2005

GENE REGULATORY NETWORKS OF AGL15 A PLANT MADS TRANSCRIPTION FACTOR

Cong Zhu
University of Kentucky

Click here to let us know how access to this document benefits you.

Recommended Citation
Zhu, Cong, "GENE REGULATORY NETWORKS OF AGL15 A PLANT MADS TRANSCRIPTION FACTOR" (2005). University of Kentucky Doctoral Dissertations. 446.
https://uknowledge.uky.edu/gradschool_diss/446

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
ABSTRACT OF DISSERTATION

Cong Zhu

The Graduate School
University of Kentucky
2005
GENE REGULATORY NETWORKS OF AGL15,
A PLANT MADS TRANSCRIPTION FACTOR

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture at the University of Kentucky

By

Cong Zhu

Lexington, Kentucky

Co-Directors: Dr. Sharyn E. Perry, Associate Professor of Agronomy and Dr. Joseph Chappell, Professor of Agronomy

Lexington, Kentucky

2005

Copyright © Cong Zhu 2005
ABSTRACT OF DISSERTATION

GENE REGULATORY NETWORKS OF AGL15, A PLANT MADS TRANSCRIPTION FACTOR

Plant embryogenesis is an intriguing developmental process that is controlled by many genes. AGAMOUS Like 15 (AGL15) is a MADS-domain transcriptional regulator that accumulates preferentially during this stage. However, at the onset of this work it was unknown which genes are regulated by AGL15 or how AGL15 is regulated. This dissertation is part of the ongoing effort to understand the biological roles of AGL15.

To decipher how AGL15 functions during plant development, a chromatin immunoprecipitation (ChIP) approach was adapted to obtain DNA fragments that are directly bound by AGL15 in vivo. Putative AGL15 targets were isolated, and binding and regulation was confirmed for one such target gene, ABF3.

In addition, microarray experiments were performed to globally assess genes that are differentially expressed between wild type and agl15 young seeds. Among them, a gene, At5g23405, encoding an HMGB domain protein was identified and its response to AGL15 was confirmed. Preliminary results suggest
that the loss-of-function of At5g23405 might have an effect on somatic embryogenesis, consistent with AGL15 repression of the expression of this gene.

Lastly, to address the question about how the regulator is regulated, the cis elements controlling the expression of AGL15 must be identified. Deletion analysis of the AGL15 promoter indicated the presence of putative positive and negative cis elements contributing to the expression of AGL15. Further analysis suggested that AGL15 regulates the expression of its own gene and this regulation may partially be explained by the direct binding of the protein to the AGL15 promoter.

The data presented in this dissertation demonstrate that ChIP can be used to identify previously unsuspected targets of AGL15. Based on ChIP, a ChIP-chip technique is being developed in the lab to allow a more global analysis of in vivo binding sites. The identification of target genes and cis elements in AGL15 promoter is a step towards characterization of the biological roles of AGL15.

KEYWORDS: AGL15, Embryogenesis, MADS transcription factor, Chromatin immunoprecipitation (ChIP), Autoregulation

Cong Zhu

April 15, 2005
GENE REGULATORY NETWORKS OF AGL15,  
A PLANT MADS TRANSCRIPTION FACTOR

By

Cong Zhu

Sharyn E. Perry  
Co-Director of Dissertation

Joseph Chappell  
Co-Director of Dissertation

Arthur G. Hunt  
Director of Graduate Studies

April 15, 2005
RULES FOR THE USE OF DISSERTATION

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgements.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name

Date
GENE REGULATORY NETWORKS OF AGL15, A PLANT MADS TRANSCRIPTION FACTOR

____________________________________

DISSERTATION

____________________________________

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture at the University of Kentucky

By
Cong Zhu

Lexington, Kentucky

Co-Directors: Dr. Sharyn E. Perry, Associate Professor of Agronomy and Dr. Joseph Chappell, Professor of Agronomy

Lexington, Kentucky

2005

Copyright © Cong Zhu 2005
ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my advisor Dr. Sharyn Perry for her excellent academic guidance, insightful advice, exceptional patience and generous support. Without her help, this dissertation would not have come to fruition. I also would like to thank other members of my dissertation committee, Drs. Joseph Chappell, Arthur Hunt, Chuck Staben. I greatly benefited from their support and I thank them for helping me on my study and devoting their time to review my dissertation. I also would like to thank Dr. Brett Spear for his kindness to serve as my outside examiner.

I would like to express my gratitude to the present and past members of the Perry lab, Ellen Harding, Dr. Huai Wang, Dr. Weining Tang and Kristine Hill. I have truly enjoyed the past five and half years because of their technical support, inspirational discussion and friendships.

I would also like to express my appreciation to Dr. Sresty Tavva from the PGEN lab for his assistance on quantitative RT-PCR experiments, Donna Wall from the UK Microarray Core Facility for her technical support on the microarray studies, Ms. Lisa Meeks and Dr. Kevin Forbes of the Hunt lab for their technical assistances.

I wish to express my sincere appreciation to my family for their support throughout my study. Most of all, I would like to thank my wife, Hua Liu for her great love, care and encouragement.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS**............................................................................................................................................iii

**LIST OF TABLES**........................................................................................................................................................... vii

**LIST OF FIGURES**................................................................................................................................................................ viii

**Chapter 1 Literature Review** ..............................................................................................................................................1

1.1 Plant embryogenesis...........................................................................................................................................................2

1.1.1 Morphogenesis...............................................................................................................................................................3

1.1.2 Maturation and desiccation................................................................................................................................................3

1.1.2.1 Proteins........................................................................................................................................................................5

1.1.2.2 Lipids.............................................................................................................................................................................5

1.1.2.3 Carbohydrates.............................................................................................................................................................5

1.1.3 Genes that are involved in plant embryogenesis.............................................................................................................6

1.1.3.1 Apical-basal patterning ..............................................................................................................................................7

1.1.3.2 Radial patterning...........................................................................................................................................................9

1.1.3.3 Maternal effect..............................................................................................................................................................10

1.1.3.4 Suspensor effect..........................................................................................................................................................10

1.1.3.5 LEC genes..................................................................................................................................................................11

1.1.4 Somatic embryogenesis....................................................................................................................................................11

1.1.5 Summary of embryogenisis...............................................................................................................................................14

1.2 AGL15, a MADS domain protein.........................................................................................................................................15

1.2.1 The MADS family...........................................................................................................................................................15

1.2.2 The MADS domain..........................................................................................................................................................17

1.2.3 MADS box genes in animals and yeasts............................................................................................................................17

1.2.4 MADS box genes in plants................................................................................................................................................20

1.2.4.1 Type I plant MADS box genes................................................................................................................................20

1.2.4.2 MIKC-type MADS box genes.....................................................................................................................................21

  Floral organ identity specification - The ABC (E) model.........................................................................................................22

  Floral meristem identity specification..................................................................................................................................23

  Flowering time control............................................................................................................................................................25

  Other developmental programs...............................................................................................................................................25

1.2.5 AGL15, A MADS box gene preferentially expressed in Embryos.....................................................................................26

1.2.5.1 Expression pattern of AGL15....................................................................................................................................26

1.2.5.2 Biological role studies of AGL15................................................................................................................................28

1.3 Research proposal and strategies........................................................................................................................................30

1.3.1 Chromatin Immunoprecipitation....................................................................................................................................31

1.3.2 Microarray..........................................................................................................................................................................31

1.3.3 Identification of important cis elements in the *AGL15* promoter..................................................................................32

1.3.4 Specific aims of this dissertation research.......................................................................................................................32
Chapter 2 Identification and analysis of downstream target genes of AGL15 by a chromatin immunoprecipitation approach .................33

2.1 Introduction ........................................................................................................34
2.2 Results ..................................................................................................................36
   2.2.1 Identification of AGL15 binding sites in vivo .................................................36
   2.2.2 In vitro selection on the ChIP population .........................................................37
   2.2.3 Analysis of the isolated downstream targets .....................................................40
   2.2.4 AGL15 interacts directly with the regulatory region of ABF4 .........................42
   2.2.5 ABF3 expression is responsive to the changes AGL15 Amounts .................45
2.3 Discussion .............................................................................................................47
2.4 Material and methods .........................................................................................52
   2.4.1 Arabidopsis embryonic culture tissue .............................................................52
   2.4.2 Arabidopsis plant growth .................................................................................52
   2.4.3 Chromatin immunoprecipitation .................................................................53
   2.4.4 Tissue fixation and nuclei isolation ...............................................................53
   2.4.5 Chromatin solubilization and preadsorption .................................................53
   2.4.6 Immunoprecipitation ....................................................................................54
   2.4.7 Elution and DNA recovery .............................................................................55
   2.4.8 DNA cloning and sequencing .......................................................................55
   2.4.9 In vitro selection of the ChIP population .........................................................56
   2.4.10 Protein analysis ............................................................................................57
   2.4.11 Enrichment test for in vivo binding of AGL15 ..............................................58
   2.4.12 RNA slot blot ...............................................................................................58
   2.4.13 Generation of transgenic plants and reporter activity quantification ..........59

Chapter 3 Microarray analysis of Arabidopsis wild type and agl15 seeds and identification of a HMGB box gene ...............................60

3.1 Introduction ..........................................................................................................61
3.2 Results ..................................................................................................................62
   3.2.1 Microarray analysis of Arabidopsis in wild type and agl15 Seeds ....................62
   3.2.2 Confirmation of the Microarray results ............................................................68
   3.2.3 Characterization of At5g23405 .....................................................................72
3.3 Discussion .............................................................................................................78
3.4 Material and methods ..........................................................................................82
   3.4.1 Arabidopsis plant germination and growth .....................................................82
   3.4.2 Arabidopsis RNA extraction ..........................................................................83
   3.4.3 Microarray ......................................................................................................83
   3.4.4 RT-PCR .........................................................................................................84
   3.4.5 Quantitative real-time RT-PCR ....................................................................85
   3.4.6 Seedling Liquid Culture ................................................................................86
LIST OF TABLES

Table 3.1a Genes up-regulated in agl15 versus Ws........................................... 64
Table 3.1b Genes down-regulated in agl15 versus Ws................................... 65
Table 3.2  QRT-PCR analysis of At5g23405 expression................................. 73
LIST OF FIGURES

Figure 1.1 Schematic diagram of the dicotyledonous plant Arabidopsis Embryogenesis..........................................................................................4
Figure 1.2 Type I (SRF type) and type II (MEF2) type MADS domain proteins.16
Figure 1.3 MEF2A-DNA complex........................................................................................................18
Figure 1.4 A revised ABC model for flower development.................................24
Figure 2.1 Outline of the chromatin immunoprecipitation (ChIP) protocol......38
Figure 2.2 Chromatin immunoprecipitation of AGL15........................................39
Figure 2.3 In vitro selection of the ChIP population...........................................41
Figure 2.4 AGL15 directly bind to the regulatory region of ABF3......................44
Figure 2.5 Response of ABF3 expression to AGL15 levels..............................46
Figure 3.1 Distribution in functional categories of genes differentially expressed between agl15 and Ws.................................................................67
Figure 3.2 Potential AGL15 binding sites in the target genes.........................69
Figure 3.3 RT-PCR analysis of four genes identified in microarray experiments..........................................................70
Figure 3.4 Alignment of the amino acid sequences...........................................74
Figure 3.5 A phylogenetic tree generated from At5g23405.1 and other HMGB family members in Arabidopsis.............................................................76
Figure 3.6 Embryonic development from the shoot apex of seedlings in culture........................................................................................................77
Figure 4.1 Deletion analysis of the AGL15 promoter in seedlings..................90
Figure 4.2 Effects of AGL15 promoter deletions on the expression pattern of AGL15:GUS.................................................................93
Figure 4.3 Response of AGL15 expression to exogenous auxin....................95
Figure 4.4 Response AGL15:GUS expression to AGL15 levels and forms....96
Figure 4.5 Binding of AGL15 to the 5' regulatory region of its own gene........99
Figure 4.6 Effect of site-directed mutagenesis of potential CArG motifs on GUS expression levels.................................................................103
CHAPTER 1 LITERATURE REVIEW
1.1 Plant embryogenesis

The life cycle of higher plants alternates between the dominant, independent sporophyte and the dependent gametophyte. During plant sexual reproduction, the male gametophytes or pollen grains, which contain two sperm cells and one vegetative nucleus when mature, are formed in the anther. The female gametophyte or embryo sac, is formed in the ovule, and typically consists of seven cells: one egg, two synergids, one central cell and three antipodal cells. The ovule is organized in such a way that the egg cell and two synergid cells are located at the micropylar end of the embryo sac, while the antipodal cells are located at the chalazal end. Similarly, the egg also has a polar organization with the nucleus and majority of cytoplasm residing at the chalazal end and a vacuole mostly occupying the micropylar end of the egg (Reiser and Fisher, 1993). Fertilization takes place in the embryo sac when the pollen tube enters through the micropyle and releases two sperm cells to combine with the egg and central cells separately. One fertilization event gives rise to the zygote, which develops into an embryo, and the second fertilization give rise to the endosperm. The life of the sporophyte thus begins.

Plant embryogenesis is the process of a single-cell zygote developing into a mature multi-cellular plant embryo. Higher plant embryogenesis is often divided into 3 stages: morphogenesis, maturation and desiccation (West and Harada, 1993). During morphogenesis, the embryonic body parts and primary tissue layers are established. Further development to a mature embryo results from changes in morphogenesis to that of maturation and desiccation. How these processes occur and how they are controlled remained open questions for investigation. Because many studies, especially molecular genetic studies, on embryogenesis were performed using the dicotyledonous plant model system Arabidopsis thaliana (Arabidopsis), this literature review is focused on Arabidopsis and dicotyledonous plants, unless indicated otherwise.
1.1.1 Morphogenesis

Morphologically, mature higher plants have a polar organization along the shoot-root axis, which may be traced back to the polar organization of the embryo sac and egg. The first division of the zygote is asymmetrical (Goldberg et al., 1994). Most of the cytoplasm is distributed to the smaller apical daughter cell near the chalazal end of the ovule while the larger basal cell that is near the micropyle is highly vacuolated. The following divisions of the apical cell give rise to embryo proper while the divisions of the basal cell give rise to the suspensor and hypophysis. Further development of the embryo proper result in the formation of the shoot apical meristem, cotyledons, hypocotyl and embryonic root such that most of the mature embryo derives from the apical cell. The suspensor provides nutrients and growth factors to the embryo and eventually degenerates, while the hypophysis forms part of the root apical meristem. Some representative stages of Arabidopsis thaliana (Arabidopsis) embryogenesis are shown in Figure 1.1. In order to develop from a zygote to a mature embryo, cells not only accumulate in numbers but also differentiate in function. During the early stages of plant development, a simple body plan of the plant is established along the apical-basal axis and the radial axis. This organization of the embryo establishes a framework for the postembryonic development of the plant.

1.1.2 Maturation and desiccation

There are not only morphological changes during embryogenesis. Embryos also go through significant metabolic program changes to establish dormancy and prepare for germination. During embryo maturation, cells in the cotyledons and hypocotyls continue to divide and expand so that the size of the embryo increases significantly. Storage protein, lipids and carbohydrates are synthesized to provide an energy source and/or structural components. At the last stage of development, the embryo becomes metabolically quiescent and highly dehydration tolerant with a gradual decrease of the water content in the embryo.
Figure 1.1 Schematic diagram of the dicotyledonous plant Arabidopsis embryogenesis

(a) Zygote underwent one cell division. ac, apical cell; bc, basal cell.

(b) 4-cell stage embryo. ep, embryo proper; s, suspensor. The apical cell underwent two longitudinal divisions to give rise to a four-celled embryo proper.

(c) Globular stage embryo. O’ line represent the boundary produced by the first set of transverse divisions of the embryo proper.

(d) Transition stage embryo. pd, protoderm; gm, ground meristem; pc, procambium. The concentric organization of the primary meristem tissues became evident during this stage.

(e) Linear cotyledon stage embryo. c, cotyledons; sa, shoot apex; h, hypocotyl; pc, procambium; ra, root apex.
1.1.2.1 Proteins

Storage proteins represent one group of the major reserves for Arabidopsis embryos (Mansfield, 1992). The two major species of storage proteins in Arabidopsis are 12S globulins and 2S albumins (Fujiwara et al., 2002). Other than 12S and 2S storage proteins, oleosins and late-embryogenesis-abundant (LEA) proteins also accumulate in mature embryos. Oleosins are associated with oil bodies that store triacylglycerol. LEA proteins are highly hydrophilic and contain a random coiled-coil moiety that may help embryos become tolerant to desiccation (Ingram, 1996). In Arabidopsis the proteins accumulate steadily throughout the course of embryogenesis and reach a maximum at the end of maturation to account for almost half of the dry matter of the embryo (Baud et al., 2002).

1.1.2.2 Lipids

The amount of lipids also increases significantly during embryogenesis (Baud et al., 2002). Lipids mainly exist in the form of triacylglycerol (TAG) stored as oil bodies to provide energy for future use during germination and seedling establishment. Aside from the increase in amount, the composition of the fatty acids changes significantly. The content of saturated fatty acids, such as palmitic acid and stearic acid drop sharply, while the amount of unsaturated fatty acids, such as oleic acid, linoleic acid, alpha-linoleic acid, eicoseinic acid increase progressively (Baud et al., 2002). These changes in lipid composition can increase the flexibility of the cell membrane and might help the embryo to cope with the dramatic water loss during desiccation.

1.1.2.3 Carbohydrates

The percentage of hexoses (glucose and fructose) in the developing embryos decreases significantly and the mature embryo contains only insignificant amounts of hexoses. At the same time, the levels of sucrose and oligosaccharides, including raffinose and stachyose steadily increase with
maturation of the embryo (Baud et al., 2002). The accumulation of sucrose will serve as both an energy source and an essential element in desiccation tolerance (Corbineau and Côme, 2000). Oligosaccharides, together with sucrose, may help protect membranes from damage by desiccation (Buitink et al., 2000). In Arabidopsis, another major carbohydrate, starch is mainly stored in places other than the embryo, specifically two cell layers of the outer integument; nonetheless, the breakdown of the storage starch in other places within the seed may be converted to oligosaccharide storage in the embryo (Baud et al., 2002).

1.1.3 Genes that are involved in plant embryogenesis

It has long been known that a large number of genes are expressed during embryogenesis (Goldberg et al., 1989) that are involved in many aspects of the three stages of embryogenesis (morphogenesis, maturation and desiccation). For the model plant Arabidopsis, apical-basal and radial embryo pattern formation is stereotypic. Genes involved in pattern formation are thought to control the formation of a specific domain of the embryo. However, as shown in the following sections, many of these so called embryo patterning genes actually play more general roles in basic cellular processes and have broader functions during plant development than specifically controlling embryo pattern formation. Nonetheless they are important for embryo development because when they are mutated, normal embryo patterning is disrupted.

How do the plant embryo cells acquire their fates and form the stereotypic patterning? More and more evidence suggests that cell fate is specified by positional information, and auxin signaling appears to play an important role (Jürgens, 2001). Members of the auxin transporter PINFORMED (PIN) family are essential for the distribution of the auxin gradient (Friml et al., 2002; Benková et al., 2003; Friml et al., 2003).

Mutation of genes PIN1 and PIN7 can disrupt the auxin gradient formation during embryogenesis and cause defects in the establishment of polarity (Friml et al., 2003). By using an in vitro culture system, Liu et al. (1993) found that inhibition of
polar auxin transport induced formation of fused cotyledons in *Brassica juncea*, a morphology that phenocopies the Arabidopsis *pin7* mutant. It was suggested that auxin polar transport is critical for the establishment of bilateral symmetry during early plant embryogenesis. Similarly, inhibition of auxin transport using inhibitors also caused disruption of apical-basal axis formation during early Brassica embryo development (Hadfi et al., 1998). It has been demonstrated that polar auxin transport is also critical for root patterning (Sabatini et al., 1999; Friml et al., 2000). These results indicate that auxin gradients and genes involved in polar auxin transport regulation play a pivotal role in embryo patterning. Screens which were performed at the seedling stage to identify mutants “missing” particular domains will be further discussed in the following paragraphs.

### 1.1.3.1 Apical-basal patterning

The apical-basal patterning includes the establishment of polarity and the formation of the 3 domains: the apical, central and basal along the apical-basal axis. Plant tissues and organs above the ground are derived from shoot apical meristem (SAM). Screens performed at the seedling stage for mutants missing particular pattern elements led to isolation of *gnom*, that is ball-shaped and lacking well-defined apical-basal domains. *GNOM* encodes a brefeldin A (BFA)-sensitive guanine nucleotide exchange factor (GEF) regulating endosomal vesicle trafficking, and has been shown to be involved in auxin transport by regulating PIN1 distribution (Steinmann et al., 1999; Geldner et al., 2003). Mutation of *GNOM* causes disruption of the apical-basal pattern formation in the early embryo (Mayer et al., 1993). Mutation of *GURKE* (*GRK*) (Torres Ruiz et al., 1996) and *PASTICCINO* (*PAS*) (Faure Jean et al., 1998; Vittorioso et al., 1998) cause defects in the formation of both the SAM and cotyledons, which could suggest that both genes are involved in apical domain formation. Both *GRK* and *PAS* were found to be allelic to *ACC1*, which encodes an acetyl-CoA carboxylase (Baud et al., 2004). ACCase catalyzes the ATP-dependent formation of malonyl-CoA, and it was suggested that the lack of malonyl-CoA is responsible for the developmental defects observed in *acc1/gk/pas3* mutants. However, the exact
molecular mechanisms controlling the apical domain organization are not clear. The generation of stem cells in the shoot apical meristem requires expression of the *WUSCHEL (WUS)* gene, which encodes a homeodomain protein whose transcript can be detected as early as the 16-cell embryo (Mayer et al., 1998). The maintenance of the stem cell population in the SAM requires a *WUS-CLAVATA (CLV)* circuit (Schoof et al., 2000). *CLV* function consists of three genes, *CLV* 1-3. *CLV3* encodes a small secreted polypeptide that can bind to the *CLV1/CLV2* receptor complex (reviewed in Clark, 2001). When the number of stem cells is increased, more *CLV3* is released from the stem cells and the *CLV3* binds to the *CLV1/CLV2* receptor kinase in underlying layers. This causes fewer cells expressing *WUS*, which in turn decreases the production of the stem cell population. On the other hand, when the number of stem cells is decreased, less *CLV3* is released and more cells express *WUS*, therefore the production of the stem cell population is increased. *SHOOT MERISTEMLESS (STM)* encodes a homeodomain transcription factor that positively regulates the formation of the SAM (Long and Barton, 1998). *STM* maintains the undifferentiated state of the stem cells of the SAM by preventing the expression of *ASYMMETRIC LEAVES1 (AS1)* in those cells (Byrne et al., 2000). Other genes, such as *ZWILLE (ZLL)/PINHEAD (PIN)* (McConnell and Barton, 1995; Moussian et al., 1998) and *ARGONAUTE1 (AGO1)* (Fagard et al., 2000), both encode proteins that act within the Arabidopsis miRNA pathway and are also important for the shoot meristem formation (Vaucheret et al., 2004; reviewed in Laux et al., 2004). *CUP-SHAPED COTYLEDON 1(CUC1), CUC2* (Aida et al., 1997) encode putative NAC-domain transcription factors that mark the boundary of SAM and cotyledons. *CUC1* and *CUC2* function upstream of *STM* and activate *STM* in the appropriate region that results in the separation of cotyledon primordia (Aida et al., 1999).

Mutation of *FACKEL (FK)* that encodes a sterol C-14 reductase, first shows defects in the development of central domain, but later affects apical and basal domains as well (Jang et al., 2000; Schrick et al., 2000). Genetic analysis of other genes including *HYDRA1*, which encodes a Δ8-Δ7 sterol isomerase
(Topping Jennifer et al., 1997) and STEROL METHYLTRANSFERASE1 (SMT1) (Diener et al., 2000) have similar phenotypes. These observations suggested that sterols might play important roles in embryo patterning.

**MONOPTEROS** (MP) (Hardtke and Berleth, 1998), BODENLOS (BDL) (Hamann et al., 1999) and AUXIN RESISTANT 6 (AXR6) (Hobbie et al., 2000) encode proteins involved in auxin signal transduction. Mutations of these genes cause abnormalities in the basal domain. **MP** encodes a putative transcription factor of auxin response factor (ARF) family, while **BDL** and **AXR6** encode nuclear proteins of the indoleacetic acid (IAA) family involved in auxin signaling via ubiquitin-mediated protein degradation pathway (reviewed in Leyser, 2002).

### 1.1.3.2 Radial patterning

Radial patterning dictates the inner-outer arrangement of different layers of cells. In Arabidopsis, both the embryonic root and hypocotyl share similar developmental programs and radial structures. Genes involved in radial patterning also have been identified. **WOODEN LEG** (WOL) (Scheres et al., 1995)/**CYTOKININ RECEPTOR1** (CRE1) (Inoue et al., 2000) encodes a two-component histidine kinase (Mahonen, 2000) that is involved in cytokinin signal transduction and controls cell division in the root primordium. Mutation of this gene leads to loss of phloem development. The periclinal division in the ground tissue is regulated by two genes, **SHOOT ROOT** (SHR) and **SCARECROW** (SCR) (Sheres et al., 1995). Both genes encode putative transcription factors that were found to be important for the asymmetric cell division and specification of the development of endodermis and cortical cells (Helariutta et al., 2000). **SHR** is expressed in the vasculature tissues and the protein product can move to the surrounding ground tissue to regulate the expression of **SCR** (Yasuda et al., 2001). Two genes, **ARABIDOPSIS THALIANA MERISTEM LAYER1** (ATML1) (Lu et al., 1996) and **PROTODERMAL FACTOR2** (PDF2) (Abe et al., 2003) both encode HD-GL2 class homeodomain transcriptional factors critical for the specification of the epidermal cells.
1.1.3.3 Maternal effect

Embryogenesis occurs within maternal tissues, and evidence from mutants in *SUSPENSOR1 (SUS1) / SHORT INTEGUMENTS1 (SIN1) / CARPEL FACTORY (CAF) / DICER-LIKE1 (DCL1)* that encode a Dicer-like protein involved in post-transcription regulation of mRNA (Jacobsen et al., 1999; Xie et al., 2003; Golden et al., 2002) suggested that there is female sporophytic effect on embryogenesis. The *Polycomb*-group (PcG) genes *MEDEA (MEA)* (Grossniklaus et al., 1998), *FERTILIZATION INDEPENDENT ENDOSPERM (FIE)* (Ohad et al., 1996) and a zinc finger protein *FERTILIZATION INDEPENDENT SEED2 (FIS2)* (Chaudhury et al., 1997; Luo et al., 1999) regulate embryo development at least in part by controlling the expression of a MADS-box gene *PHERES1 (PHE1)* (Köhler et al., 2003).

1.1.3.4 Suspensor effect

The suspensor is derived from the basal cell of the first division of the zygote and in Arabidopsis forms a file of 7-9 cells. The suspensor has the potential to develop into an embryo. Mutant analysis suggested that the capability may be inhibited by the normal developmental process of embryogenesis (Schwartz et al., 1994). Mutation of genes including *TWIN1* (Vernon and Meinke, 1994), *TWIN2* (Zhang and Somerville, 1997), *SUSPENSOR2 (SUS2)*, *SUSPENSOR3 (SUS3)* (Schwartz et al., 1994), *RASPBERRY1*, *RASPBERRY2* (Yadegari et al., 1994) and *RASPBERRY3 (RSY3)* (Apuya et al., 2002), *AtDBR1* (Wang et al., 2004a) can cause abnormal embryogenesis, which in turn disrupts suspensor development. The biochemical mechanisms of these genes vary and the stages of the embryo developmental abnormalities vary, but none of them are known to be embryo development specific regulatory factors. *TWN2* encodes a putative valyl-tRNA synthetase (Zhang and Somerville, 1997). SUS2 is a spliceosome assembly factor (Meinke, 1996). *RSY3* encodes a protein may sorted into the chloroplast thylakoid membrane. *AtDBR1* encodes a putative lariat debranching-
like enzyme involved in intron degradation. The molecular identities of RASPBERRY1, RASPBERRY2, TWIN1 and SUS3 are unknown at present.

1.1.3.5 LEC genes

During embryo maturation and desiccation, the embryo goes through biochemical and genetic reprogramming to reach a quiescent stage (Goldberg et al., 1994). Progression through this stage is subject to regulation by many transcription factors, such as LEAFY COTYLEDON (LEC) genes LEC1 (Meinke, 1992), LEC2 and FUSCA3 (FUS3) (Keith et al., 1994). The mutations of these genes result in premature exit from embryo development. The defective mutant embryos can be rescued in culture. But the cotyledons of the rescued seedlings have leaf-like traits because trichomes are present (Meinke et al., 1994). LEC genes are required for cotyledon formation and maturation processes including storage product accumulation, desiccation tolerance and dormancy maintenance (Harada, 2001). LEC1 encodes a HAP3 subunit of the CCAAT binding transcription factor (Kwong et al., 2002). Ectopic postembryonic expression of the LEC1 gene in vegetative cells induces the expression of embryo-specific genes and initiates formation of embryo-like structures (Lotan et al., 1998). Both LEC2 and FUS3 encode B3 domain transcription factors primarily expressed during embryogenesis (Stone et al., 2001; Luerssen et al., 1998). The ectopic postembryonic expression of LEC2 was also found to be able to confer embryogenic competence to vegetative cells (Stone et al.). Transient assays showed that FUS3 is sufficient to activate genes usually expressed during maturation (Reidt et al., 2000). Because the important roles LEC genes play in embryogenesis, LEC genes are suggested to be central regulators of embryogenesis (Harada, 2001).

1.1.4 Somatic embryogenesis

The research on zygotic embryogenesis has led us to understand more about plant embryo development. However, fertilization and subsequent embryo development normally occur within layers of maternal tissues. A detailed analysis
of embryogenesis has been hampered by this inaccessibility. This difficulty can be partially overcome by using somatic embryos (SE) obtained via in vitro culture. The SEs are morphologically and physiologically similar to zygotic embryos (reviewed in Zimmerman, 1993; Dodeman et al., 1997).

In plants, embryogenesis from cells other than the zygote can be induced by growth hormones or other environmental conditions. Somatic embryogenesis can be defined as the development from somatic cells or structures that follow a differentiation pattern which leads to a body pattern resembling that of zygotic embryos (Emons, 1994). The embryos derived from somatic cells can regenerate whole viable plants. This phenomenon has been first demonstrated in carrot, then later in alfalfa and many other species (Dodeman et al., 1997).

Not all somatic cells can give rise to embryos. Those that can are named embryogenic cells. How a somatic cell becomes embryogenic is not well understood. Successful generation of somatic embryos depends on many factors, including explant types, plant growth hormones such as auxin and cytokinin, light conditions and stress conditions (Gaj, 2004). Largely somatic embryogenesis is a stochastic process. Nonetheless, highly reproducible protocols for somatic embryogenesis induction in Arabidopsis have been established (Ikeda-Iwai et al., 2002; Mordhorst et al., 2002). Somatic embryos have a similar mechanism of morphogenesis as their zygotic counterpart. In zygotic embryos, the first asymmetrical division gives rise to two cells, one apical cell and one basal cell. The basal cell develops into a suspensor structure that connects the embryo proper with the maternal tissue. Despite the absence of real suspensor cells in somatic embryos, there are other similar structures that may function as suspensors in somatic embryos (Emons, 1994). In addition, in culture tissue, the cell that forms the somatic embryo is the cytoplasm-dense cell derived from the first division of the embryogenic cell, similar to the formation of the embryo proper which is derived from the apical cell that is cytoplasmically dense (Emons, 1994). Furthermore, although it is suggested that a regular pattern of embryogenic cell divisions is not required for patterning in somatic embryos, all
apical-basal and radial pattern elements were demonstrated to be present in somatic embryos, including the proper arrangement of all the structures such as shoot and root meristem, vasculature tissue and cotyledons (Mordhorst et al., 1998b). Following morphogenesis, zygotic embryos go through maturation, desiccation and become dormant at the end of the embryogenesis. Somatic embryos, on the other hand, generally do not acquire desiccation tolerance and become dormant. However, somatic embryos are able to synthesize certain storage proteins during maturation stage (Dodeman et al., 1997).

In Arabidopsis, many genes were identified to play a role in somatic embryogenesis. Ectopic expression of either \textit{LEC1} or \textit{LEC2} promotes somatic embryo formation on the vegetative tissues of the plant (Lotan et al., 1998; Stone et al., 2001). \textit{AtSERK1} encodes a leucine-rich repeat (LRR) transmembrane receptor-like kinase (RLK) that is highly expressed during somatic embryogenic cell formation in culture and during early zygotic embryogenesis. Overexpression of \textit{AtSERK1} induced a significant increase of embryo production in the culture when compared with wild type (Hecht et al., 2001b). \textit{BBM} encodes a putative transcription factor containing an AP2/ERF DNA binding domain and can also induce somatic embryogenesis in the pollen grain culture system and other postembryonic tissues (Boutilier et al., 2002). \textit{AGL15} encodes a MADS domain transcription factor that is preferentially expressed in tissues that are developing in an embryonic mode (Heck et al., 1995; Rounsley et al., 1995; Perry et al., 1996; Perry et al.). Ectopic expression of \textit{AGL15} can induce somatic embryogenesis in some contexts (Harding et al., 2003). \textit{AtGA2ox6} encodes a GA-2 oxidase that is involved in GA metabolism and constitutive expression of \textit{AtGA2ox6} can enhance somatic embryo production from shoot apices (Wang et al., 2004b).

Conversely, loss of function of some genes can induce somatic embryogenesis. The roots of \textit{pk/l} seedlings express embryonic characteristics and can form somatic embryos (Ogas et al., 1997). \textit{PKL} encodes a CHD chromatin remodeling factor that can serve as a component of transcriptional repressor complexes.
(Ogas et al., 1999). The embryos of \textit{pt/amp1, clv1} and \textit{clv3}, all have enlarged shoot apical meristems (SAM) that can produce somatic embryos in culture, and \textit{pt clv} double mutant have even larger SAM and enhanced effects on somatic embryogenesis. It was suggested that the \textit{PT} and \textit{CLV} genes act in independent pathways that control SAM size and an increased SAM may be responsible for facilitated establishment of somatic embryogenesis in \textit{Arabidopsis} (Mordhorst et al., 1998b). However, the relationship between the size of the SAM and the capacity of somatic embryogenesis induction is somewhat doubtful, because embryos lacking a SAM such as \textit{stm, wus} and \textit{zll/pnh} also can induce somatic embryogenesis (Mordhorst et al., 2002). Furthermore, it was reported that cotyledon tissue might a play major role in the formation of somatic embryos (Raghavan, 2004).

1.1.5 Summary of embryogenesis

Unlike animals, in which cell lineage plays critical role in the embryo patterning, positional cues and cell-cell communication play more important roles in the plant embryogenesis (reviewed in Laux et al., 2004). How a single zygotic cell develops into a multi-cellular whole plant consisting of various tissues and organs is still unclear. The imposition and perception of patterning information, such as auxin and cytokinin gradients in the embryo patterning remain elusive. Furthermore we still do not know how cell types are specified even though transcription factors that control cell fate in some tissues have been identified. Obviously, more work needs to be done in order to understand the developmental regulatory network involved in embryo development processes. Mutant isolation is an important way to identify genes in embryo development and is an ongoing effort. A large scale data set of genes that are required during embryogenesis have been collected. Initial analysis found that the loss-of-function of 250 \textit{EMB} (embryo) genes have embryo phenotypes and thus were suggested to be required for normal embryo development in \textit{Arabidopsis} (Tzafrir et al., 2004). Of these genes, a few encode transcription factors while most encode basal cellular function components. At the same time, many of them are
not embryo-specific (Tzafrir et al., 2004). In addition, it is common that a gene will have pleiotropic effects. These genes may provide important clues about the genetic regulation of embryogenesis; however, it also must be noted that due to redundancy of the gene functions, many genes without obvious phenotypes may not be included in this dataset, but may perform essential roles in embryogenesis.

1.2 AGL15, a MADS domain protein

AGAMOUS LIKE 15 (AGL15) encodes a protein that is a member of the MADS domain protein family. AGL15 was initially isolated by using differential display and Brassica napus tissue to identify genes that are specifically expressed during embryogenesis. The Arabidopsis ortholog of AGL15 was also isolated (Heck et al., 1995). Arabidopsis AGL15 was isolated concurrently using PCR and degenerate primers for the MADS box (Rounsley et al., 1995).

1.2.1 The MADS family

MADS box genes encode a family of eukaryotic transcriptional regulators that are widely found in yeasts, animals and plants. In Arabidopsis thaliana, there are 107 MADS box genes (Parenicová, 2003). The name MADS comes from the four founding members of the family, MINICHROMOSOME MAINTENANCE 1 (MCM1) (yeast: Saccharomyces cerevisiae), AGAMOUS (AG) (plant: Arabidopsis thaliana), DEFICIENS (DEF) (plant: Antirrinum majus) and SERUM RESPONSE FACTOR (SRF) (human: Homo sapiens) (Schwarz-Sommer et al., 1990). The MADS domain is conserved among all the family members, while the remaining sequences vary significantly. In animal and yeasts, MADS domain proteins can be grouped into two subfamilies based on their sequence similarities within the MADS domain: one is the MEF2 type (for MYOCYTE-SPECIFIC ENHANCER FACTOR2 from human), another is the SRF type (Sharrocks Andrew and Shore, 1995). Further phylogenetic analysis indicated that the two subfamilies also exist in plants and they are called Type I (SRF type) and type II (MEF2 type) (Alvarez-Buylla et al., 2000b) as diagrammed in Figure 1.2.
Figure 1.2 Type I (SRF type) and type II (MEF2) type MADS domain proteins
1.2.2 The MADS domain

MADS domain proteins can bind DNA and recognize a conserved sequence called a CArG motif with a consensus of CC(A/T)$_6$GG, although having different specificities (Shore and Sharrocks, 1995; Riechmann and Meyerowitz, 1997).

Crystal structure of SRF, MCM1-α2, MEF2A interacting with DNA ligands have been resolved (reviewed in Messenguy and Dubois, 2003). The structural studies found that the conformation of the MADS domains of these different proteins and the forms of their interaction with DNA are conserved (Messenguy and Dubois, 2003). A representative structure of the MEF2A-DNA is shown in Figure 1.3 as determined by (Huang et al., 2000). The interactions with the DNA are confined to the MADS-box (residues 1 to 58) and the MADS domain is also required for dimer formation. The MADS domain-DNA complex formation involves a certain degree of DNA bending, which is thought to play a role in correct formation of the transcription factor complexes (Messenguy and Dubois, 2003). The MADS domain-DNA complexes of both Type I and Type II MADS proteins from human and yeast have a similar conformation (Messenguy and Dubois, 2003). Like yeast and animal MADS domain proteins, plant MADS domains also bind to DNA as dimmers, recognize CArG motifs and introduce DNA bending (Riechmann and Meyerowitz, 1997).

1.2.3 MADS box genes in animals and yeasts

Proteins encoded by MADS box genes are involved in a diverse range of biological activities in eukaryotic organisms. In yeast, they are involved in cell-type specific transcription, cell cycle response and arginine metabolism. In animals and humans, they play important roles in mitogenic responses and muscle development.

In *Saccharomyces cerevisiae*, MCM1 and ARG80 belong to the Type I MADS sub family while RLM1 and SMP1 belong to Type II (Messenguy and Dubois, 2003). MCM1 plays a key role in cell-type-specific transcription and pheromone
Figure 1.3 MEF2A-DNA complex

The structure data was retrieved from Entrez's Molecular Modeling Database (MMDB) (Chen, 2003). The structure ID is 1C7U. Shown here is the MADS domain in complex with its DNA target. The primary DNA-binding elements are the two α-helices, one from each monomer. The two β-sheets and MEF2S subdomain (the α-helix C-terminal to the two β-sheets) are important for dimer formation. The N-terminal extension of the MADS domain further promotes protein-DNA interaction.
response (Shore and Sharrocks, 1995). MCM1 is essential for cell-type specific genes in the three cell types of the yeast Saccharomyces cerevisiae, the haploid a and α, and the diploid a/α types. In α-type cells, MCM1 interacts with α2 repressor to directly repress the expression of α-type cell-specific genes; on the other hand, MCM1 activates α-type cell specific genes by interacting with activator α1. In a-type cells, α-type genes are not transcribed due to the absence of the α1 activator (Shore and Sharrocks, 1995). In addition, in the yeast, MCM1 interacts with STE12 and is critical for the mating control of a and α type cells to form an a/α cell. MCM1 has pleiotropic effects in the yeast cell. MCM1 plays important roles in transcriptional regulation of cell-cycle-dependent genes, minichromosome maintenance, recombination, TY transcription, arginine metabolism, and osmotolerance (Messenguy and Dubois, 2003). ARG80 is important for arginine metabolism in yeast (Shore and Sharrock, 1995). RLM1 is involved in cell wall integrity control and SMP1 regulates osmotic stress response. In other species of yeast, MADS proteins play different roles in the cell (Messenguy and Dubois, 2003).

SRF is a type I MADS domain protein that is involved in the regulation of immediate-early genes and muscle-specific gene transcription (Shore and Sharrocks, 1995). The immediate-early genes are a group of genes that include c-fos, β-actin and junB whose transcripts are transiently induced by extracellular mitogenic stimuli without de novo protein synthesis of their transcriptional regulators (Shore and Sharrocks, 1995; Messenguy and Dubois, 2003). The regulation of these various genes requires direct or indirect interaction of SRF with other transcription factors including SRF ternary complex factors (TCF) of ETS family, homeodomain proteins, high mobility group (HMG) factor SSRP1 (Messenguy and Dubois, 2003). SRF also plays an important role in muscle development by interacting with myogenic specific transcription factors including MyoD, TEF1 and GATA4 (Messenguy and Dubois, 2003).

In mammals, there are four MEF2 genes identified named as MEF2A, MEF2B, MEF2C and MEF2D (Shore and Sharrocks, 1995). A number of homologues also
have been identified in other animals such as *Xenopus, Drosophila* and *C. elegans* (Shore and Sharrocks, 1995; Messenguy and Dubois, 2003). In humans, *MEF2A*, *B* and *D* are expressed in wide variety of tissues while *MEF2C* was found to be specifically expressed in muscle, brain and spleen tissues (Shore and Sharrocks, 1995). *MEF2* is essential for the muscle development in mouse and *Drosophila* (Messenguy and Dubois, 2003). Similar to other MADS proteins in yeast and SRF, *MEF2* also interacts with different proteins to regulate different downstream target genes (Messenguy and Dubois, 2003). *MEF2* promotes transcription of muscle-specific genes by physically interacting with Myo-D protein (Molkentin et al., 1995). The interaction of *MEF2* and a cell-specific GATA4 protein leads to activation of target genes (Morin et al., 2000).

### 1.2.4 MADS box genes in plants

In plants MADS box encoded proteins serve as homeotic regulators to specify floral organ identities and as well as having other broader roles in plant development (Riechmann and Meyerowitz, 1997). There are 107 MADS genes in *Arabidopsis* (Parenicová et al., 2003) and 71 in *Rice* (*Oryza sativa*) (De Bodt et al., 2003). Most of *Arabidopsis* MADS box genes (101 out of 107) were expressed and many can be found in different tissues (roots, leaves, inflorescences or siliques) at different stages, which would suggest that MADS domain proteins may play different roles in a broad range of plant growth and developmental stages.

#### 1.2.4.1 Type I plant MADS box genes

Type I MADS box genes can be found both in animals and plants (Alverez-Buylla et al., 2000). In plants, they are also referred to as M type MADS domain proteins (Parenicova, 2003; Kofuji, 2003). Of the 107 MADS genes in *Arabidopsis*, 67 belong to Type I and 39 belong to Type II (Parenicová et al., 2003) and one remained ambiguous and do not belong to either subfamilies. Type I MADS box genes can be further grouped into four subfamilies, Mα (25 genes), Mβ (20 genes), Mγ (16 genes) and Mδ (6 genes) (Parenicová et al., 2003). However, the
functions of most of the Type I MADS box genes are poorly understood. Only one Type I MADS box gene was functionally characterized. *PHERES1 (PHE1)*, a Type I MADS box gene plays a crucial role in embryo development. *PHE1* gene can be repressed by Polycomb Group proteins MEA/FIE and this regulation is critical for Arabidopsis seed development (Köhler et al., 2003). Type I genes are less conserved than Type II genes (Parenicová et al., 2003; Nam et al., 2004) and might be under less evolutionary constraints but functionally more redundant than their type II counterparts.

### 1.2.4.2 MIKC-type MADS box genes

In plants, MADS box genes were first identified in *Arabidopsis thaliana* and *Antirrhinum majus* (Riechmann and Meyerowitz, 1997). These original members were found to belong to the Type II subfamily. Compared with animal and yeast Type II MADS proteins, proteins encoded by most of the plant Type II genes contain three additional plant-specific domains: an intervening (I) domain (~30 a.a.), a keratin-like coiled-coil (K) domain (~70 a.a.), and a variable length C-terminal (C) domain (Figure 1.2) and they are referred to as MIKC type proteins. The MIKC type genes have been identified in most major evolutionary lineages of green plants such as angiosperms, gymnosperms, ferns, and mosses and many of them are functionally characterized (Johansen et al., 2002; De Bodt et al., 2003). It has been suggested that I and K domains be involved in protein-protein interactions, while the C domain might possess transactivation activity or be involved in ternary protein complex formation (Riechmann and Meyerowitz, 1997; Egea-Cortines et al., 1999; Honma and Goto, 2001).

In flowering plants, MIKC type MADS domain proteins are well known for their diverse roles in plant development, including floral organ identity specification, control of flowering time and many other developmental programs (Riechmann and Meyerowitz, 1997). The identification and functional characterization of several MIKC MADS box genes helped established the classic ABC model for flower development (Meyerowitz et al., 1991).
Floral organ identity specification - The ABC (E) model

The Arabidopsis flower consists of four whorls of organs (from outer to inner): whorl 1, four sepals; whorl 2, four petals; whorl 3, 6 stamens and whorl 4, two fused carpels. Molecular and genetic studies in Arabidopsis thaliana and Antirrhnum majus (snapdragon) showed that loss-of-function mutations of certain genes had defects in specific whorl(s). According to functions deduced from loss-of-function mutation phenotypes, the homeotic genes were grouped into three classes: class A, controls whorls 1 and 2; class B, controls whorls 2 and 3 and class C, controls whorls 3 and 4 (Riechmann and Meyerowitz, 1997). In Arabidopsis, class A genes includes APETALA1 (AP1) and APETALA2 (AP2); class B genes includes APETALA3 (AP3) and PISTILLATA (PI) and the class C gene is AGAMOUS (AG) (Riechmann and Meyerowitz, 1997). Except for AP2, all others are MIKC type II MADS box genes. ABC genes are necessary and sufficient for the formation of flower organs within the floral domain; mutation of any one of the ABC genes will cause abnormality of flower structure and ectopic expression of any one of the gene causes corresponding gain-of-function phenotype within the context of flower. However, they are not sufficient for floral organ formation outside of the floral domain; for example, vegetative tissues are not converted to floral organ by overexpression of ABC genes (Mizukami and Ma, 1995; Krizek and Meyerowitz, 1996). The discovery of SEPALLATA (SEP) genes (Pelaz et al., 2000) led to our further understanding of the molecular mechanisms controlling flower organ specification and revision of the ABC model (Goto et al., 2001; Honma and Goto, 2001; Theissen, 2001). Single or double mutants of the SEP1/SEP2/SEP3 didn’t show obvious phenotypes, while the triple mutant of sep1 sep2 sep3 consists of floral organs very similar to ag ap3 pi triple mutant (bc double mutant) in which all floral organs develop as sepals (Pelaz et al., 2000). The result suggested that B and C gene products are not functional in the sep triple mutant thus SEP genes are necessary for activities of the class B and C genes. Furthermore, combined action of the SEP genes, together with the A and B genes, is sufficient to convert leaves into petals (Pelaz et al., 2001). In addition, SEP4 gene also has been shown have redundant
functions to other three *SEP* genes (Ditta et al., 2004). The quadruple mutant converts all floral organs to leaves, similar to the phenotype found for *a b c* triple mutant. Therefore, the *SEP* genes are also referred to as class E genes for flower development. SEP proteins were suggested to interact with B and C function proteins in tetrameric complexes and that then can bind to the CArG motifs of target genes and control floral organ identities (Theißen, 2001). The revised ABC model became the ABCE model, in which class A+E genes alone specify sepals, classes A+B+E specify petals, classes B+C+E specify stamens and classes C+E specify carpels (Figure 1.4). Studies on petunia (*Petunia hybrida*) MADS box genes *FBP7* and *FBP11* revealed their functions in ovule development (Riechmann and Meyerowitz, 1997). In Arabidopsis, *SHATTERPROOF1* (*SHP1*), *SHP2*, and *SEEDSTICK* (*STK*) / *AGAMOUS LIKE 11* (*AGL11*) promote specification of ovule identity (reviewed in Skinner et al., 2004).

**Floral meristem identity specification**

*AP1* also functions as a floral meristem identity gene (Riechmann and Meyerowitz, 1997). Together with a non-MADS domain protein LEAFY (*LFY*) (Weigel et al., 1992), *AP1* specifies the lateral primordia to become a reproductive organ - flower rather than vegetative shoot. Ectopic expression of *AP1* triggers flower formation while the *ap1* mutant can partially convert flowers into shoots (Irish and Sussex, 1990; Bowman et al., 1993). *CAULIFLOWER* (*CAL*) (Bowman et al., 1993; Kempin et al., 1995), another MADS box gene partially redundant with *AP1*, also plays a role in floral meristem identity specification. A MADS box gene *FRUITFULL* (*FUL*) (Gu et al., 1998) also has been shown to share partially redundant function with *AP1* and *CAL* in floral organ identity specification. A recent report suggested that *AGAMOUS LIKE24* (*AGL24*) promote the formation of inflorescence meristem identity and its transcription was repressed by *LFY* and *AP1* (Yu et al., 2004).
Figure 1.4 A revised ABC model for flower development

A schematic diagram of Arabidopsis flower structure is represented, the revised ABC model is modified from Goto et al. (2001). Floral organ identities are specified by the combined activities of four classes of proteins, A, B, C and E. Class A and C are mutually exclusive so as to keep each other out of its own activity domain. The quartet model proposes that tetrameric protein complexes are formed among four MADS domain proteins and function to specify floral organ identity.
Flowering time control

Flowering time was suggested to be controlled by four main pathways: long-day photoperiod, GA, autonomous and vernalization (Jack, 2004). A MADS box gene, *FLOWERING LOCUS C* (*FLC*) (Michaels and Amasino, 1999) plays central role in flowering time control (Sheldon et al., 2000). Four other MADS box genes also have been found to play a role in the regulation of flowering time. *AGAMOUS LIKE 20* (*AGL20*) / *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) (Lee et al, 2000; Samach et al, 2000) and *FRUITFUL* (*FUL*) (Gu et al., 1998) promote flowering, while *MADS AFFECTING FLOWERING1* (*MAF1*) / *FLOWERING LOCUS M* (*FLM*) (Ratcliffe et al., 2001; Scortecci et al, 2001) and *SHORT VEGETATIVE PHASE* (*SVP*) (Hartmann et al., 2000) can inhibit flowering.

Other developmental programs

A MADS box gene *ANR1* is expressed in roots and regulates nitrate-induced root architecture arrangement (Zhang and Ford, 1998). Some other MADS box genes were expressed in different tissues, such as *AGL16* expression in trichomes and guard cells; *AGL18* expression in endosperm and pollen and *AGL19* expression exclusively in roots (Alvarez-Buylla et al., 2000a). Several other MADS genes, such as *AGL12*, *AGL79*, *AGL13*, *AGL14*, *AGL17* and *AGL21* are also preferentially expressed in roots when compared with leaves, inflorescence and siliques (Parenicocvá et al., 2003). The expression pattern differences of MADS genes might reflect their function differences during plant development. One gene, *AGAMOUS LIKE15* (*AGL15*) was found to be preferentially expressed in developing embryos (Heck et al., 1995; Rounsley et al., 1995) and its ectopic expression can promote somatic embryogenesis (Harding et al., 2003) as well as having other effects on plant vegetative and reproductive growth (Fernandez et al., 2000; Fang and Fernandez, 2002).
1.2.5 *AGL15*, A MADS box gene preferentially expressed in embryos

*AGL15* contains all four domains of the typical MIKC type MADS domain proteins, but it is quite different from other MADS proteins not only because of its expression pattern, but also because of its phylogenetic position in the MIKC subfamily. In a phylogenetic analysis based on MADS domain using genomic data of MADS box genes from animals and plants, *AGL15* itself was grouped as a clade (Alvarez-Buylla et al., 2000b). For 79 type II MADS genes from different plants, Nam et al. (2004) were able to group them into 15 clades on one tree, but *AGL15* was not able to be assigned to any of the clades. The only other MIKC protein that is loosely similar to *AGL15* is *AGL18* (Parenicová et al., 2003; Martínez-Castilla et al., 2003). The lineage difference of *AGL15* from other MIKC MADS domain proteins may suggest it performs a distinct function in plant development and evolution.

1.2.5.1 Expression pattern of *AGL15*

*AGL15* is to date the only known MIKC type MADS box gene that is preferentially expressed during embryogenesis. However, the activity of *AGL15* is not restricted to the embryo. In *Brassica napus*, by using RNA gel blot analysis it was found that *AGL15* mRNA accumulated as early as globular stage, peaked at about torpedo stage, then gradually decreased during embryo maturation. In the torpedo stage *Brassica* embryos, *AGL15* transcripts could be detected in all types of embryo cells by *in situ* hybridization (Heck et al., 1995). Immunolocalization analysis with *AGL15*-specific antibody in *Brassica* and *Arabidopsis* embryos further revealed that *AGL15* was present in all the embryo tissues, even in the egg cell before fertilization. However, *AGL15* was localized to the cytoplasm before fertilization. Shortly after fertilization, *AGL15* moved into the nucleus (Perry et al., 1996). *AGL15* was also reported to accumulate in embryonic tissues from diverse origins. Immunolocalization using *AGL15*-specific antibodies revealed that immunoreactive proteins were present in apomictic embryos from dandelion (*Taraxacum officinale*), somatic embryos derived from
microspore and organs produced in precocious germination of oilseed rape (*Brassica napus*), somatic embryos from alfalfa (*Medicago sativa*) and embryonic tissue of cotyledon-like structures formed at shoot apex in Arabidopsis mutant *xtc2* (Perry et al., 1999). These results demonstrated a correlation between presence of AGL15 or putative orthologs and development in embryonic mode.

AGL15 is also expressed in tissues other than embryos. In Arabidopsis and Brassica, AGL15 protein was detected in the nuclei of endosperm cells, suspensor cells before transition stage and even seed coat cells (Perry et al., 1996). Transcript of AGL15 also could be detected in other non-seed tissues including inflorescence apices, young floral buds, young seedlings and roots (Heck et al., 1995; Rounsley et al., 1995). Transcript levels of AGL15 in these tissues were at least 10-fold less than in developing embryos (Heck et al., 1995). AGL15 protein also has been detected by immunolocalization in all cell layers of very young shoot apical meristems and in young leaf primordia (Fernandez et al., 2000, Harding et al., 2003). In young seedlings, AGL15 is present in the shoot apex and accumulates to level comparable to that of the developing embryo; while in leaves and older shoot apical meristems, only traces of protein are detected (Fernandez et al., 2000).

After germination, AGL15 promoter activity was shown to be temporally and spatially regulated (Fernandez et al., 2000). The AGL15 promoter was active during the entire vegetative stage in shoot apical meristems, leaf primordia and leaf bases. However, the activities of the AGL15 promoter in these tissues are much lower than that of the young seedling shoot apical meristem and embryos (Fernandez et al., 2000). In addition, even though AGL15 was present in the cotyledon of embryos, after germination AGL15 promoter activity disappeared in the cotyledon while in leaves, AGL15 promoter activity was initially present in the whole young leaf, later the activity was confined to the leaf base (Fernandez et al., 2000). Furthermore, AGL15 promoter reporter activity was shown to be at the base of the young flower buds, but could not be detected in any tissues of inflorescence meristem or open flowers (Fernandez et al., 2000).
1.2.5.2 Biological role studies of AGL15

The uniqueness of AGL15 amongst the MIKC MADS domain proteins suggests that AGL15 may play different roles than other family members in plant development. In addition, the preferential expression in embryos but lower level postembryonic expression pattern suggested that AGL15 functions at various developmental stages.

Studies on the effects of ectopically expressed AGL15 supported the hypothesis that AGL15 might be important for embryogenesis (Harding et al., 2003). In Arabidopsis, zygotic embryos from transgenic plants ectopically expressing AGL15 (AGL15 driven by Cauliflower Mosaic Virus (CaMV) 35S promoter; 35S:AGL15) showed a significant increase of secondary embryo formation in culture when compared with those from wild type plants. Even more significantly, over 80% of sub-cultured secondary embryos bearing 35S:AGL15 were able to maintain embryonic state after 6-7 weeks, while less than 10% for wild type did (Harding et al., 2003). One line of the sub-cultured embryo tissue carrying 35S:AGL15 has maintained this capacity continuously for more than 8 years to date. In addition, ectopic expression of AGL15 was also shown to promote somatic embryo production from cultured seedlings in the presence of exogenous auxin (Harding et al., 2003). The constitutive expression of soybean (Glycine max) AGL15 ortholog may have a positive effect on somatic embryo production and plantlet regeneration, which potentially can be important for the transgenic soybean production that is known to have low transformation efficiency (Tang and Perry, unpublished observation). Furthermore since AGL15 is expressed throughout the embryo development and in many other tissues, it is likely that AGL15 does not play a primary role in embryo pattern formation; instead, it might function to maintain an embryonic developmental program. However, the precise role of AGL15 in zygotic embryogenesis and the mechanisms by which ectopic expression of AGL15 promotes somatic embryogenesis remain unknown except that GA may be involved (Wang et al., 2004b).
The fact that \textit{AGL15} is dynamically expressed in tissues other than embryos implies that \textit{AGL15} is also important for the postembryonic development of plants. In fact, ectopic expression of \textit{AGL15} in Arabidopsis causes some apparent changes in plant development. In plants ectopically expressing \textit{AGL15}, the petioles of cotyledons were shorter and the cotyledons are epinastic when compared with wild type. In addition, the flowering of the plants harboring 35S:\textit{AGL15} was markedly delayed compared with that of the wild type (Fernandez et al., 2000). Furthermore, in plants that constitutively express \textit{AGL15}, processes of senescence and abscission of the perianth organs are inhibited. In non-transgenic plants, sepals and petals are shed shortly after fertilization, while in plants constitutively expressing \textit{AGL15}, sepals and petals do not abscise and are retained on the plants well beyond anthesis, even retained through silique elongation and maturation stage (Fernandez et al., 2000). Further study indicated that \textit{AGL15} does not act through ethylene to have an effect on perianth senescence and abscission (Fernandez et al., 2000). The lack of abscission of the sepals and petals is not due to the defects in abscission zone development. Other than perianth senescence and abscission, processes such as fruit maturation, flowering time, silique dehiscence and seed desiccation are also delayed in the plants carrying 35S:\textit{AGL15} (Fernandez et al., 2000; Fernandez et al., 2002). Further investigation indicated that the effect of \textit{AGL15} expression on senescence and abscission is not primary, because when ectopic expression of \textit{AGL15} was specifically targeted to the two processes, the plants failed to show the same phenotype of the 35S:\textit{AGL15} plants (Fang and Fernandez, 2002). Therefore, it was hypothesized the phenotypic changes occur in plants overexpressing \textit{AGL15} reflect changes in gene regulation that occur at a stage before obvious signs of abscission or senescence appear. However, the molecular mechanisms and direct effects of \textit{AGL15} during postembryonic phases remain unsolved.
1.3 Research proposal and strategies

Much work is still needed to decipher the molecular and genetic regulatory networks controlling different programs of the plant life cycle. The broader expression pattern of *AGL15* suggests divergent processes in which *AGL15* might be involved. To better understand how *AGL15* functions during plant development, especially during plant embryogenesis, identification of downstream target genes regulated by *AGL15* is essential. An equally important task is to characterize those genes and elucidate their roles in plant development. *AGL15* then can be connected with its biological roles by definite molecular mechanisms. Target genes of other plant MADS domain proteins have been identified but fell short of definitive evidence because these experiments mostly are based on indirect genetic information (Riechmann and Meyerowitz, 1997). Only recently, target genes of AP3/PI was identified using an inducible system (Sablowski and Meyerowitz, 1998), target genes of AP3/PI were also identified using an inducible system and microarrays (Zik and Irish, 2003) and target genes of AG were identified using microarrays plus ChIP (Gomez-Mena et al., 2005).

Furthermore, the expression of *AGL15* is temporally and spatially regulated both during embryogenesis and postembryonic phases. This regulation is of importance for the biological role of *AGL15* because for many MADS domain proteins, realms of expression largely correspond to domains of function (Riechmann and Meyerowitz, 1997). In order to properly control plant developmental programs, cells may use many mechanisms such as posttranscriptional regulation, translational control, protein modification, protein degradation, nuclear localization and cell-cell movement to control the activity of a transcriptional regulator, however transcriptional control is one of the most common and most important strategies used. Therefore, it is essential to explore the mechanisms that control *AGL15* expression in order to decipher the biological roles that this special MADS domain protein may have in plant development.
1.3.1 Chromatin Immunoprecipitation

AGL15 is a MADS domain protein that functions as a transcription factor to bind to cis regulatory elements of target genes and regulate their expression. Identification of AGL15 binding sites and genes bearing these sites can provide important clues to help us understand what roles AGL15 might play in the cell. Chromatin immunoprecipitation (ChIP) has been successfully used to identify binding sites in yeast, animals and plants (Ito et al., 1997; Orlando, 2000). ChIP protocol has the ability to provide direct evidence for an AGL15-DNA interaction in vivo and it is the best available protocol for this purpose. In this approach, living cells from tissues of interest, in our case, embryonic tissue, are fixed by using formaldehyde, and chromatin is isolated and then fragmented by sonication. The solubilized chromatin fragments are isolated and the AGL15-DNA complexes are precipitated with AGL15-specific antibody and protein A-sepharose beads. The crosslinks are then reversed and the AGL15-bound DNA fragments are purified, cloned and sequenced. The obtained sequence information can be used to search against the Genbank database and identify target genes. A more detailed protocol will be described in Chapter 2.

1.3.2 Microarray

To elucidate the function of AGL15, it is essential to identify the AGL15 binding sites and corresponding genes. However, to identify which genes are regulated by AGL15, it is necessary to identify the sets of the genes that are responsive to changes in the levels or activity of AGL15. For this purpose, there are several commonly used approaches available such as Northern blotting, slot blot and RT-PCR. A recently developed approach, DNA microarray, bears the advantage of high throughput analysis on the nearly whole genome scale, and has gained tremendous popularity to monitor gene expression change under various conditions. Microarray globally measures the change of transcript levels of almost all the genes in the genome at the same time, thus can greatly speed up the process of candidate gene identification. In microarray, either cDNAs
prepared from whole organism or oligo-nucleotides corresponding to transcripts of genes in the genome are densely fixed on the slides or other supporting materials. The total RNA or mRNA from control and experimental cells are then extracted and labeled with different fluorescent dyes and hybridized with targets on the slides. The fluorescent dye signal corresponding to transcripts level are then analyzed and results obtained. Numerous experiments have been performed and abundant Arabidopsis microarray data are available at The Arabidopsis Information Resource (TAIR) (http://www.arabidopsis.org/info/expression/ATGenExpress.jsp).

1.3.3 Identification of important cis elements in the AGL15 promoter

The function of AGL15 is not determined by only what genes it regulates, but also when and how it regulates those target genes. The way AGL15 regulates its target genes depends partly on how AGL15 itself is regulated. To address the question about how the regulator is regulated, the cis elements controlling the expression of AGL15 must be identified.

1.3.4 Specific aims of this dissertation research

The overall objective of my dissertation research was to help understand the molecular mechanisms of AGL15 in embryogenesis and plant development processes. The specific aims of this dissertation research are:

1. To adapt chromatin immunoprecipitation (ChIP) on a plant system and to identify AGL15 binding sites and corresponding genes in Arabidopsis;

2. To use a microarray approach to globally monitor transcription profiles in response to AGL15 accumulation in Arabidopsis and identify candidate target genes of AGL15;

3. To characterize the regulatory cis-elements in the AGL15 promoter and characterize potential regulatory mechanisms of AGL15 expression.
CHAPTER 2

IDENTIFICATION AND ANALYSIS OF DOWNSTREAM TARGET GENES OF AGL15 BY A CHROMATIN IMMUNOPRECIPITATION APPROACH
2.1 Introduction

MADS domain proteins can bind to DNA and recognize a consensus DNA sequence of CC(A/T)$_6$GG called a CArG motif (reviewed in Shore and Sharrocks, 1995; Riechmann and Meyerowitz, 1997). However, different MADS domain proteins have different specificities in binding sites and might contribute to regulation of different groups of genes. AGL15 is a MADS domain protein that preferentially accumulates in embryos, but is also present in other tissues (Heck et al., 1995; Rounsley et al., 1995; Perry et al., 1996; Perry et al., ; Fernandez et al., 2000). The broad expression pattern suggests that AGL15 might function in various plant developmental processes. Furthermore, the AGL15 preferred binding site is different from other MADS family members with a consensus sequence of C(A/T)$_8$G (Tang and Perry, 2003), which might indicate that AGL15 regulates a different group of genes from those regulated by other MADS family members.

Although plant MADS-box genes consist of a large group of family members and extensive studies have been performed, little is known about genes that are regulated by these MADS domain proteins, especially directly regulated genes that do not belong to the MADS box family (Riechmann and Meyerowitz, 1997). Only a few studies have been reported on the identification of direct targets of MADS transcriptional regulators. Using an inducible system, Sablowski and Meyerowitz (1998) isolated three genes encoding a NAP protein and 2 unknown proteins, respectively, as targets of the MADS heterodimer AP3/PI. Recently, Zik and Irish (2002) used microarray coupled with an inducible system to identify putative direct target genes of AP3/PI. Gómez-Mena also used the same approach to identify target genes of AG (2005). In the latter case, several direct genes were confirmed using chromatin immunoprecipitation (ChIP). In our study, ChIP protocol was used to isolate direct targets of AGL15. In ChIP, various methods can be used to perform the crosslinking step, with UV light and fromaldehyde being two commonly used reagents. ChIP based on formaldehyde
crosslinking has been widely used to identify the unsuspected downstream target genes (reviewed in Kuo and David, 1999). Formaldehyde is a reactive crosslinking agent that can crosslink both protein-protein and protein-nucleic acid complexes within 2 Å radius. The crosslinking can take place in vivo by adding formaldehyde to the living cells or tissues. The nucleophilic carbon of the formaldehyde molecule can react with the amino and imino groups of the lysine, arginine and histidine and with the exocyclic amino groups of nucleic acids, such as adenine, cytosine and guanine to form a Schiff base. The advantages of using formaldehyde to do crosslinking of protein-protein and protein-nucleic acid are that the protocol does not require expensive equipment and the crosslinking can be completely reversed by modest heat treatment (68°C). To identify the DNA targets that are bound by a DNA-binding protein such as a transcription factor, the crosslinked chromatin is precipitated by using antibody that is specific to the protein of the interest. The DNA is then isolated and analyzed by PCR. In Drosophila, it has been suggested that crosslinking by formaldehyde provides an accurate guide to the interaction of proteins with their target sites in the cells (Toth and Biggin, 2000). However, in plants, ChIP had not yet been used to identify previously unsuspected targets of a transcription factor, although a related purification using columns was used by Ito et al. to identify putative target genes of AG (1997).

To understand the biological roles of AGL15, it is essential to identify genes that are directly regulated by AGL15, that is, the target gene must be confirmed both to be bound and regulated by AGL15. Therefore, the objectives of this study are: (1) to identify the direct targets of AGL15 by chromatin immunoprecipitation; (2) to confirm the binding and regulation of the target gene by AGL15.
2.2 Results

2.2.1 Identification of AGL15 binding sites \textit{in vivo}

Chromatin immunoprecipitation requires relatively large amounts of tissue that contain AGL15 in the nuclei of cells to allow for isolation of AGL15-DNA complexes. AGL15 preferentially accumulates in the nuclei of embryo or embryonic tissues from various plants including \textit{Arabidopsis thaliana} (hereafter Arabidopsis) (Rounsley et al., 1995; Heck et al., 1995; Perry et al., 1996; 1999). The reason we used Arabidopsis to isolate direct targets of AGL15 is that Arabidopsis has been well-established as a plant model system with the whole genome completely sequenced. With the whole genome sequence database, identities and locations of the ChIP isolated DNA fragments in the genome can be determined. However, Arabidopsis embryos are small and it is not possible to obtain enough tissue for a ChIP experiment, especially at earlier stages when AGL15 accumulates to the highest level. As an alternative to zygotic embryo, Arabidopsis embryonic culture tissue (ECT) was used in most cases. Initial experiments by western blot analysis of the nuclear extracts and immunohistochemical staining using anti-AGL15 serum showed that ECT accumulated similar levels of AGL15 protein in the nuclei as zygotic embryos (Wang et al., 2002b). It was also demonstrated that an embryonic development program was maintained in the ECT, as judged by morphological and molecular evidence (Harding et al., 2003). Using the protocols as described in Harding (Harding et al.), ECT is regularly sub-cultured and maintained at a sizable population so adequate tissue is readily available for ChIP experiments.

In our study, a chromatin immunoprecipitation (ChIP) approach was developed to isolate the \textit{in vivo} binding sites of AGL15. For ChIP experiments to be successful, it is essential to have antibody that is specific for the protein of the interest. The anti-AGL15 specific immune serum has been prepared and purified previously and was available for use. The antibody was characterized and demonstrated to be highly specific for AGL15 in all cases tested (Heck et al.,
1995; Perry et al., 1996; 1999). Details of the ChIP protocol are described in section 2.4.3. A schematic representation is shown in Figure 2.1. Briefly, ECT was treated with formaldehyde to stabilize the protein-protein and protein-DNA interactions in vivo by crosslinking. The nuclei were isolated after crosslinking and chromatin was solubilized by sonication. DNA fragments that bind to proteins non-specifically were removed by pretreatment with preimmune serum and protein-A Sepharose. After pre-adsorption, part of the sample was saved as total (input) for further analysis. The remaining sample was immunoprecipitated with either anti-AGL15 immune serum (I), preimmune serum (PI) or no serum control (No) with protein A-Sepharose. An aliquot of the supernatant was saved as “post-bind” for Western blot to check AGL15 immunoprecipitation efficiency. After extensively washing, the AGL15-DNA complexes were eluted from the beads. Part of the sample was saved as “eluate” for Western blot to check AGL15 protein recovery. After centrifugation, the AGL15-DNA complexes were depleted from the supernatant but pelleted with protein A-Sepharose beads when immune serum was used. Conversely, in preimmune serum or no serum controls, AGL15 remained in the supernatant (Figure 2.2a and Wang et al., 2002).

The AGL15-DNA crosslinking was readily reversed by modest heat (68 °C). An aliquot of the DNA recovered was saved for enrichment PCR. The remaining DNA sample was modified by restriction digestion, single “G” filling and ligated with linkers. The modified DNA population was PCR amplified. An aliquot of the PCR products was analyzed using agarose gel electrophoresis (Figure 2.2b). When AGL15-specific antiserum was used for immunoprecipitation, PCR products were detected (Figure 2.2b, I); on the other hand, little or no PCR products were detected when preimmune serum (Figure 2.2b, PI) or no serum (Figure 2.2b, No) controls were used in the immunoprecipitation.

2.2.2 In vitro selection on the ChIP population

To remove some of the DNA that nonspecifically precipitated with ChIP, subsequent in vitro immunoprecipitation was performed. An aliquot of the PCR
Isolate tissue, fix with formaldehyde

Isolate nuclei and solubilize chromatin with sonication

Pre-adsorption treatment with preimmune serum and protein A-Sepharose

Immunoprecipitate AGL15-DNA complexes using Anti-AGL15 serum and protein A-Sepharose and controls

Wash, elute AGL15-DNA complexes from protein A-sepharose beads

Reverse formaldehyde crosslinks by heat, remove protein, recover DNA

Sau3A I digest DNA, add linkers and PCR amplify

Clone the modified DNA fragments, sequencing and identify target sites

Figure 2.1 Outline of the chromatin immunoprecipitation (ChIP) protocol

Note: *1. An aliquot was reserved to extract total (input) DNA sample for enrichment PCR and check size of DNA fragments after sonication. *2. An aliquot was reserved unmodified to extract immunoprecipitated (I) DNA sample for enrichment PCR.
Figure 2.2 Chromatin immunoprecipitation of AGL15

(a) Western blot to monitor AGL15 during ChIP. After elution, AGL15 protein was recovered by using anti-AGL15 serum (Elute, I). Very little AGL15 remained in the soluble fraction after precipitation (Post-bind, I).

(b) PCR analysis of the ChIP isolated DNA. DNA isolated using ChIP was purified and modified with linkers for PCR amplification. PCR amplified products were present in the sample when anti-AGL15 immune serum (I) was used while little to no products were detected when preimmune serum (PI) or no serum were used (No).
population amplified from DNA fragments co-precipitated with the AGL15 protein was used in the experiment. *In vitro* selection was reiterated three times as shown in Figure 2.3a. The full-length AGL15 protein (AGL15) and a truncated version of AGL15 lacking the MADS domain (AGL15ΔM) were recovered from *E. coli* as inclusion bodies. The Coomassie brilliant blue staining of the two proteins is shown in Figure 2.3b. Anti-AGL15 serum can recognize AGL15ΔM (AGL15 antibody was prepared using AGL15 ΔM) but AGL15 lacking the MADS domain cannot bind DNA (Heck et al., 1995). Therefore AGL15ΔM serves as a negative control to monitor the general background of the non-specific DNA recovered in immunoprecipitation. As an additional control, preimmune serum was also used to “immunoprecipitate” AGL15. As shown in Figure 2.3c, both full length AGL15 and AGL15ΔM were able to be recognized by AGL15-specific antiserum and are present after elution, but AGL15 was not present in the elute sample when preimmune serum was used. On the other hand, DNA was clearly visible in the elution aliquot of the *in vitro* selected sample that was immunoprecipitated using full length AGL15 and immune serum (Figure 2.3d, AGL15+I) but not in the control samples immunoprecipitated using AGL15ΔM and immune serum, or AGL15 and preimmune serum or no protein and immune serum (Figure 2.3d, AGL15 + PI, AGL15ΔM + I and No + I). The result indicates that the DNA fragments isolated from the *in vitro* selected population requires the DNA binding domain of AGL15.

### 2.2.3 Analysis of the isolated downstream targets

The DNA fragments isolated by ChIP were modified and amplified as described, and then cloned into an appropriate vector. The fragments isolated by *in vitro* immunoprecipitation on the ChIP population were also cloned into the same vector. Both populations were propagated in *E.coli*. Plasmids were then isolated from the bacterial clones and used for sequencing. A total of 101 clones were sequenced, 47 of them were obtained from the *in vivo* ChIP population and 54 from *in vitro* selected population. The sequence information was used to search against the Arabidopsis database (NCBI, National Center for Biological
Figure 2.3 In vitro selection of the ChIP population
(a) Schematic outline of the in vitro selection procedure.
(b) Proteins used in the in vitro selection.
(c) Western blot to monitor proteins during in vitro selection. Both AGL15 and AGL15ΔM were able to be recovered in the elute.
(d) Agarose gel analysis of the specificity of the DNA recovered from in vitro selection. DNA was present when full length AGL15 and immune serum was used in the immunoprecipitation (AGL15+I) while little or no DNA was detected when precipitate using preimmune serum (AGL15+PI) or immune serum was used to precipitate AGL15ΔM or no protein in the IP.
Information, [http://www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST) to identify where fragments were located in the genome and what genes were potentially targeted. The genes identified were putative AGL15 targets. The partial collection of the 101 isolated DNA fragments were found to be located not only in the 5’ regulatory regions that might contain potential promoters of the corresponding genes, but also in other places, such as 3’ regions, intergenic regions, introns and sometimes exons. In fact, only about 38% of all the isolated fragments were found to be located in the 5’ regions. DNA fragments corresponding to the regulatory region of the genes that may be important for embryo development were isolated, such as SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE 2 (AtSERK2) that is very similar to SERK from Daucus carota (carrot) (Schmidt et al., 1997) and ADL1A that has an embryo-defective phenotype when mutated (Kang et al., 2001). We also isolated fragments corresponding to regulatory regions of genes that are involved in other aspects of plant development, such as AtGA2ox6 (DTA1, for Downstream Target of AGL15 1) that encodes a gibberellic acid (GA) metabolic enzyme GA 2-oxidase 6 (Wang et al., 2004b) and ABF3 (DTA3) that encoding a bZIP (basic leucine zipper Zinc-finger) protein that involved in ABA response (Choi et al., 2000; Kang et al., 2002). Many fragments corresponded to regulatory regions of genes of putative or unknown functions and in many cases were represented among ESTs from developing seeds or green siliques. Several putative targets were selected for further analysis and were subjected to series of experiments to determine whether the targets are true binding sites of AGL15 and whether binding of AGL15 to the target sites confers regulation on the target gene. Indeed, GA2ox6 was showed to be a direct downstream target gene of AGL15 (Wang et al., 2004b). Here another gene, ABF3 was subjected to further analysis and results obtained are described in more detail in the following sections.

2.2.4 AGL15 interacts directly with the regulatory region of ABF3

The fragment corresponding to the regulatory region of ABF3 was isolated in the ChIP population, which suggested that AGL15 binds to the fragment in vivo. To
confirm the direct association between AGL15 and its target site in the \textit{ABF3} regulatory region, enrichment PCR was performed. The assay was able to test whether the selected fragment is specifically enriched in the ChIP population. For enrichment PCR, a pair of primers was designed to specifically amplify the target gene fragment and another pair of primers was designed to amplify control DNA fragments that were not expected to be bound by AGL15, such as the DNA fragment corresponding to the coding regions of house-keeping genes, including \(\beta-2\) \textit{tubulin} (\textit{TUB2}) (Snustad et al., 1992b), \textit{elongation factor 1-\(\alpha\)} (\textit{EF1\(\alpha\)}A1) (Axelos et al., 1989) and \textit{ubiquitin extension protein} (\textit{UBQ6}) (Callis et al., 1990). Both pairs of primers were added in the same reaction to perform multiplex PCR. The DNA templates used in the PCR reaction were the DNA population co-precipitated with AGL15 in the ChIP experiment using immune serum (I), the controls templates were DNA populations obtained in the ChIP experiment using preimmune serum (PI) or no serum (No). If a DNA fragment is bound by AGL15 \textit{in vivo}, it will be represented at a higher level (i.e. enriched) in the ChIP (I) population when compared to that in the preimmune populations (I) or to total (input) DNA, relative to the unbound control.

The fragment isolated from the ChIP population was located in the 5' region of \textit{ABF3}, and is shown as a filled box in Figure 2.4a. When multiplex PCR was performed on the immune (I) population using primers specific to the region and primers specific to \textit{UBQ6}, higher amounts of the target products were observed when compared to the reference (Figure 2.4b \textit{ABF3} vs \textit{UBQ6}, lane I ). On the other hand, when multiplex PCR were performed on the input dilutions, both target and reference were amplified to a similar level (Figure 2.4b, \textit{ABF3} vs. \textit{UBQ6}, lanes input 25x, 125x, 625x). There was no canonical MADS domain protein binding site within or near the DNA fragment isolated in the ChIP experiment. Only one non-canonical binding site (CC(A/T)\textsubscript{6}GC) was present in the fragment.

Enrichment PCR was also performed on the ChIP populations derived from tissues other than ECT. \textit{AGL15} was shown to be expressed at a lower level in
Figure 2.4 AGL15 directly bind to the regulatory of \textit{ABF3}

(a) Regulatory region of \textit{ABF3}. The filled box indicates the fragment isolated by ChIP, arrows indicate the primers used in the multiplex PCR. A non-canonical CArG box is shown.

(b) Enrichment of the \textit{ABF3} regulatory region isolated by ChIP. \textit{ABF3} denotes the \textit{ABF3} regulatory region; reference was \textit{UBQ6}. PCR products sizes in bp are shown.

(c) Enrichment of the \textit{ABF3} regulatory region depends on the amount of AGL15. Enrichment PCR were performed on the populations derived from the wild type and 35S:AGL15 transgenic flower buds and inflorescence tissue using immune serum (I) or preimmune serum (PI). The target product was enriched to a significantly higher level in the ChIP sample derived from 35S:AGL15 transgenic tissues (35S:AGL15, I) than in the ChIP sample derived from wild type tissues (Wild type, I). \textit{UBQ6} was used as reference.
the flower buds and inflorescence of wild type plants (Heck et al., 1995; Fernandez et al., 2000). In 35S:AGL15 transgenic plant, AGL15 accumulated to a higher level in floral tissue than in the wild type plants (Perry et al., 1999; Fernandez et al., 2000; Wang et al., 2002). As shown in Figure 2.4c, the ABF3 target fragment was significantly enriched in the immune ChIP population derived from the 35S:AGL15 transgenic flower buds and inflorescence tissue when using UBQ6 as a reference. In the input and pre-immune precipitated populations, both ABF3 and UBQ6 products were amplified to a similar level after enrichment PCR, that is, ABF3 target fragment was not enriched in those control populations. It also can be seen from Figure 2.4c, when enrichment PCR was performed on the immune ChIP population derived from wild type plant tissues, the target fragment was enriched, but to a much less degree. The result demonstrates that the degree of enrichment of the ABF3 target fragment depends on the level of AGL15 present in the tissue.

### 2.2.5 ABF3 expression is responsive to the changes AGL15 amounts

RNA slot blots were used to assess the responsiveness of the ABF3 expression to the AGL15 accumulation levels. Probes specific to the coding regions of ABF3 and the control gene TUB2 were designed and labeled with \(^{32}\)P. Total RNA samples isolated from siliques, young leaves, top two open flowers and flower buds of wild type and 35S:AGL15 transgenic plants were prepared and applied to the blotting membrane. As shown in Figure 2.5a, ABF3 was ubiquitously expressed in all tissues tested. Abundance of ABF3 transcripts was decreased in response to increased accumulation of AGL15 (35S:AGL15). The 5’ regulatory region of ABF3 (about 3.5 kb) including the DNA fragments isolated in the ChIP population was used to generate a reporter construct and transform Arabidopsis. The lines containing one insert in homozygous state were used to cross with 35S:AGL15 transgenic plants. The segregated F\(_1\) seedlings from the cross were separated according to the presence or absence of the 35S:AGL15 transgene (Fernandez et al., 2000) and measured for GUS activity using MUG assays. As shown in Figure 2.5b, GUS activity in the seedlings with 35S:AGL15 transgene
Figure 2.5 Response of *ABF3* expression to AGL15 levels

**a)** Slot blot analysis of the *ABF3* expression levels in various tissues from wild type and 35S:AGL15 transgenic plants. *ABF3* expression levels were consistently lower in tissues where AGL15 levels were higher

**b)** *ABF3* promoter-reporter activity in 10-day old wild type and 35S:AGL15 transgenic seedlings. GUS activity was lower in the 35S:AGL15 background than in the wild type background
was lower than that of the seedlings without the transgene. The result demonstrated that the regulatory region containing the fragment isolated using AGL15-specific antibody confers response to ectopic AGL15 in vivo as assessed by a reporter gene.

2.3 Discussion

We have isolated a collection of in vivo binding sites of AGL15 using ChIP and further in vitro selection; and presented evidence that ABF3 may be a direct target gene regulated by AGL15.

Chromatin immunoprecipitation (ChIP) with ECT was used to isolate in vivo binding sites of AGL15. In addition, in vitro immunoprecipitation was performed on the isolated ChIP population. The idea of further selection on the ChIP population using in vitro immunoprecipitation was based on the fact that AGL15 can bind in a sequence specific manner to DNA fragments in gel mobility shift assays (Perry et al., 1996; Wang et al., 2002, 2004b; Tang and Perry, 2003). The assumption was that the additional in vitro immunoprecipitation selection step may reduce non-specific background of the ChIP population. This in vivo plus in vitro selection of the binding sites for DNA binding proteins has been successfully used to isolate direct targets of a Drosophila homeoprotein Engrailed (Solano et al., 2003). A total 101 unique sequences were obtained, 47 of them were from the in vivo ChIP population while 54 were from the in vitro selected ChIP population. The high complexity of the population was initially surprising, but this also has been observed in other experiments. In the experiment with Engrailed protein binding sites, from a total of more than 500 clones sequenced, 203 unique sequences were identified and only 40 were found repeated two or three times. The high complexity and low frequency of repeats suggested low redundancy of the population (Solano et al., 2003). With further sequencing from other ChIP populations, one binding sites has been isolated twice.
This preliminary result suggests that there are many AGL15 binding sites in the whole Arabidopsis genome. This may not be uncommon for eukaryotic developmental regulators. For example, a human transcription factor, E2F4 was shown to bind to the promoters of about 9% of 1444 genes studied (Ren, 2002), a nuclear factor-kappa B (NFκB) protein p65 bound to 15.5% of the 917 distinct loci on chromosome 22 (Martone et al., 2003), and in human Daundi cells, a transcription factor, c-Myc was shown to be able to bind to 15% of 4839 genes tested. Moreover, using high-density oligonucleotide arrays representing all nonrepetitive sequences on human chromosomes 21 and 22, Cawley (Cawley et al.) mapped the in vivo binding sites for three DNA binding transcription factors, Sp1, cMyc and p53, and revealed that there are approximately 12,000 binding sites for Sp1, 25,000 for cMyc and 1600 for p53 when extrapolated to the full genome. In Drosophila, studies have found several homeodomain proteins bind in vivo to most of the genes expressed during embryo development (Walter and Biggin, 1996; Liang and Biggin, 1998; Carr and Biggin, 1999). As much as 87% of genes expressed during late stages of embryogenesis are regulated by these homeodomain proteins (Liang and Biggin, 1998).

Furthermore, AGL15 binding sites were found to locate not only in 5’ regions, which traditionally are deemed as the only important regulatory regions of corresponding genes, but were also located in other places of the genes such as 3’ regions, intergenic regions, introns or even exons. In fact, among the 101 clones containing DNA fragments isolated in ChIP, only about 38% of them localized to the 5’ regions. The result is consistent with observations in other experiments. For example, it was found that only 22% of the binding sites of Sp1, cMyc and P53 were located in the 5’ region of well-characterized genes (Cawley et al., 2004). Similarly, 28% of the NF-κB-bound fragments lie within 5 kb upstream of the 5’ end (ATG) of annotated genes (Martone et al., 2003) and 25% of a human homeoprotein BARX2 binding sites located within 50 kb upstream of a gene (Stevens et al., 2004). In fact, it is a common theme that the binding sites of transcription factors are distributed throughout the genome (reviewed in Wray et al., 2003). The binding of AGL15 to sites other than 5’ regions suggested that
those regions of the gene also have regulatory functions on the transcription of target genes.

In this study, an in vivo screen and further in vitro binding selection was performed to identify novel genomic locations bound by AGL15. The specificity of ChIP was monitored by using preimmune serum in the ChIP experiment along with AGL15-specific immune serum, as shown in Figure 2.2. For the in vitro selection experiment, as shown in Figure 2.3 there is little or no DNA selected when preimmune serum was used to precipitate AGL15 or when AGL15ΔM/No protein was used in the binding, demonstrating a need for the DNA binding domain and anti-AGL15 antibody to co-immunoprecipitate DNA fragments. In addition, the binding of AGL15 to some potential targets was further confirmed by using enrichment PCR in multiple independent ChIP populations and some populations were from various tissues (Figure 2.4c and Wang et al., 2002; Wang et al., 2004b). However, for a true direct target of AGL15, the binding of AGL15 to the regulatory region must have functional consequences on gene transcription. In fact, out of about 50 potential targets assessed by RT-PCR, only 5 of them showed reproducible response to AGL15 levels. The result showed that the binding of AGL15 does not necessary have a regulatory outcome. Actually this is not uncommon since many other transcription factors have been found to be able to bind DNA but no regulation occurs (Boyd et al., 1998; Boyd and Farnham, 1999; Ren et al., 2000, Skinger and Gross, 2001; Soutoglou and Talianidis, 2002, Martone et al., 2003). One important reason for binding without regulation is that the effect of a transcription factor is highly context dependent (Fry and Farnham, 1999). A transcription factor may require other cofactors' presence to influence the transcription of a downstream target gene through protein-protein interactions. The protein-protein interaction may influence a transcription factor in many ways including formation of functional regulatory complex on the target promoter, formation of appropriate chromatin architecture, activation/repression of transcription via protein modification (reviewed in Wray et al., 2003). Therefore, it is reasonable to assume that AGL15 may need to cooperate with other protein partners in order to regulate a target gene.
Other DTAs isolated from ChIP populations generally contain canonical CArG motifs either in the form of CC(A/T)$_6$GG or C(A/T)$_8$G (the AGL15 preferred binding site, Tang and Perry, 2003) in the nearby region of the isolated fragment, such as GA2ox6 and DTA2 and the sites were shown to be critical for the binding of AGL15 to the target genes (Wang et al., 2002; Wang et al., 2004b). However, no canonical CArG motif was found either within or nearby the DNA fragment isolated corresponding to the regulatory region of ABF3. Only one non-canonical CArG motif (CCAAAAATGC) was present in the fragment isolated and was able to bind AGL15 weakly in vitro (data not shown). As shown in Figure 2.4b, however, the regulatory region isolated from the ChIP population was highly enriched in the ChIP population, which suggested that AGL15 is able to bind to the region in vivo with high affinity. The binding of AGL15 to a putative regulatory site in ABF3 without a canonical CArG motif might involve other cofactors. This is quite common for MADS proteins. A human MADS protein, SRF was able to bind to canonical CArG motifs (Shore and Sharrocks, 1995) and as a complex with other cofactors, to non-canonical CArG motifs (Riechmann and Meyerowitz, 1997). It also has been shown that SRF, MCM1 and MEF2 were able to regulate different genes when interacting with other cofactors (Shore and Sharrocks, 1995; Messenguy and Dubois, 2003).

Slot blot and the ABF3:GUS reporter construct result showed that the regulatory region containing AGL15 binding site was able to confer regulation in response to AGL15. Taken together, the results suggested that ABF3 may be a direct target of AGL15 and expression was repressed by AGL15. Although the slot blot and reporter construct experiments were mostly based on the effect of ectopically expressed AGL15, it still might be true that ABF3 is a true direct target of endogenous AGL15. This is partially supported by the observation that the regulatory region of ABF3 was able to bind to AGL15 in vivo in wild type plants as indicated by the enrichment of the fragment in the ChIP population derived from wild type open flowers and flower buds (Figure 2.4d, wild type, lane I). In addition, in wild type leaf tissues where AGL15 is lowest in accumulation (Heck et al., 1995; Fernandez et al., 2000), ABF3 was expressed at the highest level.
Both results are consistent with the idea that AGL15 directly regulates *ABF3*. To further examine the effect of AGL15 on the expression of *ABF3*, a preliminary experiment was carried out to compare the transcripts abundance of *ABF3* in both wild type and *agl15* seedlings and the result indicated that there were no obvious expression changes. The reason for the lack of expression change when AGL15 is not present may be that other proteins with redundant function to AGL15 can regulate *ABF3* in absence of AGL15. The redundancy of functions is widely documented among transcriptional regulators, especially for MADS domain proteins (Riechmann and Meyerowitz, 1997). SEPALLATA serves as a classic example. SEPALLATA (SEP) proteins have four family members and single mutations of any one of the genes or double mutations of any two gene combinations have no obvious phenotype. Only when *SEP1*, *SEP2* and *SEP3* are mutated at the same time do plants show an obvious phenotype (Pelaz et al., 2000; Honma and Goto, 2001). Quadruple mutant *sep1 sep2 sep3 sep4* has an even more prominent phenotype, which shows that some aspects of the phenotype are masked by *SEP4* (Ditta et al., 2004). AGL15 and AGL18 may share partially redundant functions as they are most related to each other within the MADS family and are expressed in the same tissues (Alvarez-Buylla et al., 2000a; Becker and Theißen; 2003; Kofuji et al., 2003). On the other hand, AGL15 might regulate the expression of target genes in an modest fashion such that many target genes including *ABF3* do not show dramatic expression changes when AGL15 is absent in the cell, as shown in the microarray results in Chapter 3. However, since AGL15 is present in the cells of a wide range of tissues, it is possible *ABF3* is regulated by AGL15 in some other tissues that have not been tested yet.

ChIP has been used in fungi, animal and now plant systems to identify downstream target genes of transcription factors (reviewed in De Bell et al., 2000; Orlando, 2000). In recent years, the approach also has been applied to plant systems, such as to identify genes associated with acetylated histones H4 in tobacco (*Nicotiana tabacum*) (Chua et al., 2004). ChIP is useful to study the gene regulatory networks and unravel the molecular mechanisms that underlie
plant development. The so called ChIP-chip approach is the combination of ChIP and microarray technology, and has gained increasing interest in recent years to map the binding sites in an high throughput fashion (reviews see Buck and Lieb, 2004; Kirmizis and Farnham, 2004). Using this approach, one can greatly speed up the process of binding sites identification for a particular DNA binding protein.

2.4 Materials and methods

2.4.1 Arabidopsis embryonic culture tissue

Developing zygotic embryos were removed from transgenic plants that are constitutively expressing AGL15. The obtained embryos were then cultured on germination medium containing Murashige-Skoog (MS) salts, vitamins, 1% (w/v) sucrose, 0.05% MES(w/v), and 0.7% agar (w/v), pH 5.6–5.7 (hereafter GM media). Secondary embryonic tissues developed on the cultured zygotic embryos. Subculturing at regular intervals of approximately 3 weeks on GM allows maintenance of the tissue in an embryonic state.

2.4.2 Arabidopsis plant growth

The seeds were sterilized 3-4 times in 70% ethanol containing 0.1% Triton X-100 for 2 min, rinsed twice with 95% ethanol, then poured and dried on a sterile Whatman No.1 filter paper in a sterile hood. Sterilized seeds of Arabidopsis thaliana ecotype Ws (Wassilewskija) wild-type plants and transgenic plants were sown on GM germination media (with 50 μg/ml kanamycin for transgenic seed, hereafter GM/Kan media) and transplanted to ProMix BX (Premier Brands, Inc., Quebec, Canada) after 7–10 days. Plants were grown at 20/18 °C under a 16-h light/8-h dark regime. Flower buds and inflorescence tissues were collected on 5-week-old plants and used for ChIP experiment or RNA extraction. For staged siliques, flowers were tagged on the day that they opened and collected at the appropriate time afterwards.
2.4.3 Chromatin immunoprecipitation

Tissue fixation and nuclei isolation

The tissue was equilibrated in MC buffer (10 mM potassium phosphate, pH 7.0, 50 mM NaCl and 0.1 M sucrose) and then fixed by adding 1% formaldehyde and incubated on ice under vacuum for one hour. After incubation, the crosslinking was stopped by adding cold glycine to 0.125 M final concentration. The tissue was washed with MC buffer, dried and flash frozen in liquid nitrogen. The frozen tissue (8-12g) was ground to powder with a mortar and pestle in liquid nitrogen. The powder was mixed thoroughly with 9-15 ml of M1 buffer (10 mM potassium phosphate, pH 7.0, 0.1 M NaCl, 10 mM β-mercaptoethanol and 1 M hexylene glycol). The resulting tissue slurry was filtered through Miracloth and then centrifugated at 1000 x g for 10 min at 4°C. The pellet was further washed for 4-5 more times with 5-7 ml M2 buffer (10 mM potassium phosphate, pH 7.0, 0.1 M NaCl, 10 mM β-mercaptoethanol, 1 M hexylene glycol, 10 mM MgCl₂ and 0.5% Triton X-100), followed by washing once with M3 buffer (10 mM potassium phosphate, pH 7.0, 0.1 M NaCl and 10 mM β-mercaptoethanol) and then resuspended in 1 ml M3. 20 µl and 100 µl of the resuspension were saved for DAPI stain to check nuclei isolation and Western blot to check for protein, respectively. 4 ml of M3 was added to the remaining suspension and the nuclei were pelleted by centrifugation at 1000 x g for 10 min at 4°C.

Chromatin solubilization and preadsorption

The nuclear pellet was resuspended in 1 ml of sonication buffer and glass beads then sonicated 10 to 15 sec. x 4 pulses with a probe sonicator (Fisher, Model 300 sonic dismembrator). The glass beads (75 to 105 µm, Sigma G-3753) were pretreated before using; first, they were rinsed in 1N HCl, and then 0.1N HCl for 30 minutes or more; after rinsing, they were washed extensively with water, then sonication buffer (10 mM potassium phosphate, pH 7, 0.1 M NaCl, 0.5% sarkosyl, 10 mM EDTA) and finally resuspended in the buffer (buffer : beads, 2:1, v/v) with addition of PMSF (200 mM stock in isopropanol) to a final concentration
of 1 mM before use. After sonication, the mix was centrifuged 12,000 x g for 5 min at 4°C and the solubilized chromatin was removed. 20 µl of the supernatant was saved as “total” (input) sample to check DNA size and for enrichment PCR. 7.5 µl of preimmune serum was added to the remaining solubilized chromatin and then incubated on a rotator for one hour at 4°C. The preadsorption mix was centrifuged at top speed in a tabletop centrifuge for 2 min at 4°C. The supernatant was moved to a new tube with 40 µl of protein A-Sepharose (Sigma P-9424, before use washed and resuspended 1:1 in with TN buffer (10 mM Tris, 7.5, 150 mM NaCl and 0.05 % azide) ) added and incubated for 1 hour at 4°C with rotation. The supernatant was removed from the sample after centrifugation at top speed for 2 min, 4 °C.

**Immunoprecipitation**

The supernatant was divided into equal aliquots for immunoprecipitation with anti-AGL15 specific sera and for preimmune sera and/or no sera controls. An equal volume of immunoprecipitation buffer (50 mM Hapes, pH 7.5, 150 mM KCl, 5 mM MgCl₂, 10 µM ZnSO₄, 1% Triton X-100, 0.05% SDS) was added to each aliquot and incubated for one to two hours at 4°C with rotation. The sample was then centrifuged at top speed, 2 min at 4°C. The top 85% of the volume was moved to a new tube with 20 µl of 50% slurry of protein A-Sepharose and incubated for one hour at 4°C with rotation. The other 15% of the sample was saved as “binding” for Western blot to verify the quality of the protein. The beads were pelleted by spinning at top speed for 1-2 min at 4°C. The supernatant was removed and saved as “post-bind” to check the depletion of the protein from the supernatant. The beads were washed with immunoprecipitation buffer (1 ml each tube) for 10 minutes at room temperature with rotation and pelleted by centrifugation at top speed for 1 min. The wash and centrifugation was repeated for 3-5 times. For the last wash, the wash and beads were moved to a new tube and then the beads pelleted. The wash was removed and the beads were ready for elution.
Elution and DNA recovery

100 µl of cold glycine elution buffer (0.1 M glycine, 0.5 M NaCl, 0.05% Tween-20, pH 2.8) was added to the beads. The sample was mixed by vortexing and pelleted in the microfuge at top speed for 1 min at room temperature. The supernatant was removed and added to a tube with 50 µl of 1 M Tris, pH 9 to neutralize. The elution and neutralization were repeated twice more to give a 450 µl total volume of the eluted sample. The eluted sample was centrifuged at top speed for 2 min at room temperature. The top 300 µl was moved to a new tube. The remaining about 150 µl in the original tube was saved as “eluted” to verify recovery of the protein. To the 300 µl elution sample, 1 µl RNase A (1 mg/ml) was added and then incubated at 37°C for 15-30 min. After RNase A treatment, proteinase K was added to final concentration of 0.5 mg/ml and incubated 6 hours at 37 °C. Next day a second aliquot of proteinase K was added and the mix was incubated at 65°C for at least 6 hours to reverse the formaldehyde crosslinks. The sample was then cooled to room temperature and chilled on ice. DNA was extracted by phenol: chloroform extraction (phenol: chloroform: isoamyl alcohol at 25:24:1). The DNA was then recovered by ethanol precipitation and can be used for enrichment PCR test or modified with linkers for PCR amplification and cloning.

DNA cloning and sequencing

1.5 µl of 10x Sau3AI buffer and 0.5 µl of Sau3AI were added to the 13 µl of the co-precipitated DNA. The digestion mix was incubated at 37°C for 2 hours then 75°C for 10 min to inactivate Sau3AI. After digestion, 1 µl of 10 mM dGTP and 0.5 µl of Klenow (Promega, Madison WI) were added and incubated at room temperature for 10 min to do the single “G” filling-in reaction. Klenow was inactivated by incubation at 75 °C for 10 min. The DNA was then subjected to phenol:chloroform extraction as before and then resuspended in 10 µl of ligation mix (1 µl of “catch linker”, 1 µl of 10 x T4 ligase buffer, 0.4 µl of T4 ligase). Catch linkers are based on a design by Kinzler and Vogelstein (1989) and consisted of
two oligonucleotides (Oligo 1: 5'- ATCGAGATATTAGAATTCTACTCA -3' and Oligo 2: 5'- GAGTAGAATTCTAATATCTC -3'). The oligo 1 was phosphorylated using standard protocols and T4 polynucleotide kinase (Promega, Madison, WI). An equal amount of oligo 2 was annealed together with oligo 1 by placing at 95 °C and let cool to room temperature. The DNA and linker mix was incubated overnight in a water bath at 16 °C to 4°C. After ligation with linker, the specificity of the DNA was checked by PCR using oligo 2. Usually the PCR was carried out for 27 cycles with annealing temperature set to 60°C. An aliquot of the PCR product was digested with EcoRI and then run on an agarose gel to check the specificity. The AGL15-specific immune sera immunoprecipitated DNA products were run on agarose gel and purified by using Geneclean Kit (Bio101 Inc., now part of QBiogene Inc., Carlsbad, CA) following manufacturer’s instruction.

2.4.4 ChIP population sequencing

The purified DNA products were ligated into the pBlueScript II SK+ (Stratagene, La Jolla, CA) vector. The cloned DNA in pBluescript II SK+ was sequenced by using either T7 Sequenase v2.0 (Amersham Life Science, Cleveland, OH) or ABI PRISM® BigDye™ Terminator (Applied Biosystems, Foster City, CA) following protocols provided by the manufacturers. The sequences were used to search against The National Center for Biotechnology Information (NCBI) using BLAST algorithm to obtain information about the cloned DNA fragments.

2.4.5 In vitro selection of the ChIP population

Approximately 500 ng of the PCR products amplified using oligo 2 from ChIP population were used for in vitro binding using full-length AGL15 and precipitated with AGL15-specific immune serum. As controls, full-length AGL15 was precipitated using preimmune serum, no protein or truncated form of AGL15 lacking MADS domain was precipitated using immune serum. AGL15 and the truncated form AGL15 were prepared, renatured and purified as described in Perry et al. (1996). The DNA and protein (or binding buffer in the case of no protein control) were mixed with 1 µg poly (dI.dC) (Amersham Biosciences Inc.,
Piscataway, NJ) in 1 x binding buffer (60mM KCl, 12mM Hepes pH7.5, 4mM Tris, pH7.5, 1mM EDTA, 1mM DTT and 10% glycerol) in a 20 µl volume. After incubation at room temperature for 30 min, 200 µl of immunoprecipitation buffer and 2.5 µl of anti-AGL15 specific immune serum was added to the reaction mix and then incubated in room temperature for 1 h with rotation. The sample was then centrifuged at top speed for 3 min. The upper 200 µl supernatant was moved to a new tube with 20 µl 50% protein A-Sepharose and then incubated for 1 h at room temperature with rotation. The sample was pelleted by centrifugation at top speed for 1 min. One ml cold (4 °C) immunoprecipitation buffer was added to the pellet and vortexed, and then centrifuged at top speed for 1 min. The pellet was washed 2 more times and moved to a new tube at the last wash. 150 µl of disassociation buffer (0.5M Tris.HCl, pH 9, 0.02M EDTA, 0.01M NaCl, 0.2% SDS) was added to resuspend the pellet and then incubated at 68 °C for 10 min. The beads were pelleted by centrifugation at top speed for 1 min. The beads and 25 µl supernatant were saved to check efficiency of immunoprecipitation. The remaining 125 µl supernatant was extracted by phenol: chloroform and precipitated by ethanol as before. The DNA was resuspended in 20 µl DD H2O, amplified by PCR and subjected to next round of in vitro selection. The process was reiterated for total of 3 rounds. After final rounds of selection, the resulted DNA was cloned and sequenced as before.

2.4.6 Protein analysis

Samples reserved during ChIP were used for protein analysis. Proteins were separated on 12.5 % (w/v) polyacrylamide denaturing gels and then blotted onto Immobilon™ PVDF Transfer Membranes (Millipore, Billerica, MA). Blots were incubated with 1: 1000 (v/v) anti-AGL15 immune serum and primary antibody and visualized using the Lumi-Glo system with the 1: 5000 (v/v) diluted secondary antibody (Kirkegaard and Perry Laboratories, Gaithersburg, MD). Blots were exposed to X-ray film (Kodak XAR5) for 1–3 min.
2.4.7 Enrichment test for \textit{in vivo} binding of AGL15

The enrichment PCR was used to test whether a target DNA fragment is indeed bound by AGL15 \textit{in vivo}. The DNA fragment was tested for enrichment in the ChIP populations as compared to a non-bound control. Oligonucleotide primers were designed to specifically amplify the select DNA fragments. As control, oligonucleotide primers were also designed to amplify portions of the coding regions of \textit{TUB2}, \textit{EF-1\alpha-A1}, or \textit{UBQ6}. In the enrichment PCR reaction, dilution of total (input) DNA or ChIP populations from immune or control immunoprecipitations were used as templates, primers for the DNA fragment and control were added together. Typically, 30-35 cycles of PCR were performed using KlenTaq1 (Ab Peptides, St. Louis, MO). PCR products were analyzed by agarose or polyacrylamide gel electrophoresis and gel images were captured using a ChemiImager (Alpha Innotech Corporation, San Leandro, CA). The images shown are inverted to better visualize the ratio of select DNA fragment to control PCR products. In this experiment, the primers used for enrichment PCR were:

\begin{align*}
\text{\textit{ABF3}}: & \quad \text{Forward: 5'}-\text{TAA CGG ATC AAC GAA TCT CGT -3'} \\
& \quad \text{Reverse: 5'}-\text{GAT ACC TGA AAG GGG TCA GA -3'} \\
\text{\textit{UBQ6}}: & \quad \text{Forward: 5'}-\text{GGT GCT AAG AAG AGG AAG AAT -3'} \\
& \quad \text{Reverse: 5'}-\text{CTC CTT CTT TCT GGT AAA CGT -3'}
\end{align*}

2.4.8 RNA slot blot

RNA samples from various tissues were collected from 5-week-old plants. Open flowers were the top two fully open flowers, collected and flash frozen in liquid nitrogen for RNA extraction. RNA samples used in the blot and probe for \textit{TUB2} were prepared by Dr. Huai Wang. Briefly, total RNA was extracted from staged tissues of wild type and \textit{AGL15} constitutive expressing plants using the hot-borate method for siliques (Wilkins and Smart, 1996) and TRlzol Reagent (Invitrogen, Carlsbad, CA) for all other tissues. mRNA was obtained using PolyATract mRNA isolation system (Promega) and applied to a Zeta-Probe GT
blotting membrane (Bio-Rad, Hercules, CA) using a vacuum manifold (Bio-Rad). DNA probes specific to \textit{ABF3} (At4g34000) or \textit{β-2 TUBULIN} (\textit{TUB2}, At5g62690) were $^{32}$P-labeled and used for blot. Slot blot was performed as described in Wang et al.(2004b). Primers used to generate gene specific probes were:

\textit{ABF3:}  
Forward: 5'- TTT GTT GCA AAC CAA CCT CA -3'  
Reverse: 5'- ACT GCT GCA ACC GTT ACT CC -3'

\textit{TUB2:}  
Forward: 5'- CTC AAG AGG TTC TCA GCA GTA -3'  
Reverse: 5'- TCA CCT TCT TCA TCC GCA GTT -3'.

\textbf{2.4.9 Generation of transgenic plants and reporter activity quantification}

The 5' region of \textit{ABF3} was amplified by PCR using Ex-Taq polymerase (Panvera, Madison, WI) from Arabidopsis ecotype Ws genomic DNA using primers. The fragment amplified was first cloned into pGEM-T Easy vector (Promega, Madison, WI), then excised and cloned into pBI121 and a reporter construct was thus generated. After confirming the sequence, the construct was used with \textit{Agrobacteria tumefaciens} GV3101 to transform Arabidopsis ecotype Ws using floral dip method (Clough and Bent, 1998). Individual reporter lines carrying one insert in homozygous state were used to do crosses with \textit{35S:AGL15} transgenic plants. The F1 seedlings with or without \textit{35S:AGL15} transgene were easily separated at the seedling stage (Fernandez et al., 2000). GUS activities of 18-25 individuals of 10 day-old seedlings were quantified using MUG assay (Gallagher, 1992). The primers used to amplify the regulatory region of \textit{ABF3} were:

Forward: 5' - CCC AAG CTT TTT TCC AAC AGT CTT G -3'  
Reverse: 5' - CGC GGA TCC TAC TCA AGC TTTCGTA -3'

The nucleotides indicated in bold denote restriction enzyme sites engineered in the primer to facilitate cloning. Forward primer contains a \textit{Hind III} site and reverse contains a \textit{BamH I} site.
CHAPTER 3

MICROARRAY ANALYSIS OF ARABIDOPSIS WILD TYPE AND
AGL15 SEEDS AND IDENTIFICATION OF AN HMGB BOX GENE
3.1 Introduction

As described in Chapter 2, novel in vivo binding sites of AGL15 can be identified in ChIP experiments. An ongoing effort in the Perry lab is to systematically isolate a more complete collection of AGL15 binding sites using ChIP-chip. ChIP-chip is microarray chip followed by ChIP experiment. In recent years, the ChIP-chip approach has been successfully used to identify the DNA fragments bound by transcription factors in high-throughput manner (Ren et al., 2000; Iyer et al., 2001; Simon et al., 2001). To map the AGL15 binding sites on the genome is a step toward identification of its direct target genes. However, AGL15 direct target gene identification cannot be based solely on binding data. The reason is that binding does not necessary result in regulation. In addition, the binding of AGL15 to a target gene does not tell us whether the expression of the target gene is activated or repressed by AGL15. Furthermore, for a binding site located between two genes, it is unclear which one gene or whether both of them is/are regulated. To address this problem, traditional approaches such as RT-PCR and Northern blot can be used. However, these are low throughput approaches and would therefore be time-consuming and laborious to perform for each AGL15 binding site. For example, our collection of a partial population already consists of over 100 putative sites. An alternative approach which is widely used and much more efficient to assess regulation by AGL15 on target genes is to use microarray analysis.

DNA microarray is a relatively new technology that was developed to analyze the expression of a large number of genes at the same time (Schena et al., 1995). Like traditional Northern blotting analysis, microarray utilizes the principle of nucleic acid (DNA or RNA) hybridization. Microarray technology can be used in many areas (Kapranov et al., 2003), including to monitor the steady-state mRNA abundance. In a sense it is like performing tens of thousands of Northern-blotting in parallel. In classical microarray, mRNAs from two samples (cell lines, tissues or other resources) are labeled and hybridized to arrays on which cDNAs or oligonucleotides of known sequences are immobilized. The hybridization signals
are analyzed and used to determine the gene expression changes between the two samples. The microarray data obtained then can be used to investigate the effect of the particular biotic or abiotic treatment of interest. In our experiments, the question to be addressed using DNA microarray is which genes are regulated by AGL15?

In this experiment, microarray technology was used to catalogue the lists of genes whose expression was changed in response to the perturbation of AGL15 protein amount in the Arabidopsis developing seeds. From the experiments, we obtained a list contains more than 640 genes that are differentially expressed between 5-8 day seeds of wild type Ws and agl15 plants. From the list, a gene encoding a HMGB domain protein was identified and further analysis was performed in an effort to characterize its function.

3.2 Results

3.2.1 Microarray analysis of Arabidopsis in wild type and agl15 seeds

Arabidopsis seeds are a difficult tissue from which to obtain good quality RNA. Several protocols were tested for RNA extraction from Arabidopsis seeds, including TRIzol Reagent Kit (Invitrogen), phenol:chloroform:isoamyl alcohol method (http://arabi4.agr.hokudai.ac.jp/ArabiE/protocols/general/general.html), a method designed by Schultz for recalcitrant plant tissue (1994) and a method developed by Gehrig for tissues rich in polyphenols and polysaccharides (Gehrig et al.). The RNA quality was assessed by spectrophotometry, RT-PCR and electrophoresis on denatured RNA gel containing MOPS and formaldehyde. From the above mentioned four protocols tested, only the last method (Gehrig et al., 2000) consistently produced high quality RNAs. The method was then used to prepare three independent RNA preparations for each genotype used in microarray experiment.
There are two formats of microarray, cDNA array and oligonucleotide array. For our study, the high coverage of genes, standardized procedure and trained technician at the UK Microarray Core Facility (University of Kentucky, Lexington) made the Arabidopsis ATH1 GeneChip® Genome Array (Affymetrix Inc., Santa Clara, CA; http://www.affymetrix.com) the platform of choice. The ATH1 array is based on high-density oligonucleotides and contains 25mer probe sets representing approximately 23,000 gene sequences on a single array.

From all the probe sets on the array, only those identified as Present by Absolute Call analysis were retained for further statistical analysis. Among all the probe sets on the chip, 16,753 genes were determined as Present, i.e., gave a mean hybridization signal above background based on the Affymetrix Microarray Suite v5.0 algorithm, on at least one of the six GeneChips used in our study. Assuming equal variance about the means of the hybridization signals on chips hybridized to the Ws and agl15 total RNA samples, a two-sample t-test was carried out on the hybridization signals for the 16,753 genes to determine the P-values. Using conventional statistical significance of P-value < 0.05, the expression levels of 375 genes were found to be significantly down-regulated in agl15 compared with Ws and 270 genes were up-regulated. To calculate the fold differences in expression between Ws and agl15, the raw expression levels of the gene on each chip were used. Using cutoff fold change of 1.4 and P-value < 0.05, the expression level of 66 genes were found to be decreased while 37 genes were increased in agl15. The two lists of these genes can be found in Table 3.1a and b. The functional categories of the genes were divided into different groups based on the annotation by the Munich Information Center for Protein Sequences (MIPS) (http://mips.gsf.de/proj/funcatDB/search_main_frame.html) Arabidopsis database. As shown in Figure 3.1a, 25 out of 37 of genes that showed increased expression in agl15 are of unknown function. Similarly, the functions of most of the genes that showed decreased expression are also unknown (49 out of 65, Figure 3.1b). The second major group of genes is involved in metabolism (13 genes total). Notably, only two genes showing differential expression between agl15 and Ws are known transcription factors. The putative MADS domain
<table>
<thead>
<tr>
<th>Affy Probeset</th>
<th>Descriptions</th>
<th>AGI number</th>
<th>P_value</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>261946_at</td>
<td>unknown protein</td>
<td>At1g64560</td>
<td>0.0138</td>
<td>4.14</td>
</tr>
<tr>
<td>254075_at</td>
<td>DREB1C</td>
<td>At4g25470</td>
<td>0.02</td>
<td>2.99</td>
</tr>
<tr>
<td>256977_at</td>
<td>hypothetical protein</td>
<td>At3g21040</td>
<td>0.0264</td>
<td>2.44</td>
</tr>
<tr>
<td>253936_at</td>
<td>STIG1 like protein</td>
<td>At4g26880</td>
<td>0.0361</td>
<td>2.40</td>
</tr>
<tr>
<td>249895_at</td>
<td>male sterility 2-like protein</td>
<td>At5g22500</td>
<td>0.0003</td>
<td>1.88</td>
</tr>
<tr>
<td>244995_at</td>
<td>ATPase a subunit</td>
<td>atpl</td>
<td>0.0401</td>
<td>1.84</td>
</tr>
<tr>
<td>265237_s_at</td>
<td>hypothetical protein</td>
<td>At2g07706</td>
<td>0.046</td>
<td>1.82</td>
</tr>
<tr>
<td>260319_at</td>
<td>hypothetical protein</td>
<td>At1g63950</td>
<td>0.0246</td>
<td>1.81</td>
</tr>
<tr>
<td>254200_at</td>
<td>putative protein</td>
<td>At4g24110</td>
<td>0.0293</td>
<td>1.75</td>
</tr>
<tr>
<td>265656_at</td>
<td>hypothetical protein</td>
<td>At2g13820</td>
<td>0.0153</td>
<td>1.74</td>
</tr>
<tr>
<td>245026_at</td>
<td>ATPase III subunit</td>
<td>At1g55390</td>
<td>0.0474</td>
<td>1.71</td>
</tr>
<tr>
<td>256162_at</td>
<td>hypothetical protein</td>
<td>At5g44610</td>
<td>0.0209</td>
<td>1.70</td>
</tr>
<tr>
<td>249009_at</td>
<td>unknown protein</td>
<td>At4g09350</td>
<td>0.049</td>
<td>1.65</td>
</tr>
<tr>
<td>244938_at</td>
<td>putative protein</td>
<td>At4g37540</td>
<td>0.0144</td>
<td>1.63</td>
</tr>
<tr>
<td>255088_at</td>
<td>putative protein heat shock protein dnaJ</td>
<td>At4g09350</td>
<td>0.049</td>
<td>1.65</td>
</tr>
<tr>
<td>254797_at</td>
<td>putative protein</td>
<td>At4g21030</td>
<td>0.0008</td>
<td>1.64</td>
</tr>
<tr>
<td>253043_at</td>
<td>putative protein predicted protein</td>
<td>At4g37540</td>
<td>0.0144</td>
<td>1.63</td>
</tr>
<tr>
<td>248916_at</td>
<td>putative protein receptor protein kinase-like protein</td>
<td>At5g45840</td>
<td>0.02</td>
<td>1.60</td>
</tr>
<tr>
<td>265420_s_at</td>
<td>hypothetical protein</td>
<td>At2g21030</td>
<td>0.0004</td>
<td>1.59</td>
</tr>
<tr>
<td>259810_at</td>
<td>Na+/H+ antiporter, putative</td>
<td>At1g49810</td>
<td>0.0099</td>
<td>1.58</td>
</tr>
<tr>
<td>250572_at</td>
<td>putative protein</td>
<td>At5g08210</td>
<td>0.0427</td>
<td>1.55</td>
</tr>
<tr>
<td>251356_at</td>
<td>putative protein</td>
<td>At3g61060</td>
<td>0.0416</td>
<td>1.54</td>
</tr>
<tr>
<td>255466_at</td>
<td>putative protein</td>
<td>At4g03000</td>
<td>0.043</td>
<td>1.54</td>
</tr>
<tr>
<td>244940_at</td>
<td>ribosomal protein S12</td>
<td>rps12.2</td>
<td>0.0412</td>
<td>1.52</td>
</tr>
<tr>
<td>257319_at</td>
<td>hypothetical protein</td>
<td>orf105a</td>
<td>0.0008</td>
<td>1.52</td>
</tr>
<tr>
<td>262128_at</td>
<td>late embryogenesis-abundant protein</td>
<td>At1g52690</td>
<td>0.0286</td>
<td>1.51</td>
</tr>
<tr>
<td>245049_at</td>
<td>ribosomal protein S16</td>
<td>rps16</td>
<td>0.0078</td>
<td>1.51</td>
</tr>
<tr>
<td>263346_at</td>
<td>En/Spm-like transposon protein</td>
<td>At2g05650</td>
<td>0.0221</td>
<td>1.50</td>
</tr>
<tr>
<td>244972_at</td>
<td>PSII 47KDa protein</td>
<td>psbB</td>
<td>0.0309</td>
<td>1.49</td>
</tr>
<tr>
<td>266201_at</td>
<td>similar to MtN3 protein</td>
<td>At2g39060</td>
<td>0.0008</td>
<td>1.48</td>
</tr>
<tr>
<td>256081_at</td>
<td>hypothetical protein</td>
<td>At1g20700</td>
<td>0.0258</td>
<td>1.46</td>
</tr>
<tr>
<td>267345_at</td>
<td>unknown protein</td>
<td>At2g44240</td>
<td>0.0236</td>
<td>1.45</td>
</tr>
<tr>
<td>266014_s_at</td>
<td>hypothetical protein</td>
<td>At2g07722</td>
<td>0.0082</td>
<td>1.45</td>
</tr>
<tr>
<td>255365_at</td>
<td>putative phosphofructokinase beta subunit</td>
<td>At4g04040</td>
<td>0.0426</td>
<td>1.45</td>
</tr>
<tr>
<td>245593_at</td>
<td>IAA14</td>
<td>At4g14550</td>
<td>0.0122</td>
<td>1.44</td>
</tr>
<tr>
<td>259764_at</td>
<td>NPR1</td>
<td>At1g64280</td>
<td>0.0198</td>
<td>1.44</td>
</tr>
<tr>
<td>F247228_at</td>
<td>trehalose-6-phosphate phosphatase</td>
<td>At5g65140</td>
<td>0.0127</td>
<td>1.41</td>
</tr>
<tr>
<td>Affy Probeset</td>
<td>Descriptions</td>
<td>AGI number</td>
<td>P_value</td>
<td>Fold change</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>248571_at</td>
<td>putative protein</td>
<td>At5g49790 0.0047</td>
<td>-5.18</td>
<td></td>
</tr>
<tr>
<td>258639_at</td>
<td>polygalacturonase (PGA3)</td>
<td>At3g07820 0.0122</td>
<td>-3.15</td>
<td></td>
</tr>
<tr>
<td>254900_at</td>
<td>hypothetical protein. RALFL28</td>
<td>At4g11510 0.0056</td>
<td>-2.97</td>
<td></td>
</tr>
<tr>
<td><strong>265007_s_at</strong></td>
<td><strong>RALF 9</strong></td>
<td><strong>At1g61566 0.0146</strong></td>
<td><strong>-2.93</strong></td>
<td></td>
</tr>
<tr>
<td>250576_at</td>
<td>cytochrome P450-like protein</td>
<td>At5g08250 0.0005</td>
<td>-2.46</td>
<td></td>
</tr>
<tr>
<td>257625_at</td>
<td>cytochrome P450</td>
<td>At3g26230 0.0275</td>
<td>-2.23</td>
<td></td>
</tr>
<tr>
<td>262122_at</td>
<td>polygalacturonase, putative</td>
<td>At1g02790 0.0178</td>
<td>-2.18</td>
<td></td>
</tr>
<tr>
<td>254395_at</td>
<td>subtilisin-like protease</td>
<td>At4g21640 0.0316</td>
<td>-2.12</td>
<td></td>
</tr>
<tr>
<td>251633_at</td>
<td>Putative metalloendopeptidase</td>
<td>At3g57460 0.0218</td>
<td>-2.11</td>
<td></td>
</tr>
<tr>
<td>248824_at</td>
<td>putative invertase inhibitor</td>
<td>At5g46940 0.0266</td>
<td>-2.08</td>
<td></td>
</tr>
<tr>
<td>257532_at</td>
<td>hypothetical protein</td>
<td>At3g04700 0.0233</td>
<td>-2.08</td>
<td></td>
</tr>
<tr>
<td><strong>266753_at</strong></td>
<td><strong>auxin-induced protein (IAA20)</strong></td>
<td><strong>At2g46990 0.0380</strong></td>
<td><strong>-2.06</strong></td>
<td></td>
</tr>
<tr>
<td>264813_at</td>
<td>putative Ca2+-dependent ser/thr protein kinase</td>
<td>At2g17890 0.0019</td>
<td>-1.81</td>
<td></td>
</tr>
<tr>
<td>257469_at</td>
<td>hypothetical protein</td>
<td>At1g49290 0.0126</td>
<td>-1.80</td>
<td></td>
</tr>
<tr>
<td>249768_at</td>
<td>receptor-like protein kinase</td>
<td>At5g24100 0.0445</td>
<td>-1.80</td>
<td></td>
</tr>
<tr>
<td>253940_at</td>
<td>putative protein</td>
<td>At4g26950 0.0455</td>
<td>-1.78</td>
<td></td>
</tr>
<tr>
<td>248943_s_at</td>
<td>putative protein</td>
<td>At5g45490 0.0162</td>
<td>-1.77</td>
<td></td>
</tr>
<tr>
<td>254511_at</td>
<td>hypothetical protein</td>
<td>At4g20220 0.0417</td>
<td>-1.75</td>
<td></td>
</tr>
<tr>
<td>258768_at</td>
<td>hypothetical protein</td>
<td>At3g10880 0.0009</td>
<td>-1.75</td>
<td></td>
</tr>
<tr>
<td>256966_at</td>
<td>pollen specific protein, putative</td>
<td>At3g13400 0.0051</td>
<td>-1.73</td>
<td></td>
</tr>
<tr>
<td>248079_at</td>
<td>unknown protein</td>
<td>At5g55790 0.0260</td>
<td>-1.71</td>
<td></td>
</tr>
<tr>
<td>257454_at</td>
<td>hypothetical Ubiquitin carboxyl-terminal hydrolase</td>
<td>At1g65170 0.0430</td>
<td>-1.68</td>
<td></td>
</tr>
<tr>
<td>252618_at</td>
<td>lipoxigenase AtLOX2</td>
<td>At3g45140 0.0469</td>
<td>-1.66</td>
<td></td>
</tr>
<tr>
<td>259269_at</td>
<td>putative pectate lyase</td>
<td>At3g01270 0.0344</td>
<td>-1.66</td>
<td></td>
</tr>
<tr>
<td>246144_at</td>
<td>dynadep light chain - like protein</td>
<td>At5g20110 0.0250</td>
<td>-1.66</td>
<td></td>
</tr>
<tr>
<td>258961_at</td>
<td>hypothetical myb-related transcriptional activators</td>
<td>At3g10580 0.0146</td>
<td>-1.66</td>
<td></td>
</tr>
<tr>
<td>262471_at</td>
<td>hypothetical protein</td>
<td>At1g50150 0.0314</td>
<td>-1.64</td>
<td></td>
</tr>
<tr>
<td>266154_at</td>
<td>putative cytochrome p450</td>
<td>At2g12190 0.0095</td>
<td>-1.63</td>
<td></td>
</tr>
<tr>
<td><strong>260688_at</strong></td>
<td><strong>hypothetical dehydrogenase-like protein</strong></td>
<td><strong>At1g17665 0.0205</strong></td>
<td><strong>-1.63</strong></td>
<td></td>
</tr>
<tr>
<td>262760_at</td>
<td>hypothetical pectinesterase</td>
<td>At1g10770 0.0034</td>
<td>-1.62</td>
<td></td>
</tr>
<tr>
<td>249118_at</td>
<td>putative protein</td>
<td>At5g43870 0.0279</td>
<td>-1.61</td>
<td></td>
</tr>
<tr>
<td>261970_at</td>
<td>glutamate decarboxylase (gad), putative</td>
<td>At1g65960 0.0449</td>
<td>-1.61</td>
<td></td>
</tr>
<tr>
<td>252914_at</td>
<td>putative dehydrin</td>
<td>At4g39130 0.0149</td>
<td>-1.58</td>
<td></td>
</tr>
<tr>
<td>Affy Probeset</td>
<td>Descriptions</td>
<td>AGI number</td>
<td>P_value</td>
<td>Fold change</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>257940_at</td>
<td>putative UDP-glucose glucosyltransferase</td>
<td>At3g21790</td>
<td>0.0059</td>
<td>-1.55</td>
</tr>
<tr>
<td>246727_at</td>
<td>major latex protein homolog - like</td>
<td>At5g28010</td>
<td>0.0365</td>
<td>-1.55</td>
</tr>
<tr>
<td>249491_at</td>
<td>germin-like protein GLP2a</td>
<td>At5g39130</td>
<td>0.0014</td>
<td>-1.54</td>
</tr>
<tr>
<td>260450_at</td>
<td>hypothetical protein</td>
<td>At1g72380</td>
<td>0.0399</td>
<td>-1.54</td>
</tr>
<tr>
<td>253678_at</td>
<td>cytidine deaminase 7</td>
<td>At4g29600</td>
<td>0.0055</td>
<td>-1.54</td>
</tr>
<tr>
<td>263052_at</td>
<td>unknown protein</td>
<td>At2g13430</td>
<td>0.0163</td>
<td>-1.53</td>
</tr>
<tr>
<td>266267_at</td>
<td>putative glutathione S-transferase</td>
<td>At2g29460</td>
<td>0.0223</td>
<td>-1.53</td>
</tr>
<tr>
<td>255877_at</td>
<td>putative PTR2 family peptide transporter</td>
<td>At2g40460</td>
<td>0.0032</td>
<td>-1.53</td>
</tr>
<tr>
<td>246238_at</td>
<td>sugar transporter like protein</td>
<td>At4g36670</td>
<td>0.0490</td>
<td>-1.52</td>
</tr>
<tr>
<td>262241_at</td>
<td>hypothetical protein</td>
<td>At1g48390</td>
<td>0.0489</td>
<td>-1.51</td>
</tr>
<tr>
<td>260551_at</td>
<td>putative trypsin inhibitor</td>
<td>At2g43510</td>
<td>0.0333</td>
<td>-1.50</td>
</tr>
<tr>
<td>262040_at</td>
<td>receptor protein kinase, putative</td>
<td>At1g80080</td>
<td>0.0464</td>
<td>-1.50</td>
</tr>
<tr>
<td>256079_at</td>
<td>hypothetical protein</td>
<td>At1g20680</td>
<td>0.0029</td>
<td>-1.49</td>
</tr>
<tr>
<td>248832_at</td>
<td>putative protein</td>
<td>At5g47170</td>
<td>0.0027</td>
<td>-1.49</td>
</tr>
<tr>
<td>252795_at</td>
<td>putative protein</td>
<td>At3g42270</td>
<td>0.0300</td>
<td>-1.49</td>
</tr>
<tr>
<td>250918_at</td>
<td>putative protein proline-rich protein APG</td>
<td>At5g03610</td>
<td>0.0141</td>
<td>-1.49</td>
</tr>
<tr>
<td>252006_at</td>
<td>purple acid phosphatase-like protein</td>
<td>At3g52820</td>
<td>0.0393</td>
<td>-1.48</td>
</tr>
<tr>
<td>258984_at</td>
<td>putative DnaJ protein</td>
<td>At3g08970</td>
<td>0.0077</td>
<td>-1.47</td>
</tr>
<tr>
<td>257568_s_at</td>
<td>hypothetical protein</td>
<td>At3g23970</td>
<td>0.0268</td>
<td>-1.47</td>
</tr>
<tr>
<td>251696_at</td>
<td>putative protein</td>
<td>At3g56590</td>
<td>0.0428</td>
<td>-1.47</td>
</tr>
<tr>
<td>256214_x_at</td>
<td>hypothetical protein</td>
<td>At1g51000</td>
<td>0.0053</td>
<td>-1.47</td>
</tr>
<tr>
<td>263745_at</td>
<td>hypothetical protein</td>
<td>At2g21450</td>
<td>0.0418</td>
<td>-1.46</td>
</tr>
<tr>
<td>266173_at</td>
<td>unknown protein</td>
<td>At2g02420</td>
<td>0.0488</td>
<td>-1.46</td>
</tr>
<tr>
<td>259926_at</td>
<td>DNA-3-methyladenine glycosylase I</td>
<td>At1g75090</td>
<td>0.0439</td>
<td>-1.44</td>
</tr>
<tr>
<td>259221_s_at</td>
<td>putative phospholipase</td>
<td>At3g03540</td>
<td>0.0055</td>
<td>-1.44</td>
</tr>
<tr>
<td>245675_at</td>
<td>hypothetical protein</td>
<td>At1g56675</td>
<td>0.0495</td>
<td>-1.43</td>
</tr>
<tr>
<td>258919_at</td>
<td>unknown protein</td>
<td>At3g10525</td>
<td>0.0189</td>
<td>-1.43</td>
</tr>
<tr>
<td>257550_at</td>
<td>hypothetical protein</td>
<td>At3g18460</td>
<td>0.0381</td>
<td>-1.42</td>
</tr>
<tr>
<td>246122_at</td>
<td>putative protein</td>
<td>At5g20380</td>
<td>0.0388</td>
<td>-1.42</td>
</tr>
<tr>
<td>247109_at</td>
<td>putative protein</td>
<td>At5g65870</td>
<td>0.0284</td>
<td>-1.42</td>
</tr>
<tr>
<td>266259_at</td>
<td>unknown protein</td>
<td>At2g27830</td>
<td>0.0324</td>
<td>-1.42</td>
</tr>
<tr>
<td>253178_at</td>
<td>putative protein</td>
<td>At4g35170</td>
<td>0.0431</td>
<td>-1.41</td>
</tr>
</tbody>
</table>
Figure 3.1 Distribution in functional categories of genes differentially expressed between agl15 and Ws

The change of expression is at least 1.4-fold change in expression between agl15 vs. Ws (P-value <0.05)
protein (including AGL15) binding sites were analyzed in the intergenic region of the genes that listed in Tables 3.1a and b. As shown in Figure 3.2. Putative MADS domain protein binding sites, called CArG motifs were over-represented in the intergenic regions of those genes that are differentially expressed between wild type and agl15, when compared to the frequency of the CArG motifs in the intergenic region of all the genes in the Arabidopsis genome. In the whole genome, 42% of the genes do not contain any predicted MADS domain protein binding sites in their intergenic regions, about 58% of the remaining genes contains one or more sites. For the genes listed in Tables 3.1a and 3.1b, only approximately 26% do not contain any CArG motif in the intergenic region, while about 74% of the genes contain one or more MADS domain protein binding sites in their intergenic region. The enrichment of the CArG motifs in the regulatory regions of the MADS domain protein target genes was also observed in other reports, such as for AP3/PI (Zik and Irish, 2003) and AG (Gomez-Mena et al., 2005). Although the presence or absence of the CArG motif does not necessarily indicates direct regulation by AGL15, the results suggests that direct AGL15 targets are recovered by microarray experiments.

3.2.2 Confirmation of the microarray results

To confirm the microarray results, 5 genes were selected to verify differential expression by RT-PCR, including At1g20700, At4g14550, At4g25470, At1g61566 and At2g46990 (bold font in Tables 3.1a and b). The transcript amounts of the preceding three genes were increased more than 1.4 fold in agl15 and the transcripts levels of the latter two genes were decreased more than 1.4 fold in agl15 when compared with Ws in 5 to 8-day-old seeds. As shown in Figure 3.3, the changes of the expression of 4 out of 5 genes tested were consistent with microarray results, i.e., those genes determined to be up-regulated in agl15 in microarray also showed increased expression in agl15 as assessed by RT-PCR. The RT-PCR results were also confirmed the decreased expression of At4g25470 and At2g46990 in agl15, the two genes determined to be down-regulated in agl15 by microarray analysis. Only one gene, At4g14550
Figure 3.2 Potential AGL15 binding sites in the target genes
Figure 3.3 RT-PCR analysis of four genes identified in microarray experiments

At1g20700 and At4g25470 were upregulated in *agl15* as determined by microarray analysis. At1g61566 and At2g46990 were downregulated in *agl15* in microarray. *EF1α-A1* was used as a control.
was not able to show consistent increased expression in \textit{agl15}.

Most of the genes that were differentially expressed between \textit{agl15} and Ws were not listed in Table 3.1 because the changes of their expression levels were less than 1.4 fold. It is possible that some of those genes could be regulated by AGL15 as well. To test this hypothesis, one gene, At5g23405 encoding a member of HMGB protein family, that is not included in Table 3.1 was further analyzed using Quantitative Real-time RT-PCR (QRT-PCR). This gene is of particular interest because one hypothesis of AGL15 function based on native and ectopic expression analysis involves a role in developmental phase transition. Proteins involved in chromatin architecture like HMG proteins and chromatin remodeling factors have been proposed to have roles in phase transition (Reyes et al., 2002). In microarray analysis, At5g23405 showed a 1.14 fold increase of expression in \textit{agl15} vs. Ws but with a P-value < 0.029. Total RNAs extracted from 8-day old seedlings of \textit{agl15} and Ws were used for QRT-PCR. \textit{AGL15} is expressed in the shoot apical meristem during vegetative development in wild type plants with detectable levels of AGL15 accumulation occurring in very young seedlings (Fernandez et al., 2000). QRT-PCR has been adapted as the standard validation method for microarray analysis (Mutch et al., 2002). The primers used in the reaction for At5g23405 and \textit{TUB2} were designed to amplify a short PCR products (90-120 bp) specific to each gene. To minimize the effect of the primer on the efficiency of amplification, the PCR efficiencies of each primer pair were also determined by performing a separate set of PCR reactions. The purpose of this experiment was to determine the expression change of At5g23405 between \textit{agl15} and Ws. The relative expression level change of the At5g23405 was determined using a improved comparative CT method (Livak and Schmittgen, 2001; Pfaffl et al., 2002). The ratio of At5g23405 expression level in \textit{agl15} compared to Ws was calculated as \( \text{Ratio} = \frac{(E_{\text{target}})^{\Delta CT_{\text{target(control-sample)}}}}{(E_{\text{ref}})^{\Delta CT_{\text{ref(control-sample)}}}} \) (Pfaffl et al., 2002). The \( E_{\text{target}} \) and \( E_{\text{ref}} \) are the PCR efficiencies of the two primer pairs for target gene (AT5g23405) and reference gene (\textit{TUB2}), separately. \( \Delta CT_{\text{target(control-sample)}} \) and \( \Delta CT_{\text{ref(control-sample)}} \) are the crossing threshold (CT) differences for the target gene and the reference
gene in the control (Ws) and sample (agl15). The PCR efficiencies of the primer pairs of the target gene and reference gene, and the relative expression level indicated as expression ration of the target gene in wild type and agl15 plants were shown in Table 3.2 were obtained using the REST© program (Pfaffl et al., 2002). Both QRT-PCR and microarray showed that the expression of At5g23405 was significantly increased in agl15 while a larger fold change was observed in QRT-PCR than that obtained in microarray. The result also demonstrated that it is possible that genes exhibiting modest differences (<1.4) between agl15 and Ws on the microarray may actually respond to AGL15.

3.2.3 Characterization of At5g23405

At5g23405 has two forms of transcripts may resulted from alternative splicing, At5g23405.1 and At5g23405.2 according to the gene annotation on the TAIR website (http://www.arabidopsis.org). The two transcripts differ from each other in that the AT5g23405.2 has a deletion of one codon which resulted in the loss of an alanine residue at position 44. At5g23405.1 encodes a protein of 149 aa. EST and cDNA evidence support that the gene is expressed (http://www.chromdb.org/, http://arabidopsis.org). Using the protein sequence to search against the NCBI conserved domain database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), we found that At5g23405 contains a HMGB-UBF_HMG-box. The protein encoded by At5g23405.1 belongs to a family of high-mobility-group B (HMG B) proteins. The HMG protein family includes 3 subfamilies, HMGA, HMGB and HMGN (Grasser, 2003). Plant HMG proteins are a group of non-histone proteins associated with chromatin and may function to regulate transcription via modulation of DNA structure and chromatin remodeling (Grasser, 1995; 2003). The HMGB proteins were formerly named as HMG1/2 and containing HMG-box domains. An alignment of the protein At5g23405.1 and other previously characterized Arabidopsis HMGB proteins is shown in Figure 3.4. The proteins encoded by At5g23405 have a basic N-terminal domain, a central HMGB-box domain and an acidic C-terminal domain. The primary structure of At5g23405 is typical for plant HMGB proteins. It also can
Table 3.2 QRT-PCR analysis of At5g23405 expression

<table>
<thead>
<tr>
<th>Genes</th>
<th>TUB2</th>
<th>At5g23405</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR Efficiencies</td>
<td>1.68</td>
<td>1.98</td>
</tr>
<tr>
<td>Ws Means&lt;sup&gt;1&lt;/sup&gt;</td>
<td>20.4</td>
<td>26.567</td>
</tr>
<tr>
<td>agl15 Means&lt;sup&gt;2&lt;/sup&gt;</td>
<td>20.233</td>
<td>24.733</td>
</tr>
<tr>
<td>Expression Ratio&lt;sup&gt;3&lt;/sup&gt;</td>
<td>3.209</td>
<td></td>
</tr>
<tr>
<td>p-Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expression Ratio-&lt;i&gt;nn&lt;/i&gt;&lt;sup&gt;4&lt;/sup&gt;</td>
<td>1.09</td>
<td>3.499</td>
</tr>
<tr>
<td>p-Values-&lt;i&gt;nn&lt;/i&gt;</td>
<td>0.406</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Note:  <sup>1</sup>Ws Means are the means of three CT values determined using Ws RNA,
<sup>2</sup>agl15 Means are the means of three CT value obtained in QRT-PCR using agl15 RNA ,  
<sup>3</sup>Expression Ratio is the relative expression ratio of At5g23405 in agl15 vs. Ws with adjustment of PCR efficiencies of both pairs of primers and normalized with reference gene TUB2.  
<sup>4</sup>Expression Rations-<i>nn</i> is the relative expression ratio of TUB2 and At5g23405 in agl15 vs. Ws with adjustment of PCR efficiencies but without normalization with reference gene.
Sequences of At5g23405.1 and other characterized Arabidopsis HMGB proteins were aligned using Clustal X 1.83 (http://www-igbmc.ustrasbg.fr/BioInfo/ClustalX/Top.html). HMGB1-6 were previously characterized (Stemmer et al., 1997; Grasser, 2003). The conserved HMGB domains are from 85 aa to 184 aa on the ruler.
be seen in the Figure 3.4 that the HMGB-box DNA binding domain is conserved among all the family members. There are 9 possible HMGB proteins in Arabidopsis. A phylogenetic tree of all the members of the Arabidopsis HMGB proteins was constructed using Clustal X and is shown in Figure 3.5. The phylogenetic analysis showed that At5g23405.1 is more closely related to HMGB6 than to any other family members. Pairwise sequence comparison revealed that the two proteins are 37% identical over the entire amino acid sequence, 52% along the HMGB domain (data not shown and Grasser, 2003). HMGB6 has been shown to have structural and functional characteristics of HMGB proteins (Grasser, 2003). It is possible that At5g23405 is also truly a member of HMGB protein family even though more evidence is required to prove this hypothesis. To explore the biological roles At5g23405, the genomic region of At5g23405 was expressed constitutively using a CaMV 35S promoter. Transgenic Arabidopsis plants overexpressing AT5G23405 were obtained. No obvious morphological phenotypes were found for the 24 T1 transgenic lines in a preliminary screening.

Two T-DNA insertion mutants were also obtained. One line was from SALK T-DNA insertion collection and a second line was obtained from the University of Wisconsin Knockout Facility. The two T-DNA insertion lines were confirmed as having one insert by antibiotic resistance and bred to be homozygosity. Lack of full-length gene transcripts for At5g23405 was confirmed using RT-PCR (data not shown). For simplicity, the two lines were named *salk* and *uw*, respectively. No phenotypic difference could be observed between the mutants and wild-type Arabidopsis plants under normal growth conditions. To test whether AT5G23405 has a role in embryo production, seeds of *Salk* and Columbia were grown in liquid culture media containing 2,4-dichlorophenoxyacetic acid (2,4-D). Using this system, Mordhorst et al. found a group of Arabidopsis mutants showed enhanced production of embryos from the SAM (1998). In addition, AGL15 was also found to be able to enhance the production of embryos from SAM in the liquid culture system (Harding et al., 2003). As shown in Figure 3.6a, 18.9% (SE=2.6%, \( n = 4 \) experiments, 909 total seedlings scored) of the Columbia seedlings produced
Figure 3.5 A phylogenetic tree generated from At5g23405.1 and other HMGB family members in Arabidopsis

The tree was drawn using neighbor joining method and the bootstrap number was set to 10,000 (Clustal X 1.83). At5g23405.1 is indicated in bold.
Figure 3.6 Embryonic development from the shoot apex of seedlings in culture

(a) Percentages of seedlings that showed embryonic development from the shoot apaxes when allowed to complete germination in liquid media containing 2,4-D. Results shown are means of four replicates of the experiment, error bars are standard errors (SE). (b) and (c) AT5G23405 mutant (salk). Red arrowheads indicate development of embryo-like tissue at the shoot apex; green arrowheads indicate lack of development at the shoot apex. Bar = 1 mm.
embryos at the SAM, while 45.8% (SE=4.4%, n=4, 528 total seedlings scored) *Salk* homozygous seedlings produced embryos at the SAM. The embryonic structure developed from shoot apex in the liquid culture is shown in Figure 3.6b while a seedling without shoot apex development are shown in Figure 3.6c. The experiments were repeated several times and similar results were obtained for both *salk* and *uw* (data not shown).

### 3.3 Discussion

Microarray has been used extensively in many aspects of the life sciences, especially expression profiling to assess the transcript accumulation of genes in different organisms. There are two basic forms of microarray, one is cDNA-based and another is oligonucleotide-based. The Arabidopsis ATH1 GeneChip® Genome Array was used in our experiment. ATH1 GeneChip® is based on oligonucleotide and each probeset consists of 11 pairs of probes, each is a 25mer oligonucleotide. According to a recent review (Jordan, 2004), in an survey conducted to compare all the existing platforms, Affymetrix GeneChip® microarray was found to be relatively more reliable than other platforms when the results were compared with quantitative PCR data to assess the differentially expressed genes.

However, similar to other biological experiments, data obtained from microarray can be variable. It is important to extract reliable information for tens of thousands genes on the Arabidopsis microarray chips. We used three biological replicates for each genotype, *Ws* and *agl15*. The use of biological replicates is important since this may decrease the chance of the fluctuation of gene expression in a particular sample in the microarray experiment (Churchill, 2002). In addition, in our experiment, statistical analysis was performed instead of using 2-fold change to determine the differentially expressed genes between the two genotypes. The drawback of using 2-fold change as the parameter to determine the differential expression of genes is that the 2-fold change of the expression of
a particular gene can occur simply by chance. The genes obtained in our experiments are statistically significant, which would be more reproducible and thus may represent the real targets of AGL15.

To strengthen the robustness of the microarray data, other than experimental design and statistical analysis, validation using independent methods is also important. In our experiment, RT-PCR and QRT-PCR were used to confirm the expression change for a few genes. It was found that microarray data were mostly supported by RT-PCR and QRT-PCR. Although the fold-change value observed in QRT-PCR was larger than that of microarray, this might due to the fact that two different tissues (young seeds vs. young seedlings) were used. Another possibility is that microarray may underestimate the magnitude of the change in gene expression and this was found to be true in some cases (Chang et al., 2000). Nevertheless, the result showed that the expression of some genes might be truly different even though the fold-change assessed by microarray is less than 2-fold or even 1.4 fold. It has been suggested that the cutoff value can be as low as 1.4, providing enough replicates are used (Zik and Irish, 2003). In the future design of the microarray experiment, more biological replicates can be used so as to have higher statistical power to more reliably recover genes that are truly regulated by AGL15.

It is intriguing that most genes identified in the microarray are of unknown function (75 out of 102), followed by genes involved in metabolism and only two genes encode potential transcription factors. It is not unusual to observe that many genes recovered by microarray as putative targets of MADS transcription factors are either related to basic cellular function (AP3/PI regulated genes, Zik and Irish, 2003) or of unknown function (Gomez-Mena et al., 2005). Therefore, to characterize the function of AGL15, it is necessary to understand the biological roles of the unknown proteins and proteins involved in metabolic pathways. Nevertheless, a few genes recovered in the microarray experiment will be discussed here in that they might be functionally relevant to AGL15 or embryogenesis.
Two genes, At1g52690 and At4g25470 were recovered. At1g52690 encodes a LEA (late embryogenesis abundant) protein while At4g25470 encodes CBF2 (CRT (C-repeat)/DRE (dehydration-responsive element) binding factor 2). Both proteins may be involved in cold tolerance and dehydration tolerance (Fowler and Tomashow, 2002). Arabidopsis embryogenesis normally goes through morphogenesis, maturation and desiccation. The two proteins might be involved in the later stages of the embryo development since later stages of embryogenesis involve in water loss. However, the seeds we used in the microarray were roughly at the morphogenesis stage and AGL15 may repress these genes at this earlier stage. In wild type seeds, AGL15 expression and the accumulation of its protein product decreases with age, which would be consistent with the fact that the expression of these genes are increased at later stages of embryogenesis (Deseny et al., 2001). Similarly, major storage proteins are synthesized during late stages of the embryogenesis, therefore the repression of ribosomal proteins (rps12.2, rps15, rps16) might suggest the inhibition of the storage protein synthesis programs before the maturation and desiccation stages. On the other hand, in agl15 the decreased expression of a pectinase gene At3g07820 (PGA3) might indicate that AGL15 plays a role in cell wall synthesis. As note by Zik and Irish (2003), cell wall-associated protein might have important roles in plant morphogenesis. PGA3 encodes polygalacturonase 3/pectinase, which can degrade pectin. Pectin has roles during cell wall formation and cell expansion. Modulation of pectin structure within the cell walls may regulate cell development (Willats et al., 2001). However, further investigation is necessary to understand how the expression of these genes are regulated by AGL15 and how they relate to seed development.

AT5G23405 was found to be differentially expressed between Ws and agl15 but well below the 2-fold cutoff. However, it was found to be significantly expressed with a ratio of more than 3-fold between Ws and agl15 as assessed using QRT-PCR, which suggested that AT5G23405 may also be a true target of AGL15. The liquid tissue culture experiment demonstrated that the lack of AT5G23405 has a postitive effect on the somatic embryo production from shoot apex. The
result is consistent with the hypothesis that AT5G23405 expression is repressed by AGL15 since ectopic expression of AGL15 can enhance somatic embryo production in the same liquid culture system (Harding et al., 2003).

Microarrays can provide the information about the differential expression of a gene between two samples. However, the genes identified in microarray may or may not be direct targets of AGL15. The expression changes of some genes may be secondary effects. To find out whether At5g23405 is a direct target of AGL15, enrichment PCR was performed to test the binding of AGL15 to two possible CArG motifs in the At5g23405 promoter region. However, the results showed neither region was significantly enriched in the ChIP population derived from embryonic tissue culture (data not shown). The enrichment PCR suggested AGL15 does not bind to the promoter of At5g23405. However, AGL15 may bind to other regions of AT5G23405, such as introns or 3' regulatory region. It is also possible that AGL15 binds to other unsuspected non-canonical CArG motifs in the promoter region. All these possibilities still remain to be examined. Alternatively, At5g23405 might be a indirect targets of AGL15.

HMGB proteins play roles in organizing the transcriptional protein complex in transcription regulation by bending DNA. It also has been shown that HMGB proteins can regulate transcription through chromatin remodeling (reviewed in Agresti and Bianchi, 2003). It has been demonstrated that mutation of certain HMGB proteins in yeast or mice caused severe growth defects (Grasser, 2003). However, the functions of plant HMGB proteins are largely unknown. Only in one case has a phenotype been associated with altered HMGB gene expression. The ectopic expression of maize HMGB1 caused subtle transient defects in root elongation in tobacco seedlings due to reduced cell division rates (Lichota et al., 2004). Part of the reason for lack of a prominent phenotype may be that many HMGB genes are ubiquitously expressed and simultaneously exist in one tissue (Yamamoto et al., 1998; O’neill et al., 1998; Stemmer et al., 1999; Wu et al., 2003; Grasser, 2003). In fact, when we searched the public database to explore the expression pattern of AT5G23405, it was found that the gene was reported to
be expressed in leaves, flowers, siliques and seedlings as tested using RNA gel blot (http://www.chromdb.org/cgi-bin/data/rnaexp.cgi?id=HMGB000012). AT5G23405 was found to be expressed ubiquitously in leaves, inflorescence, seedlings, roots and siliques in the Massively Parallel Signature Sequencing (MPSS) database (http://mpss.udel.edu/at/GeneAnalysis.php?featureName=At5g23405). The redundancy of the family members may also partly explain the lack of obvious growth phenotypes in normal conditions for loss-of-function of At5g23405. Biological systems are characteristically redundant and robust. Nevertheless, the characterization of the loss-of-function of At5g23405 mutants on the somatic embryogenesis may shed light on its function in normal growth conditions in wild type. The At5g23405 overexpressing lines and complementation lines of At5g23405 knock out line will be assessed for their abilities of making somatic embryos. Furthermore, the homozygous plants of the At5g23405 overexpressing lines and loss-of-function lines will be subjected to more biotic or abiotic treatment to reveal any specific roles At5g23405 might play in plant development.

3.4 Materials and methods

3.4.1 Arabidopsis plant germination and growth

The seeds were sterilized 3-4 times in 70% ethanol containing 0.1% Triton X-100 for 2 min, rinsed with twice with 95% ethanol, then poured and dried on a sterile Whatman No.1 filter paper in a sterile hood. Sterilized seeds of Arabidopsis thaliana ecotype Ws (Wassilewskija) or Columbia wild-type plants and transgenic plants were sown on GM germination media (GM/Kan media for transgenic seed and GM germination media supplied with 5 µg/mL BASTA (Crescent Chemical Co., Islandia, NY) for agl15 seeds) and transplanted to ProMix BX (Premier Brands, Inc., Quebec, Canada) after 7–10 days. Plants were grown at 20/18 °C under a 16-h light/8-h dark regime. Seedlings at 8 days old
were collected from GM media without transferring to ProMix BX and used for RNA extraction. For staged seeds, flowers were tagged on the day that they opened and collected at the appropriate time afterwards.

3.4.2 Arabidopsis RNA extraction

The 5 to 8-day-old Ws and agl15 seeds were collected, flash frozen and ready for RNA extraction. The method for Arabidopsis seeds RNA extraction was essentially as described in Gehrig (Gehrig et al.). Briefly, approximately 60 mg Arabidopsis seeds were ground in liquid N$_2$ with mortar and pestle, 1% (w/v) high molecular weight PEG (15,000 - 20,000 Dalton, Sigma) was added to the ground powder in 450µL RLC buffer (supplied with Qiagen RNeasy Plant Mini Kit, Qiagen Inc., Valencia, CA) and incubated in room temperature for 5-10 min, and then centrifuged for 2 min at 13,000 rpm in an Eppendorf 5415D centrifuge (Eppendorf AG, Hamburg, Germany). The rest of the steps were performed according to the instruction provided by the manufacturer for using Qiagen RNeasy Plant Mini Kit (Qiagen).

Total RNAs from seedlings were isolated using TRIzol Reagent (Invitrogen Corp., Carlsbad, CA) following manufacturer's instructions.

3.4.3 Microarray

RNA from 5 to 8-day-old Arabidopsis seeds of Ws and agl15 total RNAs were prepared as described above. Three separate preparations of RNA samples were performed for each sample, Ws and agl15. The microarray experiment was carried out using Arabidopsis ATH1 GeneChip® Genome Array (Affymetrix Inc., Santa Clara, CA). The preparation of probe, microarray hybridization, data acquisition and statistical analysis were performed at UK Microarray Core Facility (University of Kentucky, Lexington, [http://www2.mc.uky.edu/UKMicroArray/](http://www2.mc.uky.edu/UKMicroArray/)). After hybridization, results were analyzed using Affymetrix Microarray Suite v5.0 (Affymetrix Inc.). Based on the one-sided Wilcoxon signed rank test (Liu et al., 2002), the hybridization signal for each probe set was assigned an Absolute Call,
designated as being Present, Marginal, or Absent. Further analysis using student’s t-test was carried out using SAS® v.9.0 (SAS Institute Inc., Cary, NC). The probe sets that determined to be significantly different from each other between Ws and agl15 (p<0.05) were exported to Excel. The statistical analysis was done by bioinformatics group of the UK Microarray Core Facility.

The identity of the genes corresponding to each probe set on the two significant lists (significantly down- or up-regulated) was retrieved using batch query tools of the NetAffx (https://www.affymetrix.com/analysis/netaffx/batch_query.affx). Functional categories were assigned to genes using the AGI number to search the MIPS database (http://mips.gsf.de/cgi-bin/proj/thal/) and the Arabidopsis Information Resource website, TAIR (http://www.arabidopsis.org/).

3.4.4 RT-PCR

After RNA preparation, 1 µg of total RNA was first treated with DNase I (Invitrogen) and then used for RT-PCR with specific primers for each transcript tested. Primer sets corresponding to EF1α-A1 genes (At1g07920 and At1g07940, Axelos et al., 1989) or TUB2 (At5g62690) were used in RT-PCR as control. The primers used in RT-PCR are:

At1g20700: Forward: 5'- CCTCTCTTCTTACCATCCACTC – 3'
             Reverse: 5' - CCCTATACTCAACAAATGCTCA - 3'
At4g25470: Forward: 5' - GTATAATAGCCTCCACCAAGG - 3'
           Reverse: 5' - CAATTTACAGAGGAGTTCGTCA - 3'
At4g14550: Forward: 5' - CTCGTAGCTTGGAAACATACTCA - 3'
           Reverse: 5' - TAACAAACAAGGACATGTGGAT - 3'
At2g46990: Forward: 5' - ATCTAATGTCTCTTAATGGCTACC - 3'
           Reverse: 5' - ATTAGCTTTGAAATCTTCAGTCT - 3'
At1g61566: Forward: 5' - GTGGGCTTTGTCAATGTA - 3'
           Reverse: 5' - TGCTCTCGTAGTGTTCTTGG - 3'
EF1α-A1:   Forward: 5' - ACGCTCTACTTGCTTACC - 3'
Reverse: 5' - GCACCGTTCAATACCACC - 3'
At5g23405: Forward: 5' - GTCCTACTTTGTGGAGTGGA - 3'
Reverse: 5' - CTGTGTACCAATGCAAGA A - 3'
TUB2: Forward: 5' - TGGGACACAAACTCAGGCTA - 3'
Reverse: 5' - TGTTCCTCCACGTTATCCTC - 3'

3.4.5 Quantitative real-time RT-PCR

After RNA preparation, 1.0 µg of total RNA was first treated with DNase I (Invitrogen) and used for first strand cDNA synthesis. Reverse transcription was performed using M-MLV Reverse Transcriptase (Promega) according to the manufacturer's instructions. Fluorescence PCR amplifications were performed in triplicate using the LightCycler™ (Roche Molecular Biochemicals, Indianapolis, Indiana, USA). A 1.0 µl aliquot of each first strand cDNA reaction was amplified by primer pairs specific to the At5g23405 and EF1α-A1 in a 20 µl reaction containing 1x PCR buffer, dNTPs at 200 µM each, 0.2 µM of each primer, 1.0 units of Platinum® Taq DNA Polymerase (Invitrogen), and the double stranded DNA binding dye SYBR™ Green I (Molecular Probes, Eugene, Oregon, USA). At5g23405 is the gene of the interest and EF1α-A1 is the reference gene. Amplification reactions consisted of 35 cycles of denaturation at 95°C (30 seconds), annealing at 55°C (30 seconds), and extension at 72°C (30 seconds). Fluorescence signals were obtained once in each cycle by sequential fluorescence monitoring of each sample tube at the end of extension. A fractional cycle number or crossing threshold (CT) was determined from the exponential phase of the fluorescence amplification profiles using the Roche LightCycler™ software. The PCR efficiency of the primer pairs were determined in a separate set of real time RT-PCR using series of dilution of cDNA inputs as templates.

Determination of PCR efficiency and change of the relative expression level of At5g23405 in the agl15 7-day-old seedlings relative to that in the Ws seedlings was accomplished using REST© software (Pfaffl et al., 2002).
The primers used for At5g23405 and TUB2 were the same as those described in section 3.4.4.

3.4.6 Seedling Liquid Culture

Seedling liquid culture was performed essentially as described by Mordhorst et al. (1998). First, seeds were sterilized 3-4 times in 70% ethanol containing 0.1% Triton X-100 for 2 min, rinsed with twice with 95% ethanol, then poured and dried on a sterile Whitman No.1 filter paper in a hood. The sterilized seeds were chilled for 2 day at 4°C and cultured in liquid culture media as described by Mordhorst et al. (1998). Cultures were incubated on a rotary shaker at 23°C to 24°C with a 23-h-light/1-h-dark regime. After approximately 3 weeks, cultures were scored.
CHAPTER 4

CONTROL OF EXPRESSION AND AUTOREGULATION OF AGL15
4.1 Introduction

Cells respond to growth or environmental stimuli by changing the expression of certain genes in the genome. To reprogram the expression of a set of genes, cells employ a collection of transcriptional regulatory proteins to accomplish the task (Wyrick and Young, 2002). A particular phenotype or the way the organism responds to a given biotic/abiotic perturbation is the outcome of interactions among numerous such interconnected regulatory cascades. It has been shown that MADS-box genes and their products interact at both the transcriptional and post-translational levels, with examples of autoregulation, cross-regulation and complex formation reported (Goto and Meyerowitz, 1994; McGonigle et al., 1996; Hill et al., 1998; Tilly et al., 1998; Egea-Cortines et al., 1999; Honma and Goto, 2001; Lamb et al., 2002). Therefore, the function of a transcription factor not only depends on what genes it regulates but also how the transcription factor itself is regulated.

*AGL15* is the only member of the MIKC subgroup identified to date that is preferentially expressed during embryo development (Heck et al., 1995; Rounsley et al., 1995; Perry et al., 1996). However, *AGL15* is also expressed at lower levels after completion of germination in restricted sets of cells including the vegetative shoot apical meristem, leaf primordia, young flower buds, and in the bases of expanding lateral organs (rosette and cauline leaves, and floral organs) (Fernandez et al., 2000). Global expression of developmental regulators at relatively high levels, followed by restricted lower level expression is not uncommon and may represent situations where a gene was co-opted to perform a new function as systems became increasingly complex (Miklos and Rubin, 1996; DeKoter and Singh, 2000; Imai et al., 2002; Emambokus et al., 2003).

Potential target genes of AGL15 were identified by using ChIP microarray as described in Chapter 2 and 3. However, we know little about how the expression of AGL15 is regulated. Identification of *cis* elements controlling particular aspects of temporal and spatial expression should provide insight into gene
function and factors involved in control of expression. To better understand the regulation of \textit{AGL15} expression, we dissected \textit{AGL15} regulatory regions to identify \textit{cis} elements important for particular expression features using reporter constructs (Jefferson et al., 1987). The objects of this study are: (1) to identify DNA sequences that are involved in tissue specific expression pattern or expression level of \textit{AGL15} and (2) to explore the possibility that \textit{AGL15} directly regulates the expression of its own gene.

\subsection*{4.2 Results}

\subsection*{4.2.1 Identification of \textit{cis} elements important for \textit{AGL15} expression}

In order to better define regulatory regions important for correct temporal and spatial expression of \textit{AGL15}, a series of deletion constructs was generated. An initial construct consisting of 1260 bp 5' of the start codon of \textit{AGL15}, the first four codons of \textit{AGL15} translationally fused to the coding region of the $\beta$-\textit{glucuronidase (uidA, GUS)} gene, and 524 bp 3' of the stop codon of \textit{AGL15} was generated (Figure 4.1a, construct 1260; for simplicity, constructs are named according to the length of the 5' regulatory region). The length of 5' and 3' regions included in this construct was defined by identifying the intergenic regions to the coding sequence of \textit{AGL15} (At5g13790) in the Arabidopsis genome sequence database (The Arabidopsis Genome Initiative, 2000). ESTs for both predicted flanking genes (At5g13780 and At5g13800) are present in the database, indicating that they are not pseudogenes. Construct 1260 was stably introduced into \textit{Arabidopsis} ecotype \textit{Ws} plants, independent transgenic lines obtained, and GUS activity assessed. Expression of 1260 was the same as that of previously reported transformants harboring an \textit{AGL15:GUS} reporter construct that consisted of $\sim$2.5 kb 5' of the start codon and the first four codons of \textit{AGL15}, and $\sim$2.5 kb 3' of the stop codon regions (Fernandez et al., 2000, and data not shown). When the 524 bp sequence corresponding to the 3' region downstream of the \textit{AGL15} stop codon was deleted (1260 $\Delta$3'), the expression level of \textit{GUS}
Figure 4.1 Deletion analysis of the AGL15 promoter in seedlings

(a) Schematic representation of the translational fusions between different regions of the AGL15 promoter (hatched box), the first four codons of AGL15 and the GUS reporter gene (not shown). The bent arrow denotes the translational start site. Length of the promoter fragments are indicated from the 5' end to the translational initiation codon. Black boxes indicates three potential binding sites for MADS domain proteins. The white box indicates a potential cis element for response to auxin.

(b) Relative GUS activities of ten-day-old seedlings containing constructs in the deletion series compared to the 1260 construct. Average activities were obtained by analysis of 18-25 individual lines containing each construct. Error bars represent standard error.
was significantly decreased to less than half of that of 1260 as assessed using 4-methylumbelliferyl-β-D-glucuronide (MUG) assays (Gallagher, 1992) (Figure 4.1b). All other constructs included the 3' region. Thus, the 1260 bp 5' region and 524 bp 3' region of the coding region for AGL15 encompassed all the regulatory elements important for correct temporal and spatial expression of AGL15 and could serve as a starting point for identification of regions containing essential cis elements.

A series of GUS reporter constructs that contained different portions of the 5' regulatory regions of AGL15 were generated and stably introduced into Arabidopsis to identify cis elements important for the control of AGL15 expression (Figure 4.1a). The expression pattern of each construct was assessed using GUS activity quantitatively in 10-day-old seedlings of 18 to 25 independent transgenic lines (Figure 4.1b), identifying regions with positive or negative effects on the expression of the reporter constructs (Figure 4.1b and data not shown). For example, the GUS activity of seedlings harboring the 1158 transgene (deletion to –1158 bp, Figure 4.1a) was approximately one-third that of 1260 seedlings. However, further deletion of the 5' region to –1056 bp increased the GUS activity to nearly 1260 levels and further deletion to -956 bp significantly increased the level of GUS activity above that found in 1260. Deletion of the 5' region to –859 of the start codon decreased GUS activity to background levels in seedlings (Figure 4.1b, compare to Ws, wild type) and in other tissues with the exception of low-level staining observed in developing pollen (data not shown). Likewise, no GUS activity, with the exception of developing pollen, was observable upon further deletion to -444.

To examine potential elements within the 5' region downstream of –860, a series of internal deletion constructs were created. All of the internal deletion constructs contained the –1260 to –860 region to ensure that the necessary elements for expression of AGL15 were retained. An internal deletion of –859 through –756 within the –1260 5' region (construct ∆859-756, Figure 4.1a) did not significantly
change the expression level of the reporter gene compared with 1260 (Figure 4.1b). A larger deletion encompassing –859 to –639 significantly increased the expression level to about 2.3 fold that observed in 1260 transgenic seedlings (Figure 4.1b). Further internal deletions towards the translational start codon of AGL15 first decreased GUS activity to a level similar to that found for 1260 (Figure 4.1a and b; Δ859-383 and Δ859-303) and then dramatically increased expression (Δ859-207 and Δ859-128, Figures 4.1a and b).

In general, transgenes within the deletion series that showed GUS activity also showed a similar spatial pattern of staining as found for 1260 and as previously reported (Fernandez et al., 2000). GUS activity was present in the developing embryo, seedling shoot apical meristem (data not shown), and at some stages of flower development, predominantly in the abscission zone (Figure 4.2a). Like 1260 (Figure 4.2a) and the previously reported expression pattern for AGL15:GUS (Fernandez et al., 2000), for most constructs that exhibited GUS expression, no GUS activity was detected at the base of the floral organs by the time of anthesis. However, transgenic plants containing some constructs within the series showed some dramatic changes in the pattern of GUS activity in reproductive tissues. Plants containing transgenes Δ859-207 and Δ859-128 not only showed increased GUS activity in SAM of seedlings (Figure 4.1b), but also prolonged expression in the abscission zones, in open flowers and even as late as nearly mature siliques (Figures 4.2b and 4.2c, compare to 1260, Figure 4.2a).

4.2.2 AGL15 expression responds to auxin

Generally, treatment with exogenous auxin is important for induction of somatic embryogenesis (recent reviewed in von Arnold et al., 2002; Feher et al., 2003). Because AGL15 accumulates in a wide variety of embryonic tissues, including somatic embryos (Perry et al., 1999), we examined the regulatory regions of AGL15 for elements that have been reported to be involved in auxin response. A sequence similar to that corresponding to part of the NDE element from the soybean SAUR (Small Auxin-Up RNA) 15A gene promoter (Xu et al., 1997),
Figure 4.2 Effects of AGL15 promoter deletions on the expression pattern of AGL15:GUS

(a) GUS activity in 1260 floral buds. Note the absence of GUS activity in flowers postanthesis (white arrowheads).

(b) GUS activity persists in the cells at the base of a nearly mature Δ859-207 silique (blue arrowhead).

(c) GUS activity in Δ859-128 floral buds. Note the persistent GUS activity detected in the basal portion of the flowers post-anthesis (blue arrowheads).
that is involved in auxin response, was found -1050 bp 5' of the ATG start codon of AGL15. To determine if the regulatory regions of AGL15 respond to addition of exogenous auxin, and whether the NDE-like element may be involved in this response, transgenic lines with constructs 1260 (includes the partial NDE sequence) and 956 (lacks the partial NDE sequence) were assessed for auxin response by treating eight-day transgenic seedlings with the synthetic auxin 2,4-D. As shown in Figure 4.3a, GUS activity was significantly increased in the auxin treated samples compared to the non-treated controls for transgenic lines carrying 1260. However, the increase in GUS activity in response to exogenous auxin was much lower for transgenic lines carrying construct 956 and was not significantly different from untreated samples (Figure 4.3a). Additionally, semi-quantitative RT-PCR was performed to determine the accumulation amounts of endogenous AGL15 and transgenic GUS transcripts in response to exogenous auxin in 1260 and 956 plants (Figure 4.3b). The increased accumulation of GUS transcript with 2,4-D treatment was significantly higher in 1260-seedlings than in 956-seedlings, while the native AGL15 transcript in both transgenic lines showed a similar induction of expression in response to 2,4-D (Figure 4.3b).

### 4.2.3 AGL15 represses its own expression

Because MADS-domain proteins have been reported to regulate their own expression as well as to control expression of other MADS-box genes (Hill et al., 1998; Tilly et al., 1998; Lamb et al., 2002) AGL15 was evaluated for potential for feedback control of expression. AGL15:GUS reporter constructs were introduced into backgrounds where AGL15 (35S:AGL15) was constitutively expressed. A transgenic line containing a single homozygous insert of 1260 was used for crosses to plants carrying a hemizygous copy of the 35S:AGL15 transgene. The F1 seedlings from these crosses segregated 1:1 for the presence and absence of the 35S:AGL15 transgene and could be separated at the seedling stage (Fernandez et al., 2000). As shown in Figure 4.4a, decreased GUS activity was observed in the siblings carrying the 35S:AGL15 transgene compared to the
Figure 4.3 Response of AGL15 expression to exogenous auxin

(a) Fold-change of GUS activity with auxin treatment of seedlings carrying the 1260 or 956 transgenes compared to untreated seedlings. Average values and standard errors are obtained by measurements on 18 to 25 individuals from a representative line homozygous for the 1260 or 956 transgene.

(b) RT-PCR products using oligonucleotide primers to assess accumulation amounts of the GUS transcript (from the 1260 and 956 transgenes) and the endogenous AGL15 transcript in seedlings with and without 2,4-D treatment. EF1α serves as an equal loading control and products are analyzed on agarose gel.
Figure 4.4 Response AGL15:GUS expression to AGL15 levels and forms.  
(a) Arabidopsis plants homozygous for the 1260 transgene were crossed to plants hemizygous for a 35S:AGL15 transgene. The F1 plants with or without 35S:AGL15 transgene were separated and stained for GUS activity.  
(b) Relative GUS activities of AGL15:GUS (1260) in 35S:AGL15 compared to wild type ten-day-old seedlings generated as described in (a).  
(c) Relative GUS activities of AGL15:GUS (1260) in an agl15 mutant compared to wild type backgrounds.  
(d) AGL15:GUS (1260) activity in response to a form of AGL15 that includes a strong transcriptional activation domain (VP16). F1 plants were generated as described in (a) except that a hemizygous 35S:AGL15-VP16 plant was used for the cross instead of 35S:AGL15.
siblings expressing AGL15 only from the endogenous gene. MUG assays confirmed that GUS activity in 10-day-old seedlings constitutively expressing AGL15 was ~20% the level of activity found in wild type siblings (Figure 4.4b). Conversely, when the 1260 transgene was introduced into agl15 plants, GUS activity was significantly increased to about 1.5-fold (Figure 4.4c). agl15 plants contained a T-DNA insertion in the first intron, and no correctly spliced full-length transcript could be detected by RT-PCR (Lehti-Shiu, M.D. and Fernandez, D.E., University of Wisconsin-Madison, personal communication).

To further test for regulation of AGL15 by its own gene product, we examined whether AGL15 expression responds to form of AGL15. AGL15:GUS (1260) was crossed with a transgenic line that constitutively expressed a modified form of AGL15 that included a strong transcriptional activation domain (Sadowski et al., 1988; 35S:AGL15-VP16). The 35S:AGL15-VP16 transgene caused tight upward curling of cotyledons and rosette leaves. In the F1 segregating population, GUS activity was increased and present in a broader domain in the AGL15-VP16 siblings compared to wild type siblings (Figure 4.4d).

4.2.4 AGL15 directly regulates the expression its own gene

Expression of AGL15:GUS in backgrounds accumulating different amounts and forms of AGL15 indicated the existence of a negative feedback loop leading to a net repression of expression. However, autoregulation may be direct or indirect. MADS-domain proteins bind DNA sequences referred to as CArG motifs that have a canonical sequence of CC(A/T)$_6$GG, (reviewed in Riechmann and Meyerowitz, 1997). No canonical CArG motifs were identified in the 5' regulatory region of AGL15, but several non-canonical CArG motifs are present, including two with a form of C(A/T)$_8$G that AGL15 preferentially binds in vitro (Tang and Perry, 2003). These two potential cis elements are located at –1198 to –1189 and -671 to –662 and referred to as CArG1 and CArG2 respectively. A third potential binding site was predicted using MatInspector (http://www.genomatix.de), is located at -279 to –270 (CArG3) and has the form
CC(A/T)nNNGG. Both AGAMOUS (Shiraishi et al., 1993) as well as AGL15 (Tang and Perry, 2003) can bind this type of DNA sequence \textit{in vitro}.

To test whether AGL15 binds directly to its own promoter, enrichment assays were performed. This assay investigates whether a given DNA fragment is represented at higher levels in a DNA population selected by chromatin immunoprecipitation (ChIP) than in the total input DNA population relative to a non-selected control. AGL15-specific antiserum was used to select \textit{in vivo} formed AGL15-DNA complexes from a embryonic tissue culture that constitutively expresses AGL15 (Harding \textit{et al.}, 2003), the co-precipitated DNA recovered and assessed by multiplex PCR for the presence of AGL15 promoter fragments compared a portion of the coding region of β-2 TUBULIN (Snustad \textit{et al.}, 1992), that is not expected to be bound by AGL15. While the AGL15 promoter fragments and the control fragment were both present in total DNA, the AGL15 promoter fragments were enriched after ChIP selection of AGL15 binding sites as shown in Figure 4.5a. The three regions containing CArG boxes were all enriched compared to the control, suggesting that they all can be bound by AGL15 \textit{in vivo}. The two CArGs most proximal to the transcription start site were reproducibly enriched to a greater extent than was CArG1 (Figure 4.5a).

To determine whether AGL15 can bind to its own regulatory regions in tissue where AGL15 is expressed only from the endogenous promoter, we performed ChIP experiments using \textit{B. napus} developing seeds and zygotic embryos. Portions of the 5′ regulatory region of Brassica AGL15-1 and AGL15-2 are available in the database (accession numbers U22665, U22681) and longer sequence information was obtained for both genes by TAIL-PCR. 5′-promoter regions (1133 nucleotides) obtained from \textit{Brassica napus} for AGL15-1 and AGL15-2 were 55.6% identical to each other and both were over 50% identical to the Arabidopsis 5′ region when aligned using EMBOSS-needle (http://www.ebi.ac.uk/emboss/align/). Both Brassica genes contain a conserved region that includes a putative CArG3 site in a similar position and context as found in the 5′ region of Arabidopsis AGL15. A region containing a sequence
**Figure 4.5 Binding of AGL15 to the 5' regulatory region of its own gene**

(a) Mapping of *in vivo* binding sites of AGL15 across its own 5' regulatory region. Oligonucleotide primer pairs to amplify regions of the *AGL15* promoter containing potential binding sites for AGL15 and to amplify control regions not expected to be bound by AGL15 (*TUB2*) were used in multiplex PCR on total DNA diluted 125- and 625-fold and on DNA recovered by immune precipitation (I) or preimmune precipitation (PI) in ChIP with embryonic culture tissue. CArG1, 2 and 3 are as in Figure 4.1a.

(b) AGL15 binds to its own regulatory regions in *B. napus* zygotic embryos. Oligonucleotide primers to amplify the proximal 5' region of *Brassica napus* AGL15-1 as well as the coding regions for a *Brassica tubulin* (*TUB*) were used in multiplex PCR on total DNA (diluted 25-, 125- and 625-fold) and immune (I) and preimmune (PI) ChIP populations from developing *B. napus* embryos.

(c) Multiplex PCR as described in (a) for the region of the *AGL15* promoter that includes CArG3 using ChIP populations derived from wild type Ws and *agl15* seedlings.

(d) Autoradiography of EMSAs to assess the interaction of AGL15 with CArG2. A 37-bp fragment including the CArG2 site was used in EMSAs. Lane 1 contained 50 ng AGL15 with a C-terminal T7 tag, expressed and purified from *E. coli* and incubated with the radiolabeled DNA fragment. Lane 2 lacked AGL15-T7 protein. Lane 3 added 0.4 µg of anti-T7 antibody to the reaction in lane 1. Filled arrowheads indicate supershift caused by addition of the antibody. Lanes 4 through 6 contained increasing amounts of unlabeled CArG2 probe as competitor (50-, 200-, and 400-fold in excess to the radiolabeled probe). Lanes 7 through 9 contained increasing amounts of unlabeled mutated CArG2 probe where the "C" in the CArG motif was changed to a "T" as competitor (50-, 200-, and 400-fold in excess).
similar to CArG2 was also identified for AGL15-2. Sequence information did not extend far enough to identify potential CArG1 sites, and no obvious CArG2 site was identified in the sequence obtained for AGL15-1. DNA fragments containing CArG2 and 3 from AGL15-2 and CArG3 from AGL15-1 were enriched in ChIP populations derived using AGL15 immune serum and B. napus embryos or developing seeds (Figure 4.5b).

To assess whether AGL15 binds its own regulatory regions after completion of germination, enrichment assays were performed on WS and agl15 seedlings. As shown in Figure 4.5c, the 5' regulatory regions were enriched in ChIP populations derived from WS seedlings and using anti-AGL15 immune serum, but not in populations from agl15 seedlings, indicating that AGL15 can bind to upstream regulatory elements when expressed from the endogenous gene in seedlings.

Binding of AGL15 to the individual CArG motifs was further analyzed in vitro by electrophoretic mobility shift assay (EMSA). Oligonucleotides corresponding to the three CArGs were synthesized, radiolabeled and used in EMSA with AGL15 containing a T7 tag at the C-terminal end. As shown in Figure 4.5d, AGL15 was able to bind to CArG2 in vitro, but not to CArG1 or CArG3 (data not shown). Competition with unlabelled CArG2, but not with an unlabelled form of CArG2 where the C(A/T)₈G was changed to T(A/T)₈G, demonstrated sequence specificity of binding of AGL15-T7 to radiolabeled CArG2 (Figure 4.5d, lanes 4-6, unlabelled CArG2, lane 7-9, unlabelled mutant CArG2). Moreover, addition of T7 antibody decreased the abundance of the shifted band and caused appearance of a supershifted band (Figure 4.5d, lane 3).

4.2.5 Site specific mutagenesis of the CArG boxes within the AGL15 promoter

Previous work has demonstrated that the C and G in CArG motifs are highly conserved and important for in vitro binding of MADS-domain proteins (reviewed in Shore and Sharrocks, 1995; Riechmann and Meyerowitz, 1997), including
AGL15 (Perry et al., 1996; Wang et al., 2002; Tang and Perry, 2003; Wang et al., 2004b). Therefore, these bases were targeted in site-directed mutagenesis of transgenic construct 1260 to determine the impact of presence of the three CArG motifs on expression of AGL15 in vivo. No changes in the spatial pattern of GUS activity were observed (data not shown), but site-directed mutagenesis of CArG2 or CArG3 significantly changed the level of GUS activity. As shown in Figure 4.6, mutation of CArG2 caused a significant increase in GUS activity to nearly twice that observed in 1260. Conversely, the GUS activity in the CArG3 mutant was significantly decreased to about one-half of that of 1260 containing lines. Mutation of CArG1 had no significant effect on GUS activity.

4.3 Discussion

Gene expression is regulated by a complex interaction of cis-acting DNA elements and trans-acting regulatory proteins to generate the correct spatial and temporal pattern of gene expression. These interactions may lead to induction or repression of gene expression with a variety of inputs “read” by the basal transcriptional machinery, either via direct interaction or through chromatin remodeling (reviewed in Arnosti, 2003). Deletion analysis experiments are a valuable approach towards identifying regions containing relevant cis elements.

Deletion analysis of the AGL15 promoter indicated the presence of putative positive and negative cis elements contributing to expression of AGL15. The region encompassing 1260 bp 5’ of the initiation codon of AGL15 and 524 bp 3’ of the stop codon contained all of the elements to drive expression in the same spatial and temporal contexts and to a similar level as a previously reported AGL15:GUS reporter construct (Fernandez et al., 2000). Deletion of the 3’ region did not change the pattern of expression but reduced the GUS activity observed to less than half that found with inclusion of the 3’ region. This observation is consistent with a role for the 3’ end either in transcriptional control of expression level, or in post-transcriptional controls such as processing, stability of transcript or even translational control, all of which have been reported in the literature.
Figure 4.6 Effect of site-directed mutagenesis of potential CArG motifs on GUS expression levels

Site directed mutagenesis was used to generate versions of the 1260 transgene where each of the three predicted CArG motifs were mutagenized such that AGL15 would no longer be predicted to bind. 1260m1, -m2, and -m3 refer to mutagenesis of CArG1, 2 and 3 respectively. The relative level of GUS activity compared to 1260 was determined from analysis of 18 to 25 independent transgenic lines for each construct. Average values and standard errors are shown.
(reviewed in Day and Tuite, 1998; Gutiérrez et al., 1999). For the purpose of this study, the 3’ end was included for all other constructs to identify cis regions in the 5’ regulatory regions.

The region from -1260 through -859 was particularly important for expression; any deletions more proximal to the start codon resulted in nearly complete loss of reporter expression. This region included sequence similar to that corresponding to part of the NDE element from the soybean SAUR (Small Auxin-Up RNA) 15A gene promoter (Xu et al., 1997). The NDE element is generally composed of three regions, none of which is exclusively responsible for auxin response (Xu et al., 1997). The regulatory regions of AGL15 have sequences similar to two of the three elements, and the exact match, CATATG, also appears in the A1 portion of the SAUR 15A promoter that also plays a role in expression of this gene in response to auxin treatment (Xu et al., 1997). Transgenic plants with reporter transgenes that included this element showed an increase in GUS activity and in GUS transcript levels in response to auxin treatment, whereas loss of this element greatly reduced any response to auxin (Figure 4.3). The endogenous AGL15 gene responded to auxin in both transgenic lines (Figure 4.3). The small response of transgene 956 to auxin may be mediated by other elements within the promoter or the 3’ region. An auxin response factor binding site was predicted to lie within the 3’ region by searching the database of plant cis-acting regulatory DNA elements (PLACE, http://www.dna.affrc.go.jp/PLACE/wais.html, (Higo et al., 1999).

The observation that AGL15 expression is up-regulated by auxin was particularly intriguing. Auxin has been long known for its diverse and important roles in many developmental processes, (reviewed in Crozier et al., 2000; Kepinski and Leyser, 2003), including somatic embryogenesis where tissue explants are treated with exogenous auxin (e.g., in Arabidopsis thaliana; Sangwan et al., 1992; Wu et al., 1992; O’Neill and Mathias, 1993; Pillon et al., 1996; Luo and Koop, 1997; Ikeda-Iwai et al., 2002). Auxin is thought to induce embryogenic competence, but the mechanism by which this occurs is unclear (Harada et al., 1998). One
Arabidopsis system where exogenous auxin is not required involves culturing zygotic embryos with a 35S:AGL15 transgene (Harding et al., 2003). The embryogenic culture established in this system has been stably propagated without exogenous growth regulators for over 8 years to date (Harding and Perry, unpublished observation). It is intriguing to consider that exogenous auxin may not be needed in this system in part because AGL15 is constitutively expressed from a heterologous promoter. However, auxin is still required in the system of Mordhorst et al. (1998) where embryos are produced from the shoot apical meristem of seedlings that complete germination in liquid media (Mordhorst et al., 1998), perhaps reflecting the fact that some other factor(s) may be missing in this context. Recently the *Medicago truncatula* orthologue of *SOMATIC EMBRYOGENESIS RECEPTOR KINASE* (*MtSERK1*) was reported to be expressed in response to auxin within two days of culture initiation (Nolan et al., 2003), a similar timeframe as found for AGL15 (Figure 4.3). Expression of SERK has been reported in several systems as coincident with or able to promote development in an embryonic mode (Schmidt et al., 1997; Somleva et al., 2000; Hecht et al., 2001a). Ectopic expression of *WUSCHEL* in the presence of auxin also promotes embryogenesis and expression of *FUS3, LEC1* and *AGL15* are induced by the heat shock treatment that induces *WUS* expression and further up-regulated by auxin, but gene expression was assessed after two weeks of treatment by which time embryos were apparent (Gallois et al., 2004).

To identify *cis* elements important for proper expression from -859 towards the start site, a series of internal deletions were generated where the -1260 to -859 fragment was present, but deletions from -859 to the -128 were made. GUS activity increased, in some cases dramatically, or remained similar to transgene 1260 in this series (Figure 4.1). The transgene that caused the most dramatic increases in GUS activity in the seedlings (Δ859-207, approximately four-times the activity observed in 1260 seedlings) also had an obvious and dramatic effect on expression during reproductive development. Notably, GUS activity was persistent in Δ859-207, as well as the more extensive deletion, Δ859-128, in the base of flowers and siliques, even until the silique was nearly mature (Figures
This was quite different from that reported previously (Fernandez et al., 2000) and for transgene 1260 (Figure 4.2a) where GUS activity could be detected in young flower buds and at the bases of older buds, but was absent near the time of anthesis. These results may indicate elements within -302 to -206 region that repress AGL15 expression during late flower development. To test whether continuous expression of AGL15 in the bases of floral organs from its own regulatory regions was sufficient to delay senescence and/or abscission of perianth organs, AGL15 was expressed from a modified form of its own regulatory regions with deletion of -859 through -207. However, no changes in perianth retention were apparent. This agrees with Fang and Fernandez (2002), where overexpression of AGL15 within the abscission zone was insufficient to delay abscission of perianth organs.

GUS activity of transgene 1260 in seedlings, as well as other tissues such as flower buds, is dramatically decreased when in the AGL15 overexpressor (35S:AGL15) background (Figures 4.4a, 4.4b and data not shown). Conversely, the 1260 transgene yields significantly more GUS activity when present in a loss-of-function agl15 background (Figure 4.4c) indicating that the potential negative autoregulation is not simply an artifact of the overexpression construct. Finally, the 1260 reporter transgene was able to respond to the form of AGL15. A modified form of AGL15 that contains a strong transcriptional activation domain (the VP16 domain, Sadowski et al., 1988) was able to cause a dramatic increase and expansion of GUS activity within seedlings compared to wild type siblings from a genetic cross (Figure 4.4d). All of the above support a role for AGL15 in autoregulation to restrict levels of expression.

Autoregulation of AGL15 expression by its own product could be the result of direct or indirect regulation, or both. At least three potential binding sites for MADS-domain proteins are present within the 5' regulatory regions of AGL15. Two of these sites, CArG1 and 2 are of a unique C-8-G form with a longer than standard A/T-rich domain than the canonical CC-6-GG form. The C-8-G form has been shown to be preferentially recognized by AGL15 in vitro (Tang and Perry, 2006).
2003). Enrichment tests on populations of chromatin that have been selected for direct association of AGL15 by ChIP indicate that at least part of the autoregulatory effects may be due to direct regulation (Figure 4.5). 5' regulatory regions were found to be over-represented compared to total DNA or control immunoprecipitations using preimmune sera, not only in embryogenic culture tissue constitutively expressing AGL15 via a 35S:AGL15 transgene, but also in Brassica zygotic embryo and in Arabidopsis wild type seedlings where AGL15 is expressed only from its endogenous regulatory sequences. Notably, seedlings that are unable to accumulate AGL15 due to an insertional mutation show no enrichment of AGL15 5' regulatory regions (Figure 4.5c). EMSA experiments confirmed that AGL15 can bind in a sequence-specific manner to CArG2 in vitro (Figure 4.5d). Although AGL15 was not found to bind to CArG1 or CArG3 in EMSA, it should be noted that EMSA's represent an in vitro context independent of other co-factors or chromatin architecture and lack of binding in EMSA in no way negates the evidence from enrichment tests that AGL15 binds to DNA fragments including CArG1 and 3 in vivo. Other MADS-domain proteins have also been reported to bind sites in vivo that would not be recognized in vitro (Molkentin et al., 1995; Black et al., 1996).

The site-directed mutagenesis experiments are supportive of direct autoregulation of AGL15 by its own gene product. Mutation of CArG1 produced a slight, but not significant reduction of GUS activity levels (Figure 4.6). This was also the CArG motif containing - DNA fragment that was reproducibly least enriched in ChIP populations using AGL15-specific serum (Figure 4.5a and data not shown). The decrease is consistent with the 5' deletion experiments where loss of ~100 bp from the 5' end and including CArG1 decreased GUS activity. Mutation of CArG2 nearly doubles GUS activity compared to 1260, indicating a role for this site in negative regulation of expression. This finding is consistent with the internal deletion construct Δ859-639, that has deleted CArG2, resulting in a more than 2-fold increase in GUS activity compared to 1260. The next dramatic increase in GUS activity is observed when CArG3 is deleted in the Δ859-207 transgene where activity has nearly doubled again compared to Δ859-
However, site directed mutagenesis of CArG3 led to a decrease in GUS activity in seedlings, indicating that this site may be involved in up-regulation of AGL15. Transgenic plants with a reporter construct in which both CArG2 and 3 are mutated (1260m2m3) show nearly 1260 levels of GUS activity (data not shown). This type of effect, where loss of a positive regulatory factor or element may be compensated for by loss of a negative factor/element has been previously reported and appears to be relatively common (reviewed in Lee and Young, 2000). A transcriptional regulator may have stimulatory or repressive effects on gene expression at different sites (reviewed in Latchman, 2001) and it is possible that AGL15 first stimulates expression through CArG3, and as increased amounts of AGL15 accumulate, binding to other cis elements such as CArG2 lead to repression of expression. The increased and prolonged GUS activity observed for lines carrying the internal deletion construct \( \Delta 859-207 \) indicates involvement of more than just CArG2 and 3. When loss-of-function agl15 plants are transformed with the transgenic construct in which CArG2 and 3 are both mutated, the level of GUS activity is nearly two-fold that found for this construct in the wild type background (data not shown). This result indicated that there may be another relevant site for negative autoregulation of AGL15. Also supportive of this hypothesis is the fact that constitutive expression by the 35S:AGL15 transgene led to reduction of GUS activity for 1260m2m3 plants (data not shown), although indirect pathways could contribute in both of these cases as well. The \( \Delta 859-207 \) deletion caused loss of another potential, but non-canonical CArG motif more proximal to the transcriptional start site than CArG3—this CArG motif has a form CC-7-G that is recognized by AGL15 in vitro (Tang and Perry, 2003). In fact, there are at least two non-canonical putative CArG motifs 3’ to CArG3 that are conserved between Arabidopsis and the Brassica AGL15’s (data not shown).

Other MADS-box genes are involved in autoregulatory loops. In animals, the MADS domain protein MEF2 participates in a positive direct autoregulatory loop to maintain myogenic programs (Cripps et al., 2004). In Antirrhinum flowers, DEFICIENS and GLOBOSA (Schwarz-Sommer et al., 1992; Tröbner et al.,
and their counterparts in Arabidopsis APETALA3 and PISTILLATA (Hill et al., 1998; Tilly et al., 1998), form heterodimers to contribute to the expression of their own genes. In some cases, multiple CArG motifs involved in expression have been identified within the 5’ regulatory regions (Schwarz-Sommer et al., 1992; Hill et al., 1998; Tilly et al., 1998), but in other cases there are no obvious MADS-domain protein binding sites within the region responsible for autoregulation (Chen et al., 2000). This may indicate a role for indirect regulation, or the necessity for co-factors to facilitate binding perhaps to a non-canonical type CArG motif or via protein-protein interactions. As found for AGL15, multiple CArG motifs are present within the AP3 promoter that appear to act in a redundant manner within the context of the native promoter, and that individually contribute to expression or repression of AP3 (Tilly et al., 1998).

In summary, we have identified a region of the AGL15 promoter involved in auxin response, as well as regions important for cessation of expression at the bases of flower organs nearing anthesis. We have also presented evidence that AGL15 regulates its own expression and identified at least two cis elements involved in this response. Indeed, AGL15 expression fulfills the major tests for direct regulation: direct association with regulatory regions (Figure 4.5); and response of the regulatory regions to AGL15 amounts (Figure 4.4).

4.4 Experimental procedures

4.4.1 Plant material and growth conditions

Seeds of Arabidopsis thaliana ecotype Wassilewskija (Ws) were surface sterilized and sown on GM germination media and transgenic plants were germinated on GM/Kan media. Seeds were chilled for 2 days at 4 °C and then transferred to room temperature under cool-white fluorescent light. Seedlings were transplanted at 7 to 10 days to ProMix BX (Premier Brands, Inc., Quebec, Canada) and plants were grown under long day growth conditions (16-h light/8-h dark regime at 20/18 °C) in a Conviron growth chamber with fluorescent and
incandescent lights. *Brassica napus* cv. Tower plants were grown in the greenhouse. Flowers were pollinated on the day that they opened and embryos collected at approximately torpedo to early maturation stage.

### 4.4.2 Reporter constructs

#### 5' deletion constructs

A previously reported *AGL15:GUS* reporter construct (Fernandez et al., 2000) was used as template for PCR to generate the deletion series. In this construct, expression of β-glucuronidase (GUS) gene was controlled by ~2.5 kb 5' flanking sequence of the Arabidopsis *AGL15* and ~2.5 kb 3' flanking sequence (Fernandez et al., 2000) Upon completion of the genome sequence of Arabidopsis (Initiative, 2000), intergenic sequences of *AGL15* were determined to be 1260 bp 5' and 524 bp 3' of the coding region of *AGL15*. Therefore, the fragment containing 1260 bp of the *AGL15* 5' region, the first four codons of *AGL15* translationally fused to the GUS coding region, and the 524 bp of the sequence 3' to the coding region of *AGL15* was obtained by PCR using synthesized oligonucleotide primers (IDT Inc., Coralville, IA). Forward primers were designed to correspond to the 5' part of the nucleotide sequence of each deletion fragment and included a *SalI* site (construct 1260: 5'-ACGGGTCTGACAAGAAGCATGGAAACAGT-3', construct 444: 5'-ACGGGTCTGACCTGACCATTCCCTCCCT-3', construct 859: 5'-ACGGGTCTGACTTGCGGTCAACATTGTT-3', construct 1158: 5'-ACGGGTCTGACCATAAGGAAGTTTAGA-3', construct 1056: 5'-ACGGGTCTGACTATAGGAAGGCGACATT-3', construct 956: 5'-ACGGGTCTGACTTGAAGATACC3'; *SalI* site in bold). The common reverse primer (5'- CCGGAATTCGCCGCTGGCTGCTCAATTGCTCTC -3') corresponds to the 3' end of the 3' intergenic region of *AGL15* with extra bases added as *EcoRI* site. For construct 1260Δ3', the forward primers used was the same as that for construct 1260; the reverse primer used was: 5'-CCGGGAATTCCTAAACAGAGACCTTTTG -3'. PCR fragments were cloned into a
pGEM-T Easy® vector (Promega, Madison, WI), followed by SalI-EcoRI excision of inserts, gel purification, and ligation into the AGL15:GUS construct as described in Fernandez (2000), replacing the original SalI-EcoRI insertion fragment. All constructs were confirmed by sequencing. The constructs were introduced into Agrobacterium GV3101 and used to transform Arabidopsis as previously described (Wang et al., 2002).

Internal deletion constructs

Internal deletion construct Δ859-756 was obtained by inverse PCR using an intermediate construct for 1260 cloned into a pGEM-T Easy® vector (Promega) as template. One primer was designed to anneal at -860 (5'-AGGATACGACTTTGCCTCGAG -3') and another at -755 in the AGL15 promoter region with their 3’ ends facing away from each other such that when PCR was performed the fragment from -859 to -756 was deleted. The PCR product was blunted and self-ligated. The SalI-EcoRI fragment was excised and cloned into the AGL15:GUS construct of (Fernandez et al., 2000), replacing the original SalI-EcoRI fragment. The other internal deletion constructs Δ859-639, Δ859-383, Δ859-303, Δ859-207, Δ859-128, were similarly created with appropriate primers. All constructs were verified by sequencing and were introduced into Arabidopsis ecotype Ws plants as above. Primers used to generate internal deletion constructs are:

Δ859-756: 5’- CTTACACAGGCTATATATCCAAC -3’
Δ859-639: 5’- CCGGTGTTACAAAGCTGCTACT -3’
Δ859-383: 5’- CTAGTTGATAACATAATGGTAACC -3’
Δ859-303: 5’- TGTCGGTCAACATTGTTGGTC -3’
Δ859-207: 5’- GCAACACACAACATTACC -3’
Δ859-128: 5’- GCAATCTTTTGTGGTGCT -3’.

Mutagenesis constructs

Three potential cis elements that could bind MADS domain proteins called CArG
motifs were identified within the 5′ regulatory regions of AGL15. These three potential CArG motifs, referred to as CArG1, 2 and 3 were located at –1198 to –1189 (5′-CTATATAATG -3′), –671 to –662 (5′-CTATTTATTG -3′) and -279 to -270 (5′-CCAAATGTGG -3′) respectively. Nucleotides that were mutated are indicated in bold. One set of mismatch primers was designed for each site and used in mutagenesis PCR with PfuTurbo® DNA polymerase (Stratagene, La Jolla, CA), according to the manufacturer’s instruction. The plasmid template used for mutagenesis was construct 1260. The three resulting constructs are referred to as 1260m1 and 1260m2 where the C’s in CArG 1 and CArG2 were changed to T’s; and 1260m3 where the GG in CArG 3 was changed to TT. The mutated regions were confirmed by sequencing and constructs stably introduced into Arabidopsis as above. Primers used for mutagenesis of the three CArGs are:

CArG1: Forward: 5′- GCATCCGATGCTAGTTATATAATGTTGTC -3′
Reverse: 5′- GACAACATTATATACTAGCATCGGATGC -3′

CArG2: Forward: 5′- GATAAAATCTCGTTATTTATTTATTGATTTTGAG -3′
Reverse: 5′- CTCAAAATCAATAAATAAATAACGAGATTTTATC -3′

CArG3: Forward: 5′- AAATCCTCCAAATGTAAACAAAAAGGTATCATG -3′
Reverse: 5′- CATGATACCTTTTTGTTACATTTGGAGGATT -3′

4.4.3 β-Glucuronidase activity assays

For histochemical staining, Arabidopsis seedlings or tissues were immersed in GUS staining solution (Fernandez et al., 2000), vacuum infiltrated for 5 minutes and then incubated at 37 °C overnight. The chlorophyll was removed by incubating with several changes of 70% ethanol. Quantitative assays of GUS activity were performed according to Gallagher, (1992) (Gallagher, 1992), activity calculated as nmol 4-MU/min.µg of protein, and the relative activities were calculated accordingly. For each GUS reporter line, approximately 18-25 independent transgenic lines were assayed.

4.4.4 Auxin response and RT-PCR

Eight-day-old transgenic seedlings with constructs containing or lacking the
putative auxin response elements were transferred from GM plates to liquid B5 media as in (Ikeda-Iwai et al., 2002) with or without 9 µM 2,4-D, and cultured under light at ~60 rpm for 48 hours. The seedlings were collected in liquid nitrogen and stored at -80 °C for MUG assays or RNA extraction. RNA was isolated by standard protocols using TRIZol Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer’s recommendation, followed by DNase I (Invitrogen) treatment. Total RNA was used for RT-PCR with primers specific to each transcript tested. A primer set corresponding to EF1α-A1 gene (Axelos et al., 1989) was used in RT-PCR as control. Primer sequences for the genes tested are:

**AGL15:**
- Forward: 5’- CTCGAGCGCTCTCATAAACCACGACA -3’
- Reverse: 5’- GGTACCGCTTCAGGTGAGAATTTGC -3’

**GUS:**
- Forward: 5’- GGGCCAACAGTTCCTGATTA -3’
- Reverse: 5’- GAGCGTCGCAGAACATTACA -3’

**EF1-α A1:**
- Forward: 5’- ACGCTCTACTTGCTTTCACC -3’
- Reverse: 5’- GCACCGTTCCAATACCACC -3’

### 4.4.5 PCR enrichment test for evaluation of *in vivo* binding of AGL15

Multiplex PCR was used to test whether DNA fragments corresponding to regulatory region of AGL15 were enriched in DNA populations selected by chromatin immunoprecipitation (ChIP) using AGL15-specific antiserum compared to total input DNA and to control immunoprecipitations using preimmune serum. ChIP was performed as described in Wang et al. (2002) using embryonic culture tissue (Harding et al., 2003), *Brassica napus* zygotic embryos, or Arabidopsis seedlings. Oligonucleotide primers were designed to amplify DNA fragments containing the three potential CArG motifs in the Arabidopsis AGL15 5’-regulatory regions, or coding sequence from β-2 *Tubulin (TUB-2)*, (Snustad et al., 1992) as a control not expected to be bound by AGL15. Oligonucleotide primers were also designed corresponding to *Brassica napus* AGL15-1 and AGL15-2 regulatory regions and to a Brassica *Tubulin 2* coding region. PCR products were separated on a 2% (w/v) agarose gel and the image captured using a
Chemilmager (Alpha Innotech Corporation, San Leandro, CA). The primers used for enrichment PCR are:

**AGL15 CArG1:**
- Forward: 5’- GCATGGAACAGTCGTCTAGTG -3’
- Reverse: 5’- CGACTTTGCCTCGAGAAAAG -3’

**AGL15 CArG2:**
- Forward: 5’- TTTTCTCGAGGCAAAGTCGT -3’
- Reverse: 5’- CAATTTCAGAGTCCTGAAAGGA -3’

**AGL15 CArG3:**
- Forward: 5’- GGAAGAAAAGGGAAAGTAGGACC -3’
- Reverse: 5’- GAGAGAAGAGGTAGAAGGAAGAAGGAAGAAGGA -3’

**TUB2:**
- Forward: 5’- GTCCTACTTTGTGGAGTGGA -3’
- Reverse: 5’- CTGTGTACCAATGCAAGAA -3’

**Bn AGL15-1:**
- Forward: 5’- AGTGTTGAATTGCTTCGAGA -3’
- Reverse: 5’- CTATTGAAACTCCTTTTGGGG -3’

**Bn TUB:**
- Forward: 5’- CGAGAGGATCACAGCAATACAG -3’
- Reverse: 5’- GGATCCATTCCACACAAAGGAGAAGGAAGAAGGA -3’

### 4.4.6 Electrophoretic mobility shift assay (EMSA)

EMSA was performed essentially as described in Huang (Huang et al., 1993). Radiolabeled probe was generated as in Tang and Perry (2003). Protein-DNA incubations were performed in binding reaction buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM DTT, 5% glycerol, 50 µg /ml poly(dI.dC), 100 µg/ml BSA) with approximately $10^4$ cpm radiolabeled probe and 0.5 to 4 µg column purified T7 tagged AGL15 protein (Tang and Perry, 2003). Supershift was performed by addition of 0.4 µg T7 tag antibody (Novagen, Madison, WI). Competition assays were performed by addition of excess unlabeled wild type or mutated probe. The reaction mix was incubated at room temperature for 40 min, then separated on a 5% polyacrylamide gel in 0.5 x TBE. After drying the gel, the image was visualized by exposure to a phosphoimager screen (PhosphoImager 445SI-486, Molecular Dynamics, Sunnyvale, CA). The primers used to generate probe for AGL15 CArG2 EMSA are:

- Forward: 5’- GATAAAATCTCGTTATCTATTTATTGATTTTG -3’
- Reverse: 5’- GACACTCAATCAATAATAGATAACGAG -3’
A separate pair of primers was designed and used as competitor probe. The competitor primer sequences are the same as the pair shown above, only the nucleotide denoted in bold in forward primer was changed to T and the “G” denoted in reverse primer was changed to “A”.

4.4.7 Extension of promoter sequence of *Brassica napus* AGL15-2

Two *AGL15* genes are present in *Brassica napus*, *AGL15-1* and *AGL15-2* (Heck et al., 1995). However, the *AGL15-2* sequence available in the database was not long enough to compare similarity between Brassica *AGL15* and Arabidopsis *AGL15* within the 5′ regulatory region. To further extend the 5′ flanking sequence, thermal asymmetrical interlaced TAIL-PCR (Liu et al., 1995) was employed. PCR reactions were conducted on MJ research PTC 100, using Takara ExTaq enzyme (Panvera, Madison, WI). Two rounds of PCR amplifications were used to isolate DNA upstream of the known portion of the *AGL15-2* promoter. 15 pmol of the *AGL15-2* specific primer TR1 was used with 150 pmol of the partially degenerate primer AD-2 for the first PCR reaction. PCR was performed as follows: (1) 96°C for 5 min; (2) 94°C for 10 sec; (3) 65°C for 30 sec; (4) 72°C for 1 min; (5) repeat four additional cycles of steps 2 through 4; (6) 94°C for 10 sec; (7) 25°C for 3 min; (8) ramp to 72°C over 3 min; (9) 72°C for 3 min; (10) 94°C for 10 sec; (11) 65°C for 30 sec; (12) 72°C for 1 min; (13) repeat one more cycle of steps 10 through 13; (14) 94°C for 10 sec; (15) 44°C for 1 min; (16) 72°C for 1 min; (17) repeat 14 more cycles of steps 10 through 16; (18) 72°C for 3 min; and (19) 4°C.

The product from the first PCR was diluted 1:50, and 1 µl of the dilution was used for the second round of PCR. In the second PCR, 15 pmol of a second nested *AGL15* specific primer TR2 and 15 pmol of AD-2 were used. The PCR conditions were: (1) 96°C for 5 min; (2) 94°C for 10 sec; (3) 61°C for 30 sec; (4) 72°C for 1 min; (5) repeat one more cycle of steps 2 through 4; (6) 94°C for 10 sec; (7) 44°C for 1 min; (8) 72°C for 1 min; (9) repeat 17 additional cycles of steps 2 through 8; (10) 72°C for 4 min; and (11) 4°C. A second series of TAIL-PCR reactions were performed to further extend the sequence information. The products were cloned
into pGEM-T Easy® vector (Promega, Madison, WI) and sequenced. The primers used for *Brassica napus* AGL15-2 TAIL PCR are:

- AD2: 5’-NGTCGASWGANAWGAA -3’
- TR1: 5’- CGATCCTCTTTATCTCTATTTTCCCA -3’
- TR2: 5’- CCTCGATCCATCTTAATTCTTTC -3’
- TR1*: 5’- GAAACTTGAGTGTCTATCTCGAAGC -3’
- TR2*: 5’- AGCAATTCAACACTCATCCATTATA -3’.

Primers TR1 and TR2 are the primers used in the first series TAIL PCR and TR1* and TR2* are the primers used in the second series TAIL PCR.
CHAPTER 5  DISCUSSION AND PERSPECTIVES
The regulation of gene expression is highly complex and often involves networks consisting of many regulatory factors. This is true for both prokaryotes such as \textit{E. coli} (Babu and Teichmann, 2003) and eukaryotes such as \textit{S. cerevisiae} (Lee et al., 2002). A particular transcription factor usually regulates more than one target gene and the expression of a particular gene is usually regulated by more than one transcription factor. The transcription factors involved in a particular biological process usually do not act alone; instead, they interact with each other in a combinatorial fashion to control gene expression.

In plant developmental processes, such as embryogenesis, a large number of genes are expressed (Goldberg et al., 1994). \textit{AGL15} is one of the genes that is preferentially expressed during Arabidopsis embryo development but also expressed at lower level in other vegetative and reproductive tissues (Heck et al., 1995; Rounsley et al., 1995; Perry et al., 1996; Perry et al., 1999). Evidence from previous studies suggested that AGL15 may play a role during embryogenesis in promotion and maintenance of embryonic development (Heck et al., 1995; Perry et al., 1996; Harding et al., 2003). During the post-germinative growth stage, it has been suggested that AGL15 may function to regulate senescence and abscission of floral organs (Fernandez et al., 2000; Fang and Fernandez, 2002) at least when AGL15 is ectopically expressed. As a member of the MADS domain protein family, AGL15 may function as a transcriptional regulator and modulate gene expression. We are interested in understanding the roles that AGL15 plays in plant development. In an effort to elucidate the biological roles of AGL15, we performed studies to identify downstream target genes regulated by AGL15 and to characterize the regulation of the expression of \textit{AGL15} itself. This chapter summarizes results described in previous chapters and discusses some future directions.

The interactions of transcription factors with specific target DNA sequences are central to the operation of the gene regulatory networks. To identify AGL15 binding sites in the Arabidopsis genome, one approach we adapted is chromatin immunoprecipitation (ChIP) as described in Chapter 2. The Perry lab is one of
the first groups to use ChIP to isolate previously unsuspected in vivo targets of a transcription factor in plants. Furthermore, in an attempt to enrich potential true AGL15 binding sites, an in vitro immunoprecipitation (in vitro IP) following ChIP isolation was conducted. Using ChIP and in vitro IP, a collection of potential binding sites of AGL15 was obtained and corresponding target genes were identified. ABF3, a gene encoding a bZIP protein was isolated and the promoter region of ABF3 was confirmed to be associated with AGL15 in vivo. The expression of ABF3 was decreased in response to ectopic expression of AGL15.

Although the ChIP and in vitro IP protocols have been used successfully to identify AGL15 target genes, the process is time consuming and labor intensive. To identify potential AGL15 target genes, we need to isolate AGL15 bound DNA fragments by ChIP, sequence the isolated DNA fragments and search a database to identify potential target genes. Individual sites must be tested to confirm in vivo association with AGL15 and individual genes tested for response to AGL15 accumulation. A more efficient approach is to use a chromatin immunoprecipitation-microarray chip (ChIP-chip) approach. ChIP-chip is advantageous compared to conventional direct sequencing in that ChIP-chip is high throughput. Additionally, ChIP-chip allows identification of true sites from non-specific background. As noted by Buck and Lieb (2004), only ~17% of the sequenced clones are likely to be true targets assuming a 20-fold enrichment of targets and predicting that targets represent 1% of all genomic fragments. ChIP-chip has been used to study binding sites of many transcription factors in yeast and humans (reviewed in Wyrick and Young, 2002). However, genome-wide mapping of binding sites requires a microarray chip representing the whole genome, including intergenic regions. This resource is not yet commercially available for Arabidopsis. But the Perry lab has made significant progress in this direction with custom microarray chips. Continued work in this field would greatly accelerate elucidation of the AGL15 transcriptional regulatory network.

The direct sequencing ChIP or ChIP-chip can be used to identify DNA binding sites and potential target genes of AGL15; however, binding does not necessary
mean regulation of the nearby gene occurs (Boyd et al., 1998; Boyd and Farnham, 1999; Ren et al., 2000, Skinger and Gross, 2001; Soutoglou and Talianidis, 2002, Martone et al., 2003). As an alternative to individually testing the responsiveness of the expression of potential target genes by RNA gel blot, RT-PCR or other conventional yet low-throughput methods, high-throughput microarray approach can be used. As described in Chapter 3, microarray experiments were carried out to explore the genes responsive to AGL15 levels. However, most of the genes identified in the microarray encode proteins with unknown functions, which makes it difficult to understand the significance of AGL15 in regulation of those genes.

To understand the function of AGL15, it is also important to determine the roles that the products of the regulated genes play in plant development. ABF3 was shown to be involved in ABA response and drought tolerance (Choi et al., 2000; Kang et al., 2002). To explore the interaction between AGL15 and ABF3, genetic crosses were performed using 35S:AGL15 and 35S:ABF3 transgenic plants. Occasionally, the delay in floral organ senescence and abscission observed in 35S:AGL15 was “rescued” to near wild type in the presence of the 35S:ABF3 transgene. It would be interesting to conduct more complete experiments to assess under what conditions this may occur. It also would be interesting to know whether AGL15 is implicated in the ABA pathway to regulate embryo development because AGL15 is known to regulate expression of a GA 2-oxidase that is involved in GA metabolism and level of expression this gene impacts an somatic embryogenesis (Wang et al., 2004b). It is still an intriguing question how plant hormones and gene regulation are integrated to control plant development.

For a few genes that are differentially expressed between wild type and agl15, including At1g20700, At1g61566 and At2g46990, T-DNA knock out SALK lines were obtained and no obvious phenotype was observed. However, this does not mean that the genes or the regulation of their expression by AGL15 does not have biological roles. This simply may reflect the fact that we have not found the proper screening conditions yet or that other proteins of redundant functions are
present. The robustness and plasticity of the gene regulatory network may contribute the lack of phenotype when the activity of a particular gene is perturbed. Nevertheless, further phenotypical characterization should reveal the functions of these gene products in plant development. The characterization of a knock out allele of At5g23405 is such an example that lacks an obvious phenotype under normal growth conditions. However, in the liquid culture system used in the experiment as described in Chapter 3, the loss of expression of At5g23405 does have an effect on somatic embryo production from shoot apical meristem of cultured seedlings. The function of At5g23405 therefore may be characterized in more detail using this system. The next experiments would be to complement the knock out lines and assess the phenotype, to generate homozygous lines of the gain-of-function of the gene, and to further assess the phenotypes of the knock out line including generating double mutants with other members of the HMGB gene family, including HMGB6, the closest family member of At5g23405. In addition, because it is a putative chromatin associated protein, its subcellular localization should be investigated. The regulation of At5g23405 by AGL15 should be confirmed molecularly and phenotypically. Furthermore, to confirm it is a member of HMGB family, biochemical analysis of the protein also should be conducted. Many of the experiments mentioned are currently underway. Upon completion, biological roles of AGL15 and At5g23405 would become clearer.

For most of the genes encoding products of unknown function that are listed in Tables 3.1a and b, more detailed analysis using bioinformatics tools would be very helpful to explore their possible functions. The Arabidopsis microarray data performed under many experimental conditions or genotypes are publicly available through TAIR (http://www.arabidopsis.org) and the MPSS data are available through Arabidopsis MPSS database (http://mpss.udel.edu/at/?). This kind of in silico analysis may provide useful information about possible biological processes in which these genes are involved.
In our microarray experiment, two samples, Ws and agl15 were used and each sample included three replicates. The use of biological replicates and statistical analysis in the construction of the gene list is advantageous over using simple 2-fold change criteria owing to the statistical power of the strategy. To further improve the confidence in the differentially expressed genes in the list, a larger microarray experiment would contain three different samples, Ws, AGL15 loss-of-function (agl15) and AGL15 gain-of-function (35S:AGL15). Most interest would be in genes whose transcript is less abundant in agl15 compared to wild type and more abundant in 35S:AGL15 compared to wild type, because those genes might be positively regulated by AGL15. Conversely, genes with increased transcript abundance in agl15 but decreased transcript abundance in 35S:AGL15 would also be of interest, because these genes might be negatively regulated by AGL15. Additionally, it would be beneficial to have more replicates in order to more reliably recover genes whose expression is differentially expressed between the two samples, especially for those with a fold change less than two (Yue et al., 2001; Zik and Irish, 2002). Equally important, target genes should also be systematically and independently confirmed by other experimental methods such as quantitative Real-Time RT-PCR. However, both RT-PCR and microarray experiments can only measure the steady state level of the gene transcripts, which is affected by both transcription rate and turnover rate. As a complimentary approach to RT-PCR or other methods that assess steady-state transcripts, nuclear run-on experiment could be conducted to monitor the transcription rates of select target genes.

Furthermore, it would be ideal if the results from the microarray and ChIP-chip can be combined to identify direct targets of AGL15. The genes identified in both experiments would indicate what genes are bound by AGL15 and for which binding has biological consequence. In fact, the strategy has been used successfully to construct transcriptional regulatory networks in yeast (Lee et al., 2002). However, from the analysis of the initial ChIP population and microarray results, which were described in Chapter 2 and 3, no common genes have been found in both populations. Part of the reason may be that many of the targets we
isolated by ChIP may not be really bound by AGL15; instead they represent nonspecific background as discussed above. Secondly, for some of the genes that are bound by AGL15, such as ABF3 and GA2ox6, the absence of response to AGL15 in seeds in the microarray experiments may reflect lack of enough repeats to show significant difference or lack of differential expression because redundant functions are present. On the other hand, there are genes that showed response to AGL15 in microarray experiments but have not been recovered in the ChIP population. This may due to the fact that some genes recovered in the microarray experiments are indirect targets of AGL15 and they are not expected to be isolated in the ChIP population. In addition, we have very low coverage of the AGL15 binding sites in the whole genome using direct sequencing ChIP strategy. The ChIP-chip protocol currently under development in the lab will yield more reliable results and broader coverage of AGL15 binding sites and we may see common genes identified by both methods by then.

The identification of target genes of AGL15 is just part of the story. It is also important to understand how the regulator itself is regulated. Experiments in Chapter 4 were performed to address this question. Potential regions that are important for the expression of AGL15 were identified. One region may confer the response of expression of AGL15 to auxin induction and the region contains a sequence that is similar to a cis element involved in auxin response. Other sequences also have been found to be similar to other cis elements that involved in various transcriptional regulatory processes. However, it is not enough to predict possible cis elements by sequence analysis alone. Further experiments should include genetic analysis using those transcription factors known to be important for the embryogenesis or shoot apical meristem development such as LEC1 and WUS or yeast one-hybrid system.

Furthermore, we also showed evidence that AGL15 may regulate itself through direct binding to the binding sites in the promoter of its own gene (Zhu and Perry, 2005). However, it is still unknown how AGL15 differentiates among the various potential binding sites in the promoter of it own gene. AGL15 has been
demonstrated to have DNA binding activity (Perry et al., 1999; Wang et al., 2002; Tang and Perry., 2003; Wang et al., 2004; Zhu and Perry, 2005) and preferentially bind to a consensus sequence, C(A/T)$_8$G in vitro (Tang and Perry, 2003). However, how different MADS proteins specifically regulate particular developmental programs remain unknown. In C. elegans, in vitro binding affinity was found to be relevant for in vivo function (Gaudet and Mango, 2001). But other studies indicate that the MADS domain is not involved in function specificity in flower development (Krizek, 1996; reviewed in Riechmann and Meyerowitz, 1997).

As we mentioned previously, it is likely that AGL15 interacts with other transcription regulators to control gene expression. Identification of these interacting proteins will be important for understanding how AGL15 regulate expression of target genes, including modulation of specificity of binding.

Although many questions remain about AGL15’s role during plant development, tools and approaches described in this dissertation, as well as in use in the Perry lab, will allow a more complete picture of AGL15 gene regulatory networks.
REFERENCES


Raghavan, V. (2004). Role of 2,4-dichlorophenoxyacetic acid (2,4-D) in somatic embryogenesis on cultured zygotic embryos of Arabidopsis: cell expansion, cell cycling, and morphogenesis during continuous exposure of embryos to 2,4-D. Am. J. Bot. 91, 1743-1756.


Reidt, W., Wohlfarth, T., Ellerstrom, M., Czihal, a., Tewes, A., Ezcurra, I., Rask, L. and Baumlein, H. (2000). Gene regulation during late
embryogenesis: the RY motif of maturation-specific gene promoters is a
direct target of the FUS3 gene product. The Plant Journal 21, 401-408.

1291-1301.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A. and
Dynlacht, B.D. (2002). E2F integrates cell cycle progression with DNA

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I.,
Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L.,
Wilson, C.J., Bell, S.P. and Young, R.A. (2000). Genome-wide location
and function of DNA binding proteins. Science 290, 2306-2309.

Reyes, José C., Hennig, Lars and Gruissem, Wilhelm (2002) Chromatin-
Remodeling and Memory Factors. New Regulators of Plant Development
Plant Physiol. 130, 1090-1101

development. Biological Chemistry 370, 1079-1101.


is an unusually potent transcriptional activator. Nature 335, 563-564.

In vitro regeneration of Arabidopsis thaliana from cultured zygotic embryos

Monitoring of Gene Expression Patterns with a Complementary DNA

Scheres, B., Di Laurenzio, L., Willemsen, V., Hauser, M.T., Janmaat, K.,
Weisbeek, P. and Benfey, P.N. (1995). Mutations affecting the radial
organisation of the Arabidopsis root display specific defects throughout the


Wang, H., Caruso, L.V., Downie, A.B., and Perry, S.E. (2004b). The embryo MADS-domain protein AGAMOUS-Like 15 directly regulates expression of

**Walter, J., Dever, C. and Biggin, M.** (1994). Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in Drosophila embryos. Genes Dev. 8, 1678-1692.


VITA

NAME

Cong Zhu

DATE OF BIRTH

October 18, 1973

PLACE OF BIRTH

Binchuan county, Yunnan province, China

EDUCATION

Master of Science, 1998
Chinese Academy of Agricultural Sciences, Beijing, China

Bachelor of Science, 1995
Department of Biology, Nankai University, Tianjin, China

EMPLOYMENT

Research Assistant, 1999-2005
Department of Agronomy, University of Kentucky

Liu, Shen & Associates Intellectual Property Law Offices, Beijing, China

AWARDS AND HONORS

Plant Biology Travel Grant, American Society of Plant Biologists, 2004
University of Kentucky Graduate School Academic Year Fellowship, 2000
Nankai University Academic Excellency Fellowship, 1994
Nankai University Academic Excellency Fellowship, 1993

PUBLICATIONS
