Functional characterization of DUF642 genes in *Arabidopsis thaliana*

by

Shanjida Khan

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Plant Biology

Department of Biological Sciences
University of Alberta

ABSTRACT

DOMAIN OF UNKNOWN FUNCTION 642, (DUF642), is an uncharacterized protein family in the Pfam database, a large collection of protein families. In Arabidopsis thaliana there are 10 proteins that contain DUF642 domains. DUF642 appears to be specific to plants and is present in gymnosperms, monocots and dicots. The present study was designed to investigate the biochemical function and physiological role of two Arabidopsis DUF642 genes, AT5G25460 (*DGR2*) and AT5G11420 (*DUFB*). The three dimensional structures of DGR2 and DUFB proteins were predicted by I-TASSER and validated by Ramachandran plot using RAMPAGE Server. Structural models indicated that DGR2 and DUFB showed similarity to carbohydrate binding proteins with hydrolase and carbon-oxygen lyase activity, respectively. Translation fusions with reporter genes showed a punctate pattern of subcellular localization within the cytoplasm, which did not co-localize with the Sec21 Golgi marker but was immediately adjacent to each other, suggesting DGR2 localizes to the trans-Golgi network. DGR2 and DUFB were expressed heterologously in *E. coli* but sufficient purified proteins could not be obtained for downstream functional assays. Although both DGR2 and DUFB have high sequence similarity (93.4% nucleotide identity), and were presumed to be paralogs, they were expressed in complementary spatial domains according to expression patterns of reporter genes and qRT-PCR analysis. GUS reporter fusions of *DGR2* were expressed in the root apex and the later stages of lateral root primordial (LRP) where cells are dividing and elongating whereas *DUFB* was expressed in the elongating tissues of roots and not in LRP. Both genes were responsive to auxin identified by reporter gene assay where promoter region sequences of DGR2 and DUFB were fused to the GUS gene. T-DNA insertion mutants (Salk_042864 for DGR2 and Salk_094931 for

DUFB), RNAi, and *DGR2* and *DUFB* overexpressing plants showed no morphological differences from wild-type phenotypes, therefore, metabolic profiling of these mutant plants was performed by gas chromatography/mass spectrometry to reveal metabotypes. The results suggested that *DGR2* and *DUFB* both affected TCA cycle intermediates and were involved in the carbohydrate metabolism. In addition to *DGR2* and *DUFB* characterization in Arabidopsis, a portion of Ph.D. research work included metabolic profiling of developing flax seeds by GC/MS.

In summary, we have generated predicted 3D models of DGR2 and DUFB. Subcellular localization revealed that DGR2 possibly localizes to the trans-Golgi network. We speculated that DGR2 is required for cell elongation and division whereas DUFB is required for cell elongation in Arabidopsis. Metabolite profiling of the Arabidopsis T-DNA insertion mutants, RNAi and overexpression plants of *DGR2* and *DUFB* reveals metabolic phenotypes previously unidentified and illustrate perturbation of TCA cycle. Both proteins may influence growth by modifying probably pectin thereby perturbed primary metabolic processes. Both genes are expressed *E. coli* but failed to obtain purified proteins. The metabolic profiling of developing flax seeds by GC/MS identified unique metabolites and the pathways perturbation during different stages of flax seed development.

Dedication

This thesis is proudly dedicated

To my husband, Sajjad Khandker

And

To our two children, Samiha Khandker and Sakib Khandker

ACKNOWLEDGEMENT

First and foremost I want to express my heartfelt thanks and gratitude to my supervisor, Dr. Michael Deyholos. It has been an honor to be his Ph.D. student. Throughout my Ph.D. study, he has been actively interested in my research work and has always been provided patience guidance encouragement and outstanding advices. His enthusiasm and passion for his research was contagious and motivational for me. I greatly appreciate all his contributions of time, ideas, and funding which enable me to peruse my Ph.D. successfully.

I would also like to thank my committee members, Dr. Shelagh Campbell and Dr. Enrico Scarpella, for their valuable advices and encouragement. I would like to thank Dr. Enrico Scarpella who helped me and trained me on focal microscopy which enabled me to perform the subcellular localization and co-localization studies discussed in this dissertation. I would like to thank Dr. Ramanarayan Krishnamurthy who trained me how to perform metabolic profiling using GC/MS and provided suggestion during the metabolic profiling studies in my Ph.D. work. I also like to thank Erik Bergen who helped me to set up GC/MS apparatus. I would like to thank Dr. Muhammad Rahman for his suggestion and help during the heterologous protein expression and purification experiment.

I would like to thank all past and present members of Dr. Deyholos lab (Mary De Pauw, Dr. Neil Hobson, Dr. David Pinzon, and Dr. Joshua McDill, Haiyan Zhuang, Kashfia Faruque Zhu Dan, Ningyu Zhang, Anupreeti Ramdoss, Dr. Abdur Rashid, Dr. Manjeet Kumari, Leonardo Galindo, Dr. Rowshon) for their immense contribution to my professional time and good advices during my stay in this lab.

Finally, I would like to thank my family for all their support and encouragement during the tough times in the Ph.D. pursuit.

Table of Contents

List of Figures

List of Tables

List of Abbreviation

DAS	Days after sowing
GC/MS	Gas chromatography–mass spectrometry
GO	Gene Ontology
CiFP	citrine fluorescent protein
TGN	Trans-Golgi network
ER	Endoplasmic reticulum
IAA	Indole-3-acetic acid
MeJA	Methyl jasmonate
ACC	1-aminocyclopropane-1-carboxylic acid
GA	Gibberellic acid
Glc	Glucose
Gal	Galactose
Ara	Arabinose
LGalL	L-galactono-1,4-lactone
MS media	Murashige & Skoog media
T_1 T_2 T_3	Transgenic progeny generation
WT	Wild type
CDS	Coding sequence
ANOVA	Analysis of variance
SDS	Sodium dodecyl sulphate
PAGE	polyacrylamide gel electrophoresis
qRT-PCR	Quantitative real time RT-PCR
IPTG	Isopropylthio-β-galactoside
HCA	hierarchical cluster analysis
PCA	Principal component analysis
DUF	Domain of unknown function
TAIR	The Arabidopsis Information Resource
PBS	phosphate buffered saline

TBS	Tris-Buffered Saline
Col-0	Columbia
Cor	Correlation coefficient

1. Chapter1

1.1. Literature review

Advancement of genome sequencing technologies has made it possible to identify thousands of genes that contribute to different biological processes. Genome sequencing projects are adding nucleotide sequences to the publicly available databases. Among them, many genes are not characterized yet. The challenge comes in respect to characterizing those unknown genes. The uncharacterized genes can be attributed to many important functions. Therefore, it is important to know how the genes and the proteins they encode function in the organism, to get new insight into the related biological processes. In Arabidopsis, approximately half of the ~26,000 predicted genes can be matched by sequence similarity to genes of known function, and less than 10% of the original gene annotations were supported by experimental evidence (Ostergaard and Yanofsky, 2004). Bork et al. (1998) suggested protein annotation categories which were: 1. Molecular function: biochemical function performed by a protein, 2: Cellular function, such as metabolic pathways and signal transduction, 3. Phenotypic function: includes interactions of physiological system with the surrounding environment that determine the phenotyping properties and behavior of the organism. The Gene Ontology (GO) Consortium (www.geneontology.org) has developed a systematic and standardized nomenclature with the goal of annotating genes and gene products with precisely defined, common, controlled vocabulary in any organisms. Go Ontology consists of three domains-molecular function, biological process, and cellular component. (Ashburner et al. 2000; The Gene Ontology Consortium 2001). Functional annotation based on GO is now widely accepted among the scientific communities and used in biological databases annotation projects and computational analyses (Tanya et al., 2004; Rhee et al., 2008). The annotation of protein function is based on experimental procedures which mainly focused on a single gene or protein, or a small set of genes or proteins in a family. The functional annotation of protein through *in vitro* experimental determination is an expensive and time consuming procedure.

1.1.1. Domains of unknown function (DUFs)

Proteins are built with one or many domains that serve as functional entities. The classical definition of a domain denotes independent folding and tertiary structure. More recently, domains have come to be defined based on sequence alone (Sigrist et al., 2010; Kessel and Ben-Tal, 2011; Goodacre et al, 2013). There are different sequenced-based databases of protein domains, although these have substantial overlap with each other (Mulder et al., 2008) including: Interpro and Pfam databases. Pfam is composed of two types of subdivisions, Pfam-A and Pfam-B. Proteins domains in both Pfam-A and Pfam-B are defined by Hidden Markov Models (HMMs), but the Pfam-A domains are manually curated and annotated, whereas the Pfam-B domains are generated automatically and may be less reliable. To define a Pfam-A domain, a seed alignment is generated by multiple alignment of a set of representative sequences and performed manually. Next, using the seed alignment HMM is developed for database searching and alignment purposes. Finally, a full alignment is generated by searching each Pfam-A profile HMM against a large sequence collection that is based on Uniprot knowledgebase (UnioprotKB) (Sonnhammer et al., 1998; The UniProt Consortium, 2011; Finn et al., 2014). In Pfam release 27.0, there are 14,831 manually curated Pfam-A protein families.

Domains of unknown function (DUFs) are a large group of protein families in Pfam. More than 20% of protein domains are annotated as DUFs (Goodacre et al., 2013). Many DUFs are highly conserved throughout the different kingdoms of life, which suggests that they have important biological functions. DUFs are named incrementally, by adding a number after the DUF prefix, according to the order in which they are added to the Pfam database. If any member of a DUF family is characterized and assigned a function, the corresponding protein domain is removed from the list of DUFs. Figure 1-1 shows the distributions of DUFs in four super kingdoms of life. There are ~3,600 Pfam DUFs. More than 1,500 DUFs are found in eukaryotes and more than 300 DUFs are found in all domains of life except viruses (Goodacre et al., 2013). Goodacre et al. (2013) identified that 355 essential proteins in 16 model bacterial species contain 238 DUFs, most of which represent single-domain proteins that suggest biological essentiality of

DUFs. For example, DUF283 (PF03368) was found to be part of the well-characterized Dicer endonuclease, and subsequent research showed that it has similarity to double-stranded RNA-binding domains (Dlakic, 2006; Qin et al. 2010). DUF1 and DUF2 are widely distributed in bacterial signaling proteins and it has been shown that they act as enzymes to process the ubiquitous signaling molecule c-di-GMP. Therefore, DUF1 and DUF2 were renamed as GGDEF and EAL, respectively (Romling and Simm 2009). IRX15 and IRX15-L contain DUF579 domain that has been shown to be involved in xylan synthesis in Arabidopsis (Jensen et al., 2011). Wang et al. (2014) suggested that DUF1618 proteins have important roles in the development and fitness of rice. DUF26 is a plant specific protein family and 40 members of this family are present in Arabidopsis. A subgroup of plant receptor-like/Pelle kinases (RLK), which is also known as cysteine-rich receptor-like kinases (CRKs), is comprised with one or more repeats of DUF26 domains consisting of a C-X8-C-X2-C motif (Wrzaczek et al., 2010). CRKs are suggested to be involved in oxidative stress, pathogen defense (Czernic et al., 1999; Wrzaczek et al., 2010) and programmed cell death (Chen et al., 2004; Acharya et al., 2007).

Proteomics has revealed the enrichment of some DUF-containing proteins family in specific cellular compartments. In plant cell wall proteome analysis, a number of DUF-containing proteins have been identified. These include: DUF642, DUF248, DUF26, DUF538, DUF246, DUF1005 and DUF1184 (Bayer et al., 2006; Minic et al., 2007; Irshad et al., 2008). Another study with Golgi proteome characterization of Arabidopsis by Parsons et al. (2012) found 13 different domains of unknown function proteins in the Golgi compartment. These included: DUF1068, DUF579, DUF707, DUF1259, DUF23, DUF616, DUF502, DUF1640, DUF1195, DUF1682, DUF228, DUF604, and DUF106.

1.1.2. Domain of unknown function 642 (DUF642)

DUF642 is a protein family in the Pfam database (PF04862). In Arabidopsis, there are ten genes that encode DUF642 domains. DUF642 gene family is well conserved in different plant species and present in gymnosperms, monocots and dicots and absent in non-seed known plant genomes (Vázquez-Lobo, et al.,2012). Albert et al. (2005) performed phylogenetic analysis with DUF642 homologs using ESTs from 18 plant species from the TIGR plant gene indices (Quackenbush et al., 2001). Their analysis defined three clades of DUF642 genes designated as A, B and C.

Clade C contained angiosperm and gymnosperms genes. Therefore, these genes were shared by a common ancestor of angiosperms and gymnosperms. Asterid, rosid and monocot genes were found in all three clades. Magnoliid genes were included in clades A and B whereas the basal-most angiosperms (Amborella and the Nymphaeales) were found in both clades B and C. Albert et al. (2005) pointed out from their phylogenetic analyses that At5g11420 was duplicated recently. In 2012, Vázquez-Lobo et al. conducted phylogenetic analyses of DUF642 domains using nucleotide sequences and with the Maximum Likelihood (ML) and Bayesian Inference (BI) methods. They analyzed 154 sequences from 24 plant species. From their analysis, reported that DUF642 protein family is highly conserved in spermatophyte species and absent in non-seed plant genomes such as algae, mosses and probably ferns. They reported that DUF642 protein sequences shared conserved motifs that defined the family. They found eighty amino acid residues were conserved in 90% of the sequences. Gao et al. (2012) conducted GUS expression analysis using two DUF642 genes (At1g80240 and At5g25460, designated as DGR1 and DGR2, respectively) and found that these genes are responsive to L-Galactono-1,4-Lactone and both are expressed in complementary patterns to each other in tissues of the seedling. In the microarray study of Arabidopsis seedlings conducted by Nemhauser et al. (2004), it was found that transcript abundance of At5g25460 and At4g32460 increased following auxin treatment. De Pauw et al. (2007) in their microarray analysis with bast fibre producing tissues of *Cannabis sativa* identified transcripts of DUF642-containing genes enriched in the top of hemp stems which was above the snap-point, a region of first accumulation of metaxylem, cell elongation to cell wall thickening (Gorshkova et al. 2003) as compared to the bottom of stem where phloem fibre secondary walls were evident (DePauw et al., 2007).

In plant cell wall proteome studies, members of the DUF642 protein family were identified in cell wall extracts, and among N- glycosylated fractions of Arabidopsis (Bayer et al., 2006; Minic et al., 2007; Irshad et al., 2008). Three (At1g29980, At2g34510 and At5g14150) of the ten Arabidopsis DUF642 proteins were predicted to be glycosylphosphatidylinositol (GPI)-anchored proteins (Borner et al., 2003 and Dunkley et al., 2006). DUF642-containing proteins were reported to interact with cell wall polysaccharides and pectinmethylesterase in vitro ((Vázquez-Lobo, et al., 2012 and Sánchez and Buen, 2012). Sánchez and Buen (2012) identified interacting proteins of DUF642 using purified recombinant 5xHis-tagged At5g11420 and At4g32460. In

their study, they prepared a DUF642 affinity column using recombinant 5xHis-tagged At5g11420 and At4g32460 which was incubated with crude protein extracts from flowers and leaves from Arabidopsis. The results revealed that FLOR1 and AtPME3 interact *in vitro* with At5g11420 and At4g32460. However, the specific biochemical and physiological function of DUF642 is still unknown.

1.1.3. Plant cell wall

Plant cells are surrounded by a wall that is mainly composed of complex polysaccharides and glycoproteins, including enzymes and structural proteins (Rose and Lee, 2010). The cell wall provides mechanical support to plants and protects them from diverse environmental conditions. Moreover, plant cell wall polymers are precursors for biofuels and are important for natural ecosystems. Plants can have two types of cell wall, primary cell wall and secondary cell wall, which differ based on their compositions and functions. Primary plant cell walls are mainly composed of polysaccharides cellulose, hemicellulose, and pectin. On the other hand, secondary cell wall components are mainly composed of cellulose, hemicelluloses and lignin:

Cell wall proteins (CWP) are involved in many important functions such as cell wall modification, cell structure, signaling and interactions among the plasma membrane proteins**.** Most of the cell wall proteins are reported to be encoded by multigene families. Jamet et al. (2006) reviewed all available Arabidopsis cell wall proteome data to identify 281 proteins in their CWP database. About 90% of CWP are placed in categories on the basis of predicted biochemical or biological functions. However, the biochemical functions of only a few identified proteins have been experimentally demonstrated. Interestingly, 10% of the proteins have unknown function. The challenge is to elucidate their biological role within the cell wall. Jamet et al. (2006) also noted that 11 common proteins are found in 3 different organs: etiolated hypocotyls, cell cultures and rosettes. Among these are 11 proteins, two glycoside hydrolases, two PMEs, one germin, four proteins with interacting domains and one protein of unknown function. Therefore, it can be predicted that unknown proteins are major components of plant cell walls. Among these unknown proteins are many unique DUFs (Jamet et al. 2006).

1.1.4. Plant cell wall polysaccharides

Plant cell walls are made of carbohydrate polymers, primarily with load bearing cellulose-hemi-cellulose network embedded in pectin. These constituents may be degraded and modified by endogenous enzymes during plant growth and development.

1.1.4.1. Cellulose

Cellulose consists of a collection of β-1,4-linked glucan chains. Cellulose molecules may associate with each other to form crystalline cellulose by hydrogen bonding and Van der Waals forces (Nishiyama,et al., 2002). The content of cellulose in Arabidopsis cell wall ranges from 15% of leaf and 6 to 33% of stem (Zablackis et al., 1995 and Zhong et al., 2005). Cellulose is synthesized at the plasma membrane by cellulose synthase complexes (CSCs).

1.1.4.2. Hemicellulose

Hemicellulose polysaccharides constitute a substantial fraction of a mature cell wall across plant species (Carpita, 1996). Hemicelluloses contain different sugar monomers and a β-linked sugar backbone. In case of xylans, mannans, and xyloglucans, the backbone sugars are β-1,4-D-Xyl, β-1,4-D-Man, and β-1,4-D-Glc, respectively (Gilbert 2010).

1.1.4.3. Xyloglucan

Xyloglucan (XyG) is the most dominant hemicellulosic component of the primary cell wall and comprises about 20–30 % of primary cell wall of dicots (Fry and Janice. 1989; Scheller and Ulvskov 2010). One or more members of the cellulose synthase-like family C (CSLC) genes are involved in the biosynthesis of glucan backbone of XyGs (Cocuron et al. 2007; Pauly and Keegstra, 2008). CSLC genes are glycosyltransferases (GT). Transcriptional analysis of nasturtium (*Tropaeolum majus*) seeds, which produce large amounts of galactoXyG as a storage polymer, showed overexpression of *CSLC4* gene (Cocuron et al., 2007). It was reported that XyG:XylT, but not its transferase activity, is necessary for glucan synthesis. A study conducted by Cavalier et al. (2008) identified two *Arabidopsis thaliana* genes that encode xylosyltransferases, *XXT1* and *XXT2*, which are involved in xyloglucan biosynthesis *in vivo*.

Double *xxt1 xxt2* knockout mutants showed abnormal root hairs and slow growth compared to wild-type plants and lacked detectable xyloglucan in their cell walls.

1.1.4.4. Mannan

Mannas are important constituents of hemicelluloses in the secondary cell wall of gymnosperms (Pauly and Keegstra, 2008). Mannans are widely distributed in plants and comprise four types: linear mannan, glucomannan, galactomannan, and galactoglucomanan (Petkowicz et al., 2001). Heteromannans are synthesized from activated nucleotide sugars which include GDP-mannose, GDP-glucose, and UDP-galactose (Liepman et al. 2005). Dhugga et al. (2004) first identified β-mannan synthase (ManS), a member of the cellulose synthase-like family A (CSLA) in guar seeds. Recently, Wang et al. (2012) identified a ManS enzyme from the CSLA family that is involved in galactomannan biosynthesis in fenugreek endosperm. ManS enzyme preferentially use GDP-mannose as the substrate for the backbone synthesis. Other studies (Liepman et al. 2005, 2007; Suzuki et al. 2006; Goubet et al. 2009) reported that the CSLA family is involved in the synthesis of the glucomannan backbone.

1.1.4.5. Xylan

Xylan is the main hemicellulose in the secondary cell walls of eudicots and in the primary and secondary cell walls of grasses and cereals (Jensen et al., 2011). Xylan is synthesized by enzymes in the Golgi apparatus. Several enzymes have been reported in xylan biosynthesis in plants. For example, *Irregular Xylem (IRX) 9* and *IRX14* of Glycosyltransferase Family 43 (GT43) and *IRX10* of GT47 required for synthesizing the xylan backbone. Studies conducted by Lee et al. (2012) and Brown et al. (2009) revealed that mutations in these three genes cause dwarfing and a reduction in xylan content and xylosyltransferase activity. Jensen et al. (2011) identified two Arabidopsis genes encode proteins containing a Domain of Unknown Function (DUF) 579 and were designated IRREGULAR XYLEM (IRX)15 and IRX15-LIKE (IRX15-L). These genes are likely involved in xylan biosynthesis. They generated double mutant using Arabidopsis T-DNA knockout lines for the two genes and found that *irx15 irx15*-L double mutants displayed a moderate reduction in stem xylose. More recent study (Yuan et al., 2013) found the Arabidopsis DUF231 domain-containing ESK1 protein is a putative acetyltransferase

required for O-acetylation of xylan. Glycosyl hydrolases, xylanase and xylosidases are envisaged to be involved in xylan remodeling in Arabidopsis (Rennie and Scheller, 2014).

1.1.4.6. Pectin

The structural classes of the pectic polysaccharides include homogalacturonan (HG), xylogalacturonan (XGA), rhamnogalacturonan II (RG-II), and rhamnogalacturonan I (RG-I) (Fig 1-4). Among these pectin polysaccharides, rhamnogalacturonan II is the most structurally complex. HG is the most abundant polysaccharide that comprises 65% of pectin. RGI constitutes 20% to 35% (Mohnen, 2008). XGA and RGII, each comprises less than 10% (Zandleven et al., 2007; Mohnen, 2008). The glycosyltransferase (GT) enzymes use nucleotide-sugar substrates and acceptors for pectin biosynthesis. It is predicted that pectic polysaccharides are cross-linked to hemicelluloses, phenolic compounds, and to wall proteins.

1.1.4.7. Homogalacturonan

Homogalacturonan (HG) is composed of (1,4)-alpha-D-galactosyluronic acid residues and can be methyl esterified and/or acetylated (ONeill et al., 1990). Effective solubilization of an HG synthase enzyme that catalyzes the transfer of galactouronic acid to homogalacturonan and the development of an acceptor-dependent assay for α-1,4-GalA transferase was conducted by Doong and Mohnen in 1998. Later, Sterling et al. (2006) first successfully identified a pectin biosynthetic enzyme, the HG α-1,4-GalA transferase and designated it as GALACTURONOSYLTRANSFERASE1 (GAUT1). The GAUT group of proteins belongs to CAZy GT family 8 and has 15 members in Arabidopsis; it is possible that all GAUT proteins are GalA transferases (Harholt et al., 2010). GAUT7 and GAUT7 are together in a complex involved in the biosynthesis of HG (Mohnen, 2008).

1.1.4.8. Rhamnogalacturonan I

Rhamnogalacturonan I has a backbone of alternating rhamnose and galacturonic acid residues with the side chains of α-1,5-arabinans, β-1,4-galactans, and arabinogalactans. Many enzymes are required to synthesize this pectin, but few have been identified. Only two GTs have been identified as involved in RGI biosynthesis, namely ARABINAN DEFICIENT1 (ARAD1;

At2g35100) and XYLOGALACTURONAN DEFICIENT1 (XGD1; At5g33290). ARAD1 and XGD1 are placed in subgroups B and C of CAZy family GT47, respectively (Harholt et al., 2010).

1.1.4.9. Rhamnogalacturonan II

RGII is composed of 1,4-linked a-D-Gal*p*A residues and has not been studied extensively. Only a few genes are known to be involved in RGII biosynthesis, including the RGXT family in subgroup B of CAZy GT77. Members of this family with characterized functions include *RGXT1*, *RGXT2*, and *RGXT3* (Egelund et al., 2006, 2008). These proteins can transfer Xyl from UDP-Xyl onto Fucose and have the α-1,3-xylosyltransferase activity. This linkage is only present in RGII, therefore, indicating their involvement in RGII biosynthesis. Many genes must be involved in the synthesis of complex structure of RGII but only few have been identified.

1.1.5. Matrix cell wall polysaccharides biosynthesis in the Golgi

Golgi is the main organelle in which biosynthesis of cell wall polysaccharides and glycosylation of glycoproteins and glycolipids occurs. A Golgi stack has four defined regions: the cis, medial, trans Golgi, and the trans-Golgi network. Cell wall polysaccharides are continually synthesized and transported from Golgi stacks and subsequently progress to the trans-Golgi network where they are sorted and packaged into the vesicles and sent to the destination (Caffall and Mohnen, 2009).

After translation, proteins enter the secretory pathway and glycosylate during translocation into the lumen of the rough endoplasmic reticulum (RER). Glycosylation generally occurs only at asparagine residues in the sequence Asn-X-Ser/Thr, where X is any amino acid except proline (Marshall, 1974; Kaplan et al., 1987). Asparagine-linked glycan (N-glycan) undergoes sequential maturation during transit via the cis, medial, and trans cisternae involving the sequential removal and addition of sugar residues. A complex series of posttranslational enzymatic steps is involved in the formation of glycoproteins with diverse biological functions. In the glycosylation pathway in plants, biosynthesis of N-linked glycans of glycoproteins involves enzymes in each sub-Golgi compartment (Oikawa, et al., 2013). Gendre et al. (2013) reported that the secretion of cell wall polysaccharides for cell elongation occur through the *trans*-Golgi network (TGN). It was found

that ECHIDNA/Ypt localized in the trans-Golgi network and involved in the secretation of cell wall polysaccharides. (Gendre et al., 2013).

The noncellulosic polymers hemicellulose and pectin are synthesized by glycosyltransferases (GTs) that are located in the different compartments of the Golgi apparatus. GTs are type II membrane-bound proteins with a catalytic domain facing the lumen of the Golgi (Ridley et al., 2001; Sterling et al., 2001; Geshi et al., 2004). A non-esterified form of HG is synthesized in *cis* and medial Golgi cisternae, and the methylesterification occurs in both medial and *trans* compartments, and the side chains of RGI are added in *trans* cisternae (Driouich et al., 2012). Atmodjo et al. (2011) demonstrated that the association of HG biosynthesis enzymes AtGAUT1 with AtGAUT7 is required to facilitate activity of galacturonosyltransferase in Golgi membrane. GAUT7 protein is anchored in the Golgi membrane. The subcellular localization of pectin biosynthetic HG:GalAT activity was detected exclusively in Golgi-enriched fractions (Sterling et al., 2001). *In vivo*, subcellular localization of the protein containing a domain of unknown function (DUF) 579, designated as IRREGULAR XYLEM (IRX)15 revealed that IRX15 is localized in Golgi and other cytosolic bodies, possibly the trans-Golgi network (Jensen et. al., 2011). RGXT1 and RGXT2 belong to glycosyltransferases (CAZy GT-family-77) and encode cell wall (1,3)-alpha-d-xylosyltransferases are localized in Golgi and involved in the biosynthesis of pectic rhamnogalacturonan-II (Egelund J, 2006). Arabidopsis gene *ESKIMO1* (*ESK1*) is involved in xylan O-acetylation during secondary wall biosynthesis and is localized in Golgi (Yuan et al., 2013).

1.1.6. Plant growth and development

Plant growth occurs through coordinated expansion and elongation of individual cells. In cell expansion, an increase in cell size occurs in two or three directions whereas in cell elongation, expansion occurs in one direction. The increase in cell size can be 100 000 or 1 000 000 times greater than the original size of the cell (McCann et al., 2001). This enlargement of cell is induced by expansion of the vacuole and increases in cell ploidy level from endoreplication (Perrot-Rechenmann, 2010). Cell walls are born in dividing cells in the cell plates that form during cytokinesis and create a wall between daughter cells. During cell expansion and elongation, remodeling of cell wall polysaccharides occur simultaneously. Three classes of

polymers provide the strength to elongating cells: (1) the microfibrils arranged in the transverse axis, (2) the crosslinking glycans in the longitudinal axis, and (3) networks involving structural proteins or phenylpropanoid compounds, or elements of the pectin network (McCann et al., 2001). During cell growth, new cellulose microfibrils and other wall polymers must be synthesized and deposited. Golgi is the site where cell wall polysaccharides are synthesized and secreted to the elongating cell wall.

Plant growth and development are regulated by a diverse group of growth regulators, collectively called plant hormones or phytohormones. These include auxins, cytokinins, gibberellins (GA), abscisic acid (ABA), ethylene, the brassinosteroids (BRs), jasmonic acid (JA), salicylic acid (SA), polyamines, strigolactones (SL), nitric oxide (NO) and peptide hormones (Dharmasiri et al., 2013; Santner et al., 2009). Among all the plant hormones, auxins are major regulators and play important roles during the entire life span of a plant. Auxins are involved in cell division, cell elongation and cell differentiation. Indole-3-acetic acid (IAA) is the major form of auxin in higher plants. The studies with auxin responsive genes reveled that auxin responses are regulated by two large protein families named the ARF and Aux/IAA proteins (Berleth et al., 2004; Parry and Estelle, 2006). Crosstalk between the phytohormones ethylene and auxin are most widely studied and revealed their interactions are crucial for plant development (Stepanova and Alonso, 2005 and Muday et al., 2012).

1.1.7. Approaches to characterize the function of an unknown gene

Recent advances in plant genomics technologies such as genome sequencing and cDNA sequencing (RNA-seq) have contributed many unknown genes in the databases. The challenge comes in terms of characterizing the unknown genes. The classical approach of a gene annotation includes generating knockout mutant and overexpression of the target gene. In addition, many other high throughput experimental approaches can be applied to assign function to an unknown gene or a protein.

1.1.7.1. Homology search approach

In silico approaches are widely used to infer gene function. The most common approach to functional annotation is based on sequence homology (Lee et al., 2007). However, current

bioinformatics tools are not able to predict the function of roughly a one third of the genes in a plant genome. Moreover, predictions of at least 60% of some gene families have been shown to be wrong (Roberts, 2004). Goonesekere et al. (2010) annotated eight DUF genes based on a combination of sequence analysis and modeling.

1.1.7.1. Reporter genes approach

Application of reporter gene assays is useful in identifying organ-specific gene expression patterns and the subcellular localization of a target protein. Generally, a reporter gene is fused to a cis-regulatory DNA or a coding sequence. Among the available reporter genes, the most widely used in plants are β-glucuronidase (GUS) from *Escherichia coli* and GFP from the jellyfish *Aequorea victoria* and its derivatives. The choice of reporter gene depends on the type of research question. GUS is most widely used in qualitative and quantitative promoter expression analysis. β-glucuronidase is very sensitive and able detect even a very week promoter. The disadvantages of GUS assay are, it is a destructive method and can diffuse to surrounding the tissue. On the other hand, GPF is a non-destructive method and mostly does not diffuse from its place of expression

The information of spatial and temporal expression patterns is an important tool for functional annotation of an unknown gene. Many genes are active in different developmental stages and organs. Moreover, many genes are expressed conditionally. Gao et al. (2012) identified two DUF (domains of unknown function) 642 family genes (At1g80240 and At5g25460, designated as *DGR1* and *DGR2*, respectively) that are expressed in a complementary manner in the root tips. Moreover, both genes are responsive to L-Galactono-1,4-lactone (L-GalL).

Protein subcellular localization is critically important for protein function annotation. Plant cells are surrounded by wall and inside the wall contain the cytosol and the cytoplasmic organelles. After translation, proteins are translocated to target organelle to perform function. Subcellular localization of many genes is still unknown. Prediction of subcellular localizations of a target protein is helpful but not very reliable. Therefore, *in vivo* study of subcellular localization is important to identify the compartment, thereby, elucidate the function of the protein. The

DUF579 domain containing protein *IRX15*, which is involved in xylan biosynthesis, is localized in Golgi (Jensen et. al., 2011).

1.1.8. Heterologous protein expression

In vitro assays with purified recombinant protein often provide the most accurate proof about the activity for a specific enzyme (Fernandez and Vega, 2013). This requires prior knowledge of the type of reaction catalyzed and the substrate specificity of the candidate enzyme. Therefore, it is difficult to apply this technique to the candidate genes with unknown function (Prosser et al., 2014). Theoretically, production of recombinant protein is very straightforward, but in practice, difficulties arise. A wide range of expression hosts are available which include bacteria, archaea, filamentous fungi, yeasts, and protozoa. All have advantages and disadvantages and the choice of host mainly depends on the protein of interest (Demain and Vaishnav, 2009). For example, for the proteins that undergo post-transcriptional modification, a prokaryotic expression system may not be suitable (Sahdev et al., 2008). Nevertheless, many eukaryotic proteins are successfully expressed in *Escherichia coli*. *E. coli* is the most frequently used host for production of enzymes and other proteins by recombinant DNA technology. Its simplicity and fast growing rate make it an ideal candidate. However, *E. coli* presents some disadvantages for expression of eukaryotic proteins: it lacks most post-translational modifications, is less able to form disulfide bonds, and has different codon bias. An expression vector is required that must have elements necessary for protein expression. The pET vectors with the regulation of the T7 promoter are extremely popular for recombinant protein expression. Protein purification is another challenge in recombinant protein expressions. Apart from the difficulties, heterologous protein expression is considered as a powerful tool for functional and biochemical analyses of genes and gene families isolated from various organisms (Yesilirmak and Sayers, 2009).

1.1.9. Metabolic profiling approach

Metabolomics aims to provide a comprehensive non-biased, high throughput assay of complex metabolite mixtures present in a biological sample. Metabolites are the end products of cellular regulatory processes, the levels of which are the result of ultimate responses of biological systems to genetic and environmental changes (Fiehn, 2002). Metabolomics has a wide range of

applications that includes microbial biotechnology, food technology, pharmacology, toxicology, enzyme discovery, systems biology, and plant biotechnology (Gomez-Casati et al., 2013)

Even though metabolomics is considered the newest of the "omics" sciences, metabolic profiling is not new. GC-MS based approaches of metabolic profiling were developed in the 60s and 70s. Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography was published by Linus Pauling et al., in 1971. After 20 years, the first milestone in plant metabolomics was established by Sauter et al. in 1991. GS/MS based metabolic profiling methodology for the determination of metabolites in Arabidopsis was developed by Fiehn et al. in 2000. Major instrumental approaches for metabolic profiling include gas chromatography–mass spectrometry (GC-MS), liquid chromatography–mass spectrometry (LC-MS), nuclearmagnetic resonance (NMR) and capillary electrophoresis–mass spectrometry (CE-MS). But none of the methods can detect all total metabolites of a biological system. Compared to other instruments, GC-MS is considered most efficient, sensitive, and reliable tools for metabolomics (Fiehn et al., 2002). The steps include in metabolic profiling by GC-MS are comprised of sample collection, metabolites extraction, compound derivatization, instrument analysis, data analysis, and metabolite annotation and pathway analysis (Fig. 1-8) (Qiu and Reed, 2014).

Enzymes are involved in metabolic pathways and do not function in isolation. There is a complex network that results from various interacting proteins and metabolites. Therefore, *in vitro* analysis will often not provide a complete picture of the physiological function of a candidate gene (Prosser et al., 2014). *Ex vivo* metabolic profiling is a key tool for pathway discovery and novel metabolite identification in a biological system (Prosser et al., 2014; Zhang et al., 2011). Metabolites are the byproducts of substrate-enzyme reactions and the final products of gene expression and activity. So, the metabolic profiling can provide more appropriate insight about the regulatory points in the metabolic pathways of the candidate gene. There are two types of metabolites: primary which are essential for growth, development and reproduction; and secondary metabolites that are not directly involved in these processes. The metabolites in plants are much lower than the number of genes. In the model plant *Arabidopsis thaliana*, there are roughly 26,500 genes that can be estimated to produce 8,000 primary

metabolites (including intermediates) and 1,750–3,500 secondary metabolites (including intermediates). (Pichersky and Lewinsohn, 2011).

Through metabolomics, it is possible to identify pathways that are involved in the production of important metabolites for health and food. For example, genetically modified golden rice accumulates beta carotene in the endosperm (Ye et al., 2000) that was identified by metabolites analysis. Recently, an unusual isoleucine biosynthesis pathway in *Geobacter metallireducens* was discovered using combined approach of stable-isotope probing and metabolomics (Posser et al., 2014). Metabolic profiling revealed that infection of soybean by fungal pathogens results in the mobilization of carbohydrates, disturbance of the amino acid pool, and activation of isoflavonoid, a-linolenate, and phenylpropanoid biosynthetic pathways of the plant (Aliferis et al., 2014). In another study, metabolic profiling revealed overexpression of UDP-glucose pyrophosphorylase (UGPase). UGPase2 is involved in primary and secondary metabolic pathways that result in the reduction of sugar and starch levels (Payyavula et al., 2014). Metabolic profiling is also used as a tool for phenotyping genetically and environmentally modified plants. Na Jom et al. (2014) in their study with black gram identified distinct differences in metabolite profiles among three black gram varieties. They also found that climate effect changes the metabolite profiles of the black gram. Fukushima et al. (2014) conducted metabolic profiling of 50 Arabidopsis mutants and generated a database. Their analysis of data from *mur9*-and *eto1-1* mutant compared to wild type showed that both mutants exhibit a significant increase in the succinate level in the tricarboxylic acid cycle.

1.2. Conclusions and objectives

DUF642 is a domain of unknown function. The biochemical and physiological functions of DUF642-containing proteins are unknown. DUF642 family genes are well conserved in a wide range of plant species which indicates their specific functions in the plant biological processes. Plant cell wall proteomes identify DUF642-ciontaining genes in cell wall extracts. Moreover some of the members of this protein family interact with pectinmethylesterase *in vitro*.

I hypothesized that DUF642 functions in cell walls and is involved in non-cellulosic polysaccharides biosynthesis. My objective was to identify the functions of two DUF642 genes named DGR2 (At5g25460) and DUFB (At5g11420) using the model plant *Arabidopsis thaliana*.

The following approaches were carried out for the functional characterization of two DUF642 genes in Arabidopsis: (i) bioinformatics analysis (ii) mutant analysis, (iii) gene expression analysis, (iv) protein expression analysis, (iii) heterologous protein expression in E. coli and (v) metabolic profiling. My research questions were: (i) what are the expression patterns of these genes and the proteins they encode? (ii). Can I express these proteins heterologously in *E. coli* and use the purified protein to know the biochemical function of the proteins? (iii). What are the interacting proteins? (iv) In which organelle this protein is localized and (v) in which pathways these proteins function? (v) Do DGR2 and DUFB function in the same manner or complementary manners? (vi) Are DGR2 and DUFB regulated by any growth hormones?

2. **Chapter 2**

2.1. *In Silico* Studies of DUF642 Proteins in Arabidopsis

2.2. Introduction

Domains of Unknown Function (DUF) are protein families in the Pfam database that have no characterized function. DUF642 is one of the protein families in this database. DUF642 is defined by a large domain that has an average length of ~323aa, and comprises the bulk of the proteins that contain it. This domain is highly conserved in spermatophytes and is present in some predicted cell wall proteins (Vázquez-Loboa, et al., 2012). In Arabidopsis, there are 10 genes that contain DUF642 and seven of these genes also contain a galactose binding domain-like fold (Fig. 2-1).

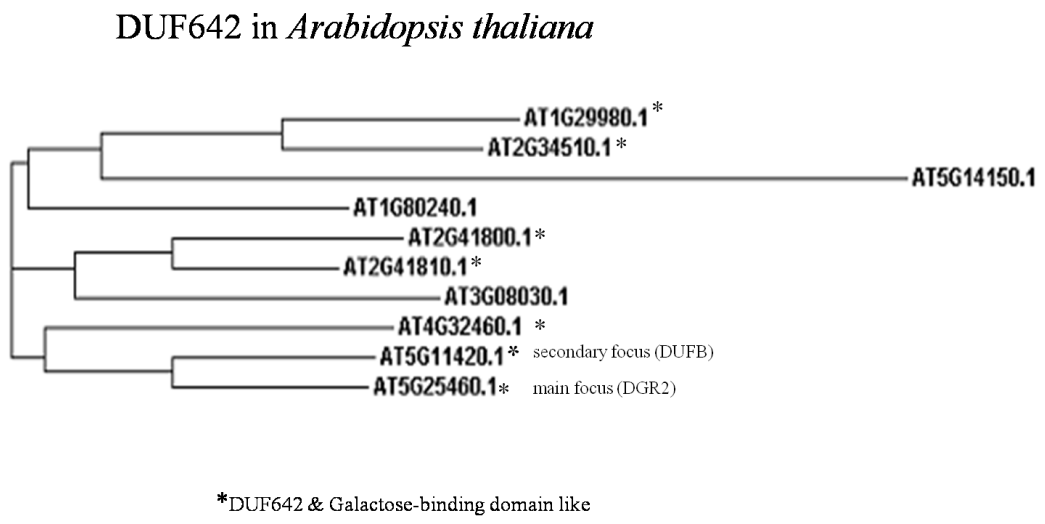


Figure 2-1. Phylogram of DUF642 protein family in *Arabidopsis thaliana* constructed with TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html) after alignment of amino acid sequences with ClustalW (http://www.ebi.ac.uk/clustalw).

The objective of the present *in silico* analysis is to make inferences about the functions of two DUF642 genes, At5g25460 (*DUF642 L-GALL RESPONSIVE* 2, *DGR2*) and At5g11420

(*DUF642 CONTAINING B, DUFB*). These genes were selected for analysis because they have high amino acid sequence similarity (93.4%) to each other. Moreover, a comparison of five cell wall proteomic studies (Bayer et al., 2006; Boudart et al. 2005; Charmont et al., 2005; Borderies et al., 2003; Chivasa et al., 2002) showed that only four out of ten DUF642 genes could be detected among the proteomes sampled; the detected proteins included At5G25460 (DGR2) and AT5G11420 (DUFB). Therefore, it was proposed that DGR2 and DUFB are expressed proteins and characterization of these two proteins may reveal more information about cell wall metabolism. The following computational analyses will be applied: homology-based 3D modeling, 3D structure validation, phylogenetic inference, promoter analysis, domain prediction, gene co-expression and network analysis.

2.3. Materials and Methods

2.3.1. Promoter analysis

An 1841 bp region from immediately upstream of the start codon of *DGR2* and a 1454 bp region from upstream of *DUFB* were searched for the presence of conserved cis-regulatory elements using the PLANT CARE database and web interface (http:// bioinformatics.psb.ugent.be/webtools plantcare html). These regions were selected based on the findings that the median intergenic region in Arabidopsis is 1.4 kb and the average length of most functional promoters in Arabidopsis could be established at 500 bp (Korkuc et al., 2014; Zhan et al., 2006). We picked a ~2 kb region upstream from the ATG to represent the *DGR2* promoter because the intergenic region was more than 2kb (~4.5 kb). For *DUFB*, the full intergenic region was used as a promoter because this region was about 1.4 kb. The same fragments were used in reporter gene assays *in vivo* (Chapter 3).

2.3.2. Protein structure prediction

The secondary structures of DGR2 and DUFB were predicted using the default options of PSIPRED (Jones 1999; Bryson et al. 2005), and the I-TASSER Standalone Package (Yang et al., 2015; Roy et al., 2010) for structural and functional prediction (see below).

2.3.3. Structure based function prediction

The functions of 3D modeled structure of DGR2 and DUFB were predicted by I-TASSER which was based on global and local similarity to template proteins in PDB with known structure and functions. The resultant global and local similarity scores were used to rank the template proteins (functional homologues) and transfer the annotation (EC numbers and gene ontology terms) based on the top scoring hit (Yadav et al., 2013). I-TASSER server predicted gene ontology (GO) terms based on functional homology score (Fh-score) for the query proteins. Each modeled protein was associated with multiple GO terms, but only those GO terms which lie in highest Fh-score category were considered. In general, GO terms elucidate the putative function of modelled structures. A consensus prediction was predicted by a consensus between a GO term and its ancestor terms having Fh-score greater than 1.0 by 1-Tasser server. Functional annotations were conducted by Enzyme Commission numbers and Gene Ontology terms (Roy et al., 2010) for DGR2 and DUFB. One of the predicted models with high C-score value and high decoy value obtained from I-TASSER were refined to obtain near to their native structure by using the high-resolution protein structure refinement, ModRefiner (Xu and Zhang, 2011).

2.3.4. Evaluation of Predicted Structures

The refined models for DGR2 and DUFB structure were evaluated using the Rampage server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.ph). The Ramachandran plot calculated phi/psi angles (ϕ,ψ) between N-C_α and C_α-C atoms of residues, and thus generated Ramachandran Plots. phi/psi plots obtained from the amino acid residues were subdivided into "favored", "allowed" and "outlier" regions.

2.3.5. Phylogenetic tree

DUF642 protein sequences from nine species were downloaded from Phytozome V9.1. These species were selected because they represented the flowering plants (angiosperms). Sequences were aligned using ClustalW (Jeanmougin et al., 1998) and a neighbor-joining tree was constructed using MEGA4 (Tamura et al., 2007) with default settings. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Tamura et al., 2007) and were in the units of the number of amino acid substitutions per site. Bootstrap values (%) on the branches were calculated as the number of times that a particular grouping of sequences appears during the bootstrap analysis.

2.3.6. Determination of signal peptide, domain assignment and topology predictions

Sequences were analyzed with SignalP (SignalP–http://www.cbs.dtu.dk/services/SignalP/; Bendtsen et al., 2004) software for predicting signal peptides and Topology predictions were performed using SOSUI transmembrane protein prediction server (http://bp.nuap.nagoya-u.ac.jp/sosui/. Functional domains were searched by InterProScan (http://www.ebi.ac.uk/InterproScan).

2.3.7. Tissue specific expression prediction

The expression profiles of *DGR2* and *DUFB* were examined *in silico* using Genevestigator (Zimmermann et al., 2004) and the Arabidopsis eFP Browser (Winter et al., 2007).

2.4. Results and Discussion

DOMAIN OF UNKNOWN FUNCTION 642 (DUF642) is large, (~323aa) conserved protein motif defined in the Pfam database (PF04862). In Arabidopsis there are 10 genes containing DUF642 domains. In this study, *in silico* analysis of two DUF642 genes in Arabidopsis, *DGR2* (At5g25460) and *DUFB* (At5g11420) was conducted. DGR2 and DUFB had 93.4% amino acid sequence similarity and 86.2% identity (Fig. 2-2). The amino acid sequence length for DGR2

and DUFB was 369 and 366, respectively (Table 2-3). Molecular weight and isoelectric point for DGR2 were 39.97 and 7.4 and these values for DUFB were 39.64 and 7.8 (Table 2-3). The proteins were predicted to be soluble and secreted proteins as predicted by SOSUI and TargetP, respectively.

```
Length: 362
Identity:     312/362 (86.2%)
Similarity:   338/362 (93.4%)
Gaps:           0/362 ( 0.0%)
Score: 1672

  5 SLSFLFVLLIATITSVICFSDGMLPNGDFELGPKPSDMKGTQVINKKAIP     54
    |...||:.......|.:.|.|||||||||||||||||||:::||.|||
  8 SFFLLFIATAMAAKSTVSFRDGMLPNGDFELGPKPSDMKGTEILNKLAIP     57

 55 SWELSGFVEYIKSGQKQGDMLLVVPAGKFAIRLGNEASIKQRLNVTKGMY    104
    :||::|||||||||.||||||||||||||||:||||||||||||.|.||||
 58 NWEVTGFVEYIKSGHKQGDMLLVVPAGKFAVRLGNEASIKQRLKVVKGMY    107

105 YSLTFSAARTCAQDERLNISVAPDSGVIPIQTVYSSSGWDLYAWAFQAES    154
    ||||||||||||||||||||||||||||||||||||||||||||||||||
108 YSLTFSAARTCAQDERLNISVAPDSGVIPIQTVYSSSGWDLYAWAFQAES    157

155 NVAEIVIHNPGEEEDPACGPLIDGVAIKALYPPRPTNKNILKNGGFEEGP    204
    :|||:||||||.||||||||||||||:::||||||||||||||||||||||
158 DVAEVVIHNPGVEEDPACGPLIDGVAMRSLYPPRPTNKNILKNGGFEEGP    207

205 YVLPNATTGVLVPPFIEDDHSPLPAWMVESLKAIKYVDVEHFSVPQGRRA    254
    .|||.:|||||:||||||||||||.||||||||:||||||||||||||||
208 LVLPGSTTGVLIPPFIEDDHSPLPGWMVESLKAVKYVDVEHFSVPQGRRA    257

255 VELVAGKESAIAQVARTVVGKTYVLSFAVGDANNACQGSMVVEAFAGKDT    304
    :||||||||||||||.|||:||||||||||||||||||:||||||||||||||
258 IELVAGKESAIAQVVRTVIGKTYVLSFAVGDANNACKGSMVVEAFAGKDT    307

305 LKVPYESRGKGGFKRASLRFVAVSTRTRVMFYSTFYSMRSDDFSSLCGPV    354
    |||||||:|.|||||||:||||||||:|:||||||||:||||||||||||||
308 LKVPYESKGTGGFKRASIRFVAVSTRSRIMFYSTFYAMRSDDFSSLCGPV    357

355 IDDVKLLSARKP    366
    ||||||:|.|||
358 IDDVKLISVRKP    369
```

Figure 2-2. Amino acid sequence alignment of DGR2 and DUFB proteins in Arabidopsis by ClustalW2 (Larkin et al.2007).

A search of DGR2 and DUFB using InterProScan identified two conserved domains: the DUF642 domain, and the galactose binding-like domain (Fig. 2-3). DGR2 had two and DUFB had one galactose binding-like domain, respectively.

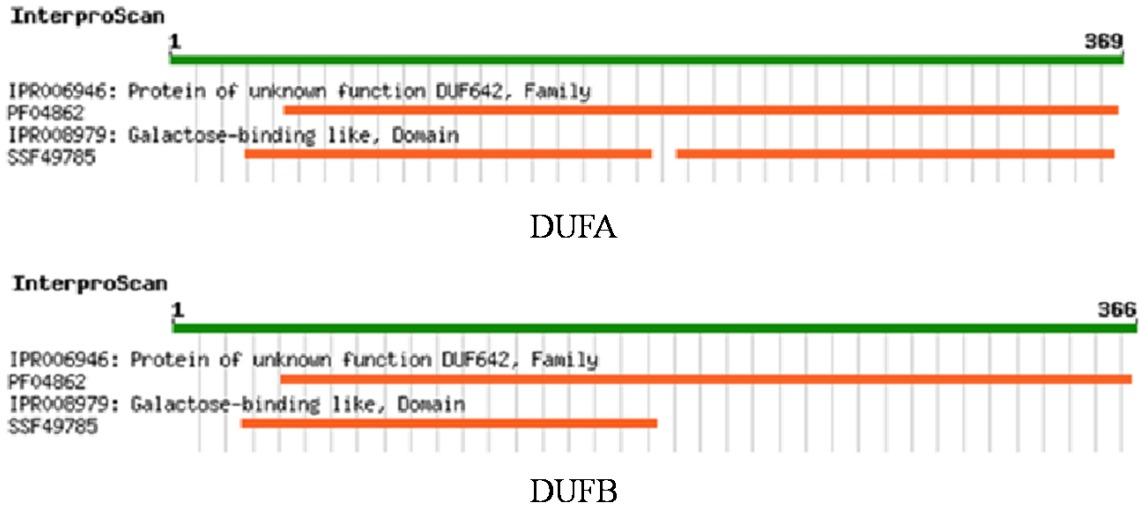


Figure 2-3. Domain assignment of DUFA and DUFB by SWISS_MODEL PHYRE.

2.4.1. Tissue specific expression prediction

A summary of transcript expression profiles of *DGR2* and *DUFB* was obtained from a survey of 98 Arabidopsis tissues, using the Genevisible (http://genevisible.com/search) server (Fig. 2-4). *DGR2* transcripts were most abundant in the petal, petiole, pedicel, replum and hypocotyl whereas *DUFB* was most abundant in petal, petiole, rosette cell, hypocotyls and shoot apical meristem. It was revealed by the eFP browser that *DGR2* and *DUFB* are expressed in the root tissues especially in stage II (Table 2-1). In stage II, cells transition from optically dense to a more transparent appearance, modulated by longitudinal expansion, which occurs about 0.30 mm from the root tip (Birnbaum et al., 2003). In the stem, both genes also more highly expressed in the apical region of the stem compared to the basal region (Table 2-1). Expression of both genes is also detected in the flowers and rosette leaves (Table 2-1). It was also revealed that expression of *DGR2* is upregulated by auxin but down regulated by GA3 and MeJA and remains unchanged by ACC treatment. On the other hand, expression of DUFB was found unchanged by GA3,

slightly decreased upon ACC and IAA treatments, and down-regulated by MeJA (Table 2-1). It was also found that the both genes are responsive to cold and heat stresses (Table 2-1).

Figure 2-4. Expression of AT5G25460 (246919_at) and AT5G11420 (250366_at) across 98 tissues tested by GENEVESTIGATOR.

Table 2-1. A summary of microarray-based expression profile of *DGR2* and *DUFB* of *A. thaliana* compiled from the eFP Browser. Values are absolute expression level in the tissues.

Gene Name	**Tissue specific expressions**	**Expression Level**
DGR2	Root Stage III Stele	239.29
	Root Stage III Cortex + Endodermis	644.17
	Root Stage II Stele	428.61
	Root Stage II Cortex + Endodermis	1153.83
	Root Stage I Stele	61.49
	Root Stage I Cortex + Endodermis	178.14
	Whole stem top	2667.6
	Whole stem bottom	18.55
	Flower	1519.4
	Rosette Leaf	597.96
	Germinated endosperm	237.36
	Expressions in response to hormones	**Expression Level**
	Control at 3 Hours, tissue: seedling	1320.14
	1µM IAA Treated at 3 Hours; tissue: seedling	1852.18
	10uM MJ Treated at 3 Hours,tissue: seedling	603.22
	1uM GA-3 Treated at 3 Hours, tissue: seedling	951.46
	10uM ACC Treated at 3 Hours, tissue: seedling	1320.43
	Expressions in response to biotic stress	**Expression Level**
	Control root after 3h	1298.88
	Cold root after 3h, 4°C, tissue: root	1521.88
	Heat root after 3h, 38°C,tissue: root	1716.88
Gene Name	**Tissue specific expressions**	**Expression Level**
DUFB	Root Stage III Stele	26.23
	Root Stage III Cortex + Endodermis	133.26
	Root Stage II Stele	70.87
	Root Stage II Cortex + Endodermis	354.63
	Root Stage I Stele	6.12
	Root Stage I Cortex + Endodermis	30.66
	Whole stem top	2004.71
	Whole stem bottom	78.94
	Flower	513.85
	Rosette Leaf	493.33
	Germinated endosperm	361.24
	Expressions in response to hormones	**Expression Level**
	Control at 3 Hours, tissue: seedling	446.81
	1µM IAA Treated at 3 Hours; tissue: seedling	373.95
	10uM MJ Treated at 3 Hours,tissue: seedling	199.08
	1uM GA-3 Treated at 3 Hours, tissue: seedling	437.87
	10uM ACC Treated at 3 Hours, tissue: seedling	311.02
	Expressions in response to biotic stress	**Expression Level**
	Control root after 3h	25.63
	Cold root after 3h	82.51

	Heat root after 3h	16.59

2.2.2. Promoter analysis

Each gene normally has a distinct combination of transcription factor binding sites (TFBSs) in its promoter region, which regulates the spatial and temporal expression of that gene (Qiu, 2003). A comparison of conserved cis-regulatory elements in upstream regions of *DGR2* and *DUFB* was performed (Table 2-2). It was found that a putative TATA box element was more frequent in the region upstream of *DGR2* than *DUFB* (Table 2-2). The number of TATA boxes could be responsible for higher levels of gene expression; however this should be tested experimentally (Zamani Babgohari et al. 2014). A number of phytohormone-responsive motifs were detected in both promoter regions. In the *DGR2* upstream region, the TGACG-motif and CGTCA-motif that are involved in MeJA-responses were detected, whereas these motifs were absent from the *DUFB* upstream region (Table 2-2). Jasmonates induce plant-defense responses and act to regulate defense-related genes (Kazan and Manners, 2008). Furthermore, an auxin responsive TGA motif (AACGAC) was present in *DGR2* promoter, but not *DUFB* (Table 2-2). This suggested that the expression of *DGR2* gene was influenced by auxin which is supported by the microarray data of the e-FP browser in section 2.2.1 & Table 1. Kim et al., (2011) reported that a TGA motif at position -693 was important for basal expression of a *Gossypium hirsutum* cellulose synthase catalytic subunit, in four promoters in cotton roots as well as in Arabidopsis roots.

Abscisic acid responsive and gibberellin-responsive elements were identified in the upstream intergenic regions of both *DGR2* and *DUFB*, (Table 2-2). A salicylic acid responsive element (TCA-element) was found only in *DGR2* (Table 2-2). Three regulatory elements associated with seed/endosperm expression (Skn-1, Ry-element, and the GCN4-motif) were present in the upstream region of *DUFB* whereas only one, GCN4_motif, was present in *DGR2* upstream region (Table 2-2). The Skn-1 motif interacts with the GCN4-motif, an AACA-motif, and an ACGT-motif to control endosperm expression (Washida et al., 1999; Taliercio, 2008). Day et al. (2008) in their transcriptome analysis of Arabidopsis endosperm revealed that *DGR2* and *DUFB* both were preferentially expressed in the endosperm, which is in agreement with the microarray

data obtained from the e-FP bowser (Section 2.2.1 & Table 2-2). Many light responsive elements were found in both the promoters (Table 2-2). Interestingly, a circadian element that is involved in circadian control was found in *DGR2* but not in *DUFB* (Table 2-2). Among the other important cis-acting elements found in *DGR2* and *DUFB* promoters were a Box-W1 (fungal elicitor responsive element), and a CAT-box (cis-acting regulatory element related to meristem expression (Table 2-2).

Table 2-2. Elements present in the promoter region of DGR2 (At5g25460) and DUFB (AT5G11420) according to the PLANT CARE database.

cis element	DGR2	DUFB	Function
3-AF1 binding site	0	1	light responsive element
AAGAA-motif	0	1	
ABRE	2	3	cis-acting element involved in the abscisic acid responsiveness
AE-box	1	0	part of a module for light response
ACE		1	cis-acting element involved in light responsiveness
ARE	2	1	cis-acting regulatory element essential for the anaerobic induction
ATC-motif	1	0	part of a conserved DNA module involved in light responsiveness
Box 4	1	2	part of a conserved DNA module involved in light responsiveness
Box I	2	2	light responsive element
Box-W1	1	1	fungal elicitor responsive element
C-repeat/DRE	1	0	regulatory element involved in cold- and dehydration-responsiveness
CAAT-box	31	33	common cis-acting element in promoter and enhancer regions
CAT-box	3	1	cis-acting regulatory element related to meristem expression
CTAG-motif	1	1	
CATT-motif	1	0	
CGTCA-motif	1	0	cis-acting regulatory element involved in the MeJA-responsiveness
F-box	2	0	
G-Box	4	2	cis-acting regulatory element involved in light responsiveness
G-box	7	5	cis-acting regulatory element involved in light responsiveness
P-box	1	1	gibberellin-responsive element

GA-motif	1	0	part of a light responsive element
GAG-motif	1	1	part of a light responsive element
GARE-motif	1	0	
MBS	1	0	MYB binding site involved in drought-inducibility
GC-motif	1	0	
GCN4_motif	1	1	cis-regulatory element involved in endosperm expression
HSE	1	0	is-acting element involved in heat stress responsiveness
H-box	0	1	cis-acting regulatory element involved in light responsiveness
I-box	0	1	part of a light responsive element
RY-element	0	1	is-acting regulatory element involved in seed-specific regulation
Skn-1_motif	2	2	cis-acting regulatory element required for endosperm expression
LAMP-element	2	0	part of a light responsive element
Sp1	0	1	light responsive element
TATA-box	48	64	core promoter element around -30 of transcription start
TC-rich repeats	0	1	cis-acting element involved in defense and stress responsiveness
TCCC-motif	0	1	part of a light responsive element
Unnamed__1	3	4	
Unnamed__3	3	4	
Unnamed__4	3	8	
W box	1	1	
box II		1	part of a light responsive element
TATC-box	1	0	is-acting element involved in gibberellin-responsiveness
TCA-element	2	0	cis-acting element involved in salicylic acid responsiveness
TCT-motif	1	0	part of a light responsive element
TGA-element	1	0	auxin-responsive element
TGACG-motif	1	0	cis-acting regulatory element involved in the MeJA-responsiveness
circadian	1	0	cis-acting regulatory element involved in circadian control
rbcS-CMA7a	2	0	part of a light responsive element

Table 2-3. Physico-chemical properties of DGR2 and DUFB.

Protein designation	AGI ID	Amino acid sequences	Molecular weight	pI
DGR2	AT5G25460	369	39.97	7.4
DUFB	AT5G11420	366	39.64	7.8

2.2.3. Protein structure prediction and function analysis

PSIPRED uses neural networking and homology information, which is collected using PSI-Blast, and which is combined with information about the properties of individual amino acids for predicting secondary structures. The secondary structure prediction by PSIpred indicated that DGR2 and DUFB were mainly made of beta sheets and coils. There was only one predicted **α-**helix, which was in the N-terminus (Fig. 2-5) of each protein. Therefore, DGR2 and DUFB are β-type proteins. Carbohydrate binding modules are mainly composed of beta sheets. Many glycosyl hydrolases are β-type proteins. It was found that the β -galactosidase from *Kluyveromyces lactis* is a β-type protein, having 22% β -turns, 14% parallel β -sheet, 25% antiparallel β -sheet, 34% unordered structure, and only 5% alpha-helix. Poch et al. (1992) reported that β -galactosidase from *K. lactis* consisted of 11 β -strands and two α–helices.

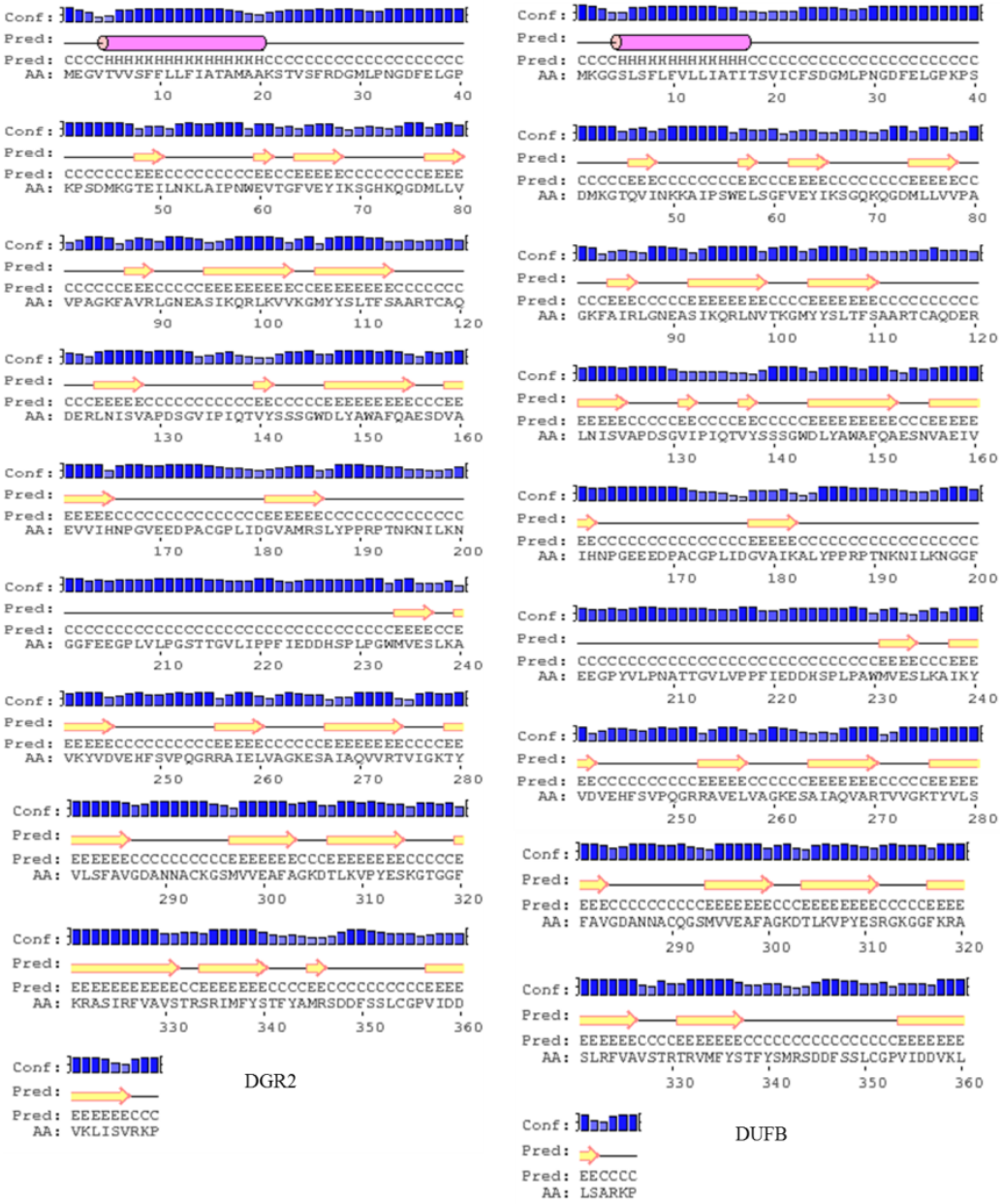


Figure 2-5. Secondary structure of DGR2 and DUFB proteins of Arabidopsis generated by Psipred server**.** Yellow arrows indicated β-sheets and pink cylinders indicated α-helices.

I-TASSER is an iterative, threading-based software suite for predicting three-dimensional (3D) atomic models of a protein, and inferring function from these models. I-TASSER first generates

structural templates from multiple threading alignments to PDB (Protein Data Bank) accessions, and then 3D models are constructed by iterative template fragment assembly simulations (Zhang, 2008). The function of the protein is then predicted by matching the 3D models with other known proteins (Zhang, 2008). The accuracy of predicted models was estimated by a confidence score (C-score). C-scores are typically in the range from −5 to 2, with higher scores representing higher confidence in the prediction. The Template Modeling score (TM-score) and root mean-square deviation (RMSD) are used to evaluate the structural similarity of two models of correct topology. A TM-score <0.17 indicates the similarity is no greater than would be expected for a random pair of models. 3D models for four DGR2 and DUFB proteins were successfully generated by different threading templates identified by LOMETS from the PDB library as shown in Table 2-4. A total of five structural models were predicted for each of DGR2 and DUFB, using I-TASSER (Tables 2-5a and 2-6a). In each case, Model 1, which was the model with the maximum C-score, was selected for further analysis. The TM-score for Model 1 of both DGR2 and DUFB was calculated to be 0.52 (Table 2-5b and 2-6b). A TM-score >0.5 indicates a model with an expected correct topology and a TM score < 0.17 means a random similarity, thus the selected models were consistent with correct topology. Cα-root mean square deviation (RMSD) and TM-score are both measures of topological similarity between the model and temple structures. The RMSD between the model and the template structure was found to be 10.2 Å for both the proteins. In general, the values of RMSD situated in the range of roughly 2~5 Å for medium-resolution models, and for high-resolution models, in the range of roughly 1–2 Å (Roy et al., 2010). The predicted structures of DGR2 and DUFB were compared to PDB using TM-align. After submitting the sequence of each query protein to I-TASSER, template proteins of similar folds from the PDB library were retrieved by LOMETS, which uses multiple threading programs. After structure assembly simulation, the top 10 proteins from the PDB that had the closest structural similarity, i.e. the highest TM-score were matched. The top ten proteins with highly similar structures based on TM and I-TASSER were identified for each of DGR2 and DUFB as shown in Table 2-7 and Table 2-8. The protein 2zxqA PDB-Hit had the top rank for both and had a TM-score of 0.821 and a coverage of 0.908 for DGR2 and 0.83 and 0.904 for DUFB. The selected models were then refined by ModRefiner server (Fig. 2-6). For the refined structures, as compared to the initial models, the values of RMSD and TM-scores were 1.043 and 0.987 for DGR2 and 1.191 and 0.974 for DUFB.

2.2.4. Evaluation of predicted structures

Structural validation of the refined 3D models of DGR2 and DUFB was done using RAMPAGE (Lovell et al., 2003). The stereochemical quality of the modeled proteins was assessed based on a Ramachandran validation score for favored regions and allowed regions (Lovell et al., 2003). In general, a score >98% denotes good stereochemical quality of the models. However, if more than 88% of the amino acid residues are in the favoured or allowed region, the protein model is sufficient for *in silico* studies (Singh et al., 2015). The percentage of residues that were in the allowed or disallowed regions of DGR2 and DUFB models in the Ramachandran plots was shown in Fig. 2-7 and Table 2-9. The percentage of residues in favored region was 82.6% and 82.7%, while residues in allowed region were 13.4% and 12.4% for DGR2 and DUFB, respectively. The residues in disallowed regions were 4.1% and 4.9% for DGR2 and DUFB, respectively. Therefore, 96% and 95.1% of the residues of DGR2 and DUFB were in the favoured or allowed regions, respectively. This indicated that the stereochemical quality of the predicted models was acceptable (Gunasekaran et al., 1996).

2.2.5. Structure-based function prediction

Biological functions of DGR2 and DUFB were annotated by COFACTOR, a structure-based method for biological function based on the I-TASSER structure prediction (Roy et al., 2012). Five enzyme homologs were identified in PDB as having similar functions to the predicted DGR2 and DUFB sequences (Table 2-11 and 2-12). Notably, 2zxqA PDB-Hit (endo-alpha-N-acetylgalactosaminidase) had the top rank with confidence scores of 0.237 and 0.231 for DGR2 and DUFB, respectively by the Enzyme Classification (EC) number prediction. Based on Gene Ontology terms (Table 2-10), molecular functions for two top ranked molecular functions for DGR2 were carbohydrate binding (GO-Score: 0.25) and hydrolase activity, hydrolyzing O-glycosyl bond (GO-Score: 0.15) and for DUFB were carbon-oxygen lyase activity (GO-Score: 0.44) and hydrolase activity, acting on glycosyl bond (GO-Score: 0.42). Top two ranked biological process predicted for both proteins were primary metabolic process (DGR2, GO-Score: 0.55; DUFB, GO-Score: 0.41) and cell adhesion (DGR2, GO-Score: 0.44; DUFB, GO-Score: 0.40). Cellular compartment of both proteins was predicted to be in the extracellular region with GO-score of 0.13 for both proteins.

Table 2-4. List of top ten templates used by I-TASSER for 3D structure predictions of DGR2 and DUFB proteins.

Gene name	PDB IDs
DGR2	2zxqA, 3ecqB, 2zexA, 2zexA, 2zxqA, 3ecqA, 2zxqA 2zexA, 2zexA, 3c7eA
DUFB	2zxqA, 2zxqA, 3ecqB, 2zexA, 2zexA, 2zxqA, 2zexA, , 3ecqA, 2zxqA, 2zxqA

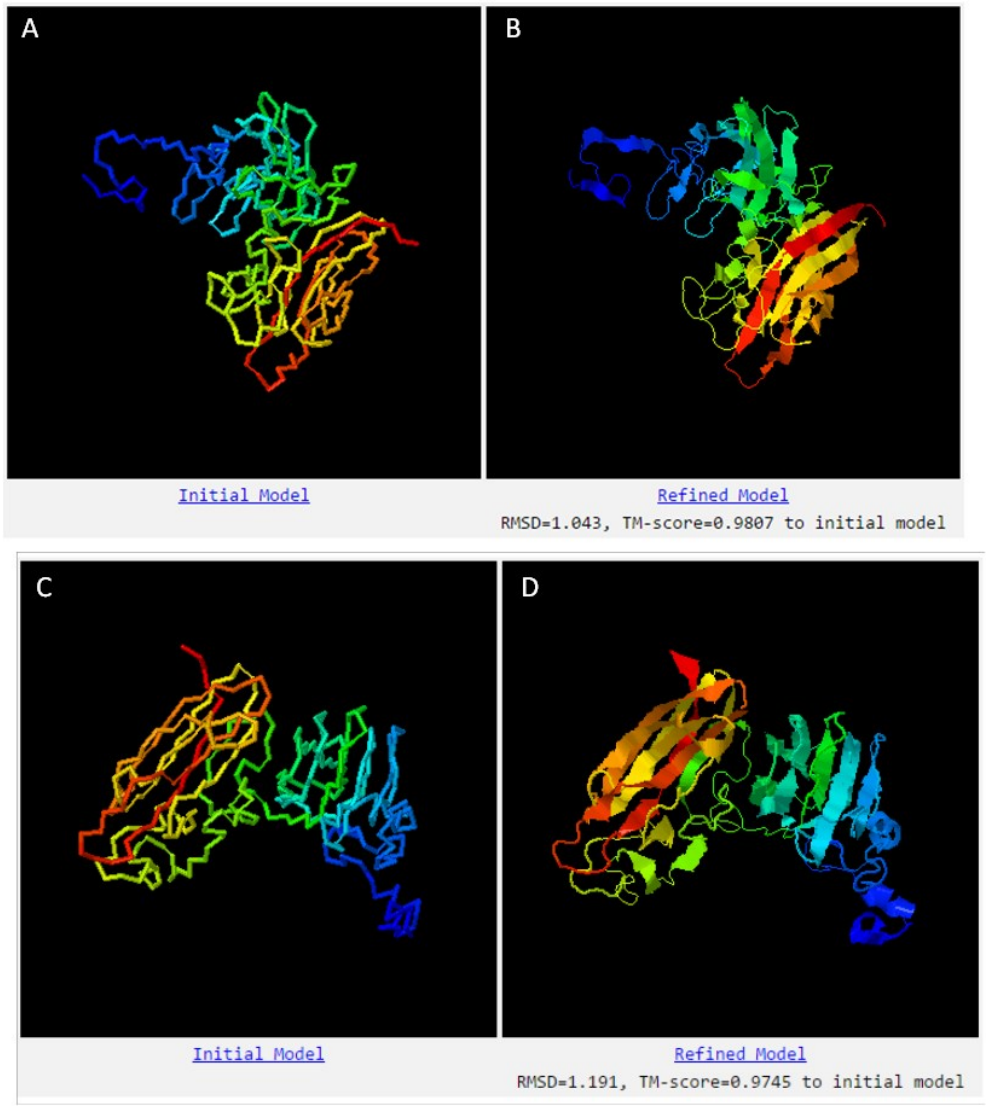


Figure 2-6. Best 3D Ribbon structures of DGR2 (A & B) and DUFB (C&D) predicted by the template-based prediction program, I-TASSER (A & C) and refined by ModRefiner (B & D).

Table 2-5a. C-scores, no. of decoys, cluster density value for different models of DGR2:

parameters	Model 1	Model 2	Model 3	Model 4	Model 5
C-scores	-1.56	-3.54	-3.08	-2.67	-3.96
No. of decoys	1680	212	363	106	100
Cluster density	0.1228	0.0170	0.0269	0.0406	0.0111

Table 2-5b. I-TASSER output for DGR2:

Protein name	TM-score	RMSD
DGR2	0.52±0.15	10.2±4.6Å

Table 2-6a. C-scores, no. of decoys, cluster density value for different models of DUFB

parameters	Model 1	Model 2	Model 3	Model 4	Model 5
C-scores	-1.63	-3.11	-2.91	-3.01	-4.62
No. of decoys	1799	115	96	86	79
Cluster density	0.1144	0.0261	0.0320	0.0290	0.0058

Table 2-6b. I-TASSER output for DUFB:

Protein name	TM-score	RMSD
DUFB	0.52±0.15	10.4±4.6Å

Table 2-7. Proteins with highly similar structure with DGR2 in Protein Data Bank (as identified by TM-align computer algorithm).

Rank	PDB Hit	PDB description	TM-score	RMSD[a]	IDNE[a]	Cov
1	2zxqA	endo-alpha-N-acetylgalactosaminidase	0.821	2.73	0.137	0.908
2	3ecqA	endo-alpha-N-acetylgalactosaminidase	0.793	2.95	0.109	0.892
3	2y8kA	arabinoxylan-specific	0.442	6.48	0.091	0.707

		xylanase				
4	1hn0A	chondroitin sulfate ABC lyase I	0.437	5.28	0.039	0.607
5	3a7q	chondroitin sulfate lyase abc	0.437	5.12	0.048	0.594
6	3zr5A	galactocerebrosidase	0.437	5.30	0.041	0.607
7	4gwmA	β metalloproteinase	0.434	4.80	0.093	0.093
8	2agsA	2,3-difluoro-KDN	0.425	6.35	0.070	0.677
9	3b69A	T cruzi trans-sialidase	0.423	6.48	0.038	0.669
10	2jgdB	2-oxoglutarate dehydrogenase	0.417	7.01	0.034	0.718

Table 2-8. Proteins with highly similar structure with DUFB in Protein Data Bank (as identified by TM-align computer algorithm).

Rank	PDB Hit	PDB description	TM-score	RMSD	IDNE	Cov
1	2zxqA	endo-alpha-N-acetylgalactosaminidase	0.830	2.63	0.123	0.904
2	3ecqA	endo-alpha-N-acetylgalactosaminidase	0.803	2.77	0.117	0.891
3	2e26A	Reelin,	0.447	5.20	0.068	0.067
4	3zr5A	galactocerebrosidase I	0.433	5.82	0.061	0.639
5	4gwmA	promeprin beta	0.431	4.70	0.074	0.555
6	2y8kA	arabinoxylan-specific xylanase.	0.429	6.86	0.069	0.730
7	2xvlA	alpha-Xylosidase	0.422	5.93	0.028	0.626
8	3zxjA	GH43 glycoside hydrolase	0.420	5.49	0.053	0.601
9	2w91A	endo-beta-d-glucosaminidases	0.418	5.53	0.045	0.596

10	1hn0A	chondroitin sulfate ABC lyase I	0.417	5.62	0.049	0.598

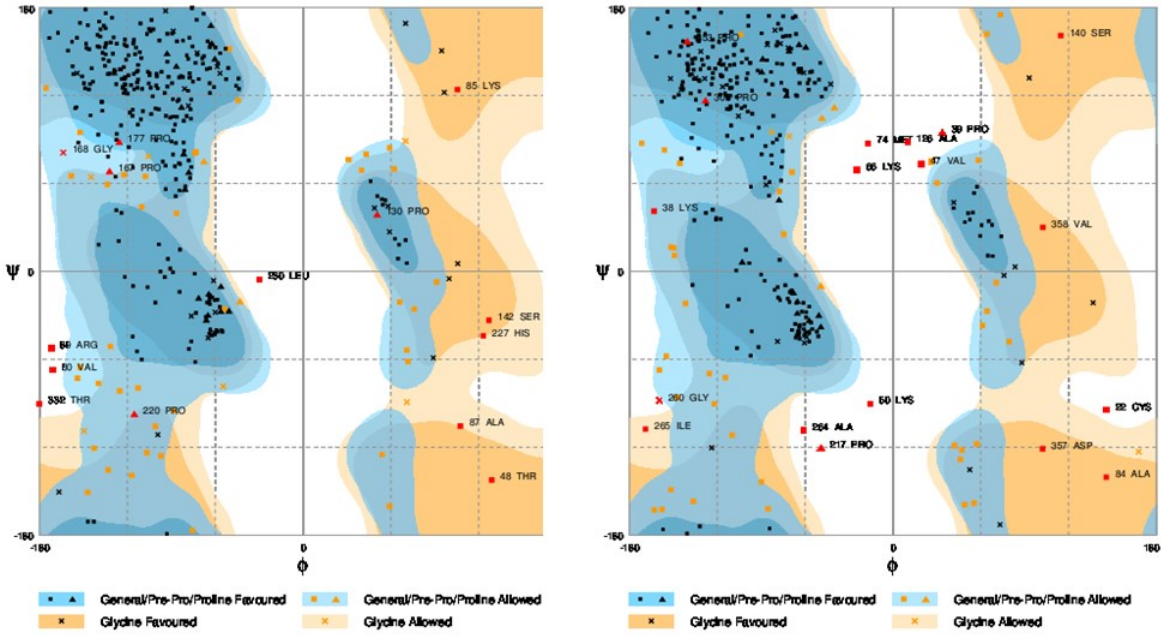


Figure 2-7. Model validation studies of DGR2 and DUFB by Ramachandran's plot using Rampage server. Dark blue and dark orange are favored regions. Light blue and light orange are allowed regions. White region is disallowed region.

Table 2-9. Ramachandran plot statistics for DGR2 and DUFB of Arabidopsis.

Protein	Residues in	Number of residues	Percentage
DGR2	Favoured region	303	82.6
	Allowed region	49	13.4
	Outlier region	16	4.1
DUFB	Favoured region	208	82.7
	Allowed region	97	12.4
	Outlier region	59	4.9

Table 2-10. Consensus prediction of gene ontology terms for DGR2 and DUFB by 1-Tasser server. The GO-Score associated with each prediction is defined as the average weight

of the GO term. It's range is (0-1) and higher values indicate more confident predictions.

DGR2	Molecular Function	GO-Score	Biological Process	GO-Score	Cellular Component	GO-Score
	GO:0030246 (carbohydrate binding)	0.25	GO:0044238 (primary metabolic process	0.50	GO:0005576 (Extracellular region)	0.13
	GO:0004553 (Hydrolase activity, hydrolyzing O-glycosyl bond)	0.14	GO:0007155 (Cell adhesion)	0.41		
	GO:0043169 (Cation binding)	0.14				
	GO:0016837 (Carbon–oxygen lyase activity, acting on polysaccharides)	0.13				
DUFB	GO:0016835 (Carbon-oxygen lyase activity)	0.44	GO:0044238 (Primary metabolic process	0.44	GO:0005576 (Extracellular region)	0.13
	GO:0016798 (Hydrolase activity, acting on glycosyl bond)	0.42	GO:0007155 (Cell adhesion)	0.40		

Table 2-11. Enzyme Commission numbers based function prediction by COFACTOR based on I-TASSER structure prediction for DGR2.

Rank	Cscore	PDB Hit	TM-score	RMSD	IDEN	Cov	EC Number	Predicted function with enzyme commission number	Active Site Residues
1	0.237	2zxqA	0.821	2.73	0.137	0.91	3.2.1.97	Endo-alpha-N-acetylgalactosaminidase.	NA
2	0.227	3ecqB	0.791	2.98	0.109	0.89	3.2.1.97	Endo-alpha-N-acetylgalacto	259

								saminidase	
3	0.130	2jgdA	0.417	7.01	0.034	0.72	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring). .	NA
4	0.130	2agsA	0.425	6.35	0.070	0.677	3.2.1.18	Exo-alpha-sialidase.	99
5	0.129	1s0kA	0.423	6.42	0.045	0.664	3.2.1.18	Exo-alpha-sialidase.	NA

Table 2-12. Enzyme : Enzyme Commission numbers based function prediction by COFACTOR based on I-TASSER structure prediction for DUFB.

Rank	Cscore	PDB	TM-score	RMSD	IDEN	Cov	EC Number	Predicted function with enzyme commission number	Active Site Residues
1	0.231	2zxqA	0.830	2.61	0.127	0.90	3.2.1.97	Endo-alpha-N-acetylgalactosaminidase.	NA
2	0.225	3ecqB	0.802	2.77	0.117	0.89	3.2.1.97	Endo-alpha-N-acetylgalactosaminidase	NA
3	0.133	1hn0A	0.446	5.24	0.042	0.62	4.2.2.20	Chondroitin-sulfate-ABC endolyase.	NA
4	0.131	1s0kA	0.420	6.44	0.050	0.67	3.2.1.18	Exo-alpha-sialidase.	82
5	0.130	2agsA	0.425	6.30	0.053	0.67	3.2.1.18	Exo-alpha-sialidase.	254

2.2.6. Phylogenetic analysis

The translated protein sequences of DUF642 genes from the different plant species were subjected to multiple sequence alignment by ClustalW tool followed by clustering by neighbor-joining (NJ). The resulting dendrogram revealed existence of three major groups of protein sequences: A, B and C. Group A was further divided into two sub-groups: A-I and A-II (Fig. 2-8). Notably, group A-I was composed of eudicots and group A-II was composed of mainly

monocots. This division indicates that a progenitor of DUF642-containing proteins likely existed in the last common ancestor of eudicots and monocots, and that the function of this protein confers advantages to both groups of species. DGR2 (ATH6) and DUFB (ATH10) showed maximum similarity with *Carica papaya* within sub-group A-I. The papaya belongs to the family Caricaceae and shared a common ancestor with *Arabidopsis thaliana* ~72 million years ago (Wikstrom et al., 2001). The species *L. usitatissimum*, *P. trichocarpa* and *M. esculenta* belong to the Linaceae family and to the order Malpighiales. The lineage leading to Arabidopsis thaliana is considered to have diverged from the lineage leading to Malpighiales between 100 and 120 MYA (Tuskan et al., 2006). *A. thaliana* and *O. sativa* last shared a common ancestor ∼150 to 200 million years ago (Jackson, 2006). The analysis of DUF642-containing proteins from nine representative species of angiosperms revealed that DUF642 domain was highly conserved. It was also noted that DGR2 (ATH6) and DUFB (ATH10) appear to be paralogous genes originating from a recent gene duplication (Fig. 2-8). Vázquez-Lobo et al. (2012) reported that DUF642 proteins are highly conserved in spermatophyte plants and their conserved motifs indicated an ancestral intragenic duplication event.

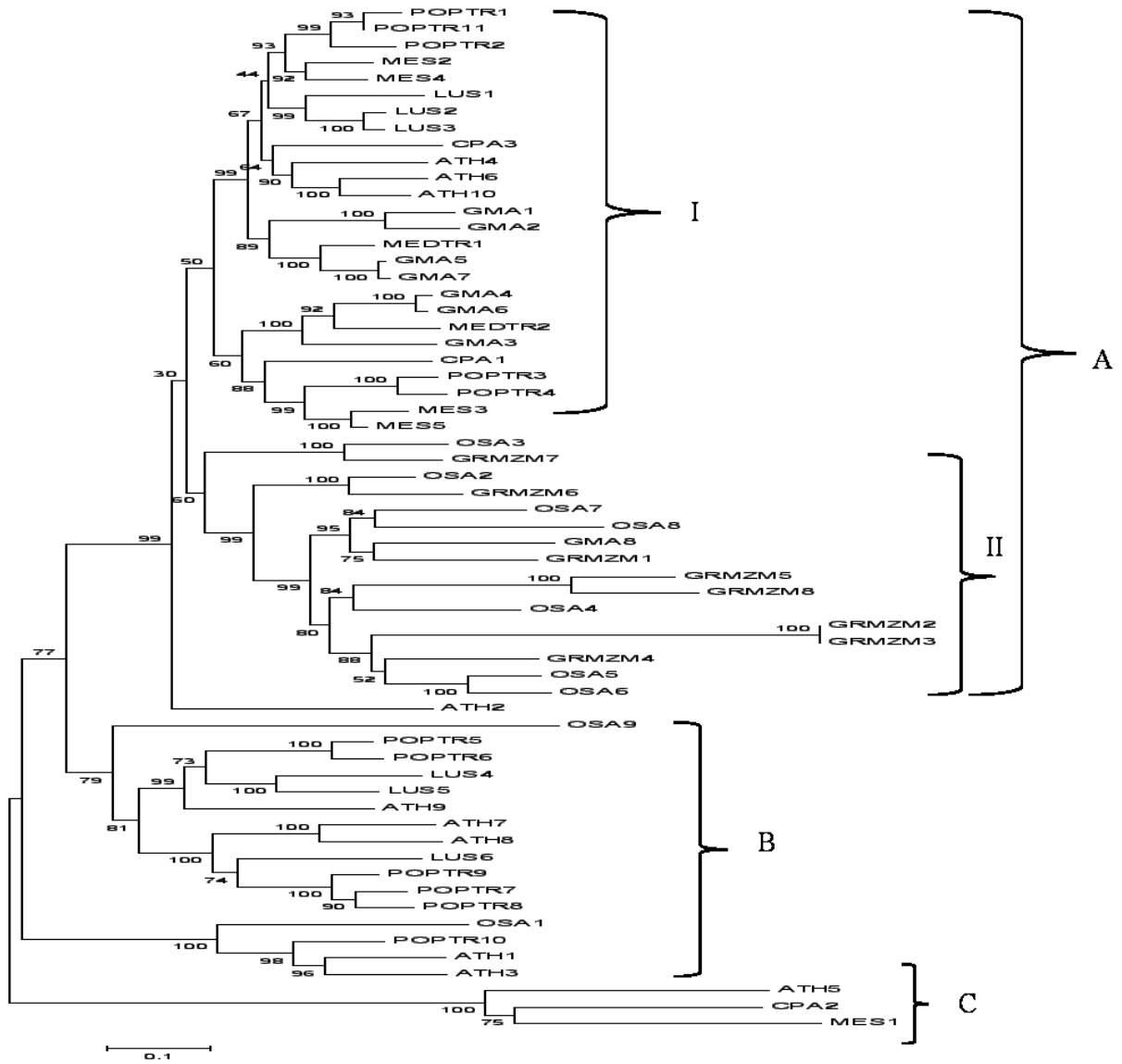
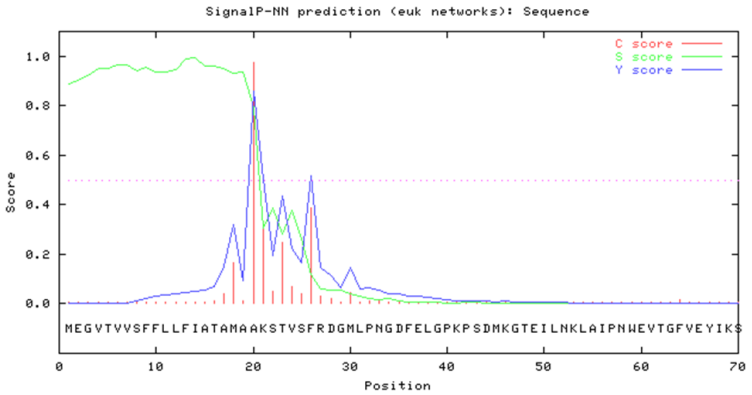


Figure 2-8. A Phylogenetic trees of DUF642 protein family in different plant species was constructed using MEGA 4 software (Tamura, 2007) using the neighbor-joining (NJ) method with 100 bootstrapping replicates. ATH (*Arabidopsis thaliana*), POPTR (*Populus trichocarpa*), LUS (*Linum usitatissimum*), OSA (*Oryza sativa*), GMA (*Glycine max*), MEDTR (*Medicago trucatula*), CPA (*Carica papaya*), GRMZM (*Zea mays*), MES (*Manihot esculenta*).

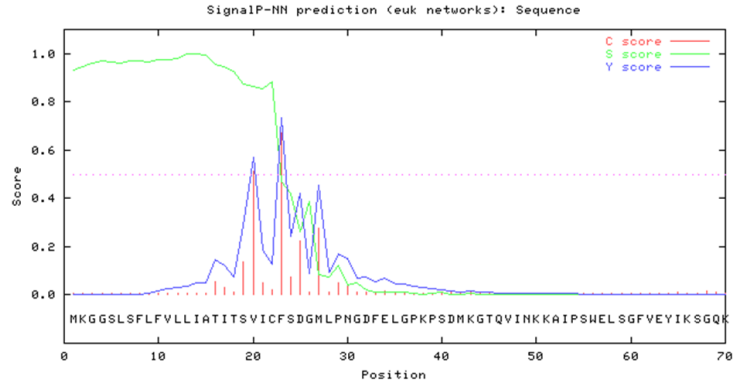
2.2.7. Subcellular localization prediction

SignalP 2.0 predicted (http://www.cbs.dtu.dk/services/SignalP/) that both proteins had signal peptides at the N-terminus. A putative signal peptide of 19 amino acid residues starting from amino acid residue 1 to 19 and a cleavage site was present between residues 19 and 20 (Fig. 2-9)

for DGR2. For DUFB, a putative signal peptide of 22 amino acid residues starting from amino acid residue 1 to 23 and a cleavage site was present between residues 22 and 23 (Fig. 2-9). TargetP 1.1 Server (Emanuelsson, 2000) and iSPORT (Bannai et al, 2002) predicted both DGR2 and DUFB to be secreted proteins. As such, the final destination of DGR2 and DUFB could be endoplasmic reticulum, Golgi apparatus or extracellular compartments. A Golgi predictor (Yuan and Teasdale, 2002) predicted DGR2 as Golgi protein and DUFB as post -Golgi protein. Proteins with index values greater than the threshold are predicted as Golgi proteins. The index values and threshold values were 22.17 and 20.00 for DGR2 and 18.47 and 20.00 for DUFB.



Sequence length = 70
Most likely cleavage site between pos. 19 and 20



Sequence length = 70
Most likely cleavage site between pos. 22 and 23

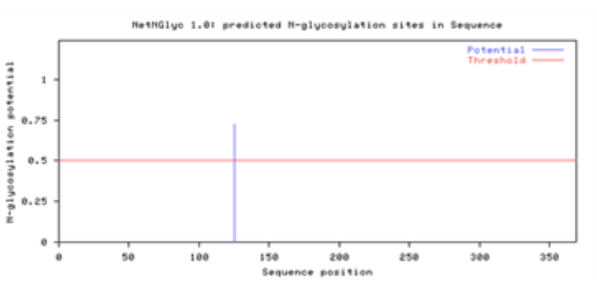
Figure 2-9. Signal peptide prediction for DGR2 and DUFB by SignalP server.

2.2.8. Predictions of glycosylation

Glycosylation is a conserved posttranslational modification that is found in all eukaryotes. In plants, secretory proteins are often glycosylated by N-linked oligosaccharides in the endoplasmic

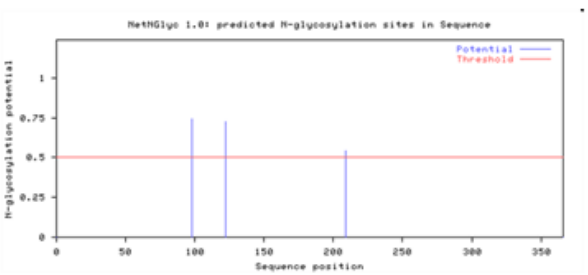
reticulum (ER) and Golgi apparatus. DGR2 and DUFB, both proteins were predicted to be N-glycosylated as determined by NetNGlyc 1.0 Server (Fig. 2-10). One N-linked (Asn-Xaa-Ser/Thr) glycosylation above the threshold level (G-score >0.5) was identified for DGR2 whereas three were identified for DUFB. NetOGlyc 4.0 Server (Steentoft et al., 2013) predicted the potential sites for O-glycosylation in DGR2 sequence at positions 186 and 214. For DUFB, not a single O-glycosylation site was predicted. *O*-linked oligosaccharides are linked to the hydroxyl group of serine or threonine via *N*-acetylgalactosamine (GalNac) or (in collagens) to the hydroxyl group of hydroxylysine via galactose (Molecular Cell Biology, 4th edition). Little is known about O-glycosylation of secreted proteins in plants. O-glycosylations can be either generated by secreted proteins or by cytosolic/nuclear proteins. The glycoproteins, cell wall extensin and sporamin are identified O-glycosylated with Ser residues with one single galactose (Cho and Chrispeels, 1976; Matsuoka et al., 1995).

Protein *N*-glycosylation



SeqName	Position	Potential	Jury agreement	N-Glyc result
Sequence	125 NISV	0.7267	(9/9)	++

AT5G25460



SeqName	Position	Potential	Jury agreement	N-Glyc result
Sequence	98 NVTK	0.7410	(9/9)	++
Sequence	122 NISV	0.7274	(9/9)	++
Sequence	209 NATT	0.5457	(4/9)	+

AT5G11420

Figure 2-10: Predicted N-linked glycosylation sites in DGR2 (AT5G25460) and DUFB (AT5G11420) protein sequences. The table inserted in figure showed output scores.

2.2.9. Gene co-expression and network analysis

Gene co expression analysis can predict the function of an unknown gene because genes in the co-expression network may be involved in similar biological processes and may play a role in similar biological functions (Liang et al., 2014). In the molecular interaction network, a node represents a gene, gene product or metabolite, and a link or edge refers to an interaction between them (Alm and Arkin, 2003). In a gene co-expression network, nodes and links represent genes and indicate their co-expression relationships, and can characterize such topological properties as small-world, hierarchically modular or scale-free (Luo et al., 2007, Liang et al., 2014).

Expression Angler identified 25 genes that were co-expressed with *DGR2* and *DUFB* based on an r-value cut off range between 0.75 and 1.00. The top 10 genes are presented in Table 2-13: Expression Angler uses gene expression data for ~22000 Arabidopsis genes generated using the ATH1 Affymetrix Whole Genome GeneChip. It was found that both *DGR2* and *DUFB* were co-expressed with genes related to cell wall metabolism, including pectin lyase-like superfamily proteins, AXR2_IAA7_indole-3-acetic acid 7, and members of the alpha-expansin gene family. (Table 2-13).

Table 2-13. Top 10 genes co-expressed with DGR2 and DUFB as identified by Expression Angler having a Pearson correlation coefficient (*r*-value) cutoff between 0.75 and 1.00.

Gene name	AGI ID	Description	*r*-value
DGR2	At5g25460	Protein of unknown function, DUF642	1.00
	At3g07010	Pectin lyase-like superfamily protein	0.835
	At3g02170	Encodes LONGIFOLIA2 (LNG2)	0.823
	At5g11420	Protein of unknown function, DUF642	0.786
	At1g23080	AXR2_IAA7_indole-3-acetic acid 7	0.783
	At2g40610	Member of Alpha-Expansin Gene Family	0.778
	At4g23820	Pectin lyase-like superfamily protein;	0.783
	At1g64390	Member of Alpha-Expansin Gene Family (GH9C2	0.756
	At3g57800	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	0.755
	At3g20820	Leucine-rich repeat (LRR) family protein;	0/752

DUFB	At5g11420	Protein of unknown function, DUF642	1.00
	At3g07010	Pectin lyase-like superfamily protein	0.786
	At5g25460	Protein of unknown function, DUF642	0.786
	At5g03120	unknown protein	0.783
	At2g40610	Member of Alpha-Expansin Gene Family	0.754
	At4g36540	Encodes the brassinosteroid signaling component BEE2	0.739
	At5g63180	Pectin lyase-like superfamily protein	0.728
	At1g76890	encodes a plant trihelix DNA-binding protein	0.725
	At3g29030	Encodes an expansin	0.723
	At3g16370	GDSL-like Lipase/Acylhydrolase superfamily protein	0.719

The ATTED-II server was used to conduct a separate analysis of Arabidopsis co-expression networks (Tables 2-14, 2-15; Figures 2-11, 2-12). ATTED-II generates co-expression networks from both microarray data and cis-elements. Gene to gene relationships in Arabidopsis having correlation coefficients between 0.60 to -0.60 were calculated**.** Three genes were directly related to DGR2 gene expression network which were DUFB (Cor: 0.60), leucine-rich repeat (LRR) family protein (Cor: 0.58) and pectin lyase-like superfamily protein (Cor: 0.56). In case of DUFB, four genes were directly related which included DGR2 (Cor: 0.60), expansin 11 (Cor: 0.58), Barwin-like endoglucanases superfamily protein (0.58), and a phototropic-responsive NPH3 family protein (Cor: 0.58).

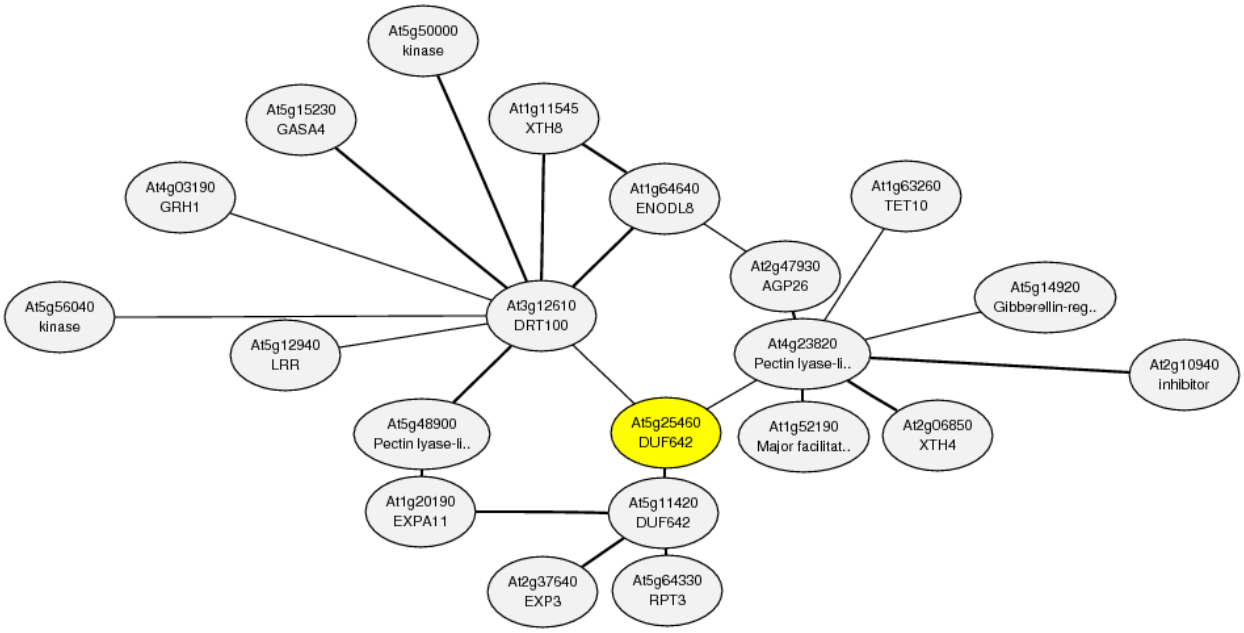


Figure 2-11: Connection of DGR2 by one node within a co-expression network generated by ATTED-II v6.

Table 2-14. Genes that are directly connected to *DGR2* in the network analysis by ATTED-II v6.

Correlation coefficient	Locus	Function
0.60	At5g11420	Protein of unknown function, DUF642
0.58	At3g12610	Leucine-rich repeat (LRR) family protein
0.56	At4g23820	Pectin lyase-like superfamily protein

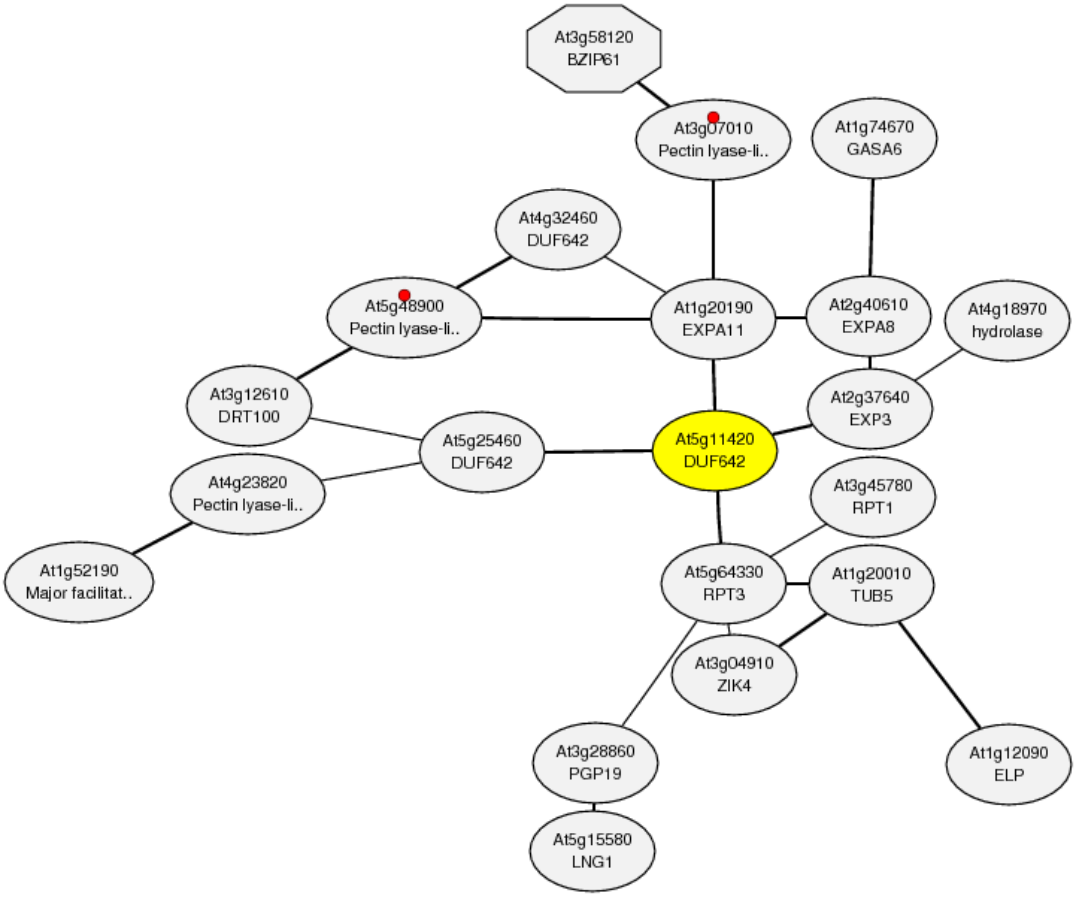


Figure 2-12: Connection DUFB by one node within a co-expression network generated by ATTED-II v6. Octagonal shaped node indicated transcription factor (TF) gene. Red dots indicated the common KEGG pathway in the network, Pentose and glucuronate interconversions (KEGG ID: ath00040).

Table 2-15. Genes that are directly connected to *DUFB* in the network analysis by ATTED-II v6.

Correlation coefficient	Locus	Function
0.60	At5g25460	Protein of unknown function, DUF642
0.58	At1g20190	expansin 11
0.55	At2g37640	Barwin-like endoglucanases superfamily protein
0.55	At5g64330	Phototropic-responsive NPH3 family protein

2.3. Conclusions

The DUF642 protein family is well-conserved throughout both monocots and eudicots. Characterization of *DGR2* and *DUFB* in Arabidopsis showed that both were predicted to be secreted proteins, based on the predicted N-glycosylation sites and N-terminal signal peptides which predicted DGR2 and DUFB were secreted proteins.

The predicted 3D structures of DGR2 and DUFB showed homology to proteins associated with primary metabolic process and cell adhesion. DGR2 was predicted to have carbohydrate binding and hydrolase activity, while DUFB protein structure was associated with carbon-oxygen lyase activity and hydrolase activity. Both were predicted to be secretory proteins.

DGR2 and DUFB were expressed in a wide range of plant tissues. It was found that many cell wall-related proteins such as pectin lyase-like superfamily protein, AXR2_IAA7_indole-3-acetic acid 7, member of alpha-Expansin gene family, glycosyl hydrolase 9C2 were co-expressed with both proteins.

The data from the sequence analysis of the promoters was relevant to the publicly available microarray dataset in the eFP browser. The promoter analysis revealed the upstream region of *DGR2* and *DUFB* promoters contained cis-regulatory elements consistent with a function in meristem and endosperm development. Microarray data obtained from the eFP browser also showed that both genes are expressed in the top of the stem and in the germinated endosperm. DGR2 promoter analysis also revealed that the expression changes during heat- and cold stresses. Phytochromes auxin and MeJA are likely involved in the differential regulation of DGR2 expression during the multiple developmental processes. These data are also in agreement with the microarray data from the eFP browser.

It was hypothesized from the *in silico* analysis that DGR2 and DUFB were cell wall proteins and involved in the metabolic processes by hydrolyzing pectin or other non-cellulosic polysaccharides and were thereby involved in a multiple developmental processes.

3. Chapter 3

3.1. Spatial and temporal expression of Two DUF642 Domain-Containing Genes in Arabidopsis

3.2. Introduction

Many conserved protein domains have been identified that have no known biochemical or physiological function. DOMAIN OF UNKNOWN FUNCTION 642 (DUF642) is one such domain defined in the Pfam database (PF04862) (http://pfam.sanger.ac.uk/, Finn et al. 2008). This large (~323 aa), conserved motif appears to be found only in seed plants (Vázquez-Lobo, et al., 2012). Proteins containing the DUF642 domain have been found in cell wall extracts of Arabidopsis and grape (Bayer et al., 2006, Irshad et al., 2008, Negri, et al., 2008) and among N-glycosylated fractions of Arabidopsis (Minic et al., 2007). Microarray analyses identified transcripts of DUF642-containing genes to be enriched in the top of hemp stems, as compared to the bottom, and in hypocotyls of Arabidopsis (DePauw et al., 2007 and Irshad et al., 2008). DUF642-containing proteins have also been reported to interact with cell wall polysaccharides and pectinmethylesterase *in vitro* (Vázquez-Lobo, et al., 2012 and Zuniga-Sanchez and Gomboa-de-Buen, 2012).

In *A. thaliana*, there are 10 genes that encode DUF642 (TAIR). All but three of these (At5g14150, At3g08030 and At1g80240) are also predicted to contain a galactose-binding domain-like fold. A recent analysis of six DUF642-containing genes of (At2g41800, At3g08030, At4g32460, DUFB, At1g80240 and At5g25460) showed that transcripts of two out of six genes (At1g80240 and At5g25460) were responsive to the ascorbic acid (AsA) precursor, L-galactono-1,4-lactone (L-GalL), and these genes were therefore named *DGR1* and *DGR2*, respectively, for *DUF642 L-GalL-RESPONSIVE* (Gao et al. 2012). Four out of six genes tested (DUFB, At3g32460, At2g41800 and At3g08030) did not increase in transcript abundance following L-GalL treatment. None of the six genes tested were responsive to ascorbic acid (AsA) treatment. Promoter::reporter fusion assays of *DGR1* (1,898 bp upstream of the

translation start site, TSS) and *DGR2* (2,078 bp upstream of the TSS) in transgenic Arabidopsis indicated that *DGR1* was expressed primarily in the root tip and developing anthers, while *DGR2* was expressed in anthers and throughout the seedling, but had a complementary pattern to *DGR1* in the root as *DGR2* was expressed in most tissues except the root tip (Gao et al. 2012). Gao et al. (2012) also heterologously expressed *DGR2* protein in *E. coli* but could not find evidence of binding of Gal to *DGR2*.

At5g25460 (*DGR2*) and DUFB (*DUFB*) are more closely related to each other than to any other DUF642-containing genes. Yet only *DGR2*, but not *DUFB*, was previously reported to be responsive to L-GalL metabolism. We were therefore motivated to further characterize the expression patterns of these two genes to better understand the full set of functions of DUF642 genes in plant growth and development.

3.3. Materials and methods

3.3.1. Plant materials and growth conditions

Surface-sterilized seeds of *A. thaliana* wild-type (Col-0) and *DGR2pro::GUS* and *DUFBpro::GUS* transgenics were sown on agar plates containing half-strength Murashige and Skoog (MS) medium (Murashige and Skoog, 1962) and 0.7 % Phytablend supplemented with 1 % sucrose, stratified for 24 h at 4 ºC and grown at 22 ºC with a 16 h light/8 h dark regime. Selection for transgenes was performed on solid MS medium supplemented with 50 µg/ml kanamycin. Selected plants were then transferred to soil and were grown under standard conditions at 23 ºC in a 16-h light/8-h dark cycle with the relative humidity of 50%. For histochemical analysis, three independent lines of T_3 generation *DGR2pro::GUS* and *DUFBpro::GUS* were used and for each line at least 30 independent plants were tested. The results of GUS staining were reported were observed in at least 95% of individuals examined.

3.3.2. Plasmid construction and plant transformation

1841 bp and 1644 bp fragments upstream of the transcription start sites of *DGR2* and *DUFB* were PCR amplified using Platinum® Taq DNA Polymerase High Fidelity (Life Technologies) using primers modified with *Bam*HI and *Hind*III restriction tags

(DGR2_Bam_3 = 5- CGC GGATCC ATT GAC GGA AGA GAG AAC

DGR2_Hind_3=5-CCC AAG CTT GAT ATATGA TAA ATATTA C

DUFB_Bam_3 = 5- CGC GGATCC TGT GGA CGA CCA AAG

DUFB_Hind_3=5- CCC AAG CTT TTA CCCTTC TTG TC), respectively. The fragments were cloned into TOPO TA Cloning® vector and inserted into *E. coli* DH5α competent cells. The plasmid miniprep was conducted using Plasmid Miniprep Kit (Qiagen). Restriction digestion of these plasmids was conducted with FastDigest enzymes, *Bam*HI and *Hin*dIII (Thermo Scientific). *DGR2pro::GUS* and *DUFBpro::GUS* fusion vectors were constructed by ligating the restriction products into the pRD420 cloning vector (Datla et al. 1992). Constructs were transformed into *Agrobacterium tumefaciens* GV3101. Plants were transformed using the floral dip method (Clough and Bent, 1998).

3.3.3. GUS histochemistry

Arabidopsis tissues (except dry seeds) were harvested and then placed in ice cold 90% acetone and vacuum infiltrated for 2 minutes. The samples were then incubated 30 minutes at -20 °C. Before infiltration in GUS staining solutions, the samples were washed twice in 50 mM $NaHPO_4$ pH 7.2. Vacuum infiltrate with GUS Staining solution (0.2 % Triton X-100, 10 mM EDTA, 50 mM $NaHPO_4$ pH 7.2, 2 mM $K_4Fe(CN)_6$, 2 mM $K_3Fe(CN)_6$, 2 mM X-gluc) for 30 min, followed by overnight incubation at 37 °C. Then the samples were fixed in 30 % ethanol for 1 h, FAA (50 % ethanol, 5 % formaldehyde, 10 % glacial acetic acid) overnight, and 70 % ethanol for final storage. The samples were photographed with an Olympus BX51 microscope, fitted with a HDCE-90D digital camera.

3.3.4. qRT-PCR analyses

Total RNA was extracted from wild type plants using the RNeasy Plant Mini Kit (Qiagen). RNA from root and root apex was extracted using RNeasy Micro Kit from Qiagen. DNA was removed from the RNA by using TURBO DNA-free™ Kit from Life technologies. The quality and quantity of extracted RNA was determined by Agilent 2100 Bioanalyzer. RNA samples were converted to cDNA using RevertAid H Minus Reverse Transcriptase (Thermo) with an oligo(dT) Primer. All PCRs were performed using three technical replicates and three biological repeats. The qRT-PCR was performed by using SYBR Green I dye reagent. All qRT-PCR was performed in an Applied Biosystems 7500 Fast system. The data as expressed as fold change ($2^{-\Delta\Delta C_T}$). UBQ10 (UBIQUITIN 10) was used as an internal reference (Tong et al., 2009 and Czechowski et al., 2005). Primer sequences used for qRT-PCR were (5' to 3'):

DGR2FWD TCAATATGGAAGGCGTCACC;

DGR2RVS CCTAGCTCGAAGTCTCCGTTT;

DUFBFWD; GTCTCTTCTCTTTACTTTGGTCGTC;

DUFBRVS: AGTCGCCGTTTGGTAACATC.

3.3.5. *DGR2* and *DUFB* expressions in seeds, during germination and growth

For analysis of embryos, dry seeds were placed in 1.5 ml microcentrifuge tubes to which was added 40 µl of GUS staining solution. With a plastic micropestle, the seed coats were disrupted so that the embryo could be released, and then an additional 960 µl of GUS staining solution was added and incubated over night at 37°C. Seeds were imbibed overnight and in the next day GUS proceeded as with the dry seeds. GUS assay also performed from 0 to 5 DAS and mature *DGR2pro::GUS* and *DUFBpro::GUS* transgenic plants.

3.3.6. Hormone treatments

4 DAS (days after sowing) Col-0 and *DGR2pro::GUS* and *DUFBpro::GUS* transgenic plant seedlings were used for hormone treatment experiments by qRT-PCR and GUS staining, respectively. The plant hormones used in this study were: 3-indoleacetic acid (IAA, Sigma), the

ethylene precursor 1-aminocyclopropane-1-carboxylic acid (ACC, CalBiochem), gibberellic acid-3 potassium salt (GA_3, Sigma) and methyl jasmonate (MeJA, Sigma). Hormones were added separately to a concentration of 10 and 100 μM in half strength liquid MS media. Seedlings at 4 DAS were placed for 16 h on the treatment medium prior to harvesting for qRT-PCR and GUS assay.

3.3.7. Sugar treatments

4 DAS Col-0 seedlings were used for sugar treatments. The sugars used in this experiment were D-Glucose (Glc), D-Galactose (Gal), D-Fructose (Fru) and D-Arabinose (Ara). Wild type and transgenic seedlings at 4 DAS were placed in the liquid half MS media that were supplemented with 5% of Glc, Gal, Fru and Ara and 10mM L-GalL for 16h for qRT-PCR and GUS histochemical assays, respectively.

3.3.8. Statistical analysis and graphics

One-way ANOVA was conducted to test whether there were any statistically significant differences in the expressions of *DGR2* and *DUFB* between different tissues and between the treated and untreated seedlings. Duncan's Multiple range test was also performed. All the statistical analyses were performed in SAS 9.1 and results were plotted using SigmaPlot 11.0.

3.4. Results

In silico analysis in chapter II showed that *DGR2* and *DUFB* shared 93.4% and 86.2% amino acid sequence similarity and identity, respectively. Both contained signal peptides at the N-terminus, and were predicted to be soluble proteins. Their predicted secondary structure consisted primarily of beta sheets and coils, and both were predicted with high confidence by PSIPRED (Section 2.2.3) to contain conserved hydrolase domain folds and galactose binding domain folds (Section 2.2.3). Based on these structural comparisons, and their phylogenetic relationships, these proteins might reasonably be expected to have similar functions and

expression patterns. We therefore compared their expression patterns using reporter gene and qRT-PCR assays as described below.

3.4.1. Reporter gene assays

We ligated genomic regions from upstream of *DGR2* and *DUFB* to the *β*-glucuronidase (GUS) reporter gene, to visualize the spatial and temporal gene expression patterns that the genomic fragments were capable of conferring in transgenic Arabidopsis plants. We referred to the resulting constructs and transgenic lines as *DGR2pro::GUS* and *DUFB* pro::GUS.

3.4.2. Reporter gene expression in seedlings

GUS expression was assayed in seeds and seedlings every day from 0 to 5 days past sowing (DAS). No staining was observed in any part of the 0 or 1 DAS embryos in any line (data not shown). However, in lines bearing the *DGR2pro::GUS* construct, expression was detected throughout cotyledons and hypocotyl at 2 and 3 DAS (Fig. 3-2). At 4 DAS, *DGR2pro::GUS* staining was absent from the cotyledons except in vascular tissues. Staining in the hypocotyl diminished basally, and was absent from the root except for a distinct cluster of cells in the root-shoot junction and the meristematic zone of the root apex. Stain was also detected in the region of the shoot apical meristem. By 5 DAS, staining of *DGR2pro::GUS* constructs disappeared almost entirely from the cotyledons and hypocotyls, and was detected only in roots, especially in the root apex, and at the root-shoot junction. Dark grown seedlings showed the same expression pattern as seedlings grown in the light (data not shown).

In *DUFB* $_{pro}$::*GUS* lines, at 2 DAS, reporter gene expression was detected only at the tips of cotyledons (Fig. 3-3). This pattern had changed abruptly by 3 DAS, at which point *DUFB* $_{pro}$::*GUS* lines stained most intensely in roots (but not in the root apex). This pattern persisted in 4 DAS and 5 DAS seedlings, in which the root was the most intensely stained tissue, except for the root apex. Some staining was also detected in the region of the shoot apex and at the tips of cotyledons at either 4 or 5 DAS in *DUFB* $_{pro}$::*GUS* lines (Fig. 3-3).

At subsequent stages of root growth, *DGR2* $_{pro}$::*GUS* continued to be expressed predominantly in the root apex, and was also detected in a distinct pattern in lateral root primodira (LRP) (Fig. 3-7). *DUFB* expressed in the roots excluding root apex and was not expressed in developing lateral root primordia (Fig. 3-7). These staining patterns were also observed in mature plants (Fig. 3-7).

3.4.3. Reporter gene expression in flowers

We next examined GUS expression in flowers. No expression of *DGR2pro::GUS* was detected in flowers. In *DUFB* $_{pro}$::*GUS* lines, GUS was detected in the abscission zone of petals, siliques and flowers (Fig. 3-4) and in developing anthers, but not in mature anthers (Fig. 3-4).

3.4.5. Reporter gene expression in response to exogenous hormone

When 4 DAS *DGR2pro::GUS* and *DUFB* $_{pro}$::*GUS* plants were incubated for 16 h in 10 μM of IAA, ACC, GA_3 or MeJA, increased staining in the cotyledons, hypocotyls, and roots compared to mock-treated plants was detected only in the IAA treated *DGR2pro::GUS* plants (data not shown). Next, both the transgenic lines were treated with 100 μM concentrations of IAA, ACC, GA_3, or MeJA. No staining differences were detected for GA_3, or MeJA in either transgenic line compared to mock-treated plants. On the other hand, in *DGR2pro::GUS* plants treated with 100 μM of IAA, increased staining again was detected in the cotyledons, hypocotyls and in the roots compared to mock treated plants. No staining differences were detected when *DGR2pro::GUS* plants were treated with 10μM of ACC but when treated with 100 μM ACC, staining was detected in the cotyledons and hypocotyls but root staining showed no difference in comparison with mock treated plant roots (Fig.3-5). In *DUFB* $_{pro}$::*GUS* plants treated with 100 μM IAA, a higher level of staining was detected in the hypocotyls and cotyledons and in the roots especially in the transition zone and elongation zone of the root apex compared to mock treated plants and when treated with 100 μM ACC, a very intense staining was detected in the elongation zone of root apex (Fig. 3-6).

3.4.6. Reporter gene expression in response to monosaccharides

We tested 4 DAS *DGR2pro::GUS* and *DUFB* $_{pro}$::*GUS* plants to test whether these were responsive to the monosaccharides: D-Glucose (Glc), D-Galactose (Gal), D-Fructose (Fru), and D-Arabinose (Ara). In *DGR2pro::GUS* plants, intense staining was detected in the hypocotyls and in the roots when treated with Gal and Ara, respectively (Fig. 3-5). No changes in the staining patterns were observed when *DGR2pro::GUS* plants that were treated with Glc (Fig 3-5) or Fru (data not shown). When *DUFB* $_{pro}$::*GUS* plants were treated with Gal, intense staining was detected in the roots, hypocotyls and in cotyledon veins and when treated with Glc and Ara, intense staining was detected in the roots compared to mock treated plants (Fig 3-6). *DUFB* $_{pro}$::*GUS* plants were also treated with L-GalL but no staining differences were observed between treated and non-treated plants (results not shown).

3.4.7. Quantitative RT-PCR analysis of seedling expression

Using Col-0 wild-type plants, we measured the transcript abundance of *DGR2* and *DUFB*. Our objective was to confirm that the patterns observed with the reporter gene fusions were representative of the native expression patterns of these genes. We measured transcript abundance in the root apex and in roots from which the apex had been removed. Plants were 5 DAS at the time of dissection and root apices were detached at a position approximately 0.4 mm from the root tip. Results showed that *DGR2* transcript abundance was significantly higher (2.5-fold, ($p < 0.001$) in the root apex compared to the rest of the root. This observation was consistent with the results of GUS staining. Conversely, no significant difference was seen in *DUFB* expression in the root apex compared to the rest of the root (Figure 3-8A). In the same seedlings, the transcripts of *DGR2* and *DUFB* were almost undetectable in the hypocotyls of 5 DAS grown in the dark (results not shown).

We extended our qRT-PCR analysis of transcripts in the root tips to 10 DAS seedlings, to see whether the observed patterns persisted later into development. The root apices were collected as described above, and additionally lateral roots were removed from the main root. As was observed at 5 DAS, a significantly ($p=<0.001$) higher level of *DGR2* transcript abundance was

measured in the root apex compared to the rest of the roots, and the abundance of *DUFB* again did not differ between the root apex and the main root (Figure 3-8B).

3.4.8. Hormone and sugar responses of *DGR2* and *DUFB* transcripts

We also used qRT-PCR to measure changes in transcript abundance induced by hormones in *DGR2* and *DUFB*. 4 DAS Col-0 plants were treated separately with 10 µM of indole acetic acid (IAA), the ethylene precursor (ACC), gibberellin (GA_3), and methyl jasmonate (MeJA) for 16 hours. *DGR2* transcript abundance significantly (P=<0.001) increased following treatment with exogenous IAA (Fig. 3-9). Ethylene likewise increased *DGR2* transcript abundance, but not as much as IAA. GA_3 did not change *DGR2* transcript abundance whereas MeJA significantly decreased *DGR2* transcript abundance (Fig. 3-9). On the other hand, *DUFB* transcripts decreased but not significantly following treatment with ACC and IAA but decreased significantly when treated with GA_3 and MeJA (Fig. 3-9). When *DGR2* $_{pro}$::*GUS* and *DUFB* $_{pro}$::*GUS* lines were treated with a higher concentration of ACC and IAA (100 µM), their transcript abundance increased significantly(P=<0.001) (Fig.3-9). This result was consistent with the GUS staining results of *DGR2* $_{pro}$::*GUS* and *DUFB* $_{pro}$::*GUS* IAA and ACC treated seedlings (Fig. 3-5 and 3-6).

The results of qRT-PCR for 16h sugar treated 4 DAS wild type plants showed that *DGR2* $_{pro}$::*GUS* transcripts significantly (P=<0.001) increased by D-Gal and D-Ara which were 2.03X and 1.62X, respectively. No significant changes in transcript abundance were detected for D-Glc and D-Fru treated plants (Fig. 3-10). On the other hand transcripts of *DUFB* $_{pro}$::*GUS* were upregulated significantly (P=0.05) with D-Glc and D-Gal (Fig. 3-10) which were 0.9X and 1.1X, respectively.

When 4 DAS Col-0 plants were treated with 10mM L-GalL, transcripts of *DGR2* were significantly upregulated (Fig 3-11) and transcripts of *DUFB* were down regulated but not significantly.

3.5. Discussion

3.5.1. Spatial expression of *DGR2* and *DUFB*

Expression of DUF642 domain-containing genes has been associated with fiber and cell wall development (DePauw et al., 2007; Bayer et al., 2006). Two DUF642-containing genes, *DGR2* and *DUFB*, are presumptive paralogs that have high sequence similarity to each other. Despite these similarities, a previous study showed that *DGR2* but not *DUFB* was induced by L-GalL (Gao et al., 2011). We were therefore motivated to further characterize the expression patterns of these genes to gain insight into the range of functions of DUF642-containing genes in plants.

The histochemical analysis of roots of transgenic plants bearing *DGR2* $_{pro}$::*GUS* or *DUFB* $_{pro}$::*GUS* indicated that *DGR2* was expressed most strongly in the root apex and developing lateral root primordia whereas *DUFB* expression in roots was excluded from the root apex and from developing lateral roots, even when it was expressed strongly in adjacent parts of the root (Fig. 3-2, 3-3 and 3-7). This was also consistent with qRT-PCR analysis of these genes in dissected roots of wild-type plants (Fig. 3-8). This contradicts the report of Gao et al (2012), based on reporter gene fusions, that *DGR2* was expressed throughout roots but was excluded from the root apex. We note that Gao et al. (2012) used an upstream region for their reporter fusion that was 237 bp longer than the one used in the present study. Further, unlike what we have reported here, Gao et al. (2012) did not confirm their results by qRT-PCR. In our results, *DGR2* and *DUFB* appeared to be differentially expressed in the meristematic zone and the transition zones, respectively. Lateral root primordia develop in eight stages through a series of cell division and expansion to create a structure similar to primary root apex (Malamy and Benfey, 1997). Stage I contains a pair of short pericycle cells lying end to end and flanked by two longer cells. Stage II forms two cell layers, an inner layer and outer layer and stage III creates a three layer lateral root primordium. In the present study, GUS staining was not detected in these stages but was detected in the later stages and in the tip of mature lateral roots of *DGR2* $_{pro}$::*GUS* transgenic plants (Fig 3-7). At the later stages, cells are involved in divisions and the cells located near the axis elongated towards the tip of the root due to the elongation of

the primordium (Szymanowska-Pułka et al., 2012). Therefore, the GUS staining patterns suggest that *DGR2* is expressed in root cells that are dividing or elongating. In the transition zone, cells leaving the apical region are proposed to be in a transitional stage of cyto-architectural re-arrangement in order to perform rapid cell elongation (Baluška , 1990,1996, 1997). IAA treatment of *DUFB* strongly increased GUS staining intensity in the transition and elongation zone of the root apex, therefore, *DUFB* expression may be correlated with cell elongation by the regulation of IAA. However, in dark grown hypocotyls, which also elongate rapidly, no expression of *DGR2* or *DUFB* transcripts was detected by qRT-PCR (data not shown).

3.5.2. Hormonal regulation of *DGR2* and *DUFB*

Reporter gene fusions and qRT-PCR assays provided evidence that transcripts of both *DGR2* and *DUFB* increased following treatment with IAA (Fig. 3-5, 3-6 and 3-9). According to qRT-PCR analysis, transcripts of *DGR2* were more sensitive than *DUFB* to IAA; *DGR2* showed a response at 10 μM IAA, whereas *DUFB* transcripts were responsive only to 100 μM IAA (Fig. 3-9). Furthermore, an increase in GUS staining intensity was detected in *DGR2* $_{pro}$::*GUS* plants when treated with 10 μM, but when *DUFB* $_{pro}$::*GUS* plants were treated with 10 μM IAA, no staining difference was detected (data not shown) until the concentration increased to 100 μM IAA. These hormones increased GUS staining intensity in transgenic reporter lines, but in no case did the hormone treatment result in ectopic GUS staining in the root tips of *DUFB* $_{pro}$::*GUS* lines (Fig. 3-6). Auxin is an important phytohormone involved in different developmental processes of plants including root and vascular development and cell division and elongation (Overvoorde et al., 2010; Perrot-Rechenmann, 2010; Berleth, 2000). The developmental expression patterns of *DGR2* $_{pro}$::*GUS* plants showed that *DGR2* was highly expressed in the root apex and lateral root primordia and was upregulated by IAA, especially in the root apex and all hypocotyl cells and cotyledons (Fig. 3-2, 3-5 and 3-7). On the other hand, *DUFB* $_{pro}$::*GUS* plants showed that *DUFB* was not expressed in the root apex and developing lateral roots primordia and upregulated by IAA treatments in the elongation zone of root apex and vascular tissue of hypocotyls and cotyledon veins (Fig. 3-3, 3-6 and 3-7). When treated with ACC, *DGR2* $_{pro}$::*GUS* plants showed GUS staining in the cotyledons and hypocotyls but not in the root apex whereas in ACC treated *DUFB* $_{pro}$::*GUS* plants, increased GUS staining was detected in the elongation zone of root apex

only (Fig. 3-5 and 3-6). Hormonal cross talk between auxin and ethylene determines the developmental cell fate and root growth (Benková and Hejátko, 2009). Muday et al. (2012) reported that auxin dependent seedling growth influenced by ethylene. Ethylene modulates auxin synthesis, transport and signaling with unique targets and responses in a range of tissues of seedlings for growth and development (Muday et al., 2012). It is therefore reasonable to conclude that *DGR2* and *DUFB* are involved in auxin and ethylene dependent root and seedling development but target different tissues. On the other hand, expression of both genes was down-regulated by the application of MeJA. This could be due to the impairment of ethylene and auxin biosynthesis by methyl jasmonate (Soto et al., 2012) and because it was evident in the present study that *DGR2* and *DUFB* expression were dependent on auxin and ethylene, so when we applied 10 μM MeJA, the expression of *DGR2* and *DUFB* were significantly ($P<0.001$) downregulated (Fig. 3-9). Application of 10 μM GA_3 was significantly ($P<0.001$) downregulated the transcript expressions of *DUFB* but not *DGR2* (Fig. 3-9). Gibberellic acid has been shown to affect abscission, mainly by promoting abscission (Medeghini-Bonatti et al., 1976) and in the present study, we have identified that DUFB involved in abscission but not DGR2.

GUS staining results revealed that *DUFB* was involved in anther development (Fig. 3-4). GUS staining of *DUFB* $_{pro}$::*GUS* was also detected in the abscission zones (AZs) of flower and siliques (Fig. 3-4). The AZ develops at the junction between the leaving organ and main plant body (Bleecker and Patterson, 1997). It was reported that ethylene promotes abscission, whereas auxin inhibits this process (Roberts et al., 2002) so the IAA and ACC responsive genes are involved in abscission so does *DUFB*. Microarray analysis with stamen abscission zones of Arabidopsis showed that many hydrolytic encoding genes and cell wall modifying enzymes upregulated prior to abscission and *DUFB* has predicted hydrolase like fold. Therefore, the results from GUS staining of the present experiment suggested that *DUFB* has a role in abscission.

3.5.3. Regulation of *DGR2* and *DUFB* by metabolites

Recognizing that L-GalL is the terminal precursor for ascorbic acid (AsA) biosynthesis in *Arabidopsis thaliana*, Gao et al. (2012) tried to determine in their study whether DGR2 and DGR1 were involved in AsA biosynthesis, but could not identify any significant difference in the AsA levels between the *dgr* mutants and the controls. They then predicted that responsiveness to L-GalL of *DGR2 c*ould be due to changing sugar concentrations. We expected that *DGR2* and *DUFB* might respond to sugars as *in silico* analysis predicted that DGR2 and DUFB contain galactose bindng-like folds and hydrolase-like folds and were identified in previous studies in the cell wall extracts. Therefore, it is possible that DUF642 domain containing genes are involved in cell wall remodeling by hydrolyzing cell wall matrix polysaccharides. The biosynthesis of plant cell wall requires nucleotide sugar interconversion enzymes, nucleotide sugar transporters, and glycosyltransferases (Reiter and Vanzin, 2001). In the present study, *DGR2* $_{pro}$::*GUS* and *DUFB* $_{pro}$::*GUS* plants, as well as Col-0 plants, were treated with Glc, Gal, Fru, and Ara for GUS staining and qRT-PCR. The qRT-PCR results showed that *DGR2* transcripts significantly ($p=<0.001$) increased in response to Gal and Ara whereas *DUFB* transcripts increased significantly ($p=<0.05$) in response to Glc and Gal. These observations were consistent with GUS staining results (Fig 3-5, 3-6, 3-10). In was reported that DUF642 protein family members interact with cell wall polysaccharides, and that two of the DUF642 encoded proteins, DUFB and At4g32460, interact with the AtPME3 catalytic domain *in vitro* (Vázquez-Lobo, et al., 2012 and Zúñiga-Sánchez and Gamboa-de Buen, 2012). Pectin consists of two major types of polysaccharides: homogalacturonan, and rhamnogalacturonan I. Rhamnogalacturonan I is a major component of pectin with a backbone of alternating rhamnose and galacturonic acid residues and side chains that include alpha-1,5-arabinans, beta-1,4-galactans, and arabinogalactans. Many enzymes are required to synthesize pectin but all not have been identified. Therefore, it is possible that DGR2 and DUFB are involved in the modification of pectin side chains during plant growth and development.

3.6. Conclusions

Despite their high sequence similarity, *DGR2* and *DUFB* were found to have distinct expression patterns that indicated these genes have non-redundant roles in plant development. *DGR2* was

expressed in the root tip, while *DUFB* was excluded from this region. Transcript expression of both genes increased following treatment with IAA, ACC, or Gal, although *DGR2* was more sensitive than *DUFB* to IAA. Furthermore, *DGR2* was induced by Ara whereas *DUFB* was induced by Glc.

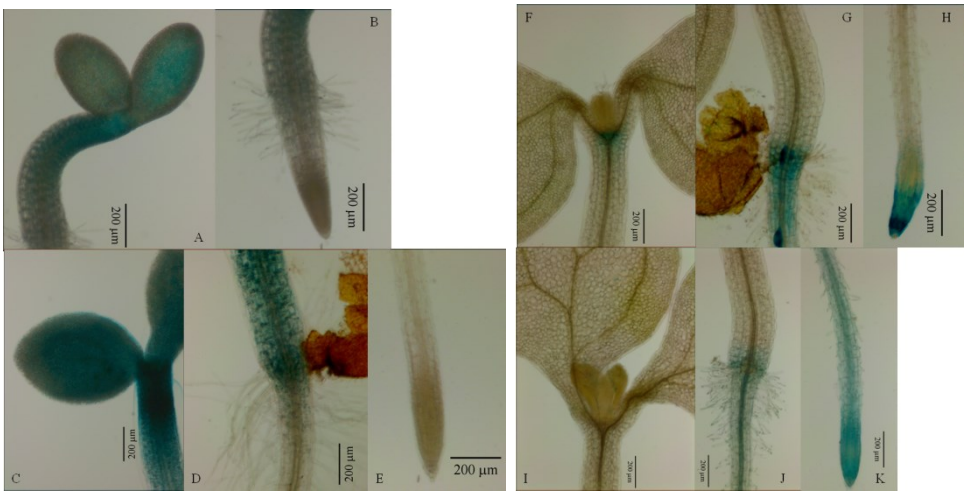


Figure 3-2. *DGR2*pro::GUS seedling expression. In *Arabidopsis* transgenic plants carrying the *Arabidopsis DGR2* promoter-*GUS* fusion gene, GUS staining was shown in the following developmental stages: 2 DAS (A&B), 3 DAS (C-E), 4 DAS (F-H) and 5 DAS (I-L) was demonstrated by histochemical staining.

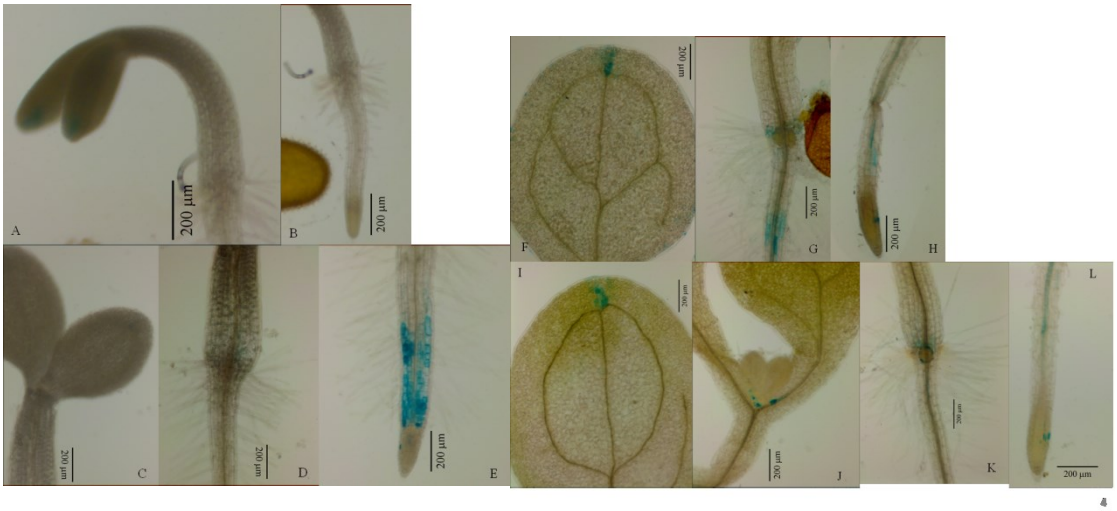


Figure 3-3. *DUFB*pro::*GUS* seedling expression. In Arabidopsis transgenic plants carrying the Arabidopsis DGR2 promoter-GUS fusion gene, GUS staining was shown in the following developmental stages: 2 DAS (A&B), 3 DAS (C-E), 4 DAS (F-H) and 5 DAS (I-L) demonstrated by histochemical staining.

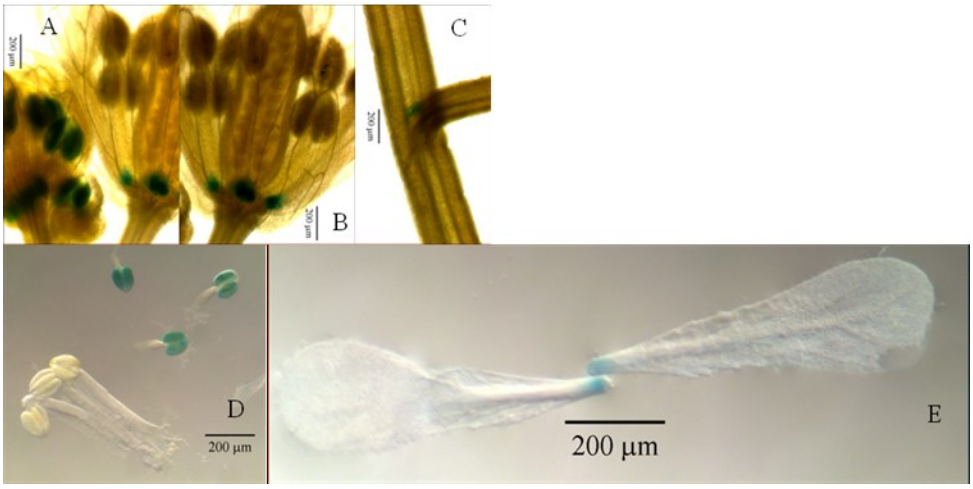


Figure 3-4. *DUFB*pro::*GUS* expressions. In *Arabidopsis* transgenic plants carrying the *Arabidopsis DUFB* promoter-*GUS* fusion gene, GUS staining was shown in the following tissues: flowers, anthers and abscission zone demonstrated by histochemical staining.

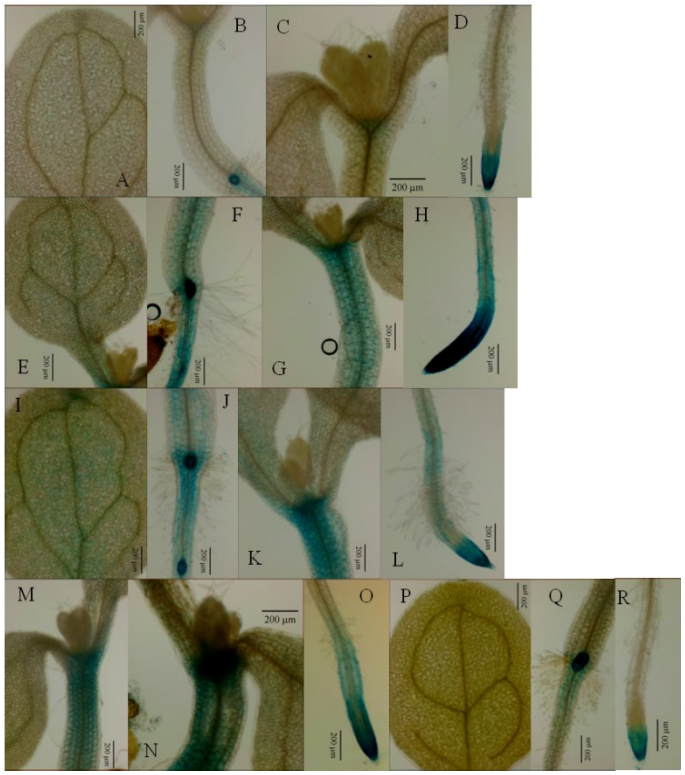


Figure 3-5. *DGR2*pro::GUS expression induced by100 µM auxin (IAA), 100 µM ethylene (ACC), 5% Gal , 5% Ara and 5% Glc. A-D: Mock; E-H: IAA; I-L: ACC; M: Gal; N & O: Ara;.P-R: Glc.

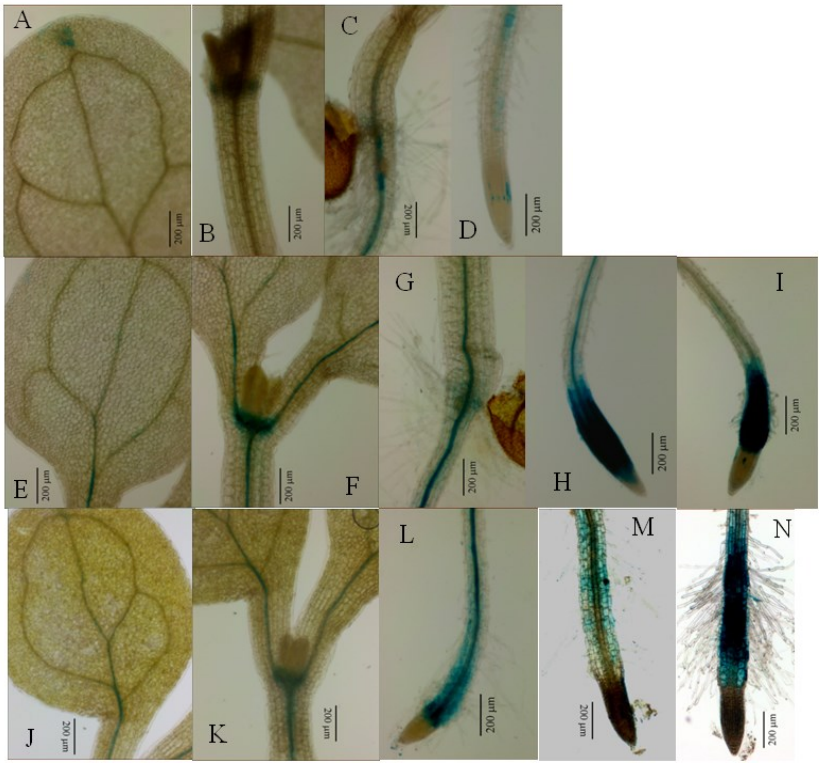


Figure 3-6. *DUFB*pro::GUS expression induced 100 µM auxin (IAA), 100 µM ethylene (ACC), 5% Gal, 5% Ara and 5% Glc. A-D*:* Mock; E-H: IAA; I: ACC; J-L: Gal; M: Ara;.N: Glc.

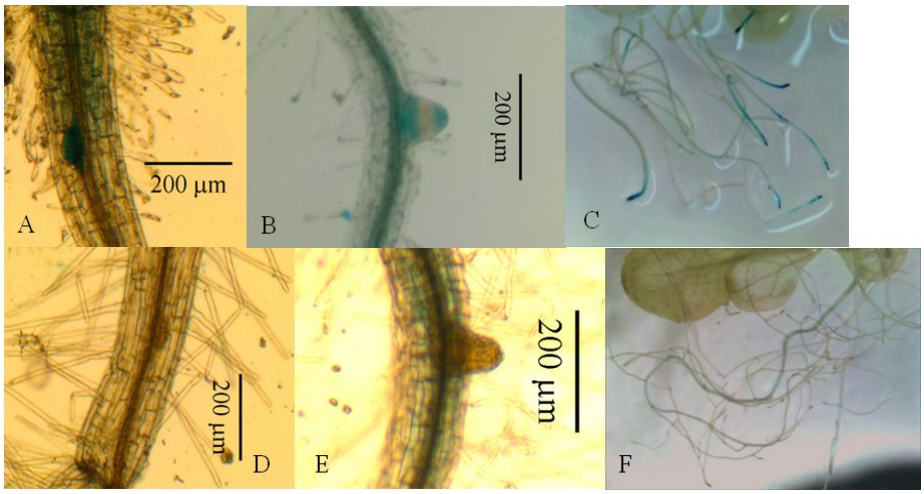


Figure 3-7. GUS expression in developing lateral root primordia and in two weeks old *DGR2pro::GUS* and *DUFB* $_{pro}$::*GUS* transgenic plants. A-C: 4 DAS, 5 DAS and 14 DAS *DGR2*; D-F: A-C: 4 DAS, 5 DAS and 14 DAS *DUFB*.

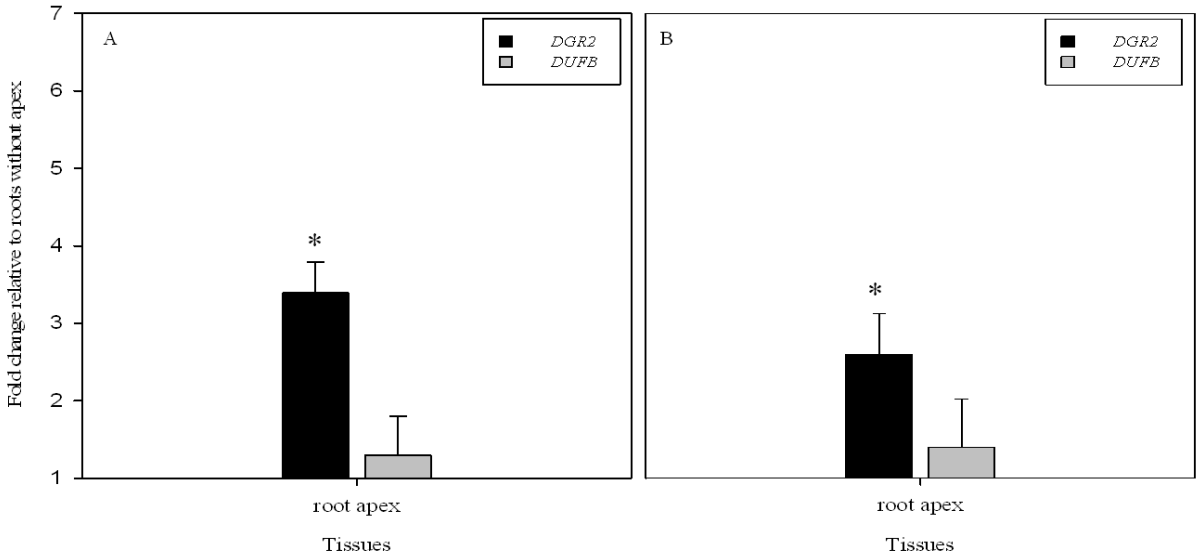


Figure 3-8. The relative expression of *DGR2* and *DUFB* in the root apex of 5 DAS (A) and 10 DAS (B) old Col-0 plants compared to roots without apex the value of which is 1. Asterisks indicate values significantly different from roots without apex (P=<0.001). Error bars on each column indicate standard deviation from three biological replicates.

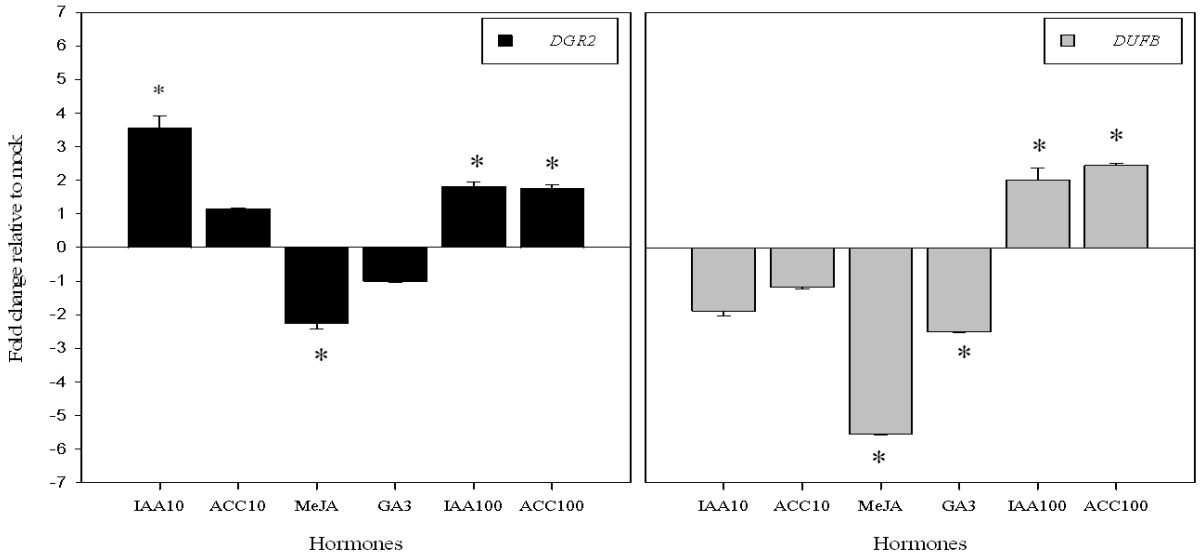


Figure 3-9. The relative expressions of *DGR2* and *DUFB* in response to the hormones compared to mock treated plants the value of which is 1. IAA10 and IAA100 are 10µM and 100 µM IAA concentrations; ACC10 and ACC100 are 10µM and 100 µM ACC concentrations. Asterisks indicate values significantly different from mock treated seedlings (P=<0.001). Error bars on each column indicate standard deviation from three biological replicates.

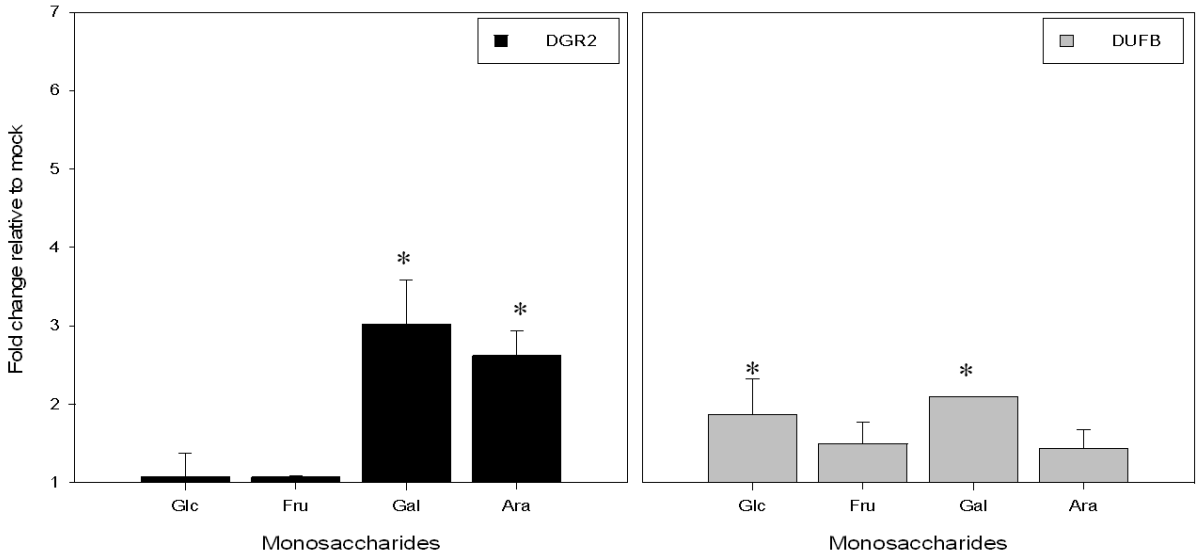


Figure 3-10. The relative expressions of *DGR2* and *DUFB* in response to the monosaccharides compared to mock treated plants the value of which is 1. Asterisks indicate values significantly different from mock treated seedlings (P=<0.001 for *DGR2* and P=<0.05 for *DUFB*). Error bars on each column indicate standard deviation from three biological replicates.

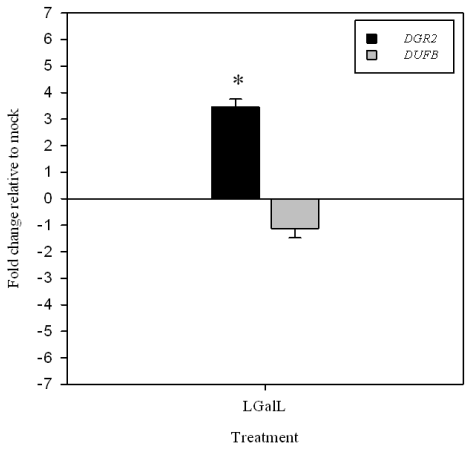


Figure 3-11. The relative expressions of *DGR2* and *DUFB* in response to L-GalL compared to mock treated plants the value of which is 1. Asterisks indicate values significantly different from mock treated seedlings (P=<0.001). Error bars on each column indicate standard deviation from three biological replicates.

4. Chapter 4

4.1. Towards systematic functional characterization of two DUF642 genes in Arabidopsis

4.2. Introduction

Studies of genes with unknown function require experimental methods to more reliably infer gene function. These methods include analysis of: transcript and protein expression patterns; protein-protein interactions; subcellular localization; functional assays with heterologously expressed proteins; and loss-of-function and overexpression mutant phenotypes.

RNAi, also termed as post-transcriptional gene silencing (PTGS), is one method to generate a loss-of-function phenotype. In this method, gene silencing is triggered when a double-stranded RNA (dsRNA) homologous to the sequence of the target mRNA is recognized by the plant cell. A protein complex called DICER encounters dsRNA and cuts it into pieces called small-interfering RNAs or siRNAs. RISC, a multiprotein complex, incorporates one strand of a small interfering RNA and guides gene silencing in a sequence-specific manner. Several vectors for RNAi are available to generate transgenic plants using *Agrobacterium tumefaciens* mediated delivery into plants. Among them, pHELLSGATE is a vector that facilitates the cloning of genes using an *in vitro* recombinase system (Wesley et al., 2013). RNAi is not always specific for a single gene, and may silence one or more genes that share highly conserved regions of sequence.

Another technique to obtain loss-of-function mutants is through insertional mutagenesis. Transposons (Sundaresan et al. 1995; Martienssen, 1998) or the T-DNA of *A. tumefacians* (Azpiroz-Leehan and Feldmann, 1997) can be used as mutagens because of their ability to insert randomly within chromosomes. For example, researchers at the Salk Institute generated indexed T-DNA insertional mutants of *Arabidopsis thaliana* with the goal of providing an insertion mutant for every identified gene in the *A. thaliana* genome (O'Malley and Ecker, 2010). Arabidopsis contains over 25,500 genes, and to date 50,090 Salk lines representing 24,535

individual genes have been made available publicly through the Arabidopsis Biological Resource Center (ABRC).

Heterologous expression systems have the potential to provide direct evidence for the biochemical function of a gene. The challenge comes in terms of synthesis of an active protein. It is especially true when it comes to an unknown function protein because there will be no established functional assay for unknown proteins. Although there are many eukaryotic heterologous expression systems available including insect, yeast, and various mammalian cells, expression in *Escherichia coli* is popular because it is fast and simple. The disadvantage of the *E. coli* system is that it lacks the eukaryotic post-translational machinery. In many cases, post translational modifications are necessary to produce a soluble and active eukaryotic protein (Farrokhi et al., 2013).

To infer the cellular function of a protein, it is useful to know its subcellular localization. Several bioinformatics tools are available for prediction of protein localization based on various protein characteristics, but these predictions must be validated experimentally. Many organelle-based proteomics studies (Gilchrist et al., 2006; Harriet et al., 2012; Albenne et al., 2013) have provided a detailed list of proteins identified in various subcellular compartments. Nevertheless, these lists are not always accurate, due to contamination that may occur during organelle isolation (Andersen et al., 2006; Gatto et al., 2010). Experimental methods for protein localization include tagging of proteins using reporter genes such as green fluorescent protein (Hanson and Köhler, 2001) and immunolocalization (Burns et al., 1994). Complementary methods must be used to confirm that a fluorescently tagged protein behaves similarly to the endogenous protein.

The present study was designed to reveal the function of two DUF642 genes, At5g25460 (*DGR2*) and At5g11420 (*DUFB*) in *A. thaliana* using the following experimental techniques: (1) analysis of tissue specific transcript expression patterns by qRT-PCR, (2) analysis of tissue

specific protein expression patterns using anti-DUF642 antibody, (3) protein-protein interactions by co-IP (4) subcellular cellular localization of DGR2 by CiFP tagged protein and co-localization study by immunohistochemistry using Sec21 specific antibody and fixed tissues from 35S::CiFP plants, (5) heterologous expression and purification of DGR2 and DUFB in *E. coli*. and (4) generation of loss-of-function and overexpression mutants and their developmental and biochemical phenotyping.

4.3. Materials and Methods

4.3.1. Plant growth conditions

4.3.1.1. Conditions in MS-agar medium

Transgenic and WT Arabidopsis seeds were surface-sterilized by 50% (v/v) commercial bleach (Javex, Clorox), cold incubated at 4°C for two days in the dark, and then sown in petri-dishes (100 x15 mm) containing ½ x MS medium (pH 5.7 by KOH) and 0.7% Phytablend (MS-agar), with 1% sucrose.

4.3.1.2. Conditions in the soil

Arabidopsis seeds were sown in pots (140 x 150 mm) containing wet potting soil (Sunshine Mix, Sungro). The pots were incubated in the dark at 4°C for two days, and then transferred to a growth chamber. Wild type and transgenic Arabidopsis plants (ecotype Col-0) were grown with day/night temperatures of 22°/19°C and a light/dark cycle of 16h/8h with approximate photosynthetic photon flux density of 160 µmol m^{-2} s^{-1}, following Doyle and Doyle (1987) with modifications.

4.3.2. RNA isolation, quality control and cDNA synthesis

Tissues from different organs that included green siliques, roots, flowers, fully expanded rosette leaves from 3-4 weeks old plants and 5 days old seedlings from Arabidopsis Col-0 grown under

conditions described in Section 4.1.1. were collected for RNA extraction. Tissues were collected in 2 ml microcentrifuge tubes and immediately frozen in liquid nitrogen. The frozen tissues were disrupted without thawing, using a plastic pestle and a microcentrifuge tube. Total RNA was extracted from the disrupted tissues using the RNeasy Plant Mini Kit (Qiagen). DNA was removed from the samples using DNA-free RNA kit (Ambion). To confirm that DNA was removed, samples were subjected to agarose gel electrophoresis using DNA-specific primers. The RNA concentration was measured on the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). RNA quality was assessed using the Agilent-2100 Bioanalyzer and RNA 6000 NanoChips (Agilent Technologies). Next, first strand cDNA synthesis was performed using 1.0 µg of total RNA from each sample, with either SuperScript III FirstStrand Synthesis System for RT-PCR (Invitrogen, USA) or RTAid H-MMLV (Fermentas), and Oligo(dT)12-18 127, or oligo(dT)18 (Fermentas) as the primer in a 20 µL reaction volume following the manufacturer's protocol.

4.3.3. Extraction of genomic DNA

4 to 6 leaves were flash frozen in liquid nitrogen and ground with a plastic pestle and a microfuge tube. Genomic DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN), according to the manufacturer's protocol. DNA concentration was determined by a NanoDrop spectrophotometer (NanoDrop ND 1000 V3 7.1).

4.3.4. Plasmid construction

4.3.4.1. Construction of RNAi plasmid

For the construction of RNAi vector constructs, standard molecular biology cloning techniques (Sambrook et al., 1989) and Gateway Recombination Technology (Invitrogen; Helliwell and Waterhouse 2003) were used. As described below, PCR primers, DUF_L124 and DUF_R399 were designed to amplify a 276bp fragment (Fig. 4-1) of a *DGR2* gene (At5g25460) from Arabidopsis genomic DNA. Each primer had two regions: one region was complementary to the *DGR2* gene (underlined in the primers) and the other part of each primer was an attB site to

allow the PCR product to be incorporated into Gateway® -compatible vector, pHellsgate using the BP clonase recombination reaction. This amplified 276bp fragment region shared 100% similarity with *DGR2* and 80% similarity with *DUFB* (Fig. 4-2) due to high sequence similarity between these two genes.

DUF_L124
*GGGGACAAGTTTGTACAAAAAAGCAGGCT*CCATCAGACATGAAAGGAACAG

The italicized region of the primer was the attB1 site. The underlined region anneals to *DGR2* (At5g25460) genomic DNA at nucleotide 124 of the CDS.

DUF_R399 *GGGGACCACTTTGTACAAGAAAGCTGGGT*ACCGGAGTCAGGTGCTACAG

The italicized region of the primer was the attB2 site. The underlined region anneals to DUF642 (At5g25460) genomic DNA at nucleotide 399 of the CDS.

The total length of the PCR product (including primers) which was approximately 340bp was amplified in a standard PCR from gDNA of Arabidopsis Col-0 plant using DUF_L124 and DUF_R399 primer pairs mentioned below. The amplified PCR products were then subjected to gel electrophoresis, and a band of expected size band was excised from the gel and purified by Wizard® SV Gel and PCR Clean-Up System kit (Promega, Cat. No. A9282). This fragment was cloned into a donor vector (pDONR222) following the BP cloning protocol (Invitrogen) to generate an entry vector, which was then used to transform *E. coli* (DH5α) chemically competent cells by the freeze/thaw method. Positive clones were selected in the presence of 50 μg/ml kanamycin. The positive entry-vector plasmid was subjected to sequencing using a vector-specific forward primer (5'-GTAAAACGACGGCCAGT)-3'). The manufacturer's BigDye terminator cycle sequencing reagent, and an AB13730 sequencer (Applied Biosystem) were used for sequencing experiments. Next, the cloned gene was subcloned into the Destination Vector, pHellsgate12 (constitutive) by the LR reaction between the Entry clone and the Destination Vector. The resulting construct was then transformed into *E. coli* (DH5α) competent cells. Positive transformants were selected on LB agar with 100 μg/ml spectinomycin. Clones were PCR screened for inserts with primers to confirm the presence of the gene insert, and then subjected to restriction digestion using Xba1 or Xho1/Kpn1 to confirm that the insert was properly positioned in the construct.

```
>At5g25460
ATGGAAGGCGTCACCGTCGTGTCTTTCTTCCTTCTTTTCATCGCCACCGCCATGGCCGCCAAGT
CCACCGTCTCCTTCCGTGACGGCATGTTACCAAACGGAGACTTCGAGCTAGGACCAAAACCATC
AGACATGAAAGGAACAGAAATACTAAACAAACTAGCAATACCAAACTGGGAAGTCACAGGATTC
GTCGAATACATTAAATCAGGACATAAACAAGGAGACATGCTTCTCGTTGTTCCCGCCGGTAAAT
TCGCTGTAAGACTTGGGAACGAAGCATCGATCAAACAAAGACTTAAAGTGGTTAAAGGAATGTA
TTACTCACTCACTTTTAGTGCTGCTAGAACTTGTGCACAAGACGAGAGACTTAACATATCTGTA
GCACCTGACTCCGGTGTGATTCCGATTCAGACGGTTTATAGTAGTAGTGGTTGGGATTTATATG
CTTGGGCGTTTCAAGCTGAGAGTGATGTTGCTGAAGTTGTGATTCATAATCCTGGTGTTGAGGA
AGATCCAGCTTGTGGTCCACTTATTGATGGTGTTGCTATGAGATCTCTTTACCCTCCTAGACCA
ACTAATAAGAACATTTTGAAAAACGGAGGATTTGAAGAAGGTCCATTAGTATTACCCGGCTCGA
CAACTGGAGTTTTGATCCCACCGTTTATAGAAGACGACCACTCTCCTTTACCTGGATGGATGGT
GGAGTCTCTCAAAGCTGTCAAGTACGTAGACGTTGAACATTTCTCAGTCCCACAGGGTCGCAGA
GCTATTGAGCTTGTAGCGGGTAAAGAGAGTGCCATCGCTCAAGTGGTTCGGACTGTCATTGGGA
AGACTTACGTGCTGTCTTTTGCGGTTGGAGACGCCAACAATGCTTGCAAAGGATCAATGGTGGT
TGAGGCTTTTGCAGGAAAAGATACACTTAAGGTCCCTTACGAGTCGAAAGGCACAGGAGGGTTT
AAACGAGCTTCTATTCGATTTGTGGCGGTTTCGACCCGATCAAGAATTATGTTCTACAGCACTT
TCTATGCCATGAGGAGCGATGATTTCTCGTCATTGTGTGGGCCTGTGATCGATGATGTCAAGCT
TATAAGCGTTCGTAAACCATAG
```

Figure 4-1. *DGR2* CDS fragment used for RNAi was shown as underlined region.

```
CLUSTAL 2.1 multiple sequence alignment


DUFB            ATGAAAGGAGGCAGCCTCTCGTTTCTCTTCGTTCTCCTAATCGCCACCATCACTTCCGTC
60
EMBOSS_001      ------------------------------------------------------------


DUFB            ATTTGCTTCAGTGACGGGATGTTACCAAACGGCGACTTTGAACTAGGACCAAAACCATCG
120
EMBOSS_001      -----------------------------------------------------CCATCA
6
                                                                     *****

DUFB            GACATGAAAGGAACGCAAGTAATAAACAAGAAGGCGATTCCTAGCTGGGAGCTTTCAGGC
180
EMBOSS_001      GACATGAAAGGAACAGAAATACTAAACAAACTAGCAATACCAAACTGGGAAGTCACAGGA
66
                **************  ** ** *******    ** ** ** * ******  *  ****

DUFB            TTCGTCGAATACATAAAGTCCGGTCAAAAACAAGGAGACATGCTTCTCGTAGTCCCGGCC
240
EMBOSS_001      TTCGTCGAATACATTAAATCAGGACATAAACAAGGAGACATGCTTCTCGTTGTTCCCGCC
126
                ************** ** ** ** ** ********************** ** ** ***
```

```
DUFB            GGAAAGTTCGCAATCCGGCTAGGCAACGAGGCATCGATCAAACAAAGACTTAACGTTACA
300
EMBOSS_001      GGTAAATTCGCTGTAAGACTTGGGAACGAAGCATCGATCAAACAAAGACTTAAAGTGGTT
186
                ** ** *****  *  * ** ** ***** ********************** **

DUFB            AAAGGAATGTATTACTCACTGACGTTCAGTGCCGCAAGGACATGTGCCCAAGACGAACGG
360
EMBOSS_001      AAAGGAATGTATTACTCACTCACTTTTAGTGCTGCTAGAACTTGTGCACAAGACGAGAGA
246
                ******************* ** ** ***** ** ** ** ***** ********  *

DUFB            CTCAACATATCGGTGGCACCTGACTCAGGCGTTATTCCTATACAGACGGTGTACAGTAGC
420
EMBOSS_001      CTTAACATATCTGTAGCACCTGACTC--------------------CGGT----------
276
                ** ******** ** ***********                    ****

DUFB            AGTGGATGGGACCTTTACGCATGGGCGTTCCAAGCCGAGAGTAACGTGGCAGAGATCGTG
480
EMBOSS_001      ------------------------------------------------------------


DUFB            ATTCATAATCCTGGTGAGGAGGAAGATCCTGCTTGTGGACCACTCATTGATGGTGTGGCA
540
EMBOSS_001      ------------------------------------------------------------


DUFB            ATCAAAGCTCTATACCCTCCTCGGCCCACCAATAAGAATATATTGAAGAACGGAGGATTT
600
EMBOSS_001      ------------------------------------------------------------


DUFB            GAAGAAGGTCCCTACGTACTCCCAAACGCAACAACCGGCGTTCTGGTTCCTCCCTTTATA
660
EMBOSS_001      ------------------------------------------------------------


DUFB            GAAGATGACCACTCTCCTTTACCCGCGTGGATGGTCGAATCACTCAAAGCCATCAAATAC
720
EMBOSS_001      ------------------------------------------------------------


DUFB            GTTGATGTCGAGCATTTCTCGGTCCCACAAGGCCGTCGAGCCGTGGAGCTAGTGGCAGGC
780
EMBOSS_001      ------------------------------------------------------------


DUFB            AAAGAAAGCGCAATCGCTCAGGTAGCTAGGACCGTTGTGGGAAAAACTTACGTGCTTTCG
840
EMBOSS_001      ------------------------------------------------------------


DUFB            TTTGCGGTTGGAGATGCTAACAATGCTTGCCAAGGATCGATGGTGGTCGAGGCATTTGCG
900
EMBOSS_001      ------------------------------------------------------------
```

```
DUFB            GGAAAAGACACTCTAAAGGTACCTTATGAGTCTCGAGGCAAAGGAGGGTTCAAACGCGCT
960
EMBOSS_001      ------------------------------------------------------------

DUFB            TCTCTACGGTTTGTGGCGGTTTCGACCCGCACAAGAGTTATGTTTTACAGCACATTTTAC
1020
EMBOSS_001      ------------------------------------------------------------

DUFB            TCGATGAGAAGCGATGATTTCTCATCACTGTGTGGGCCCGTGATCGATGATGTTAAGCTC
1080
EMBOSS_001      ------------------------------------------------------------

DUFB            CTCAGTGCTCGTAAGCCGTAA 1101
EMBOSS_001      ---------------------
```

Figure 4-2. Multiple sequence alignment of synthetic *DUFB* and *DGR2* CDS (EMBOSS_001) amplicon used in RNAi vector construction using ClustalW multiple alignment tool (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

4.3.4.2. Construction of over-expression plasmids

4.3.4.2.1. Construction of 35S::*DGR2*

The 1.1-kb NcoI-BstEII full-length *DGR2* cDNA was cloned into the NcoI-BstEII site of binary vector pCAMBIA1303 to generate 35S::*DGR2* in which *DGR2* was expressed from the cauliflower mosaic virus 35S promoter.

DUF642 CDS was amplified from cDNA, synthesized from 5 day-old wild type Arabidopsis seedlings using a Platinum®Taq DNA Polymerase (Invitrogen, Cat No. 11304-011) with the following primer pairs:

DUF_L-22 TCTCGTTCTCTCTCTCTTCCGTCAA

DUF_R+1154 AAGACCACACCGATGCATTTTC

The product of PCR amplification (1.1 kb) was cloned into the pCRII-TOPO vector (Invitrogen) and transformed into chemically competent *E. coli* (DH5α) cells. The positive clones were identified on LB medium containing 50μg/ml ampicillin. 6-7 positive colonies were then grown overnight in LB liquid medium supplemented with 50μg/ml ampicillin. The bacterial pellets were collected by centrifugation at 140,000 rpm for 2 min and plasmid was isolated using

GenElute plasmid Miniprep Kit (Sigma) and sequenced to confirm that no error was introduced by PCR. *DGR2* could not be cloned directly into pCAMBIA1303 using NcoI and BstEII, because of the presence of internal NcoI site. Therefore, site-directed mutagenesis was performed to eliminate interfering restriction sites within the *DGR2* CDS by the following mutagenic primer pairs: Primers B and C contain the complementary sequence with a point mutation that eliminated the NcoI site from the *DGR2* CDS amplicon in the PCR. Two-step PCR reactions were performed as shown in Figure 4-3.

Primer B: 5'- TTG GCG GCC ACT GCG GTG GCG A -3'

Primer C: 5'- TCG CCA CCG CAG TGG CCG CCA A -3'

Primer A: *GGCCATGG*ATGGAAGGCGTCAC

Primer D: 5'- GGGGTTACCCTA TGG TTT ACG AAC GCT -3'

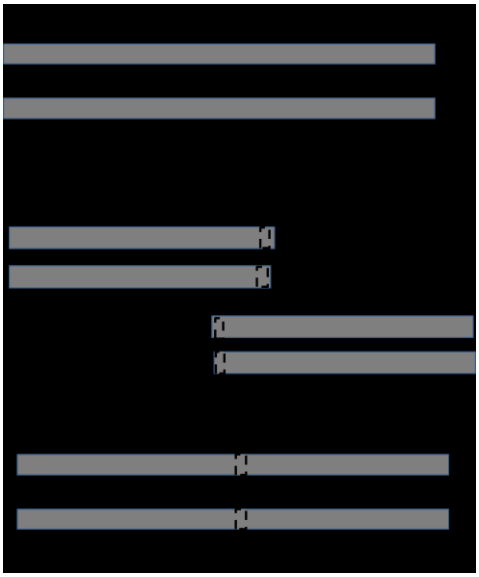


Figure 4-3. Site-directed mutagenesis by primer. First PCR reactions were performed with A & B and C & D primer pairs. Second step PCR reactions were performed with A & D primers.

Next, the amplicons from PCR1 and PCR2 were combined together and used as a template for PCR3 with the Primer A + Primer D. The amplicons were then cloned into the pCRII-TOPO vector following the Invitrogen protocol. The point mutation and open reading frame were verified by sequencing. The resultant plasmids and pCAMBIA1303 vector were digested with NcoI plus BstEII restriction enzymes for cloning the *DGR2* insert into the vector. The digested plasmids and the vectors were subjected to gel electrophoresis to obtain the *DGR2* fragment and linearized vector. Based on the expected size, the bands were excised from the gel, purified and ligated following the ligation reaction using T4 DNA ligase. The resultant construct was then transformed into *E. coli* (DH5α) cells. The positive transformants were selected in the presence of 50 µg/ml kanamycin. Plasmid was isolated from a pure culture of *E. coli* harboring 35S::*DGR2* plasmids and inserted into Agrobacterium and then in to Arabidopsisa to generate overexpression plants.

4.3.4.2.2. Construction of 35S::DUFB

The 1.1-kb NcoI_BamHI- full-length *DUFB* cDNA was cloned into the NcoI and BamHI sites of binary vector pCAMBIA1303 to generate 35S::*DUFB* in which *DUFB* was expressed from the cauliflower mosaic virus 35S promoter. The following primers were used to amplify the CDS with restriction sites to clone into the pCAMBIA1303 vector. The procedure was same as with the generation of the 35S::*DGR2* construct except that for *DUFB*, no site directed mutagenesis was required.

Forward Primer: 5'- CAT GCC ATG GAT GAA AGG AGG CAG CCT C -3'

Reverse primer: 5'- GGG GTA ACC TTA CGG CTT ACG AGC ACT -3'

4.3.4.3. Construction of translational fusion constructs

The coding sequence of *DGR2* and was cloned in the modified pCAMBIA1303 vector under the control of 35S CAMV promoter to generate 35S::DGR2:CiFP, 35S::DGR2:CiFP:KDEL and 35S::CiFP constructs.

Based on the coding sequence of *DGR2* obtained from TAIR database, a DNA fragment was synthesized by Genscript (www.genscript.com). This permitted removal of the native NcoI site from the coding sequence and addition of preferred restriction sites for the ease of cloning which were BspHI and NcoI (Fig 4-5A). The amino acid sequence of the synthesized coding sequence was the same as the amino acid sequence of the native protein (Fig. 4-5B). For cloning, the modified pCAMBIA1303 binary vector was used. The pCAMBIA1303 vector included GFP and GUS reporter genes under the control of the 35S promoter of cauliflower mosaic virus (CaMV). GFP has been reported to be strongly pH dependent in aqueous solutions and intracellular compartments in living cells (Kneen et al., 1998). Therefore, Citrine (CiFP) was introduced into the pCAMBIA1303 by removing the GFP and the GUS sequences. Citrine was described as superior to other YFPs and it is less sensitive to fluctuations in intracellular pH (Griesbeck et al., 2001). At first, the GUS and GFP peporter genes were cut out of pCAMBIA1303 as an NcoI BstEII fragment and then the vector fragment was ligated to the synthesized NcoI BstEII fragment of CiFP that was tagged with KDEL, a ER retention signal, using T4 ligase. The modified plasmid was designated as PK100. Another construct was generated by removing the KDEL sequence from the PK100 construct by digesting with AvrII and SpeI, and then self-ligating and designating it as PK89. To confirm that KDEL was removed from the PK89 construct, sequencing of plasmid was conducted using a CiFP-specific forward primer and vector specific reverse primer. Both PK89 and PK100 constructs were then cut with NcoI AfeI and we ligated the synthesized BspHI NcoI fragment of DUF642 into them to make the fusion constructs. Figure 4-4 shows a schematic representation of the plasmid construction.

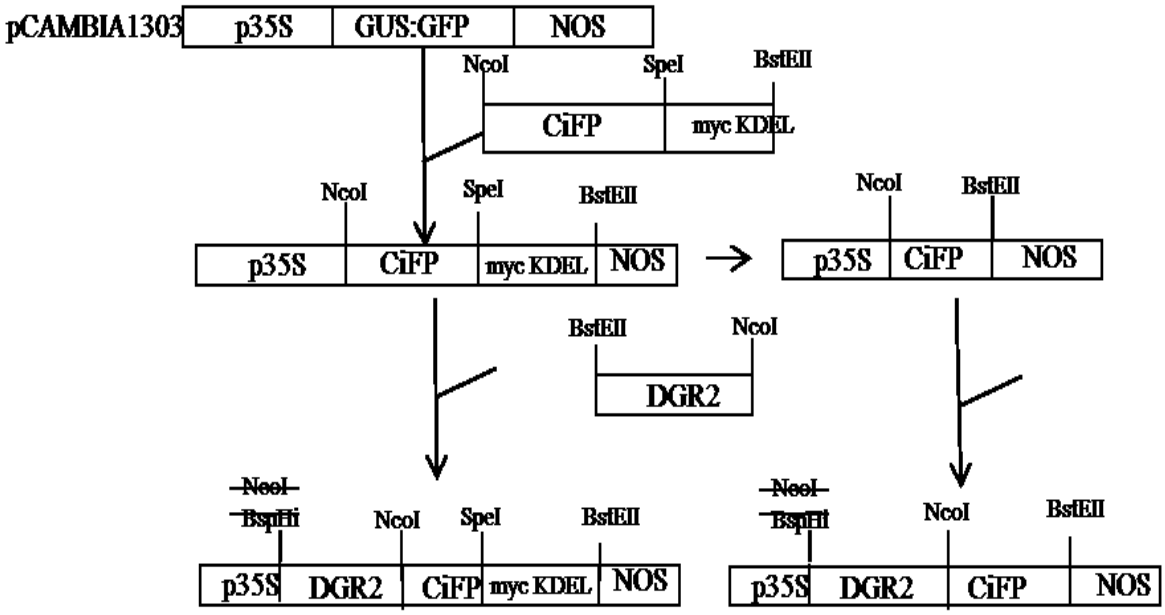


Figure 4-4. Schematic illustration of construction of translational fusion constructs.

A. CLUSTAL 2.1 multiple sequence alignment

```
DGR2                --ATGGAAGGCGTCACCGTCGTGTCTTTCTTCCTTCTTTTCATCGCCACCGCCATGGCCG 58
Snythetic_DGR2      CCATGGAAGGCGTAACCGTTGTTTCTTTCTTCCTGCTGTTCATTGCAACCGCAATGGCCG 60
                      ***********.***** ** *********** ** ***** **.*****.*******

DGR2                CCAAGTCCACCGTCTCCTTCCGTGACGGCATGTTACCAAACGGAGACTTCGAGCTAGGAC
118
Snythetic_DGR2      CGAAATCTACGGTTTCCTTCCGTGATGGCATGCTGCCGAACGGTGACTTCGAACTGGGCC
120
                    * **.** ** ** *********** ****** *.**.*****:********.**.**.*

DGR2                CAAAACCATCAGACATGAAAGGAACAGAAATACTAAACAAACTAGCAATACCAAACTGGG
178
Snythetic_DGR2      CGAAACCGTCTGACATGAAAGGCACTGAGATTCTGAACAAACTGGCGATTCCTAACTGGG
180
                    *.*****.**:***********.**:**.**:**.********.**.**:**:*******

DGR2                AAGTCACAGGATTCGTCGAATACATTAAATCAGGACATAAACAAGGAGACATGCTTCTCG
238
Snythetic_DGR2      AAGTAACCGGTTTCGTTGAATACATCAAATCTGGTCACAAACAGGGCGACATGCTGCTGG
240
                    ****.**.**:***** ******** *****:**:** *****.**.******** ** *

DGR2                TTGTTCCCGCCGGTAAATTCGCTGTAAGACTTGGGAACGAAGCATCGATCAAACAAAGAC
298
Snythetic_DGR2      TGGTGCCTGCGGGTAAATTTGCGGTGCGCCTGGGCAACGAAGCGTCTATCAAACAGCGTC
300
                    * ** ** ** ******** ** **..*.** ** ********.** ********..*:*

DGR2                TTAAAGTGGTTAAAGGAATGTATTACTCACTCACTTTTAGTGCTGCTAGAACTTGTGCAC
358
Snythetic_DGR2      TGAAAGTCGTAAAAGGCATGTACTACTCCCTGACTTTCTCTGCAGCGCGTACGTGTGCTC
360
                    * ***** **:*****.***** *****.** ***** : ***:** .*:** *****:*
```

```
DGR2                AAGACGAGAGACTTAACATATCTGTAGCACCTGACTCCGGTGTGATTCCGATTCAGACGG
418
Snythetic_DGR2      AAGACGAACGTCTGAACATTAGCGTTGCACCGGATTCCGGCGTTATCCCGATCCAGACCG
420
                    *******..*:** *****::  **:***** ** ***** ** ** ***** ***** *

DGR2                TTTATAGTAGTAGTGGTTGGGATTTATATGCTTGGGCGTTTCAAGCTGAGAGTGATGTTG
478
Snythetic_DGR2      TTTATTCTAGCTCCGGCTGGGATCTGTACGCTTGGGCCTTTCAGGCTGAGAGCGACGTGG
480
                    *****: *** :  ** ****** *.** ******** *****.******** ** ** *

DGR2                CTGAAGTTGTGATTCATAATCCTGGTGTTGAGGAAGATCCAGCTTGTGGTCCACTTATTG
538
Snythetic_DGR2      CAGAAGTTGTGATTCATAACCCGGGTGTTGAGGAAGATCCGGCTTGTGGTCCGCTGATTG
540
                    *:***************** ** *****************.***********.** ****

DGR2                ATGGTGTTGCTATGAGATCTCTTTACCCTCCTAGACCAACTAATAAGAACATTTTGAAAA
598
Snythetic_DGR2      ACGGTGTTGCTATGCGTAGCCTGTATCCGCCGCGTCCTACGAACAAAAACATCCTGAAGA
600
                    * ************.*::  ** ** ** ** .*:**:** ** **.*****  ****.*

DGR2                ACGGAGGATTTGAAGAAGGTCCATTAGTATTACCCGGCTCGACAACTGGAGTTTTGATCC
658
Snythetic_DGR2      ATGGTGGCTTTGAAGAAGGCCCGCTGGTGCTGCCGGGTTCTACTACTGGTGTGCTGATCC
660
                    * **:**.*********** **. *.**. *.** ** ** **:*****:**  ******

DGR2                CACCGTTTATAGAAGACGACCACTCTCCTTTACCTGGATGGATGGTGGAGTCTCTCAAAG
718
Snythetic_DGR2      CGCCATTCATCGAGGATGATCACAGCCCGCTGCCAGGTTGGATGGTTGAAAGCCTGAAGG
720
                    *.**.** **.**.** ** ***:  **  *.**:**:******** **.:  ** **.*

DGR2                CTGTCAAGTACGTAGACGTTGAACATTTCTCAGTCCCACAGGGTCGCAGAGCTATTGAGC
778
Snythetic_DGR2      CTGTGAAATATGTTGACGTCGAACACTTTTCTGTACCGCAGGGTCGCCGTGCGATTGAAC
780
                    **** **.** **:***** ***** ** **:**.**.*********.*:** *****.*

DGR2                TTGTAGCGGGTAAAGAGAGTGCCATCGCTCAAGTGGTTCGGACTGTCATTGGGAAGACTT
838
Snythetic_DGR2      TGGTTGCGGGCAAGGAGTCTGCGATCGCCCAGGTAGTCCGCACCGTTATCGGTAAAACCT
840
                    * **:***** **.***: *** ***** **.**.** ** ** ** ** ** **.** *

DGR2                ACGTGCTGTCTTTTGCGGTTGGAGACGCCAACAATGCTTGCAAAGGATCAATGGTGGTTG
898
Snythetic_DGR2      ACGTACTGTCTTTTGCGGTAGGCGACGCCAACAATGCTTGCAAAGGCTCTATGGTCGTTG
900
                    ****.**************:**.***********************.**:***** ****

DGR2                AGGCTTTTGCAGGAAAAGATACACTTAAGGTCCCTTACGAGTCGAAAGGCACAGGAGGGT
958
Snythetic_DGR2      AAGCTTTCGCAGGTAAAGATACCCTGAAAGTTCCGTACGAAAGCAAAGGTACCGGTGGCT
960
                    *.***** *****:********.** **.** ** *****.:  ***** **.**:** *

DGR2                TTAAACGAGCTTCTATTCGATTTGTGGCGGTTTCGACCCGATCAAGAATTATGTTCTACA
1018
```

```
Snythetic_DGR2      TCAAGCGCGCCTCCATCCGTTTTGTAGCAGTGTCCACTCGTTCCCGTATCATGTTCTATT
1020
                    * **.**.** ** ** **:*****.**.** ** ** **:**..*:** ******** :

DGR2                GCACTTTCTATGCCATGAGGAGCGATGATTTCTCGTCATTGTGTGGGCCTGTGATCGATG
1078
Snythetic_DGR2      CCACCTTCTACGCAATGCGTTCCGATGATTTCTCCAGCCTGTGCGGTCCAGTGATCGACG
1080
                     *** ***** **.***.* : ************ : . **** ** **:******** *

DGR2                ATGTCAAGCTTATAAGCGTTCGTAAACCAT---------AG---------- 1110
Snythetic_DGR2      ACGTGAAACTGATCTCTGTCCGCAAGCCGAGCGCTCTCGAGTAAGGTGACC 1131
```

B. CLUSTAL 2.1 multiple sequence alignment

```
Synthetic_DGR2      MEGVTVVSFFLLFIATAMAAKSTVSFRDGMLPNGDFELGPKPSDMKGTEILNKLAIPNWE 60
DGR2                MEGVTVVSFFLLFIATAMAAKSTVSFRDGMLPNGDFELGPKPSDMKGTEILNKLAIPNWE 60
                    ************************************************************

Synthetic_DGR2      VTGFVEYIKSGHKQGDMLLVVPAGKFAVRLGNEASIKQRLKVVKGMYYSLTFSAARTCAQ
120
DGR2                VTGFVEYIKSGHKQGDMLLVVPAGKFAVRLGNEASIKQRLKVVKGMYYSLTFSAARTCAQ
120
                    ************************************************************

Synthetic_DGR2      DERLNISVAPDSGVIPIQTVYSSSGWDLYAWAFQAESDVAEVVIHNPGVEEDPACGPLID
180
DGR2                DERLNISVAPDSGVIPIQTVYSSSGWDLYAWAFQAESDVAEVVIHNPGVEEDPACGPLID
180
                    ************************************************************

Synthetic_DGR2      GVAMRSLYPPRPTNKNILKNGGFEEGPLVLPGSTTGVLIPPFIEDDHSPLPGWMVESLKA
240
DGR2                GVAMRSLYPPRPTNKNILKNGGFEEGPLVLPGSTTGVLIPPFIEDDHSPLPGWMVESLKA
240
                    ************************************************************

Synthetic_DGR2      VKYVDVEHFSVPQGRRAIELVAGKESAIAQVVRTVIGKTYVLSFAVGDANNACKGSMVVE
300
DGR2                VKYVDVEHFSVPQGRRAIELVAGKESAIAQVVRTVIGKTYVLSFAVGDANNACKGSMVVE
300
                    ************************************************************

Synthetic_DGR2      AFAGKDTLKVPYESKGTGGFKRASIRFVAVSTRSRIMFYSTFYAMRSDDFSSLCGPVIDD
360
DGR2                AFAGKDTLKVPYESKGTGGFKRASIRFVAVSTRSRIMFYSTFYAMRSDDFSSLCGPVIDD
360
                    ************************************************************

Synthetic_DGR2      VKLISVRKPSALE 373
DGR2                VKLISVRKP---- 369
                    *********
```

Figure 4-5. Multiple sequence alignment of synthetic DGR2 and DGR2 CDS nucleotides (A) and amino acids (B) sequences using ClustalW multiple alignment tool (http://www.ebi.ac.uk/Tools/msa/clustalw2/

4.3.4.4. Vector construction for heterologous protein expression

PET-22b(+) vector (Novagen) was used for the expression of DGR2 and DUFB in *E. coli*. The vector contains an N-terminal pelB signal peptide sequence for periplasmic localization and C-terminal 6-His tag sequence. For cloning, a *DUFB* gene fragment was amplified from Arabidopsis cDNA and a *DGR2* gene fragment was amplified from the synthetic peptide generated for subcellular localization and cloned into a TOPO cloning vector. Primers with NcoI–HindIII sites for *DUFB* and NcoI-XhoI sites for *DGR2* were used for cloning into the NcoI–HindIII sites and NcoI-XhoI sites of pET-22b(+) , respectively, as shown below.

DGR2:

Fwd: 5'- CAT GCC ATG GCG AAA TCT ACG GTT TCC TT -3

Rvs: 5'- TGG CCT CGA GAG CGC TCG GCT TG -3'

DUFB:

Fwd: 5'- CAT GCC ATG GGC TTC AGT GAC GGG ATG -3'

Rvs: 5'- CCC AAG CTT CGG CTT ACG AGC ACT GAG -3'

At first the generated constructs without removing the native signal peptides of DGR2 and DUFB proteins and used for periplasmic protein expression in *E. coil*. Using those construcs we couldn't get any expressed protein. We predicted that there might be a conflict between the native signal peptide of gene and the signal peptide of the plasmid. So, the predicted signal peptides with a predicted cleavage sites between amino acid residues 19-20 and 22-23 that were present at the N-terminus of DGR2 and DUFB, respectively were removed during cloning of each gene. Therefore the forward primer was designed to anneal 60 and 66 base pairs downstream from the start codon of DGR2 and DUFB, respectively, and the reverse primer was designed to exclude the stop codon. The amplified fragments were cloned into the TOPO cloning vector. The plasmids were then inserted into DH5alpha *E. coli* competent cells by the freeze/thaw method. The positive clones were selected on ampicillin IPTG plates. 5–10 single positive clones were picked and cultured by shaking overnight at 37°C in 2 mL LB liquid media

containing100 µg/mL ampicillin. Plasmids from positive colonies were isolated from *E. coli* using a plasmid miniprep kit. The plasmid concentration was measured by Nanodrop. The plasmids were digested using the FastDigest restriction enzyme from Fermentas according to their protocol. The purified plasmids were subsequently heat-shock transformed into *E. coli* strains BL21(DE3) and BL21(DE3)pLysS from Promega. Thermo Scientific™ FastDigest™ enzymes restriction enzymes used for cloning and purchased from Life Technologies.

4.3.5. Identification of homozygous T-DNA insertion in Salk lines

Confirmed homozygous T-DNA insertion lines Salk_042864C and Salk_094931C, carrying T-DNA in *DGR2* and *DUFB* genes (Table 4-1) and their non-transformed parental control, CS7000, were identified in the TAIR database and obtained from the Arabidopsis Biological Resource Center (ABRC, Ohio, USA). All seeds were grown in soil under the conditions described in section (4.1.1.). Seeds were harvested from individual Salk lines and gDNA was extracted from the corresponding lines. Genotyping was performed by PCR screening to isolate homozygous lines for T-DNA insertion mutations. The following primers were used for genotyping and designed using the SIGnAL T-DNA verification primer design tool (http://signal.salk.edu/tdnaprimers.2.html). For each Salk line, a left primer (LP) and a right primer (RP) were designed. Using two sequence-specific primers (LP & RP) and a left T-DNA border primer (LB), two reactions: LP+RP and LB+RP were performed according to the protocol of SIGnAL.

Primers for Salk lines:

Salk_042864:

LP TTTTCAGACAATTGGCGAGAG

RP AAGTGGTTCGGACTGTCATTG

Salk_094931:

LP TTTGTTTTGTGGCTATCGAGC

RP CATGAGTTGGCATTTGTGTTG

LBb1.3:

ATTTTGCCGATTTCGGAAC

Table 4-1. SALK T-DNA insertion lines used in this study.

Gene ID	Salk line ID	Insertion number	Insertion location	Ecotype	Plasmid	Selection
At5g25460 (*DGR2*)	Salk_042864C	Single	Exon	Col-0	pROK2	Kanamycin
AT5g11420 (*DUFB*)	Salk_094931	Single	Promoter	Col-0	pROK2	Kanamycin

4.3.6. Plant transformation and genetic selection

Arabidopsis ecotype Col-0 (wild type) was transformed with the floral dip technique using *Agrobacterium tumefaciens* strain GV3101 (Clough and Bent, 1998) for generating all the transgenic lines used in this study. At first, the binary vector containing a plant selectable resistance gene was introduced into the *A. tumefaciens* strain by the freeze-thaw method (Weigel and Glazebrook, 2002) and then the floral dip method was used to transform Col-0 plants with the plasmids. Surface-sterilised seeds of the transformants were selected on solid media containing half-strength MS mediumn, 1% (w/v) sucrose, and 0.8% (w/v) agar (Phytoblend; Caisson Laboratories) supplemented with appropriate antibiotics such as 50 µg/mL of hygromycin for overexpression plants and 50 µg/mL of kanamycin for RNAi plants. The plants were self-pollinated and transgenic lines segregating ~3:1 for antibiotic resistance in the T_2 generation were selected. Homozygous plants from the T_3 generation were used for mutant phenotyping analysis.

4.3.7. Hormone treatment

Seedlings of transgenic plants, 35S::*DGR2*, 35S::*DUFB*, RNAi, Salk_*dgr2* and Salk_*b* were grown vertically on square plates containing ½ MS medium supplemented with 100 µM of 3-indoleacetic acid (IAA, Sigma) with a density of 30 seedlings per plate and were allowed to grow normally for 7 days. Images were captured under a microscope (Leica Microsystems, Wetzlar, Germany) mounted with HDCE-90D camera. Root and hypocotyl lengths were quantified using ImageJ software.

4.3.8. Primer design for qRT-PCR

The cDNA sequences of At5g25460 *(DGR2)* and At5g11420 *(DUFB)* were extracted from TAIR database and used to design primers for qRT-PCR using the Universal Probe Library (http://lifescience.roche.com/). Primers were synthesized by Integrated DNA Technologies (IDT). The specificity of primers was confirmed *in silico* by using a BLASTN search against *Arabidopsis thaliana* database in NCBI. UBQ10 (At4g05320; RG) was used as a reference gene based on the published paper (Pfaffl, 2001). The primer sequences for the genes were as follows:

AT5g25469 (*DGR2*):

Fwd: 5’-TCAATATGGAAGGCGTCACC -3’

Rvs: 5'- CCT AGC TCG AAG TCT CCG TTT -3'

TTGTGACCACCGAGAGAGCTTGGTAA (3' ‡ 5')

AT5g11420 (DUFB):

Fwd: 5'- GTC TCT TCT CTT TAC TTT GGT CGT C -3'

Rvs: 5'- AGT CGC CGT TTG GTA ACA TC -3'

UBQ10:

Fwd: 5’-GGCCTTGTATAATCCCTGATGAA-3’

Rvs: 5-AGAAGTTCGACTTGTCATTAGAAAGAAA-3'

4.3.9. Quantitative Real-Time PCR (qRT-PCR)

The qRT-PCR experiments were performed using an ABI 7500 fast real-time PCR system (Applied Biosystem) and with SYBR Green dye-based detection method. SYBR Green master mix was obtained from MBSU, University of Alberta, which included SYBR Green 1 dye, Ampli Taq Gold DNA polymerase, dNTPs with dUTP, passive reference and optimized buffer components. The reactions were carried out in 96 well plates and each well contained a 10 µl reaction volume that included 5 µl of master mix, 2.5 µl of cDNA template and 2.5 µl of forward and reverse primers mix (1.6µM of each of the primer concentration).

The specificity and amplification efficiency of all primers were validated by qRT–PCR. A standard curve was generated by performing qPCR with a serial dilution in which an aliquot of cDNA was diluted to 1/4, 1/16, 1/64, 1/256, 1/1024, and 1/4096 fold with ddH20. qPCR was conducted with all the dilutions and with a no-template control, using previously optimized primer, *UBQ10*. Primer efficiency (E) was calculated using a plot of Ct versus the log and performing a linear regression analysis. The reaction efficiency was calculated from the slope of the line using the equation: Efficiency = 10(–1/slope)-1. The validated primers were used for the transcript expression analysis with the template dilution of 1/64 in a 10 µl reaction volume. All qRT-PCRs were performed using the following program: initial denaturation for 2 min at 95°C followed by 40 cycles of 15s at 95°C and 1 min at 60°C. A dissociation curve was obtained after completion of the qPCR cycles.

The relative transcript expression patterns of *DGR2* and *DUFB* were calculated by the comparative CT method, also referred to as the ΔΔCT Method (Livak and Schmittgen, 2001). The ΔCT value was calculated by the flowing formula:

ΔCT = CT target – CT reference

The ΔΔCT value was calculated by the following formula:

ΔΔCT = ΔCT test sample – ΔCT calibrator sample

The fold differences were calculated using the following formula:

Fold change=2–ΔΔCt

4.3.10. Total protein extraction

Total protein was extracted from different tissues of rosette Arabidopsis plants that included green silique, top of the stem, bottom of the stem, flowers, rosette leaves of 3-weeks old and 5 days old Arabidopsis Col-0 seedlings. In addition, total protein was also extracted from hypocotyl tissues of 6 DAS (days after sowing) and 13 DAS plants grown in Magenta boxes under the same conditions described in Section 4.1.1, except that they were wrapped in two layers of aluminum foil. Tissues were homogenized in phosphate protein extraction buffer (100 mM sodium phosphate 25 mM Tris-HCl, 10% glycerol, 1% Triton 100, 10 mM EDTA, and protease inhibitor cocktail) and then quantified the extracted proteins by Bradford protein assay method (Bradford, 1976). Next, an aliquot of protein extracts (10 µl) of equal concentration from each sample was subjected to SDS-PAGE and Western Blotting.

4.3.11. Protein production

E. coli BL21(DE3) and BL21(DE3)pLysS cells transformed with the plasmids were struck from glycerol stocks on to LB-agar plates supplemented with 100 µg/ml ampicillin (Sigma-Aldrich, Cat No. 9518-25g) and 34 µg/ml chloroamphenicol (Sigma-Aldrich Cat No. C0378-5G) and incubated at 37°C overnight (ON). 5-6 single colonies of plasmid-containing cells were then inoculated into 5 ml of 2XYT medium supplemented with 100 µg/ml of ampicillin and 34 µg/ml in 5-6 tubes and incubated at 37°C/220 rpm for 2-3 hours. The cultures were then re-inoculated into 15 ml 2XYT supplemented with 100 µg/ml ampicillin and 34 µg/ml chloroamphenicol medium until the OD600 nm reached between 0.6-0.8. At OD600 nm 0.6-0.8, 0.1 mM, protein expression was induced by addition of 0.5 mM and 1.0 mM isopropyl-β-D-thio-

galactopyranoside (IPTG) and then incubated at 37°C, 18°C and 22°C /250 rpm for 2h, 4h and overnight (except for 37°C). Small-scale expression and purification experiments were performed for optimization before the large-scale expression (1L).

The single colonies that were used for protein expression were also tested for toxicity and plasmid viability (Table 4-2) by using the protocol described by the EMBL (http://www.ebi.ac.uk/). At first, the colonies were suspended in 200µl of water and were vigorously shaken. The suspensions were then plated on four different plates as follows:

Table 4-2. Toxicity and plasmid viability test.

Plate	Cells that grow on these plates
LB plate	All viable cells
LB plate + antibiotic	Cells that still carry the plasmid
LB plate + IPTG (1 mM)	Cells that have lost the plasmid or mutants that have lost the ability to express the target gene
LB plate + antibiotic + IPTG (1 mM)	Only mutants that retain the plasmid but have lost the ability to express the target gene

4.3.12. Cellular compartment fractionation

Cells were harvested by centrifugation for 20min at 8000rpm and 4°C and the periplasmic *E. coli* fraction was extracted via modified osmotic shock as described by Sockolosky and Szka (2013). Briefly, harvested cells were suspended in a hypertonic solution of 50 mM Tris, 20% w/v sucrose, ,pH 8 (25 mL), and a Complete EDTA-free protease inhibitor cocktail tablet (Roche). This mixture was incubated for 30 min at 4 °C. Cells were centrifuged and the supernatant was collected. Cells were re-suspended in a hypotonic solution of 5 mM $MgSO_4$ (25mL) and incubated for 30 min at 4 °C followed by an additional centrifugation. The supernatant from the hypotonic solution was combined with the supernatant from the hypertonic solution, centrifuged to remove debris, and dialyzed against 50 mM Tris-HCl, 100mM NaCl, pH.8.8 using Spectra/Por molecular porous membrane tubing of MWCO: 12-14,000, diameter: 29.0mm vol/length: 64.4

ml/cm (Spectrum Laboratories,) with three buffer changes, overnight at 4°C. At the end of the dialysis, lysate was centrifuged to remove any precipitates.

4.3.13. Protein purification

The molecular weight and isoelectric point of DGR2 were 39.97 and 7.4 and for DUFB were 39.64 and 7.8, as obtained from TAIR database. After removal of signal peptide and addition of 6-His tag, the values were predicted to 39.30 and 7.15 for DGR2 and for DUFB they were 38.35 and 7.12 as calculated by Compute pI/Mw tool(http://web.expasy.org/compute_pi/). Based on the new pI values, the pH of the buffers for purification was set to 8.8. This is because, at pH=pI, the protein had zero net charge and as a result, the proteins could be precipitate out at pH very near to pI values of the purified proteins. The periplasmic fractions containing soluble DGR2, DUFB and control were purified by Ni^{2+} affinity chromatography as follows. At first, the periplasmic extracts were incubated with Ni-NTA agarose (QIAGE, R10-22-40-42/43, S13-26-46) and rocked overnight at 4°C prior to purification. Poly-Prep columns of 9 cm high and 0.8 x 4 cm dimensions (BIO-RAD, CAT # 731-1550) which can hold up to 2 ml of chromatography support and 10 ml of sample in an integral reservoir were used for purification. The columns were prepared by equilibrating with two column volumes (CV) of equilibration buffer (50 Mm Tris-HCl, 100mM NaCl, pH 8.8). Afterward, periplasmic extract incubated with Ni-NTA-agarose was gradually added into the column and the flow through was collected. Two different washing steps were used as part of optimization as follows:

Washing procedure I

I. 50 Mm Tris-HCl, 100mM NaCl, 20mM mM imidazole, pH 8.8 by three CVs

II. 50 Mm Tris-HCl, 100mM NaCl, 40mM mM imidazole, pH 8.8 by two CVs

III. 50 Mm Tris-HCl, 100mM NaCl, 100mM mM imidazole, pH 8.8 by one CVs

The above washing buffer resulted in lower yields and higher contaminants.

Washing procedure II

I. 50 Mm Tris-HCl, 100mM NaCl, 4mM mM imidazole, pH 8.8 by one CVs

II. 50 Mm Tris-HCl, 100mM NaCl, 4M mM imidazole 10% ethanol, pH 8.8 by one CV

III. 50 Mm Tris-HCl, 100mM NaCl, 20mM mM imidazole, 1.5M NaCl, pH 8.8 by one CV

IV. 50 Mm Tris-HCl, 100mM NaCl, 20mM mM imidazole, pH 8.8 by one CV

V. 50 Mm Tris-HCl, 100mM NaCl, 40mM mM imidazole, pH 8.8 by one CV

These washing steps also could neither increase the yield nor lower the contaminants.

Bound protein was eluted using the following elution buffers:

Elution 1: 50 Mm Tris-HCl, 100mM NaCl, 100 mM imidazole-1ml

Elution II: 50 Mm Tris-HCl, 100mM NaCl, 300 mM imidazole. Three fractions were collected by adding 1 ml of Elution II buffer each time.

The fractions were pooled, concentrated by an Amicon 3 kDa MWCO Ultra Centrifugal Filter Units (Millipore, Cat No. UFC500396). Protein concentrations were measured using the Qubit fluorometer (Invitrogen, Carlsbad, CA) and Quant-iT Protein Assay Kit (Invitrogen) adapted for the Qubit fluorometer according to the manufacturer's protocol. The expected ~40kDa recombinant DGR2 and DUFB were resolved and visualized by SDS-PAGE and western blotting using the His-tag antibody (CEDARLABE, Cat. No. A00174-200) and the DUF642-antibody. The expected ~40kDa recombinant DGR2 and DUFB were resolved and visualized by SDS-PAGE and western blotting using the His-tag antibody (CEDARLABE, Cat. No. A00174-200) and the DUF642-antibody.

4.3.14. Co-Immunoprecipitation

5 days old wild type Col-0 Arabidopsis seedlings grown under conditions described in Section 4.1.1. were homogenized in phosphate protein extraction buffer and then quantified by Bradford protein assay. After protein extraction, co-immunoprecipitation was performed with DUF642

antibody and using the Pierce™ Classic Magnetic IP/Co-IP Kit (Thermo scientific, Cat. No.88828). In brief, magnetic beads were washed with ice-cold 1 mM HCl and then incubated with DUF642-antibody for 60 minutes. Extracted protein was incubated with antibody-bound beads for overnight at 4ºC. The beads were washed twice by washing buffer and quenched the reaction for 60 minutes with Quenching Buffer followed by washing the beads once with Modified Borate Buffer and once with IP Lysis/Wash Buffer. Immunoprecipitates were washed two times with Pierce IP lysis/wash buffer and once with ultrapure water. Finally, antigen was eluted using Pierce Elution buffer. For protein detection, samples were separated by 12% gradient SDS–PAGE for 10-15 minutes so that total protein can run up to ~1 cm in length. Then the ~1 cm band was excised from the gel and analyzed by MALDI-TOF mass spectrometry performed by Alberta Proteomics and Mass Spectrometry Facility Department of Biochemistry, Faculty of Medicine & Dentistry, University of Alberta. The control experiment was performed in the same way, omitting the DUF642 antibody.

4.3.15. SDS-PAGE and western blotting

4.3.15.1. SDS-PAGE

The protein samples were subjected to electrophoresis analysis (Laemmli, 1970). An aliquot of protein extracts (10 µl) of equal concentration from each sample was mixed with 6x Laemmli sample buffer (Bromophenol blue (0.25%), DTT (dithiothreitol; 0.5 M, added immediately before use), glycerol (50%) SDS (sodium dodecyl sulfate; 10%)) and was heated at 100ºC for 8-10 minutes. 15 µl of each prepared sample were loaded on a 10% acrylamide gel (stacking gel: 4.9 ml water; 2.5 ml 1.5M Tris-HCl, pH 8.8; 2.5ml acrylamide/bis 37.5:1, 40%; 100 µl 10% SDS; 50 µl10% ammonium persulfate, 20 µl TEMED; resolving gel: 980 ml water; 440 µl 1.5M Tris-HCl, pH 6.8; 300 µl acrylamide/bis 37.5:1, 40%; 18µ10% SDS; 10 µl10% ammonium persulfate; 10 µl TEMED) and run using 1X SDS running buffer (3g Tris Base; 18.8g glycine; 10 ml 10% SDS to 1L) using the Mini-PROTEAN Tetra cell electrophoresis system for protein from Bio-Rad electrophoresis at 100 V for an hour. Gels were stained with Coomassie Blue-G-250 solution (methanol:water = 1:1;10% acetic acid; 0.006% (w/v) Coomassie Blue dye) for overnight at 4ºC and destained with 50% methanol/50%water/10% acetic acid for at least two hours.

4.3.15.2. Western Blot

The expected DGR2 and DUFB bands were visualized by western blotting using a DUF642-antibody. This DUF642 antibody was raised in rabbit using the synthetic peptide, PNGDFELGPKPSDMC. This peptide is found in four DUF642 genes: At5g25460 (DGR2), MW-39.97; At5g111420 (DUFB), MW-39.64; At1g80240 , MW-40.22 and At4g32460, MW-39.82. Peptide synthesis and generation of DUF642antibody were carried out by Genscript (Piscataway, NJ). The proteins from the electrophoresis gels were transferred onto PVDF (BIO-RAD,Trans-Blot® Turbo™ RTA Transfer Kit BIO-RAD, CAT # 170-4272) using the Trans-Blot Turbo Transfer System (BIO-RAD, CAT # 170-4155) which was compatible with traditional semi-dry blotting systems at 25 V for 10 minutes. The membrane was blocked overnight at 4°C using blocking solution of 5% skim milk (BD, Cat No. 232100) in TBST (50 mM Tris.HCl, pH 7.4, 150 mM NaCl, 0.1% Tween 20). Next, the membrane was incubated with a 1:5000 dilution of primary antibody in 1X TBS (50 mM Tris.HCl, pH 7.4, 150 mM NaCl) for 2 hours. Afterward, the membranes were washed with TBST, 10 minutes each, then rinsed in TBS. After washing, a secondary antibody, goat anti-mouse IgG horseradish peroxidase (ROCKLAND, KCB003) was used at a dilution of 1:8000 in TBST containing 1% skim milk and incubated at room temperature for 45 minutes. All westerns were developed with TMB peroxidase substrate kit (VECTOR; SK-4400) according to the manufacturer's protocol.

4.3.16. Protein identification by peptide mass fingerprinting

DGR2 and DUFB peptides were identified from total proteins (Section 4.1.10), purified heterologously expressed proteins (section 4.1.13) and proteins obtained from Co-IP (section 4.1.14) by peptide mass fingerprinting using mass spectrometry. The SDS page gel bands corresponding to expected DGR2 and DUFB that showed signals in western blot analysis were cut from the gel and sent to Alberta Proteomics and Mass Spectrometry Facility Department of Biochemistry for analysis. Briefly, the excised gel bands were destained twice in 100mM ammonium bicarbonate/acetonitrile (50:50). The samples were then reduced (10mm BME in 100mm bicarbonate) and alkylated (55mM iodoacetamide in 100mm bicarbonate). After dehydration, enough trypsin (6ng/ul) was added to just cover the gel pieces and the digestion was

allowed to proceed overnight (~16 hrs.) at room temperature. Tryptic peptides were first extracted from the gel using 97% water/2% acetonitrile/1% formic acid followed by a second extraction using 50% of the first extraction buffer and 50% acetonitrile. Fractions containing tryptic peptides dissolved in aqueous 25% v/v ACN and 1% v/v formic acid were resolved and ionized by using Nanoflow HPLC (Easy-nLC II, Thermo Scientific) coupled to the LTQ XL-Orbitrap hybrid mass spectrometer (Thermo Scientific). Nanoflow chromatography and electrospray ionization were accomplished by using a PicoFrit fused silica capillary column (ProteoPepII, C18) with 100µm inner diameter (300Å, 5µm, New Objective). Peptide mixtures were injected onto the column at a flow rate of 3000 nL/min and resolved at 500 nL/min using 70 min linear gradients from 0 to 45% v/v aqueous ACN in 0.2% v/v formic acid. The mass spectrometer was operated in data-dependent acquisition mode, recording high-accuracy and high-resolution survey Orbitrap spectra using external mass calibration, with a resolution of 60 000 and m/z range of 400–2000. The fourteen most-intense, multiply-charged ions were sequentially fragmented by using collision-induced dissociation, and spectra of their fragments were recorded in the linear ion trap. After two fragmentations, all precursors selected for dissociation were dynamically excluded for 60 s. Data was processed using Proteome Discoverer 1.4 (Thermo Scientific) and the Uniprot *Arabidopsis thaliana* database was searched using SEQUEST (Thermo Scientific). Search parameters included a precursor mass tolerance of 10ppm and a fragment mass tolerance of 0.8Da. Peptides were searched with carbamidomethyl cysteine as a static modification and oxidized methionine and deamidated glutamine and asparagine as dynamic modifications.

4.3.17. Immunohistochemistry (IHC) analysis

Whole-mount immunolocalizations were performed as described by Sauer et al. (2006) using 35S::DGR2:CiFP, 35S::DGR2:CiFP:KDEL and 35S::CiFP transgenic Arabidopsis roots from three independent lines. Antibodies were diluted as follows: rabbit anti-SEC21 (1:1000 in 1% BSA in 1x PBS; Agrisera AS08 327) and Cy5-conjugated secondary anti-rabbit (1:300 in 1% BSA in 1x PBS; Jackson ImmunoResearch, Code No. 711-175-152) antibodies. In brief, root tissues were fixed in 4% paraformaldehyde by incubating for 1h in a vacuum desiccator at room temperature. After washing with water, fixed tissues were transferred to the microscope slides

and dried overnight at room temperature. The boundaries of the dried area were marked using a PAP pen (ImmEdge™ PEN, Vector, Cat No: H-4000) on the microscope slide and the material was rehydrated by 1X PBS. 2% Driselase (Sigma, Cat No. D9515-1G) was added onto the microscope slide and incubated for 60 minutes at 37ºC which facilitated the access of the antibody to the antigen by digesting cellulose and pectins. Driselase was removed by five washing steps with 1X PBS for 10 minutes each. Next, 3% IGEPEAL CA-630 (Sigma, Cat No. 18896-50ML) plus 10% DMSO was added and incubated for 1h at room temperature. IGEPEAL CA-630 was removed by 5 washing steps with IX PBS incubated 10 minutes each. After washing, tissues were blocked with 3% bovine serum albumin (BSA) for 60 minutes at room temperature. After removing the BSA, tissues were incubated with primary antibody, Sec21, for 4h at 37ºC. Primary antibody was removed from the microscope slides by washing with 1X PBS five times and incubating 10 minutes each time. After removing the primary antibody, secondary antibody was added to the microscope slides and incubated for 3h at 37ºC. After incubation, secondary antibody was washed out by 1X PBS as was performed for washing the primary antibody. The liquid was removed and a drop of ProLong®Diamond Antifade mounted (Life technologies, Cat No. P36961) medium was added to the microscope slides and covered the sample with a cover slip. The antifade medium was cured for 24h at room temperature in the dark and stored at 4ºC until the tissues were analyzed by confocal microscopy.

4.3.18. Confocal microscopy

Seedlings from three independent lines of the T2 generation bearing 35S::DGR2:CiFP, 35S::DGR2:CiFP:KDEL and 35S::CiFP translational fusions were subjected to confocal-laser scanning microscopy at 5 days after planting. Imaging was performed on a Zeiss LSM 510 Meta confocal laser scanning microscope using a 40x objective. For CiFP fluorescence analysis, the 514 nm excitation line of an argon ion laser was used, with a 520-555-nm band-pass filter. For co-localization study, three images of Cy5, CiFP and superimposition of Cy5 and CiFP were captured simultaneously. Cy5 fluorescence was analyzed with the 633 nm excitation line of a HeNe 633 laser and with a 647 - 700-nm band-pass filter. Images were processed with the LSM Image Browser ZEN lite.

4.3.19.1. Cell wall sugar analysis using GC/MS

Cell wall sugar composition of 7 DAS experimental plants was determined with GC/MS measurement according to a protocol by Foster et al (2010), and as explained briefly in the following sections. These plants were grown under normal conditions in ½ MS agar plates described in section 4.1.1. Three biological replicates and two technical replicates were used.

4.2.19.2. Cell wall isolation

Seedlings at 7 DAS were collected and lyophilized and 60-70mg of lyophilized plant materials were ground into a fine powder in a 2ml Sarstedt screw cap tube, with two 5.5mm stainless steel balls and a Retschmill. The powdered plant material of each sample was then washed with 1.5 ml of 70% ethanol by centrifuged at 10,000 rpm for 10 min, and the supernatant from the samples were removed and then the pellet of each sample was washed with 1.5 ml chloroform/methanol (1:1 v/v) followed by centrifuge at 10,000 rpm for 10 min. Pellets were again washed with 500 µl acetone to obtain alcohol insoluble residue (AIR).

4.2.19.3 Starch removal

The AIR was subsequently de-starched by re-suspending the pellet from each sample in 1.5ml of 0.1M sodium acetate pH 5.0 and heating for 20 min at 80ºC in a heating block. After cooling, destarched material was mixed with 35µl of 0.01% sodium azide (NaN_3), 35µl amylase (50ug/ml in water; E4551, Sigma) and 17µl pullulanase (E2412, Sigma) were added and incubated overnight at 37ºC in a shaking incubator. The digestion was terminated by heating the suspension at 100ºC for 10 min. The pellet was then washed in 1.5ml dH2O three times following vortex, centrifuge, and carefully removal of supernatant. Next, pellets were re-suspended by adding 500µl acetone in each sample followed by evaporation.

4.3.19.4. Weak acid hydrolysis (converting non-cellulosic polysaccharides to monosaccharides)

The non-cellulosic neutral monosaccharide composition of the wall matrix polysaccharides was obtained by treating de-starched AIR with trifluoroacetic acid. In brief, 250µl of 2M trifluoroacetic acid (TFA; 2M = 153ul of stock TFA + 847µl water) was added to 2mg of cell wall material. As an internal standard, 10µl of 5mg/ml inositol was added to each sample. The samples were then incubated for 90 min at 120ºC in a heating block and then centrifuged at 10,000 rpm for 10 min. 100ul of acidic supernatant was transferred to glass screw cap vial and evaporated under a gentle stream of air at RT (~30 min). The pellets were washed three times with 300µl 2-Propanol following vortexing, and evaporation (~20 min).

4.3.19.5. Producing alditol acetates

Monosaccharide composition of plant cell wall can be determined by quantitatively and qualitatively followed by derivatization to alditol acetates after hydrolysis and reduction (Albersheim et al., 1967; Blakene et al., 1983; Saeman et al., 1983). In this experiment, to reduce cyclic sugars to linear alditols, 200µl of 10mg/ml sodium borohydride solution in 1M ammonium hydroxide (1M = 66ul of stock NH_4OH + 934µl 100% EtOH) was added to each sample and incubated at room temperature for 90 min. The sample was neutralized by adding 150µl glacial acetic acid and the samples were evaporated under a gentle stream of air at room temperature. Next, 250µl of acetic acid/methanol (1:9 v/v) was added and evaporated, and then 250µl methanol was added followed by vortexing, and evaporation three times. To acetylate the alditols, 50µl acetic anhydride and 50µl pyridine was added to each sample and incubated at 120ºC for 20 min in a heating block. After evaporating the samples under a gentle stream of air at room temperature, the samples were washed with 200ul toluene three times and then 500µl ethyl acetate was added. Then, 2ml of dH2O was added and the sample was centrifuged at 2,000 rpm for 5 min to separate the layers (ethyl acetate was at the top).

4.3.19.6. GC-MS

For sugar analysis by GC/MS, an Agilent 5975 GC/MS with Electron Impact (EI) ion sources; an Agilent 7693 autosampler and a ChemStation instrument control and data handling system were

used. 1µl samples were injected with helium carrier gas at a flow rate of 1 ml/min. in a HP-5MS column (30m x 0.25mm x 0.25µm). The initial temperature was 110ºC for 1 min., 3ºC/min until 200C, 10min at 270C, 10C/min until 250C, and a hold at 250C for 10min. To identify the retention time of each sugar, the known concentrations of samples containing rhamnose, fucose, arabinose, xylose, mannose, glucose, and galactose separately were detected using the same method that was used for the experimental sample analysis. Moreover for quality control, blank samples were run after every five samples. Sugar contents were calculated by means of auto integration (ChemStation integrator, threshold=15) of peak area of each sugar. The sugars were identified by comparing the retention time with the retention time identified for the known sugars. The metabolites peak areas were divided by peak area of the internal standard, ribitol, to correct any recovery differences. The corrected peak areas were then normalized by dividing by the dry weight of the samples

4.3.20. Statistical analysis and graphics

One-way ANOVA was conducted using MS Excel and graphs were created using SigmaPlot 11.0.

4.4. Results

4.4.1. Tissue specific mRNA expression pattern

Tissue-specific mRNA expression analysis of *DGR2* and *DUFB* was performed using quantitative real-time PCR (qRT-PCR) in Columbia-0 wild-type plants grown under conditions described in Section 4.1.1. The primer pairs were tested for specificity by using cloned *DGR2* or cloned *DUFB* as templates in separate PCR reactions. Total RNAs were extracted from: fully expanded rosette leaves; roots; flowers; green siliques and stems from 3-4 week old plants and 5 DAS seedlings of *A. thaliana* Col-0 plants. Comparative Ct measurements (delta delta Ct) give a relative expression difference between samples, where a lower delta Ct means greater expression and were used to determine the mRNA expressions in different tissues relative to a control gene.

In this study, both *DGR2* and *DUFB* genes showed lower delta Cts (higher transcript abundance relative to a control gene) in the root, flower and seedling tissues (Fig.4-6A and Fig.4-7A) and to stem, siliques and rosette leaves showed higher delta Cts (lower transcript abundance relative to a control gene). Moreover, transcript abundance of *DGR2* was higher compared to *DUFB* (Fig. 4-6B and Fig.4-7B). Data were converted to fold-change (delta delta Ct), relative to expression levels in stem, which showed that in seedlings (5 DAS), transcripts for *DGR2* were enriched in seedlings, roots and flowers (Fig. 4-6B). A higher level of *DUFB* gene expressions was also detected in seedlings, roots and flowers compared to stems (Fig. 4-7B). The level of *DGR2* transcript expression in roots was approximately five times higher than in flower.

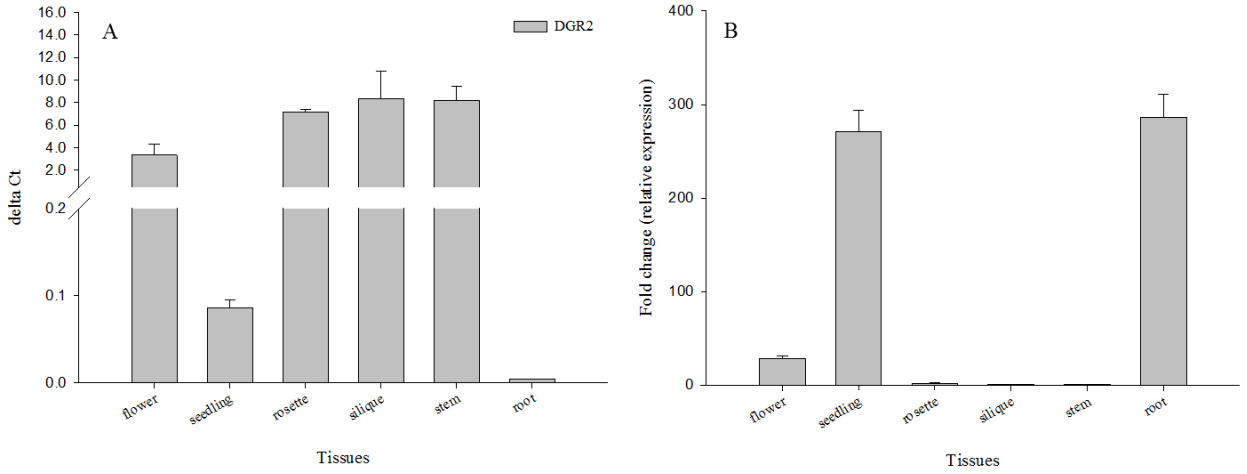


Figure 4-6. Tissue-specific *DGR2* gene expression. Mean ΔCt values (normalized against UBQ10), (A) and relative expression of DGR2 in different tissues. The expressions were relative to the stem (B). Fold change in expression was calculated using the double delta Ct method. Data were shown as the mean values of twelve individuals from three biological and three technical replicated with standard deviations illustrated as vertical bars.

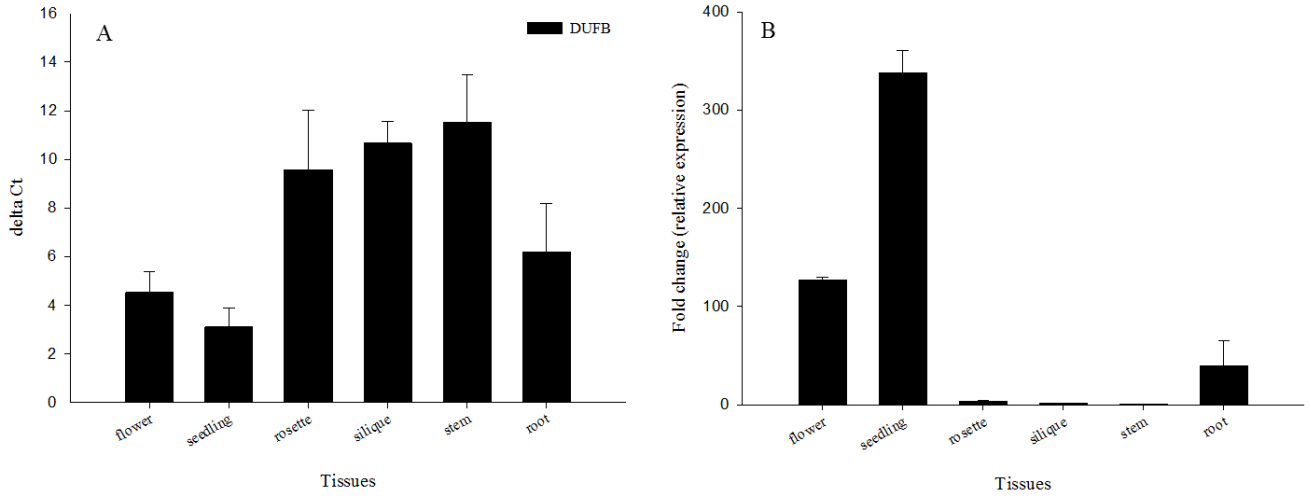


Figure 4-7: Tissue-specific *DUFB* gene expressions. Bar graph showing the mean ΔCt values (normalized against UBQ10), (A) and relative expression of *DUFB* in different tissues. The expressions were relative to the stem (B). Fold change in expression was calculated using the double delta Ct method. Data were shown as the mean values of twelve individuals from three biological and three technical replicated with standard deviations illustrated as vertical bars.

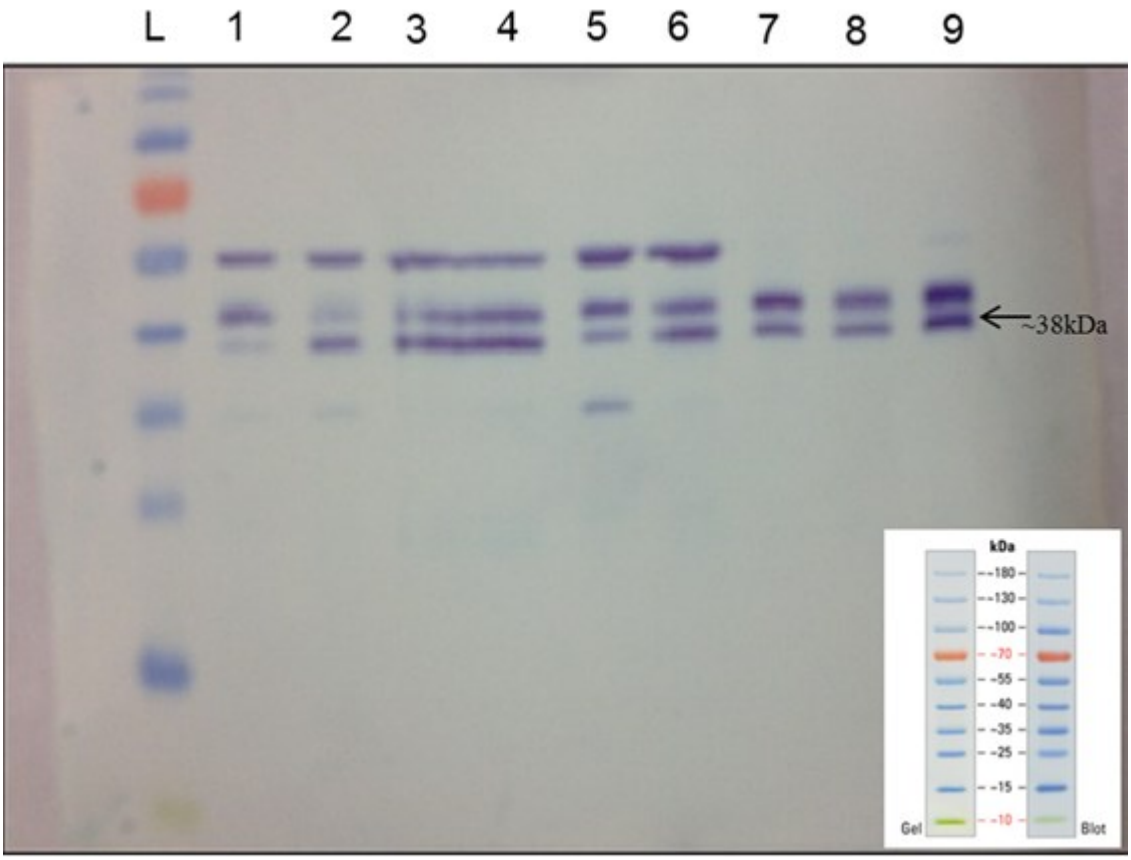
4.4.2. Tissue specific protein expression patterns

To identify the organ specific expression patterns of DGR2 and DUFB proteins, an antibody was raised against synthetic peptide (PNGDFELGPKPSDMC) of DUF642. This peptide was found in four different DUF642 proteins in Arabidopsis that included DGR2, DUFB and At1G80240 and At4G32460. Total proteins were extracted from wild type Arabidopsis plants, including whole seedlings at 5 DAS (days after sowing) and specific organs and tissues at 3-4 weeks after sowing (Figure 4-8). The organs tested were: basal region of stems, apical region of stems, rosette leaves, flowers, siliques. In addition, dark-grown hypocotyls were tested at 6 DAS and 13 DAS Arabidopsis hypocotyl has been used as a model for organ elongation because in this organ, growth occurs by cell elongation with almost no cell division (Gendreau et al., 1997; Raz and Koornneef, 2001; Saibo et al., 2003). Previous study (Kudo and MII, 2004) showed that in the dark, hypocotyls grew exponentially between 3-7 days but after 10 days, no further elongation was detected. In this study, we selected 6 DAS, dark grown hypocotyl to represent fast-elongating cells and 13 DAS to represent cells where elongation was minimal, to test whether DGR2 and DUFB are involved in cell elongation.

The predicted molecular weight of DGR2 and DUFB proteins is ~40kDa with the signal peptide. On the western blot, up to four distinct proteins bands were detected in each lane. These bands were approximately ~35kDa, ~38kDa, ~45kDa and ~55 kDa (Fig. 4-8). The ~35kDa band was detected only in lanes 1,2, and 5 (stems and rosette leaves). The ~38kDa and ~45kDa bands were detected in all tissues, and with similar intensity in each tissue, except in lane 1 (basal region of stems), where the ~38kDa band was notably less intense, and lane 2 (apical region of stem), where conversely the ~45kDa band was notably less intense. The ~55kDa band was detected in all lanes, except lanes 7-9 (dark grown hypocotyls).

To confirm that DGR2, DUFB, and other DUF642 containing proteins were present in the regions of the gels identified in western blot, three bands were cut from lane 6 (seedlings) and four bands were cut from lane 5 (rosette leaves) from a corresponding SDS page gel, Peptide mass fingerprinting was conducted using mass spectrometry. DGR2, DUFB, and At3g08030

were identified in proteins extracted from the band at ~38kDa (Fig. 4-8, indicated by arrow; Table 4-3. Two more DUF642 proteins, At4g32460 and At5g14150 were identified in the bands at ~45 kDa and above ~55kDa, respectively (Table 4-3). No DUF642 proteins were identified in the band at ~35 kDa. (Fig. 4-8).



~38kDa

Figure 4-8: Western blot analysis with DUF642 antibody. Lane L: Protein marker; 1: bottom of stems 2: top of stems, 3: siliques, 4: flowers, 5: rosettes, 6: seedlings, 7: 13d hypocotyls, 8: 13d hypocotyls 9: 6d hypocotyls. Inset: protein ladder used. An aliquot of protein extracts (10 µl) of equal concentration from each sample were loaded per lane, on 10% acrylamide SDS PAGE gel before protein transfer onto PVDF membrane.

Table 4-3. Proteins identified by mass spectrometry; MS sample refers to the gel slice cut out from the polyacrylamide gel (Figure 4-8).

Size	Accession	Description	Score	Coverage	# Unique Peptides	Pfam IDs	# AAs	MW [kDa]	calc. pI
~38kDa	Q8H168	Putative uncharacterized protein At5g11420 OS=Arabidopsis thaliana GN=At5g11420 PE=2 SV=1 - [Q8H168_ARATH]	6.63	7.65	2	Pf04862	366	39.64	7.69
	Q940Q9	Putative uncharacterized protein AT5g25460 OS=Arabidopsis thaliana PE=2 SV=1 - [Q940Q9_ARATH]	10.8	12.91	3	Pf04862	369	39.97	7.36
	Q9C6U3	Putative uncharacterized protein At3g08030 OS=Arabidopsis thaliana GN=T8G24.2 PE=4 SV=1 - [Q9C6U3_ARATH]	21.37	26.93	5	Pf04862	323	39.06	7.59
~45 kDa	Q9SUU6	Putative uncharacterized protein AT4g32460 OS=Arabidopsis	6.47	19.87	2	Pf04862	365	39.81	8.20

		thaliana GN=F8B4.160 PE=2 SV=1 [Q9SUU6_ARATH]							
~55kDa	Q9FMT6	Putative uncharacterized protein At5g14150 OS=Arabidopsis thaliana GN=At5g14150 PE=4 SV=1 [Q9FMT6 ARATH]	5.80	6.35	2	Pf04862	383	40.90	4.21

4.4.3. Co-Immunoprecipitation

Protein-protein interactions of DGR2 were assessed using Co-IP followed by mass spectrometry (MS)-based proteomic analyses. Total protein was extracted from wild type and subjected to co-immunoprecipitation (Co-IP) with anti-DUF642 antibody. The control experiment was performed without the antibody. Proteins bound to the matrix were eluted and used to perform one-dimensional SDS-PAGE. The total band obtained after running SDS-PAGE for a short period of time so that the protein bands should spread up to one cm in length. The bands from both the control and experimental samples were excised from the gel and analyzed by mass spectrometry. Several proteins were identified in both the control and experimental samples and were considered to be cross-reactive and were removed from the analysis. Two Golgi localized proteins involved in carbohydrate metabolism (beta-glucosidase (At3g18080) and beta-hexosaminidase (At1g65590), and an adenosylhomocysteinase (At4g13940) were detected in the sample (Table 4-4) with the anti-DUF642 antibody.

Table 4-4. Putative DUF642-interacting proteins identified by Co-IP and Mass spectrometry.

Accession	TAIR ID	Description	Number of unique peptides	Subcellular location
Q9LKR3	At3g18080	beta-glucosidase	3	Golgi apparatus, cell wall
Q8L7S6	At1g65590	beta-hexosaminidase	3	N-glycocylated in Golgi, located in cell membrane
O23255	At4g13940	adenosylhomocysteinase	2	Apoplast, plasmamembrane

4.4.4. Subcellular localization

In silico analysis (Section 2.2.7) predicted that DGR2 and DUFB both are secreted proteins and both proteins were identified in studies of the cell wall proteome (Bayer et al., 2006; Minic et al.,

2007; Irshad et al., 2008). Thus, evidence suggests that both proteins localize in the same cellular compartment. Therefore, we used only DGR2 for investigations of subcellular localization *in vivo*. Three different CiFP (citrine fluorescent protein) fusion constructs were generated using the vector pCAMBIA1303 and DGR2. These included a fusion construct with CiFP fused to the C-terminus of DGR2 cDNA (35S::DGR2:CiFP), a fusion construct with CiFP, in which the ER retention signal KDEL (Munro and Pelham, 1997) was attached to the C-terminus of CiFP and fused to the C-terminus of DGR2 cDNA, (35S::DGR2:CiFP:KDEL, and as a control, a fusion construct with CiPF cDNA (35S::CiFP). All three constructs were driven by the CaMV35S promoter. The subcellular localization of DGR2 was assayed *in vivo* using Agrobacterium-mediated stable expression in Arabidopsis (Fig 4-9 and 4-10) using seedlings from three independent lines of T2 generation plants and was analyzed by confocal laser-scanning microscopy.

In silico analysis predicted that DGR2 is a secreted protein (Section 2.2.5). To test whether DGR2 passed through the secretory pathway, a translational fusion was modified to include a carboxy-terminal endoplasmic reticulum (ER) retention signal, KDEL (35S::DGR2:CiFP:KDEL). As the presence of KDEL resulted in retention of this fluorescent signal in the endoplasmic reticulum (Fig. 4-9). When the KDEL signal was removed (35S::DGR2:CiFP), DGR2 was expected to be localized extracellularly, however, these transgenic plants expressing DGR2 fused CiFP showed revealed punctate intracellular fluorescence (Fig. 4-10). The punctate pattern was observed only in 35S::DGR2:CiFP plants but not in 35S::DGR2:CiFP:KDEL or 35S::CiFP (Fig. 4-9 & 4-10). Similar punctate patterns have been reported for Golgi localized proteins (Jensen et al., 2011; Scheuring et al., 2012; Gao et al., 2012).

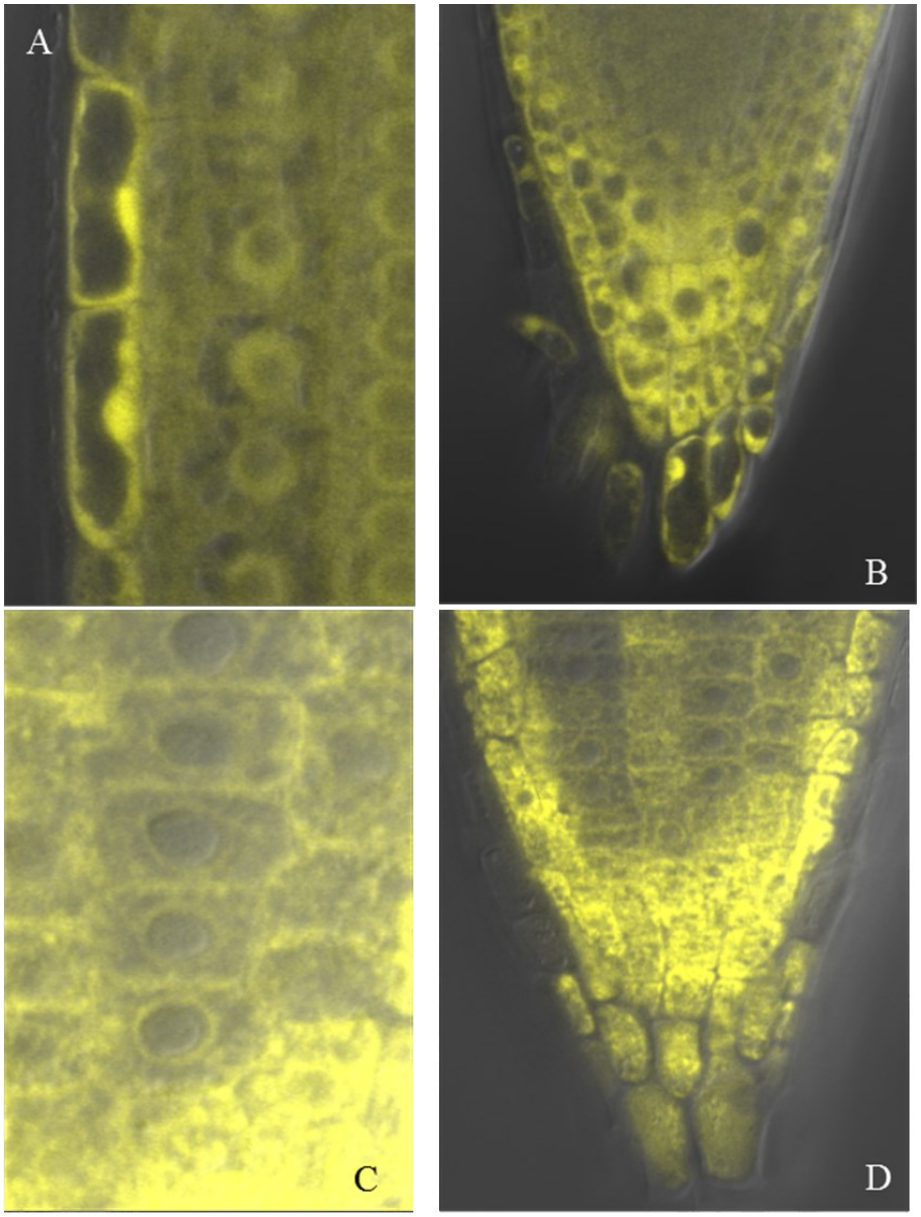


Figure 4-9. Subcellular localization of (A & B) 35S::CiFP and 35S::DGR2:CiFP:KDEL (C & D) by stable expression in *Arabidopsis thaliana*.

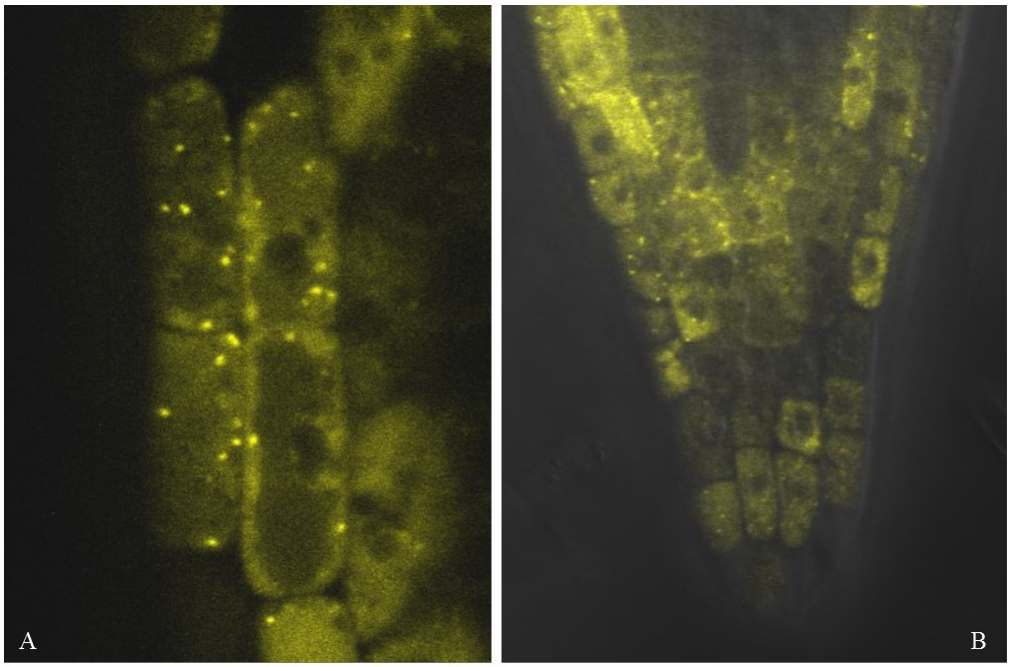


Figure 4-10. Subcellular localization of CiFP-tagged DGR2, 35S::DGR2:CiFP in Arabidopsis root tips.

4.4.5. Immunohistochemical analysis

To further analyze the subcellular localization of DGR2, we attempted to co-localize a Cy5-labeled anti- Sec21 antibody with the punctate fluorescent signal in 35S::DGR2:CiFP plants. Sec21 is an early Golgi (cis-Golgi) marker protein**.** A negative control was performed with 35S::DGR2:CiFP and 35S::CiFP fixed tissues without using a primary antibody. No Cy5 signal was detected in the negative controls, but the CiFP signal was detected in the positive control (results not shown). Immunostaining using wild type plants served as a positive control for Golgi staining. The positive control using wild type plants showed a clear punctate Cy5 signal from the Golgi (Fig. 4-11G). As shown in Figure 4-11(C and F), the fluorescent signals of anti-Sec21 and DGR2:CiFP did not overlap. On the other hand, DGR2:CiFP was found frequently in a position immediately adjacent to the anti-Sec21 signals (Fig. 4-12, A to F).

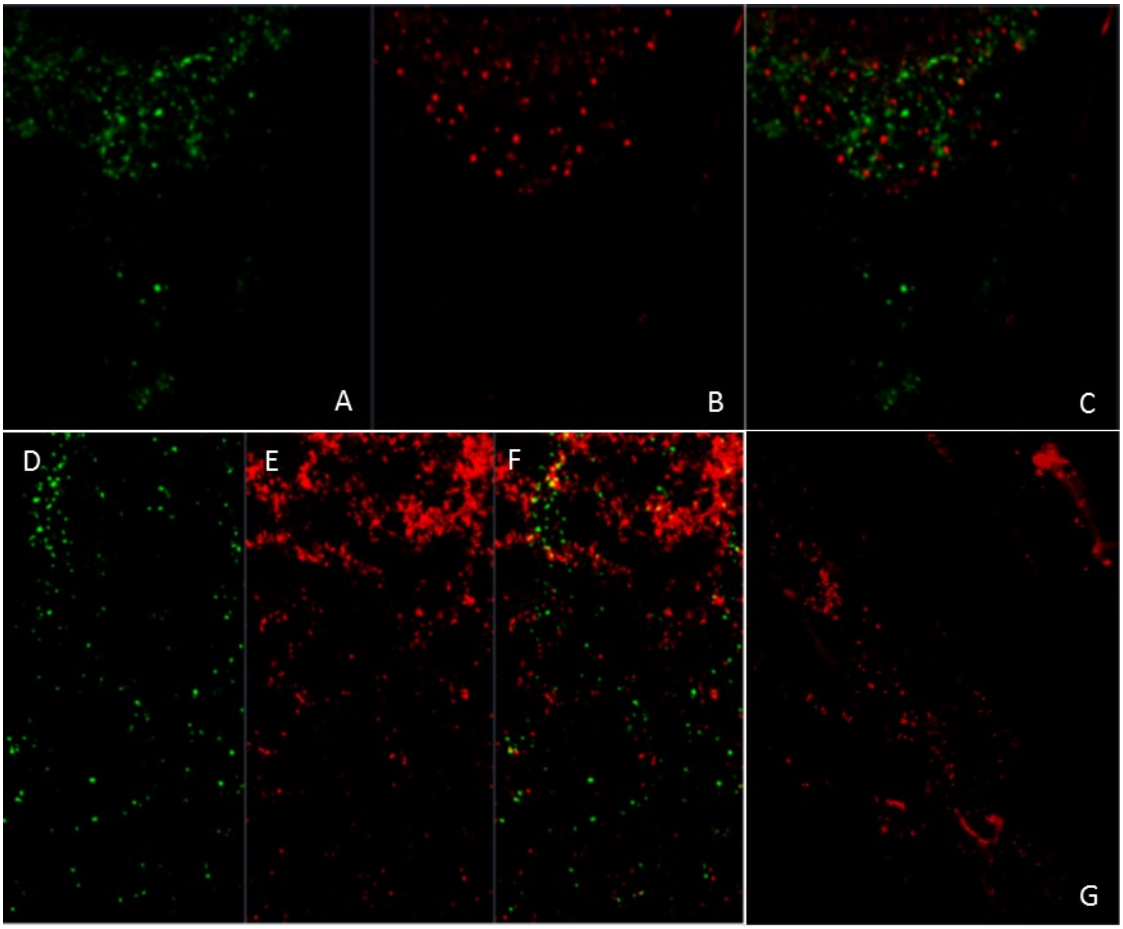


Figure 4-11. Colocalization of Golgi marker Sec21 and and CiFP-DGR2 Fusion. In root cells of transgenic Arabidopsis plants expressing DGR2 fusion was not colocalized with Golgi marker Cy5. CLSM images using CiFP specific filter (A, D); Cy5 specific filter (B, E); superimposition of CiFP and Cy5 (C, F); Control: WT plants with Sec 21 Golgi marker (G).

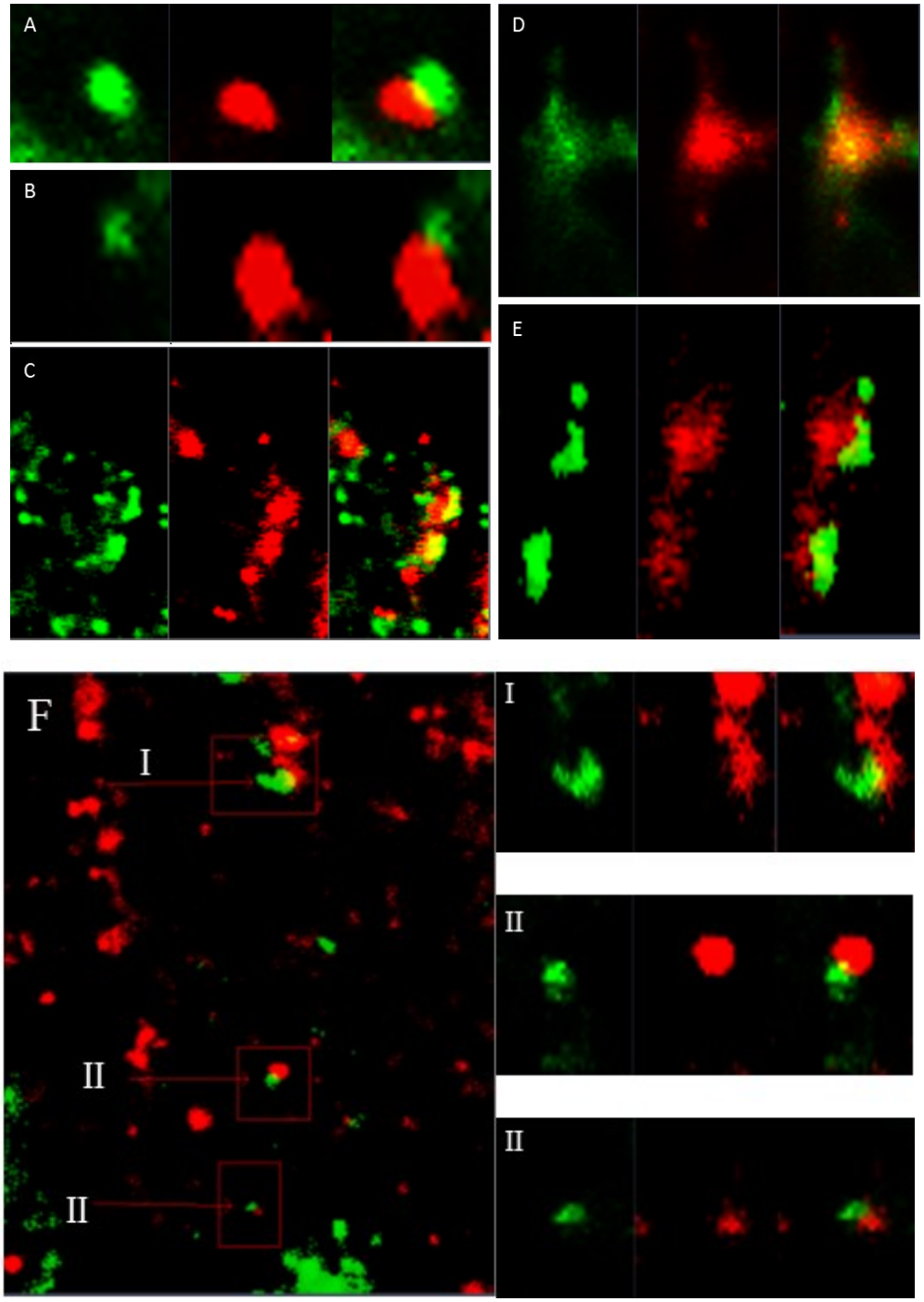


Figure 4-12. Colocalization of Golgi marker Sec21 and CiFP-DGR2 fusion. A, B, C, D, E and F showed CiFP-DGR2 and Sec21-Cy5 expression immediate opposition to each other and in root cells. I, II and III showed the co-localization of CiFP-DGR2 and Sec21-Cy5 in the root cell and the inset showed co-localization and the numbered boxes indicated regions magnified at right.

4.2.6. Heterologous protein expression

DGR2 and *DUFB* were cloned into expression vector pET22b+, which generated pET22b+DGR2 and pET22b+DUFB constructs. The plasmid contains a T7 promoter, Lac operon, N-terminal pelB signal sequence for periplasmic localization, multiple cloning sites and T7 terminator. At first the constructs were generated without removing the signal peptide from the CDS of *DGR2* and *DUFB*. No detectable target protein expression was obtained from these constructs. Subsequently, new constructs were generated in which the signal peptides were removed from both *DGR2* and *DUFB*. These expression constructs, named pET22b+DGR2 and pET22b+DUFB, respectively, were confirmed by restriction enzyme digestion and DNA sequencing. The recombinant plasmids were transformed into *E. coli* BL21(DE3) and BL21(DE3)pLysS E to optimize expression in liquid 2XYT medium. Initially, the expression of fusion proteins was induced with 0.5 and 1mM IPTG at 37ºC, 22ºC and 17ºC for 2 h, 4h, 6h and 24h. The maximum expression of the expected size (~40 kDa) was achieved after 4 h of induction with both 0.5 and 1mM IPTG. Among BL21(DE3) and BL21(DE3)pLysS, maximum expression was obtained from BL21(DE3)pLysS as detected by western blot (Fig. 4-13). The extracted proteins were reactive to rabbit-anti-DUF642 antibody and anti-His antibody as evidenced by western blot analysis (Fig. 4-14). The molecular weight of the recombinant protein is in accordance with the calculated molecular weight of recombinant DGR2 and DUFB (~40 kDa without the signal peptide and with the histidine tag.). No band was observed in the extract from the control pET22b+ (Fig. 4-14). Even after optimization of different factors such as temperature, medium, IPTG concentration and growth period, expression of His-tagged recombinant DGR2 and DUFB resulted in low protein yield.

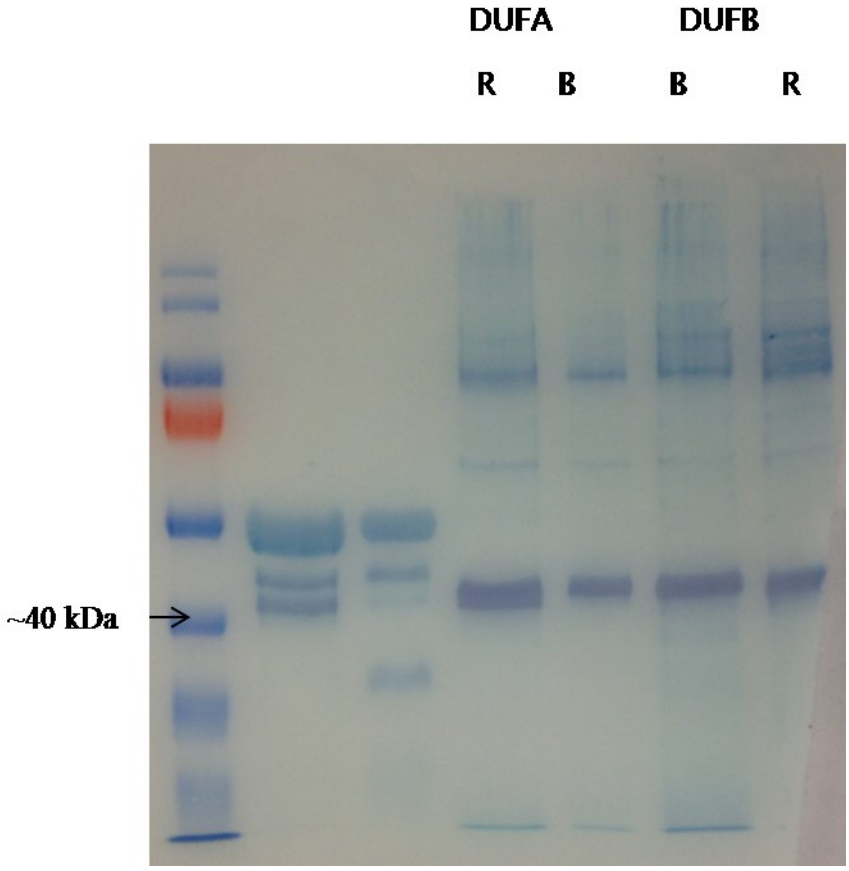


Figure 4-13. Expression of DGR2 and DUFB in B: BL21(DE3) and R: BL21(DE3)pLysS E. coli strains**.**

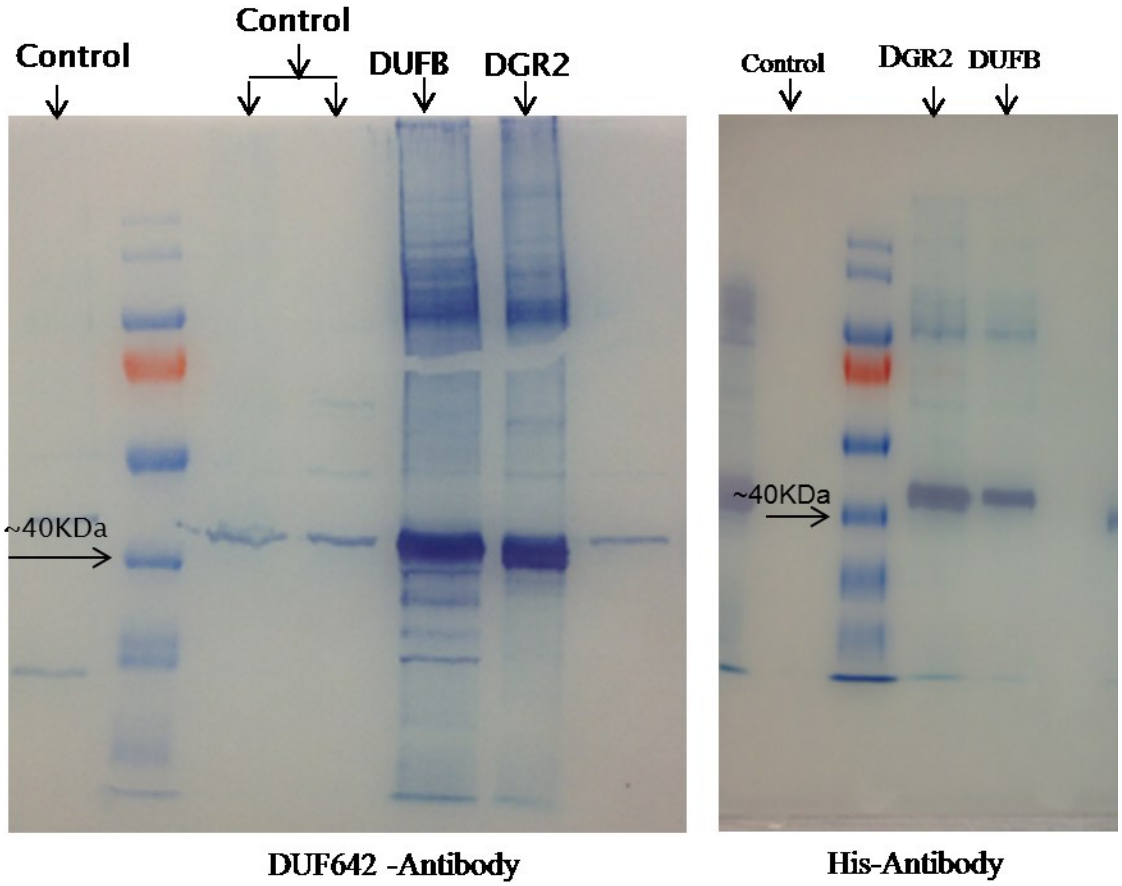


Figure 4-14. SDS-PAGE and western blot analysis of expressed DGR2 and DUFB using rabbit-anti-DUF642 antibody and anti-His antibody.

4.4.7. Protein purification

To examine functional properties of DGR2 and DUFB proteins, the recombinant proteins were purified using immobilized nickel affinity column chromatography (Ni-NTA) by exploiting the histidine tag. The purification process was based on the strong interaction of metal ions with artificial polyhistidine tags at the C-terminus of the protein of interest. Several attempts were made to purify the expressed DGR2 and DUFB but all attempts failed to recover purified protein (Fig. 4-15). Only a very small amount of the applied recombinant protein was retained and eluted with an imidazole-containing buffer as judged by SDS–PAGE analysis of various fractions. DGR2 and DUFB both expressions were very low even after optimizing various parameters. As a result, native *E. coli* proteins were co-eluted with the expressed protein even

after performed several washing steps. Identity of purified recombinant proteins was confirmed by peptide mass finger printing using Mass Spectrometry (Table 4-5).

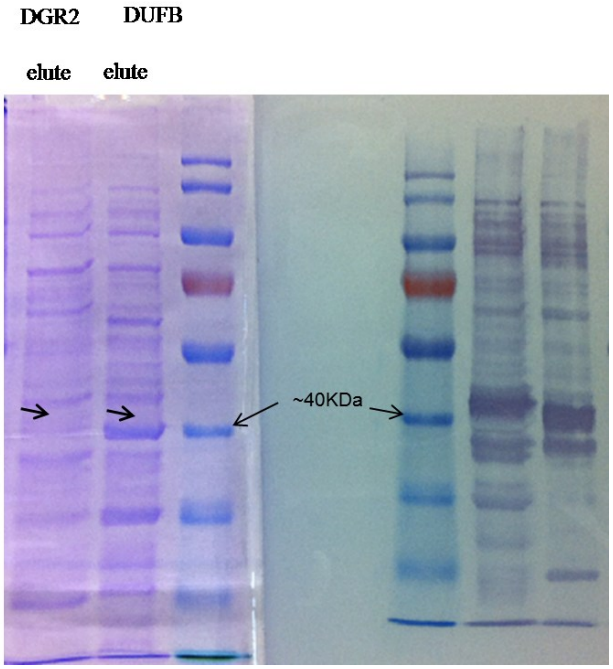


Figure 4-15. SDS-PAGE and western blot analysis of DGR2 and DUFB after purification.

Table: 4-5. Proteins identified by mass spectrometry; MS sample refers to the gel slice cut out from the polyacrylamide gel (Figure 4-15).

Accession	Description	Score	Coverage	Unique Peptides	# Peptides	# PSMs	Pfam IDs	# AAs	MW [kDa]	calc. pI
Q940Q9	**AT5g25460**/F18G18_200 OS=Arabidopsis thaliana PE=2 SV=1 - [Q940Q9_ARATH]	82.78	27.91	8	8	24	Pf04862	369	40.0	7.36
Q8H168	Putative uncharacterized protein **At5g11420** OS=Arabidopsis thaliana GN=At5g11420 PE=2 SV=1 - [Q8H168_ARATH]	53.62	37.16	8	8	16	Pf04862	358	38.5	6.46

4.4.8. Mutant analysis

To determine the effects of overexpression of *DGR2* and *DUFB* on transcript abundance, over-expression lines were developed using a constitutive CaMV35S promoter (Benfey, Chua 1990; Benfey et al., 1990; Bert et al., 1999; Jackson et al. 2002, Llyod 2003). Two independent T_3 lines were subjected to transcript analysis by qRT-PCR using primer pairs described in section 4.1.8. Results presented in Figure 4-16 demonstrated that the transcript levels in these lines were increased from 30-40 fold for *DUFB* (Fig. 4-16B) and 35 to 50 fold for *DGR2* (Fig. 4-16A) compared to wild type Col-0 plants, indicating that *DGR2* and *DUFB* were significantly up-regulated in these lines. None of the overexpression mutant lines showed morphological differences compared to wild type plant when grown under normal growth conditions. These lines were then used to conduct further biochemical analyses.

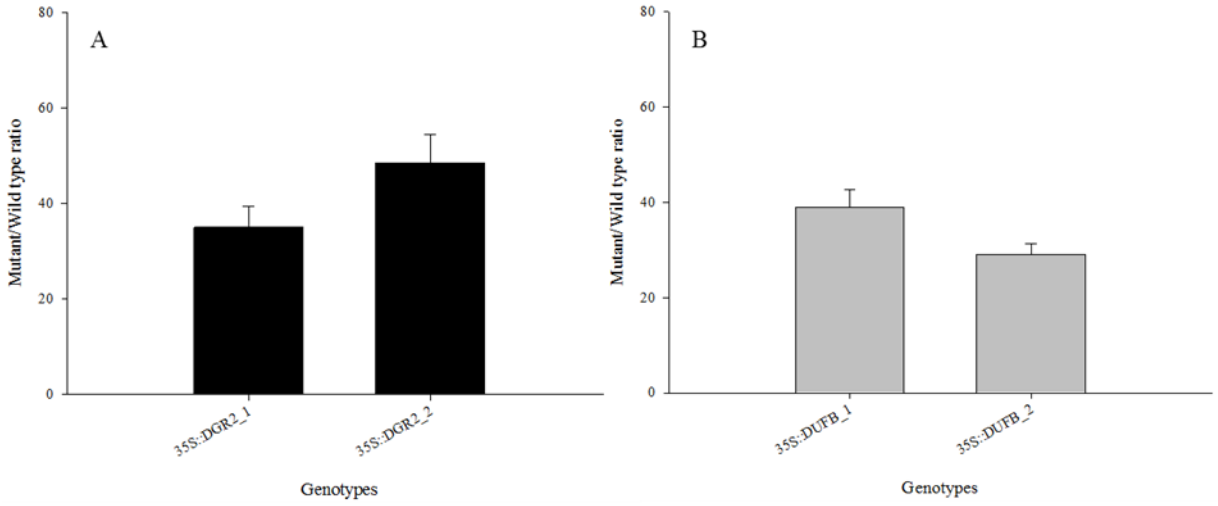


Figure 4-16. qRT-PCR analysis of the transcript abundance of mutants and WT plants. (A)Two independent lines of DGR2 (35S::DGR2-1 and 35S_DGR2-2) amplified with DGR2 specific primer mentioned in Section 4.1.8 and (B) Two independent lines of DUFB (35S::DUFB-1 and 35S_DUFB-2)) amplified with DGR2 specific primer described in Section 4.1.8. Plants were grown under the conditions in soil described in Section 4.1.1. Error bars represent standard deviations (SD) of four biological replicates and four technical replicates.

To evaluate loss-of-function phenotypes of *DGR2* and *DUFB*, an RNAi construct was produced that was targeted to *DGR2*. The 276-nucleotide targeting region within the RNAi construct, excluding the Gateway Recombination Sequence, shared 100% identity with DGR2 and 80%

identity with DUFB (refer to Materials and Methods). The T3 plants were subjected to transcript analysis by qRT-PCR using the primer pairs mentioned in Section 4.1.8. showed that transcript abundance in the RNAi lines for *DGR2* was between 10 and 20 % of the transcript abundance of wild type plants (Fig. 4-17A). Transcript abundance of *DUFB* also decreased but not as drastically as for *DGR2*, as determined by qRT-PCR which was 40% and 60% (4-17). These mutant lines didn't show any visual phenotypic differences compared to wild type plant when grown under normal growth conditions. These lines were then used to conduct further biochemical studies.

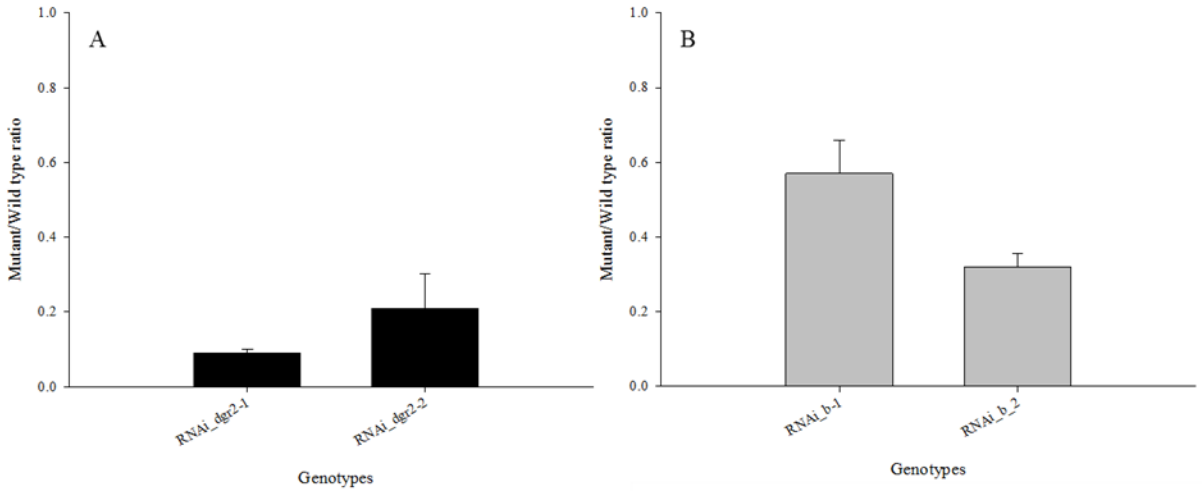


Figure 4-17. Quantitative real-time PCR (qRT-PCR) analysis of mutants and wiltdtype Col-0 plants. (A) Analysis of two independent RNAi lines(RNAi-dgr2, RNAi-dgr2-2) using DGR2 specific primers mentioned in Section 4.1.8 and (B) Analysis of two independent RNAi lines(RNAi-dgr2, RNAi-dgr2-2) using DUFB specific primers (RNAi-b-1, RNAi-b-2) mentioned in Section 4.1.8. Plants were grown under the conditions in soil described in Section 4.1.1. Error bars represent standard deviations (SD) of three biological replicates.

Homozygous lines bearing a T-DNA insertion in the exon of *DGR2* and in the promoter of *DUFB* were obtained from the Arabidopsis stock center, and the genotype was confirmed by PCR as described in the Section 4.1.5. These presumptive loss of function lines will be referred to as *salk_dgr2* and *salk_b* respectively.

4.4.9. Monosaccharide profiling

Isolation and evaluation of different cell wall polysaccharides could allow for an evaluation of a possible role of *DGR2* and *DUFB* in cell wall modification. For this analysis, T3 generations of mutant plants with two biological and three technical replicates were used. Seedlings were harvested at 7 DAS, including 35S::DGR2, 35S::DUFB, *Salk_dgr2*, *Salk_b*, RNAi plants and wild type Arabidopsis Col-0 plants grown under conditions described in Section 4.1.1. The non-cellulosic neutral monosaccharide composition of the cell wall matrix polysaccharides were obtained by treating de-starched alcohol insoluble residue (AIR) with trifluoroacetic acid and subsequent derivatization of the solubilized monosaccharide in to their corresponding alditol acetates followed by quantification by GC-MS (Foster et al., 2010). Figure 4-18 showed the monosaccharide composition of the cell walls of six different genotypes of Arabidopsis. Biochemical profiling of fucose, xylose, and mannose did not show significant differences between Col-0 and the five tested genotypes of *DGR2* and *DUFB*. The loss-of function and overexpression mutants of *DUFB* and RNAi lines showed a significant ($P>0.005$) increase in galactose content compared to Col-0 which were 1.78-fold, 1.15-fold and 1.16-fold, respectively. Glucose and arabinose increased significantly in *Salk_b* and RNAi lines compared to Col-0. However, no significant changes in the monosaccharide contents were observed in *Salk_dgr2* or 35S::DGR2 compared to Col-0.

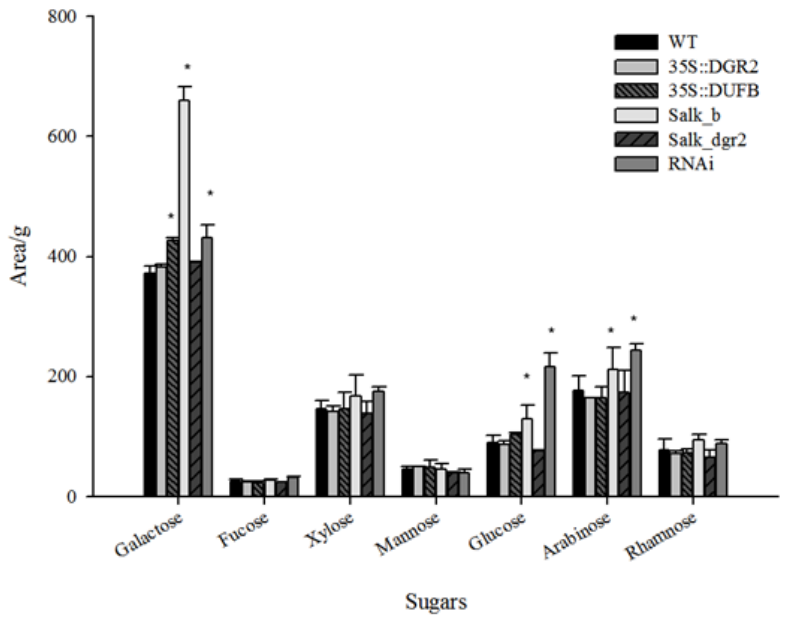


Figure 4-18. Monosaccharide composition of cell walls from six genotypes of Arabidopsis: WT, 35S::DGR2, 35S::DUFB, Salk_b, Salk_dgr2 and RNAi. . Data were shown as the means of three biological and three technical replicates with standard deviations illustrated as vertical bars The asterisk indicate P value<0.05.

4.4.10. Effects of auxin on root and hypocotyl length of DGR2 and DUFB mutants

To test the effect of auxin on root and hypocotyl length of DGR2 and DUFB mutant lines, seeds were germinated on vertically oriented growth medium supplementation with 100µM IAA. We selected these organs based on GUS expression analysis where we identified that by the application of exogenous auxin *DGR2* and *DUFB* transcripts were upregulated in the root apex. and also in the hypocotyls of *DGR2*pro:GUS seedlings. Root and hypocotyl lengths were measured after seven DAS using 30 seedlings from three different seedling plates of each mutant. The assay was repeated two times with similar results. None of the genotypes showed any difference when compared to Col-0 wild-type controls on normal medium. In the presence of IAA, root length of some *DGR2* and *DUFB* overexpression lines significantly (P<0.001) differed compared to wild-type (Fig. 4-19). Root growth of DGR2 over expression lines was inhibited drastically by IAA whereas root growth of 35S::DUFB was significantly higher compared to Col-0 and 35S::DGR2 grown on IAA medium. On the other hand, root length of *Salk_b* and *Salk_dgr2* and RNAi lines showed no significant differences in root growth compared to Col-0. Hypocotyl length was also significantly (P<0.001) affected by IAA (Fig.4-

19). Hypocotyl length of *Salk_dgr2*, RNAi and 35S::DUFB increased significantly compared to Col-0 (Fig. 4-19).

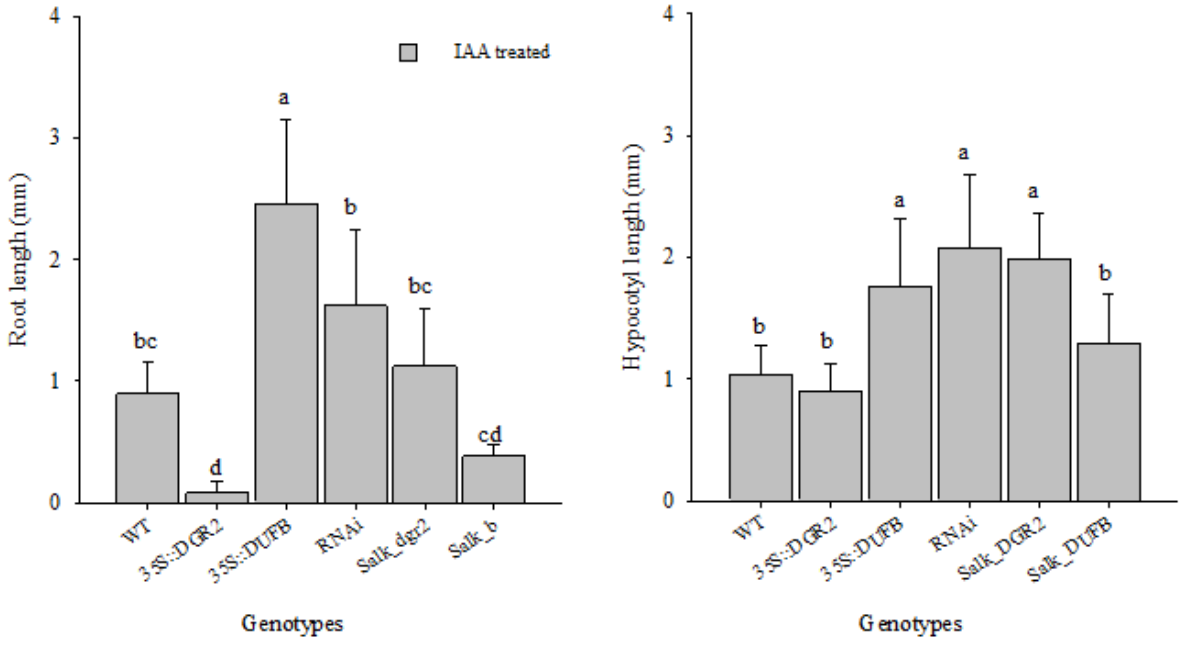


Figure 4-19. Root and hypocotyl growth on 100µM IAA for seven days. Data are means ± SD (n ≤15). Means followed by the same letter are not significantly different (P <0.001) according to Duncan Multiple Range Test.

4.5. Discussion

qRT-PCR data analysed by the delta delta Ct method showed that *DGR2* and *DUFB* transcripts were present in stems, siliques and rosettes but had highest abundance in the root and flower relative to stems (Fig. 4-6 & Fig. 4-7). This was generally consistent with the electronic fluorescent pictograph (e-FP) generated from public expression data for DGR2 and DUFB (Section 2.2.1). Protein expression analysis of DGR2 and DUFB using western blot analysis and an antibody that recognized several members of the DUF642 family showed up to four distinct bands in each tissue assayed (Fig. 4-8). Mass spectrometry identified DGR2, DUFB at the expected size near ~40 kDa, and a third DUF642 containing-protein (At3g08030) within these bands. Western blot signals were identified in all the tested tissues including rosettes, siliques and stems where mRNA of *DGR2* and *DUFB* were relatively low according to qRT-PCR (Fig. 4-6, Fig. 4-7 & Fig. 4-8). The real time quantitative PCR (qRT-PCR) and western blot analyses showed low levels of DUF642 mRNA (Fig. 4-6 & Fig. 4-7) and protein accumulations (Lane 1 and Lane 2 in Fig. 4-8) in the stem tissues. Previous studies reported that the correlation between mRNA and protein abundance was not strong (Nie et al., 2006, Futcher et al. 1999; Gygi et al. 1999; Ideker et al. 2001;Greenbaum et al. 2003; Washburn et al. 2003) because of three possible reasons: (i) translational regulation, (ii) differences in protein in vivo half-lives, and (iii) the significant amount of experimental error, including differences with respect to the experimental conditions (Greenbaum et al. 2003; Beyer et al. 2004; Nie et al., 2006). A weak signal was detected at ~38 kDa in the bottom of the stem (Fig. 4-8, Lane 1) compared to top of the stem (Fig. 4-8; Lane 2), the size of the signal where DGR2 and DUFB proteins were identified. This weak signal can be caused by low concentration of antibody or DUF642 antigen. This result was in agreement with the result of De Pauw et al. (2007) in which they identified that DUF642-containing genes enriched in the top of hemp stems compared to bottom of the stems. Expression data from e-FP browser also showed that these genes expressed highly in the top of the stem compared to bottom of the stem (Section 2.2.1). The top of the stem is a region where metaxylem, cell elongation to cell wall thickening (Gorshkova et al. 2003) takes place. At ~40 kDa, strong signal intensity was obtained in the 6 DAS (Fig 4-8; Lane 9) and 13 DAS dark grown hypocotyls (Fig 4-8; Lane 9). In 3-6 DAS dark grown hypocotyls, cell elongation occur exponentially compared to 10 DAS dark grown hypocotyls, where cell elongation almost ceased (Kudo and Mii, 2004). Contrary to these findings, in GUS study (Chapter 3), no GUS staining

was observed in the dark grown hypocotyls which was also confirmed by qRT PCR assay where mRNA transcripts of both genes were almost undetectable. There could be two possibilities, one of that western blot signals were from the DUF642 protein other than DGR2 and DUFB because the DUF642 antibody was generated against the synthetic peptide that was found in two other DUF642 including DGR2 and DUFB or because poor correlation between mRNA and their coding to protein for example due to posttranscriptional modification of protein or half life of the protein..

DGR2 was identified in the cell wall extracts in plant proteomics analysis (Bayar et al., 2006) but in the present study, confocal images of DGR2 CDS tagged with CiFP, subcellular localization study reveal punctate patterns instead of in the cell wall (Section 4.2.4). Punctate pattern was very typical for Golgi localized protein. Numerous enzymes for cell wall biosynthesis are located in the Golgi apparatus and plasma membrane (Oikawa et al., 2013). To test whether DGR2 was localized in the Golgi, colocalization study was performed using the Golgi marker sec 21 (Section 4.2.5.). Sec21 is a gene required for ER to Golgi protein transport and can be used as an early (cis) Golgi marker but not for trans Golgi (Tanaka et al., 2014; Levi et al, 2010). The results showed that DGR2 was not colocalized with the early (cis) Golgi marker, Sec21 but revealed that DGR2 was in the immediate adjacent to Sec21 (Fig. 4-11). Therefore, it is possible that DGR2 was associated with the trans-Golgi network (TGN) which is considered to be part of the Golgi apparatus and located within a ribosome-excluding Golgi matrix (Moore and Staehelin, 1988). Cai et al. (2011) in their study found that the rice secretory carrier membrane protein 1 (SCAMP1) which localized to the PM and trans-Golgi network (TGN), were largely separated from the Golgi marker Man1–RFP but co-localized with the TGN marker RFP–SYP6. TGN is a highly mobile organelle that can be closely associated with a Golgi or another TGN and can be located at a distance independently. (Viotti et al., 2010). TGN in plants is related to multiple functions which include correctly pack and transport newly synthesized proteins and carbohydrates to vacuoles or extracellular domains (Richter et al., 2009). TGN also functions as the early endosome and received the endocytosed material from the plasma membrane (Viotti et al., 2012; Dettmer et al., 2006). Present results suggested that DGR2 localized in the trans-Golgi network (TGN).

Co IP result (Table 4-4) showed that these proteins co-eluted with beta-glucosidase (At3g18080), which belongs to the glycosyl hydrolase family 1. *In silico* characterization (Chapter 2) of DGR2 and DUFB indicated that DGR2 and DUFB are polysaccharide hydrolysing enzymes. Co-IP result again suggested that DGR2 and DUFB are involved in the degradation of cell wall polysaccharides.

In this study, DGR2 and DUFB were heterologously expressed in *E. coli* but the expression levels were very low (Section 4.2.6.). The low level of expressions made it difficult to obtain purified proteins. Like other *E. coli* strains, BL21(DE3) and BL21(DE3)pLysS contains a number of host proteins that are rich in nonconsecutive histidine residues. These histidine-rich proteins co-purify during IMAC procedures rendering recombinant protein preparations impure, because DGR2 and DUFB were expressed at very low levels. The lower level of expressions of DGR2 and DUFB could be due many reasons such as protein stability, mRNA degradation or expressed proteins were toxic to the host organism.

DGR2 and DUFB expression analysis in Chapter 3 showed that both genes were responsive to auxin. Based on these observations, effects of exogenous auxin on root and hypocotyl lengths of DGR2 and DUFB were investigated (Fig. 4-19). In this study, loss and gain of function mutants were grown on ½ MS medium supplemented with 100mM auxin for 7 days. The results showed that root growth of wild type and 35::DGR2 were reduced significantly. Moreover, reduction of root length of DGR2 was even significantly lowered compared to wild type plants. On the other hand, root length of DGR2 was not reduced drastically by IAA treatment. The response of roots to exogenous auxin was inhibited elongation of root growth, but the mechanism of root growth inhibition was poorly understood (Tanimoto 2005). The findings from the present study indicated that DGR2 and DUFB are involved auxin mediated root development but through different mechanisms. It was known that some concentrations of auxin stimulate cell division, cell expansion and elongation (Perrot-Rechenmann, 2010). During cell expansion and elongation, cell wall remodeling occurs by modifying cell wall polysaccharides (Perrot-

Rechenmann, 2010, Cosgrove 2000; Kende et al. 2004; Sampedro and Cosgrove 2005). During these processes many cell wall proteins activated that included expansins, xyloglucan endotransglycosylase/hydrolases (XTHs) and endoglucanases. DGR2 and DUFB predicted to be cell wall proteins with hydrolase and galactose binding domain like fold. Therefore, it could be possible that DGR2 and DUFB are involved in cell wall remodeling during cell expansion and elongation and regulated by auxin.

Monosaccharide profiling of *DGR2* and *DUFB* mutant cell walls showed (Fig. 4-18) that the content of galactose was significantly increased in 35S::DUFB and Salk-b plants whereas glucose and arabinose was decreased significantly in *Salk_b* and RNAi plants compared to wild type plants. No significant changes in the cell wall monosaccharide contents were detected in *DGR2* mutant lines. Our data indicated that *DGR2* and *DUFB* functions in different mechanism during the cell wall modification.

In summary, DGR2 and DUFB expression was detected in all major plant organs. A punctate pattern of *in vivo* subcellular localization was revealed instead of predicted cell wall localization. The punctate pattern is typical for Golgi localization. Co-localization study with Golgi marker Sec 21 revealed that DGR2 and DUFB possibly localise to the trans-Golgi Network. DGR2 and DUFB could be expressed heterologously in *E. coli* but the low level of expression made it difficult to purify them using IMAC. An effect of auxin was observed on root and hypocotyls growth but the mode of action was not clear. Significant decrease in cell wall galactose, arabinose and glucose was identified in the *DUFB* mutant lines indicating changes in the expression levels of *DUFB* exerted on the cell wall polysaccharides.

5. Chapter 5

5.1. Metabolic profiling of different DUF642 genotypes in Arabidopsis

5.2. Introduction

Metabolites are the substrates and products of enzyme-catalyzed reactions. The term metabolites is restricted to small molecules (up to ~1000 Da), such as amino acids, lipids, and carbohydrates and does not include the proteins and nucleotides that are the concern of proteomics, genomics and transcriptomics. Primary metabolites are essential for growth, development and reproduction. Secondary metabolites are not directly involved in these processes. Examples of secondary metabolites include antibiotics and pigments.

Metabolomics is an emerging "omics" science that has tremendous potential for contributing to biology (Glassbrooket al., 2000; Roesnner et al., 2001). Metabolomics can generally be defined as the study of global metabolite profiles in a system (cell, tissue or organism) under a given set of conditions (Goodacre et al., 2004). Nicholson et al. (1999) defined metabolic profiling as "measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification". Metabolic profiling requires combined application of sample preparation, spectroscopic techniques, data acquiring, multivariate statistical analysis and biological interpretation of the final results. Sophisticated statistical algorithms can be used to compare the enormous data generated to identify similarity and differences such as Principal Component Analysis (PCA). This analysis can assist in identifying the function of unknown genes in the context of different genetic backgrounds.

Different platforms are available for metabolic profiling, but none of them are able to describe the total metabolites of the cell simultaneously. It is important to select a platform that is sensitive, reliable and fast and able to recover significant amount of metabolites. The most widely used technologies for metabolite identification are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). NMR has lower

throughput and lower sensitivity than MS and is more suited for providing structural information of the compounds. On the other hand MS, and especially gas chromatography coupled to electron-impact quadrupole mass spectrometry (GC/MS) considered one of the most mature metabolomics techniques, which is capable of detecting a large number of metabolites with high reliability (Fiehn et al., 2000b).

One of the challenges in metabolic profiling is data processing. Raw metabolomics data must undergo preprocessing before conducting any advanced statistical analysis. This includes comparisons between replications, noise reduction, and identification of peaks and compounds. Various software packages are available for this data processing, such as AMDIS software (Davies, 1998; Stein, 1999) for peak deconvolution and separation. Compounds can be identified by matching the spectra to a reference library. The most widely used library is the National Institute for Standards and Technology (NIST) database. This database is developed and maintained by the National Institutes of Health (NlH) and the U.S. Environmental Protection Agency (EPA).

Enormous progress has been made in the development of tools to identify the functions of genes in plant systems. Most prominent among these techniques are the generation of transgenic and mutant plants to the parallel analysis of mRNA profiling (Baldwin et al., 1999) and proteomics approaches (Santoni et al., 1998). The limitation of these methods in plant functional genomics is that they do not provide any direct information about how a change in mRNA or protein is coupled to a change in biological function (Fiehn et al., 2000b). The systems biology approach powered by the integration of transcriptomics, proteomics and now metabolomics, has provided a new framework for discovery. Previous studies on DUF642 genes shown how these gene products interact with cell wall polysaccharides and pectinmethylesterase *in vitro* and could be identified in the cell wall protein extracts (Boyer et al., 2006; Irshad et al., 2008; Vanquez-Lobo, et al., 2012; Sanchez and Buen, 2012,). Nevertheless, the functions of this gene family are still unknown.

Plant cell walls are composed of four major building blocks: cellulose, hemicellulose, lignin and pectin. They also contain many proteins and glycoproteins that include various enzymes and structural proteins (Rose and Lee, 2010). The building blocks of cell walls are synthesized through primary metabolism. Carbohydrates are first formed in the process of photosynthesis. Other chemical constituents are formed through different metabolic pathways such as glycolysis, citric acid cycle, amino acid synthesis. Therefore, metabolic phenotyping could provide clues to the functions of the predicted cell wall protein family, DUF642.

In this study, GC/MS was used to profile metabolites of Arabidopsis plants that varied in the abundance of transcripts of DUF642 family genes.

5.3. Materials and Methods

5.3.1. Plant material and growth conditions

Arabidopsis thaliana seeds were stratified for 48h at 4°C and were grown in petri-dish (15 X 90 mm) containing half-strength Murashige-Skoog (MS) basal salts (Caisson, MSP01-50LT) with 1% agar (Sigma, A1296-1KG) and 1% sucrose. The following T-DNA insertion lines were obtained from the Arabidopsis Biological Resource Center (www.arabidopsis.org): SALK_0442864C (*Salk_dgr2*) and SALK_094931C (*Salk_b).* Homozygosity was verified by PCR for Salk_dgr2 using the primers LP TTTTCAGACAATTGGCGAGAG and RP AAGTGGTTCGGACTGTCATTG and for Salk_b using LP TTTGTTTTGTGGCTATCGAGC and RP CATGAGTTGGCATTTGTGTTG and with insert-specific primer (LBb1.3: ATTTTGCCGATTTCGGAAC). For the construction of RNAi lines standard molecular biology (Sambrook et al., 1989) and Gateway Recombination Technology (Invitrogen, Helliwell and Waterhouse 2003) were used. A 276bp fragment of DGR2 amplified from the genomic DNA of Col-0 and was incorporated into Gateway®-compatible vector, pHellsgate-12. Overexpression lines of DGR2 and DUFB were generated using pCAMBIA 1303 containing 1.17kb *DGR2* and 1.1 kb *DUFB* cDNA under the control of

CaMV 35S promoter. These constructs were introduced into WT plants using the *Agrobacterium tumefaciens*-mediated floral dip method. Detailed methods of generating transgenic plants were described in Chapter 4.

5.3.2. Sample extraction

In this experiment, three biological replicates from each treatment were used for extraction and analysis. For transgenic lines, 35S::DGR2 and 35S::DUFB, two different lines and three biological replicates of each line were used. For extraction of metabolites, three days old Col-0 and transgenic DUF642 seedlings were collected and lyophilized then homogenized by ball mill. Aliquots of frozen powder (~5mg) were extracted using the extraction method described by Fiehn (2006). According to this method, an extraction mixture of chloroform (HPLC grade, Sigma: 366927), methanol (HPLC grade, Fisher chemical: CAS 67-56-1) and water (HPLC grade, Fisher chemical: CAS7732-18-5) was prepared at a ratio of 1:2.5:1 (v/v/v) and degassed using an ultrasonicator. In each sample, 1 ml of cold extraction solvent was added (-15°C and degassed). Ribitol (Sigma, A5502-5G) was added as an internal standard for normalization. Samples were then shaken in a rotary shaker for 10 minutes at 4°C and centrifuged for 2 minutes at 14 000 rpm. The supernatant transferred into new tubes then 400 µl of pure water was added to each samples and vortex for 10 s. The upper phase of each sample was collected as a 'polar phase' (mixture of methanol and water) in new microfuge tubes for derivatizaton. The polar phase was dried using speed vacuum concentration for 1 h and then lyophilized overnight.

5.3.3. Derivatization

The extracted metabolites samples were derivatized as described by Fiehn (2006). To each lyophilized samples, 20 µl of methoxyamine solution was added which contained 20 mg methoxyamine hydrochloride (Sigma, 226904-1G) per ml in pyridine (Sigma, 270970-100ML). Samples were then shaken for 90 minutes at 28°C followed by centrifugation for 30 s at 14000 rpm. Silylting agent *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane (Sigma, 69478-1ML-F) was added at a volume of 90 µl to each sample and was shaken for 30 minutes at 37°C.

Samples were then transferred to inserts (Agilent, 5181-1270) containing glass vials for the GC/MS.

5.3.4. GC/MS analysis

Samples were analyzed using an Agilent 7890A gas chromatograph; an Agilent 5975C MSD with Electron Impact (EI) ion sources; an Agilent 7693 autosampler*;* a ChemStation instrument control and data handling system. For GC/MS analysis, derivatized sample extract (1 µL) was injected using splitless mode and splitless single taper liner with deactivated glass wool (4 mm ID) on to a 30 m length, 0.250 mm diameter and 0.25 µm film thickness DB-5MS capillary column (Agilent J & W GC Columns). The carrier gas was helium and flow rate was 1 ml/min. The initial oven temperature was 70°C and after a 10 minutes solvent delay, oven temperature was increased to 76°C at 1°C/min. From 76°C, oven temperature was increased to 320°C at 6.1 °C/min for a final run time of 62 minutes. The analysis condition was shown in Table 5-1. Total ion chromatogram (TIC) of a sample that was run by this method shown in Fig. 5-2. Autotuning was performed using Perfluorotributylamine (PFTBA), with m/z of 69, 219 and 502. Autotuning revealed the air/water check in the system which was H_20~0.08% N_2~0.68%, O_2~0.29% CO_2~0.03% and N_2/H2O~875%. To minimize the carryover effects, liners were replaced every 15 injections and background was monitored by running blanks after every five samples and before any sets of treatment samples. To monitor the system performance, a known concentration of succinic acid and malic acid samples were run at the beginning and at the end of the experiment. An alkane standard which was a mixture of 75 µl of C10, C20-C40 (Sigma, 94234-2ML) and 25 µl of C21-C40 (Sigma, 04070-1ML was run before running the samples to test the method performance (Figure 5-1).

Chemstation software (Agilent Technologies) operated the system and validated chromatogram and spectrum output. Specific mass spectral fragments were detected in defined retention time windows using NIST MS search 2.0. Mass spectral matching with the NIST library was conducted manually and matches accepted with threshold of match >650 (maximum match is equal to 1000**).** Metabolite peak assignment was reconfirmed

by using the automated mass spectral deconvolution and identification system (AMDIS) and the National Institute of Standards and Technology (NIST) library (version 98). Peaks that were at least 70% pure were automatically assigned identity.

Table 5-1. Analytical conditions of GC/MS.

Instruments	
GC-MS	Agilent 7890A Gas Chromatography with 5975C Mass Selective Detector (GC-MSD)
Auto-sampler	Agilent 7693
Column	DB-5MS capillary column (Agilent J & W GC Columns)
Liner and syringe	Liner 4mm ID LPD (Part#5188-6568) and Agilent Autosampler Syringe (part 39302-0713)
GC	
Flow	1ml/min
Pressure	8.8085 psi
Column temperature	70 °C (1 min) – (1°C /min) - 76°C (7 min)-(6.1°C /min)-320°C (62 min)
Injection mode	Splitless
Carrier gas	He (Constant Linear Velocity)
Average velocity	36.796 cm/sec
Heater temperature	200°C
Injection volume	1 µL
Septum purge flow	3 ml/m
MS	
Source Temperature	230°C
Quad temperature	150°C
Aux-2 temperature	280°C
Electron energy	70 eV
Scanning mass range	50-550 amu

Scan speed	Normal (6.61/sec)
Solvent delay	10 minutes
EMV mode	Gain factor; Gain factor = 1 (1341V)
Time windows	65 minutes

5.3.5. Data analysis

For individual metabolites, means ±SD were calculated from three biological replicates. Metabolite contents were calculated for individual metabolites obtained from GC/MS selected ion chromatogram by means of auto integration (ChemStation integrator, threshold=15) of peak area of each metabolite. Metabolites peak areas were divided by peak area of the internal standard, ribitol, to correct any recovery differences. The corrected peak areas were then normalized by dividing with dry weight of the samples and expressed as area/g. One-way ANOVA was performed by SAS/STAT software (version 9.1; SAS Institute, NC). Multivariate statistical analysis, PCA and HCA was performed by XLSTAT2014 statistical software of $\log_{10}$ transformed data of 52 metabolites obtained from six different genotypes of DUF642.

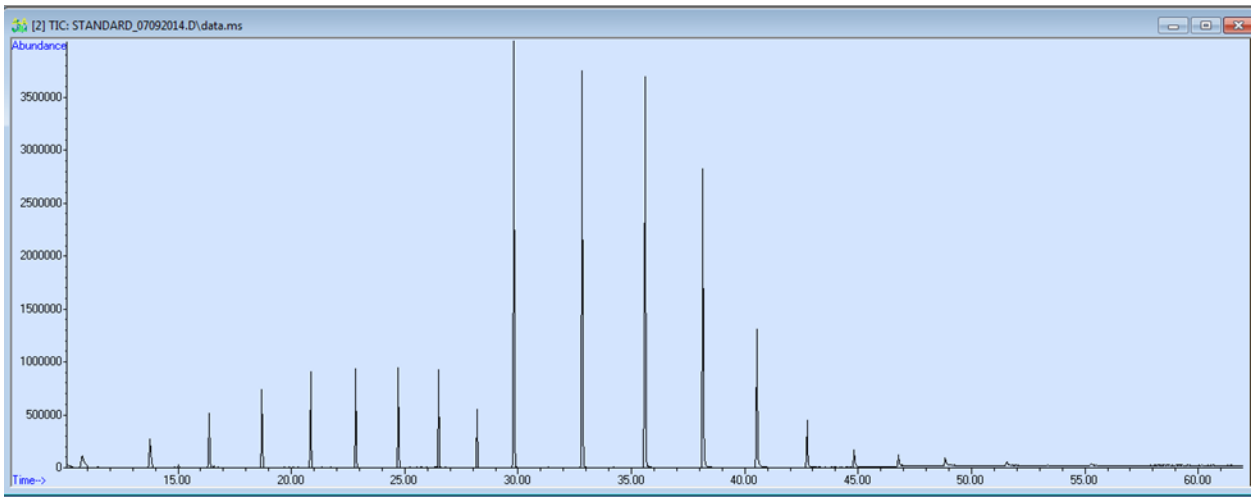


Figure 5-1. GC/MS total ion chromatogram of the alkane standard mixture.

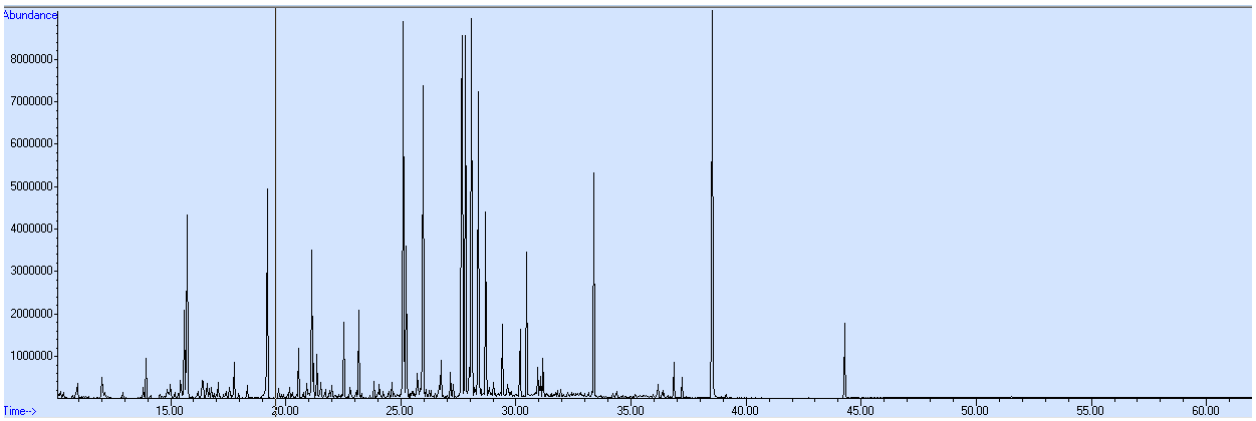


Figure 5-2. A GC/MS total ion chromatogram of transgenic Arabidopsis plants of DUF642.

5.3.6. Metabolite pathway analysis

Polar metabolites identified in the mutants to have significant differences compared to Col-0 were subjected to pathway analysis using MetaboAnalyst 2.0 (Xia et al., 2012) server. In the MetaboAnalyst 2.0, pathway analysis is performed by using impact and p-value scores. The software provided a fit coefficient (p) from the pathway enrichment analysis and an impact factor from the pathway topology analysis for each analyzed pathway. A p-value less than 0.05 and an impact score greater than 0.1 were considered to be significant. The impact score calculated the importance of the metabolites found in the pathway. It was calculated based on the sum of impact scores for each metabolite identified in a pathway, which are based on the importance of the metabolites to each given pathway.

5.4. Results

To understand the metabolic consequences of manipulating transcript abundance of DUF642 family genes in Arabidopsis, we conducted metabolic profiling of 3 DAS mutant and wild-type (Col-0) plants, using GC/MS. The 3 DAS time point was selected for metabolic profiling because previous studies showed this is when DUF642 expression is first detectable. We tested two different members of the DUF642 gene family: At5g25460 (DGR2) and At5g11420 (DUFB), as either loss of function EMS mutants (Salk-dgr2), (Salk-b), or as a transgenic loss-of-function RNAi suppression line (RNAi),

or as transgenic overexpression lines (35S::DGR2, 35S::DUFB). The characterization of transcript expression in these lines is described in Chapter 4.

5.4.1. Metabolite composition

In total, 47 metabolites were identified by GC/MS in the six different genotypes tested. The total number of metabolites and their contents varied across the genotypes (Table 5-2 and Fig. 5-3 and 5-4). Among the 47 metabolites, 38 metabolites were identified in all six genotypes (Table 5-2). Metabolite abundance was calculated for individual metabolites by means of auto integration of the peak area and normalization of the peak area with an internal standard. The normalized peak area was expressed relative to mass of starting dry weight. Therefore, metabolite content is here expressed as area/g. The identified metabolites were grouped into four compound classes, amino acid, organic acids, monosaccharide and other organic compounds. In the following paragraphs, we compare the metabolite profiles between genotypes.

Table 5-2. Identified polar metabolites in six genotypes of DUF642 mutants in Arabidopsis.

Genotypes	**number of metabolites**	**number of unique metabolites**
35S_DGR2	42	42
35S_DUFB	44	44
Col-0	44	44
RNAi	43	43
Salk_b	42	42
Salk_dgr2	42	42
Overall number of unique metabolites		**47**
Names	**total**	**elements**

35S_DGR, 35S_DUF, Col-0, RNAi, Salk_b, Salk_dgr2	38	Ribose, Succinic acid, Asparagine, Alanine, Glutamine, 2-Ketoglutaric acid, Butanoic acid, Cysteine, Ribofuranose, N-Acetylglutamine, Galacto-Hexodialdose, Pyruvic acid, Malic acid, Threonine, 2-Furanone, Inosose, Valine, Fumaric acid, Ornithine, Arabinose, Threonic acid, Gulonic acid, Norvaline, Glucose, Proline, Pyrazine, Galactose, Fructose, Gulose, Isoleucine, Phosphoric acid, Butane,Thymine, Erythrose, Ala-Thr, Lysine, Mannose, Serine
35S_DGR2, 35S_DUFB, Col-0, Salk_b, Salk_dgr2	1	Xylose
35S_DGR2, Col-0, RNAi, Salk_b, Salk_dgr2	1	Glycine
35S_DUFB, Col-0, RNAi, Salk_b, Salk_dgr2	2	N-Acetyl-L-Lysine, Aspartic acid
35S_DGR2, Col-0	1	Glutaric acid
Col-0, RNAi	1	Ribo-hexos-3-ulose
35S_DGR2, 35S_DUFB	1	Erythro-Pentopyranose
35S_DUFB, RNAi	1	Threitol
35S_DUFB	1	Fucose

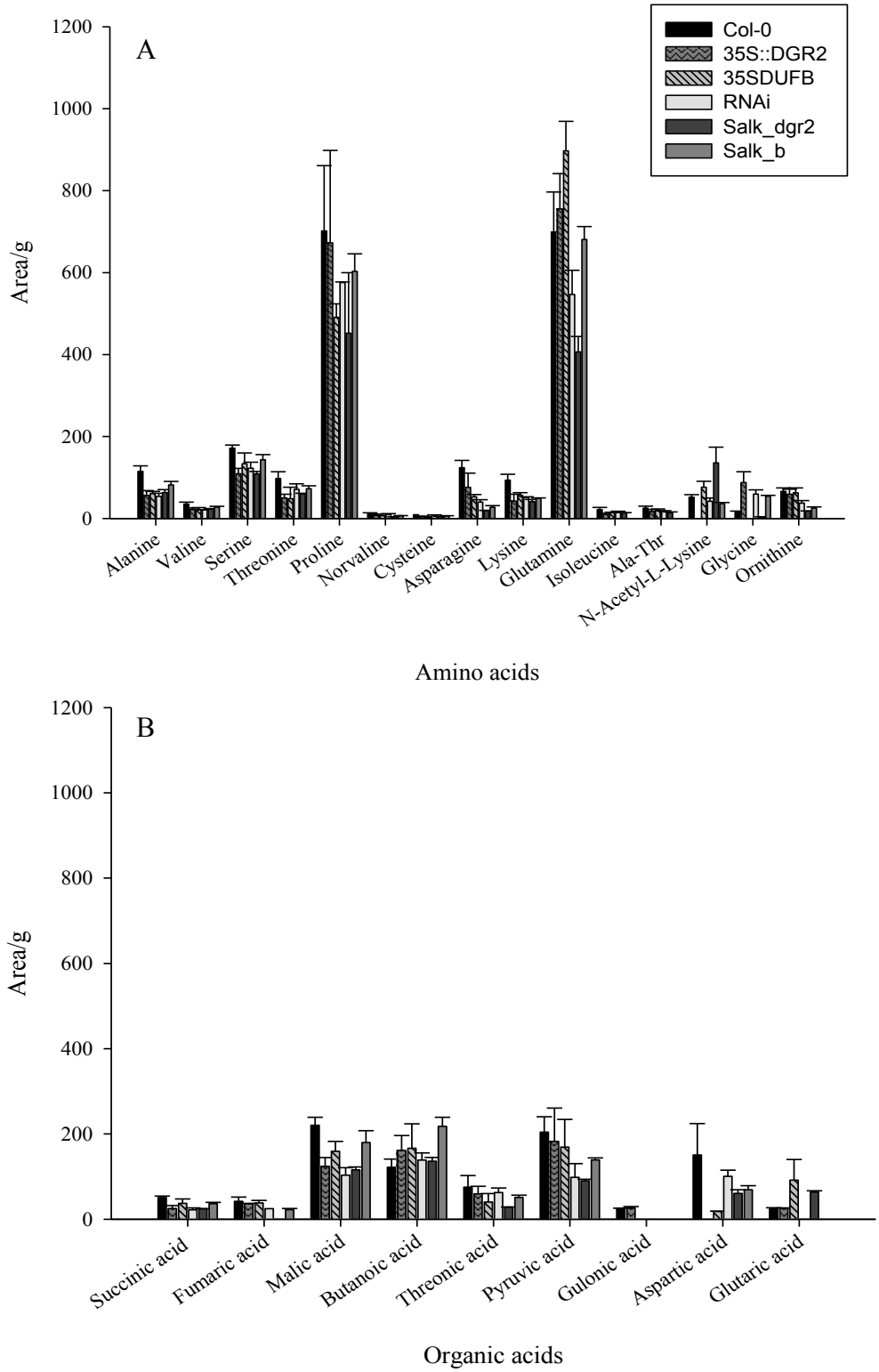


Figure 5-3. (A) Amino acids and (B) organic acids in six genotypes of DUF642 in Arabidopsis. Samples were from 3 DAS whole plants. The values were mean of three replications ± standard deviations..

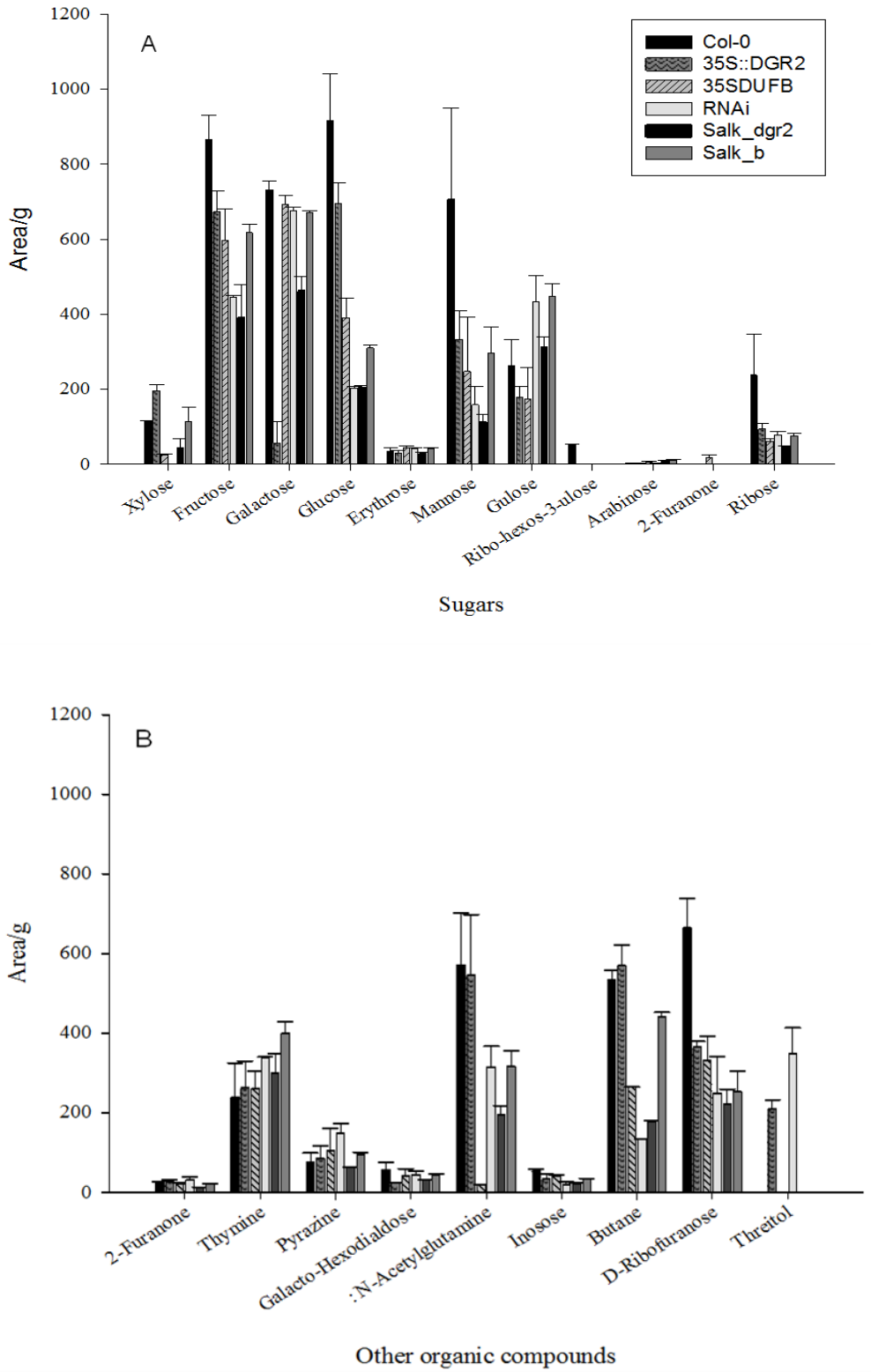
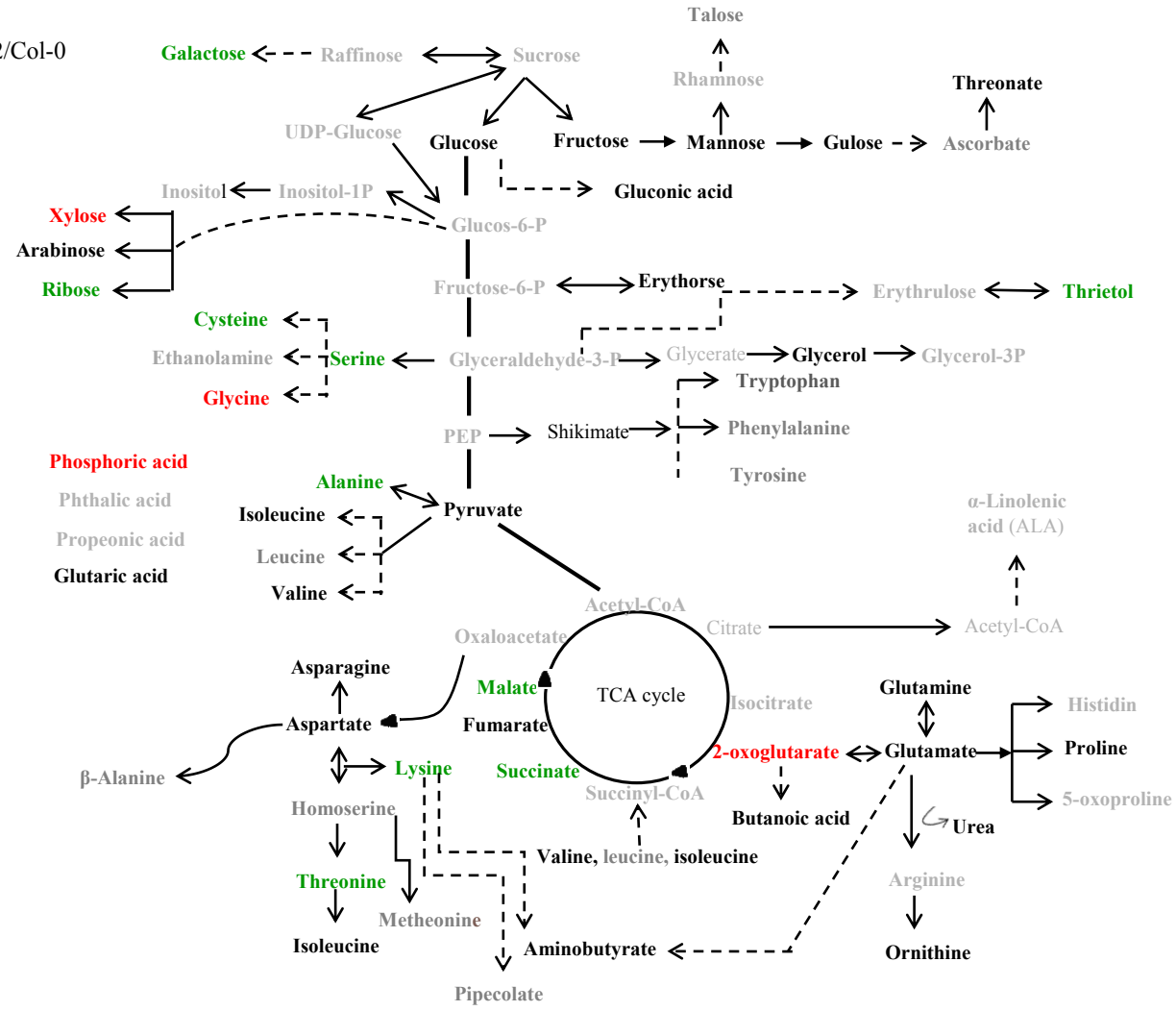


Figure 5-4. (A) Sugars and (B) other organic compounds in different genotypes of DUF642 in Arabidopsis. The samples were from 3 DAS whole plants. The values were mean of three replications ± standard deviations.

5.4.2. Profile mapping to metabolic pathways

In order to identify the perturbation of metabolic pathways, significant ($P<0.05$) metabolic changes observed between Col-0 and the mutants were highlighted in the simplified metabolic pathway maps (Fig. 5-5 and Fig. 5-6). A significant increase in xylose was observed in 35S::DGR2 (1.69-fold) whereas significant decrease in xylose was observed in 35S::DUFB (0.21-fold) and Salk_dgr2 (0.37-fold). No significant change of xylose was observed in Salk_b. Galactose was significantly decreased in 35S::DGR2 (0.07-fold) but no significant changes were observed in any other mutant lines. Mannose was significantly decreased in 35S::DUFB (0.35-fold), Salk_dgr2 (0.15-fold) and Salk_b (0.41-fold) but not in 35S::DGR2. On the other hand, arabinose was increased significantly in 35::DUFB (0.35-fold), Salk_dgr2 (0.15-fold) and Salk_b (0.41-fold) but no significant change was observed in 35S::DGR2. To confirm that different lines for same mutants were indicted the similar results of sugar modification, we analyzed the data from two different lines of 35S::DUFB and 35S::DGR2. The results revealed that the two different lines had similar sugar modification responses (Fig. 5-7). Significant increases and decreases of polyol, threitol were observed in 35S::DGR2 and Salk_dgr2, respectively. No significant changes were observed in 35S::DUFB and Salk_b. Glycine significantly increased in 35S::DGR2 (5.77-fold) but decreased significantly in Salk_dgr2 (0.25-fold). On the other hand, it was increased significantly in 35S::DUFB (0.25-fold) and decreased significantly in Salk_dgr2 (3.58-fold). In all the mutant lines, lysine decreased significantly except 35S:: DUFB, where lysine increased significantly (1.47-fold). Organic acids, malate, succinate 2-oxoglutarate and fumarate that are involved in the central metabolic pathways were significantly changed in the mutants compared to Col-0 plants. To confirm the sugar modification responses of 35S::DGR2 and 35S::DUFB, we analyzed data from two different lines of each mutant. The results were presented in Fig. 5-7. It was found that the both lines of each mutant showed similar results.

(A) DGR2/Col-0


Galactose
Raffinose
Sucrose
Talose
Rhamnose
Threonate
UDP-Glucose
Glucose
Fructose
Mannose
Gulose
Ascorbate
Inositol
Inositol-1P
Gluconic acid
Xylose
Arabinose
Ribose
Glucos-6-P
Fructose-6-P
Erythorse
Erythrulose
Thrietol
Cysteine
Ethanolamine
Serine
Glyceraldehyde-3-P
Glycerate
Glycerol
Glycerol-3P
Glycine
Tryptophan
PEP
Shikimate
Phenylalanine
Tyrosine
Phosphoric acid
Phthalic acid
Propeonic acid
Glutaric acid
Alanine
Isoleucine
Leucine
Valine
Pyruvate
α-Linolenic acid (ALA)
Acetyl-CoA
Citrate
Acetyl-CoA
Oxaloacetate
Asparagine
Malate
TCA cycle
Isocitrate
Glutamine
Histidin
Aspartate
Fumarate
2-oxoglutarate
Glutamate
Proline
β-Alanine
Lysine
Succinate
Succinyl-CoA
Butanoic acid
5-oxoproline
Homoserine
Urea
Valine, leucine, isoleucine
Threonine
Arginine
Metheonine
Isoleucine
Aminobutyrate
Ornithine
Pipecolate

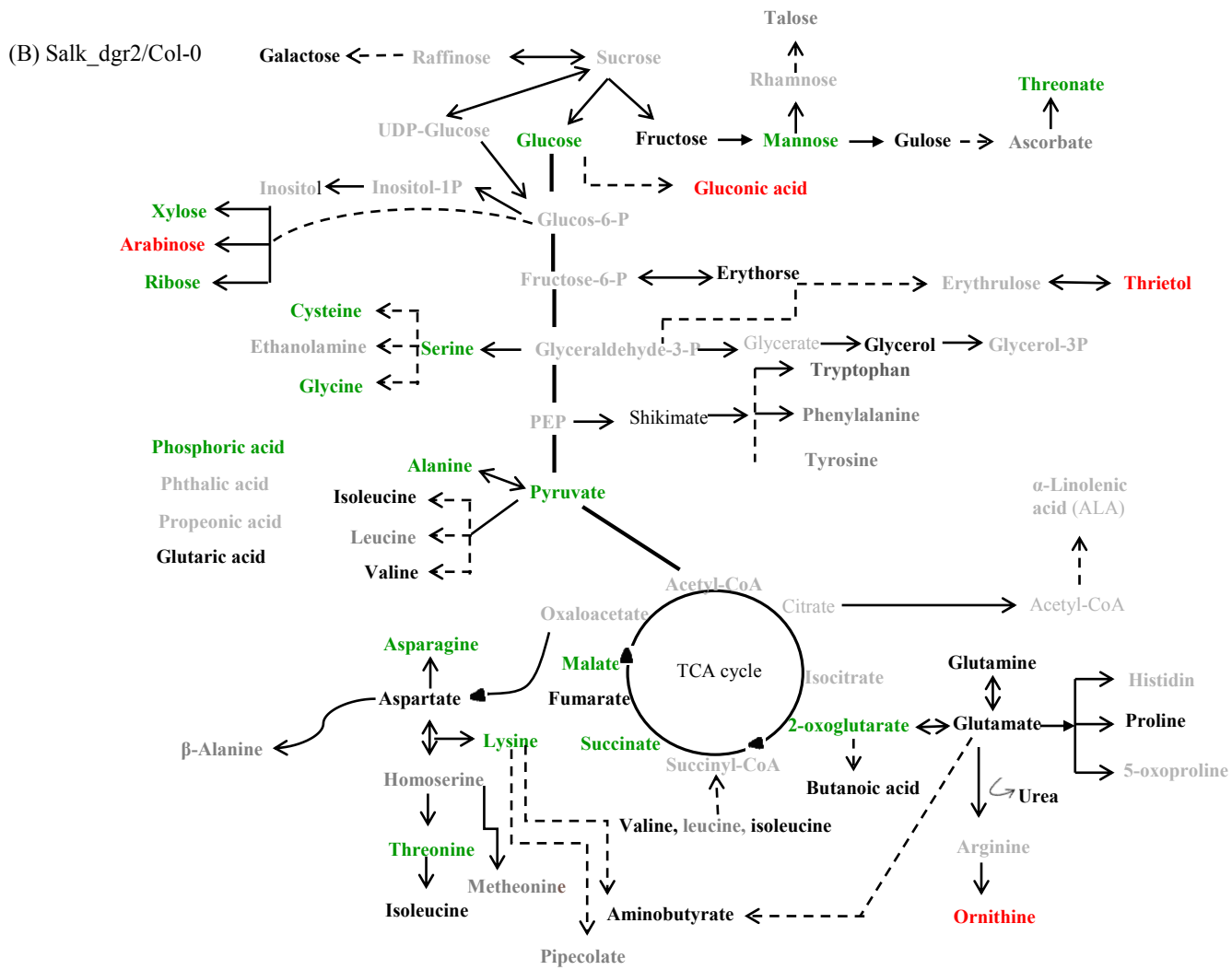
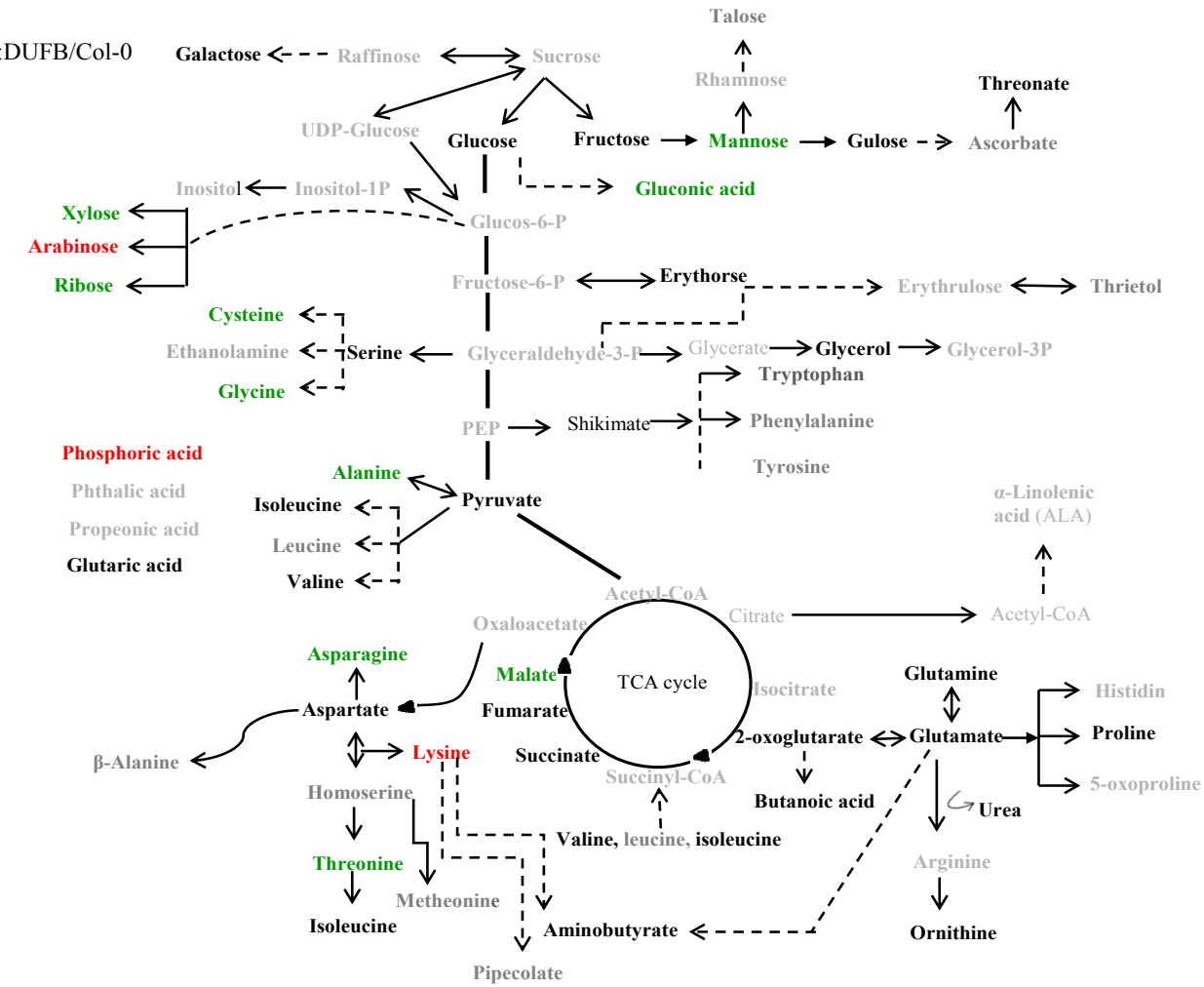


Figure 5-5. Changes in the metabolic contents in the genotypes of DUF642 mutants. The ratios between (A) 35S::DGR2 and Col-0, (B) Salk_dgr2 and Col-0. The levels of significance was set at $p<0.05$. Black letters indicated no significant change, red letters indicated increase and green letters indicated decrease, ash letters indicated not identified.

.

(A) 35S::DUFB/Col-0



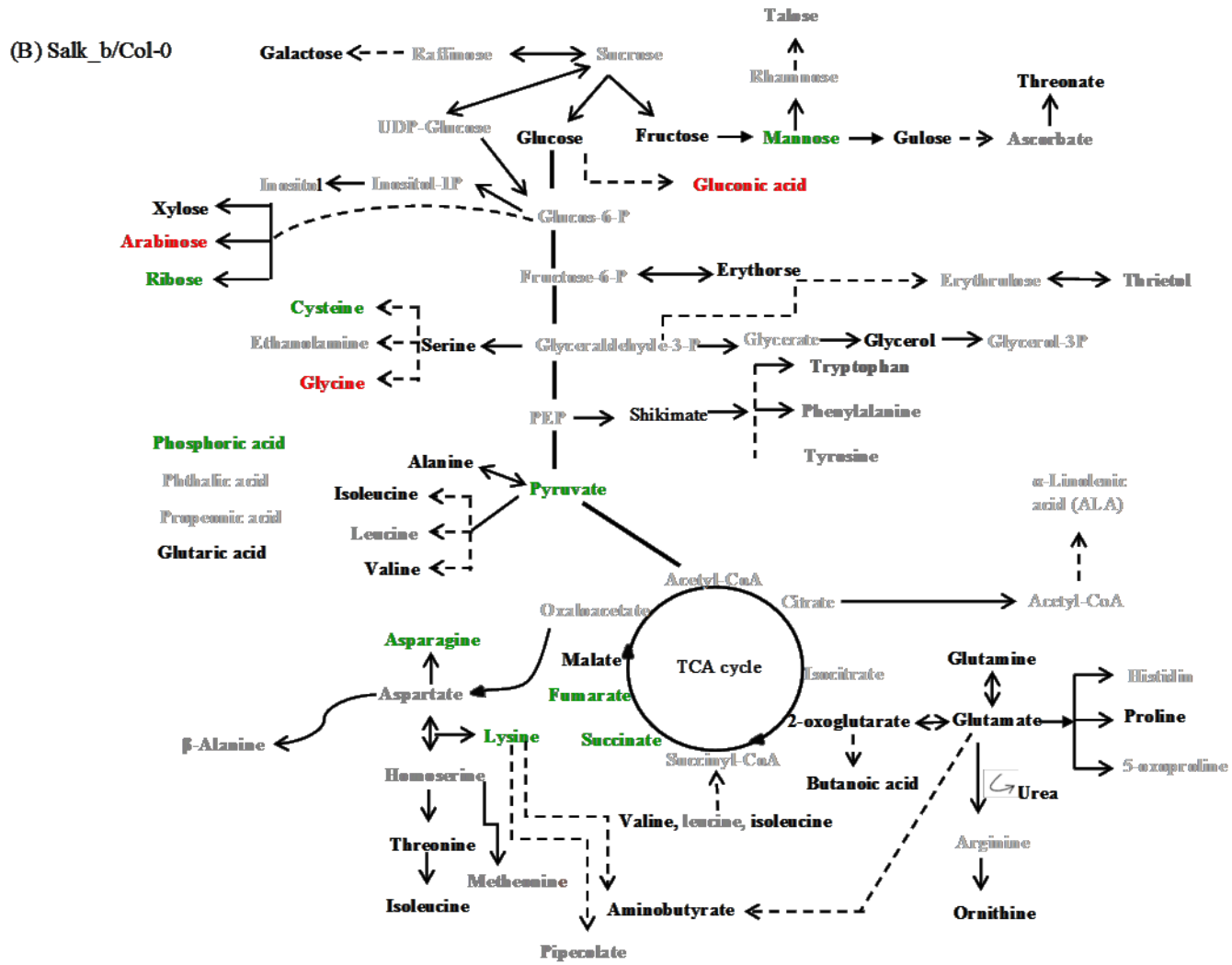


Figure 5-6. Changes in the metabolic contents in the genotypes of DUF642 mutants. The ratios between (A) 35S::DUFB and Col-0, (B) Salk_b and Col-0. The levels of significance was set at $p<0.05$. Black letters indicated no significant change, red letters indicated increase and green letters indicated decrease, ash letters indicated not identified

.

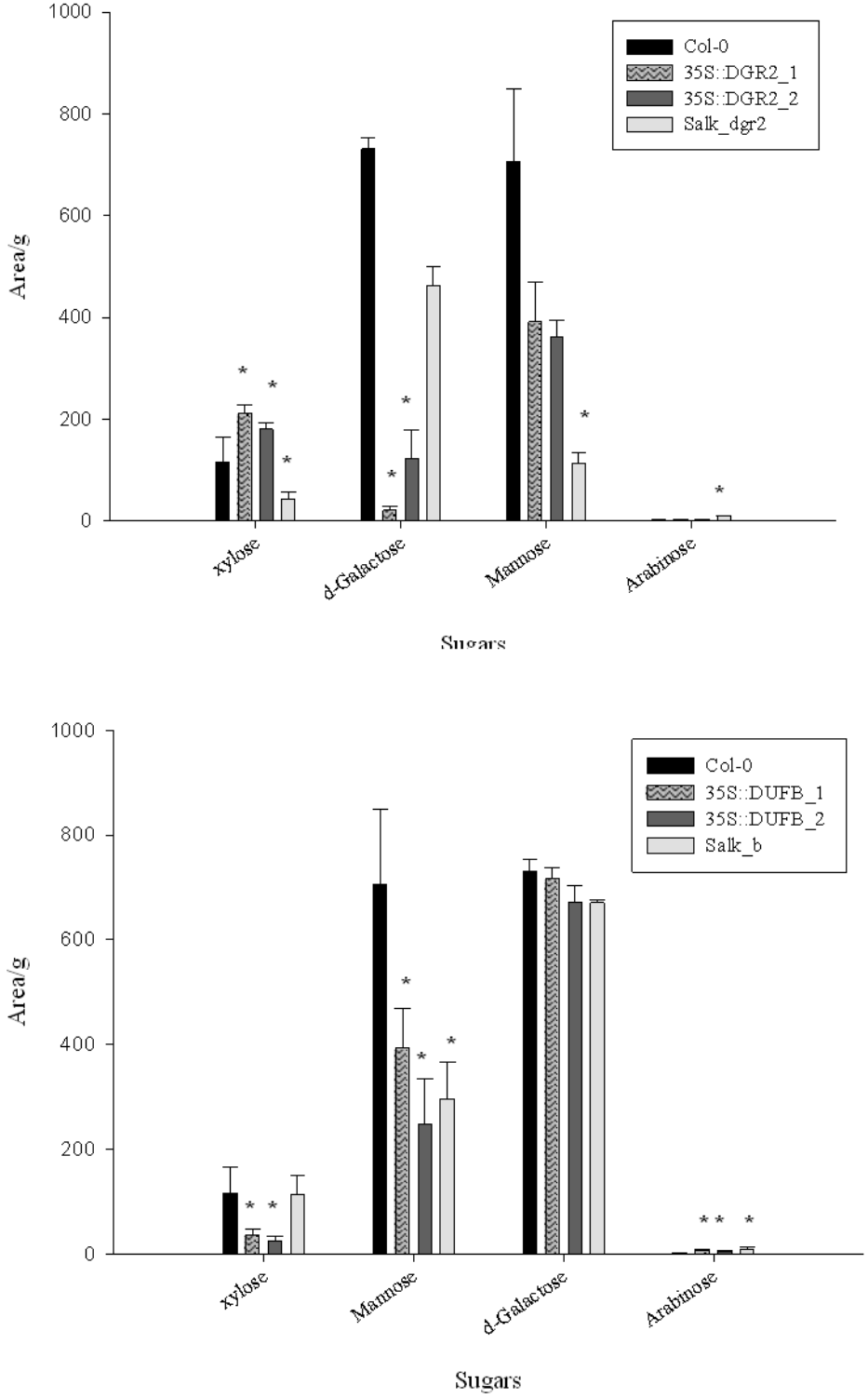
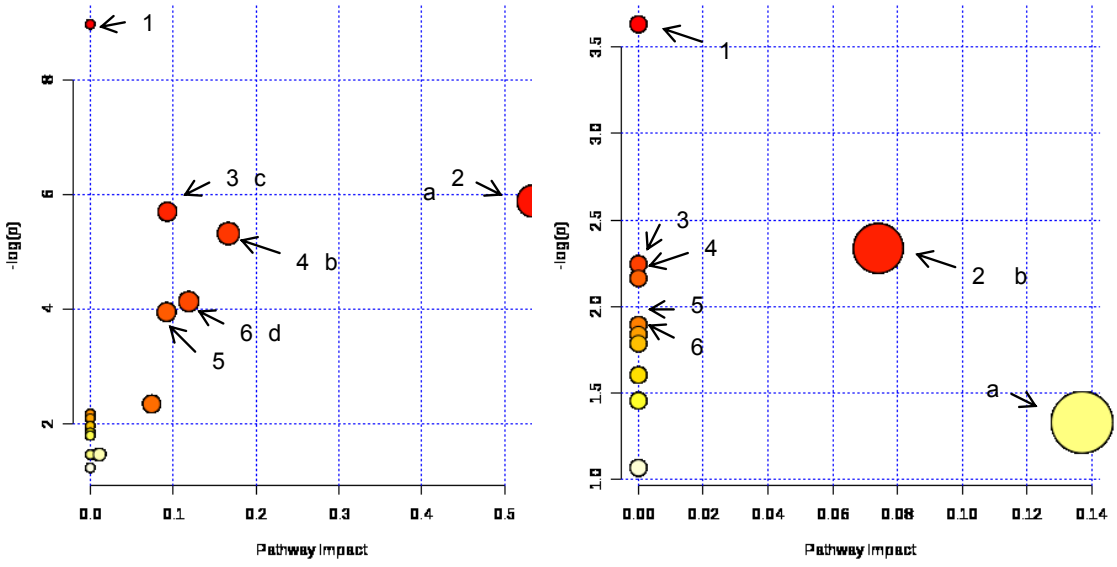


Figure 5-7. Sugar contents in two different lines of 35S::DGR2 (1 & 2) and 35S::DUFB (1 & 2) and in Salk_dgr2, Salk_b and Col-0. The values were mean of three replications ± standard deviations. * indicated significant difference from wild type at P<0.05.

5.4.3. Metabolic pathways analysis

To get a holistic view of the metabolic perturbations induced by DUF642 mutations, pathway analysis of the biochemical pathways of the Kyoto Encyclopedia of Genes and Genomics (KEGG, http://genome.jp/kegg) was conducted. The analysis was performed based on the metabolites that changed significantly between mutants and Col-0 plants. The results revealed pathways that were significantly ($p<0.05$; impact score greater than 0.1) perturbed in 35S::DGR2 samples involved cyanoamino acid metabolism, glycine, serine and threonine metabolism, aminoacyl-tRNA biosynthesis, methane metabolism, citrate cycle (TCA cycle) and alanine, aspartate and glutamate metabolism (Fig. 5-8). In contrast, the pathways that showed marked perturbations in 35S::DGR2 included aminoacyl-tRNA biosynthesis, lysine biosynthesis, cyanoamino acid metabolism and glycine, serine and threonine metabolism (Fig. 5-8). In Salk_dgr2, potentially perturbed pathways included alanine, aspartate and glutamate metabolism, aminoacyl-tRNA biosynthesis, glycine, serine and threonine metabolism, citrate cycle (TCA cycle), cyanoamino acid metabolism, methane metabolism and sulfur metabolism (Fig. 5-9). The pathways significantly influenced in Salk_b were alanine, aspartate and glutamate metabolism, aminoacyl-tRNA, citrate cycle (TCA cycle), tyrosine metabolism, butanoate metabolism, glycine, serine and threonine metabolism and lysine biosynthesis.

,



35S::DGR2 35S::DUFB

35S::DGR2

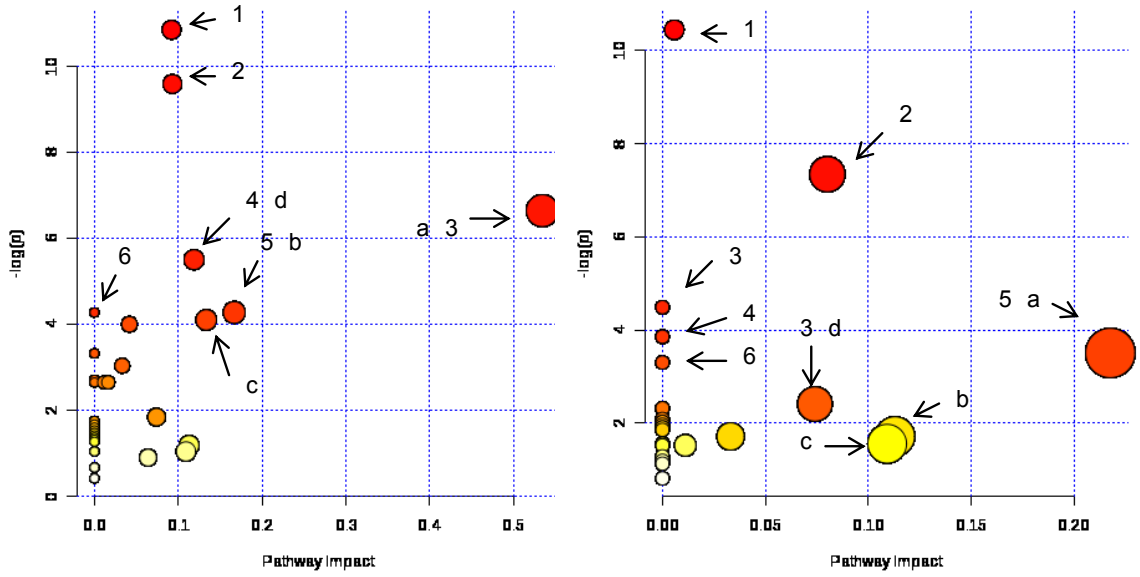
Significant pathway based on p-values and impact factor

ID	Pathway Name	Hits	p	Impact
1	Cyanoamino acid metabolism	3	1.27E-04	
2, a	Glycine, serine and threonine metabolism	3	0.0027891	0.53413
3, c	Aminoacyl-tRNA biosynthesis	4	0.0033541	0.09302
4, b	Methane metabolism	2	0.0049165	0.16667
5, d	Citrate cycle (TCA cycle)	2	0.016134	0.11875
6	Alanine, aspartate and glutamate metabolism	2	0.019393	

35S::DUFB

Significant pathway based on p-values and impact factor		**Hits**	**p**	**Impact**
	Pathway Name			
1	Aminoacyl-tRNA biosynthesis	3	0.026447	
2, b	Lysine biosynthesis	1	0.096792	0.07407
3	Cyanoamino acid metabolism	1	0.10598	
4	Pentose and glucuronate interconversions	1	0.11508	
5	Fructose and mannose metabolism	1	0.15064	
6	Pentose phosphate pathway	1	0.16792	
a	Glycine, serine and threonine metabolism	1		0.13697

Figure 5-8: Summary of pathway analysis with MetaboAnalyst: Metabolites altered in 35S::DGR2 compared to Col-0 and 35S::DUFB compared to Col-0 mapped to multiple biosynthetic pathways. Statistics for pathways with major changes based on the p value (pathways 1-6) or on high impact (pathways a-d). Colours in the pathways: light blue means metabolites are is not the data but used in the enrichment analysis, grey means the metabolite is not in the data and also excluded from enrichment analysis, from yellow to red means the metabolites are in the data with different levels of significance.



Salk_dgr2 Salk_b

Salk::dgr2

Significant pathway based on p-values and impact factor

ID	Pathway Name	Hits	p	Impact
1	Alanine, aspartate and glutamate metabolism	5	1.96E-05	
2	Aminoacyl-tRNA biosynthesis	7	6.90E-05	0.53413
3, a	Glycine, serine and threonine metabolism	4	0.0013092	0.53413
4, b	Citrate cycle (TCA cycle)	3	0.0040889	0.18175
5	Cyanoamino acid metabolism	2	0.01396	
6, d	Methane metabolism	2	0.01396	0.16667
c	Sulfur metabolism	2		0.13333

Salk_b

Significant pathway based on p-values and impact factor	**Pathway Name**	**Hits**	**p**	**Impact**
1, d	Alanine, aspartate and glutamate metabolism	3	2.8986E-5	
2, b	Citrate cycle (TCA cycle)	1	6.4532E-4	0.08011
3	Tyrosine metabolism	1	0.011212	
4	Butanoate metabolism	1	0.011212	
5	Aminoacyl-tRNA biosynthesis	1	0.021118	
6, a	Glycine, serine and threonine metabolism	1	0.029953	0.21756
c	Lysine biosynthesis	1		0.07407

Figure 5-9.) Summary of pathway analysis with MetaboAnalyst: metabolites altered in Salk_dgr2 compared to Col-0 and Salk_b compared to Col-0 mapped to multiple biosynthetic pathways. Statistics for pathways with major changes based on the p value (pathways 1-6) or on high impact (pathways a-d). Colours in the pathways: light blue means metabolites are is not the data but used in the enrichment analysis, grey means the metabolite is not in the data and also excluded from enrichment analysis, from yellow to red means the metabolites are in the data with different levels of significance.

5.4.4. Principle Component Analysis

Metabolite profiling data obtained for the six different genotypes were subjected to statistical analysis via PCA to identify metabolic phenotypes of DUF642 mutants. PCA factors 1 and 2 were used for visualization of differences between genotypes. In this analysis, there were 47 attributes (metabolites) and 6 observations (genotypes). PC1 accounted for 43.72% and PC2 accounted for 19.55% of the variation. PCA identified the existence of differences in metabolite composition among loss of function and gain of function mutants of DGR2 and DUFB plants. It was notable that all three loss-of-function mutants were in one cluster and all the gain-of-function mutants were in another cluster (Fig. 5-10). Col-0 was separated from both the loss of function and the gain of function mutant clusters. Fig. 5-10 showed that principal component PC1 gave Col-0 a positive value whereas negative values for RNAi, Salk_dgr2 and Salk_b. So, there was inverse correlation between the loss-of-function mutants and wild type plants. PCA1 also gave 35S::DGR2 and 35S::DUFB positive values, which indicates they had also inverse correlation with loss of function mutants. On the other hand, PCA2 showed that Salk_dgr2, Salk_b displayed positive values whereas 35S::DGR2 and 35S::DUFB displayed negative values (Fig. 5-10). So, according to PCA2, 35DGR2 and 35S DUFB had inverse relationships with Salk_dgr2 and Salk_b plants. Therefore, the obtained PCA revealed a strong separation of the loss of function mutants from the gain of function mutants of both genes. In order to demonstrate the major contributors of variation between the genotypes, a biplot (Fig.5-11) was generated with both the loadings and the scores for PCA1 and PCA2 in parallel. In the biplot, the variables were plotted in the form of vectors (loadings) and observations as labeled dots (scores). The top 13 compounds including galactose, mannose, xylose, arabinose, glucose, gulose, alanine, serine, lysine, succinate, 2-oxoglutarate, pyruvate, gluconate were found to be significantly different.

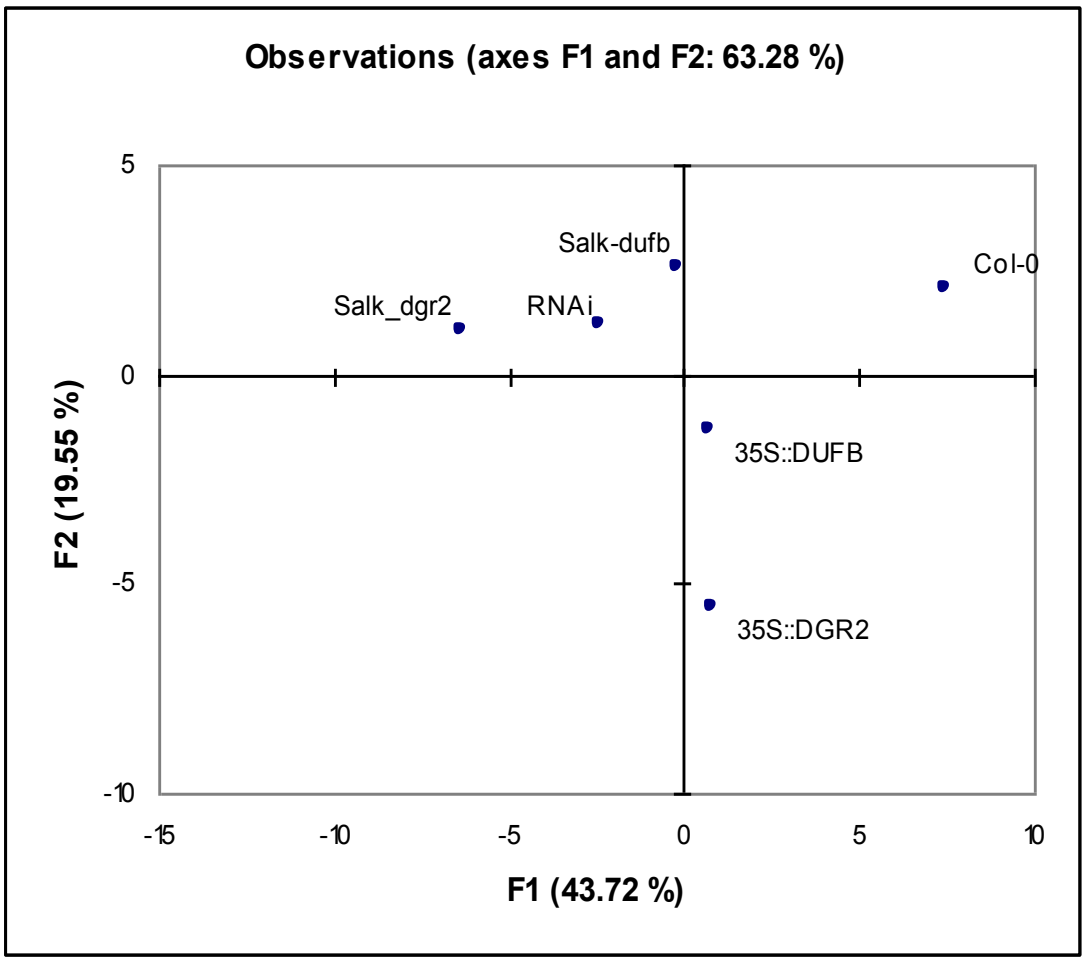


Figure 5-10. (A) Scatter plot of 6 different genotypes of DUF642 in Arabidopsis based on the first two principal component analysis (PCA) axes. The percentage of variance explained by each axes is indicated. LOG_{10} transformed value was used for the graph.

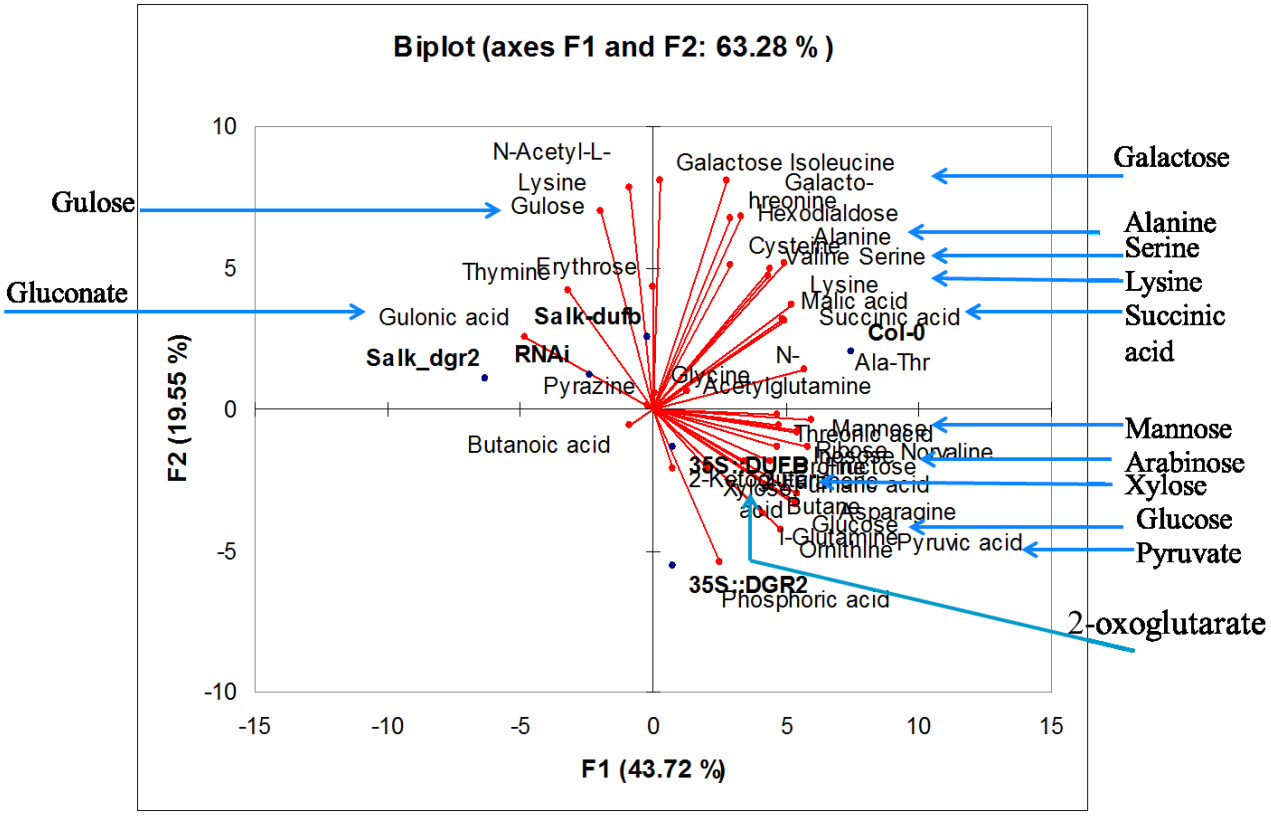


Figure 5-11. A PCA biplot of the mean centered metabolites data is shown. In the biplot both the genotypes (dots) and the variables (vectors) were shown to enable interpretation about the relations. In total, 63.28% of the variation of the metabolites data was represented by PC1 (43.72%) and PC2 (19.55%). LOG_{10} transformed value was used for the graph.

5.5. Discussion

In the context of assigning functions to an unknown gene, metabolic profiling has advantages over transcript analysis. Metabolites and proteins are functional entities whereas messenger RNAs are transmitters of genetic information, not functional cellular entities (Oliver, 2000). Metabolites can be identified as sample constituents and based on the actual biochemical status of the tissues. In the present study, metabolic profiling of different genotypes of DUF642 plants in Arabidopsis was conducted by GC/MS to elucidate the function of DUF642 genes. Three loss-of-function mutants (Salk_dgr2, Salk_b and RNAi) and two gain-of-function mutants for DGR2 and DUFB (35::DGR2 and 35S::DUFB) were studied and Col-0 plants were used as a control. In total, 47 metabolites were identified and grouped in to four functional classes: sugars, amino acids, organic acids and other organic compounds (Fig.5-3 & 5-4).

The primary role of amino acids is in protein biosynthesis. In the present study, the levels of six amino acids were altered in 35S::DGR2 and 35S::DUFB. Among these six amino acids, four were same for both mutant lines, whereas serine and glycine were altered in 35S::DGR2 and norvaline and asparagine were altered in 35S::DUFB. These findings indicated that protein biosynthesis was altered in *DGR2* and *DUFB* mutants but by different pathways. The pathway enrichment analysis revealed that the TCA cycle was one of the pathways perturbed in 35S::DGR2, Salk_dgr2 and Salk_b (Fig 5-8 and & 5-9). The TCA cycle is an important pathway for the oxidation of carbohydrates, fatty acids and amino acids, is vital for the generation of energy. Metabolite contents of TCA cycle, malate and succinate, decreased in 35S::DGR2 and Salk_dgr2 whereas malate decreased in 35S::DUFB and fumarate and succinate decreased in Salk_b plants (Fig. 5-5 and 5-6). The alteration of this pathway indicated possible disorder of primary cell metabolism in DUF642 mutants. Alanine, aspartate and glutamate metabolism pathways were altered in 35S::DGR2, Salk_dgr2 and Salk_b (Fig. 5-8 and 5-9). This was indicative of perturbation of TCA cycle which is needed for cell growth. Pathway analysis also identified the perturbation of glycine, serine and threonine metabolism in all DUF642 mutant lines. Amino acid metabolism made up the majority of affected pathways by DGR2 and DUFB

mutants (Fig. 5-8 and 5-9). The results suggested that DGR2 and DUFB might hold regulatory effects on amino acid metabolism other than glutamate and glutamine.

Principle component analysis was applied to the metabolites data set that was obtained from the six different genotypes of DUF642 in Arabidopsis. In PCA, samples were separated based on cumulative correlation of the metabolites and identified vectors that resulted in the largest separation between samples. The score plot showed that overexpression and loss of function mutant lines were clearly separated from each other (Fig. 5-10). PCA analysis indicated that many compounds contributed to distinguishing between the genotype including amino acids, sugars, organic acids and other organic compounds, which could be regarded as potential biomarkers to elucidate the functions of DUF642 genes (Fig. 5-11). The biplot (Fig 5-11) revealed that 35S::DUFB and 35S::DUFB clustered with the sugars, mannose and xylose. So, these sugars were among the main metabolites that distinguished the overexpression lines from Col-0 and the loss-of-function mutants. Xylose was significantly increased in 35S::DGR2 and decreased significantly in 35S::DUFB (Fig. 5-7). Mannose, on the other hand, decreased significantly in 35S::DUFB, Salk_dgr2 and Salk_b but not in 35S::DGR2. Furthermore, the apparent reduction of galactose content (Fig 5-7) in 35S::DUFB mutant lines might be due to the perturbation of overall galactose biosynthesis or galactose degradation pathway. Therefore, it can be pointed out that DGR2 and DUFB were involved in carbohydrate metabolism.

Threitol was significantly decreased in 35S::DGR2 and increased significantly in Salk_dgr2 (Fig. 5-5). Threitol is a sugar alcohol but its biosynthetic pathway has not been studied. Walters et al. (2009) in their study with Alaskan beetle identified that threitol was synthesized from erythrose 4-phosphate, a C4 intermediate in the phosphate pentose pathway (PPP). He also reported that in PPP, sugar phosphatase (s) preferred threitol 4-phosphate as a substrate and produce threitol over erythrose. So, it is indicative that, 35S::DGR2 affected PPP. Perturbation of glycolysis was displayed in Salk_dgr2 as indicated by decreased glucose and pyruvic acid (Fig. 5-5) contents.

Rhamnogalacturonan II (RG-II) is composed of 11 different glycosyl residues and many enzymes are involved in RG-II metabolism in the cell wall of Arabidopsis. Xylose and fucose are two of the subunits of RG-II (Yapo et al., 2011). An increase in fucose and decrease in xylose in 35S::DUFB were indicative of possible changes in cell wall structure linked with RG-II. Galactose, which is also a component of RG-II and related pectins (Yapo et al., 2011) decreased significantly, and xylose increased significantly in 35S::DGR2. It is possible that DGR2 is involved in RG-II degradation, which mostly occurs during cell wall expansion and elongation. DGR2 and DUFB both might be participated in cell wall changes by targeting pectin. In the previous study it was found that heterologously expressed DGR2 interacted with pectinmethylesterase *in vitro (*Sanchez and Buen, 2012). Pectin is synthesized in Golgi and secreted in the cell wall in highly methylesterified form and there de-esterified by pectin methylesterases (PME). Mannose, is a compoment of hemicelluloses also decreased significantly in 35S::DUFB, Salk_dgr2, and salk_b. This is again indicative that DUFB and DGR2 are involved in cell wall modification through metabolism of glycosyl residues.

5.6. Conclusions

In this study, metabolic profiling of loss-of-function and gain-of-function mutants of DGR2 and DUFB genes in Arabidopsis was conducted. By relative comparison of the metabolic alteration of these mutant lines with wild type plants and multivariate data analysis, we could elucidate the function of DUF642 genes. We could identify metabolites and metabolic pathways associated with DGR2 and DUFB genes. The results showed that TCA cycle was affected by mutations in either gene. Significant decreases in malate, fumarate and succinate were observed in the mutant lines. It could be noted that DGR2 and DUFB were involved in cell metabolism by regulating TCA cycle. In addition, it was discovered that DGR2 and DUFB had regulatory effects on amino acid metabolism except glutamine and glutamate. It was also revealed that DGR2 and DUFB had possible function in cell wall modification by targeting pectin and involved in carbohydrate metabolism. Our data also illustrated the power of metabolic profiling in finding gene function and provide a broader assessment of metabolic change following overexpression and loss of function of DGR2 and DUB genes.

6. Chapter 6

6.1. Metabolites switch of flax (Linum usitatissimum) developing embryos and mature seed

6.2. Introduction

Flax (*Linum usitatissimum*) is a multipurpose important crop and belongs to the family *Linaceae*. Two types of flax can be identified: one is linseed type, grown for oil production from the seed and another is for collecting fiber from the stem. Canada is a major flax producing country along with Argentina, India, the USA and Russia (Canadian Food Inspection Agency).

Flax seed and fiber has wide range of industrial applications. Its seeds contain about 41 to 43% oil, 30% dietary fiber and 25% protein, minerals, vitamins, and carbohydrates (Bhatty, 1995; Morris, 2001; El-Beltagi et al., 2007, Ziolkovska, 2012). Flax seed oil enriched with Linoleic (Omega-6) and Linolenic (Omega-3) polyunsaturated fatty acids (PUFA). It is the richest crop source of α-linolenic acid (ALA), a precursor of omega-3 fatty acids, and comprises approximately 55% of the total fatty acids in the seed (Carter, 1993). ALA can prevent atherosclerotic cardiovascular disease through improving lipid profiles, lowering blood pressure, inhibiting platelet aggregation and thrombosis (LeAnne et al., 2004). The lignan constituents of flaxseed consist mainly of secoisolariciresinol diglucoside (SDG, a phytoestrogen) at levels 75 - 800 times greater than any other crop (Westcott and Muir, 1996 and Thompson et al., 1997) and may be important in the treatment of breast cancer (Wang et al., 2005), prostate cancer (Demark-Wahnefried et al., 2001) and type 2 diabetes (Prasad, 2001). The seed fiber of flax is rich in pentosans and the hull fraction contains 2-7% mucilage, a source of soluble dietary fiber (Vaisey-Genser and Morris, 1997). In addition to these important attributes, flax seed also contains components that might have harmful effects on health such as cyanogenic glycoside, trypsin inhibitors, allergens and goitrogens (Cunnane and Thompson, 1995). Because of these exceptional properties, a detailed study of flax seed metabolism is especially important to understand how these valuable properties are produced during seed development.

Three major phases in plant seed development are: embryogenesis, seed maturation and desiccation or quiescence. During embryogenesis, a single-celled zygote undergoes a series of cell divisions to form a mature embryo. Seed maturation is the result of transition from maternal to filial metabolic regulation (Weber et al., 2005). A shift of metabolites and gene expression patterns was reported during previous studies of seed maturation (Fait et al., 2006; Venglat et al., 2011. Flax seed development and embryogenesis was studied by Venglat et al. (2011) in which they studied three major tissues: the diploid embryo; the triploid endosperm; and the maternal seed coat. After fertilization, the zygote undergoes cell division and forms the globular embryo which is successively transformed into heart embryo, early torpedo, late torpedo, cotyledon stage embryo with rounded cotyledon tips, mature ambryo with elongated cotyledons ((Venglat et al., 2011).

Metabolites are the end products of cellular processes and their levels represent the ultimate reflection of the response of biological systems to genetic or environmental changes (Fiehn, 2002). Metabolic profiling, in tandem with multivariate statistical analysis, has yielded valuable information such as biochemical phenotyping of Arabidopsis ecotypes and mutants (Fiehn et al., 2000a), evaluating antioxidant properties in grape (Pacifico et al. 2009), identifying metabolic shifts in pathways during strawberry fruit development (Zhang et al., 2011) and during Arabidopsis seed development and germination (Fait et al., 2006). Different platforms are available for metabolic profiling such as gas chromatography-mass spectrometry (GC-MS), nuclear magnetic resonance (NMR) spectroscopy. Gas chromatography-mass spectrometry (GC-MS)-based metabolic profiling is considered to be the most reliable for its high resolution and precision and able to detect uncommon metabolites (Fiehn et al., 2000b).

Studies were conducted to understand the complex processes of flax seed development in respect to gene expressions and proteome profiling (Venglat et al. 2011; Barvkar et al., 2012) but metabolic profiling which is the newest in the “omics” sciences was still missing. The purpose of the present study was to investigate the changes in metabolic composition during the

embryogenesis and in the mature seed of flax by gas chromatography–mass spectrometry (GC-MS) using non-targeted, quantitative profiling of polar extracts.

6.3. Materials and Methods

6.3.1. Plant growth conditions and tissue collections

Flax seeds of *Linum usitatissimum* cv CDC Bethune were grown in a growth chamber using a daily cycle of 16 hours of light (23°C) and 8 hours of dark (16°C). Flax flowers were tagged after opening and developing bolls were collected at heart stage, torpedo stage, embryonic cotyledon stage and mature green seed stage. For heart, torpedo and embryonic cotyledon stages, tissue collection was performed after confirming the developmental stage under the microscope (Olympus BX51). At first a capsule was opened, and one developing seed from the capsule was examined and staged the rest of the capsule was sampled for metabolite extraction. The dissected seeds were discarded. Immediately after collection, tissue samples were placed in liquid nitrogen. For mature green seed stage, seeds were collected 21 days after flowering. The seeds were collected from more than 16 individual plants and more than 1000 capsules were dissected to get enough tissues for metabolite extraction from required developmental stages.

6.3.2. Sample extraction

Three independent extractions of metabolites were performed with the tissue. Samples were lyophilized then homogenized by ball mill. For samples at green seed stage, a rigorous grinding procedure was implemented to get fine powder of the tissue. At first, the lyophilized tissue was grounded by dipping in liquid nitrogen and using pestle in a microfuge tube. Next, the ground powder was again homogenized by ball mill to get very fine powder of the tissue. Metabolites were extracted using the method described by Fiehn (2006). In this extraction protocol, a mixture of chloroform, methanol and water was used. This protocol was evaluated as an effective method considering reproducibility of extracting both polar and non polar metabolites (Tambellini et al., 2013). In this method, aliquots of frozen powder (~5mg) were extracted using HPLC grade extraction mixture consisting of chloroform, methanol and water at a ratio of 1:2.5:1 (v/v/v) and degassed the mixture using an ultrasonicator. In each sample, 1 ml of cold extraction solvent was

added (-15°C and degassed). Ribitol was added as an internal standard for normalization. Samples were then shaken in a rotary shaker for 10 minutes at 4°C and centrifuged for 2 minutes at 14 000 rpm. The supernatant transferred into new tubes then 400 µl of pure water was added to each samples and vortex for 10s. The upper phase of each samples were collected as a 'polar phase' (mixture of methanol and water) in new microfuge tubes for derivatizaton. The polar phase was dried using speed vacuum concentration for 1h and then lyophilized overnight.

6.3.3. Derivatization

Two-step chemical derivatization was performed on the extracted polar metabolites according to the protocol of Fiehn (2006). Briefly, oximation was carried out by dissolving the samples in 20 µl of methoxamine hydrochloride (20 mg ml^{-1} in pyridine) and incubating at 28 °C for 90 min. Next, samples were further derivatized with the addition of silylting agent, *N*-Methyl-*N*-(trimethylsilyl) trifluoroacetamide with 1% trimethylchlorosilane at a volume of 90 µl to each sample and were shaken for 30 min at 37°C. Samples were then transferred to inserts (Agilent, 5181-1270) containing glass vials for the GC/MS. Before analysis, samples were diluted (50:50) in pyridine and the dilution factor was included in the calculations.

6.3.4. GC/MS analysis

GC/MS analysis of the polar extract of metabolites was performed with an Agilent 7890A gas chromatograph and an Agilent 5975C MSD coupled to an electron impact (EI) ion sources. Derivatized sample extract (1 µL) was injected using splitless mode and splitless single taper liner with deactivated glass wool (4 mm ID) on to a 30 m length, 0.250 mm diameter and 0.25 µm film thickness DB-5MS capillary column (Agilent J & W GC Columns). Helium (99.999%) was used as the carrier gas with the flow rate at 1 ml/min. The initial oven temperature was 70°C and after a 10 minutes solvent delay, oven temperature was increased to 76°C at 1°C/min. From 76°C, oven temperature was increased to 320°C at 6.1 °C/min for a final run time of 62 minutes. Ion source filament energy 70 eV and TIC (total ion chromatogram) spectra were recorded in the mass range of 50-550 amu. The detailed analysis condition was shown in Table 5-1 in chapter 5. To minimize the carryover effects, liners were replaced every 15 injections and background was

monitored by running blank after every 5 samples and before any set of treatment samples. An alkane standard which was a mixture of 75 µl of C10, C20-C40 and 25 µl of C21-C40 was run before running the samples to test the method performance. Chemstation software (Agilent Technologies) operated the system and validated chromatogram and spectrum output. Specific mass spectral fragments were detected in defined retention time windows using NIST MS search 2.0. Mass spectral matching with the NIST library was conducted manually and matches accepted with threshold of match >650 (maximum match is equal to 1000). Metabolite peaks were assigned and validated using the automated mass spectral deconvolution and identification system (AMDIS) and the National Institute of Standards and Technology (NIST) library (version 98). Peaks that were at least 70% pure were automatically assigned identity

6.3.5. Data analysis

For individual metabolites, means ±SD were calculated from three biological replicates. Metabolite contents were calculated for individual metabolites obtained from GC/MS selected ion chromatogram by means of auto integration (ChemStation integrator, threshold=15) of peak area of each metabolite. Metabolites peak areas were divided by peak area of the internal standard, ribitol, to correct any recovery differences. The corrected peak areas were then normalized by dividing with dry weight of the samples and expressed as area/g. One-way ANOVA was performed by SAS/STAT software (version 9.1; SAS Institute, NC). PCA was performed on the log10 transformed data sets obtained from metabolic profiling of developing seed of flax with the software package XLSTAT2014. To describe the increased or decreased abundance of the various metabolites, the ratios were calculated as the means of the metabolite contents among different stages. Simplified metabolic pathways were created by taking the information from KEGG pathway database.

6.3.6. Metabolite pathway analysis

To determine the metabolic pathways which significantly altered in different developmental stages, MetaboAnalyst 2.0 (Xia et al., 2012) was used to assess the pathway enrichment within the obtained data set. Pathways were ranked based on two magnitude of significance. A p-value

score, based on metabolite set enrichment analysis and an impact factor which is calculated based on the importance of the identified metabolites within the directional network. A p-value less than 0.05 and an impact score greater than 0.1 considered to be significant. Metabolites that were significantly changes between two stages compared were used in the pathway analysis.

6.4. Results

In the present study, metabolic profiling of flax seed at different developmental stages was investigated to identify metabolic changes during seed development. Whole seeds were analysed at heart stage, torpedo stage, cotyledon stage and mature green seed (21 days after anthesis) stage. Polar phase of metabolites were analysed by GC/MS and could identified 68 metabolites after comparing with the non sample control experiment. Metabolite contents were calculated for individual metabolite obtained from GC/MS selected ion chromatogram by means of auto integration of peak area of each metabolite and corrected the peak area by dividing with the area of internal standard. The corrected peak areas were then normalized by the dry weight of the sample and the content of metabolites were expressed as area/g. The mean values of identified metabolites were presented in Table 6-1.

A Venn diagram was used to illustrate the distribution of the identified metabolites in the flax seed at different developmental stages. In total 68 metabolites were identified. Among these 28 metabolites were identified in all four stages (Fig. 6-2 and Table 6-2). These metabolites contained sugars, amino acids, organic acids and other organic compounds. The maximum number of metabolites was identified in the torpedo stage which was 54. In case of heart stage, cotyledon stage and mature green seed stage, these numbers were 50, 45 and 40, respectively (Table 2). The unique metabolites identified in heart stage were erythrotetrofuranose, xylopyranose, inosose, and glycerol. The maximum number of unique metabolites was obtained at torpedo stage and were erythro-pentitol, thiazole, methionine, azetidinone, iodo-l-tyrosine, norvaline, isoleucine, and leucine. In cotyledon stage, the unique metabolites identified were gluconic acid, piperidone and pentenoic acid. The metabolites that were identified only in the mature green seed stage were threitol, histamine, α-linolenic acid (ALA).

To associate metabolites with pathways, significant ($P<0.05$) changes in the levels of metabolites during the different stages of flax seed development were highlighted in the simplified metabolic pathway maps (Fig. 6-3). For simplicity, only metabolites involved in major metabolic pathways were mapped. The changes in metabolite contents were calculated by dividing the metabolite level in the flax seed development stages: hearts stage, torpedo stage, cotyledon stage and mature stage.

The metabolites that were not mapped and that were found in all four stages were 2-furanone, a naturally occurring lactone involved in plant defense, homoharringtonine, a naturally occurring plant alkaloid with antitumor properties involved in the inhibition of initial elongation step of protein synthesis (Lü and Wang Wang, 2014) and N-acetylglutamine. Other metabolites that were found in a subset of developmental stages, but that were not mapped, were 4-butanediamine, erythro-pentonic acid, ethylene glycol, N-acetyl-l-lysine, histamine, piperidone, pentenoic acid, erythro-pentitol, thiazole, methionine, azetidinone, iodo-l-tyrosine, norvaline, erythrotetrofuranose, xylopyranose, inosose and glycerol.

During the heart stage to the torpedo stage, significant increases in 2-oxoglutarate (1.28-fold), phenylalanine (1.64-fold), threonate (1.63-fold) and proline (1.46-fold) (Table 6-3, Fig. 6-3) were observed whereas significant decreases were observed for glutamine (0.25-fold), xylose (0.43) and phosphoric acid (0.01). . The amino acids that were detected only in cotyledon stage were leucine, isoleucine, valine, homoserine, threonine, metheonine and tyrosine. The metabolites that were detected in heart stage but not in cotyledon stage were glycerol, asparagine and lysine. During the torpedo stage to the cotyledon stage, significant increase was observed only in case of gulose (1.44-fold) but significant decreases were observed in case of fructose (0.63-fold), erythorse (0.68-fold), serine (0.07-fold), homoserine (0.24-fold), glycine (0.52-fold) and pyruvate (0.05-fold) (Table 6-3 and Figure 6-3). During cotyledon stage to the mature stage, gulose (0.14-fold), glucose (0.16-fold) and fructose (0.14-fold), pyruvate (0.24-fold) were significantly decreased whereas talose (5.86-fold), serine (6.59-fold), phosphoric acid (60.77)

and aminobutyric acid (1.23) were significantly upregulated. The levels of galactose, cysteine and alanine remained the same throughout the developing stages

To obtain a more holistic view of the metabolic alterations occurred during seed development, pathway analysis of the biochemical pathways of the Kyoto Encyclopedia of Genes and Genomics (KEGG, http://genome.jp/kegg) was conducted. The analysis was conducted based on the metabolites that changed significantly between torpedo stage and heart stage, torpedo stage and cotyledon stage and cotyledon stage and mature stage. Also in the analyses, metabolites that were identified only in the torpedo stage, cotyledon stage and mature green seed stage were included (Figure 6-7 and 6-7.1). Both concerted changes in metabolite intensity within pathways (Global Test) (Goeman and Bühlmann, 2007) and alterations of high impact were considered in the analyses. Based on p values, altered pathways identified in torpedo stage were aminoacyl-tRNA biosynthesis, valine, leucine and isoleucine biosynthesis, glucosinolate biosynthesis, nitrogen metabolism. Based on impact values, the pathways were phenylalanine, tyrosine and tryptophan biosynthesis, isoquinoline alkaloid biosynthesis, pyruvate metabolism and glycolysis. In cotyledon stage, altered pathways identified based on global test were glycine, serine and threonine metabolism, pentose phosphate pathway, alanine, aspartate and glutamate metabolism and lysine biosynthesis whereas based on impact values the altered pathways were glycine, serine and threonine metabolism, methane metabolism, pyruvate metabolism and glycolysis (Figure 6-7 and 6-7.1). Pathways found to be altered based on global test in the mature seed stage were arginine and proline metabolism, alanine, aspartate and glutamate metabolism, aminoacyl-tRNA biosynthesis and pantothenate and CoA biosynthesis. Beta-alanine metabolism, alpha-linolenic acid metabolism, isoquinoline alkaloid biosynthesis and glycine, serine and threonine metabolism had altered metabolites with high impact (Figure 6-7 and 6-7.1).

Principal component analysis was conducted to obtain a global view of the metabolic changes that occurred during flax seed development. PCA uses an n-dimensional vector approach to separate the samples based on the cumulative correlation of all metabolite data. The calculated vectors that yielded the greatest separation between samples were identified and then used to

calculate their factor scores (Zhang et al. 2010, Roessner et al., 2001). The results from the first two principle components, PC1 and PC2, which showed highest variance between samples, were plotted and shown in Fig. 6-4 and 6-5. In Fig. 6-4, only the stages of flax seed development was plotted for clear interpretation. PC1 accounted for 47.98% and PC2 accounted for 76.35% of the variation. Principal component PC1 gave heart stage, torpedo stage and cotyledon stage positive values whereas negative value for mature green seed stage. So, there were inverse correlations between the developing seed and the developing embryos. On the other hand, PCA2 showed that mature green seed stage, heart stage, torpedo stage had positive values where as cotyledon stage had negative value. So according to PCA2 mature green seed stage and cotyledon stage had inverse relationship. The biplot (Fig. 6-5) showed both the loadings and the scores for PCA1 and PCA2 in parallel. Screen plot (Fig. 6-5B) showed the eigenvalues and cumulative variance of first three components. Mature seed constituted a single cluster with the metabolites threitol histamine, α-Linolenic acid (ALA), tyrosine, phosphoric acid and urea. Heart stage and torpedo stage were in one cluster with a broad range of metabolites including sugars, amino acids, organic acids and other organic compounds. Cotyledon stage made a different cluster with the metabolites including sugars such as gulose, amino acids such as homoserine, lysine, and tryptophan and other organic acids. In order to clearly demonstrate the clustering, agglomerative hierarchical clustering (AHC) was conducted (Fig. 6-6). AHC was performed using Pearson dissimilarity and Euclidean distance matrix utilizing the Ward's linkage method and resulting dendrogram was presented in Fig. 6-6. In agreement with PCA, four different stages of flax seed development made three clusters. Group three comprised with torpedo stage and heart stage and group one and group two were comprised with mature green seed stage and cotyledon stage, respectively.

Table 6-1. Metabolite contents identified in four stages of flax seed development. Values were means of three replications with ± standard deviations.

Types	Metabolites	Mature green seed	±SD	Coty	±SD	Torpedo	±SD	Heart	±SD
	Alanine	8772.69	894.04	15613.10	1946.46	7602.06	1221.08	5741.66	683.69
	Leucine					128.77	54.60		
	Valine	3453.07	306.71	6310.63	820.34	9127.18	956.25		
	Isoleucine					9495.51	1861.98		
	Glycine	3742.63	71.58	4337.24	806.28	8319.50	994.45	11177.51	1766.86
	Serine	6177.38	528.77	937.08	426.05	13052.79	1157.26	11567.97	1131.08
	Threonine	3147.55	142.46	4149.97	554.94	4593.66	405.19		
	Methionine					178.54	20.94		
	β-Alanine	509.05	43.60	824.96	160.73	501.40	97.78	281.75	104.73
	Homoserine			273.77	98.73	1151.59	215.09		
	Proline	15167.91	1811.10	37200.96	3082.56	35978.40	2986.22	24567.00	1662.21
	Ornithine	3251.82	950.95			5346.41	1035.84	3010.96	808.25
	Glutamine	11030.55	3704.80	16765.53	3360.43	14819.69	346.86	58394.07	7941.42
	Phenylalanine	1318.94	278.02	2332.39	639.82	3627.44	667.18	2214.08	478.41
	N-Acetylglutamine	5713.07	971.15	30661.17	2935.62	32129.22	2131.85	26259.99	4573.92
	N-α-Acetyl-L-Lysine	1764.86	253.76			3014.51	83.94	2100.76	751.19
	3-Iodo-L-tyrosine					4819.65	387.93		
	Tryptophan	1565.82	406.24	3361.57	1010.86	3053.24	1.94	2283.21	798.62
	Norvaline					634.53	40.85		
	Cysteine	265.87	97.51	287.93	58.72	338.92	86.07	283.88	85.31
	Tyrosine	1858.78	497.56			3932.14	652.29		
	Asparagine	1559.87	572.98	1973.56	288.00			1146.64	297.55
	Lysine			335.13	25.82			250.01	24.77
	Pipecolic acid			81.56	0.51			288.10	21.64
Aminoacids	Piperidinecarboxylic acid					338.62	113.53		
	Phosphoric acid	11807.31	1900.72	194.28	74.40	132.11	1.87	9166.88	906.95
	Succinic acid	2276.34	757.32	3847.85	717.86	7215.39	534.13	5786.85	919.85
	Fumaric acid	614.46	22.22	825.57	120.74	1825.28	146.20	2087.08	579.26
	Malic acid	8661.02	919.90	24876.02	2630.89	66878.29	8345.31	58846.93	1603.99
	Butanoic acid			3189.13	694.02	4024.80	316.21	3392.49	504.23
	Threonic acid			2399.34	616.52	3088.83	135.99		
	Glutaric acid			19180.43	2164.48	1812.73	132.51	1566.37	94.43
	Pyruvic acid	8377.96	810.40	1040.59	159.89	19016.88	1242.11	12069.44	1412.27
	2-Ketoglutaric acid ditms	2897.65	250.35	12139.39	1747.17	14028.16	874.05	10921.01	1824.09
	Erythro-Pentonic acid					1928.05	149.97	1056.20	219.72
	α-Linolenic acid	217.78	61.66						
	Glutamic acid	11873.71	1895.76	65900.13	4378.91	75030.27	2536.75	11476.71	3239.57
Organic acids	Gluconic acid			1702.85	291.05				
	2-Furanone	275.66	44.97	613.43	67.26	727.43	50.00	482.09	77.68
	Mannose	5512.47	1010.44	8118.03	946.27	4158.14	186.12	3899.43	1006.64
	Fucose			27333.13	3355.18	45934.86	2538.38	43532.49	2508.03
	Fructose	11782.08	3417.96	84188.77	5600.45	134547.93	3950.85	105215.40	1371.07
	Butane	7329.67	797.86	80695.35	1036.35	114055.43	1096.40	94746.47	2983.18
	Galactose	46492.40	1850.86	107060.90	1436.92	120009.70	12315.45	119807.69	4572.12
	Erythrose			1022.37	36.36	1496.05	47.48	1282.53	180.70
	Glucose			64686.48	2337.58	87316.91	4587.30	68881.18	9764.70
	Gulose	27655.40	2639.21	196104.68	4583.73	136502.05	10718.56	126900.03	2338.41
	Xylose					118193.80	4196.32		
Sugars	Talose	23260.37	3926.25	3970.78	779.97	3453.06	193.17	4065.30	986.54
	Thiazole					86.89	12.77		
	l-2-Aminobutyric acid	1240.1	183.27	1104.43	208.98	1247.13	230.37	1012.86	79.33
	Urea	1148.79	12.21	803.63	100.31	1077.11	13.21	823.80	91.71
	Azetidinone					478.53	20.69		
	Thymine			6080.78	903.03	2401.94	647.46	1817.48	551.28
	butanedioate	18492.71	2420.09	4273.63	1189.98	1771.45	99.79	873.48	231.45
	1,4-Butanediamine					784.76	10.61	698.72	282.92
	Phenol			2235.09	368.95	4626.71	70.83	5433.91	970.11
	Erythro-Pentitol					3068.67	553.93		
	Ribofuranose	218478.18	8618.20			94149.66	12552.90	142042.11	5536.49
	Histamine	681.48	182.69						
	Ribose	17905.96	5121.91	49087.58	6457.20			44732.07	8879.47
	Ethylene glycol			438.53	65.64			6718.89	1733.34
	2-Piperidone			1487.26	280.61				
	Glycerol							5308.09	672.08
	Inosose							1683.15	528.06
	Erythrotetrofuranose							3610.30	1021.55
Other compounds	Xylopyranose							6383.00	1528.84

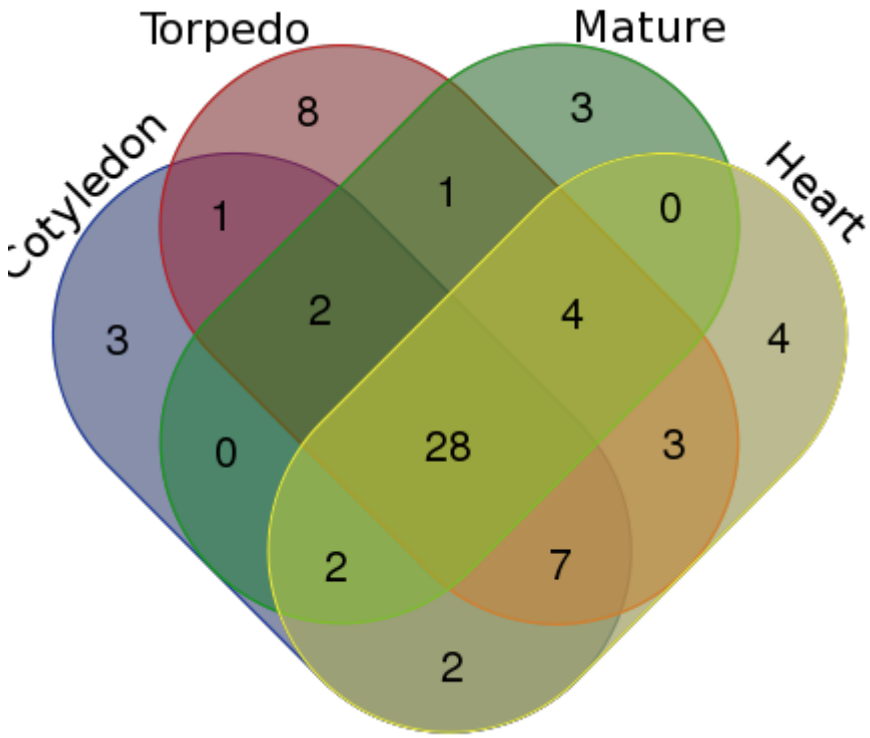


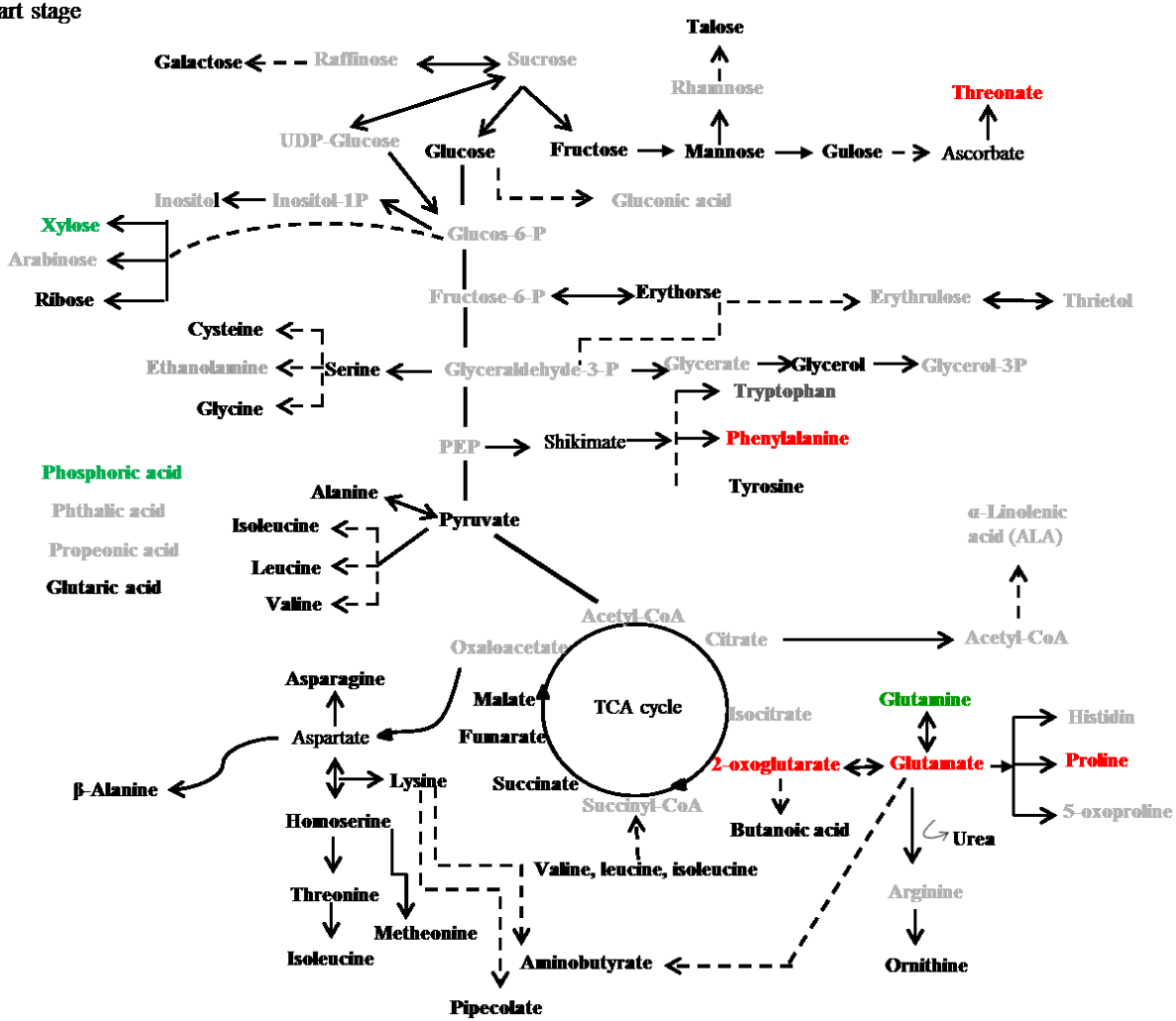
Figure 6-2: Venn diagram of the metabolites identified in the flax seed at different developmental stages. The diagram illustrated the number of the metabolites at four stages of flax seed development**.**

Table 6-2: Number of metabolites identified in different developmental stages of flax seed.

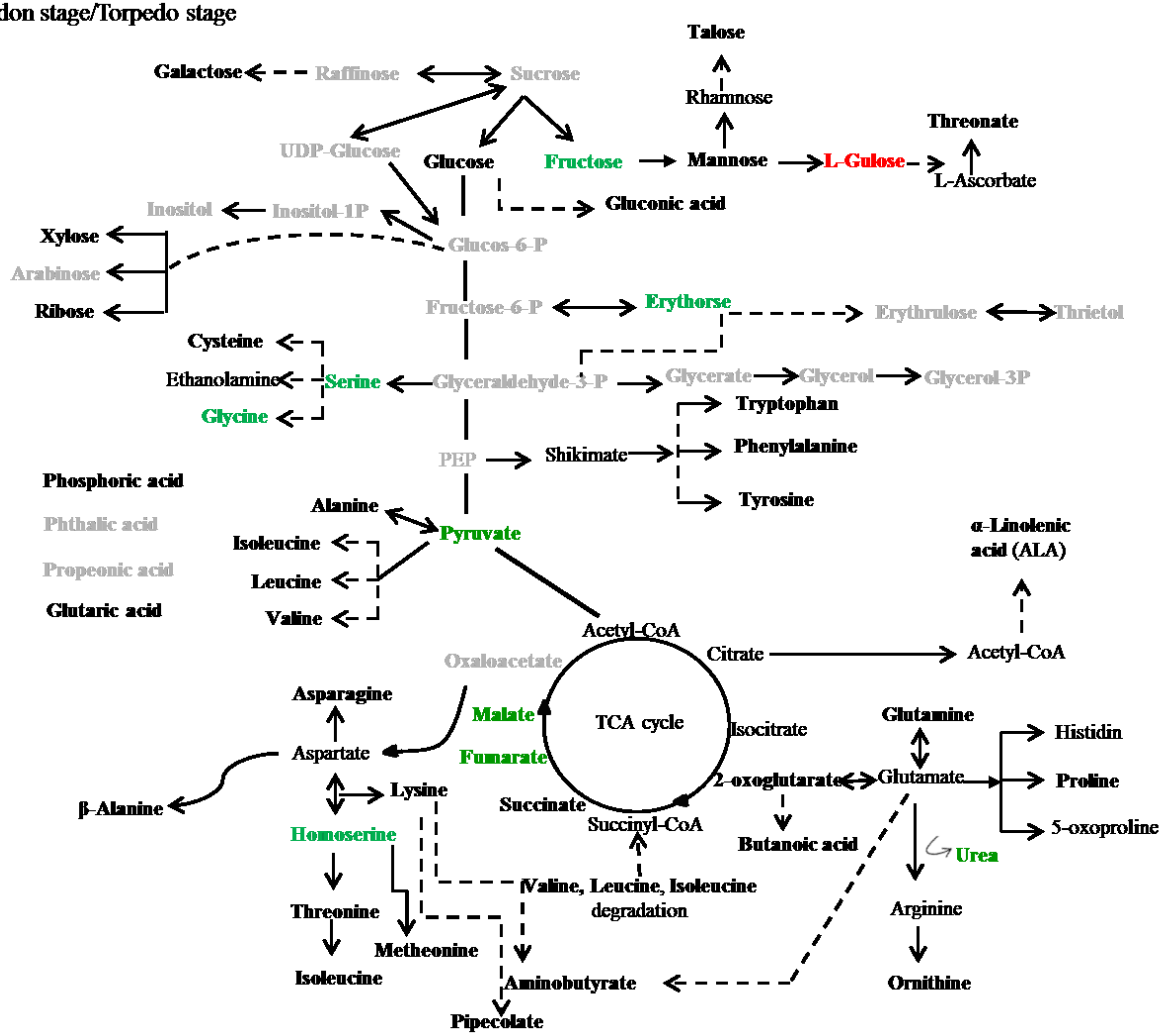
List names	**number of elements**	**number of unique elements**
Cotyledon stage	45	45
Heart stage	50	50
Mature green seed (MGS) stage	40	40
Torpedo stage	54	54
Overall number of unique elements		**68**
Names	**total**	**elements**

Cotyledon,Heart,MGS,Torpedo	28	Aminobutyric acid, Glycine, Glutamic acid, 2-Ketoglutaric acid, Cysteine, Talose, 2-Furanone, Fumaric acid, Glucose, Proline, Tryptophan, ß-Alanine, butanedioate, Phosphoric acid, Butane Mannose, Serine, Succinic acid, Alanine, Glutamine ,N-Acetylglutamine, Malic acid, Pyruvic acid, Urea, Galactose, Fructose, Gulose,Phenylalanine,
Cotyledon,Mature,Torpedo	2	Valine, Threonine
Cotyledon,Heart,Torpedo	7	Threonic acid, Thymine, Pipecolic acid, Butanoic acid, Phenol, Erythrose, Fucose
Cotyledon,Heart,MGS	2	Asparagine, Ribose
Heart, MGS,Torpedo	4	Xylose,Ribofuranose, Ornithine, N-Acetyl-L-Lysine
Cotyledon,Torpedo	1	Homoserine
Cotyledon,Heart	2	Ethylene glycol, Lysine
Mature,Torpedo	1	Tyrosine
Heart,Torpedo	3	1,4-Butanediamine, Pentanedioic acid, Erythro-Pentonic acid
Cotyledon	3	Gluconic acid, Piperidone ,Pentenoic acid
Torpedo	8	Erythro-Pentitol ,Thiazole, Methionine ,Azetidinone ,Iodo-L-tyrosine, Norvaline, Isoleucine, Leucine
MGS	3	Threitol, Histamine, a-Linolenic acid (ALA)
Heart	4	Erythrotetrofuranose, Xylopyranose, Inosose, Glycerol

A: Torpedo stage/Heart stage



B: Embryonic cotyledon stage/Torpedo stage



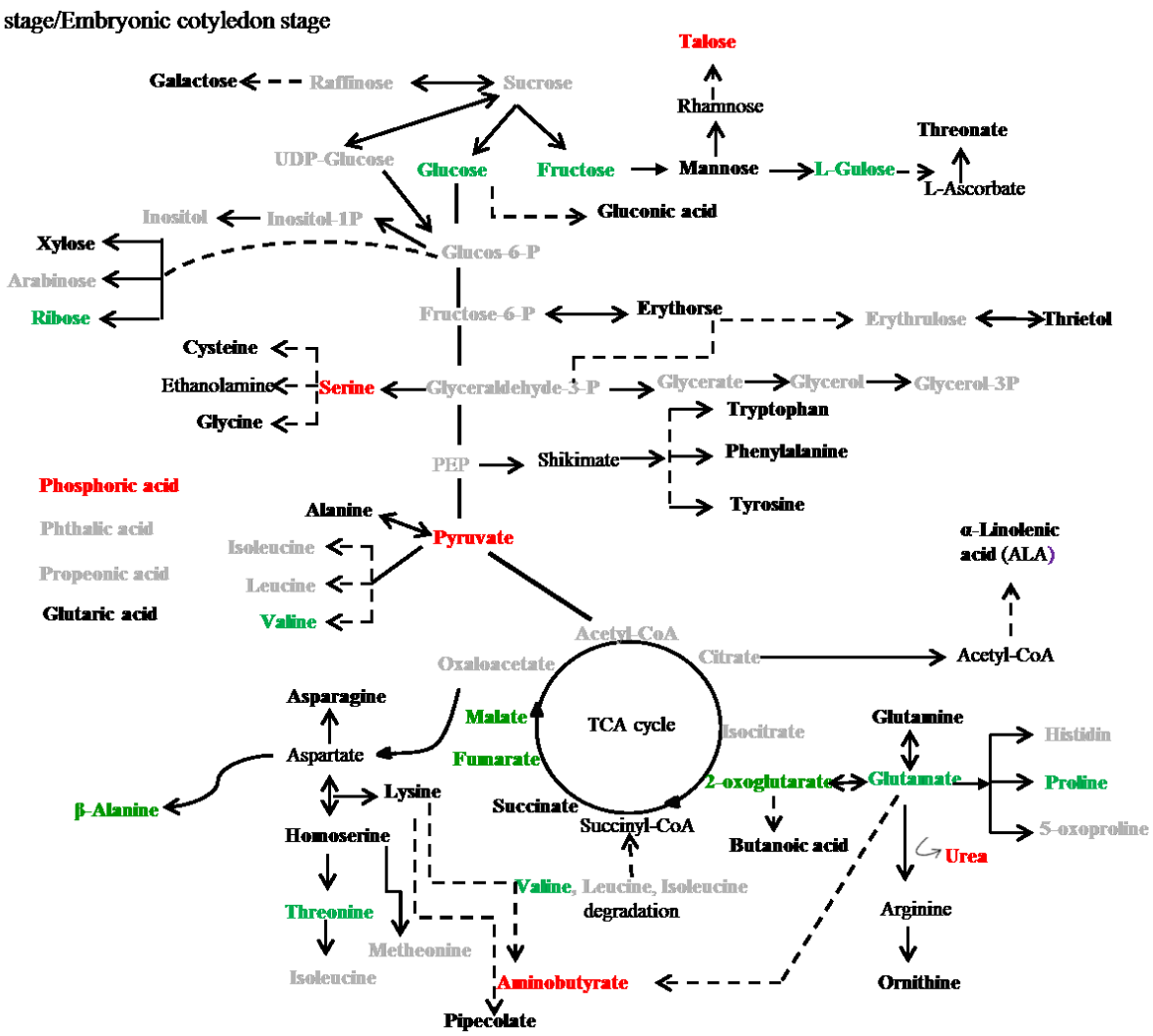


Figure 6-3: Changed in the metabolite contents during flax seed development. (A) The ratios between torpedo and heart, (B) cotyledon stage and torpedo stage, (C) cotyledon stage and mature green seed stage. The levels of significance was set at $p<0.05$. Black letters indicate no significant change, red letters indicate increase and green letters indicate decrease, ash letters indicate not identified.

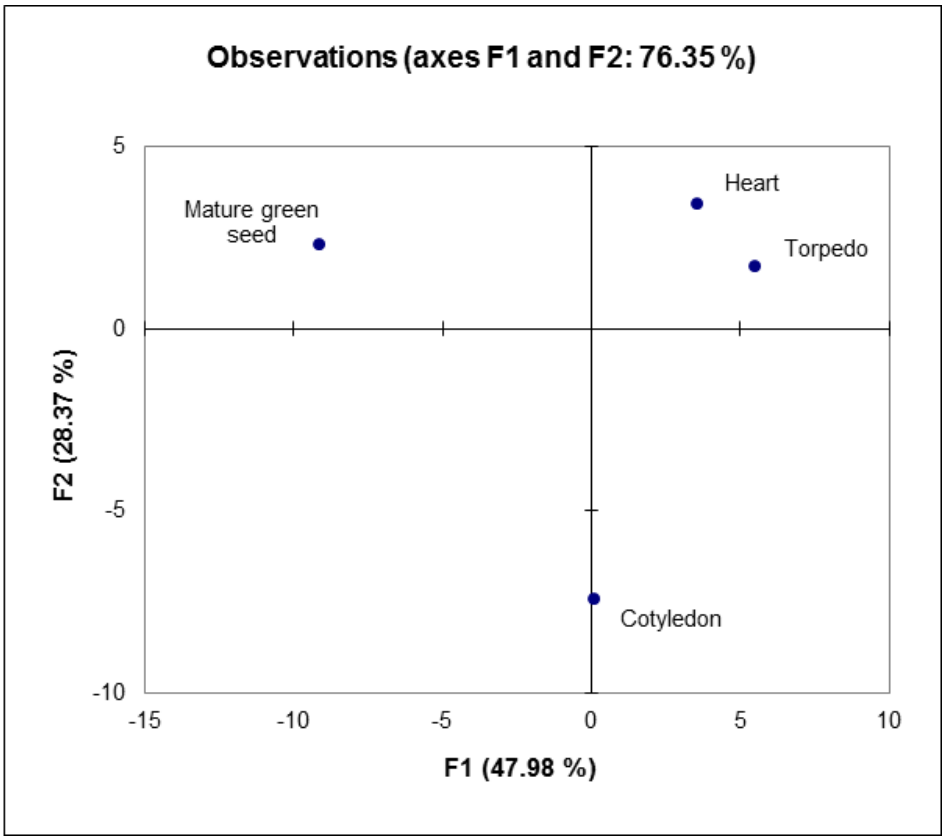
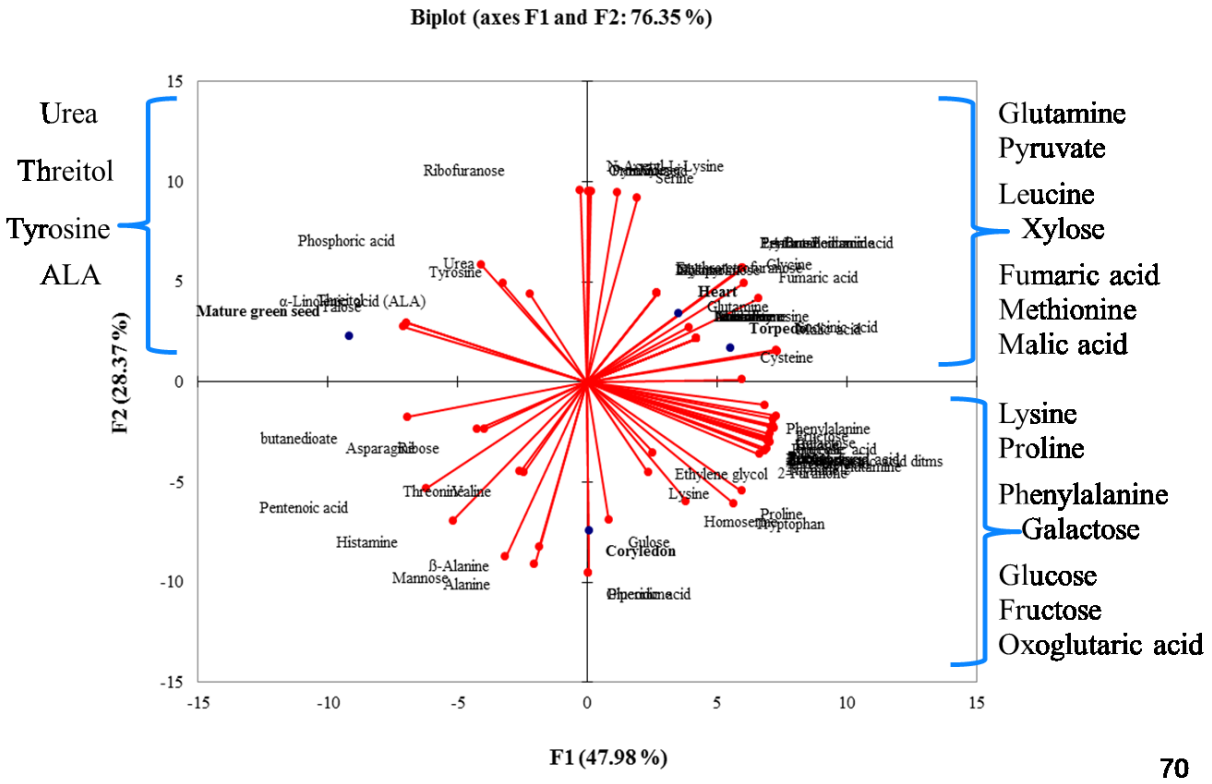
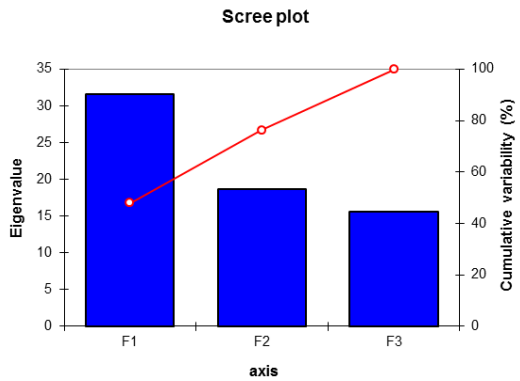


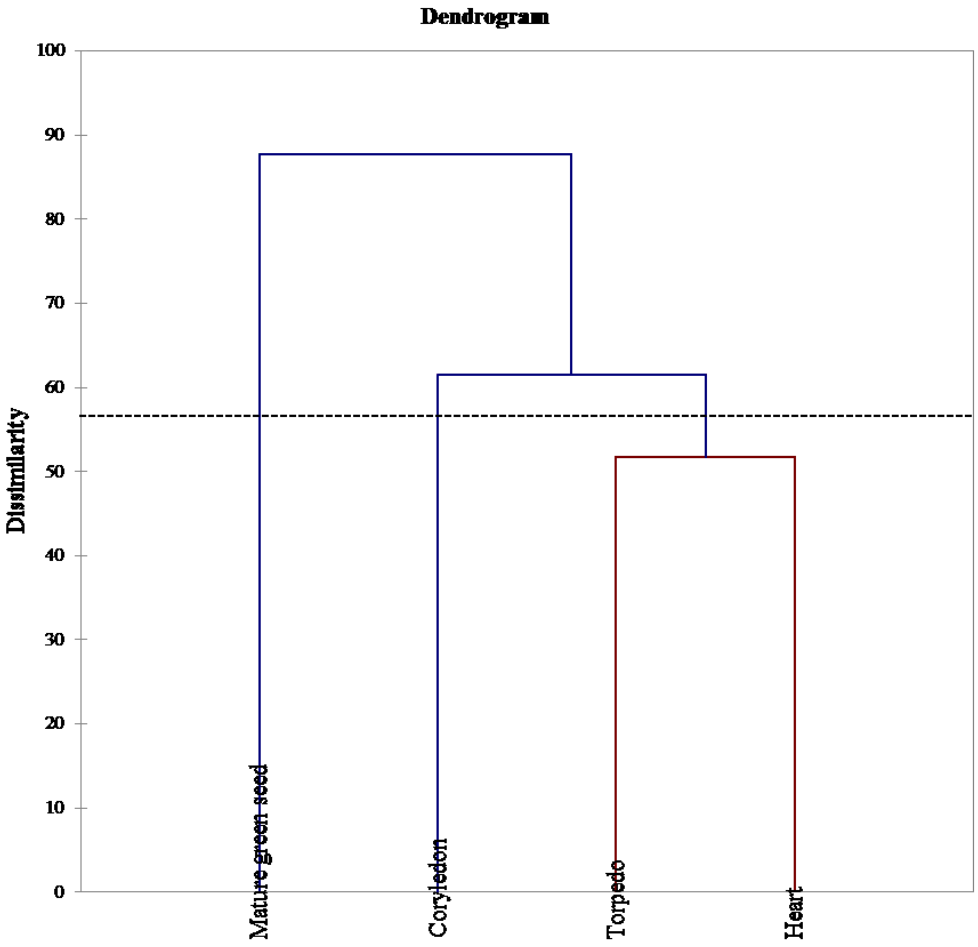
Figure 6-4. Scatter plot of different stages of flax seed based on the first two principal component analysis (PCA) axes. The percentage of variance explained by each axes was indicated.





Principal Component Analysis:			
Eigenvalues:			
	F1	F2	F3
Eigenvalue	31.666	18.722	15.612
Variability	47.980	28.367	23.654
Cumulative	47.980	76.346	100.000

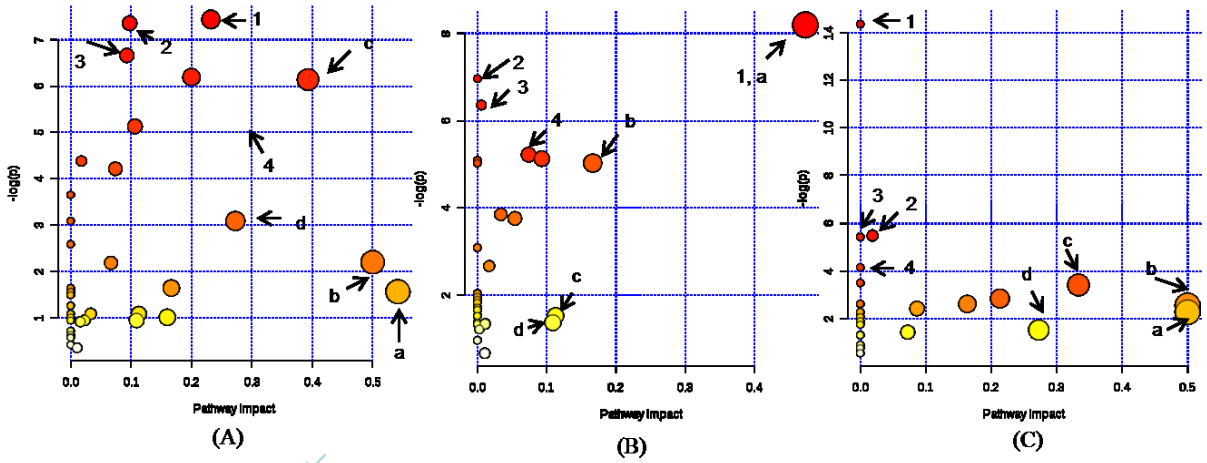
Figure 6-5. (A) PCA biplot of the mean centered metabolites data was shown. In the biplot both the developmental stages (dots) and the variables (vectors) were shown to enable interpretation about the relations. In total, 76.35% of the variation of the metabolites data was represented by PC1 (47.98%) and PC2 (28.37%). (B) Eigenvalues and cumulative variance of first three components were shown in the screen plot.



Proximity matrix (Euclidean distance):				
	Torpedo	Heart	Coryledon	ure green s
Torpedo	0	10.176	11.102	13.398
Heart	10.176	0	10.626	12.290
Coryledon	11.102	10.626	0	11.564
Mature green seed	13.398	12.290	11.564	0

Figure 6-6. Clustering result: Dendrogram used Euclidean distance metric. Cluster diagram for four stages of flax seed development classified by 52 metabolites.

Metaboanalyst Pathway Analysis:



6-7.1

	Mature green seed stage/Cotyledon stage							
	Significant pathway based on p-values							
ID	**Pathway Name**	**Hits**	**p**					
1	Arginine and proline metabolism	5	5.90E-04					
2	Alanine, aspartate and glutamate metabolism	4	6.40E-04					
3	Aminoacyl-tRNA biosynthesis	6	0.001281					
4	Pantothenate and CoA biosynthesis	3	0.002045					
	Significant pathway based on impact factor							
	Pathway Name	**Hits**	**Impact**					
a	beta-Alanine metabolism	1	0.54					
b	Isoquinoline alkaloid biosynthesis	1	0.50					
c	Glycine, serine and threonine metabolism	4	0.39					
d	alpha-Linolenic acid metabolism	2	0.27					
	Cotyledon stage/Torpedo stage				**ID**	**Torpedo stage/Heart stage**		
	Significant pathway based on p-values	**Hits**	**p**			**Significant pathway based on p-values**	**Hits**	**p**
	Pathway Name					**Pathway Name**		
1	Glycine, serine and threonine metabolism	4	2.74E-04		1	Aminoacyl-tRNA biosynthesis	8	5.92E-07
2	Pentose phosphate pathway	3	9.40E-04		2	Valine, leucine and isoleucine biosynthesis	3	0.00412
3	Alanine, aspartate and glutamate metabolism	3	0.0017253		3	Glucosinolate biosynthesis	4	0.00438
4	Lysine biosynthesis	2	0.0054009		4	Nitrogen metabolism	2	0.01557
	Significant pathway based on impact factor	**Hits**	**Impact**			**Significant pathway based on impact factor**	**Hits**	**Impact**
a	Glycine, serine and threonine metabolism	4	0.47		a	Phenylalanine, tyrosine and tryptophan biosynthes	2	0.5
b	Methane metabolism	2	0.17		b	Isoquinoline alkaloid biosynthesis	1	0.50
c	Pyruvate metabolism	1	0.11		c	Pyruvate metabolism	1	0.33
d	Glycolysis	1	0.11		d	Glycolysis	1	0.27

Figure 6-7. Summary of pathway analysis with MetaboAnalyst: metabolites altered in (a) mature green seed stage compared to cotyledon stage, (b) cotyledon stage compared to torpedo stage (c) torpedo stage compared to heart stage map to multiple biosynthetic pathways. 6.1: B. Statistics for pathways with major change based on the p value (pathways: 1-4) or on high impact (pathways: a-d). Colours in the pathways: light blue means metabolites are is not the data but used in the enrichment analysis, grey means the metabolite is not in the data and also excluded from enrichment analysis, from yellow to red means the metabolites are in the data with different levels of significance.

6.5. Discussion

Linseed has high nutritional value and its oil also has a wide range of industrial applications such as an ingredient of fine paints and varnishes. Recently, flax seed development has been described through proteomics and transcriptomics (Barvkar et al., 2012; Venglat et al., 2011). Unlike transcriptomics, metabolomics can describe the actual biochemical status of a tissue. Despite the importance of metabolomics, this is the first study to investigate the development of flax seed through non-targeted metabolic profiling.

The mature green seed stage is a reserve accumulation period. Therefore, the reductions in the levels of primary metabolites, including sugars, amino acids and organic acids (Fig. 6-3) between the embryonic cotyledon and the mature green seed stages could be involved in the synthesis of storage reserve accumulation. These patterns of change indicated that decreased metabolites were consumed in the biosynthesis of fatty acids and other amino acids that were precursor of the production of oil and storage proteins. Fait et al. (2006) conducted metabolic profiling using GC/MS in the green seed of Arabidopsis during reserve accumulation period (10 ± 1 to 17 ± 1 days after flowering). They reported the reductions of many primary metabolites including amino acids, sugars, polyols and organic acids and suggested their assimilation in the production of oil and storage proteins.

In this study, the MetaboAnalyst 2.0 server was used to conduct pathway analysis (Fig. 6-7). The server uses p-value (less than 0.05) and impact value scores (greater than 0.1) to trim the pathways that were significantly altered between the sample types. The impact score is calculated depending on the sum of the impact scores of each metabolites identified in a pathway and on the importance of the metabolites in that pathway. Based on the p-values, the first four pathways were found to be significantly altered in mature green seed stage compared to cotyledon stage involved amino acid metabolism with arginine and proline metabolism, alanine, aspartate and glutamate metabolism being affected. In addition to the amino acid metabolism, aminoacyl-tRNA biosynthesis pathway was altered. This suggested that perhaps in mature seed stage those amino acids were involved in the storage protein synthesis. In flax seed, most of the proteins are storage protein that constitutes ~23% of the flax seed (DeClercq et al., 2002). Gutierrez et al. (2006) reported that seed filling lasts from 20 to 30 days after flowering and Barvkar et al. (2012) suggested 16−30 days after anthesis is the seed filling stage. ALA was

identified only in the mature green seed stage and not during embryo development. This result was in agreement with the findings of Barvkar et al. (2012) in which they detected FAD2 and FAD3A proteins that play important role in fatty acid synthesis at 22 DAA (days after anthesis). Also, Banik et al. (2011) detected the transcript abundance of these enzymes at the similar developmental stage. Fatty acid synthesis in this stage also predicted by the reductions of glucose, malate, fumarate, and 2-oxoglutarate in mature green stage compared to cotyledon stage. In fatty acid synthesis, the carbon atoms and reducing power are provided by the citric acid cycle, and the pentose phosphate pathway, whereas glycolysis and oxidative phosphorylation provide the required ATP (Berg et al., 2002). The alterations in metabolic pathways based on impact value in mature seed stage compared to cotyledon stage showed that β -alanine metabolism had high impact value and the content of β-alanine decreased significantly in the mature seed stage (Fig. 6-3 and Fig. 6-7). β-alanine is one of the precursors of pentothenate and β-alanine betaine in plant. Pantothenate is an essential vitamin and precursor of coenzyme A (CoA). Pathway analysis also showed that pantothenate and CoA biosynthesis were altered significantly in the mature seed stage compared to the cotyledon stage. Biosynthesis of coenzyme A is a critical factor in lipid metabolism. This finding again emphasized that oil biosynthesis occurred in mature green seed stage and not during embryo development. Therefore, mature seed stage was important for fatty acid and oil accumulation.

Histamine was one of the unique metabolites identified in the mature green seed stage. Histamine has been identified in a number of plants (Smith, 1980) and remains as a free amine or as N-acetylated derivatives and as amides with organic acids (Luckner, 1990). Histamine is considered a secondary metabolite because it is not directly involved in the growth and development of plant. Notably, levels of GABA and glutamate significantly increased and decreased in mature green seed stage compared to cotyledon stage, respectively (Fig. 6-3). Gamma aminobutyrate (GABA) is produced from glutamate by glutamate decarboxylase (GAD). Fait et al. (2011) reported that conversion of glutamate to GABA by GAD during seed development plays an important role in balancing carbon and nitrogen metabolism and in storage reserve accumulation. Threitol along with ALA and histamine were identified as unique metabolites and could be considered as possible biomarkers for thr mature green seed stage of flax.

In the pathways analysis (Fig 6-7), one of the first four altered pathways identified based on p-value in torpedo stage compared to cotyledon stage was nitrogen metabolism. It is well known that nitrogen is important during embryo development. Three different stages of embryo development showed distinct metabolic profiling. In case of amino acids, Asparagine, lysine and were identified in the heart stage but not detected in the torpedo stage. On the other hand, pyruvate derived amino acids, valine, leucine, and isoleucine, as well as the aspartate-derived amino acids, homoserine, threonine, methionine were detected only in torpedo stage. Interestingly, higher number of metabolites as well as higher number of unique metabolites was obtained in torpedo stage, emphasized elaborated network pathways in the torpedo stage. Xiang et al. (2011) constructed stage transition metabolic networks using KEGG and embryo-specific global gene expression data of Arabidopsis. They found the stage transitions between quadrant and torpedo stages showed higher number of upregulated nodes than the other stages. 2-oxoglutarate and glutamate were increased and glutamine was decreased significantly in torpedo stage compared to cotyledon stage suggested possible utilization of glutamine in the formation of other amino acids. Amino acids, mainly asparagine and glutamine are the primary source of carbon and nitrogen to the developing embryo (Hsu et al., 1984; Rainbirdet al., 1984). Conversion of the amide amino acids to the other amino acids required for protein synthesis. It was reported that seed development demand increased methionine and suggested that methionine plays an important role in seed germination, priming and seed development in different plant species including flax (Barvkar et al., 2012; Shi et al., 2010; Gallardo et al., 2007 and Gallardo et al., 2001). In the present study, methionine was detected in torpedo stage but not in cotyledon and mature green seed stage. It was possible that methionine was utilized in the biosynthesis of storage protein in these stages. A potentially interesting finding from the pathways analysis in torpedo stage compared to heart stage was glucosinolate biosynthesis., Glucosinolate is a secondary metabolite and can identify in many important plant species such as Brassicales, *Ochradenus baccatus* contains glucosinolates (GLSs) (Samuni-Blank et al. 2012). Hydrolysis of glucosinolates by the enzyme glucosinolase or thioglucosidase release glucose and biologically active isothiocyanates, thiocyanates, and nitrile compounds (Halkier and Gershenzon, 2006). Flaxseed contains cyanogenic glycoside compounds that include linamarin, linustatin and neolinustatin (Shima et al., 2014). Glucosinolate biosynthesis and cyanogenic glycoside

biosynthesis share common amino acid precursors (Stotz, 2015). Moreover, the intermediate compounds such as α-nitrocarboxylic acid synthesized in glucosinolate biosynthesis pathway have been demonstrated to be an intermediate in cyanogenic glycoside pathway (Seigler 1997). Therefore, it is possible that glucosinolate biosynthesis pathway identified in the early embryonic seed development stage was involved in the generation of intermediates for the biosynthesis of cyanogenic glycoside compounds in flax seed. In the embryonic cotyledon stage, unique metabolites: gluconic acid, piperidone, and pentenoic acid were identified and could be used as possible biomarkers for this stage. Numerous pathways were suggested to be affected during the embryogenesis. Significantly altered pathways were the amino acid biosynthesis pathways, pentose phosphate pathway, nitrogen metabolism, glycolysis and pyruvate metabolism. This showed that core metabolic pathways were operated during embryogenesis. Xiang et al. (2011) also reported that embryogenesis of Arabidopsis involved glycolysis, pentose phosphate pathway, pyruvate metabolism, and carbon fixation pathways.

6.6. Conclusion

In conclusion, this study provided a global view of the complex metabolic processes that occurred during different stages of flax seed development. Each stage had its own unique metabolic profile. Metabolites identified during the embryo development were involved in the central metabolic pathways such as amino acid metabolism, glycolysis, pyruvate metabolism and pentose phosphate pathways. On the other hand, in case of mature green seed stage, in addition to the central metabolic pathways, alterations also involved biosynthesis of storage proteins and oil. Increases in ALA and β-alanine were detected in mature green seed stage but not during embryo development. Decreases in the abundance of primary metabolites including amino acids and sugars in mature green stage indicated their possible assimilation into storage compounds. Methionine was identified in the torpedo stage but was not detected in the cotyledon stage or mature green seed stage, which may indicate that it too was assimilated into storage proteins during cotyledon and mature green seed stages. Biosynthesis of GABA from glutamate was also detected in mature green seed stage. A large number of unique metabolites were identified in the torpedo stage, showing active secondary metabolites. Glucosinolate biosynthesis was predicted during embryogenesis which possibly involves in the production of cyanogenic glycosides

compounds in flax seed. Three biomarker metabolites were identified for mature green seed stage, ALA, histamine and thritol. Present results also revealed that amino acid biosynthesis played an important role in generating several classes of compounds related to the embryo development and seed quality.

The data obtained in the present study will provide a resource for future studies of these metabolites for better understanding of the molecular mechanisms that govern the development of flax seed and seed quality.

7. Chapter 7

7.1. General discussion and future directions

7.1.1. General discussion

There are ten DUF642 domain-containing genes in Arabidopsis. Previous study with two DUF642 genes *DGR2* (At5g25460) and *DGR1* (At1g80240) showed that *DGR2* was responsive to L-Galactono-1,4-lactone (L-GalL), a terminal precursor for ascorbic acid (AsA) (Gao et al., 2012). However, they didn't identify any significant difference in the AsA levels between the *dgr2* mutants and the controls, and both genes were expressed complementarily during development of Arabidopsis. Moreover, a loss of function Salk line for *DGR2* (SALK_125079) showed shorter roots and smaller rosettes than Col-0. On the other hand, antisense RNA and Salk lines (SALK_142260 and SALK_054867) of another DUF642 domain-containing protein At4g32460 resulted in very short siliques with no seeds (Zúñiga-Sánchez et al., 2014). Zúñiga-Sánchez et al., (2014) also suggested that the DUF642 proteins encoded by At4g32460 and At5g11420 could be positive regulators of PME activity during several developmental processes. The present study was designed to better characterize two DUF642 genes *DGR2* (At5g25460) and *DUFB* (At5g11420) in Arabidopsis using a range of techniques, which included protein and gene expression analysis, mutant analysis and metabolic profiling for pathway analysis. A part of the Ph. D. research work also involved metabolic profiling of developing flax seeds by GC/MS.

3D model structures of DGR2 and DUFB were generated by 1-TASSER server and the modified models were validated by Ramachandran plots using RAMPAGE server (Section 2.2.3 and Section 2.2.4. The percentage of residues in the allowed/favored region was more than 95% for both DGR2 and DUFB. Based on structural homology, DGR2 and DUFB were predicted to be involved in primary metabolic process and cell adhesion with DGR2 predicted to have carbohydrate binding and hydrolase activity whereas DUFB was predicted to have carbohydrate binding and carbon-oxygen lyase activity (Section 2.2.5).

qRT-PCR analysis showed that *DGR2* and *DUFB* were expressed at relatively high levels in root and flower tissues (Fig. 4-6B & Fig. 4-7B). Expression of both genes was also detected in the stems, siliques and rosettes but at lower abundance relative to root and flower (Fig. 4-6A & Fig. 4-7A). Publicly available gene expression data (Arabidopsis eFP browser; Winter et al., 2007) likewise indicated that *DGR2* and *DUFB* were expressed in different organs of Arabidopsis including flowers, rosettes, stems, and roots (Section 2.2.1). GUS reporter gene assays with upstream genomic regions of *DGR2* and *DUFB* showed that both genes were expressed in the roots of Arabidopsis in a complementary manner (Chapter 3). *DGR2* was expressed in roots including root tips, whereas *DUFB* was expressed in the roots excluding tips (Fig 3-2 & 3-3). Also, the *DGR2* reporter fusion was expressed at the later stages of lateral root primordium (LRP) development. At later stages of LRP, some cells are participated in cell division and others in cell elongation towards the tips coupled with the elongation of the primordium (Szymanowska-Pułka et al., 2012). On the other hand, *DUFB* reporter fusion was not expressed during LR formation (Fig. 3-7). This suggested that DGR2 is involved in lateral root formation and participates in cell division or cell elongation. Treating Arabidopsis Col-0 plants with exogenous auxin (IAA) resulted in induction of *DGR2* and *DUFB* expression (Fig. 3-9; Fig. 3-5 and Fig.3-6). Following IAA treatment, increased GUS staining intensity was observed in the root apex of *DGR2* whereas for *DUFB,* increased GUS staining was observed in the transition and elongation zone of the root apex. So, differential expression of *DGR2* and *DUFB* genes in response to exogenous auxin treatment and expression in different organs under normal growth conditions as shown by GUS, together suggested roles in auxin-regulated root development process and control of cell division and elongation, respectively. *DGR2* and *DUFB* were also induced by exogenous ACC, and that expression of both *DGR2* and *DUFB* decreased following treatment by MeJA (Fig.3-9). This could be due to the impairment of ethylene and auxin biosynthesis by methyl jasmonate (Soto et al., 2012). Mutant analysis with loss and gain of functions mutant also showed that root length was drastically reduced in 35S::*DGR2* (Fig. 4-19). From the above findings of the present study, it can be suggested that *DGR2* and *DUFB* are involved in auxin-mediated plant development.

In this study, DGR2 and DUFB recombinant proteins were heterlogously expressed in *E. coli* but sufficiently purified proteins for further biochemical analysis could not be obtained. This could be due to a low level of expression of these proteins and to co-elution with native *E.coli* proteins. The reasons for the low level of expressions of both DGR2 and DUFB either expressed proteins were toxic to *E. coli* or due to nature of the protein properties which include half life of the protein.

Co-immunoprecipitation assay results (Table 4-4) showed that these proteins co-eluted with beta-glucosidase (At3g18080), which belongs to the glycosyl hydrolase family 1. *In silico* characterization (Chapter 2) of *DGR2* and *DUFB* indicated that both are polysaccharide hydrolysing enzymes. Therefore, Co-IP result strengthens the hypothesis that DGR2 and DUFB involve in the degradation of cell wall polysaccharide. During cell division and eleongation, many glycoside hydrolases enzymes involved in remodeling of cell wall.

Previous publications have described DGR2 as being localized to cell wall (Bayer et al., 2006, Irshad et al., 2008, Negri, et al., 2008), based on cell wall proteomics. In the current *in vivo* subcellular localization study, using translational fusions to a fluorescent reporter (35S::DGR2:CiPF), a punctate pattern of fluorescent localization was observed within cytoplasm of transgenic plants (Fig. 4-10). This pattern was typical of Golgi-localized proteins. Co-localization study by immunohistochemistry using early Golgi specific marker Sec21 with fixed root tissue of 35S::DGR2:CiPF revealed that the fluorescent reporter was not fully co-localized with Sec21 but was in the immediate vicinity to Sec21 (Fig. 4-11) which indicated that DGR2 is not localized in the cis/middle Golgi but in the *trans*-Golgi network (TGN). No DUF642 proteins were detected in a previously published Golgi proteomics survey (Harriet et al., 2012). Previous cell wall proteomics studies identified DUF642 proteins in the cell wall extracts (Bayer et al., 2006, Irshad et al., 2008, Negri, et al., 2008). TGN is a highly mobile organelle and can be closely associated with a Golgi or another TGN and can be located at a distance independently and serves as a major sorting center for the biosynthetic cargo coming from the Golgi and destined either to the PM/cell wall/cell plate or to the vacuole and as an early endosome. (Otegui

et al., 2006; Viotti et al., 2010; Kang et al., 2011; Park and Jürgens, 2012). This makes it challenging to distinguish cargo proteins from TGN residents (Groen et al., 2014). Moreover, organelle-centric proteomics rely on the purification and significant enrichment of the organelle of interest, which is not achievable for many organelles. Incomplete separation of organelles leads to false discoveries, with erroneous assignments (Groen et al., 2014)

In the post-genomic era, metabolic profiling is a useful tool for elucidating gene function. In this study, metabolic phenotyping of six different genotypes of DUF642 genes were performed by GC-MS. The results revealed that amino acid biosynthesis was perturbed in *DGR2* and *DUFB* mutants. Also, DGR2 and DUFB might be inbolve in carbohydrate metabolism which were indicated by the perturbation of TCA cycle (Fig. 5-5, Fig. 5-6, Fig.5-8 and Fig. 5-9) that generates ATP for celluar functions. Principal component analysis (Fig. 5-10 and Fig. 5-11) showed distinct metabolites compositions associated with each genotype, including amino acids, sugars, organic acids and other organic compounds, which could be regarded as potential biomarkers to elucidate the functions of DUF642 genes. Metabolic profiling also revealed that 35S::DUFB plants had an increase in fucose and decrease in xylose whereas 35S::DGR2 plants were increased in xylose. These sugars are some of the minor constituents of pectin. It is possible that DGR2 and DUFB both might be participate in cell wall changes by targeting pectin. Previous *in vitro* studies revealed that At4g32460- and At5g11420-encoded proteins interact with the catalytic domain of pectin methyl esterase 3 (AtPME3, which is encoded by At3g14310). Moreover, At4g32460 antisense RNAi lines showed decreased PME activity in the leaves (Zúñiga-Sánchez et al., 2014).

Flax is important for its valuable characteristics of both linseed oil and fiber. Linseed has a wide range of industrial applications and tremendous health benefits. Metabolic switches occur during seed development (Weber et al., 2005). Therefore, metabolic profiling can provide important insights into biochemical pathways and novel metabolic intermediates of flax seed during development. Flax seeds not only contain coumpounds that benefits heat it also contain harmful compounds. So, the results could benefit metabolic pathway engineering to improve flax seed

quality. In the present study, metabolic profiling of developing flax seed (Heart stage, torpedo stage, cotyledon stage and mature green seed stage) was performed using GC/MS. The result showed that in the mature green seed stage the level of primary metabolites, including sugars, amino acids and organic acids, were decreased compared to the embryo development stages (Fig. 6-3). This was indicative that decreased metabolites were consumed in the biosynthesis of fatty acids and other amino acids that were precursor of the production of oil and storage proteins. In flax seed, most of the proteins are storage protein that constitutes ~23% of the flax seed (DeClercq et al., 2002). The major component of oils biosynthesis, ALA and β-alanine were detected only at mature green seed stage. Therefore, oil biosynthesis occurred in the mature green seed stage. Gamma aminobutyrate (GABA) is produced from glutamate by glutamate decarboxylase (GAD). In the mature green seed stage, GABA was increased significantly and glutamate decreased significantly compared to embryo development. This showed that increased amounts of GABA were synthesized at this stage. Pathway analysis showed that nitrogen metabolism was significant at torpedo stage compared to cotyledon stage, probably because a large amount of nitrogen is required at early stage of embryo development. It was also found that flax seed contained glucosinolate that formed during embryogenesis.

7.1.1.1 Concluding remarks

Our results showed that DGR2 and DUFB are expressed in all major organs which suggested that they play an important role in growth and development of Arabidopsis. GUS expression study revealed that despite high sequence similarity, these proteins functions in the complementary manners. *DGR2* was highly expressed in the root apex, contains zones of cell division and rapid elongation. *DGR2* also expressed in the later stages of lateral root primordia, stages of actively dividing and elongating cells and was upregulated by IAA, especially in the root apex. On the other hand, *DUFB* was not expressed in the root apex and developing lateral roots primordia and was upregulated by IAA treatments in the elongation zone of root apex. From these findings of GUS study, we speculated that DGR2 involves in both cell division and cell elongation whereas DUFB promotes cell elongation but not cell division. Subcellular localization of DGR2 revealed possible localization in the trans-Golgi network which indicated that DGR2 might involve in the protein sorting or endocytotic process. We successfully

generated 3D models of DGR2 and DUFB and predicted their function in primary metabolic process. We have presented metabolomics-based evidence that support the functional roles DGR2 and DUFB are coupled with the primary metabolic pathways. Metabolic profiling and pathway analysis of DGR2 and DUFB mutants showed perturbation of the TCA cycle activity and provoke considerable impact on xylose, mannose, arabinose and galactose. It is possible that both genes are involved in carbohydrate metabolism, probably pectin. Therefore, mutations of DGR2 and DUFB caused change in the carbohydrate metabolic process thus changed in the energy demands as reflected by perturbation of TCA cycle. We found that DGR2 and DUFB have predicted galactose binding domain like and hydrolase fold which lead to the hypothesis that they are enzymes with the glycoside hydrolase activities. Experimental evidence is required to prove this hypothesis which include enzyme assay with purified protein. We expressed both proteins in *E. coli* but were unable to obtain sufficiently pure protein to perform any enzyme assay. So, it is clear that future research aimed at identifying the precise enzymatic roles to prove this hypothesis which could lead to emerge exciting new models of the carbohydrate metabolism and probably cell wall biosynthesis.

We also performed metabolic profiling of flax seed to get an insight of metabolic shifts during development which contribute to the formation of its exceptional properties. We could successfully monitor the alterations in several major groups of compounds and pathway perturbation during flax seed development. Each stage of seed development has its own unique metabolic profiles with the most drastic changes occurred at the transition toward green seed stage where decreased in the abundance of primary metabolites indicated their possible assimilation into storage compounds and in the torpedo stage where majority of unique metabolites were identified.

7.1.2. Future directions

3D models of DGR2 and DUFB were generated and evaluated in this study. In addition, 3D models of other DUF642 proteins can be generated and thereby can be predicted the functions of all 10 DUF642 genes in Arabidopsis as well as can submitted these models in the Protein Data Bank (PDB). DUF642 proteins contain domain of unknown function, therefore structural homologues might not exist in the PDB database. Zhang, et al. (2009) suggested that if structural homologues do not exist, or exist but cannot be identified then models are required to be

constructed by ab initio modelling. Unfortunately, accuracy of *ab initio* modelling is low and the success is limited to small proteins (<100residues) (Zhang, et al., 2009). However, the quality of the 3D protein structures generated by 1-TASSER (homology modelling) could be improved by subjected to energy minimization which can eliminate distorted geometries by moving atoms to release internal constraints using the available online servers.

Subcellular localization of a protein is associated with its function, so it is important to know where the protein of interest is localized. *In vivo* protein localization of DGR2 showed punctate patterns and remain elusive where it was localized. It could be localized other compartment of cell that also showed punctate localization such as PVC or TNG. On the other hand, C-terminus CiFP tagging might resulted in dislocation. To identify, where DGR2 localized, the following experiments can be conducted:

-Colocalization study with Trans Golgi marker and PVC marker using fixed cells of 35S::DGR2:CiFP

-Identify subcellular localization of DGR2 by DUF642 antibody and using immunogold transmission electron microscopy method.

-Generating transgenic plants by tagging fluorescent proteins at N-terminus

In vitro studies with purified recombinant protein often provide the most definitive proof of a bona fide activity for a specific enzyme. In this study, DGR2 and DUFB were expressed heterologously in *E.coli* but no purified protein could be obtained to conduct downstream analysis, presumably because the level of protein expression was very low and histidine containing proteins of *E.coli* were co-eluted as contaminants during purification. Moreover, it is not known if DGR2 and DUFB undergo post transcriptional modification and become functional. So, instead of using *E. coli*, a eukaryotic host such as yeast or insect can be used for obtaining DGR2 and DUFB recombinant proteins.

It is still unknown whether DGR2 and DUFB are enzymes. But they are predicted to be enzymes with hydrolase and galactose binding like fold. Identifying an unknown protein's substrate is a challenge and time consuming. Metabolic profiling approaches can be used to identify the substrate of enzymes. In these approaches, cell extract can be incubated with purified DGR2 and DUFB recombinant proteins after expressing in a eukaryotic host and then metabolic profiling can be performed to observe changes in the abundance of the metabolites in the presence of the enzyme. In the present study, metabolic profiling identified precursor metabolites and elucidated pathways. Next, the combination of high-resolution GC–MS-based untargeted metabolomics with stable isotope tracing can be performed to obtain a global overview of the cellular fate of precursor metabolites identified in this study.

8. Bibliography

Acharya BR, Raina S, Maqbool SB, Jagadeeswaran G, Mosher SL, Appel HM, Schultz JC, Klessig DF, Raina R. 2007. Overexpression of CRK13, an Arabidopsis cysteine-rich receptor-like kinase, results in enhanced resistance to Pseudomonas syringae. Plant J;50(3):488-99.

Albenne C, Canut H, Jamet E. 2013. Plant cell wall proteomics: the leadership of Arabidopsis thaliana. Front Plant Sci.;4:111.

Albert VA, Soltis DE, Carlson JE, Farmerie WG, Wall PK, Ilut DC, Solow TM, Mueller LA, Landherr LL, Hu Y, Buzgo M, Kim S, Yoo MJ, Frohlich MW, Perl-Treves R, Schlarbaum SE, Bliss BJ, Zhang X, Tanksley SD, Oppenheimer DG, Soltis PS, Ma H, DePamphilis CW, Leebens-Mack JH. 2005. Floral gene resources from basal angiosperms for comparative genomics research. BMC Plant Biol.;5:5.

Aliferis KA, Faubert D, Jabaji S. 2014. A Metabolic Profiling Strategy for the Dissection of Plant Defense against Fungal Pathogens. Published online 2014 Nov 4.

Alm E, Arkin AP. 2003. Biological networks. Curr Opin Struct Biol.;13(2):193-202.

Andersen JS, Mann M.2006. Organellar proteomics: turning inventories into insights. EMBO Rep.;7(9):874-9.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet.;25(1):25-9.

Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA 3rd, Orlando R, Scheller HV, Mohnen D. 2011. Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proc Natl Acad Sci U S A;108(50):20225-30.

Azpiroz-Leehan R, Feldmann KA. 1997. T-DNA insertion mutagenesis in Arabidopsis: going back and forth. Trends Genet.;13(4):152-6.

Baldwin D, Crane V, Rice D. 1999. A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr Opin Plant Biol.;2(2):96-103.

Baluska F, Kubica S, Hauskrecht M. 1990. Postmitotic‘isodiametric’cell growth in the maize root apex. Planta;181(3): 269–274.

Baluska F, Vitha S, Barlow PW, Volkmann D. 1997. Rearrangements of F-actin arrays in growing cells of intact maize root apex tissues: a major developmental switch occurs in the post mitotic transition region. Eur J Cell Biol.;72(2):113-21.

Baluska F, Volkmann D, Barlow PW. 1996. Specialized zones of development in roots: view from the cellular level. Plant Plant Physiol.;112(1):3-4.

Banik M, Duguid S, Cloutier S. 2011. Transcript profiling and gene characterization of three fatty acid desaturase genes in high, moderate, and low linolenic acid genotypes of flax (Linum usitatissimum L.) and their role in linolenic acid accumulation. Genome.;54(6):471-83.

Bannai H1, Tamada Y, Maruyama O, Nakai K, Miyano S. 2002. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics;18(2):298-305.

Barvkar VT, Pardeshi VC, Kale SM, Kadoo NY, Giri AP, Gupta VS. 2012. Proteome profiling of flax (Linum usitatissimum) seed: characterization of functional metabolic pathways operating during seed development. J Proteome Res.;11(12):6264-76.

Bastiaan O R Bargmann, Steffen Vanneste, Gabriel Krouk, Tal Nawy, Idan Efroni, Eilon Shani, Goh Choe, Jiří Friml, Dominique C Bergmann, Mark Estelle, and Kenneth D Birnbauma, 2013. A map of cell type-specific auxin responses. Mol Syst Biol.; 9: 688.

Bayer EM, Bottrill AR, Walshaw J, Vigouroux M, Naldrett MJ, Thomas CL, Maule AJ. 2006. Arabidopsis cell wall proteome defined using multidimensional protein identification technology. Proteomics.;6(1):301-11.

Benková E, Hejátko J. 2009. Hormone interactions at the root apical meristem. Plant Mol Biol.;69(4):383-96.

Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu I, Zoeckler B, Montoya M, Miller N, Weems D, Rhee SY. 2004. Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol.;135(2):745-55.

Berg JM, Tymoczko JL, Stryer L. 2002. Biochemistry, W. H. Freeman & Co., New York:974

Berleth T, Krogan NT, Scarpella E. 2004. Auxin signals--turning genes on and turning cells around. Curr Opin Plant Biol.;7(5):553-63.

Berleth T, Mattsson J, Hardtke CS. 2000. Vascular continuity and auxin signals. Trends Plant Sci.;5(9):387-93.

Beyer A, Hollunder J, Nasheuer HP, Wilhelm T. 2004. Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale. Mol Cell Proteomics.;3(11):1083-92.

Bhatty R. 1995. Nutrient composition of whole flaxseed and flaxseed meal. In: Cunnane, S.C., Ed. L.U. Thompson, Flaxseed in Human Nutrition. AOCS Press, Champaign, USA:22–42.

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN. 2003. A gene expression map of the Arabidopsis root. Science;302(5652):1956-60.

Blaby-Haas CE, de Crécy-Lagard V., 2011, Mining high-throughput experimental data to link gene and function. Trends Biotechnol.;29(4):174-82.

Bleecker AB, Patterson SE. 1997. Last exit: senescence, abscission, and meristem arrest in Arabidopsis. Plant Cell.;9(7):1169-79.

Blilou I, Xu J, Wildwater M, Willemsen V, Paponov I, Friml J, Heidstra R, Aida M, Palme K, Scheres B. 2005. The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. Nature.;433(7021):39-44.

Boisvert FM, Lam YW, Lamont D, Lamond AI. 2010. A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. Mol Cell Proteomics.;9(3):457-70.

Bolle C, Schneider A, Leister D. 2011. Perspectives on Systematic Analyses of Gene Function in Arabidopsis thaliana: New Tools, Topics and Trends. Curr Genomics.;12(1):1-14.

Borderies G, Jamet E, Lafitte C, Rossignol M, Jauneau A, Boudart G, Monsarrat B, Esquerré-Tugayé MT, Boudet A, Pont-Lezica R. 2003. Proteomics of loosely bound cell wall proteins of Arabidopsis thaliana cell suspension cultures: a critical analysis. Electrophoresis;24(19-20):3421-32.

Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y. 1998. Predicting function: from genes to genomes and back. J Mol Biol.;283(4):707-25.

Borner GH, Lilley KS, Stevens TJ, Dupree P. 2003. Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. Plant Physiol.;132(2):568-77.

Boudart G, Jamet E, Rossignol M, Lafitte C, Borderies G, Jauneau A, Esquerré-Tugayé MT, Pont-Lezica R. 2005. Cell wall proteins in apoplastic fluids of Arabidopsis thaliana rosettes: identification by mass spectrometry and bioinformatics. Proteomics;5(1):212-21.

Brieger A, Plotz G, Hinrichsen I, Passmann S, Adam R, Zeuzem S. 2012. C-terminal fluorescent labeling impairs functionality of DNA mismatch repair proteins. PLoS One.;7(2):e31863.

Brown DM, Goubet F, Wong VW, Goodacre R, Stephens E, Dupree P, Turner SR. 2009. Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant J.;52(6):1154-68.

Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT. 2005. Protein structure prediction servers at University College London. Nucleic Acids Res.;33(Web Server issue):W36-8.

Burns N, Grimwade B, Ross-Macdonald PB, Choi EY, Finberg K, Roeder GS, Snyder M. 1994. Large-scale analysis of gene expression, protein localization, and gene disruption in Saccharomyces cerevisiae. Genes Dev.;8(9):1087-105.

Caffall KH, Mohnen D. 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr Res.;344(14):1879-900.

Cai Y, Jia T, Lam SK, Ding Y, Gao C, San MW, Pimpl P, Jiang L. 2011. Multiple cytosolic and transmembrane determinants are required for the trafficking of SCAMP1 via an ER-Golgi-TGN-PM pathway. Plant J.;65(6):882-96.

Carpita NC. 1996 Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol.;47:445-476.

Carter JF. 1993. Potential of flaxseed and flaxseed oil in baked goods and other products in human nutrition. Cereal Foods World;38:753–759.

Cavalier DM, Lerouxel O, Neumetzler L, Yamauchi K, Reinecke A, Freshour G, Zabotina OA, Hahn MG, Burgert I, Pauly M, Raikhel NV, Keegstra K. 2008. Disrupting two Arabidopsis thaliana xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component. Plant Cell.;20(6):1519-37.

Chanhui Lee, Ruiqin Zhong, and Zheng-Hua Ye. 2012. Biochemical characterization of xylan xylosyltransferases involved in wood formation in poplar. Plant Signal Behav.; 7(3): 332–337.

Charmont S, Jamet E, Pont-Lezica R, Canut H. 2005. Proteomic analysis of secreted proteins from Arabidopsis thaliana seedlings: improved recovery following removal of phenolic compounds. Phytochemistry;66(4):453-61.

Chen K, Fan B, Du L, Chen Z. 2004, Activation of hypersensitive cell death by pathogen-induced receptor-like protein kinases from Arabidopsis. Plant Mol Biol.;6(2):271-83.

Chivasa S, Ndimba BK, Simon WJ, Robertson D, Yu XL, Knox JP, Bolwell P, Slabas AR. 2002. Proteomic analysis of the Arabidopsis thaliana cell wall. Electrophoresis;23(11):1754-65.

Cho Y., Chrispeels M. 1976. Serine-O-galactosyl linkages in glycopeptides from carrot cell walls.Phytochemistry;15: 165-169.

Chow CM, Neto H, Foucart C, Moore I. 2008. Rab-A2 and Rab-A3 GTPases define a trans-golgi endosomal membrane domain in Arabidopsis that contributes substantially to the cell plate. Plant Cell.;20(1):101-23.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J.;16(6):735-43.

Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N, Wilkerson CG. 2007. A gene from the cellulose synthase-like C family encodes a beta-1,4 glucan synthase. Proc Natl Acad Sci U S A.;104(20):8550-5.

Cosgrove DJ. 2000. Loosening of plant cell walls by expansins. Nature;407(6802):321-6.

Cunnane SC and Thompson LU. 1995. In: Flaxseed in Human Nutrition. Eds. Champaign, IL: AOCS Press.:219-236.

Czechowski T1, Stitt M, Altmann T, Udvardi MK, Scheible WR. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol.;139(1):5-17.

Czernic P, Visser B, Sun W, Savouré A, Deslandes L, Marco Y, Van Montagu M, Verbruggen N. 1999. Characterization of an Arabidopsis thaliana receptor-like protein kinase gene activated by oxidative stress and pathogen attack. Plant J.;18(3):321-7.

Datla RS, Hammerlindl JK, Panchuk B, Pelcher LE, Keller W. 1992. Modified binary plant transformation vectors with the wild-type gene encoding NPTII. Gene.;122(2):383-4.

Davies, T. 1998. The new automated mass spectrometry deconvolution and identification system (AMDIS). Spectroscopy;10(3):24–27.

De Pauw MA., Vidmar JJ., Collins J., Bennett RA., Deyholos MK. 2007. Microarray analysis of bast fibre producing tissue of Cannabis sativa identifies transcripts associated with conserved and specialised processes of secondary wall development. Func. Plant Biol.;34:737–749.

DeClercq DR; Daun JK. 2002 Quality of Western Canadian Flaxseed; Canadian Grain Commission: Winnipeg, Manitoba. http://www. grainscanada.gc.ca/flax-lin/harvest-recolte/2011/hqf11-qrl11-eng.htm.

Dello Ioio R, Nakamura K, Moubayidin L, Perilli S, Taniguchi M, Morita MT, Aoyama T, Costantino P, Sabatini S. 2008. A genetic framework for the control of cell division and differentiation in the root meristem. Science.;322(5906):1380-4.

Demain AL, Vaishnav P. 2009. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv.; 27(3):297-306.

Demark-Wahnefried W, Price DT, Polascik TJ, Robertson CN, Anderson EE, Paulson DF, Walther PJ, Gannon M, Vollmer RT. 2001. Pilot study of dietary fat restriction and flaxseed supplementation in men with prostate cancer before surgery: exploring the effects on hormonal levels, prostate-specific antigen, and histopathologic features. Urology.;58(1):47-52.

Dettmer J, Hong-Hermesdorf A, Stierhof YD, Schumacher K. 2006. Vacuolar H+-ATPase activity is require

Dharmasiri S, Jayaweera T, Dharmasiri N. 2013. Plant Hormone Signalling: Current Perspectives on Perception and Mechanisms of Action. Ceylon J of Sci. (Bio. Sci.);42 (1): 1-17.

Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S, Anderson P. 2004. Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. Science.;303(5656):363-6.

Dlakić M. 2006. DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. Bioinformatics.;22(22):2711-4.

Doong RL, Mohnen D. 1998. Solubilization and characterization of a galacturonosyltransferase that synthesizes the pectic polysaccharide homogalacturonan. Plant J;13: 363–374.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh green tissue. Phytochemistry Bulletin 19.

Driouich A1, Follet-Gueye ML, Bernard S, Kousar S, Chevalier L, Vicré-Gibouin M, Lerouxel O. 2012. Golgi-mediated synthesis and secretion of matrix polysaccharides of the primary cell wall of higher plants. Front Plant Sci.;3:79.

Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P, Lilley KS. 2006. Mapping the Arabidopsis organelle proteome. Proc Natl Acad Sci U S A.;103(17):6518-23.

Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P, Lilley KS. 2008. Use and misuse of the gene ontology annotations. Nat Rev Genet.;9(7):509-15.

Egelund J, Damager I, Faber K, Olsen CE, Ulvskov P, Petersen BL. 2008. Functional characterisation of a putative rhamnogalacturonan II specific xylosyltransferase. FEBS Lett.;582(21-22):3217-22.

Egelund J, Petersen BL, Motawia MS, Damager I, Faik A, Olsen CE, Ishii T, Clausen H, Ulvskov P, Geshi N. 2006. Arabidopsis thaliana RGXT1 and RGXT2 encode Golgi-localized (1,3)-alpha-D-xylosyltransferases involved in the synthesis of pectic rhamnogalacturonan-II. Plant Cell.;18(10):2593-607.

El-Beltagi HS, Salama ZA., El-Hariri DM. 2007. Evaluation of fatty acids profile and the content of some secondary metabolites in seeds of different flax cultivars (Linum usitatissimum L.). Gen. Appl. Plant Physiol.;33(3-4):187-202.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol.;300(4):1005-16.

Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G. 2006. Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiol.;142(3):839-54.

Fait A, Nesi AN, Angelovici R, Lehmann M, Pham PA, Song L, Haslam RP, Napier JA, Galili G, Fernie AR. 2011. Targeted enhancement of glutamate-to-γ-aminobutyrate conversion in Arabidopsis seeds affects carbon-nitrogen balance and storage reserves in a development-dependent manner. Plant Physiol.;157(3):1026-42.

Farrokhi N., Firouzabadi MB and Gholami A. 2013. Functional Genomics of Plant Cell Wall Biosynthetic Enzymes: Type-II Membrane-Bound Glycosyltransferases. Int J Bot.;9(1):1-17.

Fernández FJ1, Vega MC. 2013. Technologies to keep an eye on: alternative hosts for protein production in structural biology. Curr Opin Struct Biol.;23(3):365-73.

Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L. 2000a. Metabolite profiling for plant functional genomics. Nat Biotechnol.;18(11):1157-61.

Fiehn O, Kopka J, Trethewey RN, Willmitzer L. 2000b. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Anal Chem.;72(15):3573-80.

Fiehn O. 2002. Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol.;48(1-2):155-71.

Fiehn O. 2006. Metabolite profiling in Arabidopsis. Methods in Molecular Biology.;323:439–447.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database, Nucleic Acids Res.;42(Database issue):D222-30.

Fry SC, Miller JG. 1989. Toward a working model of the growing plant cell wall: phenolic cross-linking reaction in the primary cell walls of dicotyledons. In NG Lewis, MG Paice, eds, Plant Cell Wall Polymers: Biogenesis and Biodegradation. ACS Symposium Series, 399. American Chemical Society, Washington, DC, 33-46.

Fukaki H, Tasaka M. 2009. Hormone interactions during lateral root formation. Plant Mol Biol.;69(4):437-49.

Fukushima A, Kusano M, Mejia RF, Iwasa M, Kobayashi M, Hayashi N, Watanabe-Takahashi A, Narisawa T, Tohge T, Hur M, Wurtele ES, Nikolau BJ, Saito K. 2014. Metabolomic Characterization of Knockout Mutants in Arabidopsis: Development of a Metabolite Profiling Database for Knockout Mutants in Arabidopsis. Plant Physiol.;165(3):948-961.

Futcher BG, Latter I, Monardo P, McLaughlin CS, Garrels JI. 1999. A sampling of the yeast proteome. Mol. Cell. Biol. 19: 7357–7368.

Gallardo K, Firnhaber C, Zuber H, Héricher D, Belghazi M, Henry C, Küster H, Thompson R. 2007. A combined proteome and transcriptome analysis of developing Medicago truncatula seeds: evidence for metabolic specialization of maternal and filial tissues. Mol Cell Proteomics.;6(12):2165-79.

Gallardo K, Job C, Groot SP, Puype M, Demol H, Vandekerckhove J, Job D. 2001. Proteomic analysis of arabidopsis seed germination and priming. Plant Physiol.;126(2):835-48.

Gao Y, Badejo AA, Sawa Y, Ishikawa T. 2012. Analysis of two L-Galactono-1,4-lactone-responsive genes with complementary exqin

Gao Y, Nishikawa H, Badejo AA, Shibata H, Sawa Y, Nakagawa T, Maruta T, Shigeoka S, Smirnoff N, Ishikawa T. 2011. Expression of aspartyl protease and C3HC4-type RING zinc finger genes are responsive to ascorbic acid in Arabidopsis thaliana. J Exp Bot.;62(10):3647-57.

Gatto L, Vizcaíno JA, Hermjakob H, Huber W, Lilley KS. 2010. Organelle proteomics experimental designs and analysis. Proteomics.;10(22):3957-69.

Gazzarrini S1, McCourt P. 2003. Cross-talk in plant hormone signalling: what arabidopsis mutants are telling us. Ann Bot.;91(6):605-12.

Gendre D1, McFarlane HE, Johnson E, Mouille G, Sjödin A, Oh J, Levesque-Tremblay G, Watanabe Y, Samuels L, Bhalerao RP. 2013. Golgi-localized enzyme complexes for plant cell wall biosynthesis. Trends Plant Sci.;18(1):49-58.

Gendreau E, Traas J, Desnos T, Grandjean O, Caboche M, Höfte H. 1997. Cellular basis of hypocotyl growth in Arabidopsis thaliana. Plant Physiol.;114(1):295-305.

Georg H.H. Borner, Kathryn S. Lilley, Timothy J. Stevens, and Paul Dupree, 2003. Identification of Glycosylphosphatidylinositol-Anchored Proteins in Arabidopsis. A Proteomic and Genomic Analysis. Plant Physiol.; 132(2): 568–577.

Geshi N1, Jørgensen B, Ulvskov P. 2004. Subcellular localization and topology of beta(1-->4)galactosyltransferase that elongates beta(1-->4)galactan side chains in rhamnogalacturonan I in potato. Planta.;218(5):862-8.

Gilbert HJ. 2010. The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiol.;153(2):444-55.

Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, Nagaya H, Roy L, Gosline SJ, Hallett M, Paiement J, Kearney RE, Nilsson T, Bergeron JJ. 2006. Quantitative proteomics analysis of the secretory pathway. Cell.;127(6):1265-81.

Glassbrook N, Beecher C, Ryals J. 2000. Nat Biotechnol.;18(11):1142-3.

Goeman JJ, Bühlmann P. 2007. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics.;23(8):980-7.

Goldberg RB, Beals TP, Sanders PM. 1993. Anther development: basic principles and practical applications. Plant Cell.;5(10):1217-29.

Gomez-Casati DF, Zanor MI, Busi MV. 2013. Metabolomics in plants and humans: applications in the prevention and diagnosis of diseases. Biomed Res Int.;2013:792527.

Goodacre NF, Gerloff DL, Uetz P. 2013. Protein domains of unknown function are essential in bacteria. MBio.;5(1).

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol.;22(5):245-52.

Goonesekere NC1, Shipely K, O'Connor K. 2010. The challenge of annotating protein sequences: The tale of eight domains of unknown function in Pfam. Comput Biol Chem.;34(3):210-4.

Gorshkova TA, Sal'nikova VV, Chemikosova SB, Ageeva MV, Pavlencheva NV, van Dam JEG. 2003. The snap point: a transition point in Linum usitatissimum bast fiber development. Industrial Crops and Products. 18(3), 213–222.

Goubet F, Barton CJ, Mortimer JC, Yu X, Zhang Z, Miles GP, Richens J, Liepman AH, Seffen K, Dupree P., 2009, Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. Plant J.;60(3):527-38.

Greenbaum D, Colangelo C, Williams K, Gerstein M. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol.;4(9):117.

Griesbeck O, Baird GS, Campbell RE, Zacharias DA, Tsien RY. 2001. Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. J Biol Chem.;276(31):29188-94.

Groen AJ, Sancho-Andrés G, Breckels LM, Gatto L, Aniento F, Lilley KS. 2014. Identification of Trans-Golgi Network Proteins in Arabidopsis thaliana Root Tissue. J Proteome Res.; 13(2): 763–776.

Gunasekaran K, Gomathi L, Ramakrishnan C, Chandrasekhar J, Balaram P. 1998. Conformational interconversions in peptide beta-turns: analysis of turns in proteins and computational estimates of barriers. J Mol Biol.;284(5):1505-16.

Gutierrez L, Conejero G, Castelain M, Guénin S, Verdeil JL, Thomasset B, Van Wuytswinkel O. 2006. Identification of new gene expression regulators specifically expressed during plant seed maturation. J Exp Bot.;57(9):1919-32.

Gygi SP, Rochon Y, Franza BR, Aebersold R. 1999. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol.;19(3):1720-30.

Halkier BA, Gershenzon J. 2006. Biology and biochemistry of glucosinolates. Annu Rev Plant Biol.;57:303-33.

Han W, Rhee JS, Maximov A, Lin W, Hammer RE, Rosenmund C, Südhof TC. 2005. C-terminal ECFP fusion impairs synaptotagmin 1 function: crowding out synaptotagmin 1. J Biol Chem.;280(6):5089-100.

Hanson MR, Köhler RH. 2001. GFP imaging: methodology and application to investigate cellular compartmentation in plants. J Exp Bot.;52(356):529-39.

Harholt J, Suttangkakul A, Scheller HV. 2010. Biosynthesis of Pectin. Plant Physiol.;153, 384–395.

Helliwell C, Waterhouse P. 2003. Constructs and methods for high-throughput gene silencing in plants. Methods.;30(4):289-95.

Hsu FC, Bennett AB, Spanswick RM. 1984. Concentrations of sucrose and nitrogenous compounds in the apoplast of developing soybean seed coats and embryos. Plant Physiol.;75(1):181-6.

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science.;292(5518):929-34.

Irshad M, Canut H, Borderies G, Pont-Lezica R, Jamet E. 2008. A new picture of cell wall protein dynamics in elongating cells of Arabidopsis thaliana: confirmed actors and newcomers. BMC Plant Biol.;8:94.

Ivanchenko MG, Muday GK, Dubrovsky JG. 2008. Ethylene-auxin interactions regulate lateral root initiation and emergence in Arabidopsis thaliana. Plant J.;55(2):335-47.

Jackson S, Rounsley S, Purugganan M. 2006. Comparative sequencing of plant genomes: choices to make. Plant Cell;18(5):1100-4.

Jamet E, Canut H, Boudart G, Pont-Lezica RF., 2006, Cell wall proteins: a new insight through proteomics. Trends Plant Sci. ;11(1):33-9.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. 1998. Multiple sequence alignment with Clustal X. Trends Biochem Sci.;23(10):403-5.

Jensen JK, Kim H, Cocuron JC, Orler R, Ralph J, Wilkerson CG. 2011. The DUF579 domain containing proteins IRX15 and IRX15-L affect xylan synthesis in Arabidopsis. Plant J.;66(3):387-400.

Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol.;292(2):195-202.

Kang BH, Nielsen E, Preuss ML, Mastronarde D, Staehelin LA. 2011. Electron tomography of RabA4b- and PI-4Kβ1-labeled trans Golgi network compartments in Arabidopsis. Traffic;12(3):313-29.

Kaplan HA, Welply JK, Lennarz WJ.1987. Oligosaccharyl transferase: the central enzyme in the pathway of glycoprotein assembly. Biochim Biophys Acta.;906(2):161-73.

Kazan K, Manners JM. 2008. Jasmonate signaling: toward an integrated view. Plant Physiol.;146(4):1459-68.

Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc.;4(3):363-71.

Kende H, Bradford K, Brummell D, Cho HT, Cosgrove D, Fleming A, Gehring C, Lee Y, McQueen-Mason S, Rose J, Voesenek LA. 2004. Nomenclature for members of the expansin superfamily of genes and proteins. Plant Mol Biol.;55(3):311-4.

Kessel A, Ben-Tal N. 2011. Introduction to Proteins: Structure, Function and Motion. ChemBioChem.;12(10): 1603–1604.

Kim HJ, Murai N, Fang DD, Triplett BA. 2011. Functional analysis of Gossypium hirsutum cellulose synthase catalytic subunit 4 promoter in transgenic Arabidopsis and cotton tissues. Plant Sci.;180(2):323-32.

Kneen M, Farinas J, Li Y, Verkman AS. 1998. Green fluorescent protein as a noninvasive intracellular pH indicator. Biophys J.;74(3):1591-9.

Korkuc P, Schippers JH, Walther D. 2014. Characterization and identification of cis-regulatory elements in Arabidopsis based on single-nucleotide polymorphism information. Plant Physiol.;164(1):181-200.

Kudo N, Mii M. 2004. Endoreduplication cycles during hypocotyl growth of cabbage (Brassica oleracea L.) under light and dark conditions. Plant Biotechnol.;21: 295–298.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics;23(21):2947-8.

Lee C, O'Neill MA, Tsumuraya Y, Darvill AG, Ye ZH. 2007. The irregular xylem9 mutant is deficient in xylan xylosyltransferase activity. Plant Cell Physiol.;48(11):1624-34.

Lee GJ, Sohn EJ, Lee MH, Hwang I. 2004. The Arabidopsis rab5 homologs rha1 and ara7 localize to the prevacuolar compartment. Plant Cell Physiol.;45(9):1211-20.

Lee J, Wu S, Zhang Y. 2009. Ab initio protein structure prediction. In From Protein Structure to function with Bioinformatics (London: Springer).:1-26.

Levi SK, Bhattacharyya D, Strack RL, Austin JR, Glick BS. 2010. The Yeast GRASP Grh1 Colocalizes with COPII and Is Dispensable for Organizing the Secretory Pathway. Published online 2010 Jun 21.

Liang YH, Cai B, Chen F, Wang G, Wang M, Zhong Y, Cheng ZMM. 2014. Construction and validation of a gene co-expression network in grapevine (Vitis vinifera. L.). Horticulture Research 1.

Liepman AH, Nairn CJ, Willats WG, Sørensen I, Roberts AW, Keegstra K. 2007. Functional genomic analysis supports conservation of function among cellulose synthase-like a gene family members and suggests diverse roles of mannans in plants. Plant Physiol.;143(4):1881-93.

Liepman AH, Wilkerson CG, Keegstra K. 2005. Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. Proc Natl Acad Sci U S A.;102(6):2221-6.

Linus Pauling, Arthur B. Robinson, Roy Teranishi,† and Paul Cary. 1971. Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography. Proc Natl Acad Sci U S A.;68(10): 2374–2376.

Livak KJ1, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods.;25(4):402-8.

Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins.;50(3):437-50.

Lü S, Wang J. 2014. Homoharringtonine and omacetaxine for myeloid hematological malignancies. J Hematol Oncol.; 7:2.

Luckner M. 1990. Secondary Metabolism in Microorganisms, Plants, and Animals, 3rd edn. Springer, Berlin.

Luo F, Yang Y, Zhong J, Gao H, Khan L, Thompson DK, Zhou J. 2007. Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. BMC Bioinformatics.;8:299.

Malamy JE1, Benfey PN. 1997. Organization and cell differentiation in lateral roots of Arabidopsis thaliana Development. Development.;124(1):33-44.

Marshall RD. 1974. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. Biochem Soc Symp.;(40):17-26.

Martienssen RA. 1998. Functional genomics: probing plant gene function and expression with transposons. Proc Natl Acad Sci U S A.;95(5):2021-6.

Matsuoka K, Watanabe N, Nakamura K. 1995. O-glycosylation of a precursor to a sweet potato vacuolar protein, sporamin, expressed in tobacco cells. Plant J.;8(6):877-89.

McCann, MC, Roberts K, Carpita NC. 2001. Plant Cell Growth and Elongation. s. John Wiley & Sons, New York. 1-8.

Medeghini-Bonatti P, Laudi G, and Fricano G. 1976. Studies on the Physiology of Abscission: Effect of Treatments with Gibberellic Acid. Amer. J. Bot.;63(2): 135-137

Minic Z, Jamet E, Négroni L, Arsene der Garabedian P, Zivy M, Jouanin L. 2007, A sub-proteome of Arabidopsis thaliana mature stems trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. J Exp Bot.;58(10):2503-12.

Mohnen D. 2008. Pectin structure and biosynthesis. Curr Opin Plant Biol.;11(3):266-77.

Molecular Cell Biology (4th edition). Lodish H, Berk A, Lawrence Zipursky S, Matsudaira P, Baltimore D, Darnell J. 2000. Freeman & Co., New York, NY, 2000, 1084 pp., ISBN 0-7167-3136-3.

Morris DH. 2001. Essential nutrients and other functional compounds in flaxseed. Nutr Today.;36:159–162.

Motley AM, Berg N, Taylor MJ, Sahlender DA, Hirst J, Owen DJ, Robinson MS. 2006. Functional analysis of AP-2 alpha and mu2 subunits. Mol Biol Cell. 2006 Dec;17(12):5298-308.

Muday GK, Rahman A, Binder BM. 2012. Auxin and ethylene: collaborators or competitors? Trends Plant Sci.;17(4):181-195.

Mulder NJ, Kersey P, Pruess M, Apweiler R.. 2008. In silico characterization of proteins: UniProt, InterPro and Integr8. Mol Biotechnol.;38(2):165-77.

Murashige T, Skoog F. 1962. A revised medium for rapid growth and bio-assays with tobacco tissue cultures. Physiol Plant.;15(3):473–497.

Na Jom K, Chanput W, Ngampongsai S. 2014. Effect of genetic and climatic variability on the metabolic profiles of black gram (Vigna mungo L.) seeds and sprouts. J Sci Food Agric. doi: 10.1002/jsfa.6869. [Epub ahead of print].

Negri AS, Prinsi B, Scienza A, Morgutti S, Cocucci M, Espen L. 2008. Analysis of grape berry cell wall proteome: a comparative evaluation of extraction methods. J Plant Physiol.;165(13):1379-89.

Nemhauser JL, Mockler TC, Chory J. 2004. Interdependency of brassinosteroid and auxin signaling in Arabidopsis. PLoS Biol.;2(9):E258.

Nicholson JK, Lindon JC, Holmes E. 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica.;29(11):1181-9.

Nie L, Wu G, Zhang W. 2006. Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in Desulfovibrio vulgaris: a quantitative analysis. Genetics. 2006 Dec;174(4):2229-43.

Nishiyama Y, Langan P, Chanzy H. 2002. Crystal structure and hydrogen-bonding system in cellulose Ibeta from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc.;124(31):9074-82.

Oikawa A, Lund CH, Sakuragi Y, Scheller HV. 2013. Golgi-localized enzyme complexes for plant cell wall biosynthesis. Trends Plant Sci. 2013 Jan;18(1):49-58.

Oliver S. 2000. Guilt-by-association goes global. Nature;403(6770):601-3.

O'Malley RC, Ecker JR. 2010. Linking genotype to phenotype using the Arabidopsis unimutant collection. Plant J.;61(6):928-40.

ONeill M, Albersheim P, Darvill A. 1990. The pectic polysaccharides of primary cell walls. In PM Dey, JB Harborne, eds, Methods in Plant Biochem.; 2. Academic Press, London, 415-441.

Østergaard L, Yanofsky MF. 2004. Establishing gene function by mutagenesis in Arabidopsis thaliana. Plant J.;39(5):682-96.

Otegui MS, Herder R, Schulze J, Jung R, Staehelin LA. 2006. The proteolytic processing of seed storage proteins in Arabidopsis embryo cells starts in the multivesicular bodies. Plant Cell;18(10):2567-81.

Overvoorde P, Fukaki H, Beeckman T. 2010. Auxin Control of Root Development. Cold Spring Harb Perspect Biol.; 2(6): a001537.

Pacifico S, D'Abrosca B, Scognamiglio M, Gallicchio M, Potenza N, Piccolella S, Russo A, Monaco P, Fiorentino A. 2011. Metabolic profiling of strawberry grape (Vitis × labruscana cv. 'Isabella') components by nuclear magnetic resonance (NMR) and evaluation of their antioxidant and antiproliferative properties. J Agric Food Chem.;59(14):7679-87.

Park M, Jürgens G. 2012. Membrane traffic and fusion at post-Golgi compartments. Front Plant Sci.;2:111.

Parry G, Estelle M. 2006. Auxin receptors: a new role for F-box proteins. Curr Opin Cell Biol.;18(2):152-6.

Parsons HT, Christiansen K, Knierim B, Carroll A, Ito J, Batth TS, Smith-Moritz AM, Morrison S, McInerney P, Hadi MZ, Auer M, Mukhopadhyay A, Petzold CJ, Scheller HV, Loqué D, Heazlewood JL. 2012. Isolation and proteomic characterization of the Arabidopsis Golgi defines functional and novel components involved in plant cell wall biosynthesis. Plant Physiol.;159(1):12-26.

Parsons HT, Drakakaki G, Heazlewood JL. 2012. Proteomic dissection of the Arabidopsis Golgi and trans-Golgi network. Front Plant Sci.;3:298.

Pauly M, Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. Plant J.;54(4):559-68.

Payyavula RS, Tschaplinski TJ, Jawdy SS, Sykes RW, Tuskan GA, Kalluri UC. 2014. Metabolic profiling reveals altered sugar and secondary metabolism in response to UGPase overexpression in Populus. BMC Plant Biol.;14:265.

Perrot-Rechenmann C. 2010. Cellular responses to auxin: division versus expansion. Cold Spring Harb Perspect Biol.;2(5):a001446.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat Methods. 810:785-786.

Petkowicza CL de O, Reichera F, Chanzyb H, Taravelb FR, Vuongb R. 2001. Linear mannan in the endosperm of Schizolobium amazonicum. Carbohydr Polym.;44 (2):107–112.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res.;29(9):e45.

Pichersky E, Lewinsohn E. 2011. Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol.;62:549-66.

Prasad K. 2001. Secoisolariciresinol diglucoside from flaxseed delays the development of type 2 diabetes in Zucker rat. J Lab Clin Med.;138(1):32-9.

pression during the development of Arabidopsis thaliana. Plant Cell Physiol.;53(3):592-601.

Prosser GA, Larrouy-Maumus G, de Carvalho LP. 2014. Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. EMBO Rep.;15(6):657-69.

Qin H, Chen F, Huan X, Machida S, Song J, Yuan YA. 2010. Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. RNA.;16(3):474-81.

Qiu P. 2003. Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. Biochem Biophys Res Commun.;309(3):495-501.

Qiu Y and Reed D. 2014. Chapter: Gas Chromatography in Metabolomics Study. Advances in Gas Chromatography, Edited by Dr Xinghua Guo, 02/2014; InTech., ISBN: 978-953-51-1227-3

Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J. 2001. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res.; 29(1):159-64.

Rainbird RM, Thorne JH, Hardy RW. 1984. Role of amides, amino acids, and ureides in the nutrition of developing soybean seeds. Plant Physiol.;74(2):329-34.

Raz V, Koornneef M. 2001. Cell division activity during apical hook development. Plant Physiol.;125(1):219-26.

Reiter WD, Vanzin GF. 2001. Molecular genetics of nucleotide sugar interconversion pathways in plants. Plant Mol Biol.;47(1-2):95-113.

Rennie EA and Scheller HB. 2014. Xylan biosynthesis. Curr Opin Biotechnol 26: 100-107.

Rhee SY, Wood V, Dolinski K, Draghici S. 2008. Use and misuse of the gene ontology annotations. Nat Rev Genet.;9(7):509-15.

Richter S, Voss U, Jürgens G. 2009. Post-Golgi traffic in plants. Traffic.;10(7):819-28.

Ridley BL1, O'Neill MA, Mohnen D. 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry.;57(6):929-67.

Robert C. Day, Rowan P. Herridge, Barbara A. Ambrose, and Richard C. Macknight. 2008. Transcriptome Analysis of Proliferating Arabidopsis Endosperm Reveals Biological Implications for the Control of Syncytial Division, Cytokinin Signaling, and Gene Expression Regulation. Plant Physiol.; 148(4): 1964–1984.

Roberts JA, Elliott KA, Gonzalez-Carranza ZH. 2002. Abscission, dehiscence, and other cell separation processes. Annu Rev Plant Biol.;53:131-58.

Roberts RJ. 2004. Identifying protein function--a call for community action. PLoS Biol.;2(3):E42.

Robichon C, Luo J, Causey TB, Benner JS, Samuelson JC. 2011. Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. Appl Environ Microbiol.;77(13):4634-46.

Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie A. 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell.;13(1):11-29.

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L. 2000. Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J.;23(1):131-42.

Römling U, Simm R. 2009. Prevailing concepts of c-di-GMP signaling. Contrib Microbiol.;16:161-81.

Rose JK1, Lee SJ. 2010. Straying off the highway: trafficking of secreted plant proteins and complexity in the plant cell wall proteome. Plant Physiol.;153(2):433-6.

Roy A, Kucukural A, Yang Zhang Y.2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc.; 5(4): 725–738.

Roy A, Yang J, and Zhang Y. 2012. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res.; 40(Web Server issue): W471–W477.

Růzicka K, Ljung K, Vanneste S, Podhorská R, Beeckman T, Friml J, Benková E. 2007. Ethylene regulates root growth through effects on auxin biosynthesis and transport-dependent auxin distribution. Plant Cell.;19(7):2197-212.

Sahdev S, Khattar SK, Saini KS. 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem.;307(1-2):249-64.

Saibo NJ, Vriezen WH, Beemster GT, Van Der Straeten D. 2003. Growth and stomata development of Arabidopsis hypocotyls are controlled by gibberellins and modulated by ethylene and auxins. Plant J. ;33(6):989-1000.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: A laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY.

Sampedro J, Cosgrove DJ. 2005. The expansin superfamily. Genome Biol.;6(12):242.

Samuni-Blank M, Izhaki I2, Gerchman Y3, Dearing MD4, Karasov WH5, Trabelcy B2, Edwards TM6, Arad Z7. 2014. Taste and physiological responses to glucosinolates: seed predator versus seed disperser. PLoS One.;9(11):e112505.

Sánchez E. Z., and Buen A. G. 2012. The Two DUF642 At5g11420 and At4g32460- Encoded Proteins Interact In Vitro with the AtPME3 Catalytic Domain, Protein Interactions, Dr. Jianfeng Cai (Ed.), ISBN: 978-953-51-0244-1, InTech, Available from: http://www.intechopen.com/books/proteininteractions/ the-two-duf642-at5g11420-and-at4g32460-encoded-proteins-interact-in-vitro-with-the-atpme3-catalytic.

Santner A, Calderon-Villalobos LI, Estelle M. 2009. Plant hormones are versatile chemical regulators of plant growth. Nat Chem Biol.;5(5):301-7.

Santoni V, Rouquié D, Doumas P, Mansion M, Boutry M, Degand H, Dupree P, Packman L, Sherrier J, Prime T, Bauw G, Posada E, Rouzé P, Dehais P, Sahnoun I, Barlier I, Rossignol M. 1998. Use of a proteome strategy for tagging proteins present at the plasma membrane. Plant J.;16(5):633-41.

Sauer M, Paciorek T, Benková E, Friml J. 2006. Immunocytochemical techniques for whole-mount in situ protein localization in plants. Nat Protoc.;1(1):98-103.

Sauter H, Lauer M, Fritsch H., 1991. Metabolic profiling of plants a new diagnostic technique. In: Baker DR, Fenyes JG, Moberg WK (Eds.), American Chemical Society Symposium Series No. 443. American Chemical Society, Washington DC, pp. 288–299.

Scheller HV, Ulvskov P., 2010, Hemicelluloses. Annu Rev Plant Biol.;61:263-89.

Seigler DS. 1997. Plant secondary metabolism. Boston: Kluwer Academic Publishers. 303.

Shi J, Zhen Y, Zheng RH. 2010. Proteome profiling of early seed development in Cunninghamia lanceolata (Lamb.) Hook. J Exp Bot.;61(9):2367-81.

Shima YY, Gui B, Arnison PG, Wang Y, Reaney, MJT, 2014. Flaxseed (Linum usitatissimum L.) bioactive compounds and peptide nomenclature: A review. Trends in Food Science & Technology;38(1):5–20.

Sigrist CJ, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N. 2010. PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res., 38(Database issue):D161-6.

Singh AK, Manjasetty B, Gl B, Koul S, Kaushik A, Ekka MK, Singh V, Kumaran S. 2015. Crystal Structure of Fad35R from Mycobacterium tuberculosis H37Rv in the Apo-State. PLoS One.;10(5):e0124333.

Smith TA. 1980. Plant amines. In: EA Bell, BV Charlwood, eds, Secondary Plant Products. Encycl of Plant Physiol.(New Series), Springer-Verlag, Berlin;8:433-460.

Sonnhammer EL1, Eddy SR, Birney E, Bateman A, Durbin R. 1998. Pfam: Multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res.;26(1):320-2.

Soto A, Ruiz KB, Ziosi V, Costa G, Torrigiani P. 2012. Ethylene and auxin biosynthesis and signaling are impaired by methyl jasmonate leading to a transient slowing down of ripening in peach fruit. J Plant Physiol.;169(18):1858-65.

Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H. 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J.;32(10):1478-88.

Stein, S. E. 1999. An integrated method for spectrum extraction and compound identification from gc/ms data. J. Am. Soc. Mass Spectrom.;10(8):770–871.

Stepanova AN, and Alonso JM. 2005. Ethylene signaling and response pathway: A unique signaling cascade with a multitude of inputs and outputs. Physiol. Plant.;123(2), 195–206.

Sterling JD, Atmodjo MA, Inwood SE, Kumar Kolli VS, Quigley HF, Hahn MG, Mohnen D., 2006. Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase. Proc Natl Acad Sci U S A.;103(13):5236-41.

Sterling JD1, Quigley HF, Orellana A, Mohnen D. 2001. The catalytic site of the pectin biosynthetic enzyme alpha-1,4-galacturonosyltransferase is located in the lumen of the Golgi. Plant Physiol.;127(1):360-71.

Stotz , H , Brown , P & Tokuhisa , J 2015 , ' Glucosinolate biosynthesis from amino acids ' . in J P F D'Mello (ed.) , Amino Acids in Higher Plants . CAB International, Wallingford, Oxfordshire , pp. 436-447.

Sundaresan V, Springer P, Volpe T, Haward S, Jones JD, Dean C, Ma H, Martienssen R. 1995. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes Dev.;9(14):1797-810.

Suzuki S1, Li L, Sun YH, Chiang VL. 2006. The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in Populus trichocarpa. Plant Physiol.; Plant Physiol.;142(3):1233-45.

Sveinsson S, McDill J, Wong GK, Li J, Li X, Deyholos MK, Cronk QC. 2014. Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (Linum) using transcriptomics. Ann Bot.;113(5):753-61.

Szymanowska-Pułka J, Potocka I, Karczewski J, Jiang K, Nakielski J, Feldman LJ. 2012. Principal growth directions in development of the lateral root in Arabidopsis thaliana. Ann Bot.;110(2):491-501.

Taliercio E. 2008. Isolation and Characterization of an ADP-Glucose Pyrophosphorylase Gene from Gossypium hirsutum L. J Cotton Sci.;12:273–279.

Tambellini NP, Zaremberg V, Turner RJ, Weljie AM. 2013. Evaluation of extraction protocols for simultaneous polar and non-polar yeast metabolite analysis using multivariate projection methods. Metabolites.;3(3):592-605.

Tamura K, Dudley J, Nei M and Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol.;24:1596-1599.

Tanaka H, Nodzyłski T, Kitakura S, Feraru MI, Sasabe M, Ishikawa T, Kleine-Vehn J, Kakimoto T, Friml J. 2014. BEX1/ARF1A1C is required for BFA-sensitive recycling of PIN auxin transporters and auxin-mediated development in Arabidopsis. Plant Cell Physiol.;55(4):737-49.

Tanimoto, E. 2005. Regulation of root growth by plant hormones: Roles for auxin and Gibberellin. Crit. Rev. Pl. Sci.; 24(4): 249-265.

Tanya Z. Berardini, Suparna Mundodi, Leonore Reiser, Eva Huala, Margarita Garcia-Hernandez, Peifen Zhang, Lukas A. Mueller,2 Jungwoon Yoon, Aisling Doyle, Gabriel Lander, Nick Moseyko, Danny Yoo, Iris Xu, Brandon Zoeckler, Mary Montoya, Neil Miller, Dan Weems, and Seung Y. Rhee. 2004. Functional Annotation of the Arabidopsis Genome Using Controlled Vocabularies. Plant Physiol.; 135(2): 745–755.

Thompson LU, Rickard SE, Cheung F, Kenaschuk EO, Obermeyer WR. 1997. Variability in anticancer lignan levels in flaxseed. Nutr Cancer.;27(1):26-30.

Tong Z, Gao Z, Wang F, Zhou J, Zhang Z. 2009. Selection of reliable reference genes for gene expression studies in peach using real-time PCR. BMC Mol Biol.;10:71.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science. 2006 Sep 15;313(5793):1596-604.

Vaisey-Genser M, Morris DH. 1997. Flaxseed: Health, Nutrition and Functionality. Winnipeg, MB: Flax Council of Canada.

Vázquez-Lobo A, Roujol D, Zuñiga-Sánchez E, Albenne C, Piñero D, Gamboa de Buen A, Jamet E. 2012. The highly conserved spermatophyte cell wall DUF642 protein family: phylogeny and first evidence of interaction with cell wall polysaccharides in vitro. Mol Phylogenet Evol.;63(2):510-20.

Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S 2011. Gene expression analysis of flax seed development. BMC Plant Biol.;11:74.

Viotti C, Bubeck J, Stierhof YD, Krebs M, Langhans M, van den Berg W, van Dongen W, Richter S, Geldner N, Takano J, Jürgens G, de Vries SC, Robinson DG, Schumacher K.2010. Endocytic and secretory traffic in Arabidopsis merge in the trans-Golgi network/early endosome, an independent and highly dynamic organelle. Plant Cell;22(4):1344-57.

Vuttipongchaikij S1, Brocklehurst D, Steele-King C, Ashford DA, Gomez LD, McQueen-Mason SJ. 2012. Arabidopsis GT34 family contains five xyloglucan α-1,6-xylosyltransferases. New Phytol.;195(3):585-95.

Walters KR Jr, Serianni AS, Sformo T, Barnes BM, Duman JG. 2009. A nonprotein thermal hysteresis-producing xylomannan antifreeze in the freeze-tolerant Alaskan beetle Upis ceramboides. Proc Natl Acad Sci U S A.;106(48):20210-5.

Wang L, Chen J, Thompson LU. 2005. The inhibitory effect of flaxseed on the growth and metastasis of estrogen receptor negative human breast cancer xenograftsis attributed to both its lignan and oil components. Int J Cancer.;116(5):793-8.

Wang L, Shen R, Chen LT, Liu YG. 2014. Characterization of a novel DUF1618 gene family in rice. J Integr Plant Biol.;56(2):151-8.

Wang Y, Alonso AP, Wilkerson CG, Keegstra K., 2012, Deep EST profiling of developing fenugreek endosperm to investigate galactomannan biosynthesis and its regulation. Plant Mol Biol. 2012 Jun;79(3):243-58.

Washburn MP, Koller A, Oshiro G, Ulaszek RR, Plouffe D, Deciu C, Winzeler E, Yates JR 3rd. 2003. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A.;100(6):3107-12.

Washida H, Wu CY, Suzuki A, Yamanouchi U, Akihama T, Harada K, Takaiwa F. 1999. Identification of cis-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. Plant Mol Biol.;40(1):1-12.

Weber H, Borisjuk L, Wobus U. 2005. Molecular physiology of legume seed development. Annu Rev Plant Biol.;56:253-79.

Weigel D., Glazebrook J. 2002. Arabidopsis: a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Weinheimer I, Boonrod K, Moser M, Zwiebel M, Füllgrabe M, Krczal G, Wassenegger M. 2010. Analysis of an autoproteolytic activity of rice yellow mottle virus silencing suppressor P1. Biol Chem.;391(2-3):271-81.

Wellmer F, Riechmann JL, Alves-Ferreira M, Meyerowitz EM. 2004. Genome-wide analysis of spatial gene expression in Arabidopsis flowers. Plant Cell. ;16(5):1314-26.

Wesley SV, Helliwell CA, Smith NA, Wang MB, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM. 2001. Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J. 2001 Sep;27(6):581-90.

Westcott ND, Muir AD. 1996. Variation in flax seed lignan concentration with variety, location and year. Proceedings of the 56th flax institute of the United States, Double wood Inn, Fargo, North Dakota: USA.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLoS ONE 2(8): e718.

Wolters H1, Jürgens G. 2009. Survival of the flexible: hormonal growth control and adaptation in plant development. Nat Rev Genet.;10(5):305-17.

Wrzaczek M, Brosché M, Salojärvi J, Kangasjärvi S, Idänheimo N, Mersmann S, Robatzek S, Karpiński S, Karpińska B, Kangasjärvi J. 2010. Transcriptional regulation of the CRK/DUF26 group of receptor-like protein kinases by ozone and plant hormones in Arabidopsis. BMC Plant Biol.;10:95.

Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. 2012. MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. Nucleic Acids Res.; 40(Web Server issue): W127–W133.

Xiang D, Venglat P, Tibiche C, Yang H, Risseeuw E, Cao Y, Babic V, Cloutier M, Keller W, Wang E, Selvaraj G, Datla R. 2011. Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in Arabidopsis. Plant Physiol.;156(1):346-56.

Xiao JF, Li ZS, Sun M, Zhang Y, Sun CC. 2004. Homology modeling and molecular dynamics study of GSK3/SHAGGY-like kinase. Comput Biol Chem.;28(3):179-88.

Xu D, Zhang Y. 2011. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization. Biophys J.; 101(10): 2525–2534.

Yadav DK, Shukla D, Tuteja N. 2013. Rice heterotrimeric G-protein alpha subunit (RGA1): in silico analysis of the gene and promoter and its upregulation under abiotic stress. Plant Physiol Biochem.;63:262-71.

Yapo BM. 2011. Pectin Rhamnogalacturonan II: On the "Small Stem with Four Branches" in the Primary Cell Walls of Plants. Int J Carbohydrt Chem.;2011:1.

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I. 2000. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science.;287(5451):303-5.

Yesilirmak F and Sayers Z. 2009. Heterelogous Expression of Plant Genes. Int. J. Plant Genomics. 2009: 296482.

Yuan Y, Teng Q, Zhong R, Ye ZH. 2013. The Arabidopsis DUF231 domain-containing protein ESK1 mediates 2-O- and 3-O-acetylation of xylosyl residues in xylan. Plant Cell Physiol.;54(7):1186-99.

Yuan Z, Teasdale RD. 2002. Prediction of Golgi Type II membrane proteins based on their transmembrane domains. Bioinformatics;18(8):1109-15.

Zablackis E, Huang J, Müller B, Darvill AG, Albersheim P. 1995. Characterization of the cell-wall polysaccharides of Arabidopsis thaliana leaves. Plant Physiol.;107(4):1129-38.

Zamani Babgohari M, Ebrahimie E, Niazi A. 2014. In silico analysis of high affinity potassium transporter (HKT) isoforms in different plants. Aquat Biosyst.;10:9.

Zandleven J, Sørensen SO, Harholt J, Beldman G, Schols HA, Scheller HV, Voragen AJ. 2007. Xylogalacturonan exists in cell walls from various tissues of Arabidopsis thaliana. Phytochemistry.;68(8):1219-26.

Zhan S, Horrocks J, Lukens LN. 2006. Islands of co-expressed neighbouring genes in Arabidopsis thaliana suggest higher-order chromosome domains. Plant J.;45(3):347-57.

Zhang J, Wang X, Yu O, Tang J, Gu X, Wan X, Fang C. 2011. Metabolic profiling of strawberry (Fragaria x ananassa Duch.) during fruit development and maturation. J Exp Bot.;62(3):1103-18.

Zhang L, Dong W, Zhang D, Shi G. 2010. Two-stage image denoising by principal component analysis with local pixel grouping. Pattern Recognition;43:1531–1549

Zhang Y. I-TASSER server for protein 3D structure prediction. 2008. BMC Bioinformatics.;9:40.

Zhong R, Peña MJ, Zhou GK, Nairn CJ, Wood-Jones A, Richardson EA, Morrison WH 3rd, Darvill AG, York WS, Ye ZH. 2005. Arabidopsis fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell.;17(12):3390-408.

Zhu XF, Suzuki K, Saito T, Okada K, Tanaka K, Nakagawa T, Matsuda H, Kawamukai M. 1997. Geranylgeranyl pyrophosphate synthase encoded by the newly isolated gene GGPS6 from Arabidopsis thaliana is localized in mitochondria. Plant Mol Biol.;35(3):331-41.

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. 2004. GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol.;136(1):2621-32.

Ziolkovska A. 2012. Laws of flaxseed mucilage extraction. Food Hydrocolloids.:26: 197-204.

Zúñiga-Sánchez E, Gamboa-de Buen A. 2012. The Two DUF642 At5g11420 and At4g32460-Encoded Proteins Interact In Vitro with the AtPME3 Catalytic Domain. Protein Interactions, Dr. Jianfeng Cai (Ed.), ISBN: 978-953-51-0244-1, InTech, Available from: http://www.intechopen.com/books/proteininteractions/ the-two-duf642-at5g11420-and-at4g32460-encoded-proteins-interact-in-vitro-with-the-atpme3-catalytic.

Zúñiga-Sánchez E, Soriano D, Martínez-Barajas E, Orozco-Segovia A, Gamboa-deBuen A. 2014. BIIDXI, the At4g32460 DUF642 gene, is involved in pectin methyl esterase regulation during Arabidopsis thaliana seed germination and plant development. BMC Plant Biol.;14:338.