Genes Regulating Differentiation at the Shoot Apex of Flax (*Linum usitatissimum*)

by

Ningyu Zhang

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

Doctor of Philosophy

in

Plant Biology

Department of Biological Sciences
University of Alberta

ABSTRACT

Fiber harvested from flax phloem tissue is a renewable resource with promising uses in eco-friendly composites. Most molecular and cellular research to date has focused on later stages of fiber differentiation including the development of the fiber cell wall. On the other hand, the molecular mechanisms that govern specification of fibers are largely unknown. All phloem fibers in flax are formed during primary growth. Therefore transcription factors enriched in the shoot apices are likely to govern fiber identity, and therefore fiber yield. In this study, I used RNA-Seq to compare the gene expression in the apical region (AR) of the shoot apex which contained the apical-most 0.5mm of the stem and basal region (BR), which contained the entire stem except for the apical-most 1 cm. AR included the SAM and its immediate derivatives whereas BR represented all stem and vascular tissues at later stages of differentiation. The RNA-Seq study identified 349 putative transcription factors that are preferentially expressed in the AR including 18 MYBs and nine NACs. MYBs and NACs have been revealed to be required for the vascular cell identity in other species. A total of 240 putative MYBs and 182 predicted NAC domain genes were identified within the whole-genome sequence of flax. Phylogenetic analysis of the flax NAC gene family revealed that two distinct subfamilies were largely expanded. Flax had a higher proportion of R2R3-MYB than most of other sequenced plant species. Analysis of the expression data in public database indicated that the majority of *LusMYBs* and *LusNACs* were expressed in wide range of tissues with low expression level while a few others were particularly abundant in some specific tissues. Transcript expression profiling of the *LusNACs* in the VNS subfamily in 12 different flax tissues suggested that *LusNAC28* and *LusNAC125* were highly expressed in developing fibers.

A previously uncharacterized *Arabidopsis* gene, *At3g05980*, encodes a predicted protein of 245 amino acids (27.6 kDa). This protein does not contain any annotated domains, and its predicted secondary structure consists mostly of disordered coils. It has one closely-related paralog in Arabidopsis, *At5g19340*. Homologs of *At3g05980* are found in all eudicots examined, but not in any other taxa. There are four highly conserved amino acid motifs within the protein. Using qRT-PCR and GUS reporter assays, I found that transcripts of *At3g05980* were highly expressed in immature embryos and the micropylar endosperm, as well as petals, and apices of shoots and roots, and atrichoblasts. Transcripts were highly induced by cold treatment, but not by other stress or hormone treatments. These results were consistent with expression patterns previously reported in public databases. I produced loss-of-function (LOF) mutants of this gene, using CRISPR/Cas9-mediated gene editing, as well as overexpression (OX) lines using the 35S-CaMV promoter. LOF lines were morphologically indistinguishable from wild-type, but OX lines had minor defects, including cotyledon epinasty, and slight shortening of both plant height and silique length. Neither LOF nor OX differed from WT in tolerance to freezing. In the absence of cold-treatment, LOF mutants had increased transcript abundance of the stress- and cold-responsive gene RD29, compared to WT, but expression patterns of five other cold-responsive genes were largely unchanged in LOF, compared to wild-type, both before and during cold treatment. Translational fusions of *At3g05980* with fluorescent proteins were localized to peroxisomes. However, assays of peroxisomal function, including dark growth of seedlings, and sensitivity to 2, 4-DB and IBA, were similar between LOF, OX, and WT. Furthermore, fatty acid profiling of seeds did not show any difference between the genotypes. Thus, *At3g05980* encodes a eudicot-specific, peroxisomaly localized protein with transcripts that are cold-inducible, and enriched in specific tissues (particularly rapidly growing tissues), but this gene does not appear to be required for normal

morphology, peroxisomal function, or cold tolerance responses. The immediate future task will be to examine phenotypes in double mutants of both *At3g05980* and it paralog *At5g19340*.

PREFACE

This thesis is the original work by Ningyu Zhang. Chapter 2 of this thesis has been published as follows:

Zhang, N., & Deyholos, M. K. (2016). RNASeq Analysis of the Shoot Apex of Flax (*Linum usitatissimum*) to Identify Phloem Fiber Specification Genes. *Frontiers in Plant Science*, *7*, 950. http://doi.org/10.3389/fpls.2016.00950

I conducted all experiments, and assisted in analysis and writing of the manuscript. MD designed experiments and assisted in analysis and writing of the manuscript.

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisor Dr. Michael Deyholos for his immense support, inspiration, visionary guidance, and great patience throughout the course of my study. I also would like to thank Dr. Jocelyn Hall who acted as my supervisor after Mike moved to the UBC and gave me immense help on in situ hybridization and phylogenetics. I would also thank Dr. Enrico Scarpella who supplied valuable guidance on my project including confocal microscopy and histochemical staining.

I would also like to express my deep gratitude to Neil Hobson, Shanjida Khan and Arlene Oatway, for their help in experiment performance, troubleshooting and theory discussion. I am truely grateful to Shane Carey for all the help he gave me. I deeply appreciate all other labmates (Leonardo Galindo, David Pinzon, Zhu Dan, Haiyan Zhuang, Kashfia Faruque, Lai to, Anupreeti Ramdoss, Brandi and Amanda).

Finally, I would like to thank all my family and friends for giving me encouragement and support behind me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX

LIST OF ABBREVIATION

TILLinG	targeting induced local lesions in genomes
VIGS	virus-induced gene silencing
NPA	1-naphthylphthala- mic acid
SAM	shoot apical meristem
LOF	loss of function
CDS	coding DNA sequence
EST	expressed sequence tag
GUS	beta-glucuronidase
qRT-PCR	quantitative real-time polymerase chain reaction
TAIR	The Arabidopsis Information Resource
T-DNA	transfer deoxyribonucleic acid
X-Gluc	5-bromo-4-chloro-3-indolyl-ß-D-glucuronide
$K_3 Fe(CN)_6$	potassium ferricyanide
$K_4 Fe(CN)_6$	potassium ferrocyanide
IBA	indole-3-butyric acid
IAA	indole-3-acetic acid
GC-MS	gas chromatography-mass spectrometry
OX	overexpression
2, 4-DB	4-(2,4-dichlorophenoxy)butyric acid
VND	vascular-related NAC-domain
NST	NAC secondary wall thickening promoting factor
SND	secondary wall-associated NAC domain protein
SMB	SOMBRERO
BRN	BEARSKIN
VNS	VND-, NST/SND-, SMB-related proteins
GO	Gene Ontology
G-layer	gelatinous-layer
Gn-layer	galactan-enriched layer
ANOVA	analysis of variance
WT	wild-type
Col-0	Columbia-0
CT	threshold cycles
DBD	DNA-binding domain
DAS	days after sowing
ABA	abscisic acid
IAA	3-indoleacetic acid
BA	6-benzylaminopurine
MeJA	methyl jasmonate
BR	brassinosteroid
ACC	1-aminocyclopropane-1-carboxylic acid

GA3	gibberellic acid-3 potassium salt
PTS	peroxisome targeting signal
Basta	glufosinate-ammonium or phosphinothricin
GFP	green fluorescent protein
CiFP	citrine fluorescent protein
CA	cold acclimated conditions
NA	nonacclimated conditions
COR	cold-regulated gene

Chapter 1. Literature review

1.1 Flax

Flax (*Linum usitatissimum*) is a eudicot crop grown primarily in temperate regions of the world (Rubilar et al, 2010). It belongs to the family Linaceae and the order Malpighiales. *Linum* is composed of approximately 180 species (McDill et al., 2009; Sveinsson et al., 2014). As a slender herbaceous plant, flax can grow up to 1.2 meters tall. It bears lanceolate leaves and blue flowers. Its fruit is a small, round, dry capsule 5-9 mm in diameter, containing up to ten brown or yellow seeds (depending on cultivar type). Flax seeds have a glossy surface, and are typically 4-6 mm length (Nôžková et al., 2014).

Flax is grown for its either stem phloem (bast) fibers or its seeds. Due to their great length and high tensile strength, flax phloem fibers are currently used as a valuable material in production of textiles, high-quality papers and reinforcing composite polymers (Deyholos, 2006). Flax seeds are enriched in a number of components that are beneficial for our health, such as dietary fiber which benefits our digestive health, and omega-3 fatty acids which can improve our brain function (Carter, 1993; Rubilar et al., 2010; Rabetafika et al., 2011). Flax seed is also the richest source of lignan, which is beneficial for cardiovascular system and has reported anticancer function. Flax seed oil is also an important ingredients of paints, varnishes and linoleum (Singh et al., 2011).

Domesticated flax is thought to have been derived from *Linum bienne* Mil, a wild flax species (Diederichsen & Hammer, 1995; Fu & Allaby, 2010; Uysal et al., 2010). They share certain common characteristics such as blue flowers, strong stems and 15 pairs of chromosomes. These

two species can be crossed and the progenies are fertile. The botanical origin of flax is believed to be either the Indian subcontinent or the Mediterranean East (Vavilov, 1951).

1.1.1 Linseed or fiber flax

Flax cultivated for seeds and fibers are usually of different varieties, and they are named linseed and fiber flax, respectively. Through divergent selection for thousands of years, linseed and fiber flax have gained considerably different morphology, physiology, anatomy and agronomic properties. Linseed cultivars are usually shorter, more branched and produce more and larger seeds. On the other hand, fiber flax tends to be taller and less branched, but produces more and higher quality fibers. Linseed cultivars are grown in the continental climate region of Canada, China, India, the United States and Argentina while fiber flax cultivars are grown in the cool climate areas such as some areas of China, Russia and Western Europe (Reddy et al., 2009). Linseed cultivars produce fibers as well but these fibers are undesirable due to their low yield, inferior quality and short length. In fact, fibers are deemed a nuisance for linseed varieties since they are prone to be stuck in the harvesting or processing machine. Recently, developing a use for linseed straw has become an active area of research.

1.1.2 Cultivation history of flax

Flax is one of the oldest plants domesticated by humans. The earliest evidence of fiber flax use is 30,000-year old knotted wild flax fibers discovered in Dzudzuana Cave, located in the foothills of the Caucasus, Georgia (Kvavadze et al., 2009). By contrast, linseed flax is assumed to have been originally cultivated as food resource in Fertile Crescent region, based on discovery of seeds with increased size were found at the Tell Ramad archeological site in Syria (Vanzeist & Bakkerheeres, 1975). Over the last two centuries, flax cultivation had experienced a dramatic

decline due to the rise in cotton, jute cultivation and appearance of synthetic fibers and oils. In the early 2000s, cultivation of flax resurged in part because some biologically active components in its seeds were proven to be beneficial to human health (Deyholos, 2006).

1.1.3 Flax as a research model

Flax is not only valued for its industrial application and health benefits, but it has also been used as a research model to study plant cell growth, phloem development and cell wall formation. For instance, members of the flax genus (*Linum* spp.) have been used historically as models for the study of shoot apical meristems (SAMs; Esau, 1942). In contrast to fibers produced in many other plant species, flax phloem fiber cells undergo a extensive, intrusive elongation and they are large and grouped into bundles and are therefore easier to isolate. In addition, the cell elongation and secondary cell wall thickening of flax phloem fibers are spatially and temporally separate (Gorshkova et al., 2003; Gorshkova et al., 2005). Additionally, flax has some other desirable traits that make it attractive to scientists: i) it is treated as a diploid, with a small genome (approximately 373 Mb). The flax genome was sequenced in 2012 through whole-genome shotgun sequencing, releasing sequences of 43,384 putative genes, which could be aligned to 93% of the published flax ESTs and 86% of *Arabidopsis thaliana* genes, suggesting a good coverage (Wang et al., 2012); ii) The growth cycle of flax is relatively short, around 100 days including a vegetative period of 45-50 days, 15-25 days of flowering and a maturation period of 30-40 days; iii) Flax is highly self-pollinating. The outcrossing rate is as low as 0.3 to 2.0% under normal circumstances and remains 1 to 5% even when the flax plants are grown in close proximity (Dillman, 1938).

Various forward or reverse genetic approaches are available in flax, providing important tools for gene function analysis. Since its initial application to flax two decades ago, agrobacterium-induced

transformation has become an indispensable tool in flax functional genomics research (McHughen, 1989). An EMS mutant population with high mutant rate (1/41 kb) was generated in flax and a TILLinG (Targeting Induced Local Lesions IN Genomes) platform based on endonuclease ENDO1 was developed. This population contains a total of 4,894 independent M2 families, of which 10,839 individual plants from 4,033 M2 families have been phenotyped and 1,552 families (38.5%) were visually abnormal. All the available flax mutant phenotype data can be found in UTILLdb (http://urgv.evry.inra.fr/UTILLdb). Other next generation sequencing-based mutant identification approaches are being developed in this and other EMS populations (Chantreau et al., 2013; Galindo-González et al., 2015). Recently, a VIGS (Virus-Induced Gene Silencing) method has been reported in flax, which will accelerate the functional characterization of individual candidate genes (Chantreau et al., 2015).

Meanwhile, numerous studies describing transcript profiling data and proteomic data associated with flax fiber differentiation, seed development or stress responses have been released (Roach & Deyholos, 2007; Roach & Deyholos, 2008; Yu et al., 2014; Dash et al., 2014; Hotte & Deyholos, 2008; Day et al., 2013; Hradilová et al., 2010). A high-resolution consensus genetic map has been established for flax from three mapping populations which include 770 ordered markers in 15 linkage groups spanning 1,551 cM. On average, there is one marker per 2.0 cM. 670 molecular markers from the consensus genetic map has been anchored to the flax physical map and 204 of the 416 flax fingerprinted contigs were covered (Cloutier et al., 2010; Cloutier et al., 2012).

1.2 Flax phloem fibers

1.2.1 Plant fiber cells

Fibers are present in many vascular plants and are defined as sclerenchyma with an elongated shape (the ratio between cell length and diameter ranging from 50 to 2000 or even more), tapered ends, and a secondary cell wall up to 15 μm in thickness. The main role of plant fiber cells is to provide mechanical support for the plant body (Snegireva et al., 2015). Fibers can be found in various organs such as root, stem, leaves and seeds. Those existing in the primary body may be derived from the procambium such as in cereals, palms, reeds and bamboo, or from the ground meristem such as in the outer interfascicular sectors of the *Arabidopsis* pith. Others in secondary plant body are derived from vascular cambium (Esau, 1965). Fibers are one of the longest plant cells. Generally, the longest fibers are those produced in the primary phloem while the shortest are those present in the secondary xylem (Fahn, 1982; Chernova & Gorshkova, 2007). The quality of fibers is primarily determined by their strength and flexibility, which again depend on their cell length and cell wall composition. The main cell wall components of plant fibers are cellulose, hemicellulose and lignin, and the quantities of these components vary between different plant species, different plant parts and plant ages. Cellulose is the strongest and stiffest component of fiber (Ramamoorthy et al., 2015).

Phloem fibers are the most commercially valuable fibers. Phloem fibers contain more cellulose (up to 90%) and much less lignin and xylan than other types of fibers, resulting in its higher tensile strength and flexibility. Phloem fibers are mostly used in the production of textile while recently there has been a surge in using phloem fibers to replace the fiberglass in composites. The reason lies in the fact that natural fiber based composites have lower density, better mechanical and

acoustic properties, higher processing properties and neutral ecobalance. The major phloem fiber crops are flax, ramie, hemp, jute and kenaf (Ramamoorthy et al., 2015).

1.2.2 Flax phloem fiber differentiation

Flax phloem fibers are found in the primary phloem poles of the stele, and they are arranged into bundles of 12-40 cells (Ageeva et al., 2005). Flax fibers are unique because of their great length and extremely thick cell wall (Gorshkova & Morvan, 2006). Phloem fibers of flax originate from the shoot apical meristem during primary growth. When the procambium is first formed, the cells which will become future phloem fiber cannot be distinguished. In flax, sieve tubes and companion cells mature earlier than fibers and after their maturation, fibers continue to become widen and elongate. The elongating and expanding fibers gradually intrude between surrounding cells, which may damage sieve tubes and companion cells. Flax phloem fibers differentiate in a gradient along the length of the stem, and their development can be divided into three general stages: (i) specification; (ii) cell elongation; (iii) cell wall thickening (Gorshkova et al., 2003).

1.2.2.1 Ontogeny and specification of flax phloem fibers

Specification of phloem fibers occurs in the apical-most 0.5 mm of the flax stem since young phloem fibers can be anatomically distinguished at 0.4-0.5 mm from the stem tip (Ageeva et al., 2005). How fiber cell identity is specified remains elusive. This is partially due to the fact that studying the fiber initiation with the classic biological methods is difficult. When we can see fiber cells, the cellular factors specifying fiber identity may have already completed their activity. Currently, biochemical or molecular-genetic markers of early phloem fibers are not available, and therefore identification of the fibers at their earliest developmental stages has to rely on their

characteristic positions and morphological features including elongated shape and broader diameter.

However, through studying model plants such as Arabidopsis, poplar and zinnia cell culture, a lot of information has been achieved about the molecular network regulating vascular initiation, and these findings mainly focus on the procambium or cambium establishment and specification of xylem or phloem as a whole or xylem differentiation (Reviewed by Ohashi-Ito and Fukuda 2014). Briefly, the canalization of auxin fluxes results in the procambial cell specification, and the procambial cells divide periclinally to give rise to the procambium tissue. The procambium tissue undergoes a series of differentiation events and forms specialized xylem and phloem cells.

Auxin flow initiates procambial cell differentiation

Exogenous application of the hormone auxin has long been known to trigger vascular tissue initiation, and later in 1981, a model named 'canalization of auxin flow hypothesis' was proposed. This model proposed that auxin produced in apical meristems initially moves towards the root and in undifferentiated cells through diffusion. This directional auxin flow induces some cellular changes of the recipient cells which allow rapid auxin flow. The canalization of auxin flow in narrow cell files then establishes a local auxin maximum and initiates vascular tissues formation (Jacobs, 1952; Sachs, 1969; Sachs, 1981). This hypothesis was later confirmed by many molecular genetic studies (Ruthardt et al., 2005; Scarpella et al.,2006; Wenzel et al., 2007; Wenzel et al., 2007).

In Arabidopsis, studying mutants with defects in vascular tissue formation or patterning showed that many genes previously reported to be involved in auxin biosynthesis, transport and signaling were critical for vascular tissue initiation or patterning (Reviewed in Caño-Delgado et al.,2010).

For example, loss-of-function mutation in MONOPTEROS (MP) gene, also known as auxin response factor 5, led to a highly reduced leaf vein system and misaligned tracheary elements in the inflorescence stems and leaves (Berleth & Jürgens, 1993; Przemeck et al., 1996). MP activates the expression of PINFORMED1 (PIN1) gene, an auxin efflux carrier. Mutation of *PIN1* or treatment of plants with NPA, an inhibitor of auxin efflux carriers, altered the vein patterns in leaves (Mattsson et al., 1999; Sieburth, 1999).

Further research supports the importance of *MP* in vascular differentiation. *MP* up-regulates the expression of *ATHB8* (a HD-ZIP III transcription factor), a positive regulator of procambial and cambial cell proliferation as well as xylem differentiation (Baima et al., 2001). Four other HD-ZIPIII transcription factors activated by MP are *ATHB15/CORONA* (*CNA*), *PHAVOLUTA* (*PHV*), *PHABULOSA* (*PHB*), and *REVOLUTA* (*REV*). These act redundantly with *ATHB8* to promote xylem mother cell proliferation (Ohashi-Ito & Fukuda, 2010). MP expression is stimulated by auxin (Wenzel et al., 2007). Apart from this, *MP* induces the expression of TARGET OF MP 5 (TMO5), TMO7, and TMO6, transcription factors promoting procambial cells initiation in embryo (Schlereth et al., 2010). TMO5 forms dimeric complex with LONESOME HIGHWAY (LHW). The TMO5- LHW complex immediately follow the feedback auxin signaling loop comprising PIN1, MP and ATHB8 during the initiation of procambium precursor cell differentiation (Ohashi-Ito & Fukuda, 2010).

Cytokinin signaling acts to regulate the balance between procambial maintenance and xylem/phloem differentiation

Cytokinin plays a key role in promoting procambial cell formation and maintenance. *Wooden leg (wol*), a Arabidopsis mutant of a cytokinin receptor, forms additional protoxylem vessels through procambial cell differentiation (Mähönen et al., 2006a). This gene encodes a histidine kinase

known as CRE1 or ATHK4, which is preferentially expressed in the procambium (Mähönen et al., 2000). Mutation of three cytokinin receptors (ATHK2, ATHK3 and ATHK4) simultaneously results in ectopic protoxylem formation (Mähönen et al., 2006a). CKI1, another histidine kinase, was also revealed to mediate the procambium/cambium activity through cytokinin signaling pathway (Hejátko et al., 2009). Cytokinin receptors act together with downstream components, such as histidine-containing phosphotransfer factors. A cytokinin signaling inhibitor, AHP6 (ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN6) was revealed to be specifically expressed in the protoxylem and positively regulate protoxylem differentiation (Mähönen et al., 2006b). On the other hand, AHP6 was found to be an MP-targeted gene and therefore it is auxin-dependent (Bishopp et al., 2011).

A mobile peptide hormone is involved in procambial /cambial cell maintenance

TDIF (TRACHEARY ELEMENT DIFFERENTIATION INHIBITORY FACTOR), a phloem cell-produced peptide encoded by *CLE41* and *CLE44*, moves apoplastically to procambial/cambial cells where it is perceived by TDR (TDIF RECEPTOR), a leucine-rich repeat (LRR) –receptor like kinase. After TDR perceives TDIF, expression of WOX4 (WUS-related homeobox4) is activated. WOX4, a transcription factor expressed in procambial of Arabidopsis and cambial cells in *Populus*, plays a central role in regulating procambial or cambial cell maintenance (Schrader et al., 2004; Ji et al., 2010). The TDIF-TDR complex is also involved in the prohibition of xylem differentiation (Hirakawa et al. , 2011).

Differentiation of procambial cells into xylem cells

Brassinosteroids were reported to promote xylem differentiation by stimulating the expression of HD-ZIP III transcription factors (Ohashi-Ito & Fukuda, 2003; Motose et al., 2004). As mentioned previously, several HD-ZIP III transcription factors, including *PHB*, *PHV*, *REV*, *CNA* and *ATHB8,*

act redundantly to positively regulate the xylem specification from procambial cells. Transcripts of these HD-ZIP III transcription factors are known to be negatively regulated by miRNA165 and miRNA166, whereas SHORT ROOT (SHR)-SCARECROW (SCR) transcription factor complex induces the expression of miRNA165 and miRNA166.

Initiation of xylem cell differentiation

Several NAC domain transcription factors are known to regulate xylem differentiation. Specifically, VND6 (Vascular-related NAC-domain 6) and VND7 induce the metaxylem and protoxylem vessel differentiation, respectively. Overexpression of VND6 and VND7 induce the ectopic differentiation metaxylem and protoxylem vessel respectively, from both vascular cells and nonvascular tissues (Kubo et al., 2005). Likewise, SND1 (Secondary Wall-associated NAC Domain Protein1) and NST1 (NAC Secondary Wall Thickening Promoting Factor1) genes in NAC domain transcription factor family are proven master regulators of the initiation of fiber differentiation (Mitsuda et al., 2007).

Phloem development

In comparison with the great progress obtained in understanding the regulation of xylem development, far less is known about the specific regulatory factors involved in the developmental commitment to phloem cell fates. ALTERED PHLOEM DEVELOPMENT (APL) is the first identified phloem development regulator. Mutation of *APL* led to formation of xylem-like cells at the phloem positions and ectopic expression of this gene inhibited the xylem development (Bonke, Thitamadee, Mähönen, Hauser, & Helariutta, 2003). It indicated that *APL* gene positively regulated phloem differentiation while negatively regulated the xylem differentiation. *NAC45* and *NAC86* are two known target genes of APL produced during phloem differentiation (Furuta et al., 2014). Two polar membrane-associated proteins, OCTOPUS (OPS) and BREVIS RADIX (BRX)

were found to promote sieve element identity and its maintenance (Rodriguez-Villalon et al., 2014). The CLAVATA3/EMBRYO SURROUNDING REGION45 (CLE45) peptide is a negative regulator of protophloem differentiation and it functions by interacting with BARELY ANY MERISTEM3 (BAM3) receptor-like kinase (Depuydt et al., 2013).

1.2.2.2 Phloem fiber elongation

After specification, flax fibers elongate extensively to become one of longest plant cells (around 77 mm in some varieties; Mohanty et al., 2000). At the early stages of fiber elongation, fibers grow symplastically with the surrounding tissues. Flax phloem fibers undergoing symplastic growth have several characteristics: i) they have flat ends and cell diameters that are approximately 4-7 μm; ii) they have an elongated shape and readily transmit light because of their large vacuole; iii) they usually have elongated nuclei and may be multinucleate due to the occurrence of karyokinesis (Ageeva et al., 2005). At the end of symplastic growth, flax phloem fiber grows to approximately 70-100 μm in length (Snegireva et al., 2010).

Later, fibers undergo an extensive cell elongation through intrusive growth. During intrusive growth, fibers grow faster in longitudinal orientation than the surrounding cells. Therefore, they intrude the surrounding cells and penetrate the middle lamella (Esau, 1965). Fibers penetrate the neighboring cells by their ‘knees’. Similar ‘knees’ are formed at both ends, implying that during the intrusive growth, flax phloem fibers elongate in both directions. Flax phloem fibers with this type end can be first identified at 300-500 μm below the shoot apex. During the intrusive growth, diameters of flax phloem fibers also increase several fold and the total cell volume may increase many thousand fold (Gorshkova et al., 2012). The intrusive growth of flax phloem fibers start before the surrounding cells finish their symplastic growth and the whole internode stop elongating

(Ageeva et al., 2005). The intrusive elongation of flax phloem fibers occurs through diffused growth, during which their whole surface expand (Ageeva et al., 2005; Gorshkova et al., 2003).

Fiber cell intrusive elongation involves two processes: intensive cell vacuolization and cell wall extension. Cell enlargement could be initiated by changes of turgor pressure or cell wall extensibility. However, there is still not yet definitive data to indicate whether changes in turgor pressure or cell wall extensibility are the determinant factors of initiation or termination of intrusive fiber elongation (Gorshkova et al., 2012).

The hormone gibberellin has long been known to promote the differentiation and elongation of fibers in both xylem and phloem. However, exogenous application of gibberellin promotes the elongation of other cells and internode elongation to the same extent as the fibers, suggesting that ability of gibberellin to the promoting elongation is not specific to the fibers (Gorshkova et al., 2012). Recently, gibberellin was found to have a specific function in fiber elongation and it was reported to upregulate the genes encoding enzymes involved in pectin degradation in aspen. The intrusive growth of fibers is accompanied by the splitting of middle lamellae which resembles wound reactions induced by pathogen attack, while the wound effect is not induced during fiber intrusive growth. Fibers undergoing intrusive elongation does not express wound response marker genes and some genes non-specifically induced by wound, like chitinases and β-1,3-glucanases genes, are not significantly induced during fiber intrusive growth (Roach & Deyholos, 2007; Snegireva et al., 2010; Gorshkova et al., 2012).

1.2.2.3 Secondary cell wall thickening of flax phloem fibers

Flax phloem fiber cell elongation lasts for several days. After this, fibers start to deposit thick secondary cell walls. These two stages are spatially separated. The transition point between them is called the snap point, which is a location along the stem. The snap point can be first identified in 3 week-old plants by manual detection. Fibers below the snap point have higher mechanical strength which make flax stems harder to be manually torn. As plant grow, the snap point migrates apically but it finally disappears when the stem growth ceases and plants start to flower. This occurs in 7 week-old flax (Gorshkova et al., 2003; Ageeva et al., 2005; Snegireva et al., 2010).

The outermost fibers in flax stems are the first to develop secondary cell walls, and this process occurs even before all traces of protophloem sieve elements have disappeared. The secondary cell walls of flax phloem fibers are of the gelatinous type and at the early stage of deposition, are composed of two layers: an inner heterogenous and loosely packed galactan-enriched layer (Gn-layer); and an outer, more homogenous gelatinous-layer (G-layer). Later, the Gn-layer is gradually transformed into the G-layer. When mature, flax phloem fibers are almost completely composed of G-layers (Gorshkova & Morvan, 2006; Gorshkova et al., 2004).

The gelatinous secondary cell wall has several other characteristics: i) it contains a high amount of crystalline cellulose, usually 80-90%. This property is partially attributed to the modification of galactan. During the transition from Gn-layer to G-layer, galactan undergo partial hydrolysis and become a relatively smaller molecule which is tightly bound to cellulose microfibrils (Mikshina et al., 2009; Gurjanov et al., 2008). Additionally, Roach et al. found that the β-galactosidase activity within the precursor Gn-layer is a determining factor for this process (Roach et al., 2011); ii)

cellulose microfibrils in G-layer are almost parallel to the fibers' longitudinal axis; iii) when mature, the thickness of G-layer can reach 10 μm or more while cell wall of general plant cells is only 0.1-1 μm in thickness; and iv) they do not contain or only contain trace quantity of xylans and lignin. All these unique characteristics contribute to the high mechanical strength of flax phloem fibers (Gorshkova et al., 2010).

1.3 Gene expression pattern in plant shoot apical meristem

In higher plants, all of the above-ground structures are generated from the shoot apical meristem (SAM), which serves as a stem cell reservoir. The SAM can be generally divided into three different regions: the central zone (CZ) present at the summit of the shoot apex, the peripheral zone (PZ) surrounding the CZ and the underlying rib zone (RZ). Stem cells are located in the CZ, where they produce daughter cells by asymmetric cell divisions. One of their daughter cells remains as a stem cell while the other one will be displaced into the PZ where it will differentiate into various specialized cell types and will be recruited into lateral organogenesis. In the past decade, shoot apex transcriptomes have been described in various plant species, including maize, pea, soybean, rice, *Arabidopsis* and chickpea, but the transcriptome analysis of the flax shoot apex is still lacking (Emrich et al., 2006; Wong et al., 2008; Wang et al., 2014; Haerizadeh et al.,2009; Jiao et al., 2009; Yadav et al., 2009; Singh & Jain, 2014). Moreover, although extensive transcript profiling data about flax fiber development has been published, these have all focused on later stages of development (Day et al., 2005; Roach & Deyholos, 2007; Roach & Deyholos, 2008).

1.4 This research

The goal of this research is to identify key regulators of phloem fiber specification, which, based on examples (e.g. xylem specification), are likely to include particular transcription factors (Kubo

et al., 2005; Yamaguchi et al, 2010a; Yamaguchi et al, 2010b). Because anatomical data indicates that phloem fiber specification occurs very near the SAM, we targeted the shoot apex for this analysis. We compared the gene expression patterns in the apical-most 0.5 mm of the shoot apex to the mature and mature tissues located more basally within the stem (Chapter 2). Additionally, because NAC and MYB transcription factors have been found to play an important role in plant vascular cell differentiation, I performed a phylogenetic and expression analysis of NAC and MYB transcription factors in flax (Chapter 3 and Chapter 4; Yamaguchi et al., 2008; Ohashi-Ito et al., 2010; Zhong et al., 2007a; McCarthy et al., 2009). Finally, I also characterized an *Arabidopsis* gene of unknown function (Chapter 5) that has a flax ortholog that was enriched in the shoot apical meristem as reported in Chapter 1.

1.4.1 The importance of this research

This study will provide insight into the transcriptome of the flax shoot apex and may also point to candidate transcription factors that govern the specification of flax phloem fiber identity. Even though it is difficult, exploring the molecular mechanisms underlying fiber initiation may have enormous economic impact and it will also increase our understanding about cell differentiation and tissue patterning.

Chapter 2. RNA-Seq analysis of the shoot apex of flax (*Linum usitatissimum*) to identify phloem fiber specification genes

2.1 Introduction

All of the post-embryonic, above-ground structures of seed plants are generated from the shoot apical meristem (SAM), which acts as a reservoir of stem cells. Members of the flax genus (*Linum* spp.) have been used historically as models for the study of SAMs (Esau, 1942). Cultivated flax (*Linum usitatissimum*) is grown in more than 50 countries for its seeds or its stem phloem (bast) fibers (Rubilar et al., 2010). Due to prolonged intrusive growth, and a highly crystalline cellulosic secondary wall, flax phloem fibers are among the longest and strongest cells in plants (Mohanty et al., 2000). In flax, all phloem fibers are derived from primary growth in the shoot apex. Specification of phloem fibers occurs in the apical-most 0.5 mm of the stem, since young phloem fibers can be anatomically distinguished starting 0.4–0.5 mm from the shoot apex (Gorshkova et al., 2003). The molecular mechanisms that govern fiber identity are almost entirely unknown (Gorshkova et al., 2012). Also, in contrast to the significant progress obtained in the past decade toward understanding xylem differentiation, information about the phloem fiber differentiation is very scarce (Rybel et al., 2016). In the past decade, shoot apex transcriptomes have been described in various plants, including maize, pea, soybean, rice, Arabidopsis and chickpea, but none of these produce significant primary phloem fibers (Ohtsu et al., 2007; Wong et al., 2008; Haerizadeh et al., 2009; Jiao et al., 2009; Yadav et al., 2009; Wang et al., 2014). Most molecular and cellular research on flax fiber has thus far focused on later stages of development (Day et al., 2005; Roach and Deyholos, 2007; Feinart et al., 2010).

Differential transcript expression data from the region of the shoot apex in which fiber specification occurs would complement other approaches (e.g., mutant screening) aimed at understanding primary phloem fiber differentiation.

2.2 Materials and methods

2.2.1 Plant materials

Flax (i.e., linseed) plants (*L. usitatissimum* L. cv. CDC Bethune; Rowland et al., 2002) were grown in potting mix in an environmental chamber at 22°C, with a cycle of 16 h light and 8 h dark, as previously described (Wang et al., 2012). Fourteen days after germination (Figure 2-1A), approximately 0.5 mm of the apical-most part of each stem (the apical region, AR) was dissected under a Leica S6D stereo microscope, all visible leaf primordia were removed, and the tissue was frozen in liquid nitrogen. A representative dissection, visualized under an environmental scanning electron microscope, is shown in Figure 2-1B, and transverse sections of a shoot apex, corresponding to the apical and basal-most tissues sampled, are shown in Figures 2-1C, 2-1D. Shoot apices were similarly dissected from approximately 200 plants and pooled prior to each RNA extraction. After collecting the shoot apex, the remainder of the stem (i.e., the basal region, BR) from 1 cm below the shot apex to the stem base was also dissected, stripped of leaves, visible lateral branches and axillary meristems, and frozen in liquid nitrogen. In this way, mature stems from at least six plants were pooled for each RNA extraction. For RNA-Seq of the AR, samples were harvested from four biological replicates (i.e., four sets of plants that were grown spatially and temporally independently from each other), and tissues obtained from two biologically independent replicates were used for the BR. For qRT-PCR, three additional, independent

biological replicates (i.e., different plants than those used for RNA-Seq were obtained from each of the AR and BR).

2.2.2 RNA extraction, sequencing and data processing

RNA from each biological replicate (Section Plant Materials) was extracted separately. RNeasy Micro Kit and RNeasy Plant Mini Kit were used to isolate RNA from the AR and BR samples, respectively. Extracted RNA was then digested with TURBO DNA-free™ Kit to remove DNA contamination and their quality was evaluated using a RNA 6000 Nano chip on an Agilent 2100 Bioanalyzer. Total RNA was delivered to the service provider, BGI, where each biological replicate was sequenced separately. Oligo(dT)-coupled magnetic beads were used to isolate poly-A+ mRNA, which was used as a template for cDNA synthesis primed by random hexamers, followed by second strand synthesis using *E. coli* DNA PolI. Double-stranded cDNA (Qiaquick PCR Purification Kit), was sheared with a nebulizer, end repaired, and ligated to Illumina PE adapter oligos, and the products size-selected by gel purification to produce 200 bp fragments. These were PCR amplified through 15 cycles to prior to sequencing using an Illumina HiSeq 2000 with 90 bp, paired-end reads. The quality of the sample during processing prior to sequencing was monitored using the Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System. Because the sequencing output for samples AR2, AR3, and AR4 was slightly lower than expected (9.6 million reads output per sample), additional aliquots of each of these three samples were sequenced in three additional runs. Raw reads from all runs were filtered to remove adapter sequences, contamination, and low-quality reads, and the filtered raw reads were deposited in the SRA archive. Each of the nine paired read files were uploaded to SRA in fastq format.

To quantify the relative abundance of transcripts in the shoot apex (AR) as compared to the remainder of the stem (BR), the clean sequencing reads described in Section RNA Extraction and Sequencing were mapped to the flax reference genome (Wang et al., 2012; downloaded from Phytozome 9 as Lusitatissimum_200.fa) using Tophat2 (Trapnell et al., 2012), and the accepted hits were used as input for cufflinks, with default parameters. All potential splicing isoforms were treated by cufflinks as representing the same transcript.The resulting assemblies were merged and with the reference genome annotation (downloaded from Phytozome 9 as Lusitatissimum_200_gene.gff3) with cuffmerge, and finally Cuffdiff was used to calculate normalized differential transcript abundance between the samples.

2.2.3 qRT-PCR

Reference genes used in the qRT-PCR analysis were selected by comparing the expression stability of nine housekeeping genes (Listed in the Appendix 1) in the AR and BR following the previous description (Huis et al., 2010). Real-time PCR was performed in Applied Biosystems 7500 Fast Real-time PCR System following the manufacturer's protocol. Each amplification reaction was 10 µl and consisted of 0.4 µM of each primer, 5 µl SYBR Green Master Mix and 2.5 µl 16-fold diluted cDNA. Threshold cycles (C_T) were determined through 7500 Fast Software. The PCR program used was as follows: 95°C for 2 min, 40 cycles of 95°C for 10 s and 60°C for 30 s, then 72°C for 30 s and 72°C for 3 min; fluorescence data was collected at 60°C. Data were analyzed using the $2^{-\Delta\Delta}C_T$ method (Kenneth & Schmittgen, 2001). Primer sequences used are listed in the Appendix 2.

2.2.4 Gene Ontology analysis of the differentially expressed genes

Gene Ontology enrichment was performed for the AR preferentially expressed genes and BR preferentially expressed genes by the Singular Enrichment Analysis (SEA) in agriGO V2.0 using the following parameters: hypergeometric test, Yekutieli multi-testing adjustment, significance level 0.05, 5 minimum mapping entries, Plant Slim GO (Tian et al., 2017). All the flax transcripts in Phytozome v11.0 were used as background (Goodstein et al., 2012).

2.3 Results

2.3.1 Analysis of gene expression in the AR and BR of flax stem

Regulators governing phloem fiber specification are assumed to be expressed in the shoot apical meristem because these regulators should operate before phloem fibers can be anatomically distinguished. To investigate the expression patterns of genes in the shoot apical meristem, I investigated gene expression in two different segments dissected from whole flax stems from which leaves and leaf primordia have been removed: (i) the apical region (AR), which was comprised of the 0.5 mm apical-most stem segment, and (ii) the basal region (BR), which comprised the region from 1.05 cm below the shoot apex to the base of stem. The AR was expected to contain cells undergoing specification as fibers, while the BR was expected to contain fibers at various stages of differentiation. A total of four biological replicates of AR and two biological replicates of BR were sequenced. After sequencing, the adapter sequences, contamination, and low-quality reads were filtered. As a result, a total of 9.6 to 22 million high-quality clean reads were obtained from each sample and these clean reads were then mapped to the flax genome by Tophat2 (Table 2-1; Wang et al., 2012).

Furthermore, we qualified the relative abundance of transcripts in the shoot apex (AR) as compared to the remainder of the stem (BR) by cuffdiff and we found 1791 transcripts and 2011 transcripts were specifically expressed in the AR and BR respectively, while 38,044 transcripts were expressed in both AR and BR. Moreover, transcripts for 6207 genes were revealed to be significantly ($q < 0.05$) more abundant in AR compared to BR, and 4405 of these were enriched at least 2-fold in the AR. Conversely, transcripts for 8388 genes were significantly ($q < 0.05$) more abundant in BR compared to AR, and 7901 of these were enriched at least 2-fold in the BR.

2.3.2 Quantitative real-time PCR analysis of differential transcript abundance

To evaluate the accuracy of the differential transcript expression measurements that we obtained, we used qRT-PCR to measure transcript abundance in independently grown replicates of the same tissues that were used for RNA-Seq. In order to select an appropriate reference gene for the qRT-PCR, GeNorm was used to determine the expression stability of nine commonly used reference genes among tissues assayed in our study (Huis et al., 2010). GADPH and ETIF5A were found to be the most stable, and ETIF5A gene chosen arbitrarily from this pair as the internal control (Appendix 1). Thirteen genes were selected for qRT-PCR, as an independent validation of the accuracy of the RNA-Seq results (Figure 2-2). These genes were selected in part because they were all transcription factors from gene families that could be potentially associated with early differentiation events in the shoot apex including specification of vascular/phloem identity (Zhao et al., 2005; Kalve et al., 2014; Rybel et al., 2016). As shown in Figure 2-2, the RNA-Seq and qRT-PCR analysis showed highly consistent expression patterns for the 13 genes tested. We therefore conclude that that RNA-Seq data presented here accurately represents differences in transcript expression between the shoot apical region (AR) and the bulk of the stem (BR).

2.3.3 GO enrichment analysis of differentially expressed transcripts

2.3.3.1 AR preferentially expressed genes

To further understand the function of the differential genes, Gene Ontology (GO) enrichment analysis was performed for the AR preferentially expressed genes and BR preferentially expressed genes respective. 61 significantly enriched GO terms were identified in the AR preferentially expressed genes based on FDR<0.05, including 21 in terms of biological process, 13 in terms of molecular function, 27 in terms of cellular component (Figure 2-3a; Figure 2-3b).

These 21 enriched GO terms in biological process mainly belong to three big categories: metabolic process, cellular process and developmental process (Figure 2-3a). The specific metabolic process overrepresented in the AR preferentially expressed genes were nitrogen compound metabolic process (*p*-value=9.97e-43), translation (*p*-value=4.77e-38), and DNA metabolic process (*p*-value=8.19e-19) whereas the cellular process terms mainly pointed toward cell cycle (*p*-value=3.2e-17) and cellular metabolic process (*p*-value=7.02e-28) which was again pointed toward the translation and DNA metabolic process. Furthermore, the enriched GO terms in developmental process categories were anatomical structure development and multicellular organism development (Figure 2-3a).

In terms of molecular function, the predominant GO terms were structural molecular activity (*p*-value=3.63e-38) and nucleic binding (*p*-value=3.63-38) including DNA and RNA-binding (Figure 2-3a). The 27 overrepresented cellular component GO terms included many high level GO terms which defined very great range (Figure 2-3b). However, through examine the specific localizations under these high-level terms on the hierarchical graph generated by agriGO2, I found that the gene

products of AR-enriched genes were mainly located in the ribosome (*p*-value=1.34e-42), cytoskeleton (*p*-value=2.46e-10) and nucleus (*p*-value=1.17-32).

2.3.3.2 BR preferentially expressed genes

On the other hand, 11 significantly enriched GO terms were identified for the genes preferentially expressed in BR. The most predominant biological process GO term was photosynthesis (*p*-value=9.85e-20) while GO terms including generation of precursor metabolites and energy, lipid metabolic process, localization establishment as well as transport were also revealed to be significantly overrepresented in the BR enriched genes (Figure 2-4). As shown in Figure 2-4, two GO terms in cellular components (thylakoid and membrane) and molecular function (transporter activity and catalytic activity) were overrepresented in the BR-enriched genes.

2.3.4 Transcription factors significantly more enriched in the AR

Analysis through PlantTFDB predicted that 373 and 437 transcriptions were preferentially expressed in AR and BR respectively, including 27 AR-specific genes and 58 BR-specific genes (Jin et al., 2017). These transcription factors belonged to 46 families and 11 families had members preferentially expressed in AR but not BR, including ARR-B, BBR-BPC, CPP, E2F/DP, FAR1, GRF, HB-PHD, S1Fa-like, SRS, STAT and LFY (Figure 2-5; Appendix 3). Notably, the flax genome is predicted to encode only two STAT transcription factors in total, and both were found to be AR-enriched (Figure 2-5). Furthermore, genes in AP2, B3, GeBP and NF-YC family were also highly upregulated in the AR (Figure 2-5). In contrast, bZIP, C2H2, Dof, WRKY, NAC, ERF and the HSF families were significantly enriched in the BR (Figure 2-5). Inspection of all AR-enriched transcription factors found that 49 transcription factors encoding genes were at least 16-fold more enriched in the AR compared to BR (Table 2-2).

2.4 Discussion

In this study, we compared the gene expressions in the AR and BR of the flax stem by RNA-Seq. The aim of this analysis was to identify transcriptional regulators of phloem fiber specification, considering that phloem fiber cell identity specification occurs in the shoot apical meristem. Inspection of the data showed that several markers of shoot apex tissues were highly enriched in the AR sample. For example, *PROTODERMAL FACTOR 1* (*PDF1*) transcripts have been reported to be expressed exclusively in the L1 layer of meristems and the protoderm of organ primordia (Abe et al., 1999). In our results, transcripts of putative *PDF1* genes (Lus10007351, Lus10031390, Lus10010941) were at least 19.5-fold more abundant in AR than BR (Zhang & Deyholos, 2016). Similarly, *CUP-SHAPED COTYLEDON* (*CUC*) genes are required for SAM function and organ separation (Hasson et al., 2011). Transcripts of three putative *CUC* genes (Lus10041924, Lus10005537, Lus10013205) were at least 45-fold more abundant in AR than BR; two other putative CUC genes (Lus10037106, Lus10003458) were not detected in either sample. As a third example, the SHOOT MERISTEMLESS (STM) transcription factor is essential for SAM formation and maintenance (Endrizzi et al., 1996); a putative STM gene (Lus10030003) was 4.8-fold enriched in the AR sample compared to BR. Conversely, several markers of late differentiation were more enriched in the BR compared to the AR. For example, *CELLULOSE SYNTHASE A* (*CESA*) genes *CESA4, CESA7*, and *CESA8* are associated with secondary wall synthesis (Chantreau et al., 2015). We observed transcripts of flax genes annotated as *CESA4* (Lus10008225, Lus10008226), and *CESA8* (Lus10007296, Lus10029245) to be at least 125-fold enriched in the BR compared to the AR (no CESA7 genes were identified in the original flax genome annotation used in this study). Another well-established marker of xylem differentiation is *XYLEM CYSTEINE PROTEINASE-2* (*XCP2*; Avci et al., 2008). The two putative

flax *XCP2* genes (Lus10030722, Lus10013204) were enriched 106-fold in the BR compared to the AR. Thus, expression of at least some well-known markers of early and late stem development were observed in patterns that matched expectations.

GO enrichment analysis indicated that genes involved in DNA metabolism and the cell cycle were over-represented in the AR preferentially expressed genes. This is related to the active cell division found in the shoot apical meristem and this finding was consistent with what was previously reported (Yadav et al., 2009). Besides, genes involved in translation were also revealed to be significantly enriched in the AR. I found among the 273 AR-enriched genes involved in translation, 233 had the same molecular function: structural constituent of ribosome. This was reasonable since AR included many constantly dividing meristematic cells and ribosomes were therefore largely abundant in the AR required for protein synthesis. Besides, as reported previously in pea shoot apical meristems, 'nucleus' and 'ribosome' were overrepresented cellular component classifications and 'nucleic acid binding' and ' structure molecule activity' were overrepresented molecular functions for the enriched genes (Liang et al., 2009).

GO enrichment suggested that the BR preferentially-expressed genes were dominated by genes associated with photosynthesis and the thylakoid compartment (Figure 2-4). Photosynthesis-related genes have been reported to have lower transcript abundance in the pea shoot apical meristem compared to the non-meristematic tissues (Wong et al., 2008). It was indicated that only non-meristem cells in plants have the photosynthetic machinery (Fleming, 2006). Meristem cells are heterotrophic since they only contain proplastids, which lack the thylakoid structure of functional chloroplasts and they do not contain chlorophyll and express the proteins required for

photosynthesis (Fleming, 2006). As found in previous report, genes involved in 'transport' and 'generation of precursor metabolites and energy' or encode products in membrane were also significantly enriched in the meristem containing AR compared to the nonmeristematic tissue BR (Liang et al., 2009). Checking the specific AR-enriched genes in 'generation of precursor metabolites and energy' categories found that most genes in these categories were also related to the photosynthesis.

This study found 373 transcription factors significantly more enriched in the AR compared to the BR and 49 of them were 16 times more abundant. Based on the function of their Arabidopsis orthologs, some of these 49 genes might be involved in the flax shoot apical meristem formation (e.g. *Lus10041924*), shoot apical meristem maintenance (e.g. *Lus10002657*, *Lus10016809*, *Lus10032098*, *Lus10026432*, *Lus10005282*, *Lus10013960* and *Lus10001238*), epidermal cell fate determination (*Lus10014933*, *Lus10023568*, *Lus10007643*), and floral organ development (e.g. *Lus10039214*, *Lus10035029* and Lus10016732). However, the function of most of the transcription factors significantly enriched in the AR were not yet characterized and these genes may also have an important function related to meristem maintenance or organogenesis. Meanwhile, the 349 AR-enriched genes should contain some transcriptional regulators of flax phloem fiber specification. Further characterization of these genes will be necessary.

2.5 Conclusions

This study has compared the transcriptomic difference between the AR and BR region of flax and this will improve our understanding of SAM function and maintenance in general. Transcripts of 90% of genes were detected in both AR and BR. 14,595 (35%) genes were differentially expressed

between AR and BR. A total of 6207 transcripts (including 373 transcription factors) were significantly more abundant in the AR. These genes deserve further investigation to uncover the molecular mechanisms underlying primary phloem fiber differentiation.

2.6 Figures and tables

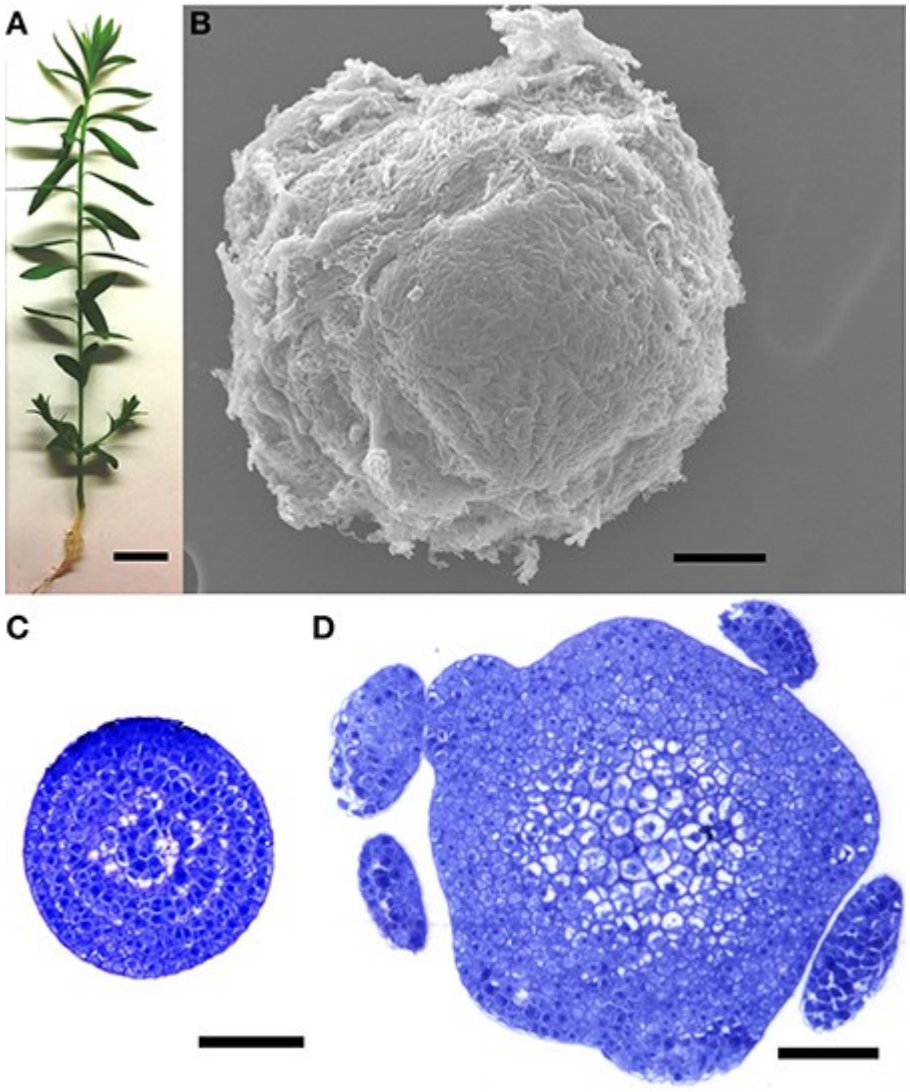


Figure 2-1 Plant tissues used for RNA-Seq library construction. (A) A 14-day plant at the time of dissection. (B) Environmental scanning electron micrograph of an unfixed, dissected shoot apical region (AR), representative of the tissue used for RNA extractions. (C, D) Transverse sections through the apical (C) and basal (D) limits of the shoot apical region (AR), showing extent of morphological differentiation at time of RNA extraction. Plants used for RNA extraction did not contain the leaf primordia seen in (D). Scale bars (A) 1 cm; (B–D) 50 μm.

Table 2-1 A summary of the RNA-Seq data.

Sample	SRA Accession	read orientation	clean reads	mapped reads	of mapped reads, # aligned to multiple loci	of mapped reads, # discordant
AR1	SRR1056618	left end	18737601	17521982	1322985	
		right end	18737601	17401016	1307125	
		pair	18737601	16547868	1206086	277915
AR2	SRR1056620	left end	9653298	9055491	1137842	
		right end	9653298	8954883	1123304	
		pair		8514534	1065943	239296
	SRR1056621	left end	9655838	9056753	1142941	
		right end	9655838	8952853	1127935	
		pair		8510856	1069985	238369
AR3	SRR1056622	left end	9647100	9068825	808752	
		right end	9647100	8960836	795159	
		pair		8526247	739690	206678
	SRR1056623	left end	9652208	9070825	806840	
		right end	9652208	8961837	793980	
		pair		8525212	738322	206728
AR4	SRR1056624	left end	9659902	9041639	781706	
		right end	9659902	8924493	767587	
		pair		8493377	712478	177096
	SRR1056625	left end	9666281	9048163	781482	
		right end	9666281	8927476	766501	
		pair		8496385	711739	176566
BR1	SRR1038482	left end	18811289	17715907	1282625	
		right end	18811289	17526289	1260744	
		pair		16878680	1191166	301986
BR2	SRR1421513	left end	22066254	20798802	1704912	
		right end	22066254	20813734	1708057	
		pair		19897998	1610722	383886

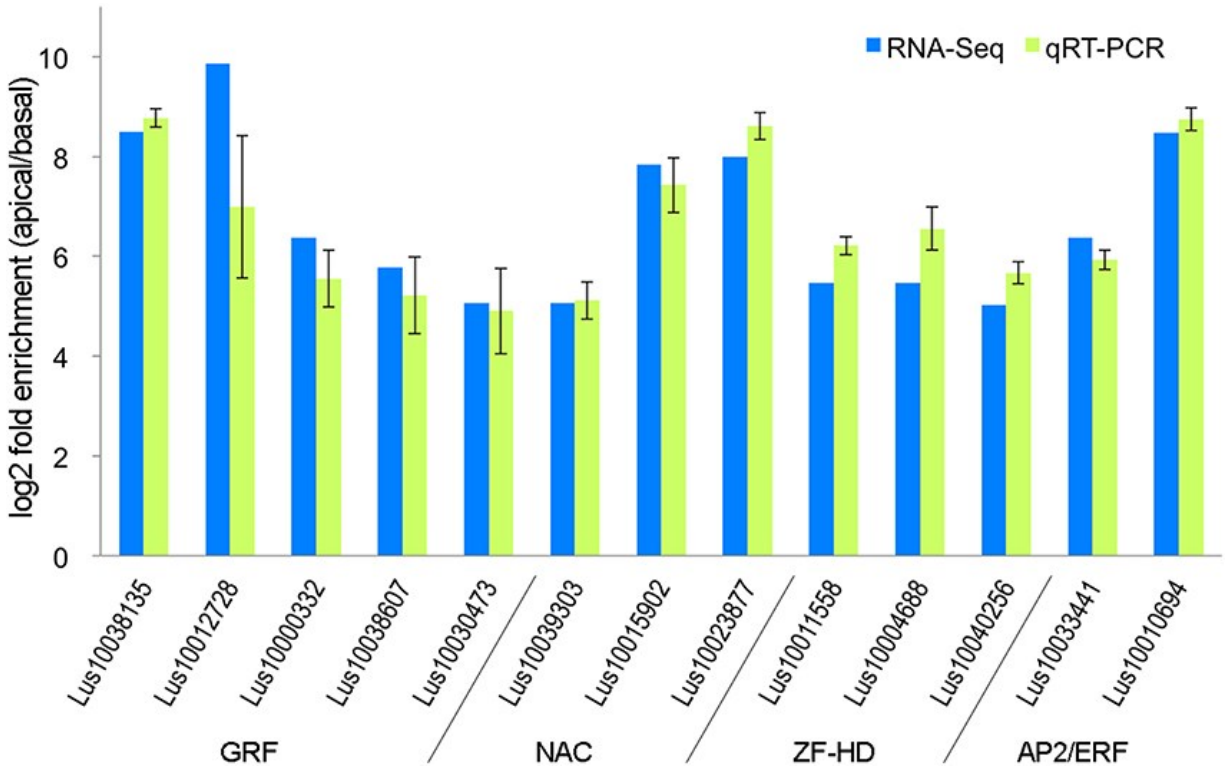


Figure 2-2 Ratio of transcript abundance in the stem apical region (AR) compared to the basal region (BR), as measured by qRT-PCR and RNA-Seq on independently grown tissues.

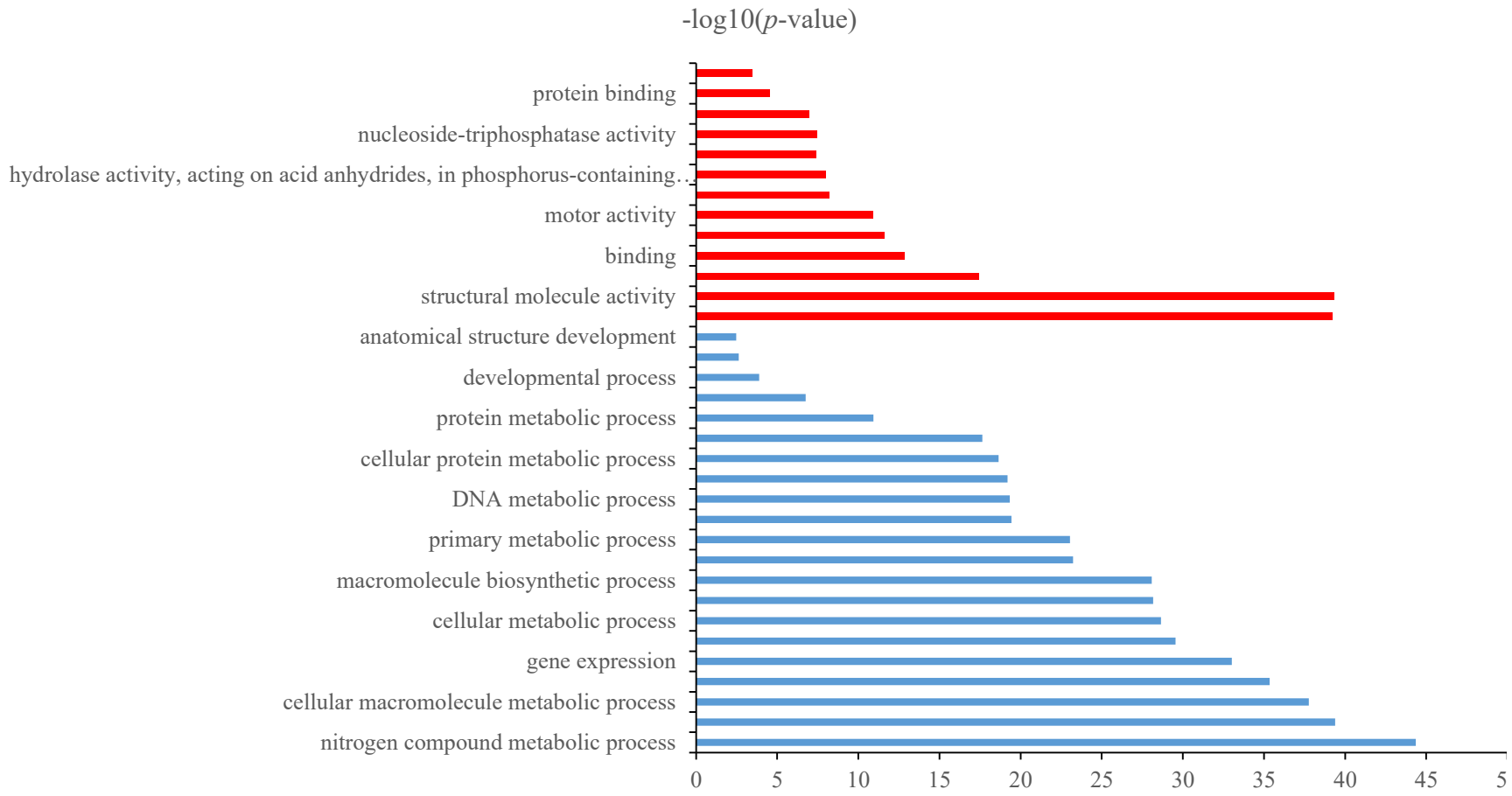


Figure 2-3a. GO terms (Biological Process and Molecular Function) significantly enriched in the AR preferentially expressed genes.

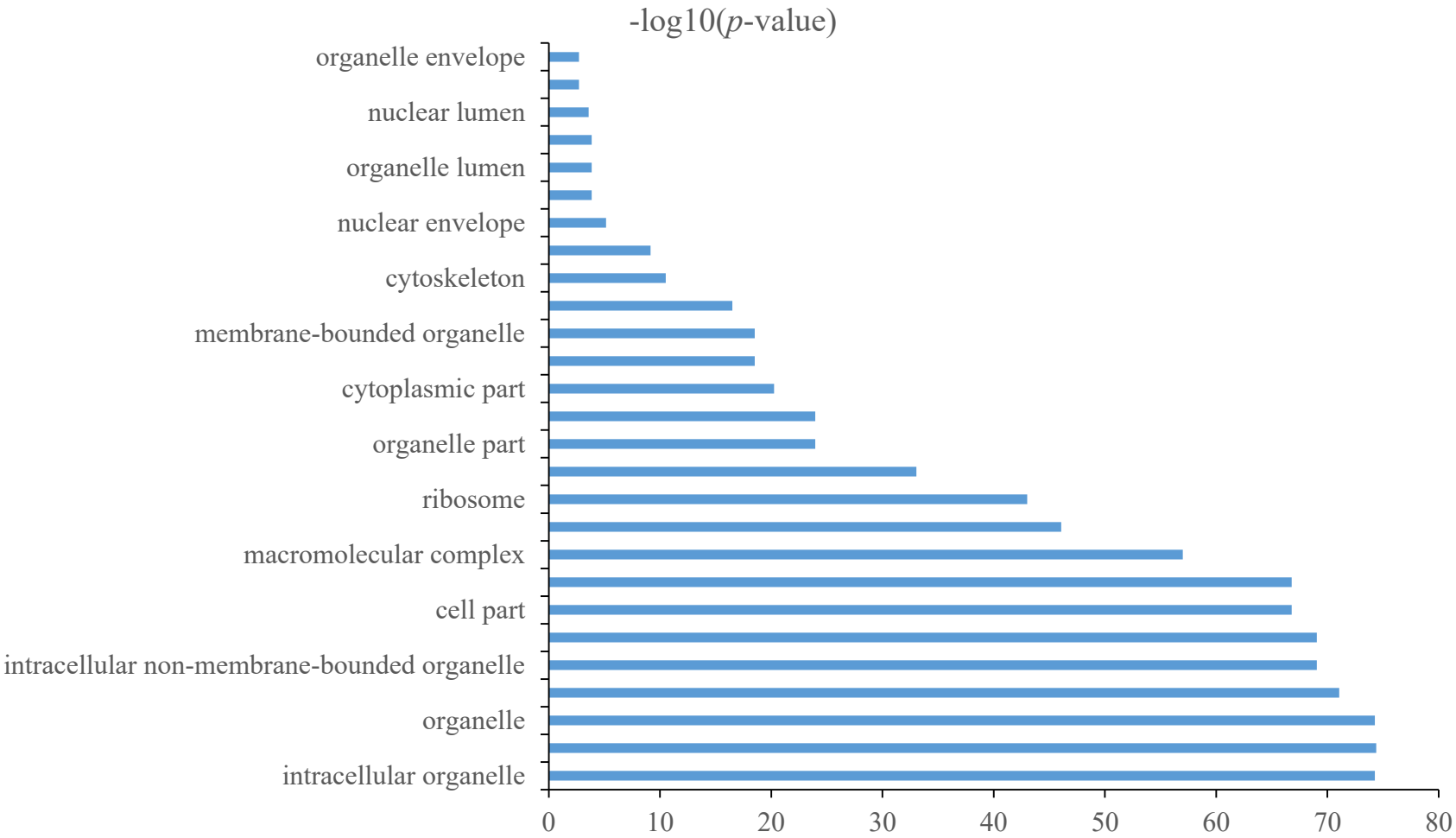


Figure 2-3b. GO enrichment of the AR preferentially expressed genes in terms of cellular component.

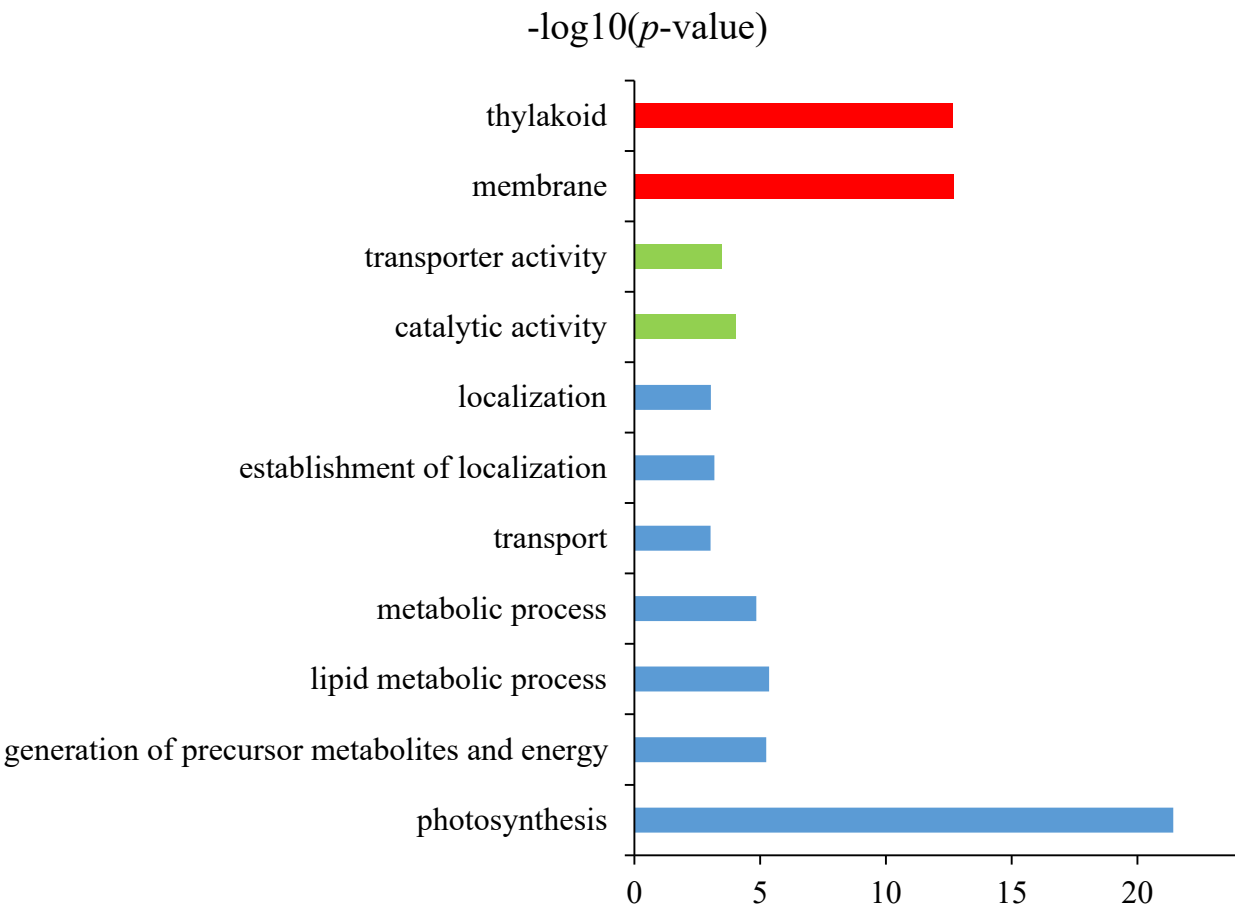


Figure 2-4 GO enrichment of the BR preferentially expressed genes in terms of cellular component (red bars), molecular function (green bars) and biological processes (blue bars).

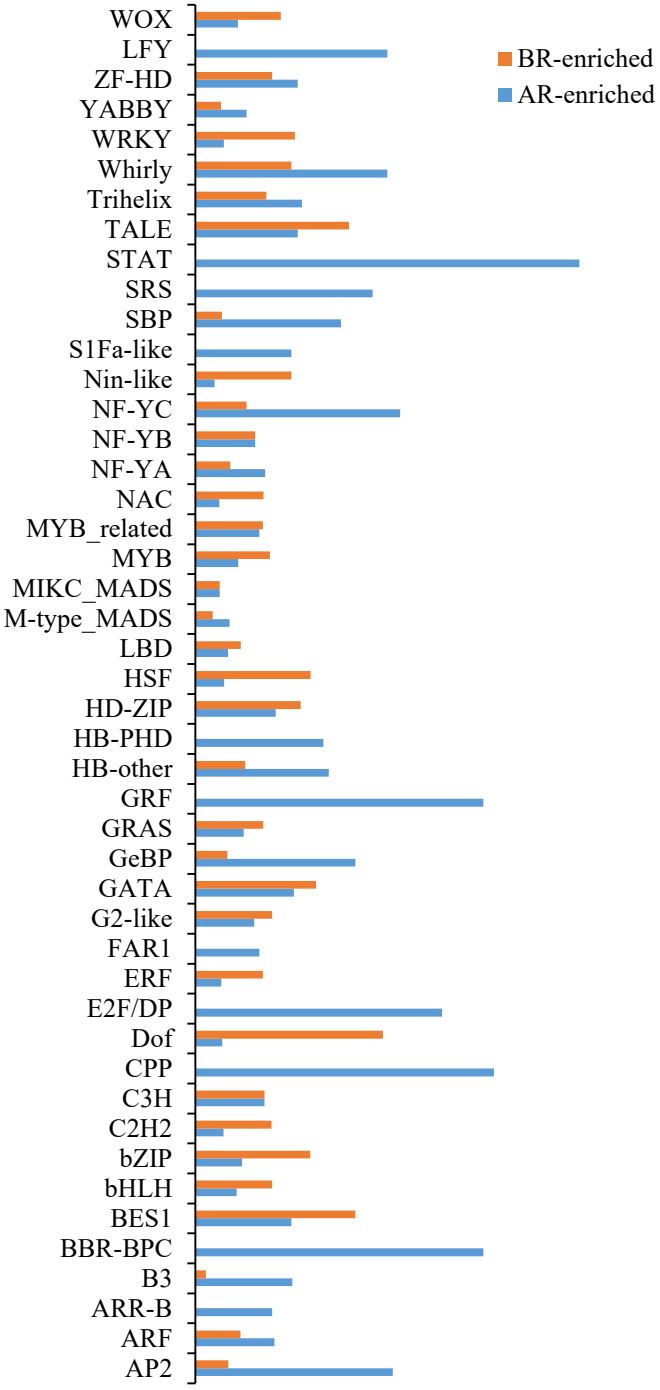


Figure 2-5 Differential expression patterns of different transcription factor families in flax AR and BR.

Table 2-2 Transcription factors with over 16-fold more abundant in AR than BR.
Note: 'inf' indicates infinity.

TF ID	Family	FPKM (AR)	FPKM (BR)	log2(fold_change AR/BR)	q_value
Lus10002657	AP2	17.8611	0.0529	-8.3992	0.0343
Lus10038135	ZF-HD	83.1918	0.5731	-7.1816	0.0072
Lus10037668	GRF	81.3636	0.6441	-6.9810	0.0154
Lus10040453	MYB_related	1907.7400	18.3992	-6.6961	0.0003
Lus10015902	bHLH	45.8242	0.5461	-6.3909	0.0005
Lus10033441	GRF	67.5686	1.0211	-6.0482	0.0003
Lus10007147	ZF-HD	135.6130	2.2800	-5.8943	0.0003
Lus10013205	NAC	21.6260	0.4813	-5.4898	0.0021
Lus10011559	GRF	88.8187	2.1356	-5.3781	0.0003
Lus10004688	TALE	94.6156	3.3132	-4.8358	0.0003
Lus10016809	M-type_MADS	8.5552	0.3061	-4.8048	0.0062
Lus10039303	B3	40.8492	1.6464	-4.6329	0.0003
Lus10032098	B3	18.0971	0.8001	-4.4994	0.0003
Lus10026432	TALE	59.5062	2.6867	-4.4692	0.0003
Lus10017434	B3	26.5536	1.2609	-4.3964	0.0003
Lus10019275	GRF	97.2257	4.6230	-4.3944	0.0003
Lus10011558	GRF	95.1001	4.6204	-4.3634	0.0003
Lus10035093	G2-like	12.3215	0.6329	-4.2830	0.0022
Lus10001238	TALE	16.2423	0.8568	-4.2446	0.0024
Lus10040256	TALE	132.3690	7.1855	-4.2033	0.0003
Lus10039214	MYB	7.6462	0.4154	-4.2022	0.0296
Lus10014302	ZF-HD	2.8567	0.1718	-4.0559	0.0377
Lus10037670	AP2	4.1253	0.0000	inf	0.0003
Lus10000747	B3	1.1021	0.0000	inf	0.0003
Lus10012046	B3	12.7802	0.0000	inf	0.0003
Lus10012226	ERF	4.1387	0.0000	inf	0.2108
Lus10014345	ERF	3.1392	0.0000	inf	0.0003
Lus10015653	ERF	37.4187	0.0000	inf	0.0003
Lus10032882	GRAS	4.6761	0.0000	inf	0.0003
Lus10014380	GRF	115.8940	0.0000	inf	0.0756
Lus10030800	HD-ZIP	3.8459	0.0000	inf	0.0003
Lus10009336	LBD	2.5880	0.0000	inf	0.0003
Lus10016732	LFY	1.0584	0.0000	inf	0.0003
Lus10028214	M-type_MADS	10.0320	0.0000	inf	0.2251
Lus10035029	M-type_MADS	2.2205	0.0000	inf	0.4331

Lus10016139	MYB	3.3814	0.0000	inf	0.0003
Lus10018518	MYB	2.8838	0.0000	inf	0.0003
Lus10021428	MYB	10.6303	0.0000	inf	0.0003
Lus10038092	MYB	30.9376	0.0000	inf	0.0003
Lus10007643	MYB_related	6.8005	0.0000	inf	0.0089
Lus10014933	MYB_related	2.9425	0.0000	inf	0.0606
Lus10023568	MYB_related	1.8089	0.0000	inf	0.0661
Lus10041924	NAC	11.7809	0.0000	inf	0.0003
Lus10018283	Trihelix	41.1782	0.0000	inf	0.0003
Lus10027398	Trihelix	3.6627	0.0000	inf	0.0003
Lus10031672	Trihelix	5.3463	0.0000	inf	0.0003
Lus10005282	WOX	2.2659	0.0000	inf	0.0003
Lus10013960	WOX	5.5696	0.0000	inf	0.0003

Chapter 3. Genomic-wide characterization of the MYB transcription factor superfamily in flax

3.1 Introduction

MYB domain proteins are one of the biggest transcription factor families in plants. In *Arabidopsis*, 9% of transcription factors belong to this family (Riechmann et al., 2000). This gene family has a one-billion-year-old history and is represented in genomes of all major eukaryotic lineages (Kranz et al., 2000). The oncogene v-MYB, a determinant of avian myeloblastosis, is the first named MYB transcription factor (Klempnauer et al., 1982). Three types of MYB-related genes (c-MYB, A-MYB and B-MYB) were subsequently discovered in vertebrates, and they were revealed to play important roles in cell differentiation, proliferation and apoptosis (Weston, 1998). The first characterized MYB gene in plants is *Zea mays C1*, which regulates anthocyanin biosynthesis (Avila et al., 1993).

MYB proteins are defined based on the conserved MYB DNA-binding domain (MYB DBD) at their N-terminus. The MYB DBD is composed of imperfect repeats of an approximately 50- 53 amino acid region and each repeat forms a helix-helix-loop-helix secondary structure that binds to the major groove of the target DNA (Lipsick, 1996; Stracke et al., 2001). Several conserved Trp residues present in the MYB DNA-binding domain are important for its specific binding to target DNA (Nagadoi et al., 1995). However, the sequences at the C-terminus of MYB proteins are highly divergent (Jiang et al., 2004b; Kranz et al., 1998).

The MYB superfamily is classified into four major types based on the number of MYB repeats: 1R-MYB (also called MYB-related proteins), 2R-MYB (R2R3-MYB proteins), 3R-MYB (R1R2R3-MYB proteins), and 4R-MYB (atypical-MYB), containing one, two, three and four repeats of the MYB motif, respectively. All the known MYBs in animals are 3R-MYB. Most higher plant genomes described to date have approximately five 3R-MYB genes, which have a conserved role in cell cycle regulation (Ito, 2005; Haga et al., 2007; Rosinski & Atchley, 1998; Kranz et al., 1998; Dubos et al., 2010). However, 2R-MYB is the major MYB type in plants, with most genomes encoding at least 100 of these genes (e.g. 124 in *Arabidopsis*, and 192 in poplar; Wilkins et al., 2009). These are involved in diverse biological and physiological processes, such as cell morphogenesis, meristem formation, cell cycle regulation, hormone signaling, secondary metabolism, abiotic and biotic stress responses (Baumann et al., 2007; Ito et al., 2001; Abe et al., 2003; Dai et al., 2007; Deluc et al., 2006; Johnson & Dowd, 2004; De Vos et al., 2006). 4R-MYB is the smallest group, and has only one member in many plant genomes. Furthermore, the function of plant 4R-MYB is still unknown (Dubos et al., 2010). MYB-related proteins were suggested to be involved in circadian regulation, cellular morphologies, secondary metabolism, organ morphogenesis, phosphate starvation as well as chloroplast development (Lu et al., 2009; Pesch & Hülskamp, 2009; Simon et al., 2007; Dubos et al., 2008; Kerstetter et al., 2001; Waters et al., 2009; Rubio et al., 2001).

The MYB transcription factor family has been comprehensively analyzed in many plant species, such as Arabidopsis, soybean, rice, grape and poplar (Stracke et al., 2001; Du et al., 2012; Yanhui

et al., 2006; Matus et al., 2008; Du et al., 2013; Wilkins et al., 2009). By contrast, only limited information has been obtained about MYBs in flax even though genomic sequences of flax have been released (Huis et al., 2010; Wang et al., 2012). In addition, our lab is interested in understanding the mechanisms regulating flax cell wall formation and vascular differentiation, which may also involve MYBs. For example, Arabidopsis MYB46 and MYB83 were revealed to function as key regulators of cellulose, hemicellulose, and lignin biosynthesis in vessels and xylary fibers (Zhong et al., 2007a; Zhong & Ye, 2012). In this study, I have performed a genome-wide identification of MYB domain protein in flax and analyzed the gene structures, phylogeny and expression patterns of 2R-, 3R and 4R MYBs. A separate, large group of 'MYB-related proteins' were beyond the scope of this study. I am specifically interested to learn whether MYBs have roles in flax phloem fiber cell identity specification. We assumed that transcriptional regulators of flax phloem fiber cell specification should be abundant in the shoot apex, therefore I have investigated the *LusMYBs* that showed preferential expressions in the AR compared to the BR from the RNA-seq dataset described in Chapter 2 of this thesis. Further functional studies of these genes (e.g., through mutant analysis) may help to decode the genetic basis of primary phloem fiber identity. Taken together, this study may provide important clues for future research on the functions of MYB in flax growth and development.

3.2 Material and methods

3.2.1 Materials

Refer to Section 2.2.1 for the methods to grow plants, collect samples, extract RNA and prepare cDNA.

3.2.2 Genomic-wide identification of MYB transcription factors in flax genome

All the 43,384 protein sequences in the flax whole genome shotgun assembly were downloaded from Phytozome v. 11.0 (https://phytozome.jgi.doe.gov/pz/portal.html; Wang et al., 2012; Goodstein et al., 2012). Protein sequences of all the *Arabidopsis* MYB transcription factors were obtained from TAIR 9.0 (https://www.arabidopsis.org/; Yanhui et al., 2006; Stracke et al., 2001). BLASTP program in BLAST package 2.3.0+ (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST) was used to query *Arabidopsis* MYB transcription factors against the 43,384 predicted flax proteins downloaded from phytozome v 11.0 (https://phytozome.jgi.doe.gov/pz/portal.html). Hits with E-values > 10^{-10} and redundant hits were manually removed. The resulting protein sequences were further analyzed through PROSITE server (http://prosite.expasy.org/prosite.html) to confirm the presence of MYB domain (Castro et al., 2006). Any proteins with non-MYB conserved domains were excluded. The molecular weight, isoelectric point and amino acid lengths were calculated using Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/protein_mw.html).

3.2.3 Phylogenetic analysis

Sequences of *Arabidopsis thaliana* and *Populus trichocarpa* 2R-, 3R *and* 4R-MYB proteins were downloaded from TAIR (https://www.arabidopsis.org/) and a previous study conducted by Chai (Chai et al., 2014) respectively. The full-length amino acid sequences of flax, *Arabidopsis* and poplar MYB transcription factors were aligned using the multiple sequence alignment program MUSCLE with the default parameters and the phylogenetic tree was constructed using neighbor-joining method using Mega 5.0 with the following parameters: Poisson correction, uniform rates

and pairwise gap deletion mode (Edgar, 2004; Tamura et al., 2011) . The bootstrap value applied was 1000. The phylogenetic tree was then rooted at the mid-point.

3.2.4 Meta-analysis of flax MYB gene expression

3.2.4.1 EST identification

The coding sequences (CDS) of *LusMYBs* were used as queries to search the flax EST database (accessed Mar 2017; 286,856 sequences) in NCBI by BLASTn. Only ESTs with at least 95% identity to *LusMYB* CDS were selected.

3.2.4.2 Microarray

Microarray datasets with accession numbers GSE21868 and GSE29345 were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). GSE21868 measured transcribed abundance in leaves (L), roots (R), stem inner tissue at vegetative stage (SIV) or green capsule stage (SIGC), stem outer tissue at vegetative stage (SOV) or green capsule stage (SOGC), as well as embryos of 10, 20 and 40 days after flowering (designated as E1, E2 and E3 respectively; Fenart et al., 2010). Transcript expression was also compared between two different flax cultivars, Drakkar and Belinka. The former genotype produces better fibers and has higher resistance to *Fusarium oxysporum* (a fungal pathogen; Fenart et al., 2010). GSE29345 compared expression of genes in different parts of flax stem, including internal stem tissues of either the whole stem (WSI), upper stem (USI), middle stem (MSI), or lower stem (LSI); and external stem tissues of the whole stem (WSE), upper stem (USE), middle stem (MSE), and lower stem (LSE; Huis et al., 2012). Probes used in these microarray studies were designed based on EST sequences (https://urgi.versailles.inra.fr/Species/Flax/Download-sequences). These ESTs were queried

against the 187 putative *LusMYB* gene coding sequences (CDS) by BLASTN. Only those with 90% length match to the *LusMYB* CDS and the sequence identities, not less than 95% were considered. The cutoff E-value was 10^{-10}. Heat maps were created using the mean RMA-normalized, averaged gene-level signal intensity (log2) values of all the biological replicates by MultiExperiment Viewer (MeV v4.9, http://www.tm4.org/ -mev.html). Genes were hierarchically clustered with Pearson correlation and the single clustering method. The log2 signal values have been mean-centered before clustering. This involves taking the mean expression value for each gene or transcript, and subtracting it from each expression value for that gene or transcript. The mean value will then be zero.

Expression of flax MYB genes were also investigated in an unpublished microarray dataset performed in our laboratory studying gene expression profiles in five stages of flax stem development (To, 2013). Tissues were collected from 3 weeks old flax plants from which all leaves had been removed. Stem segments of 1 cm were dissected from five different parts of flax stem: the shoot apex (T1), 1 cm stem segment above the snap-point (T2), at the snap point (T3) and below the snap point (T4) and the 1 cm stem segment from bottom of flax stem in which phloem fiber cells have deposited thick secondary cell wall (T5). Probes were aligned to the published whole genome shotgun assembly of flax by BLASTN analysis and only those 100% identical to the flax MYB genes CDS were analyzed (Wang et al., 2012). Log-normalized signal intensities were used to make heat maps by MeV v4.9 (http://www.tm4.org/mev.html). To find the

differentially expressed *LusMYBs* in at least one of these five segments, a two-way ANOVA with Turkey's multiple comparisons was performed.

3.2.4.3 RNA-seq

Expression patterns of *LusMYBs* were analyzed in the normalized RNA-seq dataset published by Kumar et al (Kumar et al., 2013). In addition, I compared the expression values of putative flax MYB genes in the RNA-seq data I have described in Chapter 2 of this thesis.

3.2.4.4 qRT-PCR

qRT-PCR analysis was conducted to confirm expression of several AR-abundant MYB genes discovered in the RNA-seq analysis. Three independently grown replicates of AR and BR were utilized in analysis and *ETIF5A* was used as an internal control. Refer to the Section 2.2.3 for the method to conduct qRT-PCR. Data was analyzed using $2^{-\Delta\Delta CT}$ (Livak & Schmittgen, 2001) method. Primer sequences used are displayed in the Appendix 2.

3.3 Results

3.3.1 Identification of MYB transcription factors in flax genome

To identify MYB transcription factors in the flax genome, BLASTP was run locally to query the *Arabidopsis* MYB domain proteins against the 43,384 putative flax proteins (Wang et al., 2012; Stracke et al., 2001; Yanhui et al., 2006). The PROSITE program was then used to check the presence of complete MYB DBDs in each protein (Sigrist et al., 2009). From this process, a total of 240 putative flax MYB transcription factors were identified, including 53 encoding MYB-related protein, 179 encoding 2R-MYB, 7 encoding 3R-MYB and 1 encoding 4R-MYB (Appendix

4). Additionally, nine MYB-domain-containing proteins that also contained non-MYB domains were excluded from this study. The predicted *LusMYBs* were distributed on 137 separate scaffolds and consisted of 191 to 1350 amino acids, with molecular weight of 21.22 to 149.23 kDa. The isoelectric point ranged from 4.37 to 10.76 (Appendix 5). A similar range of MYB protein size was reported in apple (*Malus domestica*; Cao et al., 2013).

3.3.2 Phylogenetic analysis

To classify the predicted MYBs into groups based on similarities in their amino acid sequences, I have constructed a Neighbour-Joining phylogenetic dendrogram using the full-length amino acid sequences of MYB proteins from *Arabidopsis*, flax and *Populus* (Figure 3-1). *Populus trichocarpa* is a related taxon of flax in the Malpighiales order, and *Arabidopsis,* a more distantly related species, was selected because *Arabidopsis* MYBs were well characterized (Katiyar et al., 2012; Yanhui et al., 2006). Since the total number of MYBs from these three species was too big to display in the phylogenetic dendrogram, I chose not to include MYB-related genes in the subsequent analyses. Based on the dendrogram in figure 3-1 and data from *Arabidopsis*, I clustered the MYB family proteins into 18 clades (Dubos et al., 2010a). All 18 clades included representatives from all three species, with the exception of clade 11 which did not include any *Arabidopsis* MYB proteins, indicating that MYBs in this clade may have been obtained in Malpighiales after divergence from the last common ancestor with *Arabidopsis* or they might have been lost from *Arabidopsis* during the evolution (Table 3-1). This pattern suggested that genes in this clade might have a specialized function in Malpighiales. We also noticed that clade 12 was largely expanded in flax and *Populus*. In flax, 160 out of the 187 MYBs appeared as duplicate

pairs in the phylogenetic tree, which is consistent with a recent (5-9 MYA) whole-genome duplication event in flax (Wang et al., 2012).

3.3.3 Meta-analysis of flax MYB gene expression

I collected information on transcript expression of LusMYB genes from existing data sources including EST libraries, microarrays, and RNA-Seq experiments. These data are summarized in the Table 3-2.

3.3.3.1 Identification of LusMYB ESTs in the NCBI

To find transcriptional evidence for the putative flax MYB genes, and explore their expression patterns across tissues, I searched flax ESTs datasets available at NCBI. ESTs were found for 71 out of the 186 flax MYB genes and only a single EST representative was detected for half of them (Table 3-3). The greatest number of ESTs were found for *LusMYB139* (65) and *LusMYB140* (63) and their ESTs were only detected in seeds. High numbers of ESTs were also found for *LusMYB108* (15) and *LusMYB75* (13). The vast majority of the ESTs for these two genes were detected in cotyledon embryo and torpedo stage seed coat. A few (1-2) ESTs of *LusMYB108* and *LusMYB75* were detected in endosperm, fiber-enriched tissue, and mature embryo but not in other EST libraries. Meanwhile we found nine ESTs were found for *LusMYB44*, and eight of these were detected in the seed coat at torpedo stage. An EST of *LusMYB90* was found only in leaves while four MYBs (*LusMYB83*, *LusMYB43*, *LusMYB42* and *LusMYB87*) only had ESTs present in stems. Another subset of MYB genes (*LusMYB101*, *LusMYB182*, *LusMYB184*, *LusMYB174*, and *LusMYB183*) only had ESTs observed in stem peels. *LusMYB47*, *LusMYB48* and *LusMYB131* were

only detected in flowers, whereas *LusMYB33* and *LusMYB81* only had EST detected in fiber-enriched tissue.

3.3.3.2 Expression of LusMYBs in microarray datasets

To further characterize patterns of LusMYB expression, I investigated flax MYB genes in two previously published microarray datasets (GEO accessions GSE21868 and GSE29345). These experiments measured global transcript abundance during embryo and stem development, also compared expression in stems of two fiber-type cultivars (Belinka, Drakkar) that differ in fiber quality and disease resistance. From these data, expression profiles of 22 LusMYB genes were obtained (Figure 3-2; Figure 3-3; Fenart et al., 2010; Huis et al., 2012). As shown in Figure 3-2, *LusMYB56* was enriched in the seeds 10-15 days after flowering and in the stem, it was more abundant in the xylem enriched internal stem tissues (Figure 3-2; Figure 3-3). By contrast, *LusMYB147* was specifically enriched in more mature seeds, 20-30 days and 40-50 days after flowering (Figure 3-2). Within the stem, *LusMYB147* accumulated more transcripts in the internal tissues of the upper stem (Figure 3-3). *LusMYB172* was also highly expressed in the seeds 40-50 days after flowering although its transcripts in the leaves were also abundant (Figure 3-2). Within the stem, *LusMYB172* expression was obviously much higher in the phloem-enriched external tissues of the whole stem, upper stem, middle stem or lower stem (Figure 3-3). Three *LusMYBs* (*LusMYB45*, *LusMYB174* and *LusMYB76*) showed particularly high expression levels in the inner tissues of the flax stem at both the vegetative stage and green capsule stage (Figure 3-2). On the other hand, the remaining 15 flax *MYBs* did not seem to be enriched in any one tested tissue (Figure 3-2; Figure 3-3). However, within the stem, *LusMYB182* appeared to have high expression levels

in outer tissues of the upper and middle part of flax stem and internal tissues of lower stem (Figure 3-3).

Meanwhile, as demonstrated in the Figure 3-2, several flax MYB genes showed differential expression levels in Drakkar and Belinka, such as three obviously more abundant in Drakkar (*LusMYB36*, *LusMYB45* and *LusMYB181*) and two more enriched in Belinka (*LusMYB161*, *LusMYB90*).

I have also investigated the expression of LusMYB genes in an unpublished microarray study which explored gene expression patterns in five stages of flax stem development. Probes used in this study were designed based on a draft of flax genome and after alignment, 326 of them were aligned to 163 *LusMYBs*. I retrieved the expression data for these 326 probes and searched these data for those with differential expression in at least one of the five distinct stem segments. As a result, only seven probes corresponding to LusMYB gene showed differential expression in at least one of the five different segments representing five different developing stages of flax stem. Three out of these seven LusMYB genes (*LusMYB127*, *LusMYB129* and *LusMYB113*) showed similar expression patterns, with expression peaks in the stem segment collecting from just above the snap point and further down the stem (Figure 3-4; Table 3-4a,b). The snap point was a defined transition region on flax stem. Flax phloem fiber in the stem below this region started to deposit thick secondary cell wall (Gorshkova et al., 2003). Likewise, *LusMYB148* was also enriched in the stem just above the snap point but its expression was lowest in the stem just below the snap point. *LusMYB118* was enriched in the shoot apex, while *LusMYB33* showed peak expression in a more

mature stem tissue (4 to 5 cm below the shoot apex). Moreover, *LusMYB51* was enriched in the most mature tissue analyzed in this study, and phloem fibers in which have already possessed thick secondary cell wall.

3.3.3.3 RNA-seq

Analysis of a previously published RNA-seq data (Kumar et al., 2013) indicated that most MYBs showed very low transcript abundance while a few MYBs specifically accumulated very high transcript abundance in globular and heart embryo (*LusMYB140*, *LusMYB139* and *LusMYB54*), anther (*LusMYB9*, *LusMYB145*, *LusMYB156*, *LusMYB131*, *LusMYB130*, *LusMYB129* and *LusMYB165*), root (*LusMYB10*) and leaf (*LusMYB111* and *LusMYB81*; Figure 3-5a; 3-5b; 3-5c; Appendix 6).

Expression of the 187 putative LusMYB genes were examined in the RNA-seq data described in Chapter 2 of this thesis. Among these 187 *LusMYBs*, 18 were significantly ($q<0.05$) enriched in the AR compared to the BR and 12 of them were above 2-fold more abundant in AR. In addition, three flax MYBs were only detected in the AR but not in the BR (Table 3-5). By contrast, 33 *LusMYBs* were significantly ($q<0.05$) more highly expressed in the BR compared to the AR. Among them, 21 were above two times more abundant in the BR and 11 were not detected in the AR (Table 3-6).

3.3.3.4 Verification of LusMYB gene expression in the AR and the BR by qRT-PCR

Transcript abundance of eight MYB genes that showed at least two times more abundance in the AR of the RNA-seq analysis were confirmed by qRT-PCR. As shown in Figure 3-6, quantitative real time –PCR further revealed that all these eight *MYBs* were enriched in the AR compared to the BR. Attempt to measure the abundance of *LusMYB36* by qRT-PCR failed due to the lack of a specific primer.

3.4 Discussion

MYB transcription factors are broadly represented in eukaryotes and they have large numbers and diverse functions in plants. In this study, I have identified 240 MYB domain proteins from flax including 53 MYB-related proteins, 179 R2R3-MYB, seven R1R2R3-MYB and one 4R-MYB. As observed in other plants, flax has many more R2R3-MYBs than other MYB types (Wong et al., 2016; Wang et al., 2015; Wilkins et al., 2009; Zhai et al., 2016). The number of R2R3-MYB in flax (179) is expanded compared to *Arabidopsis* (126) and is close to *Populous trichoparpa* (192), however, flax (74.58%) has higher proportion of R2R3-MYB than *Arabidopsis* (55.02%) and *Populus trichoparpa* (46.83%; Stracke et al., 2001; Dubos et al., 2010). The proportion of R2R3-MYB genes in flax appeared to be higher than all the other plants in which MYB have been genomic-widely characterized except Asian pear (Appendix 7). The *LusMYBs* with two, three or four repeats were characterized in this study. These LusMYBs were revealed to consist of 191 to 1350 amino acids, with molecular weight of 21.22 to 149.23 kDa. The isoelectric point ranged from 4.37 to 10.76. These ranges are comparable to the findings in other plant species (Katiyar et al., 2012; Yanhui et al., 2006; Cao et al., 2013; He et al., 2016).

Expression of flax MYB transcripts were investigated through a scrutiny of publically available ESTs, microarray and RNA-Seq database. Through these analysis, I have found experimental evidence for transcriptions of all the putative 187 *LusMYBs* and the vast majority of *LusMYBs* were expressed at a very low transcriptional levels. This was consistent with the generally low transcript abundance of transcription factors.

3.4.1 MYBs and flax seed development

Some flax MYBs might have a role in seed development. *LusMYB139* and *LusMYB140* had the highest number of ESTs detected among all the *LusMYBs*, and their ESTs were observed in seeds only. We found most of their ESTs were derived from the globular embryo, heart embryo or torpedo embryo (Table 3-3). RNA-seq analysis confirmed that these two genes were preferentially transcribed in embryos at globular and heart stage. In the phylogenetic tree, *LusMYB139* and *LusMYB140* clustered together as duplicated genes in Clade 8 (Figure 3-1). Their orthologue in *Arabidopsis* (*AtMYB103*) is a transcriptional regulator of anther development, cell wall thickening in xylem tissues and the syringyl lignin biosynthesis (Zhu et al., 2010; Zhong et al., 2008). Based on the expression profiles of *LusMYB139* and *LusMYB140*, we assumed that these two MYBs might play some roles in the early stages of flax seed development. Additionally, both microarray and RNA-seq analysis indicated that *LusMYB147* and *LusMYB172* were exclusively enriched in mature seeds, an indicative of their roles in late stage of seed development (Figure 3-2; Appendix 6). In addition, *LusMYB56* was suggested to be specifically enriched in the young seeds (10-15

days after flowering; Figure 3-2). Therefore, I hypothesized that this gene might be also associated with seed development.

3.4.2 MYBs and flax xylem differentiation

Expression profiles of three MYBs (*LusMYB174*, *LusMYB45* and *LusMYB76*) indicated that they might play roles in xylem differentiation. Among all the tested EST libraries, *LusMYB174* had only a single EST observed and it was derived from the stem peels and RNA-seq analysis again suggested that it was enriched in flax stem (Table 3-3; Figure 3-5c). Besides, analysis in both microarray datasets GSE21868 and GSE29345 showed that *LusMYB174* was preferentially transcribed in the inner part of flax stem (Figure 3-2; Figure 3-3). Some *Arabidopsis* MYBs in the same clade as *LusMYB174* were known to be involved in lignin, xylan and cellulose biosynthesis (Lee et al., 2009; Zhong et al., 2008). Beyond that, microarray and RNA-seq analysis showed that transcripts of *LusMYB45* and *LusMYB76* were particularly accumulated in the inner tissues of the flax stem at both the vegetative stage and green capsule stage (Figure 3-2, 3-3; Appendix 6). *Arabidopsis* orthologs of *LusMYB76* (*AtMYB46* and *AtMYB83*) were reported to regulate the lignin and secondary cell wall biosynthesis (Zhong et al., 2007a; McCarthy et al., 2009; Sakamoto & Mitsuda, 2015) and orthologs of *LusMYB45* (*AtMYB43* and *AtMYB20*) were also involved in lignin biosynthesis (Zhao & Dixon, 2011). I assumed that these flax MYBs might regulate the transcription of cell wall-related genes during flax stem xylem formation.

3.4.3 MYBs might be involved in flower development.

ESTs of three MYBs (*LusMYB47*, *LusMYB48* and *LusMYB131*) were only detected in flowers but not the other EST libraries and the phylogenetic analysis placed all three genes in Clade 2 (Table

3-3; Figure 3-1). This result motivated me to check the expression of all flax members in this clade. RNA-seq analysis suggested that eight out of the 18 genes in the clade 2 (*LusMYB47*, *LusMYB48*, *LusMYB131*, *LusMYB49*, *LusMYB130*, *LusMYB129*, *LusMYB96*, *LusMYB95*) were specifically enriched in the anther (Figure 3-5a, 3-5b, 3-5c). Many *Arabidopsis* genes in this clade (*AtMYB21*, *AtMYB24* and *AtMYB57*, *AtMYB81*, *AtMYB33*, *AtMYB65*, *AtMYB120*, *AtMYB97* and *AtMYB101*) were reported to be involved in anther/pollen development (Cheng et al., 2009). Meanwhile, *AtMYB78* and *AtMYB108*, two Clade 2 members and their cotton and tomato orthologs were revealed to play important roles in plant pathogen defense (Mandaokar & Browse, 2008; Mengiste et al., 2003; Cheng et al., 2016; Abuqamar et al., 2009). *MYB108* was also reported to be involved in the jasmonate-mediated stamen and pollen maturation in *Arabidopsis* (Mandaokar & Browse, 2008). *LusMYB146* and *LusMYB145*, flax orthologs of *AtMYB78* and *AtMYB108*, were found to be significantly induced by *Fusarium oxysporum f. sp. lini* and they both showed a preferential expression in anther (Galindo-González & Deyholos, 2016). I predict that these flax *MYBs* possess roles in anther development and biotic stress.

3.4.4 Some MYBs were selected as candidates of fiber cell identity determination regulator.

An *Arabidopsis* MYB gene has been reported to regulate vascular cell specification (Bonke et al., 2003). 18 LusMYBs were found to be significantly ($q<0.05$) enriched in the AR compared to the BR by RNA-seq and some of these 18 genes were potential to act as transcriptional regulators of flax fiber specification. I have summarized functions of their *Arabidopsis* orthologs in Appendix 8. Six AR-enriched MYBs belonging to the 3R-MYB type have a conserved role in cell-cycle regulation (Appendix 8). Their abundance in the AR should not be linked to phloem fiber

specification since intense cell division and mitosis occur in shoot apex. *LusMYB26*, a member of clade 11, was not detected in stem and therefore more-enriched in the AR than BR (Figure 3-1; Table 3-5). I found Clade 11 appeared to be Malpighiales-specific since it contained 11 flax MYBs, 9 poplar MYBs and no Arabidopsis MYBs (Figure 3-1; Table 3-1). However, although the RNA-seq data published by Kumar et al. showed that *LusMYB26* was not expressed in stem, *LusMYB26* accumulated its highest expression in root (Appendix 6). There is no other data available for us to make inference about functions of this clade. Out of these 18 genes, *LusMYB36* and *LusMYB181* were suggested to be preferentially accumulated in two contrasting varieties, Drakkar and Belinka (Drakkar produces better quality fibers than Belinka; Figure 3-2). I have attempted to check cellular localization of these MYB candidates by in situ hybridization but I failed to obtain specific signals.

3.5 Conclusions

A total of 240 putative MYB genes were identified from flax genome. They were clustered into 18 distinct groups. Flax had a higher proportion of R2R3-MYB than most of other sequenced plant species. Through analysis of the expression data in public database, this study had found experimental evidence for transcriptions of all the putative 187 flax MYBs. The majority of *LusMYBs* were expressed in wide range of tissues with low expression level while a few others were particularly abundant in some specific tissues. The large size of MYB family in flax suggests that they have diversified functions, however, to further examine their biological function in flax development, analyses with knock out mutants will be necessary.

3.6 Figures and tables

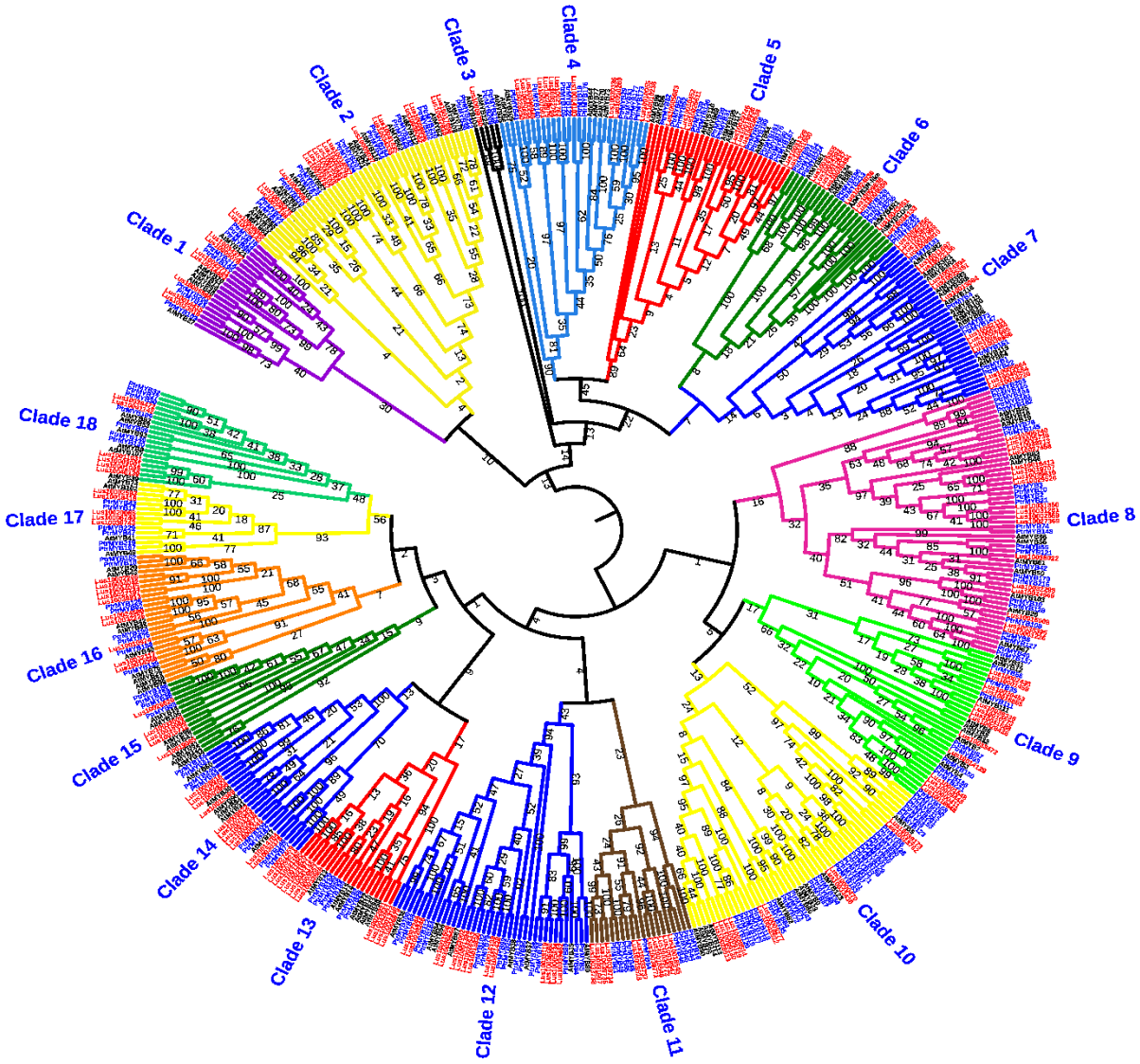


Figure 3-1 Dendrogram of MYBs. Full length of MYB protein sequences from *Arabidopsis thaliana* (AtMYBs), *Populus trichocarpa* (PtMYBs) and *Linum usitatissimum* (LusMYBs) were used in the analysis. MUSCLE was used to conduct multiple sequence alignment and the dendrogram was constructed using Neighbor-Joining algorithm by MEGA 5 (Edgar, 2004; Tamura et al., 2011). Bootstrap test was applied and replicated 1,000 times. The leaf labels of LusMYBs, AtMYBs and PtrMYBs were denoted in red, black and blue respectively.

Table 3-1 Membership details of each LusMYB subgroup. *Lus*: *Linum usitatissimum*; *At*: *Arabidopsis thaliana*; *Ptr*: *Populus trichocarpa;*

Clade	*Lus*	*At*	*Ptr*
1	7	6	8
2	18	15	15
3	1	2	2
4	12	7	10
5	8	8	13
6	9	6	10
7	16	12	7
8	18	12	25
9	14	9	10
10	10	10	31
11	11	0	9
12	16	8	15
13	8	7	6
14	13	6	8
15	5	7	4
16	10	9	8
17	5	4	6
18	6	6	7

Table 3-2 Data sources of the *LusMYBs* expression profiles demonstrated in this study.

Data Type	**Description**	**Reference**
EST	286,856 Sanger sequenced ESTs isolated from: embryos at five stages of development (globular, heart, torpedo, cotyledon and mature stages); seed coats at globular and torpedo stages; endosperm (pooled globular to torpedo stages); flower; leaf; etiolated seedlings; three stem tissues including the outer fiber-bearing tissues of flax stems harvested at the mid-flowering stage; stem and stem peel (consisting of epidermis, cortical tissue, phloem, developing fiber and cambial tissues) harvested from four-weeks-old flax;	(Venglat et al., 2011; Day et al., 2005)
Microarray	GSE21868; oligonucleotide probes hybridized to RNA from: roots (R), leaf sample at green capsule stage (L), stem inner tissue at vegetative stage (SIV), stem outter tissue at vegetative stage (SOV), stem inner tissue at green capsule stage (SIGC), stem outter tissue at green capsule stage (SOGC), embryos of 10, 20 and 40 days after flowering (designated as E1, E2 and E3 respectively); stems of Belinka and Drakkar (two fiber-type cultivars that differ in fiber quality and disease resistance);	(Fenart et al., 2010)
	GSE29345; oligonucleotide probes hybridized to RNA from: internal tissues of the whole stem (WSI), upper stem (USI),	

Microarray	middle stem (MSI), or lower stem (LSI); external tissues of the whole stem (WSE), upper stem (USE), middle stem (MSE), and lower stem (LSE);	(Huis et al., 2012)
Microarray	oligonucleotide probes hybridized to RNA from: the shoot apex (T1); 1cm stem segment above the snap-point (T2); 1cm stem segment at the snap point (T3); 1cm stem segment below the snap point (T4); and the 1 cm stem segment from bottom of flax stem, in which phloem fiber cells have deposited thick secondary cell wall (T5);	Unpublished
RNA-Seq	Compare transcript expression patterns in two segments of the vegetative stem of 14d flax plants, from which all visible leaves had been removed. The segments were: (i) the apical region (AR) of the shoot apex, which contained the apical-most 0.5 mm of the stem, including the SAM and its immediate derivatives; and (ii) the basal region (BR), which contained the entire stem except for the apical-most 1 cm, and therefore represented all stem and vascular tissues at later stages of differentiation as compared to the AR;	Chapter 2 of this thesis
RNA-Seq	investigate transcript abundances in embryos at five stages of development (globular, heart, torpedo, cotyledon and mature	

	stages); seed; anther; ovary; mature flower; root; stem; leave; etiolated seedlings;	(Kumar et al., 2013)

Table 3-3 EST profiles of LusMYB genes. GE: globular embryo; HE: heart embryo; TE: torpedo embryo; CE: cotyledon embryo; ME: mature embryo; EN: endosperm; GC: globular seed coat; TC: torpedo seed coat; ES: etiolated seedling; LE: leaf; ST: stem; PS: stem peel; FL: flower; F: fiber enriched tissue at mid-flowering stage (Venglat et al., 2011; Day et al., 2005)

Gene Name	GE	HE	TE	CE	ME	EN	GC	TC	ES	LE	ST	PS	FL	F	Total
LusMYB139	15	18	15		3	3	8	3							65
LusMYB140	16	16	16		2	2	8	3							63
LusMYB108				6	1	1		6						1	15
LusMYB75				5		1		5						2	13
LusMYB175	1	2	3			2						3			11
LusMYB44							1	8							9
LusMYB125							1	1	1	2		1		1	7
LusMYB10							1	1				2	2		6
LusMYB9							1					2	3		6
LusMYB172			4								1				5
LusMYB95			1			2	1	1							5
LusMYB163			1			1		2							4
LusMYB37	1	2	1												4
LusMYB36	1	2	1												4
LusMYB105				3	1										4
LusMYB187			3			1									4
LusMYB171			4												4
LusMYB83											3				3
LusMYB110					2							1			3
LusMYB147				1	1									1	3
LusMYB111					1							1		1	3
LusMYB148				1	1									1	3
LusMYB128								1	1			1			3
LusMYB47													2		2
LusMYB45							1	1							2
LusMYB181			2												2
LusMYB158						2									2
LusMYB5							2								2
LusMYB25											1	1			2
LusMYB101												2			2
LusMYB4												1	1		2
LusMYB179			1			1									2
LusMYB88								1			1				2

LusMYB81														2	2
LusMYB186		1													1
LusMYB141		1													1
LusMYB97		1													1
LusMYB112			1												1
LusMYB89						1									1
LusMYB35								1							1
LusMYB90										1					1
LusMYB48													1		1
LusMYB55		1													1
LusMYB54	1														1
LusMYB43											1				1
LusMYB42											1				1
LusMYB182												1			1
LusMYB31				1											1
LusMYB8					1										1
LusMYB33														1	1
LusMYB56		1													1
LusMYB21		1													1
LusMYB184												1			1
LusMYB87											1				1
LusMYB162			1												1
LusMYB7					1										1
LusMYB100								1							1
LusMYB174												1			1
LusMYB138			1												1
LusMYB64								1							1
LusMYB131													1		1
LusMYB183												1			1
LusMYB161			1												1
LusMYB34								1							1
LusMYB126		1													1
LusMYB94		1													1
LusMYB166	1														1
LusMYB106						1									1
LusMYB29				1											1
LusMYB119					1										1

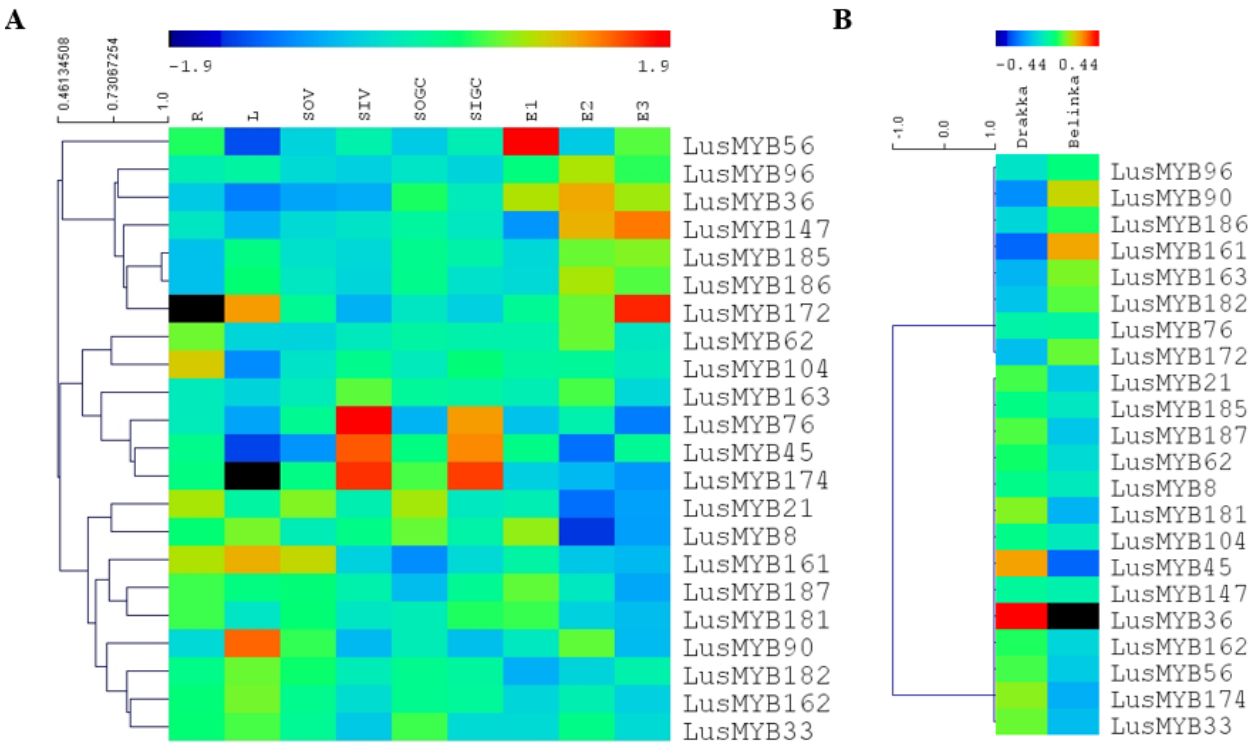


Figure 3-2 Expression profiles of flax *MYBs* in previously published microarray dataset GSE21868 (Aug et al., 2015). Red indicated high abundance while blue indicated low abundance. A:Tissues analyzed were root (R), leaf (L), outer stem tissues at the vegetative stage (SOV), outer stem tissues at the green capsule stage (SOGC), inner stem tissues at the vegetative stage (SIV) , inner stem tissues at the green capsule stage (SIGC), seeds at 10-15 days after flowering (E1), 20-30 days after flowering (E2) and 40-50 days after flowering (E3; Aug et al., 2015). B: Expressions of *LusMYBs* were compared in two contrasting flax cultivars, Drakkar and Bellinka respectively. The heat map was generated using RMA-normalized, average log2 signal values by MEV (Multi Experiment Viewer. http://www.tm4.org/mev). Genes were hierarchically clustered based on the expression pattern using Pearson Correlation distance matrix and the single clustering method. The log2 signal values has been mean-centered before clustering. This involves taking the mean expression value for each gene or transcript, and subtracting it from each expression value for that gene or transcript. The mean value will then be zero.

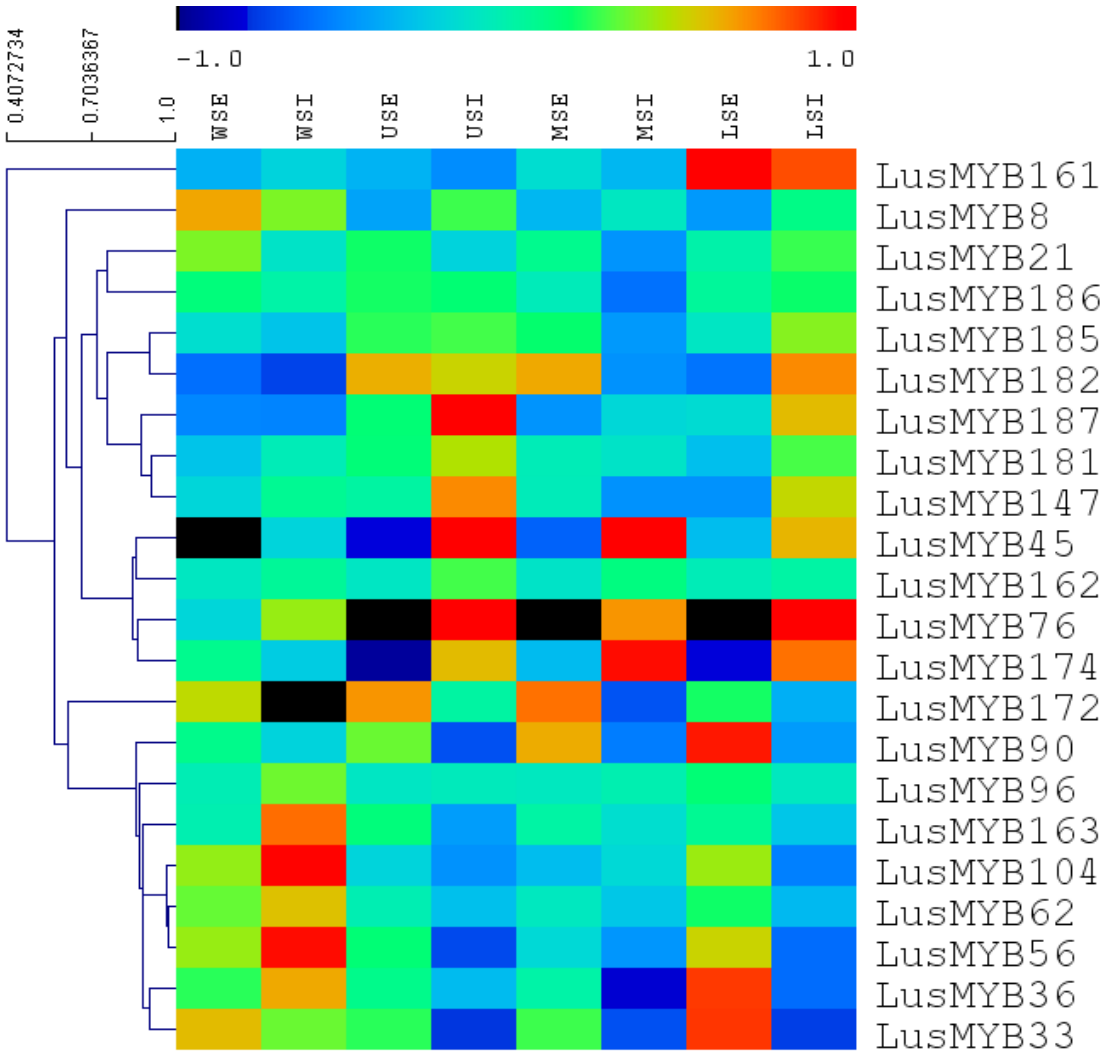


Figure 3-3 Transcript abundance of *LusMYBs* in previously published microarray dataset GSE29345 (Huis et al., 2012). Red indicated high abundance while blue indicated low abundance. WSE: external (i.e. phloem and cortex enriched) tissues of the whole stem; WSI: internal tissues of the whole stem; USE: external tissues of the upper stem; USI: internal tissues of the upper stem; MSE: external tissues of the middle stem; MSI: internal tissues of the middle stem; LSE: external tissues of the lower stem; LSI: internal tissues of the lower stem; The heat map was generated using RMA-normalized, average log2 signal values by MEV (Multi Experiment Viewer. http://www.tm4.org/mev). Genes were hierarchically clustered based on the expression pattern using Pearson Correlation distance matrix and the single clustering method. The log2 signal values has been mean-centered before clustering. This involves taking the mean expression value for each gene or transcript, and subtracting it from each expression value for that gene or transcript. The mean value will then be zero.

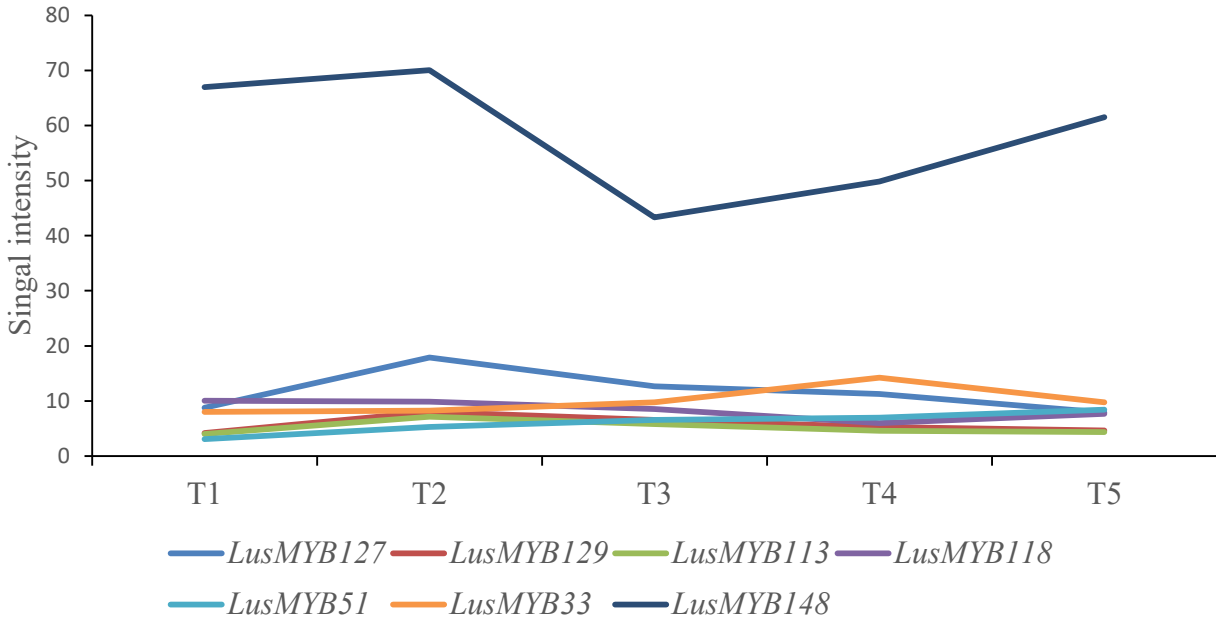


Figure 3-4 *LusMYBs* showed differential expression in at least one of the five different segments examined in flax stem microarray. Data were obtained from (To, 2013).

Table 3-4a Signal intensities of the seven *LusMYBs* showed differential expression in at least one of the five different segments. Data were obtained from (To, 2013)

Gene name	T1	T2	T3	T4	T5
LusMYB127	8.73597	17.89426	12.67487	11.28712	7.876578
LusMYB129	4.188584	8.019055	6.556034	5.282254	4.613198
LusMYB113	3.995833	7.124993	5.822584	4.558507	4.332315
LusMYB118	10.05013	9.854898	8.53373	5.958522	7.642319
LusMYB51	3.096649	5.290823	6.573896	6.985857	8.417082
LusMYB33	8.012695	8.23443	9.730631	14.24072	9.745688
LusMYB148	66.9445	70.03235	43.30971	49.85328	61.50595

Table 3-4b Statistical details for the seven genes differentially expressed in one of the five studied flax tissues in flax stem microarray study. Data were obtained from (To, 2013). A two-way ANOVA test was conducted followed by a Tukey's multiple comparisons test using GraphPad Prism 7.00. * denotes *p*-value between 0.01-0.05; **denotes *p*-value between 0.001-0.01; ***denotes *p*-value between 0.0001-0.001; ****denotes *p*-value <0.0001; ns (not significant) denotes *p*-value >0.05;

	LusMYB127	*LusMYB129*	*LusMYB113*	*LusMYB118*	*LusMYB51*	*LusMYB33*	*LusMYB148*
	g10550.t1|sl-954-989	g1704.t1|sl-1608-1643	g18848.t1|sl-783-818	g34592.t1|sl-1012-1047	g47442.t1|sl-90-125	g21765.t1|sl-745-781	g32699.t1|sl-642-681
T1 vs. T2	****	***	**	ns	ns	ns	*
T1 vs. T3	***	ns	ns	ns	**	ns	****
T1 vs. T4	ns	ns	ns	***	***	****	****
T1 vs. T5	ns	ns	ns	ns	****	ns	****
T2 vs. T3	****	ns	ns	ns	ns	ns	****
T2 vs. T4	****	*	ns	***	ns	****	****
T2 vs. T5	****	**	*	ns	**	ns	****
T3 vs. T4	ns	ns	ns	ns	ns	****	****
T3 vs. T5	****	ns	ns	ns	ns	ns	****
T4 vs. T5	**	ns	ns	ns	ns	****	****

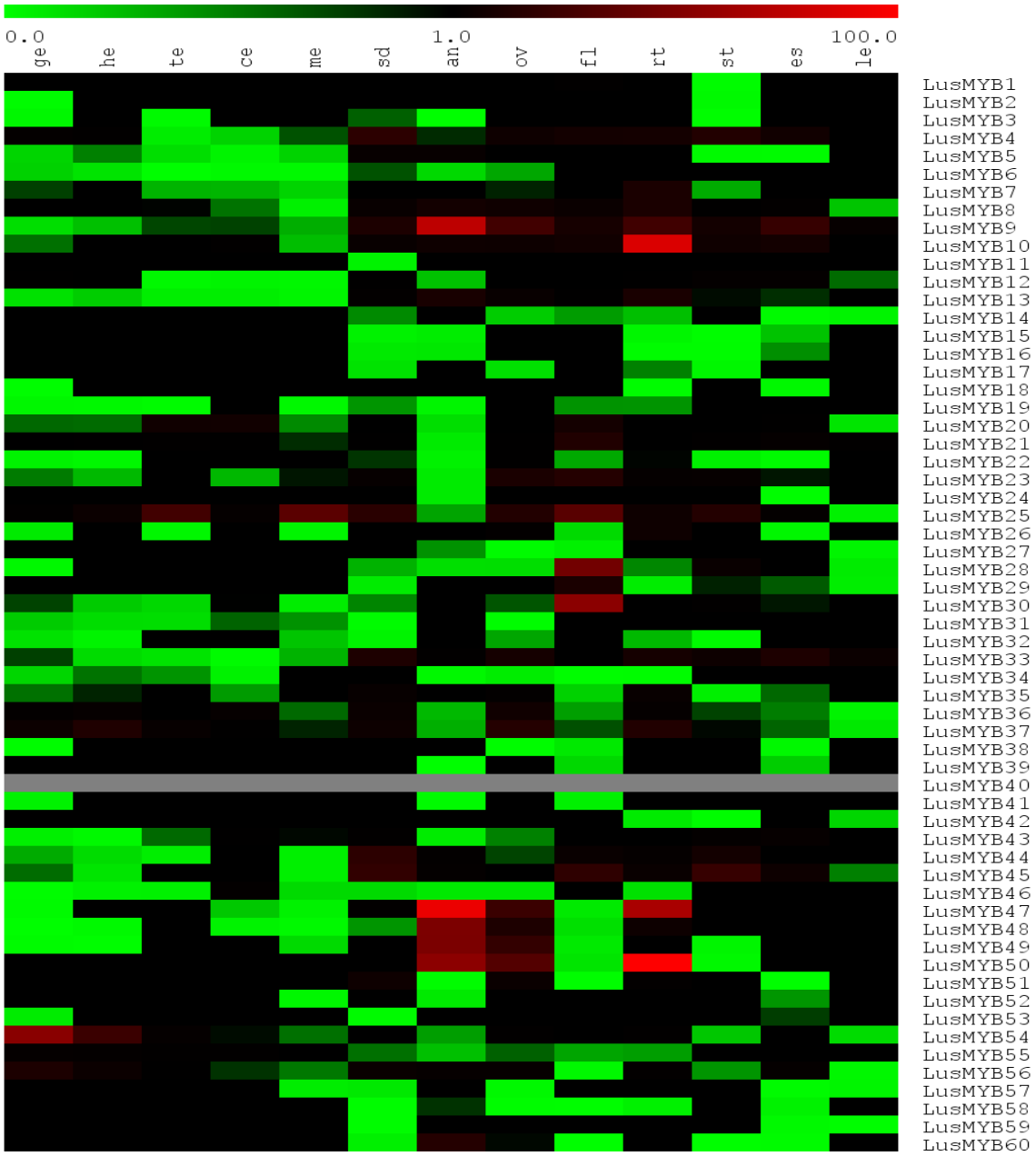


Figure 3-5a Expression profiles of *LusMYB1-60* in 13 different tissues (Kumar et al., 2013). Tissues examined including globular embryo (ge), heart embryo (he), torpedo embryo (te), cotyledon embryo (ce), mature embryo (me), seeds (sd), anthers (an), ovaries (ov), mature flower (fl), root (rt), stem (st), etiolated seedlings (es), leaves (le). Red indicated high abundance while green indicated low abundance. Genes with no expression in all the tested tissues were shown as grey.

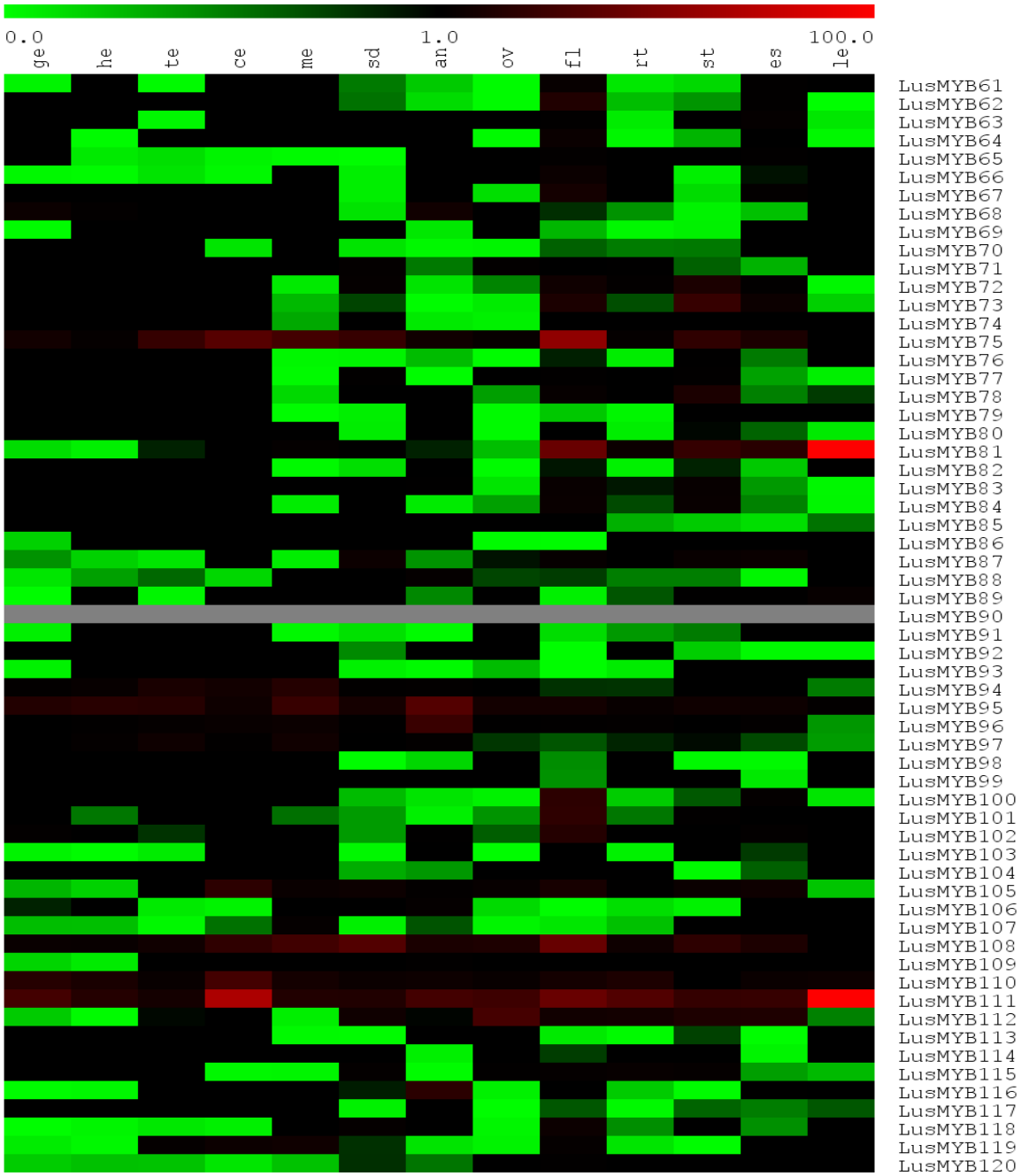


Figure 3-5b Expression profiles of *LusMYB61-120* in 13 different tissues (Kumar et al., 2013). Tissues examined including globular embryo (ge), heart embryo (he), torpedo embryo (te), cotyledon embryo (ce), mature embryo (me), seeds (sd), anthers (an), ovaries (ov), mature flower (fl), root (rt), stem (st), etiolated seedlings (es), leaves (le). Red indicated high abundance while green indicated low abundance.Genes with no expression in all the tested tissues were shown as grey.

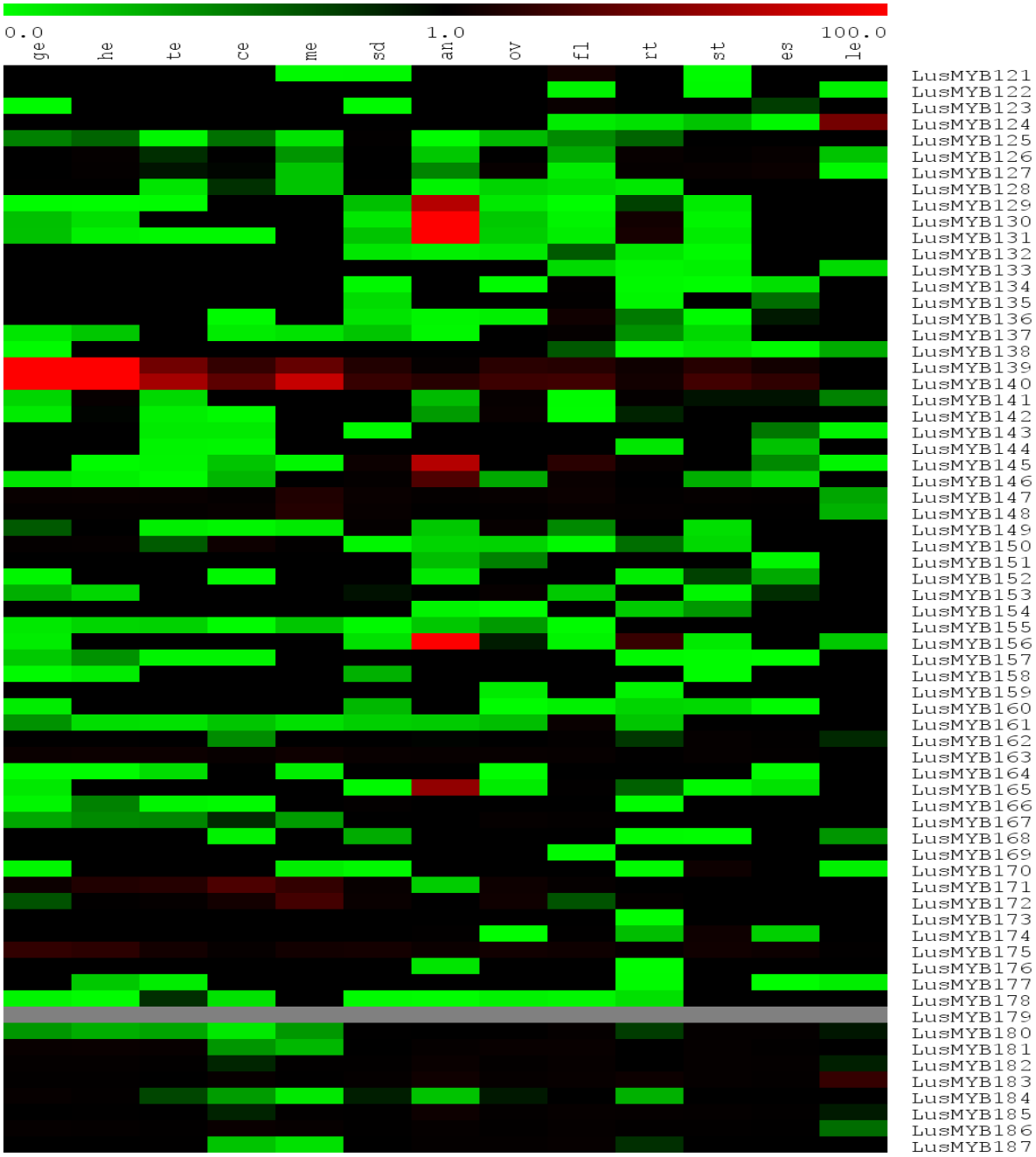


Figure 3-5c Expression profiles of *LusMYB121-187* in 13 different tissues (Kumar et al., 2013). Tissues examined including globular embryo (ge), heart embryo (he), torpedo embryo (te), cotyledon embryo (ce), mature embryo (me), seeds (sd), anthers (an), ovaries (ov), mature flower (fl), root (rt), stem (st), etiolated seedlings (es), leaves (le). Red indicated high abundance while green indicated low abundance.Genes with no expression in all the tested tissues were shown as grey.

Table 3-5 18 flax MYB genes were significantly more abundant in the AR compared to the BR. Data were obtained from the Chapter 2 of this thesis. NA: no data.

Gene Name	FPKM (AR)	FPKM (BR)	log2(fold_change AR/BR)	q_value
LusMYB34	7.65	0.42	4.2	0.03
LusMYB36	22.33	1.63	3.78	0
LusMYB149	4.7	0.49	3.25	0.018
LusMYB141	45.83	6.52	2.81	0
LusMYB35	8.99	1.46	2.62	0.002
LusMYB142	39.44	6.72	2.55	0
LusMYB187	21.91	6.6	1.73	0
LusMYB181	18.83	5.85	1.69	0
LusMYB102	9.24	3.45	1.42	0.013
LusMYB172	49.79	19.73	1.34	0
LusMYB171	40.27	17.91	1.17	0.024
LusMYB175	17.16	8.01	1.1	0.003
LusMYB179	19.39	11.25	0.79	0.016
LusMYB180	7.84	4.68	0.75	0.013
LusMYB162	7.64	4.96	0.62	0.038
LusMYB61	10.63	0	NA	0
LusMYB26	2.88	0	NA	0
LusMYB66	3.38	0	NA	0

Table 3-6 33 putative flax *MYBs* were significantly enriched in the BR compared to the AR. Data were obtained from the Chapter 2 of this thesis. NA: no data.

Gene Name	FPKM (AR)	FPKM (BR)	log2(fold_change AR/BR)	q_value
LusMYB87	0.46	28.97	-5.98	0.002
LusMYB114	0.32	10.69	-5.07	0.048
LusMYB10	1.19	32.7	-4.78	0
LusMYB44	0.96	21.01	-4.45	0.005
LusMYB81	4.22	58.67	-3.8	0
LusMYB9	3.66	32.44	-3.15	0
LusMYB4	7.09	51.76	-2.87	0
LusMYB28	1.08	7.74	-2.85	0.007
LusMYB107	0.9	5.94	-2.72	0.001
LusMYB125	1.56	8.03	-2.37	0.001
LusMYB43	3.54	16.9	-2.26	0
LusMYB128	1.53	7.29	-2.26	0.001
LusMYB117	0.51	2.31	-2.17	0.03
LusMYB46	0.62	2.65	-2.1	0.04
LusMYB127	4.67	16.36	-1.81	0
LusMYB108	16.01	51.98	-1.7	0
LusMYB75	12.31	39.15	-1.67	0
LusMYB101	3.16	9.88	-1.64	0.003
LusMYB126	2.99	9.26	-1.63	0.003
LusMYB120	6.39	18.31	-1.52	0
LusMYB12	6.39	16.86	-1.4	0.005
LusMYB140	27.34	42.56	-0.64	0.025
LusMYB174	0	17.45	NA	0
LusMYB78	0	14.28	NA	0
LusMYB170	0	13.25	NA	0
LusMYB84	0	11.59	NA	0

LusMYB82	0	7.62	NA	0
LusMYB113	0	7.58	NA	0
LusMYB76	0	5.1	NA	0
LusMYB80	0	3.98	NA	0
LusMYB29	0	3.34	NA	0
LusMYB154	0	3.3	NA	0
LusMYB79	0	2.66	NA	0

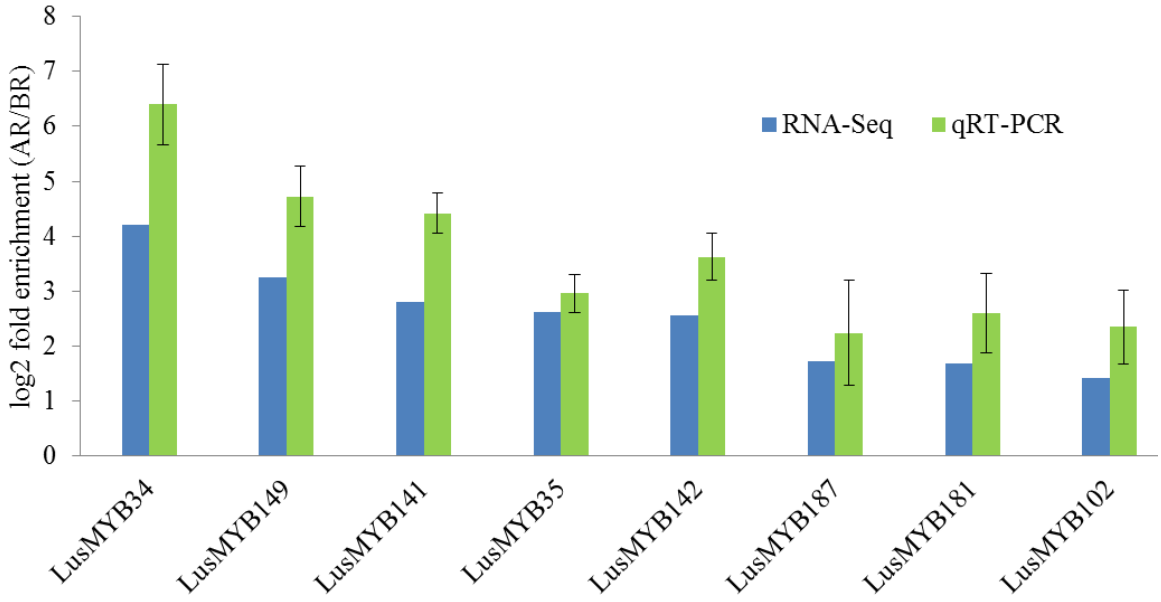


Figure 3-6 Ratio of transcript abundance of eight *LusMYBs* in the AR compared to the BR, as measured by qRT-PCR and RNA-seq on independently grown tissues.

Chapter 4. Genome-wide characterization of the NAC transcription factor family in flax

4.1 Introduction

The NAC (NAM, ATAF1/2 and CUC2) domain gene family is a group of plant-specific transcription factors with a conserved NAM domain in the N-terminus (Ernst et al., 2004; Olsen, et al., 2005). It is one of the largest transcription factor families in the plant kingdom, containing 105 genes in *Arabidopsis thaliana*, 140 genes in rice (*Oryza sativa*) and 163 genes in poplar (*Populus trichocarpa*; Ooka et al., 2003; Jensen et al., 2010; Hu et al., 2010). NAC proteins are important for various aspects of plant growth and development, including: plant shoot apical meristem development (Takada et al., 2001; Hibara et al., 2006), floral organ formation (Sablowski & Meyerowitz, 1998), lateral root development (Xie et al., 2000; He et al., 2005), seed development (Sperotto et al., 2009), leaf senescence (Guo & Gan, 2006), embryo development (Duval et al., 2002), cell cycle control (Kim et al., 2006), nutrient remobilization (Uauy et al., 2006), shoot branching determination (Mao et al., 2007), hormone signaling (Xie et al., 2000) and response to various abiotic stresses (Puranik et al., 2012) and biotic stress (Wang et al., 2009).

Several NAC proteins in the VNS (VND-, NST /SND- *and* SMB-related proteins) subfamily have been found to regulate differentiation of xylem vessels and fiber cells in *Arabidopsis* and some other plant species (Kubo et al., 2005; Ohtani et al., 2011; Hussey et al., 2011). For example, *VASCULAR-RELATED NAC-DOMAIN6* (*VND*) genes *VND6* and *VND7* genes regulate

metaxylem and protoxylem vessel formation, respectively, in the *Arabidopsis* primary root (Kubo et al., 2005; Yamaguchi et al., 2008).

Based on the above information, we assumed that some NACs may be involved in phloem fiber cell identity determination in flax. Additionally, although large amounts of information have been uncovered about NAC domain proteins, most of this research studied NACs in model plants such as *Arabidopsis*, rice and poplar (Olsen et al., 2005; Jensen et al., 2010; Yamaguchi et al., 2010; Zhong et al., 2010; Ohtani et al., 2011; Nakashima et al., 2007). In contrast, very limited knowledge has been obtained about NAC proteins in flax. To facilitate future studies of NAC-mediated gene regulation in flax, and possible crop improvement through manipulating flax fiber differentiation, I sought to perform genome-wide identification of flax NAC domain genes and characterize this family through analysis of its phylogeny and expression profiles.

4.2 Materials and methods

4.2.1 Sequences identification

To identify the NAC proteins in flax, I used *Arabidopsis* NAC protein sequences retrieved from TAIR (https://www.arabidopsis.org/) as queries in BLAST alignments against the 43,384 predicted flax proteins available from Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). BLAST package 2.3.0+ was used and only sequences with e-values less than 10^{-10} were saved for further analysis. The redundant sequences were then manually removed. To further confirm that these sequences represented NAC proteins, all the putative sequences were then analyzed by the Hmmsearch program in HMMER3 and Pfam program (http://pfam.xfam.org/) to validate the presence of a NAM domain (Pfam02365) at the N-terminus of amino acid sequences (Finn et al., 2015a). The amino acid length, molecular weight and isoelectric point of putative flax NAC

proteins were calculated using Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/protein_iep.html; Stothard, 2000).

4.2.2 Phylogenetic analysis

The full-length sequences of flax, *Arabidopsis,* and poplar NAC proteins were aligned by MAFFT 7.0 and IQ-TREE was applied to construct a phylogenetic tree using the maximum likelihood method (Katoh & Standley, 2013; Nguyen et al., 2015). The best-fit substitution model was automatically chosen by IQ-TREE by Bayesian (BIC). The branch supports were assessed by bootstraping with 1000 replicates (Minh et al., 2013). The tree was rooted at the midpoint.

4.2.3 Tissue-specific expression analysis

4.2.3.1 EST

EST libraries in NCBI were queried by BLASTn to find evidence for the transcription of putative flax *NACs* (accessed May 2017; 286,856 sequences). Only ESTs showing identity >99% to the coding sequences were accepted.

4.2.3.2 Microarray

Microarray datasets with accession numbers GSE21868 and GSE29345 were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). GSE21868 examined expression in leaf (L), roots (R), stem inner tissue (xylem enriched) at vegetative stage (SIV) or green capsule stage (SIGC), stem outer tissue (phloem fibers and cortex enriched) at vegetative stage (SOV) or green capsule stage (SOGC), as well as seeds at young (10-15 days after flowering, designated as E1), middle (20-30 after flowering, designated as E2) or mature stage (40-50 days after flowering, designated as E3; Fenart et al., 2010). This project also compared gene expression

between two contrasting flax cultivars, Drakkar and Belinka. Drakkar produces better fibers and has higher resistance to *Fusarium* (a fungal pathogen) than Belinka (Fenart et al., 2010).

GSE29345 compared gene expression in external and internal tissues of the whole flax stem (abbreviated as WSE and WSI respectively), of the upper stem (abbreviated as USE and USI respectively), middle stem (abbreviated as MSE and MSI respectively) and lower stem (LSE and LSI respectively). Probes used in these two microarray studies were designed based on the EST sequences (https://urgi.versailles.inra.fr/Species/Flax/Download-sequences; Huis et al., 2012). EST probes were queried against putative NAC gene coding sequences (CDS) by running a local BLASTN program. Only those with 90% length match to the CDS and the sequence identities not less than 95% were considered. The cutoff E-value was 10^{-10}. Heat maps were then created using the mean RMA-normalized, averaged gene-level signal intensity ($\log_2$) values of all the biological replicates by MultiExperiment Viewer (MeV v4.9, http://www.tm4.org/ -mev.html). Genes were hierarchically clustered with Pearson correlation.

Transcript abundance of genes was also examined in a microarray dataset produced by our lab. This microarray study compared the gene expression levels in five 1-cm segments collected from the stem of 3-weeks-old flax, including 0-1 cm (T1), 2-3 cm (T2), 3-4 cm (T3), 4-5 cm (T4) and 8-9 cm (T5) from the shoot apex. Phloem fibers in T1, T2 and T3 were undergoing intrusive growth while in T4 and T5 demonstrating thick secondary cell wall. Probes were designed based on a published draft of flax genome (Wang et al., 2012). Probes were aligned to the coding sequences of predicted flax NACs and only those with identity greater than 95% and E-value less than 10^{-10} were used in the further analysis (To, 2013). Two-way ANOVA with Tukey's multiple

comparisons was then performed to find the *LusNACs* with differential expression in at least one of these five segments.

4.2.3.3 RNA-Seq

Transcript patterns of NACs were analyzed in a previously published RNA-Seq dataset (Kumar et al., 2013). Additionally, I have analyzed the expression patterns of *NACs* in the AR and BR using the RNA-Seq data I presented in Chapter 2 of this thesis.

4.2.3.4 qRT-PCR

Transcript abundance of selected *NACs* in the AR and BR were compared through quantitative real time-PCR. Flax plant (*L. usitatissimum* cv. CDC Bethune) growth, tissue collection, RNA extraction, cDNA synthesis as well as qRT-PCR performance and data analysis were the same as described in Section 2.2.1.

I also checked the expression patterns of VNS subfamily members across 12 different tissues by qRT-PCR. Five tissues were collected from one-month-old plants, including early fibers (EF), early xylem (X), roots (R), leaves (L), early cortical peels (ECP). The other seven tissues were collected from two-month-old plants, including senescent leaves (SL), late cortical peels (LCP), late fibers (LF), late xylem (LX), flowers (F), flower buds (FB), green bolls (GB). The flax cultivar CDC Bethune was used. GADPH was used as an endogenous control (Huis et al., 2010). Primer sequences used were listed in Appendix 2.

4.3 Results

4.3.1 Identification of NACs in flax genome

Through BLASTP analysis, a total of 182 putative flax NAC proteins distributed on 126 separate scaffolds were identified. These proteins consisted of 56 to 677 amino acids, with an average length of 345 amino acids. The isoelectric point ranged from 4.21 to 10.65 (Appendix 10). All these 182 putative flax NAC proteins were confirmed by HMMER3 and Pfam to contain a NAM domain (Pfam domain PF02365).

4.3.2 Phylogenetic analysis

To classify the putative flax NAC proteins based on sequence similarity, I constructed a maximum-likelihood phylogenetic dendrogram using protein sequences of NACs from flax, *Arabidopsis* and poplar (Figure 4-1). *Arabidopsis* was selected since it was a commonly used model plant and currently most functional information of NAC transcription factors has been obtained from studies in *Arabidopsis* (Shahnejat-Bushehri et al., 2016; Lee et al., 2017; Nakano et al., 2015). On the other hand, poplar was chosen because it was in the same order (Malpighiales) as flax and the genome sequences of poplar have been well annotated (Tuskan et al., 2006). VT+F+G4 was suggested by IQ-TREE as the best-fit substitution model for these sequences and therefore it was applied to construct the phylogenetic dendrogram (Abascal et al., 2005). Based on the phylogenetic analysis and data from *Populus*, I divided NAC domain proteins into 17 separate clades (Table 4-1; Hu et al., 2010). Clade 8 was the biggest one and had 31 members in flax, 25 in poplar and 12 in *Arabidopsis*. Flax members were represented in all the clades excepted clade 3, which comprised a single *Arabidopsis* member (*ANAC006*; Table 4-1). Clade 1 and 4 appeared largely expanded in flax (Table 4-1). Interestingly, clade 2 had around 20 *NACs* from flax and poplar

respectively with no *Arabidopsis* representatives, suggesting that *NACs* in this clade might have acquired important functions in Malpighiales (Table 4-1). Additionally, we found most flax *NACs* appeared in pairs, which were probably produced by the previously reported genome duplication (Figure 4-1; Wang et al., 2012).

Based on the phylogenetic tree, the flax VNS subfamily had 17 members in flax, 13 in *Arabidopsis* and 16 in poplar (Clade 12 in the Figure 4-1). The 17 flax VNS genes included eight *VNDs* (*LusNAC136, LusNAC28, LusNAC125, LusNAC42, LusNAC20, LusNAC46, LusNAC10 and LusNAC160*), six *NSTs* (*LusNAC151, LusNAC36, LusNAC161, LusNAC146, LusNAC66* and *LusNAC164*) and three *SMBs* (*LusNAC89*, *LusNAC122* and *LusNAC61*).

4.3.3 Meta-analysis of LusNAC gene expression

Studying spatial and temporal expression patterns of genes can supply useful information about their functions. To make some inferences about functions of flax NAC genes, I explored their expression abundance in existing EST, microarray and RNA-Seq datasets. The data sources investigated are described in the Table 3-2.

4.3.3.1 ESTs of *LusNACs*

To find out which of the putative *LusNAC*s were transcribed, I searched for ESTs of each *LusNAC* in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). As a result, ESTs were identified for 49 out of the 182 putative flax NAC genes, and ESTs of *LusNACs* were observed in all the sampled tissues (Table 4-1). This result suggests that *LusNACs* are involved in a great range of developmental and physiological process. However, only a few ESTs were detected for most of the 49 LusNACs except *LusNAC104*, which had 28 ESTs detected, of which 27 were derived from libraries of developing embryo with the remaining one from endosperm (Table 4-1).

4.3.3.2 *LusNACs* expression analysis in publicly available microarray datasets

As described above, only a minority of *LusNAC* genes were represented in the publically available EST databases. To obtain a more complete picture of NAC expression patterns in flax, I further performed a comprehensive expression analysis of *LusNAC* genes in two previously published microarray datasets, GSE21868, and GSE29345 (Aug et al., 2015; Huis et al., 2012). Expression profiles of 36 flax NAC genes were obtained while no data were found for the remaining *LusNACs* (Figure 4-2; Figure 4-3). This low coverage was reasonable since probes of these arrays were designed based on the ESTs but not the genomic data. Through analyzing gene expression in GSE21868, we found a number of *LusNACs* were enriched in specific tissues. For example, six *LusNACs* showed preferential expression in the flax inner stem tissues, including *LusNAC46*, *LusNAC160*, *LusNAC87*, *LusNAC66*, *LusNAC31* and *LusNAC121* (Figure 4-2). *LusNAC66*, *LusNAC31* and *LusNAC121* were abundant in the inner stem tissue at the vegetative stages while *LusNAC87* was enriched in the inner part of the stem from green capsule stage (Figure 4-2). On the other hand, *LusNAC46* and *LusNAC160* were enriched in the inner tissues of the flax stem at both stages (Figure 4-2).

Furthermore, many *LusNACs* displayed especially high expression levels in leaves, including *LusNAC5*, *LusNAC16*, *LusNAC39*, *LusNAC143*, *LusNAC29*, *LusNAC25*, *LusNAC33* and *LusNAC126* (Figure 4-2). Moreover, *LusNAC26* was found to be particularly abundant in the seeds at 10-15 days after flowering and *LusNAC137* had the highest transcription abundance in root (Figure 4-2). Meanwhile, *LusNAC29* had apparently more transcript abundance in the flax cultivar Belinka compared to Drakkar (Figure 4-2).

Microarray dataset GSE29345 explored the expression patterns of genes in six parts of flax stem, including: external tissues of the upper stem (USE), the middle stem (MSE), the lower stem (LSE) and the whole stem (WSE); internal tissues of the upper stem (USI), the middle stem (MSI), the lower stem (LSI) and the whole stem (WSI). The external tissues of flax stems are phloem and cortex enriched while the internal tissues of flax stems are xylem-enriched. In flax, the stem tissues and cell walls show a developmental gradient along the length of the stem from the top to the bottom. In the internal part of flax stem, this developmental gradient consists of the formation of primary xylem and then layers of secondary xylem, which each successive layer of xylem tissue undergoing secondary cell wall thickening as well as extensive lignification. In contrast, in the external part of flax stem, the upper tissues were associated with phloem fiber elongation and the middle part was the start point of secondary cell wall formation (Gorshkova & Morvan, 2006). A heatmap constructed using the expression data from this analysis indicated that *LusNAC29* transcripts were enriched exclusively in the inner tissues of the lower part of the stem. As found in the dataset GSE21868, transcripts of *LusNAC46*, *LusNAC160*, *LusNAC87*, *LusNAC66*, *LusNAC31* and *LusNAC121* were particularly enriched in the inner part of flax stem as compared to the external part (Figure 4-2). Among them, *LusNAC31* accumulated the highest transcription abundance in the internal tissue of the upper stem, while *LusNAC31, LusNAC106* and *LusNAC46* were found to be enriched in the inner tissue of the entire length of the stem (Figure 4-3).

I also analyzed putative flax NAC transcription factors in a microarray produced by our lab that investigated the transcript abundance of genes in five different 1-cm regions of flax stem. Expression data for 128 out of the 182 putative LusNACs were checked in this study (probes were not present for the remainder of the other 54 flax NAC genes (data not shown)). Among the 128 LusNACs detected, only three genes (*LusNAC182*, *LusNAC67* and *LusNAC161*) showed

differential expressions in at least one tissue (Table 4-4). This study showed that *LusNAC182* was obviously enriched in the shoot apex, with decreasing expression as the stem got mature (Figure 4-4). *LusNAC161* had highest expression level in the stem just below the snap point whereas *LusNAC67* was most abundant just around the snap point (Figure 4-4).

4.3.3.3 RNA-Seq

A previously reported RNA-Seq study measured the transcript expression of *LusNACs* in 13 flax tissues and 167 putative *LusNACs* were detected in at least one of the tissues examined (Kumar et al., 2013). Overall, the *LusNACs* showed diversified expression patterns among these tissues. A majority of the *LusNACs* were expressed in all or many of the tissues tested but their respective transcript abundance was rather low (Figure 4-5a; 4-5b; 4-5c; Appendix 11). However, several *LusNACs* accumulated very high transcript abundance in specific tissues, including four *LusNACs* (*LusNAC32*, *LusNAC68*, *LusNAC115* and *LusNAC128*) that were clearly enriched in the mature embryo, seven specifically enriched in the anther (*LusNAC175*, *LusNAC51*, *LusNAC62*, *LusNAC133, LusNAC26, LusNAC31* and *LusNAC63*), three particularly abundant in the flower (*LusNAC43, LusNAC141* and *LusNAC166)* and two exclusively enriched in leaf *(LusNAC95 and LusNAC163*; Figure 4-5a; 4-5b; 4-5c; Appendix 11). Additionally, I found 43 of the 162 *NACs* tested (26.5%) peaked their transcript abundance in embryo, 35 (21.6%) in flower and 29 (17.9%) in anther and 21 (13%) in stem (Figure 4-5a; 4-5b; 4-5c; Appendix 11).

We were interested to find flax *NACs* with transcript expression patterns that correlated with flax phloem fiber specification. To find potential candidates, we analyzed expressions patterns of *NACs* in the RNA-Seq data described in Chapter 2 of this thesis. As a result, nine *LusNACs* were found to be significantly more enriched in the apical region (AR) compared to the basal region (BR) by

the RNA-Seq (Table 4-5). Among them, seven *NACs* were above two-fold enriched in AR and transcripts of one *NAC* (*LusNAC65*) was only detected in the AR but not in the BR (Table 4-5). Inversely, 30 *NACs* were revealed to be significantly more abundant in the BR compared to the AR, among which, 17 were above 2-fold BR-enriched and nine *LusNACs* were detected in the BR but not in the AR (Table 4-6). *NACs* not detected in the AR or BR indicated that they might either not be transcribed or transcribed at low abundance in the corresponding tissue. Alternatively, they might be transcribed during a different developmental stage.

4.3.3.4 qRT-PCR

4.3.3.4.1 Confirm the expression patterns of several AR-enriched *LusNACs* by qRT-PCR

As described above, nine *LusNACs* were found to be more abundant in the AR compared to the BR by RNA-Seq (Table 4-5). Because the AR was expected to contain genes that regulated flax phloem fiber specification, I wanted to use qRT-PCR to confirm expression of genes in this region. I was only able to measure transcript abundance for eight of the nine genes identified by qRT-PCR since no gene-specific primers were obtained for *LusNAC65*. I measured transcript abundance in three regions: the AR, BR, and 1 cm below the AR. I checked the 1 cm region between the AR and BR because genes enriched in this area are expected to be related to phloem fiber cell elongation but not cell specification. As indicated in Figure 4-6, qRT-PCR analysis suggested that all of eight NACs tested showed preferential expression in the AR as compared to either the BR or the 1 cm segment below AR.

4.3.3.4.2 Analysis of the expression patterns of *LusVNS* genes in 12 flax tissues by qRT-PCR

Due to the important roles of VNS genes in vascular differentiation and secondary cell wall development, I have further investigated their expression profiles in 12 different flax tissues by

qRT-PCR. Overall, genes in this subfamily showed diverse expression patterns (Figure 4-7). All the VNS subfamily members were detected in roots while leaves had the lowest numbers of VNS genes detected (only 13 out of 17 genes were detected; Figure 4-7).

Among the flax *VNDs*, *LusNAC136* was specifically expressed in the late xylem whereas *LusNAC28* and *LusNAC125* were most abundant in early fibers. *LusNAC10*, *LusNAC160*, *LusNAC46* and *LusNAC20* were preferentially expressed in the early xylem while *LusNAC42* had a low expression level in all the tested tissues except in root. Meanwhile, all the *LusVNDs* except *LusNAC136* were enriched in root (Figure 4-7).

All six flax *NST/SNDs* appeared to be enriched in the root, vascular tissues of the stem and reproductive tissues examined (Figure 4-7). Specifically, *LusNAC66* and *LusNAC146* were exclusively abundant in the green bolls while *LusNAC164*, *LusNAC161*, *LusNAC36* and *LusNAC151* showed high expression levels in flowers, green bolls, flower buds and roots. Beyond these, *LusNAC161* and *LusNAC36* were also found to be enriched in the xylem. The difference was *LusNAC161* was enriched in both early and late xylem while *LusNAC36* was abundant in the early xylem only (Figure 4-7).

All the three *LusSMB* genes were most enriched in the roots and they showed overall lower expression levels compared to *LusVNDs* and *LusNSTs*. Different with *LusVNDs* and *LusNSTs,* the flax genes in the SMB family tended to be expressed only in some tissues (Figure 4-7). For example, *LusNAC89* was only detected in roots and late xylem whereas *LusNAC122* was only transcribed in roots, early cortical peels, late cortical peels, green bolls and flower buds. Transcripts of *LusNAC61* were not detected in early fibers and leaves (Figure 4-7).

Combining the data in Figure 4-1 and 4-7, I found that the duplicated gene pairs in VNS subfamily tended to have consistent expression patterns with respect to the tissues tested, like *LusNAC28/LusNAC125*; *LusNAC46/LusNAC20*; *LusNAC36/LusNAC151*; *LusNAC10/LusNAC 160*. This suggested that genes produced through genome duplication tend to maintain their functions during evolution. Meanwhile, some duplicated gene pairs showed very different expressions patterns (such as *LusNAC122 /LusNAC89*), suggesting that after duplication they might have experienced sub-functionalization or neofunctionalization.

4.4 Discussion

NAC domain proteins are plant-specific transcription factors that play important roles in many aspects of plant development as well as environmental responses. Here, we have identified 182 putative NAC domain proteins in the flax genome, one of the largest known NAC families (Jin et al., 2014; Shao et al., 2015). Through phylogenetic analysis, they were classified into 17 different clades (Figure 4-1). Clade 1 and clade 4 were apparently expanded in flax compared to *Arabidopsis* and poplar, indicating that *NACs* in these two clades might have evolved some lineage-specific roles. To date, the *Arabidopsis* and poplar genes in these two clades have not been functionally characterized.

4.4.1 Flax genes in the VNS subfamily

The dendrogram indicated that the flax VNS subfamily had 17 members, with 8 *VND*s, 6 *NST*s and 3 *SMB*s (Clade 12 of Figure 4-1). The gene numbers in each group were comparable to those found in poplar, which had eight *VNDs*, four *NSTs* and four *SMBs* respectively (Yao et al., 2012). Comparative genomics identified VNS genes in many plant species and found their number was

not significantly associated with genome size or the presence of woody tissues (Zhu et al., 2012). On the other hand, Yao et al. compared the number of *VNDs*, *NSTs*, and *SMBs* in 19 plant species and found that 17 species had more *VNDs* than *NSTs* and *SMBs* (Yao et al., 2012).

I checked the expression patterns of these 17 *LusVNSs* in 12 flax tissues. Among them, *LusNAC136* was specifically expressed in the xylem, and its *Arabidopsis* ortholog, *VND7* was suggested to regulate the xylem differentiation in root (Figure 4-7; Kubo et al., 2005). This suggested that *LusNAC136* may have a specific role in regulating stem xylem differentiation. Beyond *LusNAC136*, four other VND genes appeared to be involved in xylem tissue differentiation in both flax stem and root, including *LusNAC10, LusNAC160, LusNAC46* and *LusNAC20* (Figure 4-7). In contrast, two other VND genes including *LusNAC28* and *LusNAC125* might be involved in regulating secondary cell wall formation in stem phloem fiber since both of them had highest expression levels in stem phloem fiber (Figure 4-7). As with the *LusVNDs*, many flax NST genes were also enriched in the stem vascular tissues (*LusNAC164, LusNAC161and LusNAC36*) and roots (*LusNAC164, LusNAC161, LusNAC36, LusNAC151*). However, we found all the flax genes in the NST group were strongly expressed in the reproductive tissues, for instance, some were revealed to be enriched in the flower (*LusNAC164, LusNAC146, LusNAC36 and LusNAC151*), green capsules (*LusNAC164, LusNAC146, and LusNAC66*) and flower buds (*LusNAC164, LusNAC161, LusNAC36 and LusNAC151*). These expression patterns were consistent with the previous findings. In *Arabidopsis*, genes in the VND groups were preferentially expressed in xylem vessels and they regulated the root and shoot xylem vessel cell differentiation, while genes in NST groups were suggested to regulate the differentiation of secondary cell wall containing cells other than xylem vessels, including interfascicular fiber (*NST1* and *NST3*), anther endothecium (*NST1* and *NST2*) and silique cells (NST1 and NST3; Zhong et al., 2006; Zhong et

al., 2007b; Mitsuda et al., 2007; Kubo et al., 2005; Mitsuda et al., 2005; Mitsuda & Ohme-Takagi, 2008). However, in poplar, rice and maize, both VND and NST genes were expressed in vessels and fibers (Zhong et al., 2010; Zhong et al., 2011). By contrast, the three SMB-related genes in flax had overall low transcript abundance; they were shown to be obviously more enriched in the root. The Arabidopsis SMB-related genes were revealed to be expressed in the root cap and induce the ectopic secondary cell wall deposition when overexpressed (Willemsen et al., 2008; Bennett et al., 2010). Meanwhile, compared to the xylem and phloem fiber expression of poplar VNDs and NSTs, the poplar SMB group genes were only expressed in root tissues (Zhong et al., 2010; Ohtani et al., 2011). Altogether, flax VNDs might be involved in vascular tissue differentiation in root and stem while NSTs might be involved in the secondary cell walls of many tissues. However, SMBs might have a role in flax root. All these indicated that flax VNS genes might have conserved roles with their homologs in other plant species.

4.4.2 *LusNACs* with a potential role in phloem fiber specification

The RNA-Seq analysis described in Chapter 2 identified 9 NACs that were more abundant in the AR of flax stems compared to the BR (Table 4-5). Their enrichment in the AR was confirmed by qRT-PCR (Figure 4-6). These NACs that were enriched in the shoot apex may be associated with specification of phloem fiber cell identity, or with many other processes. For instance, *LusNAC93* and *LusNAC65* are orthologues of CUC (CUP-SHAPED COTYLEDON) protein, a transcriptional regulator of postembryonic shoot meristem formation and organ boundary formation (Hibara et al., 2006; Burian et al., 2015). *LusNAC50* and its duplicated gene *LusNAC27* were both found to be more enriched in the AR compared to BR. However, their *Arabidopsis* ortholog, SOG1 has been reported to be required in actively dividing cells since they acted as a master regulator of DNA damage (Yoshiyama et al., 2009). Among the remaining six AR-enriched *NACs*,

LusNAC100 and *LusNAC120* were duplicated genes and were therefore considered to share conserved functions. *LusNAC158* belongs to clade 2 in the phylogenetic dendrogram, which consisted of 21 flax genes and 17 poplar genes but no *Arabidopsis* representatives.

4.4.3 Other *LusNACs* possible to be involved in the flax stem vascular tissue differentiation

Four other LusNACs (*LusNAC87*, *LusNAC66, LusNAC31* and *LusNAC121*) have been found to be specifically enriched in the inner tissues of flax stem, suggesting that they might have a role in the flax stem xylem tissue development (Figure 4-2).

By analyzing NAC genes in an unpublished microarray study of five 1-cm segments collecting from different positions of flax stem, I found *LusNAC182*, *LusNAC67* and *LusNAC161* were specifically enriched in certain stem segments. *LusNAC182* was most abundant in 0-1 cm below the shoot apex (Figure 4-4). Since the flax phloem fiber cells in this area were undergoing intrusive elongation, we assumed that *LusNAC182* might be involved in phloem fiber cell elongation. NAC domain transcription factors have been found to be involved in cell expansion through transcriptional regulation of genes such as cellulose synthase and aquaporins (Pei et al., 2013; Jiang et al., 2014). The turgor pressure change was one main process of flax phloem fiber elongation and aquaporin genes were highly expressed in fiber-forming tissues including flax phloem fiber (Snegireva et al., 2010; Roach & Deyholos, 2008; Roach & Deyholos, 2008). *LusNAC67* was most enriched in T3 (3-4 cm from the shoot apex) which corresponded to the snap point (Figure 4-4). The snap point is a mechanically-definable region in the flax stem that is considered a transition point of phloem fiber development (Gorshkova et al., 2003). Phloem fibers in the stem above this point grow intrusively and do not deposit secondary cell walls, whereas phloem fibers in the stem below this point had thick secondary cell wall. I proposed that *LusNAC67*

might have a role in regulating secondary cell wall formation in flax phloem fiber cells. Although *LusNAC161* was significantly more abundant in T4 than other areas of flax stem, I found this gene was more abundant in the xylem tissue of the flax stem and root compared to phloem by qRT-PCR analysis (Figure 4-4; Figure 4-7). This might indicate that this gene is involved in both phloem fiber and xylem secondary cell wall deposition in flax stem.

4.4.4 *LusNACs* might be related to embryo development

In this study, I found some *LusNACs* that were specifically enriched in embryos. *LusNAC104* had the most ESTs identified and a vast majority of them were detected from libraries of embryo (Figure 4-2). RNA-Seq again showed that this gene was enriched in embryo (Appendix 11). However, EST identification indicated that this gene was most abundant in the embryos at the torpedo stage while RNA-Seq analysis revealed that this gene had highest expression level in the embryos at the heart stage (Table 4-2; Appendix 11). This indicated LusNAC104 might be involved in the flax embryo development. Its *Arabidopsis* homologous genes (*NTL9*) is reported to regulate plant defense response and other characterized *Arabidopsis* genes in the same clade are involved in the pathogen defense (e.g.*ANAC091*), cold stress (*ANAC062*) or cell differentiation (*ANAC068*) but no embryo-related functions have yet been reported (Donze et al., 2014; Kim & Park, 2007; Seo & Park, 2010). Meanwhile, the microarray data GSE21868 indicated that *LusNAC26* was specifically expressed in embryos at 10-15 days after flowering, indicating that it might have a role in early stage of embryo development. Furthermore, four other *LusNACs* were specifically expressed in the mature embryo, including *LusNAC32, LusNAC68, LusNAC115* and *LusNAC128* (Figure 4-5a, b, c). This indicated that flax NACs might be involved in different stages of embryo development. In the expression data obtained from Kumar's RNA-Seq study, I found five flax NACs (*LusNAC26*, *LusNAC51*, *LusNAC62*, *LusNAC133* and *LusNAC175)* specifically

enriched in anther and three NACs (*LusNAC43*, *LusNAC141* and *LusNAC166*) exclusively abundant in flower (Kumar et al., 2013). The involvement of NACs in embryogenesis and floral development have been reported before. For example, the *NAM* (*no apical meristem*) gene was revealed to be required for the pattern formation in embryos and flowers and CUC genes were reported to be involved in the shoot apical meristem formation (Souer et al., 1996; Takada et al., 2001; Vroemen, 2003). A tomato NAC gene (SINAM2) was involved in flower-boundary morphology and a rose NAC (RhNAC100) was suggested to control the cell expansion in flower petals (Hendelman et al., 2013; Pei et al., 2013).

In total I have found evidence for transcription of 180 of 182 predicted LusNACs with the exception of Lus10005917 and Lus10037106. These two could either be pseudogenes or genes with some spatial or temporal expression patterns not covered in the analyzed datasets.

4.5 Conclusions

This study has identified 182 putative NAC genes from the flax genome. They were clustered into 17 distinct clades and two clades (Clade 1 and Clade 4) were found to be largely expanded in the flax. Using a combination of EST, microarray, RNA-Seq and qRT-PCR data, experimental evidence was found for 180 putative *LusNACs*. The expression data listed in this study may provide useful information for function annotation of this gene family in flax.

4.6 Figures and tables

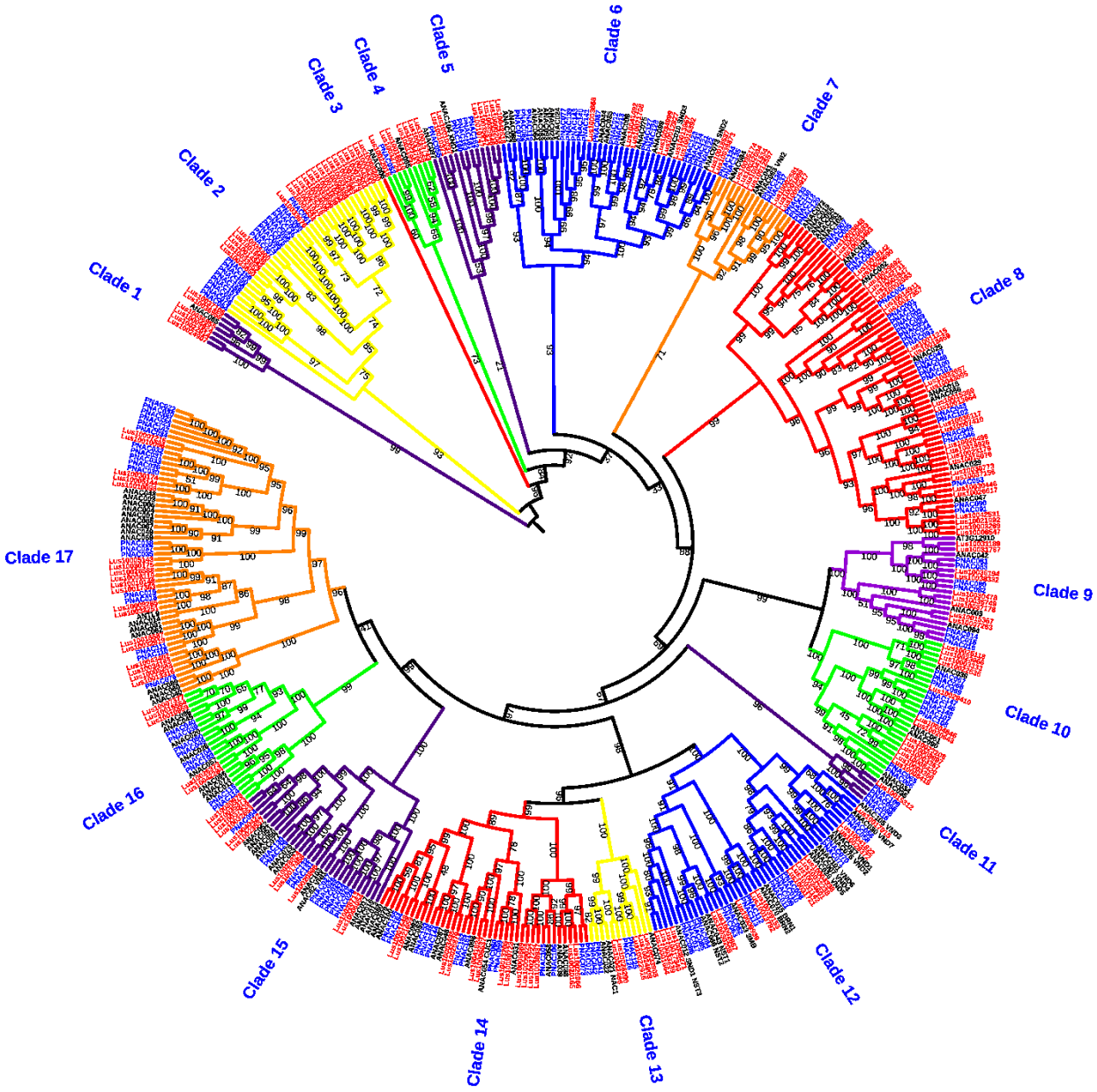


Figure 4-1 Maximum-likelihood phylogenetic tree of NAC domain-containing proteins from flax (red leaves), *Arabidopsis* (black leaves), and poplar (blue leaves). The full-length amino acid sequences were used to construct a phylogenetic tree using IQ-TREE (Nguyen et al., 2015). The numbers labeled on each node were bootstrap values.

Table 4-1 Membership details of each LusNAC clade. *Lus*: *Linum usitatissimum; At: Arabidopsis thaliana; Ptr:* poplar *trichocarpa;*

Clade	***Lus***	***At***	***Ptr***
1	6	1	0
2	21	0	17
3	0	1	0
4	7	1	1
5	6	2	5
6	10	14	17
7	7	3	6
8	31	12	25
9	9	4	7
10	13	5	10
11	1	0	4
12	17	13	16
13	5	3	4
14	14	13	13
15	9	11	11
16	7	8	8
17	19	16	19

Table 4-2 Number of flax *NACs* ESTs in various tissues. Tissues examined are as follows: Globular embryo (GE), Heart embryo (HE), Torpedo embryo (TE), Cotyledon embryo (CE), Mature embryo (ME), Endosperm (EN), Globular seed coat (GC), Torpedo seed coat (TC), Etiolated seedling (ES), Leaf (LE), Stem (ST), Stem peel (PS), Flower (FL), Fiber enriched tissue at mid-flowering stage (F) (Venglat et al., 2011; Day et al., 2005).

Gene Name	**GE**	**HE**	**TE**	**CE**	**ME**	**EN**	**GC**	**TC**	**ES**	**LE**	**ST**	**PS**	**FL**	**F**	**Total**
LusNAC104	1	3	13	8	2	1									28
LusNAC92			1	1			3	1		1					7
LusNAC32	1	1		1			2	1							6
LusNAC114			1				4							1	6
LusNAC68		1		1										3	5
LusNAC111				2										3	5
LusNAC128					1		2				1			1	5
LusNAC169											3	2			5
LusNAC5			1					2						1	4
LusNAC79		1	1			1	1								4
LusNAC95			2		1									1	4
LusNAC140											3			1	4
LusNAC180						3			1						4
LusNAC44														3	3
LusNAC66					1	1		1							3
LusNAC119														3	3
LusNAC145		1				1					1				3
LusNAC158						1	2								3
LusNAC40										1	1				2
LusNAC47													2		2
LusNAC49												1		1	2
LusNAC51								2							2
LusNAC70								2							2
LusNAC73			1									1			2

LusNAC115					2										2
LusNAC156					1									1	2
LusNAC166								1				1			2
LusNAC179														2	2
LusNAC9	1														1
LusNAC16														1	1
LusNAC25												1			1
LusNAC31														1	1
LusNAC34						1									1
LusNAC48							1								1
LusNAC80												1			1
LusNAC96						1									1
LusNAC100			1												1
LusNAC118														1	1
LusNAC124							1								1
LusNAC126				1											1
LusNAC130							1								1
LusNAC135												1			1
LusNAC139										1					1
LusNAC142														1	1
LusNAC146								1							1
LusNAC153					1										1
LusNAC161								1							1
LusNAC165	1														1
LusNAC175												1			1

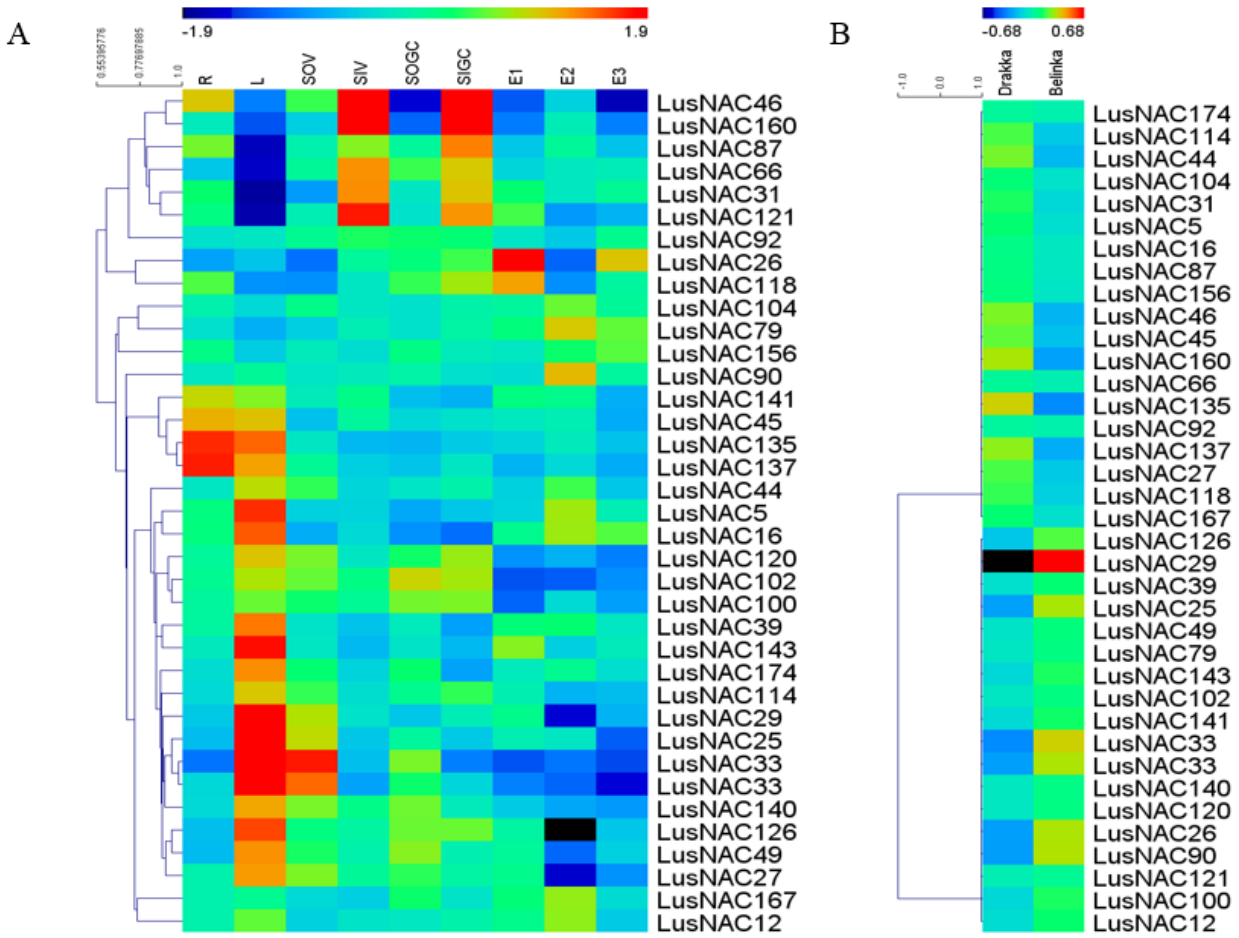


Figure 4-2 Transcript abundance of *LusNACs* in previously published microarray dataset (GSE21868) (Aug et al., 2015). Red indicates high abundance whereas blue indicates low abundance. A: Tissues analyzed including: root (R); leaf (L);outer stem tissues at the vegetative phase (SOV); inner stem tissues at the vegetative phase (SIV); outer stem tissues at the green capsule phase (SOGC); inner stem tissues at the green capsule phase (SIGC); Seeds at 10-15 days after flowering (E1); Seeds 20-30 days after flowering (E2); Seeds 40-50 days after flowering (E3); B: Expressions of *LusNACs* were compared in two contrasting flax cultivars, Drakkar and Bellinka. RMA-normalized, average log2 signal values were used to prepare a heat map by MEV (MultiExperiment Viewer; Howe et al., 2010). Genes were clustered using Pearson Correlation distance matrix by single clustering method. The signal values for each gene were mean centered before clustering. This involves taking the mean expression value for each gene and subtracting it from each expression value for that gene. The mean value will be zero.

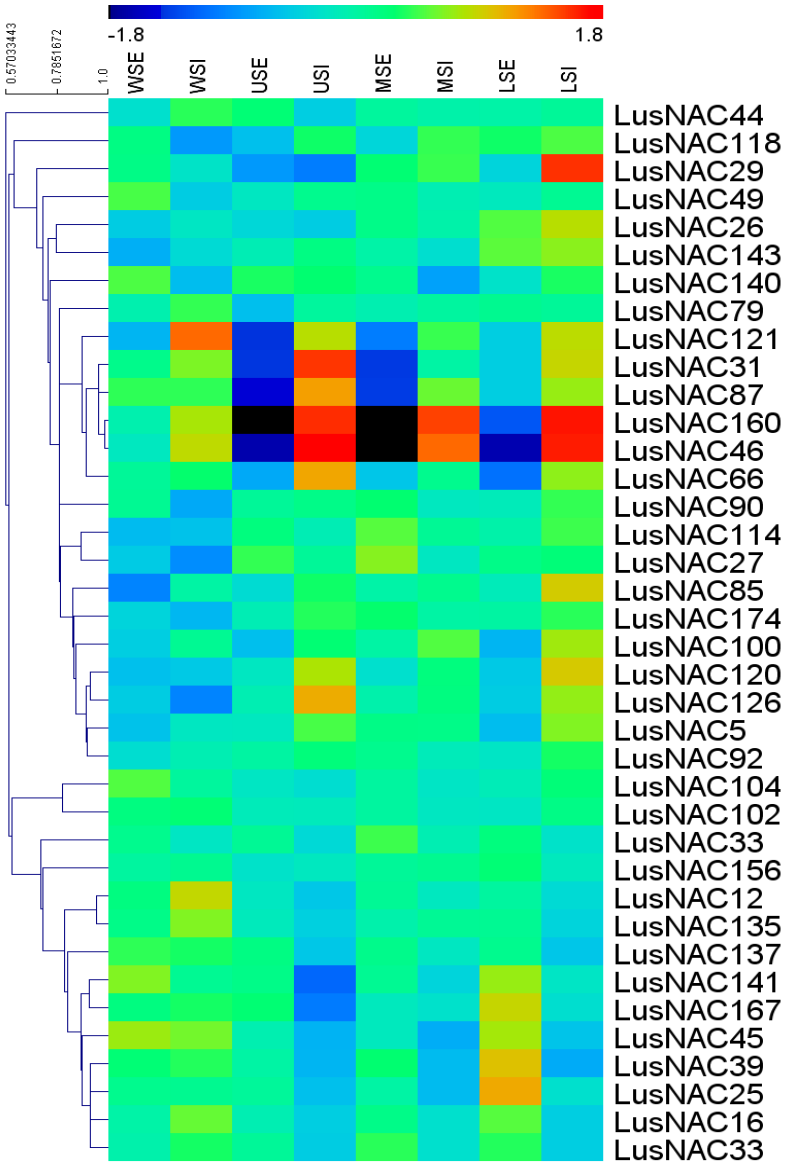


Figure 4-3 Transcript abundance of *LusNACs* in previously published microarray dataset (GSE29345; Huis et al., 2012). RMA-normalized, average log2 signal values were used to produce a heat map. Red indicated high abundance while blue indicated low abundance. WSE: external (i.e. phloem and cortex enriched) tissues of the whole stem; WSI: internal tissues of the whole stem; USE: external tissues of the upper stem; USI: internal tissues of the upper stem; MSE: external tissues of the middle stem; MSI: internal tissues of the middle stem; LSE: external tissues of the lower stem; LSI: internal tissues of the lower stem; RMA-normalized, average log2 signal values were used to prepare a heat map by MEV (MultiExperiment Viewer; Howe et al., 2010). Genes were clustered using Pearson Correlation distance matrix by single clustering method. The signal values for each gene were mean centered before clustering. This involves taking the mean expression value for each gene and subtracting it from each expression value for that gene. The mean value will be zero.

Table 4-3 Transcript abundance of *LusNAC* probes with differential expression in at least one out of the five stem tissues examined. Data was obtained from (To, 2013)

Probe Name	Gene name	T1	T2	T3	T4	T5
g35937.t1|sl-1141-1181	*LusNAC161*	0.35443	0.92048	3.69146	5.20481	3.60091
g42595.t1|sl-802-837	*LusNAC182*	10.5536	8.12987	6.49118	4.95178	1.73903
g1479.t1|sl-636-671	*LusNAC67*	2.79748	7.74971	10.1025	5.52504	5.86915

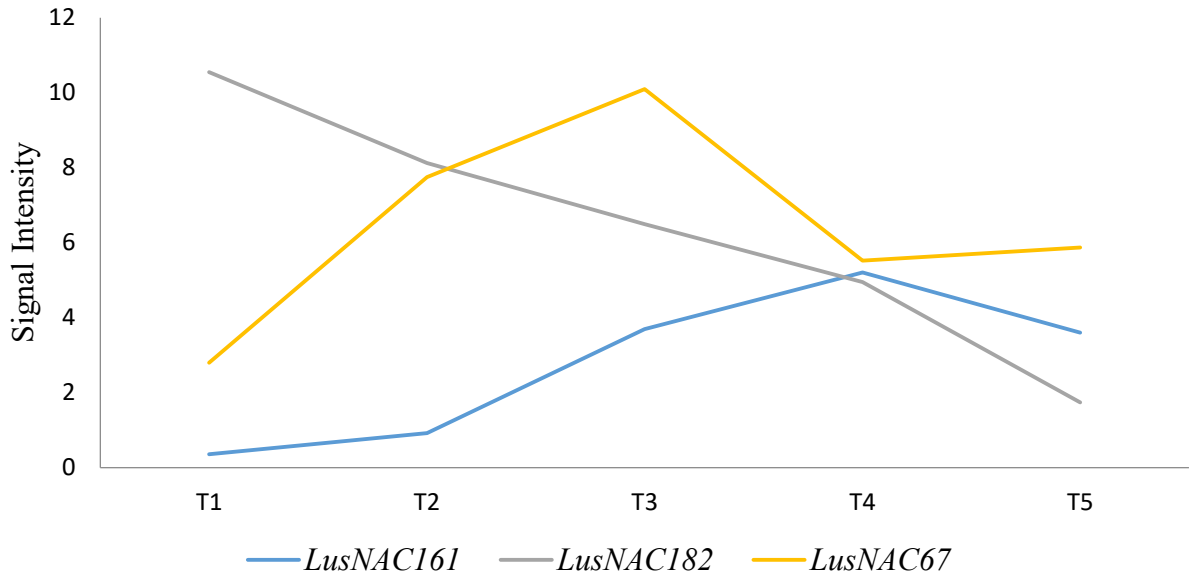


Figure 4-4 *LusNACs* with differential expression in at least one out of the five stem tissues examined. Data was obtained from (To, 2013).

Table 4-4 Significance analysis for the *LusNACs* among the five 1-cm segments studied in flax stem microarray study. Data was obtained from (To, 2013).

	LusNAC161	*LusNAC182*	*LusNAC67*
	g35937.t1|sl-1141-1181	g42595.t1|sl-802-837	g1479.t1|sl-636-671
T1 vs. T2	ns	ns	***
T1 vs. T3	ns	*	****
T1 vs. T4	**	***	ns
T1 vs. T5	ns	****	ns
T2 vs. T3	ns	ns	ns
T2 vs. T4	**	ns	ns
T2 vs. T5	ns	****	ns
T3 vs. T4	ns	ns	**
T3 vs. T5	ns	**	**
T4 vs. T5	ns	ns	ns

A two-way ANOVA test was conducted followed by a Tukey's multiple comparisons test using GraphPad Prism 7.00. * denotes *p*-value between 0.01-0.05; **denotes *p*-value between 0.001-0.01; ***denotes *p*-value between 0.0001-0.001; ****denotes *p*-value <0.0001; ns (not significant) denotes *p*-value >0.05;

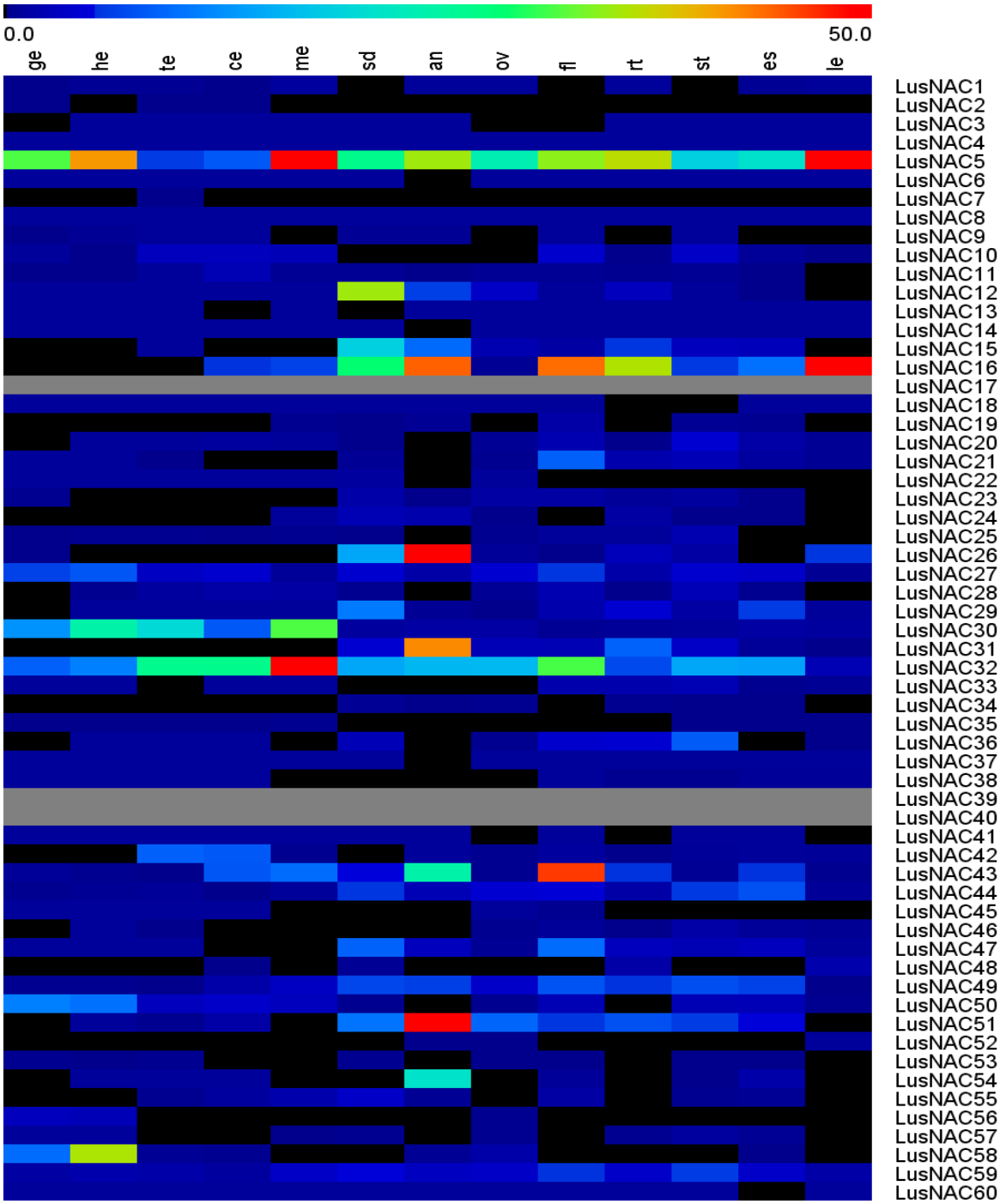


Figure 4-5a Expression profiles of *LusNAC1-60* in 13 different tissues (Kumar et al., 2013). They were as follows: globular embryo (ge), heart embryo (he), torpedo embryo (te), cotyledon embryo (ce), mature embryo (me), seeds (sd), anthers (an), ovaries (ov), mature flower (fl), root (rt), stem (st), etiolated seedlings (es), leaves (le). The Mev_4_9_0 was applied to draw the heat map (Howe et al., 2010). Red indicated high expression whereas blue indicated low expression. Genes with no expression in all the tested tissues were shown as grey.

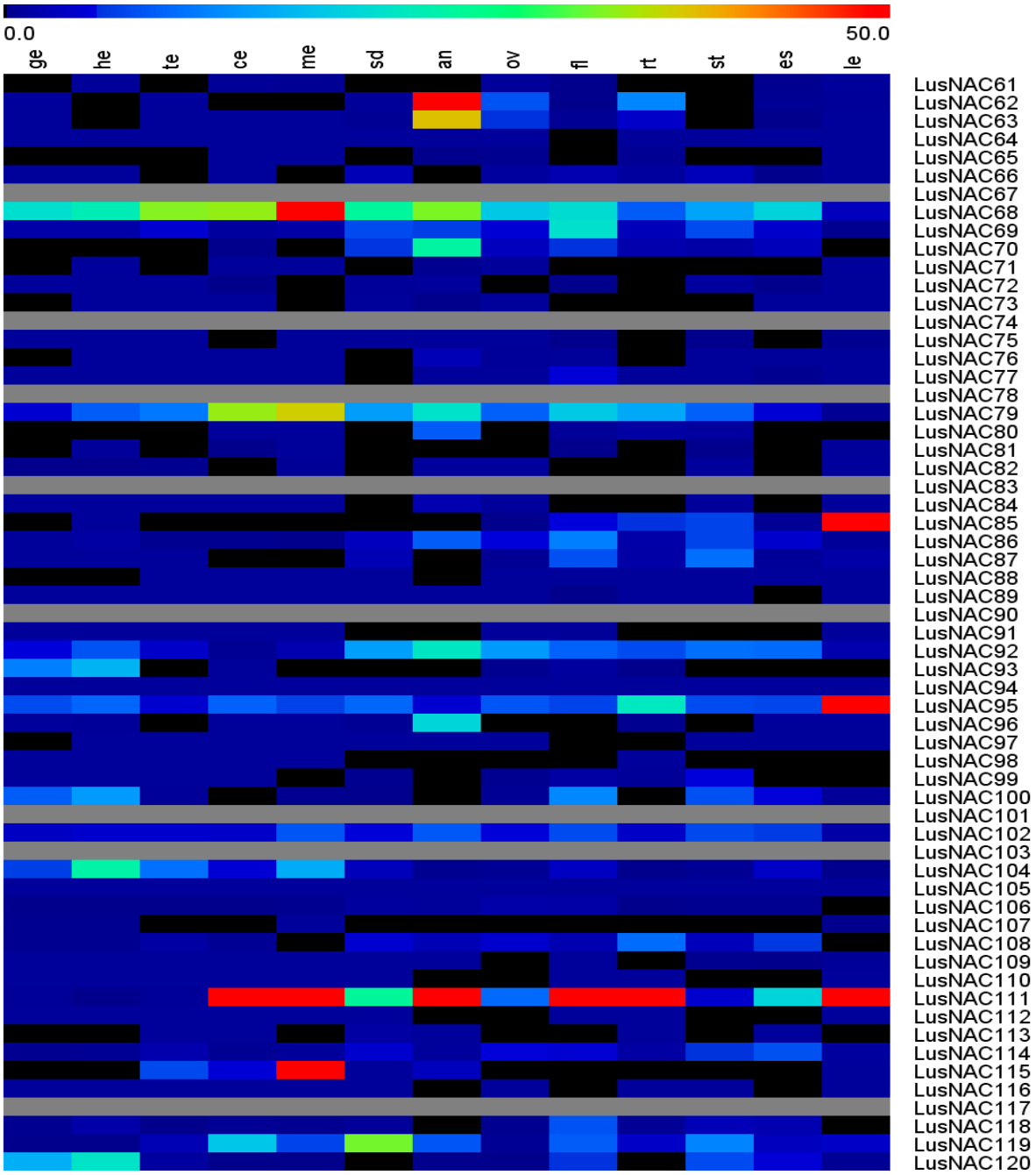


Figure 4-5b Expression profiles of *LusNAC61-120* in 13 different tissues (Kumar et al., 2013). They were including: globular embryo (ge), heart embryo (he), torpedo embryo (te), cotyledon embryo (ce), mature embryo (me), seeds (sd), anthers (an), ovaries (ov), mature flower (fl), root (rt), stem (st), etiolated seedlings (es), leaves (le). The Mev_4_9_0 was applied to draw the heat map (Howe et al., 2010). Red indicated high expression whereas blue indicated low expression. Genes with no expression in all the tested tissues were shown as grey.

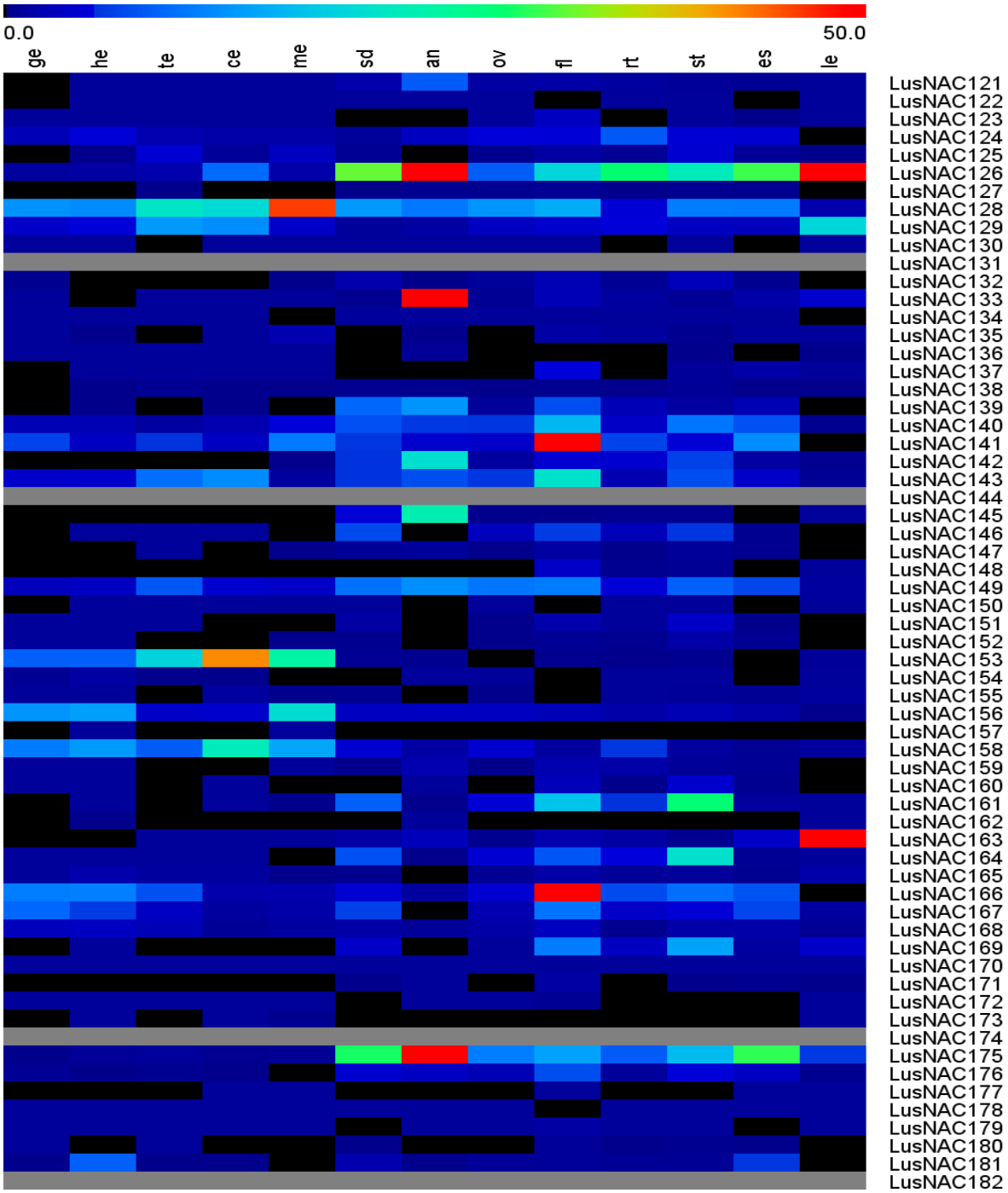


Figure 4-5c Expression profiles of *LusNAC120-182* in 13 different tissues (Kumar et al., 2013). They were including: globular embryo (ge), heart embryo (he), torpedo embryo (te), cotyledon embryo (ce), mature embryo (me), seeds (sd), anthers (an), ovaries (ov), mature flower (fl), root (rt), stem (st), etiolated seedlings (es), leaves (le). The Mev_4_9_0 was applied to draw the heat map (Howe et al., 2010). Red indicated high expression whereas blue indicated low expression. Genes with no expression in all the tested tissues were shown as grey.

Table 4-5 *LusNACs* with significant more transcripts in the AR compared to the BR. Data was obtained from Chapter 2 of this thesis.

Gene Name	FPKM (AR)	FPKM (BR)	log2(fold_change AR/BR)	q_value
LusNAC93	21.63	0.48	5.49	0.002
LusNAC158	22.43	3.17	2.82	0
LusNAC50	17.88	3.06	2.55	0
LusNAC100	23.1	5.45	2.08	0
LusNAC120	24.44	7.08	1.79	0
LusNAC27	19.44	6.19	1.65	0
LusNAC92	55.1	24.19	1.19	0
LusNAC114	18.71	10.46	0.84	0.005
LusNAC65	11.78	0	NA	0

Table 4-6 *LusNACs* significantly more enriched in the BR compared to the AR. Data was obtained from Chapter 2 of this thesis.

Gene Name	FPKM (AR)	FPKM (BR)	log2(fold_change AR/BR)	q_value
LusNAC169	0.68	27.07	-5.32	0.001
LusNAC119	1.1	41.48	-5.24	0
LusNAC83	0.68	17.25	-4.67	0
LusNAC99	0.31	7.18	-4.52	0.034
LusNAC10	0.65	10.25	-3.99	0.001
LusNAC125	0.74	11.27	-3.93	0
LusNAC181	0.45	4.35	-3.27	0.03
LusNAC132	0.71	6.69	-3.23	0
LusNAC28	0.48	4.14	-3.09	0.01
LusNAC29	0.24	1.78	-2.88	0.046
LusNAC19	0.31	2.12	-2.77	0.01
LusNAC180	0.35	2.21	-2.64	0.045
LusNAC16	2.83	16.91	-2.58	0
LusNAC126	14.26	80.21	-2.49	0
LusNAC140	3.11	14.86	-2.26	0
LusNAC143	4.76	15.2	-1.68	0
LusNAC175	29.77	80.5	-1.43	0
LusNAC86	4.03	7.76	-0.95	0.02
LusNAC69	6.35	11.46	-0.85	0.011
LusNAC59	4.29	7.13	-0.73	0.045
LusNAC49	9.15	14.46	-0.66	0.029
LusNAC12	0	2.42	NA	0
LusNAC33	0	2.97	NA	0
LusNAC36	0	21.29	NA	0
LusNAC38	0	3.03	NA	0
LusNAC40	0	5.82	NA	0
LusNAC74	0	1.48	NA	0
LusNAC109	0	3.92	NA	0
LusNAC151	0	11.27	NA	0
LusNAC161	0	65.15	NA	0

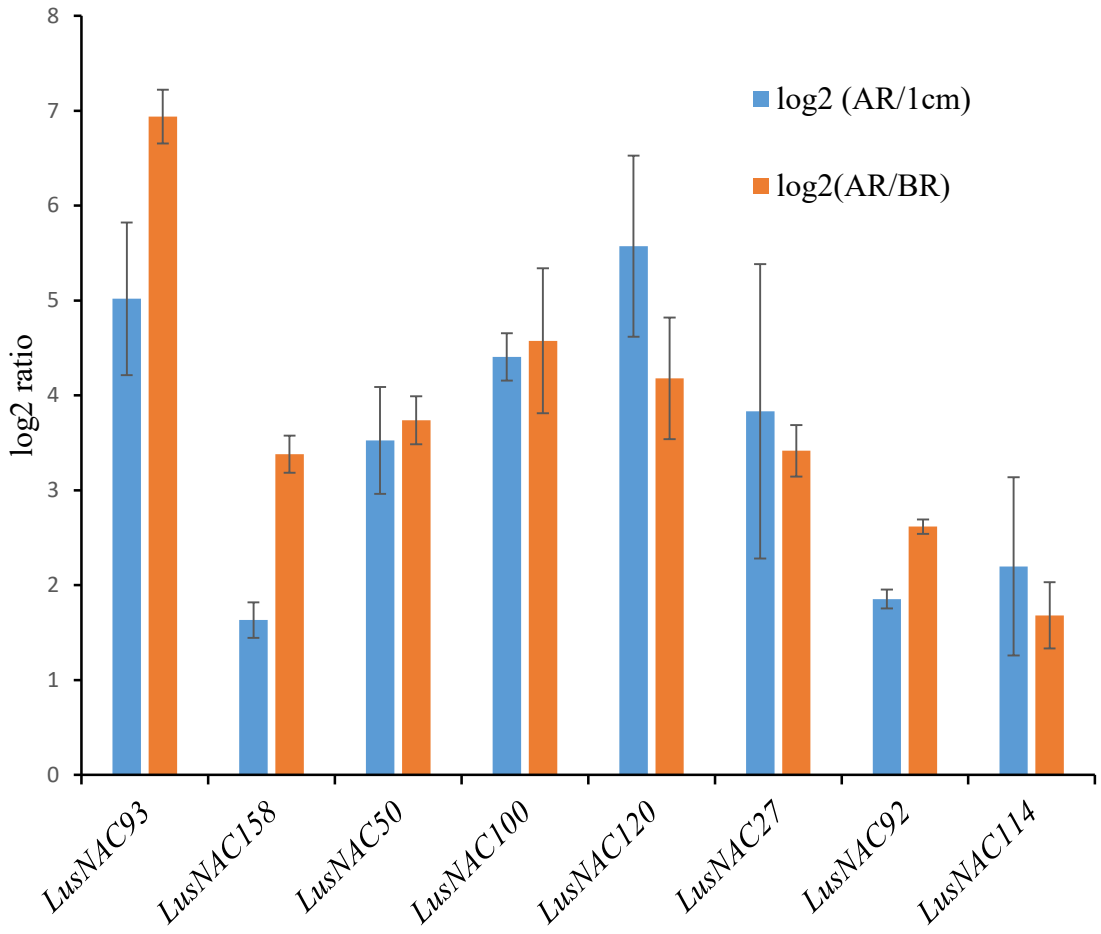


Figure 4-6 Validation of the expressions of eight selected AR-enriched *LusNACs* by qRT-PCR. Error bars denoted standard derivations.

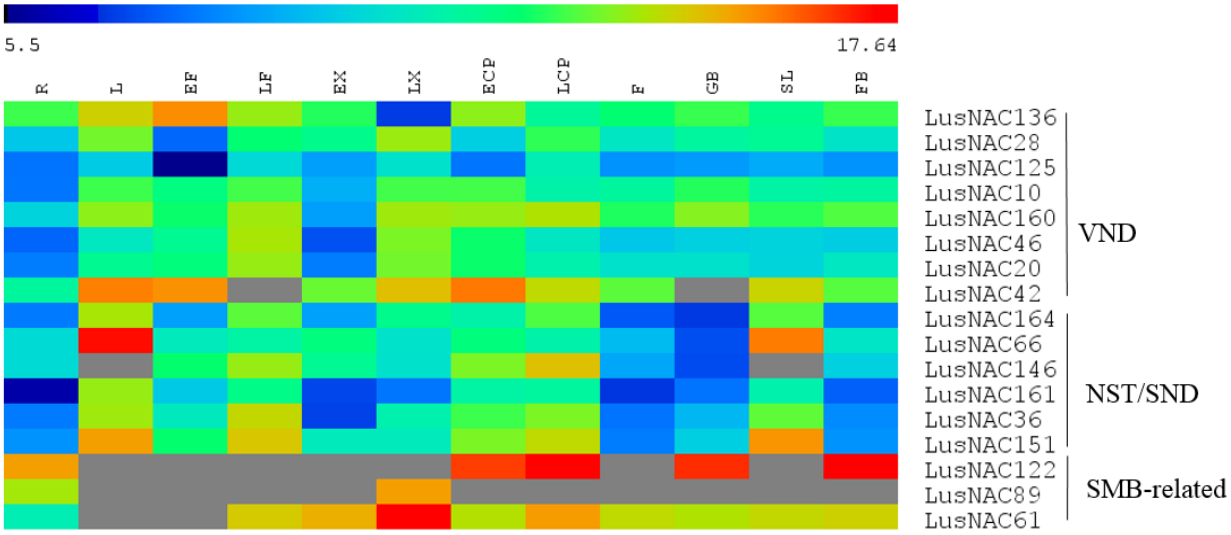


Figure 4-7 Transcript abundance of VND, NST/SND and SMB orthologue genes in 12 different tissues analyzed by qRT-PCR. Delta- C_T (C_T of target gene minus C_T of endogenous controls) values were used to produce a heat map. Blue indicates high expression level while red indicates low abundance. Grey indicated no transcripts detected. R: roots; L: leaves; EF: early fibers; LF: late fibers; EX: early xylem; LX: late xylem; ECP: early cortical peels; LCP: late cortical peels; F: flowers; GB: green bolls; SL: senescent leaves; FB: flower buds ;

Chapter 5. Functional analysis of an uncharacterized *Arabidopsis* gene, *At3g05980*.

5.1 Introduction

In my study of flax shoot apex transcript expression (Zhang & Deyholos, 2016), I identified a predicted flax gene, *Lus10041215*, which had transcripts that were 53 times more abundant in the shoot apex as compared to the remainder of the stem. The 207 aa protein encoded by *Lus10041215* does not contain any conserved domains that have been annotated in either Pfam or NCBI's Conserved Domain Database (Finn et al., 2015b; Marchler-Bauer et al., 2015). However, *Lus10041215* has been assigned to an unnamed PANTHER protein family, PTHR31722:SF2, which includes two genes from Arabidopsis, *At5g19340* and *At3g05980*, as well as genes from several other eudicots (Mi et al., 2017). Indeed, when *Lus10041215* is used to query the Arabidopsis proteome, *At5g19340* and *At3g05980* are the best BLASTP matches (e-value 5.4 x 10^{-20}; 3.9 x 10^{-19}, respectively). Because functional genetic analysis in *Arabidopsis* is faster and easier than in flax, I chose to characterize *At3g05980* in *Arabidopsis.* This gene was selected because *At3g05980* was shown by the microarray data in eFP Browser to be enriched in the shoot apical meristem, whereas *At5g19340* abundance in the shoot apical meristem was relatively low.

5.2 Materials and methods

5.2.1 Plant materials

All the seeds used in this study were in the Columbia (Col-0) background. Arabidopsis plants were grown at 22°C with a cycle of 16 h light and 8 h dark. The surface-sterilized *Arabidopsis* seeds were vernalized for 4 days at 4°C in darkness before being sown on 1/2 X MS medium. The 1/2 X

MS medium as referred to throughout this thesis contains 1/2 strength Murashige & Skoog (MS) basal medium, plus 0.7% (w/v) agar and 1% (w/v) sucrose.

5.2.2 In silico analysis

The gene structure, amino acid length, molecular weight and isoelectric point of At3g05980 were obtained from the Arabidopsis Information Resource (TAIR; Garcia-Hernandez et al., 2002). Signal peptide and transmembrane domain analyses were conducted by SignalP 4.0 and TMHMM Server v 2.0 respectively (Petersen et al., 2011; Krogh et al., 2001). The presence of annotated conserved domains was checked using ScanProsite and Pfam (Sigrist et al., 2009; Coggill, et al., 2015). The subcellular localization of At3g05980 was predicted using several commonly used web servers, including PSORT, WoLF PSORT, Plant-mPLoc, TargetP, MultiLoc2, SUBA3 and YLoc (Nakai & Horton, 1999; Horton et al., 2007; Chou & Shen, 2010; Emanuelsson et al., 2007; Blum et al., 2009; Hawkins & Bodén, 2006; Briesemeister et al., 2010; Tanz et al., 2013).

5.2.2.1 Homologs Identification and conservation analysis

BLASTP was used to align At3g05980 to the predicted proteins from 64 Viridiplantae genomes available at Phytozome v12.1 database using the default settings, except that the e-value threshold was set at $<10^{-6}$ (Goodstein et al., 2012). Multiple sequence alignment of all the At3g05980 homologs was conducted using ClustalW and MAFFT with default parameters and full protein sequences (Edgar, 2004; Katoh et al., 2009). The conserved motifs among these protein sequences were identified using the MEME suite (Bailey et al., 2009).

5.2.2.2 Phylogenetic analysis

A neighbor-joining phylogenetic tree was constructed using MEGA5 from the multiple sequence alignment produced by MAFFT described above (Tamura et al., 2011) and the Dayhoff amino acid

substitution model. Dayhoff model was selected in this study since the most widely used amino acid substitution matrices are based on this model (Henikoff & Henikoff, 1992). Gaps or positions missing residues were deleted from pairwise distance estimate. Default values were used for the remaining parameters. Branch support was determined using bootstrap with 1000 replicates run under same search parameters.

5.2.2.3 *At3g05980* expression prediction

In silico expression profiles of *At3g05980* were extracted from the eFP Brower 2.0 in the Bio-Analytic Resource for Plant Biology server (BAR) and Genevestigator (Zimmermann et al., 2005; Waese et al., 2017).

5.2.2.4 Co-expressed analysis

The names of the top 300 genes co-expressed with *At3g05980* were obtained from ATTED-II and input into the Bingo application in Cytoscape v 3.5.1 to conduct Gene Ontology enrichment analysis (Obayashi et al., 2007; Shannon et al., 2003).

5.2.2.5 *cis*-acting regulatory elements prediction

The entire upstream intergenic region upstream of the initiation codon of *At3g05980* (2,799 bp) was input into PLACE, PlantCARE and AGRIS for *cis*-acting regulatory element identification (Kenichi et al., 1999; Rombauts et al., 1999; Yilmaz et al., 2011). The *Arabidopsis* genome sequence was downloaded from Phytozome v12.1 and the occurrence of each cis-element in *Arabidopsis* genome was counted using Bioperl scripts (Stajich et al., 2002). A one-tailed Z-test was used to determine whether a cis-element was significantly enriched in the *At3g05980* promoter compared to the whole genome (The Arabidopsis Genome Initiative, 2000). The formula used was as follows: $Z = \frac{(F_p - F_g)}{(\frac{F_g \times (1-F_g)}{N_p})}$, where F_p

indicated the frequency of a certain cis-element in the promoter sequence, F_g was the frequency of a certain cis-element in the genome, and N_p was the length of the promoter fragment.

5.2.2.6 Protein 3D structure and function prediction

The 3D structure of At3g05980 was predicted using I-TASSER, and this server predicted the function of this protein based on the top-ranked 3D model (Zhang, 2008).

5.2.3 At3g05980 expression pattern analysis

5.2.3.1 Promoter:: GUS fusion study

All the intergenic DNA sequence upstream of the start codon of *At3g05980* (2799 bp) was amplified from *Arabidopsis* wild-type (Col-0) plants using primers with HindIII and BamHI restriction sites (HindIII-At3g05980promoterF: CCCAAGCTTGGTTATAATATTTTATGTGG; BamHI-At3g05980promoterR: CGCGGATCCTTCTTCTATTGTGATGAAG). The resulting PCR product was purified with Wizard® SV Gel and PCR Clean-Up System (Promega) and then subcloned into the TOPO TA Cloning® vector before transformed into *E. coli* Top10 competent cells. Plasmids were extracted using Plasmid Miniprep Kit (Qiagen) and then digested with HindIII and BamHI. The digestion products were then cloned into the same site of the pRD420 vector (Datia et al., 1992). The construct was then introduced into *Arabidopsis* wild-type plants (Col-0) through *Agrobacterium tumefaciens* GV3101 by floral dip (Clough & Bent, 1998). *At3g05980pro:: GUS* transgenic seeds were selected on 1/2 X MS medium containing 50 µg/ml kanamycin. Expression of the *GUS* gene was studied in the T_2 generation of the *At3g05980pro:: GUS* transgenic plants. More than 10 progeny of each of 10 independent primary transformants were

analyzed. pRD410 transformants carrying a CaMV 35S:uidA fusion and pRD420 transformants carrying uidA with no promoter were used as positive and negative controls, respectively.

For histochemical GUS staining, seedlings or tissues of transgenic plants were vacuum infiltrated in ice-cold 90% (v/v) acetone for 2 minutes before incubation at -20°C for 30 min. Samples were then washed twice with 50 mM $NaHPO_4$ (pH7.2) and incubated in 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_3 $Fe(CN)_6$, 2 mM X-gluc) at 37°C for 2 days. After successive incubation in 30% ethanol (one hour) and FAA (50 % ethanol, 5 % formaldehyde, 10 % glacial acetic acid) overnight, tissues were transferred into 70% ethanol for final storage. Samples were then observed with an Olympus BX51 microscope and photographed with an HDCE-90D digital camera.

5.2.3.2 qRT-PCR

For qRT-PCR testing of hormone responsiveness of *At3g05980*, Arabidopsis Col-0 wild-type plants of seven days after sowing (DAS) were incubated in 10 µM ABA (abscisic acid), 5 µM IAA (3-indoleacetic acid), 5 µM BA (6-benzylaminopurine), 10 µM MeJA (methyl jasmonate), 1 µM BR (brassinosteroid), 20 µM ACC (1-aminocyclopropane-1-carboxylic acid), 1 µM GA3 (gibberellic acid-3 potassium salt). For qRT-PCR testing of the responsiveness of *At3g05980* to salt, osmotic and cold stress, 7 DAS (days after sowing) wild-type Arabidopsis plants were transferred to the liquid MS medium (half-strength MS basal medium plus 1% sucrose) with 0.2 M NaCl or 0.3 M mannitol. Cold treatment was done by transferring seven DAS plants to the MS liquid medium and incubated in the 4°C. Samples for all the treatment were collected at 1 h, 3 h, 6 h, 12 h and 24 h after incubation.

For qRT-PCR analysis of tissue-specific *At3g05980* expression, shoot apices and inflorescence apices were dissected from *Arabidopsis* WT plants at 18 DAS and 23 DAS respectively, under a dissecting microscope. The Arabidopsis at 18 DAS had ten visible rosette leaves and 1-4 floral primordia at stage 3-5 (Smyth et al., 1990). The shoot apices sample contained some leaf tissues or floral primordia that could not be dissected entirely from the shoot apex. Four DAS seedlings had two cotyledons but true leaves had not emerged, while 7 DAS seedling had the first two true leaves visible. Rosette leaves, cauline leaves, siliques (green siliques) and four flower samples (stage 12, 13, 14 and 15/16) were taken from one-month-old plants. Flower stages were assigned according to Cai's definition (Cai & Lashbrook, 2008). Roots were collected from plants at 18 DAS. Three biological replicates of each tissue were collected in liquid nitrogen and stored at -80 °C until use. RNA was isolated using RNeasy Plant Mini Kit and then treated with TURBO DNA-free™ Kit to remove DNA. A NanoDrop 1000 spectrophotometer and an Agilent 2100 Bioanalyzer were used to check the RNA quality. The cDNA was then synthesized with a First Strand cDNA Synthesis Kit and Oligo (dT) 18 primer. Each PCR reaction had three biological replicates and three technical replicates. Real-time PCR was performed in an Applied Biosystems 7500 Fast Real-time PCR System. Each amplification reaction was 10 µl and consisted of 0.4 µM of each primer, 0.25 X SYBR Green, 1 X ROX, 0.075 U Platinum Taq, 0.2 mM dNTPs and 2.5 µl 16 X diluted cDNA. Threshold cycles (C_T) were determined through 7500 Fast Software. The *Arabidopsis* Actin 2 and EF-1a genes were used as endogenous controls (Czechowski, 2005). Each sample had three biological replicates and three technical replicates. Data were analyzed using the $\Delta\Delta C_T$ method (Zhang et al., 2015). Primers sequences used in this study were as follows:

At3g05980qPCRAS: TTTAGAGACGGTTTCAAAGACG;

At3g05980qPCRS: GAGAAGGAGATACGAGGTCCAA;

At3g05980qPCRAS2: AGGACAGTGTCGTCTTTGTCTCC;

At3g05980qPCRS2: TTCGCTGCGTCCTCAAGTGAAC;

Actin2AS: TGAGAGATTCAGATGCCCAGAA;

Actin2S: TGGATTCCAGCAGCTTCCAT;

EF1AAS: TGAGCACGCTCTTCTTGCTTTCA;

EF1AS: GGTGGTGGCATCCATCTTGTTACA;

The specificity of primers was checked using BLASTN to align them to the *Arabidopsis* genome sequences in the NCBI nucleotide database and by examining the migration of the PCR products using agarose gel electrophoresis.

5.2.4 Subcellular localization

To create transgenic plants overexpressing a GFP-At3g05980 fusion protein, the CDS (coding DNA sequence) of At3g05980 was first PCR amplified from the cDNA of WT Arabidopsis using BamHI and XbaI incorporating primers (BamHI -At3g05980CDSF:5' CGCGGATCCTCATGGTTTTAGAGACGGTTTC 3'; XbaI-At3g05980CDSR:5' CTAGTCTAGACTAGGCGCGTCTCTCTACT 3'). The amplicon was then digested with BamHI and XbaI, and was ligated into the BamHI and XbaI double-digested pCsGFPBT vector. The resulting constructs were then transformed into *A. tumefaciens* GV3101 through freeze-thaw method and the positive transformants were transferred into WT *Arabidopsis* plants through floral dipping method (Clough & Bent, 1998). Transgenic plants were selected on the 1/2 X MS medium

containing 50 ng/µL Hygromycin B. T_3 plants from six independent lines were used for localization analysis.

For peroxisome localization, a binary peroxisome marker plasmid (Clone name: PX-RB) was obtained from TAIR and transformed into homozygous 35S:: GFP: At3g05980 transgenic plants using *A. tumefaciens* GV3101. In this marker, mCherry (a red fluorescent protein) includes the peroxisome targeting signal 1 (PTS1, Ser-Lys-Leu) located at its C-terminus (Shaner et al., 2004; Reumann, 2004). This marker uses a 35S promoter with dual enhancer elements (Nelson et al., 2007). The transgenic plants were selected on 1/2 X MS medium supplemented with 10 µg/ml Basta (also known as glufosinate-ammonium or phosphinothricin). T_2 transgenic plants were observed using confocal laser scanning microscope.

A modified pCAMBIA1303 vector (pCAMBIA1303m) described previously was used to create transgenic plants overexpressing a At3g05980-CiFP fusion protein (Khan, 2015). A DNA fragment was synthesized by Genescript, which had NcoI and AfeI restriction sites incorporated to the 5' end and 3' end of the *At3g05980* coding sequence. The synthesized DNA fragment was then digested with NcoI and AfeI and inserted into the NcoI and AfeI double digested pCAMBIA1303m vector by ligation. The ligation product was again transformed into the *E. coli* (TOP10) competent cells. Clones grown on the LB medium added with 50 µg/ml kanamycin were propagated. Plasmids were then extracted and sequenced to confirm that the CDS of At3g05980 protein was inserted in-frame to the N-terminus of the CiFP and no mutation occurred. The correct fusion constructs were then transformed into *A. tumefaciens* GV3101 through freeze-thaw method and the positive transformants were transferred into WT *Arabidopsis* plants through floral dipping

method (Clough & Bent, 1998). Transgenic plants were selected on the 1/2 X MS medium containing 50 ng/µl Hygromycin B. T_3 seeds of six independent lines were used for localization analysis. T_3 plants from six independent lines were used for localization analysis.

5.2.5 Overexpression plasmid construction

The CDS of *At3g05980* was amplified from cDNA of Arabidopsis seedlings using NcoI and BstEII tagged primers. The PCR product was cloned into the pCRII-TOPO vector and transformed into *E. coli* (TOP10) competent cells. Ampicillin (50 µg/ml) was used to select positive clones. The selected positive clones were then grown overnight in liquid LB medium (37°C) and plasmids extracted from these clones were sequenced to confirm that no mutation had occurred. The confirmed plasmids as well as pCAMBIA1303 vector were double digested by NcoI and BstEII. Digested products were separated on a 2% agarose gel. The At3g05980 CDS fragment as well as a modified vector were excised from the gel and purified using a Wizard® SV Gel and PCR Clean-Up System. T4 DNA ligase was then used to clone the At3g05980 CDS fragment into the pCAMBIA1303 vector. The ligation product was again transformed into the *E. coli* (TOP10) competent cells. Clones grown on the LB medium added with 50 µg/ml kanamycin were propagated. Plasmids were then extracted and sequenced to confirm that the At3g05980 fragments were correctly inserted into the pCAMBIA1303 vector. The correct plasmid were then transformed into *Agrobacterium* GV3101 competent cell by electroporation and into the *Arabidopsis* by floral dipping (Narusaka et al., 2010). Transgenic plants were selected on the 1/2 X MS medium containing 50 ng/µl Hygromycin B. Homozygous plants from the T_3 generation of three independent transgenic lines were used for phenotyping analysis. qRT-PCR was performed to check the relative transcript abundance of At3g05980 in each transgenic lines compared to the wild type by the $\Delta\Delta C_T$ method as described in the section 2.2.3. Floral buds of WT plants and each

of the overexpression lines were sampled from four-week-old plants and EF-1a was used as the endogenous control (Czechowski, 2005).

5.2.6 Identification of the homozygous T-DNA insertional mutants

Two T-DNA insertional mutant lines were obtained from ABRC: SALK_024489 and SAIL_1054_G02 (Alonso, 2003). Genotyping was performed using two-primer PCR and the non-transformed parent control was used as a control. One PCR reaction was performed using LP+RP and another PCR reaction using LB+RP. A product was obtained in the LP+RP reaction for WT or HZ lines, with no product for HM lines. Meanwhile, no product was obtained for the HM or HZ lines in the LB + RP reactions.

Primers used for SAIL_1054_G02 were:

LP: TAGAACCAAAACGAGTGGTCC

RP: AAGGAGATACGAGGTCCAAGC

LB2: GCTTCCTATTATATCTTCCCAAATTACCAATACA

Primers used for SALK_024489 were:

LP: GGAAGCAATTTACCTTCGGAG

RP: TTTGTCCATACCCAATAGTTTGC

LBb1.3: ATTTTGCCGATTTCGGAAC

5.2.7 Creation of *At3g05980* mutant by CRISPR-Cas9

The CRISPR-PLANT online platform was used to design sgRNA targets for *At3g05980* (Xie et al., 2014). Those closest to the start codon (<100 bp) were sent to the Cas-OFFinder for off-target

prediction and to the CRISPRscan for editing efficiency prediction (Bae et al., 2014; Moreno-Mateos et al., 2015). The sgRNA targets used in this analysis were as follows:

At3g05980 target 1: GAGATACGAGGTCCAAGCAACGG

At3g05980 target 2: TGTAAGGAAGATGTCGTCAAAGG

At3g05980 target 3: TCATCTGATTTATCTGACGGTGG

At3g05980 target 4: ATCTTCCTTACACATTACGGGGG

Two constructs were made to generate an *At3g05980* single mutant following the previous description, and each construct had two *At3g05980* sgRNA targets inserted into the pHEE401 vector (Xing et al., 2014). Construct one had *At3g05980* target 1 and target 2 inserted whereas the construct two had *At3g05980* target 3 and target 4 cloned into the pHEE401 vector.

These constructs were then transformed into *Arabidopsis* wild-type plants (Col-0) through Agrobacterium strain GV3101, using the floral dip method (Clough & Bent, 1998). Seeds of T_0 plants were selected on 1/2 X MS medium with 25 mg/L Hygromycin, and resistant seedlings were grown into soil. Genomic DNA of T_1 plants was extracted using DNeasy Plant Mini Kit. I amplified and sequenced the fragments flanking the target sites by PCR using gene-specific primers, to confirm presence of the intended gene edits.

5.2.8 Freezing assay

For cold-acclimation (CA) treatments, 14 DAS Arabidopsis seedlings grown in the ½ X MS culture plates were cultivated in a 4 °C chamber (16 h light/ 8 h dark) for 3 days before freezing treatment. For nonacclimated (NA) treatment, 17 DAS Arabidopsis seedlings were treated by freezing directly. A programmable freezer was used to do the freezing treatment. Plants were

maintained at 0°C for 1 h and then the temperature was reduced by 1°C / h until the target temperature (described in the figure legend) reached. After the freezing treatment, plants were recovered in a 4°C chamber without light for 12 h and then grown for in normal growing condition (22°C with 16 h light) for another 3 days. The survival rates were then determined by counting the plants with emerging green leaves (Jiang et al., 2017).

5.2.9 Electrolyte leakage test

Whole seedlings were used in the electrolyte leakage test as described previously (Ding et al., 2015). Briefly, all seedlings following freezing treatment were collected in a conical screw-cap polypropylene tube with 8 ml deionized water. The electrical conductivity (EC) was measured (S_0). Samples were gently shaken at room temperature for 15 min before measuring the EC again (S_1). The samples were boiled at 100 °C for 30 min and shaken at room temperature for another 20 min before measuring the EC again (S_2). Electrolyte leakage was calculated using the following formula: $(S_1 - S_0)/ (S_2 - S_0)$.

5.2.10 Expression of cold-regulated genes

Expression of six cold-regulated genes was compared in the WT and the At3g05980 loss-of-function mutant. 10 DAS WT and mutant plants grown in 1/2 X MS medium were treated at 4°C with 16 h light/8 h dark. Total RNA was extracted from the whole seedlings. The cDNA synthesis as well as qRT-PCR performance were same as described in the section 5.3.3.2.

Actin2 was used as the reference gene. Primers used were as follows:

CBF1-qF: GGAGACAATGTTTGGGATGC;

CBF1-qR: CGACTATCGAATATTAGTAACTCC;

CBF2-qF: CGACGGATGCTCATGGTCTT;

CBF2-qR: TCTTCATCCATATAAAACGCATCTTG;

CBF3-qF: TTCCGTCCGTACAGTGGAAT;

CBF3-qR: AACTCCATAACGATACGTCGTC;

KIN1-qF: TGCCTTCCAAGCCGGTCAGA;

KIN1-qr: AGGCCGGTCTTGTCCTTCAC;

RD29A-qF: GCCGAGAAACTTCAGATTGG;

RD29A-qR: CCATTCCTCCTCCTCCTTTC;

COR47-qF: CCGAGCACGAGACACCAAC;

COR47-qr: TCCACGATCCGTAACCTCTGTT;

Actin2qF: TGAGAGATTCAGATGCCCAGAA;

Actin2qr: TGGATTCCAGCAGCTTCCAT;

5.2.11 Seed fatty acid profiling, auxin analogs sensitivity assay and sucrose dependence assay

Dry mature seeds were used for fatty acid determination. Seed fatty acids were extracted and analyzed as previously described (Poirier et al., 1999). Basically, fatty acids were first converted into FA methyl esters in methanol solution containing 1M HCl for 2 h at 80°C. The fatty acids in seeds were subsequently measured using GC-MS. For auxin analog sensitivity, seeds were plated on 1/2 X MS medium with 0.2 µg/ml 2,4-DB, 30 uM IBA or no hormone. Hormone concentrations were selected based on (Park et al., 2013; Footitt et al., 2002). Plates were grown at 22°C with 16 h light for 7 days before checking the root length. For the sucrose dependence assay, seeds were plated on 1/2 X MS medium or on medium without 1% sucrose. Plates were transferred to the dark for 7 days before photo was taken.

5.3 Results

5.3.1 *In silico* analysis of At3g05980

At3g05980 consists of a single exon encoding a predicted protein of 245 amino acids and 27.6 kDa with isoelectric point of 10.2. No signal peptide, transmembrane domain or any annotated functional domain was detected in its protein sequences. *In silico* analysis through several commonly used web-based algorithms predicted that this protein might be targeted to the cytoplasm, chloroplast, nucleus, mitochondrion or plastid (Table 5-1).

5.3.1.1 Homologs identification

Querying the At3g05980 amino acid sequence against the predicted proteins in the 64 sequenced plant species available at Phytozome v.12.1 by BLASTP (e-value $<10^{-6}$) identified a total of 63 presumed homologs of *At3g05980*, and these were found in all of the 37 eudicots analyzed (Goodstein et al., 2012). Most species had one or two copies, except *Kalanchoe laxiflora* which had four. However, no apparent homologs of *At3g05980* were detected in the surveyed genomes of monocots, bryophytes and green algae species. This absence indicates that *At3g05980* may be specific to eudicots. Meanwhile, keyword searching indicated that only these 63 proteins were annotated as members of PTHR31722: SF2 in Phytozome v12.1. As shown in the multiple sequence alignment, multiple highly conserved motifs exist in the protein sequence of *Lus10041215* as well as its distant homologs (Figure 5-1). *At3g05980* had only one paralog in *Arabidopsis, At5g19340.* These two Arabidopsis proteins shared 76.3% and 66.1% similarity and identity, respectively. This implied that they might have conserved functions.

5.3.1.2 Phylogenetic analysis

A neighbor-joining dendrogram was constructed from the protein sequences of *At3g05980* homologs, which is consistent with grouping into three broad clades: Clade I, II and III (Figure 5-2). Whereas there was little support for the backbone, many derived clades were well supported (e.g., 100% bootstrap). In Clade I, three flax genes (*Lus10041215*, *Lus10002455* and *Lus10010529*) formed a well-supported clade, suggesting that this group of genes likely originated from a duplication that occurred after flax had diverged from the other species that were analyzed. However, *Arabidopsis* genes showed a different pattern: *At3g05980* and *At5g19340* were place in separate subclades in Clade II. Both genes were placed with orthologs from *Arabidopsis lyrata*, *Arabidopsis halleri*, *Boechera stricta*, *Capsella rubella*, *Capsella grandiflora*, *Brassica rapa* and *Eutrema salsugineum*. This pattern suggested that in Brassicaceae, genes in this family had duplication events that occurred prior to the divergence of these species (Figure 5-2).

5.3.1.3 Conservation Analysis

To investigate the sequence conservation of this family, I analyzed multiple sequence alignments of the *At3g05980* homologs. Several highly conversed regions were identified among these genes (Appendix 12). Analysis by MEME Suite defined four conserved motifs within these proteins (Figure 5-2; Bailey et al., 2009). Sequence logos of these four conserved motifs are shown in the Figure 5-3 (Crooks et al., 2004). It was noted that all of the homologs contained four motifs, with the exception of two flax proteins, Lus10010529 (lacking motif 1 and motif 2) and Lus10041215 (lacking motif 2; Figure 5-2).

5.3.1.4 Expression prediction

To make inferences about the function of *At3g05980*, I analyzed its microarray-derived expression pattern using data from BAR (Waese et al., 2017). Microarray data indicated that this gene was ubiquitously detected in all the 47 tissues tested, and it was preferentially accumulated in shoot apices, petals, developing seeds, and roots (Figure 5-4a). In root, transcripts of this gene were most abundant in atrichoblast cells (Table 5-2a). GENEVESTIGATOR analyed the expression of this gene across 111 tissues and revealed that the ten tissues with the most abundant transcripts are: root epidermal atrichoblast, root epidermis, petal, axillary shoot, root hair, replum, lateral root cap, phloem and mesophyll cell (Figure 5-4b). Meanwhile, transcripts of this gene were reported to be repressed by exogenous application of the hormone ABA, and up-regulated by cold (Table 5-2b; 5-2c). In addition, osmotic stress reduced its expression in shoots while stimulating its transcript accumulation in root to a small extent (Table 5-2c). UV-B and wounding were also reported to possess a minor inhibitory and stimulating role on the transcription of *At3g05980* in shoot (Table 5-2c).

5.3.1.5 Co-expression analysis

Co-expressed genes might be involved in similar or related biological processes. I used ATTED-II to identify 300 genes that are co-expressed with *At3g05980* (Obayashi et al., 2007). Gene ontology (GO) enrichment analysis of these genes indicated that cell differentiation, carbohydrate metabolism and lipid metabolism-related genes and genes with catalytic activity and transferase activity were overrepresented among these (Figure 5-5; Obayashi et al., 2009).

5.3.1.6 *Cis-element* prediction

I explored *cis*-elements in the entire upstream intergenic region upstream of *At3g05980* in three commonly used plant cis-acting regulatory elements databases (PLACE, Plant CARE and AGRIS) and compared the frequency of each cis-element in the $At3g05980_{PRO}$ to its frequency in the whole *Arabidopsis* genome (Higo, 1998; Rombauts et al., 1999; Yilmaz et al., 2011; The Arabidopsis Genome Initiative, 2000). Based on this analysis, I found 26 cis-elements that were significantly enriched (p-value ≤ 0.05) in the upstream intergenic region of *At3g05980* and many of them were involved in the abiotic stress, including five related to dehydration and cold (DRE; MYCATERD1; MYCATRD22; ATHB2_BINDING_SITE_MOTIF; ABRELATERD1, two related to salt stress (GT1GMSCAM4; ATHB2_BINDING_SITE_MOTIF), one associated with wounding (QARBNEXTA; Table 5-3; Yamaguchi-Shinozaki, 1994; Chen et al., 2002; Simpson et al., 2003; Abe et al., 1997; Sessa et al., 1993; Yoo et al., 2010; Elliott & Shirsat, 1998). Moreover, some cis-elements involved in the hormone signaling were revealed to be more abundant in this promoter, such as ARFAT (involved in auxin response), ABRE (involved in ABA response), GADOWNAT (involved in GA response) and QARBNEXTA (involved in JA response; Ulmasov et al., 1999; Hobo et al., 1999; Huang et al., 2008; Elliott & Shirsat, 1998). Notably, four light responsive cis-elements including BOX_4, CCA1ATLHCB1, ATC-MOTIF, G-BOX (CUF1) element were also enriched in this region (Table 5-3; Wang, 1997; Xie et al., 2003; Kawagoe et al., 1994). Two root expression associated (SP8BFIBSP8BIB, ROOTMOTIFTAPOX1) and one endosperm expression related cis-element (AACACOREOSGLUB1) were also enriched in this region (Table 5-3; Ishiguro & Nakamura, 1992; Elmayan & Tepfer, 1995; Wu et al., 2000).

5.3.1.7 Protein structure and function prediction

The function of a protein is determined largely by its sequence and three-dimensional (3D) structure. To predict a possible structure of At3g05980, I used the iterative threading assembly refinement (I-TASSER; Roy et al., 2010). Based on the secondary structure predicted by I-TASSER server, At3g05980 comprised five α-helices with two β-strand, while a large proportion of this protein was predicted to be coil (Figure 5-6). The top-scoring model of At3g05980 created by I-TASSER were shown in the Figure 5-6, but this was not judged to be significant, since it had a confidence score (C-score) of -4.17, TM-score of 0.27 and RMSD value of 16 Å (Figure 5-6). For comparison, a C-score of -1.5 or higher is expected to produce the correct topology 90% of the time and the TM-score > 0.5 usually has an accurate fold (Roy et al., 2010). To effectively predict the ligand-binding site and functional important residues, RMSD value of a model needs to be in the range of 1–2 Å and 2–5 Å respectively (Roy et al., 2009). I have also predicted the 3D structure of At3g05980 using Phyre2 (Kelley et al., 2015), but the best prediction had an overall low confidence, and was not considered to be relevant to further analysis (data not shown).

5.3.2 Tissue expression pattern analysis

5.3.2.1 qRT-PCR

I used qRT-PCR to measure the abundance of At3g05980 transcripts in various tissues, including 4 DAS seedlings, 7 DAS seedlings, 14 DAS shoot apices and 21 DAS inflorescence apices, roots, rosette leaves, cauline leaves, stems, siliques as well as flowers at stages 12, 13, 14, and 15/16. Transcripts of this gene were expressed in all of the surveyed tissues and showed highest expression in the inflorescence apices dissected from 23 DAS *Arabidopsis* plants. Transcripts were also highly abundant in 14 DAS shoot apices, flowers at each of the stages tested, and siliques. In these examined flower stages, *At3g05980* had its highest transcript abundance in flowers at stage

13. I attempted to check the transcript abundance of this gene in the vegetative shoot apex from 7 DAS seedlings, but I failed in RNA extraction. However, the transcript levels of *At3g05980* were low in 4 DAS seedlings, 7 DAS seedlings, roots, rosette leaves, cauline leaves and stems (Figure 5-7).

5.3.2.2 Promoter-GUS fusion study

I examined the expression pattern of At3g05980 during plant development using a promoter-GUS reporter fusion. The entire 2,799 bp upstream intergenic region of At3g05980 was fused to the β-Glucuronidase (GUS) reporter gene and was transformed into wild-type Arabidopsis plants (Col-0). Histochemical analysis was analyzed in the T_2 transgenic plants of 29 transgenic lines. At least ten individuals from each line were examined and patterns representing most of the observed individuals were presented here. In seedlings, GUS activity was only detected in hydathodes, stipules and roots. Within roots, GUS activity was detected at the tip of the radicle immediately after germination (Figure 5-8). In 2 DAS and 3 DAS seedlings, GUS activity was observed in the root cap, elongation zone and artrichoblast cells in the maturation zone, but not the root apical meristem (Figure 5-8). In 4 DAS seedlings, GUS maintained its expression in the elongation zone and maturation zone but disappeared from the root cap. Meanwhile, the distal part of the meristematic zone was also stained (Figure 5-8). By 8 DAS, GUS expression was only detected in the meristematic zone and artrichoblast cells in the elongation zone of the primary root tip. Meanwhile, GUS activity was also observed in the lateral root primordia and elongating lateral root (Figure 5-8). In flowers, GUS activity was only observed in the petals and filaments of the opening flower (Figure 5-9). Meanwhile, I observed GUS staining in the embryos proper and suspensor at globular stage embryos as well as in the micropylar endosperm of the mature green embryos (Figure 5-10). However, I have not detected staining in embryos at other stages.

5.3.3 Responses of *At3g05980* to plant hormones

I used qRT-PCR to measure abundance of At3g05980 transcripts in response to exogenous application of several hormones (ABA, IAA, GA_3, BA, BR, ACC, MeJA). Seedlings of 7 DAS wild-type *Arabidopsis* plants were incubated in liquid ½ X MS media supplemented with each of these hormones at 22°C (16h light/ 8h dark) and I measured transcript expression of *At3g05980* after incubation for 0 h, 1 h, 3 h, 6 h, 12 h and 24 h. I used concentrations of each hormone within ranges typically used in similar experiments in the literature (Austin et al., 2016; Okushima et al., 2005; Armstrong et al., 2004; Yang et al., 2017; Zhang et al., 2014; Ruzicka et al., 2009). Transcription of this gene was inhibited by ABA 24 h after treatment but was not significantly affected by any of the other hormones applied (Figure 5-11).

5.3.4 Response of *At3g05980* to abiotic stresses

I have analyzed the expression of *At3g05980* gene in response to salt (200 mM NaCl), osmotic (300 mM mannitol) and cold stress (4°C) by qRT-PCR. For salt and osmotic stress, seedlings of 7 DAS wild-type *Arabidopsis* plants were incubated in liquid MS media supplemented with 200 mM NaCl or 300 mM mannitol at 22°C (16h light/ 8h dark). For cold treatment, 7 DAS seedlings were incubated in a 4°C with continuous light. Then I measured transcript expression of *At3g05980* under all three abiotic stresses after incubation for 0 h, 1 h, 3 h, 6 h, 12 h and 24 h. I found that the expression of this gene was significantly induced by salt and cold, although NaCl only altered its expression slightly (Figure 5-12). Following cold treatment, the expression level of *At3g05980* enhanced rapidly and reached a peak at 6 h to 13- fold and then reduced gradually to 6-fold at 24 h.

5.3.5 Subcellular localization of At3g05980

The identification of the native compartment of a protein is important for understanding its role. I examined the subcellular localization of At3g05980 protein in Arabidopsis roots and root hairs by fusing the coding sequence of At3g05980 protein in-frame to the C-terminus of GFP (green fluorescent protein) in the pCsGFPBT vector or the N-terminus of CiFP (citrine fluorescent protein) in the pCAMBIA 1303 vector. Both constructs were expressed under control of the cauliflower mosaic virus (CaMV) 35S promoter. *A. tumifaciens* carrying the 35S:: GFP: At3g05980 fusion construct or the 35S:: At3g05980: CiFP fusion construct were used to infiltrate the flowers of wild-type Col-0 plants, with unfused, 35S:: GFP or 35S:: CiFP infiltrated in parallel as controls. T_3 generation progeny (n=10) of three independent transformants of each construct were examined using fluorescence microscopy. As expected, uniformly distributed green fluorescence and citrine fluorescence were observed in cells expressing 35S:: GFP and 35S:: CiFP constructs respectively (data not shown). However, no fluorescence signal was detected in the transgenic plants expressing 35S:: At3g05980: CiFP fusion construct. In contrast, a punctate fluorescence pattern was found in 35S:: GFP: At3g05980 transgenic plants, and it appeared that the organelle labeled was small and round (Figure 5-13). The morphology and size of this labeled organelle was consistent with the peroxisome (Muench & Mullen, 2003). Therefore, I co-expressed a peroxisome marker construct mCherry: PST1 in T_3 homozygous plants of 35S:: GFP: At3g05980 transgenic lines. Confocal laser scanning microscopy of root tip observation suggested that the GFP: At3g05980 fusion was consistently co-localized with the mCherry: PST1 peroxisome marker (Figure 5-13; Nelson et al., 2007).

5.3.6 Functional genetic analysis of At3g05980

5.3.6.1 Morphology of At3g05980 overexpression lines

I created transgenic plants that expressed At3g05980 under control of the constitutive 35S promoter. T3 plants (n=12) of three different lines were studied, in which transcript expression of At3g05980 gene had increased from 62- to 130-fold compared to wild-type (Figure 5-14). The overexpression lines exhibited some changes in morphology, including epinasty of cotyledons and leaves, shorter plant height, shorter silique length as well as abnormal silique morphology (Figure 5-15). These morphological defects were also seen in the 35S:: At3g05980: GFP and 35S:: CiFP: At3g05980 transgenic plants (data not shown).

5.3.6.2 Create *At3g05980* mutants by the CRISPR/Cas9 system

I have analyzed the two T-DNA insertional mutant lines of At3g05980 obtained from ABRC. According to the information found in TAIR, the mutant line SAIL_1054_G02 and SALK_024489 were predicted to have T-DNA inserted in the 460bp and 622 bp upstream of the At3g05980 start codon, respectively. Genotyping showed that plants in SAIL_1054_G02 did not contain a T-DNA insertion while the transcript level analysis of the homozygous SALK_024489 plants by RT-PCR indicated that this line was an overexpression line even though the homozygous SALK_024489 plants did not show the morphological defects as shown in the 35S:: At3g05980 overexpression lines I described in the section 5.3.5.1. (Figure 5-16).

Because none of the available insertion lines had verifiable loss-of-function for At3g05980, I used the CRISPR/Cas9 system to generate loss-of-function mutants for At3g05980. Two constructs were created and each contained two sgRNAs. The construct one had two sgRNAs which targeted the coding sequence 36 -59 bp from the start codon and 114-137 bp from the start codon

respectively. Similarly, the construct two had two sgRNAs which targeted the coding sequence 63 -86 bp from the start codon and 103-126 bp from the start codon respectively. I obtained heritable homozygous single mutants among T_2 progeny, and used Sanger sequencing to verify the disruptions in their coding regions (Figure 5-12). Three multiple alleles were analyzed.These mutants (designated as *At3g05980- CR* in this thesis) did not show any discernible morphological or growth defects compared to the WT plants.

5.3.6.3 Freezing assay and electrolyte leakage assay

Because *At3g05980* transcripts were found to be strongly induced by cold treatment (Table 5-2c; Figure 5-12), I performed freezing sensitivity assays and electrolyte leakage assays using CRISPR-Cas9 loss-of-function mutants (*At3g05980-CR*), as well as 35S::At3g05980 transgenic plants. It is known that exposure to chilling (0–15 °C) and non-freezing temperature can increase the freezing tolerance of plants such as Arabidopsis that evolved in temperate climates. This process is called cold acclimation (Thomashow, 1999). I measured the responsiveness of plants to cold under both cold acclimated (CA) and nonacclimated (NA) conditions. For the nonacclimated (NA) freezing assay, plants were grown at 22 °C with 16 h of light until 17 DAS, at which point they were directly subjected to freezing treatment. For the cold-acclimated (CA) freezing assay, plants were grown at 22 °C with 16 h light until 14 DAS and then grown in 4°C cold chamber with 16 h light for three days before being subjected to the freezing treatment. The freezing treatment for both NA and CA assay were conducted as follows: plants were maintained under at 0 °C for 1 h, and temperatures were then dropped by 1 °C /h until -5°C or -6°C for NA assay and -10°C for CA. These temperatures were chosen based on Jiang's report (Jiang et al., 2017). After freezing, I counted the seedling survival rate and checked the electrolyte leakage rate, which are indicators of the cell membrane damage under stress. We found that under both NA and CA conditions, the survival

rate and the electrolyte leakage rate of the mutants were not significantly different from WT plants (Figure 5-18; 5-19).

5.3.6.4 Expression of stress-responsible genes in *At3g05980*

Although overall freezing tolerance was not changed in the At3g05980 mutant, it was possible that cold-related signaling pathways were altered. The ICE1-CBFs-COR (cold-regulated gene) signaling pathway is the most important and best characterized cold signaling pathway in plants (Chinnusamy et al., 2007). Three CBF genes (*CBF1*, *CBF2* and *CBF3*) are encoded in Arabidopsis and play an important role in the cold-responsive network by binding to CRT/DRE cis-elements (A/GCCGAC) in the promoters of COR genes and regulating their expression (Maruyama et al., 2004). A recent transcriptome study indicated that mutation of CBFs significantly altered the expression of over 3000 CORs under cold treatment (Shi et al., 2017). ICE1 encodes a MYC transcription factor and activates the transcription of CBF through binding to the MYC recognition cis-elements (CANNTG) in their promoter (Chinnusamy et al., 2003). To check if the involvement of At3g05980 in cold at the molecular level, I compared the expression level of six cold-responsive genes (including *CBF1, CBF2 CBF3, RD29A, KIN1, AND COR47*) in the *At3g05980-CR* mutant and WT seedlings under cold by qRT-PCR (Figure 5-20). Expression of all the marker genes tested was induced in WT after cold treatment, consistent with the previous studies (Figure 5-20; Kurkela & Franck, 1990; Gilmour et al., 1998; Yamaguchi-Shinozaki & Shinozaki, 1993; Gilmour et al., 1992). Meanwhile, these genes were also induced in the *At3g05980-CR* mutant, although the induction levels of some genes were slightly (but significantly) changed compared with those in the wild type (Figure 5-20). I found that *CBF1* and *CBF3* showed higher expression in *At3g05980-CR* compared to WT after undergoing cold treatment for 48 h (Figure 5-20). Meanwhile, the expression of *RD29A* was significantly increased in the *At3g05980-CR* mutant (compared to wild-

type) under both normal conditions and cold (Figure 5-20). However, I found that induction of *KIN1*, *COR47,* as well as *CBF2* in the *At3g05980-CR* mutant was comparable to that in the wild-type plants (Figure 5-20).

5.3.6.5 Assays of peroxisome function

Peroxisomes are primarily associated with β-oxidation of fatty acids in plants, an essential process to convert stored fatty acids into sucrose, especially during early seedling establishment (Graham, 2008). Mutants with compromised fatty acid oxidation have short hypocotyls when grown in the dark in the absence of sucrose (Baker et al., 2006). Furthermore, 2,4-DB (2,4-dichlorophenoxybutyric acid) and IBA (indole-3-butyric acid), two auxin analogues have been used to select for defects in fatty acid β-oxidation as well. The genotypes with compromised fatty acid β-oxidation are resistant to the inhibitory effect of exogenous 2,4-DB and IBA on growth (Zolman et al., 2000; Hayashi et al., 1998). The CTS gene encodes a peroxisomal ATP binding cassette (ABC) transporter protein and transports the fatty acid into peroxisome (Russell et al., 2000). *cts* mutants have reduced germination potential and are resistant to auxin analogues 2,4-DB and IBA (Footitt et al., 2002; Zolman et al., 2001b; Hayashi, 2002). From the publicly available microarray data, I noted that the transcript abundance of At3g05980 was increased in the *cts* mutant (Waese et al., 2017).

Having shown that At3g05980 is localized in the peroxisome (Figure 5-13), and that *At3g05980* transcript abundance is affected in peroxisomal mutants, I was interested in learning whether mutants of *At3g05980* showed defects under dark growth, or when treated with 2,4-DB or IBA. I observed that hypocotyl growth of the *At3g05980-CR* and 35S: At3g05980 on sucrose-free medium in the dark and found no measurable changes compared to WT (Figure 5-21). Furthermore,

no significant difference in the sensitivity of *At3g05980-CR* and 35S: At3g05980 to 2,4-D and IBA was detected (Figure 5-22).

Both microarray and GUS assays indicated At3g05980 transcript abundance was enriched in developing seeds (Figure 5-10; Waese et al., 2017). Co-expression analysis also predicted that this gene might be involved in lipid metabolism (Figure 5-5). To test whether At3g05980 might affect the composition of lipids in mature embryos, I analyzed the fatty acid profile of dry, mature seeds of *At3g05980-CR* and found no significant difference compared to WT seeds (Figure 5-23).

5.4 Discussion

Lus10041215 was an uncharacterized flax gene that was expressed 53 times more in the shoot apex compared to the remainder of the stem, and had no functional annotation (Zhang & Deyholos, 2016). This motivated me to characterize one of its Arabidopsis homologs, *At3g05980*. I found that the *At3g05980* gene family was restricted to eudicots and encoded predicted proteins normally containing four uncharacterized conserved motifs (Figure 5-2; Appendix 12). Its homologs were detected in all the sequenced eudicot genomes published in Phytozome v12.1. This suggests that it encodes a function required in eudicots.

5.4.1 Tissue-specific expression patterns

I checked the expression of *At3g05980* by qRT-PCR and found that *At3g05980* was transcribed in all the tested tissues but had higher transcript abundance in shoot apex, unopened flower buds, flowers and siliques (Figure 5-7). This pattern was generally consistent with what we found from the microarray data in the electronic fluorescent pictograph (e-FP) which suggested that

At3g05980 was expressed ubiquitously in a great range of tissues and it had highest expression level in the petals, shoot apex and developing seeds (Figure 5-4a; Schmid et al., 2005).

I also characterized the tissue-specific expression profiles of At3g05980 through promoter-GUS fusions and again this gene was expressed in the flowers, including petals and filaments (Figure 5-10). Furthermore, the promoter of this gene also derived the GUS expression in the globular embryo as well as the micropylar endosperm of the mature green embryos (Figure 5-11). This was partially consistent with the findings in the eFP Brower, which showed that *At3g05980* had a relatively high expression level in the embryo at the linear cotyledon stage and the micropylar endosperm of the mature green embryos (Waese et al., 2017). Furthermore, both the eFP-Brower and promoter-GUS assay indicated that this gene was strongly expressed in the root elongation zone and the artichoblast cells in the maturation zone of root (Table 5-2a; Figure 5-9). Also, the microarray data obtained from the GENEVESTIGATOR indicated that At3g05980 gene was most abundant in the root atrichoblast, petals and lateral root caps (Figure 5-5b). However, I found most GUS transgenic lines analyzed did not show GUS staining in the shoot apex and flower buds, even though both microarray and qRT-PCR study indicated that At3g05980 had high transcript abundance in the shoot apex (Figure 5-9; 5-10).

The differences between the GUS pattern and the other expression data suggested that not all of the cis-elements required for the native expression of At3g05980 were included in the fragment cloned upstream of the GUS reporter. It had been reported that the cis-elements regulating a gene's transcription may also be located downstream or even within the transcribed region (Kertész et al., 2006; Barrett et al., 2012). Discrepancies between qRT-PCR and GUS analysis may also result from the differences in the sensitivity of these two techniques.

5.4.2 Morphology of At3g05980 overexpression lines

To elucidate the function of At3g05980, I ectopically expressed this gene in WT plants under the CaMV 35S promoter and the overexpression lines showed some morphological differences in cotyledon shape, leaf shape, silique morphology and plant height compared to the WT plants (Figure 5-10). Additionally, I found similar phenotypes in the plants expressing 35S:: GFP: At3g05980 (data not shown). Based on these phenotypes, it appeared that gain-of-function mutation of this gene might have either changed the cell proliferation or cell expansion rate. At3g05980 was indeed highly expressed in the tissues with active cell proliferation (e.g: embryo, shoot apex) or cell expansion (e.g: petal, the elongation zone of root, filament). However, the observed morphological defects were not consistent with their expression patterns. The promoter-GUS assay showed that this gene was not expressed in leaves and cotyledons (data not shown). Also, both the public available microarray data and our qRT-PCR analysis showed that this gene had a very low expression level in leaf and cotyledon (Figure 5-5a; 5-8). In this study, a constitutive promoter (the CaMV 35S) was utilized. Ectopic overexpression in this way may confer novel activity on a particular protein or cause a protein with normal activity to be expressed in the wrong tissues or at an inappropriate time. A T-DNA insertional mutant line of *At3g05980* (SALK_024489) proved to be an overexpression lines but no morphological defects was found (Figure 5-17). This might be due to the difference in transcript levels: The expression level of At3g05980 was increased around five times in SALK_024489 whereas in the overexpression lines it was elevated 62 to 130 times (Figure 5-17; 5-15).

By comparison, the loss-of-function mutant of At3g05980 did not show any discernable morphological defects. This may be due to the functional redundancy with the Arabidopsis paralog, At5g19340. These two proteins shared 76.3% sequence similarity and 66.1% sequence identity.

Microarray-based expression profiles in the eFP Brower indicated that both At3g05980 and At5g19340 showed relatively weak expression throughout plant organs and developmental stages and these two genes showed very similar expression patterns, therefore it was possible that they might have overlapping functions (Waese et al., 2017).

5.4.3 The At3g05980 in cold stress

Both microarray and qRT-PCR studies indicated that At3g05980 was cold-induced (Table 5-2c; Figure 5-13). Data the GENEVESTIGATOR indicated that *At3g05980* gene was significantly upregulated in the ICE1 mutant (Zimmermann et al., 2005). ICE1 was an important regulator of cold-induced transcriptome and freezing tolerance (Chinnusamy et al., 2003). Meanwhile, four most common cold-related cis-elements (DRE-LIKE_PROMOTER_MOTIF; ABRE; MYCATERD1; MYCATRD22; G-box) were overrepresented in the promoter of At3g05980 gene (Figure 5-3). Therefore, I assumed that the At3g05980 gene was involved in cold stress. I first checked the freezing tolerance of At3g05980 overexpression lines and loss-of-function mutants under both cold acclimated and non-acclimated condition, but both of them showed comparable level of freezing tolerance as the WT plants (Figure 5-19; 5-20). Then I checked the expressions of several cold-regulated genes in the At3g05980 loss-of-function mutants. As a result, expression of three cold-regulated genes were found to be significantly upregulated in the mutant compared to the WT (Figure 5-22a). Two of them encoded important regulators of cold stress, *CBF1* and *CBF3*. The other gene induced was *RD29A,* a COR gene directly regulated by *CBFs* (Liu, 1998). We noted that *RD29A* was induced in the *At3g05980* mutant even in an unstressed environment, and this phenomenon has been reported before: both *CBF1* and *RD29A* were induced in 35S: CBF1 transgenic plants under unstressed condition, and these plants also showed a dwarf phenotype under normal conditions (Liu, 1998). However, we noted that both *CBF1* and *CBF3* were

significantly upregulated in the *Ag3g05980* mutant only after 48 h cold treatment, this suggested that the induction of *RD29A* in the At3g05980 was not induced by the upregulating of *CBF1* and *CBF3*. Although *CBFs* were suggested to regulate the transcription of *RD29A*, mutations that either upregulated or downregulated *RD29A* expression without altering the CBF transcription level have been reported, whereas some other mutants were revealed to have *CBF* expression level changed but not *RD29A* (Zhu et al., 2005; Lee, 2002; Hojoung et al., 2002; Dong et al., 2009). The *RD29* gene was reported to be induced not only by cold but also by drought, osmotic stress, high salt and ABA (Yamaguchi-Shinozaki & Shinozaki, 1993). Both microarray and qRT-PCR analysis revealed that *At3g05980* expression was reduced by the exogenous application of ABA (Table 5-2b; Figure 5-12). We noted that the ABA response cis-element (ABRE) was also significantly enriched in the At3g05980 promoter (Table 5-3). These indicated that At305980 might be involved in the CBF-dependent cold signaling pathway or ABA-dependent cold signaling pathway but it only affected the expression of cold signaling components to a small extent and this effect was not dramatic enough to alter the freezing tolerance. These two signaling pathways were found to be not completely independent (Knight et al., 2004). *At3g05980* was upregulated by cold after 3 h treatment and it was proposed that genes upregulated earlier after cold might be encode transcription factors or components required for signaling in response to cold or for chilling tolerance (Figure 5-12; Knight & Knight, 2012).

5.4.4 The At3g05980 and fatty acid β-oxidation

This study localized the At3g05980 protein in the peroxisome through translational fusion with the GFP (Figure 5-14). Prediction through three common used approaches, PSORT, PROSITE and PeoxiP suggested that this protein was not a PTS1-containing protein. Possible explanations are as follows: These PTS1 prediction methods had limitations and they were either restrictive missing

known peroxisomal protein or rather permissive with too many false positive results (Brocard & Hartig, 2006). Meanwhile, all these methods are based on experimentally verified peroxisomal proteins, which just represent a limited set and they were reported to fail to recognize some unusual but verified targeting signal (Lametschwandtner et al., 1998; Kragler et al., 1998). The peroxisome prediction by PSORT and PROSITE are only based on the C-terminal tripeptide of submitted protein which may return incorrect results (Geraghty et al., 1999). Peroxisomes were first characterized in mammalian tissues in the 1960s and their first discovered function was cleaning up the peroxide produced by other organelles (Duve & Baudhuin, 1966). Now, the function of peroxisomes have been found to extend far beyond reactive oxygen metabolism, with roles in processes including fatty acid β-oxidation, the glyoxylate cycle, detoxification, photorespiration, primary carbon metabolism, secondary metabolism, development, biosynthesis of salicylic acid, biotic and abiotic stress (reviewed in Olsen, 1998; Hu et al., 2012). Peroxisome mutants were revealed to have seedling establishment limitations due to impaired seed storage oil utilization during germination (Zolman et al., 2000). Forward genetic screens revealed that the naturally occurring auxin IBA was converted to the active IAA (the principal form of auxin) in peroxisomes and this process was critical for lateral root formation in developing seedlings (Zolman et al., 2000; Zolman & Bartel, 2004). The hormone JA biosynthesis was suggested to require three rounds of peroxisomal β-oxidation and a peroxisome biogenesis protein has been discovered to affect photomorphogenesis (Creelman & Mullet, 1997; Hu, 2002). Peroxisomes were also revealed to be involved in the metabolism of the branched-chain amino acids, propionate and isobutyrate (Zolman et al., 2001a; Lucas et al., 2007). In this study, I found mutation of *At3g05980* gene did not change the fatty acid profiles in the seed and the sensitivity of plants to the inhibitory effects of exogenous 2,4-D and IBA on root elongation (Figure 5-24). Meanwhile, the mutant did not

exhibit any developmental defects in the absence of exogenous sucrose (Figure 5-23). All these observations indicated the *At3g05980* gene does not function in peroxisomal β-oxidation. However, the lack of phenotype is possibly also due to functional redundancy. Alternatively, this gene might be involved in other peroxisomal processes. For example, peroxisomal metabolism has been shown to play a role in cold stress signaling as well as plant tolerance to cold stress (Dong et al., 2009).

5.5 Conclusions

At3g05980 encodes an unknown Arabidopsis gene. This gene is present in eudicots only and contains four conserved motifs. In silico analysis suggested that many cold-related cis-elements were overrepresented in its promoter. GO enrichment of the predicted co-expression genes indicated that this gene might be related to the cell differentiation, carbohydrate metabolism and lipid metabolism-related and it might have catalytic activity and transferase activity. Expression profiling conducted in this study indicated that *At3g05980* is highly expressed in the petals, shoot apex, roots of young seedlings as well as embryos at the globular stage and the micropylar endosperm at the mature green stage. Protein of At3g05980 was targeted to the peroxisome. Overexpression of At3g05980 showed morphological defects in leaf shape, cotyledon shape, silique morphology, short silique and short plant height whereas mutation of this gene did not induce any discernable phenotype. Meanwhile, loss-of-function mutation of *At3g05980* had no impact on the freezing tolerance of Arabidopsis but slightly altered the expression of some cold-regulated genes. Furthermore, At3g05980 has no effect on peroxisomal β-oxidation capacity.

5.6 Figures and tables

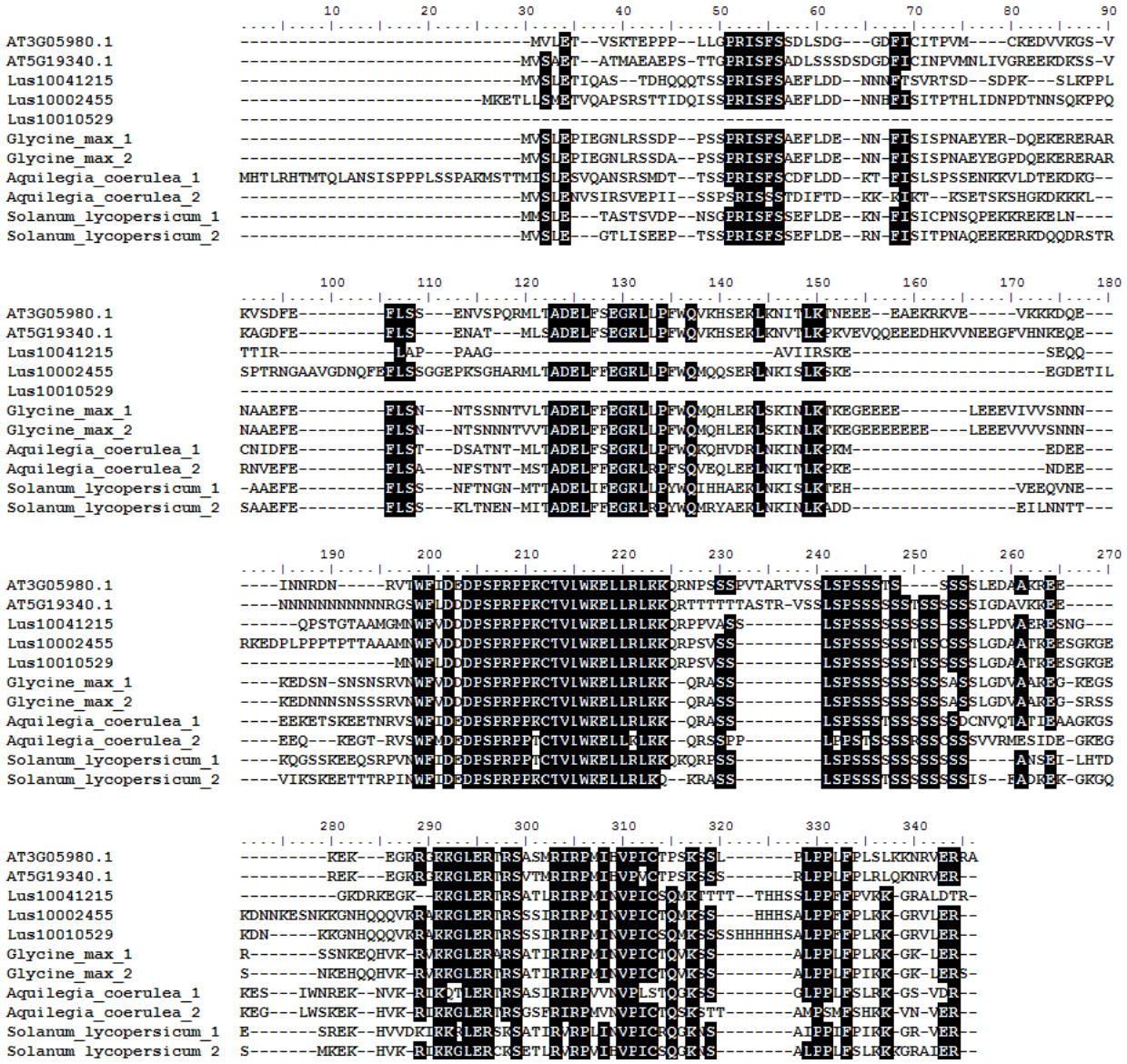


Figure 5-1 Multiple sequences alignments of At3g05980 homologs from several plant species. Residues with > 75% identity were shadowed.

Table 5-1 Subcellular localization of At3g05980 predicted by several commonly used programs (Horton et al., 2007; Chou & Shen, 2010; Emanuelsson et al., 2007; Blum et al., 2009; Hawkins & Bodén, 2006; Briesemeister et al., 2010).

Tools	**Predicted Localizations**
PSORT	cytoplasm
WoLF PSORT	chloroplast
Plant-mPLoc	nucleus
TargetP	not in mitochondrial or chloroplast
MultiLoc2	cytoplasm
SUBA4	nucleus
YLoc	nucleus

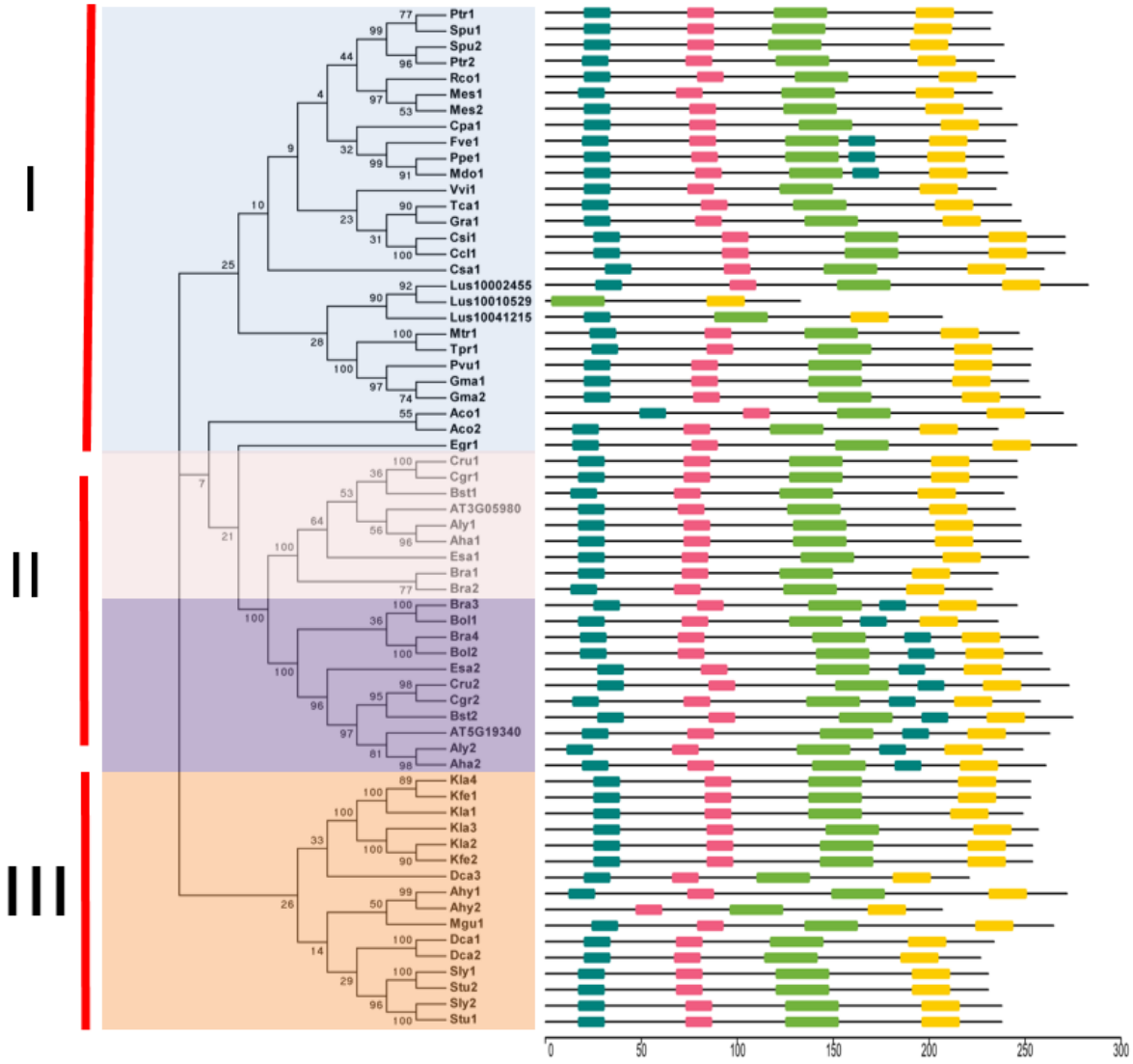


Figure 5-2 The unrooted phylogenetic dendrogram of At3g05980 and its homologs identified from Phytozome v12.1 as well as motifs discovered in At3g05980. Deduced amino acid sequences were aligned with MAFFT (Katoh et al., 2009). The phylogenetic dendrogram was created using the Neighbor-joining method, following the Dayhoff model of amino acid substitutions (Grishin, 1995). The numbers at the branch points represented bootstrap values. The full species names were listed in the Appendix 13. The discovered conserved motifs are displayed on the right-hand side as different colored boxes.

Figure 5-3 Sequence logos of the discovered motifs in At3g05980 and its homologs. A: motif1; B: motif2; C: motif3; D: motif4.

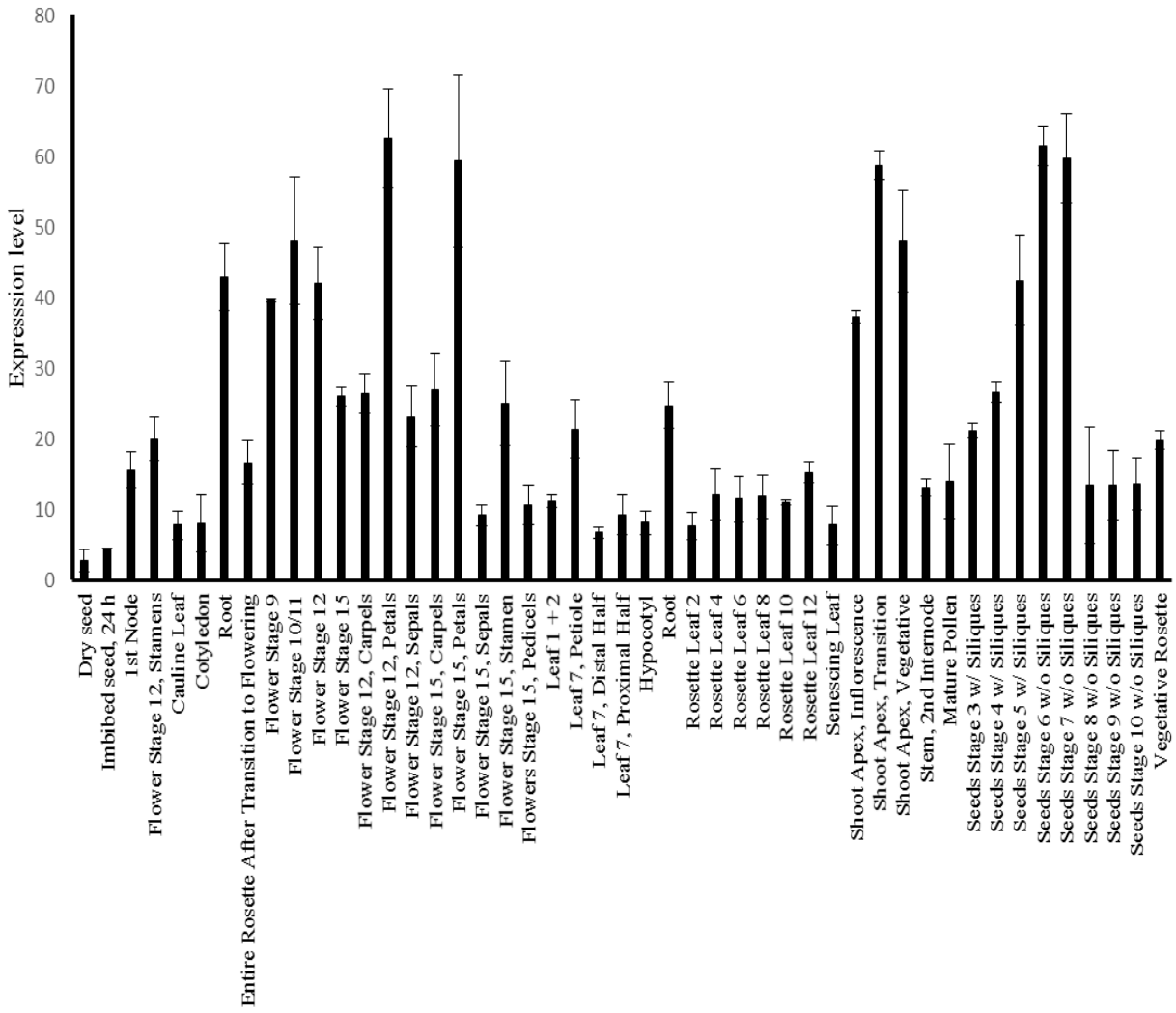


Figure 5-4a Microarray-derived expression profiles of At3g05980 gene across various tissues. Data were retrieved from The Bio-Analytic Resource for Plant Biology (http://bar.utoronto.ca/welcome.htm; Waese et al., 2017).

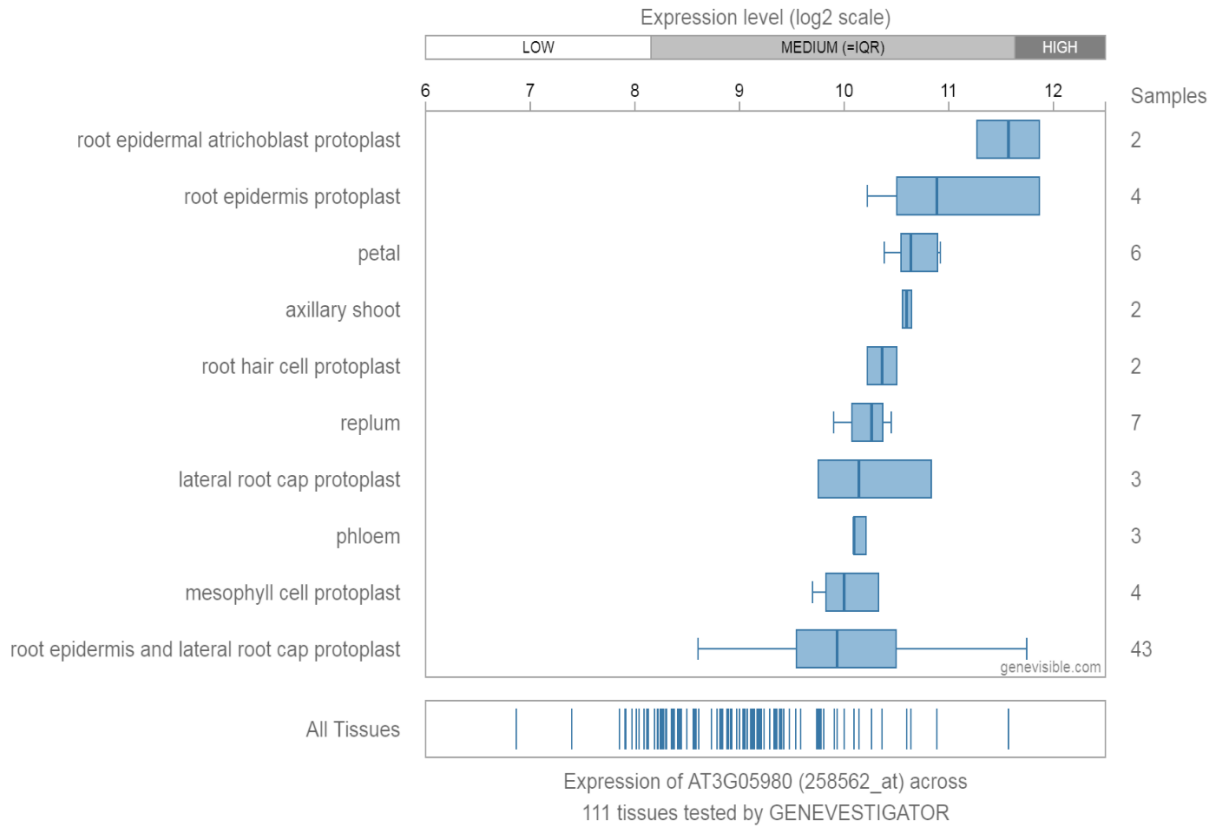


Figure 5-4b Expression profiles of At3g05980 gene across 111 various tissues obtained from Genevisible (https://genevisible.com/search; Hruz et al., 2008).

Table 5-2a Tissue specific expression pattern of At3g05980 obtained from eFP Broswer (Waese et al., 2017).

Tissues	Absolute expression level
Root Stage I Cortex + Endodermis	30.65
Root Stage I Epidermal Artrichoblasts	110.1
Root Stage I Stele	21.16
Root Stage II Cortex + Endodermis	35.34
Root Stage II Epidermal Artrichoblasts	126.96
Root Stage II Stele	24.4
Root Stage III Cortex + Endodermis	44.09
Root Stage III Epidermal Artrichoblasts	158.3
Root Stage III Stele	30.43

Table 5-2b Transcript level changes of At3g05980 in response to exogenous hormones application. Data were obtained through microarray analysis and extracted from the eFP Brower (Waese et al., 2017). The time point with significantly reduced (<0.5 fold) expression is highlighted in green,

Treatment	Fold change (relative to the Mock treatment)		
	0.5 h	1 h	3 h
10uM ACC	1.22	1.97	0.7
1uM IAA	0.92	1.01	0.53
10uM ABA	1.41	1.2	0.28
10uM MeJA	0.92	1.46	0.67
1uM GA-3	0.91	1.56	0.58
10uM BL	1.45	1.64	0.84

Table 5-2c Transcript level of At3g05980 in Arabidopsis shoot and root responding to various abiotic stresses. Data were obtained through microarray analysis and extracted from the eFP Brower (Waese et al., 2017). The time point with significant reduced (<0.5 fold) or upregulated expression level (>2 fold) were highlighted with green and red respectively.

		Fold change (relative to the Mock treatment)					
	Treatment	0.5 h	1 h	3 h	6 h	12 h	24 h
Shoot	Cold	0.91	1.06	0.66	2.09	5.05	2.43
	Osmotic	1.3	0.86	0.55	0.78	0.49	1.17
	Salt	1.51	1.1	0.77	1.1	0.8	0.82
	Oxidative	1.12	1.21	0.57	0.69	0.97	0.84
	UV-B	1.03	1.54	0.72	0.35	0.73	0.85
	Wounding	2.02	1.06	0.94	0.76	0.96	1.46
	Drought	1.32	1.28	0.87	0.85	0.66	0.85
	Heat	1.45	0.32	0.82	0.98	0.83	0.79
Root	Cold	1.33	1.39	1	1.16	1.03	0.54
	Osmotic	2.46	2.1	1.14	0.79	0.52	0.56
	Salt	1.65	1.13	0.76	0.6	0.66	0.59
	Oxidative	1.04	0.71	1.22	1.17	0.84	0.51
	UV-B	0.98	0.67	1.07	0.92	0.89	0.73
	Wounding	0.83	0.83	0.95	0.73	0.95	0.68
	Drought	1.51	1.27	1.3	0.78	0.77	0.63
	Heat	0.96	0.87	0.84	0.78	0.53	0.52

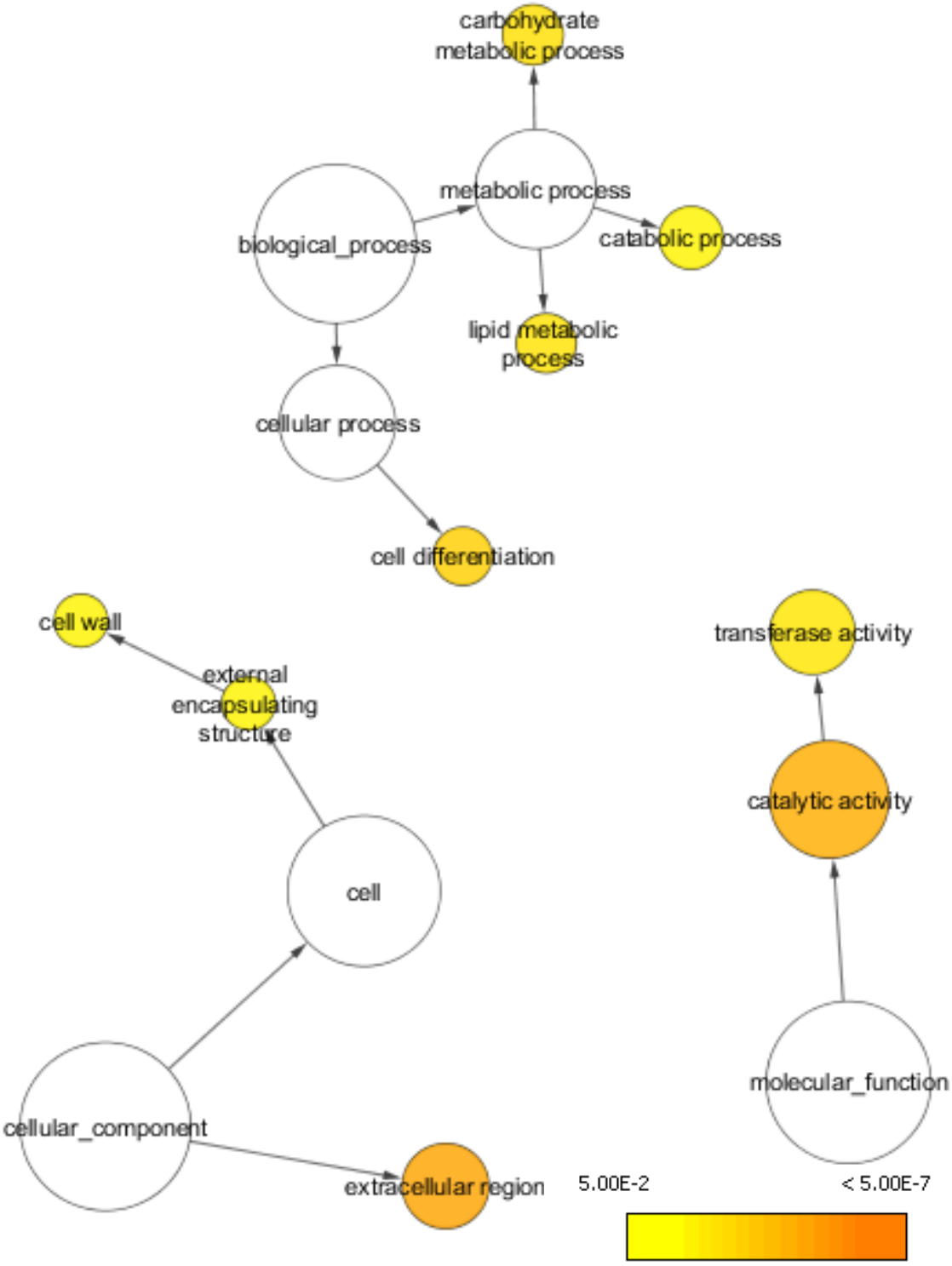


Figure 5-5 Gene Ontology (GO) enrichment of the 300 Arabidopsis genes predicted to be co-expressed with At3g05980 by ATTED-II (Obayashi et al., 2009).GO enrichment was conducted by the Bingo application in Cytoscape v 3.5.1 (Shannon et al., 2003). Significantly enriched GO slim categories were highlighted with different colors representing different levels of significance. The size of each circle is correlated to the number of genes.

Table 5-3 *Cis* elements overrepresented in the promoter of At3g05980 (*p*-value ≤0.05).

Cis element	Sequence	*p*-value	Description
DRE-LIKE_PROMOTER_MOTIF	TGCCGACAT	0	drought and cold response elements
BP5OSWX	CAACGTG	2.67238 E-07	OsBP-5 (a MYC protein) binding site in WX promoter
ZDNAFORMINGATCAB1	ATACGTGT	2.18E-05	Z-DNA-forming sequence' found in the Arabidopsis chlorophyll a/b binding protein gene (cab1) promoter
CCA1_BINDING_SITE_MOTIF	AAAAATCT	0.001	specify circadian phase;rhythmic transcription
MYCATERD1	CATGTG	0.002	MYC recognization sequence for expression of erd1 (early responsive to dehydration)
MYCATRD22	CACATG	0.002	Binding site for MYC(rd22BP1) in Arabidopsis dehydration responsive gene (rd22)
SP8BFIBSP8BIB	TACTATT	0.002	root-specific responsive elements; one of SPBF binding site (SP8b) sporamin
BOX_4	ATTAAT	0.003	part of a conserved DNA module involved in light responsiveness
ARFAT	TGTCTC	0.005	ARF (auxin response factor) binding site in the promoter of auxin-responsive gene
CCA1ATLHCB1	AAMAATCT	0.006	leaf-specific responsive elements; related to regulation by phytochrome
ATC-MOTIF	AGTAATCT	0.007	part of a conserved DNA module involved in light responsiveness
TATABOX5	TTATTT	0.008	a functional TATA element
GT1GMSCAM4	GAAAAA	0.01	salt-related cis-acting element
MARTBOX	TTWTWTTW TT	0.01	T-box', motif found in SAR (scaffold attached region; or MAR)

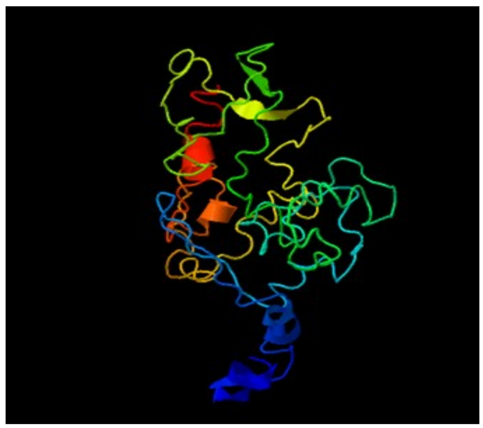
BELLRINGER/REPLUMLESS/PENNYWISE_BS1_IN_AG	AAATTAAA	0.011	related to floral and inflorescence meristems
POLASIG2	AATTAAA	0.017	Poly-A'signal found in rice alpha-amylase;
CAREOSREP1	CAACTC	0.018	promoter region of a cystein proteinase (REP-1) gene in rice
ABRE	TACGTG	0.022	involved in ABA responsiveness
G-BOX(CUF1) element	CACGTA	0.022	early senescence of rice flag leaf; cis-Acting regulatory element involved in light and cold responsiveness
ATHB2_BINDING_SITE_MOTIF	TAATAATTA	0.023	Dehydration, high salinity and low temperature responsive
GADOWNAT	ACGTGTC	0.025	GA-responsive element
POLASIG3	AATAAT	0.028	Plant Poly-A signal ; Consensus sequence for plant polyadenylation signal
AACACOREOSGLUB1	AACAAAC	0.046	involved in controlling the endosperm-specific expression
QARBNEXTA	AACGTGT	0.046	JA-responsive element or wounding (or wounding and tensile stress responsive element)
ABRELATERD1	ACGTG	0.046	induction by dehydration stress and dark-induced senescence
ROOTMOTIFTAPOX1	ATATT	0.049	root-specific responsive elements

A

```
                     20                  40                  60                  80                  100
                      |                   |                   |                   |                   |
Sequence   MVLETVSKTEPPPLLGPRISFSSDLSDGGDFICITPVMCKEDVVKGSVKVSDFEFLSSENVSPQRMLTADELFSEGKLLPFWQVKHSEKLKNITLKTNEEEEAEKRKVEVKKKD
Prediction CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCSSSCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Conf.Score 9334553457899999951562545446666666665444443334577767715644787665655647887517815241102455544435544566555666666655543
           H:Helix; S:Strand; C:Coil
```

```
      120                 140                 160                 180                 200                 220                 240
       |                   |                   |                   |                   |                   |                   |
KDQEINNRDNRVTWFIDEDPSPRPPKCTVLWKELLRLKKQRNPSSSPVTARTVSSLSPSSSTSSSSSLEDAAKREEKEKEGKRGKKGLERTRSASMRIRPMIHVPICTPSKSSLPLPPLFPLSLKKNRVERRA
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCCCCSSSSSCCCCCCCCCCCCCCCCCCCCCHHHHCCCHHCCC
4311246541000145556688788775005777434532336788444656666776776776433121002232112413331356555578556554446887776677676897772454316300059
```

B



C-score=-4.17
Estimated TM-score = 0.27±0.08
Estimated RMSD = 16.0±3.1Å

Figure 5-6 (A) Predicted secondary structure of the At3g05980 protein generated by I-TASSER (Zhang, 2008); α-helices (H) and β-strands (S) were highlighted in red and blue respectively. The letter C indicated coil. The confidence score for each residue ranging 0 to 9 was demonstrated in the next row. (B) The best 3D model of At3g05980 generated by I-TASSER. The confidence score (C-score) in the range of -5 to 2 was a measurement of the model quality. Higher C-score indicated better quality and models with C-score > -1.5 had a correct fold. RMSD, root mean square deviation (in the range of 0 to 30 Å) and TM-score (in the range of 0 to 1) were estimates of the model accuracy (the structural similarity between the predicted model and the native structure). TM-score < 0.17 means two randomly picked proteins and TM-score > 0.5 means two proteins have similar fold.

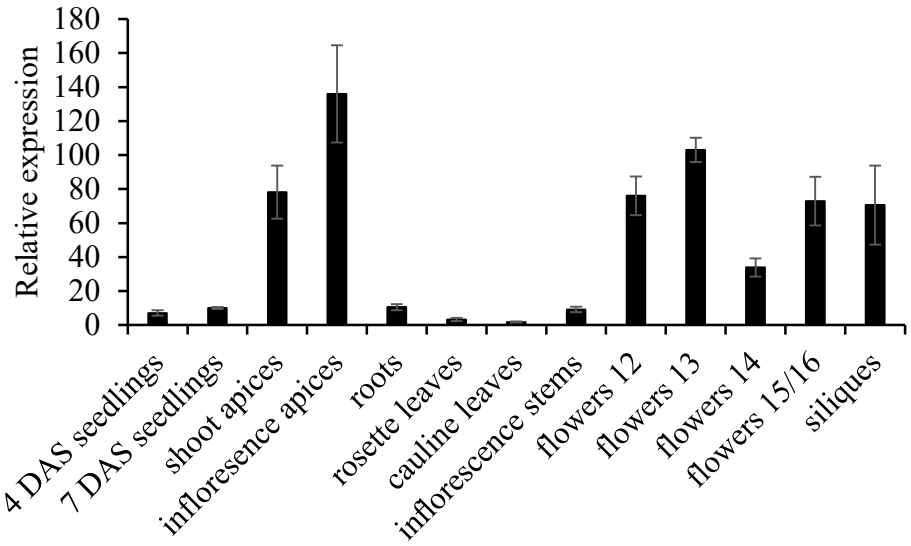
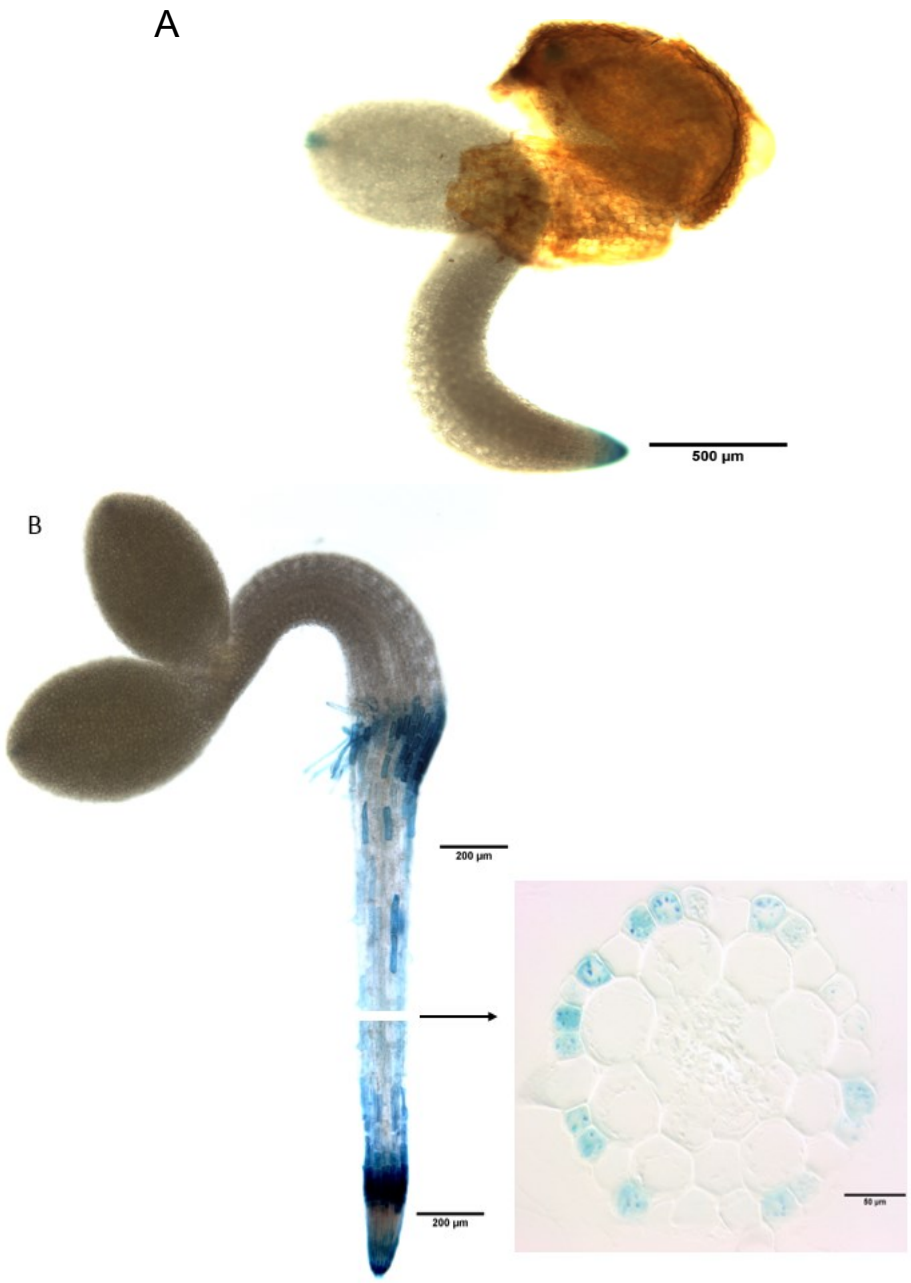


Figure 5-7 Expression patterns of At3g05980 gene in different Arabidopsis tissues. Shoot apices and root were dissected from 18 days plants. Inflorescence apices were dissected from 23 days plants. Rosette leaves, cauline leaves, flowers and siliques were collected from four-weeks-old plant. Shoot apices samples may contain some leaf or floral primordial leftover. Flowers were named according to Cai's definition (Cai & Lashbrook, 2008). EF1A and ACTIN2 were used as endogenous control (Czechowski, 2005). Error bars indicated the standard derivations.


A
500 µm
B
200 µm
200 µm


C
200 µm
200 µm
200 µm
200 µm
D
200 uM
200 uM
200 uM
200 uM
200 uM

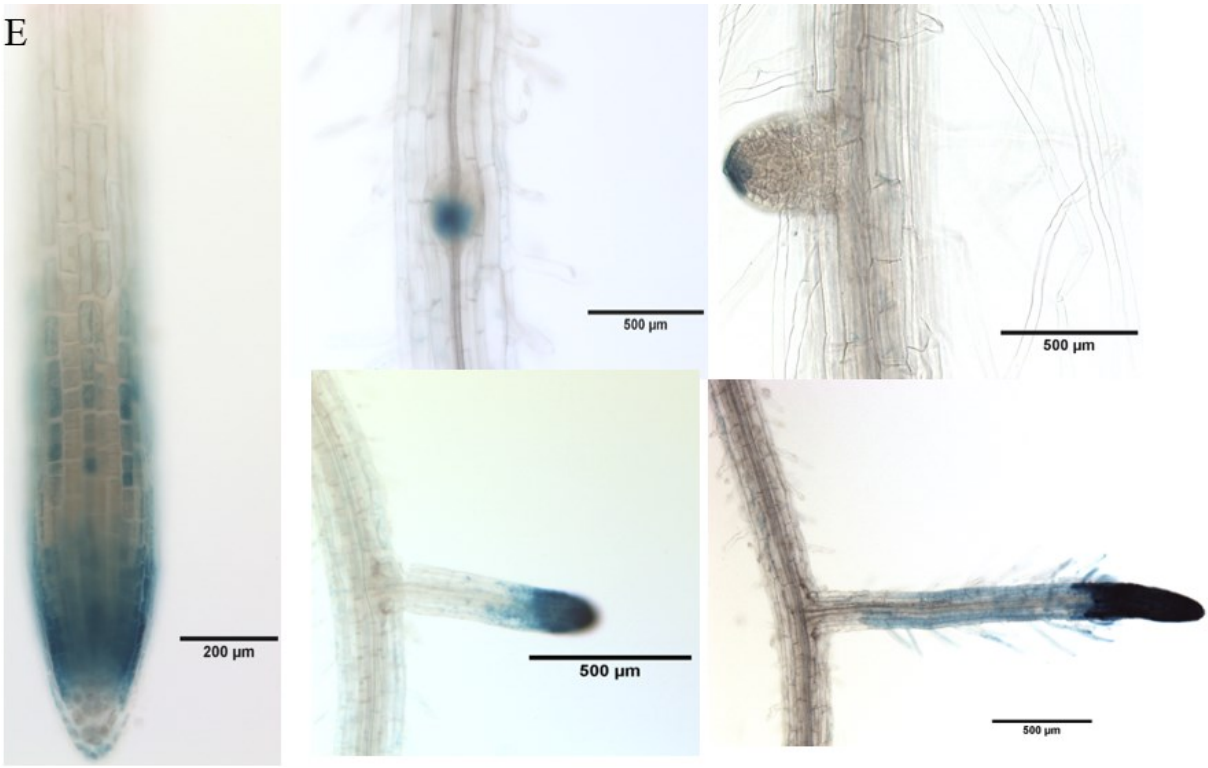


Figure 5-8 At3g05980 expression in seedlings. GUS activity in 1-day-old seedlings (A), 2-day-old seedlings (B), 3-day-old seedlings (C), 4-day-old seedlings (D) and 8-day-old seedlings (E).

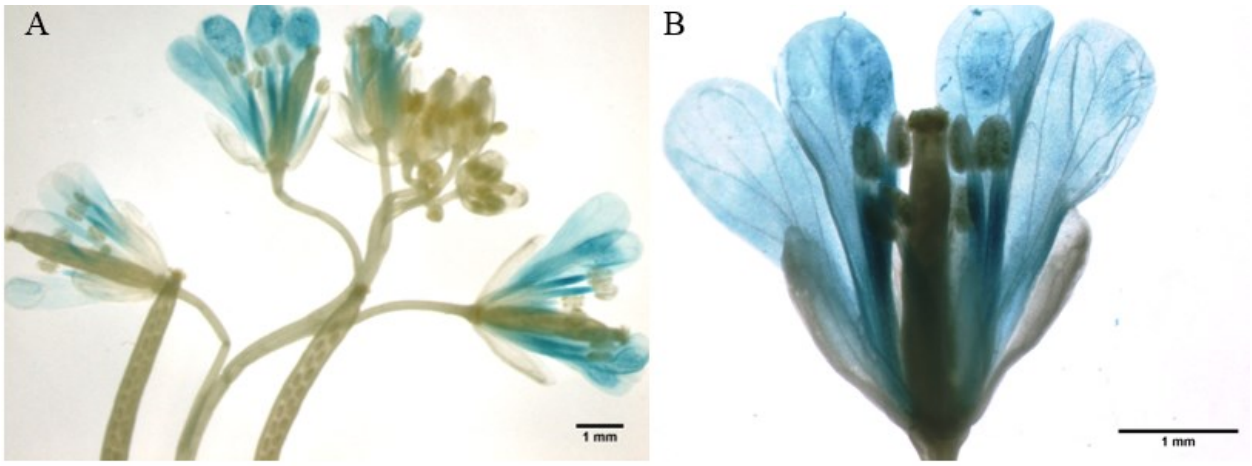


Figure 5-9 At3g05980 expression in flowers.

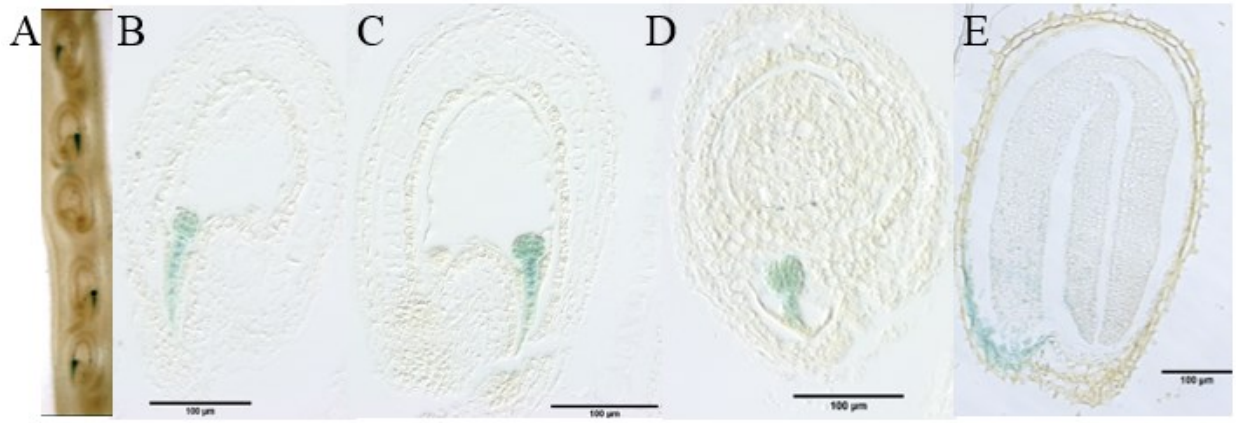


Figure 5-10 At3g05980 expression in developing seeds.

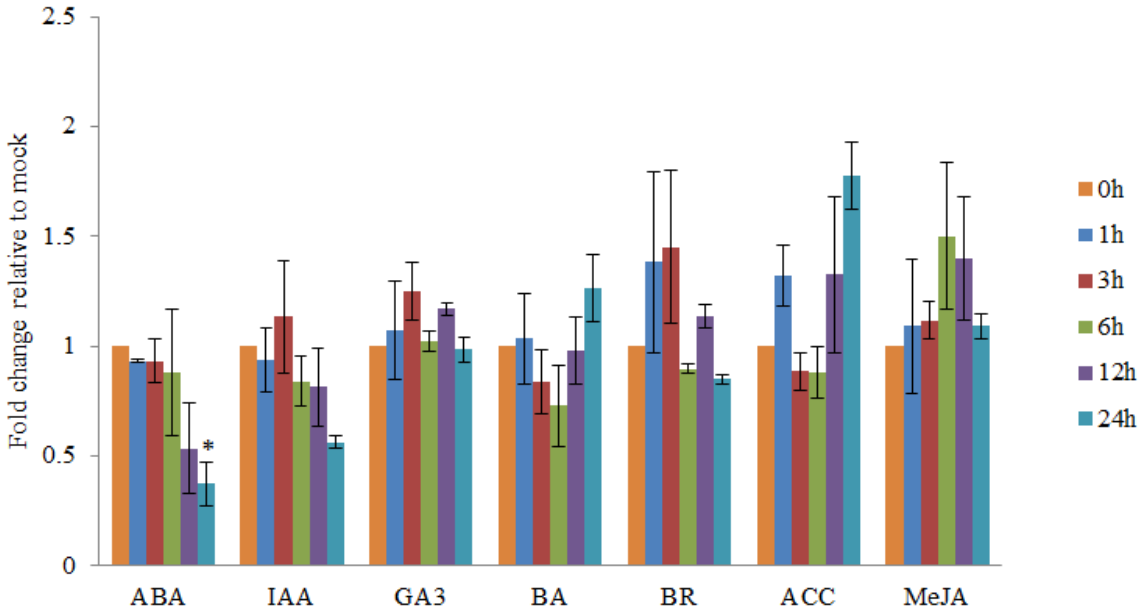


Figure 5-11 Effects of hormones on the transcript level of the At3g05980 gene. Seven DAS Arabidopsis seedlings were transferred and maintained in the ½ X MS liquid medium supplement with the following hormones: ABA (abscisic acid):10 µM; IAA (3-indoleacetic acid):5 µM; BA (6-benzylaminopurine):5 µM; MeJA (methyl jasmonate):10 µM; BR (brassinosteroid):1 µM; ACC (1-aminocyclopropane-1-carboxylic; EF-1A was used as the endogenous control (Czechowski, 2005); Gene expression levels in seedlings were measured by RT-PCR. Error bars represented the standard derivations. The asterisk indicates a significant change ($p < 0.05$, student's t-test).

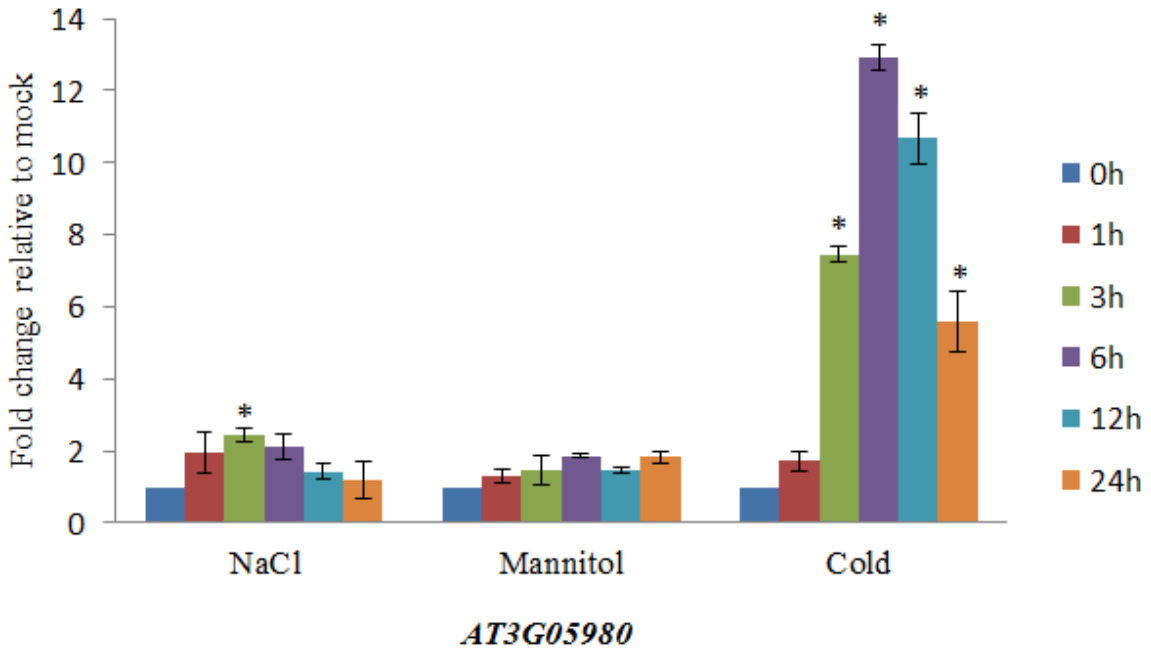


Figure 5-12 Responsiveness of At3g05980 gene to several abiotic stresses checked by qRT-PCR. EF-1A was used as the endogenous control (Czechowski, 2005). Error bars represented the standard derivations. NaCl: 200mM; Mannitol: 300mM; Cold: 4 °C. * *p*-value <0.05 (Student's t-test).

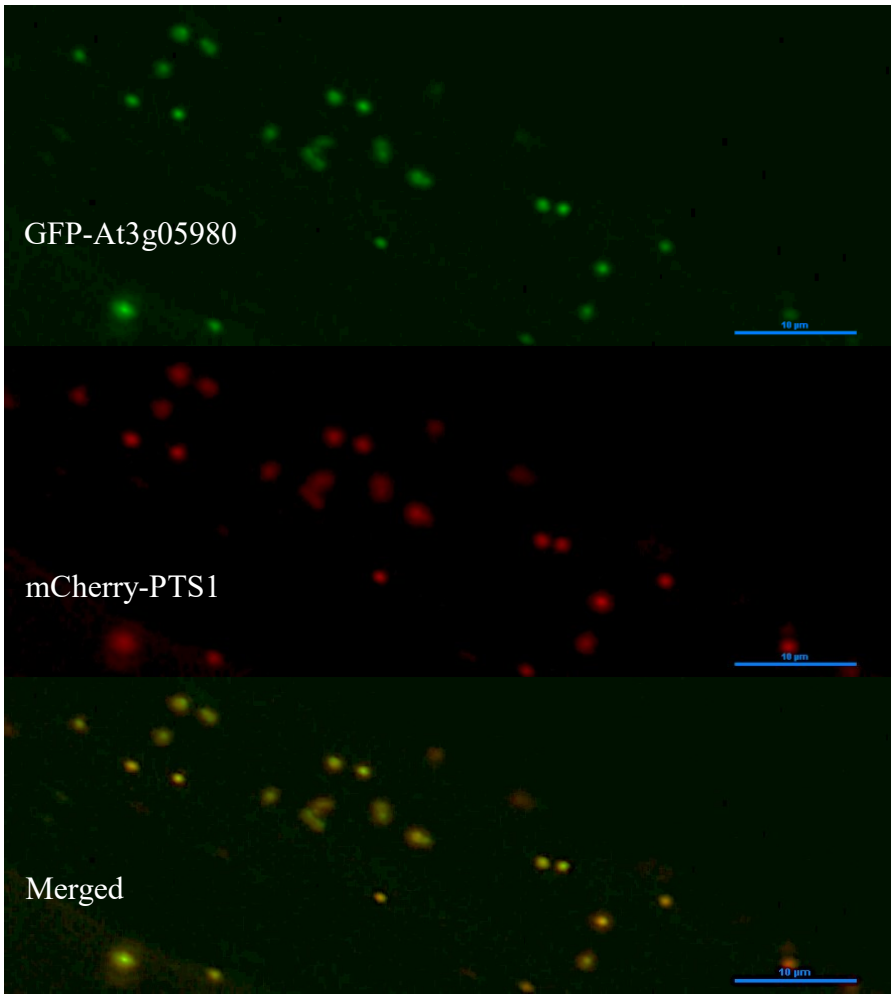


Figure 5-13 Subcellular localization of At3g05980. Cells shown are root tip cells of a plant coexpressing mCherry-PST1 and GFP-At3g05980.

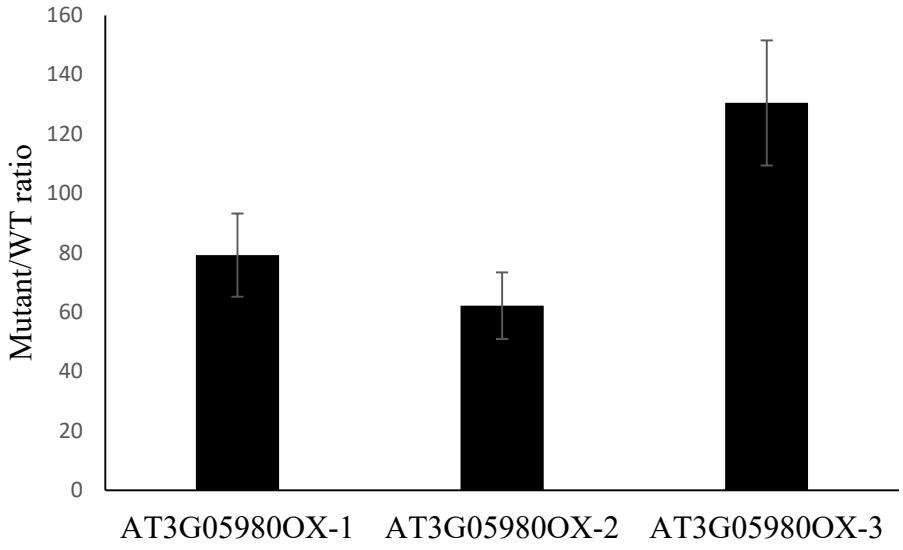


Figure 5-14 Transcript abundance of At3g05980 in 35S:: At3g05980 transgenic lines checked by qRT-PCR; Floral buds were sampled from four-weeks-old plants and EF-1a was used as the endogenous control (Czechowski, 2005). The error bars indicated the standard derivations of three biological replicates.

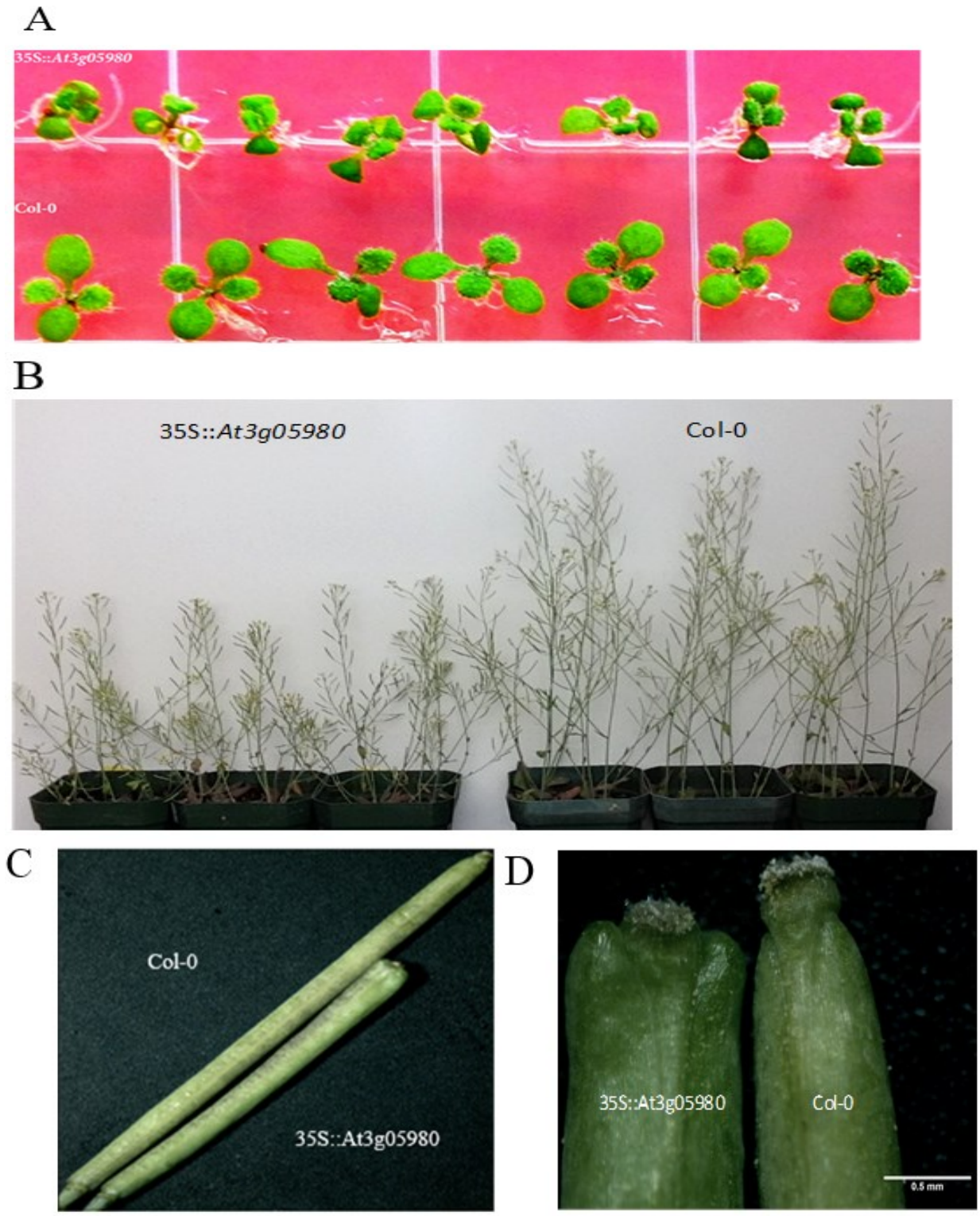


Figure 5-15 Morphology of 35S:: At3g05980 transgenic plants. A: 8 DAS seedlings grown on 1/2 X MS medium. B: Two-months-old 35S::At3g05980 and WT plants; C and D: Siliques of 35S::At3g05980 and WT plants.

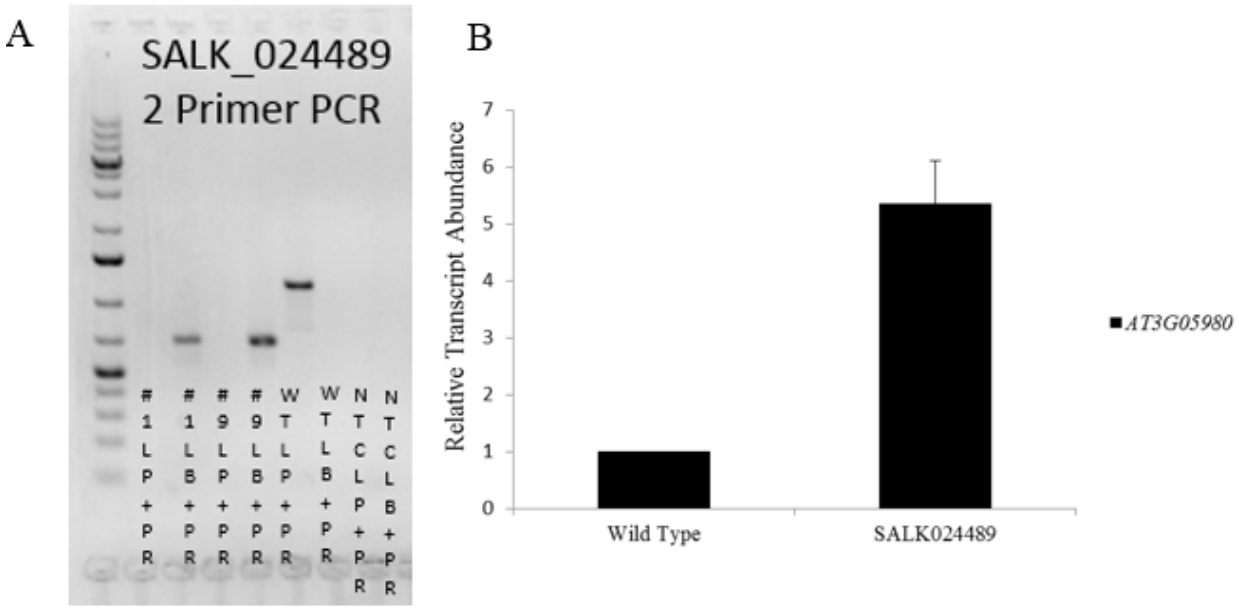


Figure 5-16 A T-DNA line of At3g05980 characterized in this study. A: Two homozygous SALK_024489 plants (#1 and #9) identified by two-primer PCR. Both genomic DNA from WT Col-0 and water (NTC) was used as control. First lane: 1kp DNA ladder. B: Relative transcript abundance of At3g05980 in SALK024489 compared to the WT Col-0 checked by qRT-PCR.

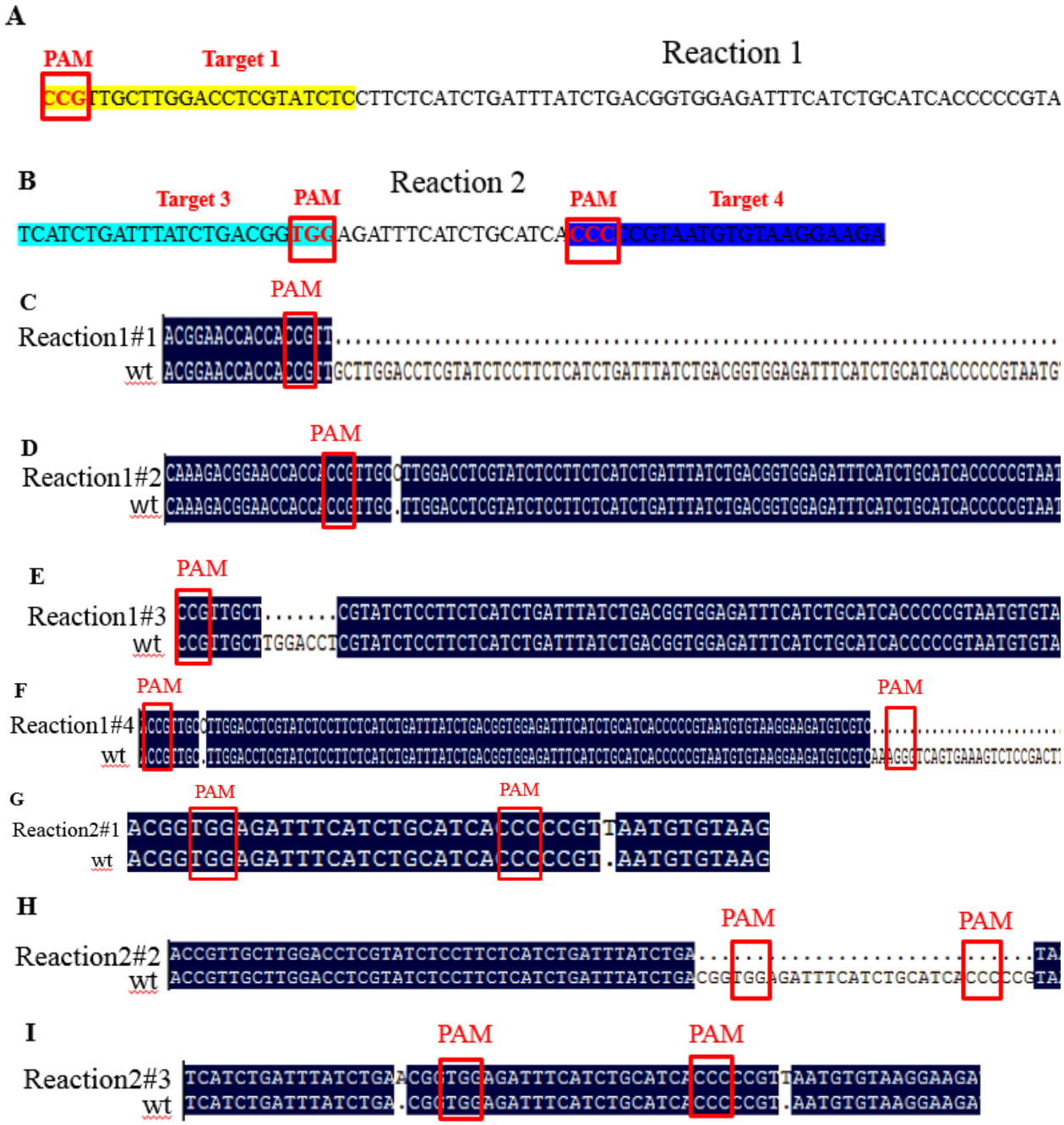
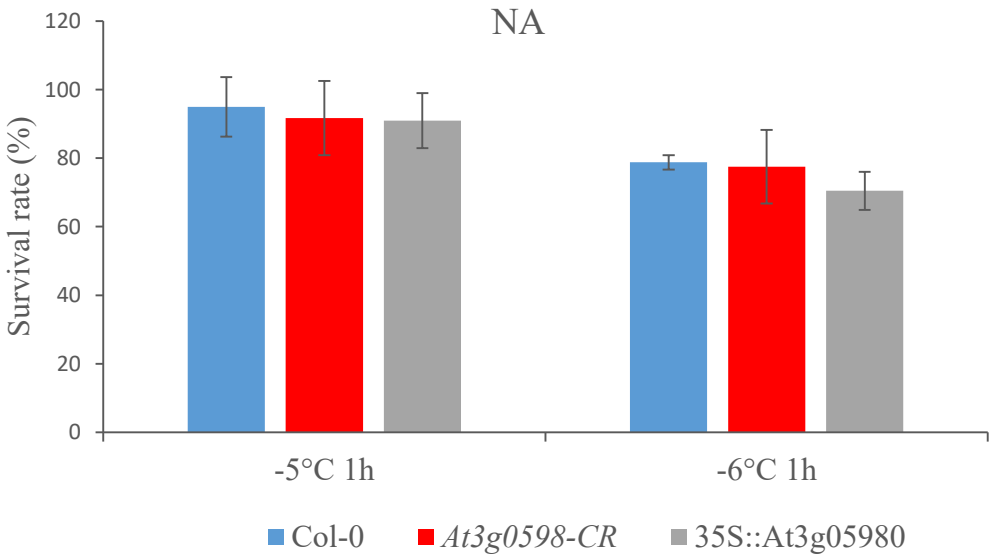


Figure 5-17 At3g05980 single gene editing created by CRISPR-Cas9 system. A: sequence of the target site for the construct one. B: sequence of the target site for the construct two. C, D, E, F: Representatives of editing generated in At3g05980 by construct one. G, H, I: Representatives of editing generated in At3g05980 by construct two.

A



B

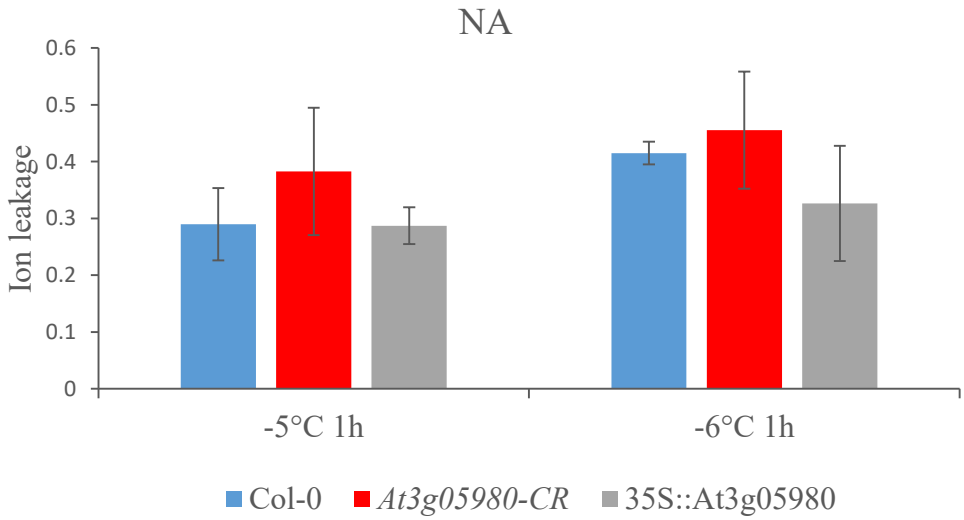
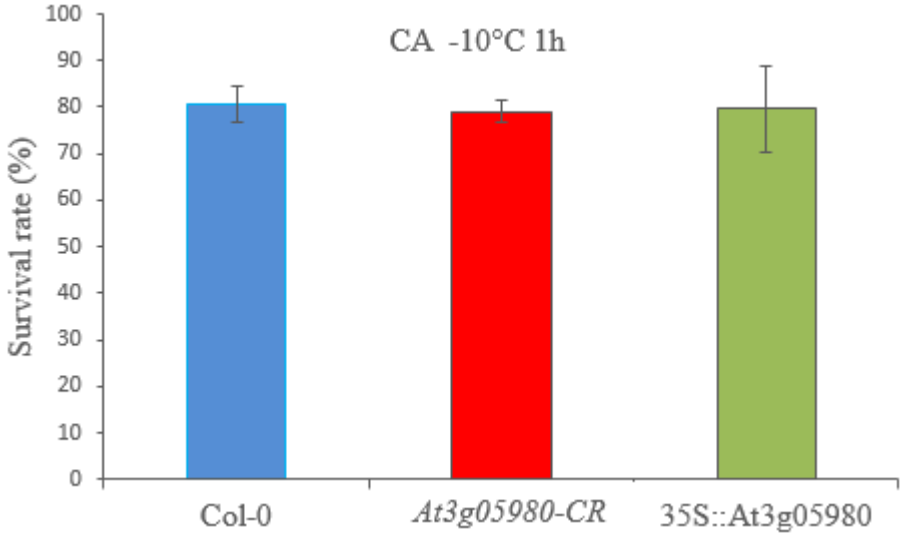


Figure 5-18 The nonacclimated (NA) freezing phenotype: survival rate (A) and ion leakage (B) of two-weeks-old At3g05980 mutants. Error bars indicates the standard derivations of three biological replicates.

A



B

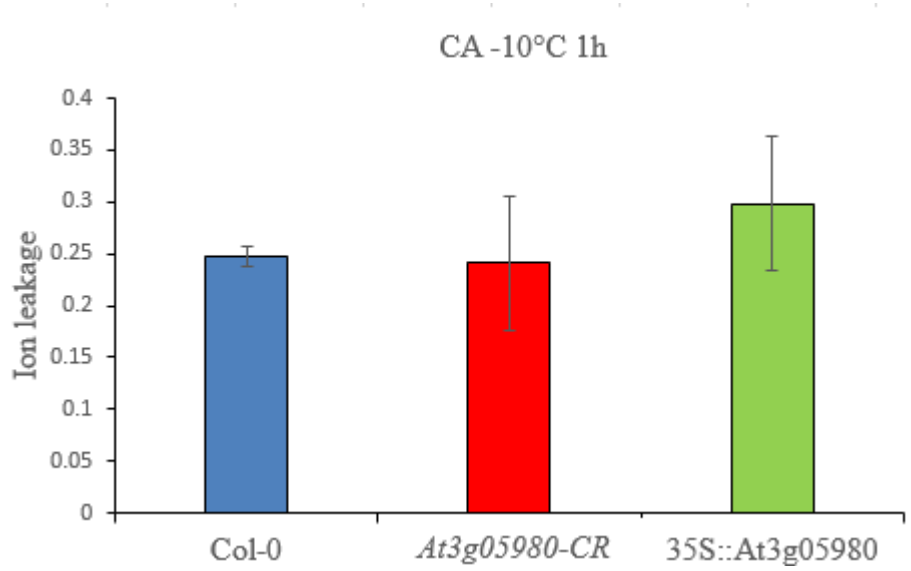


Figure 5-19 The cold-acclimated (CA) freezing phenotype: survival rate (A) and ion leakage (B) of two-weeks-old At3g05980 mutants. Error bars indicates the standard derivations of three biological replicates.

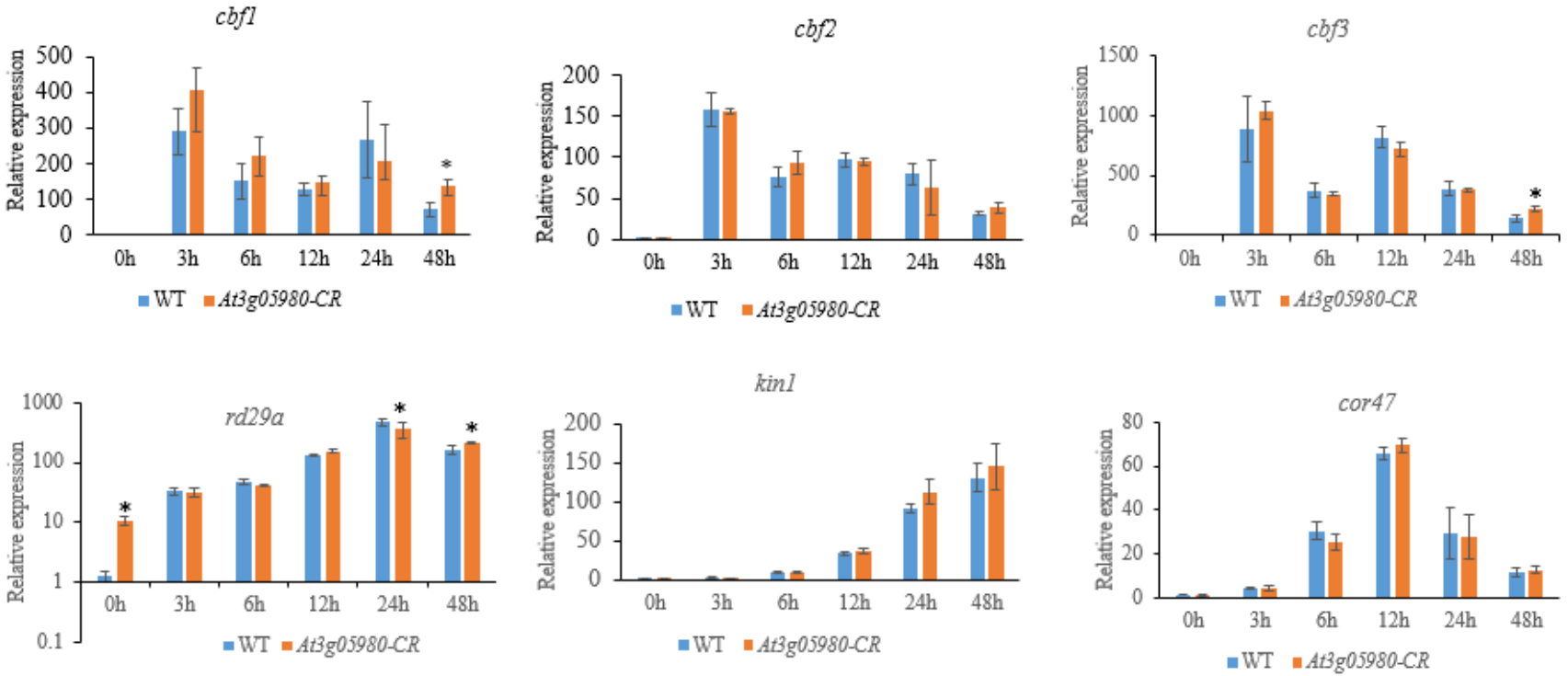


Figure 5-20 Compare the transcript levels of several cold-regulated genes in *At3g05980*-CR and WT by qRT-PCR. 10 DAS WT and *At3g05980* mutant seedlings were treated at 4°C for the indicated time. Actin2 was applied as the endogenous control (Czechowski, 2005). Error bars indicated standard derivation of four biological replicates. * p-value<0.05 (Student's t-test).

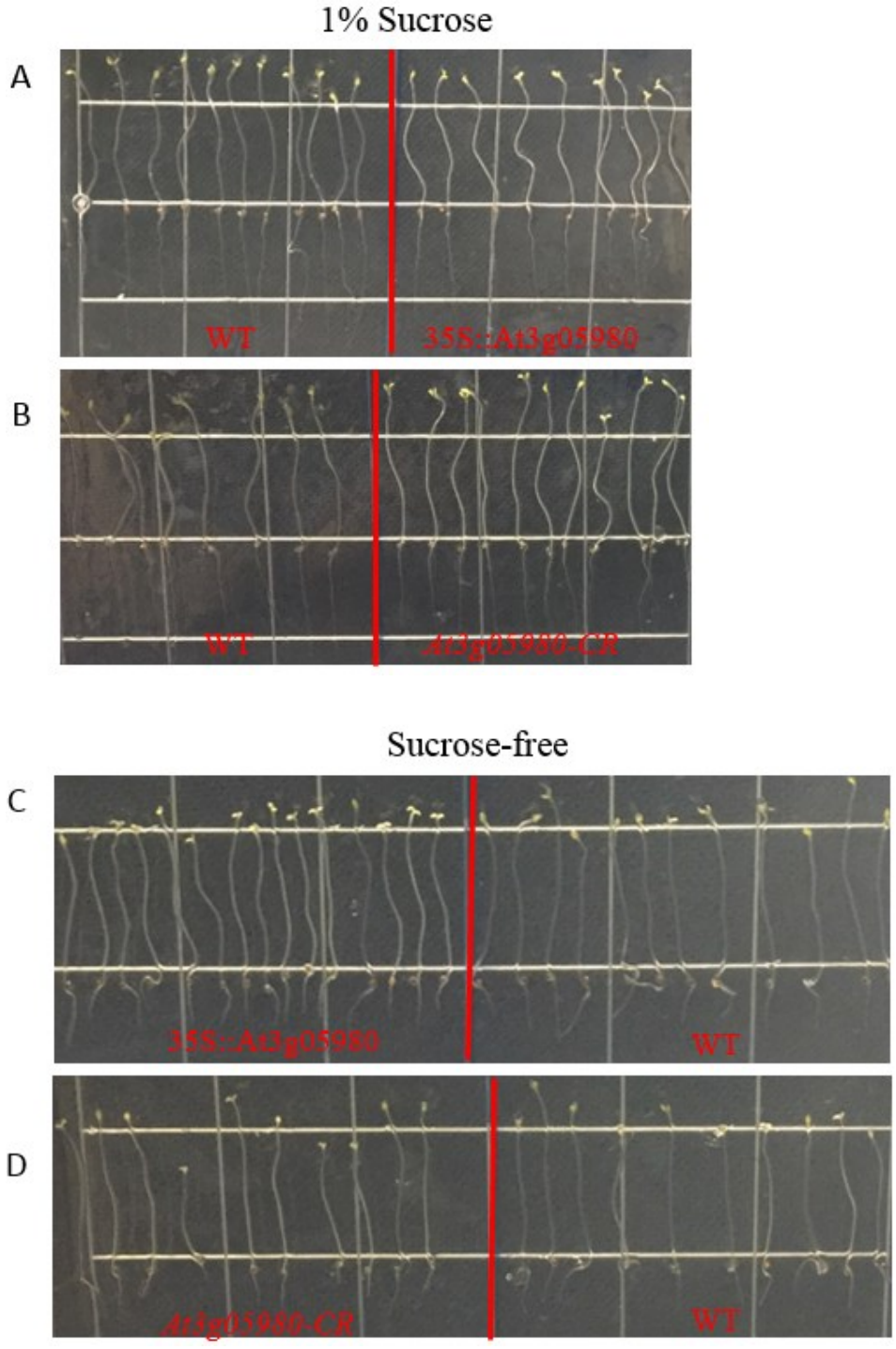


Figure 5-21 Phenotyping of At3g05980 mutants on 1/2 X MS medium supplemented with 1% sucrose or without sucrose under dark for 7 days.

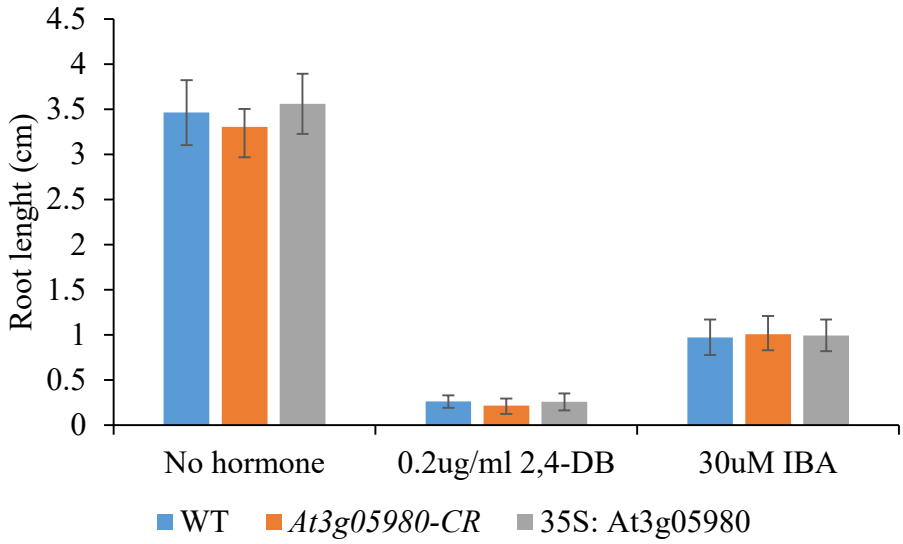


Figure 5-22 Comparison of root growth of WT, *At3g05980*-CR and 35S: At3g05980 on 1/2 X MS medium with no added hormone or medium containing 0.2 µg/ml 2,4-DB or 30uM IBA after growing seven days at 22°C with 16 h light. Error bars indicates standard derivations (n=18).

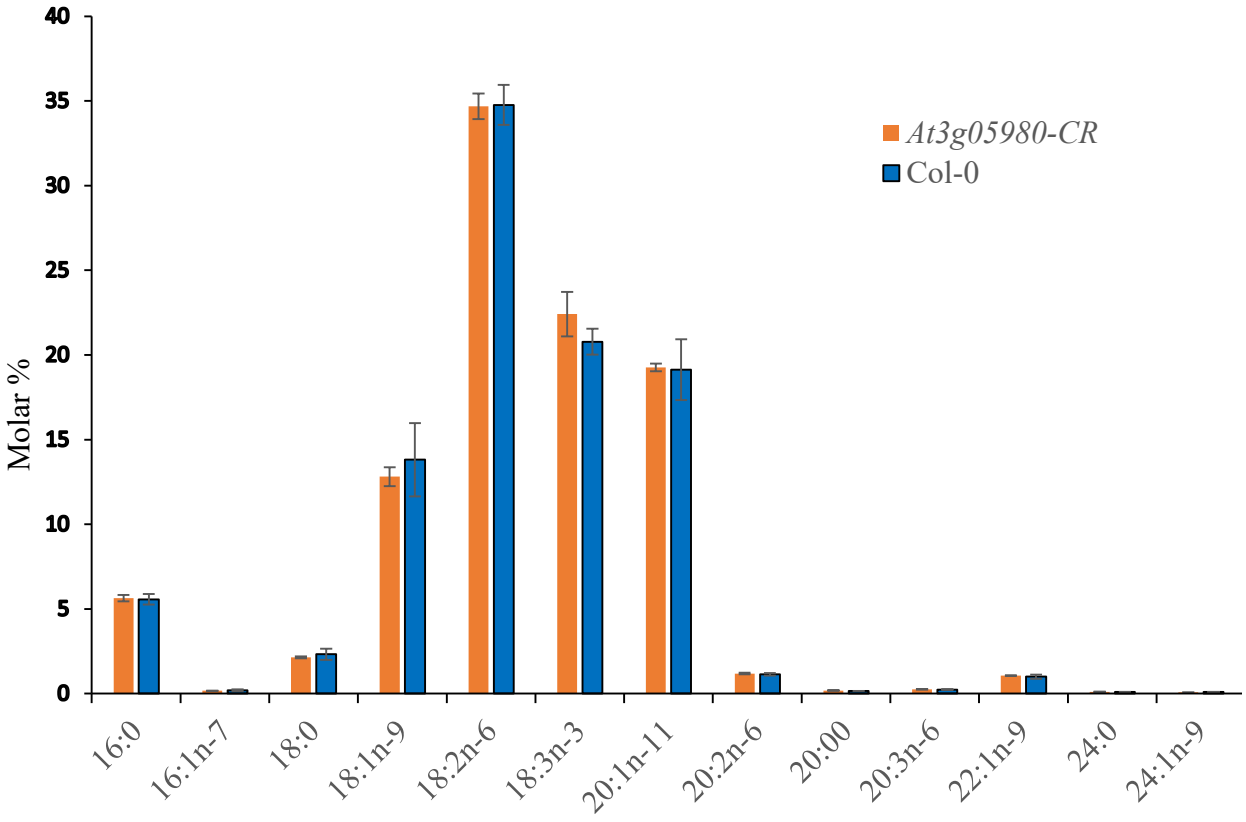


Figure 5-23 Fatty acid profiles in the dry seeds of wild-type Col-0 and *At3g05980*-CR. Error bars indicates standard derivations of three biological replicates.

Chapter 6. General Discussion

6.1 Potential transcriptional regulators of phloem fiber specification

In this study, my first objective was to find transcriptional regulators of flax phloem fiber specification. To date, the genetic basis of primary phloem fiber identity in any species is unknown. We hypothesized that the transcription factors controlling phloem fiber specification should have a higher expression level in the AR (the apical-most 0.5 mm of the stem) compared to the BR (1cm below the shoot apex to the base of the stem) based on the following: 1) all the phloem fibers in flax stem are derived from the shoot apical meristem; 2) the first visible phloem fibers in flax stem were identified around 0.5 mm from the shoot apex (Gorshkova et al., 2003); 3) the transcriptional regulators that control fiber cell fate are assumed to complete their activity before we can see fiber cells (Gorshkova et al., 2012). I first used RNA-Seq to compare the gene expressions in the AR and BR (Chapter 2). As a result, 6207 genes were found to be preferentially expressed in the AR compared to the BR and among them, 349 genes were predicted to encode transcription factors including 27 AR uniquely expressed genes (Chapter 2). Meanwhile, a total of 49 transcription factors were found to have at least 16 times more abundance in the AR compared to the BR and many of them were reported to be involved in the stem identity specification, shoot apical meristem formation and maintenance as well as epidermal cell identity specification in *Arabidopsis*. Even so, many of these AR –enriched transcription factors were not characterized yet in any plant species and some of them might have a role in the shoot apical meristem formation or organogenesis. Meanwhile, these 349 AR-enriched transcription factors may contain some transcriptional regulators of flax phloem fiber specification.

Studies in *Arabidopsis* and a few other plant species showed that some NAC and MYB transcription factors played key roles in plant vascular differentiation (Grant et al., 2010; Wang, et al., 2009; Xu et al., 2014). I predicted that some NAC or MYB transcription factors preferentially expressed in the AR might be involved in the phloem fiber cell specification. Based on the RNA-Seq analysis, we found that 18 *LusMYBs* were significantly enriched in the AR (Table 3-5). To make inference about their functions, I have searched the *Arabidopsis* orthologues of these 18 *LusMYBs* (Appendix 8). The *Arabidopsis* orthologs of six AR-preferentially expressed *LusMYBs* including *LusMYB187* (*AtMYB3R2*), *LusMYB181* (AtMYB3R2), *LusMYB180* (*ATMYB3R1)*, *LusMYB162* (*ATMYB3R1)*, *LusMYB175* (*AtMYB3R4)* and *LusMYB179* (*AtMYB3R5)* were reported to be involved in the cell cycle regulation (Haga et al., 2011; Haga et al., 2007; Saito et al., 2015). Besides, *LusMYB34* and *LusMYB36* were duplicated genes and *LusMYB35* was their closest paralog. The *Arabidopsis* ortholog of these three genes, *AtMYB17*, was reported to be an important meristem identity regulator from vegetative growth to flowering (Zhang et al., 2009; Pastore et al., 2011). Similarly, two duplicated genes *LusMYB172* and *LusMYB171* were both found enriched in the AR and their *Arabidopsis* ortholog, *AtMYB91*, was revealed to function in the leaf proximodistal axis specification (Hay, 2006). Two other AR-enriched *LusMYBs*, *LusMYB141* and *LusMYB142*, formed a duplicated gene pair and their *Arabidopsis* ortholog (*ATMYB105)* was known to be involved in the boundary specification, meristem initiation and maintenance, and organ patterning while the *Arabidopsis* ortholog of another AR-enriched duplicated gene pair (*LusMYB61*/ *LusMYB66)*, *AtMYB36*, was suggested to promote differentiation of the endodermis (Lee et al., 2009; Liberman et al., 2015; Fernández-Marcos et al., 2017). The remaining three AR-enriched *LusMYBs* were *LusMYB26*, *LusMYB149* and *LusMYB102* and their orthologs have not been functionally characterized in any species. Although

the expression level of *LusMYB26* was the lowest among these 18 AR-enriched *LusMYBs*, the transcript of *LusMYB26* was only detected in the AR but not in the BR and this gene belong to a clade consisted only flax and populous genes but not *Arabidopsis* genes.

Similarly, we found nine *LusNACs* that were preferentially expressed in the AR from the RNA-Seq analysis, including: *LusNAC93*, *LusNAC158*, *LusNAC50*, *LusNAC100*, *LusNAC120*, *LusNAC27*, *LusNAC92*, *LusNAC114* and *LusNAC65* (Table 4-5). *LusNAC93* accumulated 45 times more transcript abundance in the AR compared to the BR and its *Arabidopsis* ortholog was *AtCUC3*, an important transcriptional regulator of shoot apical meristem formation, axillary meristem initiation and organ separation (Hibara et al., 2006; Raman et al., 2008). *LusNAC158* was seven-fold more enriched in the AR. Blast search in TAIR10 indicated that the best Blast hit of this gene in *Arabidopsis* genome was *VND2*. However, the phylogenetic dendrogram I constructed had divided *LusNAC158* in a clade without VNS genes (Figure 4-1). Based on the phylogenetic dendrogram, *LusNAC158* was a member of clade 2, which consisted 21 flax genes, 17 *populus* genes but none *Arabidopsis* gene (Figure 4-1). Genes in this subfamily have not been functionally characterized yet and the large number of flax genes in this family indicate that they might be important for flax development. *LusNAC50* and *LusNAC27* were 5.8 times and 3.1 times more enriched in the AR respectively and they were duplicated genes. Their *Arabidopsis* ortholog was SOG1, which was suggested to govern multiple responses to DNA damage (Yoshiyama et al., 2014). Similarly, two duplicated genes, *LusNAC100* and *LusNAC120*, expressed 4.2-fold and 3.5-fold in the AR compared to the BR respectively and the function of their *Arabidopsis* ortholog was not yet characterized. However, one of their closest *Arabidopsis* homologs in the same clade (*VNI2*) was shown to negatively regulate xylem vessel formation, therefore, it would be necessary to study

one of these two genes. Although *LusNAC92* was 2.3-fold more enriched in the AR and its *Arabidopsis* ortholog *AtNAC50* was involved in flower time control (Ning et al., 2015). *LusNAC65* was only detected in the AR but not in the BR, although its expression level in the AR was lowest among these nine AR-enriched *LusNACs* (inferred based the FPKM value). Meanwhile, the *Arabidopsis* orthologous gene of *LusNAC65*, *AtCUC1*, was revealed to be an important regulator of shoot apical meristem formation and auxin-mediated lateral root formation (Lee et al., 2015; Spinelli et al., 2011).

6.2 Characterization of flax NAC and MYB gene family

In addition to vascular differentiation, NAC and MYB transcription factors were also reported to be important for many other aspects of plant development (Zhong et al., 2007a; Legay et al., 2010; Wang et al., 2014). Therefore, I have performed a genomic-wide identification and expression profiling of MYB and NAC transcription factors from flax. This study identified 240 putative *MYBs* and 182 putative *NACs* from the flax genome and they were divided into 18 and 17 clades respectively (Figure 3-1; Figure 4-1). The identified *LusMYBs* included 53 *MYB-related genes*, 179 *2R-MYBs*, seven *3R-MYBs* and one *4R-MYBs* (Appendix 4). I have checked the expressions of *LusMYBs* of *2R-, 3R-* and *4R*-type and *LusNACs* in publicly available EST, microarray and RNA-Seq data. As a result, I found *LusMYB76*, *LusMYB45*, *LusMYB174, LusNAC46*, *LusNAC160*, *LusNAC87*, *LusNAC66*, *LusNAC31*, *LusNAC121* might be involved in the flax xylem differentiation since they were all specifically expressed in the xylem tissue of the flax stem (Figure 3-2; Figure 3-3; Figure 4-2; Figure 4-3). Additionally, we found *LusMYB90*, *LusMYB36* and *LusMYB33* might be related to the secondary cell wall formation in flax stem phloem fiber cells since these three genes appeared to be more preferentially expressed in the external part of the flax stem compared to the inner part and they had a higher expression in the lower part of the

flax external stem (Figure 3-3). Meanwhile, we noted that *LusMYB36* showed apparently higher expression level in the Drakkar than Belinka, the former one was a flax variety with better fibers (Figure 3-2).

This study also revealed that *LusNAC182, LusMYB118, LusMYB127, LusMYB129, LusMYB113* and *LusMYB148* were significantly enriched in the top part of flax stem in which phloem fiber were undergoing intrusive elongation, indicating that they might be related to phloem fiber cell elongation (Figure 3-4; Figure 4-4). In contrast, *LusNAC67* was most abundant around the snap point and *LusNAC161*, *LusMYB51* and *LusMYB33* were most abundant in the stem below the snap point, suggesting that they might be involved in the secondary cell deposition in the phloem fiber cells of flax stem (Figure 3-4; Figure 4-4).

Moreover, through comparison the transcript expressions of *LusVNDs* in 12 different tissues by qRT-PCR, I found *LusNAC28* and *LusNAC125*, were enriched in the phloem fibers, while *LusNAC136*, the ortholog of *Arabidopsis VND7*, was preferentially accumulated in the xylem tissues (Figure 4-7). This suggested that *LusNAC28* and *LusNAC125* might be associated with the phloem fiber development whereas LusNAC136 might be involved in the xylem development. Considering that the fibers I used to qRT-PCR analysis was collected from the lower part of the flax stem, I assumed that *LusNAC28* and *LusNAC125* might be related to the secondary cell wall formation in the phloem fibers.

As a summary, this study havhase proposed several candidate genes for further study of the flax phloem fiber cell specification and secondary cell wall deposition. To further determine whether they had the functions proposed here, we need to study the loss-of-function mutants for these genes.

6.3 Functional analysis of an uncharacterized *Arabidopsis* gene, *At3g05980*.

In Chapter 2, we found an uncharacterized flax gene, *Lus10041215,* was abundant in the shoot apex. In Chapter 5, I studied one of its *Arabidopsis* homologs, *At3g05980*, and uncovered several of its characteristics. First, *At3g05980* was found to be conserved in eudicot but not present in other taxa, suggestive of a eudicot-specific role. Second, four highly conserved motifs were discovered in the protein sequence of *At3g05980*. Third, *At3g05980* mRNA was preferentially expressed in the shoot apices, root apices, atrichoblasts, petals, young developing embryos and micropylar endosperms. Fourth, *At3g05980* transcript was greatly induced by cold, but not by salt, drought or hormone treatments including ABA, IAA, GA3, BA, BR, ACC, MeJA. These results were generally consistent with the expression patterns previously reported in public databases. *At3g05980* was shown to be enriched in the shoot apices by both the microarray data in the eFP browser and my qRT-PCR study, however, this study indicated that the intergenic sequence upstream the start codon of this gene could not drive the GUS staining in the shoot apices. To further confirm the expression of *At3g05980* in the shoot apices, a different method such as in situ hybridization should be applied. Fifth, overexpression of *At3g05980* lead to minor morphological defects, including cotyledon epinasty, and slight shortening of both plant height and silique length. However, loss-of-function mutation in this gene did not induce any discernable morphological abnormality. We also found that the freezing tolerance of At3g05980 overexpression lines and loss-of-function mutants were not significantly different from the WT. However, the expression of *RD29* was increased in the loss-of-function mutants, under either normal or stressed condition.

Lastly, protein of At3g05980 was targeted to the peroxisome while function loss of At3g0980 or transcript increase of At3g05980 had no effect on the peroxisomal β-oxidation.

An important next step would be to phenotype the double mutants of *At3g05980* and its paralog *At5g19340*, in terms of plant morphology, freezing tolerance, chilling stress, peroxisomal β-oxidation, root (both primary and lateral) development as well as root hair. However, to make better inference of the role of *At3g05980*, it may be necessary to compare the transcriptomes of the loss-of-function mutants and WT to examine the pathways involved. Beyond that, it may be also necessary to study the protein expression patterns by developing At3g05980 protein specific antibody.

6.4 Conclusions

This study investigated the transcriptome of flax shoot apices and identified genes enriched in this region. This will improve our understanding of the shoot apices in general and help to define the genetic mechanisms of phloem fiber specification. Additionally, this study has expanded our understanding about the NAC and MYB transcription factors in flax and identified several NAC and MYB which were potentially associated with the phloem fiber differentiation in flax stem. Furthermore, this study also gained some insight about an uncharacterized *Arabidopsis* gene, *At3g05980*. Although we did not get conclusive information about its specific function, the phenomena observed in this study indicated that this gene might be related to the cold stress and root development.

References

Abascal, F., Zardoya, R., & Posada, D. (2005). **ProtTest: Selection of best-fit models of protein evolution.** *Bioinformatics, 21*(9), 2104–2105. https://doi.org/10.1093/bioinformatics/bti263

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2003). **Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling.** *The Plant Cell, 15*(1), 63–78. https://doi.org/10.1105/tpc.006130

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., & Shinozaki, K. (1997). **Role of arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression.** *Plant Cell, 9*(10), 1859–1868. https://doi.org/10.1105/tpc.9.10.1859 [doi]\r9/10/1859 [pii]

Abuqamar, S., Luo, H., Laluk, K., Mickelbart, M. V., & Mengiste, T. (2009). **Crosstalk between biotic and abiotic stress responses in tomato is mediated by the AIM1 transcription factor.** *Plant Journal, 58*(2), 347–360. https://doi.org/10.1111/j.1365-313X.2008.03783.x

Ageeva, M. V., Petrovská, B., Kieft, H., Sal'nikov, V. V., Snegireva, A. V., Van Dam, J. E., Van Veenendaal W. L., Emons, A. M., Gorshkova, T. A., & Van Lammeren, A. A. (2005). **Intrusive growth of flax phloem fibers is of intercalary type.** *Planta, 222*(4), 565–574. https://doi.org/10.1007/s00425-005-1536-2

Alonso, J. M. (2003). **Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana.** *Science, 301*(5633), 653–657. https://doi.org/10.1126/science.1086391

Armstrong, J. I., Yuan, S., Dale, J. M., Tanner, V. N., & Theologis, A. (2004). **Identification of inhibitors of auxin transcriptional activation by means of chemical genetics in Arabidopsis.**

Proceedings of the National Academy of Sciences of the United States of America, 101(41), 14978–83. https://doi.org/10.1073/pnas.0404312101

Aug, T., July, T., April, T., Jan, T., May, T., Ho, M., & Jan, T. (2015). **Development and validation of a flax (*Linum usitatissimum* L.) gene expression oligo microarray.** *BMC Genomics, 27*(1), 288–311. https://doi.org/10.1186/1471-2164-11-592

Austin, R. S., Hiu, S., Waese, J., Ierullo, M., Pasha, A., Wang, T. T., & Provart, N. J. (2016). **New BAR tools for mining expression data and exploring Cis-elements in Arabidopsis thaliana.** *Plant Journal, 88*(3), 490–504. https://doi.org/10.1111/tpj.13261

Avila, J., Nieto, C., Cañas, L., Benito, M. J., & Paz-Ares, J. (1993). **Petunia hybrida genes related to the maize regulatory C1 gene and to animal myb proto-oncogenes.** *The Plant Journal : For Cell and Molecular Biology.* https://doi.org/10.1046/j.1365-313X.1993.03040553.x

Bae, S., Park, J., & Kim, J. S. (2014). **Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.** *Bioinformatics, 30*(10), 1473–1475. https://doi.org/10.1093/bioinformatics/btu048

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., & Noble, W. S. (2009). **MEME Suite: Tools for motif discovery and searching.** *Nucleic Acids Research, 37*(SUPPL. 2). https://doi.org/10.1093/nar/gkp335

Baima, S., Possenti, M., Matteucci, A., Wisman, E., Altamura, M. M., Ruberti, I., & Morelli, G. (2001). **The Arabidopsis ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems.** *Plant Physiol, 126*(2), 643–655. https://doi.org/10.1104/pp.126.2.643

Baker, A., Graham, I. A., Holdsworth, M., Smith, S. M., & Theodoulou, F. L. (2006). **Chewing the fat: β-oxidation in signalling and development.** *Trends in Plant Science.* https://doi.org/10.1016/j.tplants.2006.01.005

Barrett, L. W., Fletcher, S., & Wilton, S. D. (2012). **Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements.** *Cellular and Molecular Life Sciences.* https://doi.org/10.1007/s00018-012-0990-9

Baumann, K., Perez-Rodriguez, M., Bradley, D., Venail, J., Bailey, P., Jin, H., & Martin, C. (2007). **Control of cell and petal morphogenesis by R2R3 MYB transcription factors.** *Development, 134*(9), 1691–1701. https://doi.org/10.1242/dev.02836

Bennett, T., Van Den Toorn, A., Sanchez-Perez, G. F., Campilho, A., Willemsen, V., Snel, B., & Scheres, B. (2010). **SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in Arabidopsis.** *Plant Cell, 22*(3), 640–654. https://doi.org/10.1105/tpc.109.072272

Berleth, T., & Jürgens, G. (1993). **The role of the monopteros gene in organising the basal body region of the Arabidopsis embryos.** *Trends in Genetics, 9*(9), 299. https://doi.org/10.1016/0168-9525(93)90246-E

Bishopp, A., Help, H., El-Showk, S., Weijers, D., Scheres, B., Friml, J., & Helariutta, Y. (2011). **A mutually inhibitory interaction between auxin and cytokinin specifies vascular pattern in roots.** *Current Biology, 21*(11), 917–926. https://doi.org/10.1016/j.cub.2011.04.017

Blum, T., Briesemeister, S., & Kohlbacher, O. (2009). **MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction.** *BMC Bioinformatics, 10*(1), 1–11. https://doi.org/10.1186/1471-2105-10-274

Bonke, M., Thitamadee, S., Mähönen, A. P., Hauser, M. T., & Helariutta, Y. (2003). **APL regulates vascular tissue identity in Arabidopsis.** *Nature, 426*(6963), 181–186. https://doi.org/10.1038/nature02100

Briesemeister, S., Rahnenführer, J., & Kohlbacher, O. (2010). **YLoc-an interpretable web server for predicting subcellular localization.** *Nucleic Acids Research, 38*(SUPPL. 2). https://doi.org/10.1093/nar/gkq477

Brocard, C., & Hartig, A. (2006). **Peroxisome targeting signal 1: Is it really a simple tripeptide?** *Biochimica et Biophysica Acta - Molecular Cell Research*. https://doi.org/10.1016/j.bbamcr.2006.08.022

Burian, A., Raczyńska-Szajgin, M., Borowska-Wykręt, D., Piatek, A., Aida, M., & Kwiatkowska, D. (2015). **The CUP-SHAPED COTYLEDON2 and 3 genes have a post-meristematic effect on Arabidopsis thaliana phyllotaxis**. *Annals of Botany, 115*(5), 807–820. https://doi.org/10.1093/aob/mcv013

Byrne, M. E., Barley, R., Curtis, M., Arroyo, J. M., Dunham, M., Hudson, A, & Martienssen, R. A. (2000). **Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis.** *Nature, 408*(6815), 967–971. https://doi.org/10.1038/35050091

Cai, S., & Lashbrook, C. C. (2008). **Stamen Abscission Zone Transcriptome Profiling Reveals New Candidates for Abscission Control: Enhanced Retention of Floral Organs in Transgenic Plants Overexpressing Arabidopsis ZINC FINGER PROTEIN2.** *Plant Physiology, 146*(3), 1305–1321. https://doi.org/10.1104/pp.107.110908

Caño-Delgado, A., Lee, J.-Y., & Demura, T. (2010). **Regulatory mechanisms for specification and patterning of plant vascular tissues.** *Annual Review of Cell and Developmental Biology*, 26, 605–637. https://doi.org/10.1146/annurev-cellbio-100109-104107

Cao, Z. H., Zhang, S. Z., Wang, R. K., Zhang, R. F., & Hao, Y. J. (2013). **Genome Wide Analysis of the Apple MYB Transcription Factor Family Allows the Identification of MdoMYB121 Gene Confering Abiotic Stress Tolerance in Plants.** *PLoS ONE, 8*(7). https://doi.org/10.1371/journal.pone.0069955

Carter, J. F. (1993). **Potential of flaxseed and flaxseed oil in baked goods and other products in human nutrition.** *CEREAL FOODS WORLD, 38*(10), 753–759.

Chai, G., Wang, Z., Tang, X., Yu, L., Qi, G., Wang, D., & Zhou, G. (2014). **R2R3-MYB gene pairs in Populus: Evolution and contribution to secondary wall formation and flowering time.** *Journal of Experimental Botany, 65*(15), 4255–4269. https://doi.org/10.1093/jxb/eru196

Chantreau, M., Chabbert, B., Billiard, S., Hawkins, S., & Neutelings, G. (2015). **Functional analyses of cellulose synthase genes in flax (*Linum usitatissimum*) by virus-induced gene silencing.** *Plant Biotechnology Journal.* https://doi.org/10.1111/pbi.12350

Chantreau, M., Grec, S., Gutierrez, L., Dalmais, M., Pineau, C., Demailly, H., & Hawkins, S. (2013). **PT-Flax (phenotyping and TILLinG of flax): development of a flax (Linum usitatissimum L.) mutant population and TILLinG platform for forward and reverse genetics.** *BMC Plant Biology, 13*(1), 159. https://doi.org/10.1186/1471-2229-13-159

Chen, W., Provart, N. J., Glazebrook, J., Katagiri, F., Chang, H. S., Eulgem, T., & Zhu, T. (2002). **Expression Profile Matrix of Arabidopsis Transcription Factor Genes Suggests Their**

Putative Functions in Response to Environmental Stresses. *The Plant Cell, 14*(3), 559–574. https://doi.org/10.1105/tpc.010410

Cheng, H. Q., Han, L. B., Yang, C. L., Wu, X. M., Zhong, N. Q., Wu, J. H., & Xia, G. X. (2016). **The cotton MYB108 forms a positive feedback regulation loop with CML11 and participates in the defense response against Verticillium dahliae infection.** *Journal of Experimental Botany, 67*(6), 1935–1950. https://doi.org/10.1093/jxb/erw016

Cheng, H., Song, S., Xiao, L., Soo, H. M., Cheng, Z., Xie, D., & Peng, J. (2009). **Gibberellin acts through jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote stamen filament growth in Arabidopsis.** *PLoS Genetics, 5*(3). https://doi.org/10.1371/journal.pgen.1000440

Chernova, T. E., & Gorshkova, T. A. (2007). **Biogenesis of plant fibers.** *Russian Journal of Developmental Biology, 38*(4), 221–232. https://doi.org/10.1134/S1062360407040054

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B. H., Hong, X., Agarwal, M., & Zhu, J. K. (2003). **ICE1: A regulator of cold-induced transcriptome and freezing tolerance in arabidopsis.** *Genes and Development, 17*(8), 1043–1054. https://doi.org/10.1101/gad.1077503

Chinnusamy, V., Zhu, J., & Zhu, J. K. (2007). **Cold stress regulation of gene expression in plants.** *Trends in Plant Science, 12*(10), 444–51. https://doi.org/10.1016/j.tplants.2007.07.002

Chou, K. C., & Shen, H. Bin. (2010). **Plant-mPLoc: A top-down strategy to augment the power for predicting plant protein subcellular localization.** *PLoS ONE, 5*(6). https://doi.org/10.1371/journal.pone.0011335

Clough, S. J., & Bent, A. F. (1998). **Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*.** *Plant Journal, 16*(6), 735–743. https://doi.org/10.1046/j.1365-313X.1998.00343.x

Cloutier, S., Ragupathy, R., Miranda, E., Radovanovic, N., Reimer, E., Walichnowski, A., & Banik, M. (2012). **Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.).** *Theoretical and Applied Genetics, 125*(8), 1783–1795. https://doi.org/10.1007/s00122-012-1953-0

Cloutier, S., Ragupathy, R., Niu, Z., & Duguid, S. (2010). **SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits.** *Molecular Breeding, 28*(4), 437–451. https://doi.org/10.1007/s11032-010-9494-1

Creelman, R. A., & Mullet, J. E. (1997). **BIOSYNTHESIS AND ACTION OF JASMONATES IN PLANTS.** *Annual Review of Plant Physiology and Plant Molecular Biology, 48*(1), 355–381. https://doi.org/10.1146/annurev.arplant.48.1.355

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). **WebLogo: A sequence logo generator.** *Genome Research, 14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Czechowski, T. (2005). **Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis.** *Plant Physiology, 139*(1), 5–17. https://doi.org/10.1104/pp.105.063743

Dai, X., Xu, Y., Ma, Q., Xu, W., Wang, T., Xue, Y., & Chong, K. (2007). **Overexpression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic Arabidopsis.** *Plant Physiology, 143*(4), 1739–51. https://doi.org/10.1104/pp.106.094532

Dash, P. K., Cao, Y., Jailani, A. K., Gupta, P., Venglat, P., Xiang, D., & Datla, R. (2014). **Genome-wide analysis of drought induced gene expression changes in flax (*Linum usitatissimum*).** *GM Crops & Food, 5*(2), 106–19. https://doi.org/10.4161/gmcr.29742

Datia, R. S. S., Hammerlindl, J. K., Panchuk, B., Pelcher, L. E., & Keller, W. (1992). **Modified binary plant transformation vectors with the wild-type gene encoding NPTII.** *Gene, 122*(2), 383–384. https://doi.org/10.1016/0378-1119(92)90232-E

Day, A., Fénart, S., Neutelings, G., Hawkins, S., Rolando, C., & Tokarski, C. (2013). **Identification of cell wall proteins in the flax (*Linum usitatissimum*) stem.** *Proteomics, 13*(5), 812–825. https://doi.org/10.1002/pmic.201200257

Day, A., Addi, M., Kim, W., David, H., Bert, F., Mesnage, P., & Hawkins, S. (2005). **ESTs from the fiber-bearing stem tissues of flax (*Linum usitatissimum* L.): Expression analyses of sequences related to cell wall development.** *Plant Biology, 7*(1), 23–32. https://doi.org/10.1055/s-2004-830462

De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., & Hulo, N. (2006). **ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.** *Nucleic Acids Research, 34*(WEB. SERV. ISS.). https://doi.org/10.1093/nar/gkl124

De Duve, C., & Baudhuin, P. (1966). **Peroxisomes (microbodies and related particles).** *Physiological Reviews, 46*(2), 323–357.

De Vos, M., Denekamp, M., Dicke, M., Vuylsteke, M., Van Loon, L., Smeekens, S. C., & Pieterse, C. M. (2006). **The *Arabidopsis thaliana* Transcription Factor AtMYB102 Functions in**

Defense Against the Insect Herbivore Pieris rapae. *Plant Signaling & Behavior, 1*(6), 305–311. https://doi.org/10.4161/psb.1.6.3512

Deluc, L., Barrieu, F. C., Marchive, C., Lauvergeat, V., Decendit, A., Richard, T., & Hamdi, S. (2006). **Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway.** *Plant Physiology, 140*(2), 499. https://doi.org/10.1104/pp.105.067231

Depuydt, S., Rodriguez-Villalon, A., Santuari, L., Wyser-Rmili, C., Ragni, L., & Hardtke, C. S. (2013). **Suppression of Arabidopsis protophloem differentiation and root meristem growth by CLE45 requires the receptor-like kinase BAM3.** *Proceedings of the National Academy of Sciences, 110*(17), 7074–7079. https://doi.org/10.1073/pnas.1222314110

Deyholos, M. K. (2006). **Bast fiber of flax (*Linum usitatissimum* L.): Biological foundations of its ancient and modern uses.** *Israel Journal of Plant Sciences, 54*(4), 273–280. Retrieved from http://www.tandfonline.com/doi/abs/10.1560/IJPS_54_4_273

Diederichsen, A., & Hammer, K. (1995). **Variation of cultivated flax (*Linum usitatissimum* L. subsp. *usitatissimum*) and its wild progenitor pale flax (*subsp. angustifolium* (Huds.) Thell.).** *Genetic Resources and Crop Evolution, 42*(3), 263–272. https://doi.org/10.1007/BF02431261

Dillman, A. C. (1938). **Natural Crossing in Flax.** *Agronomy Journal, 30*(4), 279. https://doi.org/10.2134/agronj1938.00021962003000040002x

Ding, Y., Li, H., Zhang, X., Xie, Q., Gong, Z., & Yang, S. (2015). **OST1 kinase modulates freezing tolerance by enhancing ICE1 stability in *Arabidopsis*.** *Developmental Cell, 32*(3), 278–289. https://doi.org/10.1016/j.devcel.2014.12.023

Dong, C.-H., Zolman, B. K., Bartel, B., Lee, B., Stevenson, B., Agarwal, M., & Zhu, J. K. (2009). **Disruption of *Arabidopsis* CHY1 reveals an important role of metabolic status in plant cold stress signaling.** *Molecular Plant, 2*(1), 59–72. https://doi.org/10.1093/mp/ssn063

Donze, T., Qu, F., Twigg, P., & Morris, T. J. (2014). **Turnip crinkle virus coat protein inhibits the basal immune response to virus invasion in Arabidopsis by binding to the NAC transcription factor TIP.** *Virology, 449*, 207–214. https://doi.org/10.1016/j.virol.2013.11.018

Du, H., Feng, B. R., Yang, S. S., Huang, Y. B., & Tang, Y. X. (2012a). **The R2R3-MYB Transcription Factor Gene Family in Maize.** *PLoS ONE, 7*(6), e37463. https://doi.org/10.1371/journal.pone.0037463

Du, H., Wang, Y. Bin, X, Y., Liang, Z., Jiang, S. J., Zhang, S. S., & Tang, Y. X. (2013). **Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants.** *DNA Research, 20*(5), 437–448. https://doi.org/10.1093/dnares/dst021

Du, H., Yang, S. S., Liang, Z., Feng, B. R., Liu, L., Huang, Y. B., & Tang, Y. X. (2012b). **Genome-wide analysis of the MYB transcription factor superfamily in soybean.** *BMC Plant Biology, 12*(1), 106. https://doi.org/10.1186/1471-2229-12-106

Dubos, C., Le Gourrierec, J., Baudry, A., Huep, G., Lanet, E., Debeaujon, I., & Lepiniec, L. (2008). **MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*.** *Plant Journal, 55*(6), 940–953. https://doi.org/10.1111/j.1365-313X.2008.03564.x

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). **MYB transcription factors in *Arabidopsis*.** *Trends in Plant Science.* https://doi.org/10.1016/j.tplants.2010.06.005

Duval, M., Hsieh, T. F., Kim, S. Y., & Thomas, T. L. (2002). **Molecular characterization of AtNAM: A member of the Arabidopsis NAC domain superfamily.** *Plant Molecular Biology, 50*(2), 237–248. https://doi.org/10.1023/A:1016028530943

Edgar, R. C. (2004). **MUSCLE: Multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Research, 32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Elliott, K. A., & Shirsat, A. H. (1998). **Promoter regions of the extA extensin gene from *Brassica napus* control activation in response to wounding and tensile stress.** *Plant Molecular Biology, 37*(4), 675–687. https://doi.org/10.1023/A:1005918816630

Elmayan, T., & Tepfer, M. (1995). **Evaluation in tobacco of the organ specificity and strength of the rolD promoter, domain A of the 35S promoter and the 35S2 promoter.** *Transgenic Research, 4*(6), 388–396. https://doi.org/10.1007/BF01973757

Emanuelsson, O., Brunak, S., von Heijne, G., & Nielsen, H. (2007). **Locating proteins in the cell using TargetP, SignalP and related tools.** *Nature Protocols, 2*(4), 953–971. https://doi.org/10.1038/nprot.2007.131

Emrich, S. J., Barbazuk, W. B., Li, L., & Schnable, P. S. (2006). **Gene discovery and annotation using LCM-454 transcriptome sequencing.** *Genome Research, 17*: 69–73. https://doi.org/10.1101/gr.5145806

Ernst, H. A., Nina O, A., Skriver, K., Larsen, S., & Lo Leggio, L. (2004). **Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors.** *EMBO Reports, 5*(3), 297–303. https://doi.org/10.1038/sj.embor.7400093

Fahn, A. (1982). **Plant anatomy.** Retrieved January 6, 2016, from http://eds.b.ebscohost.com/eds/detail/detail?vid=6&sid=268000f0-2fcd-4b97-a96e-

4ab4f4dbbb16%40sessionmgr120&hid=117&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=alb.433421&db=cat03710a

Fenart, S., Ndong, Y. P. A., Duarte, J., Rivière, N., Wilmer, J., Van Wuytswinkel, O., & Thomasset, B. (2010). **Development and validation of a flax (*Linum usitatissimum* L.) gene expression oligo microarray.** *BMC Genomics, 11*(1), 592. https://doi.org/10.1186/1471-2164-11-592

Fernández-Marcos, M., Desvoyes, B., Manzano, C., Liberman, L. M., Benfey, P. N., Del Pozo, J. C., & Gutierrez, C. (2017). **Control of Arabidopsis lateral root primordium boundaries by MYB36.** *New Phytologist, 213*(1), 105–112. https://doi.org/10.1111/nph.14304

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., & Eddy, S. R. (2015a). **HMMER web server: 2015 Update.** *Nucleic Acids Research, 43*(W1), W30–W38. https://doi.org/10.1093/nar/gkv397

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., & Bateman, A. (2015b). **The Pfam protein families database: towards a more sustainable future.** *Nucleic Acids Research, 44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

Fleming, A. J. (2006). **The co-ordination of cell division, differentiation and morphogenesis in the shoot apical meristem: A perspective.** *In Journal of Experimental Botany ,57*, 25–32. https://doi.org/10.1093/jxb/eri268

Footitt, S., Slocombe, S. P., Larner, V., Kurup, S., Wu, Y., Larson, T., & Holdsworth, M. (2002). **Control of germination and lipid mobilization by COMATOSE, the Arabidopsis homologue of human ALDP.** *EMBO Journal, 21*(12), 2912–2922. https://doi.org/10.1093/emboj/cdf300

Fu, Y. B., & Allaby, R. G. (2010). **Phylogenetic network of Linum species as revealed by non-coding chloroplast DNA sequences.** *Genetic Resources and Crop Evolution, 57*(5), 667–677. https://doi.org/10.1007/s10722-009-9502-7

Furuta, K. M., Yadav, S. R., Lehesranta, S., Belevich, I., Miyashima, S., Heo, J. O., & Helariutta, Y. (2014). **Plant development. Arabidopsis NAC45/86 direct sieve element morphogenesis culminating in enucleation.** *Science (New York, N.Y.), 345*(6199), 933–937. https://doi.org/10.1126/science.1253736

Galindo-González, L., & Deyholos, M. K. (2016). **RNA-seq Transcriptome Response of Flax (*Linum usitatissimum* L.) to the Pathogenic Fungus *Fusarium oxysporum f.* sp. *lini*.** *Frontiers in Plant Science, 7*, 1766. https://doi.org/10.3389/fpls.2016.01766

Galindo-González, L., Pinzón-Latorre, D., Bergen, E. A., Jensen, D. C., & Deyholos, M. K. (2015). **Ion Torrent sequencing as a tool for mutation discovery in the flax (*Linum usitatissimum* L.) genome.** *Plant Methods, 11*(1), 19. https://doi.org/10.1186/s13007-015-0062-x

Garcia-Hernandez, M., Berardini, T. Z., Chen, G., Crist, D., Doyle, A., Huala, E., & Zhang, P. (2002). **TAIR: A resource for integrated Arabidopsis data.** *Functional and Integrative Genomics.* https://doi.org/10.1007/s10142-002-0077-z

Geraghty, M. T., Bassett, D., Morrell, J. C., Gatto, G. J., Bai, J., Geisbrecht, B. V, & Gould, S. J. (1999). **Detecting patterns of protein distribution and gene expression in silico.** *Proceedings of the National Academy of Sciences of the United States of America, 96*(March), 2937–2942. https://doi.org/10.1073/pnas.96.6.2937

Gilmour, S. J., Artus, N. N., & Thomashow, M. F. (1992). **cDNA sequence analysis and expression of two cold-regulated genes of *Arabidopsis thaliana*.** *Plant Mol Biol, 18*(1), 13–21.

Gilmour, S. J., Zarka, D. G., Stockinger, E. J., Salazar, M. P., Houghton, J. M., & Thomashow, M. F. (1998). **Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression.** *Plant Journal, 16*(4), 433–442. https://doi.org/10.1046/j.1365-313X.1998.00310.x

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., & Rokhsar, D. S. (2012). **Phytozome: A comparative platform for green plant genomics.** *Nucleic Acids Research, 40*(D1). https://doi.org/10.1093/nar/gkr944

Gorshkova, T. A., Chemikosova, S. B., Sal'nikov, V. V., Pavlencheva, N. V., Gur'janov, O. P., Stolle-Smits, T., & Van Dam, J. E. G. (2004). **Occurrence of cell-specific galactan is coinciding with bast fiber developmental transition in flax.** *Industrial Crops and Products, 19*, 217–224. https://doi.org/10.1016/j.indcrop.2003.10.002

Gorshkova, T. A., Sal'nikov, V. V., Chemikosova, S. B., Ageeva, M. V., Pavlencheva, N. V., & van Dam, J. E. G. (2003). **The snap point: a transition point in *Linum usitatissimum* bast fiber development.** *Industrial Crops and Products, 18*(3), 213–221. https://doi.org/10.1016/S0926-6690(03)00043-8

Gorshkova, T. A., Ageeva, M., Chemikosova, S., & Salnikov, V. (2005). **Tissue-specific processes during cell wall formation in flax fiber.** *Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology, 139*(1), 88–92. https://doi.org/10.1080/11263500500056070

Gorshkova, T. A., Brutch, N., Chabbert, B., Deyholos, M., Hayashi, T., Lev-Yadun, S., & Pilate, G. (2012). **Plant Fiber Formation: State of the Art, Recent and Expected Progress, and Open**

Questions. *Critical Reviews in Plant Sciences, 31*(3), 201–228. https://doi.org/10.1080/07352689.2011.616096

Gorshkova, T., Gurjanov, O. P., Mikshina, P. V., Ibragimova, N. N., Mokshina, N. E., Salnikov, V. V., & Chemikosova, S. B. (2010). **Specific type of secondary cell wall formed by plant fibers.** *Russian Journal of Plant Physiology, 57*(3), 328–341. https://doi.org/10.1134/S1021443710030040

Gorshkova, T., & Morvan, C. (2006). **Secondary cell-wall assembly in flax phloem fibers: Role of galactans.** *Planta, 223*(2), 149–158. https://doi.org/10.1007/s00425-005-0118-7

Graham, I. A. (2008). **Seed Storage Oil Mobilization.** *Annual Review of Plant Biology, 59*(1), 115–142. https://doi.org/10.1146/annurev.arplant.59.032607.092938

Grant, E. H., Fujino, T., Beers, E. P., & Brunner, A. M. (2010). **Characterization of NAC domain transcription factors implicated in control of vascular cell differentiation in Arabidopsis and *Populus***. *Planta, 232*(2), 337–352. https://doi.org/10.1007/s00425-010-1181-2

Grishin, N. V. (1995). **Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites.** *Journal of Molecular Evolution, 41*(5), 675–679. https://doi.org/10.1007/BF00175826

Guo, Y., & Gan, S. (2006). **AtNAP, a NAC family transcription factor, has an important role in leaf senescence.** *Plant Journal, 46*(4), 601–612. https://doi.org/10.1111/j.1365-313X.2006.02723.x

Gurjanov, O. P., Ibragimova, N. N., Gnezdilov, O. I., & Gorshkova, T. A. (2008). **Polysaccharides, tightly bound to cellulose in cell wall of flax bast fiber: Isolation and identification.** *Carbohydrate Polymers, 72*(4), 719–729. https://doi.org/10.1016/j.carbpol.2007.10.017

Haerizadeh, F., Wong, C. E., Singh, M. B., & Bhalla, P. L. (2009). **Genome-wide analysis of gene expression in soybean shoot apical meristem.** *Plant Molecular Biology, 69*(6), 711–727. https://doi.org/10.1007/s11103-008-9450-1

Haga, N., Kato, K., Murase, M., Araki, S., Kubo, M., Demura, T., & Ito, M. (2007). **R1R2R3-Myb proteins positively regulate cytokinesis through activation of KNOLLE transcription in *Arabidopsis thaliana*.** *Development, 134*(6), 1101–1110. https://doi.org/10.1242/dev.02801

Haga, N., Kobayashi, K., Suzuki, T., Maeo, K., Kubo, M., Ohtani, M., & Ito, M. (2011). **Mutations in MYB3R1 and MYB3R4 Cause Pleiotropic Developmental Defects and Preferential Down-Regulation of Multiple G2/M-Specific Genes in Arabidopsis.** *Plant Physiology, 157*(2), 706–717. https://doi.org/10.1104/pp.111.180836

Hawkins, J., & Bodén, M. (2006). **Detecting and sorting targeting peptides with neural networks and support vector machines.** *Journal of Bioinformatics and Computational Biology, 4*(1), 1–18. https://doi.org/10.1142/S0219720006001771

Hay, A. (2006). **ASYMMETRIC LEAVES1 and auxin activities converge to repress BREVIPEDICELLUS expression and promote leaf development in Arabidopsis.** *Development, 133*(20), 3955–3961. https://doi.org/10.1242/dev.02545

Hayashi, M. (2002). **Ped3p is a Peroxisomal ATP-Binding Cassette Transporter that might Supply Substrates for Fatty Acid beta-Oxidation.** *Plant and Cell Physiology, 43*(1), 1–11. https://doi.org/10.1093/pcp/pcf023

Hayashi, M., Toriyama, K., Kondo, M., & Nishimura, M. (1998). **2,4-Dichlorophenoxybutyric acid-resistant mutants of Arabidopsis have defects in glyoxysomal fatty acid beta-oxidation.** *The Plant Cell, 10*(2), 183–195. https://doi.org/10.1105/tpc.10.2.183

He, Q., Jones, D. C., Li, W., Xie, F., Ma, J., Sun, R., & Zhang, B. (2016). **Genome-Wide Identification of R2R3-MYB Genes and Expression Analyses During Abiotic Stress in *Gossypium raimondii*.** *Scientific Reports, 6*(1), 22980. https://doi.org/10.1038/srep22980

He, X. J., Mu, R. L., Cao, W. H., Zhang, Z. G., Zhang, J. S., & Chen, S. Y. (2005). **AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development.** *The Plant Journal, 44*(6), 903–916. https://doi.org/10.1111/j.1365-313X.2005.02575.x

Hejátko, J., Ryu, H., Kim, G. T., Dobesová, R., Choi, S., Choi, S. M., & Hwang, I. (2009). **The histidine kinases CYTOKININ-INDEPENDENT1 and ARABIDOPSIS HISTIDINE KINASE2 and 3 regulate vascular tissue development in Arabidopsis shoots.** *The Plant Cell, 21*(7), 2008–21. https://doi.org/10.1105/tpc.109.066696

Hendelman, A., Stav, R., Zemach, H., & Arazi, T. (2013). **The tomato NAC transcription factor SlNAM2 is involved in flower-boundary morphogenesis.** *Journal of Experimental Botany, 64*(18), 5497–5507. https://doi.org/10.1093/jxb/ert324

Henikoff, S., & Henikoff, J. G. (1992). **Amino acid substitution matrices from protein blocks.** *Proceedings of the National Academy of Sciences, 89*(22), 10915–10919. https://doi.org/10.1073/pnas.89.22.10915

Hibara, K. I., Karim, M. R., Takada, S., Taoka, K. I., Furutani, M., Aida, M., & Tasaka, M. (2006). **Arabidopsis CUP-SHAPED COTYLEDON3 Regulates Postembryonic Shoot Meristem and Organ Boundary Formation.** *The Plant Cell Online, 18*(11), 2946–2957. https://doi.org/10.1105/tpc.106.045716

Higo, K. (1998). **PLACE: a database of plant cis-acting regulatory DNA elements.** *Nucleic Acids Research, 26*(1), 358–359. https://doi.org/10.1093/nar/26.1.358

Higo, K., Ugawa, Y., Iwamoto, M., & Korenaga, T. (1999). **Plant cis-acting regulatory DNA elements (PLACE) database: 1999.** *Nucleic Acids Research.* https://doi.org/10.1093/nar/27.1.297

Hirakawa, Y., Kondo, Y., & Fukuda, H. (2011). **Establishment and maintenance of vascular cell communities through local signaling.** *Current Opinion in Plant Biology, 14*(1), 17–23. https://doi.org/10.1016/j.pbi.2010.09.011

Hobo, T., Asada, M., Kowyama, Y., & Hattori, T. (1999). **ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent.** *Plant Journal, 19*(6), 679–689. https://doi.org/10.1046/j.1365-313X.1999.00565.x

Hojoung, L., Guo, Y., Ohta, M., Xiong, L., Stevenson, B., & Zhu, J. K. (2002). **LOS2, a genetic locus required for cold-responsive gene transcription encodes a bi-functional enolase.** *EMBO Journal, 21*(11), 2692–2702. https://doi.org/10.1093/emboj/21.11.2692

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). **WoLF PSORT: Protein localization predictor.** *Nucleic Acids Research, 35*(SUPPL.2). https://doi.org/10.1093/nar/gkm259

Hotte, N. S. C., & Deyholos, M. K. (2008). **A flax fiber proteome: identification of proteins enriched in bast fibers.** *BMC Plant Biology, 8*, 52. https://doi.org/10.1186/1471-2229-8-52

Howe, E., Holton, K., Nair, S., Schlauch, D., Sinha, R., & Quackenbush, J. (2010). **MeV: MultiExperiment viewer.** *In Biomedical Informatics for Cancer Research,* 267–277. https://doi.org/10.1007/978-1-4419-5714-6_15

Hradilová, J., Rehulka, P., Rehulková, H., Vrbová, M., Griga, M., & Brzobohatý, B. (2010). **Comparative analysis of proteomic changes in contrasting flax cultivars upon cadmium exposure.** *Electrophoresis, 31*(2), 421–31. https://doi.org/10.1002/elps.200900477

Hu, J. (2002). **A Role for Peroxisomes in Photomorphogenesis and Development of *Arabidopsis*.** *Science, 297*(5580), 405–409. https://doi.org/10.1126/science.1073633

Hu, J., Baker, A., Bartel, B., Linka, N., Mullen, R. T., Reumann, S., & Zolman, B. K. (2012). **Plant Peroxisomes: Biogenesis and Function.** *The Plant Cell, 24*(6), 2279–2303. https://doi.org/10.1105/tpc.112.096586

Hu, R., Qi, G., Kong, Y., Kong, D., Gao, Q., & Zhou, G. (2010). **Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*.** *BMC Plant Biology, 10*, 145. https://doi.org/10.1186/1471-2229-10-145

Huang, D., Wu, W., Abrams, S. R., & Cutler, A. J. (2008). **The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors.** *Journal of Experimental Botany, 59*(11), 2991–3007. https://doi.org/10.1093/jxb/ern155

Huis, R., Hawkins, S., & Neutelings, G. (2010). **Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.).** *BMC Plant Biology, 10*, 71. https://doi.org/10.1186/1471-2229-10-71

Huis, R., Morreel, K., Fliniaux, O., Lucau-Danila, A., Fenart, S., Grec, S., & Hawkins, S. (2012). **Natural Hypolignification Is Associated with Extensive Oligolignol Accumulation in Flax Stems.** *Plant Physiology, 158*(4), 1893–1915. https://doi.org/10.1104/pp.111.192328

Hussey, S. G., Mizrachi, E., Spokevicius, A. V, Bossinger, G., Berger, D. K., & Myburg, A. A. (2011). **SND2, a NAC transcription factor gene, regulates genes involved in secondary cell**

wall development in *Arabidopsis* fibers and increases fiber cell area in *Eucalyptus*. *BMC Plant Biology, 11*(1), 173. https://doi.org/10.1186/1471-2229-11-173

Ishiguro, S., & Nakamura, K. (1992). **The nuclear factor SP8BF binds to the 5′-upstream regions of three different genes coding for major proteins of sweet potato tuberous roots.** *Plant Molecular Biology, 18*(1), 97–108. https://doi.org/10.1007/BF00018460

Ito, M. (2005). **Conservation and diversification of three-repeat Myb transcription factors in plants.** *Journal of Plant Research.* https://doi.org/10.1007/s10265-005-0192-8

Ito, M., Araki, S., Matsunaga, S., Itoh, T., Nishihama, R., Machida, Y., & Watanabe, A. (2001). **G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors.** *The Plant Cell, 13*(8), 1891–1905. https://doi.org/10.1105/TPC.010102

Jacobs, W. P. (1952). **The Role of auxin in the differentiation of xylem round a wound.** *Americam Journal of Botanics, 39*, 301. https://doi.org/10.2307/2438258

Jensen, M. K., Kjaersgaard, T., Nielsen, M. M., Galberg, P., Petersen, K., O'Shea, C., & Skriver, K. (2010). **The *Arabidopsis thaliana* NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling.** *The Biochemical Journal, 426*(2), 183–196. https://doi.org/10.1042/BJ20091234

Ji, J., Shimizu, R., Sinha, N., & Scanlon, M. J. (2010). **WOX4 Promotes Procambial Development.** *Plant Signaling & Behavior, 5*(7), 916–20. https://doi.org/10.1104/pp.109.149641

Jiang, B., Shi, Y., Zhang, X., Xin, X., Qi, L., Guo, H., & Yang, S. (2017). **PIF3 is a negative regulator of the CBF pathway and freezing tolerance in *Arabidopsis*.** *Proceedings of the National Academy of Sciences of the United States of America, 114*(32), E6695–E6702. https://doi.org/10.1073/pnas.1706226114

Jiang, C., Gu, J., Chopra, S., Gu, X., & Peterson, T. (2004a). **Ordered origin of the typical two- and three-repeat Myb genes.** *Gene, 326*(1–2), 13–22. https://doi.org/10.1016/j.gene.2003.09.049

Jiang, C., Gu, X., & Peterson, T. (2004b). **Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*.** *Genome Biology, 5*(7), R46. https://doi.org/10.1186/gb-2004-5-7-r46

Jiang, G., Jiang, X., Lü, P., Liu, J., Gao, J., & Zhang, C. (2014). **The rose (*Rosa hybrida*) NAC transcription factor 3 gene, RhNAC3, involved in ABA signaling pathway both in rose and *Arabidopsis*.** *PLoS ONE, 9*(10). https://doi.org/10.1371/journal.pone.0109415

Jiao, Y., Tausta, S. L., Gandotra, N., Sun, N., Liu, T., Clay, N. K., & Nelson, T. (2009). **A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies.** *Nature Genetics, 41*(2), 258–263. https://doi.org/10.1038/ng.282

Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2017). **PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants.** *Nucleic Acids Research, 45(*D1), D1040–D1045. https://doi.org/10.1093/nar/gkw982

Jin, J., Zhang, H., Kong, L., Gao, G., & Luo, J. (2014). **PlantTFDB 3.0: A portal for the functional and evolutionary study of plant transcription factors.** *Nucleic Acids Research, 42*(D1). https://doi.org/10.1093/nar/gkt1016

Johnson, E. T., & Dowd, P. F. (2004). **Differentially enhanced insect resistance, at a cost, in *Arabidopsis thaliana* constitutively expressing a transcription factor of defensive metabolites.** *Journal of Agricultural and Food Chemistry, 52*(16), 5135–5138. https://doi.org/10.1021/jf0308049

Kamiya, T., Borghi, M., Wang, P., Danku, J., Kalmbach, L., Hosmani, P., Naseer, S., Fujiwara, T., Geldner, N., & Salt, D., E. (2015). **The MYB36 transcription factor orchestrates Casparian strip formation.** *Proceedings of the National Academy of Sciences of the United States of the America, 112*(33), 10533.

Esau, K. (1942). **Vascular Differentiation in the Vegetative Shoot of Linum . I . The Procambium.** *American Journal of Botany, 29*(9), 738–747.

Esau, K. (1965). **Plant Anatomy.** 2nd edition. John Wiley& Sons Inc. New York. Retrieved from http://eds.b.ebscohost.com/eds/detail/detail?vid=4&sid=268000f0-2fcd-4b97-a96e-4ab4f4dbbb16%40sessionmgr120&hid=117&bdata=JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=alb.812124&db=cat03710a

Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., & Bansal, K. C. (2012). **Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis.** *BMC Genomics, 13*, 544. https://doi.org/10.1186/1471-2164-13-544

Katoh, K., Asimenos, G., & Toh, H. (2009). **Multiple alignment of DNA sequences with MAFFT.** *Methods in Molecular Biology, 537*, 39–64. https://doi.org/10.1007/978-1-59745-251-9_3

Katoh, K., & Standley, D. M. (2013). **MAFFT multiple sequence alignment software version 7: Improvements in performance and usability.** *Molecular Biology and Evolution, 30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kawagoe, Y., Campell, B. R., & Murai, N. (1994). **Synergism between CACGTG (G-box) and CACCTG cis-elements is required for activation of the bean seed storage protein β-phaseolin gene.** *The Plant Journal, 5*(6), 885–890. https://doi.org/10.1046/j.1365-313X.1994.5060885.x

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). **The Phyre2 web portal for protein modeling, prediction and analysis.** *Nature Protocols, 10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053

Kerstetter, R. A., Bollman, K., Taylor, R. A., Bomblies, K., & Poethig, R. S. (2001). **KANADI regulates organ polarity in Arabidopsis.** *Nature, 411*(6838), 706–709. https://doi.org/10.1038/35079629

Kertész, S., Kerényi, Z., Mérai, Z., Bartos, I., Pálfy, T., Barta, E., & Silhavy, D. (2006). **Both introns and long 3′-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants.** *Nucleic Acids Research, 34*(21), 6147–6157. https://doi.org/10.1093/nar/gkl737

Kim, Y. S., Kim, S. G., Park, J. E., Park, H. Y., Lim, M. H., Chua, N. H., & Park, C. M. (2006). **A membrane-bound NAC transcription factor regulates cell division in *Arabidopsis*.** *The Plant Cell, 18*(11), 3132–3144. https://doi.org/10.1105/tpc.106.043018

Kim, Y. S., & Park, C. M. (2007). **Membrane regulation of cytokinin-mediated cell division in *Arabidopsis*.** *Plant Signaling & Behavior, 2*(1), 15–16.

Klempnauer, K. H., Gonda, T. J., & Bishop, J. M. (1982). **Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: The architecture of a transduced oncogene.** *Cell, 31*(2 PART 1), 453–463. https://doi.org/10.1016/0092-8674(82)90138-6

Knight, H., Zarka, D. G., Okamoto, H., Thomashow, M. F., & Knight, M. R. (2004). **Abscisic acid induces CBF gene transcription and subsequent induction of cold-regulated genes via the CRT promoter element.** *Plant Physiology, 135*(3), 1710–1717. https://doi.org/10.1104/pp.104.043562

Knight, M. R., & Knight, H. (2012). **Low-temperature perception leading to gene expression and cold tolerance in higher plants.** *New Phytologist.* https://doi.org/10.1111/j.1469-8137.2012.04239.x

Kobayashi, K., Suzuki, T., Iwata, E., Nakamichi, N., Suzuki, T., Chen, P., & Ito, M. (2015). **Transcriptional repression by MYB3R proteins regulates plant organ growth.** *The EMBO Journal, 34*(15), 1992–2007. https://doi.org/10.15252/embj.201490899

Koichiro, T., Daniel, P., Nicholas, P., Glen, S., Masatoshi, N., & Sudhir, K. (2011). **MEGA 5.** *Molecular Biology and Evolution.*

Kragler, F., Lametschwandtner, G., Christmann, J., Hartig, A., & Harada, J. J. (1998). **Identification and analysis of the plant peroxisomal targeting signal 1 receptor NtPEX5.** *Proceedings of the National Academy of Sciences of the United States of America, 95*(22), 13336–41. https://doi.org/10.1073/pnas.95.22.13336

Kranz, H. D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., & Weisshaar, B. (1998). **Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*.** *Plant Journal., 16*(2), 263–276. https://doi.org/10.1046/j.1365-313x.1998.00278.x

Kranz, H., Scholz, K., & Weisshaar, B. (2000). **c-MYB oncogene-like genes encoding three MYB repeats occur in all major plant lineages.** *Plant Journal, 21*(2), 231–235. https://doi.org/10.1046/j.1365-313X.2000.00666.x

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). **Predicting transmembrane protein topology with a hidden Markov model: application to complete**

genomes. *Journal of Molecular Biology, 305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., & Demura, T. (2005). **Transcription switches for protoxylem and metaxylem vessel formation.** *Genes & Development, 19*(16), 1855–1860. https://doi.org/10.1101/gad.1331305

Kumar, S., Jordan, M. C., Datla, R., & Cloutier, S. (2013). **The LuWD40-1 Gene Encoding WD Repeat Protein Regulates Growth and Pollen Viability in Flax (*Linum Usitatissimum* L.).** *PLoS ONE, 8*(7), 1–10. https://doi.org/10.1371/journal.pone.0069124

Kurkela, S., & Franck, M. (1990). **Cloning and characterization of a cold- and ABA-inducible Arabidopsis gene.** *Plant Molecular Biology, 15*(1), 137–144. https://doi.org/10.1104/pp.111.179838

Kvavadze, E., Bar-Yosef, O., Belfer-Cohen, A., Boaretto, E., Jakeli, N., Matskevich, Z., & Meshveliani, T. (2009). **30,000-Year-Old Wild Flax Fibers.** *Science, 325*(5946), 1359–1359. https://doi.org/10.1126/science.1175404

Lametschwandtner, G., Brocard, C., Fransen, M., Van Veldhoven, P., Berger, J., & Hartig, A. (1998). **The difference in recognition of terminal tripeptides as peroxisomal targeting signal I between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent to it.** *Journal of Biological Chemistry, 273*(50), 33635–33643. https://doi.org/10.1074/jbc.273.50.33635

Lee, B. H. (2002). **A Mitochondrial Complex I Defect Impairs Cold-Regulated Nuclear Gene Expression.** *The Plant Cell. 14*(6), 1235–1251. https://doi.org/10.1105/tpc.010433

Lee, B. H., Jeon, J. O., Lee, M. M., & Kim, J. H. (2015). **Genetic interaction between GROWTH-REGULATING FACTOR and CUP-SHAPED COTYLEDON in organ separation.** *Plant Signaling & Behavior, 10*(2), e988071. https://doi.org/10.4161/15592324.2014.988071

Lee, D. K., Geisler, M., & Springer, P. S. (2009). **LATERAL ORGAN FUSION1 and LATERAL ORGAN FUSION2 function in lateral organ separation and axillary meristem formation in *Arabidopsis*.** *Development, 136*(14), 2423–2432. https://doi.org/10.1242/dev.031971

Lee, M. H., Jeon, H. S., Kim, H. G., & Park, O. K. (2017). **An Arabidopsis NAC transcription factor NAC4 promotes pathogen-induced cell death under negative regulation by microRNA164.** *New Phytologist, 214*(1), 343–360. https://doi.org/10.1111/nph.14371

Legay, S., Sivadon, P., Blervacq, A. S., Pavy, N., Baghdady, A., Tremblay, L., & Grima-Pettenati, J. (2010). **EgMYB1, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in *Arabidopsis* and poplar.** *New Phytologist, 188*(3), 774–786. https://doi.org/10.1111/j.1469-8137.2010.03432.x

Li, X., Xue, C., Li, J., Qiao, X., Li, L., Yu, L., & Wu, J. (2016). **Genome-Wide Identification, Evolution and Functional Divergence of MYB Transcription Factors in Chinese White Pear (*Pyrus bretschneideri*).** *Plant and Cell Physiology, 57*(4), 824–847. https://doi.org/10.1093/pcp/pcw029

Li, Z., Peng, R., Tian, Y., Han, H., Xu, J., & Yao, Q. (2016). **Genome-Wide Identification and Analysis of the MYB Transcription Factor Superfamily in *Solanum lycopersicum*.** *Plant & Cell Physiology, 57*(8), 1657–77. https://doi.org/10.1093/pcp/pcw091

Liang, D., Wong, C. E., Singh, M. B., Beveridge, C. A., Phipson, B., Smyth, G. K., & Bhalla, P. L. (2009). **Molecular dissection of the pea shoot apical meristem.** *Journal of Experimental Botany, 60*(14), 4201–4213. https://doi.org/10.1093/jxb/erp254

Liberman, L. M., Sparks, E. E., Moreno-Risueno, M. A., Petricka, J. J., & Benfey, P. N. (2015). **MYB36 regulates the transition from proliferation to differentiation in the *Arabidopsis* root.** *Proceedings of the National Academy of Sciences, 112*(39), 12099–12104. https://doi.org/10.1073/pnas.1515576112

Lipsick, J. S. (1996). **One billion years of Myb.** *Oncogene, 13*(2), 223–235.

Liu, Q. (1998). **Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low-Temperature-Responsive Gene Expression, Respectively, in *Arabidopsis*.** *The Plant Cell, 10*(8), 1391–1406. https://doi.org/10.1105/tpc.10.8.1391

Livak, K. J., & Schmittgen, T. D. (2001). **Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method.** *Methods, 25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Lu, S. X., Knowles, S. M., Andronis, C., Ong, M. S., & Tobin, E. M. (2009). **CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL Function Synergistically in the Circadian Clock of Arabidopsis.** *Plant Physiology, 150*(2), 834–843. https://doi.org/10.1104/pp.108.133272

Lucas, K. A., Filley, J. R., Erb, J. M., Graybill, E. R., & Hawes, J. W. (2007). **Peroxisomal metabolism of propionic acid and isobutyric acid in plants.** *Journal of Biological Chemistry, 282*(34), 24980–24989. https://doi.org/10.1074/jbc.M701028200

Mähönen, A. P., Higuchi, M., Törmäkangas, K., Miyawaki, K., Pischke, M. S., Sussman, M. R., & Kakimoto, T. (2006a). **Cytokinins regulate a bidirectional phosphorelay network in *Arabidopsis*.** *Current Biology : CB, 16*(11), 1116–22. https://doi.org/10.1016/j.cub.2006.04.030

Mähönen, A. P., Bishopp, A., Higuchi, M., Nieminen, K. M., Kinoshita, K., Törmäkangas, K., & Helariutta, Y. (2006b). **Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development.** *Science (New York, N.Y.), 311*(5757), 94–98. https://doi.org/10.1126/science.1118875

Mähönen, A. P., Bonke, M., Kauppinen, L., Riikonen, M., Benfey, P. N., & Helariutta, Y. (2000). **A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root.** *Genes and Development, 14*(23), 2938–2943. https://doi.org/10.1101/gad.189200

Mandaokar, A., & Browse, J. (2008). **MYB108 Acts Together with MYB24 to Regulate Jasmonate-Mediated Stamen Maturation in *Arabidopsis*.** *Plant Physiology, 149*(2), 851–862. https://doi.org/10.1104/pp.108.132597

Mao, C., Ding, W., Wu, Y., Yu, J., He, X., Shou, H., & Wu, P. (2007). **Overexpression of a NAC-domain protein promotes shoot branching in rice.** *New Phytologist, 176*(2), 288–298. https://doi.org/10.1111/j.1469-8137.2007.02177.x

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., & Bryant, S. H. (2015). **CDD: NCBI's conserved domain database.** *Nucleic Acids Research, 43*(D1), D222–D226. https://doi.org/10.1093/nar/gku1221

Maruyama, K., Sakuma, Y., Kasuga, M., Ito, Y., Seki, M., Goda, H., & Yamaguchi-Shinozaki, K. (2004). **Identification of cold-inducible downstream genes of the Arabidopsis DREB1A/CBF3**

transcriptional factor using two microarray systems. *Plant Journal, 38*(6), 982–993. https://doi.org/10.1111/j.1365-313X.2004.02100.x

Mattsson, J., Sung, Z. R., & Berleth, T. (1999). **Responses of plant vascular systems to auxin transport inhibition.** *Development, 126*(13), 2979–2991.

Matus, J. T., Aquea, F., & Arce-Johnson, P. (2008). **Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes.** *BMC Plant Biology, 8*(1), 83. https://doi.org/10.1186/1471-2229-8-83

McCarthy, R. L., Zhong, R., & Ye, Z. H. (2009). **MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in *Arabidopsis*.** *Plant and Cell Physiology, 50*(11), 1950–1964. https://doi.org/10.1093/pcp/pcp139

McDill, J., Repplinger, M., Simpson, B. B., & Kadereit, J. W. (2009). **The Phylogeny of *Linum* and Linaceae Subfamily Linoideae, with Implications for Their Systematics, Biogeography, and Evolution of Heterostyly.** *Systematic Botany, 34*(2), 386–405. https://doi.org/10.1600/036364409788606244

McHughen, A. (1989). **Agrobacterium mediated transfer of chlorsulfuron resistance to commercial flax cultivars.** *Plant Cell Reports, 8*(8), 445–9. https://doi.org/10.1007/BF00269045

Mengiste, T., Chen, X., Salmeron, J., & Dietrich, R. (2003). **The BOTRYTIS SUSCEPTIBLE1 gene encodes an R2R3MYB transcription factor protein that is required for biotic and abiotic stress responses in *Arabidopsis*.** *The Plant Cell, 15*(11), 2551–65. https://doi.org/10.1105/tpc.014167

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., & Thomas, P. D. (2017). **PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.** *Nucleic Acids Research, 45*(D1), D183–D189. https://doi.org/10.1093/nar/gkw1138

Mikshina, P. V., Chemikosova, S. B., Mokshina, N. E., Ibragimova, N. N., & Gorshkova, T. A. (2009). **Free galactose and galactosidase activity in the course of flax fiber development.** *Russian Journal of Plant Physiology, 56*(1), 58–67. https://doi.org/10.1134/S1021443709010099

Minh, B. Q., Nguyen, M. A. T., & Von Haeseler, A. (2013). **Ultrafast approximation for phylogenetic bootstrap.** *Molecular Biology and Evolution, 30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K., & Ohme-Takagi, M. (2007). **NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*.** *The Plant Cell, 19*(1), 270–280. https://doi.org/10.1105/tpc.106.047043

Mitsuda, N., & Ohme-Takagi, M. (2008). **NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity.** *Plant Journal, 56*(5), 768–778. https://doi.org/10.1111/j.1365-313X.2008.03633.x

Mitsuda, N., Seki, M., Shinozaki, K., & Ohme-takagi, M. (2005). **The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence.** *The Plant Cell, 17*(November), 2993–3006. https://doi.org/10.1105/tpc.105.036004.1

Mohanty, A. K., Misra, M., & Hinrichsen, G. (2000). **Biofibers, biodegradable polymers and biocomposites: An overview.** *Macromolecular Materials and Engineering*, 276–277. https://doi.org/10.1002/(SICI)1439-2054(20000301)276:1<1::AID-MAME1>3.0.CO;2-W

Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J. D., Fernandez, J. P., Mis, E. K., Khokha, M. K., & Giraldez, A. J. (2015). **CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo.** *Nature Methods, 12*(10), 982–988. https://doi.org/10.1038/nmeth.3543

Motose, H., Sugiyama, M., & Fukuda, H. (2004). **A proteoglycan mediates inductive interaction during plant vascular development.** *Nature, 429*(6994), 873–878. https://doi.org/10.1038/nature02613

Muench, D. G., & Mullen, R. T. (2003). **Peroxisome dynamics in plant cells: A role for the cytoskeleton.** *Plant Science.* https://doi.org/10.1016/S0168-9452(02)00426-0

Nagadoi, A., Morikawa, S., Nakamura, H., Enari, M., Kobayashi, K., Yamamoto, H., & Nishimura, Y. (1995). **Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain.** *Structure, 3*(11), 1217–1224. https://doi.org/10.1016/S0969-2126(01)00257-X

Nakai, K., & Horton, P. (1999). **PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization.** *Trends in Biochemical Sciences.* https://doi.org/10.1016/S0968-0004(98)01336-X

Nakano, Y., Yamaguchi, M., Endo, H., Rejab, N. A., & Ohtani, M. (2015). **NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants.** *Frontiers in Plant Science, 6*(May), 288. https://doi.org/10.3389/fpls.2015.00288

Nakashima, K., Tran, L. S. P., Van Nguyen, D., Fujita, M., Maruyama, K., Todaka, D., & Yamaguchi-Shinozaki, K. (2007). **Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice**. *Plant Journal, 51*(4), 617–630. https://doi.org/10.1111/j.1365-313X.2007.03168.x

Narusaka, M., Shiraishi, T., Iwabuchi, M., & Narusaka, Y. (2010). **The floral inoculating protocol: A simplified *Arabidopsis thaliana* transformation method modified from floral dipping.** *Plant Biotechnology, 27*(4), 349–351. https://doi.org/10.5511/plantbiotechnology.27.349

Nelson, B. K., Cai, X., & Nebenführ, A. (2007). **A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants.** *Plant Journal, 51*(6), 1126–1136. https://doi.org/10.1111/j.1365-313X.2007.03212.x

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). **IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies.** *Molecular Biology and Evolution, 32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Ning, Y. Q., Ma, Z. Y., Huang, H. W., Mo, H., Zhao, T. T., Li, L., & He, X. J. (2015). **Two novel NAC transcription factors regulate gene expression and flowering time by associating with the histone demethylase JMJ14.** *Nucleic Acids Research, 43*(3), 1469–1484. https://doi.org/10.1093/nar/gku1382

Nôžková, J., Remeselníková, K., & Bjelková, M. (2014). **Characterization and evaluation of flax seeds (*Linum usitatissimum* L.) on selected genotypes.** *Journal of Central European Agriculture, 15*(1), 193–207. https://doi.org/10.5513/JCEA01/15.1.1434

Nurmberg, P. L., Knox, K. A., Yun, B. W., Morris, P. C., Shafiei, R., Hudson, A., & Loake, G. J. (2007). **The developmental selector AS1 is an evolutionarily conserved regulator of the plant**

immune response. *Proceedings of the National Academy of Sciences of the United States of America, 104*(47), 18795–18800. https://doi.org/10.1073/pnas.0705586104

Obayashi, T., Hayashi, S., Saeki, M., Ohta, H., & Kinoshita, K. (2009). **ATTED-II provides coexpressed gene networks for *Arabidopsis*.** *Nucleic Acids Research, 37*(SUPPL. 1). https://doi.org/10.1093/nar/gkn807

Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., & Ohta, H. (2007). **ATTED-II: A database of co-expressed genes and *cis* elements for identifying co-regulated gene groups in *Arabidopsis*.** *Nucleic Acids Research, 35*(SUPPL. 1). https://doi.org/10.1093/nar/gkl783

Ohashi-Ito, K., & Fukuda, H. (2003). **HD-zip III homeobox genes that include a novel member, ZeHB-13 (Zinnia)/ATHB-15 (*Arabidopsis*), are involved in procambium and xylem cell differentiation.** *Plant & Cell Physiology, 44*(12), 1350–1358. https://doi.org/10.1093/pcp/pcg164

Ohashi-Ito, K., & Fukuda, H. (2010). **Transcriptional regulation of vascular cell fates.** *Current Opinion in Plant Biology, 13*(6), 670–676. https://doi.org/10.1016/j.pbi.2010.08.011

Ohashi-Ito, K., & Fukuda, H. (2014). **Initiation of vascular development.** *Physiologia Plantarum, 151*(2), 142–146. https://doi.org/10.1111/ppl.12111

Ohashi-Ito, K., Oda, Y., & Fukuda, H. (2010). ***Arabidopsis* vascular-related NAC-domain6 directly regulates the genes that govern programmed cell death and secondary wall formation during xylem differentiation.** *The Plant Cell, 22*(10), 3461–3473. https://doi.org/10.1105/tpc.110.075036

Ohtani, M., Nishikubo, N., Xu, B., Yamaguchi, M., Mitsuda, N., Goué, N., & Demura, T. (2011). **A NAC domain protein family contributing to the regulation of wood formation in poplar.**

The Plant Journal : For Cell and Molecular Biology, 67(3), 499–512. https://doi.org/10.1111/j.1365-313X.2011.04614.x

Okushima, Y., Mitina, I., Quach, H. L., & Theologis, A. (2005). **AUXIN RESPONSE FACTOR 2 (ARF2): A pleiotropic developmental regulator.** *Plant Journal, 43*(1), 29–46. https://doi.org/10.1111/j.1365-313X.2005.02426.x

Olsen, A. N., Ernst, H. A., Leggio, L. L., & Skriver, K. (2005). **NAC transcription factors: Structurally distinct, functionally diverse.** *Trends in Plant Science.* https://doi.org/10.1016/j.tplants.2004.12.010

Olsen, L. J. (1998). **The surprising complexity of peroxisome biogenesis.** *Plant Molecular Biology, 38*(1–2), 163–189. https://doi.org/10.1023/A:1006092830670

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., & Kikuchi, S. (2003). **Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*.** *DNA Research, 10*, 239–247.

Park, S., Gidda, S. K., James, C. N., Horn, P. J., Khuu, N., Seay, D. C., & Dyer, J. M. (2013). **The α/β Hydrolase CGI-58 and Peroxisomal Transport Protein PXA1 Coregulate Lipid Homeostasis and Signaling in *Arabidopsis*.** *The Plant Cell, 25*(5), 1726–1739. https://doi.org/10.1105/tpc.113.111898

Pastore, J. J., Limpuangthip, A., Yamaguchi, N., Wu, M. F., Sang, Y., Han, S. K., & Wagner, D. (2011). **LATE MERISTEM IDENTITY2 acts together with LEAFY to activate APETALA1.** *Development, 138*(15), 3189–3198. https://doi.org/10.1242/dev.063073

Pei, H., Ma, N., Tian, J., Luo, J., Chen, J., Li, J., & Gao, J. (2013). **An NAC transcription factor controls ethylene-regulated cell expansion in flower petals.** *Plant Physiology, 163*(2), 775–791. https://doi.org/10.1104/pp.113.223388

Pesch, M., & Hülskamp, M. (2009). **One, two, three models for trichome patterning in *Arabidopsis*?** *Current Opinion in Plant Biology.* https://doi.org/10.1016/j.pbi.2009.07.015

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nature Methods, 8*(10), 785–6. https://doi.org/10.1038/nmeth.1701

Poirier, Y., Ventre, G., & Caldelari, D. (1999). **Increased flow of fatty acids toward beta-oxidation in developing seeds of *Arabidopsis* deficient in diacylglycerol acyltransferase activity or synthesizing medium-chain-length fatty acids.** *Plant Physiology, 121*(82), 1359–1366. https://doi.org/10.1104/pp.121.4.1359

Przemeck, G. K., Mattsson, J., Hardtke, C. S., Sung, Z. R., & Berleth, T. (1996). **Studies on the role of the *Arabidopsis* gene MONOPTEROS in vascular development and plant cell axialization.** *Planta, 200*(2), 229–237. https://doi.org/10.1007/BF00208313

Puranik, S., Sahu, P. P., Srivastava, P. S., & Prasad, M. (2012). **NAC proteins: Regulation and role in stress tolerance.** *Trends in Plant Science.* https://doi.org/10.1016/j.tplants.2012.02.004

Rabetafika, H. N., Van Remoortel, V., Danthine, S., Paquot, M., & Blecker, C. (2011). **Flaxseed proteins: Food uses and health benefits.** *International Journal of Food Science and Technology.* https://doi.org/10.1111/j.1365-2621.2010.02477.x

Ramamoorthy, S. K., Skrifvars, M., & Persson, A. (2015). **A review of natural fibers used in biocomposites: plant, animal and regenerated cellulose fibers.** *Polymer Reviews, 55*(1), 107–162. https://doi.org/10.1080/15583724.2014.971124

Raman, S., Greb, T., Peaucelle, A., Blein, T., Laufs, P., & Theres, K. (2008). **Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in *Arabidopsis thaliana*.** *Plant Journal, 55*(1), 65–76. https://doi.org/10.1111/j.1365-313X.2008.03483.x

Reddy, V. D., Rao, K. V, Reddy, T. P., Title, B., Information, A., Information, A., & History, P. (2009). **Compendium of transgenic crop plants.** *In Compendium of Transgenic Crop Plants*, 191–198. https://doi.org/10.1002/9781405181099.k0108

Reumann, S. (2004). **Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses.** *Plant Physiology*, 135(2), 783–800. https://doi.org/10.1104/pp.103.035584

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., & Yu, G. (2000). **Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes.** *Science (New York, N.Y.), 290*(5499), 2105–2110. https://doi.org/10.1126/science.290.5499.2105

Roach, M. J., & Deyholos, M. K. (2007). **Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fiber-bearing phloem tissues.** *Molecular Genetics and Genomics, 278*(2), 149–165. https://doi.org/10.1007/s00438-007-0241-1

Roach, M. J., & Deyholos, M. K. (2008). **Microarray analysis of developing flax hypocotyls identifies novel transcripts correlated with specific stages of phloem fiber differentiation.** *Annals of Botany, 102*(3), 317–330. https://doi.org/10.1093/aob/mcn110

Roach, M. J., Mokshina, N. Y., Badhan, A., Snegireva, A. V, Hobson, N., Deyholos, M. K., & Gorshkova, T. A. (2011). **Development of cellulosic secondary walls in flax fibers requires beta-galactosidase.** *Plant Physiology, 156*(3), 1351–1363. https://doi.org/10.1104/pp.111.172676

Rodriguez-Villalon, A., Gujas, B., Kang, Y. H., Breda, A. S., Cattaneo, P., Depuydt, S., & Hardtke, C. S. (2014). **Molecular genetic framework for protophloem formation.** *Proceedings of the National Academy of Sciences, 111*(31), 11551–11556. https://doi.org/10.1073/pnas.1407337111

Rombauts, S., Déhais, P., Van Montagu, M., & Rouzé, P. (1999). **PlantCARE, a plant cis-acting regulatory element database.** *Nucleic Acids Research.* https://doi.org/10.1093/nar/27.1.295

Rosinski, J. A., & Atchley, W. R. (1998). **Molecular evolution of the Myb family of transcription factors: Evidence for polyphyletic origin.** *Journal of Molecular Evolution, 46*(1), 74–83. https://doi.org/10.1007/PL00006285

Roy, A., Kucukural, A., & Zhang, Y. (2010). **I-TASSER: a unified platform for automated protein structure and function prediction.** *Nature Protocols, 5*(4), 725–738. https://doi.org/10.1038/nprot.2010.5

Roy, A., Srinivasan, N., & Gowri, V. S. (2009). **Molecular and structural basis of drift in the functions of closely-related homologous enzyme domains: Implications for function annotation based on homology searches and structural genomics.** *In Silico Biology, 9*(1–2). https://doi.org/10.3233/ISB-2009-0379

Rubilar, M., Gutiérrez, C., Verdugo, M., Shene, C., & Sineiro, J. (2010). **Flaxseed as a source of functional ingredients.** *Journal of Soil Science and Plant Nutrition, 10*(3), 373–377. https://doi.org/10.4067/S0718-95162010000100010

Rubio, V., Linhares, F., Solano, R., Mart'in, A. C., Iglesias, J., Leyva, A., & Paz-Ares, J. (2001). **A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae.** *Genes & Development, 15*(16), 2122–2133. https://doi.org/10.1101/gad.204401.availability

Russell, L., Larner, V., Kurup, S., Bougourd, S., & Holdsworth, M. (2000). **The *Arabidopsis* COMATOSE locus regulates germination potential.** *Development, 127*(17), 3759–67.

Ruthardt, N., Petra, J., Morris, D. A., Emans, N., Ju, G., Paciorek, T., & Friml, J. (2005). **Auxin inhibits endocytosis and promotes its own efflux from cells.** *Nature, 435*(7046), 1251–6. https://doi.org/10.1038/nature03633

Ruzicka, K., Simaskova, M., Duclercq, J., Petrasek, J., Zazimalova, E., Simon, S., & Benkova, E. (2009). **Cytokinin regulates root meristem activity via modulation of the polar auxin transport.** *Proceedings of the National Academy of Sciences, 106*(11), 4284–4289. https://doi.org/10.1073/pnas.0900060106

Sablowski, R. W. M., & Meyerowitz, E. M. (1998). **A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA.** *Cell, 92*(1), 93–103. https://doi.org/10.1016/S0092-8674(00)80902-2

Sachs, T. (1969). **Polarity and the induction of organized vascular tissues.** *Annals of Botany, 33*(2), 263–275. Retrieved from http://aob.oxfordjournals.org/cgi/content/abstract/33/2/263

Sachs, T. (1981). **The control of the patterned differentiation of vascular tissues.** *Advances in Botanical Research, 9*(C), 151–262. https://doi.org/10.1016/S0065-2296(08)60351-1

Saito, T., Fujikawa, H., Haga, N., Suzuki, T., Machida, Y., & Ito, M. (2015). **Genetic interaction between G2/M phase-specific transcription factor MYB3R4 and MAPKKK ANP3 for**

execution of cytokinesis in *Arabidopsis thaliana*. *Plant Signaling and Behavior, 10*(3). https://doi.org/10.4161/15592324.2014.990817

Sakamoto, S., & Mitsuda, N. (2015). **Reconstitution of a secondary cell wall in a secondary cell wall-deficient Arabidopsis mutant.** *Plant and Cell Physiology, 56*(2), 299–310. https://doi.org/10.1093/pcp/pcu208

Salih, H., Gong, W., He, S., Sun, G., Sun, J., & Du, X. (2016). **Genome-wide characterization and expression analysis of MYB transcription factors in *Gossypium hirsutum*.** *BMC Genetics, 17*(1), 129. https://doi.org/10.1186/s12863-016-0436-8

Scarpella, E., Friml, J., Marcos, D., & Berleth, T. (2006). **Control of leaf vascular patterning by polar auxin transport.** *Genes & Development, 20*, 1015–1027. https://doi.org/10.1101/gad.1402406

Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E. H., & Weijers, D. (2010). **MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor.** *Nature, 464*(7290), 913–916. https://doi.org/10.1038/nature08836

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., & Lohmann, J. U. (2005). **A gene expression map of *Arabidopsis thaliana* development.** *Nature Genetics, 37*(5), 501–506. https://doi.org/10.1038/ng1543

Schrader, J., Nilsson, J., Mellerowicz, E., Berglund, A., Nilsson, P., Hertzberg, M., & Sandberg, G. (2004). **A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity.** *Plant Cell, 16*(9), 2278–2292. https://doi.org/10.1105/tpc.104.024190

Seo, P. J., & Park, C. M. (2010). **A membrane-bound NAC transcription factor as an integrator of biotic and abiotic stress signals.** *Plant Signaling & Behavior, 5*(5), 481–483. https://doi.org/10.4161/psb.11083

Sessa, G., Morelli, G., & Ruberti, I. (1993). **The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities.** *The EMBO Journal, 12*(9), 3507–17.

Shahnejat-Bushehri, S., Tarkowska, D., Sakuraba, Y., & Balazadeh, S. (2016). **Arabidopsis NAC transcription factor JUB1 regulates GA/BR metabolism and signalling.** *Nature Plants, 2*(3), 16013. https://doi.org/10.1038/nplants.2016.13

Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., & Tsien, R. Y. (2004). **Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.** *Nature Biotechnology, 22*(12), 1567–1572. https://doi.org/10.1038/nbt1037

Khan, S. (2015). **Functional characterization of DUF642 genes in *Arabidopsis thaliana*.** University of Alberta.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., & Ideker, T. (2003). **Cytoscape: A software Environment for integrated models of biomolecular interaction networks.** *Genome Research, 13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shao, H., Wang, H., & Tang, X. (2015). **NAC transcription factors in plant multiple abiotic stress responses: progress and prospects.** *Frontiers in Plant Science, 6*. https://doi.org/10.3389/fpls.2015.00902

Shi, Y., Huang, J., Sun, T., Wang, X., Zhu, C., Ai, Y., & Gu, H. (2017). **The precise regulation of different COR genes by individual CBF transcription factors in *Arabidopsis thaliana***. *Journal of Integrative Plant Biology, 59*(2), 118–133. https://doi.org/10.1111/jipb.12515

Sieburth, L. E. (1999). **Auxin is required for leaf vein pattern in *Arabidopsis*.** *Plant Physiology, 121*(4), 1179–1190. https://doi.org/10.1104/pp.121.4.1179

Sigrist, C. J. A., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A., & Hulo, N. (2009). **PROSITE, a protein domain database for functional characterization and annotation.** *Nucleic Acids Research, 38*(SUPPL.1). https://doi.org/10.1093/nar/gkp885

Simon, M., Lee, M. M., Lin, Y., Gish, L., & Schiefelbein, J. (2007). **Distinct and overlapping roles of single-repeat MYB genes in root epidermal patterning.** *Developmental Biology, 311*(2), 566–578. https://doi.org/10.1016/j.ydbio.2007.09.001

Simpson, S. D., Nakashima, K., Narusaka, Y., Seki, M., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2003). **Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence.** *Plant Journal, 33*(2), 259–270. https://doi.org/10.1046/j.1365-313X.2003.01624.x

Singh, K. K., Mridula, D., Rehal, J., & Barnwal, P. (2011). **Flaxseed: a potential source of food, feed and fiber.** *Critical Reviews in Food Science and Nutrition, 51*(March 2014), 210–222. https://doi.org/10.1080/10408390903537241

Singh, V. K., & Jain, M. (2014). **Transcriptome profiling for discovery of genes involved in shoot apical meristem and flower development.** *Genomics Data, 2*, 135–138. https://doi.org/10.1016/j.gdata.2014.06.004

Smyth, D. R., Bowman, J. L., & Meyerowitz, E. M. (1990). **Early flower development in *Arabidopsis*.** *Plant Cell, 2*(8), 755–767. https://doi.org/10.1105/tpc.2.8.755

Snegireva, A., Chernova, T., Ageeva, M., Lev-Yadun, S., & Gorshkova, T. (2015). **Intrusive growth of primary and secondary phloem fibers in hemp stem determines fiber-bundle formation and structure.** *AoB PLANTS, 7*(0), plv061-. https://doi.org/10.1093/aobpla/plv061

Snegireva, A. V., Ageeva, M. V., Amenitskii, S. I., Chernova, T. E., Ebskamp, M., & Gorshkova, T. A. (2010). **Intrusive growth of sclerenchyma fibers.** *Russian Journal of Plant Physiology, 57*(3), 342–355. https://doi.org/10.1134/S1021443710030052

Soler, M., Camargo, E. L. O., Carocha, V., Cassan-Wang, H., San Clemente, H., Savelli, B., & Grima-Pettenati, J. (2015). **The Eucalyptus grandis R2R3-MYB transcription factor family: Evidence for woody growth-related evolution and function.** *New Phytologist, 206*(4), 1364–1377. https://doi.org/10.1111/nph.13039

Souer, E., Van Houwelingen, A., Kloos, D., Mol, J., & Koes, R. (1996). **The no apical Meristem gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries.** *Cell, 85*(2), 159–170. https://doi.org/10.1016/S0092-8674(00)81093-4

Sperotto, R. A., Ricachenevsky, F. K., Duarte, G. L., Boff, T., Lopes, K. L., Sperb, E. R., & Fett, J. P. (2009). **Identification of up-regulated genes in flag leaves during rice grain filling and characterization of OsNAC5, a new ABA-dependent transcription factor.** *Planta, 230*(5), 985–1002. https://doi.org/10.1007/s00425-009-1000-9

Spinelli, S. V., Martin, A. P., Viola, I. L., Gonzalez, D. H., & Palatnik, J. F. (2011). **A Mechanistic Link between STM and CUC1 during Arabidopsis Development.** *Plant Physiology, 156*(4), 1894–1904. https://doi.org/10.1104/pp.111.177709

Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., & Birney, E. (2002). **The Bioperl toolkit: Perl modules for the life sciences.** *Genome Research, 12*(10), 1611–1618. https://doi.org/10.1101/gr.361602

Stothard, P. (2000). **The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.** *BioTechniques, 28*(6).

Stracke, R., Werber, M., & Weisshaar, B. (2001). **The R2R3-MYB gene family in *Arabidopsis thaliana*.** *Current Opinion in Plant Biology.* https://doi.org/10.1016/S1369-5266(00)00199-0

Sun, Y., Zhou, Q., Zhang, W., Fu, Y., & Huang, H. (2002). ***ASYMMETRIC LEAVES1*, an *Arabidopsis* gene that is involved in the control of cell differentiation in leaves.** *Planta, 214*(5), 694–702. https://doi.org/10.1007/s004250100673

Sveinsson, S., McDill, J., Wong, G. K. S., Li, J., Li, X., Deyholos, M. K., & Cronk, Q. C. B. (2014). **Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (*Linum*) using transcriptomics.** *Annals of Botany, 113*(5), 753–761. https://doi.org/10.1093/aob/mct306

Takada, S., Hibara, K., Ishida, T., & Tasaka, M. (2001). **The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation.** *Development (Cambridge, England), 128*(7), 1127–35.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). **MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance,**

and maximum parsimony methods. *Molecular Biology and Evolution, 28*(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

Tanz, S. K., Castleden, I., Hooper, C. M., Vacher, M., Small, I., & Millar, H. A. (2013). **SUBA3: A database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis.** *Nucleic Acids Research, 41*(D1). https://doi.org/10.1093/nar/gks1151

The Arabidopsis Genome Initiative. (2000). **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature, 408*(6814), 796–815. https://doi.org/10.1038/35048692

Thomashow, M. F. (1999). **PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms.** *Annual Review of Plant Physiology and Plant Molecular Biology, 50*(1), 571–599. https://doi.org/10.1146/annurev.arplant.50.1.571

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., & Su, Z. (2017). **AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update.** *Nucleic Acids Research, 45*(W1), W122–W129. https://doi.org/10.1093/nar/gkx382

To, L. T. (2013). **Genetics of seed coat and stem development in flax (*Linum usitatissimum* L.).** Retrieved from https://era.library.ualberta.ca/files/k643b212r#.WiwtI7Q-fOQ

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., & Rokhsar, D. (2006). **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science (New York, N.Y.), 313*(5793), 1596–604. https://doi.org/10.1126/science.1128691

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., & Dubcovsky, J. (2006). **A NAC Gene regulating senescence improves grain protein, zinc, and iron content in wheat.** *Science (New York, N.Y.), 314*(5803), 1298–301. https://doi.org/10.1126/science.1133649

Ulmasov, T., Hagen, G., & Guilfoyle, T. J. (1999). **Dimerization and DNA binding of auxin response factors.** *The Plant Journal : For Cell and Molecular Biology, 19*(3), 309–319. https://doi.org/10.1046/j.1365-313X.1999.00538.x

Uysal, H., Fu, Y. B., Kurt, O., Peterson, G. W., Diederichsen, A., & Kusters, P. (2010). **Genetic diversity of cultivated flax (*Linum usitatissimum* L.) and its wild progenitor pale flax (*Linum bienne* Mill.) as revealed by ISSR markers.** *Genetic Resources and Crop Evolution, 57*(7), 1109–1119. https://doi.org/10.1007/s10722-010-9551-y

Vanzeist, W., & Bakkerheeres, J. A. H. (1975). **Evidence for linseed cultivation before 6000 BC.** *Journal of Archaeological Science, 2*(3), 215–219. Retrieved from http://www.rug.nl/research/portal/publications/evidence-for-linseed-cultivation-before-6000-bc(ff0b4514-93dc-42f2-a7c7-9be42c1ad436).html

Vavilov, N. (1951). **The origin, variation, immunity and breeding of cultivated plants.** *Chronica Botanica, 13*(2990), 1–366. https://doi.org/10.1126/science.115.2990.433-a

Venglat, P., Xiang, D., Qiu, S., Stone, S. L., Tibiche, C., Cram, D., & Datla, R. (2011). **Gene expression analysis of flax seed development.** *BMC Plant Biology, 11*(1), 74. https://doi.org/10.1186/1471-2229-11-74

Vroemen, C. W. (2003). **The *CUP-SHAPED COTYLEDON3* Gene Is Required for Boundary and Shoot Meristem Formation in Arabidopsis.** *The Plant Cell Oneline, 15*(7), 1563–1577. https://doi.org/10.1105/tpc.012203

Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., & Provart, N. J. (2017). **ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology.** *The Plant Cell*, tpc.00073.2017. https://doi.org/10.1105/tpc.17.00073

Wang, L., Cao, C., Ma, Q., Zeng, Q., Wang, H., Cheng, Z., & Wang, Y. (2014). **RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications.** *BMC Plant Biology, 14*, 169. https://doi.org/10.1186/1471-2229-14-169

Wang, M., Qi, X., Zhao, S., Zhang, S., & Lu, M. Z. (2009). **Dynamic changes in transcripts during regeneration of the secondary vascular system in *Populus tomentosa* Carr. revealed by cDNA microarrays.** *BMC Genomics, 10*, 215. https://doi.org/10.1186/1471-2164-10-215

Wang, S., Li, E., Porth, I., Chen, J. G., Mansfield, S. D., & Douglas, C. J. (2014). **Regulation of secondary cell wall biosynthesis by poplar R2R3 MYB transcription factor PtrMYB152 in Arabidopsis.** *Scientific Reports, 4*, 5054. https://doi.org/10.1038/srep05054

Wang, X., Basnayake, B. M., Zhang, H., Li, G., Li, W., Virk, N., & Song, F. (2009). **The *Arabidopsis* ATAF1, a NAC transcription factor, is a negative regulator of defense responses against necrotrophic fungal and bacterial pathogens.** *Molecular Plant-Microbe Interactions : MPMI, 22*(10), 1227–1238. https://doi.org/10.1094/MPMI-22-10-1227

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., & Deyholos, M. K. (2012). **The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads.** *The Plant Journal, 72*(3):461-73.https://doi.org/10.1111/j.1365-313X.2012.05093.x

Wang, Z., Tang, J., Hu, R., Wu, P., Hou, X. L., Song, X. M., & Xiong, A. S. (2015). **Genome-wide analysis of the R2R3-MYB transcription factor genes in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) reveals their stress and hormone responsive patterns.** *BMC Genomics, 16*, 17. https://doi.org/10.1186/s12864-015-1216-y

Wang, Z. Y. (1997). **A Myb-related transcription factor is involved in the phytochrome regulation of an Arabidopsis Lhcb gene.** *The Plant Cell Online, 9*(4), 491–507. https://doi.org/10.1105/tpc.9.4.491

Waters, M. T., Wang, P., Korkaric, M., Capper, R. G., Saunders, N. J., & Langdale, J. A. (2009). **GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*.** *The Plant Cell Online, 21*(4), 1109–1128. https://doi.org/10.1105/tpc.108.065250

Wenzel, C. L., Schuetz, M., Yu, Q., & Mattsson, J. (2007). **Dynamics of MONOPTEROS and PIN-FORMED1 expression during leaf vein pattern formation in *Arabidopsis thaliana*.** *Plant Journal, 49*(3), 387–398. https://doi.org/TPJ2977 [pii]\r10.1111/j.1365-313X.2006.02977.x

Weston, K. (1998). **Myb proteins in life, death and differentiation.** *Current Opinion in Genetics and Development.* https://doi.org/10.1016/S0959-437X(98)80065-8

Wilkins, O., Nahal, H., Foong, J., Provart, N. J., & Campbell, M. M. (2009). **Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors.** *Plant Physiology, 149*(2), 981–93. https://doi.org/10.1104/pp.108.132795

Willemsen, V., Bauch, M., Bennett, T., Campilho, A., Wolkenfelt, H., Xu, J., & Scheres, B. (2008). **The NAC Domain transcription factors FEZ and SOMBRERO control the orientation of cell division plane in *Arabidopsis* root stem cells.** *Developmental Cell, 15*(6), 913–922. https://doi.org/10.1016/j.devcel.2008.09.019

Wong, C. E., Bhalla, P. L., Ottenhof, H., & Singh, M. B. (2008). **Transcriptional profiling of the pea shoot apical meristem reveals processes underlying its function and maintenance.** *BMC Plant Biology, 8*, 73. https://doi.org/10.1186/1471-2229-8-73

Wong, D. C., Schlechter, R., Vannozzi, A., Holl, J., Hmmam, I., Bogs, J., & Matus, J. T. (2016). **A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation.** *DNA Research, 23*(5), 451–466. https://doi.org/10.1093/dnares/dsw028

Wu, C. Y., Washida, H., Onodera, Y., Harada, K., & Takaiwa, F. (2000). **Quantitative nature of the prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: Minimal cis-element requirements for endosperm-specific gene expression.** *Plant Journal, 23*(3), 415–421. https://doi.org/10.1046/j.1365-313X.2000.00797.x

Xie, K., Zhang, J., & Yang, Y. (2014). **Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops.** *Molecular Plant.* https://doi.org/10.1093/mp/ssu009

Xie, Q., Frugis, G., Colgan, D., & Chua, N. H. (2000). ***Arabidopsis*** **NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development.** *Genes and Development, 14*(23), 3024–3036. https://doi.org/10.1101/gad.852200

Xie, Y., Liu, Y., Meng, M., Chen, L., & Zhu, Z. (2003). **Isolation and identification of a super strong plant promoter from cotton leaf curl Multan virus.** *Plant Molecular Biology, 53*(1–2), 1–14. https://doi.org/10.1023/B:PLAN.0000009257.37471.02

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H.Y., Han, C. Y., Liu, B., & Chen, Q. J. (2014). **A CRISPR/Cas9 toolkit for multiplex genome editing in plants.** *BMC Plant Biology, 14*(1), 327. https://doi.org/10.1186/s12870-014-0327-y

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., & Demura, T. (2014). **Contribution of NAC transcription factors to plant adaptation to land.** *Science (New York, N.Y.), 343(*6178), 1505–8. https://doi.org/10.1126/science.1248417

Yadav, R. K., Girke, T., Pasala, S., Xie, M., & Reddy, G. V. (2009). **Gene expression map of the *Arabidopsis* shoot apical meristem stem cell niche.** *Proceedings of the National Academy of Sciences of the United States of America, 106*(12), 4941–4946. https://doi.org/10.1073/pnas.0900843106

Yamaguchi-Shinozaki, K. (1994). **A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress.** *The Plant Cell Online, 6*(2), 251–264. https://doi.org/10.1105/tpc.6.2.251

Yamaguchi-Shinozaki, K., & Shinozaki, K. (1993). **Characterization of the expression of a desiccation-responsive rd29 gene of *Arabidopsis thaliana* and analysis of its promoter in transgenic plants.** *Molecular & General Genetics, 236*(2–3), 331–340. https://doi.org/10.1007/BF00277130

Yamaguchi, M., Goué, N., Igarashi, H., Ohtani, M., Nakano, Y., Mortimer, J. C., & Demura, T. (2010). **VASCULAR-RELATED NAC-DOMAIN6 and VASCULAR-RELATED NAC-DOMAIN7 effectively induce transdifferentiation into xylem vessel elements under control of an induction system.** *Plant Physiology, 153*(3), 906–14. https://doi.org/10.1104/pp.110.154013

Yamaguchi, M., Kubo, M., Fukuda, H., & Demura, T. (2008). **Vascular-related NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in *Arabidopsis* roots and shoots.** *Plant Journal, 55*(4), 652–664. https://doi.org/10.1111/j.1365-313X.2008.03533.x

Yamaguchi, M., Ohtani, M., Mitsuda, N., Kubo, M., Ohme-Takagi, M., Fukuda, H., & Demura, T. (2010). **VND-INTERACTING2, a NAC domain transcription factor, negatively regulates xylem vessel formation in *Arabidopsis*.** *The Plant Cell, 22*(4), 1249–1263. https://doi.org/10.1105/tpc.108.064048

Yang, L., Teixeira, P. J., Biswas, S., Finkel, O. M., He, Y., Salas-Gonzalez, I., & Dangl, J. L. (2017). **Pseudomonas syringae type III effector HopBB1 promotes host transcriptional repressor degradation to regulate phytohormone responses and virulence.** *Cell Host and Microbe, 21*(2), 156–168. https://doi.org/10.1016/j.chom.2017.01.003

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., & Li-Jia, Q. (2006). **The MYB transcription factor superfamily of *Arabidopsis*: Expression analysis and phylogenetic comparison with the rice MYB family.** *Plant Molecular Biology, 60*(1), 107–124. https://doi.org/10.1007/s11103-005-2910-y

Yao, D. X., Wei, Q., Xu, W. Y., Syrenne, R. D., Yuan, J. S., & Su, Z. (2012). **Comparative genomic analysis of NAC transcriptional factors to dissect the regulatory mechanisms for cell wall biosynthesis.** *BMC Bioinformatics, 13*(Suppl 15), 12. https://doi.org/10.1186/1471-2105-13-s15-s10

Yilmaz, A., Mejia-Guerra, M. K., Kurz, K., Liang, X., Welch, L., & Grotewold, E. (2011). **AGRIS: The *Arabidopsis* gene regulatory information server, an update.** *Nucleic Acids Research, 39*(SUPPL. 1). https://doi.org/10.1093/nar/gkq1120

Yoo, C. Y., Pence, H. E., Jin, J. B., Miura, K., Gosney, M. J., Hasegawa, P. M., & Mickelbart, M. V. (2010). **The *Arabidopsis* GTL1 transcription factor regulates water use efficiency and**

drought tolerance by modulating stomatal density via transrepression of SDD1. *The Plant Cell, 22*(December), 4128–4141. https://doi.org/10.1105/tpc.110.078691

Yoshiyama, K., Conklin, P. A., Huefner, N. D., & Britt, A. B. (2009). **Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage.** *Proceedings of the National Academy of Sciences of the United States of America, 106*(31), 12843–8. https://doi.org/10.1073/pnas.0810304106

Yoshiyama, K. O., Kimura, S., Maki, H., Britt, A. B., & Umeda, M. (2014). **The role of SOG1, a plant-specific transcriptional regulator, in the DNA damage response.** *Plant Signaling & Behavior, 9*, e28889. https://doi.org/10.4161/psb.28889

Yu, Y., Huang, W., Chen, H., Wu, G., Yuan, H., Song, X., & Guan, F. (2014). **Identification of differentially expressed genes in flax (*Linum usitatissimum* L.) under saline-alkaline stress by digital gene expression.** *Gene, 549*(1), 113–122. https://doi.org/10.1016/j.gene.2014.07.053

Zhai, R., Wang, Z., Zhang, S., Meng, G., Song, L., Wang, Z., & Xu, L. (2016). **Two MYB transcription factors regulate flavonoid biosynthesis in pear fruit (*Pyrus bretschneideri* Rehd.).** *Journal of Experimental Botany, 67*(5), 1275–1284. https://doi.org/10.1093/jxb/erv524

Zhang, G. B., Yi, H.Y., & Gong, J. M. (2014). **The *Arabidopsis* Ethylene/Jasmonic Acid-NRT Signaling Module Coordinates Nitrate Reallocation and the Trade-Off between Growth and Environmental Adaptation.** *The Plant Cell, 26*(10), 3984–3998. https://doi.org/10.1105/tpc.114.129296

Zhang, J. D., Ruschhaupt, M., & Biczok, R. (2015). **ddCt method for qRT – PCR data analysis.** *Bioconductor*, 1–8.

Zhang, N., & Deyholos, M. K. (2016). **RNASeq analysis of the shoot apex of flax (*Linum usitatissimum*) to identify phloem fiber specification genes.** *Frontiers in Plant Science, 7*, 950. https://doi.org/10.3389/fpls.2016.00950

Zhang, Y. (2008). **I-TASSER server for protein 3D structure prediction.** *BMC Bioinformatics, 9*(1), 40. https://doi.org/10.1186/1471-2105-9-40

Zhang, Y., Cao, G., Qu, L. J., & Gu, H. (2009). **Characterization of *Arabidopsis* MYB transcription factor gene AtMYB17 and its possible regulation by LEAFY and AGL15.** *Journal of Genetics and Genomics, 36*(2), 99–107. https://doi.org/10.1016/S1673-8527(08)60096-X

Zhao, Q., & Dixon, R. A. (2011). **Transcriptional networks for lignin biosynthesis: More complex than we thought?** *Trends in Plant Science.* https://doi.org/10.1016/j.tplants.2010.12.005

Zhong, R., Demura, T., & Ye, Z. H. (2006). **SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*.** *The Plant Cell Online, 18*(11), 3158–3170. https://doi.org/10.1105/tpc.106.047399

Zhong, R., Lee, C., McCarthy, R. L., Reeves, C. K., Jones, E. G., & Ye, Z. H. (2011). **Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors.** *Plant and Cell Physiology, 52*(10), 1856–1871. https://doi.org/10.1093/pcp/pcr123

Zhong, R., Lee, C., & Ye, Z. H. (2010). **Functional characterization of poplar wood-associated NAC domain transcription factors.** *Plant Physiology, 152*(2), 1044–1055. https://doi.org/10.1104/pp.109.148270

Zhong, R., Lee, C., Zhou, J., Mccarthy, R. L., & Ye, Z. H. (2008). **A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*.** *Plant Cell, 20*(10), 2763–2782. https://doi.org/10.1105/tpc.108.061325

Zhong, R., Richardson, E. A., & Ye, Z. H. (2007a). **The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*.** *The Plant Cell Online, 19*(9), 2776–2792. https://doi.org/10.1105/tpc.107.053678

Zhong, R., Richardson, E. A., & Ye, Z. H. (2007b). **Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*.** *Planta, 225*(6), 1603–1611. https://doi.org/10.1007/s00425-007-0498-y

Zhong, R., & Ye, Z. H. (2012). **MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes.** *Plant and Cell Physiology, 53*(2), 368–380. https://doi.org/10.1093/pcp/pcr185

Zhou, C., Chen, Y., Wu, Z., Lu, W., Han, J., Wu, P., & Wu, G. (2015). **Genome-wide analysis of the MYB gene family in physic nut (*Jatropha curcas* L.).** *Gene, 572*(1), 63–71. https://doi.org/10.1016/j.gene.2015.06.072

Zhu, J., Verslues, P. E., Zheng, X., Lee, B., Zhan, X., Manabe, Y., & Bressan, R. A. (2005). **HOS10 encodes an R2R3-type MYB transcription factor essential for cold acclimation in plants.** *Proceedings of the National Academy of Sciences of the United States of America, 102*(28), 9966–9971. https://doi.org/10.1073/pnas.0503960102

Zhu, J., Zhang, G. Q., Chang, Y. H., Li, X. C., Yang, J., Huang, X. Y., & Yang, Z. N. (2010). **AtMYB103 is a crucial regulator of several pathways affecting *Arabidopsis* anther**

development. *Science China Life Sciences, 53*(9), 1112–1122. https://doi.org/10.1007/s11427-010-4060-y

Zhu, T., Nevo, E., Sun, D., & Peng, J. (2012). **Phylogenetic analyses unravel the evolutionary history of nac proteins in plants.** *Evolution, 66*(6), 1833–1848. https://doi.org/10.1111/j.1558-5646.2011.01553.x

Zimmermann, P., Hennig, L., & Gruissem, W. (2005). **Gene-expression analysis and network discovery using Genevestigator.** *Trends in Plant Science.* https://doi.org/10.1016/j.tplants.2005.07.003

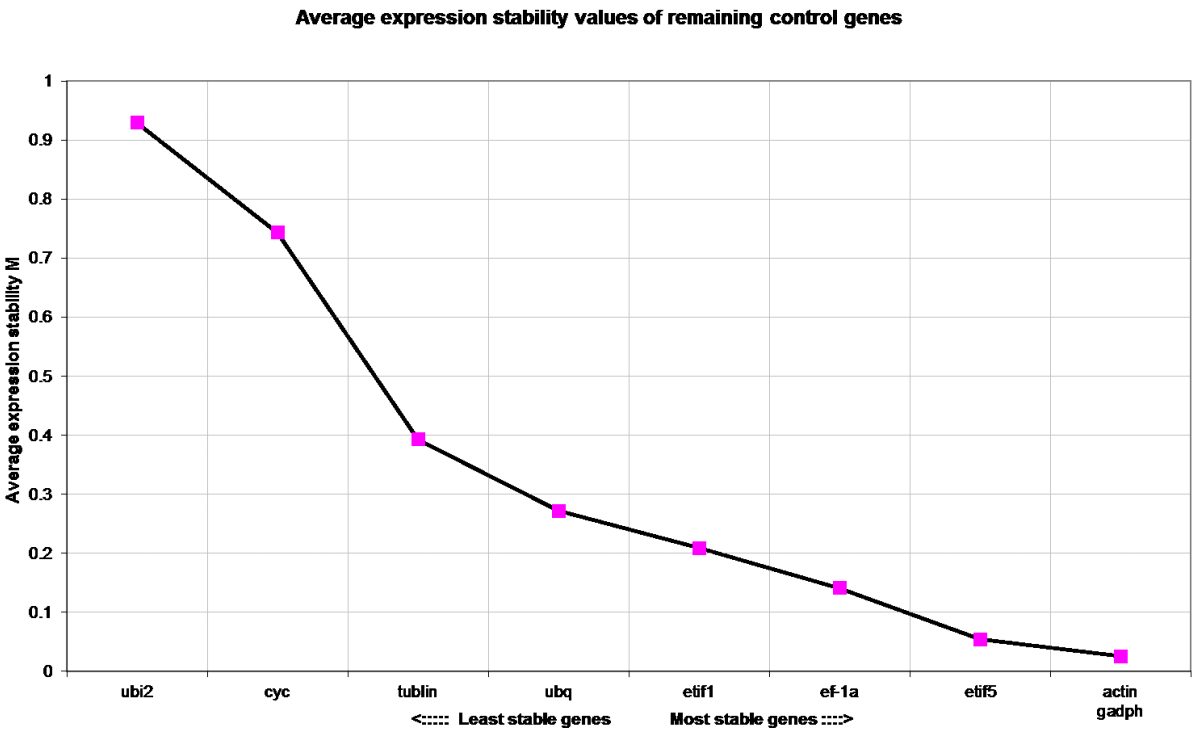
Zolman, B. K., & Bartel, B. (2004). **An *Arabidopsis* indole-3-butyric acid-response mutant defective in PEROXIN6, an apparent ATPase implicated in peroxisomal function.** *Proceedings of the National Academy of Sciences of the United States of America, 101*(6), 1786–91. https://doi.org/10.1073/pnas.0304368101

Zolman, B. K., Monroe-Augustus, M., Thompson, B., Hawes, J. W., Krukenberg, K. A., Matsuda, S. P., & Bartel, B. (2001a). **chy1, an *Arabidopsis* mutant with impaired beta-oxidation, is defective in a peroxisomal beta-hydroxyisobutyryl-CoA hydrolase.** *The Journal of Biological Chemistry, 276*(33), 31037–46. https://doi.org/10.1074/jbc.M104679200

Zolman, B. K., Silva, I. D., & Bartel, B. (2001b). **The *Arabidopsis* pxa1 mutant is defective in an ATP-binding cassette transporter-like protein required for peroxisomal fatty acid beta-oxidation.** *Plant Physiology, 127*(3), 1266–78. https://doi.org/10.1104/pp.010550.et

Zolman, B. K., Yoder, A., & Bartel, B. (2000). **Genetic analysis of indole-3-butyric acid responses in *Arabidopsis thaliana* reveals four mutant classes.** *Genetics, 156*(3), 1323–1337.

Appendix



Appendix 1. Average expression stability values (M) of nine flax common used reference genes in flax shoot apices and mature stem. GeNorm was used to calculate the M values. The reference gene with lowest average expression stability is most stable in the examined

Appendix 2. qRT-PCR primers used in this study.

Gene ID	Forward	Reverse
Lus10038135	CCCATTCCAGTAAACGCTTC	AGAAGGAGAAAGAGGGGGATT
Lus10012728	CGCAGCGTATTACCACCATA	CCGAACCTCCTTGTCCTTG
Lus10000332	GGCGAGGAGTTGCAAGAA	TCCACAGCAATGTGAGTCATC
Lus10038607	ATTTGGCTCGGCACTTACC	TAAAGCTGCAACGTCGTGAG
Lus10030473	GGCCAACCCAAACGAAAT	CCTTCTGATCGGTGGTGAA
Lus10039303	GCCGGAAATGTATGTGTTTTC	ACCACTGCACTGACTGTTGC
Lus10015902	TGGCCTCCTCCAGCTAGATA	GAATCCCGGAATCCCAGTAG
Lus10023877	ATGGCGAAACCAACATGAGT	TGGAATCTTCCCAGATGGAT
Lus10011558	GCGAACTCGACACAAAACCT	AAGAGGACCACCACCCATC
Lus10004688	AATCCAAGCGTCGGGAAT	TGGCCATAAAACTGGTTGCT
Lus10040256	CGAATCAGAGCAAAAAGCTGA	TCGTCCGTTTATTTGCGATAC
Lus10033441	ACTACTGAACATCAGTCTCACCAGA	TCCAGAAGGAGGAGTAGGATGT
Lus10010694	AACTTCCACCGCAAACAAAC	GGGATTGTGGTGGTGATTATG
LusMYB34	ATTCCGCAACATCAGGGTC	GGGTAGCCATCATAGTAGTGAGTGT
LusMYB149	GGGAGCAGCTGCAACAGTA	CCCAATCCAGCCATTGTT
LusMYB141	GCAAACTTGTTCCATAACCAGA	TTGATTATTCCTCTCCCACCA
LusMYB35	TTCCGCAACATTAAGGTCAAGT	AGTAGCCGTCATAGTAGCGAGTG
LusMYB142	CAGCAAGCTTGTTTCACCAG	TGATTATTCCTCTCCCACCATT
LusMYB187	TGTTCTCTGACGCTCAAACC	GCGAGTTTTCCATGCAACTT
LusMYB181	GATGGCGTAATTGGGAATCTT	GAGATTTCCATCCCGAAGGT
LusMYB102	GGCTGCGTTGGTGTAATCA	GTCCTCAGAGGCGGAGAAA
LusNAC136	CAAGGCTGTTGTGTCGAAGA	GATTTTGGAGGCGGTATTCA
LusNAC28	ACTGCGTTTCTCGACGATTC	CGGCAGAGAGTTAGGGCTTT
LusNAC125	ACAGCAGGGCAGTAGCTTGT	GAAGCTCGTTGAGGAAGCT
LusNAC10	AATGACGGATTGGAGAGTGC	GTTCGATGCGGTTCTGATCT
LusNAC160	GTGACGGATTGGAGAGTGCT	TCCTCCTCCTCGTCCTGAT
LusNAC46	AGCGATCAAGAGCAAGTGGT	AAACGAGGACGAAGGAGAC
LusNAC20	CAACAATGTCTCCCCTTCGT	CGATCTCGCAGGTTGATGTA
LusNAC42	GCAAGATTGGAACGGATGAT	TGTTGCTCGGTTTGTACGAG
LusNAC61	GTGGATTTGACGGGTCCAT	CGGCGGCTACTGATTCTG
LusNAC151	GATGGTCGTTGCGACTTTTT	TGTGACTCACCCGGTTTGTA
LusNAC36	TTTTCTACAAAGGCCGTGCT	TTCTGTCCAGTGTCGTCGAG
LusNAC161	AGGGTGGGTGGTGTGTAGAG	TTGATGATGAGCTCGTGAG
LusNAC146	GCAGGGGATCATGTGAATCT	GAGGTCGATCTTGTCGGAA
LusNAC66	AAGAAATACCCGACCGGAAC	TCAACCCAATCCTTCTCCTG
LusNAC164	TGATTGGATCATGCACGAGTA	TCCGGGGTTCGAGTTAATAG

LusNAC122	CCGCAGAACGAGTGGTATTT	TCCTCATCCCGATTTTCTTG
LusNAC89	GGTTCAAACAACCACACCAA	GCTTCCTAAGGCATGGTGAT
LusGADPH	AGGTTCTTCCCGCTCTCAAT	CCTCCTTGATAGCAGCCTTG

Appendix 3. Predicted transcription factors enriched in the AR. 'inf' is the abbreviation of infinity.

TF ID	Family	Lus_id	FPKM (AR)	FPKM (BR)	log2(fold_change AR/BR)	q_value
Lus10002657	AP2	Lus10002657	17.8611	0.052907	-8.39915	0.034259
Lus10004990	AP2	Lus10004990	25.7466	14.4466	-0.833656	0.005019
Lus10007719	AP2	Lus10007719	169.654	25.3011	-2.74532	0.000269
Lus10015055	AP2	Lus10015055	10.3313	2.6839	-1.94462	0.000269
Lus10018124	AP2	Lus10018124	4.07314	1.97636	-1.0433	0.006044
Lus10018655	AP2	Lus10018655	244.958	25.9819	-3.23695	0.000269
Lus10019331	AP2	Lus10019331	50.6888	27.3715	-0.88899	0.001704
Lus10019905	AP2	Lus10019905	9.54602	2.98857	-1.67545	0.000269
Lus10023165	AP2	Lus10023165	8.59631	1.24517	-2.78738	0.000269
Lus10026477	AP2	Lus10026477	6.98672	0.85289	-3.03418	0.000269
Lus10036141	AP2	Lus10036141	7.56907	3.57026	-1.08408	0.000725
Lus10041595	AP2	Lus10041595	12.9967	5.32051	-1.2885	0.000269
Lus10011730	AP2	Lus10011730	88.9252	8.75037	-3.34518	0.000269
Lus10039650	AP2	Lus10039650	9.15107	1.36756	-2.74234	0.000269
Lus10000965	AP2	Lus10000965	21.7225	5.386	-2.01191	0.000269
Lus10040140	AP2	Lus10040140	19.0402	4.63484	-2.03846	0.000269
Lus10005264	ARF	Lus10005264	55.7724	20.0599	-1.47523	0.000269
Lus10007440	ARF	Lus10007440	32.6074	18.5173	-0.816328	0.003266
Lus10010969	ARF	Lus10010969	25.6537	14.9091	-0.782973	0.004241
Lus10012421	ARF	Lus10012421	25.0888	5.27059	-2.25101	0.000269
Lus10013942	ARF	Lus10013942	65.7997	18.0831	-1.86344	0.000269
Lus10024320	ARF	Lus10024320	17.9592	5.26439	-1.77039	0.000505
Lus10031354	ARF	Lus10031354	25.0581	14.2122	-0.818146	0.00224
Lus10005340	ARR-B	Lus10005340	50.2697	9.1962	-2.45058	0.000269
Lus10037719	ARR-B	Lus10037719	21.5104	13.751	-0.645503	0.018807
Lus10041020	ARR-B	Lus10041020	28.9721	8.52058	-1.76564	0.000269
Lus10000368	B3	Lus10000368	11.6847	3.50485	-1.73719	0.002925
Lus10006483	B3	Lus10006483	11.1275	3.08218	-1.85211	0.000269
Lus10007522	B3	Lus10007522	14.6798	1.16766	-3.65214	0.000269
Lus10009688	B3	Lus10009688	40.7717	5.20336	-2.97005	0.000269
Lus10009764	B3	Lus10009764	40.8384	7.3497	-2.47417	0.000269
Lus10011245	B3	Lus10011245	6.1828	2.40134	-1.36442	0.000505
Lus10014044	B3	Lus10014044	3.54437	0.937032	-1.91936	0.01244
Lus10015266	B3	Lus10015266	15.3101	2.7441	-2.48008	0.000269
Lus10017434	B3	Lus10017434	26.5536	1.26091	-4.39637	0.000269
Lus10018440	B3	Lus10018440	4.76038	2.9279	-0.701209	0.037159

Lus10019870	B3	Lus10019870	37.451	2.37531	-3.97882	0.000269
Lus10019873	B3	Lus10019873	36.6353	10.4168	-1.81433	0.000269
Lus10021006	B3	Lus10021006	56.5031	29.0648	-0.959056	0.029298
Lus10023691	B3	Lus10023691	1.19463	0.34088	-1.80922	0.048109
Lus10023844	B3	Lus10023844	13.4622	8.10728	-0.73163	0.048402
Lus10025533	B3	Lus10025533	2.06475	0.547635	-1.91468	0.002588
Lus10026067	B3	Lus10026067	10.4939	3.06507	-1.77555	0.001518
Lus10026921	B3	Lus10026921	56.1271	4.20421	-3.73879	0.000269
Lus10032098	B3	Lus10032098	18.0971	0.800097	-4.49944	0.000269
Lus10032315	B3	Lus10032315	3.76516	0.809813	-2.21705	0.000269
Lus10032748	B3	Lus10032748	0.277807	0.136322	-1.02706	1
Lus10036045	B3	Lus10036045	89.2962	21.6978	-2.04105	0.000269
Lus10039303	B3	Lus10039303	40.8492	1.64638	-4.63294	0.000269
Lus10012389	BBR-BPC	Lus10012389	80.2922	46.0393	-0.802393	0.042979
Lus10018060	BBR-BPC	Lus10018060	30.6233	17.1025	-0.840427	0.006323
Lus10024313	BBR-BPC	Lus10024313	96.5043	64.0233	-0.591996	0.03601
Lus10031078	BBR-BPC	Lus10031078	47.3483	21.6948	-1.12596	0.000269
Lus10040427	BBR-BPC	Lus10040427	82.5209	50.7372	-0.701717	0.015205
Lus10042056	BBR-BPC	Lus10042056	25.8309	11.9898	-1.10729	0.000269
Lus10014327	BES1	Lus10014327	61.7901	24.1219	-1.35703	0.220962
Lus10018842	BES1	Lus10018842	15.0228	8.12355	-0.886975	0.00224
Lus10026036	BES1	Lus10026036	65.8808	36.232	-0.862594	0.002066
Lus10000332	bHLH	Lus10000332	13.1521	1.02384	-3.68323	0.000269
Lus10001271	bHLH	Lus10001271	22.3723	1.39989	-3.99833	0.000269
Lus10002160	bHLH	Lus10002160	113.378	12.8452	-3.14184	0.000269
Lus10005999	bHLH	Lus10005999	1.34543	0.295892	-2.18493	0.042584
Lus10007101	bHLH	Lus10007101	2.25153	0.69274	-1.70052	0.047656
Lus10009475	bHLH	Lus10009475	30.6653	9.30722	-1.72018	0.000269
Lus10013284	bHLH	Lus10013284	10.3233	4.90988	-1.07215	0.005317
Lus10014726	bHLH	Lus10014726	60.4831	28.6468	-1.07816	0.000933
Lus10015902	bHLH	Lus10015902	45.8242	0.546053	-6.39093	0.000505
Lus10018761	bHLH	Lus10018761	20.9282	1.4992	-3.80319	0.000269
Lus10021846	bHLH	Lus10021846	19.25	4.47285	-2.10559	0.000269
Lus10024631	bHLH	Lus10024631	21.3018	4.79903	-2.15016	0.000269
Lus10024811	bHLH	Lus10024811	9.74829	1.77022	-2.46122	0.000269
Lus10029950	bHLH	Lus10029950	31.1151	5.93322	-2.39073	0.000269
Lus10032267	bHLH	Lus10032267	57.896	19.8608	-1.54354	0.000269
Lus10032542	bHLH	Lus10032542	26.5535	12.1863	-1.12364	0.000269

Lus10038939	bHLH	Lus10038939	16.9008	10.4708	-0.690716	0.017857
Lus10039631	bHLH	Lus10039631	50.4157	14.3166	-1.81618	0.000269
Lus10041592	bHLH	Lus10041592	7.97184	2.5743	-1.63074	0.001886
Lus10042017	bHLH	Lus10042017	10.1045	5.02536	-1.00771	0.009266
Lus10043199	bHLH	Lus10043199	15.8946	8.00948	-0.988759	0.001704
Lus10002028	bZIP	Lus10002028	12.2918	5.66529	-1.11747	0.012699
Lus10002900	bZIP	Lus10002900	27.0229	6.22844	-2.11724	0.000505
Lus10005146	bZIP	Lus10005146	18.0022	7.47255	-1.2685	0.000269
Lus10008150	bZIP	Lus10008150	22.4997	8.91692	-1.33529	0.000269
Lus10008927	bZIP	Lus10008927	40.9753	16.7693	-1.28893	0.012569
Lus10008929	bZIP	Lus10008929	17.7848	9.64198	-0.883248	0.004241
Lus10014324	bZIP	Lus10014324	20.1405	10.8543	-0.89183	0.007053
Lus10019376	bZIP	Lus10019376	24.2502	13.9006	-0.80285	0.006189
Lus10024204	bZIP	Lus10024204	25.3632	15.1374	-0.744619	0.018807
Lus10024847	bZIP	Lus10024847	7.26988	3.17317	-1.19601	0.005019
Lus10028889	bZIP	Lus10028889	25.3739	15.3857	-0.721754	0.011791
Lus10034296	bZIP	Lus10034296	22.7699	3.44208	-2.72578	0.000269
Lus10041475	bZIP	Lus10041475	36.0449	2.30484	-3.96706	0.000269
Lus10003681	C2H2	Lus10003681	21.7771	10.956	-0.991089	0.003928
Lus10007474	C2H2	Lus10007474	4.20508	1.47271	-1.51366	0.023873
Lus10014974	C2H2	Lus10014974	2.95139	1.20588	-1.2913	0.02364
Lus10019482	C2H2	Lus10019482	13.2703	3.81489	-1.79849	0.000269
Lus10026497	C2H2	Lus10026497	34.5553	17.9862	-0.942019	0.001134
Lus10028951	C2H2	Lus10028951	4.95831	1.51963	-1.70613	0.012964
Lus10031838	C2H2	Lus10031838	4.78714	1.47291	-1.70049	0.00719
Lus10033148	C2H2	Lus10033148	6.62503	3.00966	-1.13833	0.011925
Lus10037989	C2H2	Lus10037989	6.30202	2.01855	-1.64249	0.000269
Lus10043332	C2H2	Lus10043332	9.75769	2.72307	-1.84131	0.000269
Lus10000486	C3H	Lus10000486	33.6058	13.5819	-1.30703	0.000269
Lus10004573	C3H	Lus10004573	35.1927	16.4099	-1.10071	0.000505
Lus10007941	C3H	Lus10007941	25.2425	15.0198	-0.748991	0.009796
Lus10013476	C3H	Lus10013476	15.3217	9.54911	-0.682138	0.022509
Lus10014490	C3H	Lus10014490	7.4565	4.55335	-0.71157	0.040266
Lus10019481	C3H	Lus10019481	31.3856	18.3486	-0.774432	0.005317
Lus10025973	C3H	Lus10025973	26.6264	17.0562	-0.642565	0.021639
Lus10028950	C3H	Lus10028950	6.00136	3.71211	-0.693047	0.026912
Lus10030063	C3H	Lus10030063	4.36019	2.50142	-0.801645	0.039252
Lus10035248	C3H	Lus10035248	21.6093	9.30515	-1.21555	0.030962

Lus10035460	C3H	Lus10035460	24.7377	12.3243	-1.0052	0.000269
Lus10002033	CPP	Lus10002033	20.6463	3.23572	-2.67372	0.000269
Lus10002895	CPP	Lus10002895	19.7321	2.83941	-2.79689	0.000269
Lus10006604	CPP	Lus10006604	22.2638	8.8122	-1.33713	0.000269
Lus10009494	CPP	Lus10009494	6.22397	1.96193	-1.66556	0.000269
Lus10011693	CPP	Lus10011693	8.1995	3.72857	-1.13691	0.010454
Lus10023656	CPP	Lus10023656	8.39024	3.48974	-1.26559	0.000725
Lus10039358	CPP	Lus10039358	35.0556	12.2365	-1.51846	0.000269
Lus10005677	Dof	Lus10005677	7.34024	3.20034	-1.1976	0.00647
Lus10014001	Dof	Lus10014001	43.9974	28.0264	-0.650635	0.042979
Lus10020314	Dof	Lus10020314	4.643	1.69803	-1.45119	0.010058
Lus10004217	E2F/DP	Lus10004217	8.6327	3.29397	-1.38998	0.003598
Lus10014423	E2F/DP	Lus10014423	10.6867	5.52932	-0.950642	0.015205
Lus10016972	E2F/DP	Lus10016972	5.20186	1.64735	-1.65888	0.000269
Lus10021298	E2F/DP	Lus10021298	3.32101	1.41262	-1.23325	0.008862
Lus10023926	E2F/DP	Lus10023926	8.18061	3.42811	-1.25479	0.006763
Lus10029421	E2F/DP	Lus10029421	14.0578	2.38342	-2.56026	0.000269
Lus10032439	E2F/DP	Lus10032439	7.21715	0.898498	-3.00584	0.002413
Lus10033151	E2F/DP	Lus10033151	4.8278	2.27264	-1.087	0.015205
Lus10042941	E2F/DP	Lus10042941	7.36074	0.934743	-2.97721	0.001328
Lus10011319	ERF	Lus10011319	4.55112	1.77091	-1.36173	0.000269
Lus10016245	ERF	Lus10016245	4.03393	1.87482	-1.10544	0.016042
Lus10016827	ERF	Lus10016827	43.8834	13.7551	-1.67371	0.000269
Lus10032353	ERF	Lus10032353	4.80218	1.45211	-1.72553	0.012182
Lus10033938	ERF	Lus10033938	8.25895	1.87418	-2.1397	0.001134
Lus10037487	ERF	Lus10037487	11.3703	4.82798	-1.23578	0.01371
Lus10038607	ERF	Lus10038607	47.1958	3.18951	-3.88725	0.000269
Lus10020226	FAR1	Lus10020226	31.3077	18.666	-0.746107	0.021738
Lus10007132	G2-like	Lus10007132	84.1973	19.6685	-2.09789	0.000269
Lus10011660	G2-like	Lus10011660	8.70046	1.34509	-2.69339	0.000269
Lus10016676	G2-like	Lus10016676	34.3139	16.2243	-1.08063	0.000725
Lus10029607	G2-like	Lus10029607	14.6522	8.97437	-0.70723	0.0182
Lus10030989	G2-like	Lus10030989	28.7231	15.3113	-0.907619	0.001886
Lus10032746	G2-like	Lus10032746	10.5431	4.10853	-1.3596	0.008455
Lus10035043	G2-like	Lus10035043	14.1469	3.6492	-1.95483	0.012828
Lus10035093	G2-like	Lus10035093	12.3215	0.632916	-4.28302	0.00224
Lus10036758	G2-like	Lus10036758	1.52564	0.688628	-1.14762	0.611186
Lus10037169	G2-like	Lus10037169	60.7345	19.6132	-1.63069	0.000269

Lus10002412	GATA	Lus10002412	38.4802	23.5859	-0.706192	0.028322
Lus10020684	GATA	Lus10020684	14.884	4.81875	-1.62703	0.006615
Lus10021466	GATA	Lus10021466	85.2777	46.4483	-0.876542	0.003928
Lus10028301	GATA	Lus10028301	16.3549	2.81979	-2.53606	0.003598
Lus10029863	GATA	Lus10029863	16.846	4.07705	-2.04681	0.003765
Lus10031464	GATA	Lus10031464	3.16536	0.7161	-2.14414	0.031175
Lus10037398	GATA	Lus10037398	15.2611	1.69493	-3.17056	0.000269
Lus10037721	GATA	Lus10037721	55.1948	29.7086	-0.893649	0.009137
Lus10041810	GATA	Lus10041810	46.5335	29.0681	-0.678835	0.023072
Lus10002794	GeBP	Lus10002794	101.029	37.355	-1.4354	0.000269
Lus10004772	GeBP	Lus10004772	74.2337	40.9571	-0.857962	0.002066
Lus10005506	GeBP	Lus10005506	51.7187	24.6067	-1.07163	0.000269
Lus10007188	GeBP	Lus10007188	20.9319	6.12406	-1.77314	0.044436
Lus10018859	GeBP	Lus10018859	52.2376	23.8631	-1.13031	0.000269
Lus10004353	GRAS	Lus10004353	43.9259	18.8497	-1.22053	0.000269
Lus10006322	GRAS	Lus10006322	11.3765	3.86339	-1.55811	0.000269
Lus10010462	GRAS	Lus10010462	19.8154	7.69618	-1.36441	0.000269
Lus10011542	GRAS	Lus10011542	27.0204	12.5784	-1.1031	0.000269
Lus10012554	GRAS	Lus10012554	1.70599	0.574125	-1.57117	0.013955
Lus10024014	GRAS	Lus10024014	14.0901	6.6888	-1.07486	0.000505
Lus10028934	GRAS	Lus10028934	42.0179	17.3442	-1.27655	0.000269
Lus10029592	GRAS	Lus10029592	12.9722	5.00042	-1.3753	0.000269
Lus10039709	GRAS	Lus10039709	3.84595	0.402354	-3.2568	0.009402
Lus10040284	GRAS	Lus10040284	4.8274	2.77859	-0.796893	0.025565
Lus10041740	GRAS	Lus10041740	48.6467	17.6777	-1.46041	0.000269
Lus10008268	GRF	Lus10008268	6.75094	1.91052	-1.82112	0.047168
Lus10009533	GRF	Lus10009533	90.6053	10.5288	-3.10526	0.000269
Lus10011558	GRF	Lus10011558	95.1001	4.62038	-4.36336	0.000269
Lus10011559	GRF	Lus10011559	88.8187	2.13561	-5.37814	0.000269
Lus10019274	GRF	Lus10019274	80.1183	5.62643	-3.83184	0.000269
Lus10019275	GRF	Lus10019275	97.2257	4.62303	-4.39443	0.000269
Lus10020352	GRF	Lus10020352	147.971	12.4219	-3.57436	0.000269
Lus10033236	GRF	Lus10033236	10.1621	1.34172	-2.92105	0.000933
Lus10033441	GRF	Lus10033441	67.5686	1.02108	-6.04819	0.000269
Lus10037668	GRF	Lus10037668	81.3636	0.644086	-6.98099	0.015447
Lus10009816	HB-other	Lus10009816	9.69615	4.10754	-1.23914	0.000269
Lus10013684	HB-other	Lus10013684	23.9424	13.3448	-0.843287	0.008862
Lus10017688	HB-other	Lus10017688	38.5931	20.7485	-0.895335	0.007478

Lus10017944	HB-other	Lus10017944	36.3068	17.6527	-1.04035	0.001328
Lus10018634	HB-other	Lus10018634	13.162	9.0416	-0.541732	0.049386
Lus10024826	HB-other	Lus10024826	9.55683	5.89195	-0.697788	0.014702
Lus10039870	HB-other	Lus10039870	21.3454	12.85	-0.732153	0.006044
Lus10040921	HB-other	Lus10040921	11.0418	5.37293	-1.03919	0.000269
Lus10018741	HB-PHD	Lus10018741	23.2578	9.00685	-1.36862	0.000269
Lus10021064	HD-ZIP	Lus10021064	21.8551	10.4094	-1.07008	0.001886
Lus10007849	HD-ZIP	Lus10007849	13.8676	5.64204	-1.29743	0.002925
Lus10004759	HD-ZIP	Lus10004759	13.8362	3.74063	-1.8871	0.000269
Lus10023159	HD-ZIP	Lus10023159	182.187	71.4942	-1.34953	0.000269
Lus10007650	HD-ZIP	Lus10007650	13.6199	3.22262	-2.07941	0.000269
Lus10006765	HD-ZIP	Lus10006765	13.0918	3.39828	-1.94579	0.000269
Lus10020059	HD-ZIP	Lus10020059	13.3096	2.10707	-2.65915	0.000269
Lus10031321	HD-ZIP	Lus10031321	6.16656	3.18453	-0.953386	0.008313
Lus10038449	HD-ZIP	Lus10038449	63.0696	23.1685	-1.44478	0.000269
Lus10023357	HD-ZIP	Lus10023357	54.9254	19.3472	-1.50535	0.000269
Lus10031892	HD-ZIP	Lus10031892	8.43449	2.12476	-1.989	0.000269
Lus10011941	HSF	Lus10011941	4.44504	2.10751	-1.07666	0.029509
Lus10036062	HSF	Lus10036062	19.3628	9.8738	-0.971608	0.004241
Lus10042646	HSF	Lus10042646	54.3264	17.8197	-1.60818	0.000269
Lus10003789	LBD	Lus10003789	1.99061	0.567196	-1.81129	0.03847
Lus10011906	LBD	Lus10011906	3.80441	0.329837	-3.52785	0.025457
Lus10023591	LBD	Lus10023591	84.9812	10.2166	-3.05623	0.000269
Lus10016809	M-type_MADS	Lus10016809	8.55524	0.306089	-4.80479	0.006189
Lus10026613	M-type_MADS	Lus10026613	9.3578	2.80586	-1.73772	0.000269
Lus10027404	MIKC_MADS	Lus10027404	65.2698	13.5583	-2.26724	0.000269
Lus10031665	MIKC_MADS	Lus10031665	52.2961	14.3946	-1.86117	0.000269
Lus10033187	MIKC_MADS	Lus10033187	123.848	83.41	-0.570274	0.048313
Lus10011687	MYB	Lus10011687	21.9107	6.60439	-1.73014	0.000269
Lus10021762	MYB	Lus10021762	9.23962	3.44576	-1.42301	0.01308
Lus10022136	MYB	Lus10022136	18.8325	5.84628	-1.68763	0.000269
Lus10024392	MYB	Lus10024392	40.2698	17.905	-1.16934	0.02376
Lus10025355	MYB	Lus10025355	49.7874	19.7324	-1.33522	0.000269
Lus10026611	MYB	Lus10026611	39.4448	6.72053	-2.55319	0.000269
Lus10027459	MYB	Lus10027459	8.99345	1.46278	-2.62016	0.00224
Lus10030378	MYB	Lus10030378	22.3264	1.62844	-3.77719	0.000269
Lus10030452	MYB	Lus10030452	45.8304	6.51718	-2.81399	0.000269

Lus10034133	MYB	Lus10034133	17.1577	8.01194	-1.09863	0.002925
Lus10036453	MYB	Lus10036453	4.70366	0.494854	-3.24871	0.017971
Lus10037898	MYB	Lus10037898	7.64026	4.96034	-0.623183	0.03847
Lus10038623	MYB	Lus10038623	7.84214	4.67747	-0.745519	0.013324
Lus10039214	MYB	Lus10039214	7.64617	0.415381	-4.20223	0.029615
Lus10043451	MYB	Lus10043451	19.3934	11.2455	-0.786218	0.015928
Lus10004489	MYB_related	Lus10004489	15.2946	4.15443	-1.8803	0.000269
Lus10012209	MYB_related	Lus10012209	2.85322	0.949171	-1.58785	0.001328
Lus10012602	MYB_related	Lus10012602	13.9989	8.50102	-0.719602	0.03002
Lus10014653	MYB_related	Lus10014653	9.98965	6.24788	-0.677069	0.040579
Lus10017319	MYB_related	Lus10017319	6.47396	3.52099	-0.878666	0.028421
Lus10020117	MYB_related	Lus10020117	12.6128	5.94679	-1.0847	0.000505
Lus10026522	MYB_related	Lus10026522	13.0719	3.3091	-1.98196	0.000269
Lus10031893	MYB_related	Lus10031893	15.4744	5.02099	-1.62384	0.000269
Lus10033961	MYB_related	Lus10033961	20.464	10.2756	-0.99386	0.001134
Lus10038846	MYB_related	Lus10038846	85.2884	52.6424	-0.696125	0.014702
Lus10040453	MYB_related	Lus10040453	1907.74	18.3992	-6.69608	0.000269
Lus10042209	MYB_related	Lus10042209	11.7557	6.09393	-0.947912	0.005164
Lus10007216	NAC	Lus10007216	18.7073	10.4557	-0.83931	0.005019
Lus10013205	NAC	Lus10013205	21.626	0.481261	-5.48981	0.002066
Lus10020794	NAC	Lus10020794	230.557	117.636	-0.970786	0.000269
Lus10021708	NAC	Lus10021708	17.8751	3.05968	-2.54649	0.000269
Lus10032004	NAC	Lus10032004	24.4355	7.07754	-1.78766	0.000269
Lus10035174	NAC	Lus10035174	23.1034	5.45486	-2.08249	0.000269
Lus10035400	NAC	Lus10035400	22.4263	3.16971	-2.82277	0.000269
Lus10037939	NAC	Lus10037939	55.1023	24.1878	-1.18783	0.000269
Lus10038670	NAC	Lus10038670	17.3744	9.577	-0.859319	0.001704
Lus10041492	NAC	Lus10041492	19.4372	6.19295	-1.65012	0.000269
Lus10021259	NF-YA	Lus10021259	19.9682	11.7725	-0.762281	0.020455
Lus10031505	NF-YA	Lus10031505	9.34034	2.5335	-1.88234	0.000269
Lus10004867	NF-YB	Lus10004867	28.2466	2.14767	-3.71723	0.002759
Lus10020621	NF-YB	Lus10020621	41.6559	14.0219	-1.57084	0.000269
Lus10023167	NF-YB	Lus10023167	16.6832	5.89025	-1.502	0.000269
Lus10027242	NF-YB	Lus10027242	65.3726	23.2564	-1.49106	0.000269
Lus10038952	NF-YB	Lus10038952	56.1689	22.0094	-1.35165	0.000269
Lus10008701	NF-YC	Lus10008701	23.4199	6.16286	-1.92606	0.04802
Lus10021934	NF-YC	Lus10021934	75.3695	37.847	-0.993802	0.015809
Lus10026118	NF-YC	Lus10026118	33.0766	12.5752	-1.39522	0.000269

Lus10026780	NF-YC	Lus10026780	28.4624	10.5037	-1.43816	0.000269
Lus10030657	NF-YC	Lus10030657	54.5632	5.80089	-3.23358	0.000269
Lus10030832	NF-YC	Lus10030832	28.2339	3.48821	-3.01687	0.00224
Lus10041221	NF-YC	Lus10041221	58.0473	13.9305	-2.05898	0.000269
Lus10041638	NF-YC	Lus10041638	33.9554	16.9333	-1.00378	0.006763
Lus10023049	Nin-like	Lus10023049	2.77547	0.817441	-1.76355	0.001704
Lus10023345	S1Fa-like	Lus10023345	12.7145	5.11411	-1.31392	0.000269
Lus10003126	SBP	Lus10003126	45.2711	7.9427	-2.51089	0.000269
Lus10006411	SBP	Lus10006411	27.2988	16.782	-0.701926	0.019052
Lus10007984	SBP	Lus10007984	11.7213	5.22278	-1.16624	0.000269
Lus10012020	SBP	Lus10012020	69.513	19.6781	-1.82069	0.000269
Lus10016275	SBP	Lus10016275	55.6692	13.7061	-2.02206	0.000269
Lus10018610	SBP	Lus10018610	10.226	4.37347	-1.22539	0.034988
Lus10021034	SBP	Lus10021034	76.3218	10.0246	-2.92856	0.000269
Lus10021141	SBP	Lus10021141	7.38151	2.08235	-1.8257	0.000269
Lus10021614	SBP	Lus10021614	17.9974	9.74403	-0.885197	0.003431
Lus10023818	SBP	Lus10023818	9.4356	1.17525	-3.00514	0.01072
Lus10028181	SBP	Lus10028181	9.20218	0.620645	-3.89013	0.001886
Lus10003416	SRS	Lus10003416	7.63626	1.66981	-2.19318	0.003765
Lus10009697	SRS	Lus10009697	11.867	1.44753	-3.03529	0.000269
Lus10024305	SRS	Lus10024305	8.14942	2.19952	-1.88951	0.002066
Lus10028352	SRS	Lus10028352	4.45621	1.28063	-1.79896	0.009402
Lus10036032	SRS	Lus10036032	16.035	2.85176	-2.49129	0.000933
Lus10041802	SRS	Lus10041802	12.1542	2.23714	-2.44173	0.031705
Lus10005584	STAT	Lus10005584	3.69525	1.37827	-1.42281	0.002066
Lus10013716	STAT	Lus10013716	7.18105	3.11944	-1.20291	0.000725
Lus10021452	TALE	Lus10021452	27.2668	2.62122	-3.37883	0.000269
Lus10026432	TALE	Lus10026432	59.5062	2.68665	-4.46916	0.000269
Lus10016110	TALE	Lus10016110	19.0657	2.0601	-3.21019	0.000269
Lus10004688	TALE	Lus10004688	94.6156	3.31321	-4.83578	0.000269
Lus10040256	TALE	Lus10040256	132.369	7.18552	-4.20333	0.000269
Lus10030003	TALE	Lus10030003	78.1272	16.2361	-2.26662	0.000269
Lus10042102	TALE	Lus10042102	17.8971	1.34246	-3.73677	0.000269
Lus10001238	TALE	Lus10001238	16.2423	0.856825	-4.24461	0.002413
Lus10008643	Trihelix	Lus10008643	11.3275	3.69969	-1.61435	0.000269
Lus10008988	Trihelix	Lus10008988	29.0333	19.7089	-0.558858	0.04376
Lus10009184	Trihelix	Lus10009184	22.5344	11.3722	-0.98661	0.010319
Lus10014375	Trihelix	Lus10014375	7.43473	2.36314	-1.65357	0.000505

Lus10015924	Trihelix	Lus10015924	27.1425	17.0855	-0.667783	0.01244
Lus10023872	Trihelix	Lus10023872	1.9985	0.197269	-3.34068	0.031814
Lus10027718	Trihelix	Lus10027718	15.3247	6.05119	-1.34056	0.000269
Lus10035570	Trihelix	Lus10035570	6.02376	2.25069	-1.42029	0.002588
Lus10035582	Trihelix	Lus10035582	18.7892	8.12173	-1.21004	0.000269
Lus10036723	Whirly	Lus10036723	48.8766	17.4724	-1.48407	0.000269
Lus10037206	Whirly	Lus10037206	32.7113	13.8425	-1.24069	0.000269
Lus10014745	WRKY	Lus10014745	75.6406	21.6982	-1.80158	0.000269
Lus10019898	WRKY	Lus10019898	37.5685	3.1142	-3.59259	0.000269
Lus10021999	WRKY	Lus10021999	7.86614	3.60393	-1.12608	0.016649
Lus10022150	WRKY	Lus10022150	18.8242	12.0215	-0.646962	0.023523
Lus10024864	WRKY	Lus10024864	16.2956	8.73738	-0.899212	0.007889
Lus10033857	WRKY	Lus10033857	58.8224	16.6158	-1.82381	0.000269
Lus10036268	WRKY	Lus10036268	15.5627	10.6049	-0.553355	0.047469
Lus10042538	WRKY	Lus10042538	15.1971	6.1034	-1.31611	0.002588
Lus10019407	YABBY	Lus10019407	28.4356	11.5631	-1.29816	0.002759
Lus10030105	YABBY	Lus10030105	10.1762	0.636055	-3.9999	0.002925
Lus10005244	ZF-HD	Lus10005244	30.4133	14.5787	-1.06084	0.017971
Lus10007147	ZF-HD	Lus10007147	135.613	2.27999	-5.89432	0.000269
Lus10014302	ZF-HD	Lus10014302	2.85666	0.171756	-4.05589	0.037744
Lus10038135	ZF-HD	Lus10038135	83.1918	0.573075	-7.18157	0.00719
Lus10037670	AP2	Lus10037670	4.12534	0	inf	0.000269
Lus10000747	B3	Lus10000747	1.10208	0	inf	0.000269
Lus10012046	B3	Lus10012046	12.7802	0	inf	0.000269
Lus10012226	ERF	Lus10012226	4.13865	0	inf	0.210832
Lus10014345	ERF	Lus10014345	3.13922	0	inf	0.000269
Lus10015653	ERF	Lus10015653	37.4187	0	inf	0.000269
Lus10032882	GRAS	Lus10032882	4.6761	0	inf	0.000269
Lus10014380	GRF	Lus10014380	115.894	0	inf	0.075569
Lus10030800	HD-ZIP	Lus10030800	3.84587	0	inf	0.000269
Lus10009336	LBD	Lus10009336	2.58798	0	inf	0.000269
Lus10016732	LFY	Lus10016732	1.05836	0	inf	0.000269
Lus10028214	M-type_MADS	Lus10028214	10.032	0	inf	0.22514
Lus10035029	M-type_MADS	Lus10035029	2.22046	0	inf	0.43309
Lus10016139	MYB	Lus10016139	3.38138	0	inf	0.000269
Lus10018518	MYB	Lus10018518	2.88378	0	inf	0.000269
Lus10021428	MYB	Lus10021428	10.6303	0	inf	0.000269

Lus10038092	MYB	Lus10038092	30.9376	0	inf	0.000269
Lus10007643	MYB_related	Lus10007643	6.80047	0	inf	0.008862
Lus10014933	MYB_related	Lus10014933	2.94253	0	inf	0.060625
Lus10023568	MYB_related	Lus10023568	1.80891	0	inf	0.066097
Lus10041924	NAC	Lus10041924	11.7809	0	inf	0.000269
Lus10018283	Trihelix	Lus10018283	41.1782	0	inf	0.000269
Lus10027398	Trihelix	Lus10027398	3.66274	0	inf	0.000269
Lus10031672	Trihelix	Lus10031672	5.34632	0	inf	0.000269
Lus10005282	WOX	Lus10005282	2.26594	0	inf	0.000269
Lus10013960	WOX	Lus10013960	5.56958	0	inf	0.000269

Appendix 4. List of putative flax MYBs and their Arabidopsis orthologs.

Gene Symbol	Gene ID	*Arabidopsis* Ortholog	*Arabidopsis* Ortholog Description	E-value
R2R3-MYB				
LusMYB1	Lus10038062	AT2G47190.1	*ATMYB002*	4.10E-70
LusMYB2	Lus10009996	AT2G47190.1	*ATMYB002*	7.18E-69
LusMYB3	Lus10033438	AT1G22640.1	*ATMYB003*	3.80E-60
LusMYB4	Lus10028435	AT4G38620.1	*ATMYB004*	3.17E-79
LusMYB5	Lus10039173	AT3G13540.1	*ATMYB005*	3.14E-75
LusMYB6	Lus10013762	AT3G13540.1	*ATMYB005*	1.05E-74
LusMYB7	Lus10000411	AT4G09460.1	*ATMYB006*	2.57E-100
LusMYB8	Lus10016948	AT4G09460.1	*ATMYB006*	2.42E-96
LusMYB9	Lus10001548	AT4G09460.1	*ATMYB006*	6.40E-93
LusMYB10	Lus10009448	AT4G09460.1	*ATMYB006*	2.24E-92
LusMYB11	Lus10000470	AT2G16720.1	*ATMYB007*	3.92E-60
LusMYB12	Lus10041888	AT2G16720.1	*ATMYB007*	5.73E-79
LusMYB13	Lus10014129	AT2G16720.1	*ATMYB007*	3.45E-76
LusMYB14	Lus10033473	AT2G16720.1	*ATMYB007*	4.52E-81
LusMYB15	Lus10040139	AT5G16770.2	*ATMYB009*	3.79E-69
LusMYB16	Lus10001093	AT5G16770.2	*ATMYB009*	8.41E-68
LusMYB17	Lus10033737	AT5G16770.2	*ATMYB009*	4.01E-47
LusMYB18	Lus10011031	AT5G16770.2	*ATMYB009*	1.14E-64
LusMYB19	Lus10036336	AT2G47460.1	*ATMYB012*	8.31E-68
LusMYB20	Lus10002435	AT2G47460.1	*ATMYB012*	2.68E-62
LusMYB21	Lus10001458	AT2G47460.1	*ATMYB012*	1.29E-63
LusMYB22	Lus10010273	AT2G47460.1	*ATMYB012*	2.77E-67
LusMYB23	Lus10033889	AT2G31180.1	*ATMYB014*	1.23E-72
LusMYB24	Lus10042561	AT2G31180.1	*ATMYB014*	1.00E-65
LusMYB25	Lus10003557	AT2G31180.1	*ATMYB014*	5.52E-70
LusMYB26	Lus10018518	AT2G31180.1	*ATMYB014*	1.55E-48
LusMYB27	Lus10022021	AT2G31180.1	*ATMYB014*	8.09E-69
LusMYB28	Lus10011820	AT2G31180.1	*ATMYB014*	5.49E-69
LusMYB29	Lus10041145	AT3G23250.1	*ATMYB015*	1.19E-72
LusMYB30	Lus10021185	AT3G23250.1	*ATMYB015*	1.17E-70
LusMYB31	Lus10021871	AT3G23250.1	*ATMYB015*	1.35E-73
LusMYB32	Lus10026620	AT5G15310.1	*ATMYB016*	3.09E-95
LusMYB33	Lus10033003	AT5G15310.2	*ATMYB016*	5.75E-98

LusMYB34	Lus10039214	AT3G61250.1	*ATMYB017*	1.42E-61
LusMYB35	Lus10027459	AT3G61250.1	*ATMYB017*	1.26E-70
LusMYB36	Lus10030378	AT3G61250.1	*ATMYB017*	9.59E-105
LusMYB37	Lus10014784	AT3G61250.1	*ATMYB017*	3.26E-110
LusMYB38	Lus10005740	AT4G25560.1	*ATMYB018*	7.85E-57
LusMYB39	Lus10039213	AT4G25560.1	*ATMYB018*	9.08E-58
LusMYB40	Lus10027458	AT4G25560.1	*ATMYB018*	6.76E-58
LusMYB41	Lus10005739	AT5G52260.1	*ATMYB019*	5.83E-56
LusMYB42	Lus10004042	AT1G66230.1	*ATMYB020*	2.44E-77
LusMYB43	Lus10004043	AT1G66230.1	*ATMYB020*	1.98E-80
LusMYB44	Lus10038913	AT1G66230.1	*ATMYB020*	6.15E-81
LusMYB45	Lus10027197	AT1G66230.1	*ATMYB020*	1.66E-81
LusMYB46	Lus10002296	AT1G66230.1	*ATMYB020*	1.38E-79
LusMYB47	Lus10022259	AT3G27810.1	*ATMYB021*	9.39E-78
LusMYB48	Lus10013081	AT3G27810.1	*ATMYB021*	4.50E-51
LusMYB49	Lus10032129	AT5G40350.1	*ATMYB024*	4.78E-68
LusMYB50	Lus10014557	AT5G40350.1	*ATMYB024*	1.86E-68
LusMYB51	Lus10015608	AT3G13890.2	*ATMYB026*	1.86E-72
LusMYB52	Lus10023918	AT3G53200.1	*ATMYB027*	1.01E-52
LusMYB53	Lus10014415	AT3G53200.1	*ATMYB027*	2.84E-53
LusMYB54	Lus10005245	AT3G28910.1	*ATMYB030*	1.70E-81
LusMYB55	Lus10030677	AT3G28910.1	*ATMYB030*	3.43E-90
LusMYB56	Lus10015369	AT3G28910.1	*ATMYB030*	1.71E-92
LusMYB57	Lus10039462	AT3G28470.1	*ATMYB035*	1.41E-86
LusMYB58	Lus10036660	AT3G28470.1	*ATMYB035*	3.01E-64
LusMYB59	Lus10005834	AT3G28470.1	*ATMYB035*	1.13E-86
LusMYB60	Lus10033119	AT3G28470.1	*ATMYB035*	2.00E-65
LusMYB61	Lus10021428	AT5G57620.1	*ATMYB036*	7.72E-76
LusMYB62	Lus10013830	AT5G57620.1	*ATMYB036*	3.02E-67
LusMYB63	Lus10006978	AT5G57620.1	*ATMYB036*	5.46E-76
LusMYB64	Lus10001394	AT5G57620.1	*ATMYB036*	7.36E-74
LusMYB65	Lus10001316	AT5G57620.1	*ATMYB036*	8.13E-76
LusMYB66	Lus10016139	AT5G57620.1	*ATMYB036*	3.70E-74
LusMYB67	Lus10023002	AT2G36890.1	*ATMYB038*	3.11E-74
LusMYB68	Lus10014569	AT5G14340.1	*ATMYB040*	3.31E-75
LusMYB69	Lus10032117	AT5G14340.1	*ATMYB040*	8.41E-77
LusMYB70	Lus10031607	AT4G28110.1	*ATMYB041*	4.24E-54

LusMYB71	Lus10033738	AT4G28110.1	*ATMYB041*	2.28E-54
LusMYB72	Lus10032226	AT4G12350.1	*ATMYB042*	1.49E-87
LusMYB73	Lus10024589	AT4G12350.1	*ATMYB042*	2.17E-86
LusMYB74	Lus10010974	AT5G16600.1	*ATMYB043*	3.54E-67
LusMYB75	Lus10010238	AT5G67300.1	*ATMYB044*	1.12E-70
LusMYB76	Lus10039610	AT5G12870.1	*ATMYB046*	7.17E-67
LusMYB77	Lus10031850	AT5G12870.1	*ATMYB046*	1.29E-65
LusMYB78	Lus10002559	AT5G12870.1	*ATMYB046*	2.51E-66
LusMYB79	Lus10027369	AT5G12870.1	*ATMYB046*	7.51E-65
LusMYB80	Lus10029520	AT5G12870.1	*ATMYB046*	5.55E-66
LusMYB81	Lus10005886	AT3G46130.1	*ATMYB048*	4.54E-84
LusMYB82	Lus10029746	AT1G17950.1	*ATMYB052*	1.85E-60
LusMYB83	Lus10031326	AT1G17950.1	*ATMYB052*	9.25E-64
LusMYB84	Lus10031900	AT1G17950.1	*ATMYB052*	2.74E-67
LusMYB85	Lus10039734	AT5G65230.1	*ATMYB053*	1.03E-55
LusMYB86	Lus10039735	AT5G65230.1	*ATMYB053*	1.47E-52
LusMYB87	Lus10038022	AT4G01680.3	*ATMYB055*	3.54E-87
LusMYB88	Lus10005683	AT5G17800.1	*ATMYB056*	8.27E-57
LusMYB89	Lus10020343	AT1G08810.1	*ATMYB060*	1.51E-80
LusMYB90	Lus10009522	AT1G08810.1	*ATMYB060*	8.96E-81
LusMYB91	Lus10009037	AT1G68320.1	*ATMYB062*	8.43E-71
LusMYB92	Lus10034338	AT1G68320.1	*ATMYB062*	3.71E-82
LusMYB93	Lus10041435	AT1G68320.1	*ATMYB062*	1.48E-80
LusMYB94	Lus10026787	AT3G11440.1	*ATMYB065*	2.83E-87
LusMYB95	Lus10026142	AT3G11440.1	*ATMYB065*	8.63E-108
LusMYB96	Lus10008685	AT3G11440.1	*ATMYB065*	3.10E-108
LusMYB97	Lus10036103	AT3G11440.1	*ATMYB065*	1.88E-92
LusMYB98	Lus10038395	AT3G12720.1	*ATMYB067*	9.14E-67
LusMYB99	Lus10001226	AT3G12720.1	*ATMYB067*	3.68E-71
LusMYB100	Lus10026543	AT5G65790.1	*ATMYB068*	9.18E-67
LusMYB101	Lus10032764	AT2G23290.1	*ATMYB070*	5.37E-54
LusMYB102	Lus10021762	AT2G23290.1	*ATMYB070*	2.18E-53
LusMYB103	Lus10014453	AT3G24310.1	*ATMYB071*	5.63E-66
LusMYB104	Lus10023711	AT3G24310.1	*ATMYB071*	5.46E-66
LusMYB105	Lus10030336	AT4G37260.1	*ATMYB073*	2.85E-73
LusMYB106	Lus10040940	AT4G37260.1	*ATMYB073*	7.10E-53
LusMYB107	Lus10007503	AT4G37260.1	*ATMYB073*	5.33E-55

LusMYB108	Lus10010055	AT4G37260.1	*ATMYB073*	2.50E-75
LusMYB109	Lus10028979	AT4G37260.1	*ATMYB073*	9.45E-52
LusMYB110	Lus10014103	AT4G37260.1	*ATMYB073*	8.27E-61
LusMYB111	Lus10010260	AT4G37260.1	*ATMYB073*	7.28E-82
LusMYB112	Lus10019085	AT5G07700.1	*ATMYB076*	2.49E-68
LusMYB113	Lus10016413	AT3G08500.1	*ATMYB083*	3.16E-49
LusMYB114	Lus10019707	AT3G08500.1	*ATMYB083*	2.61E-49
LusMYB115	Lus10031281	AT3G08500.1	*ATMYB083*	7.04E-70
LusMYB116	Lus10028248	AT3G49690.1	*ATMYB084*	1.24E-80
LusMYB117	Lus10039646	AT3G49690.1	*ATMYB084*	3.01E-76
LusMYB118	Lus10011606	AT3G49690.1	*ATMYB084*	1.06E-76
LusMYB119	Lus10007248	AT3G49690.1	*ATMYB084*	1.84E-81
LusMYB120	Lus10030494	AT2G02820.2	*ATMYB088*	6.12E-147
LusMYB121	Lus10041142	AT1G34670.1	*ATMYB093*	1.92E-100
LusMYB122	Lus10018546	AT1G34670.1	*ATMYB093*	1.79E-52
LusMYB123	Lus10036472	AT1G34670.1	*ATMYB093*	7.05E-98
LusMYB124	Lus10039771	AT1G34670.1	*ATMYB093*	7.34E-56
LusMYB125	Lus10042200	AT3G47600.1	*ATMYB094*	6.19E-93
LusMYB126	Lus10024218	AT3G47600.1	*ATMYB094*	1.71E-85
LusMYB127	Lus10002056	AT3G47600.1	*ATMYB094*	8.42E-86
LusMYB128	Lus10008616	AT5G62470.2	*ATMYB096*	3.15E-85
LusMYB129	Lus10027189	AT2G32460.1	*ATMYB101*	1.89E-71
LusMYB130	Lus10040063	AT2G32460.2	*ATMYB101*	3.63E-63
LusMYB131	Lus10035275	AT2G32460.2	*ATMYB101*	3.01E-68
LusMYB132	Lus10018418	AT4G21440.1	*ATMYB102*	5.28E-98
LusMYB133	Lus10018547	AT4G21440.1	*ATMYB102*	6.91E-53
LusMYB134	Lus10020085	AT4G21440.1	*ATMYB102*	1.43E-97
LusMYB135	Lus10039743	AT4G21440.1	*ATMYB102*	1.97E-94
LusMYB136	Lus10006740	AT4G21440.1	*ATMYB102*	1.25E-88
LusMYB137	Lus10002593	AT4G21440.1	*ATMYB102*	1.23E-94
LusMYB138	Lus10039772	AT4G21440.1	*ATMYB102*	9.31E-52
LusMYB139	Lus10032298	AT1G63910.1	*ATMYB103*	6.36E-81
LusMYB140	Lus10024669	AT1G63910.1	*ATMYB103*	8.15E-81
LusMYB141	Lus10030452	AT1G69560.1	*ATMYB105*	2.65E-74
LusMYB142	Lus10026611	AT1G69560.1	*ATMYB105*	2.17E-73
LusMYB143	Lus10015712	AT3G01140.1	*ATMYB106*	4.61E-86
LusMYB144	Lus10019086	AT3G01140.1	*ATMYB106*	4.40E-88

LusMYB145	Lus10037818	AT3G06490.1	*ATMYB108*	4.71E-87
LusMYB146	Lus10017096	AT3G06490.1	*ATMYB108*	1.01E-82
LusMYB147	Lus10028250	AT3G55730.1	*ATMYB109*	5.61E-91
LusMYB148	Lus10040239	AT3G55730.1	*ATMYB109*	2.49E-91
LusMYB149	Lus10036453	AT5G49330.1	*ATMYB111*	1.21E-71
LusMYB150	Lus10016855	AT5G49330.1	*ATMYB111*	1.18E-69
LusMYB151	Lus10003277	AT1G66370.1	*ATMYB113*	2.98E-46
LusMYB152	Lus10009130	AT1G66370.1	*ATMYB113*	4.54E-60
LusMYB153	Lus10042522	AT1G66370.1	*ATMYB113*	3.39E-53
LusMYB154	Lus10028513	AT1G66370.1	*ATMYB113*	1.55E-63
LusMYB155	Lus10028514	AT1G66370.1	*ATMYB113*	2.34E-53
LusMYB156	Lus10009129	AT1G66370.1	*ATMYB113*	3.56E-51
LusMYB157	Lus10022256	AT3G27785.1	*ATMYB118*	5.54E-46
LusMYB158	Lus10013084	AT3G27785.1	*ATMYB118*	1.60E-52
LusMYB159	Lus10005079	AT3G30210.1	*ATMYB121*	2.69E-62
LusMYB160	Lus10034372	AT3G30210.1	*ATMYB121*	1.68E-56
LusMYB161	Lus10009780	AT3G60460.1	*ATMYB125/DUO1*	1.68E-66
LusMYB162	Lus10037898	AT4G32730.1	*ATMYB3R1*	2.95E-76
LusMYB163	Lus10024036	AT5G41020.1	*ATMYB3R- like*	3.17E-106
LusMYB164	Lus10002384	AT4G18770.1	*AtMYB98*	5.12E-74
LusMYB165	Lus10003001	AT5G16770.1	*AtMYB9*	7.21E-65
LusMYB166	Lus10005864	AT5G58850.1	*AtMYB119*	8.71E-56
LusMYB167	Lus10012847	AT2G02820.1	*AtMYB88*	6.00E-129
LusMYB168	Lus10018220	AT5G58850.1	*AtMYB119*	1.62E-76
LusMYB169	Lus10018545	AT1G34670.1	*AtMYB93*	4.38E-49
LusMYB170	Lus10018936	AT1G17950.1	*AtMYB52*	2.15E-62
LusMYB171	Lus10024392	AT2G37630.1	*AtMYB91*	7.00E-144
LusMYB172	Lus10025355	AT2G37630.1	*AtMYB91*	6.90E-125
LusMYB173	Lus10027695	AT4G18770.1	*AtMYB98*	2.02E-86
LusMYB174	Lus10028638	AT1G17950.1	*AtMYB52*	1.13E-62
LusMYB175	Lus10034133	AT5G11510.1	*AtMYB3R4*	6.07E-17
LusMYB176	Lus10039966	AT4G18770.1	*AtMYB98*	3.20E-79
LusMYB177	Lus10040684	AT5G58850.1	*AtMYB119*	8.11E-73
LusMYB178	Lus10042111	AT4G18770.1	*AtMYB98*	2.31E-78
LusMYB179	Lus10043451	AT5G02320.1	*AtMYB3R5*	1.48E-17
R1R2R3-MYB				
LusMYB180	Lus10038623	AT4G32730.1	*ATMYB3R1*	9.86E-108

LusMYB181	Lus10022136	AT4G32730.1	*ATMYB3R1*	0
LusMYB182	Lus10025351	AT3G09370.1	*ATMYB3R3*	5.58E-138
LusMYB183	Lus10024394	AT5G02320.2	*ATMYB3R5*	8.01E-132
LusMYB184	Lus10008010	AT5G02320.2	*ATMYB3R5*	2.97E-71
LusMYB185	Lus10009008	AT3G18100.2	*ATMYB4R1*	5.60E-136
LusMYB186	Lus10009636	AT3G18100.2	*ATMYB4R1*	7.90E-151
4R-MYB				
LusMYB187	Lus10011687	AT4G32730.1	*ATMYB3R1*	9.51E-125

Appendix 5. Overview of putative LusMYBs.

Genes	Genomic contig	MW(kDa)	PI	aa length
LusMYB1	scaffold475	31.38	5.62	278
LusMYB2	scaffold1630	32	8.11	281
LusMYB3	scaffold488	33.37	7.68	299
LusMYB4	scaffold413	24.86	6.78	218
LusMYB5	scaffold34	30.35	6.41	270
LusMYB6	scaffold1168	30.53	6.89	270
LusMYB7	scaffold1615	30.92	8.56	278
LusMYB8	scaffold235	30.9	8.84	279
LusMYB9	scaffold232	31.14	8.05	282
LusMYB10	scaffold981	29.96	8.21	273
LusMYB11	scaffold3042	32.67	7.38	293
LusMYB12	scaffold272	25.09	6.52	222
LusMYB13	scaffold1247	30.26	9.9	271
LusMYB14	scaffold701	25.94	9.11	233
LusMYB15	scaffold86	24.11	9.56	212
LusMYB16	scaffold210	38.38	6.37	344
LusMYB17	scaffold701	41.51	5.07	360
LusMYB18	scaffold1035	41.62	6.75	368
LusMYB19	scaffold57	40.01	4.91	363
LusMYB20	scaffold989	35.68	6.99	321
LusMYB21	scaffold133	35.75	7.08	322
LusMYB22	scaffold732	39.7	4.7	361
LusMYB23	scaffold222	32.56	4.61	297
LusMYB24	scaffold67	23	9.78	199
LusMYB25	scaffold669	33.16	4.81	302
LusMYB26	scaffold1308	29.61	6.66	269
LusMYB27	scaffold87	22.59	9.99	198
LusMYB28	scaffold610	28.54	4.85	255
LusMYB29	scaffold280	26.45	6.23	236
LusMYB30	scaffold11	28.42	4.98	254
LusMYB31	scaffold164	25.07	6.79	223
LusMYB32	scaffold617	31.27	9.86	288
LusMYB33	scaffold51	42.51	5.95	384
LusMYB34	scaffold33	31.85	6.51	290
LusMYB35	scaffold96	31.75	5.73	288
LusMYB36	scaffold217	33.4	6.71	303

LusMYB37	scaffold184	32.1	6.11	289
LusMYB38	scaffold1036	31.49	5.03	279
LusMYB39	scaffold33	33.3	6.9	295
LusMYB40	scaffold96	32.99	6.64	293
LusMYB41	scaffold1036	29.76	6.09	266
LusMYB42	scaffold808	38.19	4.45	342
LusMYB43	scaffold808	39.36	4.37	355
LusMYB44	scaffold34	27.18	6.23	237
LusMYB45	scaffold472	27.31	6.99	237
LusMYB46	scaffold2280	40.47	4.43	364
LusMYB47	scaffold225	27.5	8.5	242
LusMYB48	scaffold242	53.92	9.94	477
LusMYB49	scaffold42	23.76	8.14	209
LusMYB50	scaffold107	24.08	7.47	211
LusMYB51	scaffold630	41.86	6.37	374
LusMYB52	scaffold177	26.66	5.07	228
LusMYB53	scaffold176	26.7	5.38	230
LusMYB54	scaffold773	33.84	9.88	310
LusMYB55	scaffold373	33.66	9.48	309
LusMYB56	scaffold635	35.83	8.57	323
LusMYB57	scaffold33	36.58	6.62	326
LusMYB58	scaffold57	41.15	5.03	371
LusMYB59	scaffold256	36.59	6.77	328
LusMYB60	scaffold306	40.78	5.04	366
LusMYB61	scaffold612	36.27	8.15	323
LusMYB62	scaffold618	38.32	8.26	340
LusMYB63	scaffold1004	35.39	6.79	311
LusMYB64	scaffold1851	32.96	8.67	293
LusMYB65	scaffold3345	35.71	6.79	315
LusMYB66	scaffold344	36.69	8.37	328
LusMYB67	scaffold355	32.2	8.56	286
LusMYB68	scaffold107	31.83	6.13	277
LusMYB69	scaffold42	32.33	6.67	281
LusMYB70	scaffold863	29.57	5.04	261
LusMYB71	scaffold701	29.6	4.83	262
LusMYB72	scaffold291	33.82	5.03	303
LusMYB73	scaffold349	34.28	4.75	308
LusMYB74	scaffold286	22.27	8.28	195

LusMYB75	scaffold468	33.15	7.88	305
LusMYB76	scaffold15	43.56	6.37	385
LusMYB77	scaffold783	38.25	6.46	342
LusMYB78	scaffold134	41.47	7	365
LusMYB79	scaffold472	40.87	7.74	357
LusMYB80	scaffold55	43.71	6.13	385
LusMYB81	scaffold1158	28.26	10.04	246
LusMYB82	scaffold418	35.65	8.27	310
LusMYB83	scaffold977	39.63	6.73	343
LusMYB84	scaffold783	40.9	6.51	353
LusMYB85	scaffold15	23.19	5.77	208
LusMYB86	scaffold15	27.48	4.61	240
LusMYB87	scaffold475	45.6	7.28	432
LusMYB88	scaffold911	35.49	9.83	318
LusMYB89	scaffold641	34.67	5.22	309
LusMYB90	scaffold1331	34.33	5.42	306
LusMYB91	scaffold883	27.53	6.36	246
LusMYB92	scaffold310	38.47	4.79	330
LusMYB93	scaffold272	39.5	5.25	338
LusMYB94	scaffold361	53.49	5.22	488
LusMYB95	scaffold319	133.73	5.88	1217
LusMYB96	scaffold1635	75.21	6.88	682
LusMYB97	scaffold76	52.43	5.22	478
LusMYB98	scaffold28	33.22	6.58	294
LusMYB99	scaffold1649	33.77	7.06	299
LusMYB100	scaffold617	37.91	8.17	338
LusMYB101	scaffold82	29.16	8.38	268
LusMYB102	scaffold74	29.37	7.95	270
LusMYB103	scaffold218	34.57	9.05	305
LusMYB104	scaffold505	34.86	9.42	304
LusMYB105	scaffold217	37.97	9.09	356
LusMYB106	scaffold280	24.86	7.13	225
LusMYB107	scaffold1519	26.37	5.2	242
LusMYB108	scaffold621	33.28	7.37	307
LusMYB109	scaffold540	26.71	6.6	244
LusMYB110	scaffold1247	26.4	6.8	235
LusMYB111	scaffold161	36.58	9.34	342
LusMYB112	scaffold30	42.06	5.14	380

LusMYB113	scaffold179	30.97	7.19	280
LusMYB114	scaffold420	31.43	6.79	282
LusMYB115	scaffold977	37.24	6.51	332
LusMYB116	scaffold327	39.39	7.24	355
LusMYB117	scaffold15	40.42	6.6	358
LusMYB118	scaffold262	40.42	6.56	362
LusMYB119	scaffold338	41.78	7.28	376
LusMYB120	scaffold917	51.51	5.29	465
LusMYB121	scaffold280	34.6	7.64	313
LusMYB122	scaffold1308	25.6	6.25	225
LusMYB123	scaffold57	34.74	7	314
LusMYB124	scaffold15	25.42	5.95	222
LusMYB125	scaffold123	39.8	7.29	374
LusMYB126	scaffold165	32.72	8.04	292
LusMYB127	scaffold752	36.47	7.31	330
LusMYB128	scaffold1686	43.04	8.84	399
LusMYB129	scaffold472	60.88	6.06	561
LusMYB130	scaffold12	52.08	9.77	460
LusMYB131	scaffold151	61.71	5.43	567
LusMYB132	scaffold251	37.1	6.1	333
LusMYB133	scaffold1308	29.46	4.86	255
LusMYB134	scaffold23	36.99	6.79	328
LusMYB135	scaffold15	21.22	10.41	191
LusMYB136	scaffold204	48.21	5.6	436
LusMYB137	scaffold1999	36.48	6.42	327
LusMYB138	scaffold15	29.18	4.7	252
LusMYB139	scaffold291	57.22	7.38	513
LusMYB140	scaffold349	57.86	7.5	518
LusMYB141	scaffold917	41.46	8.41	375
LusMYB142	scaffold617	42.96	8.58	388
LusMYB143	scaffold430	36.14	8.54	318
LusMYB144	scaffold30	35.83	8.96	314
LusMYB145	scaffold196	37.1	6.52	337
LusMYB146	scaffold216	28.68	10.42	261
LusMYB147	scaffold327	43.51	6.32	407
LusMYB148	scaffold86	43.26	6.19	406
LusMYB149	scaffold57	33.5	6.11	302
LusMYB150	scaffold153	33.89	5.82	312

LusMYB151	scaffold885	27.56	5.65	247
LusMYB152	scaffold1536	34.66	9.6	304
LusMYB153	scaffold67	27.56	4.61	248
LusMYB154	scaffold413	33.97	9.76	297
LusMYB155	scaffold413	31.04	9.21	272
LusMYB156	scaffold1536	32.61	8.8	288
LusMYB157	scaffold225	43.44	9.78	388
LusMYB158	scaffold242	38.15	10.76	341
LusMYB159	scaffold1311	26.17	8.48	228
LusMYB160	scaffold310	25.57	9.19	224
LusMYB161	scaffold271	32.53	5.32	288
LusMYB162	scaffold475	92.73	6.08	839
LusMYB163	scaffold353	64.01	9.69	552
LusMYB164	scaffold2788	47.52	6.62	415
LusMYB165	scaffold599	81.11	6.38	732
LusMYB166	scaffold1158	43.66	8.84	403
LusMYB167	scaffold1313	52.78	6.21	474
LusMYB168	scaffold163	42.45	10	388
LusMYB169	scaffold1308	30.63	4.83	264
LusMYB170	scaffold103	32.41	9.46	287
LusMYB171	scaffold16	41.76	9.67	365
LusMYB172	scaffold339	38.42	9.29	336
LusMYB173	scaffold2	62.04	6.18	540
LusMYB174	scaffold346	32.75	9.22	290
LusMYB175	scaffold292	106.73	5.17	949
LusMYB176	scaffold12	46.4	5.91	408
LusMYB177	scaffold156	37.7	9.98	341
LusMYB178	scaffold123	46.99	7.45	411
LusMYB179	scaffold25	29.94	9.94	257
LusMYB180	scaffold37	92.91	5.47	839
LusMYB181	scaffold371	113.28	5.37	1020
LusMYB182	scaffold339	56.2	9.51	508
LusMYB183	scaffold16	56.23	9.31	508
LusMYB184	scaffold517	48.75	9.64	436
LusMYB185	scaffold883	99.94	8.35	886
LusMYB186	scaffold169	107.63	8.9	957
LusMYB187	scaffold476	149.23	5.12	1350

Appendix 6. Transcript levels of *LusMYBs* across tissues checked by RNA-seq (Kumar et al., 2013). ge: globular embryo; he: heart embryo; te: torpedo embryo; ce: cotyledon embryo; me: mature embryo; sd: seeds; an: anthers; ov: ovaries; fl: mature flower; rt: root; st: stem; es: etiolated seedlings; le: leaves; max: the highest expression level among these tissues;

	he	te	ce	me	sd	an	ov	fl	rt	st	es	le	max in
LusMYB1	0	0	0	0	0	0	0	2.41074	0	0.01142	0	0	fl
LusMYB2	0	0	0	0	0	0	0	1.44203	0	0.023147	0	0	fl
LusMYB3	0	0.018662	0	0	0.621366	0.003921	0	0	0	0.007606	0	0	sd
LusMYB4	2.22626	0.067196	0.169033	0.674547	18.6601	0.834379	6.96194	8.95193	9.26278	14.6285	8.34412	1.66295	sd
LusMYB5	0.495259	0.126793	0.041268	0.171063	4.43177	4.24313	1.59453	0	1.25421	0.016078	0.01224	0	sd
LusMYB6	0.105259	0.002932	0.022073	0.022973	0.676343	0.145155	0.344033	0	1.42058	0	0	0	rt
LusMYB7	1.21512	0.291642	0.283238	0.166434	1.07165	1.60099	0.862968	1.38344	11.6985	0.324953	1.04902	0	rt
LusMYB8	1.07346	0.98729	0.545102	0.052717	4.74655	8.77769	7.2871	5.64568	11.8194	1.85448	2.86014	0.227183	rt
LusMYB9	0.245089	0.724939	0.740551	0.317024	12.5404	75.0004	26.682	9.98458	26.4896	8.98845	22.4482	3.95499	an
LusMYB10	1.27645	1.4874	2.33788	0.252839	4.81436	7.15208	6.49989	7.99489	86.1707	7.49991	9.14886	1.61485	rt
LusMYB11	0	0	0	0	0.045173	0	0	0	0	0	0	0	sd
LusMYB12	1.29018	0.027432	0.040476	0.030476	2.01923	0.237865	1.31128	1.36069	1.91132	3.17537	2.96217	0.5772	st
LusMYB13	0.192283	0.06364	0.064164	0.035704	2.7961	11.0386	4.70048	2.1036	11.309	0.953848	0.81883	0	rt
LusMYB14	0	0	0	0	0.458608	0	0.198353	0.386839	0.249663	0	0.018973	0.04266	sd
LusMYB15	0	0	0	0	0.050909	0.067374	0	1.53045	0.031652	0.005076	0.234529	0	fl
LusMYB16	0	0	0	0	0.081563	0.087499	0	1.38183	0.008341	0.009064	0.43537	0	fl
LusMYB17	0	0	0	0	0.107418	0	0.116873	0	0.491872	0.028391	0	0	rt
LusMYB18	0	0	0	0	0	0	0	1.19866	0.011721	0	0.0271	0	fl
LusMYB19	0.023558	0.015281	0	0.024798	0.414475	0.035506	1.08555	0.421167	0.416219	0	0	0	ov
LusMYB20	0.580118	7.79954	8.02429	0.457507	1.63091	0.125874	1.04561	9.19479	1.65937	1.33682	1.90755	0.101622	fl
LusMYB21	2.03901	2.29057	2.11529	0.827399	1.91871	0.076803	1.92566	13.9762	1.64507	2.21253	3.36842	2.47807	fl
LusMYB22	0.023828	0	0	0	0.79294	0.050667	1.85646	0.333635	0.981296	0.051526	0.01401	0	ov
LusMYB23	0.272666	1.52734	0.275866	0.917543	3.8699	0.076075	11.7132	15.5503	3.70369	3.90992	0.916431	0	fl
LusMYB24	0	0	0	0	0	0.076596	0	0	0	0	0.007783	0	an
LusMYB25	5.9264	26.9241	3.66159	36.5325	17.6374	0.354953	14.6829	35.623	7.20934	15.2822	3.12399	0.052373	me

LusMYB26	0	0.02775	0	0.038359	1.67267	0	2.86763	0.142288	7.06412	0	0.028217	0	rt
LusMYB27	0	0	0	0	0	0.423545	0.013083	0.053997	0	0	0	0.037069	an
LusMYB28	0	0	0	0	0.301355	0.124141	0.125777	45.8761	0.466796	5.15045	1.27116	0.07385	fl
LusMYB29	0	0	0	0	0.074226	0	0	11.7506	0.070439	0.867227	0.644349	0.064415	fl
LusMYB30	0.202582	0.152833	0.993463	0.070392	0.47802	0	0.663655	55.4693	1.32996	2.36406	0.91092	0	fl
LusMYB31	0.132383	0.12694	0.611714	0.423691	0.010813	0	0.005808	0	0	0	0	0	ce
LusMYB32	0.041923	0	0	0.226545	0.042618	0	0.34657	0	0.2734	0.014564	0	0	ov
LusMYB33	0.136939	0.105379	0.013888	0.297455	12.9413	2.47735	9.80889	1.22805	9.54771	6.53098	12.8306	5.90486	sd
LusMYB34	0.547429	0.418243	0.049039	0	1.7493	0.023599	0.068306	0.006294	0.007862	0	0	0	sd
LusMYB35	0.862199	1.18642	0.389004	0	4.27613	1.16457	2.7615	0.172665	4.72822	0.058464	0.588728	0	rt
LusMYB36	3.93386	1.14237	3.79657	0.58515	5.60636	0.273947	7.73472	0.372649	3.48033	0.734717	0.507476	0.041219	ov
LusMYB37	13.6052	4.15848	1.89879	0.862354	6.64522	0.312662	15.5996	0.684335	14.3167	0.969475	0.607744	0.08524	ov
LusMYB38	0	0	0	0	0	0	0.008099	0.085719	0	0	0.024628	0	fl
LusMYB39	0	0	0	0	0	0.000628	0	0.150939	0	0	0.193022	0	es
LusMYB41	0	0	0	0	0	0.013232	0	0.053639	0	0	0	0	fl
LusMYB42	0	0	0	0	0	0	0	0	0.075736	0.006413	0	0.159753	le
LusMYB43	0.021139	0.587305	1.84691	0.972522	2.20118	0.073673	0.493034	1.70071	1.50803	3.28907	4.02334	0	es
LusMYB44	0.137248	0.066536	0	0.055904	18.6688	2.93471	0.732589	4.99491	3.65993	8.80047	1.36768	0	sd
LusMYB45	0.100138	0	0	0.055904	20.7727	3.01416	1.14512	19.2644	4.93953	22.5684	6.74533	0.495651	st
LusMYB46	0.049783	0.051603	2.71222	0.142012	0.143058	0.085847	0.083501	1.56585	0.116809	0	0	0	ce
LusMYB47	0	0	0.212556	0.049668	2.168819	93.16685	24.20028	0.081938	67.38815	0	0	0	an
LusMYB48	0.020384	0	0.043186	0.021866	0.414526	49.75425	12.74716	0.121249	6.82234	0	0	0	an
LusMYB49	0.006843	0	0	0.143274	1.38402	49.3696	21.3421	0.079561	1.19065	0.037355	0	0	an
LusMYB50	0	0	0	0	1.33806	55.8199	33.6539	0.098073	114.894	0.024555	0	0	rt
LusMYB51	0	0	0	0	7.0291	0.008557	5.41546	0.004557	2.5764	0	0.00343	0	sd
LusMYB52	0	0	0	0.029414	0	0.080596	0	0	0	0	0.410472	0	es
LusMYB53	0	0	0	0	0.017081	0	0	0	0	0	0.758295	0	es
LusMYB54	23.9739	3.40894	0.956341	0.559642	1.69972	0.377989	2.30427	0	2.38196	0.218597	1.53957	0.133728	ge
LusMYB55	3.15631	2.530865	2.237343	1.414829	0.557592	0.243258	0.616326	0.35171	0.3771	0	0	0	ge
LusMYB56	5.82187	2.12678	0.802613	0.527438	5.06912	4.05021	4.0293	0.026989	2.59365	0.412589	3.49463	0.030338	ge

LusMYB57	0	0	0	0.069113	0.085516	0	0.02404	0	0	0	0.01451	0.036881	sd
LusMYB58	0	0	0	0	0.008779	0.802531	0.008168	0.004603	0.053818	0	0.048113	0	an
LusMYB59	0	0	0	0	0.010151	0	0	0	0	0	0.021825	0.005057	es
LusMYB60	0	0	0	0	0.058277	15.52875	0.970651	0.00247	1.334142	0.001462	0.01439	0	an
LusMYB61	0	0.016971	0	0	0.518574	0.20541	0.015121	4.43619	0.088169	0.147608	2.26978	0	fl
LusMYB62	0	0	0	0	0.551745	0.135315	0.012588	14.0285	0.252134	0.414195	2.29901	0.00185	fl
LusMYB63	0	0.02898	0	0	0	0	0	2.33059	0.072363	0	3.35551	0.092819	es
LusMYB64	0.015274	0	0	0	0	0	0.019335	5.35018	0.032099	0.288936	1.08086	0.024851	fl
LusMYB65	0.079577	0.121777	0.041011	0.019538	0.005679	0	0	2.89724	0	0	2.4987	0	fl
LusMYB66	0.013301	0.10518	0.025695	0	0.059423	1.09518	0	5.73549	0	0.058585	0.931244	0	fl
LusMYB67	0	0	0	0	0.060527	0	0.124204	9.21493	0	0.137075	2.87388	0	fl
LusMYB68	2.65119	0	0	0	0.109646	7.72381	0	0.816768	0.420113	0.04395	0.241434	0	an
LusMYB69	0	0	0	0	1.29509	0.089511	0	0.281847	0.026099	0.051483	1.16466	0	sd
LusMYB70	0	0	0.100367	0	0.10351	0.026659	0.035394	0.615735	0.513584	0.523366	0	0	fl
LusMYB71	0	0	0	0	3.08786	0.522993	2.64668	1.61736	1.85842	0.619522	0.295606	1.15837	sd
LusMYB72	0	0	0	0.081947	3.70927	0.104883	0.481854	7.50925	2.64758	12.163	2.52728	0.031366	st
LusMYB73	0	0	0	0.272886	0.732207	0.010913	0.071274	12.1005	0.690668	22.6872	6.16717	0.18152	st
LusMYB74	0	0	0	0.335087	0	0.080674	0.053554	0	0	0	0	0	me
LusMYB75	3.40742	22.0682	34.8135	27.534	22.5353	7.64009	4.17877	57.6311	3.35667	19.9896	12.1835	1.20902	fl
LusMYB76	0	0	0	0.027536	0.04145	0.259993	0.008108	0.875145	0.069724	1.7865	0.5196	0	st
LusMYB77	0	0	0	0.017694	2.26171	0.009557	1.51306	2.19285	1.58091	2.5507	0.360933	0.056463	st
LusMYB78	0	0	0	0.147399	1.21548	0	0.375013	3.73937	1.28222	12.1476	0.496848	0.771761	st
LusMYB79	0	0	0	0.008406	0.059642	1.53397	0.017777	0.212762	0.031731	1.123692	0	0	an
LusMYB80	0	0	0	0	0.066762	0	0.016215	1.21027	0.065319	0.973695	0.609985	0.08097	fl
LusMYB81	0.057082	0.868712	1.0118	2.94589	1.97089	0.867016	0.254283	41.8016	5.85005	21.8703	17.1196	210.037	le
LusMYB82	0	0	0	0.019922	0.123853	0	0.014968	0.917588	0.05294	0.864234	0.20955	0	fl
LusMYB83	0	0	0	0	1.19278	1.97645	0.100098	4.63112	0.906964	4.31128	0.396544	0.028793	fl
LusMYB84	0	0	0	0.068157	1.71887	0.077056	0.370537	5.06402	0.708265	4.86469	0.486285	0.0319	fl
LusMYB85	0	0	0	0	0	0	0	0	0.305387	0.201345	0.123536	0.546386	le
LusMYB86	0	0	0	0	0	0	0.013129	0.008139	0	0	0	0	ge

LusMYB87	0.175617	0.097623	0	0.063423	6.19036	0.419425	0.929074	4.02954	1.26301	4.80688	5.44711	0	sd
LusMYB88	0.36889	0.591794	0.155597	0	1.53668	3.45548	0.728441	0.761162	0.503829	0.50385	0.03224	0	an
LusMYB89	0	0.03597	0	0	1.15939	0.465549	1.01811	0.056744	0.66621	1.98247	2.08104	4.18132	le
LusMYB91	0	0	0	0.026629	0.114276	0.014721	0	0.129849	0.398959	0.514524	0	0	st
LusMYB92	0	0	0	0	0.457773	0	1.53496	0.005327	1.64987	0.18899	0.018852	0.01572	rt
LusMYB93	0	0	0	0	0.052335	0.019399	0.259044	0.005168	0.073297	0	0	0	ov
LusMYB94	4.25337	11.2579	8.68264	15.5152	2.43855	2.55848	1.49792	0.803982	0.797773	1.25156	1.02141	0.516703	me
LusMYB95	17.0335	15.8909	6.85256	23.0751	9.79048	33.6243	9.03203	8.92854	4.52994	7.39286	6.64022	2.70078	an
LusMYB96	2.4935	3.68114	4.87354	5.95074	2.00844	23.8036	1.99464	2.17631	2.59001	1.25613	2.61836	0.407037	an
LusMYB97	3.21555	6.84278	2.29035	7.61969	1.31135	2.58277	0.785785	0.665374	0.858367	0.959581	0.71217	0.384668	me
LusMYB98	0	0	0	0	0.009992	0.154982	0	0.445549	0	0.027124	0.023042	0	fl
LusMYB99	0	0	0	0	0	0	0	0.425934	0	0	0.085442	0	fl
LusMYB100	0	0	0	0	0.259218	0.101996	0.050719	18.6231	0.196938	0.653636	3.07014	0.10124	fl
LusMYB101	0.529882	0	1.36298	0.559097	0.388852	0.056724	0.417747	19.3861	0.530364	2.46704	1.02926	0	fl
LusMYB102	1.02431	0.798369	1.39684	0	0.391754	2.16458	0.630469	15.6765	1.18245	1.56928	2.60396	0	fl
LusMYB103	0.009577	0.074261	0	0	0.029526	0	0.008233	1.76464	0.049623	0	0.771	0	fl
LusMYB104	0	0	0	0	0.321356	0.401348	0	1.03558	0	0.016773	0.623133	0	fl
LusMYB105	0.175817	1.48566	18.4761	5.0734	6.31334	2.66281	4.35254	10.4551	1.96534	5.29386	6.9544	0.223953	ce
LusMYB106	1.01489	0.080817	0.044511	0	1.25262	3.67673	0.142088	0.008924	0.12271	0.049481	0	0	an
LusMYB107	0.248923	0.023607	0.555637	3.36006	0.056224	0.670595	0.036775	0.091387	0.241816	0	0	0	me
LusMYB108	4.61144	7.93402	20.0121	26.4282	33.322	11.5511	12.7533	41.1891	7.14378	18.9057	12.2839	1.43594	fl
LusMYB109	0.075614	0	0	0	0	2.30833	0	0	0	0	0	0	an
LusMYB110	11.147	4.56796	27.5	9.2883	5.93721	6.66837	4.56726	9.52295	12.5667	1.33903	5.29692	6.10664	ce
LusMYB111	15.5618	9.5346	68.3623	14.4331	13.874	27.4509	25.3607	42.0459	33.0815	21.7361	22.6035	180.315	le
LusMYB112	0.013485	0.968774	0	0.062479	7.38514	0.980941	28.2473	8.29957	9.14512	12.6838	13.2838	0.49358	ov
LusMYB113	0	0	0	0.022589	0.012001	0	0	0.07911	0.016591	0.744385	0.004903	0	st
LusMYB114	0	0	0	0	0	0.056106	0	0.760094	0	1.46091	0.042011	0	st
LusMYB115	0	0	0.027693	0.03667	3.07158	0.00992	2.48936	3.43566	4.13825	3.71002	0.374276	0.267369	rt
LusMYB116	0.026286	1.34905	1.49332	1.42684	0.887231	17.7834	0.052835	1.48613	0.201617	0.010208	1.84861	0	an
LusMYB117	0	0	0	0	0.054725	0	0.005905	0.657903	0.020086	0.606793	0.50911	0.65544	fl

LusMYB118	0.03634	0.089311	0.054986	0	3.5111	0	0.034403	6.32522	0.439258	1.01299	0.43492	0	fl
LusMYB119	0.026455	3.49479	6.12335	7.53602	0.810381	0.105335	0.044126	3.43757	0.098665	0.008295	0	0	me
LusMYB120	0.239419	0.232453	0.094453	0.235962	0.813089	0.53752	1.035726	2.301231	0	0	0	0	fl
LusMYB121	0	0	0	0.01969	0.024214	0	0	8.87554	0	0.012605	1.0352	0	fl
LusMYB122	0	0	0	0	0	0	0	0.04224	0	0.022964	0	0.054437	le
LusMYB123	0	0	0	0	0.017105	0	0	5.01696	0	0	0.757701	0	fl
LusMYB124	0	0	0	0	0	0	0	0.072799	0.114241	0.230062	0.026683	46.521	le
LusMYB125	0.604096	0.034249	0.557311	0.109332	2.45124	0.008557	0.264881	0.457296	0.60456	0	0	0	sd
LusMYB126	3.54378	0.816857	1.08579	0.414732	1.49272	0.214622	1.39678	0.336029	4.12833	2.41029	3.74936	0.229756	rt
LusMYB127	3.266	0.921778	0.979525	0.23105	1.68277	0.47669	4.24228	0.07458	3.6834	4.22405	5.56255	0.020087	es
LusMYB128	1.90532	0.10174	0.808938	0.228682	2.10549	0.045114	0.167529	0.132421	0.086987	1.1617	1.20646	0	sd
LusMYB129	0.007151	0.009013	0	0	0.245246	71.7218	0.082796	0.002823	0.74374	0.070879	0	0	an
LusMYB130	0.118517	0	0	1.16419	0.089366	201.612	0.211094	0.039755	8.2335	0.037548	0	0	an
LusMYB131	0.053458	0.024938	0.024452	0	0.230091	142.047	0.191655	0.065347	10.0396	0.07064	0	0	an
LusMYB132	0	0	0	0	0.093984	0.058227	0.085999	0.64464	0.097677	0.017074	1.60337	0	es
LusMYB133	0	0	0	0	0	0	0	0.140988	0.032103	0.054855	0	0.138127	fl
LusMYB134	0	0	0	0	0.026168	0	0.013334	2.8585	0.021393	0.006758	0.119737	0	fl
LusMYB135	0	0	0	0	0.134441	0	0	2.25144	0.032014	0	0.564878	0	fl
LusMYB136	0	0	0.017627	0	0.102086	0.033226	0.055927	7.90907	0.518405	0.012659	0.90118	0	fl
LusMYB137	0.217939	0	0.074219	0.109199	0.226305	0.023554	0	3.41323	0.430774	0.1466	0	0	fl
LusMYB138	0	0	0	0	0	0	0	0.654112	0.013542	0.078656	0.022489	0.32734	fl
LusMYB139	137.79	45.1163	21.45	40.5281	14.3681	3.51142	17.6795	16.064	7.10733	16.2851	8.7632	1.45026	ge
LusMYB140	211.688	65.1046	37.2628	78.4916	22.3641	16.3881	25.0017	26.021	8.62687	26.6786	19.7224	1.71661	ge
LusMYB141	3.725	0.148602	0	0	1.13792	0.264412	3.89048	0.009085	3.09315	0.927669	0.92614	0.48835	ov
LusMYB142	0.981743	0.060367	0.018061	0	1.47947	0.388892	4.58689	0.004357	0.861943	1.70782	0	0	ov
LusMYB143	0	0.077782	0.077401	0	0.016838	0	0	0	0	0	0.530318	0.035853	es
LusMYB144	0	0.04183	0.034406	0	0	0	1.13737	0	0.0922	0	0.238817	0	ov
LusMYB145	0.012873	0.03572	0.22493	0.028579	6.07362	71.587	1.13641	17.5746	3.55293	1.05042	0.462412	0.038831	an
LusMYB146	0.055534	0.009739	0.284704	0	3.20758	32.3712	0.336938	5.14511	1.11825	0.310333	0.14484	0	an
LusMYB147	5.19043	4.7375	2.94241	13.4423	4.96622	3.31805	3.9291	5.66515	2.35947	3.58253	2.62319	0.345902	me

LusMYB148	2.88137	3.11168	4.27733	15.1391	4.41811	2.02657	2.36956	4.08464	3.23353	2.48765	1.88789	0.300561	me
LusMYB149	1.23047	0.060803	0.034591	0.08228	2.97579	0.214637	3.77312	0.459623	1.40834	0.122721	0	0	ov
LusMYB150	3.63795	0.642219	6.51515	2.36206	0.034656	0.160874	0.169243	0.021455	0.564728	0.148085	0	0	ce
LusMYB151	0	0	0	0	0	0.254361	0.491447	0	2.48872	0	0.031954	0	rt
LusMYB152	0	0	0.027043	0	0	0.085262	0	0	0.070471	0.741362	0.332324	0	st
LusMYB153	0.155297	0	0	0	0.930384	1.14054	3.77175	0.211036	2.23507	0.032525	0.821912	0	ov
LusMYB154	0	0	0	0	0	0.045115	0.007467	0	0.203841	0.402183	0	0	st
LusMYB155	0.157767	0.175715	0.015598	0.219735	0.020445	0.216526	0.394834	0.013286	0	0	0	0	ov
LusMYB156	0	0	0	0	0.115708	186.731	0.889928	0.04119	22.5235	0.082002	0	0.20903	an
LusMYB157	0.425499	0.040851	0.040363	0	1.78669	0	0	0	0.02331	0.010984	0.00328	0	sd
LusMYB158	0.071018	0	0	0	0.301968	0	0	0	0	0.010138	0	0	sd
LusMYB159	0	0	0	0	0	0	0.077246	0	0.047147	0	0	0	ov
LusMYB160	0	0	0	0	0.277502	0	0.010925	0.053892	0.160914	0.146332	0.01317	0	sd
LusMYB161	0.128868	0.1125	0.223841	0.119854	0.188159	0.202653	0.253795	5.26566	0.22223	1.23969	0	0	fl
LusMYB162	1.83042	1.57507	0.447982	1.43934	1.04023	0.98592	1.60844	2.90963	0.783344	3.34322	2.36086	0.841053	st
LusMYB163	6.35453	6.35632	4.48773	6.91229	4.53472	4.30484	4.41204	3.91616	1.84848	3.48111	2.85926	1.09737	me
LusMYB164	0.010068	0.128946	0	0.06306	0	0	0.00494	0	0	0	0.011914	0	te
LusMYB165	0	0	0	0	0.022966	58.3431	0.071282	2.31174	0.622383	0.021184	0.110674	0	an
LusMYB166	0.490327	0.038094	0.019307	0	2.80692	0	0	0	0.010496	0	0	0	sd
LusMYB167	0.45752	0.472638	0.832935	0.386921	1.748104	1.101713	3.493432	2.811205	0	0	0	0	ov
LusMYB168	0	0	0.047844	0	0.322889	0	0	0	0.012048	0.005492	0	0.410196	le
LusMYB169	0	0	0	0	0	0	0	0.01427	0	0	0	0	fl
LusMYB170	0	0	0	0.065715	0.020212	0	0	1.50309	0.01606	6.60103	0	0.056653	st
LusMYB171	13.96228	16.02141	31.17891	21.16062	4.361705	0.192564	6.440065	3.541655	0	0	0	0	ce
LusMYB172	2.17722	3.470935	8.895155	27.4812	5.263395	1.492715	7.37373	0.673816	3.166445	0	0	0	me
LusMYB173	0	0	0	0	0	0	0	0	0.00767	0	0	0	rt
LusMYB174	0	0	0	0	1.10357	2.35782	0.007701	1.86918	0.252884	7.16291	0.180424	0	st
LusMYB175	18.2009	8.88424	3.95414	8.46159	9.48524	5.8853	8.31685	6.99778	3.34504	7.7807	6.11198	2.5455	ge
LusMYB176	0	0	0	0	0	0.100526	0	0	0.010341	0	0	0	an
LusMYB177	0.215174	0.067048	0	0	1.0327	0	0	0	0.014321	0	0.003835	0.063856	sd

LusMYB178	0.010181	0.816424	0.112431	0	0.015554	0.007633	0.049021	0.032509	0.12116	0	0	0	te
LusMYB180	0.317632	0.349179	0.08339	0.401712	1.80593	1.39425	1.89355	3.35328	0.761214	2.97946	3.31775	0.91069	fl
LusMYB181	4.97586	4.24849	0.408778	0.275606	1.45595	3.34897	4.49931	4.60617	1.15826	3.57137	2.31699	1.13275	he
LusMYB182	3.02679	2.23148	0.853716	2.40003	2.41711	4.92303	2.68787	4.06991	1.38146	3.5038	3.03913	0.877744	an
LusMYB183	2.4006	1.56005	3.87637	1.94693	3.26942	7.2419	3.58116	4.34108	6.28501	3.58657	2.76739	22.1844	le
LusMYB184	2.715315	0.723926	0.388194	0.088417	0.879695	0.234325	0.902759	1.44513	0.30358	0	0	0	ge
LusMYB185	1.97975	1.77032	0.86134	2.24176	2.12747	6.64099	2.60473	3.82764	3.43704	3.10997	2.23464	0.896744	an
LusMYB186	3.18745	2.19342	4.92714	4.08037	1.85738	3.42257	2.30877	3.54254	2.36177	2.39587	2.06836	0.57247	ce
LusMYB187	1.76273	1.901405	0.231251	0.116919	1.250151	2.990928	3.105382	3.9558	0.815685	0	0	0	fl

Appendix 7. Compositions of MYB genes in various plant species. ND: not determined;

	MYB-related	R2R3-MYB	3R-MYB	Atypical-MYB	Total	% of R2R3-MYB	Reference
Solanum lycopersicum	ND	122	4	1	ND	-	(Li et al. 2016)
Gossypium raimondii	ND	205	ND	ND	ND	-	(He et al., 2016)
Jatropha curcas	ND	123	4	1	ND	-	(Zhou et al., 2015)
Pyrus bretschneideri	22	105	2	0	129	81.40%	(Li et al., 2016)
Linum usitatissimum	53	179	7	1	240	74.58%	
Gossypium hirsutum	145	360	15	2	524	68.70%	(Salih et al., 2016)
Vitis vinifera	57	118	5	1	181	65.19%	(Wong et al., 2016)
Arabidopsis thaliana	64	126	5	1	198	55.02%	(Dubos et al., 2010)
Brassica rapa	191	256	11	9	467	54.82%	(Wang et al., 2015)
Eucalyptus grandis	151	189	7	3	350	54.00%	(Soler et al., 2015)
Oryza sativa	106	109	5	1	221	49.32%	(Jiang et al., 2004b)
Zea mays	169	157	0	0	326	48.16%	(Du et al., 2012)
Glycine max	265	244	6	2	517	47.20%	(Du et al., 2012)
Populus trichocarpa	213	192	5	0	410	46.83%	(Wilkins et al., 2008)
Solanum tuberosum	196	197	4	4	401	49.13%	(Wang et al., 2015)
Volvox carteri	9	15	3	2	29	51.72%	
Carica papaya	99	108	4	2	213	50.70%	
Cucumis sativus	166	147	10	3	326	45.09%	
Selaginella moellendorffii	54	47	3	1	105	44.76%	
Aquilegia coerulea	145	115	6	7	273	42.12%	
Physcomitrella patens	110	68	9	3	190	35.79%	
Ostreococcus lucimarinus	19	11	4	1	35	31.43%	
Micromonas pusilla	20	10	4	2	36	27.78%	
Coccomyxa subellipsoidea	15	7	5	1	28	25.00%	

Appendix 8. *Arabidopsis* orthologs of AR-enriched LusMYB genes. NA: not available.

Gene name	Arabidopsis ortholog	Function
LusMYB187	*AtMYB3R1*	cell cycle regulation; diverse roles in plant development: double mutant of myb3r1 myb3r4 causes pleiotropic developmental defects, such as dwarfism, irregular morphology of seedling and embryo, and production of polyploid offspring (Haga et al., 2011; Haga et al., 2007)
LusMYB181	*AtMYB3R1*	
LusMYB180	*ATMYB3R1*	
LusMYB162	*ATMYB3R1*	
LusMYB175	*AtMYB3R4*	
LusMYB179	*AtMYB3R5*	regulate cell cycle (Haga et al., 2007; Kobayashi et al., 2015)
LusMYB34	*AtMYB17*	is a target of the meristem identity regulator LEAFY (LFY) and plays a role in the meristem identity transition from vegetative growth to flowering (Zhang et al., 2009; Pastore et al., 2011)
LusMYB36	*AtMYB17*	
LusMYB35	*AtMYB17*	
LusMYB172	*AtMYB91*	specification of the leaf proximodistal axis, mediate stem cell function, and function as a regulator of plant immune response (Byrne et al., 2000; Hay, 2006; Sun et al., 2002; Nurmberg et al., 2007)
LusMYB171	*AtMYB91*	
LusMYB141	*AtMYB105*	boundary specification, meristem initiation and maintenance, and organ patterning (Lee et al., 2009)
LusMYB142	*AtMYB105*	
LusMYB61	*ATMYB036*	promote differentiation of the endodermis during root development and it also promotes the development the Casparian band (Liberman et al., 2015; Fernández-Marcos et al., 2017; Kamiya et al., 2015)
LusMYB66	*ATMYB036*	
LusMYB26	*ATMYB014*	NA
LusMYB149	*AtMYB111*	NA
LusMYB102	*ATMYB070*	NA

Appendix 9. List of putative LusNACs and their *Arabidopsis* orthologs.

Gene Symbol	Gene ID	Arabidopsis Ortholog	Arabidopsis Locus Description	E-value
LusNAC1	Lus10036749	AT1G26870.1	*ANAC009*	4E-98
LusNAC2	Lus10003668	AT5G14490.1	*ANAC085*	1E-84
LusNAC3	Lus10015554	AT5G17260.1	*ANAC086*	1.1E-27
LusNAC4	Lus10005917	AT5G13180.1	*ANAC083*	7E-24
LusNAC5	Lus10017458	AT5G04410.1	*ANAC078*	5E-121
LusNAC6	Lus10014342	AT3G10490.1	*ANAC052*	6.8E-08
LusNAC7	Lus10026496	AT1G61110.1	*ANAC025*	2.7E-77
LusNAC8	Lus10015389	AT1G61110.1	*ANAC025*	1.9E-69
LusNAC9	Lus10003458	AT4G27410.2	*ANAC072*	1.6E-12
LusNAC10	Lus10031142	AT1G12260.1	*ANAC007*	2E-102
LusNAC11	Lus10030175	AT4G35580.1	*ANTL9*	7.6E-38
LusNAC12	Lus10042531	AT3G04070.1	*ANAC047*	7.6E-94
LusNAC13	Lus10009924	AT3G17730.1	*ANAC057*	4.9E-12
LusNAC14	Lus10034700	AT3G18400.1	*ANAC058*	1.7E-16
LusNAC15	Lus10024908	AT4G28530.1	*ANAC074*	4.1E-61
LusNAC16	Lus10018142	AT1G01720.1	*ANAC002*	4E-140
LusNAC17	Lus10022018	AT1G69490.1	*ANAC029*	1.3E-08
LusNAC18	Lus10010098	AT4G01550.1	*ANAC069*	1E-22
LusNAC19	Lus10007377	AT1G65910.1	*ANAC028*	6E-131
LusNAC20	Lus10032238	AT1G12260.1	*ANAC007*	3E-136
LusNAC21	Lus10003848	AT5G08790.1	*ANAC081*	1.4E-43
LusNAC22	Lus10021992	AT3G04070.1	*ANAC047*	2.8E-86
LusNAC23	Lus10034999	AT4G29230.1	*ANAC075*	2E-125
LusNAC24	Lus10030446	AT1G69490.1	*ANAC029*	3E-101
LusNAC25	Lus10004846	AT5G22380.1	*ANAC090*	5.1E-92
LusNAC26	Lus10032657	AT3G15510.1	*ANAC056*	2E-103
LusNAC27	Lus10041492	AT1G25580.1	*ANAC008*	6E-166
LusNAC28	Lus10041822	AT2G18060.1	*ANAC037*	4E-114
LusNAC29	Lus10003269	AT3G04070.1	*ANAC047*	6.4E-98
LusNAC30	Lus10022965	AT2G24430.2	*ANAC038*	7E-10
LusNAC31	Lus10013967	AT4G28500.1	*ANAC073*	7E-106
LusNAC32	Lus10033251	AT4G35580.1	*ANTL9*	1E-130
LusNAC33	Lus10025118	AT2G17040.1	*ANAC036*	4.9E-80
LusNAC34	Lus10026617	AT1G69490.1	*ANAC029*	2E-99
LusNAC35	Lus10014911	AT5G08790.1	*ANAC081*	4.6E-31
LusNAC36	Lus10002687	AT2G46770.1	*ANAC043*	4E-121
LusNAC37	Lus10031951	AT2G18060.1	*ANAC037*	0.00444

LusNAC38	Lus10023208	AT2G02450.1	*ANAC034/35*	7E-107
LusNAC39	Lus10003367	AT2G02450.2	*ANAC034*	3E-99
LusNAC40	Lus10008419	AT2G02450.1	*ANAC034/35*	7E-106
LusNAC41	Lus10003334	AT1G61110.1	*ANAC025*	2.3E-06
LusNAC42	Lus10004338	AT1G62700.1	*ANAC026*	1.6E-94
LusNAC43	Lus10029692	AT1G01720.1	*ANAC002*	2.1E-88
LusNAC44	Lus10010096	AT4G17980.1	*ANAC071*	5.5E-26
LusNAC45	Lus10031189	AT2G43000.1	*ANAC042*	2.3E-81
LusNAC46	Lus10024601	AT1G12260.1	*ANAC007*	7E-139
LusNAC47	Lus10026879	AT5G18270.2	*ANAC087*	2E-106
LusNAC48	Lus10006119	AT4G35580.2	*ANAC018*	6.9E-25
LusNAC49	Lus10006054	AT1G34190.1	*ANAC017*	3.6E-90
LusNAC50	Lus10021708	AT1G25580.1	*ANAC008*	3E-159
LusNAC51	Lus10021659	AT5G61430.1	*ANAC100*	4E-113
LusNAC52	Lus10023179	AT1G61110.1	*ANAC025*	4.9E-79
LusNAC53	Lus10033281	AT2G24430.2	*ANAC038*	1.3E-07
LusNAC54	Lus10026200	AT1G56010.2	*ANAC021/22*	1E-69
LusNAC55	Lus10036773	AT1G69490.1	*ANAC029*	4E-100
LusNAC56	Lus10030723	AT1G76420.1	*ANAC031*	1.6E-70
LusNAC57	Lus10035648	AT5G64530.1	*ANAC104*	2.7E-67
LusNAC58	Lus10005537	AT5G53950.1	*ANAC098*	3E-103
LusNAC59	Lus10032724	AT3G17730.1	*ANAC057*	7.4E-12
LusNAC60	Lus10036955	AT3G04070.2	*ANAC002*	2.4E-05
LusNAC61	Lus10009939	AT1G79580.3	*ANAC033*	1.9E-94
LusNAC62	Lus10011215	AT1G61110.1	*ANAC025*	1.8E-95
LusNAC63	Lus10018469	AT1G61110.1	*ANAC025*	1.1E-98
LusNAC64	Lus10036959	AT3G04070.2	*ANAC002*	2.9E-11
LusNAC65	Lus10041924	AT5G53950.1	*ANAC098*	3E-93
LusNAC66	Lus10033239	AT1G32770.1	*ANAC012*	2.8E-92
LusNAC67	Lus10031767	AT2G43000.1	*ANAC042*	1.1E-73
LusNAC68	Lus10028824	AT5G04410.1	*ANAC078*	2E-116
LusNAC69	Lus10032919	AT5G24590.2	*ANAC091*	1.6E-94
LusNAC70	Lus10001648	AT5G61430.1	*ANAC100*	1E-114
LusNAC71	Lus10036117	AT1G69490.1	*ANAC029*	6.6E-77
LusNAC72	Lus10032653	AT3G10480.1	*ANAC050*	4.2E-27
LusNAC73	Lus10031937	AT5G08790.1	*ANAC081*	018
LusNAC74	Lus10008420	AT2G02450.2	*ANAC034*	2E-99
LusNAC75	Lus10035373	AT4G17980.1	*ANAC071*	1.3E-91
LusNAC76	Lus10024907	AT5G62380.1	*ANAC101*	3.7E-05

LusNAC77	Lus10038937	AT3G18400.1	*ANAC058*	1.1E-81
LusNAC78	Lus10007204	AT5G46590.1	*ANAC096*	0.0015
LusNAC79	Lus10015312	AT1G71930.1	*ANAC030*	4.1E-06
LusNAC80	Lus10020643	AT5G22380.1	*ANAC090*	3.7E-95
LusNAC81	Lus10030978	AT4G17980.1	*ANAC071*	1.4E-89
LusNAC82	Lus10037178	AT1G26870.1	*ANAC009*	6.2E-97
LusNAC83	Lus10010959	AT1G65910.1	*ANAC028*	2E-122
LusNAC84	Lus10033650	AT3G03200.1	*ANAC045*	7E-08
LusNAC85	Lus10003333	AT2G17040.1	*ANAC036*	1.6E-92
LusNAC86	Lus10018810	AT2G33480.2	*ANAC041*	4.2E-11
LusNAC87	Lus10018637	AT4G28500.1	*ANAC073*	2E-117
LusNAC88	Lus10033699	AT3G04060.1	*ANAC046*	4.5E-06
LusNAC89	Lus10039153	AT4G10350.1	*ANAC070*	1E-105
LusNAC90	Lus10033676	AT4G27410.2	*ANAC072*	2E-12
LusNAC91	Lus10026966	AT2G24430.2	*ANAC038*	6.6E-92
LusNAC92	Lus10037939	AT3G10480.1	*ANAC050*	8E-135
LusNAC93	Lus10013205	AT1G76420.1	*ANAC031*	3.5E-81
LusNAC94	Lus10010371	AT1G32510.1	*ANAC011*	1.8E-14
LusNAC95	Lus10023537	AT3G10480.1	*ANAC050*	2E-106
LusNAC96	Lus10015743	AT3G12910.1	*ANAC042*	4.8E-11
LusNAC97	Lus10003847	AT1G61110.1	*ANAC025*	3.9E-59
LusNAC98	Lus10010294	AT3G17730.1	*ANAC057*	1.5E-09
LusNAC99	Lus10022636	AT2G17040.1	*ANAC036*	4.4E-87
LusNAC100	Lus10035174	AT5G14000.1	*ANAC084*	6.9E-28
LusNAC101	Lus10004531	AT2G24430.2	*ANAC038*	7.6E-05
LusNAC102	Lus10017353	AT4G35580.1	*ANTL9*	1.3E-73
LusNAC103	Lus10005144	AT4G35580.3	*ANTL9*	8.3E-38
LusNAC104	Lus10010148	AT4G35580.1	*ANTL9*	1E-102
LusNAC105	Lus10013964	AT1G61110.1	*ANAC025*	1.1E-69
LusNAC106	Lus10015367	AT1G26870.1	*ANAC009*	5.5E-92
LusNAC107	Lus10019926	AT1G61110.1	*ANAC025*	2.3E-76
LusNAC108	Lus10022915	AT4G28530.1	*ANAC074*	1.2E-89
LusNAC109	Lus10008897	AT2G02450.1	*ANAC034/35*	3E-99
LusNAC110	Lus10023966	AT2G17040.1	*ANAC036*	1.6E-85
LusNAC111	Lus10025690	AT1G01720.1	*ANAC002*	8E-142
LusNAC112	Lus10025078	AT1G79580.3	*ANAC033*	1.3E-12
LusNAC113	Lus10030174	AT4G35580.3	*ANTL9*	3E-41
LusNAC114	Lus10007216	AT1G54330.1	*ANAC020*	7.2E-27
LusNAC115	Lus10013316	AT2G27300.1	*ANAC040*	2.7E-87

LusNAC116	Lus10033652	AT1G62700.1	*ANAC026*	3.1E-08
LusNAC117	Lus10022914	AT1G61110.1	*ANAC025*	8.3E-07
LusNAC118	Lus10041534	AT5G14000.1	*ANAC084*	4.2E-31
LusNAC119	Lus10037156	AT1G69490.1	*ANAC029*	2E-99
LusNAC120	Lus10032004	AT5G14000.1	*NAC084*	1.3E-29
LusNAC121	Lus10033905	AT4G28530.1	*ANAC074*	1.5E-87
LusNAC122	Lus10013782	AT4G10350.1	*ANAC070*	1E-104
LusNAC123	Lus10009029	AT3G18400.1	*ANAC058*	8.3E-94
LusNAC124	Lus10040422	AT3G10490.2	*ANAC051*	7.4E-75
LusNAC125	Lus10028372	AT2G18060.1	*ANAC037*	2E-113
LusNAC126	Lus10002581	AT5G13180.1	*ANAC083*	1.1E-90
LusNAC127	Lus10033279	AT2G24430.2	*ANAC038*	9.6E-10
LusNAC128	Lus10008285	AT4G35580.1	*ANTL9*	3E-135
LusNAC129	Lus10008200	AT3G03200.1	*ANAC045*	0.0026
LusNAC130	Lus10026588	AT4G35580.1	*ANTL9*	7.9E-26
LusNAC131	Lus10034183	AT2G27300.1	*ANAC040*	2.6E-82
LusNAC132	Lus10012927	AT4G29230.1	*ANAC075*	8E-112
LusNAC133	Lus10042466	AT1G56010.2	*ANAC021/22*	4E-83
LusNAC134	Lus10031639	AT3G55210.1	*ANAC063*	1.1E-05
LusNAC135	Lus10036194	AT2G43000.1	*ANAC042*	1.1E-70
LusNAC136	Lus10017915	AT1G71930.1	*ANAC030*	1.4E-93
LusNAC137	Lus10038332	AT2G43000.1	*ANAC042*	7.3E-74
LusNAC138	Lus10009858	AT5G62380.1	*ANAC101*	1.3E-07
LusNAC139	Lus10003435	AT5G18270.2	*ANAC087*	7E-101
LusNAC140	Lus10028713	AT1G34190.1	*ANAC017*	1E-110
LusNAC141	Lus10020883	AT5G08790.1	*ANAC081*	7E-108
LusNAC142	Lus10015392	AT4G28500.1	*ANAC073*	5E-104
LusNAC143	Lus10015587	AT5G24590.2	*ANAC091*	3E-96
LusNAC144	Lus10003366	AT2G02450.2	*ANAC034*	1.6E-51
LusNAC145	Lus10043095	AT3G15510.1	*ANAC056*	1E-107
LusNAC146	Lus10008271	AT1G32770.1	*ANAC012*	2.3E-81
LusNAC147	Lus10042284	AT3G17730.1	*ANAC057*	6E-119
LusNAC148	Lus10002083	AT5G08790.1	*ANAC081*	1.7E-67
LusNAC149	Lus10020896	AT5G09330.4	*ANAC082*	4.7E-82
LusNAC150	Lus10029410	AT5G22380.1	*ANAC090*	3E-37
LusNAC151	Lus10030205	AT2G46770.1	*ANAC043*	7E-121
LusNAC152	Lus10035647	AT5G64530.1	*ANAC104*	2.7E-67
LusNAC153	Lus10005204	AT5G22290.1	*ANAC089*	2.2E-90
LusNAC154	Lus10030478	AT5G39820.1	ANAC094	8.6E-97

LusNAC155	Lus10000206	AT5G64530.1	ANAC104	2.7E-67
LusNAC156	Lus10019638	AT1G12260.1	ANAC007	8.6E-10
LusNAC157	Lus10015076	AT1G61110.1	ANAC025	1E-78
LusNAC158	Lus10035400	AT4G36160.1	ANAC076	9.4E-11
LusNAC159	Lus10010747	AT5G64530.1	ANAC104	9.4E-65
LusNAC160	Lus10031721	AT1G12260.1	ANAC007	2E-102
LusNAC161	Lus10017340	AT2G46770.1	ANAC043	4E-105
LusNAC162	Lus10027357	AT5G08790.1	ANAC081	6.4E-49
LusNAC163	Lus10006547	AT3G04070.1	ANAC047	2E-100
LusNAC164	Lus10001664	AT1G32770.1	ANAC012	1E-103
LusNAC165	Lus10008240	AT1G33060.1	ANAC014	3.2E-05
LusNAC166	Lus10033493	AT5G08790.1	ANAC081	1E-106
LusNAC167	Lus10042518	AT3G44290.1	ANAC060	5.5E-26
LusNAC168	Lus10027227	AT3G18400.1	ANAC058	1.1E-85
LusNAC169	Lus10039873	AT4G28500.1	ANAC073	4E-119
LusNAC170	Lus10037106	AT3G04070.2	ANAC002	2.2E-12
LusNAC171	Lus10007263	AT1G26870.1	ANAC009	1.2E-93
LusNAC172	Lus10020165	AT2G24430.2	ANAC038	6.7E-91
LusNAC173	Lus10043402	AT2G27300.1	ANAC040	5.8E-84
LusNAC174	Lus10005143	AT1G33060.2	ANAC014	4.1E-39
LusNAC175	Lus10001809	AT5G13180.1	ANAC083	1.8E-90
LusNAC176	Lus10010037	AT1G65910.1	ANAC028	6.6E-24
LusNAC177	Lus10024006	AT1G79580.3	ANAC033	2.3E-14
LusNAC178	Lus10007410	AT1G69490.1	ANAC029	3.8E-77
LusNAC179	Lus10009669	AT3G18400.1	ANAC058	9.9E-93
LusNAC180	Lus10026373	AT3G17730.1	ANAC057	4E-117
LusNAC181	Lus10012557	AT5G14000.1	ANAC084	4.7E-31
LusNAC182	Lus10042731	AT1G01720.1	ANAC002	4E-145

Appendix 10. Overview of putative flax NAC domain proteins. Data source: Phytozome v.12 (https://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al., 2012).

Gene Name	Genomic contig	Pfam domain	MW(kDa)	aa length	PI
LusNAC1	scaffold31	PF02365	50.96	457	7.96
LusNAC2	scaffold734	PF02365	49.69	450	7.34
LusNAC3	scaffold860	PF02365	13.51	118	7.51
LusNAC4	scaffold26	PF02365	6.39	56	9.93
LusNAC5	scaffold1253	PF02365	58.69	531	4.21
LusNAC6	scaffold275	PF02365	22.08	187	10.36
LusNAC7	scaffold617	PF02365	40.55	354	5.56
LusNAC8	scaffold635	PF02365	37.76	337	5.73
LusNAC9	scaffold80	PF02365	24.33	214	5.66
LusNAC10	scaffold977	PF02365	38.41	328	5.96
LusNAC11	scaffold217	PF02365	68.87	617	4.58
LusNAC12	scaffold67	PF02365	43.29	382	9.21
LusNAC13	scaffold200	PF02365	20.82	178	7.04
LusNAC14	scaffold9	PF02365	23.55	207	8.64
LusNAC15	scaffold473	PF02365	30.77	269	6.23
LusNAC16	scaffold112	PF02365	34.13	304	6.93
LusNAC17	scaffold87	PF02365	12.93	116	10.14
LusNAC18	scaffold722	PF02365	18.5	162	9.99
LusNAC19	scaffold302	PF02365	74.69	677	6.28
LusNAC20	scaffold291	PF02365	46.86	402	6.98
LusNAC21	scaffold706	PF02365	11.66	101	6.52
LusNAC22	scaffold87	PF02365	44.71	396	8.85
LusNAC23	scaffold29	PF02365	58.25	524	6.62
LusNAC24	scaffold917	PF02365	31.6	275	9.04
LusNAC25	scaffold1821	PF02365	27.95	250	6.14
LusNAC26	scaffold140	PF02365	42.43	382	8.47
LusNAC27	scaffold272	PF02365	46.96	420	4.63
LusNAC28	scaffold272	PF02365	41.6	363	5.45
LusNAC29	scaffold885	PF02365	42.68	376	9.77
LusNAC30	scaffold355	PF02365	39.23	357	4.31
LusNAC31	scaffold820	PF02365	35.54	325	7.36
LusNAC32	scaffold488	PF02365	62.81	564	4.79
LusNAC33	scaffold305	PF02365	36.78	317	7.08
LusNAC34	scaffold617	PF02365	31.85	277	8.84

LusNAC35	scaffold2022	PF02365	46.68	415	5.19
LusNAC36	scaffold1347	PF02365	48.39	430	6.7
LusNAC37	scaffold42	PF02365	22.53	198	4.52
LusNAC38	scaffold98	PF02365	39.13	335	7.25
LusNAC39	scaffold203	PF02365	60.33	531	6.6
LusNAC40	scaffold61	PF02365	51.81	450	8.05
LusNAC41	scaffold1120	PF02365	18.27	158	6.52
LusNAC42	scaffold1134	PF02365	37.71	326	7.03
LusNAC43	scaffold418	PF02365	31.23	276	5.42
LusNAC44	scaffold722	PF02365	52.58	461	6.56
LusNAC45	scaffold977	PF02365	41.16	362	5.6
LusNAC46	scaffold349	PF02365	46.44	399	6.91
LusNAC47	scaffold651	PF02365	38.84	354	6.85
LusNAC48	scaffold983	PF02365	12.9	113	4.57
LusNAC49	scaffold821	PF02365	62.53	559	4.53
LusNAC50	scaffold208	PF02365	51.28	458	4.52
LusNAC51	scaffold208	PF02365	37.59	336	8.12
LusNAC52	scaffold325	PF02365	40.73	354	5.34
LusNAC53	scaffold488	PF02365	40.13	361	4.37
LusNAC54	scaffold898	PF02365	44.32	386	6.87
LusNAC55	scaffold31	PF02365	34.23	297	8.96
LusNAC56	scaffold373	PF02365	42.29	383	6.55
LusNAC57	scaffold2324	PF02365	33.53	298	9.34
LusNAC58	scaffold82	PF02365	42.37	381	4.95
LusNAC59	scaffold31	PF02365	16.93	148	10.06
LusNAC60	scaffold200	PF02365	27.97	246	8.76
LusNAC61	scaffold1376	PF02365	38.55	356	8.59
LusNAC62	scaffold251	PF02365	38.2	354	9.13
LusNAC63	scaffold31	PF02365	20.63	181	8.45
LusNAC64	scaffold123	PF02365	23.6	203	9.51
LusNAC65	scaffold488	PF02365	47.03	420	6.15
LusNAC66	scaffold783	PF02365	24.64	216	5.2
LusNAC67	scaffold540	PF02365	61.23	548	4.28
LusNAC68	scaffold51	PF02365	60.64	542	6.27
LusNAC69	scaffold2252	PF02365	37.72	336	7.86
LusNAC70	scaffold76	PF02365	39.13	349	5.16
LusNAC71	scaffold140	PF02365	10.09	87	10.65
LusNAC72	scaffold42	PF02365	37.92	332	4.54

LusNAC73	scaffold61	PF02365	44.78	391	7.34
LusNAC74	scaffold151	PF02365	40.21	353	6.09
LusNAC75	scaffold473	PF02365	18.74	164	10.05
LusNAC76	scaffold34	PF02365	35.07	304	6.65
LusNAC77	scaffold674	PF02365	42.39	376	4.47
LusNAC78	scaffold924	PF02365	26.52	237	5.67
LusNAC79	scaffold77	PF02365	27.84	250	5.4
LusNAC80	scaffold261	PF02365	40.15	352	6.24
LusNAC81	scaffold462	PF02365	52.77	474	7.51
LusNAC82	scaffold286	PF02365	67.34	606	6.79
LusNAC83	scaffold701	PF02365	16.84	146	9.63
LusNAC84	scaffold1120	PF02365	35.15	303	7.42
LusNAC85	scaffold22	PF02365	42.08	382	5.35
LusNAC86	scaffold1308	PF02365	34.63	308	8.11
LusNAC87	scaffold701	PF02365	52.31	464	4.4
LusNAC88	scaffold34	PF02365	48.29	423	7
LusNAC89	scaffold701	PF02365	54	487	8.5
LusNAC90	scaffold651	PF02365	41.58	376	9.4
LusNAC91	scaffold475	PF02365	45.36	404	6.2
LusNAC92	scaffold372	PF02365	43.86	395	6.59
LusNAC93	scaffold740	PF02365	16.45	143	7.87
LusNAC94	scaffold1216	PF02365	46.19	411	7.07
LusNAC95	scaffold430	PF02365	22.32	199	6.02
LusNAC96	scaffold706	PF02365	38.4	336	4.85
LusNAC97	scaffold732	PF02365	48.3	426	4.48
LusNAC98	scaffold59	PF02365	35.98	309	7.23
LusNAC99	scaffold43	PF02365	28.61	258	8.22
LusNAC100	scaffold406	PF02365	43.89	390	4.53
LusNAC101	scaffold511	PF02365	55	497	4.89
LusNAC102	scaffold370	PF02365	28.83	249	5.04
LusNAC103	scaffold587	PF02365	64.43	575	5.43
LusNAC104	scaffold820	PF02365	38.35	343	5.78
LusNAC105	scaffold635	PF02365	38.94	344	8.36
LusNAC106	scaffold1491	PF02365	40.75	355	6.31
LusNAC107	scaffold8	PF02365	35.34	308	6.32
LusNAC108	scaffold311	PF02365	38.83	332	7.12
LusNAC109	scaffold177	PF02365	37.02	321	8.89
LusNAC110	scaffold145	PF02365	34.12	304	7.94

LusNAC111	scaffold305	PF02365	58.6	522	4.59
LusNAC112	scaffold217	PF02365	29.18	252	4.92
LusNAC113	scaffold674	PF02365	59.14	516	4.8
LusNAC114	scaffold812	PF02365	44.97	404	9.61
LusNAC115	scaffold701	PF02365	30.87	270	8.86
LusNAC116	scaffold8	PF02365	17.91	159	9.68
LusNAC117	scaffold272	PF02365	27.1	247	8.76
LusNAC118	scaffold462	PF02365	33.96	296	8.87
LusNAC119	scaffold42	PF02365	28.82	258	8.23
LusNAC120	scaffold222	PF02365	36.09	318	6.51
LusNAC121	scaffold1168	PF02365	43.29	381	7.44
LusNAC122	scaffold883	PF02365	39.52	348	7.01
LusNAC123	scaffold86	PF02365	43.58	389	4.75
LusNAC124	scaffold413	PF02365	41.08	361	5.86
LusNAC125	scaffold1179	PF02365	28.2	250	9.48
LusNAC126	scaffold488	PF02365	43.02	390	4.88
LusNAC127	scaffold489	PF02365	62	554	4.83
LusNAC128	scaffold157	PF02365	24.11	211	8.37
LusNAC129	scaffold617	PF02365	12.6	109	4.91
LusNAC130	scaffold292	PF02365	45.87	418	8.54
LusNAC131	scaffold434	PF02365	47.71	427	6.86
LusNAC132	scaffold123	PF02365	43.25	375	6.72
LusNAC133	scaffold863	PF02365	53.53	475	4.22
LusNAC134	scaffold27	PF02365	31.32	271	9.22
LusNAC135	scaffold116	PF02365	35.55	305	6.64
LusNAC136	scaffold28	PF02365	32.09	278	7.8
LusNAC137	scaffold546	PF02365	44.57	393	4.96
LusNAC138	scaffold543	PF02365	38.96	356	6.24
LusNAC139	scaffold346	PF02365	62.85	566	4.57
LusNAC140	scaffold711	PF02365	34.01	299	5.89
LusNAC141	scaffold635	PF02365	31.13	280	7.93
LusNAC142	scaffold233	PF02365	61.22	542	4.58
LusNAC143	scaffold203	PF02365	14.08	118	4.36
LusNAC144	scaffold25	PF02365	41.86	379	8.62
LusNAC145	scaffold489	PF02365	22.44	194	9.28
LusNAC146	scaffold123	PF02365	34.38	299	5.22
LusNAC147	scaffold575	PF02365	71	631	6.65
LusNAC148	scaffold711	PF02365	43.46	398	4.87

LusNAC149	scaffold360	PF02365	32.28	280	9.58
LusNAC150	scaffold217	PF02365	48.26	431	6.56
LusNAC151	scaffold464	PF02365	22.18	192	4.81
LusNAC152	scaffold104	PF02365	44.17	398	9.15
LusNAC153	scaffold917	PF02365	48.04	429	8.23
LusNAC154	C8375105	PF02365	22.18	192	4.81
LusNAC155	scaffold420	PF02365	27.2	236	5.49
LusNAC156	scaffold54	PF02365	40.65	354	5.54
LusNAC157	scaffold151	PF02365	26.27	233	6.3
LusNAC158	scaffold94	PF02365	22.32	194	4.79
LusNAC159	scaffold783	PF02365	37.43	320	5.87
LusNAC160	scaffold511	PF02365	43.76	393	6.32
LusNAC161	scaffold472	PF02365	54.16	479	6.36
LusNAC162	scaffold1202	PF02365	42.82	377	9.42
LusNAC163	scaffold2739	PF02365	45.71	411	6.26
LusNAC164	scaffold157	PF02365	41.21	369	5.24
LusNAC165	scaffold701	PF02365	34.18	300	5.9
LusNAC166	scaffold67	PF02365	56.4	506	5.1
LusNAC167	scaffold472	PF02365	34.43	301	6.68
LusNAC168	scaffold15	PF02365	35.17	316	8.32
LusNAC169	scaffold462	PF02365	20.7	180	8.31
LusNAC170	scaffold105	PF02365	39.65	352	7.31
LusNAC171	scaffold454	PF02365	43.63	397	9.04
LusNAC172	scaffold25	PF02365	36.56	331	5.16
LusNAC173	scaffold370	PF02365	52.49	470	5.8
LusNAC174	scaffold648	PF02365	27.85	246	9.62
LusNAC175	scaffold621	PF02365	32.78	287	5.91
LusNAC176	scaffold177	PF02365	65.29	588	4.38
LusNAC177	scaffold736	PF02365	39	347	5.57
LusNAC178	scaffold169	PF02365	39.16	343	7.07
LusNAC179	scaffold898	PF02365	34.78	304	5.22
LusNAC180	scaffold6	PF02365	28.02	253	8.01
LusNAC181	scaffold67	PF02365	33.49	296	7.96

Appendix 11. Transcript abundances of *LusNACs* across tissues retrieved from a publicly available RNA-Seq dataset (Kumar et al., 2013). ge: globular embryo; he: heart embryo; te: torpedo embryo; ce: cotyledon embryo; me: mature embryo; sd: seeds; an: anthers; ov: ovaries; fl: mature flower; rt: root; st: stem; es: etiolated seedlings; le: leaves; max: the highest expression level among these tissues;

Gene Name	**ge**	**he**	**te**	**ce**	**me**	**sd**	**an**	**ov**	**fl**	**rt**	**st**	**es**	**le**
LusNAC1	0.336718	0.773295	0.717747	0.278184	1.26622	0.101946	0	0	0.089229	0	0.025228	0.704738	0
LusNAC2	0.198453	0.168349	0.304662	0.31014	0.018185	0.104568	0.006853	0.048033	0.177855	0.027298	0.155084	0.102499	0.015501
LusNAC3	0.039315	0	0	0	0	0	0	0.03578	0.028545	0	0	0	0
LusNAC4	0	0	0	0	0	0	0	0	0	0	0	0	0
LusNAC5	31.9223	41.4472	5.92709	8.09198	55.7279	26.9197	35.4217	23.8429	34.5879	36.7334	19.1654	21.5307	146.827
LusNAC6	0	0	0	0	0	0	0.028368	0	0	0	0	0	0
LusNAC7	0.068996	0.098818	0.265001	0.054832	0.015259	0.043299	0.023223	0.067504	0.063412	0.03899	0.091307	0.094642	0.03044
LusNAC8	0	0	0	0	0	0	0	0	0	0	0	0	0
LusNAC9	0.335957	0.7256	0	0	0.015336	0.724685	0.728707	0.038041	0	0.058019	0	0.00702	0.015995
LusNAC10	0	0.252024	3.03141	3.12023	2.58136	0.068392	0.036469	0.089271	3.98687	0.199265	3.65327	0.837141	0.219059
LusNAC11	0.384058	0.384358	0.996655	2.402332	0.604272	0.617508	0.356644	0.602945	0.65288	0.466933	0.734243	0.375659	0.048093
LusNAC12	0	0	0	0	0	35.443	6.15347	3.65539	0	3.04562	1.1082	0.212671	0.025137
LusNAC13	0	0	0	0.055185	0	0.087342	0	0	0	0	0	0	0
LusNAC14	0	0	0	0	0	0	0.01886	0	0	0	0	0	0
LusNAC15	0.037343	0.030435	0	0.013916	0.060678	19.0315	9.53092	2.34555	1.53083	5.56716	2.9964	2.80426	0.087446
LusNAC16	0.05044	0.014593	0.036565	5.31528	6.40533	29.0035	44.4137	0.70885	43.7316	36.2307	5.8435	10.1714	146.93
LusNAC17	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC18	0	0	0	0	0	0	0	0	0	0.027643	0.037319	0	0
LusNAC19	0.113374	0.061511	0.020515	0.096523	0.414136	0.228895	0.601337	0.176203	1.7532	0.149952	0.762194	0.406292	0.016604
LusNAC20	0.020517	0	1.0619	1.2127	1.0714	0.202894	0.008652	0.633586	2.24879	0.379931	4.43397	1.73828	0.613778
LusNAC21	0	0	0.294803	0.092289	0.015556	0.64152	0.065719	0.529737	8.73471	1.97074	2.65032	1.35508	0.601178
LusNAC22	0	0	0	0	0	1.30382	0.145389	0	0.024293	0.160049	0.185666	0.031311	0.024337

LusNAC23	0.403918	0.045202	0.107408	0.05931	0.160757	1.92433	0.285859	1.42576	1.37357	0.839193	1.35505	0.372126	0.102648
LusNAC24	0.046747	0.119525	0.094793	0.148676	1.09146	2.39687	2.06922	0.304615	0.177214	1.53259	0.23772	0.291439	0.096914
LusNAC25	0.250472	0.457578	0.473391	0.224055	0.428064	0.258423	0.187359	0.445078	1.127431	0.999724	2.170597	0.087316	0.086889
LusNAC26	0.366574	0.158946	0.045653	0.068789	0.050186	14.7255	64.012	0.97794	0.225366	2.75449	1.53279	0.15031	5.68277
LusNAC27	6.45069	7.85187	3.45763	4.17052	0.884668	4.14484	1.7062	4.40846	5.51612	1.85938	4.23086	3.8777	0.672621
LusNAC28	0.027058	0.398017	1.26771	1.61702	1.44031	0.429806	0.036985	0.631303	2.28614	0.297077	2.71701	0.509237	0.04704
LusNAC29	0.007399	0	0	0	0	10.952	0.629594	0.347612	2.07151	4.32347	1.40059	5.99571	1.03117
LusNAC30	12.8821	24.5436	20.2243	8.07044	31.8539	1.19253	1.38391	1.39046	0.676587	0.818571	1.14799	1.38818	1.26165
LusNAC31	0.014208	0.032188	0.056047	0.094521	0.041504	4.32076	42.3252	2.63813	2.54193	8.94024	3.53958	0.830401	0.302293
LusNAC32	8.64584	11.2953	26.7441	26.9893	49.8655	14.8867	16.2781	16.5368	31.638	6.9766	14.7495	14.3472	2.53863
LusNAC33	0	0	0.017364	0	0	0.088077	0.055444	0.05685	2.08725	1.96221	2.39396	0.473571	0.721023
LusNAC34	0.066421	0.095795	0.091	0.035399	0.070102	0.690731	0.245972	0.518256	0.101871	0.436305	0.22557	0.356964	0.091633
LusNAC35	0.232341	0.284597	0.227786	0.207882	0.498452	0.123315	0.012286	0.135811	0.056924	0.186344	0.272814	0.251592	0.231308
LusNAC36	0.043005	0	0	0	0.075999	2.67354	0.056312	0.533685	4.07558	4.35204	8.32725	0.171633	0.314339
LusNAC37	0	0	0	0	0	0	0.021185	0	0	0	0	0	0
LusNAC38	0	0	0	0	0.018138	0.019441	0.041287	0.006411	0	0.433067	0.574525	0.791868	0.912433
LusNAC39	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC40	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC41	0	0	0	0	0	0	0	0.021331	0	0.0263	0	0	0.09183
LusNAC42	0.005466	0.185881	8.751665	8.16154	0.490415	0.158852	0	0.421407	1.194737	0.257253	0.935139	1.013144	0.937547
LusNAC43	0.804446	0.469483	0.34084	8.03473	9.70335	5.13939	24.7026	0.545883	46.4301	5.45153	0.615838	5.34895	0.559159
LusNAC44	0.570475	0.603525	0.877399	0.364288	1.16202	5.52193	2.45433	4.50952	4.72575	1.61846	5.8341	7.46481	0.911313
LusNAC45	0	0	0	0	0.016538	0.023862	0.10389	0	0.408494	0.174106	0.064206	0.150315	0.023484
LusNAC46	0.029941	0	0.24988	0.045717	0.069372	0.163172	0.022298	0.393942	0.876178	0.248658	1.756968	0.947828	0.689828
LusNAC47	0	0	0	0.023535	0.016982	8.8952	3.1268	0.673012	9.68344	3.10993	2.51094	2.98565	0

LusNAC48	0.003257	0.024527	0.011438	0.230567	0.001768	0.745581	0.022465	0.050067	0.037422	1.57883	0.110292	0.032723	2.05517
LusNAC49	0.541894	0.429664	0.390177	1.91426	3.34468	6.75897	6.17626	3.72951	7.72688	5.42058	7.27011	6.35896	0.371423
LusNAC50	11.2549	10.1189	3.02668	3.81238	2.96208	0.507706	0.157144	0.418804	2.40238	0.134027	2.45885	2.57701	0.432617
LusNAC51	0.054106	0	0.403086	1.68826	0.024178	10.2597	72.6065	9.06093	5.49024	7.58301	6.04071	5.00553	0.042436
LusNAC52	0.013911	0.045218	0.129377	0.053518	0.047384	0.029987	0.247582	0.367733	0.091476	0.13076	0.02388	0.014389	0
LusNAC53	0.436604	0.348087	0.481697	0.073078	0.179902	0.528925	0.110649	0.250063	0.227111	0.157849	0.280426	0.230585	0.188569
LusNAC54	0.007173	0	0	0	0.015336	0.058163	21.7195	0.048209	0.887775	0.065123	0.339476	1.89579	0.000105
LusNAC55	0.013598	0.01836	0.398494	1.20022	2.02072	3.65288	0.639753	0.01613	1.47064	0.121479	0.506189	0.661971	0.05109
LusNAC56	3.05458	2.56936	0.148794	0.020318	0.070873	0.06178	0.008032	0.414556	0.076283	0.099183	0.068244	0.181779	0.048336
LusNAC57	0	0	0.117273	0.184408	0.264543	0.466222	0.014258	0.518194	0.132907	0.421948	1.1849	0.641494	0.000491
LusNAC58	9.6615	36.0066	0.665189	0.503653	0.016683	0.181939	0.760871	1.84532	0.0444	0.168264	0.030179	0.199423	0.063633
LusNAC59	1.38003	2.03062	1.77731	1.20698	3.59258	4.95178	3.25782	3.55798	5.35389	3.76472	5.89408	3.79374	1.57027
LusNAC60	0	0	0	0	0	0	0	0	0	0	0	0.012351	0
LusNAC61	0.025047	0	0.193145	0.968759	0.594264	0.062524	0.001431	0	0.275865	0.047124	0.034691	0.562468	0
LusNAC62	0	0.049696	0	0.139235	0.165658	0.92091	71.9192	7.70656	0.372065	11.862	0.182434	0.616535	0.787992
LusNAC63	0	0.020655	0	1.10841	1.12254	0.726467	38.9113	5.4307	0.60043	3.81431	0.14835	0.265229	0
LusNAC64	0	0	0	0	0	0	0	0	0.012446	0	0	0	0
LusNAC65	0.000316	0.033809	0.071724	0	0	0.188101	0.226806	0.524108	0.057669	0.69305	0.044853	0.041306	0
LusNAC66	0	0	0.018026	0	0.013904	2.65228	0.035296	1.245	2.33801	1.21731	2.81417	0.294551	0
LusNAC67	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC68	21.0154	23.397	34.1915	34.7942	74.3488	25.6842	33.9117	18.408	20.6491	8.39784	14.5553	19.4539	3.04225
LusNAC69	1.7927	1.88629	4.46002	1.36571	1.71977	7.13806	6.24577	4.59959	21.2799	2.76314	7.18071	3.93719	0.429288
LusNAC70	0.004063	0.033454	0.018891	0.352512	0.020104	5.68092	25.188	3.10074	5.29835	2.08464	1.62214	2.89074	0.040813
LusNAC71	0.035705	0	0.030908	0	0	0.015511	0.531525	0	0.014895	0.021759	0.006254	0.011184	0
LusNAC72	0	0	0	0.215744	0.159436	0.977343	0	0.077399	0.33239	0.10088	1.22676	0.284435	0

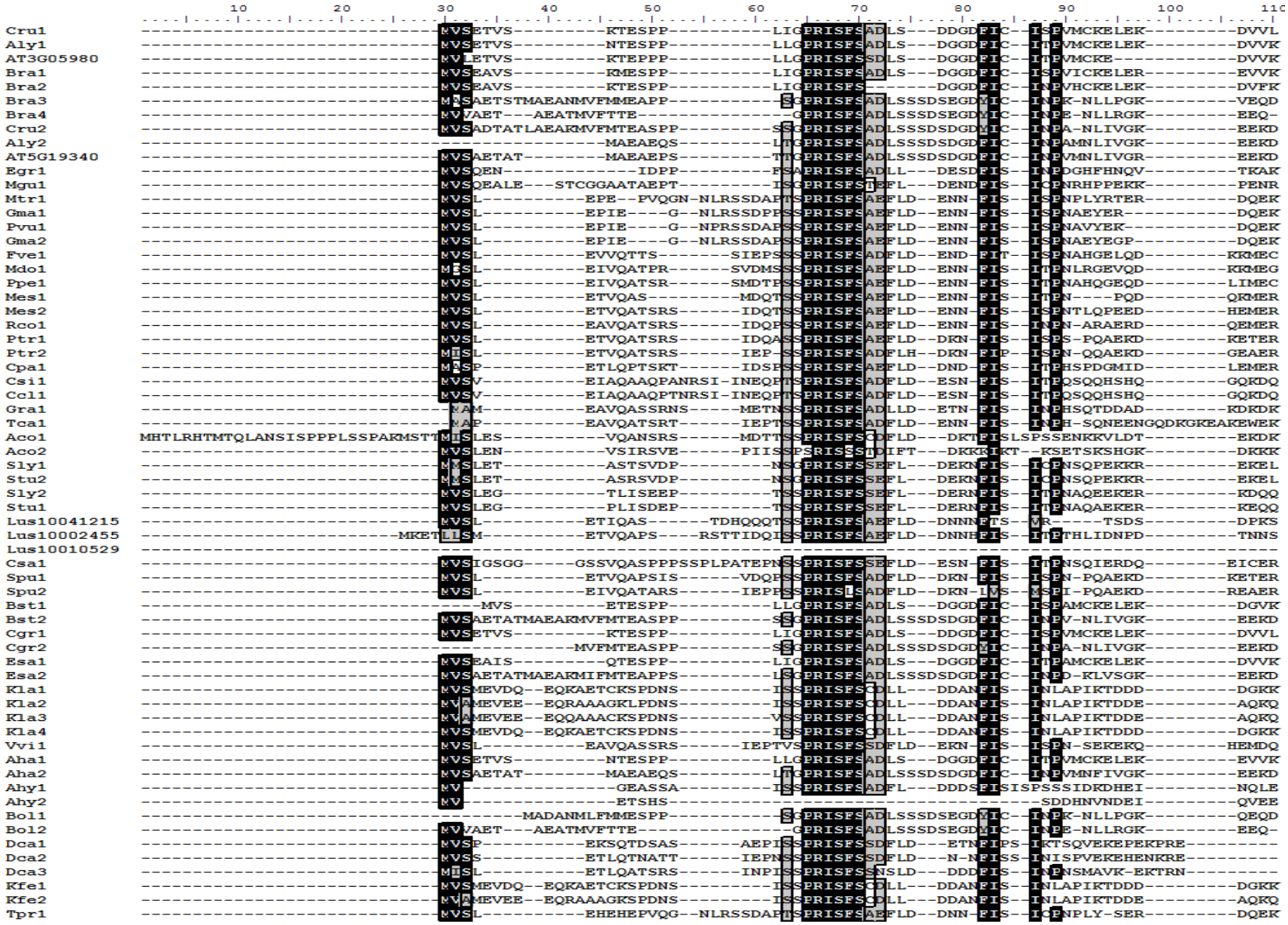
LusNAC73	0.017176	0	0	0	0.03667	0	0.35209	0	0.041802	0.016519	0.013165	0	0
LusNAC74	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC75	0	0	0	0.118659	0	0	0	0	0.284122	0.03387	0.289443	0.09404	0.428655
LusNAC76	0.085819	0	0	0	0	0.087635	2.72971	0.840943	0	0.170723	0	0	0
LusNAC77	0	0	0	0	0	0.00618	0	0	4.84763	0	0	0.462612	0
LusNAC78	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC79	4.34824	8.43382	10.752	34.9504	37.8693	13.8019	21.6154	8.6987	18.3131	14.9109	8.65125	4.56614	0.643663
LusNAC80	0.03678	0.024456	0.077807	0	0	0.075405	8.39239	0.063793	0.911401	1.48712	1.19106	0.026352	0.00805
LusNAC81	0.051284	0	0.063713	0.247383	0	0.027738	0.045953	0.018088	0.37839	0.096275	0.201114	0.176357	0
LusNAC82	0.286264	0.316235	0.413137	0.126241	0.784444	0.00646	0	0	0.045318	0.026301	0	0.179197	0
LusNAC83	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC84	0	0	0	0	0	0.075628	2.28359	0	0.029509	0.026703	0	0.153028	0
LusNAC85	0.031879	0	0.066684	0.159463	0.068563	0.156815	0.140772	0.38481	5.17515	5.35145	6.66505	0.646185	230.404
LusNAC86	1.10237	1.44074	0.424978	0.421757	0.277697	3.10187	8.5423	5.12787	11.266	1.80073	6.57954	4.16269	0.891831
LusNAC87	0	0	0	0.026612	0.140563	2.41492	0.050068	0.707159	7.51789	1.57385	9.82153	0.859992	1.66448
LusNAC88	0.021387	0.016837	0	0	0	0	0.006611	0	0	0	0	0	0
LusNAC89	0	0	0	0	0	0	0	0	0.203051	0	0	0.147931	0
LusNAC90	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC91	0	0	0	0	0	0.135988	0.067701	0	1.16366	0.008446	0.020564	0.165464	0
LusNAC92	5.24209	7.79015	3.8136	0.755906	2.27038	14.0876	22.3348	13.4741	8.99503	7.11304	9.9019	9.43929	1.99021
LusNAC93	11.2674	15.9768	0.109382	0	0.014929	0.036265	0.034367	0.455896	1.26434	0.204671	0.168116	0.104999	0.046857
LusNAC94	0	0	0	0	0	0	0	0	0	0	0	0	0
LusNAC95	7.23166	9.02056	4.24151	8.88296	6.41031	9.37338	4.31422	7.91871	6.60803	22.6104	7.12714	6.67733	71.2886
LusNAC96	1.14409	0.785258	0.056915	0	0	0.590996	19.8766	0.070137	0.010608	0.472076	0.022294	0	0
LusNAC97	0.008465	0	0	0	0	0	0	0	0.097344	0.008999	0	0	0

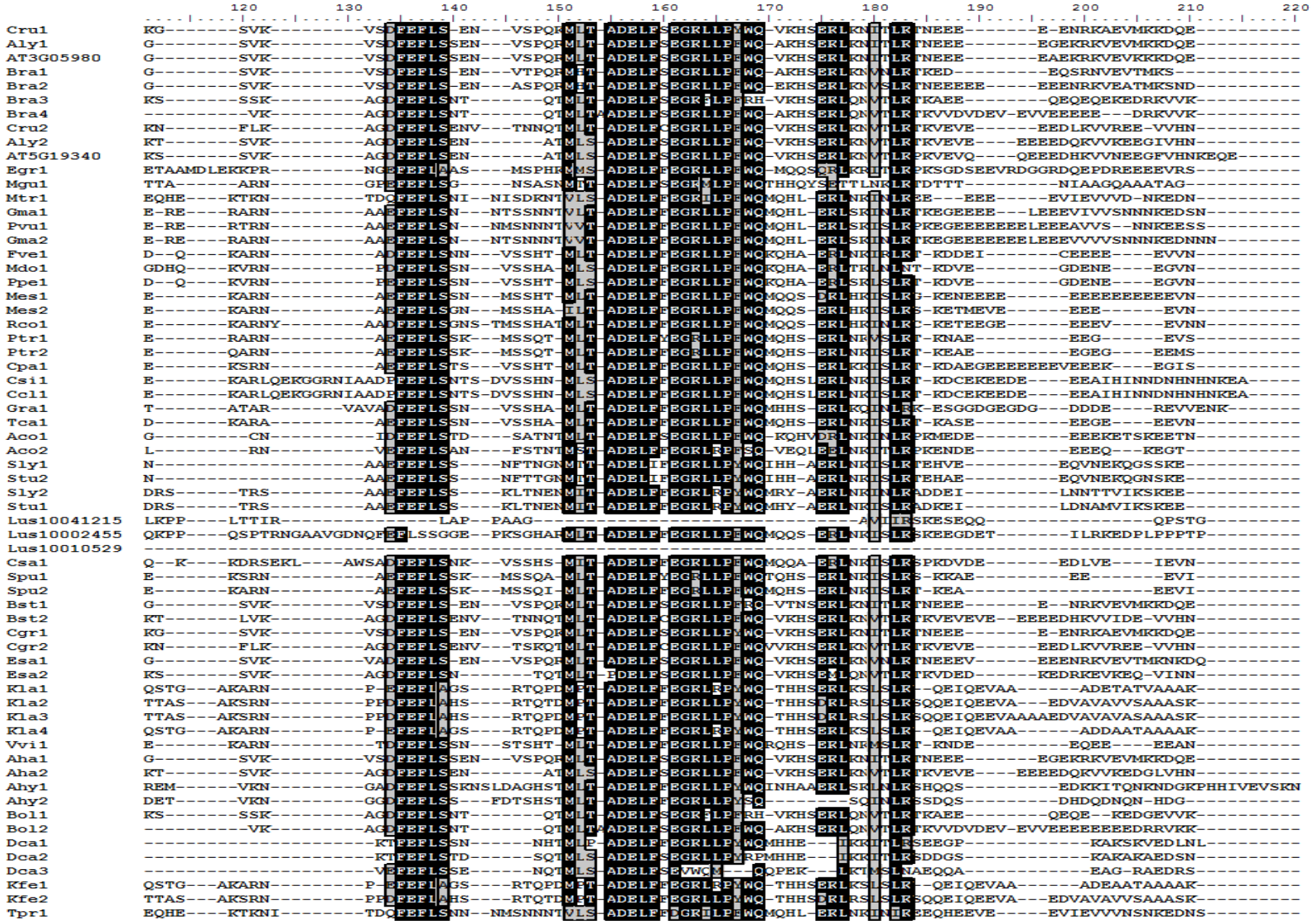
LusNAC98	0	0	0	0	0	0.033468	0.076935	0.004614	0.05686	0	0.090207	0.021598	0.029676
LusNAC99	0	0	0	0	0.131344	0.433124	0.056643	0.402446	1.40585	0.827547	5.23623	0.124714	0.097829
LusNAC100	8.44947	13.6678	0.964074	0.034888	0.414623	0.221223	0.032718	0.721517	12.0734	0.130775	7.50472	4.94523	0.953278
LusNAC101	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC102	3.52647	4.01496	4.02398	3.76456	7.91101	5.08085	8.20162	5.08536	7.18463	3.55376	7.00064	6.04903	1.70063
LusNAC103	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC104	6.10794	24.82495	10.08935	4.63047	15.2105	2.896799	0.480198	0.590004	3.41053	0.271102	0.643723	3.57529	0.312089
LusNAC105	0	0	0	0	0	0	0	0	0	0	0	0	0
LusNAC106	0.27088	0.260517	0.275537	0.329486	0.430622	1.24252	1.03536	1.83333	1.6522	0.290285	0.50662	0.443902	0.025896
LusNAC107	0.520882	0.446202	0.050113	0.029341	0	0.026709	0.122891	0.079738	0.110804	0.130771	0.062515	0.176144	0.357401
LusNAC108	0.461062	0.546061	1.39504	0.631287	0.10216	4.4756	2.38508	4.27486	2.20859	9.63177	2.77817	5.71384	0.054868
LusNAC109	0	0	0	0	0	0	0	0.012967	0	0.070048	0.404005	0.279992	0.848759
LusNAC110	0	0	0	0	0	0	0.03272	0.006764	0	0	0.058407	0.010629	0
LusNAC111	0.843051	0.213798	0.80053	76.8857	90.9638	26.0494	80.6445	9.52955	52.863	95.997	4.24403	19.7772	248.377
LusNAC112	0	0	0	0	0	0	0.005766	0.01238	0	0	0.020214	0.002329	0
LusNAC113	0.0301	0.03671	0	0	0.016083	1.43108	0	0.016625	0.087315	1.0065	0.061225	0	0.028762
LusNAC114	0.550262	0.654911	2.08311	0.614693	1.04536	4.24985	1.16541	5.04231	4.44029	1.426	5.39276	7.6172	1.22353
LusNAC115	0.097978	0.136296	6.967985	4.836765	50.5597	1.158807	2.947425	0.018915	0.060024	0.039195	0.141999	0.010999	0
LusNAC116	0	0	0	0	0	0	0.001822	0	0.00078	0	0	2.39E-05	0
LusNAC117	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC118	0.563368	1.86684	0.243521	0.702075	0.579822	0.870558	0.187144	0.57929	7.689553	0.604269	2.721655	2.233038	0.035367
LusNAC119	0.255803	0.295922	2.39406	17.8629	6.52716	33.5289	7.91386	0.67843	8.52022	3.54769	11.6959	2.97153	3.64062
LusNAC120	15.5985	21.9078	1.19403	0.22204	0.269206	0.182137	0.29174	0.197795	5.32009	0.102874	7.22621	4.69437	0.685842
LusNAC121	0.009051	0	0	0	0	1.96406	8.38313	1.51048	1.90806	1.27899	0.867927	0.702711	0
LusNAC122	0.007284	0	0	0	0	0	0	0	0.185643	0	0	0.06924	0

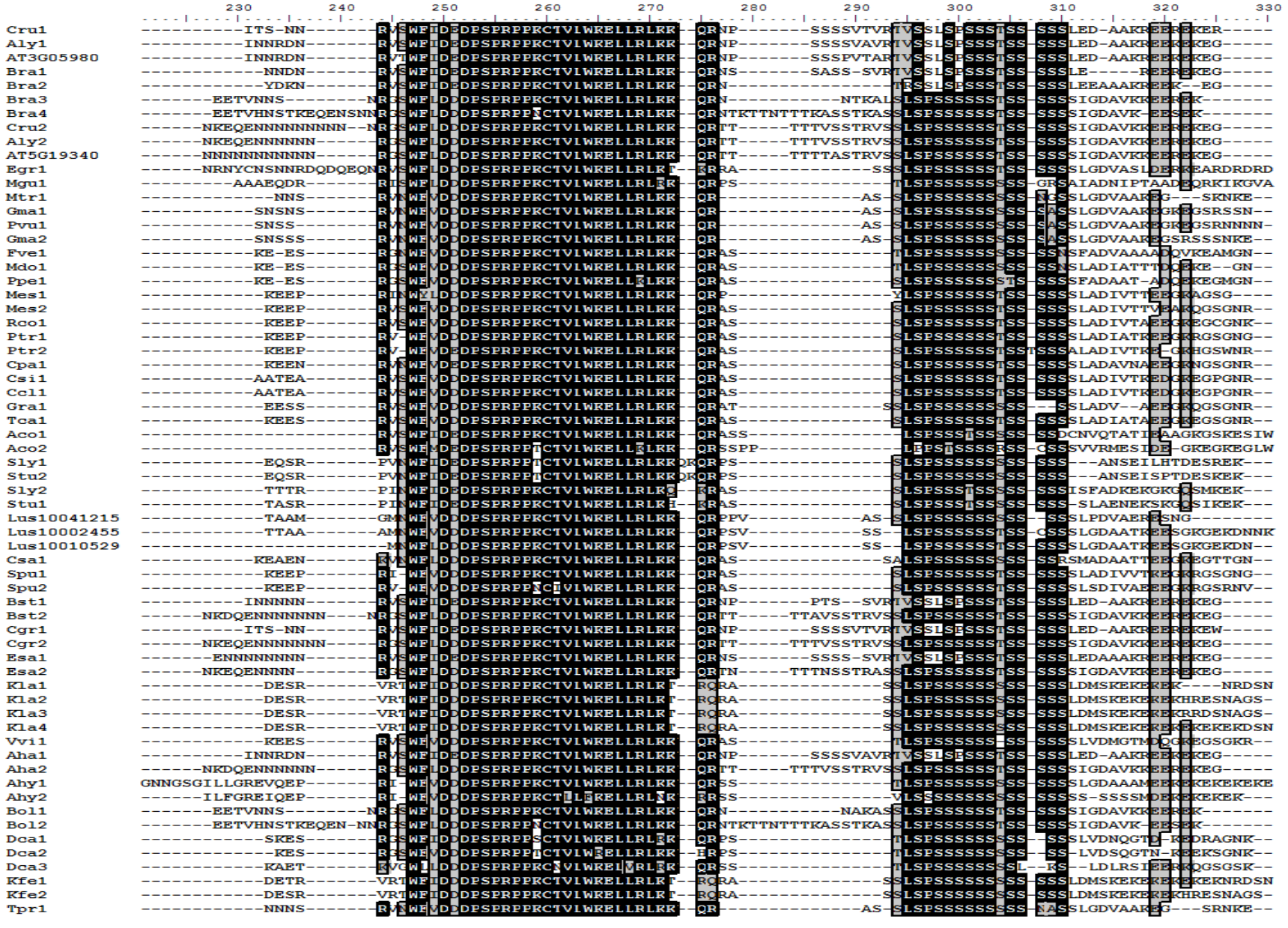
LusNAC123	0	0	0	0	0	0.083904	0.033779	0	3.47448	0.025118	0	0.30155	0
LusNAC124	2.57574	4.80524	2.32942	1.78007	1.83409	0	3.28979	4.99027	4.76269	8.11825	4.60163	4.45445	0.007772
LusNAC125	0.026757	0.251733	4.40722	0.959962	3.22518	0.591801	0.028335	0.572571	1.40992	0.703168	4.34327	0.828986	0.272227
LusNAC126	1.24634	1.55175	2.02576	9.4782	1.95064	32.7994	80.7422	8.46139	19.8645	28.8607	23.269	31.3117	450.197
LusNAC127	0.123177	0.140965	0.296579	0.046576	0.130378	0.267356	0.520445	0.409241	0.768331	0.252211	0.462338	0.418114	0.170776
LusNAC128	12.8563	12.3121	22.0161	20.4055	46.1406	13.3854	10.7164	13.0259	15.2006	4.7662	11.0648	10.942	2.13312
LusNAC129	3.89861	5.01571	13.453	12.4258	3.89448	1.07985	1.49558	3.33201	3.993	5.28253	3.15683	2.99714	19.8752
LusNAC130	0	0	0.002225	0	0	0	0	0	0	0.012034	0	0.00105	0
LusNAC131	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC132	0.466244	0.105214	0.100051	0.099364	0.342171	2.13429	0.252255	1.02027	2.52665	0.70305	2.88879	0.496569	0.175979
LusNAC133	0	0.011334	0	0	0	0.436803	53.7919	0.779636	2.68208	1.30517	0.616471	1.85523	3.94075
LusNAC134	0	0	0	0	0.062427	0	0	0	0	0	0	0	0.01348
LusNAC135	0	0.378777	0.04058	1.15136	2.21534	0.116784	0.223947	0.055994	1.70592	1.19484	0.400388	1.2472	0
LusNAC136	0	0	0	0	0	0.19142	0.81633	0.123799	0.064094	0.140585	0.196574	0.141143	0.223016
LusNAC137	0.060709	0	0	0	0	0.014979	0.017505	0.008952	4.95313	0.161399	0.8298	1.70732	0
LusNAC138	0.08447	0.352806	0.735152	0.230327	0.331891	0.500701	0.480101	0.28527	0.469993	0.294902	0.813883	0.379169	0.223339
LusNAC139	0.137754	0.229866	0.146751	0.223006	0.077719	9.24838	13.1317	1.15204	7.26095	2.70298	1.26888	2.47985	0.053131
LusNAC140	2.43342	2.42712	1.25208	2.34781	5.03879	7.48792	5.78999	5.52887	16.4356	3.68078	10.4515	7.51206	0.493533
LusNAC141	6.59555	3.37372	5.42769	3.35796	10.9319	5.59234	3.94441	3.74855	75.2371	6.57683	4.59373	12.4511	0.089853
LusNAC142	0.010603	0.016198	0.083718	0.11709	0.203301	5.40556	21.1187	1.36673	4.41115	4.16354	6.37929	1.55172	0.500532
LusNAC143	3.74875	4.0222	9.87681	12.2561	1.28296	5.34017	7.44938	5.49318	21.5315	2.28528	7.26269	3.88325	0.770935
LusNAC144	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC145	0.187615	0.195924	0.017693	0.08472	0.112899	4.89977	23.8199	0.320617	0.322584	0.403648	0.458648	0.195466	0
LusNAC146	0.179719	0	0	0	0.036654	7.05589	0.006534	3.01986	5.91296	2.89943	5.51611	0.442605	0.014317
LusNAC147	0.067189	0.09594	0	0.138873	0.296691	0.827323	0	0.25872	1.41263	0.309796	0.895117	0.440883	0.187748

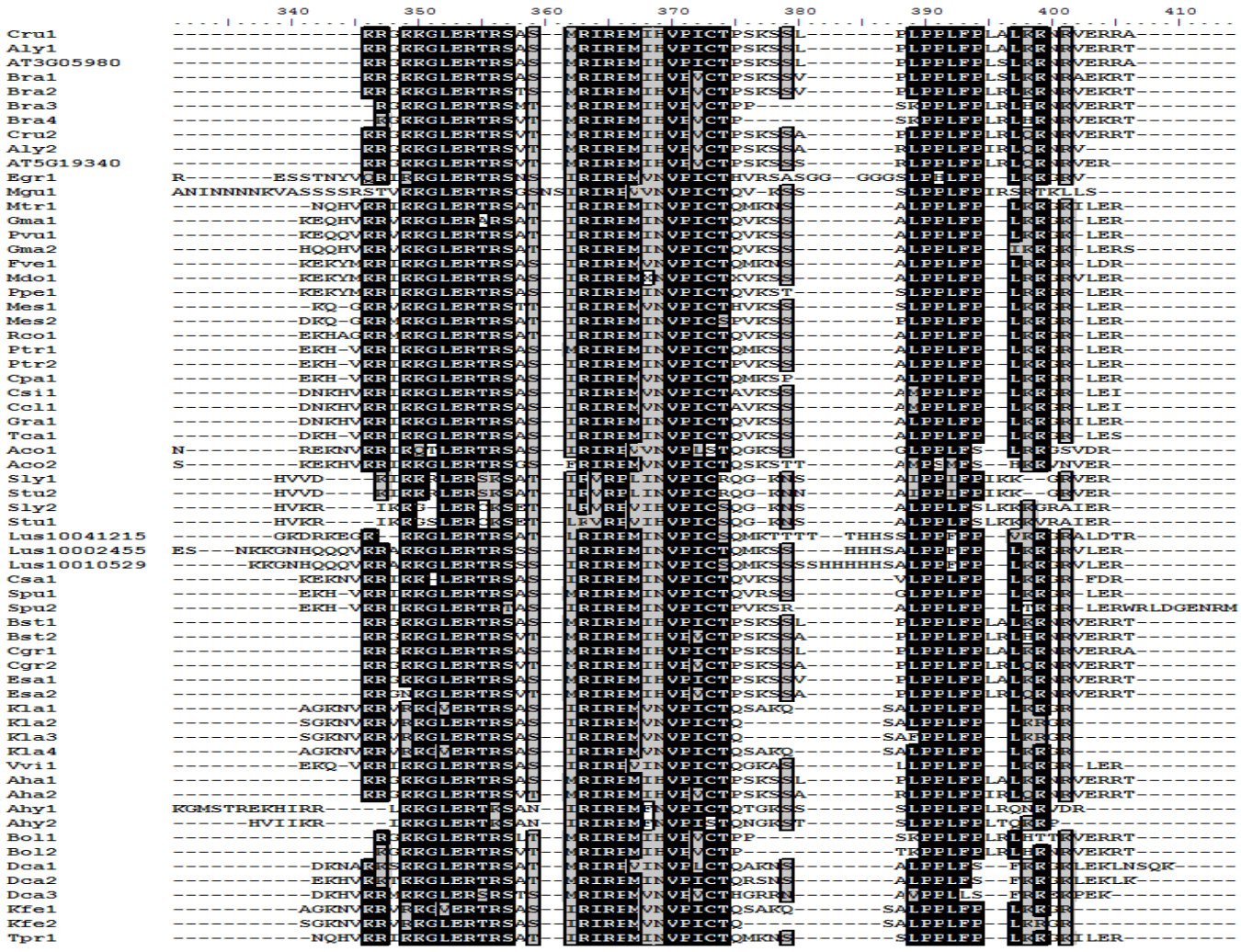
LusNAC148	0.012313	0.006279	0.015693	0.028661	0.008803	0.065912	0.144345	0.060163	3.7164	0.404971	0.619471	0.178447	0
LusNAC149	3.07278	3.34458	7.98016	3.9833	3.71044	10.1076	12.5523	10.4233	11.0256	4.86026	8.68764	6.69956	1.33834
LusNAC150	0.060621	0	0	0	0	0	0.022592	0	0.019778	0	0	0.05195	0
LusNAC151	0	0	0	0.017861	0.068167	1.52442	0.01861	0.263835	2.14214	1.00161	3.71344	0.237464	0.055445
LusNAC152	0	0	0.117273	0.184408	0.264543	0.466222	0.014258	0.250899	0.61663	0.421948	1.5997	0.641494	0.000491
LusNAC153	8.79386	8.66732	19.4423	42.2898	24.9166	0.67414	0.399119	0.141921	0.530101	0.242387	0.459659	0.161901	0.848354
LusNAC154	0.752394	1.38325	0.272448	0.202857	0.037429	0.019591	0	0	0.052366	0	0	0.029766	0
LusNAC155	0	0	0.117273	1.24156	0.264543	0.466222	0.014258	0.250899	0.132907	1.05397	0.946186	0.641494	1.26625
LusNAC156	13.1308	13.9611	3.67057	4.06523	20.7063	3.30902	3.29053	3.15549	2.70069	1.94454	2.45662	1.93729	0.388462
LusNAC157	0.001987	0	0.133714	0.053967	0	0.010176	0.048157	0.02737	0.094588	0.00395	0.014691	0.164316	0.033196
LusNAC158	10.85715	13.34995	8.362045	23.32785	14.8573	4.33178	1.450553	4.11986	1.339353	5.50311	1.291811	0.713633	1.176936
LusNAC159	0	0	0.049094	0.066312	0	0.345534	2.35192	0.226296	2.20016	1.91527	0.855739	0.64484	0.04232
LusNAC160	0	0	0.034544	0	0.067196	0.043861	0	0.04605	2.84832	0.212414	4.26753	0.479259	0.179844
LusNAC161	0.064345	0	0.020997	0	0.225272	8.78469	0.224579	4.70292	17.836	5.4439	28.4738	1.18059	0
LusNAC162	0.149755	0.328539	0.051982	0.042951	0.19234	0.017197	0	0.020869	0.013555	0.013789	0.024079	0.029113	0
LusNAC163	0.007376	0.011263	0	0	0	1.59297	2.82335	0.481748	2.33453	1.54851	0.442029	3.6245	256.185
LusNAC164	0	0	0	0	0.071281	7.59054	0.209532	4.35124	7.91619	5.19937	21.3167	0.704213	0
LusNAC165	1.15819	1.98823	1.3623	1.20616	0.240728	0.486671	0.150505	0.754052	1.66953	0.846449	0.989587	0.502528	1.79468
LusNAC166	11.0501	11.2377	7.60416	2.06888	2.06996	4.33178	1.34299	4.5013	97.0978	7.06157	9.82386	7.73345	0.057692
LusNAC167	9.3075	5.90858	3.43686	1.22872	1.78105	6.33988	0.156055	2.19509	10.2741	3.73137	4.90173	6.67046	1.19598
LusNAC168	3.008669	3.52704	2.708755	0.96151	1.24019	1.726045	0.86421	1.710225	3.17938	0.554695	1.842666	1.930719	0.75008
LusNAC169	0.008631	0	0.00825	0.111359	0.186331	3.64835	0.085166	1.1608	11.025	3.21254	14.318	1.31251	3.85803
LusNAC170	0	0	0	0	0	0	0	0	0	0	0	0	0
LusNAC171	0.008003	0.076987	0.038764	0.159443	0.044409	0.353727	0	0.006029	1.55399	0.072884	0.242794	0.43179	0.215063
LusNAC172	0	0	0	0	0	0.037958	0	0	0.615195	0.00835	0.006331	0.066498	0

LusNAC173	0.008528	0	0.059365	1.05356	0.495881	0.027674	0.038741	0.03919	0.06399	0.01599	0.057264	0.038777	0
LusNAC174	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
LusNAC175	0.299025	0.880591	1.15501	0.540015	0.705099	29.7062	52.6712	11.0391	14.4234	8.33102	16.7564	30.5908	5.72805
LusNAC176	0.705853	0.341298	0.467321	0.23034	0.149072	4.2692	3.52713	2.47017	7.40648	1.15679	5.04863	3.2807	0.57188
LusNAC177	0.008875	0.013574	0.008483	0	0	0.004432	0.005033	0.006603	0	0.004691	0.003381	0	0
LusNAC178	0	0	0	0	0	0	0	0	0.065412	0	0	0	0
LusNAC179	0	0	0	0	0	0.009466	0	0	1.52005	0	0	0.148462	0
LusNAC180	0	0.083378	0	0.036647	0.066991	0.235123	0.046666	0.110926	0.876766	0.2598	0.465098	0.259689	0.034827
LusNAC181	0.384554	8.67205	0.29267	0.318218	0.030333	1.99009	0.379229	0.955812	0.93794	0.663744	0.691327	5.68034	0.033854
LusNAC182	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A


Cru1
Aly1
AT3G05980
Bra1
Bra2
Bra3
Bra4
Cru2
Aly2
AT5G19340
Egr1
Mgu1
Mtr1
Gma1
Pvu1
Gma2
Fve1
Mdo1
Ppe1
Mes1
Mes2
Rco1
Ptr1
Ptr2
Cpa1
Csi1
Ccl1
Gra1
Tca1
Aco1
Aco2
Sly1
Stu2
Sly2
Stu1
Lus10041215
Lus10002455
Lus10010529
Csa1
Spu1
Spu2
Bst1
Bst2
Cgr1
Cgr2
Esa1
Esa2
Kla1
Kla2
Kla3
Kla4
Vvi1
Aha1
Aha2
Ahy1
Ahy2
Bol1
Bol2
Dca1
Dca2
Dca3
Kfe1
Kfe2
Tpr1




230
240
250
260
270
280
290
300
310
320
330
Cru1
Aly1
AT3G05980
Bra1
Bra2
Bra3
Bra4
Cru2
Aly2
AT5G19340
Egr1
Mgu1
Mtr1
Gma1
Pvu1
Gma2
Fve1
Mdo1
Ppe1
Mes1
Mes2
Rco1
Ptr1
Ptr2
Cpa1
Csi1
Ccl1
Gra1
Tca1
Aco1
Aco2
Sly1
Stu2
Sly2
Stu1
Lus10041215
Lus10002455
Lus10010529
Csa1
Spu1
Spu2
Bst1
Bst2
Cgr1
Cgr2
Esa1
Esa2
Kla1
Kla2
Kla3
Kla4
Vvi1
Aha1
Aha2
Ahy1
Ahy2
Bol1
Bol2
Dca1
Dca2
Dca3
Kfe1
Kfe2
Tpr1



Appendix 12. Multiple sequences alignments of all the At3g05980 homologs identified from Phytozome v12.1 by ClustalW (Goodstein et al., 2012; Thompson et al., 2002). Residues with >75% identity were shadowed. The full species names were listed in the Appendix 13.

Appendix 13. Abbreviations of plant species names used in this thesis.

Abbreviation	Species Name	Abbreviation	Species Name
Mes	*Manihot esculenta*	*Cpa*	*Carica papaya*
Rco	*Ricinus communis*	*Gra*	*Gossypium raimondii*
Lus	*Linum usitatissimum*	*Tca*	*Theobroma cacao*
Spu	*Salix purpurea*	*Csi*	*Citrus sinensis*
Ptr	*Populus trichocarpa*	*Ccl*	*Citrus clementina*
Mtr	*Medicago truncatula*	*Egr*	*Eucalyptus grandis*
Pvu	*Phaseolus vulgaris*	*Stu*	*Solanum tuberosum*
Gma	*Glycine max*	*Sly*	*Solanum lycopersicum*
Csa	*Cucumis sativus*	*Mgu*	*Mimulus guttatus*
Ppe	*Prunus persica*	*Kla*	*Kalanchoe laxiflora*
Mdo	*Malus domestica*	*Aco*	*Aquilegia coerulea*
Fve	*Fragaria vesca*	*Vvi*	*Vitis vinifera*
Ath	*Arabidopsis thaliana*	*Aha*	*Arabidopsis halleri*
Aly	*Arabidopsis lyrata*	*Ahy*	*Amaranthus hypochondriacus*
Bst	*Boechera stricta*	*Bol*	*Brassica oleracea capitata*
Cgr	*Capsella grandiflora*	*Dca*	*Daucus carota*
Cru	*Capsella rubella*	*Kfe*	*Kalanchoe fedtschenkoi*
Esa	*Eutrema salsugineum*	*Tpr*	*Trifolium pratense*
Bra	*Brassica rapa*		

Appendix 14.

Lus10041215 CDS

ATGGTGTCTTTAGAAACAATTCAAGCATCCACCGATCACCAACAGCAAACATCAAG

CCCCCGAATCTCATTCTCCGCCGAGTTTCTCGACGACAACAACAACTTCACCTCCGT

CCGTACCTCTGACTCAGATCCAAAATCCCTCAAACCGCCGCTGACTACGATTCGACT

CGCACCACCTGCGGCTGGGGCGGTTATTATCCGGAGCAAGGAATCAGAGCAGCAGC

AGCCGTCGACGGGGACGGCTGCGATGGGGATGAATTGGTTCGTGGACGACGATCCG

TCGCCGCGGCCGCCAAAGTGTACTGTTCTGTGGAAGGAATTGCTGAGGCTTAAGAA

GCAGAGGCCGCCTGTTGCTTCGTCTCTTTCGCCTTCTTCGTCCTCGTCTTCGTCATCGT

CGTCAAGCTCGCTGCCGGATGTAGCGGAGAGAGAAAGCAACGGTGGTAAGGATAGA

AAAGAGGGGAAGAAGAAAGGGCTGGAGAGGACGAGATCGGCGACTCTTAGGATTA

GGCCGATGATCAATGTCCCCATTTGTAGCCAGATGAAGACCACCACCACCACCCATC

ATTCTTCATTGCCGCCATTTTTCCCGGTTAAGAAAGGCAGAGCATTAGATACTAGAT

AA

Amino acid sequences of Lus10041215 homologs used in the multiple sequence alignment

>Capsella rubella 1

MVSETVSKTESPPLIGPRISFSADLSDDGDFICISPVMCKELEKDVVLKG

SVKVSDFEFLSENVSPQKMLTADELFSEGKLLPYWQVKHSEKLKNITLKT

NEEEEENRKAEVMKKDQEITSNNRVSWFIDEDPSPRPPKCTVLWKELLRL

KKQRNPSSSSVTVRTVSSLSPSSSTSSSSSLEDAAKREEREKERKRGKKG

LERTRSASMRIRPMIHVPICTPSKSSLPLPPLFPLALKKNRVERRA

>Arabidopsis lyrata 1

MVSETVSNTESPPLLGPRISFSADLSDGGDFICITPVMCKELEKDVVKGS

VKVSDFEFLSSENVSPQRMLTADELFSEGKLLPFWQAKHSEKLKNITLKT

NEEEEGEKRKVEVMKKDQEINNRDNRVSWFIDEDPSPRPPKCTVLWKELL

RLKKQRNPSSSSVAVRTVSSLSPSSSTSSSSSLEDAAKREEKEKEGKRGK

KGLERTRSASMRIRPMIHVPICTPSKSSLPLPPLFPLALKKNRVERRT

>At3g05980.1

MVLETVSKTEPPPLLGPRISFSSDLSDGGDFICITPVMCKEDVVKGSVKV

SDFEFLSSENVSPQRMLTADELFSEGKLLPFWQVKHSEKLKNITLKTNEE

EEAEKRKVEVKKKDQEINNRDNRVTWFIDEDPSPRPPKCTVLWKELLRLK

KQRNPSSSPVTARTVSSLSPSSSTSSSSSLEDAAKREEKEKEGKRGKKGL

ERTRSASMRIRPMIHVPICTPSKSSLPLPPLFPLSLKKNRVERRA

>Brassica rapa 1

MVSEAVSKMESPPLIGPRISFSADLSDGGDFICISPVICKELEREVVKGS

VKVSDFEFLSENVTPQRMHTADELFSEGKLLPFWQAKHSEKLKNVNLKTK

EDEQSRNVEVTMKSNNDNRVSWFIDEDPSPRPPKCTVLWKELLRLKKQRN

SSASSSVRTVSSLSPSSSTSSSSSLEREEREKEGKRGKKGLERTRSASMR

IRPMIHVPVCTPSKSSVPLPPLFPLSLKKNRAEKRT

>Brassica rapa 2

MVSEAVSKTESPPLIGPRISFSDGGDFICINPVHCKELEKDVFKGSVKVS

DFEFLSENASPQRMHTADELFSEGKLLPFWQEKHSEKLKNVSLKTNEEEE

EEEENRKVEATMKSNDYDKNRVSWFIDEDPSPRPPKCTVLWKELLRLKKQ

RNTRSSLSPSSSTSSSSSLEEAAAKREEKEGKRGKKGLERTRSTSMRIRP

MIHVPVCTPSKSSVPLPPLFPLRLKKNRVEKRT

>Brassica rapa 3

MASAETSTMAEANMVFMMEAPPSGPRISFSADLSSSDSEGDYICINPKNL

LPGKVEQDKSSSKAGDFEFLSNTQTMLTADELFSEGKFLPFRHVKHSEKL

QNVTLKTKAEEQEQEQEKEDRKVVKEETVNNSNRGSWFLDDDPSPRPPKC

TVLWKELLRLKKQRNNTKALSLSPSSSSSSTSSSSSSIGDAVKKEEREKR

GKKGLERTRSMTMRIRPMIHVPVCTPPSKPPLFPLRLHKNKVERRT

>Brassica rapa 4

MVVAETAEATMVFTTEGPRISFSADLSSSDSEGDYICINPENLLRGKEEQ

VKAGDFEFLSNTQTMLTAADELFSEGKLLPFWQAKHSEKLQNVTLKTKVV

DVDEVEVVEEEEEDRKVVKEETVHNSTKEQENSNNRGSWFLDDDPSPRPP

NCTVLWKELLRLKKQRNTKTTNTTTKASSTKASSLSPSSSSSSTSSSSSS

IGDAVKEESEKKGKKGLERTRSVTMRIRPMIHVPVCTPSKPPLFPLRLHK

NRVEKRT

>Capsella rubella 2

MVSADTATLAEAKMVFMTEASPPSSGPRISFSADLSSSDSDGDYICINPA

NLIVGKEEKDKNFLKAGDFEFLSENVTNNQTMLTADELFCEGKLLPFWQV

KHSEKLKNVTLKTKVEVEEEDLKVVREEVVHNNKEQENNNNNNNNNNRGS

WFLDDDPSPRPPKCTVLWKELLRLKKQRTTTTTVSSTRVSSLSPSSSSSS

TSSSSSSIGDAVKKEEREKEGKRGKKGLERTRSVTMRIRPMIHVPVCTPS

KSSAPLPPLFPLRLQKNRVERRT

>Arabidopsis lyrata 2

MAEAEQSLTGPRISFSADLSSSDSDGDFICINPAMNLIVGKEEKDKTSVK

AGDFEFLSENATMLSADELFSEGKLLPFWQVKHSEKLKNVTLKTKVEVEE

EEEDQKVVKEEGIVHNNKEQENNNNNNRGSWFLDDDPSPRPPKCTVLWKE

LLRLKKQRTTTTTVSSTRVSSLSPSSSSSSTSSSSSSIGDAVKKEEREKE

GKRGKKGLERTRSVTMRIRPMIHVPVCTPSKSSARLPPLFPIRLQKNRV

>AT5G19340.1

MVSAETATMAEAEPSTTGPRISFSADLSSSDSDGDFICINPVMNLIVGRE

EKDKSSVKAGDFEFLSENATMLSADELFSEGKLLPFWQVKHSEKLKNVTL

KPKVEVQQEEEDHKVVNEEGFVHNKEQENNNNNNNNNNNRGSWFLDDDPS

PRPPKCTVLWKELLRLKKQRTTTTTTASTRVSSLSPSSSSSSTSSSSSSI

GDAVKKEEREKEGKRGKKGLERTRSVTMRIRPMIHVPVCTPSKSSSRLPP

LFPLRLQKNRVER

>Eucalyptus grandis

MVSQENIDPPFSAPRISFSADLLDESDFISINPDGHFHNQVTKAKETAAM

DLEKKPRNGEFEFLAASMSPHKMMSADELFFEGKLLPFWQMQQSQRLKRI

TLKPKSGDSEEVRDGGRDQEPDREEEEVRSNRNYCNSNNRDQDQEQNRVS

WFLDDDPSPRPPKCTVLWKELLRLKTKRRASSSLSPSSSSSSTSSSSSSL

GDVASLDERKEARDRDRDRESSTNYVQRIRKGLERTRSNSIRIRPMVNVP

ICTHVRSASGGGGGSLPHLFPLKKGRV

>Mimulus guttatus 1

MVSQEALESTCGGAATAEPTISGPRISFSTEFLDENDFISICPNRHPPEK

KPENRTTAARNGPEFEFLSGNSASNMTTADELFSEGKMLPFWQTHHQYSE

TTLNKLKTDTTTNIAAGQAAATAGAAAEQDRRISWFLDDDPSPRPPKCTV

LWKELLRLRKQRPSTLSPSSSSSSSSSGRSAIADNIPTAADEQRKIKGVA

ANINNNNKVASSSSRSTVKKGLERTRSGSNSIRIRPVVNVPICTQVKSSS

LPPLFPIRSRTKLLS

>Medicago truncatula

MVSLEPEPVQGNNLRSSDAPTSPRISFSAEFLDENNFISISPNPLYRTER

DQEKEQHEKTKNTDQFEFLSNINISDKNTVLSADELFFEGKILPFWQMQH

LEKLNKINLKEEEEEEVIEVVVDNKEDNNNSRVNWFVDDDPSPRPPKCTV

LWKELLRLKKQRASSLSPSSSSSSSSSNGSSLGDVAAKEGSKNKENQHVK

RIKKGLERTRSATIRIRPMINVPICTQMKNSALPPLFPLKKGKILER

>Glycine max 1

MVSLEPIEGNLRSSDPPSSPRISFSAEFLDENNFISISPNAEYERDQEKE

RERARNAAEFEFLSNNTSSNNTVLTADELFFEGKLLPFWQMQHLEKLSKI

NLKTKEGEEEELEEEVIVVSNNNKEDSNSNSNSRVNWFVDDDPSPRPPKC

TVLWKELLRLKKQRASSLSPSSSSSSSSSSASSLGDVAAKEGKEGSRSSN

KEQHVKRVKKGLERARSATIRIRPMINVPICTQVKSSALPPLFPLKKGKL

ER

>Phaseolus vulgaris

MVSLEPIEGNPRSSDAPSSPRISFSAEFLDENNFISISPNAVYEKDQEKE

RERTRNAAEFEFLSNNMSNNNTVVTADELFFEGKLLPFWQMQHLEKLSKI

SLKPKEGEEEEEEELEEEAVVSNNKEESSSNSSRVNWFVDDDPSPRPPKC

TVLWKELLRLKKQRASSLSPSSSSSSSSSASSLGDVAAKEGKEGSRNNN

NKEQQVKRVKKGLERTRSATIRIRPMINVPICTQVKSSALPPLFPLKKGK

LER

>Glycine max 2

MVSLEPIEGNLRSSDAPSSPRISFSAEFLDENNFISISPNAEYEGPDQEK

ERERARNAAEFEFLSNNTSNNNTVVTADELFFEGKLLPFWQMQHLEKLSK

INLKTKEGEEEEEEELEEEVVVVSNNNKEDNNNSNSSSRVNWFVDDDPSP

RPPKCTVLWKELLRLKKQRASSLSPSSSSSSSSSASSLGDVAAKEGSRS

SSNKEHQQHVKRVKKGLERTRSATIRIRPMINVPICTQVKSSALPPLFPI

KKGKLERS

>Fragaria vesca

MVSLEVVQTTSSIEPSSSPRISFSADFLDENDFITISPNAHGELQDKKME

CDQKARNADFEFLSNNVSSHTMLTADELFFEGKLLPFWQKQHAERLNKIR

LKTKDDEICEEEEEVVNKEESRGNWFVDDDPSPRPPKCTVLWKELLRLKK

QRASTLSPSSSSSSSSSSSNSFADVAAAADQVKEAMGNKEKYMKRIKKGL

ERTRSASIRIRPMVNVPICTQMKNSALPPLFPLRKGRLDR

>Malus domestica

MGSLEIVQATPRSVDMSSSPRISFSAEFLDENNFISITPNLRGEVQDKKM

EGGDHQKVRNPDFEFLSSNVSSHAMLSADELFFEGKLLPFWQKQHAERLT

KLNLNTKDVEGDENEEGVNKEESRGSWFVDDDPSPRPPKCTVLWKELLRL

KKQRASTLSPSSSSSSSSSSSNSLADIATTTDQEKEGNKEKYMKRIKKGL

ERTRSASIRIRPMXNVPICTXVKSSALPPLFPLRKGRVLER

>Prunus persica

MVSLEIVQATSRSMDTPSSPRISFSAEFLDENNFISITPNAHQGEQDLIM

ECDQKVRNPEFEFLSSNVSSHTMLSADELFFEGKLLPFWQKQHAERLSKL

SLKTKDVEGDENEEGVNKEESRGSWFVDDDPSPRPPKCTVLWKELLKLKK

QRASSLSPSSSSSSSTSSSSSFADAATADQEKEGMGNKEKYMKRIKKGLE

RTRSASIRIRPMINVPICTQVKSTSLPPLFPLRKGRLER

>Manihot esculenta 1

MVSLETVQASMDQTSSPRISFSAEFLDENNFISITPNPQDQKMEREKARN

AEFEFLSSNMSSHTMLTADELFFEGKLLPFWQMQQSDKLHKISLKGKENE

EEEEEEEEEEEEEVNKEEPRINWYLDDDPSPRPPKCTVLWKELLRLKKQR

PYLSPSSSSSSTSSSSSSLADIVTTEEGKAGSGKQGKRVKKGLERTRSTT

IRIRPMVNVPICTHVKSSSLPPLFPLKKGRLER

>Manihot esculenta 2

MVSLETVQATSRSIDQTSSPRISFSAEFLDENNFISISPNTLQPEEDHEM

EREKARNAEFEFLSGNMSSHAILTADELFFEGKLLPFWQMQQSEKLHKIS

LKSKETMEVEEEEEVNKEEPRVSWFVDDDPSPRPPKCTVLWKELLRLKKQ

RASSLSPSSSSSSTSSSSSSLADIVTTVEAKQGSGNRDKQGKRMKKGLER

TRSATIRIRPMINVPICSPVKSSPLPPLFPLKKGRLER

>Ricinus communis

MVSLEAVQATSRSIDQPSSPRISFSAEFLDENNFISINPNARAERDQEME

REKARNYAADFEFLSGNSTMSSHATMLTADELFFEGKLLPFWQMQQSEKL

HKINLKCKETEEGEEEEVEVNNKEEPRVSWFVDDDPSPRPPKCTVLWKEL

LRLKKQRASSLSPSSSSSSTSSSSSSLADIVTAEEGKEGCGNKEKHAGKR

MKKGLERTRSATIRIRPMINVPICTQVKSSALPPLFPLKKGRLER

>Populus trichocarpa 1

MVSLETVQATSRSIDQASSPRISFSAEFLDDKNFISISPSPQAEKDKETE

RERARNAEFEFLSSKMSSQTMLTADELFYEGRLLPFWQMQHSEKLNKVSL

KTKNAEEEGEVSKEEPRVWFVDDDPSPRPPKCTVLWKELLRLKKQRASSL

SPSSSSSSTSSSSSSLADIATKEEGKRGSGNGEKHVKRIKKGLERTRSAS

MRIRPMINVPICTQMKSSALPPLFPLKKGRLER

>Populus trichocarpa 2

MISLETVQATSRSIEPSSPRISFSADFLHDKNFIPISPNQQAEKDGEAER

EQARNAEFEFLSSKMSSQTMLTADELFFEGRLLPFWQMQHSEKLNKISLK

TKEAEEGEGEEMSKEEPRVWFVDEDPSPRPPKCTVLWKELLRLKKQRASS

LSPSSSSSSTSSTSSSALADIVTKEGKHGSWNREKHVKRIKKGLERTRSA

SIRIRPMINVPICTPVKSSALPPLFPLKKGRLER

>Carica papaya

MASPETLQPTSKTIDSPSSPRISFSAEFLDDNDFISITPHSPDGMIDLEM

EREKSRNAEFEFLSTSVSSHTMLTADELFFEGKLLPFWQMQHSEKLKKIS

LKTKDAEGEEEEEEEVEEEKEGISKEENRVNWFVDEDPSPRPPKCTVLWK

ELLRLKKQRASSLSPSSSSSSTSSSSSSLADAVNAEEGKNGSGNREKHVK

RIKKGLERTRSASIRIRPMVNVPICTQMKSPALPPLFPLKKGRLER

>Citrus sinensis

MVSVEIAQAAQPANRSIINEQPTSPRISFSADFLDESNFISITPQSQQHS

HQGQKDQEKARLQEKGGRNIAADPFEFLSNTSDVSSHNMLSADELFFEGK

LLPFWQMQHSLEKLNKISLKTKDCEKEEDEEEAIHINNDNHNHNKEAAAT

EARVSWFVDDDPSPRPPKCTVLWKELLRLKKQRASSLSPSSSSSSTSSSS

SSLADIVTKEDGKEGPGNRDNKHVKRIKKGLERTRSASIRIRPMVNVPIC

TAVKSSAMPPLFPLKKGRLEI

>Citrus clementina

MVSVEIAQAAQPTNRSIINEQPTSPRISFSADFLDESNFISITPQSQQHS

HQGQKDQEKARLQEKGGRNIAADPFEFLSNTSDVSSHNMLSADELFFEGK

LLPFWQMQHSLEKLNKISLKTKDCEKEEDEEEAIHINNDNHNHNKEAAAT

EARVSWFVDDDPSPRPPKCTVLWKELLRLKKQRASSLSPSSSSSSTSSSS

SSLADIVTKEDGKEGPGNRDNKHVKRIKKGLERTRSASIRIRPMVNVPIC

TAVKSSAMPPLFPLKKGRLEI

>Gossypium raimondii

MAMEAVQASSRNSMETNSSPRISFSADLLDETNFISINPHSQTDDADKDK

DKTATARVAVADFEFLSSNVSSHAMLTADELFFEGKLLPFWQMHHSEKLK

QINLRKESGGDGEGDGDDDEREVVENKEESSRVSWFVDDDPSPRPPKCTV

LWKELLRLKKQRATSSLSPSSSSSSSSSSLADVAEEGKQGSGNRDNKHV

KRIKKGLERTRSASIRIRPMINVPICTQVKSSALPPLFPLKKGRILER

>Theobroma cacao

MAPEAVQATSRTIEPTSSPRISFSADFLDENNFISINPHSQNEENGQDKG

KEAKEWEKDKARAAEFEFLSSNVSSHAMLTADELFFEGKLLPFWQMQHSE

KLNKISLKTKASEEEGEEEVNKEESRVSWFVDDDPSPRPPKCTVLWKELL

RLKKQRASSLSPSSSSSSTSSSSSSLADIATAEEGKEGSGNRDKHVKRIK

KGLERTRSASIRIRPMINVPICTQVKSSALPPLFPLKKGRLES

>Aquilegia coerulea 1

MHTLRHTMTQLANSISPPPLSSPAKMSTTMISLESVQANSRSMDTTSSPR

ISFSCDFLDDKTFISLSPSSENKKVLDTEKDKGCNIDFEFLSTDSATNTM

LTADELFSEGKLLPFWQKQHVDRLNKINLKPKMEDEEEEKETSKEETNRV

SWFIDEDPSPRPPKCTVLWKELLRLKKQRASSLSPSSSTSSSSSSSDCNV

QTATIEAAGKGSKESIWNREKNVKRIKQTLERTRSASIRIRPVVNVPLST

QGKSSGLPPLFSLRKGSVDR

>Aquilegia coerulea 2

MVSLENVSIRSVEPIISSPSRISSSTDIFTDKKKIKTKSETSKSHGKDKK

KLRNVEFEFLSANFSTNTMSTADELFFEGKLRPFSQVEQLEELNKITLKP

KENDEEEEQKEGTRVSWFMDEDPSPRPPTCTVLWKELLKLKKQRSSPPLP

PSTSSSSRSSCSSSVVRMESIDEGKEGKEGLWSKEKHVKRIKKGLERTRS

GSFRIRPMVNVPICTQSKSTTAMPSMFSHKKVNVER

>Solanum lycopersicum 1

MMSLETASTSVDPNSGPRISFSSEFLDEKNFISICPNSQPEKKREKELNA

AEFEFLSSNFTNGNMTTADELIFEGKLLPYWQIHHAEKLNKISLKTEHVE

EQVNEKQGSSKEEQSRPVNWFIDEDPSPRPPTCTVLWKELLRLKKQKQRP

SSLSPSSSSSSSSSSSANSEILHTDESREKHVVDKIKKRLERSKSATIR

VRPLINVPICRQGKNSAIPPIFPIKKGRVER

>Solanum tuberosum 2

MMSLETASRSVDPNSGPRISFSSEFLDEKNFISICPNSQPEKKREKELNA

AEFEFLSSNFTTGNMTTADELIFEGKLLPYWQIHHAEKLNKISLKTEHAE

EQVNEKQGNSKEEQSRPVNWFIDEDPSPRPPTCTVLWKELLRLKKQKQRP

SSLSPSSSSSSSSSSANSEISPTDESKEKHVVDKIKKRLERSKSATIR

VRPLINVPICRQGKNNAIPPIFPIKKGRVER

>Solanum lycopersicum 2

MVSLEGTLISEEPTSSPRISFSSEFLDERNFISITPNAQEEKERKDQQDR

STRSAAEFEFLSSKLTNENMITADELFFEGKLRPYWQMRYAEKLNKINLK

ADDEILNNTTVIKSKEETTTRPINWFIDEDPSPRPPKCTVLWKELLRLKQ

KRASSLSPSSSTSSSSSSSSISFADKEKGKGQSMKEKHVKRIKKGLERCK

SETLRVRPVIHVPICSQGKNSALPPLFSLKKKGRAIER

>Solanum tuberosum 1

MVSLEGPLISDEPTSSPRISFSSEFLDERNFISITPNAQAEKERKEQQDR

STRSAAEFEFLSSKLTNENMITADELFFEGKLRPYWQMHYAEKLNKISLK

ADKEILDNAMVIKSKEETASRPINWFIDEDPSPRPPKCTVLWKELLRLKH

KRASSLSPSSSTSSSSSSSSSLAENEKSKGQSIKEKHVKRIKKGSLERCK

SETLKVRPVIHVPICSQGKNSALPPLFSLKKKVRAIER

>Lus10041215

MVSLETIQASTDHQQQTSSPRISFSAEFLDDNNNFTSVRTSDSDPKSLKP

PLTTIRLAPPAAGAVIIRSKESEQQQPSTGTAAMGMNWFVDDDPSPRPPK

CTVLWKELLRLKKQRPPVASSLSPSSSSSSSSSSSLPDVAERESNGGKD

RKEGKKKGLERTRSATLRIRPMINVPICSQMKTTTTTHHSSLPPFFPVKK

GRALDTR

>Lus10002455

MKETLLSMETVQAPSRSTTIDQISSPRISFSAEFLDDNNHFISITPTHLI

DNPDTNNSQKPPQSPTRNGAAVGDNQFEFLSSGGEPKSGHARMLTADELF

FEGKLLPFWQMQQSERLNKISLKSKEEGDETILRKEDPLPPPTPTTAAAM

NWFVDDDPSPRPPKCTVLWKELLRLKKQRPSVSSLSPSSSSSSTSSCSSS

LGDAATKEESGKGEKDNNKESNKKGNHQQQVKRAKKGLERTRSSSIRIRP

MINVPICTQMKSSHHHSALPPFFPLKKGRVLER

>Lus10010529

MNWFLDDDPSPRPPKCTVLWKELLRLKKQRPSVSSLSPSSSSSSTSSSSS

SLGDAATKEESGKGEKDNKKGNHQQQVKRAKKGLERTRSSSIRIRPMINV

PICSQMKSSSSHHHHHSALPPFFPLKKGRVLER

>Cucumis sativus

MVSIGSGGGSSVQASPPPSSPLPATEPNSSPRISFSSEFLDESNFISITP

NSQIERDQEICERQKKDRSEKLAWSADFEFLSNKVSSHSMITADELFFEG

KLLPFWQMQQAERLNKISLKSPKDVDEEDLVEIEVNKEAENKVNWFLDDD

PSPRPPKCTVLWKELLRLKKQRASSALSPSSSSSSSSSSSRSMADAATTE

EGKEGTTGNKEKNVKRIKKLERTRSASIRIRPMINVPICTQVKSSVLPPL

FPLKKGRFDR

>Salix purpurea 1

MVSLETVQAPSISVDQPSSPRISFSADFLDDKNFISISPNPQAEKDKETE

REKSRNAEFEFLSSKMSSQAMLTADELFYEGRLLPFWQTQHSEKLNKISL

KSKKAEEEEVIKEEPRIWFVDDDPSPRPPKCTVLWKELLRLKKQRASSLS

PSSSSSSTSSSSSSLADIVVTKEGKRGSGNGEKHVKRIKKGLERTRSASM

RIRPMINVPICTQVRSSGLPPLFPLKKGRLER

>Salix purpurea 2

MVSLEIVQATARSIEPPSSPRISLSADFLDDKNLVSMSPIPQAEKDREAE

REKARNAEFEFLSSKMSSQIMLTADELFFEGRLLPFWQMQHSEKLNKISL

KTKEAEEVIKEEPRVWFVDDDPSPRPPNCIVLWKELLRLKKQRASSLSPS

SSSSSTSSSSSSLSDIVAEEEGKRGSRNVEKHVKRIKKGLERTRTASIRI

RPMINVPICTPVKSRALPPLFPLTKGRLERWRLDGENRM

>Boechera stricta 1

MVSETESPPLLGPRISFSADLSDGGDFICISPAMCKELEKDGVKGSVKVS

DFEFLSENVSPQKMLTADELFSEGKLLPFRQVTNSEKLKNITLKTNEEEE

NRKVEVMKKDQEINNNNNRVSWFIDEDPSPRPPKCTVLWKELLRLKKQRN

PPTSSVRTVSSLSPSSSTSSSSSLEDAAKREEREKEGKRGKKGLERTRSA

SMRIRPMIHVPICTPSKSSLPLPPLFPLALKKNRVERRT

>Boechera stricta 2

MVSAETATMAEAKMVFMTEASPPSSGPRISFSADLSSSDSDGDFICINPV

NLIVGKEEKDKTLVKAGDFEFLSENVTNNQTMLTADELFCEGKLLPFWQV

KHSEKLKNVTLKTKVEVEVEEEEEDHKVVIDEVVHNNKDQENNNNNNNNR

GSWFLDDDPSPRPPKCTVLWKELLRLKKQRTTTTAVSSTRVSSLSPSSSS

SSTSSSSSSIGDAVKKEEREKEGKRGKKGLERTRSVTMRIRPMIHVPVCT

PSKSSAPLPPLFPLRLHKNRVERRT

>Capsella grandiflora 1

MVSETVSKTESPPLIGPRISFSADLSDDGDFICISPVMCKELEKDVVLKG

SVKVSDFEFLSENVSPQKMLTADELFSEGKLLPYWQVKHSEKLKNITLKT

NEEEEENRKAEVMKKDQEITSNNRVSWFIDEDPSPRPPKCTVLWKELLRL

KKQRNPSSSSVTVRTVSSLSPSSSTSSSSSLEDAAKREEREKEWKRGKKG

LERTRSASMRIRPMIHVPICTPSKSSLPLPPLFPLALKKNRVERRA

>Capsella grandiflora 2

MVFMTEASPPSSGPRISFSADLSSSDSDGDYICINPANLIVGKEEKDKNF

LKAGDFEFLSENVTSKQTMLTADELFCEGKLLPFWQVVKHSEKLKNVTLK

TKVEVEEEDLKVVREEVVHNNKEQENNNNNNNRGSWFLDDDPSPRPPKCT

VLWKELLRLKKQRTTTTTVSSTRVSSLSPSSSSSSTSSSSSSIGDAVKKE

EREKEGKRGKKGLERTRSVTMRIRPMIHVPVCTPSKSSAPLPPLFPLRLQ

KNRVERRT

>Eutrema salsugineum 1

MVSEAISQTESPPLIGPRISFSADLSDGGDFICITPAMCKELEKDVVKGS

VKVADFEFLSENVSPQRMLTADELFSEGKLLPFWQVKHSEKLKNVNLKTN

EEEVEEENRKVEVTMKNKDQENNNNNNNNRVSWFIDEDPSPRPPKCTVLW

KELLRLKKQRNSSSSSSVRTVSSLSPSSSTSSSSSLEDAAAKREEREKEG

KRGKKGLERTRSASMRIRPMIHVPICTPSKSSVPLPPLFPLALKKNRVER

RT

>Eutrema salsugineum 2

MVSAETATMAEAKMIFMTEAPPSLSGPRISFSADLSSSDSDGDFICINPD

KLVSGKEEKDKSSVKAGDFEFLSNTQTMLTPDELFSEGKLLPFWQVKHSE

MLQNVTLKTKVDEDKEDRKEVKEQVINNNKEQENNNNRGSWFLDDDPSPR

PPKCTVLWKELLRLKKQRTNTTTNSSTRASSLSPSSSSSSTSSSSSSIGD

AVKKEEREKEGKRGNKGLERTRSVTMRIRPMIHVPVCTPSKSSAPLPPLF

PLRLQKNRVERRT

>Kalanchoe laxiflora 1

MVSMEVDQEQKAETCKSPDNSISSPRISFSCDLLDDANFISINLAPIKTD

DDDGKKQSTGAKARNPEFEFLAGSRTQPDMPTADELFFEGKLRPYWQTHH

SEKLKSLSLKQEIQEVAAADETATVAAAKDESRVRTWFIDDDPSPRPPKC

TVLWKELLRLKTRQRASSLSPSSSSSSSSSSSLDMSKEKEKEKNRDSNA

GKNVKRVRKGVERTRSASIRIRPMVNVPICTQSAKQSALPPLFPLKKGR

>Kalanchoe laxiflora 2

MVAMEVEEEQRAAAGKLPDNSISSPRISFSCDLLDDANFISINLAPIKTD

DEAQKQTTASAKSRNPPDFEFLAHSRTQTDMPTADELFFEGKLLPYWQTH

HSDKLRSLSLKSQQEIQEEVAEDVAVAVVSAAASKDESRVRTWFIDDDPS

PRPPKCTVLWKELLRLKTRQRASSLSPSSSSSSSSSSSLDMSKEKEKEK

HRESNAGSSGKNVKRVRKGLERTRSASIRIRPMVNVPICTQSALPPLFPL

KRGR

>Kalanchoe laxiflora 3

MVAMEVEEEQQAAACKSPDNSVSSPRISFSCDLLDDANFISINLAPIKTD

DEAQKQTTASAKSRNPPDFEFLAHSRTQPDMPTADELFFEGKLLPYWQTH

HSDKLRSLSLKSQQEIQEEVAAAAEDVAVAVASAAASKDESRVRTWFIDD

DPSPRPPKCTVLWKELLRLKTRQRASSLSPSSSSSSSSSSSLDMSKEKE

KEKRRDSNAGSSGKNVKRVRKGLERTRSASIRIRPMVNVPICTQSAFPPL

FPLKRGR

>Kalanchoe laxiflora 4

MVSMEVDQEQKAETCKSPDNSISSPRISFSCDLLDDANFISINLAPIKTD

DDDGKKQSTGAKARNPEFEFLAGSRTQPDMPTADELFFEGKLRPYWQTHH

SEKLKSLSLKQEIQEVAAADDAATAAAAKDESRVRTWFIDDDPSPRPPKC

TVLWKELLRLKTRQRASSLSPSSSSSSSSSSSLDMSKEKEKEKEKEKEK

DSNAGKNVKRVRKGVERTRSASIRIRPMVNVPICTQSAKQSALPPLFPLK

KGR

>Vitis vinifera 1

MVSLEAVQASSRSIEPTVSPRISFSSDFLDEKNFISISPNSEKEKQHEMD

QEKARNTDFEFLSSNSTSHTMLTADELFFEGKLLPFWQRQHSEKLNKMSL

KTKNDEEQEEEEANKEESRVSWFVDDDPSPRPPKCTVLWKELLRLKKQRA

STLSPSSSSSSSSSSSLVDMGTMDQGKEGSGKREKQVKRIKKGLERTRS

ASIRIRPVINVPICTQGKASLLPPLFPLKKGRLER

>Arabidopsis halleri 1

MVSETVSNTESPPLLGPRISFSADLSDGGDFICITPVMCKELEKEVVKGS

VKVSDFEFLSSENVSPQRMLTADELFSEGKLLPFWQVKHSEKLKNITLKT

NEEEEGEKRKVEVMKKDQEINNRDNRVSWFIDEDPSPRPPKCTVLWKELL

RLKKQRNPSSSSVAVRTVSSLSPSSSTSSSSSLEDAAKREEKEKEGKRGK

KGLERTRSASMRIRPMIHVPICTPSKSSLPLPPLFPLALKKNRVERRT

>Arabidopsis halleri 2

MVSAETATMAEAEQSLTGPRISFSADLSSSDSDGDFICINPVMNFIVGKE

EKDKTSVKAGDFEFLSENATMLSADELFSEGKLLPFWQVKHSEKLKNVTL

KTKVEVEEEEEDQKVVKEDGLVHNNKDQENNNNNNRGSWFLDDDPSPRPP

KCTVLWKELLRLKKQRTTTTTVSSTRVSSLSPSSSSSSTSSSSSSIGDAV

KKEEREKEGKRGKKGLERTRSVTMRIRPMIHVPVCTPSKSSARLPPLFPI

RLQKNRVERRT

>Amaranthus hypochondriacus 1

MVGEASSAISSPRISFSADFLDDDSFISISPSSSIDKDHEINQLEREMVK

NGADFEFLSSKNSLDAGHSTMLTADELFFEGKLLPYWQINHAAEKLSKLN

LKSHQQSEDKKITQNKNDGKPHHIVEVSKNGNNGSGILLGREVQEPRIWF

VDDDPSPRPPKCTVLWKELLRLKKQRSSTLSPSSSSSSSSSSSSLGDAA

AMEEKEKEKEKEKEKGMSTREKHIRRLKKGLERTKSANIRIRPMFNVPIC

TQTGKSSSLPPLFPLRQNKVDR

>Amaranthus hypochondriacus 2

MVETSHSSDDHNVNDEIQVEEDETVKNGGDFEFLSSFDTSHSTMLTADEL

FFEGKLLPYSQSQINLKSSDQSDHDQDNQNHDGILFGREIQEPRIWFVDD

DPSPRPPKCTLLFKELLRLNKRRSSVLSSSSSSSSSSSSSSSSSSSMDE

KEKEKEKHVIIKRIKKGLERTKSANIRIRPMFNVPISTQNGKSTSLPPLF

PLTQKKP

>Brassica oleracea capitata 1

MADANMLFMMESPPSGPRISFSADLSSSDSEGDYICINPKNLLPGKQEQD

KSSSKAGDFEFLSNTQTMLTADELFSEGKFLPFRHVKHSEKLQNVTLKTK

AEEQEQEKEDGEVVKEETVNNSNRGSWFLDDDPSPRPPKCTVLWKELLRL

KKQRNNAKASSLSPSSSSSSTSSSSSSIGDAVKKEEREKRGKKGLERTRS

LTMRIRPMIHVPVCTPPSKPPLFPLRLHTTKVERRT

>Brassica oleracea capitata 2

MVVAETAEATMVFTTEGPRISFSADLSSSDSEGDYICINPENLLRGKEEQ

VKAGDFEFLSNTQTMLTAADELFSEGKLLPFWQAKHSEKLQNVTLKTKVV

DVDEVEVVEEEEEEEEDRRVKKEETVHNSTKEQENNNRGSWFLDDDPSPR

PPNCTVLWKELLRLKKQRNTKTTNTTTKASSTKASSLSPSSSSSSTSSSS

SSIGDAVKEESEKKGKKGLERTRSVTMRIRPMIHVPVCTPTKPPLFPLRL

HKNRVEKRT

>Daucus carota 1

MVSPEKSQTDSASAEPISSPRISFSSDFLDETNFIPSIKTSQVEKEPEKP

REKTFEFLSSNNHTMLPADELFFEGKLLPYWQMHHEIKKITLRSEEGPKA

KSKVEDLNLSKESRGSWFIDDDPSPRPPSCTVLWKELLRLRKQRPSTLSP

SSSSSSSSSSSLVDNQGTDKEDRAGNKDKNAKKSKKGLERTRSATMRIR

PVINVPLCTQAKNSALPPLFSFKKGKLEKLNSQK

>Daucus carota 2

MVSSETLQTNATTIEPNSSPRISFSSDFLDNNFISSINISPVEKEHENKR

EKTFEFLSTDSQTMLSADELFSEGKLLPYRPMHHEIKKITLKSDDGSKAK

AKAEDSNKESRGSWFVDDDPSPRPPTCTVLWRELLRLKKHRPSTLSPSSS

SSSSSSSSLVDSQGTNKEEKSGNKEKHVKKTKKGLERTRSATMRIRPMIN

VPICTQRSNSALPPLFSFKKGKLEKLK

>Daucus carota 3

MISLETLQATSRSINPISSPRISFSSNSLDDDDFISINPNSMAVKEKTRN

VEFEFLSSENQTMLSADELFSEVWQMQQPEKLKTMSLNAEQQAEAGRAED

RSKAETKVGWLLDDDPSPRPPKCNVLWKELVRLRKQRSSTLSPSSSSSSS

SLKSLDLRSIEERKQGSGSKDKHVKRMKKGLERSRSTSMRIRPMVNVPVC

THGRRNAVPPLLSFRKEKPEK

>Kalanchoe fedtschenkoi 1

MVSMEVDQEQKAETCKSPDNSISSPRISFSCDLLDDANFISINLAPIKTD

DDDGKKQSTGAKARNPEFEFLAGSRTQPDMPTADELFFEGKLRPYWQTHH

SEKLKSLSLKQEIQEVAAADEAATAAAAKDETRVRTWFIDDDPSPRPPKC

TVLWKELLRLKTRQRASSLSPSSSSSSSSSSSLDMSKEKEKEKEKEKNR

DSNAGKNVKRVRKGVERTRSASIRIRPMVNVPICTQSAKQSALPPLFPLK

KGR

>Kalanchoe fedtschenkoi 2

MVAMEVEEEQRAAAGKSPDNSISSPRISFSCDLLDDANFISINLAPIKTD

DEAQKQTTASAKSRNPPDFEFLAHSRTQTDMPTADELFFEGKLLPYWQTH

HSDKLRSLSLKSQQEIQEEVAEDVAVAVVSAAASKDESRVRTWFIDDDPS

PRPPKCTVLWKELLRLKTRQRASSLSPSSSSSSSSSSSLDMSKEKEKEK

HRESNAGSSGKNVKRVRKGLERTRSASIRIRPMVNVPICTQSALPPLFPL

KRGR

>Trifolium pratense 1

MVSLEHEHEPVQGNLRSSDAPTSPRISFSAEFLDDNNFISICPNPLYSER

DQEKEQHEKTKNITDQFEFLSNNNMSNNNTVLSADELFFDGKILPFWQMQ

HLEKLNKINIKEEQHEEVEEVIEVVVNSNKEDNSNNNSRVNWFVDDDPSP

RPPKCTVLWKELLRLKKQRASSLSPSSSSSSSSNASSLGDVAAKEGSRN

KENQHVKRIKKGLERTRSATIRIRPMINVPICTQMKNSSLPPLFPLKKGK

ILER