockdown mutant (F₃) didn't show any difference in the phenotype (Figure 43A).

3.16. At3g29075 protein profiling in wild type and mutants

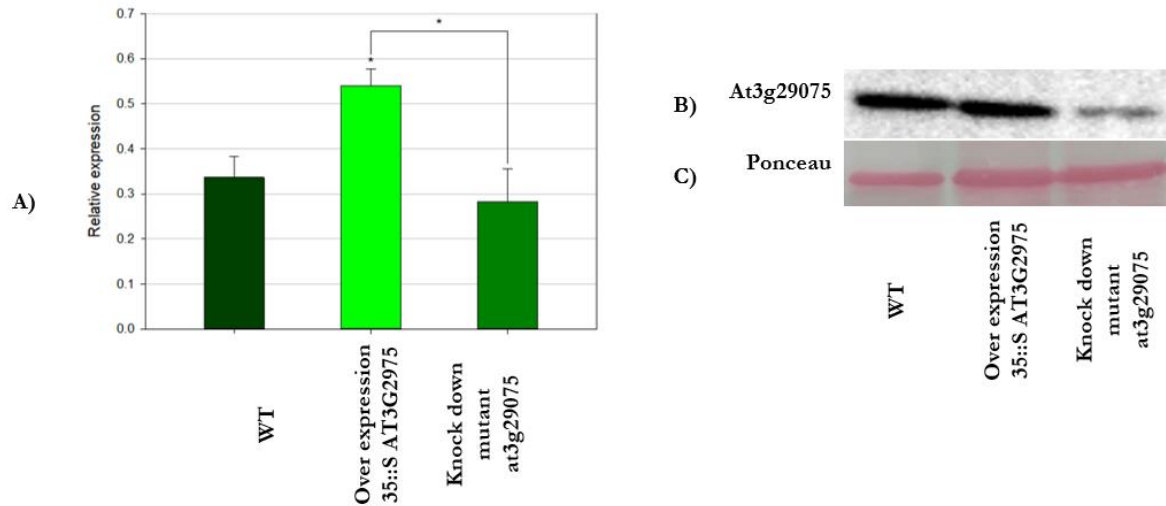


Figure 44: Expression of the *At3g29075* protein level in wild type and mutants. A) Mean values of relative expression levels of the *At3g29075* gene in three technical repetitions and quantified by using the ImageJ program. The star indicates the levels of significance in comparison to the untreated sample (one-way ANOVA, Holm-Sidak method): * $p < 0.05$. B) Immuno-detection of protein *At3g29075*. C) Ponceau stain of Rubisco used as a loading control.

To evidence, the difference in the protein expression level affects the phenotype of the plant, the wild type, *at3g29075* knockdown mutant (F₂ generation), and *AT3G29075* overexpression line (F₂ generation) grown on short-day conditions for four weeks. The total protein content extracted from the plant samples followed by protein estimation via Bradford assay. The 5 μ g of total protein loaded for three biological replicates on the gel.

The western blot showed that the *AT3G29075* overexpression line has the highest level of *At3g29075* protein expression. However, a significant level of protein *At3g29075* expression pattern increased in the *AT3G29075* overexpression line (Figure 44). The *at3g29075* knockdown line showed a lower expression of protein than the wild type plant.

3.17. Intracellular localization of At3g29075 - Transient transformation

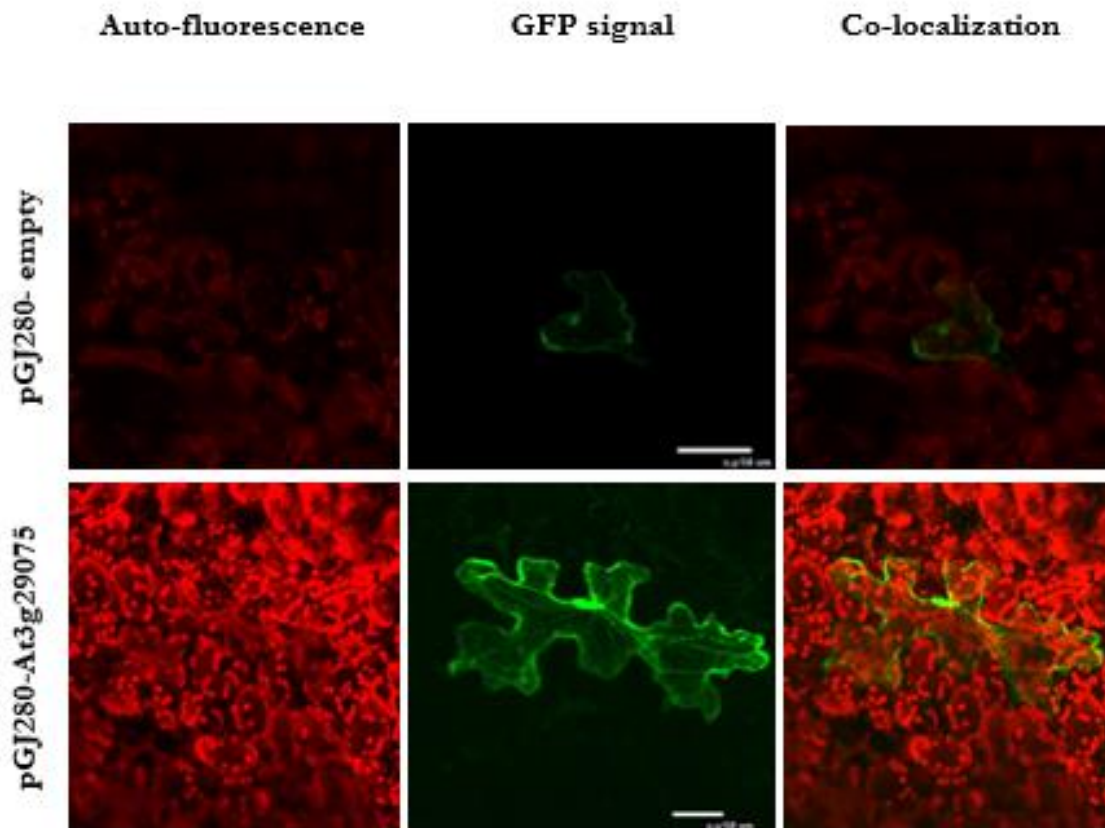


Figure 45: Microscopic analysis of the intracellular localisation of the *At3g29075* gene after particle bombardment. Top panels display the localisation of free GFP from the empty vector pGJ280. The panels at the bottom show the localisation of *At3g29075*. The white bar represents 50 μm .

A fusion construct was expressing *At3g29075*-GFP generated to investigate its intracellular localisation (Nasr, 2015). There were no clear pictures have taken to manifest the identification of the intracellular localisation of *At3g29075*, but the microarray data indicated a cytosolic expression of the protein. To validate these findings, the construct pGJ280-*At3g29075* used in a repetitive transient transformation assay by particle gun bombardment (see section 2.18.1).

The expression of pGJ280-*At3g29075* elicits in resembling patterns like the free GFP that localised in cytoplasm and nucleus of plant cells (Tsien, 1998). Based on bioinformatics predictions, a wide subcellular distribution with an accumulation of *At3g29075* in the cytosol and nucleus detected (Figure 45).

4. Discussion

The objective of this study was to characterise the unknown gene and the corresponding protein At3g29075 in *A. thaliana*. The expression of the At3g29075 was upregulated depending on the dehydration conditions. The At3g29075 protein has a very unusual amino acid composition. Structure predictions of the At3g29075 protein propose that it belongs to the disordered proteins. Further results illustrate the regulation of At3g29075 by phosphatidic acid, the principal product of the PLD-pathway. Protein overexpression studies indicate a possible involvement of the At3g29075 in early flowering processes.

4.1. Gene analysis of At3g29075

The gene At3g29075 is located on chromosome 3 of *A. thaliana*. The gene model consists of two exons with 851bp of coding sequence (CDS), which results in 294 amino acid residues (Figure 5). The *PLD α 1* gene is located in chromosome 5; it contains transcription region of 1559 bp coding for two exons separated by an intron close to the C-terminal end (Supplementary figure 29). The PLD α 1 protein consists of 381 aa. According to the TAIR database, the PLD α 1 gene is the best match candidate with 60% sequence similarity to the At3g29075 gene. The C-terminal part makes the PLD α 1 and At3g29075 unique; however, the N-terminal part consists of 49% sequence similarity (Supplementary figure 25).

4.2. At3g29075 is intrinsically disordered

Based on the unexpected (Wootton, 1994) amino acid composition (Lys-21.1%, Asp-19.4%) and structure of the At3g29075 protein, it can be categorised under Intrinsically Disordered Proteins (IDPs) (Van Der Lee et al., 2014). The At3g29075 protein has nine imperfect tandem repeats in its C-terminal region (Figure 8). In addition to that, the order/disorder prediction (DISCO-*) from “*RaptorX property*” website showed all the 294 aa (100%) of At3g29075 are disordered (Supplementary figure: 27). Jorda et al., (2010) analysed 51,685 repeats of 33,151 proteins, which represent 9.1% of all the protein in the Swissport release of January 2009 (364,403 sequences); the abundance of predicted intrinsic disorder in these sequences were analysed with several computational tools. The results of this analysis are, tandem repeats repeat-containing proteins are highly disordered. Furthermore, tandem repeats have a higher percentage of disordered residues than the entire tandem repeat-containing sequences. The At3g29075 protein was detected at a MW of ~35 kDa by using immunoblot analysis (the theoretical MW is 34.5 kDa). However, the

immunoblot analysis of the PLDrp1 total protein and phosphoprotein extracts exhibited at MW of ~51 kDa in contrast to the calculated theoretical MW (theoretical MW 43.5 kDa) (Ufer, 2015). Even though, the PLDrp1 and the At3g29075 proteins share 60% of sequence similarities and are also classified under intrinsically disordered protein class, due to their tandem repeats, they displayed altered migration speed of the protein in SDS-PAGEs.

Depending on their unique amino acid composition, intrinsically disordered proteins bind less SDS than structured proteins and often migrate at a 1.2-1.8-fold higher molecular weight than calculated from their sequences (Tomba, 2003). In addition to that, both the PLDrp1 and the At3g29075 proteins exhibit additional characteristics of IDPs, such as phosphorylation sites (Laugesen et al., 2006). The post-translational modification might be controlled via IDPs, thus results in disorder-to-order transitions (Zhang et al., 2013). Phosphorylation, as a regulatory switch of these conformational stages, has been observed for IDPs (Bah et al., 2015). Due to the PLDrp1 protein aggregation under *in vitro* condition, the circular dichroism spectroscopy could not detect a clear structure of the PLDrp1 protein (Ufer, 2015). Evidence for IDPs involved in protein aggregation has been described earlier (Zhang et al., 2013) and supports the classification of the PLDrp1 as intrinsically unstructured. An electrostatic surface model of the At3g29075 protein was predicted by the Charmm - Bioinformatic tool (Figure 38). However, the production of recombinant At3g29075 protein was insufficient to attempt the prediction of the At3g29075 protein's clear structure using circular dichroism spectroscopy. Optimising the At3g29075 recombinant protein production in the future, might aid the possibility to study the structure of the At3g29075 protein under in-vitro condition.

4.3. Phosphorylation of At3g29075

The At3g29075 protein's isoelectric point (pI) was predicted to be 6.02 and holds the phosphorylation site at position S₁₇₉ (Cruz et al., 2019; Bhaskara et al., 2017; Walton et al., 2016). Thus, providing primary evidence of phosphorylation site of the At3g29075 protein. Whereas the PLDrp1 protein consists of very acidic (pI 4.7), and carries a phosphorylation site at position S₃₃₉ (Laugesen et al., 2006). The study of phospho-enriched proteins verified the phosphorylation of the PLDrp1 protein via a VH1 kinase (Braun, 2017), and a CPK3 kinase (Mehlmer, 2008); it also suggests non-specific activation of the PLDrp1 protein by different kinases (for example, based on the developmental stage) in *A. thaliana* (Ufer, 2015). The most common post-translational modifications of proteins are phosphorylation, and it also plays a vital role in several cellular processes, regulating and

signalling networks (Cooper, 2019). Therefore, it is possible that the phosphorylation of the PLDrp1 influences its function and affects downstream targets (Ufer, 2015). The dehydration stress-triggered phosphorylation of the PLDrp1 protein (Ufer, 2015). Upon dehydration stress the At3g29075 protein demonstrated upregulation of the protein expression in wild-type plants (Figure 21). Mutating the phosphorylation site S₁₇₉ might alter the function of the At3g29075 protein and it might help to recognize its physiological role. Further studies based on the phosphorylation state of the At3g29075 in different environmental conditions and various tissues in wild type and AT3G29075 over expression and at3g29075 knockdown mutant plants possibly provide additional knowledge.

4.4. Regulation of At3g29075

The amount of any protein synthesised in the cell was regulated by the level of gene expression (Cooper, 2019). A further level of control is then obtained by regulation of protein function, which allows the cell to regulate not only the amounts but also the activities of the proteins (Cooper, 2019). The PLD α 1 regulates the PLDrp1 protein; protein and transcript levels of the PLDrp1 depends on the presence of PLD α 1 (Ufer, 2015). However, no direct physical interaction was found. Possibly, the PLDrp1 is regulated by PA, the enzymatic product of PLD α 1 (Ufer et al., 2017). Similarly, the protein At3g29075 might be regulated by members of the PLD family via binding to their principal product PA. *In silico* analysis of co-expression patterns have revealed a closely associated co-expression of the PLDrp1 and the At3g29075 proteins (Figure 11). The co-expression of two genes specifies complementary transcriptional regulation, a potential functional relationship, or involvement in the same protein complex or pathway (Weirauch, 2011). Hence, it empowers further research on the connection between the PLDrp1 and the At3g29075 proteins.

4.5. Expression of At3g29075 during different developmental stages correlated to phosphatidic acid

Previous studies illustrated that no direct interaction between the PLDrp1 and the PLD α 1 was detected (Ufer et al., 2017). The PA is an essential signalling molecule in plants that regulates a wide range of proteins and associated pathways (Munnik, 2001). The PLDrp1 protein was characterised under downstream regulation of the PLD α 1 via PA binding (Ufer et al., 2017). When it represents to the PLD family members, PLD α 1 is the most abundant member and thus results in the determined PA production under non-stress conditions (Wang, 2014). Unexpectedly PA levels in *A. thaliana* leaves were not reduced

when the PLD α 1 was absent (Devaiah et al., 2006). This suggests a PA production by other PLD isoforms or formation of PA through the PLC-pathway. Most of the generated PAs are not only represented in signalling pathways but also assists as a component for the biosynthesis of structural phospholipids in the cell (Munnik, 2001). The different molecular forms of PA species exhibit diverse affinities to proteins (Li et al., 2009; Vergnolle et al., 2016). So far, two different PA species such as PA 18:3/18:2 and PA 18:2/18:2 are identified under PLD pathway (Hou et al., 2015; Vergnolle et al., 2005).

Changes in the concentration of PA molecules might influence a massive variation of downstream responses. It is important to profile the PLD isoform-specific PA species production data to resolve the PA mediated protein regulation mechanisms. Due to the lowest concentration of PA (<0.2%) in the form of multiple species and also the quick turnover of PA species hinders the appropriate profiling of PA contents in the cell. Moreover, PLD activity is robustly regulated based on the various developmental stages and also depending on the stress response (Tao et al., 2017; Wang, 2014). Therefore, the recognition of interaction between PA species and downstream effects is challenging.

Nevertheless, development-specific analysis exposed different expression levels of the PLDrp1 in leaf extracts of 1 to 6-week old *A. thaliana* plants that may be explained by altered PA levels or PA compositions (Ufer, 2015). Similar results were shown via the development-specific expression of a putative PLDrp1-like protein in *Camellia sinensis* (Li et al., 2015). The At3g29075 protein also illustrated various expression levels on the different developmental stages of *A. thaliana* (Figure 16). Therefore, the different expression levels of PA species during the developmental stages of the plant might also have regulated the expression pattern of the At3g29075. Further studies on the mutant lines during developmental stage might provide additional information to support this finding.

4.6. Tissue-specific expression of At3g29075 via phosphatidic acid regulation

The protein At3g29075 is not expressed in dry seeds (Supplementary figure 2). During the seed germination process in hydration translation shift (HTS)(1-Day), no protein upregulation was observed; however, in germination translation shift (GTS)(2,3-Day) the At3g29075 protein level was upregulated (Figure 17, Supplementary figure 2). Therefore, the HTS and GTS might have controlled the PA expression process that directly upregulated the At3g29075 protein. Katagiri et al.,(2005) investigated a possible role of PA during seed germination, they germinated *A. thaliana* seeds under *in vivo* condition, and the

PA production was analysed during the germination process by them; the results showed the level of PA was maximum at 12 h (HTS) and decreased after 96 h (GTS) during seed germination. It shows the different expression level of PA during the seed germination process under the different translation shifts (HTS and GTS).

Prominent down-regulation of the PLDrp1 protein levels have been detected in seeds (Ufer, 2015). The different expression level of the At3g29075 protein was found in *A. thaliana* tissues (Figure 14), which illustrates that prominent upregulation of the At3g29075 in the shoot, yet there was no protein expression found in the roots. However, the PLDrp1 protein displayed contradictory higher expression in roots (Ufer, 2015). A two-fold higher expression of the PLDrp1-like protein in buds from *Camellia sinensis* related to their young expanding leaves was observed (Li et al., 2015). The activity of the PLD isoforms are tissue-specific (Bargmann and Munnik, 2006); therefore results in different PA concentrations and composition. So, the expression pattern of the PLDrp1 and the At3g29075 protein was regulated variously due to their different binding capacity to the PA molecule species. Profiling of tissue-specific lipid concentrations in wild-type *A. thaliana* plants showed PA levels that are in support with these results (Devaiah et al., 2006). The highest relative concentration of PA is found in shoots (~2%) and roots (~10%), while the lipid share of PA in leaves (0.2%) is enormously reduced, also identified similar distribution patterns for individual PA species in tissue-specific analyses (Devaiah et al., 2006). Based on the earlier findings, different molecular forms of PA exhibit diverse affinities to proteins (Guo et al., 2011), which activate or inhibit distinct signalling pathways. It seems that both PA concentration and PA composition can influence expression levels of the PLDrp1 and the At3g29075 genes in different tissues of *A. thaliana*.

4.7. Regulation of At3g29075 under stress condition

The protein At3g29075 is expressed in different levels based on the stress condition the plant has undergone. During the dehydration stress condition, the At3g29075 protein had exponentially increased when the relative water content decreased (Figure 20 & Figure 21). The dehydration process results in the activation of many stress-signalling pathways via upregulating NOS, ROS, ABA, inositolphosphates and lipid-signalling molecules (Hou et al., 2016). Consequently, phospholipase C and distinct PLD isoforms are activated upon dehydration (Hirayama et al., 1995; Wang et al., 2014; Hong et al., 2008). Thus, the production of specific PA species increased and results in the activation of dehydration-associated signalling pathways (Hong et al., 2008; Munnik et al., 2000; Doherty et al., 2017).

The different PA-producing enzymes that relatively contribute overlapping functions to aid dehydration tolerance to plants by controlling different complex cellular and physiological response (Wang et al., 2006; Hong et al., 2008, 2010; Wang et al., 2014).

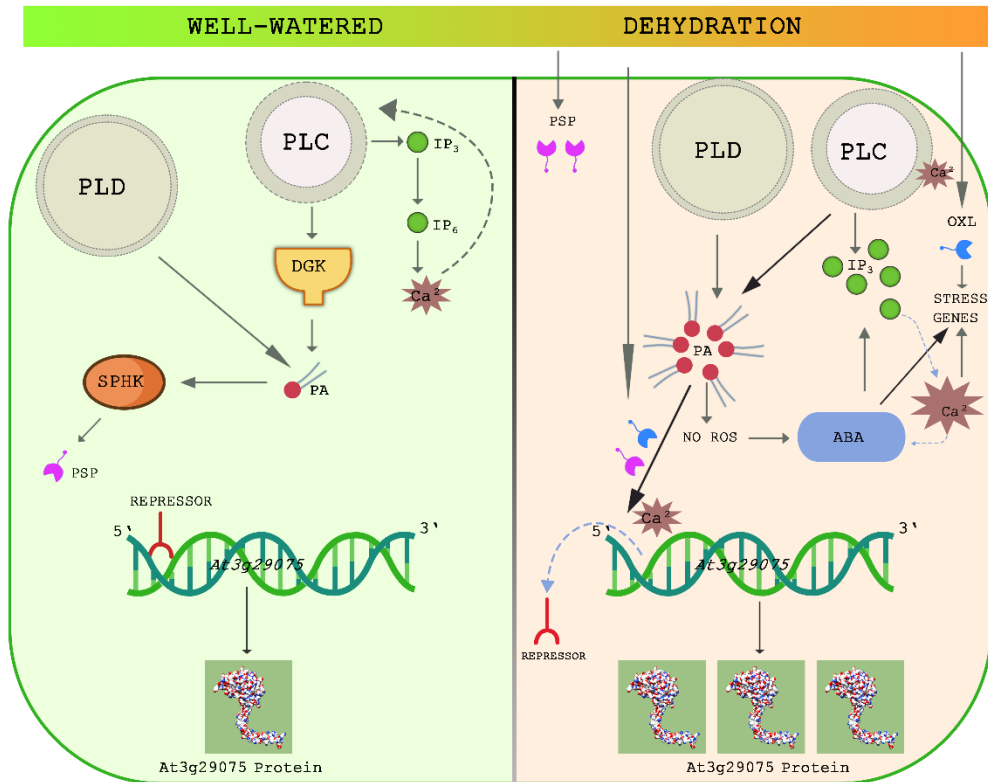


Figure 46: Model for the regulation of expression of *At3g29075* in response to dehydration in *Arabidopsis thaliana*. Under well water conditions, a repressor prevents transcription of *At3g29075*. Dehydration promote PLC and PLD activity, resulting in the increased production of PA species. Additionally, production of other signalling molecules, such as OXL, PSP, NO, ROS and IP_3 is promoted by oxidative stress and has positive feedback on the generation of PA and stress genes. Elevated levels of distinct PA species in combination with other signalling molecules displace the repressor from *At3g29075*, resulting in increased protein biosynthesis. Arrows indicate positive regulatory events. The role of other signalling molecules is indicated to keep the figure simple. ABA-abscisic acid; DGK-Diacylglycerol kinase; IP_3 -inositol trisphosphate; NO-nitric oxide, OXL-oxylipins; PA-phosphatidic acid; PLC-phospholipase C; PLD-phospholipase D; PSP-phytosphingosine phosphate; ROS-reactive oxygen species; SPHK-sphingolipid kinase. (adapted and modified from Ufer, 2017).

Therefore, the previous studies showed that the PLD α 1 regulation under dehydration stress was mediated via PA molecules (Ufer, 2015). Furthermore, the upregulation of the *At3g29075* protein upon dehydration stress also indicates the possible interaction of PA molecules (Figure 20, Figure 38). However, the *At3g29075* protein didn't express a significant level of changes in controlled light conditions and also under ABA treatment (Figure 28 & Figure 30). Whereas, under salt stress condition, the expression pattern of the *At3g29075* is slightly downregulated (Figure 25). The different PA molecule species

are involved in various stress mechanisms (Zhu, 2002); therefore, the upregulation of the At3g29075 under dehydration condition shows that the PA molecule which binds to the At3g29075 protein is most likely involved in the dehydration stress (Figure 46). Further, specific PA binding studies have to be done to provide additional support to this finding.

4.8. The Cloning and production of recombinant At3g29075 protein

In the processes of producing the At3g29075 full-size recombinant protein, several different expression vectors were used to create a full-size At3g29075 clone (Table 4). It was challenging to clone the full-size of At3g29075 gene in the expression system, due to their tandem repeats in the C-terminal part and its toxicity. So, the tandem repeats undergone the deletion to prevent cell death; however, every single full-size clone of the At3g29075 expressed different sequence deletions (Supplementary figure-16, 18, 19, 20). The early studies state that when the bacteria detect the recombinant protein to be a toxin, it undergoes different mechanism to stop protein production as a survival mechanism (Doherty et al., 1993; Dong et al., 1995). In the case of the PLDrp1 full-size cloning, there were no such difficulties found (Ufer, 2015). Even though both proteins share the sequence similarity on the N-terminal part, the C-terminal part makes the PLDrp1 and the At3g29075 proteins unique. However, the tandem repeats in the C-terminal part of the At3g29075, turns out to be toxin for the bacterial strain (Figure 9).

To provide additional support to my findings, I would like to state the following facts, those were observed during the cloning process. In the case of entry vectors/donor vectors, the sequencing results showed complete sequences of the full-size At3g29075 gene (the results are not shown). However, when it was subcloned into the expression system to overexpress the protein, the sequence was deleted. This raised a question “what happens when it was cloned without C-terminal part?”. Therefore, N-terminal parts were cloned separately, and there were no difficulties observed (Figure 36), and also the C-terminal part of the cloning was achieved successfully (Nasr, 2015). Thus, I could clearly state that the full-size protein is a toxin to the bacteria. The protein has to be full-size to examine its potential function (Cooper, 2019).

In addition to that, there were very few colonies appeared on the selection media in case of the full-size At3g29075 clone in the expression vectors. Earlier studies on the overproduction of seven membrane proteins in an *Escherichia coli*-bacteriophage T₇ RNA polymerase expression system was investigated. In all seven cases, when the expression of

the target membrane protein was induced, most of the BL21(DE3) host cells died; Similar effects were also observed with expression vectors for ten globular proteins (Miroux and Walker, 1996).

So, the first full-size At3g29075 clone obtained via pQLinkHD gateway expression system was used for the recombinant protein production (Supplementary figure 23). Even though the full-size At3g29075 cloning was successful, it was challenging to induce the recombinant protein in the host bacterial strain. Also, the T₇ transcription system is highly repressible, leaky transcription of the extremely toxic gene is still lethal to the host *E. coli* (O'Connor and Timmis, 1987; Li et al., 2009; Kong, 2000). Several different bacterial strains were used (Table 4) and also different induction times, and different concentration of IPTG was used in this process (Figure 32, Figure 33). The findings (example: deletion, few colonies on the plate, low expression of the protein) are evident the toxicity of the full-size At3g29075 gene; hence, the bacterial strains are unable to induce the high quantity of the full-size At3g29075 protein even under different conditions (Figure 34). Although several proteins have been successfully expressed to very high levels in BL21(DE3), very often significant over-production cannot be achieved because of the toxicity of the target protein, which may even cause bacterial cell death (Dumon-Seignovert et al., 2004; Doherty et al., 1993).

4.9. Interaction of At3g29075 with phosphatidic acid

Over the past years in plants and animals, the signalling role of phosphatidic acid (PA) in various physiological processes has been studied thoroughly and resulted in the identification of numerous PA-binding proteins (Chakraborty and Jiang, 2013; Hou et al., 2016; Jang et al., 2012; Julkowska et al., 2015; Munnik, 2001). However, no particular PA-binding motif has been recognised, and it might not exist. Recently an “electrostatic/hydrogen-bond switch” model (Testerink and Munnik, 2011), which elucidates the molecular basis of PA binding attributed to the unique properties of PA has been proposed: the essential amino acid residues such as, lysine and arginine were found to increase the chance of PA binding to its phosphate group. Through converting the electrostatic properties, the phosphate of PA turns into an ideal docking site for necessary amino acids. The conical shape of PA provides less stability to the plasma membrane than the cylindrical shape of structural phospholipids (Kooijman et al., 2003; Testerink and Munnik, 2011). Hence, PA is an appropriate membrane lipid for the insertion into positively charged membrane protein domains (Kooijman et al., 2007). Due to the absence

of a PA-binding motif, the identification of putative binding sites in proteins of interest is difficult. Therefore, independent lipid binding assays have been studied to assess the binding of PA to the N, C-terminal and full-size of At3g29075 protein (section 3.13.). Protein-lipid-overlay assays showed strong PA-specific binding affinity of the At3g29075; however, its N-terminal protein fragments displayed decreased, C-terminal part showed faint affinities towards PA (Figure 37). These findings were confirmed by liposome-binding assays that consider the cellular state of lipids as liposomes (Figure 38). Similar results were observed for the PLD α 1 protein interaction with PA (Ufer et al., 2017).

The recent studies illustrated that arginine and lysine residues are essential for the PA-binding process (Kooijman et al., 2007; Testerink and Munnik, 2011; Julkowska et al., 2015; Petersen et al., 2012). Whereas, the PLD α 1 displayed high numbers of lysine and arginine amino acids, Nt- PLD α 1 has a reduced lysine concentration (Ufer, 2015); the results showed higher binding affinity of the full-size PLD α 1 protein compared to its N-terminal fragment of the PLD α 1 (Ufer, 2015). However, the C-terminal At3g29075 protein part contains more lysine amino acids, yet it showed weak affinity (Figure 38) to PA. Thus, it seems like a full-size functioning protein is required for the proper binding.

The PA formation in response to drought (Katagiri et al., 2001; Testerink and Munnik, 2011) and up-regulation of the PLD α 1 in *pld α 1* knock-out plants upon dehydration indicates a PA- interaction under *in vivo* conditions (Ufer, 2015). The slight increase in the phosphorylation level of the PLD α 1 in wild-type leaves and the observed PA binding specificity *in vitro*, support an interaction of PA and PLD α 1 *in vivo* (Ufer, 2015). The sequence similarity between the PLD α 1 and the At3g29075 proteins and the results observed on dehydration stress showed upregulation of the At3g29075 (Figure 21) and also PA binding of the At3g29075 protein (Figure 37 & Figure 38) demonstrates it might be one of the eligible candidates of the PLD down-regulating proteins. Additional PA specific binding assays might reveal more evidence for regulation of the At3g29075 protein under PLD.

4.10. At3g29075 gene involvement in the flowering mechanism

The theoretical foundation of using overexpression as a genetic tool to study gene expression became a predominant method. Increasing the copy number of the entire gene or part of the gene, results in mutant phenotypes (Prelich, 2012). Even incremental changes in copy numbers can produce mutant phenotypes; thus, balanced gene expression is essential (Showe and Onorato, 1978). A natural extension of this concept is that intentional

overexpression of individual genes might be a useful tool for connecting genes to biological pathways (Prelich, 2012).

Therefore, the AT3G29075 overexpression line was generated during this study, and the phenotypes were studied with *at3g29075* knockdown mutant and wildtype. The study showed that the overexpression of gene *At3g29075* stimulated the early flowering in the *A. thaliana* (Figure 43).

The protein-protein interaction partners of *in silico* analysis showed that the *At3g29075* gene interacts with CDKB2;2 kinase (Figure 10). The cyclin-dependent kinases CDKB2;1 and CDKB2;2 are essential for both normal cell cycle progression and meristem organisation (SAM) (Andersen et al., 2008). There is no clear evidence or insight of knowledge acquired on how the *At3g29075* gene could be involved in the flowering mechanism. However, the early studies state that CDKB2;2 kinase also plays a role in the developmental programs in response to hormonal stimuli (Andersen et al., 2008). There is a possibility that the enhanced *At3g29075* protein activity might have triggered a switch in the flowering pathway via interacting with CDKB2;2 kinases in *A. thaliana*. Post-translational modification of proteins is indispensable in the regulation of all aspects of plant development, including flowering (Mulekar et al., 2012); thus, it might be one of the possible reasons for the early flowering of the AT3G29075 overexpression lines. Further studies have to be done to provide more evidence.

4.11. Plant Lysine-rich proteins: A new classification for uncommon and beneficial lysine-rich motif

Lysine (symbol **Lys** or **K**) is an α -amino acid that is used in the biosynthesis of proteins. There are 17,647 lysine-rich proteins listed on the unreviewed category in UniportKB (2019) and 2877 lysine-rich proteins as reviewed. In human 1264 proteins are identified as lysine-rich; in mouse 856; rat 268, bovine 164; When it comes to *A. thaliana* there are 120 proteins listed as lysine-rich proteins of which 116 proteins are reviewed and 4 of them are unreviewed. Very few proteins have been identified with a lysine-rich motif. Therefore, there was no classification system based on lysine motifs created so far. The *Arabidopsis thaliana* unknown protein *At3g29075* contains 21.1% lysine and -20 lysine-rich motifs were identified (Figure 47) in addition to 5 different glycine-rich motifs. The number of lysine-rich motifs are higher than the glycine-rich motifs. There comes a question “why not to categories this protein under the lysine-rich protein?”, and thus, this new proposal for

lysine-rich protein (LRPs) classification has been created. The lysine-rich protein can be classified under two different categories,

4.11.1. Classification based on the clustering of amino acids for protein secondary structure prediction (Zheng, 2004)

4.11.2. Classification based on the functional diversity of plant lysine-rich protein superfamily (Mangeon et al., 2010; Sachetto-Martins et al., 2000)

```
>At3g29075
MPYYTNDNDNDVDDFTEYDPMPSYGGYDITVTYGRSIPPSDETCYPLSSLGDAFEYQRPNFSSNHSSAYDDQALKTEYSSY
ARPGPVGSGSDFGRKPNNSGYGGRTEVEYGRKTESEHGSGYGGRIESDYVKPSYGGHEDDGGDDGHKKHSGKDYDDGDEKSK
KKEKKEKKKDKKKDGNNSEDEFKKKKKKEQYKEHHDDDDYDEKKKKKKDYNDDEKKKKKHYNDDDDDEKKKKHNYN
DDDDEKKKKKEYHDDDEDKKKKHYDNDDDEKKKKDKHRDDDEKKKKDKHHKGGHD
```

Figure 47: The lysine-rich repeats of the At3g29075 protein. Lysine amino acids are written in red colour. Lysine clusters are marked based on Zheng (2004) classification of amino acid clusters.

4.11.1. Classification based on the clustering of amino acids for protein secondary structure prediction

Methods for predicting the secondary structure of a protein from its amino acid sequence have been developed for three decades. Besides neural network models and nearest-neighbour methods, the Chou-Fasman/GOR statistical method is well established and commonly used. A simple hidden Markov model with higher-order correlations included for the secondary structure prediction. Zheng (2004) proposed several schemes for the amino acid alphabet reduction to incorporate residue correlations in the model. While the model is much simpler than the Bayesian segmentation model, its performance is still competitive. The possible motif groups that can be identified are triplet and quintuplet amino acid state for the secondary structure prediction.

A) Triplet state

A simplified version of the model can be constructed in the frame of the Chou-Fasman propensity scheme. Clustering of the amino acids for the protein secondary structure prediction (Table 5).

Table 5: Triplet state of lysine-rich amino acid clustering.

S. No	State	S. No	State
1.	KKK	4.	KXX
2.	KKX	5.	XXK
3.	XKK	6.	XKX

Possible clusters of triplet form of lysine with other amino acids creating a secondary structure of the protein. K- lysine amino acid, X- any amino acid.

B) Quintuplet states

Based on the Zheng classifications, the more realistic model uses quintuplets clustering of amino acids for the secondary structure predictions (Table 6).

Table 6: The quintuplet lysine-rich amino acid clustering.

S. No	State	S. No	State	S. No	State	S. No	State
1	KKKKK	6	KXXKK	11	KXKKX	16	XXKKX
2	KKKKX	7	KKXXK	12	KXKKX	17	XXKKX
3	XKKKK	8	XKXKK	13	XXXKK	18	XKXKK
4	KKKXX	9	XKKXK	14	KXXKK	19	XKXKK
5	XXKKK	10	XKKKX	15	KKXXX	20	KXKXX

The quintuplet state structure of the lysine (K) amino acids clusters with (X) any other amino acid and forming a cluster and creating the secondary structure of the protein.

These are the primary level of possible triplet and quintuplet, lysine amino acids clustering for the protein secondary structure prediction; however, the legitimacy of these predictions has to be verified with further studies and calculations.

4.11.2. Classification based on the functional diversity of the plant lysine-rich protein superfamily

There is a greater interest in identifying functional features of lysine-rich proteins. The occurrence of quasi-repetitive lysine-rich peptides has been reported in different organisms. However, the majority of the lysine-rich proteins are plant-based. The classification of lysine-rich protein superfamily based on its functional diversity (Mangeon et al., 2010; Sachetto-Martins et al., 2000) can be further subclassified under three more sub-divisions such as

- Expression of LRPs
- Subcellular localisation of LRPs
- Function of LRPs

Table 7: *Arabidopsis thaliana* proteins with possible lysine-rich repeats.

S. No	Gene name	Lys composition %	Number of amino acids	Sequence repeat	Signal peptide	Modulation	Tissue expression	Sub-cellular location
1	<i>At3g29075</i>	21.1	294	XKKX, XKKK, KKKX, KKKK, XXXK, KKXK	-	Dehydration, Salt stress	Not expressed in roots	Cytoplasmic, Nucleus
2	<i>At5g44610</i>	17.9	168	XKKX	-	ABA, Cold, Gibberellin, Salt stress	Developing root hairs, elongating pollen tubes	Plasma membrane, Cytoplasm

3	<i>At3g07610</i>	11.5	1027	KXXX, XXXX, XXXX, KXXX	-	Demethylating, Negative regulation of gene silencing	—	Nucleus
4	<i>At4g30860</i>	9.3	497	KKXX, KXXX, XXXX	-	Histone methyltransferase activity (H3-K36 specific); metal ion binding	Roots, flowers and buds	Nucleus
5	<i>At4g02020</i>	7.5	856	KXXX, KXXX, XXXX, XXXX	-	SS-RNA binding, methyl transferase activity, response to ABA	—	Nucleus
6	<i>At3g48430</i>	7.1	1360	KXXX, KXXX, XXXX, XXXX	-	Leaf development, vegetative to reproductive phase transition	Root tips, meristem, leaves, siliques	Nucleus
7	<i>At5g43990</i>	6.3	717	XKXX, XXXX	-	Zinc ion binding, RNA directed methylation	—	Nucleus
8	<i>At3g13490</i>	5.1	602	KXXX, KXXX	-	ATP binding; DNA binding; lysine-tRNA ligase activity; metal ion binding; tRNA binding	—	Chloroplast, Cytosol, Mitochondrion
9	<i>At2g23130</i>	5.4	185	XKXX, KXXX	-	-	Flowers, leaves stems, roots	Cell membrane
10	<i>At2g36570</i>	4.8	672	XKXX, KXXX	-	ATP binding; protein kinase activity	Roots, leaves, stems and flowers.	Cell membrane
11	<i>At5g14420</i>	4.1	468	XXXX, KXXX	-	Metal ion binding; ABA-activated signalling pathway; Auxin metabolic process; Cytokinin metabolic process; negative regulation of response to water deprivation	—	Nucleus, Plasma membrane
12	<i>At3g56580</i>	2.5	320	KXXX	-	Repressed by drought and osmotic stresses	Seedlings and flowers	Cytoplasm

The *At3g29075* protein was compared with the top 11 list out of 120 Uniport Lysine-rich protein results (2019) with a filter of *Arabidopsis thaliana*. - No information available (2019). K- Lysine; X – any other amino acid.

The primary level analysis of *A. thaliana* lysine-rich protein sequences and data from Uniport (2019) revealed that most of the top 12 proteins on the list have no signal peptides and most of them are located in the nucleus and cytoplasm and few of them are located in the cell membrane. Various tissue-specific expression levels were illustrated on the corresponding proteins in the *Tair* and *Uniport* database. Different genes documented on this list have several functions. There was one unique feature identified on this data analysis was the majority of the lysine-rich repeats are located in the C-terminal part of the protein. The *At3g29075* protein has 21.1% of lysine; so far, this is the highest of lysine consisting protein that was identified from the top 120 lysine-rich *Arabidopsis thaliana* data from the Uniport (Table 7). Developing a lysine-rich protein classification seems to be important criteria to understand their expression, localisation, function and structure. Moreover, this

new lysine classification might help to select the appropriate lysine resource for future commercial lysine-rich product development.

Disclaimer: This is the primary level of an idea to create, “the new lysine-rich protein classification” system. However, the new classification of lysine-rich protein requires extensive data analysis (~20,000 lysine-rich proteins); therefore, this proposal for the lysine-rich classification might be updated based on the results obtained in the future.

5. Summary

PLD α 1 mutant and WT protein samples were analysed in the 2D gel that revealed a missing protein spot in PLD α 1, which was identified as the unknown At5g39570 protein (Anke Kuhn, 2009). Later on, the unknown At5g39570 protein was characterised as PLD regulated protein 1 (PLDrp1) through the protein-lipid binding study (Ufer et al., 2017). BLASTn analysis of PLDrp1 protein retrieved the unknown At3g29075 protein as a homolog, which has a 49% sequence similarity in the N-terminal part (Ufer, 2015). Preliminary bioinformatics analysis exhibited that At3g29075 as class V glycine-rich protein, with 294 aa, and an estimated molecular weight of 34 kDa. The amino acid composition analysis shows also a high lysine (21.1%) content.

To avoid the PLDrp1 detection with At3g29075 antiserum, due to their N-terminal sequence homology, an antiserum was raised against the C-terminal part of the At3g29075 protein (Nasr, 2015). The PLDrp1 antibody displayed the homology of both proteins by detecting both PLDrp1 and At3g29075 proteins. The C-terminal At3g29075 antibody only binds to the At3g29075 protein. The expression analysis upon dehydration condition showed the At3g29075 protein expression was upregulated at a significant level. Hence it is engaged in abiotic stress.

PA acts as a secondary messenger in plant stress signalling, and the PLDrp1 protein displayed the binding to PA (Ufer 2017). Based on the sequence similarity with the PLDrp1 protein, the At3g29075 recombinant protein was examined with lipid and liposome binding assays. Both assays exhibited the strong binding with full-size At3g29075 and the N-terminal part of the protein. The study also exhibited a weak binding to the C-terminal half of the At329075 protein. Thus, At3g29075 protein also associates with PA.

The phenotyping of the AT3G29075 overexpression and at3g29075 knockdown mutant lines with WT in *A. thaliana* illustrated early flowering in the AT3G29075 overexpression line. Thus, it demonstrates the involvement of the At3g29075 gene in the flowering mechanism. The subcellular localisation of the At3g29075 protein was determined by the PGJ280 vector containing the At3g29075-GFP fusion construct. Transient transformation by using particle bombardment in *A. thaliana* leaves revealed the subcellular expression of the At3g29075 protein in the cytosol and the nucleus.

6. Outlook

The observation of this thesis contributes, the regulation and function of the protein At3g29075 as a novel target of phospholipase D and its cleavage product phosphatidic acid. The recognition of a specific PA-binding domain is in the main objective of research for several years, yet the binding of PA to target proteins and applies its function is a mystery. The results attained in this study propose that PA-binding takes place in the N-terminal half of the protein. Yet, the C-terminal fragment was unable to solubilisation entirely; therefore, it would be necessary to study the C-terminal part with feasible methods. The generation of constructs with partial sequences of At3g29075 and sequential study of PA-binding affinities might aid to recognise the PA-binding domain of this protein. The lipid-binding assays should be repeated with specific PA species to find out the specific PA binding mechanism of the At3g29075 protein.

During water limiting condition, the PA species might result in the upregulation thus result in the upregulation of the At3g29075 protein level in dehydration stress. Distinct PA-species were previously described to affect target proteins by translocation and disorder-to-order transitions (Hou et al. 2015). However, the stable transformation of At3g29075-GFP and consequent *in vivo* observation of the protein At3g29075 in response to abiotic stress may reveal a translocation of the protein and help to unravel its function.

By monitoring at3g29075 knock-down mutant, AT3G29075 overexpression line with WT upon standard and stress conditions might reveal the functional characteristics of the At3g29075. RNA sequencing of AT3G29075 overexpression line will illustrate altered gene-expression levels in mutant plants and may indicate to the potential involvement of the protein in physiological pathways.

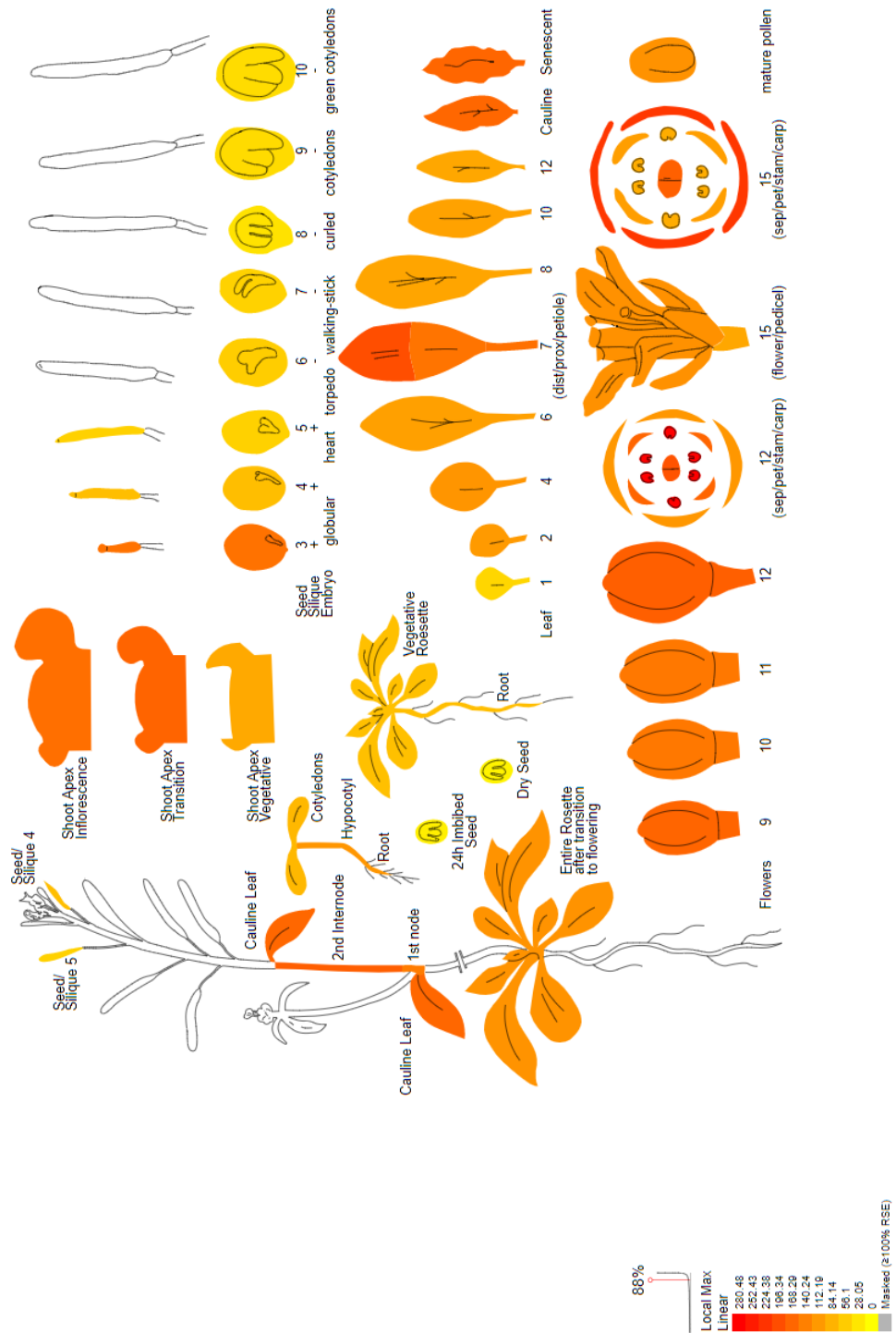
Further, protein-protein interaction assays are mandatory to identify its interaction partners. The generation stable AT3G29075 overexpression line with GFP might help to do identify the interaction partners via immunoprecipitation of GFP fusion method.

The difference in the At3g29075 expression at transcription and translation was observed during this study. To get the insights of translation regulation process, it is necessary to study the polysome occupancy of the At3g29075 gene upon different stress condition.

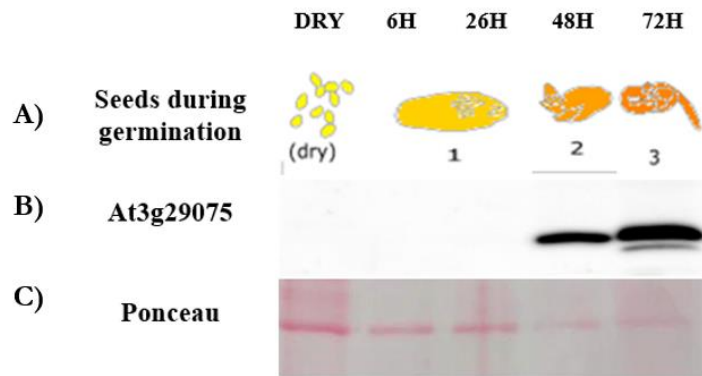
Last but not least, the second unknown protein (20kDa) was detected during the seed germination process by C-terminal At3g29075 antibody. Due to the low quantity, it was

hard to locate the correct protein spot in the 2D-gel for the MALDI-TOF analysis. Optimising a feasible method, to locate the unknown protein spot in the 2D gel, thus, helps to identify the unknown protein via MALDI-TOF. Eventually, that supports to determine the correlation between At3g29075 protein and the unknown protein.

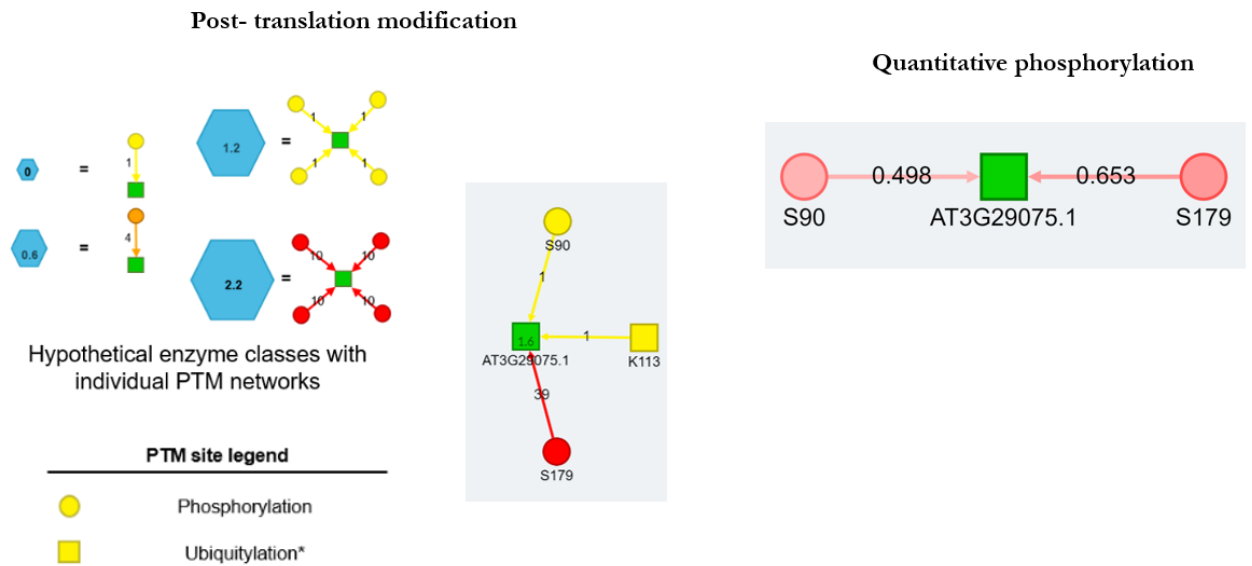
7. Supplementary data



Supplementary figure 1: Tissue-specific expression analysis of *At3g29075* based on microarray data in *Arabidopsis thaliana*. Data and pictures obtained from the BAR ePlant.



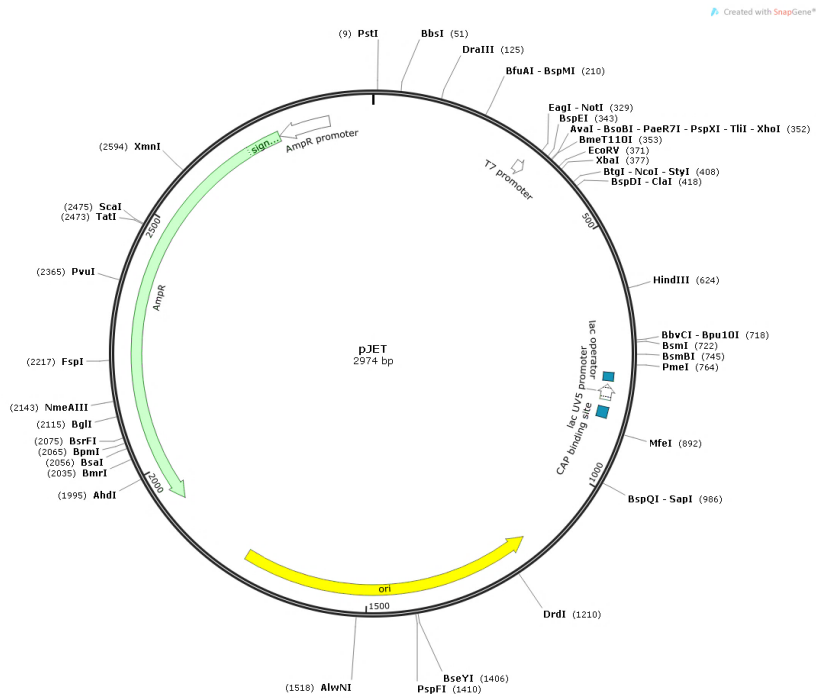
Supplementary figure 2: Immunodetection of At3g29075 in dry seed to a seedling at 72 h after the start of imbibition (HAI) in *Arabidopsis thaliana*. A) the *in-silico* analysis At3g29075 expression level during the seed germination process. B) represents immune detection of protein At3g29075 during seed germination. C) Ponceau stain of the protein samples.



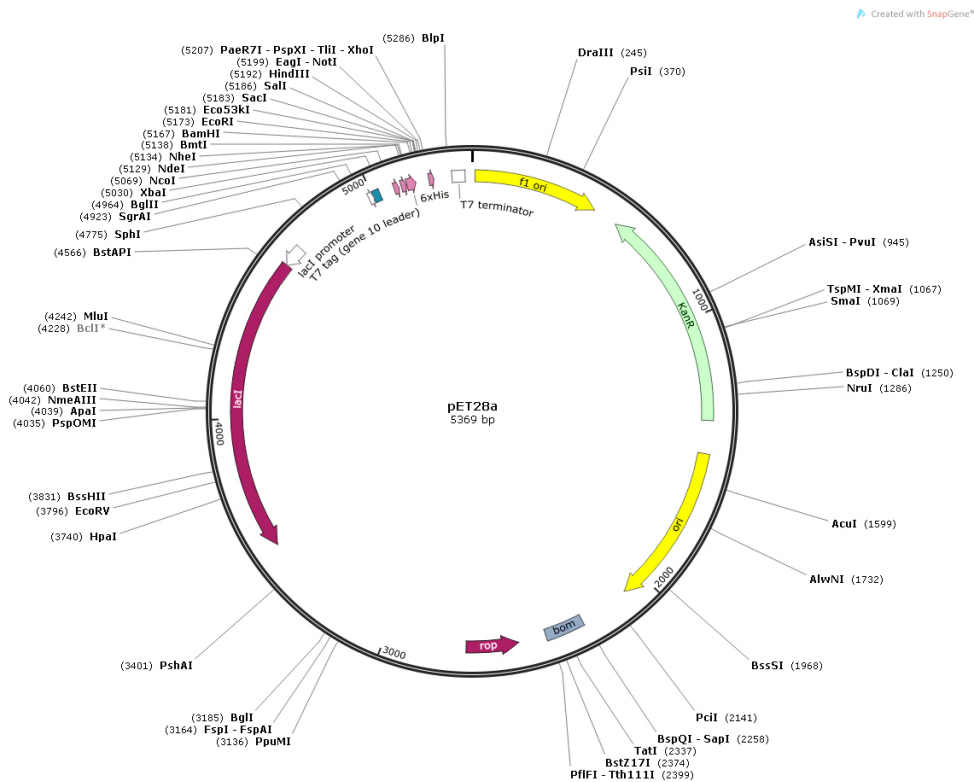
Supplementary figure 3: Post-translation modification sites of At3g29075 protein (left) and the quantitative value of phosphorylation values of the At3g29075 PTM sites (right). Source: FAT-PTM.

Vector maps

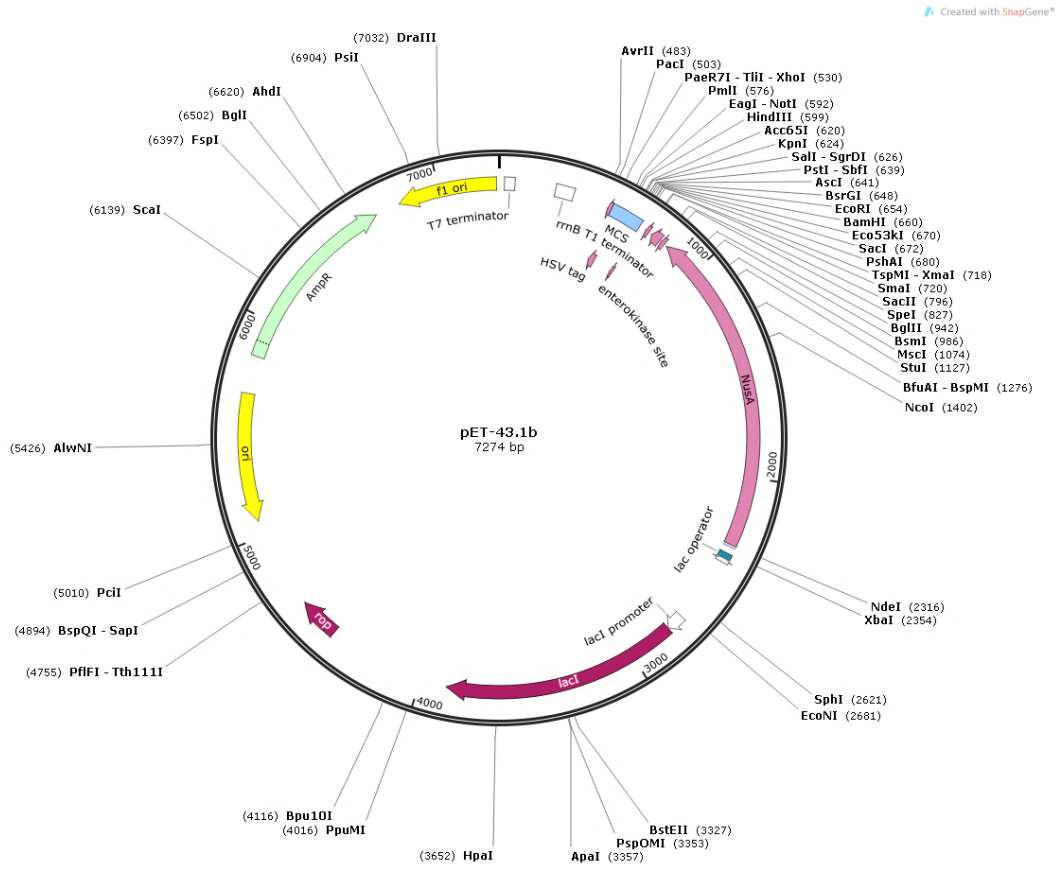
For further information, see section 2.1.6.



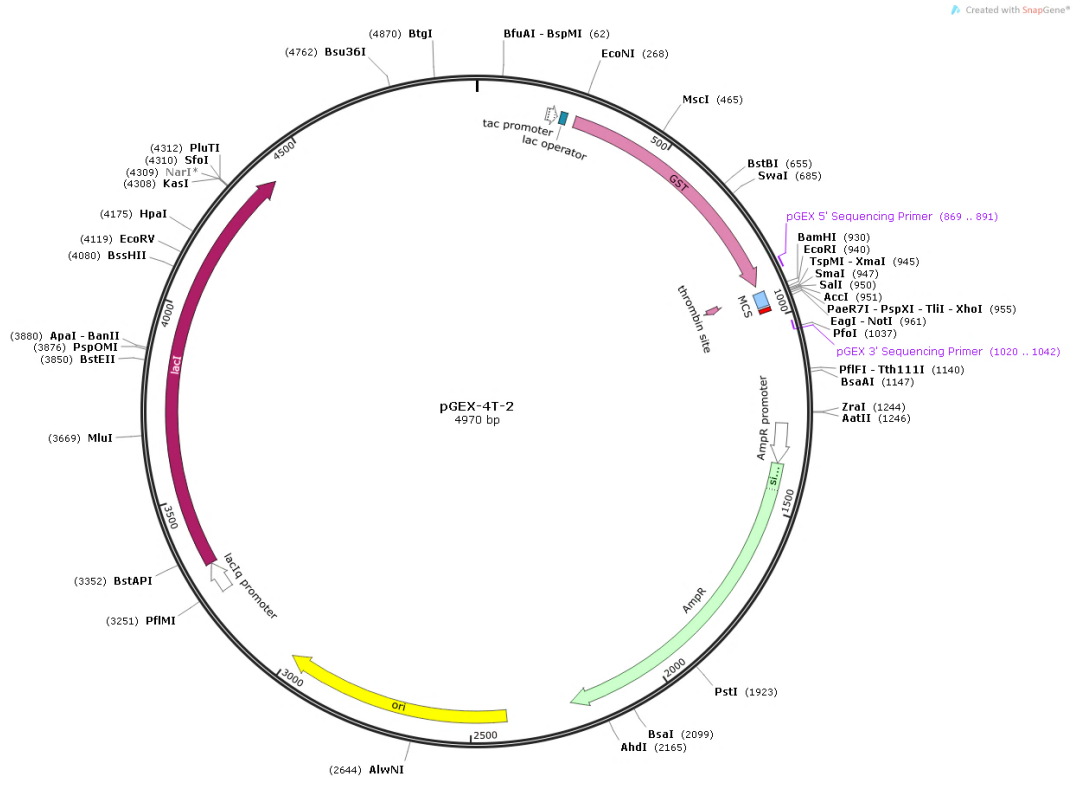
Supplementary figure 4: Schematic diagram of the pJET1.2 entry vector map.



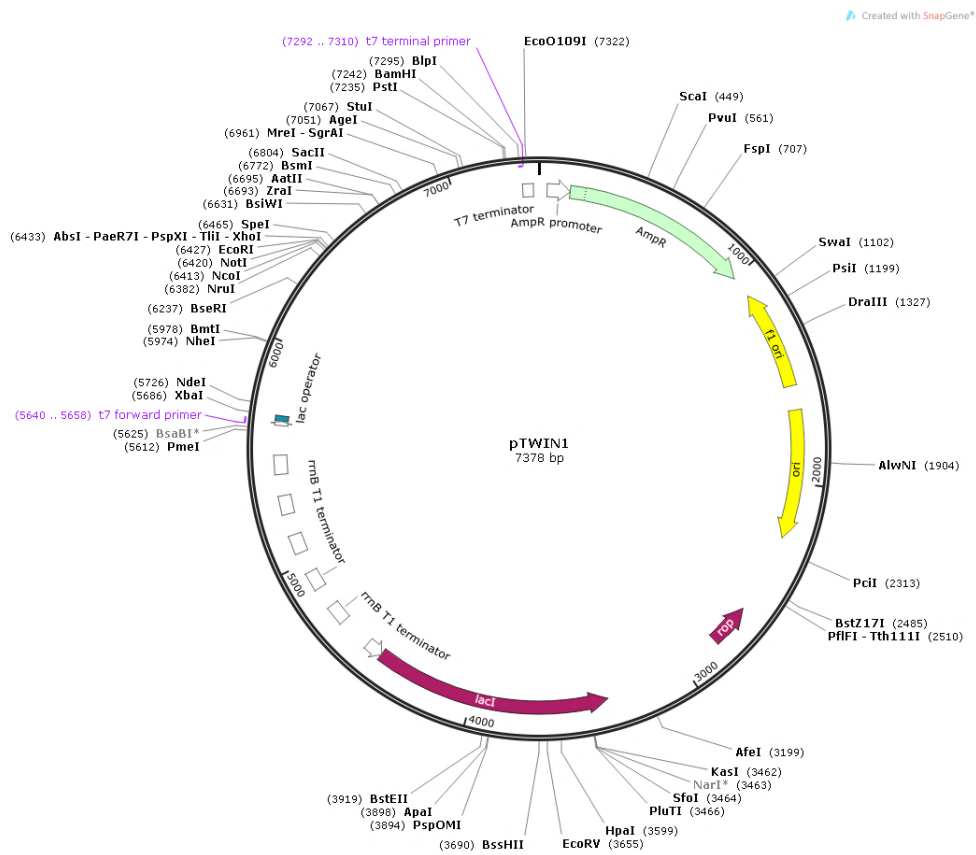
Supplementary figure 5: pET28a an expression vector map with 6x-his-tag.



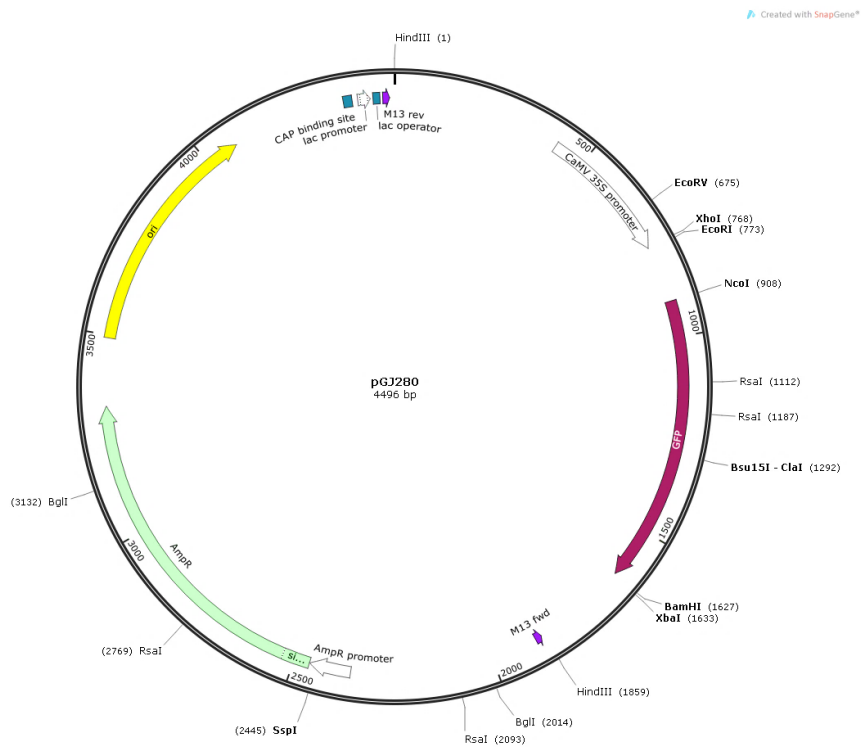
Supplementary figure 6: pET43.b expression vector map with his-tag.



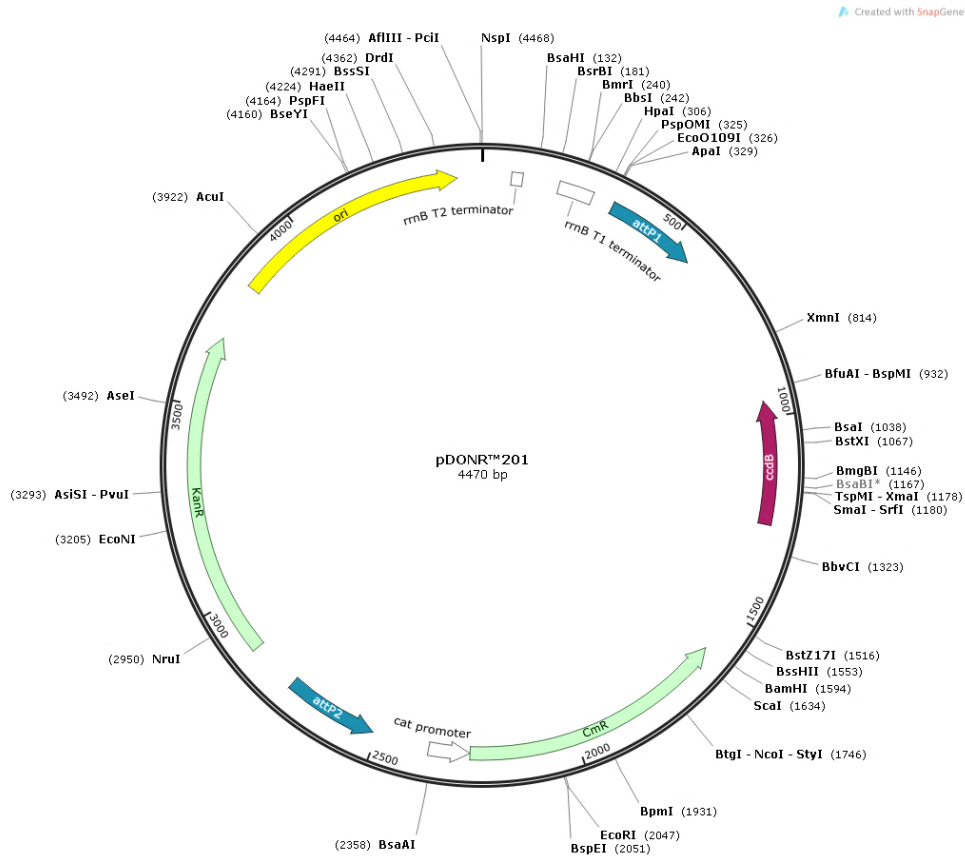
Supplementary figure 7: pGEX4T-2 expression map with GST tag.



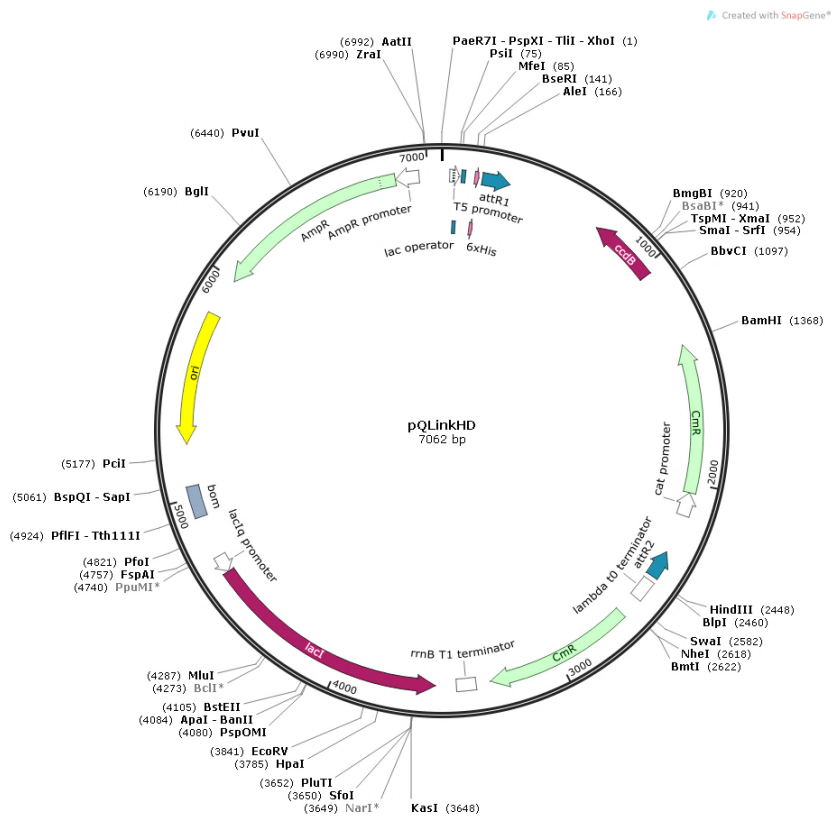
Supplementary figure 8: pTwin1 expression vector map without a tag system.



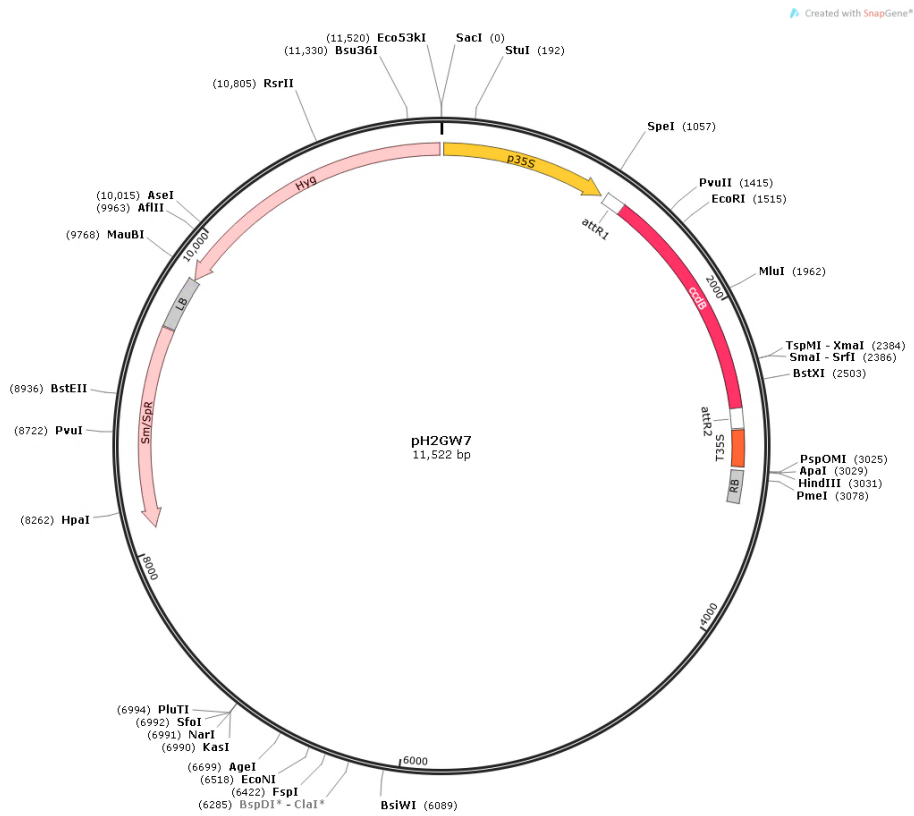
Supplementary figure 9: pGJ280 expression vector map with GFP tag for localisation.



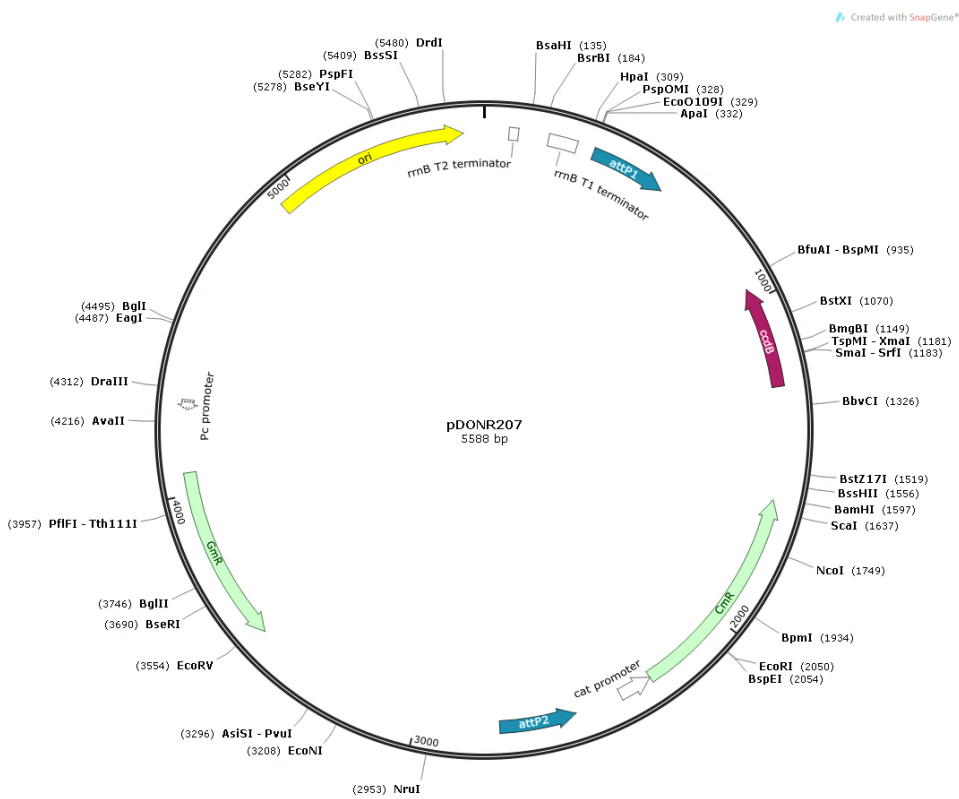
Supplementary figure 10: pDONR™201 gateway entry vector map.



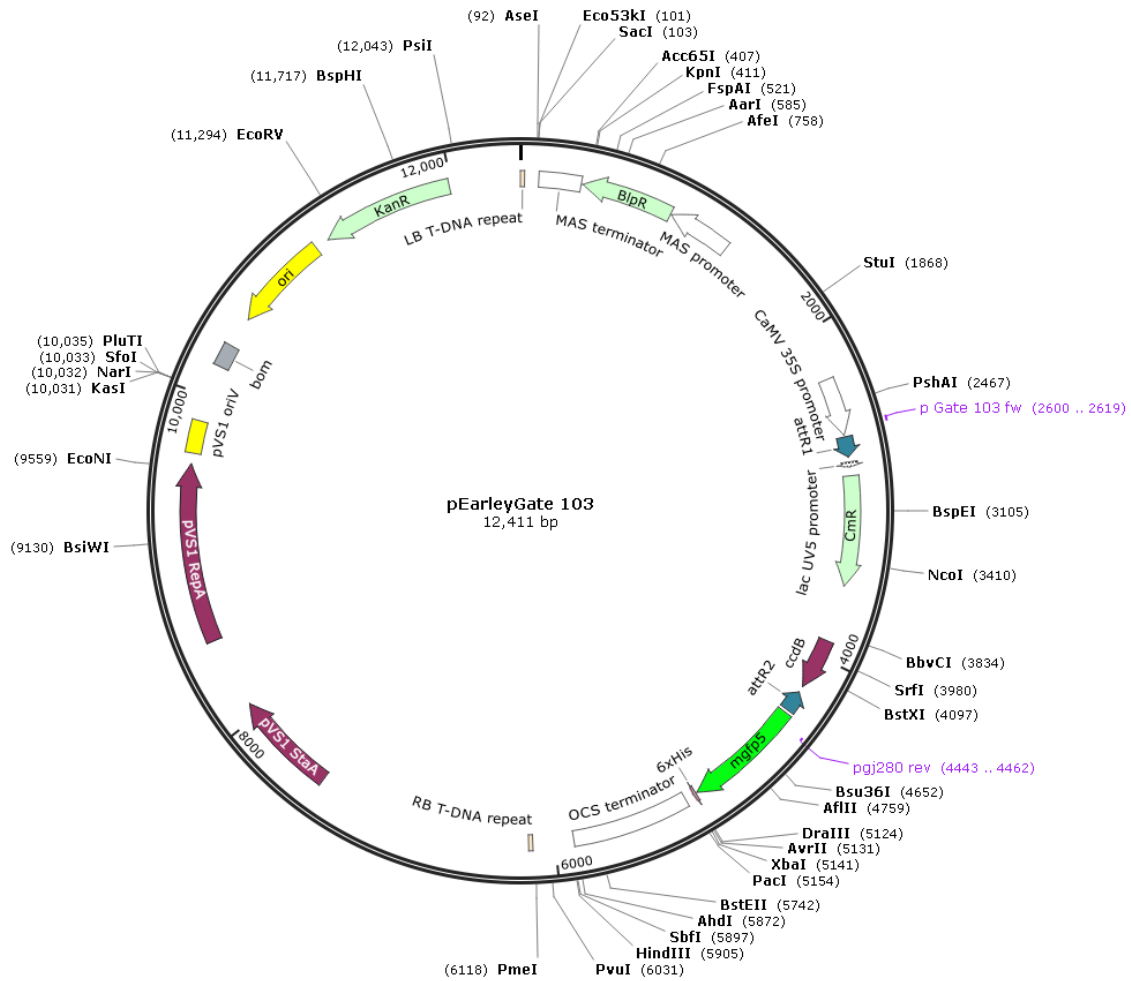
Supplementary figure 11: pQLinkHD gateway expression vector with His-tag.



Supplementary figure 12: pH2GW7 gateway expression vector map with 35S promoter.



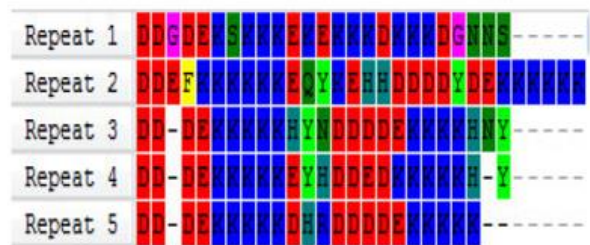
Supplementary figure 13: pDONRtm207 gateway entry vector map.



Supplementary figure 14: The pEarleyGate103 gateway expression vector map consists of GFP-tag.

```

MPYYTNDNDNDVDDFTEYDPMFYSGGYDITVITYGRSI
PPSDETCYPLSSLGDAFEYQRPNFSSNHDSAYDD
QALKTEYSSYARPGFVSGSDFGRKPNISYGGRTTEV
EYGRKTESEHSGYGGRIESDYVKPSYGGHEDDGDD
GHKKHSGKDY DDGDEKSKKKEKSKKDKKKDGNNSE
DDEFKSKKKEQYKEHHDDDDYDEKSKKKNYNDDE
EKKKKKHYNDDDEKSKKKNYNDDEKSKKKEYHD
DEDKSKKKNYNDDEKSKKKNYNDDEKSKKKNYND
DEKSKKKNYNDDEKSKKKNYNDDEKSKKKNYND
HHKHGD
    
```



Supplementary figure 15: Tandem repeats in At3g29075 are marked in Green

cDNA At3g29075.str from 1 to 885
 Alignment to
 pet28a-At3g29075.str sequence.str from 1 to 771

Matches (|): 725
 Mismatches (#): 1
 Gaps (): 204
 Unattempted (.): 0

```

      *      *      *      *      *      *      *      *      *      *
1  ~~ATGCCGTATTACCCAACGACGACAATGACGTCGACGATTTACCCGAATACGATCCGATGCCTTATAGTGGAGGCTACGACATCACCGTGACATACGG 98
   |||||
1  CCATGGCGTATTACCCAACGACGACAATGACGTCGACGATTTACCCGAATACGATCCGATGCCTTATAGTGGAGGCTACGACATCACCGTGACATACGG 100
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
99  CCGTTCAATCCACCGTCCGACGAGACTTGTACCCCTCTCTCCTCTCTCCGGCGACGCCCTTGTAGTATCAGCGACCTAATTTCTTCTAACCACGAT 198
   |||||
101  CCGTTCAATCCACCGTCCGACGAGACTTGTACCCCTCTCTCCTCTCTCCGGCGACGCCCTTGTAGTATCAGCGACCTAATTTCTTCTAACCACGAT 200
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
199  TCTTCTGCTTATGACGACCAAGCTCTTAAAAACCGAGTACAGTAGCTATGCACGACCCGGACCCGTTGGATCTGGATCTGATTTTGGCCGGAAACCTAATT 298
   |||||
201  TCTTCTGCTTATGACGACCAAGCTCTTAAAAACCGAGTACAGTAGCTATGCACGACCCGGACCCGTTGGATCTGGATCTGATTTTGGCCGGAAACCTAATT 300
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
299  CTGGATATGGAGGGAGAACCGGAGGTTGAGTATGGCCGAAAACTGAATCGGAGCATGGATCTGGCTATGGTGAAGAATTGAGAGCGATTACGTGAAGCC 398
   |||||
301  CTGGATATGGAGGGAGAACCGGAGGTTGAGTATGGCCGAAAACTGAATCGGAGCATGGATCTGGCTATGGTGAAGAATTGAGAGCGATTACGTGAAGCC 400
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
399  TAGCTATGGCGGTCACGAGGATGATGGTGACGATGGTCACAAAAACATAGTGGTAAGGATTATGATGATGGAGATGAGAAGAGTAAGAAGAAGGAGAAG 498
   |||||
401  TAGCTATGGCGGTCACGAGGATGATGGTGACGATGGTCACAAAAACATAGTGGTAAGGATTATGATGATGGAGATGAGAAGAGTAAGAAGAAGGAGAAG 500
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
499  GAGAAGAAGAAGGATAAGAAGAAGATGGTAATAACTCTGAAGATGATGAGTTTAAGAAGAAGAAGAAGAAGAGCAGTACAGGAGCATCATGATGATG 598
   |||||
501  GAGAAGAAGAAGGATAAGAAGAAGATGGTAATAACTCTGAAGATGATGAGTTTAAGAAGAAGAAGAAGAAGAGCAGTACAGGAGCATCATGATGATG 600
      *      *      *      *      *      *      *      *      *      *

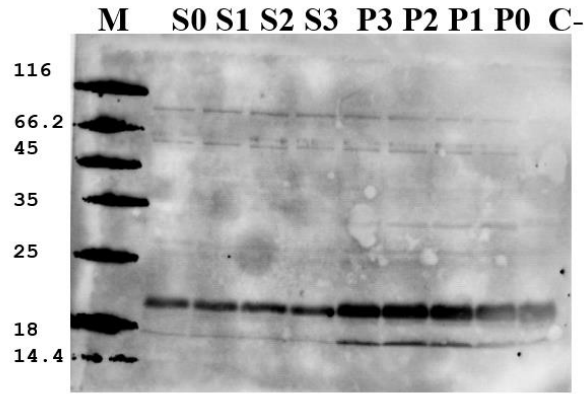
      *      *      *      *      *      *      *      *      *      *
599  ATGATTATGATGAGAAGAAGAAGAAGAAAGACTATAATGATGATGATGAGAAGAAGAAGAAGCATTATAATGATGATGATGATGAGAAGAAGAA 698
   |||||
601  ATGATTATGATGAGAAGAAGAAGAAAGAA---AGACTATAATGATGATGATGAGAAG----- 653
      *      *      *      *      *      *      *      *      *      *

      *      *      *      *      *      *      *      *      *      *
699  GAAGCATAATTACAATGATGATGATGATGAGAAGAAGAAGAAGAGGAGTATCATGATGATGAGGATAAGAAGAAGAAGCACTATGATAATGATGAT 798
   -----
654  ----- 654

      *      *      *      *      *      *      *      *      *      *
799  GATGAGAAGAAGAAGAAGGATCATCGTGATGATGATGATGAGAAGAAGAAGAAGGATAAGCACCACAAGGGAC----- 877
   |||||
654  -----AAGAAGAAGGATCATCGTGATGATGATGATGAGAAGAAGAAGAAGGATAAGCACCACAAGGGACTCGAGCACCACCACCACC 741
      *      *      *      *      *      *      *      *      *      *

      *
878  A-TGA-----CT-----AA 885
   |||||
742  ACtGAGAtCCGGCTGctAACAAGCCCGAa 771
  
```

Supplementary figure 16: Sequence alignment of At3g29075 cDNA with pET28a-At3g29075 insert shows 204 bp shorter.



Supplementary figure 17: Immunodetection of pET28a-At3g29075 recombinant protein with C-terminal At3g29075 antibody. M-Marker, C-Control, S0-S3: Soluble protein fractions before and after induction, P0-P3: Insoluble protein fractions before and after induction.

```

cDNA At3g29075.str from 1 to 885
Alignment to
pet28at3g -At3g29075.str-- Matches:645; Mismatches:2; Gaps:282; Unattempted:0

      *      *      *      *      *      *      *      *      *      *      *      *
1>ATGCCGATTACACCAACGACGCAATGACGTGACAGGATTTCACCGAATACGATCCGATGCCCTTATAGTGGAGGCTACGACATCACCGTGACATCGGCC>100
1>A-----AT-----ACg-----Ac-----c-----cgat-----a-----cctTATAGTGGAgGCTACGACATCACCGTGACAt|||c||gcC>54

      *      *      *      *      *      *      *      *      *      *      *      *
101>GTTCAATTCACCGTCCGACGAGACTTGTACCCCTCTCTCCTCTCTCCGGCGACGCCCTTTGAGTATCAGCGACCTAATTTCTCTTAACCCAGGATTC>200
55>GTTCAATTCACCGTCCGACGAGACTTGTACCCCTCTCTCCTCTCTCTCCGGCGACGCCCTTTGAGTATCAGCGACCTAATTTCTCTTAACCCAGGATTC>154

      *      *      *      *      *      *      *      *      *      *      *      *
201>TTCTGCTTATGACGCCAAGCTCTTAAAACCGAGTACAGTAGCTATGCACGACCCGGACCCGTTGGATCTGGATCTGATTTTGGCCGGAAACCTAATTTCT>300
155>TTCTGCTTATGACGCCAAGCTCTTAAAACCGAGTACAGTAGCTATGCACGACCCGGACCCGTTGGATCTGGATCTGATTTTGGCCGGAAACCTAATTTCT>254

      *      *      *      *      *      *      *      *      *      *      *      *
301>GGATATGGAGGGAGAACCGAGGTTGAGTATGGCCGGAAAACCTGAATCGGACATGGATCTGGCTATGGTGGAGAATTGAGAGCGATTACGTGAAGCCTA>400
255>GGATATGGAGGGAGAACCGAGGTTGAGTATGGCCGGAAAACCTGAATCGGACATGGATCTGGCTATGGTGGAGAATTGAGAGCGATTACGTGAAGCCTA>354

      *      *      *      *      *      *      *      *      *      *      *      *
401>GCTATGGCGGTACAGAGGATGATGGTGACGATGGTCAAAAAAACAATAGTGGTAAGGATTATGATGATGGAGATGAGAAGAGTAAGAAGAAGGAGAAGGA>500
355>GCTATGGCGGTACAGAGGATGATGGTGACGATGGTCAAAAAAACAATAGTGGTAAGGATTATGATGATGGAGATGAGAAGAGTAAGAAGAAGGAGAAGGA>454

      *      *      *      *      *      *      *      *      *      *      *      *
501>GAAGAAGAGGATAAGAAGAAGATGGTAATAACTCTGAAGATGATGAGTTTAAAGAAGAAGAAGAAGAGCAGTACAAGGAGCATCATGATGATGAT>600
455>GAAGAAGAGGATAAGAAGAAGATGGTAATAACTCTGAAGATGATGAGTTTAAAGAAGAAGAAGAAGAGCAGTACAAGGAGCATCATGATGATGAT>554

      *      *      *      *      *      *      *      *      *      *      *      *
601>GATTATGATGAGAAGAAGAAGAAGAAGAACTATAATGATGATGATGAGAAGAAGAAGAAGCATTATAATGATGATGATGATGAGAAGAAGAAGA>700
555>GATTATGATGAGAAGAAGC----->572

      *      *      *      *      *      *      *      *      *      *      *      *
701>AGCATAATTACAATGATGATGATGATGAGAAGAAGAAGAAGAGGAGTATCATGATGATGAGGATAAGAAGAAGAAGAGCCTATGATAATGATGATGA>800
572>----->572

      *      *      *      *      *      *      *      *      *      *      *      *
801>TGAGAAGAAGAAGAAGGATCATCGTGATGATGATGATGAGAAGAAGAAGAAGGATAAGCACCACAAGGGACATGACT--A--A----->885
573>-----AAGAGAAGGATCATCGTGATGATGATGATGAGAAGAAGAAGAAGGATAGCACCACAAGGGACATGACTAAGCACCACCACC>662

      *      *      *
885>----->885
663>ACCACtGAGAtCCGGCTGctAcaagnCC>691

```

Supplementary figure 18: Sequence alignment of At3g29075 cDNA with pET28a-At3g29075 insert shows 282 bp shorter.

cDNA At3g29075.str from 1 to 885

Alignment to

sequenceing result with t7 terminator.str-- Matches:672; Mismatches:16; Gaps:199; Unattempted:0

```

      *      *      *      *      *      *      *      *      *      *
1>~ATGCCGTATTACACCAACGACGACAATGACGTCGACGATTTcACCGAATACGATCCGATGCCTTATAGTGGAGGCTACGACATCACCGTGACATACGG>98
1>CCATGCGTATTACACCAACGACGACAATGACGTCGACGATttcCcGAATACGATCCGATGCCTTATAGTGGAGGCTACGACATCACCGTGACATACGG>100

      *      *      *      *      *      *      *      *      *      *
99>CCGTTCAATTCACCGTCCGACGAGACTTGTACCCTCTCTCCTCTCTCCGGCGACGCCCTTTGAGTATCAGCGACCTAATTTCTCTTAACCACGAT>198
101>CCGTTC AATTCACCGTCCGACGAGACTTGTACCCTCTCTCCTCTCTCCGGCGACGCCCTTTGAGTATCAGCGACCTAATTTCTCTTAACCACGAT>200

      *      *      *      *      *      *      *      *      *      *
199>TCTTCTGCTTATGACGACCAAGCTCTTAAAACCGAGTACAGTAGCTATGCACGACCCGGACCCGTTGGATCTGGATCTGATTTTGGCCGGAAACCTAATT>298
201>TCTTCTGCTTATGACGACCAAGCTCTTAAAACCGAGTACAGTAGCTATGCACGACCCGGACCCGTTGGATCTGGATCTGATTTTGGCCGGAAACCTAATT>300

      *      *      *      *      *      *      *      *      *      *
299>CTGGATATGGAGGGAGAACGGAGGTTGAGTATGGCCGGAAAACCTGAATCGGAGCATGGATCTGGCTATGGTGAAGAATTGAGAGCGATTACGTGAAGCC>398
301>CTGGATATGGAGGGAGAACGGAGGTTGAGTATGGCCGGAAAACCTGAATCGGAGCATGGATCTGGCTATGGTGAAGAATTGAGAGCGATTACGTGAAGCC>400

      *      *      *      *      *      *      *      *      *      *
399>TAGCTATGGCGGTACGAGGATGATGGTGACGATGGTCACAAAAACATAGTGGTAAGGATTATGATGATGGAGATGAGAAGAGTAAGAAGAAGGAGAAG>498
401>TAGCTATGGCGGTACGAGGATGATGGTGACGATGGTCACAAAAACATAGTGGTAAGGATTATGATGATGGAGATGAGAAGAGTAAGAAGAAGGAGAAG>500

      *      *      *      *      *      *      *      *      *      *
499>GAGAAGAAGAAGGATAAGAAGAAAGATGGTAATAACTCTGAAGATGATGAGTTTAAAGAAGAAGAAGAAGAGCAGTACAAGGAGCATCATGATGATG>598
501>GAGAAGAAGAAGGATAAGAAGAAAGATGGTAATAACTCTGAAGATGATGAGTTTAAAGAAGAAGAAGAAGAGCAGTACAAGGAGCATCATGATGATG>600

      *      *      *      *      *      *      *      *      *      *
599>ATGATTATGATGAGAAGAAGAAGAAGAAAGACTATAATGATGATGATGAGAAGAAGAAGAAGCAATTATAATGATGATGATGATGAGAAGAAGAA>698
601>ATGATTATGATGAGAAGAAGAAGAAGAAAG---GA-T-----aAGCA----->637

      *      *      *      *      *      *      *      *      *      *
699>GAAGCATAATTACAATGATGATGATGATGATGAGAAGAAGAAGAAGAGGAGTATCATGATGATGAGGATAAGAAGAAGAAGAAGCACTATGATAATGATGAT>798
638>---C-----C-----TC---GA---c-----GAGAGAGAGAGACAC---tGA----->665

      *      *      *      *      *      *      *      *      *      *
799>GATGAGAAGAAGAAGAAGGATCATCGTGATGATGATGATGAGAAGAAGAAGAAGGATAAGCACCACAAGGACATGACTAA>885
666>-----GATC---CG---G---tG---t---A---CA-----AAGC---Cc-----c-----aa>690

```

Supplementary figure 19: Sequence alignment of At3g29075 cDNA with pET43b-At3g29075 insert shows 199 bp shorter.

cDNA At3g29075.str from 1 to 885

Alignment to

pGEX4T2-At3g29075sequencing result.str-- Matches:872; Mismatches:1; Gaps:149; Unattempted:0

```
1>~AT-----G--C----->4
1>GAATCctCAACacaACATAtACAAAACAAACGAAACCAAGCAATCAAGCATTCTACTTCTATTGCGCAATTTAAATCATTTCTTTAAAGCAAAAGCA>100

* * * * *
5>-----CGTATTACACCAACGACGACAATGACGTCGACGATTCACCGAATACGATCCGATGCCT>63
101>AnTTTCTGaAAATTTTCACCAAnTTACGAACGATAGCCATGGCGTATTACACCAACGACGACAATGACGTCGACGATTCACCGAATACGATCCGATGCCT>200

* * * * *
164>TATAGTGGAGGCTACGACATCACCCTGACATACGGCCGTTCAATTCACCGTCCGACGAGACTTGTACCCCTCTCTCTCTCTCCGGCGACGCCTTTG>163
201>TATAGTGGAGGCTACGACATCACCCTGACATACGGCCGTTCAATTCACCGTCCGACGAGACTTGTACCCCTCTCTCTCTCTctCCGGCGACGCCTTTG>300

* * * * *
164>AGTATCAGCGACCTAATTTCTCTTAACCACGATTCTTCTGCTTATGACGACCAAGCTCTTAAAACCGAGTACAGTAGCTATGCACGACCCGGACCCCGT>263
301>AGTATCAGCGACCTAATTTCTCTTAACCACGATTCTTCTGCTTATGACGACCAAGCTCTTAAAACCGAGTACAGTAGCTATGCACGACCCGGACCCCGT>400

* * * * *
264>TGGATCTGGATCTGATTTTGGCCGGAAACCTAATTCGGATATGGAGGGAGAACCAGGAGTTGAGTATGGCCGGAAAACCTGAATCGGAGCATGGATCTGGC>363
401>TGGATCTGGATCTGATTTTGGCCGGAAACCTAATTCGGATATGGAGGGAGAACCAGGAGTTGAGTATGGCCGGAAAACCTGAATCGGAGCATGGATCTGGC>500

* * * * *
364>TATGGTGAAGAATTGAGAGCGATTACGTGAAGCCTAGCTATGGCCGTCACGAGGATGATGGTGACGATGGTCACAAAAACATAGTGGTAAGGATTATG>463
501>TATGGTGAAGAATTGAGAGCGATTACGTGAAGCCTAGCTATGGCCGTCACGAGGATGATGGTGACGATGGTCACAAAAACATAGTGGTAAGGATTATG>600

* * * * *
464>ATGATGGAGATGAGAAGAGTAAGAAGAAGGAGAAGGAGAAGAAGGATAAGAAGAAGATGGTAATAACTCTGAAGATGATGAGTTTAAGAAGAAGAA>563
601>ATGATGGAGATGAGAAGAGTAAGAAGAAGGAGAAGGAGAAGAAGGATAAGAAGAAGATGGTAATAACTCTGAAGATGATGAGTTTAAGAAGAAGAA>700

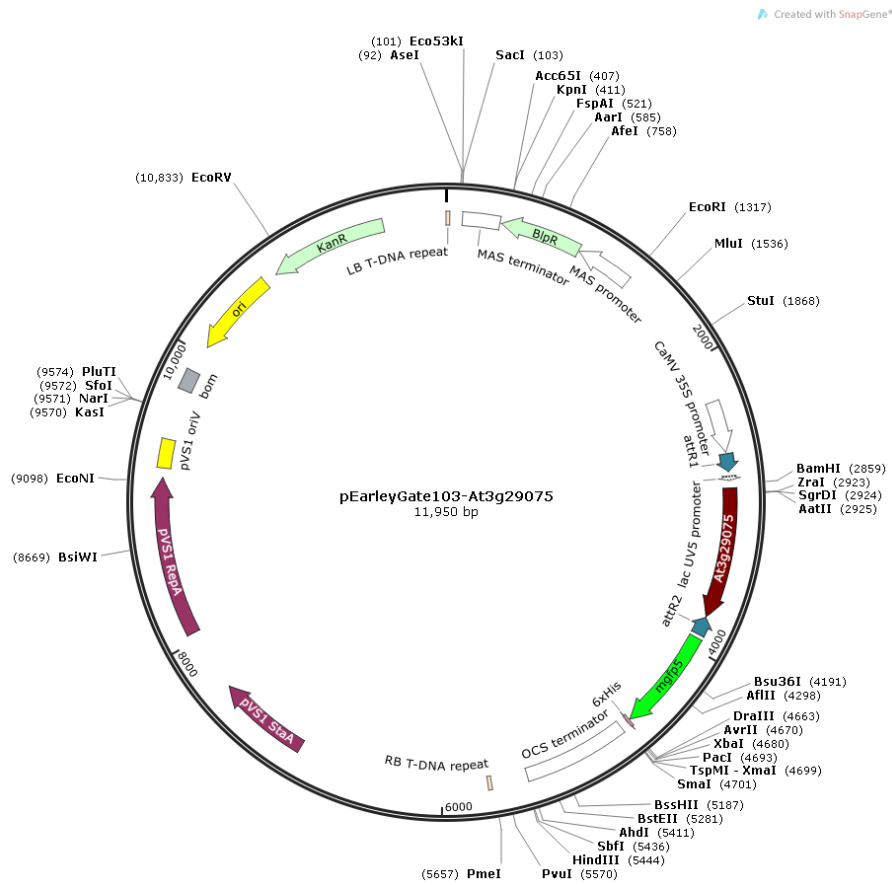
* * * * *
564>GAAGAAAGAGCAGTACAAGGAGCATCATGATGATGATGATTATGATGAGAAGAAGAAGAAGAAGAAAGACTATAATGATGATGATGAGAAGAAGAAGAA>663
701>GAAGAAAGAGCAGTACAAGGAGCATCATGATGATGATGATTATGATGAGAAGAAGAAGAAGAAGAAAGACTATAATGATGATGATGAGAAGAAGAAGAA>800

* * * * *
664>AAGCATTATAATGATGATGATGAGAAGAAGAAGCATAATTACAATGATGATGATGATGAGAAGAAGAAGAAGAGGATATCATGATGATGAGG>763
801>AAGCATTATAATGATGATGATGAGAAGAAGAAGCATAATTACAATGATGATGATGATGAGAAGAAGAAGAAGAGGATATCATGATGATGAGG>900

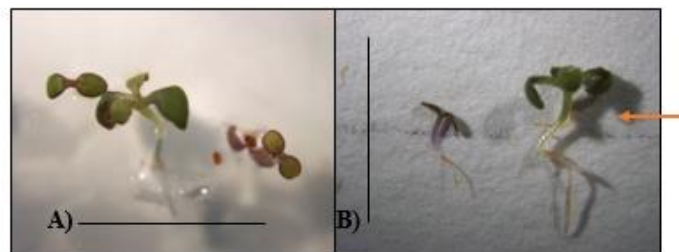
* * * * *
764>ATAAGAAGAAGAAGCACTATGATAATGATGATGATGAGAAGAAGAAGAAGGATCATCGTATGATGATGATGAGAAGAAGAAGAAGGATAA>863
901>ATAAGAAGAAGAAGCACTATGATAATGATGATGATGAGAAGAAGAAGAAGGATCATCGTATGATGATGATGAGAAGAAGAAGAAGGATAA>1000

* *
864>GCACCACAAGGGACATGACTAA>885
1001>GCACCc-----GA-----g~~~~~>1010
```

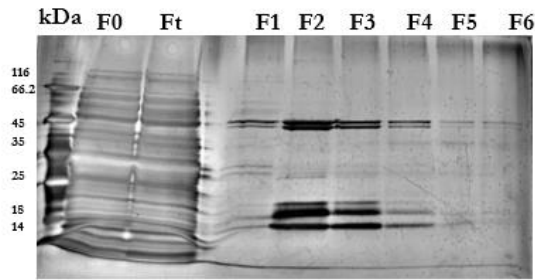
Supplementary figure 20: Sequence alignment of At3g29075 cDNA with pGEX4T2-At3g29075 insert shows 149 bp shorter.



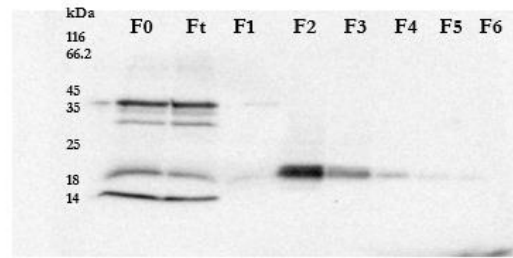
Supplementary figure 21: Construct of At3g29075-GFP in pEarleyGate103 gateway expression system.



Supplementary figure 22: Rapid selection of pEarlygate103-A3g29075-GFP overexpression line via Basta. Col-0 after rapid selection procedure (A); phosphinothricin acetyltransferase (basta) -resistant Arabidopsis seedling after rapid selection procedure (B). Size bars 1 cm. There was no GFP expression observed under the fluorescent microspore.

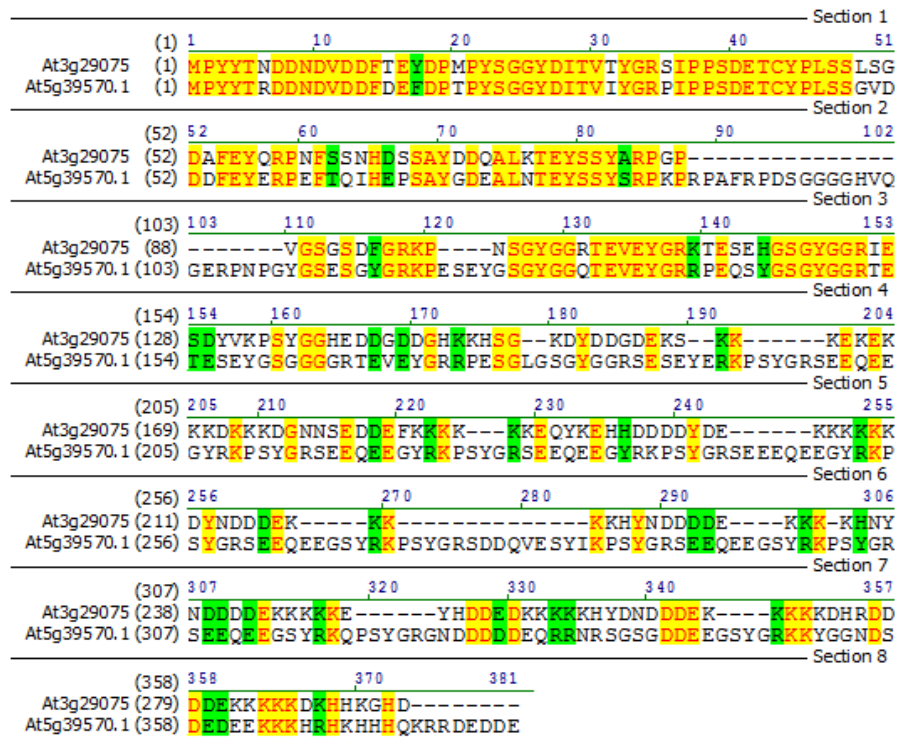


a) Silver staining - soluble fraction



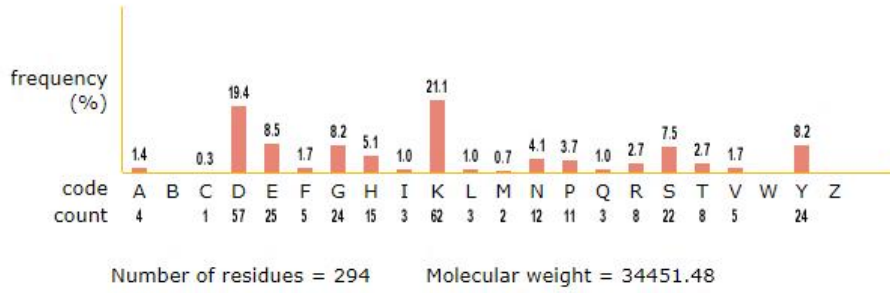
b) Immunodetection of At3g29075 1min 1:2000

Supplementary figure 24: The Ct-At3g29075 recombinant protein purification. A) Silver staining of soluble fraction from purified fractions. B) Immunodetection of recombinant Ct-at3g29075 protein fraction with Ct-At3g29075 antibody.



Supplementary figure 25: Alignment of PLDrp1 and At3g29075 protein.

COMPOSITION:



CALCULATION NOTES:

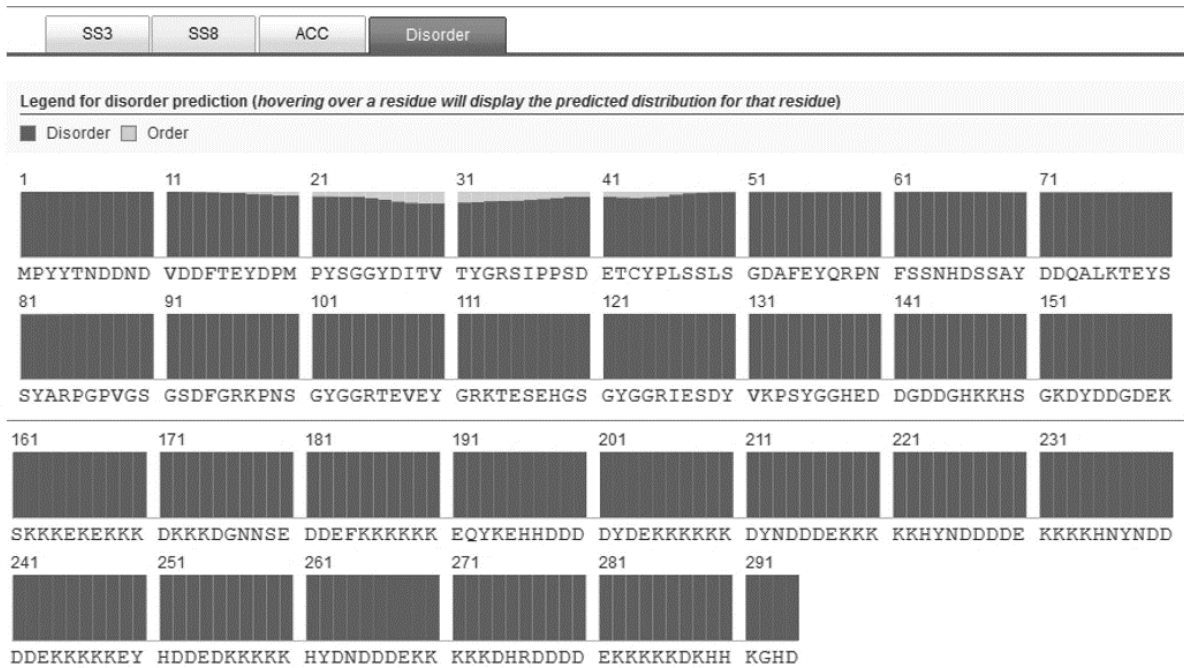
1. Molecular weight = $\text{sum of individual residues weights} - \text{water molecular weight} \times (\text{number of residues} - 1)$

where, *water molecular weight* = 18.015;

2. For each residue, the table gives the molecular weight:

A	Ala	89.09	G	Gly	75.07	N	Asn	132.12	V	Val	117.15
B	Asx	132.61	H	His	155.16	P	Pro	115.13	W	Trp	204.23
C	Cys	121.15	I	Ile	131.17	Q	Gln	146.15	Y	Tyr	181.19
D	Asp	133.10	K	Lys	146.19	R	Arg	174.20	Z	Glx	146.64
E	Glu	147.13	L	Leu	131.17	S	Ser	105.09			
F	Phe	165.19	M	Met	149.21	T	Thr	119.12	X	else	128.16

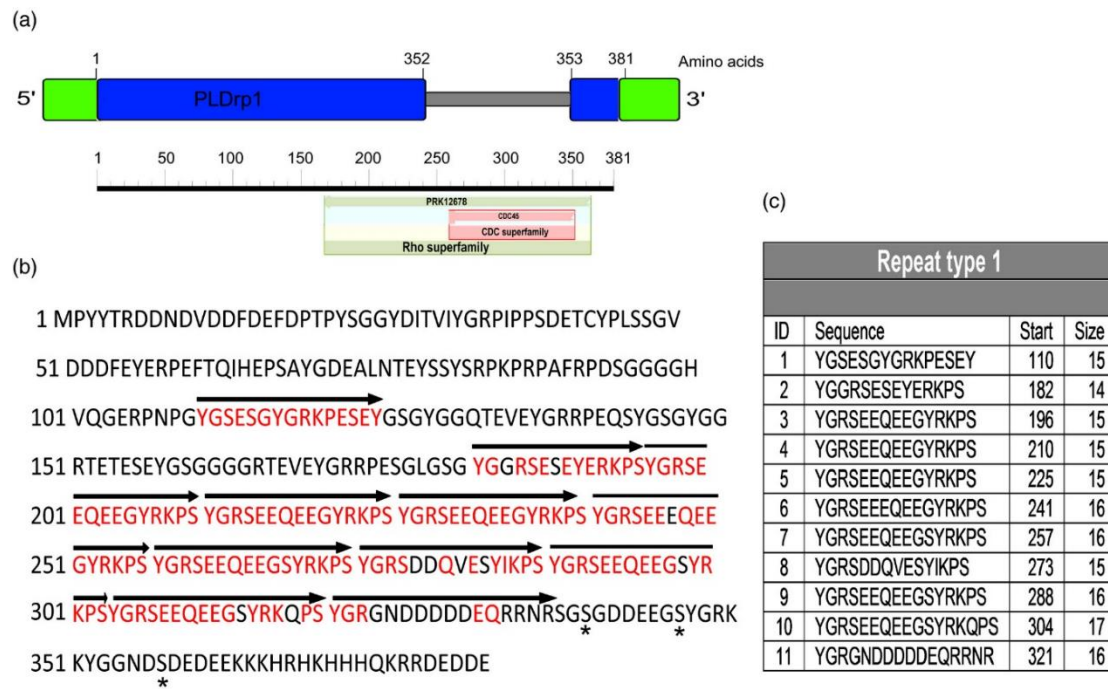
Supplementary figure 26: Amino acid composition of At3g29075 protein.



Supplementary figure 27: Disorder structural prediction of At3g29075 protein by using Raptor X property. The results of order/disorder region (DISCO) prediction showed 294 aa under the disordered region.

Supplementary table 28: List of genes involved in Gene co-expression network of At3g29075

Rank	Gene	Description
N/A	AT3G29075	Glycine-rich protein
1	LIL3:2	Chlorophyll A-B binding family protein
2	PAB8	Polyadenylate-binding protein 8
3	AT5G39570	Uncharacterized protein
4	MED19B	Probable mediator of RNA polymerase II transcription subunit 19b
5	AT3G58600	Adaptin ear-binding coat-associated protein 1 NECAP-1
6	AT2G44430	DNA-binding bromodomain-containing protein
7	AT5G52280	Myosin heavy chain-related protein
8	AT1G50950	Thioredoxin protein with a domain of the unknown function (DUF1692)
9	TPLATE	Protein TPLATE
10	AT3G50550	Unknown protein
11	RPOA DNA	Directed RNA polymerase subunit alpha
12	AT1G02330	Hepatocellular carcinoma-associated antigen 59
13	AT3G52220	Kinase phosphorylation domain
14	OHP2	One-helix protein 2
15	CRR6	Chlororespiratory reduction 6
16	AT5G24500	Unknown protein
17	AT5G11600	Unknown protein
18	AT3G09360	Cyclin/brf1-like tbp-binding protein
19	VAL1	B3 domain-containing transcription repressor VAL1
20	CYCT1-5	Cyclin-t1-5



Supplementary figure 29: Structure and amino acid sequence of phospholipase D-regulated protein1 (PLDrp1). (a) The gene model of PLDrp1 consists of two exons (blue) separated by an intron (grey). Untranslated regions are marked in green. (b) The translated protein sequence can be separated into two halves. Phosphorylated serine residues are marked by an asterisk. The C-terminal half of the protein contains 10 tandem repeats (indicated with the arrow) that are dominated by charged amino acids. The glycine-rich N-terminal region contains only one repeat unit. (c) Alignment of the tandem repeats. (Adopted from Ufer, 2017).

OAP03041.1	MPYYTNDNDVDDFTEYDMPYSGGYDITVTYGRSIPPSEETCYPLSSLSGDAFEYQRPN	60
At3g29075	MPYYTNDNDVDDFTEYDMPYSGGYDITVTYGRSIPPSEETCYPLSSLSGDAFEYQRPN	60

OAP03041.1	FSSNHDSAYDDQALKTEYSSYARPGPVGSGSDFGRKPNISGYGGRTEVEYGRKTESEHGS	120
At3g29075	FSSNHDSAYDDQALKTEYSSYARPGPVGSGSDFGRKPNISGYGGRTEVEYGRKTESEHGS	120

OAP03041.1	GYGGRIESDYVKPSYGGHEDDGDGHKKHSGKDYDDGDEKSKKKEKSKKDKKKDGMNSE	180
At3g29075	GYGGRIESDYVKPSYGGHEDDGDGHKKHSGKDYDDGDEKSKKKEKSKKDKKKDGMNSE	180

OAP03041.1	DDEFKKKKKKEQYKEHHDDDDYDEKSKKKKDYNDDEKSKKKKHYNDDDEKSKKXKSKK	240
At3g29075	DDEFKKKKKKEQYKEHHDDDDYDEKSKKKKDYNDDEKSKKKKHYNDDDEKSKK-----	232

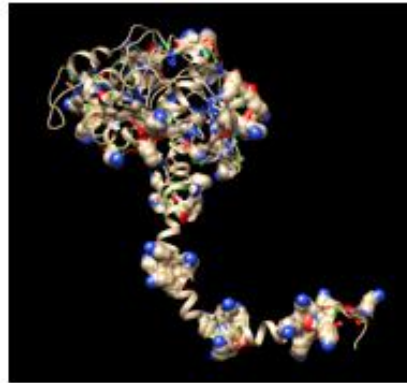
OAP03041.1	KHYNDDDEKSKKKKHYNDDDEKSKKKKEYHDEDEKSKKKKHYNDDDEKSKKKDHRDDDEK	300
At3g29075	-----KKNHYNDDDEKSKKKKEYHDEDEKSKKKKHYNDDDEKSKKKDHRDDDEK	282
*::*****		
OAP03041.1	KKKKDKHHKGGH	312
At3g29075	KKKKDKHHKGGH	294

Supplementary figure 30: Protein Blast for C terminal part of the At3g29075 showed 88% identical ortholog protein *AXX17_At3g31880* from *Arabidopsis thaliana* Ecotype *Landsberg erecta*.

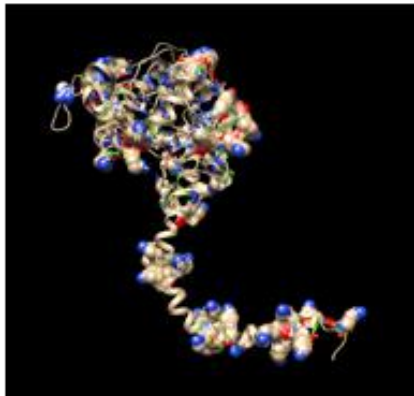
A) Arg – 7 out of 8 aa



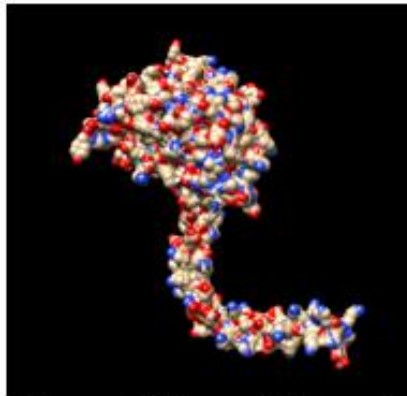
B) Lysine – 57 out of 66 aa



C) Lys and Arg - 64 aa



D) At3g29075 protein



Source: Chimera 1.14 version

Supplementary figure 31: Electrostatic surface model of At3g29075 protein. A) The Arginines are highlighted in the At3g29075 model; 7 out of 8 arginines are located in the positively charged surface of the At3g29075 protein. B) 57 out of 66 lysines are located in the positively charged At3g29075 protein's surface; 13 positively charged clusters (Zheng, 2004) are made out of 57 lysines in the At3g29075 protein's surface. C) In total 64 aa (57 lysines and 7 arginines) are located in the positively charged surface of the At3g29075 protein. D) The electrostatic surface model of complete At3g29075 protein. Blue colour represents positive clusters. Red colour represents the negative clusters. Source: Chimera 1.14.

```
>At3g29075
MPYYTNDNDVDDFTEYDMPYSGGYDITVTYGRSIPPSDETCLPLSSLSGDAFEYQRPNFSSNHDSSAYDD
QALKTEYSSYARPGPVGSGSDFGRKPNNGYGGRTVEVEYGRKTESEHGSGYGGRIESDYVKPSYGGHEDDG
DDGHKKHSGKDYDDGDEKSKKKEKEKKDKKKDGNNSEDDFKKKKKEQYKEHHDDDDYDEKK
KKKDYNDDEKKKKHYNDDEKHKHNYNDDEKKKKKEYHDEDEKKKKHYNDDEK
KKKDHRRDDDEKKKKKDKHHKGGHD
```

Supplementary figure 32: Lysine clusters of At3g29075 protein are marked with yellow colour. Lysine amino acids are written in red colour. Arginine amino acids are written in green colour. Lysine clusters are marked based on Zheng 2004 classification of amino acid clusters.

8. Reference

- Aloulou, A., Ali, Y. Ben, Bezzine, S., Gargouri, Y., and Gelb, M.H. (2012). Phospholipases: An Overview. In, pp. 63–85.
- Andersen, S.U., Buechel, S., Zhao, Z., Ljung, K., Novák, O., Busch, W., Schuster, C., and Lohmann, J.U. (2008). Requirement of B2-type cyclin-dependent kinases for meristem integrity in *Arabidopsis thaliana*. *Plant Cell* **20**: 88–100.
- Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L.E., and Forman-Kay, J.D. (2015). Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**: 106–109.
- Bargmann, B.O.R. and Munnik, T. (2006). The role of phospholipase D in plant stress responses. *Curr. Opin. Plant Biol.* **9**: 515–522.
- Basu, S., Ramegowda, V., Kumar, A., and Pereira, A. (2016). Plant adaptation to drought stress [version 1; referees: 3 approved]. *F1000Research* **5**.
- Bhaskara, G.B., Wen, T.N., Nguyen, T.T., and Verslues, P.E. (2017). Protein phosphatase 2Cs and microtubule-associated stress protein 1 control microtubule stability, plant growth, and drought response. *Plant Cell* **29**: 169–191.
- Braun, V.A. (2017). Characterisation of a MAPKK kinase from *Cratogeomys plantagineum* Interaction of VIK kinases and LEA proteins Dissertation.
- Bray, E.A. (2004). Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. In *Journal of Experimental Botany*, pp. 2331–2341.
- Bray, E.A. (1997). Plant responses to water deficit. *Trends Plant Sci.* **2**: 48–54.
- Chakraborty, M. and Jiang, X.-C. (2013). *Lipid-mediated Protein Signaling* D.G.S. Capelluto, ed (Springer Netherlands).
- Cooper, G. (2019). *The Cell - Molecular Approach*- 8th edition. Sinauer Assoc. Inc.
- Cruz, E.R., Nguyen, H., Nguyen, T., and Wallace, I.S. (2019). Functional analysis tools for post-translational modification: a post-translational modification database for analysis of proteins and metabolic pathways. *Plant J.*
- Devaiah, S.P., Roth, M.R., Baughman, E., Li, M., Tamura, P., Jeannotte, R., Welti, R., and Wang, X. (2006). Quantitative profiling of polar glycerolipid species from organs of wild-type *Arabidopsis* and a PHOSPHOLIPASE D α 1 knockout mutant. *Phytochemistry* **67**: 1907–1924.
- Ding, Y., Liu, N., Virilouvet, L., Riethoven, J.J., Fromm, M., and Avramova, Z. (2013). Four distinct types of dehydration stress memory genes in *Arabidopsis thaliana*. *BMC Plant Biol.* **13**.
- Doherty, A.J., Connolly, B.A., and Worrall, A.F. (1993). Overproduction of the toxic protein, bovine pancreatic DNaseI, in *Escherichia coli* using a tightly controlled T7-promoter-based vector. *Gene* **136**: 337–340.
- Doherty, C. et al. (2017). MATI, a novel protein involved in the regulation of herbivore-associated signaling pathways. *J. Exp. Bot.* **0747**: 1–7.
- Domínguez-González, I., Vázquez-Cuesta, S.N., Algaba, A., and Díez-Guerra, F.J. (2007). Neurogranin binds to phosphatidic acid and associates to cellular membranes. *Biochem. J.* **404**: 31–43.
- Dong, H., Nilsson, L., and Kurland, C.G. (1995). Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J. Bacteriol.* **177**: 1497–1504.
- Dumon-Seignovert, L., Cariot, G., and Vuillard, L. (2004). The toxicity of recombinant proteins in *Escherichia coli*: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* **37**: 203–206.
- Guo, L., Mishra, G., Taylor, K., and Wang, X. (2011). Phosphatidic acid binds and stimulates

- arabidopsis sphingosine kinases. *J. Biol. Chem.* **286**: 13336–13345.
- Guo, L. and Wang, X.** (2012). Crosstalk between Phospholipase D and Sphingosine Kinase in Plant Stress Signaling. *Front. Plant Sci.* **3**: 51.
- Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R., Meybeck, A., and Rome, F.** (2011). Food losses and food waste: extent, causes and prevention. (Rome).
- Harrison, S.J., Mott, E.K., Parsley, K., Aspinall, S., Gray, J.C., and Cottage, A.** (2006). A rapid and robust method of identifying transformed *Arabidopsis thaliana* seedlings following floral dip transformation. *Plant Methods* **2**: 1–7.
- Hejgaard, J. and Boisen, S.** (2009). High-lysine proteins in Hiproly barley breeding: Identification, nutritional significance and new screening methods. *Hereditas* **93**: 311–320.
- Hejgaard, J. and Boisen, S.** (1980). High-lysine proteins in Hiproly barley breeding: Identification, nutritional significance and new screening methods. *Hereditas* **93**: 311–320.
- Hirayama, T., Ohto, C., Mizoguchi, T., and Shinozaki, K.** (1995). A gene encoding a phosphatidylinositol-specific phospholipase C is induced by dehydration and salt stress in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 3903–3907.
- Hong, Y., Devaiah, S.P., Bahn, S.C., Thamasandra, B.N., Li, M., Welti, R., and Wang, X.** (2009). Phospholipase D ϵ and phosphatidic acid enhance *Arabidopsis* nitrogen signaling and growth. *Plant J.* **58**: 376–387.
- Hong, Y., Pan, X., Welti, R., and Wang, X.** (2008). Phospholipase D α 3 Is Involved in the Hyperosmotic Response in *Arabidopsis*. *Plant Cell* **20**.
- Hong, Y., Zhang, W., and Wang, X.** (2010). Phospholipase D and phosphatidic acid signalling in plant response to drought and salinity. *Plant, Cell Environ.* **33**: 627–635.
- Hou, Q., Ufer, G., and Bartels, D.** (2016). Lipid signalling in plant responses to abiotic stress. *Plant Cell Environ.* **39**: 1029–1048.
- Izumi, Y., Chibata, I., and Itoh, T.** (1978). Herstellung und Verwendung von Aminosäuren. *Angew. Chemie* **90**: 187–194.
- Jang, J.H., Lee, C.S., Hwang, D., and Ryu, S.H.** (2012). Understanding of the roles of phospholipase D and phosphatidic acid through their binding partners. *Prog. Lipid Res.* **51**: 71–81.
- Jia, M., Wu, H., Clay, K.L., Jung, R., Larkins, B.A., and Gibbon, B.C.** (2013). Identification and characterization of lysine-rich proteins and starch biosynthesis genes in the opaque2 mutant by transcriptional and proteomic analysis. *BMC Plant Biol.* **13**: 60.
- Johansson, O.N., Fahlberg, P., Karimi, E., Nilsson, A.K., Ellerström, M., and Andersson, M.X.** (2014). Redundancy among phospholipase D isoforms in resistance triggered by recognition of the *Pseudomonas syringae* effector AvrRpm1 in *Arabidopsis thaliana*. *Front. Plant Sci.* **5**: 1–9.
- Joint** (2007). Protein and amino acid requirements in human nutrition (Geneva).
- Joosen, R.V.L., Kodde, J., Willems, L.A.J., Ligterink, W., and Hilhorst, H.W.M.** (2010). The Germinator automated germination scoring system. *Seed Test. Int.* **140**: 4–8.
- Julkowska, M.M., McLoughlin, F., Galvan-Ampudia, C.S., Rankenberg, J.M., Kawa, D., Klimecka, M., Haring, M.A., Munnik, T., Kooijman, E.E., and Testerink, C.** (2015). Identification and functional characterization of the *Arabidopsis* Snf1-related protein kinase SnRK2.4 phosphatidic acid-binding domain. *Plant. Cell Environ.* **38**: 614–624.
- Katagiri, T., Ishiyama, K., Kato, T., Tabata, S., Kobayashi, M., and Shinozaki, K.** (2005). An important role of phosphatidic acid in ABA signaling during germination in *Arabidopsis thaliana*. *Plant J.* **43**: 107–117.
- Katagiri, T., Takahashi, S., and Shinozaki, K.** (2001a). Involvement of a novel *Arabidopsis* phospholipase D, AtPLDdelta, in dehydration-inducible accumulation of phosphatidic acid in stress signalling. *Plant J.* **26**: 595–605.
- Kong, H.** (2000). Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res.* **28**: 3216–3223.

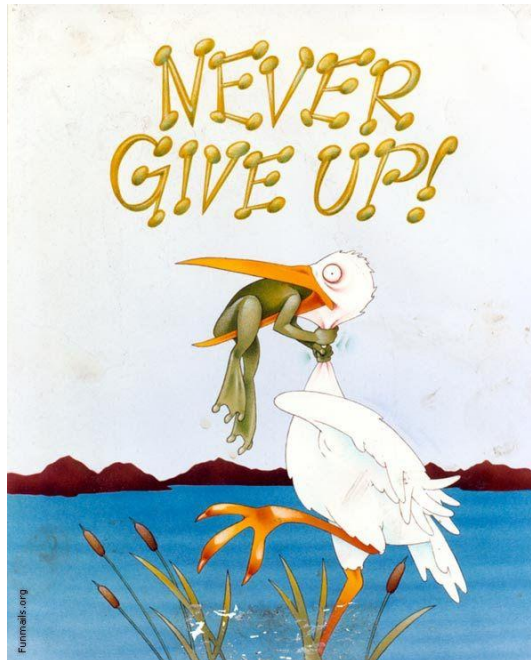
- Kooijman, E.E., Chupin, V., de Kruijff, B., and Burger, K.N.J.** (2003). Modulation of membrane curvature by phosphatidic acid and lysophosphatidic acid. *Traffic* **4**: 162–174.
- Kooijman, E.E., Tieleman, D.P., Testerink, C., Munnik, T., Rijkers, D.T.S., Burger, K.N.J., and De Kruijff, B.** (2007). An electrostatic/hydrogen bond switch as the basis for the specific interaction of phosphatidic acid with proteins. *J. Biol. Chem.* **282**: 11356–11364.
- Koornneef, M. and Meinke, D.** (2010). The development of *Arabidopsis* as a model plant. *Plant J.* **61**: 909–921.
- Laugesen, S., Messinese, E., Hem, S., Pichereaux, C., Grat, S., Ranjeva, R., Rossignol, M., and Bono, J.J.** (2006). Phosphoproteins analysis in plants: A proteomic approach. *Phytochemistry* **67**: 2208–2214.
- Van Der Lee, R. et al.** (2014). Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**: 6589–6631.
- Li, M., Hong, Y., and Wang, X.** (2009). Phospholipase D- and phosphatidic acid-mediated signaling in plants. *Biochim. Biophys. Acta* **1791**: 927–35.
- Li, Q., Li, J., Liu, S., Huang, J., Lin, H., Wang, K., Cheng, X., and Liu, Z.** (2015). A comparative proteomic analysis of the buds and the young expanding leaves of the tea plant (*Camellia sinensis* L.). *Int. J. Mol. Sci.* **16**: 14007–14038.
- Li, W., Li, M., Zhang, W., Welti, R., and Wang, X.** (2004). The plasma membrane-bound phospholipase Ddelta enhances freezing tolerance in *Arabidopsis thaliana*. *Nat. Biotechnol.* **22**: 427–433.
- Liu, Q., Zhang, C., Yang, Y., and Hu, X.** (2010). Genome-wide and molecular evolution analyses of the phospholipase D gene family in Poplar and Grape. *BMC Plant Biol.* **10**: 117.
- Liu, X., Ma, D., Zhang, Z., Wang, S., Du, S., Deng, X., and Yin, L.** (2019). Plant lipid remodeling in response to abiotic stresses. *Environ. Exp. Bot.* **165**: 174–184.
- Mangeon, A., Junqueira, R.M., and Sachetto-Martins, G.** (2010). Functional diversity of the plant glycine-rich proteins superfamily. *Plant Signal. Behav.* **5**: 99–104.
- Mehlmer, M.N.** (2008). Ca²⁺ Dependent Protein Kinases in *Arabidopsis thaliana*.
- Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D., and Koornneef, M.** (1998b). *Arabidopsis thaliana*: A model plant for genome analysis. *Science* (80-.). **282**: 662+679–68.
- Miles, R.D. and Chapman, F.A.** (2008). The Concept of Ideal Protein in Formulation of Aquaculture Feeds 1.
- Miller, W.A. and Dinesh-Kumar, S.P.** (2019). A new mechanism for translational control in plants. *FEBS J.* **286**: 3775–3777.
- Miroux, B. and Walker, J.E.** (1996). Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**: 289–298.
- Mulekar, J.J., Bu, Q., Chen, F., and Huq, E.** (2012). Casein kinase II α subunits affect multiple developmental and stress-responsive pathways in *Arabidopsis*. *Plant J.* **69**: 343–354.
- Munnik, T.** (2001). Phosphatidic acid: An emerging plant lipid second messenger. *Trends Plant Sci.* **6**: 227–233.
- Munnik, T., Meijer, H.J.G., Riet, B. Ter, Hirt, H., Frank, W., Bartels, D., and Musgrave, A.** (2000). Hyperosmotic stress stimulates phospholipase D activity and elevates the levels of phosphatidic acid and diacylglycerol pyrophosphate. *Plant J.* **22**: 147–154.
- Nakashima, K., Yamaguchi-Shinozaki, K., and Shinozaki, K.** (2014). The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front. Plant Sci.* **5**.
- O'Connor, C.D. and Timmis, K.N.** (1987). Highly repressible expression system for cloning genes that specify potentially toxic proteins. *J. Bacteriol.* **169**: 4457–4462.
- Ohlogge, J. and Browse, J.** (1995). Lipid biosynthesis. *Plant Cell* **7**.
- Ovie, S.O. and Eze, S.S.** (2013). Lysine Requirement and its Effect on the Body Composition of *Oreochromis niloticus* Fingerlings. *J. Fish. Aquat. Sci.* **8**: 94–100.

- Petersen, J., Eriksson, S.K., Harryson, P., Pierog, S., Colby, T., Bartels, D., and Röhrig, H.** (2012). The lysine-rich motif of intrinsically disordered stress protein CDeT11-24 from *Craterostigma plantagineum* is responsible for phosphatidic acid binding and protection of enzymes from damaging effects caused by desiccation. *J. Exp. Bot.* **63**: 4919–4929.
- Phospholipase D and phosphatidic acid signalling in plant response to drought and salinity.** - PubMed - NCBI.
- Pinosa, F., Buhot, N., Kwaaitaal, M., Fahlberg, P., Thordal-Christensen, H., Ellerstrom, M., and Andersson, M.X.** (2013). Arabidopsis phospholipase Ddelta is involved in basal defense and non-host resistance to powdery mildew fungi. *Plant Physiol.* **163**: 896–906.
- Potocký, M., Pleskot, R., Pejchar, P., Vitale, N., Kost, B., and Zárský, V.** (2014). Live-cell imaging of phosphatidic acid dynamics in pollen tubes visualized by Spo20p-derived biosensor. *New Phytol.* **203**: 483–94.
- Prelich, G.** (2012). Gene overexpression: Uses, mechanisms, and interpretation. *Genetics* **190**: 841–854.
- Qin, C. and Wang, X.** (2002). The Arabidopsis phospholipase D family. Characterization of a calcium-independent and phosphatidylcholine-selective PLD zeta 1 with distinct regulatory domains. *Plant Physiol.* **128**: 1057–1068.
- Reinders, J. and Sickmann, A.** (2007a). Modificomics: Posttranslational modifications beyond protein phosphorylation and glycosylation. *Biomol. Eng.* **24**: 169–177.
- Röhrig, H., Colby, T., Schmidt, J., Harzen, A., Facchinelli, F., and Bartels, D.** (2008). Analysis of desiccation-induced candidate phosphoproteins from *Craterostigma plantagineum* isolated with a modified metal oxide affinity chromatography procedure. *Proteomics* **8**: 3548–3560.
- Ruano, M.L.F., Pérez-Gil, J., and Casals, C.** (1998). Effect of acidic pH on the structure and lipid binding properties of porcine surfactant protein A: Potential role of acidification along its exocytic pathway. *J. Biol. Chem.* **273**: 15183–15191.
- Sachetto-Martins, G., Franco, L.O., and De Oliveira, D.E.** (2000). Plant glycine-rich proteins: A family or just proteins with a common motif? *Biochim. Biophys. Acta - Gene Struct. Expr.* **1492**: 1–14.
- Showe, M.K. and Onorato, L.** (1978). Kinetic factors and form determination of the head of bacteriophage T4. *Proc. Natl. Acad. Sci. U. S. A.* **75**: 4165–4169.
- Singh, A., Bhatnagar, N., Pandey, A., and Pandey, G.K.** (2015). Plant phospholipase C family: Regulation and functional role in lipid signaling. *Cell Calcium* **58**: 139–146.
- Soitamo, A.J., Piippo, M., Allahverdiyeva, Y., Battchikova, N., and Aro, E.M.** (2008). Light has a specific role in modulating Arabidopsis gene expression at low temperature. *BMC Plant Biol.* **8**.
- Tao, X., Jia, N., Cheng, N., Ren, Y., Cao, X., Liu, M., Wei, D., and Wang, F.Q.** (2017). Design and evaluation of a phospholipase D based drug delivery strategy of novel phosphatidyl-prodrug. *Biomaterials* **131**: 1–14.
- Testerink, C. and Munnik, T.** (2011). Molecular, cellular, and physiological responses to phosphatidic acid formation in plants. *J. Exp. Bot.* **62**: 2349–61.
- Tompa, P.** (2003). Intrinsically unstructured proteins evolve by repeat expansion. *BioEssays* **25**: 847–855.
- Tsien, R.Y.** (1998). The green fluorescent protein. *Annu. Rev. Biochem.* **67**: 509–544.
- Tuteja, N.** (2007). Abscisic Acid and Abiotic Stress Signaling. *Plant Signal. Behav.* **2**: 135–138.
- Ufer, G.** (2015). Proteins under the control of phospholipase D in *Arabidopsis thaliana*.
- Ufer, G., Gertzmann, A., Gasulla, F., Röhrig, H., and Bartels, D.** (2017a). Identification and characterization of the phosphatidic acid-binding *A. thaliana* phosphoprotein PLDrp1 that is regulated by PLD α 1 in a stress-dependent manner. *Plant J.* **92**: 276–290.
- Vergnolle, C., Vaultier, M., Tacconnat, L., Renou, J., Kader, J., Zachowski, A., Ruelland,**

- E., Vergnolle, C., Vaultier, M., Tacconat, L., Renou, J., and Kader, J.** (2016). The Cold-Induced Early Activation of Phospholipase C and D Pathways Determines the Response of Two Distinct Clusters of Genes in Arabidopsis Cell Suspensions Published by : American Society of Plant Biologists (ASPB) Stable URL : <http://www.jstor.org/st>. *Plant Physiol* **139**: 1217–1233.
- Walton, A. et al.** (2016). It's time for some “site”-seeing: Novel tools to monitor the ubiquitin landscape in Arabidopsis thaliana. *Plant Cell* **28**: 6–16.
- Wang, C. and Wang, X.** (2001). A novel phospholipase D of *Arabidopsis thaliana* that is activated by oleic acid and associated with the plasma membrane. *Plant Physiol.* **127**: 1102–1112.
- Wang, G., Ryu, S., and Wang, X.** (2012). Plant phospholipases: an overview. *Methods Mol. Biol.* **861**: 123–137.
- Wang, X.** (2014). Phospholipases in Plant Signaling. Springer **20**: 232pp.
- Wang, X., Devaiah, S.P., Zhang, W., and Welti, R.** (2006). Signaling functions of phosphatidic acid. *Prog. Lipid Res.* **45**: 250–278.
- Wang, X., Guo, L., Wang, G., and Li, M.** (2014). PLD: Phospholipase Ds in Plant Signaling. In, pp. 3–26.
- Watson, V.H. and Foglesong, R.H.** (2007). Catfish Protein Nutrition: Revised CONTENTS.
- Weirauch, M.T.** (2011). Gene Coexpression Networks for the Analysis of DNA Microarray Data. In *Applied Statistics for Network Biology: Methods in Systems Biology* (Wiley-VCH), pp. 215–250.
- Wootton, J.C.** (1994). Sequences with “unusual” amino acid compositions. *Curr. Opin. Struct. Biol.* **4**: 413–421.
- Young, V.R. and Pellett, P.L.** (1994). Plant proteins in relation to human protein and amino acid nutrition. *Am. J. Clin. Nutr.* **59**.
- Zachgo, E.A., Wang, M.L., Dewdney, J., Bouchez, D., Camilleri, C., Belmonte, S., Huang, L., Dolan, M., and Goodman, H.M.** (1996). A physical map of chromosome 2 of *Arabidopsis thaliana*. *Genome Res.* **6**: 19–25.
- Zhang, T., Faraggi, E., Li, Z., and Zhou, Y.** (2013). Intrinsically Semi-disordered State and Its Role in Induced Folding and Protein Aggregation. *Cell Biochem. Biophys.* **67**: 1193–1205.
- Zheng, W.M.** (2004). Clustering of amino acids for protein secondary structure prediction. *J. Bioinform. Comput. Biol.* **2**: 333–342.
- Zhu, J.-K.** (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **53**: 247–273.

Acknowledgements

In this Ph.D. thesis, I felt it has been extended and gruelling, smooth sailing sometimes, full of thorns, impeding the pace and resulting indented and reduced aspirations, procrastination, tiredness, temptations to quit. There was my favourite quote, “NEVER GIVE UP” kept me going!



Source: funmails.org

Consequently, many individuals made the completion of this thesis possible, more than can be acknowledged by names, whose contributions provided pillars of direct and indirect supports throughout the journey to its present shape.

At this moment, I heartily express my gratitude to all those kind souls who stood by me.