Functional characterization of two plant type I MADS-box genes in *Arabidopsis thaliana* – *AGL40* and *AGL62*

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

Master of Science in Plant Biology

At Massey University, Palmerston North, New Zealand

Ryohei Kaji

2008

Abstract

MADS-box transcription factors (TF) are a family of evolutionary conserved genes found across various eukaryotic species. Characterized by the conserved DNA binding MADS-box domain, MADS-box TF has been shown to play various roles in developmental processes. MADS-box genes can be based on MADS-box structural motifs divided into type I and type II lineages. In plants very limited functional characterization have been achieved with type I genes MADS-box genes.

In this project we attempted to functionally characterize 2 closely related members of the type I lineage MADS-box genes *AGL40* and *AGL62* and give further support to the hypothesis that plant type I MADS-box genes are also crucial to normal plant development. Based on our expression domain characterization assay using AGL62: GUS fusion construct, we have shown expression of *AGL62* in various tissues but especially strong in developing seeds, pollen and seedling roots and shoots. The web based microarray data suggesting that AGL62 may have a function in seed, pollen and seedling development backed up this result.

Interestingly when we carried out PCR based genotyping with segregating population of heterozygous *AGL62* T-DNA insertion lines (*agl62/+*) to identify the homozygous T-DNA insertion lines we detected no homozygous T-DNA insertion line indicating loss-of-function of AGL62 may be lethal to plant.

With reference to the *AGL62* expression in pollen, seed and seedling root and shoot, we carried out phenotypic assay on each of these tissues in *agl62/+* background to investigate whether there was any phenotypic defect observed. Significant reduction in number of seeds was observed in *agl62/+* indicating possible role of AGL62 in seed development. Our microscopic observation of seeds from *agl62/+* plants showed defective embryos and confirmed that AGL62 plays a role in seed development.

Our data on AGL62 is the first report that confirms AGL62's involvement in plant development and can be a ground work for further works on functional characterization of other members of plant type I MADS-box genes.

Acknowledgements

First and foremost I would like to thank my supervisor Dr Barbara Ambrose for her help, support and encouragements. I especially appreciate your patience with me for the times when we encountered a problem. Without your support I would not have been able to finish this thesis. I have learnt so many things from your advices both in and out of the laboratory. I will always be grateful for the times I spent in MADS-house laboratory.

I would also like to say thanks to the people in MADS-house laboratory both present and past. Thank you very much Kalika for your skills, knowledge and your dedication. You have taught me many things including discipline. Thank you Arti for being very nice to me and being my friend in the lab. I enjoyed working with you. I thank Xiuwen for her technical supports. You have been very nice teacher and taught me many skills that I would have otherwise not learned. Thank you very much Emilio for taking care of my plants all the time. Your support greatly helped finish my experimental part of the project. I thank all of you again for making the times I spent at MADS-house such an exciting and enjoyable moments.

To my family, mom, dad and my brother, thanks for all the patience and encouragement that you gave me. I really appreciate your support. Finally to my brothers and sisters in church, thanks for all the love, support and discipline you gave me. You guys have been really close friends for all these times and I am looking forward to spending more time with you guys. My special thanks to Gary for disciplining me all these times and leading me. I really enjoy having time together with you.

Table of contents

Chapter 1: Introduction

Chapter 2: Materials and Methods

Chapter 3: AGL62::GUS temporal & spatial expression pattern

Abbreviations

AGL	AGAMOUS-LIKE X
Agrobacterium	*Agrobacterium tumefaciens*
Arabidopsis	*Arabidopsis thaliana*
bp	base pair
°C	degrees Celsius
cDNA	complementary DNA
DAF	day(s) after flowering
DAP	days(s) after plating
Da	Dalton
DNA	deoxyribonucleic acid
dNTP	deoxy-nucleotide-triphosphate
E. coli	*Eschrchia coli*
EDTA	ethylene diamine tetra acetate
GFP	green florescent protein
GUS	β-glucuronidase
Kb	kilo base
L	liter
LB	Luria-Bertani
MOPS	3-(N-morpholino) propanesulfonic acid
MQ	Milli-Q
MS	Murashige & Skoog
mg	milligram
min	minute
ml	milliliter
mmol	millimol
ng	nanogram
nmol	nanomol
PCR	polymerase chain reaction
pmol	picomol

RNA	ribonucleic acid
Rpm	revolutions per minute
RT-PCR	reverse transcriptase PCR
T1	first generation after transformation
T-DNA	transfer DNA
TE	Tris-EDTA
Tris	tris (hydroxymethyl) aminomethane
µg	microgram
µl	microliter
µm	micrometer
µmol	micromole
UV	ultraviolet
V	volts
WT	wild type
X-gluc	5-bromo 4-chloro 3-indolyl glucuronide

List of figures

List of tables

Chapter 1 Introduction

1.1 MADS-box genes

MADS-box genes are an evolutionary conserved gene family found in a range of organisms from yeast to humans to plants and have a role in regulation of various developments (Baum, 1998; Vergara-Silva et al., 2000; Kofuji et al., 2003). MADS-box genes code for transcription factors (TF) and named so after an acronym from *MINICHROMOSOME MAINTENANCE 1* (*MCM1*), *AGAMOUS* (*AG*), *DEFICIENS* (*DEF*) and *SERUM RESPONSE FACTOR* (*SRF*) which are the first genes identified in the family (De Bodt et al., 2003b; De Bodt et al., 2003a). Changes in expression pattern of TF genes usually result in changes of expression pattern of downstream genes and may cause global gene switch in some cases. MADS-box genes are shown to have important roles in terms of evolution of body plans as well as various other developmental processes (Becker and Theissen, 2003).

MADS-box genes can be divided into 2 types – type I and type II – based on their structural motifs of the conserved DNA binding MADS-box domain (Figure 1.1). The conserved MADS-domain characteristic of MADS-box genes is known to bind DNA via the CarG box sequence (CC(A/T)6GG) (Johansen et al., 2002; De Bodt et al., 2003b). In type II, the approximately 58 amino acid long MADS-box domains have a structural motif similar to that of *MEF2-like* genes in animals (Figure 1.1). In plants, type II MADS-box genes have been extensively studied which showed they play various important roles in plant development (De Bodt et al., 2003a). Type I MADS-box genes have the SRF-like MADS-box domain at their N-terminus (Figure 1.1). In plants, little is known about the role type I MADS-box genes play in development compared to well studied type II lineage (De Bodt et al., 2003a). Both type I and II MADS-box genes are present in fungal, animal and plant kingdom and phylogenetic analysis of MADS-box genes in all the 3 kingdoms showed that MADS-box genes were already present in last common ancestor of animal, fungal and plant species (Martinez-Castilla and Alvarez-Buylla, 2004). Animal/Fungal MADS-box genes also have either SRF-like MADS-box domain or MEF2-like MADS-box domain respectively and have been reported to play important roles in developmental processes (De Bodt et al., 2003a). (Figure 1.1).

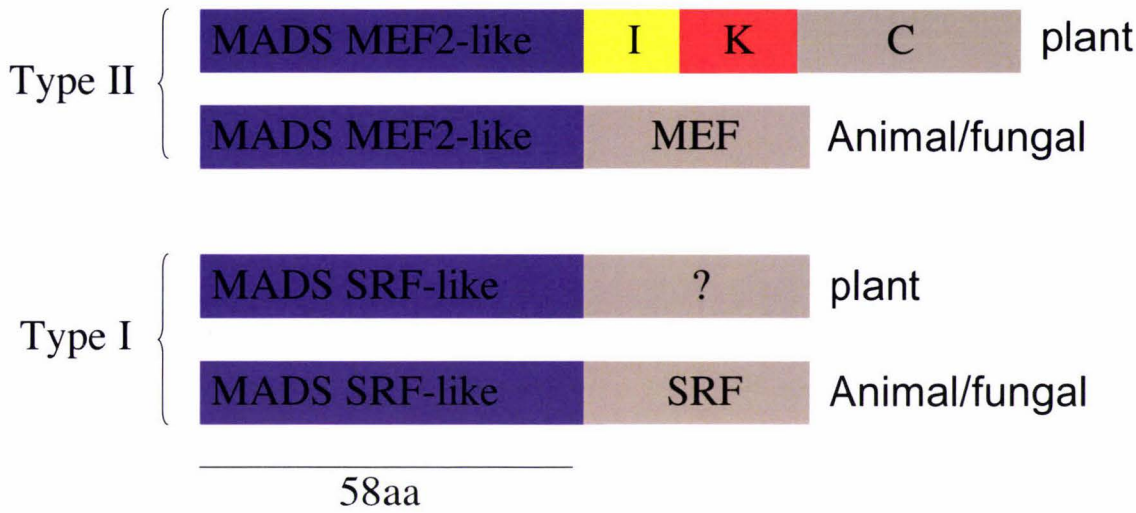

Figure 1.1 MADS-box genes in fungal, animal and plant species. Both type I and type II lineages were identified in all 3 kingdoms. Type I lineages are defined by their conserved SRF like MADS-box DNA binding domain while type II lineages are defined by their MEF-2 like MADS-box domain. All MADS-box genes also have less conserved C-terminal domain. In plants, function of C-terminal domain of type I MADS-box genes are yet unidentified and therefore denoted here as? There are extra I and K domains between the MADS-box domain and C-terminal domain in plant type II MADS-box genes.

1.2 *Arabidopsis thaliana* as plant model organism

Arabidopsis thaliana is a small flowering plant that belongs to a Brassicaceae (mustard) family (Figure 1.2), which includes many of the cultivated species such as cabbage. Although *Arabidopsis* has no major agronomic value, it has been used as a plant model organism for more than a decade for several reasons. *Arabidopsis* has a small size and relatively short life cycle of approximately 6weeks making it easy for researchers to culture large number of them in restricted space (http://www.arabidopsis.org/). It is also amenable to transformation by *Agrobacterium* resulting in production of many mutant lines (Clough and Bent, 1998). In 2000, whole *Arabidopsis* genome has been sequenced leading to identification of many new genes (Arabidopsis Genome, 2000). Currently, it is thought 125Mb *Arabidopsis* genome contains approximately 25000 genes, of which 5% encodes a transcription factor (TF) (Arabidopsis Genome, 2000). There are 11 TF family present in *Arabidopsis* and one of the family MADS-box family is thought to be homologous to animal Homeo box genes in terms of their functional roles in development (De Bodt et al., 2003a). Since the genome sequencing in 2000, 108 MADS-

box genes have been identified in *Arabidopsis* (Martinez-Castilla and Alvarez-Buylla, 2004). Examples of MADS-box genes include *AGAMOUS* (*AG*) and *APETALLA1* (*AP1*) which are the floral organ identity genes (Becker and Theissen, 2003). In *Arabidopsis*, all functionally uncharacterized MADS-box genes are named *AGAMOUS-LIKE* (*AGL*) X after *AG* –the first plant MADS-box genes to have its function characterized (De Bodt et al., 2003a). Since there are much less MADS-box genes present in animal and fungal species, much of the information about function of MADS-box genes so far came from studies in *Arabidopsis* (De Bodt et al., 2003a). Coupled with its genome sequence data, availability to many mutant lines and amenability to many biochemical and molecular biological assays, *Arabidopsis thaliana* therefore works as an ideal plant organism to be studying function of MADS-box genes.

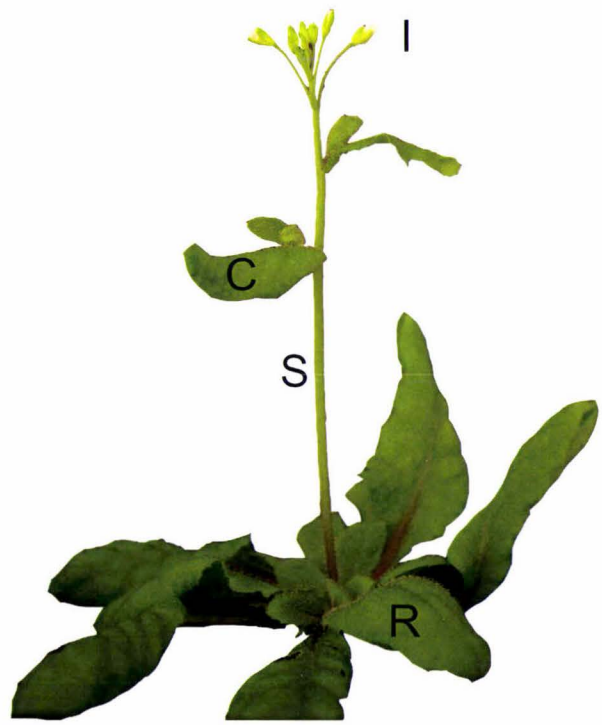

Figure 1.2 Basic body plan of *Arabidopsis thaliana*. 3 weeks old *Arabidopsis* Columbia Wild Type (WT). Arabidopsis plant is a small plant but consists of all major plant organs including root (not shown), rosetta leaves (R), stem (S), cauline leaves (C), and inflorescence (I).

1.3 Type II plant MADS-box genes

Plant type II MADS-box genes have been extensively studied and found to be involved in various aspects of plant development. (Johansen et al., 2002; De Bodt et al., 2003b). In plant type II MADS-box genes, there are additional C-terminal domains including K, I, and C–domains following the conserved N-terminal MADS-box domain (fig 1.1). The K (keratin-like) domain is well conserved between plant type II MADS-box genes and consists of approximately 80 amino acids that form a hydrophobic coiled-coil structure involved in protein – protein interaction to form dimer or multimer with other TFs (Yang et al., 2003; Yang and Jack, 2004). The I (intervening) domain is less conserved compared to MADS or K domain and is involved in specification of protein-protein interaction (Yang and Jack, 2004). The least conserved C-terminal domain is thought to be involved in Transactivation of target genes similar to other clades of MADS-box genes (Yang and Jack, 2004). Because of these MADS-box, I, K and C domains, plant type II MADS-box genes are also known as MIKC type MADS-box genes (Martinez-Castillo and Alvarez-Buylla, 2004). The I, K and C domains are only present in plant type II MADS-box genes indicating I, K and C domains have developed specifically in plant type II lineages after diversion of type I and type II lineages in plants (Figure 1.1) (Martinez-Castilla and Alvarez-Buylla, 2004).

1.3.1 Plant type II MADS-box genes and flower development

Type II MADS-box genes are particularly important in flower development although not restricted to it (Irish, 2003). Currently, there are over 250,000 known angiosperm species (flowering plants) present all over the world and the flowers come in a variety of sizes, shapes, and colors (Irish, 2003). However, despite these diversities in floral appearance, their basic organization remains the same in many flowers (Irish, 2003). Flowers arise from a floral meristem that gives rise to 4 floral organs – sepal, petal, stamen (male reproductive organ) and carpel (female reproductive organ). These floral organs arise in an orderly manner around the floral meristem axis (Lawton-Rauh et al., 2000). Sepals are formed first followed by petals and then stamens then carpels (Lawton-Rauh et al., 2000). Analysis of homeotic floral mutants which had floral organ identity miss-specification phenotype led to the identification of several type II plant MADS-box genes that are classified into 3 subclasses – A, B, and C class genes. *APETALA1* (*AP1*) represents A

class gene, B class genes include *PISTILLATA (PI)* and *APETALA3 (AP3)* and *AGAMOUS* (*AG*) represents C class gene. This in turn led to the formulation of a genetic model termed the ABC model that explains how a combination of A, B, and C classes of type II MADS-box genes are involved in specification of identity of 4 whorl organs of the flower – sepal, petal, stamen, and carpel (Figure 1.3) (Pelaz et al., 2001; Irish, 2003). In the ABC model, A class genes alone specify sepals, A and B class genes together specify petals, B and C class genes act together to specify stamens and C class genes alone specify carpels (fig. 1.3) (Purugganan et al., 1995; Lawton-Rauh et al., 2000; Pelaz et al., 2001). These ABC class genes are mutually exclusive and are restricting expression of each other in wild type (WT) to regulate their expression domains (Purugganan et al., 1995). Therefore, when one class of genes are knocked out in floral homeotic mutants, the other class of genes are over-expressed and result in conversion of one organ to another (homeotic mutation) (Figure 1.3).

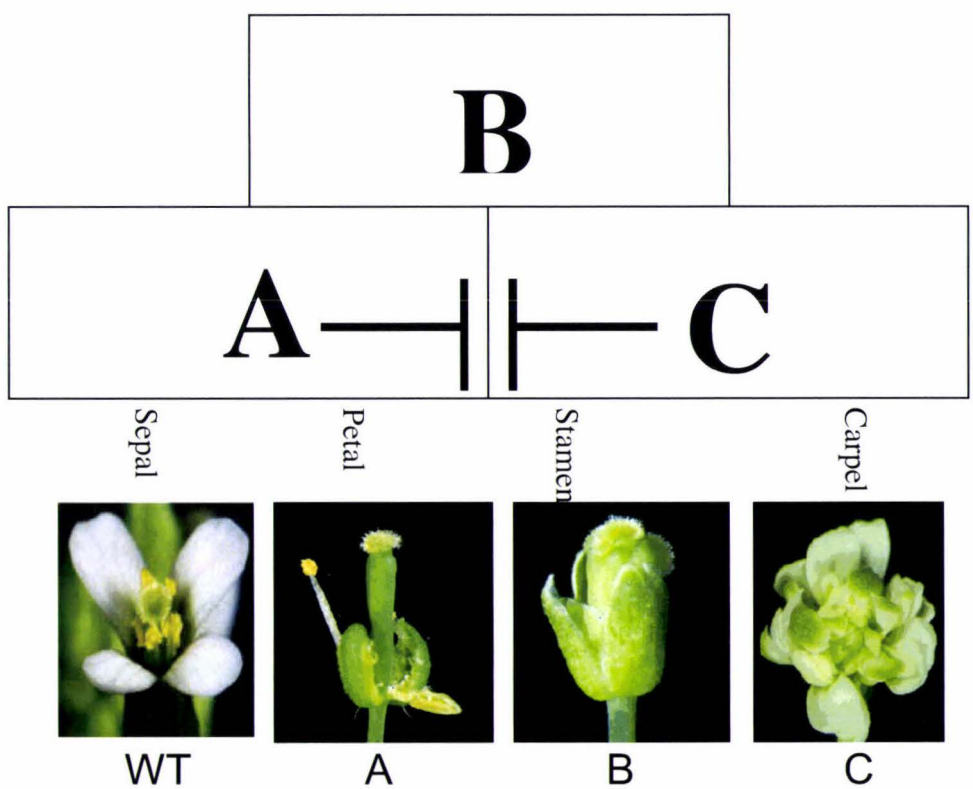

Figure 1.3 ABC floral organ identity model. Photos taken from www.Arabidopsis.org/ In ABC floral organ identity model, A class gene alone specify sepal, A and B class genes together specify petal, B and C class genes specify stamen and C class gene alone specify carpel (Top diagram). WT flowers have all 4 floral organs –petal, sepal, stamen and carpel (Bottom Figure WT). On the other hand, A class mutants lack sepal and petal (Figure A), B class mutant lack petal and stamen (Figure B) and C class mutant lacks stamen and carpel (Figure C)

1.3.2 Plant type II MADS-box genes outside the flower

In addition to floral organ identity, type II MADS-box genes have regulatory roles in flowering time of plants (Rounsley et al., 1995). The flowering pathway in plant requires quite a complex network of gene interactions. Several plant types II MADS-box genes have been isolated whose mutations resulted in altered flowering time. These genes include *SOC (SUPRESSOR OF CONSTANS), AGL24 (AGAMOUS-LIKE 24)* and *FLC (FLOWERING LOCUS C)* (Michaels et al., 2003). Isolation of such genes gave strong evidence that type II MADS-box genes are involved in control of flowering time as well.

The importance of type II MADS-box genes in plant development is not restricted to floral organ specification (Rounsley et al., 1995; Alvarez-Buylla, 2001; Ng and Yanofsky, 2001). As well as there regulatory role in reproductive organ development, MADS-box genes play important roles in development of vegetative tissues such as roots, stems, and leaves (Alvarez-Buylla, 2001). Hence, the MADS-box gene family has diverse roles throughout plant development. For example, a PCR based cloning approach resulted in the identification of *AGL16 (AGAMOUS-LIKE 16), AGL18 (AGAMOUS-LIKE 18)*, and *AGL19 (AGAMOUS-LIKE19)* and yeast-2-hybrid system identified *AGL27 (AGAMOUS-LIKE 27)* and *AGL31 (AGAMOUS-LIKE 31)* that are closely related to *FLC* (Alvarez-Buylla, 2001). RT-PCR showed that *AGL16* is highly expressed in rosette leaves and moderately expressed in roots and stems (Alvarez-Buylla, 2001). Further, *in situ* hybridization showed that *AGL16* is expressed in guard cells and trichomes in both abaxial and adaxial epidermis of rosette leaves (Alvarez-Buylla, 2001). These experiments indicated that *AGL16* may have a role in the regulation of stomata development (Alvarez-Buylla, 2001). Evolution of stomata was one of a key events during the early evolution of land plants which is consistent with the idea that MADS-box genes played an important role in the evolution of body plans in plants (Baum, 1998; Kellogg, 2004). *AGL19* is specifically expressed in root and not in any other tissues indicating a role of *AGL19* in root development (Alvarez-Buylla, 2001). Within root, this gene is expressed in columella, lateral root cap, and epidermal cells of the meristematic regions (Alvarez-Buylla, 2001). The root cap plays a central role in perception of environmental cues such as gravity and is important for gravitropic responses in plant.

AGL27 and *AGL31* are similar in sequence and expression pattern to *FLC* and are expressed in most of the plant tissues including root, leaf, stem, flower and in silique (Alvarez-Buylla, 2001). Similarity of *AGL27* and *AGL31* in sequence and expression pattern to *FLC* suggest that theses genes may have redundant activities to each other indicating *AGL27* and *AGL31* may have a function in controlling flowering time as well. However, *FLC* also seems to have at least some independent roles because single *flc* loss-of-function mutant shows a clear early flowering phenotype which is not observed in *agl27* nor *agl31* single mutants (Michaels et al., 2003). Other examples of MADS-box genes that have function in development of plant outside the flower include *SHATTERPROOF1 (SHP1), SHATTERPROOF2 (SHP2), FRUITFUL (FUL)*, and *TRANSPARENT TESTA16 (TT16)* that are involved in seed development (Rounsley et al., 1995; Alvarez-Buylla, 2001).

1.3.3 Plant type II MADS-box genes in other plant species

Completion of genome sequencing in *Arabidopsis thaliana* in 2000 identified approximately 26000 genes in this plant model organism of which just over 100 of the identified genes are MADS-box genes (Kofuji et al., 2003; Parenicova et al., 2003; Martinez-Castilla and Alvarez-Buylla, 2004). Because of a high level of conservation in the DNA binding MADS-box domain, isolation of MADS-box genes from other distantly related plant species such as maize, rice, and orchid plants were possible by using degenerate primers that bind to the conserved MADS-box domain at N-terminal (Mena et al., 1995; Becker et al., 2000; Jia et al., 2000b). Identification of MADS-box genes in these monocot plants indicated that MADS-box genes were already present before the divergence of monocots and dicots (Jia et al., 2000a). Fact that the ABC model of floral organ identity is conserved between monocots and dicots supports the above statement (Ambrose et al., 2000). In line with the ancestral presence of MADS-box genes in plants, MADS-box genes have also been identified in lower plant species such as moss (*P. patens*), fern, and gymnosperms (*G. gnemon*) (Winter et al., 1999; Henschel et al., 2002; Theissen and Becker, 2004). MADS-box genes are also functional in monocot plant development. This is of an agricultural importance as well as the scientific importance because important cereals including rice plant belong to the monocot (Jia et al., 2000b). Rice (*Oryza sativa)* genome sequencing has been completed recently and this enabled the

comparative genomics approach to compare MADS-box genes from dicots (*Arabidopsis thaliana*) to that of monocots (*Oryza sativa*) (Jia et al., 2000b). This resulted in the identification of 71 MADS-box genes in *Oryza sativa* but the actual number of MADS-box genes in rice is thought to be higher as genome annotation in rice is far from completion at this stage (Jia et al., 2000a). Examples of MADS-box genes identified in rice plant include *FDRMADS6* and *FDRMADS7* which showed high homology with *AP1* of *Arabidopsis thaliana* and *AP1* orthologues from other plant species (Jia et al., 2000a). FDRMADS6 protein showed 62% identity with AP1 in its first 154 amino acids while FDMADS7 protein showed 52% identity with AP1 in its first 231 amino acids (Jia et al., 2000a). Because *FSMADS6* and *FSMADS7* have high a level of similarity with *AP1*, an A class MADS-box gene involved in specification of floral organs, they are likely to have a function in flower development as well. Expression pattern analysis by RT PCR showed *FSMADS6* is exclusively expressed in inflorescence and no signal was detected in the vegetative tissues including root, leaf or stem (Jia et al., 2000a). On the other hand, *FSMADS7* is expressed mainly in inflorescence but weaker expression was also observed in root and shoot tissues as well (Jia et al., 2000a). Because MADS-box genes tend to be expressed in a cell where they function, this result suggested that MADS-box genes may have functions outside the flower as well in monocots such as *Oryza sativa (Jia et al., 2000b)*.

1.4 Animal/Fungal type I MADS-box genes

In plants, only little is accomplished in terms of functional characterization of type I MADS-box gene. However, in animal and fungal kingdoms, type I MADS-box have been shown to play various important roles in development (Treisman and Ammerer, 1992). Examples of animal and fungal type I MADS-box genes include *SRF* and *MCM*1 (Treisman and Ammerer, 1992).

1.4.1 Serum Response Factor

SERUM RESPONSE FACTOR (*SRF*), a type I MADS-box genes in mammals, is a transcription factor (TF) that regulate cell cycle phase transitions and muscle development (Treisman and Ammerer, 1992). SRF is a ubiquitously expressed TF protein that consists of 508 amino acids and has a molecular weight of 64Kda (Arsenian et al.,

1998). It regulates expression of many cellular immediate early genes such as c-fos in response to serum growth factor (Treisman, 1992). In response to serum, SRF recognizes a specific DNA sequence termed Serum Response Element (SRE) through its interaction with AT rich CarG-box sequence (CC(A/T)6GG) present at the DNA element to attach itself to the target genes and regulate expression of the targets (Treisman, 1992). *SRF* gene has been cloned and X-ray crystal structure of SRF MADS-box domain bound to SRE has been solved (Treisman, 1992; Treisman and Ammerer, 1992). Together, these data were used to map DNA binding, dimerization and transactivation domains in SRF. These mapping experiments mapped DNA binding and dimerization domain to the N-terminal MADS-domain respectively at positions 133 –222 and 168 – 222 and transactivation domain at C-terminal (Wynne and Treisman, 1992). GAL4-SRF chimerical transcription factor protein reporter assay experiment surprisingly showed that GAL4-full SRF construct could not activate expression of reporter gene (Wynne and Treisman, 1992). However, the same experiment showed when SRF is deleted at N-terminal up to position 203, the construct constitutively activated expression of reporter gene even in absence of the serum growth factor (Wynne and Treisman, 1992). This indicated presence of repressor domain at the N-terminal of SRF which over-lap with the MADS-box DNA binding and dimerization domain (Wynne and Treisman, 1992). Although MADS domain and repressor domain over-lap, DNA binding and dimerization domain were not responsible for the inhibition of SRF activity since mutations that affect DNA binding or dimerization did not affect inhibitory activity (Wynne and Treisman, 1992). Repressor domain function to inhibit SRF dependent transactivation of target genes when SRF is not bound to SRE allowing repression of the gene expression when transactivation is unnecessary (Davis et al., 2002). This discovery opened a way for application of the repressor domain to further study the function of SRF (Davis et al., 2002). Deletion of SRF trans-activation domain at C-terminal resulted in expression of truncated SRF that is capable of competing with WT SRF for binding to the target DNA elements but lacks the ability to upregulate transcription (Davis et al., 2002). Hence the truncated SRF functioned as a dominant negative mutant of SRF and was used in many studies to show the importance of SRF in various muscle-related gene expression (Davis et al., 2002).

It has been shown that SRF interacts with various accessory proteins via its MADS dimerization domain including homeo domain proteins phox1/Mhox and Nkx2.5, NF-kB, ATF6, myogenic bHLH factors, and HMG-I family of non-histone nuclear proteins (Marais et al., 1992; Marais et al., 1993). *In vitro* experiment replacing dimerization domain of SRF with that of MCM1 resulted in a chimerical SRF that can recruit MCM1 accessory protein STE12 (Wynne and Treisman, 1992). Depending these accessory proteins, SRF can function as both activator or repressor of its target genes (Wynne and Treisman, 1992). Extensive studies have been carried out with one interesting accessory protein family of SRF called Ets domain accessory family whose members have been shown to form ternary complex with SRF at SRE (Marais et al., 1993). Hence they are also called ternary complex factors (TCF) (Marais et al., 1993). In SRE region, there is a motif called Ets motif (GGA(A/T)) adjacent to the CarG box sequences and Ets domain proteins SAP-1 and Elk-1 have been shown to bind to the DNA through interaction with the Ets motif (Marais et al., 1993). SAP-1 and Elk-1 possess N-terminal Ets domain that bind to the Ets motif (Marais et al., 1993). These proteins also contain conserved 21amino acid regions called B-box domain located 50 residues C-terminal to the Ets domain and this B-box domain mediate ternary complex formation with SRF (Marais et al., 1993). There have been reports though, that some SREs do not contain the Ets motifs which indicate possibility that different SRF accessory proteins exist such as bHLH, NF-kB and so on (Marais et al., 1993).

Phosphorylation may also play a role in regulating SRF mediated gene expression as reported (Marais et al., 1992; Xi and Kersh, 2002). SRF itself is a phospho-protein and contains at least 4 phosphorylation sites that are targeted by Casein Kinase II (CKII) (Marais et al., 1992). It was shown that SRF have 2 phosphorylated forms – nascent non-phosphorylated form with molecular weight of 64Kda and fully phosphorylated 67Kda form (Marais et al., 1992). Upon phosphorylation, SRF is activated and carries out its function of regulating the gene expression (Marais et al., 1992). SRF have been shown to act both as activator or repressor of expression depending on its phosphorylation state (Marais et al., 1992). However, role of phosphorylation in regulation of SRF-mediated transactivation is somewhat obscure because recent research by other groups has shown that mutations that disrupt these phosphorylatioin sites did not affect transactivation ability of SRF (Iyer et al., 2003). Though phosphorylation may not directly affect activity

of SRF, it may have a role in indirectly regulating SRF-mediated transactivation by phosphorylating the accessory proteins of SRF (Bebien et al., 2003). For instance, Ets domain family proteins SAP-1 and Elk-1 possess MAP kinase target site at their C-terminal and when phosphorylated at the C-terminal, interaction between these proteins and SRF are induced to form a ternary complex which is capable of turning on the expression of target genes (Marais et al., 1993).

In vivo mechanism of how SRF activate transcription is yet to be discovered but SRF has been reported to interact with general transcription factor TFIIF, homeodomain pHOX1, and chromatin remodeling protein CBP (Arsenian et al., 1998). These 3 proteins could be interacting with SRF at SRE in response to serum to up-regulate the expression of target genes (Arsenian et al., 1998). Alternatively there could be yet unidentified protein factors that interact with SRF to induce expression in response to serum (Xi and Kersh, 2002).

There are many cases where SRF is shown to play role in developmental processes (Simon et al., 1997; Escalante and Sastre, 1998). For example, mice *srf* null mutant have shown that SRF is required for mesoderm formation in mice embryo and that disruption of this gene in mice results in embryonic lethal phenotype or dies within couple days of birth due to severe defect in muscle development (Arsenian et al., 1998). In *Dictyosterium*, SRFA have been shown to play role in final step of spore differentiation and in Drosophila, dSRF plays a role in development of trachea (Simon et al., 1997; Escalante and Sastre, 1998) .

1.4.2 Miniature chromosome 1

MCM1 (Miniature chromosome 1) is a yeast homologue of SRF and has been shown to determine mating type differentiation after activation by pheromone induced signal transduction pathway (Treisman and Ammerer, 1992). In yeasts, there are 3 cell types – haploid a, α, and diploid a/α (Treisman and Ammerer, 1992). These cells differ from each other in terms of their mating specificity and ability to form spore (Treisman and Ammerer, 1992). These differences are due to regulation of cell-type specific gene expression by MCM1 (Treisman and Ammerer, 1992). In α-cell, MCM1 forms a hetero-tetramer complex with homeodomain protein MAT α-2 (Treisman and Ammerer, 1992). MCM1-α2 complex binds 31bp partially symmetrical conserved DNA sequence located upstream of transcriptional start site of a-specific genes and repress their expressions

(Treisman and Ammerer, 1992). In contrast, in a-cell, MCM1 interacts with α-1 homeo-domain protein to positively regulate the transcription of a-cell specific genes to promote a-cell fate (Treisman and Ammerer, 1992).

MCM1 shares 72% identity with SRF over the conserved MADS-domain but binds slightly variant consensus sequence CC(C/T)AA(A/T)NNGG instead of the CarG box sequence (Wynne and Treisman, 1992). In addition, accessory proteins of SRF and MCM1 are quite diverse and some are shown to be specific to MCM1 or SRF while some other accessory proteins are capable of binding to both MCM1 and SRF (Wynne and Treisman, 1992). Like SRF, MCM1 is also capable of functioning as activator or repressor of transcription of target genes depending on the interacting partner such as stated in above example of interaction with α1 and α2 (Treisman and Ammerer, 1992).

In summary, type I MADS-box genes are playing various functions in wide range of developmental processes in animal and fungal kingdom (Treisman and Ammerer, 1992). Molecular mechanisms of how type I MADS-box genes are regulating transcription of target genes are not yet identified but their accessory proteins have been shown to play significant role in the regulation such that types of the interacting partner can signal type I MADS-box genes to function either as transctivator or repressor (Treisman and Ammerer, 1992).

1.5 Plant type I MADS-box genes

Recently functional characterization have been achieved with 2 members of plant type I MADS-box clade supporting the hypothesis that type I MADS-box genes have a functional role in plant development (Kohler et al., 2003; Portereiko et al., 2006). These 2 genes are *AGL80* and *AGL37* (*PHERES1*) (Kohler et al., 2003; Portereiko et al., 2006). Both AGL80 and PHERES1 were shown to be important for seed development in plants (Kohler et al., 2003; Portereiko et al., 2006).

1.5.1 *AGL80*

In angiosperms (flowering plants), seed formation initiates with a fusion of male gametes and female gametes (Faure et al., 2002). In Arabidopsis, double fertilization takes place in which 2 sperm nuclei (n) produced from male gametes fertilizes egg cell (n) and central cell (2n) of female gamete to produce diploid embryo (2n) and triploid endosperm

(3n) respectively (Faure et al., 2002). Embryo will eventually give rise to mature plant while endosperm is responsible for transporting maternally derived nutrient to embryo for proper embryo development (Faure et al., 2002). In last decade, number of genes involved in female gametophyte development have been isolated including *FIE, FIS2* and *MEA* (Luo et al., 2000).

In *agl80* mutant, endosperm development was absent even when the mutant was pollinated with WT pollen suggesting role of AGL80 (AGAMOUS-LIKE 80) in female gametophyte development (Portereiko et al., 2006). Consistent with the above statement, further phenotypic analysis showed central cell in *agl80* mutant plant had much smaller vacuole and nucleolus compared to the WT central cell that is characterized by a large vacuole and nucleolus (Portereiko et al., 2006). Further, AGL80::GFP reporter gene system showed expression of AGL80::GFP in central cell and in early developmental stage endosperm 3 days after pollination (Portereiko et al., 2006). In addition, looking at expressions of endosperm expressed genes *DME*, *FIS2* and *DD46* in *agl80* mutant back-ground, it showed expression of *DME* and *DD46* were disrupted in *agl80* back-ground while *FIS2* expression remained intact. In summary AGL80 is required for expression of at least some of the genes essential for endosperm development and knocking out the function of AGL80 results in defective endosperm development.

1.5.2 *PHERES1*

PHERES1 (*PHE1*) is another plant type I MADS-box gene whose role has been suggested in seed development (Kohler et al., 2003). *PHE1* was first identified by microarray assay as an up-regulated gene in *fis* class gene mutants indicating *PHE1* as a down-stream target of FIS protein complex (Kohler et al., 2003). FIS class proteins are a group of proteins that interact with each other to form PcG protein complex and repress expression of target genes and there have been many reports about the importance of epigenetic modification caused by FIS protein complex in seed development (Kohler and Makarevich, 2006). In WT, *PHE1* is not expressed in central cell before pollination but the expression is induced 1 – 2 days after pollination (DAP) in seeds containing pre-globular stage embryo (Kohler et al., 2005). In *fis* class mutants, *PHE1* expression is observed much earlier, already expressed directly after pollination (0DAP) because repression from FIS protein complex is absent in the *fis*-class mutants (Kohler et al.,

2005). Subsequent ChIP assay showed FIS class proteins MEA and FIE can bind *PHE1* promoter and confirmed *PHE1* as direct target of FIS protein Pc-G complex (Kohler et al., 2005). When *PHE1* expression was down-regulated in *mea* mutant back-ground, seed abortion phenotype of *mea* mutant was rescued confirming role PHE1 plays in seed development (Kohler et al., 2005).

Examples of *AGL80* and *PHE1* further supports the hypothesis that plant type I MADS-box genes do play a role in development. Interestingly both *AGL80* and *PHE1* belong to same sub-clade of type I MADS-box clade (Parenicova et al., 2003) and since both of them have function in seed development, research is underway to study the role of other members within this subclade in seed development (Portereiko et al., 2006).

1.6 Seed development

In their life cycles, plants alternate a haploid (n) gametophytic phase and diploid (2n) sporophytic phase (Berger, 2003). In *Arabidopsis thaliana*, sporophytic phase consists most of the plant's life-time and haploid gametophytic phase is reduced to only short time at the reproductive organ developmental stage before fertilization (Berger, 2003). Once Arabidopsis plants reach reproductive state, the diploid plants produce special cell lineages that undergo meiosis to produce mega-gametophyte (ovule) and micro-gametophyte (pollen) (McCormick, 1993; Drews and Yadegari, 2002). In ovule development, meiosis is followed by 3 syncytial divisions and cellularization to produce the mature 7-celled ovule consisting of egg cell, central cell, 2 synergid cells and 3 antipodal cells (Faure et al., 2002). In pollen development, meiosis is followed by two rounds of mitosis – pollen mitosis I (PMI) and pollen mitosis II (PMII) – to give rise to 2 sperm cells (McCormick, 1993).

Following the formation of these female and male gametophytes, a process termed double fertilization occurs which is a fusion of 2 haploid gametes and represents the end of gametophytic phase (Faure et al., 2002). After pollen grain is dropped onto style, pollen grain grows pollen tube to deliver 2 sperm cells to the ovule (Faure et al., 2002). As sperm cells are delivered to ovule, 3 antipodal cells that acted as a guide marker become degenerated and ovule now contains 4 cells (Faure et al., 2002). Once sperm cells are delivered to ovule, one of this male gamete fertilizes egg cell to produce diploid embryo (2n) (Faure et al., 2002). The other male gamete fertilizes central cell (2n) and

give rise to triploid endosperm (3n) that acts to deliver maternally produced nutrients to the embryo for embryo growth (Faure et al., 2002).

Endosperm development is mainly divided into 2 phases - synctium division and cellularization (Berger, 2003). Synctium division is a process in which mitotic division occurs without cell wall formation to produce multi-nucleate endosperm (Berger, 2003). Experiment has shown that 5 hours after pollination (H.A.P), one of the two synergid cells degenerate as pollen sperm cells are released into ovule (Berger, 2003). At this stage, change in egg cell polarity occurs and egg cell nucleus moves away from central cell (Faure et al., 2002). Then central cell nucleus becomes elongated along micropyle-chalazal axis (Faure et al., 2002). Six to seven HAP, central cell nucleus contains 2 nucleoli with the one in a micropylar position always being smaller than the one in chalazal position (Faure et al., 2002). After this stage, 2 nucleoli can also be observed in egg cell. Again they observed significant difference in size between the bigger nucleoli and smaller nucleoli (Faure et al., 2002). In central cell, 2 nucleoli have not changed in size or position at this stage (Faure et al., 2002). Fertilized central cell undergo first nuclear division as early as 7 HAP along the micropyle-chalazal axis (Faure et al., 2002). The outcome of this division is a 2-nucleate endosperm with its nuclei being adjacent to each other which can be observed at 8 to 9 HAP (Faure et al., 2002). Second and third nuclear division follows 12 and 24 HAP producing 8-nucleate endosperm at the end of third nuclear division (Faure et al., 2002). Mitotic divisions without cell wall formation continues to take place until 3 – 4 DAP at which stage multinucleate endosperm containing over 100 nuclei have developed representing the end of synctium division (Faure et al., 2002).

1.6.1 PcG complex mediated epigenetic control on seed development

Until recently only little was known about the molecular mechanisms underlying fertilization. However the recent sequencing of entire genome of *Arabidopsis* coupled with improvement in isolation of mutants has given further insight into molecular processes that take place during fertilization (Arabidopsis Genome, 2000; Dresselhaus, 2006). When pollen sperm cell (n) fertilizes egg cell (n), 2 nuclei of these gametes need to fuse in a same cell-cycle phase in order to avoid aneuploidy and to allow proper subsequent embryo growth (Dresselhaus, 2006). Therefore, there must be some sort of

cell cycle regulation in both the male and female gametes (Dresselhaus, 2006). Unlike in animal gametes, regulation of cell cycle in plant gametes remains largely unknown (Faure et al., 2002). Recently it was suggested that in *Arabidopsis*, sperm cells enter new S-phase after PMII and the sperm cells are likely to be in G2 phase when they are delivered to ovule (Dresselhaus, 2006). Since two nuclei from male and female gametes must fuse in same cell-cycle phase, simplest assumption was that female gametes are also in G2 phase at the time of fertilization (Dresselhaus, 2006). In female gametes, cell-cycle arrest is mediated by polycomb group (PcG) complex (Drews and Yadegari, 2002). These PcG complex genes include *FERTILIZATION INDEPENDENT SEED* (*FIS*), *MEDEA* (*MEA*), and *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*) and loss-of-function mutation of these genes all result in autonomous initiation of cell division in central cell without the fertilization (Luo et al., 2000). The PcG complex genes are shown to be involved in epigenetic control of gene expression (Luo et al., 2000). Many evidences point to the importance of epigenetic control of gene expression on seed development (Luo et al., 2000; Makarevich et al., 2006). Particularly the regulation of imprinted gene expression in developing endosperm by methylation was shown to play significant role in seed development in *Arabidopsis* (Luo et al., 2000). Polycomb group (Pc-G) genes initially identified in *Drosophila melanogastor* are group of genes that maintain repression of target gene expression by methylating the target histone lysine residue and hence remodeling the chromatin structure (Muller et al., 2002). In *Drosophila*, pc-g mutants failed to maintain transcriptional repression of homeo-box genes (Muller et al., 2002). Pc-G genes assemble in 2 complexes to carry out their function of repressing expression of target genes (Muller et al., 2002). First, Polycomb Repressive Complex 2 (PRC2) or E(Z)/ESC complex that consists of 4 Pc-G proteins – Enhancer of Zeste (E(z)), Extra Sex Comb (ESC), Supressor of Zeste 12 (Su(z)12) and NURF-55 – bind to histone H3 and methylate lysine27 residue (H3K27) to create epigenetic mark (Muller et al., 2002). The second complex PRC1 is then recruited to H3K27 via interaction of its component with the methylated H3K27 (Muller et al., 2002). PRC1 complex also consists of 4 proteins – Polycomb (PC), Polyhomeotic (PH), Posterior Sex Combs (PSC), and dRing (Muller et al., 2002). After binding to target gene via methylated H3K27, PRC1 interacts with SWI/SNF chromatin remodeling complex to modify chromatin structure in a way that transcription initiation is blocked (Muller et al., 2002).

In plants, Pc-G genes were first identified in a genetic screen for mutants with defective seed development (Luo et al., 2000). Several Pc-G genes have been isolated since and presence of PRC2 like complex was confirmed (Makarevich et al., 2006). However, there is still no evidence for existence of PRC1-like complex in plants yet (Makarevich et al., 2006). To date, 3 PRC2 like complexes have been characterized FIS – complex, CLF complex and VRN complex (Reyes and Grossniklaus, 2003). Interestingly all 3 plant PcG complexes seem to have MADS-box genes as their target (Michaels et al., 2003). For example, VRN complex regulates expression of *FLC* and have a role in flowering time and CLF complex repress the expression of floral organ identity gene *AG (*Michaels *et al., 2003)*. Of these 3 complexes, FIS complex has been shown to play important role in gametophyte and early seed development (Makarevich et al., 2006). FIS complex consists of 3 proteins – MEA, FIS2, and FIE – and these 3 proteins are homologues of E(z), Su(z)12, and ESC respectively (Luo et al., 2000). All 3 *mea*, *fis2*, and *fie* mutants show common phenotypes – autonomous endosperm development in absence of fertilization, arrest of embryo development at heart stage and failure of mutant seeds to develop beyond endosperm cellulalization (Luo et al., 2000). These common phenotypes indicated that MEA, FIE, and FIS2 work in a same pathway to suppress various aspects of seed development in absence of fertilization and that fertilization inactivates FIS complex to initiate seed development (Luo et al., 2000). Since MEA and FIE are homologues of E(z) and ESC respectively and E(z) and ESC physically interact in *Drosophila* to repress target genes' expression, one can postulate that MEA and FIE do the same in *Arabidopsis* as well (Pien and Grossniklaus, 2007). Yeast-2-hybrid assay was carried out and it was shown MEA and FIE do indeed interact with each other (Kohler et al., 2003). The assay also showed FIS2 does not physically interact with neither MEA nor FIE (Kohler et al., 2003). This implies yet unidentified component in FIS complex that attaches FIS2 to the rest of complex (Kohler et al., 2003).

1.7 Plant type I MADS-box mutants

Plant type I and type II MADS-box genes have a few more differences apart from their MADS-domain structures (Parenicova et al., 2003). One difference is their distribution on the chromosomes (Johansen et al., 2002). *Arabidopsis thaliana* genome is composed of 5 chromosomes (Arabidopsis Genome, 2000). Type II plant MADS-box genes are evenly

distributed across all the 5 chromosomes whereas most plant type I MADS-box genes are located on chromosomes I and V (Parenicova et al., 2003). It has been estimated that up to 83% of the plant type I MADS-box genes are located on these 2 chromosomes (Parenicova et al., 2003).

For many eukaryotic transcription factor families, gene duplication that gave rise to them occurred predominantly between different chromosomes (Parenicova et al., 2003). Plant type I and II MADS-box genes differ in this aspect as well (Parenicova et al., 2003). Analysis of closely related members of type II and type I plant MADS-box genes showed that 53% of the type II probably originated from gene duplications between 2 different chromosomes (Parenicova et al., 2003). On the other hand, 82% of the type I genes are thought to have risen by internal gene duplication within the chromosome (Parenicova et al., 2003). In other eukaryotes, recent gene duplication occurred more frequently within chromosomes and hence this difference may be suggesting diversity originated more recently in subclades of plant type I MADS-box gene (Becker and Theissen, 2003; De Bodt et al., 2003b; Martinez-Castilla and Alvarez-Buylla, 2004). Consistent with this, type I plant MADS-box genes are further classified into several subclades by several researchers (Kofuji et al., 2003; Parenicova et al., 2003; Martinez-Castilla and Alvarez-Buylla, 2004). For example, Parenicova et al divided plant type I MADS-box genes into Mα, Mβ, Mδ, and Mγ subgroups (Parenicova et al., 2003).

Also in terms of exon/intron numbers, plant type II MADS-box genes typically contain 7 to 9 exons while type I groups tend to have only 1 or 2 exons (Nam et al., 2003).

On top of all these differences, type II MADS-box genes are much more well characterized than type I clades in plants (Alvarez-Buylla, 2001; De Bodt et al., 2003b; de Folter et al., 2005). This is quite Surprising since there are much more type I plant MADS-box genes than type II (Johansen et al., 2002). It may be possible that because plant type II MADS-box genes have been known for a longer time than the type I, more experiments including reverse genetic analysis have been carried out with the type II (Ostergaard and Yanofsky, 2004). However, statistically it is unlikely that this is the only reason for the unequal identification of mutants between type I and type II genes (Parenicova et al., 2003). One possible reason for not being able to identify type I MADS-box mutant in plant is that type I plant MADS-box genes are either non-functional or pseudogenes (Parenicova et al., 2003). There is evidence that at least one

plant type I MADS-box gene (*At5g49490*) is a processed pseudogene (De Bodt et al., 2003a). However, Expression profile and protein – protein interaction studies have shown that most type I MADS-box genes are expressed in various tissues of plants and are interacting with other proteins which strongly rejects this idea (Johansen et al., 2002; de Folter et al., 2005). Also phylogenetic studies and bioinformatics showed that homologous genes have been identified in rice (*Oryza sativa*) which indicates that MADS-box genes were present before the divergence of dicots and monocots (Kofuji et al., 2003; Martinez-Castilla and Alvarez-Buylla, 2004). Because type I MADS-box genes are expressed, conserved between species, and are capable of encoding group specific protein domains, it is unlikely that plant type I MADS-box genes are non-functional or pseudogenes. Another possibility of not being able to identify the type I MADS-box phenotypic mutant is that loss-of-function mutation in these genes give embryonic lethal phenotype (Parenicova et al., 2003). However, several T-DNA or transposon insertion mutants have been identified in the type I genes which rejected this idea (Parenicova et al., 2003). Redundancy could be another reason for not identifying type I MADS-box mutant in plants (Pelaz et al., 2001; Pinyopich et al., 2003). In type II plant MADS-box genes, several very clear examples of redundancy was observed. In Arabidopsis flower formation, B and C class genes controlling petal, stamen and carpel identity are functionally dependent on 3 very similar MADS-box genes, *SEP1, SEP2*, and *SEP3* and only when all of the 3 *SEP* genes were knocked out in *sep1 sep2 sep3* triple mutant was there a loss of petal, stamen, and carpel observed (Pelaz et al., 2000; De Bodt et al., 2003a). Another example of redundancy was observed in *SHP1* and *SHP2* (Rounsley et al., 1995; Liljegren et al., 2000). Single mutants of either *shp1* or *shp2* showed phenotype indistinguishable from that of WT but when these mutants were crossed to produce the *shp1 shp2* double mutant, disturbance in dehiscence zone development was observed in the double mutant fruit which resulted in a failure to release seeds (De Bodt et al., 2003a). These examples indicated that redundant activity is a common phenomenon in plant type II MADS-box genes, which could also occur in type I MADS-box genes.

1.8 AGL40 & AGL62

Of approximately just over 100 plant MADS-box genes, about 60 of them belong to type I lineage (Parenicova et al., 2003). As previously mentioned, type I lineage can be further sub-divided into several subclades based on their sequence similarity (Parenicova et al., 2003). Both *AGL40* and *AGL62* belong to Må subclade and share expression domain in inflorescence and sillique (fruit) (Parenicova et al., 2003). Because MADS-box genes tend to function in a site of expression (De Bodt et al., 2003a), their sequence similarity and similar expression pattern indicates *AGL40* and *AGL62* are most closely related to each other and that they may have redundant activities. Yeast 2 hybrid assay showed that AGL40 and AGL62 can physically interact with each other in embryo further pointing out to the possibility of AGL40 and AGL62 playing role in embryo development (de Folter et al., 2005).

1.9 Aim & Hypothesis

So far, only little functional characterization has been achieved with plant type I MADS-box clade (De Bodt et al., 2003a). However, functional characterization has been achieved extensively in other clades of MADS-box genes (De Bodt et al., 2003a) and knowing that plant type I MADS-box genes share the conserved MADS-box domain with other MADS-box gene clades, one can hypothesize that plant type I MADS-box genes also play a role in plant development. Recent findings that AGL80 and PHE1 play a role in seed development in plants also support this hypothesis (Kohler et al., 2005; Portereiko et al., 2006).

Therefore our aims in this project are

1) To give further evidence of plant type I MADS-box genes playing role in plant development and further the understanding of a role plant type I MADS-box genes play in plant development by studying functions of AGL40 and AGL62 – 2 most closely related members within plant type I MADS-box clade – using T-DNA knock out lines.

2) Identify spatial and temporal expression pattern of *AGL40* and *AGL62 in planta* and use the information obtained for functional characterization of *AGL40* and *AGL62*