Understanding the gene expression patterns and regulatory network of epidermal and sub-epidermal cell population enriched transcription factors in *Arabidopsis thaliana* shoot apical meristem

SHIVANI BHATIA

Thesis submitted for the partial fulfillment of the degree of

DOCTOR OF PHILOSOPHY

Department of Biological Sciences

Indian Institute of Science Education and Research
Mohali-140306

May, 2019

CERTIFICATE

The work presented in this thesis has been carried out by me under the supervision of Dr. Ram Kishor Yadav, at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or full for a degree, a diploma, or a fellowship to any other university or institute.

Wherever contributions of other people have been involved, every effort has been made to indicate this clearly, with due acknowledgment of collaborative research and discussions. This thesis is a bonafide record of original work done by me.

Date:

Place: Mohali

Shivani Bhatia

In my capacity as the supervisor of the candidate's thesis work, I certify that above statements made by the candidate are true to best of my knowledge.

Dr. Ram Kishor Yadav
(Supervisor)

Dedicated to my brother and my husband

Acknowledgements

First and foremost, I would like to thank the **Almighty** for all the opportunities so far in my life and giving me a chance to fulfill all my dreams and prove my worth. In no amount of words, I can thank God for giving me such a blessed life.

I would like to express my sincere gratitude towards my supervisor **Dr. Ram Kishor Yadav** for his timeless efforts and guidance throughout my stay in his lab. His mentorship and motivation will always be appreciated because without his constant push this thesis wouldn't have been possible. His ideas and suggestions have always helped me in building this project and taking it to new heights.

Besides my advisor, I am also grateful to my committee members, **Dr. Shravan Kumar Mishra** and **Dr. Kuljeet Sandhu** for their critical feedback and valuable suggestions towards my project. Dr. Shravan Kumar Mishra and his lab members have been a great help to me for my Y1H study. And Dr. Kuljeet Sandhu has always helped me with his insightful comments in the bio-informatics area of my research. My heartfelt thanks to both my committee members.

My special thanks to **Dr. Siobhan Brady**, University of California, Davis for providing me with the library of preys for carrying out my Y1H screen. Without her help, this project couldn't have reached the same height as it is today.

I would also like to thank all my fellow lab members **Prince Saini, Shalini Yadav, Sonal Yadav, Harish Pareek, Anshul Kaushik** and **Dr. Monika Mahajan** who have always been a support throughout the course of my Ph.D. All the stimulating discussions in the lab and lab meetings have always helped me in my research. I am also thankful to all the master students, **Asha, Mahima, Medha, Rimpy and Jayesh** who have worked with me with full dedication and contributed largely towards the completion of my project. Jayesh has contributed largely towards the bio-informatics analysis part and network analysis.

I would like to acknowledge **DBT IYBA** fellowship for financial support and **IISER, Mohali** for providing me fellowship throughout my course of Ph.D.

I owe my special thanks to my brother, **Dheeraj Bhatia** who has been the greatest support for me throughout my Ph.D., as well as my life. Without his constant efforts, blessings and encouragement, I wouldn't have reached where I am today. My Mom and Dad, my sisters and my sister-in-law have also been immensely supportive and always affectionate towards me. One of the biggest support for me has been my best friend, **Harjeet Kaur**, without whom I wouldn't have imagined completing my thesis. She always believed in me and guided me best during my rough days.

Finally, I would like to thank my beloved husband **Aakash Khatri** for all his love, support, encouragement and patience towards me. He truly has been a source of motivation for me ever since I met him. And my sincere thanks to my in-laws for their sacrifice and cooperation with me. It would have been impossible to accomplish this objective without their support. I feel really blessed to have them all in my life.

List of publications

1) Malhan, D., **Bhatia, S.,** & Yadav, R. K. (2015). Genome wide gene expression analyses of Arabidopsis shoot stem cell niche cell populations. *Plant Signaling & Behavior*, *10*(4), e1011937. https://doi.org/10.1080/15592324.2015.1011937

2) **Bhatia, S.,** Pareek, H., Sundaram, J., Saini, P., Mahajan, M., Yadav, S., Yadav, S., Sahu, S., and Yadav, RK. Deciphering the regulatory network of epidermal and sub-epidermal enriched transcription factors in the shoot apical meristem of *Arabidopsis* thaliana (Manuscript under preparation).

Synopsis

Introduction

In *Arabidopsis thaliana,* meristems are established during embryogenesis at the two opposite ends of the embryo, and are termed as the root apical meristem (RAM) and the shoot apical meristem (SAM). All above ground organs, such as leaves and flowers are formed on the flanks of SAM in the peripheral zone (PZ). SAM is dynamic in nature and produce organs infinitely due to the presence of pluripotent stem cells in the central zone (CZ). These cells receive signals from the niche cells situated in the rib meristem (RM). The RM is positioned just below the CZ in the SAM and maintains the stem cells. The daughters of stem cells can descend both in PZ and RM by cell division. Accordingly, SAM can be subdivided into three zones. Plant histologists indeed detected these zones based upon the cell division activity and cytoplasmic appearance of CZ, PZ and RM cell types. These studies also revealed that in addition to zonation, plant SAMs can be classified further into distinct cell layers (Leyser and Furner, 1992). In eudicots, SAM comprises of three cell layers, while in monocots, it is made up of only two cell layers. The epidermal / L1 and sub-epidermal / L2 cell layer together form tunica, and the L3 forms the corpus. Interestingly, cells in the tunica undergo anticlinal cell division pattern, whereas cell in the corpus divide both anticlinally and periclinally (Satina et al., 1940). The zonation and cell layering organization is always reiterated in newly formed flower and axillary meristems. In the past, most of the genetic studies were conducted to understand stem cell specification and organ formation in SAM. None of the past studies were directed systematically to understand the regulatory mechanisms that controls this well-orchestrated architect of SAM.

Since plants make organs post embryonically, many events related to SAM zonation and layering specification can be studied using modern tools. A large number of recent studies have focused on understanding the cell and tissue specification using transcriptomics to understand the organogenesis in apical meristems (Birnbaum et al., 2003; Brady et al., 2007; Jiao et al., 2009). To understand the large scale datasets, network studies are being devised to unravel the physical interactions between the transcription factor proteins and their target genes (Brady et al., 2011a; Jones et al., 2014; Mukhtar et al., 2011). Network based studies have greatly improved our understanding of cell fate specification in stele (Brady et al., 2011b). In plants, xylem cell

specification occurs due to secondary cell wall biosynthesis. A recent study has shown the role of salt in inducing the cell wall biosynthesis (Taylor-Teeples et al., 2014). In shoot, organ boundaries are critical for the formation of new organs and axillary meristem initiation. Gene centered network study revealed the regulatory hierarchy among the transcription factors (TFs) regulating axillary meristem initiation in Arabidopsis SAM (Tian et al., 2014). However, there is no information on transcriptional gene regulatory network operating in the cell layers of SAM. In this study, I identified the TFs enriched in different cell layers of the shoot and made a network for epidermal and sub-epidermal cell type enriched TFs using systems biology approach.

Identification of cell type specific TFs

To identify the important factors involved in the specification of epidermal and sub epidermal cell layers, gene expression data generated from FACS sorted cells for epidermal (*pHMG:H2B-YFP*), sub-epidermal (*pHDG4:H2B-YFP*) and corpus (*pWUS:H2B-YFP*) cell populations were analyzed. A total of 1456 genes were found to be enriched in three cell layers of the shoot after applying ≥ 1.5-fold cut off ($P < 0.05$). Of the 1456 genes, 535 were enriched in the L1 layer, 256 in the L2 layer and 665 in the corpus (Yadav et al., 2014). Of the 535 L1 layer enriched transcripts, 44 encode TFs. Two hundred fifty six transcripts were found to be enriched in L2 cell layer and among them 21 encode for TFs (Yadav et al., 2014). Taken together, 65 TFs are enriched in the L1 and the L2 cell layer of shoot.

Studying the spatio-temporal expression pattern of epidermal and sub-epidermal TFs

TF driven regulatory networks control many cellular processes in multicellular organisms. To build such networks, it is necessary to understand their spatio-temporal expression patterns. In this study, I created transcriptional fusion constructs for 43 TFs by taking a 3 kb promoter fragment upstream of start codon of TF gene. I raised plants after transforming them with expression constructs using the floral dip method. When the transgenic lines were screened for H2B-YFP protein in confocal microscope, I found expression for ~58% of them (25/43). However, two promoter constructs were mis-expressed as they had sporadic expression within the SAM and it did not corroborate the cell type specific microarray data. The expression pattern for these TFs was also checked in the early embryos and 3 days old seedlings. As expected, epidermal cell layer identity genes start expressing very early during development. In the globular stage of the embryo,

genes such as *ATML1, HDG12, AT1G75710* show expression throughout the embryo. Conversely, *NF-YA5* expression is limited to the epidermis in the globular stage embryo, suggesting that *AtML1* has function in other cell types of the embryo, apart from L1 cells. Nevertheless, some of the TFs continue to express in the inner cells, and perhaps due to repression by upstream factors, the expression of these genes get restricted to epidermal cell layer later on. Yet, no expression was observed for any of the L2 layer specific marker gene such as *HDG4* and *HDG7*. Their expression commences in few cells in the 3 days old seedlings. These observations suggest that sub epidermal cell layer / L2 doesn't get specified during embryogenesis and it is possibly specified post embryonically.

By analyzing the data from 25 reporter constructs, I concluded nine major patterns of expression for these TFs in the shoot. Some of the TFs were expressed uniformly in the epidermal and sub-epidermal cell layer of shoot. *HDG4* and *HDG7* are expressed in the sub-epidermal layer of the shoot, but completely missing from the embryo. *WRKY11, WRKY22* and *AT2G31730* were expressed in the PZ, however, were absent from the CZ of the shoot. *ABF1*, a bZIP TF, was present in epidermal as well as in deeper layers of the shoot but was absent in the deeper layer of flower primordia. *EGL3* showed expression in the epidermal cell layer of sepal primordia in stage 3 flower. *APETALA1* (*AP1*) showed expression only in the floral primordia and completely missing from the shoot. The expression pattern of *AP1* is correlated with its function in floral meristem identity specification. From this study, I concluded the following things: 1) Patterns of gene expression are established right from embryonic stages of development and continue to show similar patterns of expression up to the adult shoot. 2) The upstream noncoding sequence was enough to capture the mRNA expression pattern in ~58% of the cases as reported by microarray studies or in situ hybridization. 3) Nine major patterns of gene expression were observed. It is possible that these patterns are established by distinct cis-regulatory modules, which are conserved and drive expression of the majority of the genes that are expressed in shoot. However, post-transcriptional regulation can alter their effects and may require investigation to fully comprehend the complex nature of gene regulation in multicellular organisms.

Y1H assay

Next, I carried out the Y1H screen using the gene promoters to understand the nature of cis-regulatory module and are they really conserved for a set of TFs that display similar spatial and temporal expression pattern. The 3kb promoter fragment of epidermal and sub-epidermal enriched TFs was used to identify potential prey TFs. A total of 49 promoters were successfully screened against a library of 327 TF prey proteins. TFs that show cell type specific expression as well as widely expressed in the shoot were chosen as preys to map the protein-DNA interactions. From the Y1H screen, a total of 165 interactions were concluded among 37 promoter fragments and 53 protein preys. Among the total baits screened, one interacting partner was found for 76% (37 out of 49) of the genomic regulatory elements and from the total prey library, 16% of TFs bound to at least one genomic regulatory element. The network consists of 80 nodes, with 10 nodes acting as both baits and preys, 27 nodes that are unique DNA elements and 43 nodes that act as unique preys. Nodes are connected through edges that represents a physical interaction between the two nodes.

Studying knock-out mutants of shoot enriched TFs

Next important question to address was the regulatory relationship between these interacting TFs and characterization of the loss-of-function mutants of these TFs. Loss of function mutants were identified for 45 out of 65 TFs enriched in the epidermal and sub-epidermal cell types of the shoot. Of the 38 alleles validated, 18 of them showed no transcripts in the mutant background, demonstrating them to be true null allele of the TF. Except *at1g75710* gene locus encoding a C2H2 zinc finger TF, none of the TF mutants revealed phenotypic changes. Conceivably it could be due to the genetic redundancy, which can be resolved by making higher order mutants for the closely related TFs.

***In-vivo* validation of the network**

To validate the regulatory relationship between the upstream regulator and downstream target, wherever possible, null alleles for upstream TFs were identified. In parallel, for a few TFs, over expression lines were either obtained from the stock center or created in the lab. These lines were characterized, and then were used for validating the regulatory relationship among the interacting partners. Of the 38 upstream regulators tested in this study, molecular phenotypes were observed

for DEWAX 55% (5/9), ARF9 67% (2/3), WKRY54 67% (2/3), ANAC082 100% (2/2), AtHB34 67% (2/3), AZF2 50% (1/2), ARF12 25% (1/4), HMG 45% (5/11) and GRF3 100% (1/1) target genes. Taken together, this study revealed 5 potential activators and 16 potential repressors.

GRF3 binds to the promoter of *HDG12*

GRF3 was found to bind on *HDG12* 3kb promoter in the eY1H screen. Based on the DAP-seq information about GRF binding sites, 5 binding sites were predicted for GRF on the *HDG12* promoter. To know, through which binding site GRF3 binds to *HDG12,* the promoter fragment was chopped into smaller fragments of 500bps and tested again in Y1H. I found out that GRF3 binds to the *HDG12* promoter on the region 500bp upstream of ATG. For *in-vivo* validation of the interaction, qRT-PCR experiment was done in the *grf123* mutant and *35S:GRF3* over-expression background and *HDG12* was found to be positively regulated by GRF3. Also, *hdg12* mutant has smaller leaf size in comparison to WT because of the small pavement cell size. Similarly, *grf3* single and higher order *grf123* mutant plants show smaller leaf size due to decrease in cell proliferation and cell expansion. Over expressing *HDG12* in *hdg12* mutant background rescued the smaller leaf phenotype by increasing the cell size, however without affecting the cell division. GRF3 expresses broadly in the shoot, in L1, L2, L3, however, HDG12 expression in the shoot is limited only to the epidermal cell layer. This clearly suggest that GRFs exert their influence on *HDG12* in the epidermal cell layer. Future studies will unravel the regulatory mechanism.

Summary

In summary, I have obtained spatio-temporal expression patterns for the epidermal and sub-epidermal enriched TFs. The 5' upstream non-coding sequences were found to control the transcription in 58% of the TFs studied. For the remaining TFs, expression could not be captured probably due to lack of sufficient regulatory elements within the upstream 3kb region. I concluded 165 interactions from the Y1H assay for 37 promoters. Thirty-eight interactions were validated in planta, and for 21 of them, I found a molecular phenotype. *AtGRF* mediated regulatory node influences cell size in the epidermal cell type. GRFs directly bind to the *HDG12* regulatory elements and drives its expression. GRFs are also enriched in the CZ cells, and overlap with *HDG12* expression. Future studies will focus on unraveling the transcriptional regulation of *HDG12* by GRFs and impact on shoot development.

List of tables

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of abbreviations

%	Percent
mg	Miligram
M	Molar
ml	Mililitre
mM	Millimolar
μM	Micromolar
min	Minute
GRN	Gene regulatory network
PCR	Polymerase chain reaction
SAM	Shoot apical meristem
bHLH	basic helix loop helix
TF	Transcription factor
WT	Wild type
Y1H	Yeast-one-hybrid

Table of contents

CHAPTER 1

INTRODUCTION

1.1 *Arabidopsis thaliana:* The model organism

Arabidopsis thaliana, thale cress, is a member of the Brassicaceae family. Arabidopsis is found growing naturally in continents such as north America, Europe and Asia. Naturally, it is a winter annual plant whose seeds germinate in autumn, reaches to rosette stage in the winters and flower in spring. Agriculturally important plants such as wheat, barley, tomato were used historically for research purposes. But the process of understanding the fundamental plant processes was rather slow with these model organisms (Meinke et al., 1998). In 1943, a professor at the University of Frankfurt, described several advantages of using *Arabidopsis* as an effective research model. Several advantages that make *Arabidopsis* a suitable model system, are: (A) Short generation time of about six weeks, when grown under optimal conditions. (B) Self-pollinating plant, which can produce a lot of siliques. (C) Small in size, thereby allowing the growth of plants in small growth area. (D) Easy to generate stable transgenics using *Agrobacterium* mediated transformations. (E) Smallest genome size among known angiosperm species. This makes it suitable to be used as a model plant. Over the past few years, thousands of mutant transgenic lines have been generated for Arabidopsis by the collective efforts of a few laboratories (Alonso et al., 2003; Rosso et al., 2003). These mutant lines carry transfer-DNA (T-DNA), which has got incorporated randomly within the genome and thus wherever it landed into a functional gene it disrupted its function. Many interesting mutant phenotypes related to various aspects of plant growth and development has been identified using this resource. Mutant phenotype based studies have helped in the characterization of many genes in *Arabidopsis* and their seed stocks are available from the stock center. Knowledge gained from *Arabidopsis* has also been used to facilitate understanding in various other species and to engineer crop plants for improving cold, salt and drought tolerance in them and increasing their yields. For example, *AtNHX1* gene from *Arabidopsis,* when over-expressed, resulted in salt tolerant *Brassica napus* and tomato (Zhang and Blumwald, 2001; Zhang et al., 2001). Over-expressing *Arabidopsis* CBF genes in canola, resulted in increased freezing tolerance (Jaglo et al., 2001). Similarly, over-expressing *ZmCBF* gene in maize, also resulted in increased freezing tolerance in maize (Chaiappetta 2002). Therefore, *Arabidopsis* plays a critical role in understanding biological mechanisms related to growth and development as well as stress tolerance in plants.

1.2 Meristems

The most striking feature of plants, as opposed to animals, is the post embryonic growth and development. Development of the seed plants depends on the activity of the meristematic /

undifferentiated cells present in the actively growing tissues in the plant body. Two primary meristems are established at opposite poles during embryogenesis in Arabidopsis thaliana (Barton and Poethig, 1993; Dolan et al., 1993). Because of their position at the tip of the plant, these meristems are termed as apical meristems. From the root apical meristem (RAM), arises the whole root system and from the shoot apical meristem (SAM) arises the above ground organs of the plant, such as leaves, flower and stem (Dolan et al., 1993; Leyser and Furner, 1992). Stem cells are present within these meristems that can divide infinitely throughout the life of the plant. The stem cell niche maintains the fate of the stem cells as long as the plant is alive, and at the same time integrates internal and external signals to regulate the rate of organogenesis (Schofield, 1978). Stem cell niches, to maintain the stem cell fate both in plants and animals, secrete signal molecules and transcription factors. These factors have been characterized extensively to understand the mechanism of stem cell self-renewal and differentiation.

1.3 Organization of the shoot apical meristem and stem cell maintenance

In Arabidopsis the SAM is present between the two cotyledons of the seedlings. All the above ground organs in plants arise from the activity of SAM. In the vegetative phase SAMs are flat but upon transition to reproductive phase they appear dome shaped structure. The structure of the SAM can be divided into various layers and zones (Leyser and Furner, 1992). The first two layers of the SAM are termed epidermal / L1 cell layer and the sub-epidermal / L2 cell layer in eudicots. Both epidermal and sub-epidermal cell layers display anticlinal cell division pattern and termed as tunica (Satina et al., 1940) (Schmidt 1925). The inner corpus / L3 cell layer undergoes both anticlinal and periclinal divisions. The stem cells are located in the central zone present at the tip of the shoot. Stem cells give rise to progenitors that amplify further into daughters and eventually enter in the peripheral zone where they get incorporated into organ primordia. Below the central zone, is located the organizing centre (OC) / niche. Cells of the organizing centre expresses a homeodomain transcription factor, WUSCHEL (WUS), which moves to the central zone cells and activates the expression of *CLAVATA3* (*CLV3)* in the stem cells. *CLV3* is a 13 amino acid long signal peptide that is processed from a long preproprotein of 96 amino acids (Sharma et al., 2003). The mature CLV3 peptide is arabinosylated post translationally (Ohyama et al., 2009). *CLV3* binds to its receptors CLAVATA1 (CLV1), CLAVATA2-CORYNE (CLV2-CRN) and in a non-cell autonomous manner, represses the transcription of *WUS,* thereby controlling stem cell proliferation in SAM (Brand et al., 2000; Muller et al., 2006; Reddy and Meyerowitz, 2005). Thus, there is a

negative feedback loop between *WUS* and *CLV3*, which helps in maintaining the stem cell number and regulates their proliferation in the shoot (Daum et al., 2014; Mayer et al., 1998).

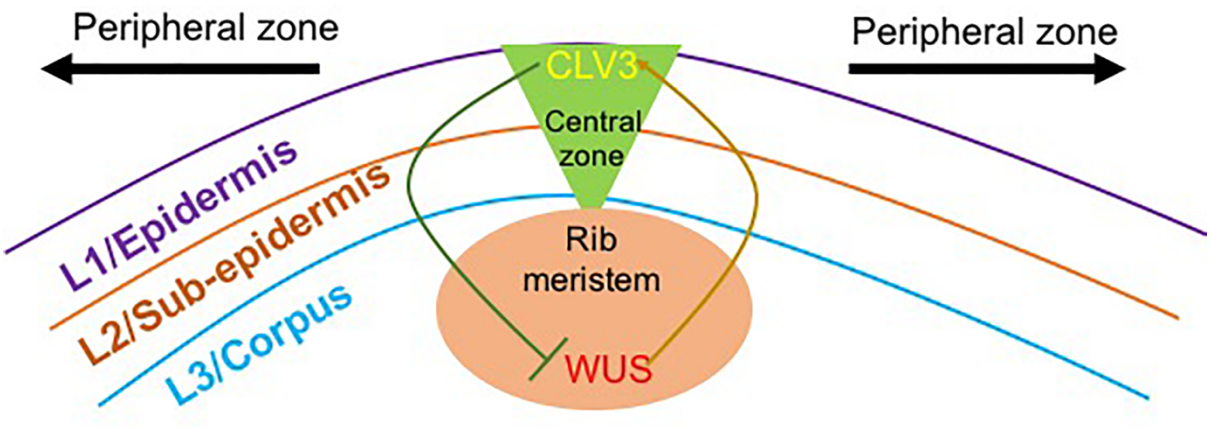


Figure 1.1: Representation of Shoot apical meristem. Schematic representation of the shoot apical meristem (longitudinal section) of *Arabidopsis thaliana.* Stem cells are located in the central zone. Rib meristem region is present below the central zone and helps in maintenance of the stem cells via WUS-CLV3 feedback loop. Cells from the central zone differentiate and enter into the peripheral zone, from where the organ primordia arise.

1.4 Gene expression patterns of transcription factors

Understanding the spatio-temporal expression patterns of TFs is important for having a deeper understanding of their gene regulatory modules. Earlier, gene trapping systems using *GUS* and *luciferase* reporter genes were used to probe the gene activity. For a large number of genes from TFs, protein kinases to metabolic enzymes, the gene activities have been determined using the gene trap system (Koo et al., 2007). Different reporter genes such as *GFP*, *GUS* and *luciferase* have been used in gene trap systems to visualize gene expression at tissue level. GFP based reporters are used mainly to study sub-cellular localizations (Haseloff and Amos, 1995; Moriguchi et al., 2005). GUS is used as a reporter gene because of its high stability and fine resolution in histochemical staining (Jefferson et al., 1987; Lindsey et al., 1993). However, GUS cannot be used for studying conditionally or temporally regulated expression patterns because of its low turnover rate. Recently, *Luciferase* based reporter system has been developed that can screen real time gene expressions with high sensitivity and speed (Millar, 1992; Mocharla et al., 1987; Thompson et al., 1991). However, it cannot

be used for monitoring cell or tissue specific gene expression. For studying gene expressions in adult plants and in spatio-temporal manner, GFP based reporters stand out the best.

For a lot of genes, the sequences present upstream of the gene are responsible for modulating the expression patterns, however there are evidences that suggest the role of other sequences within the transcribed regions in controlling the expression patterns of genes (Hong, 2003; Ito et al., 2003). For example, correct expression of a floral homeotic gene *AGAMOUS* requires the 3kb intron present at second position within the gene. But, there are studies which indicate that for a large number of genes, upstream sequences are sufficient to drive their endogenous expression (Lee et al., 2006). In a study by Lee et al. (2006), GFP transcriptional and translational fusions were constructed for 61 TFs enriched in root cell types. A 3kb promoter sequence upstream of the translational start site was used to drive the endogenous expression. Out of the 61 TFs screened, for 80% of them, the upstream 3 kb promoter fragment was able to recapitulate the native mRNA expression as predicted by the micro-array study (Lee et al., 2006), suggesting that the regulatory elements required for endogenous expression reside within the upstream 3kb elements. Also, there are reports that suggest, that 85% of the binding sites, for known TFs in humans, fall within the upstream 3kb region (Gifford et al., 2005). Traditional techniques such as RNA *in situ* hybridization can reveal expression patterns of a gene, however, it will not reveal information about the regulatory elements driving the endogenous expression.

In this study, I have used 3kb promoter fragments to construct transcriptional fusions for 44 shoot enriched TFs. Chapter 2 of the thesis discusses in detail the expression pattern associated with each TF and the expression patterns observed for various shoot enriched TFs. Out of the 44 reporter constructs made, expression for 50% of them was captured in shoot using the YFP reporter. Two of them were found to be mis-expressed and were unable to recapitulate the expression as expected from the microarray data for shoot cell types (Yadav et al., 2014). For the remaining TF promoters, YFP expression was not captured in shoot.

1.5 Large scale studies of transcription factors and gene expression

Growth of the plants is influenced by a large number of environmental factors such as duration of light and dark during the day, availability of water, pest infections and so on. Plant growth in all, relies on integrating these signals as a whole and leading to controlled gene expression, leading to various developmental outcomes. Lately, a large number of

studies have focused on techniques such as high throughput sequencing of whole transcriptomes and RNA-Seq, to understand transcriptional signatures in a spatio-temporal manner and under certain set of environmental conditions (Birnbaum, 2003; Brady et al., 2007; Jiao et al., 2009; Kilian et al., 2007; Schmid et al., 2005). Other network based approaches have made efforts to understand the genetic and physical interactions between the genes, to make sense out of the huge data generated with the help of high throughput sequencing and RNA-Seq like techniques (Arabidopsis Interactome Mapping Consortium, 2011; Brady et al., 2011a; Jones et al., 2014; Mukhtar et al., 2011).

1.6 Gene regulatory networks

In gene networks, genes are represented by nodes, which are connected via edges. The edge signifies interactions between the two nodes or genes. Interaction can be of either protein-protein or protein DNA. Past studies have shown that network based approaches are a powerful tool to decipher the regulatory module acting within the plant, and can be useful in establishing better understanding of plant as a whole system. A large number of network based studies have focused on building transcriptional regulatory networks, as transcription factors play an important role in controlling gene expression (Albert et al., 2014; Brady et al., 2011a; Busch et al., 2010; Chang et al., 2013; Franco-Zorrilla and Solano, 2014; Franco-Zorrilla et al., 2014; Helfer et al., 2011; Immink et al., 2012; Kaufmann et al., 2010).

For mapping transcription factor binding to the promoter of genes, various techniques such as mapping of DNaseI hypersensitive sites, Chromatin immuno-precipitation (ChIP) coupled with sequencing (ChIP-Seq) and Y1H can be used. But these methods have a limitation of highlighting only physical interaction between a TF and its gene and not gene regulation. Therefore, experimental evidences are further required to place these interacting partners in context of regulatory networks.

1.7 Validation of the predicted interactions *In-Planta*

The interactions predicted through high throughput screenings need to be validated *in vivo* to have a better understanding of their biological significance and to avoid the false positives usually generated in large data sets. The Y1H data signifies the binding of an upstream regulator on the promoter of a downstream target gene. Genetic analysis is one of the methods that can be used for validating these interactions in planta. Both morphological and molecular phenotypes need to be considered while validating. A morphological phenotype

relates with defects in plant growth caused due to changes in the level of upstream transcriptional regulator that also altered the expression of downstream interacting partner. However, molecular phenotype relates only with changes in the level of transcript caused by binding of TF of interest on the promoter of downstream target gene and does not result in a phenotype. Molecular phenotypes can be studied using real time quantitative reverse transcription PCR (qRT-PCR) experiments. And these real time qRT-PCR experiments can be carried out in either knock-outs or over-expressing lines or transient inducible lines for the upstream regulators of interest.

SHORTROOT (*SHR*) expresses in the immature vasculature cells of Arabidopsis root and is an ideal example of a TF showing both morphological and molecular phenotype. *SHR* is involved in the endodermis development, xylem formation and maintaining the quiescent center fate (Benfey et al., 1993; Carlsbecker et al., 2010a; Cui et al., 2007; Helariutta et al., 2000; Levesque et al., 2006; Mähönen et al., 2000; Nakajima et al., 2001; Sozzani et al., 2010). *shr* mutant shows a phenotypic defect of missing endodermis and *SHR* is considered to be an important TF because its loss of function causes defect in root development (Benfey et al., 1993). SHR is also known to bind to the promoter of *SCARECROW (SCR)*, a downstream target gene, which expresses only in endodermis. Analysis of *SCR* transcript in *shr* mutant revealed a significant downregulation of *SCR* levels in the mutant background, thereby suggesting a positive regulation of *SCR* by SHR (Helariutta et al., 2000). However, this regulation doesn't provide evidence of direct binding of SHR on *SCR* promoter and its regulation. Therefore, a ChIP-PCR with antibody against the SHR protein was done and direct binding of SHR on *SCR* was demonstrated (Cui et al., 2007). Above described methods, such as ChIP-qPCR and real time qRT-PCR can be potentially used to validate the Y1H predicted interactions. However, a single TF may or may not give both morphological and molecular phenotypes because of genetic redundancy in plants. In order to get both morphological and molecular phenotypes, one needs to screen higher order mutants.

Brady et al., conducted a Y1H study to map the network of TFs involved in stele development in root (Brady et al., 2011a). Mutants of 30 TFs were analyzed in this study for morphological defects. However, only 5 out of 30 (17%) of the mutants displayed morphological phenotypes, possibly due to redundancy. At the molecular level, altered expression of the target genes was detected for 54% of the interactions tested in Y1H. To understand the role of spatially restricted TFs in development of ground tissue, Sparks et al.

(2016) used Y1H based approach to find out the upstream regulators that bind to the *SHR* and *SCR* promoters. Mutants of both *SCR* and *SHR* show defects in radial patterning in root and shorter roots. Based on the gene regulatory network, it was predicted that upstream regulators of *SHR* and *SCR* might show similar morphological defects. Mutants of 29 different TFs were examined, however, they showed radial patterning defects with low penetrance. And none of the mutants showed a single ground layer, a feature of *shr* and *scr* mutants. Following this, real time qRT-PCR was carried out to discover molecular phenotypes associated with the network. 78% of *SHR* upstream regulators and 45% of the *SCR* upstream regulators showed molecular phenotypes, thus providing directionality to the network and providing better understanding of gene expression associated with development of ground tissue (Sparks et al., 2016).

1.8 Gene regulatory network studies in *Arabidopsis*

In *Arabidopsis*, Y1H has been exploited by researchers to build regulatory networks of TFs in various parts of the plant. Gene regulatory networks generated by using above mentioned techniques can serve as powerful tools for making testable hypotheses. Gene functions for genes under study can be looked into literature, however, network studies in some instances provide the missing links. For example, mutants of *VND6/VND7* and HD-ZIP III family have been reported in the past to show xylem related phenotypes. Using Y1H network approach, a link was established between *VND7* and *REVOLUTA (REV),* a HD-ZIP III gene. And the role of *VND7* and *REV* was established in secondary cell wall biosynthesis (Taylor-Teeples et al., 2015).

1.8.1 Gene regulatory network for *Arabidopsis* Root

1.8.1.1 Gene regulatory network for *Arabidopsis* secondary cell wall biosynthesis

Secondary cell wall is made up of lignin, cellulose and hemi-cellulose and is found in fibers, xylem and anther cells. Secondary cell wall provides structural support to the xylem cells and makes them water-proof for transporting water throughout the plant body. A role of NAC domain TFs and HD-ZIP III TFs was implicated in secondary cell wall biosynthesis in *Arabidopsis thaliana. VASCULAR-RELATED NAC DOMAIN 6 (VND6)* AND *VASCULAR-RELATED NAC DOMAIN 7 (VND7)* were considered important for xylem vessel formation. Also, HD-ZIP III TF, *PHABULOSA (PHB)* was implicated to regulate vessel formation (Carlsbecker et al., 2010b). However, there was no comprehensive knowledge of gene

networks that are involved in secondary cell wall biosynthesis. Taylor-Teeples et al. (2015) did a screening of promoters of genes involved in cellulose, hemicellulose and lignin biosynthesis, and are expressed in xylem cells. A highly connected network for xylem cell type enrich TFs was made. Interestingly, the two previously reported pathways for xylem specification, one via *VND6/VND7* and the other via HD-ZIP III TFs, were linked by this study. VND7 was determined as an upstream regulator of HD-ZIP III gene *REVOLUTA (REV)*. *REV* is known to be involved in xylem formation by regulating cell patterning within vasculature. In this study, *REV* was shown to be repressed by VND7 as xylem differentiation gets initiated. Also, it was ascertained from this study, that REV binds to the promoter of downstream genes involved in lignin biosynthesis and causes their repression. Therefore, when VND7 gets activated, it causes repression of *REV*, thereby allowing the synthesis of lignin post cell patterning (Taylor-Teeples et al., 2015).

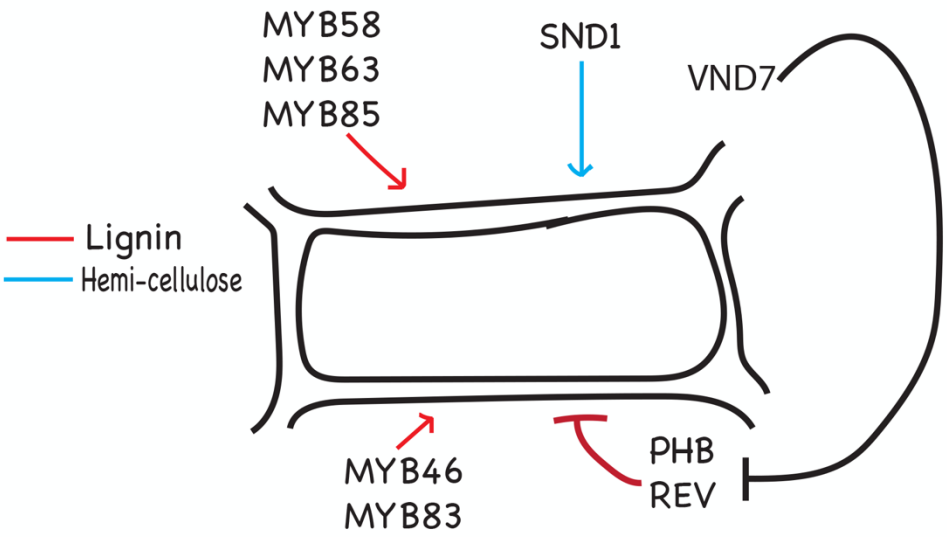


Figure 1.2: Gene regulatory network for secondary cell wall biosynthesis (adapted from Taylor-Teeples et al., 2015).

1.8.1.2 Gene regulatory network for root stele enriched TFs

Stele forms the central part of the root and includes phloem, xylem, procambium and pericycle cell types. Brady et al. (2011b) conducted a study that was focused on constructing the gene regulatory network of TFs enriched in either one or more cell types of the stele tissue in root and also miRNAs that target these TFs (Brady et al., 2011b). This was an attempt to understand the interactions that are relevant spatially. A total of 167 TF proteins were used as preys and 65 promoters as DNA baits. In total, 46 protein-DNA interactions

were concluded between 16 promoters and 21 TFs. Protein-protein interactions were also determined between these 167 TFs enriched in stele, and a total of 25 interactions were concluded between 26 TFs. By using ChIP-qPCR, authors confirmed the binding of VND7 on the promoter of *AT3G43430* and OBP2 binding on promoters of *PHB* and *PHV*. *In-vivo* validation of 59% of the total interactions was confirmed either by real time qRT-PCR or ChIP-qPCR experiments.

1.8.2 Regulatory network for shoot organ boundary enriched TFs

1.8.2.1 Gene regulatory network of organ boundary genes, to understand hierarchy of transcription factors controlling axillary meristem initiation

Shoot apical meristem of the plant contains self-renewing stem cells present at the tip of the shoot. Leaves and flowers are produced from the peripheral zone of the shoot, where differentiated cells reside. Organ primordia get separated from the shoot by a boundary domain, in which the growth of the cells gets arrested (Aida and Tasaka, 2006; Rast and Simon, 2008; Shuai et al., 2002). These organ boundaries harbor the axillary meristems, which have the developmental potential same as that of the SAM. In order to understand the process of axillary meristem initiation, Tian et al. (2014) made a gene regulatory network of genes enriched in these boundary domains, using the Y1H assay (Tian et al., 2014). Promoters of genes that regulate boundary formation and axillary meristems, such as *LAS* (Greb et al., 2003), *CUC2* (Hibara et al., 2006; Raman et al., 2008), *SHOOT MERISTEMLESS (STM)* (Grbić and Bleecker, 2000; Long and Barton, 2000) were chosen as baits in the Y1H assay. TFs enriched in organ boundaries in shoot and TFs with low expression in the boundary domain were chosen as preys. A total of 180 interactions were concluded between 103 TFs and 23 promoter elements. Through this gene regulatory network, authors were able to link most of the previously known key regulators of axillary meristem initiation. Important interactions were captured from Y1H and also validated by real time qRT-PCR experiments. CUC2 was revealed to activate the expression of *LAS* and in turn *CUC2* was found to be getting activated by RAX1 and DRN (Figure 1.3). These regulatory networks were also supported by previous genetic experiments (Hibara et al., 2006; Raman et al., 2008). Another important TF, DORNROSCHEN (DRN) bound to *CUC2* in Y1H and caused its activation. This was also supported by the *drn-1* mutant phenotype in which axillary meristem initiation was compromised, a phenotype similar to *cuc2* mutants. SPL9 and SPL15 interactions with *LAS* promoter were also captured in Y1H. SPL9 and

SPL15 caused suppression of *LAS* and was also evident from the *spl9-4 spl15-1* mutant which formed more axillary meristems in the axils of cauline leaves. This suggests that SPL suppression of *LAS* expression is responsible for controlling axillary meristem initiations in the axils of cauline leaves. Therefore, by using Y1H technique a huge gene regulatory network could be made and a significant number of interactions taking place in-vivo could be captured in a heterologous system, making it a suitable tool for generating large amounts of relevant data.

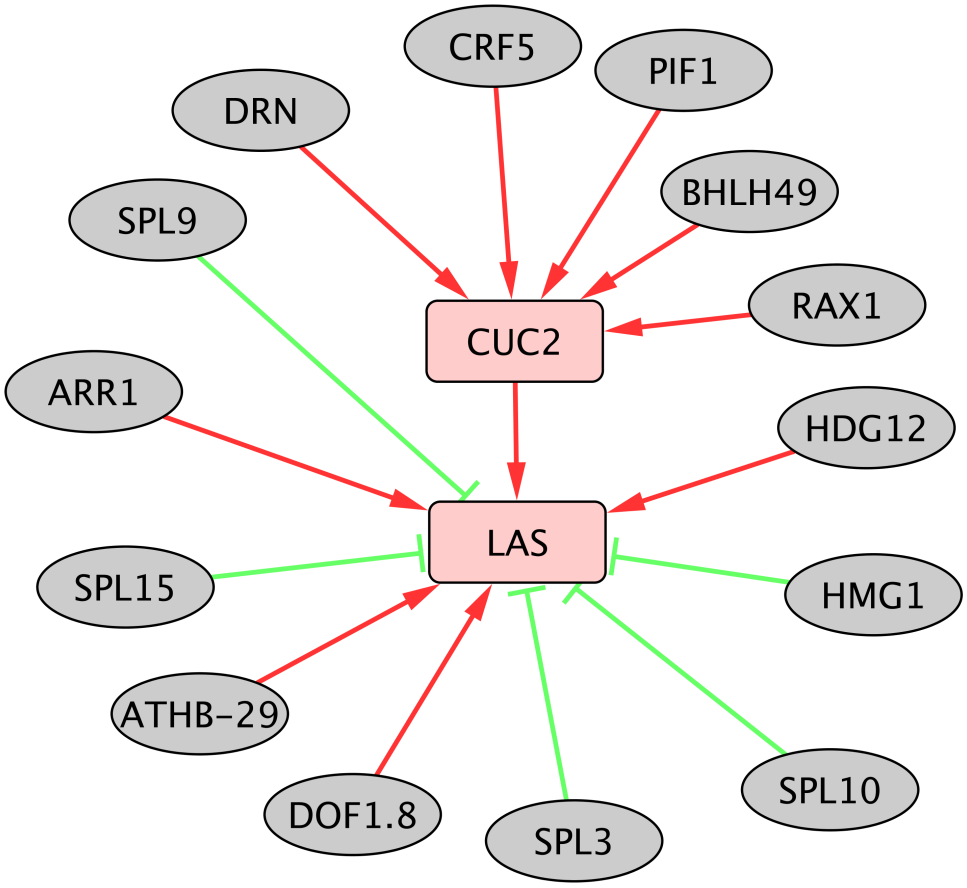


Figure 1.3: Gene regulatory network for organ boundary enriched TFs (adapted from Tian et al., 2014).

In chapter 3 of this study, I presented the findings of regulatory network constructed using the epidermal and sub-epidermal enriched TFs. Y1H assay revealed the regulatory networks operating within the epidermal and sub-epidermal cell types of the shoot. Promoters of TF genes were screened against TF protein prey library. The prey library had both narrowly as well as broadly expressed TF proteins. Since, the molecular phenotypes are often associated with the interactions predicted in the network, regulation of the downstream target genes by the upstream regulators predicted in the gene regulatory network established by real time

qRT-PCR analysis in the mutant and over-expression lines of the upstream transcription factors. Phenotypes associated with the TFs are also important indicator of their biological function. In chapter 4 of this study, I focused on studying the knockout mutant of epidermal and sub-epidermal enrich TFs. T-DNA insertions were determined for 43 TFs and their transcript levels were determined in mutant and WT background by semi-quantitative RT-PCR to ascertain true null allele. In chapter 5 of this thesis, I explored the role of GRFs in regulating *HDG12*, a downstream target gene, which I identified in Y1H screen. I found that GRF2 and GRF3 directly bind to *HDG12* promoter. GRFs are known for their role in regulating cell proliferation and thus affect leaf size and root meristem in Arabidopsis, however, their role in cell growth and expansion is not yet clear. We investigated the loss of function *hdg12* mutant phenotype and found larger leaf size in comparison to WT (Harish and Ram Yadav unpublished data). *HDG12* influences cell size in epidermal cell layer and its expression is restricted to the epidermal cells only. In contrast, GRFs are expressed broadly. Taken together, I dissected the role of GRFs in coordinating cell proliferation and cell expansion across epidermal and sub-epidermal cell layers.

1.9 References

Aida, M., and Tasaka, M. (2006). Genetic control of shoot organ boundaries. Current Opinion in Plant Biology *9*, 72-77.

Albert, N.W., Davies, K.M., Lewis, D.H., Zhang, H.B., Montefiori, M., Brendolise, C., Boase, M.R., Ngo, H., Jameson, P.E., and Schwinn, K.E. (2014). A Conserved Network of Transcriptional Activators and Repressors Regulates Anthocyanin Pigmentation in Eudicots. Plant Cell *26,* 962-980.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science *301,* 653-657.

Arabidopsis Interactome Mapping Consortium (2011). Evidence for network evolution in an *Arabidopsis* interactome map. Science *333,* 601-607.

Barton, M.K., and Poethig, R.S. (1993). FORMATION OF THE SHOOT APICAL MERISTEM IN ARABIDOPSIS-THALIANA - AN ANALYSIS OF DEVELOPMENT IN THE WILD-TYPE AND IN THE SHOOT MERISTEMLESS MUTANT. Development *119,* 823-831.

Benfey, P.N., Linstead, P.J., Roberts, K., Schiefelbein, J.W., Hauser, M.-T., and Aeschbacher, R.A. (1993). Root development in Arabidopsis: four mutants with dramatically altered root morphogenesis. Development *119,* 57-70.

Birnbaum, K. (2003). A Gene Expression Map of the Arabidopsis Root. Science *302,* 1956-1960.

Brady, S.M., Song, S., Dhugga, K.S., Rafalski, J.A., and Benfey, P.N. (2007). Combining expression and comparative evolutionary analysis. The COBRA gene family. Plant Physiology *143,* 172-187.

Brady, S.M., Zhang, L., Megraw, M., Martinez, N.J., Jiang, E., Yi, C.S., Liu, W., Zeng, A., Taylor-Teeples, M., Kim, D., et al. (2011a). A stele-enriched gene regulatory network in the Arabidopsis root. Molecular Systems Biology *7.*

Brady, S.M., Zhang, L., Megraw, M., Martinez, N.J., Jiang, E., Yi, C.S., Liu, W., Zeng, A., Taylor-Teeples, M., Kim, D., et al. (2011b). A stele-enriched gene regulatory network in the Arabidopsis root. Molecular Systems Biology *7.*

Brand, U., Fletcher, J.C., Hobe, M., Meyerowitz, E.M., and Simon, R. (2000). Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. Science *289,* 617-619.

Busch, W., Miotk, A., Ariel, F.D., Zhao, Z., Forner, J., Daum, G., Suzaki, T., Schuster, C., Schultheiss, S.J., Leibfried, A., et al. (2010). Transcriptional control of a plant stem cell niche. Developmental Cell *18,* 841-853.

Carlsbecker, A., Lee, J.-Y., Roberts, C.J., Dettmer, J., Lehesranta, S., Zhou, J., Lindgren, O., Moreno-Risueno, M.A., Vatén, A., Thitamadee, S., et al. (2010a). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature *465,* 316-321.

Chang, K.N., Zhong, S., Weirauch, M.T., Hon, G., Pelizzola, M., Li, H., Carol Huang, S.S., Schmitz, R.J., Urich, M.A., Kuo, D., et al. (2013). Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. Elife *2013.*

Cui, H., Levesque, M.P., Vernoux, T., Jung, J.W., Paquette, A.J., Gallagher, K.L., Wang, J.Y., Blilou, I., Scheres, B., and Benfey, P.N. (2007). An Evolutionarily Conserved Mechanism Delimiting SHR Movement Defines a Single Layer of Endodermis in Plants. Science *316,* 421-425.

Daum, G., Medzihradszky, A., Suzaki, T., and Lohmann, J.U. (2014). A mechanistic framework for noncell autonomous stem cell induction in Arabidopsis. Proceedings of the National Academy of Sciences *111,* 14619-14624.

Dolan, L., Janmaat, K., Willemsen, V., Linstead, P., Poethig, S., Roberts, K., and Scheres, B. (1993). Cellular Organisation of the Arabidopsis-thaliana Root. Development *119,* 71-84.
Franco-Zorrilla, J.M., and Solano, R. (2014). High-throughput analysis of protein-DNA binding affinity. Methods in Molecular Biology *1062,* 697-709.

Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. Proceedings of the National Academy of Sciences *111,* 2367-2372.

Gifford, D.K., Melton, D.A., Murray, H.L., Zucker, J.P., Boyer, L.A., Guenther, M.G., Cole, M.F., Young, R.A., Jenner, R.G., Kumar, R.M., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122,* 947-956.

Grbić, V., and Bleecker, A.B. (2000). Axillary meristem development in Arabidopsis thaliana. Plant Journal *21,* 215-223.

Greb, T., Clarenz, O., Schafer, E., Muller, D., Herrero, R., Schmitz, G., and Theres, K. (2003). Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. Genes and Development *17,* 1175-1187.

Haseloff, J., and Amos, B. (1995). GFP in plants. Trends in Genetics *11,* 328-329.

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.T., and Benfey, P.N. (2000). The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. Cell *101,* 555-567.

Helfer, A., Nusinow, D.A., Chow, B.Y., Gehrke, A.R., Bulyk, M.L., and Kay, S.A. (2011). LUX ARRHYTHMO encodes a nighttime repressor of circadian gene expression in the Arabidopsis core clock. Current Biology *21,* 126-133.

Hibara, K. -i., Karim, M.R., Takada, S., Taoka, K. -i., Furutani, M., Aida, M., and Tasaka, M. (2006). Arabidopsis CUP-SHAPED COTYLEDON3 Regulates Postembryonic Shoot

Meristem and Organ Boundary Formation. THE PLANT CELL ONLINE *18,* 2946-2957.

Hong, R.L. (2003). Regulatory Elements of the Floral Homeotic Gene AGAMOUS Identified by Phylogenetic Footprinting and Shadowing. THE PLANT CELL ONLINE *15,* 1296-1309.

Immink, R.G.H., Posé, D., Ferrario, S., Ott, F., Kaufmann, K., Valentim, F.L., Folter, S. De, Wal, F. Van Der, Dijk, A.D.J. Van, Schmid, M., et al. (2012). Characterization of SOC1's Central Role in Flowering by the Identification of Its Upstream and Downstream Regulators. Plant Physiology *160,* 433-449.

Ito, T., Sakai, H., and Meyerowitz, E.M. (2003). Whorl-specific expression of the SUPERMAN gene of Arabidopsis is mediated by cis elements in the transcribed region. Current Biology *13,* 1524-1530.

Jaglo, K.R., Kleff, S., Amundsen, K.L., Zhang, X., Haake, V., Zhang, J.Z., Deits, T., and Thomashow, M.F. (2001). Components of the Arabidopsis C-Repeat / Dehydration-Responsive Element Binding Factor Cold-Response Pathway Are Conserved in Brassica napus and Other Plant Species . Plant Physiology *127*, 910–917.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. The EMBO Journal *6,* 3901-3907.

Jiao, Y., Tausta, S.L., Gandotra, N., Sun, N., Liu, T., Clay, N.K., Ceserani, T., Chen, M., Ma, L., Holford, M., et al. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nature Genetics *41,* 258-263.

Jones, A.M., Xuan, Y., Xu, M., Wang, R.-S., Ho, C.-H., Lalonde, S., You, C.H., Sardi, M.I., Parsa, S.A., Smith-Valle, E., et al. (2014). Border control--a membrane-linked interactome of Arabidopsis. Science *344,* 711-716.

Kaufmann, K., Nagasaki, M., and Jáuregui, R. (2010). Modelling the molecular interactions in the flower developmental network of arabidopsis thaliana. In Silico Biology *10,* 125-143.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress expression data set: Protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant Journal *50,* 347-363.

Koo, J., Kim, Y., Kim, J., Yeom, M., Lee, I.C., and Nam, H.G. (2007). A GUS/luciferase fusion reporter for plant gene trapping and for assay of promoter activity with luciferin-dependent control of the reporter protein stability. Plant Cell Physiology *48,* 1121-1131.

Lee, J.-Y., Colinas, J., Wang, J.Y., Mace, D., Ohler, U., and Benfey, P.N. (2006). Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. Proceedings of the National Academy of Sciences *103,* 6055-6060.

Levesque, M.P., Vernoux, T., Busch, W., Cui, H., Wang, J.Y., Blilou, I., Hassan, H., Nakajima, K., Matsumoto, N., Lohmann, J.U., et al. (2006). Whole-genome analysis of the short-root developmental pathway in Arabidopsis. PLoS Biology *4,* 739-752.

Leyser, H.M.O., and Furner, I.J. (1992). Characterisation of three shoot apical meristem mutants of Arabidopsis thaliana. Development *116,* 397.

Lindsey, K., Wei, W., Clarke, M.C., McArdle, H.F., Rooke, L.M., and Topping, J.F. (1993). Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants. Transgenic Research *2,* 33-47.

Long, J., and Barton, M.K. (2000). Initiation of Axillary and Floral Meristems in Arabidopsis. Developmental Biology *218,* 341-353.

Mähönen, A.P., Bonke, M., Kauppinen, L., Riikonen, M., Benfey, P.N., and Helariutta, Y. (2000). A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. Genes and Development *14,* 2938-2943.

Mayer, K.F.X., Schoof, H., Haecker, A., Lenhard, M., Jürgens, G., and Laux, T. (1998). Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem. Cell *95,* 805-815.

Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D., and Koornneef, M. (1998). Arabidopsis thaliana: a model plant for genome analysis. Science *282.*

Millar, A.J. (1992). A Novel Circadian Phenotype Based on Firefly Luciferase Expression in Transgenic Plants. Plant Cell Online *4*, 1075–1087.

Mocharla, R., Mocharla, H., and Hodes, M.E. (1987). A novel, sensitive fluorometric staining techique for the detection of DNA in RNA preparations. Nucleic Acids Research *15*, 10589.

Moriguchi, K., Suzuki, T., Ito, Y., Yamazaki, Y., Niwa, Y., and Kurata, N. (2005). Functional Isolation of Novel Nuclear Proteins Showing a Variety of Subnuclear Localizations. THE PLANT CELL ONLINE *17,* 389-403.

Mukhtar, M.S., Carvunis, A.-R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M.T., et al. (2011). Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. Science *333,* 596-601.

Muller, R., Borghi, L., Kwiatkowska, D., Laufs, P., and Simon, R. (2006). Dynamic and compensatory responses of Arabidopsis shoot and floral meristems to CLV3 signaling. THE PLANT CELL ONLINE *18,* 1188-1198.

Nakajima, K., Sena, G., Nawy, T., and Benfey, P.N. (2001). Intercellular movement of the putative transcription factor SHR in root patterning. Nature *413,* 307-311.

Ohyama, K., Shinohara, H., Ogawa-Ohnishi, M., and Matsubayashi, Y. (2009). A glycopeptide regulating stem cell fate in Arabidopsis thaliana. Nature Chemical Biology *5,* 578-580.

Raman, S., Greb, T., Peaucelle, A., Blein, T., Laufs, P., and Theres, K. (2008). Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in Arabidopsis thaliana. Plant Journal *55,* 65-76.

Rast, M.I., and Simon, R. (2008). The meristem-to-organ boundary: more than an extremity of anything. Current Opinion in Genetics and Development *18,* 287-294.

Reddy, G. V, and Meyerowitz, E.M. (2005). Stem-cell homeostasis and growth dynamics can be uncoupled in the Arabidopsis shoot apex. Science *310,* 663-667.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Molecular Biology *53,* 247-259.

Satina, S., Blakeslee, A.F., and Avery, A.G. (1940). Demonstration of the Three Germ Layers in the Shoot Apex of Datura by Means of Induced Polyploidy in Periclinal Chimeras. American Journal of Botany *27,* 895.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., Lohmann, J.U., Scholkopf, B., et al. (2005). A gene expression map of Arabidopsis thaliana development. Nature Genetics *37,* 501-506.

Schofield, R. (1978). The relationship between the spleen colony-forming cell and the haemopoietic stem cell. Blood Cells *4,* 7-25.

Sharma, V.K., Ramirez, J., and Fletcher, J.C. (2003). The Arabidopsis CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. Plant Molecular Biology *51,* 415-425.

Shuai, B., Reynaga-Peña, C.G., and Springer, P.S. (2002). The lateral organ boundaries gene defines a novel, plant-specific gene family. Plant Physiology *129,* 747-761.

Sozzani, R., Cui, H., Moreno-Risueno, M.A., Busch, W., Van Norman, J.M., Vernoux, T., Brady, S.M., Dewitte, W., Murray, J.A.H., and Benfey, P.N. (2010). Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. Nature *466,* 128-132.

Sparks, E.E., Drapek, C., Gaudinier, A., Li, S., Ansariola, M., Shen, N., Hennacy, J.H., Zhang, J., Turco, G., Petricka, J.J., et al. (2016). Establishment of Expression in the SHORTROOT-SCARECROW Transcriptional Cascade through Opposing Activities of Both Activators and Repressors. Developmental Cell *39,* 585-596.

Taylor-Teeples, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., Lamothe, R., et al. (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature *517,* 571-575.

Thompson, J.F., Hayes, L.S., and Lloyd, D.B. (1991). Modulation of firefly luciferase stability and impact on studies of gene regulation. Gene *103,* 171-177.

Tian, C., Zhang, X., He, J., Yu, H., Wang, Y., Shi, B., Han, Y., Wang, G., Feng, X., Zhang, C., et al. (2014). An organ boundary-enriched gene regulatory network uncovers regulatory hierarchies underlying axillary meristem initiation. Molecular Systems Biology *10,* 755-755.

Yadav, R.K., Tavakkoli, M., Xie, M., Girke, T., and Reddy, G.V. (2014). A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. Development *141*, 2735–2744.

Zhang, H.-X., and Blumwald, E. (2001). Transgenic salt-tolerant tomato plants accumulate salt in foliage but not in fruit. Nature Biotechnology *19*, 765.

Zhang, H.X., Hodson, J.N., Williams, J.P., and Blumwald, E. (2001). Engineering salt-tolerant Brassica plants: characterization of yield and seed oil quality in transgenic plants with increased vacuolar sodium accumulation. Proceedings of the National Academy of Sciences *98*, 12832–12836.

CHAPTER 2

Understanding the spatio-temporal expression pattern of epidermal and sub-epidermal cell layer enriched transcription factors by promoter reporter studies

2.1 Introduction

Sequence of the whole genome of *Arabidopsis thaliana* was released in 2000 by an international consortium (Arabidopsis Genome Initiative, 2000). After the completion of genome sequencing, the goal was set to understand the functions of all genes that are encoded by this tiny genome (~140 Mb). The U.S. National Science Foundation (NSF) launched a program known as Arabidopsis2010 in 2001 with the goal of understanding the function of all genes by the year 2010 (Chory et al., 2010). At the whole genome level, expression of genes and annotation of their transcriptional start sites (TSSs) were determined by ATH1 gene chip and by tilling array studies, respectively (Laubinger et al., 2008; Stolc et al., 2005; Yamada et al., 2003). By these initiatives, rapid progress was made in cataloguing the entire repertoire of genes and their expression profiles in different plant tissues and organs using the ATH1 gene chip (Birnbaum et al., 2003; Casson et al., 2005). In addition to this, ATH1 gene chip was also used for identifying plant's responses to abiotic and biotic stresses (Kilian et al., 2007, 2012; Wanke et al., 2010). Later on, these studies were extended at single cell type resolution. First, to understand the cell type specific gene expression patterns (Birnbaum et al., 2003; Brady et al., 2007; Yadav et al., 2014a) and second to identify the cell type specific responses in response to nutrient deprivation, abiotic stress etc. (Dinneny et al., 2008).

In parallel, to capture the in-vivo expression of individual genes at genome scale, reporter constructs were made in such a way, that when they get integrated immediately after the promoter thereby reporting the native expression of a gene within the plant tissues. Moreover, when the insertion occurs close to enhancers they can also help in identifying the putative cis-elements that are recognised by trans acting factors, which are required to regulate the spatio-temporal expression pattern of target genes. Both enhancer trap and gene trap approaches were exploited in plants and animals for discovering novel genes and their expression patterns in vivo at genome scale (Springer, 2000) (Sundaresan et al., 1995). Enhancer trap lines also gave first evidence where the function of non-coding part of the genome was established in regulating the genes at genome scale.

Despite a large body of work, we are unable to answer several basic questions related to gene regulation in plants. For example, how distinct expression patterns are established and maintained in plants at tissue and cell type level. A large number of studies in plants have revealed that major elements required for regulating expression of a gene usually lie within

the first few kilo base region upstream of the translational start site (Li et al., 1994; Manners et al., 1998; Peiter et al., 2007; Santamaria et al., 2001). Some studies have shown that the regulatory elements might also be located in other regions such as in introns and 3' untranslated regions, although, such illustrations are few and far between (Hong, 2003; Larkin et al., 1993; Peiter et al., 2007). Lee et al. in (2006) identified the TFs enriched in distinct cell types of root tissue and analysed the expression patterns for more than sixty TF genes, by taking 2-3 kb upstream non-coding regions, to ascertain whether these upstream regulatory elements were sufficient to recapture the native mRNA expression reported by digital in situs. In this study, they observed that in the 80% cases the 2-3 kb upstream promoter regions were able to recapitulate gene's expression pattern at cellular level and it was comparable to that of native expression level reported by digital in situs. More recently, studies looking at the DnaseI hypersensitive sites within the *Arabidopsis*, rice, tomato, and maize genome revealed that in these plants majority of DnaseI sites are located within 3 kb region upstream of the gene (Maher et al., 2017). Therefore, to establish the function of cell type specific TFs and to discover the critical regulatory elements that drive their expression, promoter reporter constructs are preferred. The transgenic lines carrying the reporter constructs can also be used for checking expression of these genes in certain specific cell types, such as trichomes and stomata, which can be missed in high throughput studies such as microarray.

Multicellular organisms are made up of discrete cell types, which are specified by distinct sets of TFs via distinct regulatory networks from a common pool of stem cells. To achieve this, master regulatory TF expresses in few cells, and initiates a regulatory cascade regulating several downstream target genes. To map such regulations, it is important to understand the transcriptional output of all the genes for a given network in vivo by promoter reporter-based studies. The role of major players in establishing cell identity both in animals and plants were studied either by in-situ hybridization experiments or by driving the expression of reporter gene under its own promoter. Reporter can be a fluorescent protein such as GFP or YFP or histological colourings, such as GUS. Transcription of a gene results in its expression, this may result in accumulation of gene product that ultimately would lead to cell specification. Therefore, studying expression of a gene is often considered a powerful tool for studying the function of the gene in the context of network biology. However, gene expression reporters have their own intrinsic drawbacks, such as, each gene might be regulated by a complicated network. And also, the pattern of expression might reflect cell identity but it may not define

it. Therefore, one needs to combine various markers in order to infer a cell's identity.

To dissect the regulation of epidermal and sub-epidermal enriched TFs and to understand how upstream regulators modulate their expression in different zones of SAM, in space and time, I identified 65 TF genes enriched in these cell types by analysing the cell population microarray data (Yadav et al., 2014). For the majority of these TF genes, endogenous expression pattern is not yet established in the SAM. Next, I created transcriptional reporter constructs of 43 TFs by amplifying the 3 kb DNA fragment upstream of translation start site and screened them *in planta*. I investigated the in vivo expression of these reporters in this study by laser confocal microscopy to establish their role in epidermal and sub-epidermal cell type specific functions. My expression analysis of 43 epidermal and sub-epidermal enriched shoot TFs shows nine dominant patterns of expression in the SAM (Figure 2.14). Few of them display spatially restricted expression patterns in epidermal and sub-epidermal cell layer, respectively, suggesting that they are likely to get regulated by the same upstream regulators.

2.2 Materials and methods

2.2.1 Plant work

2.2.1.1 Plant growth media

Arabidopsis thaliana seeds were germinated on sterile Murashige and Skoog agar media containing 1X Murashige and Skoog salts, 1% (w/v) sucrose, 0.1% (w/v) MES. All the ingredients were dissolved in water and mixed well and the pH of the solution was adjusted to 5.8 using KOH. In the end, agar was added to the solution up to 0.8% (w/v) and was autoclaved. Kanamycin or hygromycin was added after the media was cooled to 40-50^{o}C.

2.2.1.2 Plant growth

The seeds were sown on Murashige and Skoog (MS) medium containing 0.8% Bacto agar (Himedia, India), 1% (w/v) sucrose and 0.1% (w/v) MES. *Arabidopsis thaliana* seeds were kept for vernalization at 4^{o}C for 4 days, and then transferred to plant growth chambers (Conviron, Canada and Percival Scientific, USA). Seedlings were grown at 22^{o}C under long day conditions of 16 hr light and 8 hr dark in ~ 70% humidity under 130 μmol light (Philips fluorescent tube lights). For adult plant growth, seedlings were transferred after 7-10 days into pots containing

soil/compost. The soil was prepared by mixing soilrite mix (KELTECH Energies Ltd. India), compost and perlite in the ratio of 3:1:1/2.

2.2.1.3 *Arabidopsis thaliana* transformation

Agrobacterium tumefaciens strain GV3101 was used for plant transformation experiments (Koncz and Schell, 1986). pSoup is a helper plasmid that can co-exist in the agrobacterium strain along with the pGreen plasmid in the presence of tetracycline. pSoup provides the replicase function for the replication of origin of pGreen plasmid. For promoter: reporter constructs, a Gateway compatible pGreen0229 plasmid backbone was used as previously reported by Yadav et al. (2014). *Arabidopsis thaliana* (L.) Heynh var. Landsberg (*erecta* mutant) was used as WT strain for generating promoter::reporter transgenic lines. For each promoter, gateway cloning was performed and the resultant promoter::reporter binary vector was transformed using the floral dip method (Clough and Bent, 1998) into WT background.

2.2.1.4 Seed sterilization

Seeds were sterilized in 1.5ml micro centrifuge tubes to avoid bacterial and fungal contamination on MS plates. Seeds were sterilized first by washing in 70% ethanol, containing 0.02% (v/v) Triton X-100, for one minute, followed with this, seeds were treated with 0.4% NaOCl (MERCK 1.93607.1021) containing 0.02% (v/v) Triton X-100 (Sigma T 8787), for three minutes. Seeds were washed at least thrice with sterilized water and transferred onto respective MS agar plates under sterile conditions.

2.2.2 Electro-competent cells preparation

To prepare electro-competent cell for *E.coli* and *A.tumefaciens* strains, single colony of each bacterial strain was inoculated in 5ml LB and grown overnight at 37^{o}C and 30^{o}C, respectively. In the *Agrobacterium tumefaciens* culture antibiotics were added (gentamycin 25 µg/ml to keep Ti plasmid resistance, rifampicin 50 µg/ml for chromosomal resistance). This pre-inoculum was used to inoculate the secondary culture at 1% final concentration. This culture was grown to an OD_{600} of about 0.8- 1 for *Agrobacterium* and 0.35-0.4 for *E.coli* and harvested at 4000rpm at 4 °C for 15 min. Afterwards 5 washing steps with decreasing volumes of ice-cold sterile 10% glycerol were followed before final resuspension in 1-2 ml of ice-cold sterile 10 % glycerol. Aliquots of 40 µl were made and cells were frozen at –80 °C.

The protocol for making electro competent cells was adopted from the Bio-Rad Gene Pulser Xcell Instruction Manual (# 165 2660).

2.2.3 Construction of promoter reporters in binary vector

Cloning promoter DNA fragments into binary plant vector

Genomic DNA was isolated from WT L*er* ecotype adult plants, using CTAB method (Thompson, 1980). Forward primer containing 5' CACC overhang was used in conjunction with reverse primer to amplify the 3000 bp promoter DNA fragment upstream of the translational start site. PCR amplification of promoter DNA fragment was carried out using Phusion high fidelity DNA polymerase from New England Biolabs (NEB), USA, http://www.neb.com). PCR products for each promoter were purified and cloned in the gateway vector pENTR/D/TOPO to make the entry clone. Sanger sequencing confirmed the integrity of clones for the promoter of interest. Entry vector containing promoter was used for setting up the LR-recombination reaction with *pGreen::H2B-YFP* destination vector. The resultant promoter::reporter binary vector clones were confirmed by colony PCR and restriction digestion. The oligonucleotides used for amplifying promoters are given in (Table 2.1). To perform the LR recombination reaction directly using pDONR/P4/P1R bait vector into plant transformation vector, we cloned the attR4: ccdB: attL1 fragment from pMW2 vector into *Xho* I and *Nhe* I sites of pGreen 0229 H2B-YFP. This modified vector was used for generating some of the *promoter:H2B-YFP* transcriptional fusions.

2.2.4 Molecular Biology techniques

All molecular biology techniques were performed according to Sambrook et al. (1989), unless otherwise stated in the text.

2.2.4.1 Restriction enzymes and high-fidelity DNA polymerase

Enzymes used in this study for molecular biology procedures involving manipulation of DNA or RNA, like restriction digestion, ligation, DNA amplification and reverse transcription of RNA were procured from NEB (USA) and BioRad (USA). All enzymes were used according to manufacturer's instructions.

2.2.4.2 Nucleic acid purification

2.2.4.2.1 Bacterial plasmid DNA isolation

Single bacterial colony was picked with a sterile tip and inoculated in 5ml LB containing the required antibiotics. The culture was grown overnight at 37^{o}C. Cells were harvested by centrifugation at maximum speed for one minute. Plasmids were then extracted with AccuPrep plasmid isolation kit (Bioneer, Korea) on spin columns as provided and instructed by the manufacturer. Elution of plasmid DNA was performed with warm sterile distilled water.

2.2.4.2.2 Purification of DNA fragments from agarose gel

DNA fragments were excised from gel and were heated at 65^{o}C in the QG buffer provided with the kit. The heat melted contents were passed through the spin columns according to the protocol. Gel purification kits were used from Bioneer. Manufacturer's instructions were followed at each step. Final elution was done with warm sterile distilled water.

2.2.4.2.3 Polymerase chain reaction (PCR) methods

All oligonucleotides used in this study for PCR amplification, genotyping and sequencing reactions were designed using Vector NTI advance^{TM} 11 suite (Invitrogen, USA). All oligonucleotides used in this study are mentioned in the text. PCR reactions were prepared in standard PCR thin walled tubes of 0.2ml. Suitable buffer provided with the enzyme was used at a final concentration of 1X. Primers were stored in a concentration of 100µM, with working concentration of 10µM. The dNTPs were used at a working concentration of 2.5mM. For colony or plasmid PCRs, to confirm cloning and T-DNA genotyping, homemade Taq Polymerase was used in a total volume of 15 or 25µl. Templates were added individually to each tube and the remaining components were added as master-mix. High fidelity Phusion polymerase from NEB (USA) was used to avoid mutations in PCR based cloning steps. The PCR reaction mix was initially heated at 95^{o}C for 3 minutes for initial denaturation of the template, followed by a final denaturation step at 95^{o}C for 30 seconds. Annealing was done at 55^{o}C for 30 seconds, followed by extension at 72^{o}C at the rate of 1kb/min. Step 2 to 4 were repeated for 29 cycles for Phusion polymerase PCRs and for 35 cycles for colony PCR reactions. And a final extension step was carried out at 72^{o}C for 10 minutes. Standard cloning methods were performed as described by Sambrook and Russell (2001). Plasmids

were checked by restriction digestion and after fixing the identity, clones were confirmed by sequencing.

2.2.5 Screening of promoter reporter lines

Each promoter reporter construct was transformed into WT L*er*, and T0 seeds were collected. Appropriate selective agent such as BASTA or Kanamycin was used to select primary transformants. To find out representative T1 plants that show expression pattern comparable to native mRNA expression as reported by in situ hybridization studies, shoot apices of all T1 rescued plants were dissected under the dissecting scope for individual construct (Leica KL300 LED). To reduce the workload on confocal microscope, first the dissected shoot apices were visualized in the upright epifluorescence microscope Axio Imager.Z2 (Zeiss, Germany) using long working distance water dipping objective. The representative T1 lines selected were further subjected to detailed expression analysis using confocal imaging.

2.2.6 Confocal imaging of inflorescence meristems and image processing

The confocal images for reporter transgenic lines were acquired with Leica SP8 upright confocal microscope. Four-week-old shoot apices of T1 plants were hand-picked with fine forceps, and dissected under Leica dissecting microscope after inserting them straight in specially designed rectangular boxes filled with 1.5% agar. To image the expression of fluorescent reporter within the inflorescence meristem, old flower buds were clipped off carefully under the dissecting microscope. To visualize the cell outline, shoot apices were either stained with FM4-64 (100 μg/ml, T3166, Invitrogen) or Propidium Iodide (100 μg/ml, P21493, Invitrogen) in dH_2O containing silwett (0.02%) for 15-20 mins. The inflorescence meristems were scanned in confocal microscope. YFP fluorochrome was excited with Argon laser (488/515) and emission spectra was collected by setting the band pass filter between (BP 500-535). The cell membrane marker dyes (FM4-64 and PI) were excited with Argon laser and emission spectra was collected through a long pass filter (BP 610-672nm). The confocal image stacks were converted in to three-dimensional (3D) image using Leica software. The side image of the SAM was generated using the side view function of Leica software.

2.2.7 *In-situ* hybridization

For carrying out *in situ* hybridization experiments, protocols used earlier were adopted with slight modifications (Grossniklaus et al., 1998; Long et al., 1996). The tissue fixation and embedding of inflorescence meristem was carried out as per the protocol described (http://plantlab.caltech.edu/protocols.html). For carrying out *in situ*, shoots and flowers of healthy and robust WT L*er* plants were fixed in 4% formaldehyde and embedded in molten paraplast. Sections of approximately 8µm thickness were prepared using a Leica microtome (Leica, Germany) and affixed onto ProbeOn Plus slides from Fisher Biotechnology, slides were kept at 42^{o}C overnight for baking. The sections were then deparaffinised using histoclear, and rehydrated gradually in decreasing concentration of ethanol. Slides were rinsed in SSC (150mM NaCl, 15mM Na Citrate) and treated with Proteinase K (1µg/ml, P-2308, SIGMA). Glycine PBS (2mg/ml) was then used to stop the Proteinase K digestion, following which the slides were post-fixed in 4% paraformaldehyde for 10 min. and acetylated using acetic anhydride in triethanolamine buffer for 10 min. The slides were once again washed with PBS, followed by a series of dehydration for 30 seconds each in 30% ethanol, followed by 60%, 80%, 90%, 95% and twice in 100% ethanol. The probes were then diluted in hybridization buffer (50% formamide, 300mM NaCl, 10mM Tris (pH 7.5), 1mM EDTA (pH 8), 50% dextran sulfate, 1X Denhardts and 100mg/ml tRNA), and applied to the tissue sections. The optimal concentration for different probes was determined by trial experiments. Hybridization was done at 50-55^{o}C overnight. After hybridization, the slides were washed in 0.2X SSC for 2 hrs at 55^{o}C. Slides were then washed in PBS and checked for the presence of colour. For colour detection, slides were blocked for 45 mins with 1% Boehringer block (100mM Tris (pH 7.5), 150mM NaCl). The slides are then washed with BSA wash solution (1% BSA in 100mM Tris (pH 7.5), 150mM NaCl, 0.3% triton X-1000, for 45 mins. Next, Anti-DIG antibody is applied to the sections. The Anti-DIG antibody linked to alkaline phosphatase was diluted at 1:1250 in the BSA wash solution. The slides were then incubated for 2 hrs at room temperature, and washed 4 times with BSA wash solution, 15 mins each. Before applying the substrate for colour reaction, the slides were equilibrated in TNM-50 (100mM Tris pH 9.5, 100mM NaCl, 50mM $MgCl_2$) for 10 mins. NBT-BCIP was used as substrate. The reaction was stopped with 1X TE after 24 hrs to 3 days. The slides were mounted with 50% glycerol, covered with coverslip and sealed with nail paint. The slides were then observed in Zeiss Axio Imager.Z2 microscope in DIC and images were captured. The images were then processed in Adobe Photoshop CC 2015. The

probes were amplified from the CDS of respective genes cloned in pENTR/D/TOPO vector. For generating sense probes, CTCGGGCCCTAATACGACTCACTATAGG and GCCAACTTTGTACAAGAAAGCTGGGTCG primer pair was used. And for generating anti-sense probes, GACCATGTAATACGACTCACTATAGGGGATATCAG and AAAGCAGGCTCCGCGGCCGCCCCCTTCACC primer pair was used. *In vitro* transcription was carried out with DIG-UTP (DIG labelling mix, Roche diagnostics) using T7 RNA polymerase.

2.3 RESULTS

2.3.1 Selection of target genes for promoter reporter analysis

In this study, I wanted to uncover the role of TFs in the development of SAM by studying their spatio-temporal expression pattern. The TFs were identified from published microarray data where individual cell populations of SAM were marked with fluorescent protein based reporters, and were sorted to establish a cell type specific gene expression map (Yadav et al., 2014). This way a gene expression map was prepared by pooling gene expression data from ten cell populations. This map covered, epidermal, sub-epidermal, niche cells / organizing centre, stem cells, and differentiated cell types of peripheral zone and rib meristem including vasculature. A total of 1456 genes were found to be enriched > 1.5-fold in the three cell layers of the shoot, namely, L1 cell layer / epidermal, L2 cell layer /sub-epidermal and the L3 layer / corpus. Of the 1456 genes, 535 were in the L1 layer, 256 in the L2 layer and 665 in the L3 / niche cells (Yadav et al, 2014). In the epidermal cell layer, of the 535 genes, 44 encode for TFs. In the sub-epidermal cell layer, of the 256 genes, 21 encode for TFs, while in the corpus of the 665 genes, 52 encode for TFs (Yadav et al, 2014). This study focuses on the TFs that are enriched exclusively in the epidermal and the sub-epidermal cell types of SAM.

2.3.2 Generating promoter reporter lines

In the first attempt towards understanding the function of epidermal and sub-epidermal cell type enriched TFs, I investigated their endogenous expression in the SAM by making promoter reporter fusions. To make transcriptional fusions, 3 kb region upstream of the translational start site was chosen for each of these TF promoters. Of the 65 TFs enriched in L1 and L2 cell layer, transcriptional fusion constructs were made for 43 TFs in pGreen 0229 gateway compatible vector having a nuclear localized histone H2B fused translationally with yellow fluorescent protein (H2B-YFP). The primers used for amplifying 3 kb promoter fragment are given in Table 2.1. For some of the genes, reporter constructs were made by setting up a LR reaction between P4P1R vector having the 3 kb promoter and the modified pGreen0229 vector (having attR4-attL1 sites). The pGreen0229 vector was modified in the lab to fasten the cloning of promoter reporter constructs. Of the 44 reporter fusions made, one was not transformed *in planta* because the MAS5 expression value for this gene was less than 100 in the cell population data. And there is a likelihood of not getting the expression of such a gene using the promoter reporter. Out of the 43 promoter reporters screened *in planta*,

promoters of 23 genes were active in the SAM, however, for 2 (*WRKY25* and *CBF1*), I noticed very weak expression restricted to a few cells. To ensure the robustness and to get representative integrity of visualized expression patterns for each construct, several independent transgenic lines were screened for each reporter construct (Table 2.2). Some of the lines did not show any expression, probably due to the lack of regulatory elements within the 3 kb promoter region used or integration of the transgene into the heterochromatin region of the genome.

Table 2.1: Primers used for amplifying 3 kb promoter fragments

Gene ID	Forward and reverse primers
AT5G64060	CACCAAAATGTTTGTCACATCCAGGTCATG
	TGGAAGACAAGGGGAAAACTTTCACTACAC
AT5G52170	CACCTATACAAGCTTGACACCCGTGAAAGT
	TTTCCCTCTCTGCAACCAAATTAAATAGTC
AT4G14770	CACCAGATAGATAGCCTAATTTTACAAAAC
	TTTCCAACACACAAACAAAAAAAATCAC
AT2G30250	CACCTTTCCTGGATCCATCGATCATTCACT
	GATGGTCTTTAATAAAGGAGACAACT
AT2G31730	CACCATATCTATGAATTTACCCACATGGT
	TCCTTGTCATCAAACAAAACAAAAATGA
AT5G61190	CACCTTCAAAATCAAAGTGTAGAAACAG
	GGAGATACTAGAGAGAGAAGAGAC
AT4G25490	CACCTTCTAAAAATCTTATTCCTCTGAA
	TGATCAGAGTACTCTGTTTCAAGAAACT
AT1G54160	CACCTGTATAACAATACAACACAGCTTA
	TGTCTTCAAATCTTTACTTCAATAAACT
AT1G75710	CACCAATTAATGAGTTGTGTAGAGAAAT
	TGAAAGAAGAAAGAAACTCTTGAGAATG
AT2G20180	CACCTCAAGTAGACAATAACTTTCTAAG
	ATCTCTCTCTCTACAAAGATGATGATAA
AT2G27050	CACCGCTAGGTGGCTAAAGACAGATGCA
	GTCTCTTCCACCACAATCAAGAACAGAG
AT3G47600	CACCTCTGTTCTAGAAACTATTAGGAAA
	CTCTATCTTCTCTATGTAAATATCACTT
AT4G01250	CACCCATTAAAGCCCAATCCAGATATAA
	TGAATTTGGTTACTCAGGGGAGACAAAA
AT4G04890	CACCTTCTGTAATAAATACGTATTAGAA
	TGTTATGGATGATTGACTATGATCACTC
AT4G16610	CACCAAGTCTTTTTTTTTTCTTCTTCTTA
	TTTGCCAAAAGAAACAAGAACAGAACAA
AT4G21750	CACCTTGGAACTTACGTAGTTTACATGC
	GATGATGATGGATGCCTATCAATTTTTG
AT5G46880	CACCAACATGCATCATCGCGGAAAAAGT
	CACTTGGAAGATTTAAAAAAACCTAATG
AT5G49330	CACCTACCAATTTCGGTGGCCCATCTAT
	TGCTTCTCGGTCTCTTCTGTCAAAAAGA
AT5G54630	CACCGAAGCACCAGCCTTCTTAGGGAG
	TGTTGTTGTCGATATCTCTCTCTTTCTT
AT5G57660	CACCGCATGCTCATCATATTATTTATCA

	TAACTATAACTTTTTTTATTTTCTTGAG
AT2G38340	CACCACGATTGAAGATAGGCTTTCCTCCAT
	TGGAAAAACACAACACGTACAAACTGTAG
AT1G49720	CACCAATGCTTATTATTTCATGCATTTG
	ACTTCTTTTCTGTTTCACAACGGAATCA
AT2G28810	CACCATTTTTTAATCAGGTTGCATATTA
	GAATATGTTTTGGAACTTTGAAAAAAGA

2.3.3 Expression pattern of transcriptional fusions (*promoter*:*H2B-YFP*) of different TF family members of *Arabidopsis thaliana*

2.3.3.1 Apetala2/ Ethylene Responsive Binding Protein (AP2/EREBP) transcription factor family

The AP2/EREBP family of TFs is unique to the plants and is one of the biggest families in *Arabidopsis*. This family has a total of 169 members, which are sub-divided into 4 sub-groups, AP2 sub-family, RAV sub-family, DREB and ERF sub-family. A total of 9 AP2/EREBP family members are enriched in the epidermal and sub-epidermal layers of the shoot, namely, *CBF1/AT4G25490*, *CRF6/AT3G61630*, *DEWAX/AT5G61590*, *DREB19/AT2G38340*, *DWARF AND DELAYED FLOWERING1 (DDF1)/AT1G12610*, *ERF9/AT5G44210*, *RAP2.4/AT1G22190*, *SHN2/AT5G25390*, and *AT1G64380*.

DEHYDRATION-RESPONSIVE ELEMENT (DRE) BINDING PROTEIN 1 / C-REPEAT BINDING FACTORS (DREB1 / CBF1) / AT4G25490

DREB1 / *CBF1* is a member of DREB sub-family A1 of AP2/ERF family of TFs. There are six members in this sub-family, and contains one AP2 domain. ***DREBs/CBFs*** are involved in response to low temperatures and increases the freezing tolerance of plants in long term. The microarray data from various shoot cell types indicate that the transcript of this gene accumulates highest in vasculature tissue (Xylem pole and phloem), and in the epidermal layer of the shoot apex. Analysis of plants transformed with *pDREB1B/CBF1:H2B-YFP*, however, did not reveal any expression in the epidermal layer of the shoot. Interestingly, the nuclear localized H2B-YFP appeared in random cells in the shoot apex without any apparent pattern specific to shoot cell type or tissue (Figure 2.1 A). Also, analysis of different stages of embryos of *pCBF1:H2B-YFP* lines, revealed only patchy expression of the gene in the heart stage and late stage embryos (Figure 2.1 C). This suggests that the transcript of *DREB1B* / *CBF1* is either unstable at normal temperatures or is not induced at sufficiently high level, therefore it becomes difficult to capture its expression using GFP based reporter in the

absences of critical threshold. In contrast, reporter made using GUS gene, worked well in such cases. A recent study found that *DREB1A* is induced by rapid and slow temperature decrease, and the clock genes CCA1/LHY1 mainly regulates the expression of *DREB1B* promoter (Kidokoro et al., 2017). To find out the role of cold in the induction of *DREB1B* / *CBF1*, the line showing patchy expression was subjected to cold treatment for 48 hrs. The random expression of YFP was detected in a greater number of cells in cold treated plants in comparison to the control plants (Figure 2.1 B). However, increase in the number of GFP positive cells after cold treatment was not quantified in this study and needs to be addressed in the future studies.

DREB19 / AT2G38340

DREB19 is a member of the DREB sub-family of the AP2/ERF family of TFs. Maximum expression of this TF, according to the cell type specific microarray data, was found in the sub-epidermal cell layer, followed by the CZ cells. However, the expression of this TF in the other cell types of the shoot was fairly low. Plants carrying the *pDREB19:H2B-YFP* transgene did not show YFP fluorescence.

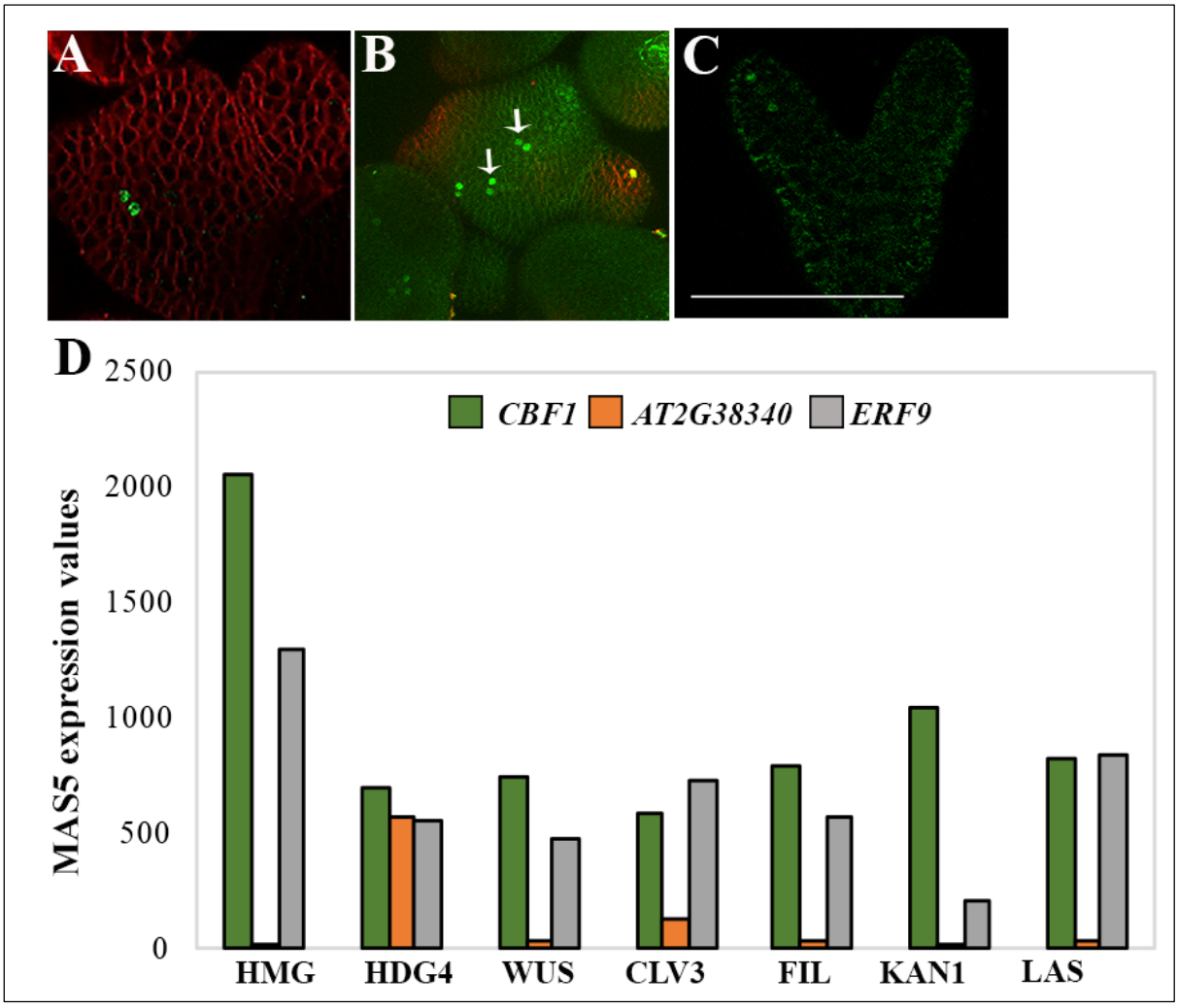


Figure 2.1: Expression pattern of AP2/EREBP TF family members. (A) Confocal image showing the top view of SAM of *Arabidopsis,* carrying the transgene of *pCBF1:H2B-YFP.* The H2B-YFP appears in random cells. (B) Confocal image showing more number of YFP positive cells (white arrows) after subjecting the *pCBF1:H2B-YFP* transgenic line to 48 hrs. of cold treatment. (C) Confocal image of heart stage embryo showing expression of *CBF1*.. (D) Histogram showing the MAS5 expression values of *CBF1, DREB19* and *ERF9* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.2 ARID Transcription Factor Family

A total of 11 members are present in this family in *Arabidopsis.* The ARID (A-T rich interaction domain) DNA binding domain is conserved in all eukaryotes. A few ARID proteins also contain an additional motif, known as the HMG-box. One ARID family member, *ATHMGB15*, is enriched in the epidermal layer of the shoot.

ARID-HMG DNA-BINDING PROTEIN 15* (*ATHMGB15*) /*AT1G04880

ATHMGB15 locus encodes an ARID-HMG DNA-binding protein and belongs to HMGB family of TFs. HMG box proteins are associated with chromatin and interact with DNA and nucleosomes in catalysing the changes in DNA topology. The *Arabidopsis* genome encodes 15 HMG-box proteins that are subdivided in to four groups, HMGB-types, ARID-HMG proteins, 3xHMG proteins and the structure-specific recognition 1 (SSRP1). ARID-HMG proteins contain one HMG box and an A/T-rich interaction domain (ARID). *Arabidopsis* genome encodes four ARID-HMG-like proteins. *ARID-HMG1 (AT1G76110)* and *HMGB15* mRNA expression pattern was reported in the shoot apical meristem by in situ hybridization studies (Yadav et al., 2009). Cell type specific microarray data obtained from the sorted cells clearly shows expression of *HMGB15* in the epidermal cell layer. A similar expression for *pAT1G04880:H2B-YFP* construct was captured using the fluorescent protein marker in the SAM (Figure 2.2). However, *HMGB15* expression pattern recorded using the *in-situ* studies was recorded in a narrow stripe in the epidermal layer in comparison to *pAT1G04880:H2B-YFP*.

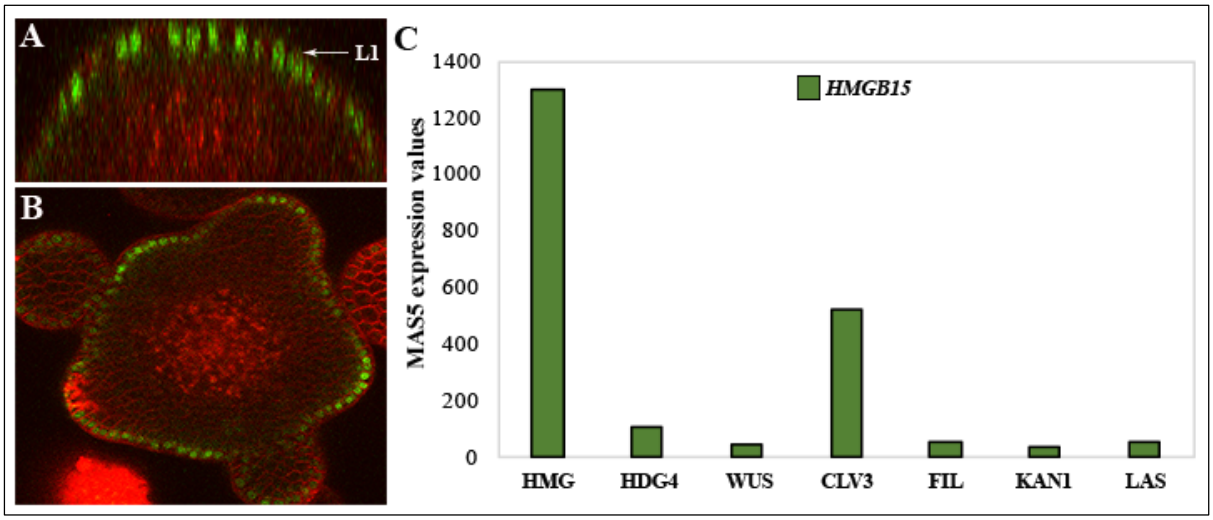


Figure 2.2: Expression pattern of ARID TF family member, *HMGB15.* (A, B) Confocal image showing the side view and top view of SAM of *Arabidopsis,* carrying the transgene *pAtHMGB15:H2B-YFP.* The white arrow in A, shows the expression of *HMGB15* in the L1 layer. (C) Histogram showing the MAS5 expression values of *AtHMGB15* in the various shoot cell types (Yadav et al., 2014). Cell outline highlighted by FM4-64 dye (red).

2.3.3.3 Basic Helix Loop Helix Transcription Factor Family

bHLH TF family is one of the largest TF families in the plant and has high degree of conservation among all kingdoms. There are 225 members in this family in *Arabidopsis* and all the members possess a bHLH domain that is required for DNA binding and dimerization. Transcripts of five bHLH genes that are enriched in the epidermal and sub epidermal cell layer of shoot apical meristem of *Arabidopsis* are as follows; *ENHANCER OF GLABRA 3* (*EGL3*) / *AT1G63650, PHYTOCHROME INTERACTING FACTOR 3- LIKE 5 (PIL5) / AT2G20180, AT2G31730, AT4G00480, AT4G01460*.

ENHANCER OF GLABRA 3* (*EGL3*) / *AT1G63650

EGL3 / AT1G63650 belongs to bHLH family of TFs. Previous studies showed the role of *EGL3* in trichome and root hair development. *egl3* enhances the *glabara3* (*gl3*) phenotypes of trichome loss on the surface of leaf and stem. To find out the role of *EGL3* in shoot epidermal layer, transgenic lines were created using 3 kb promoter fragment. Different stage embryos were dissected to study the expression of *EGL3* in the early stages of development. *EGL3* was not found to express in the globular stage embryo, however, it expresses in the epidermal layer of heart and torpedo stage embryo (Figure 2.3 J, K). Confocal imaging of the SAM and early flower primordia revealed expression of *EGL3* in the epidermal layer of emerging sepal primordia, however, there was no expression detected in the SAM. Interestingly, close analysis of confocal images of early flower primordia revealed that *EGL3* is not expressed in the stage-II flower epidermis, but it starts showing up expression in the early stage-III flower when the sepal primordia emerge (Figure 2.3 C). Subsequently, the expression of *EGL3* persists in sepal epidermis through stage-III and later stages of flower development. The expression analysis of *EGL3* reporter reflects its stated function in trichome development. It will be interesting to see whether the *gl3/egl3* mutant have problem in sepal primordia initiation.

PHY-INTERACTING FACTOR 1 / PHYTOCHROME INTERACTING FACTOR 3-LIKE 5* (*PIF1* / *PIL5*) / *AT2G20180

PIF1 / *PIL5* is a novel Myc-related bHLH TF that interacts with the phytochromes. *PIF1* / *PIL5* activates transcription in the dark, and plays negative role in phytochrome mediated signalling. This gene expresses in the epidermal cell type of the shoot as evident from the gene expression map of shoot cell types. *PIF3* was the first TF belonging to this family that

was identified in yeast-two-hybrid screen based on its interaction with phytochrome B as the bait. Several other homologs of this family *PIF4*, *PIF5* also act as negative regulators of phyB mediated light signalling. Quadruple *pif* mutant displays constitutive photomorphogenesis phenotype in the dark, indicating the redundancy in their function. Confocal imaging of the 3 kb promoter, could not capture any expression of this gene in the shoot apex CZ. I screened 4 lines and found the expression was clearly visible in the flower primordia cells (Figure 2.3 D, E, F), but its expression was completely absent from the shoot apex. Not only in the inflorescence meristem, but this gene expresses right from the very early stages of development. *PIL5* expresses in the hypocotyl region of the heart stage and torpedo stage embryo (Figure 2.3 L, M), consistent with function of different *PIFs* in skotomorphogenesis. Previous studies looking the role of *PIF* in light signalling have shown strong expression of *PIF1* in the *Arabidopsis* seedlings using GUS reporter, however, they did not focus on the shoot apex. *PIFs* are considered important players in the skotomorphogenesis and therefore, it was expected that they will be expressed in all tissues of plant. Epidermal cell layer enrichment of *PIF1* is supported by the promoter reporter data, indicating the importance of *PIF1* in regulating expression of target genes in epidermal cell layer in response to light.

bHLH / AT2G31730

AT2G31730 belongs to the bHLH superfamily of proteins. Of the 19 T1 lines screened for this reporter, 18 showed expression in the shoot and flower meristem. The expression of the reporter was observed in the shoot and flower meristem, and this data clearly establishes the expression pattern of *AT2G31730* in the PZ cells. The cell type specific microarray study detected highest expression in the sub-epidermal cell layer. Interestingly, *AT2G31730* also expresses in the epidermal cell layer, and in corpus but not in the RM (Figure 2.3 G, I). The 3D image reconstructed from the confocal stacks show expression of this gene in the emerging organ primordia, and later on in the emerging flower primordia. This expression recedes from the CZ cells in the flower meristem of stage-3 flower primordia (Figure 2.3 H).

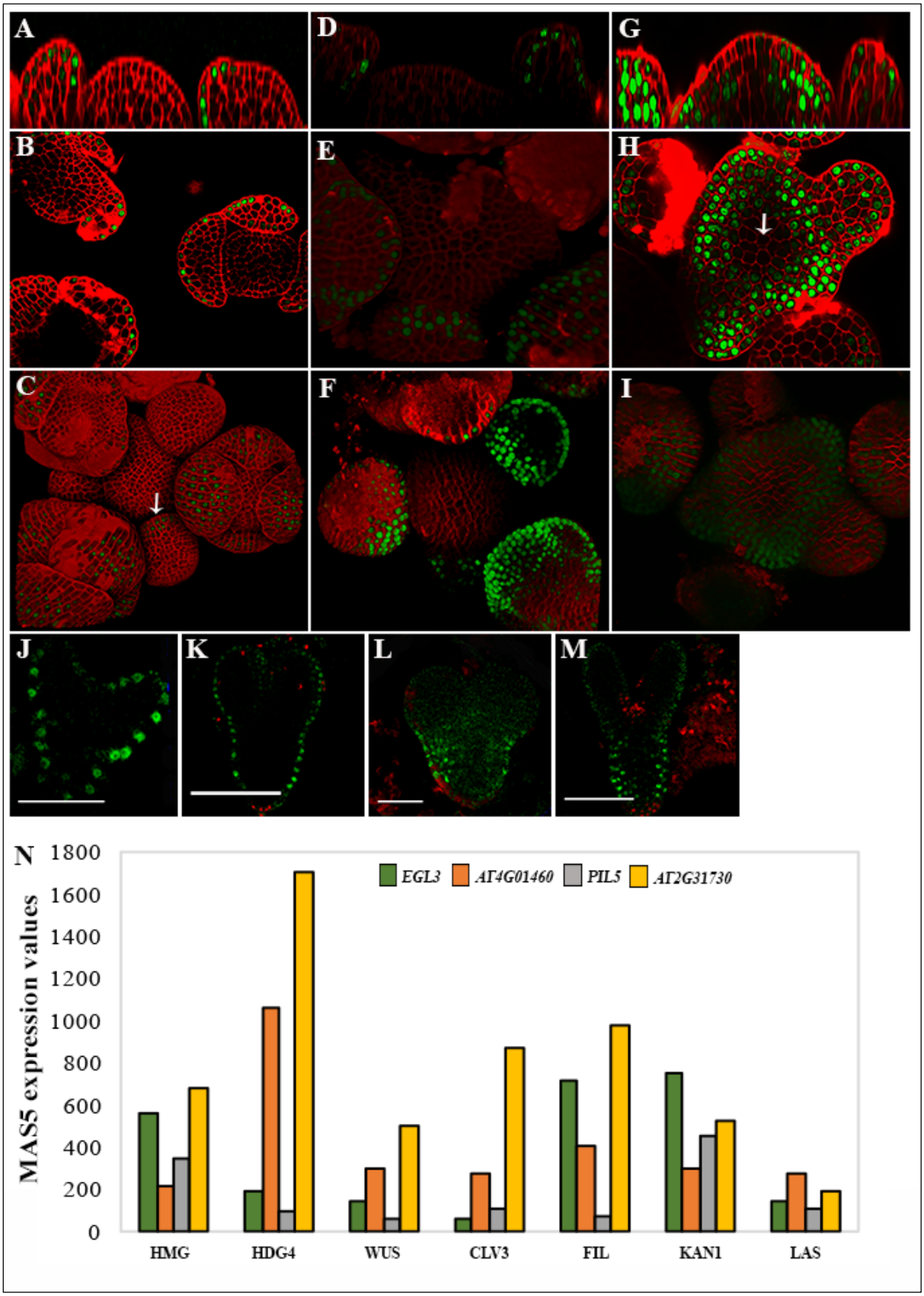


Figure 2.3: Expression pattern of bHLH TF family members. (A, D, G) Side view of WT L*er* shoot apex showing the expression pattern of *pEGL3:H2B-YFP, pPIL5:H2B-YFP, pAT2G31730:H2B-YFP.* Transverse section of SAM and 3D reconstructed top view

represented respectively in (B, E, H) and (C, F, I) for *pEGL3:H2B-YFP, pPIL5:H2B-YFP* and *pAT2G31730:H2B-YFP*. The white arrow in H shows the absence of the expression of *AT2G31730* from the CZ of the shoot. (J, K) Heart stage and torpedo stage embryos, showing the expression of *EGL3* in the epidermal cell layer. (L, M) Heart stage and torpedo stage embryos, showing the expression of *PIL5* in the hypocotyl region of the embryo. (N) Histogram showing the MAS5 expression values of *EGL3, PIL5* and *AT2G31730* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.4 bZIP Transcription Factor Family

Members of this family possess a DNA binding and a dimerization bZIP domain. A stretch of 25 amino acids basic region is present closer to the leucine zipper in bZIP proteins. This family comprises of 127 members. *ABF1/AT1G49720, bZIP52/AT1G06850* are enriched in the shoot cell population data.

ABSCISIC ACID RESPONSIVE ELEMENT-BINDING FACTOR 1* (*ABF1 / AtbZIP35*) / *AT1G49720,

bZIP family TFs are involved in diverse biological processes such as pathogen defence, light and stress signalling, seed development and flower development. *ABF1* falls in the group A of bZIP TFs and members of this group contain a bZIP domain and three casein kinase II phosphorylation sites. Group A TFs are characterized for their role in stress signalling. *ABF1* expression is induced besides abscisic acid (ABA) by cold and heat stress in *Arabidopsis* (Lee, 2005). Although, transcript of *ABF1* was enriched significantly in the epidermal layer but equally high expression of this TF was also observed in the sub-epidermal cell layer and corpus cell population data. Analysis of *ABF1* expression using the 3 kb promoter in the different stages of embryo revealed no expression in the globular and heart stage embryos. However, the torpedo stage embryos did show some expression in the adaxial side of the cotyledon and some parts of the hypocotyl, although the expression was not very uniform (Figure 2.4 D). This does not give a clear idea about the expression of *ABF1* in the early stages of development, however, it expresses broadly in the SAM cell layers. In the flower primordia, expression of this TF is more dynamic. In the stage I and II flower primordia, *ABF1* is active in all three cell layers, however in the later stages, e.g. in stage II and III of flower development expression of reporter is confined to epidermal cell layer. Though it is hard to see this dynamic behaviour of *ABF1* expression in cell population microarray data.

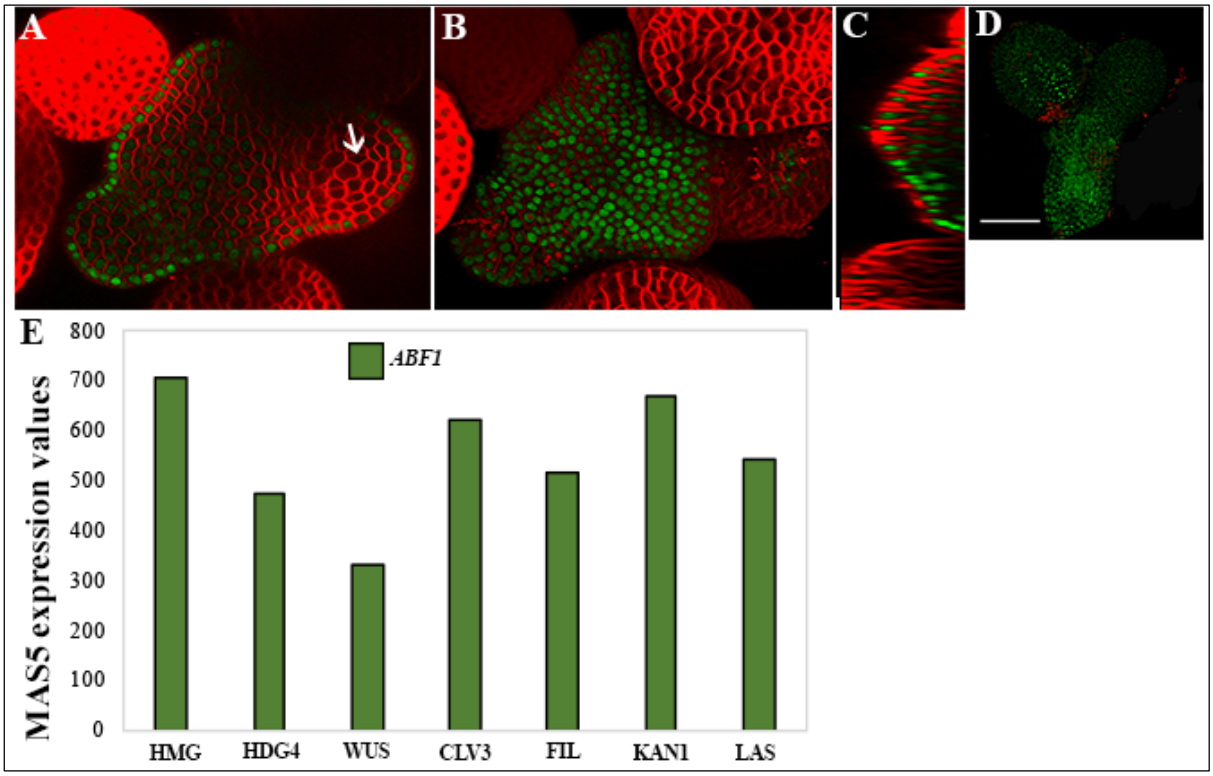


Figure 2.4: Expression pattern of bZIP TF family member, *ABF1.* Transverse section of WT L*er* shoot apex (A), 3D reconstructed top view (B) and side view (C) respectively, showing the expression pattern of *pABF1:H2B-YFP*. The white arrow in A, shows the absence of *ABF1* expression from the stage III flower primordia. (D) Late torpedo stage embryo, showing the expression of *ABF1*. (E) Histogram showing the MAS5 expression values of *ABF1* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.5 C2C2_CO-like Transcription Factor Family

There are a total of 17 members in this family. The first member identified from this family was *CONSTANS*, which has a role in induction of flowering in long photoperiods. One member of this family, *B-BOX DOMAIN PROTEIN 6*/*BBX6*/*Constans like 5*/*COL5*/*AT5G57660* is enriched in the epidermal cell type of the shoot.

CONSTANS like 5 (COL5)/ BBX6

Cell type specific micro-array data revealed a significantly high expression of this gene in epidermal cell type, abaxial organ boundary and shoot xylem vasculature cell type. However, no expression of this gene was captured in the shoot using the 3 kb promoter fragment.

2.3.3.6 C2C2_DOF Transcription Factor Family

This family of TFs is unique to the plants and there are 36 members in this family in *Arabidopsis.* These family proteins contain a Dof domain (DNA binding with one finger), which has a single C_2-C_2 zinc-finger. Four members of this family, *COG1/AT1G29160*, *DOF2.4/AT2G37590, OBP2/AT1G07640* and *AT2G28810*, are enriched in the epidermal and sub-epidermal layers of the shoot.

Dof type zinc finger / *AT2G28810*

AT2G28810 is a Dof type Zn-finger protein, where Dof refers to "DNA binding with one finger". Dof TFs have a conserved DNA binding domain at the N-terminus and activation domain at C-terminus of the protein. Highest expression of this transcription factor was found in the phloem cell population of the stem, represented by S17 cell type. However, the expression of the gene in the other cell types of the shoot was very low, with MAS5 values less than 100. The plant lines screened for YFP expression did not show any fluorescence in the shoot cell types.

2.3.3.7 C2H2 Transcription Factor Family

C2H2 family consists of zinc-finger proteins that are involved in a variety of functions from DNA binding to protein-protein interactions. There are 211 members in this family, out of which 7 are enriched in the epidermal or sub-epidermal cell types of the shoot. The genes encoding for these TFs are, *AZF3/AT5G43170*, *HDA3/AT3G44750*, *STOP1/AT1G34370*, *AT1G75710*, *AT4G16610*, *AT5G54630* and *AT5G61190*.

C2H2-like Zinc-Finger / AT1G75710

AT1G75710 locus encodes a C2H2 like zinc-finger TF. Gene expression data from the shoot cell types revealed the enrichment of this gene in the epidermal cell layer and in the differentiated cells of PZ (Figure 2.5 L). Expression pattern of *pAT1G75710:H2B-YFP* reporter was captured using confocal microscopy. This zinc finger gene starts to express very early during the development. It is expressed throughout the globular stage embryo proper, however, the expression limits to the epidermal layer from the heart stage embryos and later (Figure 2.5 D, E, F). Furthermore, side view of SAM reconstructed from confocal images confirms the expression of *AT1G75710* in the L1 meristematic layer of the shoot apical meristem (Figure 2.5 A). Also, the 3D top view of confocal images shows the expression of this transcription factor in the emerging flower primordia of the SAM (Figure 2.5 C).

However, the expression of this TF is absent from the central region of the SAM, where *CLV3* is normally expressed. *pAT1G75710:H2B-YFP* expression was also found in the mature flowers. In the flower primordia, the expression was more restricted towards the region of upcoming floral organs.

AT4G16610

AT4G16610, is C2H2 Zn-finger protein. In silico analysis of *Arabidopsis* genome revealed 176 genes encoding for zinc finger proteins. According to the microarray data published by Yadav et al (2014) for shoot cell types, this TF expresses maximally in the L1 layer of the shoot, with a MAS5 value of 344 in the HMG cell type (Figure 2.5 L). A total of eight transgenic lines were screened for the reporter expression, however, none of them showed expression in the shoot.

C2H2-like Zinc-Finger / AT5G54630

AT5G54630 belongs to the C2H2 Zinc Finger family of TFs in *Arabidopsis.* Analysis of T1 transgenic lines carrying *pAT5G54630:H2B-YFP* construct revealed expression of this TF in the epidermal cell layer of the SAM. Interestingly, the H2B-YFP expression was detected in the peripheral zone epidermal cell layer including the emerging organ anlagen, excluding the epidermal cell layer from CZ. Of the 15 T1 lines screened, 12 revealed similar transcriptional activity of the *AT5G54630* promoter. The reporter expression shows overlap with expression pattern reported in the microarray study by Yadav et al. (2014). Other than the inflorescence meristem, expression of the gene was also studied in the different stages of embryo development. Unlike *AT1G75710, AT5G54630* does not express in the entire globular stage embryo, but the expression is restricted only to the epidermal layer of the apical end of the embryo (Figure 2.5 G). Also, in the heart stage, expression is limited to the epidermal layer of the apical end and is completely absent from the hypocotyl region (Figure 2.5 H).

C2H2 like Zn-finger / *AT5G61190*

AT5G61190 is a member of C2H2 type Zinc finger protein family with a putative glycosyl hydrolase function. Nine independent transgenic lines (T1) were screened to check the YFP fluorescence, however, none of the plant line showed expression of H2B-YFP. Interestingly the annotation of this gene suggests that it is a putative endonuclease or glycosyl hydrolase that has C2H2-zinc finger domain. Glycosyl hydrolases are involved in hydrolysis of

glycosidic bonds in complex sugars. This family of transcription factors with AP2 like domain on the one hand and on the other hand a glycosyl hydrolase domain, is still not studied in detail.

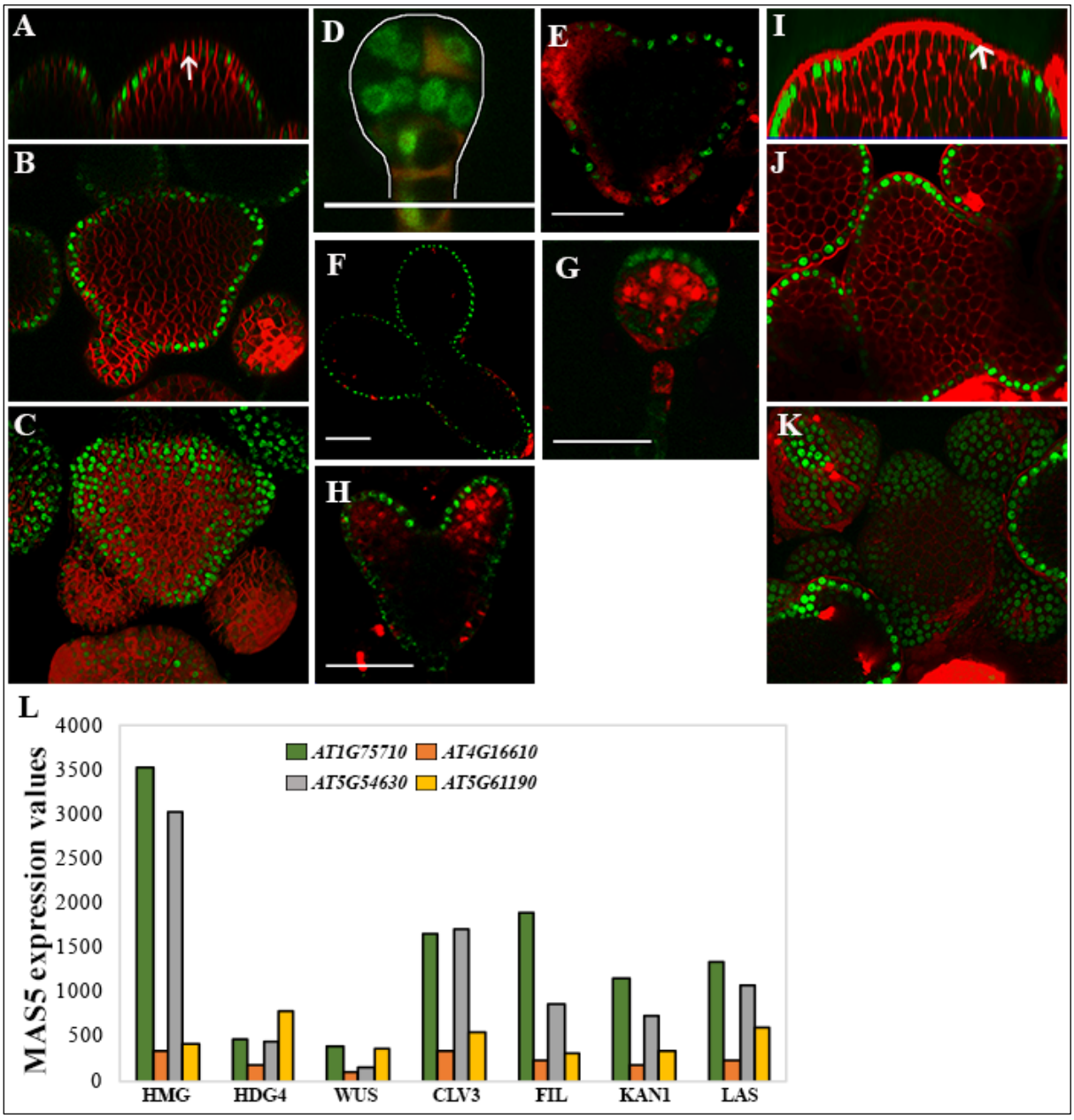


Figure 2.5: Expression pattern of C2H2 TF family members. (A, B, C) Side view of WT L*er* shoot apex, transverse section and 3D reconstructed top view, respectively, showing the expression pattern of *pAT1G75710:H2B-YFP*. (D, E, F) 16-cell stage, heart stage and bent cotyledon stage embryo, showing the expression of *pAT1G75710:H2B-YFP* respectively. (G, H) Globular and heart stage embryos respectively, showing the expression of *pAT5G54630:H2B-YFP*. (I, J, K) Side view of WT L*er* shoot apex, transverse section and 3D reconstructed top view, respectively, showing the expression pattern of *pAT5G54630:H2B-YFP*. The white arrows in A and I shows the less reporter activity in the cells of the central

zone. (L) Histogram showing the MAS5 expression values of *pAT1G75710:H2B-YFP, pAT4G16610:H2B-YFP, pAT5G54630:H2B-YFP, pAT5G61190:H2B-YFP* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.8 CCAAT_HAP2 Transcription Factor Family

CCAAT box binding factors (CBF) or Heme associated proteins (HAPs) or Nuclear Factory Y (NF-Y) transcription factor family is found in all eukaryotes. NF-Y family consists of three sub-families, NF-YA, NF-YB and NF-YC. These proteins bind to the DNA only as heterotrimeric complexes. Some members of this family, such as *NF-YB9/LEC1* are known to play important roles in embryogenesis. In *Arabidopsis*, there are 43 members in this family and three of them are enriched in the shoot, *NF-YA2/AT3G05690, NF-YA5/AT1G54160, NF-YA10/AT5G06510.*

NUCLEAR FACTOR Y, SUBUNIT A5* (*NF-YA5*) / *AT1G54160

NUCLEAR FACTOR Y, SUBUNIT A5 (*NF-YA5*) / *AT1G54160* is a member of the CCAAT-binding transcription factor (CBF-B/NF/YA) family. According to Yadav et al, (2014), *NF-YA5* expression is enriched in the L1-meristematic region of the shoot. It also expresses in the xylem pole cells of SAM vascular tissue. The expression pattern of *pNF-YA5:H2B-YFP* was imaged with confocal laser-scanning microscopy. Confocal images revealed H2B-YFP expression right from the embryonic stages of development. *NF-YA5* starts to express right from the 8-cell stage. In the 16-cell stage, expression could be visualized through-out the embryo, however, expression in the apical cells was weak in comparison to the expression in the basal cells. In the globular stage embryos, the gene expression limits only to the epidermal layer. In the heart stage and torpedo stage embryos also, *NF-YA5* expresses in the epidermal layer of the cotyledons as well as the hypocotyl region (Figure 2.6 D, E, F, G). In the SAM also, expression was limited to the epidermal layer (Figure 2.6). From the confocal stacks, a 3-dimensional (3D) image was constructed using LAS-X software to visualize the top view of shoot apex. The *NF-YA5* promoter is active throughout the epidermal cell layer in the shoot. Of the 19 T1 plant lines screened, 17 were positive for H2B-YFP expression in the epidermis of shoot.

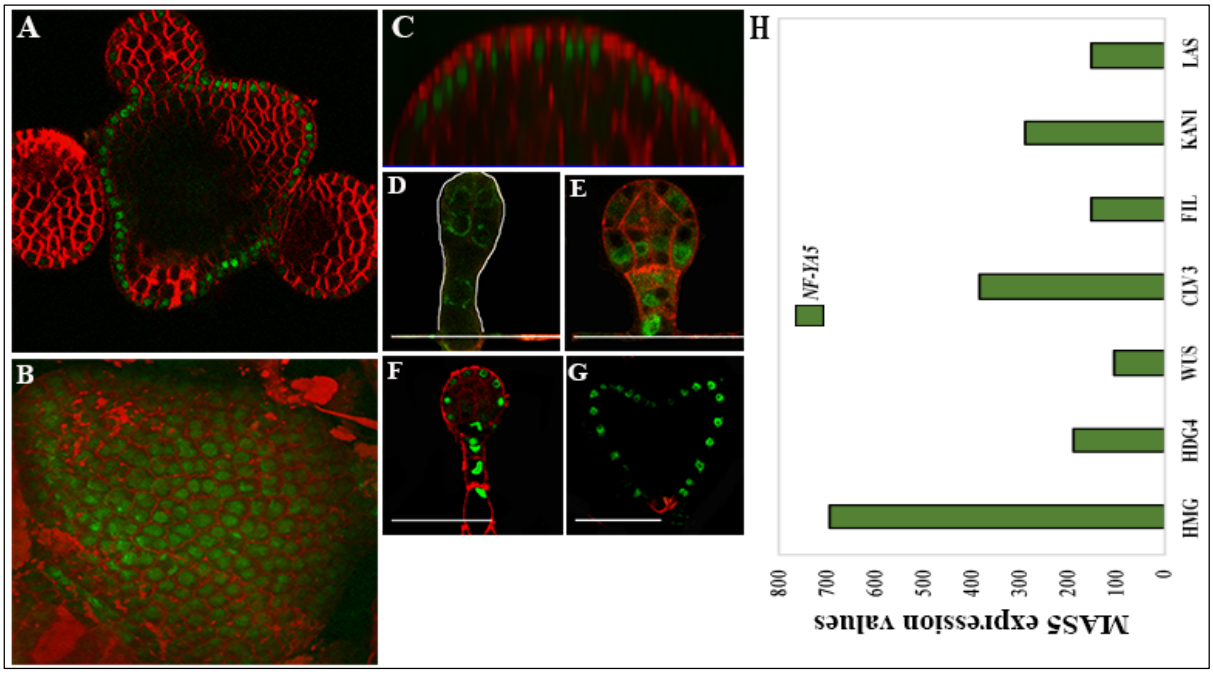


Figure 2.6: Expression pattern of CCAAT_HAP2 TF family member, *NF-YA5.* (A, B, C) Transverse section of WT L*er* shoot apex, 3D reconstructed top view and side view, respectively, showing the expression pattern of *pNF-YA5:H2B-YFP*. (D, E, F, G) 8-cell, 16-cell, globular and heart stage embryos showing the expression of *NF-YA5*. (H) Histogram showing the MAS5 expression values of *NF-YA5* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.9 CPP (Cysteine rich polycomb like proteins) Transcription Factor Family

CPP family of proteins contain a Cys rich domain, which also shares homology with Cys rich region present in some of the polycomb proteins. There are 8 members in this family, with one member, *TCX2/AT4G14770*, enriched in the sub-epidermal cell type.

AT4G14770/ TCX2

TCX2 is a TSO 1 like CX2 protein. In the radial domain maximum transcriptional activity of *TCX2* is recorded in the xylem / AtHB8 cell type. In the cell layers, highest expression of *TCX2* was detected within the sub-epidermal cell layer. Interestingly, expression of TCX2 is not restricted to any particular cell type. Expression of this TF was captured in the flower primordia using 3 kb promoter fragment in the *pTCX2:H2B-YFP* reporter construct. However, the YFP glow was completely missing from the shoot apical meristem. Examination of the embryos of *pTCX2:H2B-YFP* lines also did not reveal a clear expression of the gene in the early stages of development. A few late torpedo stage embryos showed the expression of the gene in the vasculature of the cotyledon (Figure 2.7 C). However, this does not give a real indication of the true expression pattern of the gene. Further closer

examination of leaf tissue revealed *TCX2* expression in the leaf epidermal cells. Cells undergoing active cell division were positive for nuclear localized YFP fluorescence. The relatively larger cells in the leaf epidermis mature into pavement cells, and these cells did not show expression of *TCX2*. Expression of *TCX2* was observed in the meristemoid mother cells that undergo asymmetric cell division to produce meristemoid and pavement cells. The stomatal guard mother cells display strong expression of *TCX2*, however, in the fully mature guard cells *TCX2* promoter was not active. It is possible to speculate a role for *TCX2* in promoting cell division. To better understand the expression of *TCX2* in shoot, I cloned a 6 kb promoter fragment and dipped in planta. However, the expression of *TCX2* could not be captured in the shoot despite using a 6 kb promoter, suggesting a complex regulation of the same *in-vivo*.

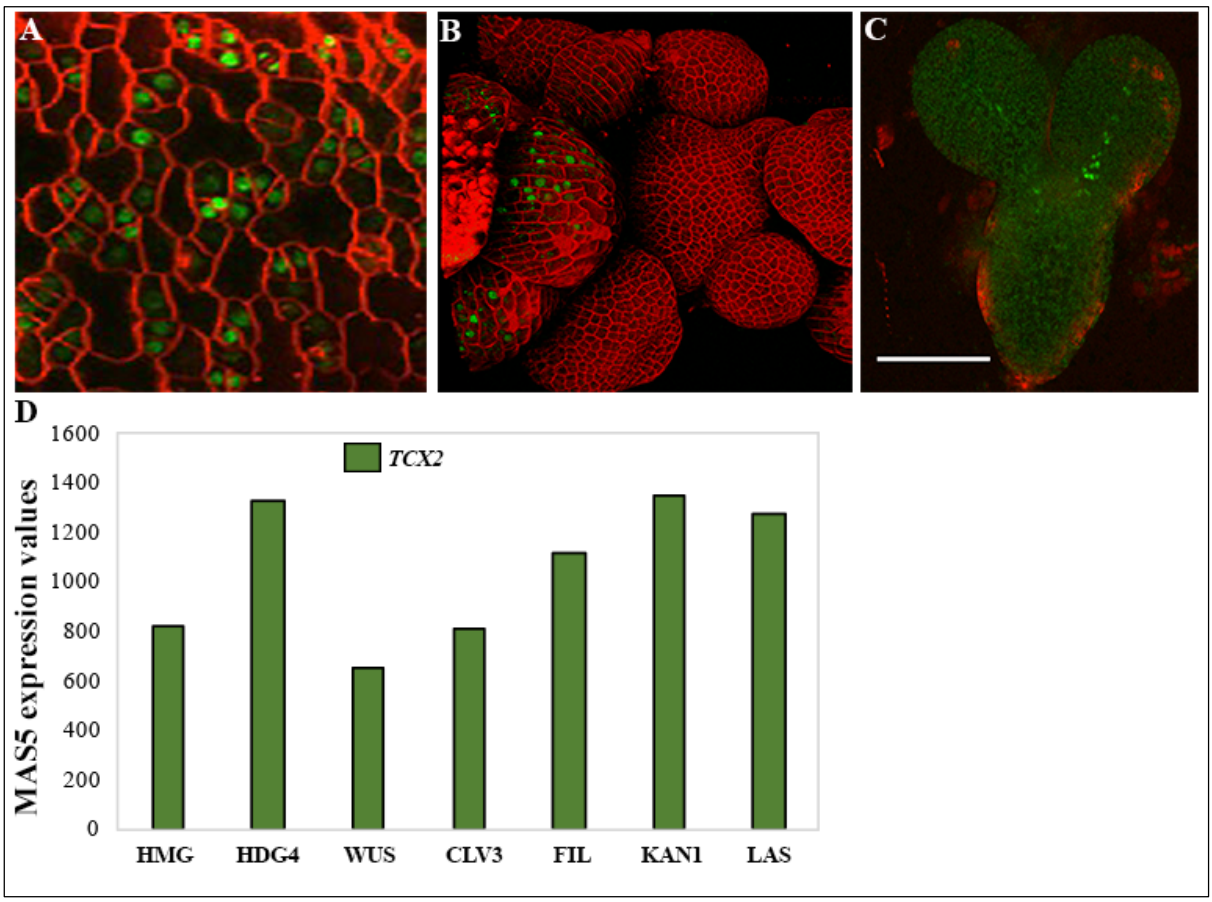


Figure 2.7: Expression pattern of CPP TF family member, *TCX2*. (A) Confocal image of epidermal layer of leaf of the transgenic plant line carrying *pTCX2:H2B-YFP* construct. *TCX2* expression was observed in meristemoid cells of leaf epidermis. (B) 3D reconstructed top view showing the expression of *TCX2* in sepals of mature flower. (C) Histogram showing the MAS5 expression values of *TCX2* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.10 EIL Transcription Factor Family

This family of proteins is involved in ethylene signalling in plants. There are a total of 6 members in this family, with *EIL1/AT2G27050* being enriched in the epidermal layer of the shoot.

ETHYLENE-INSENSITIVE3-LIKE 1 (EIN3 / EIL1) / AT2G27050

EIN3 / EIL1 is a member of the EIL family of transcription factors. *EIN3 / EIL1* is critical for ethylene response in *Arabidopsis*. Similar to *ethylene response 1* (*etr1*) mutant, *eil1/ein3* mutants are insensitive to exogenous ethylene. Genetic and molecular studies have put *EIN3* / *EIL1* down stream of *ETR1* receptor. Interestingly, EIN3 / EIL1 also interacts with PIF1. The cell type specific microarray studies revealed *EIN3/EIL1* enrichment in epidermal cell layer. To investigate the expression of *EIN3/EIL1* in the shoot apex, promoter reporter was made by amplifying the 3 kb regulatory elements upstream of TSS. Five T1 lines were screened for expression, and two lines were found positive for YFP fluorescence. Side view and 3D reconstructed top view of confocal images, reveal the expression of the gene in the epidermal layer of the SAM (Figure 2.8 A, B). Also, the expression of the gene was studied in the embryonic stages of development. The gene was not found to express in the globular and heart stage embryos. However, late torpedo stage embryos showed few positive cells on the adaxial side of the cotyledon (Figure 2.8 D). However, expression was not there in the entire cotyledon but only limited to the basal end of the cotyledon. This does not give a clear indication of the expression of this gene in the early stages, but, its expression is limited to the epidermal layer in the inflorescence meristem.

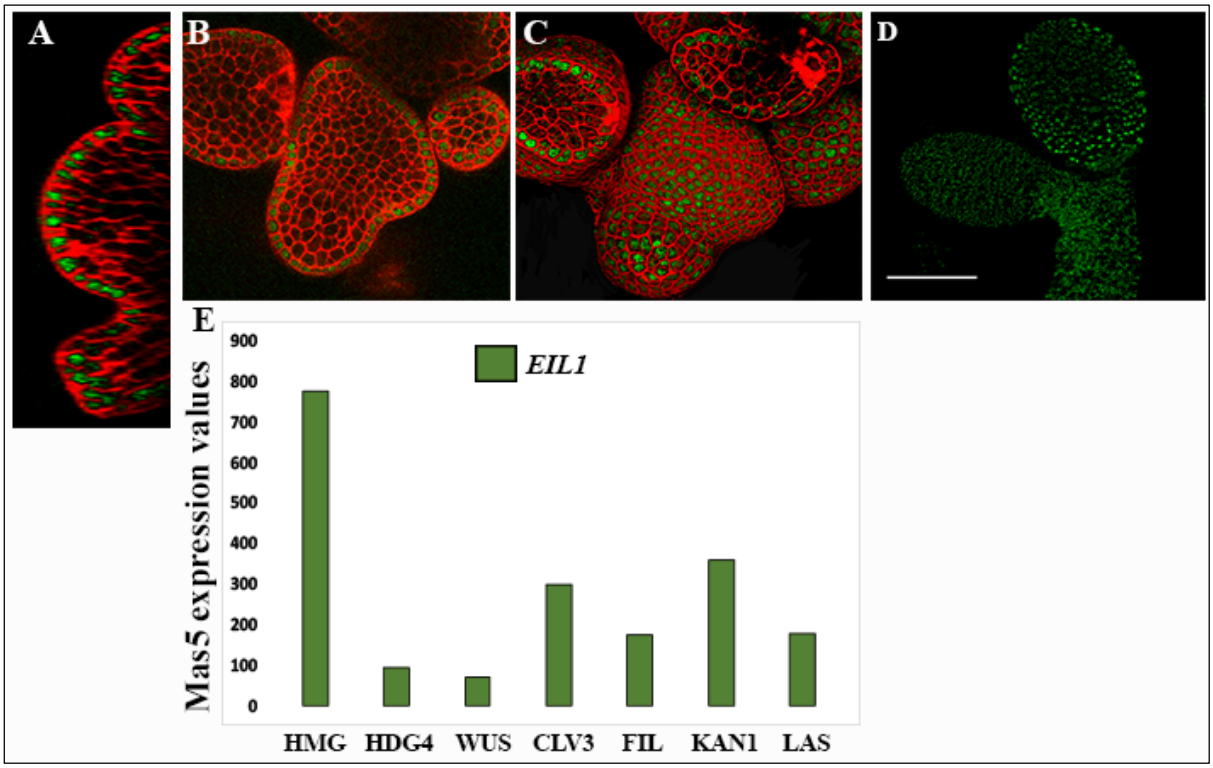


Figure 2.8: Expression pattern of EIL TF family member, *EIL1*. (A, B, C) Side view, transverse section of WT L*er* shoot apex and 3D reconstructed top view, respectively, showing the expression pattern of *pEIL1:H2B-YFP*. (D) Late torpedo stage embryo showing the expression of *EIL1*. (E) Histogram showing the MAS5 expression values of *EIL1* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.11 Homeobox Transcription Factor Family

This family is unique to plants because of the presence of leucine zipper close to the homeodomain. Based on homeodomain sequence homology, this family is sub-divided into four, from HD-ZIP I to HD-ZIP IV. Some proteins also possess additional domains other than the HD, such as homeodomain leucine zipper (HD-ZIP), plant homeodomain with a finger domain (PHD-HD), bell domain (Bell), Zn-finger with homeodomain (ZF-HD), Knotted homeobox (KNOX) and WUSCHEL homeobox containing (WOX). This family has 109 members in total, with 9 of them enriched in the shoot. Genes encoding for these 9 members are, *ATH1/AT4G32980*, *ATHB-2/AT4G16780*, *ATML1/AT4G21750*, *HDG2/AT1G05230*, *HDG4/AT4G17710*, *HDG5/AT5G46880*, *HDG7/AT5G52170*, *HDG12/AT1G17920*, and *PDF2/AT4G04890*.

ARABIDOPSIS thaliana HOMEOBOX GENE 1* (*ATH1*) / *AT4G32980

ATH1 is a member of the BELL sub-family of homeobox TF family. The role of this gene has been suggested in SAM maintenance by physically interacting with other important players of development, such as *STM, BP* and *KNAT6.* Over-expressing *ATH1* is known to severely delay flowering in the C24 accession of *Arabidopsis.* Using the 3 kb promoter, expression of *ATH1* was captured in very early stages of development. It was not found to express in the globular stage embryos. However, in the heart stage embryos, *ATH1* expression overlapped with the *WUS* domain (Figure 2.9 D). In the torpedo stage embryos, expression was found on the adaxial side of the cotyledons (Figure 2.9 E). In the inflorescence meristem, the expression of *ATH1* was recorded in the shoot in the rib-meristem zone, where *WUS* is expressed. Also, *ATH1* expression appears in the epidermal layer of the incipient primordia and extends into deeper layers as the organ primordia grow (Figure 2.9 A, B, C). In the fully mature flowers, *ATH1* has a much broader expression domain. According to the published shoot cell type data, *ATH1* expression was recorded in the epidermal layer, as well as differentiated cell types of the shoot. Taken together, the reporter pattern well corroborates with the cell type specific data.

ARABIDOPSIS thaliana MERISTEM LAYER1* (*AtML1*) / *AT4G21750

AtML1 belongs to the HD-ZIP IV family, and is known to play a pivotal role in maintaining the epidermal cell layer identity. In *Arabidopsis*, *AtML1* acts redundantly with *PDF2*, and maintains the epidermal cell layer identity in early embryo and seedling. Previous studies have shown the expression both these TFs in the epidermal cell layer, which is also supported by the cell type specific microarray study. In accordance with the known function of this gene in controlling epidermal layer specification, it starts to express very early during development. Initially, the expression of *ATML1* is there in all the cells of the globular stage embryo (Figure 2.9 I). However, later on in the heart stage and torpedo stage embryos, as the epidermal layers gets developed, the expression gets limited only to that layer (Figure 2.9 J, K). Confocal microscopy of transgenic lines generated using a 3 kb promoter fragment revealed the fluorescence signal of YFP in epidermal layer of the inflorescence meristem, as previously reported using 3.3 kb promoter element. This gene was also found to be active in the epidermal layers of the flowers, as can be seen from the side view and 3D reconstructed confocal image (Figure 2.9 F, G, H).

HOMEODOMAIN GLABROUS 4* (*HDG4*) / *AT4G17710

HDG4 belongs to HD-ZIP IV family of TFs. Cell type microarray study showed the enrichment of the *HDG4* transcript in the sub-epidermal cell layer (Figure 2.9 L), which is also confirmed by an earlier study by *in situ* hybridization. A 2 kb promoter DNA fragment of *HDG4* was fused transcriptionally and dipped in WT L*er* background. T1 lines were screened for transgene expression. Most lines showed expression of *pHDG4* in the sub-epidermal cell layers of the adult shoot. Analysis of the expression of this gene in the early embryonic stages did not reveal any YFP positive cells, indicating that the L2 layer gets formed later during the development and is absent in the embryonic stages. Also, imaging of the 3 days old seedlings (data not shown), showed very few YFP positive cells, suggesting that even in the seedling stage, there are only a few cells that have L2 like identity.

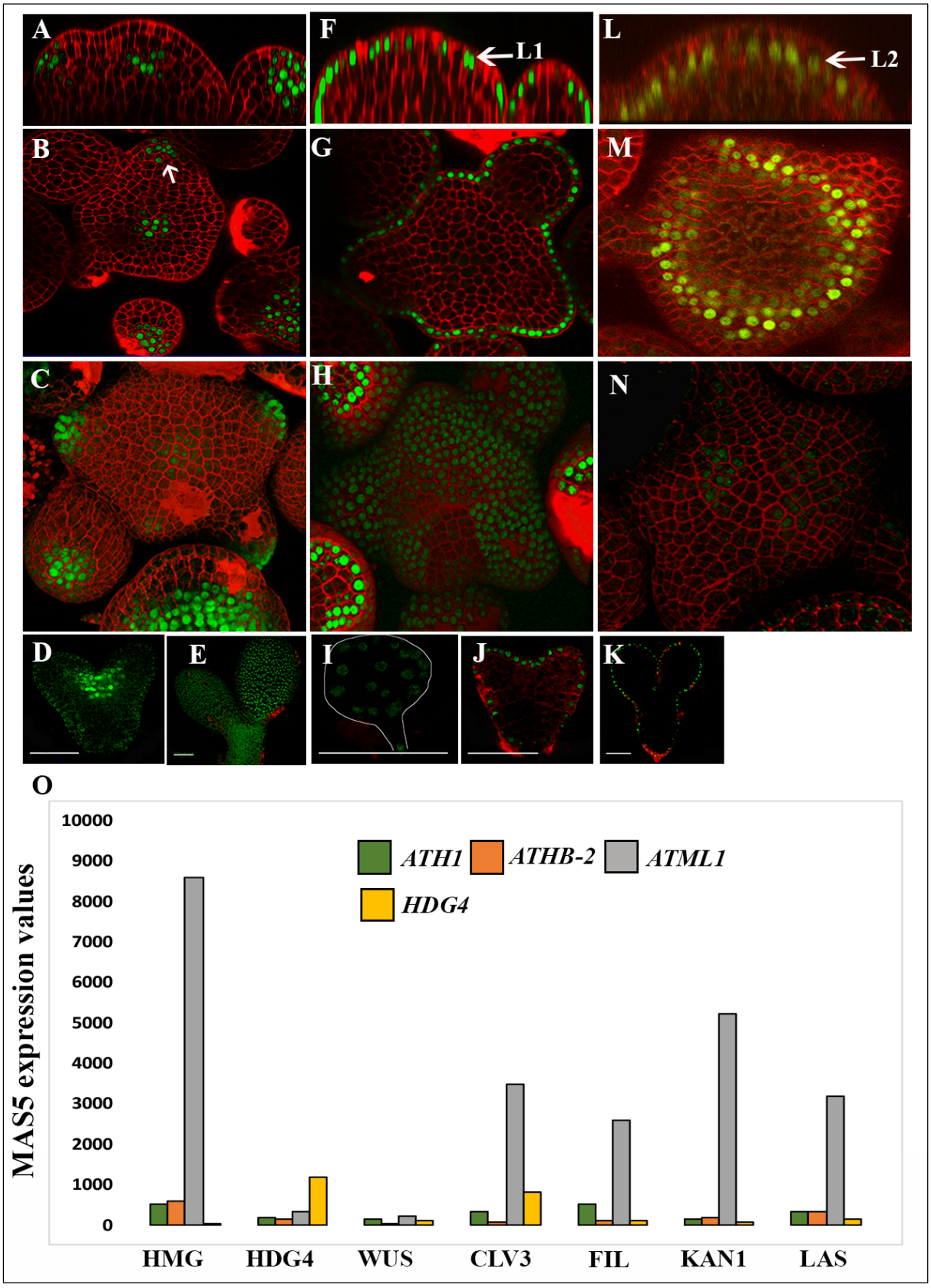


Figure 2.9: Expression pattern of homeobox TF family members. (A, B, C) Side view, transverse section of WT L*er* shoot apex and 3D reconstructed top view, respectively, showing the expression pattern of *pATH1:H2B-YFP*. The white arrow in B, shows the expression of *ATH1* in the epidermal layer of the incipient primordia. (D, E) Heart stage and

torpedo stage embryos showing the expression of *ATH1*. (F, G, H) Side view, transverse section of WT L*er* shoot apex and 3D reconstructed top view, respectively, showing the expression pattern of *pATML1:H2B-YFP*. The white arrow in D, shows the expression of the transgene in L1 layer. (I, J, K) Globular, heart and torpedo stage embryos showing the expression of *ATML1*. (L, M, N) Side view and transverse section of WT L*er* shoot apex showing the expression pattern of *pHDG4:H2B-YFP*. The white arrow in G, shows the expression of *HDG4* in the sub-epidermal layer. (O) Histogram showing the MAS5 expression values of *ATH1, ATHB-2, ATML1* and *HDG4* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

HOMEODOMAIN GLABROUS5* (*HDG5*) / *AT5G46880

HDG5 is a HD-ZIP IV family of TFs. In the gene expression map, its transcript level was found to be enriched in the epidermal layer of the shoot (Figure 2.10 V). A previous study using the GUS reporter showed high expression of *HDG5* in the epidermal cell layer of inflorescence meristem. This pattern gradually attenuated in the deeper layer of SAM (Nakamura et al., 2006). Nevertheless, a 3 kb promoter fragment fused transcriptionally to H2B-YFP showed the expression only in the epidermal layer of shoot apex (Figure 2.10 A, B, C). Of the 11 lines screened for this construct, 10 showed expression in the epidermal cell layer of inflorescence meristem. Confocal imaging of globular, heart and torpedo stage embryos also showed the activity of this reporter in the epidermal cell layer (Figure 2.10 D, E, F).

HOMEODOMAIN GLABROUS 7* (*HDG7*) / *AT5G52170

HDG7 is a member of the HD-ZIP IV family of transcription factors in *Arabidopsis*. In a previous study, GUS based reporter was made using 1.54 kb DNA fragment to see the expression of this gene. However, it failed to show any specific patterns associated to sub-epidermal cell layer as captured using cell type specific microarray study. It is still not understood what function this factor has in shoot development; however, its expression pattern is unique in comparison to other closely related HD-ZIP IV family members (Figure 2.10 G, H, I). A closer look at the cell type specific microarray data revealed that this TF is not highly expressed in sub-epidermal cell layer and the MAS5 values (137) for this TF is very low in comparison to *HDG1* (2126) and *HDG4* (1164). Also, examination of this gene in the embryonic stages did not give any positive result, just like another L2 specific TF, *HDG4*. Indicating that L2 layer does not get formed in the embryonic stages.

HOMEODOMAIN GLABROUS 12* (*HDG12*) / *AT1G17920

HDG12 is a member of the HD-ZIP IV family of TFs. The role of *HDG12* has been shown in literature to be involved in cell differentiation and trichome development (Nakamura et al., 2006). *HDG12* starts to express very early during development. It is expressed throughout the globular stage embryo, but gets restricted to the epidermis from heart stage onwards (Figure 2.10 M, N, O). *HDG12* expression starts showing up from the 4 cell stage and becomes restricted to the protoderm from the globular stage onwards (Horstman et al., 2015). Confocal imaging of transgenic lines carrying the reporter construct, *pHDG12:H2B-YFP*, revealed the expression of this gene in the epidermal cell type of the shoot (Figure 2.10 J, K, L). Other than shoot, *HDG12* is also expressed in the pavement cells of the leaf epidermis. Earlier studies have also shown the high activity of this gene in the epidermal layer of shoot as well as developing embryos.

PROTODERMAL FACTOR 2* (*PDF2*) / *AT4G04890

PDF2 belongs to the HD-ZIP IV TF family. The role of *PDF2* is well established in maintaining the epidermal cell identity right from very early stages of embryo development. Shoot expression map revealed the dominant expression of this gene in the epidermal layer of the shoot (Figure 2.10 P, Q), which is also supported by previous studies. Confocal imaging of transgenic lines generated using a 3 kb promoter fragment, could capture the expression of this gene in the L1 layer of the shoot. Also, 3D reconstructed image shows the expression of this gene in the entire epidermal layer of the shoot and flower primordia (Figure 2.10 R). Other than the shoot, *PDF2* along with its closest homolog, gets expressed in the embryo. Expressing in all the cells of the globular stage embryos, the gene gets limited to the protoderm from heart stage onwards (Figure 2.10 S, T, U).

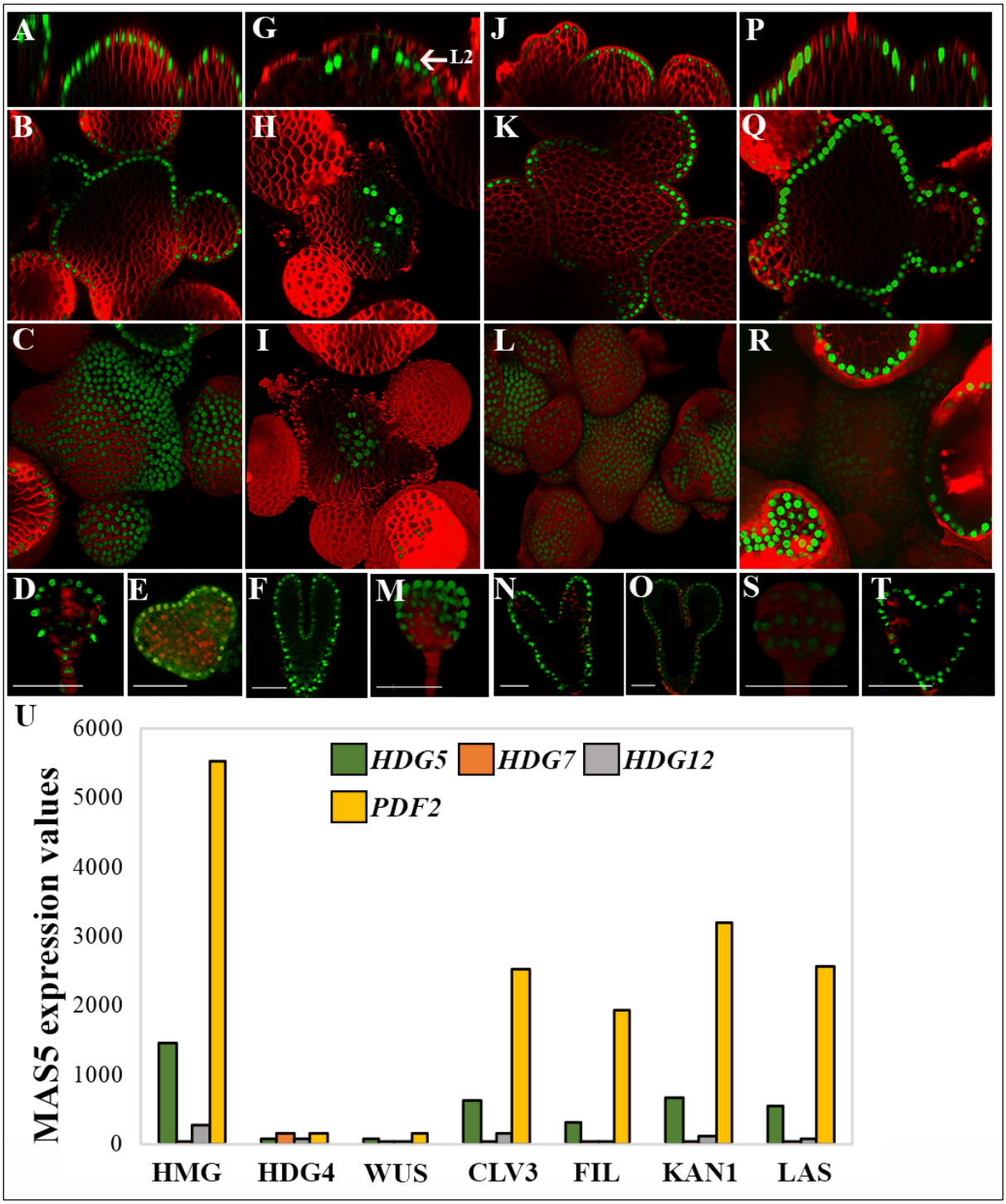


Figure 2.10: Expression pattern of homeobox TF family members. (A, B, C) Side view of WT L*er* shoot apex, transverse section and 3D reconstructed top view, respectively, showing the expression pattern of *pHDG5:H2B-YFP*. (D, E, F) Globular, heart and torpedo stage embryos showing the expression of *HDG5*. (G, H, I) Side view, transverse section of WT L*er* shoot apex and 3D reconstructed top view, respectively, showing the expression pattern of *pHDG7:H2B-YFP*. The white arrow in D indicates the L2 specific expression of *HDG7*. (J, K, L) Side view, transverse section of WT L*er* shoot apex and 3D reconstructed top view, respectively, showing the expression pattern of *pHDG12:H2B-YFP*. (M, N, O) Globular, heart and torpedo stage embryos showing the expression of *HDG12*. (P, Q, R) Side view, transverse section of WT L*er* shoot apex and 3D reconstructed top view, respectively,

showing the expression pattern of *pPDF2:H2B-YFP*. (S, T) Globular and heart stage embryos showing the expression of *PDF2*. (U) Histogram showing the MAS5 expression values of *ATML1, HDG4, HDG5, HDG7, HDG12, PDF2* in L1, L2, L3, CLV3 and differentiated cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.12 MADS Transcription Factor Family

Plants MADS box genes have been long known to play role in floral development. MADS box motif appears to be structurally conserved and is composed of 56 amino acids. MADS box consists of 2 regions, a N-terminal region rich in basic and hydrophilic residues involved in DNA binding and C-terminal region rich in hydrophobic residues. Three members of this family are enriched in the epidermal and sub-epidermal cell types of the shoot, namely *APETALA1/AT1G69120*, *CAULIFLOWER/AT1G26310*, *SEPELLATA2 /AT3G02310*.

APETALA1 (AP1) / AT1G69120

AP1 / *AT1G69120* is a MADS domain transcription factor, that is involved in floral meristem identity and floral organ identity specification. Previous studies have shown the expression of *AP1* in the flower meristem of early Stage-I, II and III, however, it was not clear from these studies whether its relative enrichment is more in the epidermal layer in comparison to sub-epidermal and corpus cell layer as reported by cell type specific microarray studies. Confocal imaging of flower meristem primordia revealed that the domain of *AP1* expression in epidermal layer is broad in comparison to L2 and L3 cell layer (Figure 2.11). This finding supports enrichment of *AP1* in epidermal cell layer, however, it is completely absent from the SAM. In the radial domain, Yadav et al. (2014) reported the maximum expression of this gene in the *CLV3* domain in comparison to other cell layers. *AP1* expression overlaps with *CLV3* expression in the flower meristem, therefore, it is highly enriched in the CZ. Combining the cell type specific microarray studies with reporter analysis could help us in identification of genes that are exclusively part of early flower primordia.

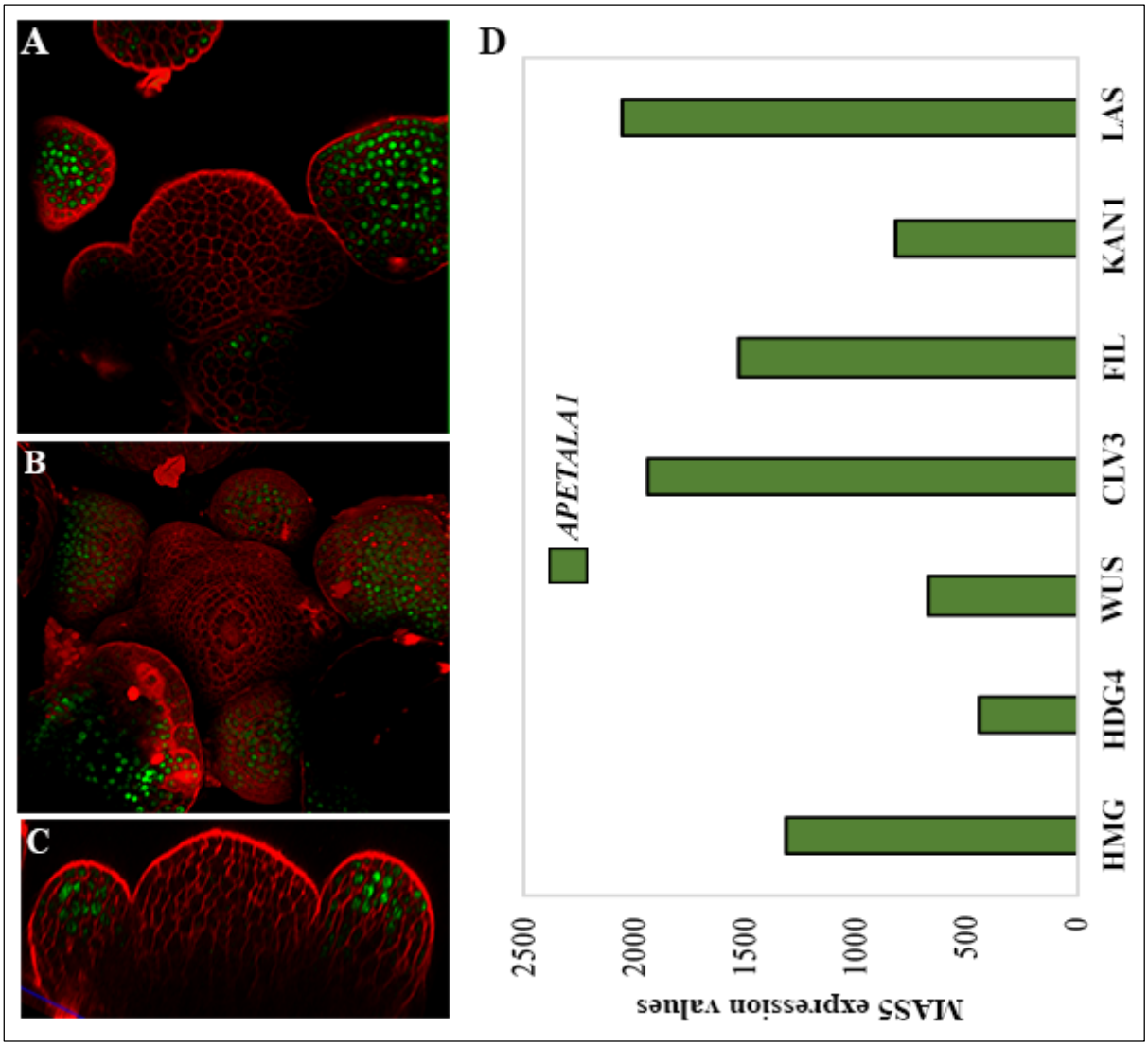


Figure 2.11: Expression pattern of MADS TF family member, *API.* (A) Confocal image of transverse section of WT L*er* shoot apex, showing the expression of *APETALA1* in the flowers. (B) 3D reconstructed top view of shoot of transgenic plant carrying the construct *pAPETALA1:H2B-YFP*. (C) Side view of shoot representing the *AP1* expression in flower primordia. (D) Histogram showing the expression pattern of *APETALA 1* in the L1, L2, L3, CLV3 and peripheral zone cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

2.3.3.13 MYB Transcription Factor Family

One-eighth of the TFs in *Arabidopsis* belong to the MYB super-family. This superfamily consists of 168 members, which are sub-divided into three families, R1R2R3 family, R2R3 family and myb-related. MYB proteins are involved in a variety of functions in plants, ranging from trichome initiation to responding to salt and dehydration stresses. Six members of the MYB family, *MYB4/AT4G38620*, *MYB30/AT3G28910*, *MYB94/AT3G47600*, *MYB96/AT5G62470*, *MYB111/AT5G49330* and *AT5G04760*, are enriched in the epidermal layer of the shoot.

MYB94 / AT3G47600

MYB proteins are divided into four classes based on the number of repeats. In plants, the MYB repeats are expanded up to four times. *MYB94* TF belongs to R2R3-type MYB domain class. The R2R3-MYB class is predominant one in plants. R2R3-MYB TFs have N-terminal DNA binding domain while the activation or repression domain is located at the C terminus of the protein. In *Arabidopsis*, one hundred twenty-six genes encode R2R3-MYB type TFs, 64 genes encode 1R-MYB type TFs, 5 genes encode 3R-MYB TFs and one gene encodes 4R-MYB TF (Stracke et al., 2001). Previous studies have shown *MYB94* expression ubiquitously throughout the plant body except the root tissue. The enrichment of *MYB94* transcript was observed higher in the epidermal peel of stem tissue in comparison to inner cell layer (Lee and Suh, 2015a). Cell type specific microarray studies have shown the expression of *MYB94* exclusively in the epidermal cell layer of shoot apex. Six independent lines were screened to see the *pMYB94:H2B-YFP* expression in the shoot apex, however, none was positive for H2B-YFP.

MYB111 / AT5G49330

MYB111 is a transcription factor belonging to the R2R3 MYB protein family in *Arabidopsis.* The expression of this TF was reported in the epidermal cell layer as well as in the *CLV3* domain (CZ) of the shoot apex by Yadav et al. (2014). However, the 3 kb promoter fragment failed to show any expression. Of the 12 T1 transformants selected on BASTA, none were positive for the YFP expression in the shoot.

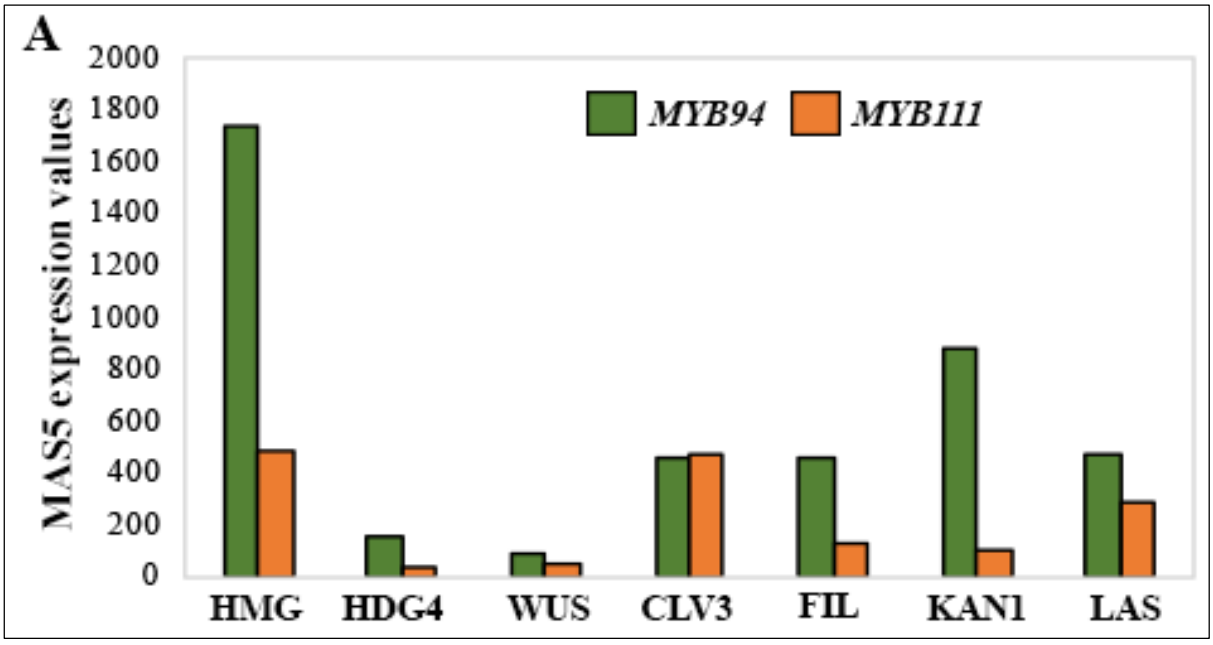


Figure 2.12: Expression pattern of MYB TF family members. (A) Histogram showing the expression values of *MYB94* and *MYB111* in the L1, L2, L3, CLV3 and peripheral zone cell types of the shoot (Yadav et al., 2014).

2.3.3.14 NAC Transcription Factor Family

The members of this family possess a NAC domain, which is present at the N-terminal and is divided into sub-domains from A to E. Members of this family play various roles in plant development. A total of 5 NAC family members are expressed in the epidermal and sub-epidermal layers of the *Arabidopsis* shoot apical meristem, *ANAC010/SND3/AT1G28470*, *ANAC028/AT1G65910*, *ANAC073/AT4G28500*, *ANAC075/AT4G29230* and *ANAC103/AT5G64060*.

ANAC10 / AT1G28470

ANAC010 is a member of the NAC family of TFs. Studies have shown the role of this gene in secondary cell wall biosynthesis (Zhou et al., 2014). The shoot cell type microarray data revealed very low expression value of this gene in the shoot, with highest expression recorded in the sub-epidermal layer, with a MAS5 value of 70. For expression using YFP reporter, ten independent T1 transgenic lines were screened, however, none of them showed expression in the shoot, suggesting that GFP reporters might not be a good choice for such low expressing genes and reporters based on staining such as GUS would be better.

ANAC103 / AT5G64060

ANAC103 is a member of the *Arabidopsis* NAC family of proteins. This gene showed a high expression in the *HDG4* and *CLV3* cell types in the microarray data from Yadav et al. (2014). However, out of the 17 plants screened, none of them showed a positive YFP expression in the shoot apex.

2.3.3.15 WRKY Transcription Factor Family

Members of this family possess one or two WRKY domains of 60 amino-acid residues with a conserved WRKYGQK motif. WRKY domain is used for binding to the DNA and depending on the number of WRKY domains and features of their zinc finger like motifs, WRKY proteins have been divided into different groups. TFs with two WRKY domains fall in-group I, while the one WRKY domain containing fall in-group II. Both group I and group II members have a similar finger motif, C-X_{4-5} -C-X_{22-23} -H-X_{13}-H. WRKY proteins having a distinct finger motif, C_2-HC, fall in-group III. WRKY family consists of 71 members out of them six are enriched in the shoot, *WRKY3/AT2G03340, WRKY11/AT4G31550, WRKY17/AT2G24570, WRKY22/AT4G01250, WRKY25/AT2G30250, WRKY54/AT2G40750*.

WRKY11 / AT4G31550

WRKY11 is a member WRKY family of TFs, unique to the plants. Expression pattern of *WRKY11* was similar to *WRKY22. WRKY11* was also found to be expressed in the peripheral region of the shoot with no expression in the CZ (Figure 2.13 A, B, C). Microarray study by Yadav et al. (2014) reported expression of *WKRY11* and *WRKY17* in the epidermal cell layer (Figure 2.13 T). However, the reported expression does not support the microarray study. Other than the inflorescence meristem, the expression of *WRKY11* was also found in the embryos. It is expressed in all the cells of the globular stage embryo (Figure 2.13 D). In the heart stage and torpedo stage embryos, expression limits to the epidermis and few inner layers of the embryo, with maximum expression in the epidermis (Figure 2.13 E, F)

WRKY17 / AT2G24570

WRKY17 belongs to the group II-D of the WRKY family. *WRKY17* along with *WRKY11* are known to act as negative regulators of basal resistance to *Pseudomonas syringae.* Transgenic lines carrying the *pWRKY17:H2B-YFP* construct were screened for expression and confocal images revealed the expression of this TF in the sub-epidermal layer of the shoot and flowers (Figure 2.13 G, H, I). However, very few cells of the SAM were positive for YFP in

comparison to the flowers. According to the cell type specific microarray data, maximum expression of this gene has been reported in the epidermal layer, followed by differentiated cell type. The expression achieved by using the 3 kb promoter does not match with the cell type data, probably due to lack of sufficient regulatory elements within this region. However, using the 3 kb promoter, expression of this gene was also found in the early stages of development, with expression in the epidermal layer of the globular stage embryos (Figure 2.13 J). With the development of the embryo, expression also expanded to the inner layers, with more expanded expression in the future cotyledons (Figure 2.13 K, L).

WRKY22 / AT4G01250

WRKY22 is a member of WRKY family of TFs. It is one of the largest families of TFs and found solely in plants. In *Arabidopsis*, seventy-two genes encode WRKY like TFs. WRKY proteins contain conserved amino acid sequence WRKYGQK and zinc-finger-like motifs. WRKY TFs can act either as activator or repressors. A role of *WRKY22* is established in dark induced leaf senescence (Zhou et al., 2011). The expression of *pWRKY22:H2B-YFP* transgene was looked under the confocal microscope in the positive lines. It was expressed right from the embryonic stages. In the globular stage embryos, interestingly, the expression is limited only to the basal end of the embryo (Figure 2.13 P). But in the heart and later stage embryos, the expression expanded to the apical end as well and expresses in the epidermal as well as inner cell layers (Figure 2.13 Q, R). However, the expression was very little towards the adaxial end of the future cotyledons. The activity of *pWRKY22* was broad in the SAM. Strong expression of *pWRKY22* in the peripheral zone of the shoot and flower organ primordia was observed, however, in the CZ expression of *pWKRY22* was missing, where *CLV3* is normally expressed (Figure 2.13 M, N, O). In the cell type microarray study Yadav et al. (2014) reported highest expression of this gene in the epidermal layer, and not very strong expression in the other tissue layers (Figure 2.12 T).

WRKY25 / AT2G30250

WRKY25 is a WRKY family TF, which is unique to the plants. The cell type microarray study reported expression of *WRKY25* in the sub-epidermal cell layer. Multiple lines were screened using confocal microscopy to image the expression of *WRKY25* promoter activity. Transgenic plant lines carrying the *pAT2G30250:H2B-YFP* reporter showed only a few H2B-YFP positive cells in the shoot apex (Figure 2.13 S). The expression of *WRKY25* promoter was irregular with respect to reporter expression in the inflorescence meristem as well as the

embryos.

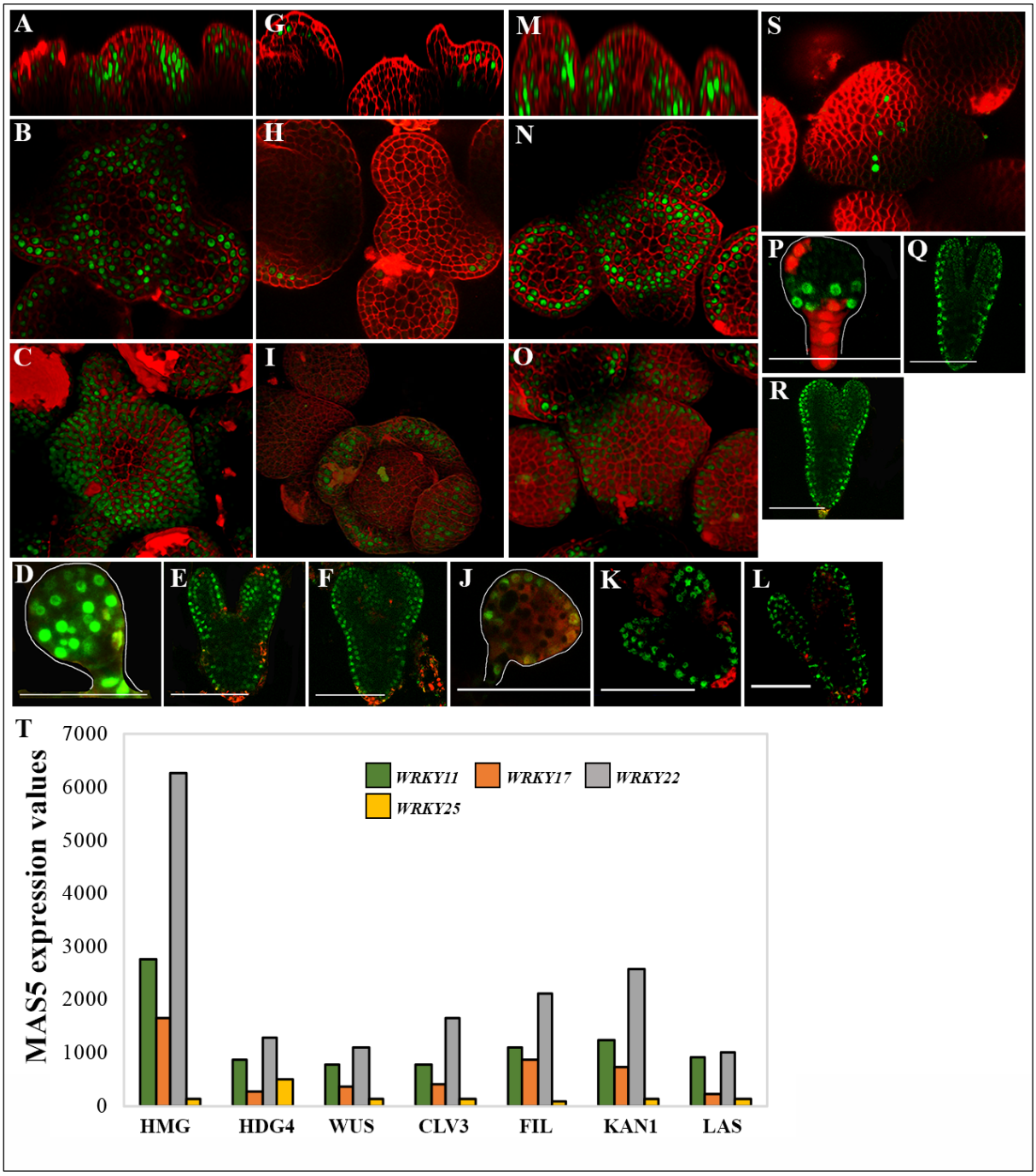


Figure 2.13: Expression patterns of WRKY TF family members. (A, B, C) Confocal image of side view, transverse section of WT L*er* shoot apex and 3D reconstructed top view, respectively, showing the expression pattern of *pWRKY11:H2B-YFP*. (D, E, F) Globular, heart and torpedo stage embryos showing the expression pattern of *WRKY11*. (G, H, I) Side view, transverse section and 3D reconstructed top view of WT L*er* shoot showing the expression of *pWRKY17*. (J, K, L) Globular, heart and torpedo stage embryos showing the expression pattern of *WRKY17*. (M, N, O) Side view, transverse section of WT L*er* shoot apex and 3D reconstructed top view, respectively, showing the expression pattern of *pWRKY22:H2B-YFP*. (P, Q, R) Globular, heart and torpedo stage embryos showing the expression pattern of *WRKY22*. (S) Transverse section of SAM, showing the expression of *pWRKY25:H2B-YFP* in random cells in the shoot. (T) Histogram showing the expression

values of *WRKY11, WRKY17, WRKY22* and *WRKY25* in the L1, L2, L3, CLV3 and peripheral zone cell types of the shoot (Yadav et al., 2014). Cell outlines highlighted by FM4-64 dye (red).

The total number of plant lines screened for each construct and the number of positive plants obtained are mentioned in the table below.

Table 2.2: Total number of plants screened and positive for each construct

Construct name	TF Name	No. of transgenic plants screened	No. of plants that expressed the transgene	Plant line followed
pAT4G25490::H2B-YFP	CBF1	4	2	#31
pAT2G38340::H2B-YFP	DBEB19	6	0	
pAT1G04880::H2B-YFP	HMG15B	20	12	#X
pAT1G63650::H2B-YFP	EGL3	5	5	#6
pAT2G20180::H2B-YFP	PIL5	4	3	#X
pAT2G31730::H2B-YFP	bHLH	19	17	#6
pAT1G49720::H2B-YFP	ABF1	7	4	#22
pAT5G57660::H2B-YFP	COL5	3	0	
pAT2G28810::H2B-YFP	Dof type Zinc Finger	8	0	
pAT1G75710::H2B-YFP	C2H2 Zinc Finger	9	7	#3
pAT4G16610::H2B-YFP	C2H2 Zinc Finger	8	0	
pAT5G54630::H2B-YFP	C2H2 Zinc Finger	15	12	#15
pAT5G61190::H2B-YFP	C2H2 Zinc Finger	9	0	
pAT1G54160::H2B-YFP	NF-YA5	19	17	#19
pAT4G14770::H2B-YFP	TCX2	8	5	#4
pAT2G27050::H2B-YFP	EIL1	12	2	#5
pAT4G21750::H2B-YFP	AtML1	4	3	#2
pAT4G32980::H2B-YFP	ATH1	10	10	#3
pAT4G17710::H2B-YFP	HDG4	17	6	#1
pAT5G46880::H2B-YFP	HDG5	11	10	#7
pAT5G52170::H2B-YFP	HDG7	30	7	#16
pAT1G17920::H2B-YFP	HDG12	7	6	#3
pAT4G04890::H2B-YFP	PDF2	6	5	#7
pAT4G16780::H2B-YFP	ATHB-2	5	0	
pAT1G69120::H2B-YFP	APETALA1	6	6	#5
pAT3G47600::H2B-YFP	MYB94	6	0	
pAT5G49330::H2B-YFP	MYB111	12	0	
pAT1G28470::H2B-YFP	ANAC010	10	0	
pAT5G64060::H2B-YFP	ANAC103	20	0	
pAT4G31550::H2B-YFP	WRKY11	13	11	#7
PAT2G24570::H2B-YFP	WRKY17	9	6	#B
pAT4G01250::H2B-YFP	WRKY22	15	6	#44
pAT2G30250::H2B-YFP	WRKY25	10	3	#11
pAT4G01460::H2B-YFP	bHLH	20	0	
pAT1G29160::H2B-YFP	Dof type Zinc Finger	5	0	
PAT1G64380::H2B-YFP	DREB	6	0	
PAT1G65910::H2B-YFP	ANAC028	1	0	
PAT3G16940::H2B-YFP	Calmodulin	7	1	#33
PAT5G44210::H2B-YFP	ERF9	4	0	
PAT5G14960::H2B-YFP	DEL2	11	0	
PAT1G16070::H2B-YFP	AtTLP8	2	2	#1

PAT4G01460::H2B-YFP	OBP2	4	0	
PAT3G02310::H2B-YFP	SEPALLATA2	2	0	

2.4 Discussion

2.4.1 3 kb upstream regulatory elements are sufficient to drive endogenous expression for a large number of TFs

In this study, I examined the expression pattern of TF genes whose transcripts are enriched ≥ 1.5 fold in the epidermal and sub-epidermal cell types of the shoot. Transcriptional fusions of the gene promoters were made using a 3 kb DNA fragment except *HDG4* (where a 2 kb fragment was used as promoter). In total, 43 promoters were cloned and used for making transcriptional fusion constructs. Transgenic lines were rescued successfully for all the constructs. Multiple lines for each construct were screened to assess the representative tissue / cell type specific expression pattern of promoter reporter. Of the 43 constructs screened, H2B-YFP expression was detected for 25 of them. A quantitative assessment of H2B-YFP expression revealed that the promoter reporter expression closely resembled to that of digital expression pattern recorded using cell type specific microarray study for 23 constructs. In two cases the reporter gave random nuclear localized YFP expression and in the rest, it did not glow at all. Interestingly, the genes, whose promoter resulted in random expression, were responsive to abiotic stresses. One reason could be that these genes get up regulated in response to abiotic stress; therefore, in normal condition either their mRNAs are unstable or below the detection threshold limit. The latter affects when the reporter is based on fluorescent protein because to see the expression of a fluorescent protein a threshold of these molecules is needed in the cell. To test the role of stress on induction and stability, transgenic plant lines carrying the promoter reporter construct when subjected to stress showed the expression in few more cells. For 12 promoters, we did not observe a visible expression pattern in shoot apex although the digital expression pattern predicted presence of their transcripts. The simplest explanation for this could be that the 3 kb DNA fragment upstream of start codon, used for making majority of the promoter reporter constructs, did not have sufficient cis regulatory elements to drive the transcription of that gene in the relevant cell types. However, for 50% of the genes, using the 3 kb promoter fragment did result in the expression of the gene in the expected cells and tissues types. Although, recent studies looking at the Dnase-I hypersensitive sites within the genome of *Arabidopsis*, maize, rice and

tomato found that majority of TF binding sites are present in the 3 kb region upstream of transcription start site (Liu et al., 2017).

For ARID family gene, *ATHMGB15,* native expression in epidermal layer was captured in most of the transgenic lines screened and also well corroborated with the cell type specific microarray data. Four bHLH family TFs were also studied for native expression in the shoot. Out of four, for three of them the YFP glow was captured in the shoot. *AT2G31730* expression was very nicely captured in the peripheral region of the shoot using the 3 kb promoter and also matched the cell type specific data predicted expression. *EGL3* and *PIL5* TFs showed only organ specific expression but absent from the shoot. This might hint at complex regulation of these TFs in the shoot or requirement for enhancer elements, which sit far away from the upstream 3 kb fragments used in the study. *In situ* experiments can be done in the future for these genes to exactly nail down the status of expression of these TFs in the SAM. And for *AT1G01460*, the native expression could not be established, hinting at the need to use a longer promoter fragment or context dependent expression of this gene.

For C2H2 TF family, epidermis specific expression was captured in the shoot apical meristem for two TFs, *pAT1G75710* and *pAT5G54630*. But the expression of both the TFs was missing from the central zone of the shoot. *In situ* hybridization was also attempted for both these TFs to study their expression pattern. Using *in situ,* the expression was detected in the epidermal layer of the shoot and the organ primordia (Harish Kumar and Ram Yadav unpublished data). However, in *in situ,* a significant expression was detected in the central zone, which was very weak or absent in the YFP reporter. This might be because the gene is expressed at low levels in the central zone and could not be detected via YFP due to levels below the threshold. However, this modulating nature of this TF might have some *in vivo* significance and might be getting captured only in the YFP reporter in comparison to *in situ*, where the tissue gets fixed. Also, the null mutant for *AT1G75710* shows defects in organ formation, such as abnormal floral organs and defects in the phyllotactic pattern. Based on its expression pattern and phenotype of the mutant plant, it can be hypothesized that this TF might be playing a crucial role in organ formation in the plant, although, experimental evidences are still required to conclude the function of this gene. Because of the co-expression of *AT1G75710* and *AT5G54630* in similar regions of the shoot and 58% sequence similarity between the two TFs, it might be true, that the two TFs might work in conjunction to carry out their functions. To validate this, a double mutant analysis of these two TFs is

under study.

A CCAAT family member, *NF-YA5* also showed native expression in the epidermal layer of the shoot. Also using GUS reporter, previous studies have reported the expression of *NF-YA5* in the leaf tissues with significant expression in the vascular system of the leaf and also the guard cells. Floral tissues and root vascular system was also found to be positive for GUS staining (Li et al., 2008).

Among homeobox family members, expression could be captured for 8 out of the 9 TFs tested. Except *ATHB-2,* all the other TFs screened, showed positive expression in different domains of the shoot. Studies in the past have also reported the expression of these homeobox TFs in various parts of the plant body using the GUS reporter (Nakamura et al., 2006). Apart from shoot, some of these TFs also start expressing from very early stages of embryo development (Horstman et al., 2015) (Takada and Jürgens, 2007).

MYB family members, *MYB94* and *MYB111* were also studied for endogenous expression pattern in the shoot. However, using a 3 kb promoter fragment did not reveal any shoot specific expression for both these TFs. But the expression of *MYB94* has been reported in literature in other parts of the plant using a GUS reporter. GUS activity was observed in the aerial parts of 7 day old seedling, stem cortex and the epidermis, vasculature of the stem, in leaves and trichomes and sepals, anther and anther filaments (Lee and Suh, 2015b). This expression pattern was observed using a 4.2 kb promoter region, suggesting that the 3 kb promoter used for making the fluorescent reporter construct might not contain sufficient regulatory elements to promote its native expression in the shoot. For *MYB111* also, GUS expression has been reported in literature in the apex of the cotyledons, primary leaves, region of apical meristem, origin of lateral meristem and the root tip (Stracke et al., 2007). This GUS expression was captured using a 1074bp promoter fragment, suggesting that GUS might work as a better reporter as compared to YFP, for some of the TFs.

Out of six WRKY family members enriched in shoot, expression pattern was studied for four of them. For 2 of them, *WRKY11* and *WRKY22,* native expression could be captured. However, two of them, *WRKY25* and *WRKY17* showed mis-expression. For *WRKY25,* patchy expression was captured in the shoot and the expression for *WRKY17* did not match well with the predicted cell type data. Suggesting that the 3 kb promoter region of these genes might

not carry sufficient cis-elements to drive their endogenous expression. The native expression of *WRKY17,* however, could be captured in the peripheral zone of the shoot using *in situ* hybridization.

2.4.2 Epidermis specific transcription factors start to express early on during development

Analysis of transgenic lines, carrying reporter constructs, revealed expression of a large number of TFs from very early during development and TFs show really interesting patterns of expression. For example, *ATML1* and *PDF2,* the TFs important for epidermal layer specification, express in all the cells of the globular stage embryo. However, gets limited to the protoderm from heart stage onwards. Similar observation has been reported in the past for *ATML1* (Takada and Jürgens, 2007). Other TFs, such as *HDG12, HDG5, NF-YA5* also showed expression in the epidermal layer. This suggests that some of the TFs are also required in the early stages of development and continue to express from very early to the reproductive stages of the plant. *PIL5,* a phytochrome interacting factor, is expressed in the hypocotyl region of the embryo, which is in accordance with the known function of the gene in skotomorphogenesis. However, no L2 specific marker was found to be active in the embryonic stages, hinting at the fact that sub-epidermal layer is not yet formed that early during development and comes up later on as the plant grows.

2.4.3 Transcription factor expression gets modulated by environmental cues

Expression pattern in shoot was studied for two AP2/EREBP family members using the YFP reporter. For both the genes, native expression could not be established. *DREB19:H2B-YFP* construct did not show any glow in the shoot, whereas for *CBF1,* the reporter was patchy and observed only in a few cells. *DREB1A / CBF1* is involved in responding to cold temperatures and abscisic acid. Therefore, one possibility to capture its spatiotemporal expression at maximum threshold is by providing suitable stimuli or conditions in which the network of upstream regulator would become fully activated, resulting in its expression. The same was tested by exposing the *CBF1* reporter to cold stress for 48 hrs and it increased the YFP expression to more number of cells. Studying activity of this TF using GUS reporter in the past, also revealed the expression of this TF in other parts of the plant. During the early stages, the expression was restricted to hypocotyl, cotyledons and roots. In the fully developed transgenic lines, no GUS staining was observed under control conditions.

However, when the plants were exposed to cold treatment, GUS staining could be observed in the sepals, siliques and leaves (Novillo et al., 2007).

DREB19 is also a drought responsive gene and the gene might get transcribed only under stress condition, thereby explaining the absence of YFP signal in transgenic lines under normal growth conditions. These genes are good example of the fact that not all TFs function in developmental pathways and a significant number of them might participate in environmental/stress responses. Also, transcriptional activity of *DREB19* using GUS reporter, has shown its expression in the region of the stem from where the leaves emerge and the xylem tissue of the roots. Suggesting, that this gene might be involved in leaf emergence and in regulation of genes involved in uptake of nutrients by xylem tissue (Krishnaswamy et al., 2011). Collectively, these evidences from literature suggest that the YFP expression should possibly also be checked in the seedling stage instead of the adult shoot.

2.4.4 Nine different transcriptional patterns associated with shoot TFs

A large number of reporters analysed in this study indicate nine different transcriptional patterns associated with shoot enriched transcription factors. TFs such as *ATML1, HDG5, PDF2, NF-YA5, HMG, HDG12* are expressed only in the epidermal cell layer of the SAM (Figure 2.14 A). *HDG4* and *HDG7* mark the sub-epidermal cell layer, but the expression of *HDG7* is more restricted as compared to *HDG4,* which is expressed more uniformly throughout the sub-epidermal cell layer (Figure 2.14 B). Interestingly, the activity of two TFs, *AT5G54630* and *AT1G75710* was associated with the epidermal cell layer but was missing from the central region where stem cells reside, revealing a different transcriptional module as compared to the TFs that are more uniformly expressed throughout the L1 (Figure 2.14 D). A few TFs were found to be expressed only in the differentiated cells of the peripheral zone. *AT2G31730, WRKY11,* and *WRKY22* are expressed in the peripheral zone but completely missing from the CZ / *CLV3* domain (Figure 2.14 F). The regulatory module impinging on the promoter of these genes might be having repressive affect in the CZ but in the PZ the upstream regulators might be promoting the expression of these TFs. Another interesting transcriptional pattern observed in the shoot was associated with *ABF1*, a bZIP family TF. This TF expresses in the epidermis and lower layers of the shoot, however absent from the lower layers of the late stage organ primordia (Figure 2.14 H). Hinting at a regulatory module in the promoter of this gene that causes its repression in the cells of the

lower layers with the increasing age of the organs. Opposing to the expression pattern of *ABF1* gene, an interesting expression profile was associated with a few TFs, such as *PIL5* and *EGL3,* which were missing from the shoot but expressed in the epidermal layer of the emerging organs only (Figure 2.14 G). Suggesting, a new set of regulators that are different than the shoot ones. *AP1* is shown to play critical role in the floral meristem and floral organ identity and its expression appears in the early flower primordia at Stage-I. *AP1* is activated by *LEAFY* (*LFY*) in the emerging organ primordia. Therefore, *LFY* acts upstream of *AP1*, and *LFY* expression always appears prior to *AP1* in the flower. Moreover, *LFY* expresses in the shoot periphery and gives the newly emerging lateral organ primordia floral identity. Expression of *TCX2* was also observed in the sepals of late stage-IV primordia suggesting the role of this TF in sepal maturation.

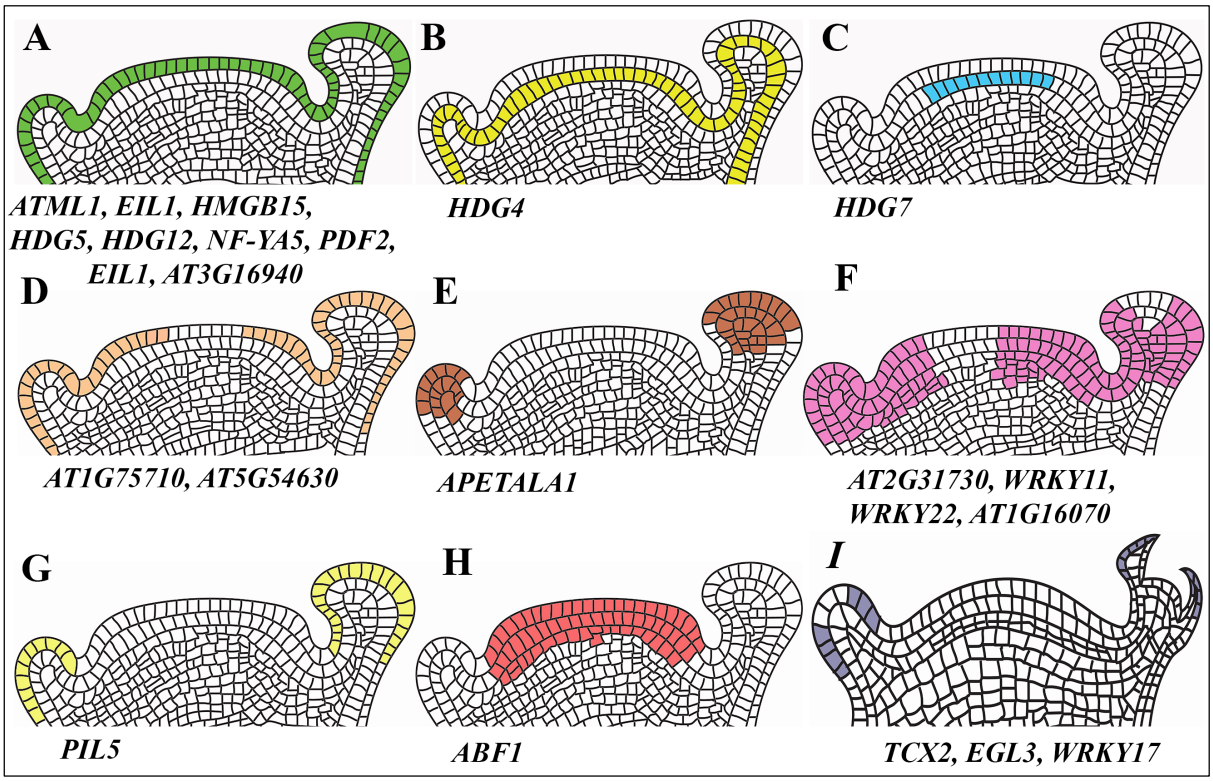


Figure 2.14: Expression patterns associated with shoot transcription factors. Diagrams showing the relative expression profiles associated with some of the shoot enriched transcription factors, observed using a 3 kb promoter sequence. (A) Light green shows epidermis specific expression. (B) Blue shows sub-epidermal layer specific expression. (C) Dark green shows expression in WUS domain of shoot and in multiple layers of flowers. (D) Light brown shows epidermis specific expression missing from the *CLV3* region. (E) Brown shows flower specific expression. (F) Pink shows expression in multiple layers of shoot and flowers, however missing from *CLV3* domain. (G) Yellow represents expression specific to flower epidermis. (H) Light red represents expression in multiple layers only in shoot. (I)

Dark grey represents expression specific to the sepal primordia.

In summary, I investigated the regulation of 43 TFs by making transcriptional reporter (H2B-YFP) fusions. I cloned the 3 kb upstream fragment of each TF including the 5'UTR. These promoter constructs were studied *in planta* by visualizing the endogenous H2B-YFP using confocal microscopy. Of the 43 reporters created 58% (25/43) showed expression in planta. By considering the visualized expression patterns, we concluded nine major expression patterns in the shoot apical meristem and early flowers. For a few selected TFs, where reporter did not show any expression, we conducted *in situ* hybridization studies on the tissue sections. I found the transcripts of these TFs as predicted by microarray studies. This study provides further validation of the cell population microarray data where the cells were sorted for recording their transcripts at high spatial resolution. H2B-YFP based reporters discovered the regulation of 22 TFs at cellular level resolution. This resource is valuable and would help us in validating the TF gene regulatory networks in future.

2.5 References

Arabidopsis, T.G.I. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408,* 796-815.

Birnbaum, K. (2003). A Gene Expression Map of the Arabidopsis Root. Science *302,* 1956-1960.

Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science *318,* 801-806.

Casson, S., Spencer, M., Walker, K., and Lindsey, K. (2005). Laser capture microdissection for the analysis of gene expression during embryogenesis of Arabidopsis. Plant Journal *42*, 111–123.

Chory, J., Ecker, J.R., Briggs, S., Caboche, M., Coruzzi, G.M., Cook, D., Dangl, J., Grant, S., Guerinot, M. Lou, Henikoff, S., et al. (2010). National Science Foundation-Sponsored Workshop Report : " The 2010 Project " Functional Genomics and the Virtual Plant . A Blueprint for Understanding How Plants Are Built and How to Improve Them. Plant Physiology *123*, 423–425.

Clough, S.J., and Bent, A.F. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant Journal *16*, 735–743.

Dinneny, J., Long, T., Wang, J., Jung, J., Mace, D., Pointer, S., Barron, C., Brady, S., Schiefelbein, J., and Benfey, P. (2008). Cell Identity Mediates the Response of Arabidopsis Roots to Abiotic Stress. Science *320,* 942-945.

Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science *280,* 446-450.

Hong, R.L. (2003). Regulatory Elements of the Floral Homeotic Gene AGAMOUS Identified by Phylogenetic Footprinting and Shadowing. THE PLANT CELL ONLINE *15,* 1296-1309.

Horstman, A., Fukuoka, H., Muino, J.M., Nitsch, L., Guo, C., Passarinho, P., Sanchez-Perez, G., Immink, R., Angenent, G., and Boutilier, K. (2015). AIL and HDG proteins act antagonistically to control cell proliferation. Development *142*, 454–464.

Initiative, T.A.G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. The Arabidopsis Genome Initiative. Nature *408,* 796-815.

Kidokoro, S., Yoneda, K., Takasaki, H., Takahashi, F., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2017). Different Cold-Signaling Pathways Function in the Responses to Rapid and Gradual Decreases in Temperature. The Plant Cell *29,* 760-774.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress

expression data set: Protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant Journal *50,* 347-363.

Kilian, J., Peschke, F., Berendzen, K.W., Harter, K., and Wanke, D. (2012). Prerequisites, performance and profits of transcriptional profiling the abiotic stress response. Biochimica Biophysica Acta - Gene Regulatory Mechanisms *1819,* 166-175.

Koncz, C., and Schell, J. (1986). The promoter of TI-DNA gene5 controls the tissue-specific expression of chimeric genes carried by a novel type of Agrobacterium binary vector. Molecular and General Genetics *204*, 383–396.

Krishnaswamy, S., Verma, S., Rahman, M.H., and Kav, N.N. V (2011). Functional characterization of four APETALA2-family genes (RAP2.6, RAP2.6L, DREB19 and DREB26) in Arabidopsis. Plant Molecular Biology *75,* 107-127.

Larkin, J., Oppenheimer, D., Pollock, S., and Marks, M. (1993). Arabidopsis GLABROUS1 Gene Requires Downstream Sequences for Function. The Plant Cell *5,* 1739.

Laubinger, S., Zeller, G., Henz, S.R., Sachsenberg, T., Widmer, C.K., Naouar, N., Vuylsteke, M., Schölkopf, B., Rätsch, G., and Weigel, D. (2008). At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis thaliana. Genome Biology *9.*

Lee, B. -h. (2005). The Arabidopsis Cold-Responsive Transcriptome and Its Regulation by ICE1. Plant Cell Online *17*, 3155–3175.

Lee, S.B., and Suh, M.C. (2015a). Cuticular wax biosynthesis is up-regulated by the MYB94 transcription factor in arabidopsis. Plant Cell Physiology *56,* 48-60.

Li, H.M., Altschmied, L., and Chory, J. (1994). Arabidopsis mutants define downstream branches in the phototransduction pathway. Genes and Development *8,* 339-349.

Li, W.-X., Oono, Y., Zhu, J., He, X.-J., Wu, J.-M., Iida, K., Lu, X.-Y., Cui, X., Jin, H., and Zhu, J.-K. (2008). The Arabidopsis NFYA5 Transcription Factor Is Regulated Transcriptionally and Posttranscriptionally to Promote Drought Resistance. PLANT CELL ONLINE *20,* 2238-2251.

Liu, Y., Zhang, W., Zhang, K., You, Q., Yan, H., Jiao, Y., Jiang, J., Xu, W., and Su, Z. (2017). Genome-wide mapping of DNase i hypersensitive sites reveals chromatin accessibility changes in Arabidopsis euchromatin and heterochromatin regions under extended darkness. Scientific Reports *7*, 1–15.

Long, J.A., Moan, E.I., Medford, J.I., and Barton, M.K. (1996). A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis. Nature *379,* 66-69.

Maher, K.A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D., Zumstein, K., Woodhouse, M., Bubb, K.L., Dorrity, M.W., et al. (2017). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. Plant Cell *30,* 15-36.

Manners, J.M., Penninckx, I.A.M.A., Vermaere, K., Kazan, K., Brown, R.L., Morgan, A., Maclean, D.J., Curtis, M.D., Cammue, B.P.A., and Broekaert, W.F. (1998). The promoter of the plant defensin gene PDF1.2 from Arabidopsis is systemically activated by fungal pathogens and responds to methyl jasmonate but not to salicylic acid. Plant Molecular Biology *38,* 1071-1080.

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., and Takahashi, T. (2006). Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in Arabidopsis1[W]. Plant Physiology *141*, 1363–1375.

Novillo, F., Medina, J., and Salinas, J. (2007). Arabidopsis CBF1 and CBF3 have a different function than CBF2 in cold acclimation and define different gene classes in the CBF regulon. Proceedings of the National Academy of Sciences *104*, 21002–21007.

Peiter, E., Montanini, B., Gobert, A., Pedas, P., Husted, S., Maathuis, F.J.M., Blaudez, D., Chalot, M., and Sanders, D. (2007). A secretory pathway-localized cation diffusion facilitator confers plant manganese tolerance. Proceedings of the National Academy of Sciences *104,* 8532-8537.

Santamaria, M., Thomson, C.J., Read, N.D., and Loake, G.J. (2001). The promoter of a basic PR1-like gene, AtPRB1, from Arabidopsis establishes an organ-specific expression pattern and responsiveness to ethylene and methyl jasmonate. Plant Molecular Biology *47,* 641-652.

Springer, P.S. (2000). Gene traps: tools for plant development and genomics. Plant Cell *12*, 1007–1020.

Stolc, V., Samanta, M.P., Tongprasit, W., Sethi, H., Liang, S., Nelson, D.C., Hegeman, A., Nelson, C., Rancour, D., Bednarek, S., et al. (2005). Identification of transcribed sequences in Arabidopsis thaliana by using high-resolution genome tiling arrays. Proceedings of the National Academy of Sciences *102,* 4453-4458.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. Current Opinion in Plant Biology *4,* 447-456.

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., and Weisshaar, B. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling. Plant Journal *50*, 660–677.

Sundaresan, V.A., Springer, P., Volpe, T., Haward, S., Jones, J.D.G., Dean, C., and Martienssen, R. (1995). Patterns of gene action revealed by enhancer trap and gene trap transposable elements. Genes and Development *9*, 1797–1810.

Takada, S., and Jürgens, G. (2007). Transcriptional regulation of epidermal cell fate in the Arabidopsis embryo. Development *134*, 1141–1150.

Thompson, W. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research *8,* 4321-4326.

Wanke, D., Berendzen, K.W., Kilian, J., and Harter, K. (2010). Insights into the Arabidopsis

Abiotic Stress Response from the AtGenExpress Expression Profile Dataset. In Plant Stress Biology: From Genomics to Systems Biology, 197-225.

Yadav, R.K., Girke, T., Pasala, S., Xie, M., and Reddy, G.V. (2009). Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. *106*, 1–6.

Yadav, R.K., Tavakkoli, M., Xie, M., Girke, T., and Reddy, G.V. (2014). A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. Development *141*, 2735–2744.

Yamada, K., Lim, J., Dale, J., Chen, H., Shinn, P., Palm, C.J., Al., E., and Ecker, J.R. (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome. Science *302,* 842-846.

Zhou, J., Zhong, R., and Ye, Z.H. (2014). Arabidopsis NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels. PLoS One *9.*

Zhou, X., Jiang, Y., and Yu, D. (2011). WRKY22 transcription factor mediates dark-induced leaf senescence in Arabidopsis. Molecules and Cells *31,* 303-313.

CHAPTER 3

Gene Centered Protein-DNA Regulatory Network For Epidermal and Sub-Epidermal Cell Type Enriched Transcription Factors

3.1 Introduction

Plants need to adapt to diverse environmental conditions, and at the same time has to coordinate their development and timing of reproduction. The temperatures to which the plant is exposed, amount of light, and other factors from the environment are also perceived and integrated as signals, which lead to various developmental outcomes. Thus, developmental outcomes are a result of tightly regulated gene expression. Techniques like microarray, RNAseq and high throughput sequencing technologies help us to generate large amount of transcriptomics data which reveal the transcriptional signatures during various environmental conditions and developmental stages. To digitize the development, it is important to connect the data points obtained from the large-scale studies into formal regulatory connections that constitutes the body of a plant. Therefore, the next important step in biology is to make sense out of these big data sets. One possible way of making sense out of these data sets, is to make use of network based approaches that help map genetic or physical interactions (Brady et al., 2011a; Dreze et al., 2011; Jones et al., 2014; Mukhtar et al., 2011). In gene networks, genes are represented as nodes and the interaction between the genes is depicted as edges. A variety of meaningful interactions can be described by these edges, such as interactions between proteins and interaction between transcription factors and regulatory regions in their downstream target DNA. Information in various biological contexts can be incorporated in these gene regulatory networks and can be very helpful in understanding the plant processes such as growth and development.

According to the central dogma of molecular biology a gene first gets transcribed into RNA molecule, which then translates into the protein. Among the RNAs, mRNA molecule is the key. The mRNA is read by the ribosomal machinery, which is made up of rRNA and proteins, and gets translated into protein. Proteins play major biological functions in an organism. What will happen to a cell if protein molecules are present in the cell whole day? Many biological activities within the cell are carried out during day and night. However, few activities are precisely regulated, in other words they are carried out either during the day and night time or specific development stage only. Understanding the mechanism behind regulation of genes that are expressed in specific condition or time of the day will elucidate the mechanism of growth and development in an organism. Transcription factors (TFs) constitute an important part of this machinery because TFs help in reading the information from DNA. A TF can act either alone as a monomer or form dimer with itself as well as with other TFs. TFs can also recruit co-activators, which assist in activating transcription,

however, when the same TF recruit a co-repressor it inhibits transcription of a target gene. Nowadays, gene expression data are produced routinely in many laboratories around the globe to understand not only the role of external and internal signals but also to understand the role of developmental regulators. Invention of high throughput approaches help us in building predictive networks based on the co-expression of targets and upstream regulator, which can be tested *in vivo*. However, from the recent studies it has become clear that many predictive networks built based on the expression data alone do not holds true in many instances when tested in vivo. Therefore, more direct approaches are needed for identifying binding of TFs to their cognate promoters. Experimental techniques such as mapping of DNase I hypersensitive sites, Chromatin immunoprecipitation coupled with sequencing (ChIP-Seq) and heterologous systems such as yeast-one-hybrid (Y1H) are a few examples whereby one can map physical interactions between transcription factors and their targets promoters.

Transcription factor centered approach

This approach focuses mainly on downstream targets of a TF. Generally, a TF is chosen, if it is known to be a master regulator of biological relevance then the mutant would show a phenotype. Understanding its interaction with downstream targets could help us to elucidate its function better, and also could assist in dissecting its role further in regulating distinct aspects of network that are interdependent but hard to discern based on the genetic evidence. In such cases high-quality antibodies will be employed by experimenter to determine the binding of the TF to target gene promoters.

Chromatin immunoprecipitation (ChIP) is the most commonly used method for a transcription factor in a protein-centered approach for identifying targets in vivo. In this technique, proteins get cross-linked with DNA when they are fixed in formaldehyde. Nuclei are isolated from the tissue, and sheared by sonication. Then, the TF protein of interest is immuno-precipitated using anti-TF antibody. The DNA bound to the TF-protein will be de-cross-linked, and will be used for making the library, which will be sequenced using high throughput sequencing methods. The DNA sequence reads will be analyzed and aligned to the genome. This will give an idea about genome wide distribution and complexity of TF target sites. The quality of binding sites and their relevance would depend upon quality of antibody, and the skills of experimenter. Often it has been seen that the antibodies are not

specific to the TF of interest despite spending time and money. Recent studies also suggest the quality of ChIP-seq data will also depend upon the abundance of TF in the given tissue. TFs that are expressed widely and more in quantity give high quality data. However, TFs that are expressed in few cell types or lowly generally are not good to begin with. A major problem faced by plant researchers adopting ChIP as method of choice is unavailability of high quality antibodies against majority of the transcription factors. To overcome these limitations, gene centered approaches are being used.

Gene centered approach

This approach focuses on target genes in an attempt to identify their upstream regulators. One of the most commonly used in vitro technique is, yeast one hybrid (Y1H). Y1H can be used to determine, which TF binds to the DNA segment of interest. With the availability of complete transcription factor library of *Arabidopsis* and maize, high throughput Y1H studies can be planned to find out the potential targets at genome scale (Burdo et al., 2014; Gaudinier et al., 2011; Petricka et al., 2012).

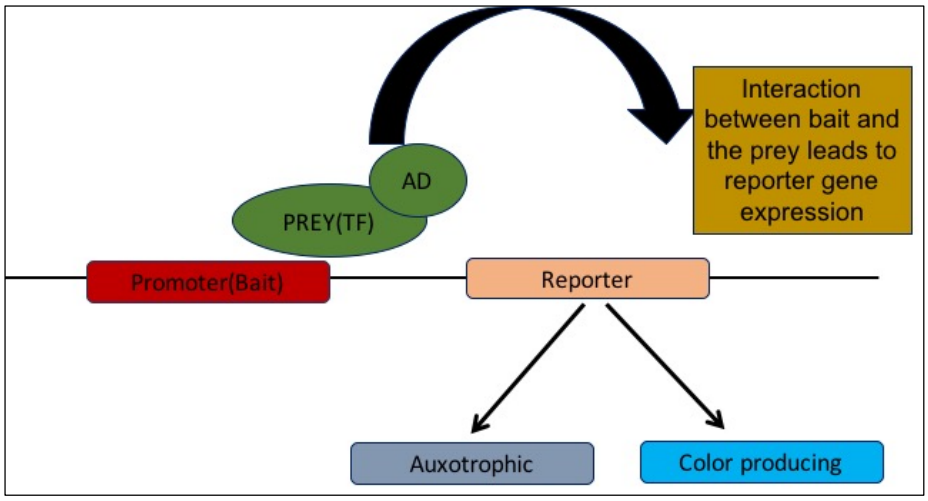


Figure 3.1: Schematic representation of Y1H assay. Promoter is cloned upstream of the reporter gene. Reporter can be either auxotrophic or color producing. Transcription factor protein is translationally fused to the activation domain of GAL4-AD. Interaction between the prey protein and promoter bait leads to the activation of the reporter.

A large number of research groups exploited this method for mapping gene regulatory network of TFs, not only in *Arabidopsis* but also in *C. elegans* and *Drosophila.* Taylor-Teeples et al (2015) described in detail the network for secondary cell wall biosynthesis and xylem cell specification in *Arabidopsis* root. Similarly, a protein interaction network has been studied by Deplancke et al (2006) in *C. elegans*. Y1H based mapping is made easy due to availability of high throughput devices. The identified interactions can be validated in *planta* in leaf mesophyll protoplast transient assays. Information generated through experiments when integrated, results into functional module that might be relevant for the cell and tissue specialization and for their function.

Gene regulatory networks and their significance

Gene expression datasets provide information only about the abundance of mRNAs and not about its regulation (Gaudinier and Brady, 2016). A network derived from gene expression data will be incomplete until all the edges connecting the nodes get meaningful interpretation. Contrary to this, gene regulatory networks take into account physical interactions between upstream regulators and their cognate targets. This can be gene-gene interaction or protein-protein interaction (de Matos Simoes et al., 2013). A gene regulatory network can serve as a 'blueprint' of molecular interactions. When such blueprints are generated for all the genes of an organism and integrated with gene expression data, it results into a comprehensive network. This really becomes powerful to explain the biological cross talk among the different gene products and molecular pathways. For example, a protein-DNA interaction network involves upstream regulator, such as, a transcription factor that binds to non-coding DNA elements and causes either activation or repression of downstream target gene. The physical information derived from the interaction when integrated with the expression data, explains the rationale behind the regulation.

Other than providing physical interaction map, gene regulatory networks can also be used for functional characterization of genes or TFs by exploiting various features of the networks. For example, if one or more TFs share common downstream targets, there is high likelihood that these TFs might be involved in similar pathways or functions. The In-degree (number of TFs binding to a particular promoter) and Out-degree (number of promoters a TF binds to) distribution in a network can be significant in making testable hypothesis or assigning functions to a gene. Also, the master regulators can be deciphered in gene regulatory networks by identifying the TF that binds to maximum number of targets.

If an increase in the protein level of a TF leads to increase in the mRNA level of downstream target gene it suggests an activating relationship, and vice-versa is possible too (Fuxman Bass et al., 2016). To confirm the above described regulatory relationship, one needs an ectopic expression line. Contrary, it is possible to test this hierarchical relationship in a loss of function mutant background of upstream regulator. Moreover, one could use this information to probe further binding sites in the target promoter. For example, if an upstream novel TF binds to multiple downstream target genes, the promoters of these genes can be aligned, and conserved *cis*-regulatory modules can be inferred (Fuxman Bass et al., 2016). However, for such functional characterization, it is important that gene regulatory networks be made at large scale, including all the relevant TFs and the downstream genes.

3.2 Materials and Methods

3.2.1 Molecular biology techniques

3.2.1.1 TF promoter cloning to make the bait plasmid

Genomic DNA was isolated from WT L*er* ecotype plants using CTAB method (Thompson, 1980). GATEWAY compatible attB4 and attB1R tail containing primers were designed to amplify the 3000 bp region of promoter upstream of translational start site. PCR products were recombined with the pDONR P4-P1R vector using BP clonase II. The resulting pENTR clone containing the promoter fragment was recombined with pMW2 (Y1H vector containing the *HIS3* reporter gene) and pMW3 (Y1H vector having the *LacZ* reporter gene), respectively, using LR clonase (Deplancke et al., 2006a). Integration of both *HIS*3 and *LacZ* reporters were confirmed by PCR and verified by sequencing. pMW2 and pMW3 clones resulting from the LR reaction were preferably linearized with following restriction enzymes; *Xho*I, *Afl*II, *Nsi*I, *BseR*I, *Nco*I, *Apa*I and *Stu*I, respectively. The linearized vectors were transformed into Ym4271 strain of yeast. Transformants were selected on yeast minimal medium lacking histidine and uracil, respectively. Integration of bait DNA within the yeast genome was confirmed by PCR.

3.2.1.2 TF open reading frame cloning

Total RNA was isolated from shoot apices of *ap1-1; cal1-1* mutant plants using QIAGEN RNeasy kit. First strand cDNA synthesis was carried out by Superscript III First Strand cDNA Synthesis kit from Invitrogen. Forward primers containing a stretch of four

nucleotides (CACC) towards the 5' end were designed to amplify the full-length CDS. The 5' CACC enables directional cloning of the insert in the pENTR/D/TOPO vector. The cloned CDS of selected TFs were sequenced fully to confirm the integrity of insert. A number of cDNA clones were obtained either from the Arabidopsis Biological Resource Centre (ABRC) Ohio or from RIKEN Japan (Lee et al., 2006). The CDS insert of TFs cloned in to pENTR vector was sub cloned into pDEST-AD 2μ destination vector by LR Clonase to make the preys. The clones were confirmed by sequencing. pDEST-AD 2μ clones were transformed into Yα1867 yeast strain. Transformants were selected on minimal media lacking tryptophan amino acid.

3.2.1.3 RNA extraction

For RNA isolation, 20mg tissue was harvested in RNase free centrifuge tubes. The tissue was then grinded in liquid N_2 and lysis buffer was added to it. The mix was then centrifuged to settle the debris and supernatant was loaded onto the spin columns provided with the kit. Several washes of wash buffer were given as per the manufactures' instruction. Final elution was done in 30μl RNase free water. ReliaPrep RNA miniprep kit from Promega was used for RNA isolation. Manufacturer's instructions were followed at each step.

3.2.1.4 cDNA synthesis

RNA amounts were equalized and reverse transcription was carried out using iScript super mix from Bio-Rad. 2μg RNA was used for synthesizing cDNA. cDNA samples were diluted prior to setting up RT-PCR reactions.

3.2.1.5 qRT-PCR from seedlings

RNA was isolated from 5 days old WT and homozygous mutant or over-expression seedlings, using the ReliaPrep RNA kit, and cDNA synthesized using the iScript super mix kit from BioRad. cDNA was synthesized from 2μg of RNA. Two biological replicates across three technical replicates each of mutant and WT Col-0 were used for each experiment. Primer efficiency was calculated for each primer pair and values represented by the efficiency corrected quantification (ECQ) model. Q-PCR experiments were carried out using the BioRad CFX96 TM Real-Time machine. PP2A was used as an internal control in these experiments. Primer sequences used for qRT-PCR validation are described in text.

3.2.1.6 Preparation of yeast competent cells

For yeast competent cells preparation, single yeast colony was picked to setup primary culture in 20 ml YAPD media and grown overnight at 30^{o}C. Next day, the volume of primary inoculum was adjusted in such a way that in the 200 ml secondary culture the inoculum should rise up to 0.2 OD. The culture was incubated at 30^{o}C for 4-5 hours, until the OD reaches between 0.6-1.0. The cells were spin down at 4000 rpm for 5 min at room temperature. The pellet was washed with 100 ml distilled water and cells were pelleted again at 4000 rpm for 5 min at room temperature. Next, the pellet was dissolved in 1/10th volume of SORB (100mM LiOAc, 10mM Tris HCl pH 8.0, 1mM EDTA pH 8.0, 1mM Sorbitol). The cells were spin down at 4000 rpm for 5 min and dissolved in 1440µl SORB and 160µl salmon sperm DNA. Re-suspended cells were aliquoted into 15µl volume and stored in -80^{o}C.

3.2.1.7 Yeast transformations

For yeast transformation 3µl plasmid DNA was added into 15µl aliquot of competent cells, along with 40% PEG. PEG was added up to the six times of total volume of the cells and plasmid. The mix was then incubated at 30^{o}C for 30 min, followed by incubation at 42^{o}C for 30 min. The cells were then kept on ice for 5 min and centrifuged at 4000 rpm for 5 min. The pellet was dissolved in water and plated on respective drop out media plates.

3.2.1.8 Yeast plasmid isolation/shuttle prep

For yeast shuttle preparation, 1.5 ml culture was taken and centrifuged at 5000 rpm for 5 min. The pellet was dissolved in 400 µl lysis buffer (2% Triton X-100, 10% SDS, 5M NaCl, 1M Tris pH 8.0, 0.5M EDTA) . Following this, 400 µl Phenol chloroform isoamyl alcohol (PCI) was added and also glass beads were added till the lower aqueous meniscus. The cells were lysed in thermomixer at maximum frequency for 2 min. Then 400 µl water was added and vortexed. Culture was centrifuged for 15 min at maximum speed at 4^{o}C. To the supernatant, 400 µl isopropanol was added and centrifuged at maximum speed for 15 min at 4^{o}C. The pellet was then washed with 50 µl, 70% ethanol and again spin down at maximum speed for 15 min. The pellet was then dried thoroughly and dissolved in 20 µl water.

3.2.2 Imaging

Nikon D-5100 camera was used for taking pictures of yeast plates. Adobe Photoshop software (Adobe systems Inc.) was used for image manipulations such as brightness and contrast.

3.3 Results

3.3.1 Cell type specific TFs used as DNA baits

The SAM of higher plants comprises of two distinct clonal cell layers, tunica and corpus (Leyser and Furner, 1992). In dicots, tunica is made up of L1 / epidermal and the L2 / sub-epidermal cell layer, wherein cells follow anticlinal cell division pattern, and thus newly divided cell remain in the same cell layer. In monocot, tunica is made up of a single cell layer. The corpus cells participate both in anticlinal and periclinal cell division patterns. The epidermal cell fate specification takes place in the embryogenesis and the transcriptome of these cells was determined first, by tagging them with fluorescent proteins, such as GFP, YFP, dsRed etc. and then sorting the cells using fluorescent activated cell sorter. Gene expression data was obtained from shoot cell types and analyzed. A total of 1456 genes were found to be enriched in three cell layers of the shoot after applying ≥ 1.5 fold cut off. Out of 1456 genes, 535 were enriched in L1 layer, 256 in the L2 layer and 665 in the WUS domain (Yadav et al., 2014). Of the 535 L1 layer enriched transcripts, 44 encode for TFs, of the 256 L2 layer enriched transcripts, 21 encode for TFs (Yadav et al., 2014). These 65 transcription factors were selected to study further. Out of 65 TFs, a 3 kb promoter fragment above the translation start site was amplified successfully for 41 TF genes. The amplified fragments were cloned in pDONR P4-P1R and subsequently sub cloned into pMW2 vector to make baits. Baits of eight other TF promoters were also included in this study. Taken together, a total of 49 baits were cloned in pMW2 vector, and these baits were screened against the prey library of shoot enriched TFs. The heatmap (Figure 3.2) shows the relative expression level of TF transcripts across the SAM cell types.

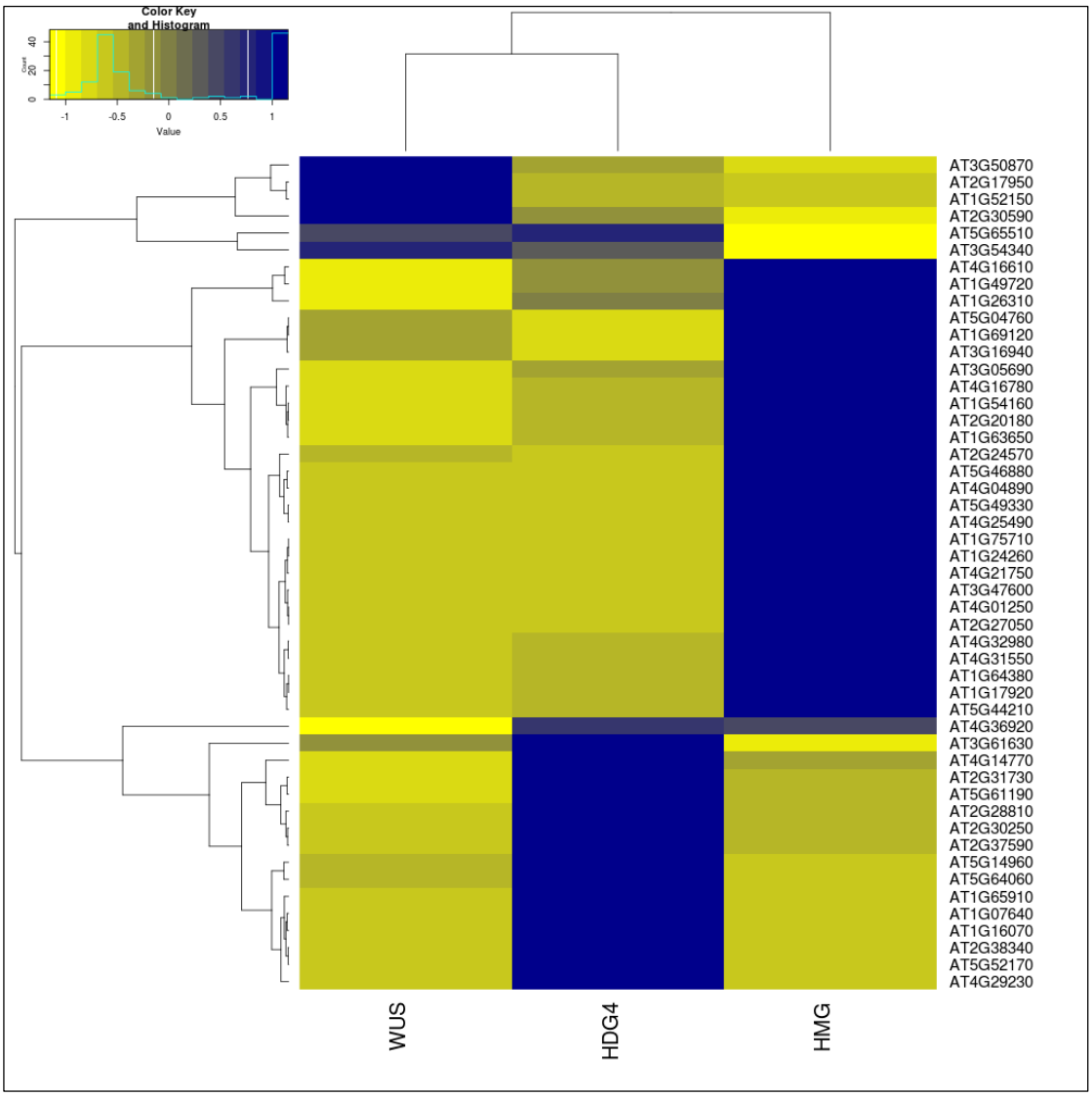


Figure 3.2: Bait transcription factor expression in various cell types. Heatmap representation of MAS5 expression values, across HMG, WUS and HDG4 cell types of the shoot, of transcription factor genes used as baits in Y1H study (Yadav et al., 2014). Alongside is given the color key for expression values, where blue color represents high expression value and yellow color represents low expression value.

The below table lists the primers used for cloning promoter baits used in this study.

Table 3.1 Primers used for bait cloning

Gene ID	Forward and reverse primers
AT5G64060	GGGGACAACTTTGTATAGAAAAGTTGCTAAAATGTTTGTCACATCC
	GGGGACTGCTTTTTTGTACAAACTTGCTGGAAGACAAGGGGAAAACTT
AT5G52170	GGGGACAACTTTGTATAGAAAAGTTGCTTATACAAGCTTGACACCC
	GGGGACTGCTTTTTTGTACAAACTTGCTTTCCCTCTCTGCAACCAAATTAAA
AT5G14960	GGGGACAACTTTGTATAGAAAAGTTGCTCACATATCTATGAATTTACC
	GGGGACTGCTTTTTTGTACAAACTTGCTCCTTGTCATCAAACAAA
AT4G29230	GGGGACAACTTTGTATAGAAAAGTTGCTACTTTGTGATAGGAAATG
	GGGGACTGCTTTTTTGTACAAACTTGCCTCAATCTCGAATATCTT
AT4G14770	GGGGACAACTTTGTATAGAAAAGTTGCTAGATAGATAGCCTAATTTTAC
	GGGGACTGCTTTTTTGTACAAACTTGCTTTCCAACACACAAACAA
AT3G61630	GGGGACAACTTTGTATAGAAAAGTTGCTTATAGTAAAATGTAACCTGTCG
	GGGGACTGCTTTTTTGTACAAACTTGCAGAGAGAGAGAGAGAGAG
AT2G38340	GGGGACAACTTTGTATAGAAAAGTTGCTCCCACGATTGAAGATAGG
	GGGGACTGCTTTTTTGTACAAACTTGCTGGAAAAACACAACACGT
AT2G37590	GGGGACAACTTTGTATAGAAAAGTTGCTTTTGTTGAAGAAAGCTAAGGATTGG
	GGGGACTGCTTTTTTGTACAAACTTGCGGTTATTCTCTTTTGATT
AT2G30250	GGGGACAACTTTGTATAGAAAAGTTGCTTTTCCTGGATCCATCGAT
	GGGGACTGCTTTTTTGTACAAACTTGCGATGGTCTTTAATAAAGG
AT2G28810	GGGGACAACTTTGTATAGAAAAGTTGCTTATTTTTTAATCAGGTTG
	GGGGACTGCTTTTTTGTACAAACTTGCGAATATGTTTTGGAACTT
AT2G31730	GGGGACAACTTTGTATAGAAAAGTTGCTCACATATCTATGAATTTACC
	GGGGACTGCTTTTTTGTACAAACTTGCTCCTTGTCATCAAACAAA
AT2G28810	GGGGACAACTTTGTATAGAAAAGTTGCTTATTTTTTAATCAGGTTG
	GGGGACTGCTTTTTTGTACAAACTTGCGAATATGTTTTGGAACTT
AT1G65910	GGGGACAACTTTGTATAGAAAAGTTGCTAGCTGTTTACTTGGGGAG
	GGGGACTGCTTTTTTGTACAAACTTGCGTCCACTGCCTTATACAA
AT1G16070	GGGGACAACTTTGTATAGAAAAGTTGCTCTTCTGTTCCGTCCTCTC
	GGGGACTGCTTTTTTGTACAAACTTGCGTTTTTTTTCCCTTTCTGG
AT1G07640	GGGGACAACTTTGTATAGAAAAGTTGCTTTAATTGACAAGGAATCAG
	GGGGACTGCTTTTTTGTACAAACTTGCTGTTCCTTCTACCCTTTT
AT5G61190	GGGGACAACTTTGTATAGAAAAGTTGCTCTTCAAAATCAAAGTGTAG
	GGGGACTGCTTTTTTGTACAAACTTGCCACCGGAGATACTAGAGA
AT4G25490	GGGGACAACTTTGTATAGAAAAGTTGCTTTCTAAAAATCTTATTCCT
	GGGGACTGCTTTTTTGTACAAACTTGCTGATCAGAGTACTCTGTTTCAAGAAA
AT1G17920	GGGGACAACTTTGTATAGAAAAGTTGCTGCTGTTGATGGGAAACCT
	GGGGACTGCTTTTTTGTACAAACTTGCAGTAATGTGCAGACGAAAAAAAAC AGTAACAAA
AT1G26310	GGGGACAACTTTGTATAGAAAAGTTGCTAACCCAAACTTTGTCTTGAGTA
	GGGGACTGCTTTTTTGTACAAACTTGCTTCTCTTAAAATAACTAA
AT1G49720	GGGGACAACTTTGTATAGAAAAGTTGCTAATGCTTATTATTTCATG
	GGGGACTGCTTTTTTGTACAAACTTGCACTTCTTTTCTGTTTCAC
AT1G54160	GGGGACAACTTTGTATAGAAAAGTTGCTATGTATAACAATACAACACAGC
	GGGGACTGCTTTTTTGTACAAACTTGCTGTCTTCAAATCTTTACT
AT1G63650	GGGGACAACTTTGTATAGAAAAGTTGCTTGGTCAATAATTTTAACGG
	GGGGACTGCTTTTTTGTACAAACTTGCTGTTTCTTCATCCCCATGTAACAAA
AT1G64380	GGGGACAACTTTGTATAGAAAAGTTGCTTATTGGATTGGGGTTTTT
	GGGGACTGCTTTTTTGTACAAACTTGCTTTTTCTCAACCAGACGT
AT1G75710	GGGGACAACTTTGTATAGAAAAGTTGCTAAATTAATGAGTTGTGTA

	GGGGACTGCTTTTTTGTACAAACTTGCTGAAAGAAGAAAGAAACT
AT2G20180	GGGGACAACTTTGTATAGAAAAGTTGCTCATCAAGTAGACAATAAC
	GGGGACTGCTTTTTTGTACAAACTTGCATCTCTCTCTCTACAAAGATGATGA TAATGG
AT2G24570	GGGGACAACTTTGTATAGAAAAGTTGCTGTTTATTAAAAAATTAATCATTA
	GGGGACTGCTTTTTTGTACAAACTTGCGATGAGAAACCAGAGGAG
AT2G27050	GGGGACAACTTTGTATAGAAAAGTTGCTAGCTAGGTGGCTAAAGACAGATG
	GGGGACTGCTTTTTTGTACAAACTTGCGTCTCTTCCACCACAATC
AT3G05690	GGGGACAACTTTGTATAGAAAAGTTGCTTTATGTTTGCCAACGATC
	GGGGACTGCTTTTTTGTACAAACTTGCCTCCAAATTCCAATTACA
AT3G47600	GGGGACAACTTTGTATAGAAAAGTTGCTATCTGTTCTAGAAACTATTAGG
	GGGGACTGCTTTTTTGTACAAACTTGCCTCTATCTTCTCTATGTA
AT4G01250	GGGGACAACTTTGTATAGAAAAGTTGCTCCATTAAAGCCCAATCCA
	GGGGACTGCTTTTTTGTACAAACTTGCTGAATTTGGTTACTCAGG
AT4G04890	GGGGACAACTTTGTATAGAAAAGTTGCTTTCTGTAATAAATACGTA
	GGGGACTGCTTTTTTGTACAAACTTGCTGTTATGGATGATTGACT
AT4G16610	GGGGACAACTTTGTATAGAAAAGTTGCTCAAAGTCTTTTTTTTTTCTTCTTCTTAAT
	GGGGACTGCTTTTTTGTACAAACTTGCTTTGCCAAAAGAAACAAGAA
AT4G16780	GGGGACAACTTTGTATAGAAAAGTTGCTCTTCTATTTAAAAAAAATATA GACAAAAGTTTT
	GGGGACTGCTTTTTTGTACAAACTTGCCTTCTGTTGAACTTTCTC
AT4G21750	GGGGACAACTTTGTATAGAAAAGTTGCTTTGGAACTTACGTAGTTTA
	GGGGACTGCTTTTTTGTACAAACTTGCGATGATGATGGATGCCTA
AT4G31550	GGGGACAACTTTGTATAGAAAAGTTGCTTATAGTTCATCTTCCAAG
	GGGGACTGCTTTTTTGTACAAACTTGCGATGATTTCTTGGTCTGA
AT4G32980	GGGGACAACTTTGTATAGAAAAGTTGCTTCTCTCAAGGGATGTCTT
	GGGGACTGCTTTTTTGTACAAACTTGCTGGGTTTCTATGAAACTG
AT5G44210	GGGGACAACTTTGTATAGAAAAGTTGCTGATTCCAGACAACCCGGA
	GGGGACTGCTTTTTTGTACAAACTTGCTGCTTATAAAAGAAGAGA
AT5G46880	GGGGACAACTTTGTATAGAAAAGTTGCTACAACATGCATCATCGCG
	GGGGACTGCTTTTTTGTACAAACTTGCCACTTGGAAGATTTAAAA
AT5G49330	GGGGACAACTTTGTATAGAAAAGTTGCTTACCAATTTCGGTGGCCC
	GGGGACTGCTTTTTTGTACAAACTTGCTGCTTCTCGGTCTCTTCT
AT1G69120	GGGGACAACTTTGTATAGAAAAGTTGCTCAATATAATGTTTAACAT
	GGGGACTGCTTTTTTGTACAAACTTGCTTTTGATCCTTTTTTAAG
AT3G16940	GGGGACAACTTTGTATAGAAAAGTTGCTATCAATTCCAATTCGAAG
	GGGGACTGCTTTTTTGTACAAACTTGCCGTGATTTCACACTCGTT
AT5G04760	GGGGACAACTTTGTATAGAAAAGTTGCTTTAGACTCGACGCTTTAA
	GGGGACTGCTTTTTTGTACAAACTTGCAATCGTTGATGATTACTC
AT5G65510	GGGGACAACTTTGTATAGAAAAGTTGCTACATCCTGACAATTTCAA
	GGGGACTGCTTTTTTGTACAAACTTGCCAAAAAAGATTGTAACTT
AT3G50870	GGGGACAACTTTGTATAGAAAAGTTGCTATTCAAGAGCAGAATCTCTCT
	GGGGACTGCTTTTTTGTACAAACTTGCCGCTCTTACACACTTCTC
AT1G52150	GGGGACAACTTTGTATAGAAAAGTTGCTTTGGTTGTTATAGAAAAA
	GGGGACTGCTTTTTTGTACAAACTTGCTACTCCTCAGCAAAACTCTTCTT
AT2G30590	GGGGACAACTTTGTATAGAAAAGTTGCTTTTTTGCATTTGTTTCTA
	GGGGACTGCTTTTTTGTACAAACTTGCCTAAGAACCCTAATTTTT
AT2G17950	GGGGACAACTTTGTATAGAAAAGTTGCTGGGTTTTATTTTGGAATTG
	GGGGACTGCTTTTTTGTACAAACTTGCGTGTGTTTGATTCGACTTT

3.3.2 Broadly expressed and cell type specific TFs were chosen as preys

Using gene expression profile data sets for ten different cell populations of the shoot, differentially expressed genes were identified. Out of 2,266 TFs estimated in *Arabidopsis*, 1,225 are present in the shoot. Out of 1,225, 65 TFs are expressed with ≥ 1.5-fold in the epidermal and sub-epidermal layers of the shoot as compared to other cell types in the shoot. Three hundred and twenty-seven shoot enriched TFs were used as preys against the 49 DNA baits. Other than cell type specific TFs, broadly expressed TFs are also known to play a role in regulating gene expression. For this, cell type specific microarray data was mined, to determine TFs that are expressed more broadly in one or more cell types of the shoot. Both broadly and narrowly expressed TFs were chosen as preys in the Y1H as it would allow for identification of protein-DNA interactions that correspond to both transcriptional activation and suppression. A total of 327 shoot expressed (in individual cell types or multiple cell types) transcription factors were chosen as protein preys in the Y1H matrix assays. The heatmap (Figure 3.3) below gives a visual representation of the expression of these TFs, chosen as preys, in different cell types of the shoot. The blue color in the heatmap relates to high expression of the TF in that particular cell type, whereas yellow color indicates low expression of the TF in the given cell type.

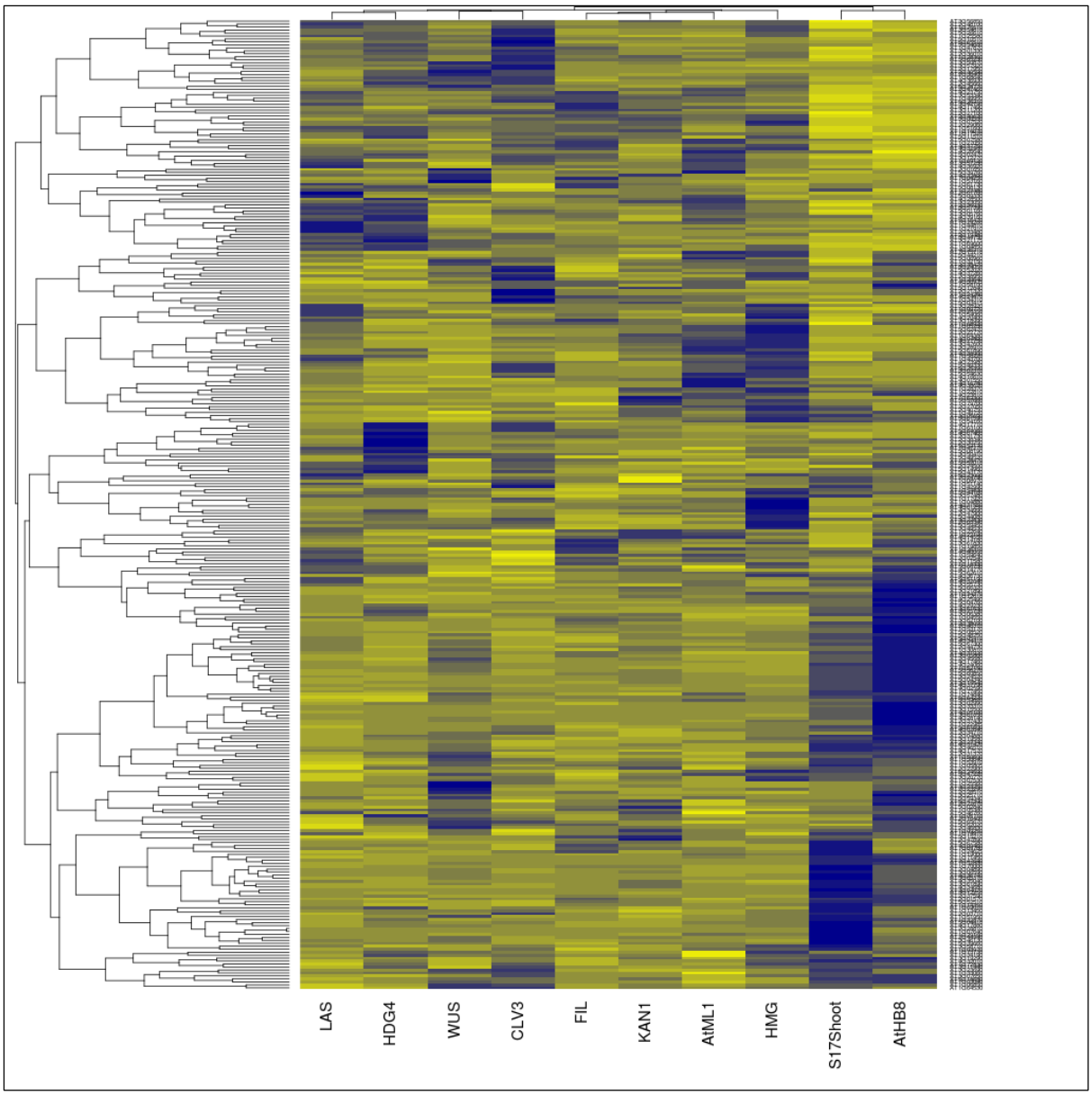


Figure 3.3: Prey transcription factor expression in various cell types. Heatmap representation of MAS5 expression values of TFs, across different cell types of the shoot, used as preys in Y1H study. Preys which are expressed narrowly in the shoot, as well as preys broadly expressed across the shoot are represented in the heatmap. The blue color in the heatmap relates to high expression of the TF in that particular cell type, whereas yellow color indicates low expression of the TF in the given cell type.

The below table lists the primers used for cloning protein preys used in this study.

Table 3.2 Primers used for prey cloning

Gene ID	Forward primer	Reverse primer
AT5G64060	CACCATGGGGAAAACTAACTTGGC	ATCGTCCTTAGTCTGACCGT
AT5G52170	CACCATGAATGGCGATCTTGAAGT	AGCAGGTATACGAAGAGCGG
AT5G14960	CACCATGGATTCTCTCGCTCTCGC	TTTCTCCCGACCAAACTCTT
AT2G38340	CACCATGGAAAAGGAAGATAACGG	GAATCTGAAATACTCAAAATATG
AT2G28810	CACCATGGTTTTCTCATCCGTCTC	CATAAGATGCTGGTGATGAT
AT1G16070	CACCATGGCTGGTTCGAGAAAAGT	AACAGTACAACAAAGCTTGG
AT4G01460	CACCATGAGTGGATTGATGAGTTT	AGTTCGACTAAGATTTGACC
AT4G25490	CACCATGAACTCATTTTCAGC	GTAACTCCAAAGCGACAC
AT1G05230	CACCGTACGTATAACTCCATCTATATAT	CTTTTCTTATACTCCGGAAACTGAAAGCAA
AT1G17920	CACCGCTGTTGATGGGAAACCTGTGGTT	AGTAATGTGCAGACGAAAAAAAACAGTA
AT1G26310	CACCAACCCAAACTTTGTCTTGAGTAGA	TTCTCTTAAAATAACTAAATTATATGAA
AT1G34370	CACCATCAAATTATGAAAATCATAATTA	TTTTTAGTTCAAGATCTTGTTTTTCAATT
AT1G64380	CACCTATTGGATTGGGGTTTTTAATTCA	TTTTTCTCAACCAGACGTTTCGTGGAGT
AT1G75710	CACCAATTAATGAGTTGTGTAGAGAAAT	TGAAAGAAGAAAGAAACTCTTGAGAATG
AT2G03340	CACCGTGAAAGGGATTTTGCTTAGAGGG	CGATTCTTGATTTTGAGAACAAAAAGGA
AT2G40750	CACCGGAGTCATGTCTTTTAATTTTAGA	TTTTTTCTCACAAAGATACAATGAATAT
AT3G02310	CACCCGCTCTAACCAACTGAGCTAATGG	CTTTCCTCACCCCAAAAAATCCCTATCTT
AT3G05690	CACCTTATGTTTGCCAACGATCCAATCG	CTCCAAATTCCAATTACAAAAAGTGAAA
AT3G28910	CACCATTTGTACTTCCCTTTTAATAATA	TATGATCTTGAACTCCCTTAATTAACTA
AT3G47600	CACCTCTGTTCTAGAAACTATTAGGAAA	CTCTATCTTCTCTATGTAAATATCACTT
AT4G04890	CACCTTCTGTAATAAATACGTATTAGAA	TGTTATGGATGATTGACTATGATCACTC
AT4G21750	CACCTTGGAACTTACGTAGTTTACATGC	GATGATGATGGATGCCTATCAATTTTTG
AT4G32980	CACCATGGACAACAACAACAAC	TTTATGCATTGCTTGGCTC
AT5G49330	CACCATGGGAAGGGCTCCG	CTCCAATGTTATCTCTC
AT5G54630	CACCATGGAAGATTCAAGC AAA	GGGCTTGCAAATGACCAC
AT5G57660	CACCATGGGATTCGGCTTAGAG	GAACGTTGGTACGACAC
AT5G62470	CACCATGGGAAGACCACCTTG	GAACATCCCTTCTTGTCCTGC
AT1G69120	CACCATGGGAAGGGGTAGG	TGCGGCGAAGCAGCCAAG
AT5G04760	CACCATGGCGTCAAGTCAGTGG	CATCCGAAACCCAAAATC
AT4G17710	CACCATGGAAACGAAAGACAAGAA	ATTACCGACGTCGTTGACGA
AT5G65230	CACCATGGGAAGATCTCCTAGCTCAGATG	AGATTGATAAGAAATGTCTGGAAAC
AT3G57920	CACCATGGAGTTGTTAATGTGTTCGGGTC	AAGAGACCAATTGAAATGTTGAGGA
AT1G68640	CACCATGCAGAGCAGCTTCAAAACCGTTC	GTCTCTAGGTCTGGCTAACCATAGA
AT2G34710	CACCATGATGATGGTCCATTCGATGAGCA	AACGAACGACCAATTCACGAACATG
AT2G21320	CACCATGCGAATTTTGTGTGATGCTTGTG	TTCATGCTCGTGATTGTTTGCGTTA
AT2G03710	CACCATGGGAAGAGGGAAAGTTGAGCTG A	GACCATCCATCCAGGGAAGAATCCA
AT1G24260	CACCATGGGAAGAGGGAGAGTAGAATTG A	AATAGAGTTGGTGTCATAAGGTAAC
AT2G45660	CACCATGGTGAGGGGCAAAACTCAGATG A	CTTTCTTGAAGAACAAGGTAACCCA
AT3G50870	CACCATGATGCAGACTCCGTACACTACTT	TCTGGTAAAGTCATGGACAAGACTT
AT1G06850	CACCATGGAGAAATCAGATCCTCCACCAG	ATAGGCAGAGCTACTCTCACTAGCA
AT2G17950	CACCATGGAGCCGCCACAGCATCAGCATC	GTTCAGACGTAGCTCAAGAGAAGCG
AT1G04880	CACCATGGCATCAAGCTCTTGTC	GTTCTGCTCAGCAGTCACC

3.3.3 Y1H assays for defining gene regulatory network

To understand the cross talk between broadly and narrowly expressed transcription factors expressed in the shoot and their role in growth and development of the plant, gene regulatory network was generated using Y1H assay. A total of 327 shoot enriched TFs were identified to be used as preys in the Y1H matrix. The DNA element, 3 kb upstream of the start codon, was successfully cloned for 49 TFs as baits. A matrix of 327 preys was spotted in 1536 format, on –TRP media, using singer robot. The baits were spread as lawn on –HIS media. Baits and preys were then mated and allowed to grow on complete media, YAPD. Following this, diploid yeast cells were selected on a double drop out media, -HIS –TRP. The diploid yeast cells were then plated on –HIS –TRP media containing varying amounts of 3-amino-1,2,4-triazole (3-AT). 3-AT is a competitive inhibitor of imidazole glycerol-phosphate dehydratase, an enzyme that catalyzes the sixth step of histidine biosynthesis (Brennan and Struhl, 1980; Joung et al., 2000). In the presence of 3-AT, yeast cells have to produce above the basal level imidazole glycerol-phosphate dehydratase in order to survive. Thus, histidine is used as an auxotrophic marker in this Y1H assay. If an interaction has occurred between a prey protein (TF), which is fused to activation domain of GAL4, and downstream DNA bait element, then, imidazole glycerol-phosphate dehydratase enzyme will be produced by the activation of the reporter gene. The enzyme will participate in the biosynthesis of histidine amino acid. If the binding of prey protein on DNA element is strong then it would be able to drive the transcription of reporter gene even at very high concentrations of 3-AT. If there is no binding of prey protein on DNA element or weak binding, then yeast cells would not grow beyond 5-10mM of 3-AT. Thus, the increasing concentrations of 3-AT would serve perfect titer, against which interactions can be judged as weak, moderate and strong, depending upon the growth. Figure 3.4 A represents the scheme employed for carrying out high throughput Y1H assay. An interaction was setup in the form of four spots for a biological replicate. If three spots grew in the presence of 3-AT on –HIS –TRP plate, then it was counted as a successful interaction. This exercise was repeated thrice, and prey-bait combinations that successfully produced growth twice were counted as interacting partners.

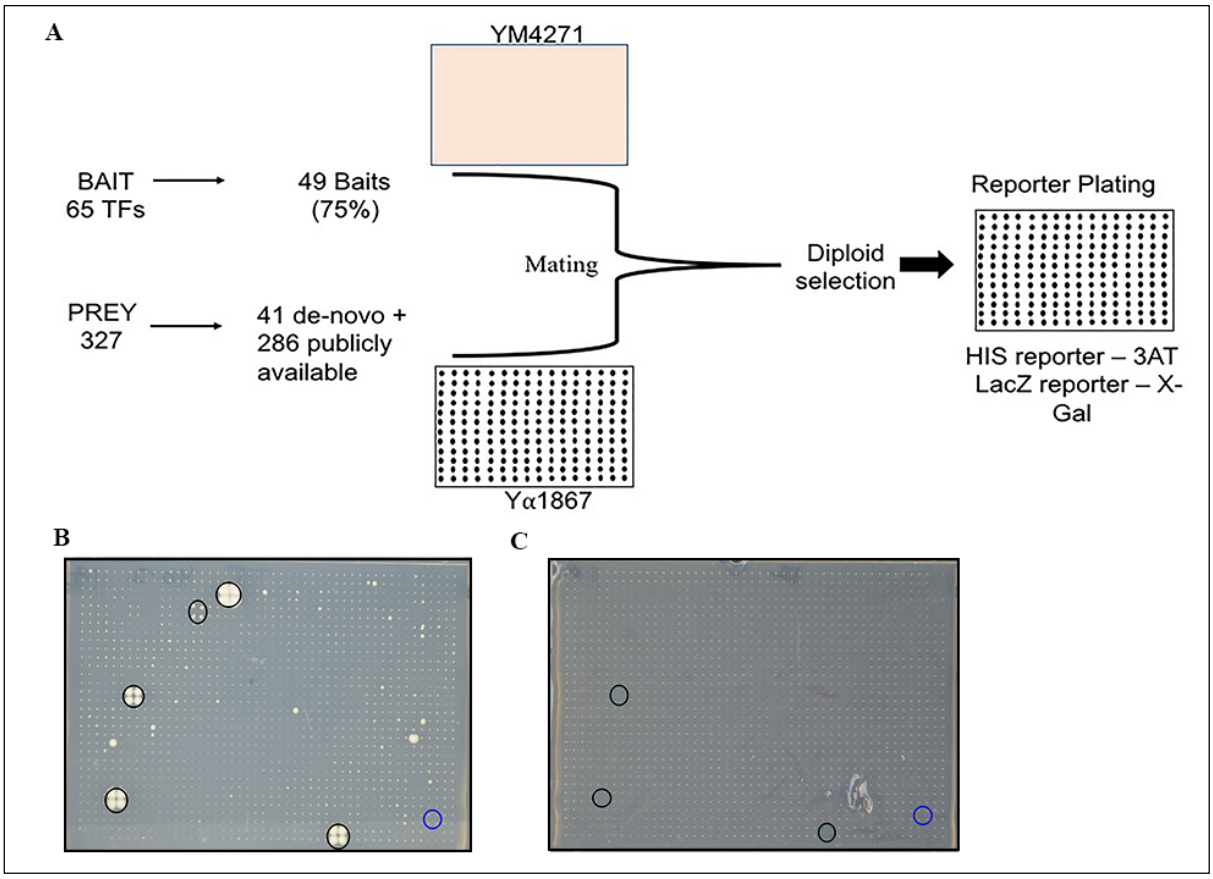
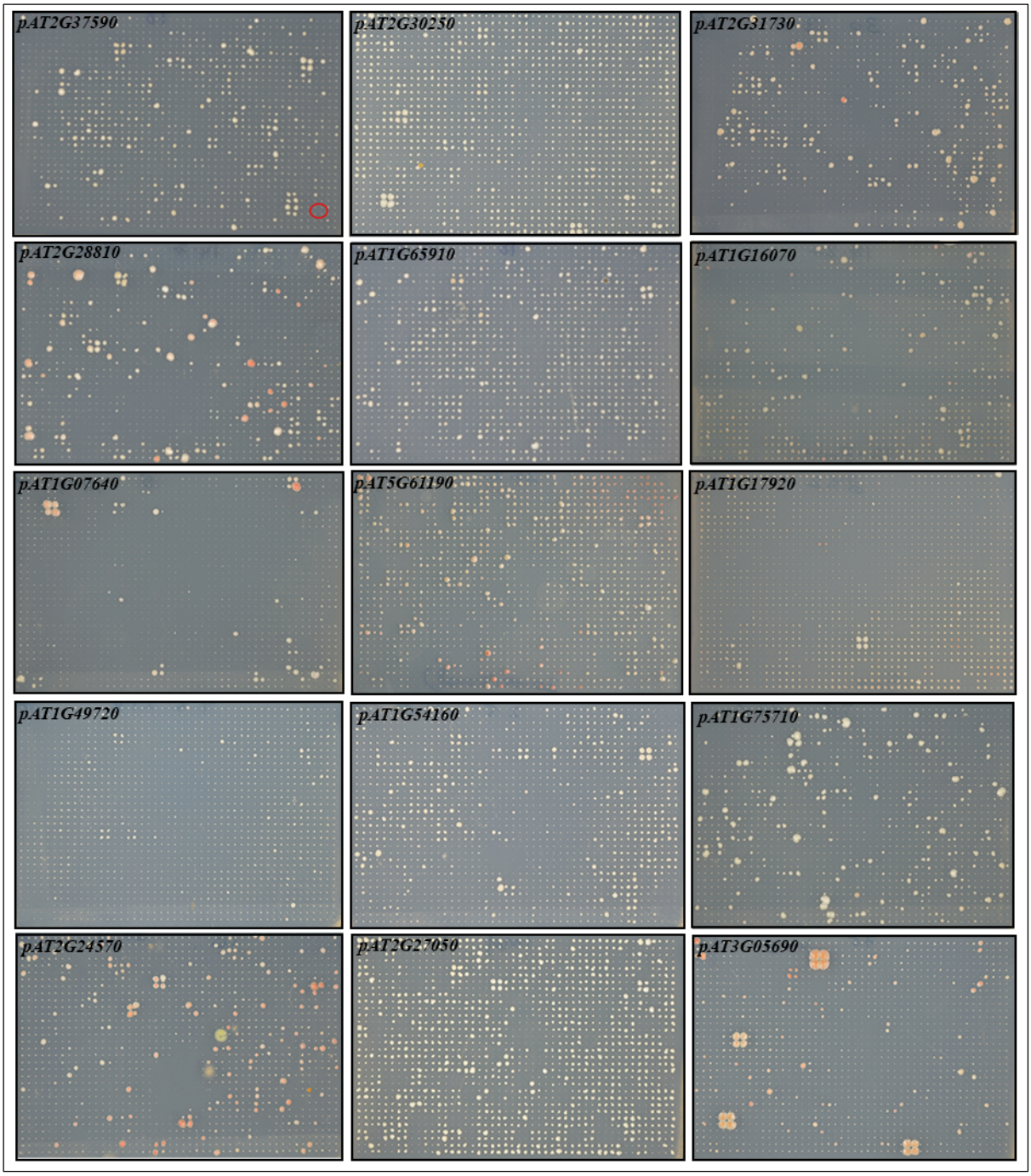


Figure 3.4: Y1H assay for shoot enriched transcription factors. (A) Schematic representation of high throughput Y1H screen carried out using the robotics facility. A total of 49 baits were screened against 327 prey proteins. Baits and preys were transformed into the yeast strain Ym4271 and Yα1867, respectively. The two strains were allowed to mate and then diploids were selected, followed by plating of diploid cells on appropriate selective media, to conclude interactions. (B, C) Robotic plates showing the interactions from Y1H screen, using *HIS3* and *LacZ* reporter genes, respectively. The blue circle represents the negative control on the plate i.e. bait mated with an empty prey, and the black circles represent actual interactions that were concluded for a specific bait.

The below figure also shows the robotic plates for 15 baits, from which the interactions were



concluded.

Figure 3.5: Y1H screen in 1536 array. Robotic plates showing the interactions concluded using *HIS3* auxotrophic marker in Y1H screen. The red circle in plate1 marks the position of negative control on every plate.

Another color producing marker, *LacZ* was used for Y1H assays. However, in the past, protein-DNA interaction screens in *Drosophila* revealed that the set of positive interactions from *LacZ* reporter completely overlapped with the positive readouts from *HIS3* marker (Hens et al., 2011a). Also, Gubelmann et al, showed in (2013) that *LacZ* reporter is less sensitive as compared to *HIS3* because it failed to capture some of the interactions which were found to be relevant *in vivo* in mammalian system (Gubelmann et al., 2013).

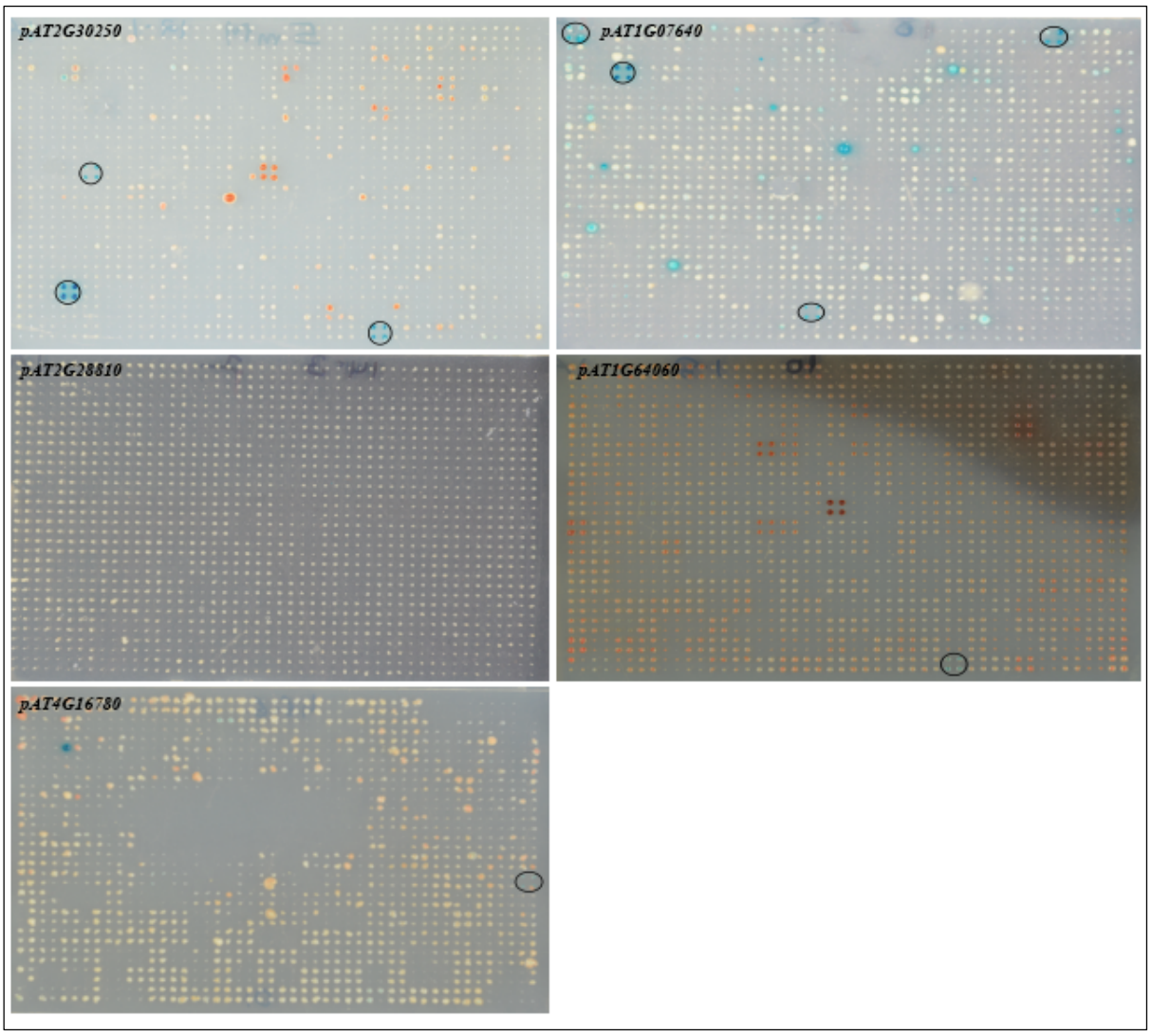


Figure 3.6: Y1H screen using *LacZ* reporter. Robotic plates showing the interactions concluded using *LacZ* color producing marker in Y1H screen.

The Y1H for this study was attempted for nine baits using *LacZ* reporter and for none of the baits, *LacZ* was able to capture all the interactions concluded using *HIS3* reporter. For one of the baits, *LacZ* was able to capture 50% of the interactions concluded using *HIS3*, however for five baits no interaction was captured using *LacZ* reporter. Suggesting that *HIS3* is a better reporter as compared to *LacZ.* For *ANAC103* bait, no interaction was captured using the *HIS3* reporter. However, using *LacZ* reporter gene, one interaction with ARF9 was captured. And for *ATHB-2* bait, 8 interactions were concluded using *HIS3* reporter. However, for the same bait, only one interaction with AT1G27050 was concluded using *LacZ* reporter, which could not be captured using the auxotrophic *HIS3* marker. But for both the baits, *ANAC103* and *ATHB-2,* the interactions were captured very weakly using the *LacZ* reporter. Therefore, in this study, *lacZ* reporter was not opted for all the baits.

Below table shows a comparison between the list of interactions for a set of baits, that were captured using *HIS3* and *LacZ* reporters.

Table 3.3: A comparison between *HIS3* and *LacZ* reporter system based on the prey interaction.

Bait	Prey	*HIS3*	*LacZ*
WRKY25	ANAC075	Yes	No
WRKY25	AT2G28810	Yes	No
WRKY25	OBP2	Yes	No
WRKY25	AT1G06850	Yes	No
WRKY25	DAG1	Yes	No
WRKY25	CAMTA2	Yes	No
WRKY25	ARF9	Yes	Yes
WRKY25	IAA18	Yes	No
WRKY25	ARF12	Yes	No
WRKY25	ARF18	Yes	No
AT2G28810	ANAC103	Yes	No
AT2G28810	DDF1	Yes	No
AT2G28810	BODENLOS	Yes	No
AT2G28810	WRKY54	Yes	No
AT2G28810	PERIANTHA	Yes	No
AT2G28810	CAMTA2	Yes	No
AT2G28810	MONOPOLE	Yes	No
AT2G28810	AT1G06850	Yes	No
OBP2	ANAC075	Yes	No
OBP2	AT2G37590	Yes	Yes
OBP2	AT2G28810	Yes	Yes
OBP2	OBP2	Yes	No
OBP2	DEWAX	Yes	No
OBP2	DAG1	Yes	Yes
OBP2	CAMTA2	Yes	No
OBP2	ANAC103	Yes	Yes

AT5G64060	**ARF9**	**No**	**Yes**
AT4G16780	**AT1G27050**	**No**	**Yes**

A total of 49 DNA baits were screened against 327 prey proteins. Of the 49 baits, three showed very high auto-activation, and they were not considered further. By taking into account remaining 46 baits and 327 prey proteins, I had setup 15,042 (TF promoters' x TF proteins) protein-DNA interactions (PDIs). Analysis of this data revealed 165 positive interactions involving 37 DNA baits and 53 TFs. At least one interacting transcription factor was identified for 76% of the regulatory genomic regions and the majority of these regulatory genomic regions were bound by more than one transcription factor (Figure 3.7). Also, ~ 16% of transcription factors bound at least one regulatory genomic DNA. And one self-regulatory loop was identified, where the TF bound to its own promoter. The number of concluded interactions are relatively low because only ~ 27% (327 out of 1225 shoot expressed TFs) of the TFs expressed in the shoot were taken as prey proteins in the Y1H screen, reflecting low coverage. In a similar study in *C. elegans,* prey library representing ~ 70% of all the TFs were used for screening and far more number of biologically relevant interactions were captured (Arda et al., 2010; Deplancke et al., 2006b; Martinez et al., 2008; Vermeirssen et al., 2007), reflecting the need of large coverage for such screens.

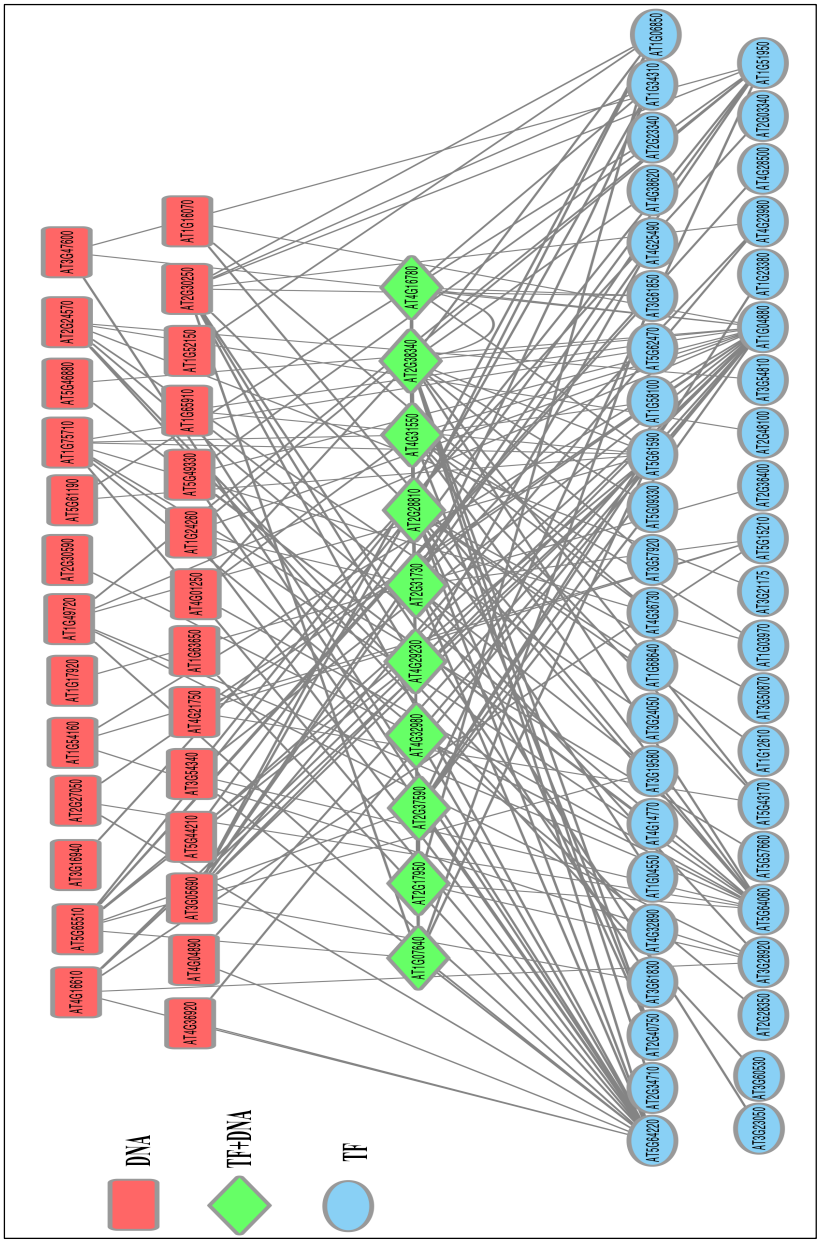


Figure 3.7: Protein-DNA interaction network of transcription factors made using Cytoscape. The nodes in red represent the TFs used as DNA baits. The nodes in blue represent the TFs used as protein preys. And the nodes in green represent the TFs which were used as both DNA baits and protein preys. The nodes are connected through edges, representing an interaction between them. The network consists of a total of 165 interactions between 39 DNA baits and 53 protein preys.

3.3.4 Probable functions of TFs involved in PDIs

Protein-DNA interactions were identified from the Y1H assay and the GO analysis was performed for TFs associated with these PDIs. TFs identified in these PDIs, correspond to various physiological and developmental functions as suggested by Gene Ontology searches. GO terms associated with these PDIs were 'organ development', 'meristem development', 'post-embryonic development', 'response to abiotic stimulus', 'response to salt stress', 'response to auxin stimulus'. Noteworthy, the GO terms enriched for the identified PDIs are quite similar to the GO categories enriched in various cell types of the shoot (Yadav et al., 2014).

3.3.5 Network analysis

In large-scale network studies, when the TF binding data was combined with the expression profile data, it was observed that 30-40% genes perturbed in response to a factor of interest are true targets (Fuxman Bass et al., 2016; Hens et al., 2011b). The remaining factors are not regulated directly. Perhaps, they are regulated indirectly. In order to understand the complex behavior displayed by individual TF node in the network, we did the network analysis of Y1H data and generated simple topology of the network by calculating the in degree and out degree relationship among the nodes. Moreover, based on the cell population expression profile, we analyzed the co-expression of bait and prey genes to find out whether they are having significant overlap in terms of their mRNA expression.

3.3.5.1 In-degree and Out-degree distribution of the network

In-degree refers to the number of TFs binding to a given promoter and Out-degree refers to the number of promoters a given TF binds to. The network of 165 interactions concluded from Y1H assay in this study, consists of 80 nodes; 10 nodes acting as both preys and baits, 27 nodes that are unique DNA elements and 43 nodes that act as unique preys. There are 4 TFs, which have an Out-degree of more than 10 (Shivani, Jayesh and Ram Yadav unpublished data). These TFs belong to different families, such as ARID, CAMTA, NAC and AP2/ERF. DEWAX, an AP2/ERF family TF binds to the cis elements in 10 distinct TF promoters. The role of *DEWAX* has been established in negatively regulating the cuticular wax biosynthesis (Go et al., 2014a). CAMTA2 binds to the promoters of 19 TFs. The role of *CAMTA2* has been demonstrated in increasing the freezing tolerance of the plants by rapid induction of *CBF* (CRT/DRE binding factor) genes (Kim et al., 2013). Other two TFs that

show a high out-degree are HMGB15 and ANAC103. Not much is known about the role of both of these TFs, except one report which suggests the role of HMGB15 in pollen tube development via interaction with AGL66 and AGL104 (Xia et al., 2014). And *ANAC103* is involved in regulating unfolded protein response via bZIP60 (Sun et al., 2013). The remaining TFs exhibit an out-degree from one to six, with maximum of them involved in interaction with a single downstream promoter.

Two promoters show an In-degree of ten or more. These promoters are *AT2G38340, AT2G30250,* belonging to AP2-ERF and WRKY family, respectively. Promoters having such high in-degree indicate that they are regulated by multiple upstream regulators and their *in-vivo* regulation might be much more complex than it might appear. These two promoters also share two common upstream regulators, namely bZIP52 and CAMTA2. Other than this, 6 promoters have an In-degree of 8 and the remaining 29 promoters have an In-degree varying between 1 to 7. The promoters of *ANAC075*, *HDG12*, *EGL3*, *ATHB-15*, *WRKY21*, *AT3G16940* have only one upstream regulator binding on them. Either it reflects the possibility of them being tightly regulated by a single upstream regulator or it shows the incomplete nature of the study due to low coverage of TFs used as prey. This creates a scope of further increasing the scale of the study by including all *Arabidopsis* shoot TFs in the screen. Also, to increase the sensitivity of Y1H screen, smaller promoter fragments can be used and it might result in better detection of interacting partners. For one of the promoters, *HDG12,* the study was taken further to understand its *in-vivo* relevance. And role of *HDG12* was established in controlling leaf growth (Harish and Ram Yadav unpublished data) and development in conjunction with its upstream regulator GRF3. The remaining 29 promoters show 2 to 7 upstream TFs binding to them. This also reveals their regulation by multiple upstream TFs. And more *in-vivo* relevant interactions can be concluded for these promoters also by screening against a large prey library or chopping the promoter fragments into smaller pieces.

In the current study, I was unable to visualize any clear regulatory module among the nodes. The number of TF promoters is also very limited and it does not allow any meaningful interpretation in terms of their apparent behavior at systems level. Moreover, majority of the nodes show single input modules (SIM). The absence of feed-forward loops (FFL) and dense overlapping regulons (DOR) in the current study indicate incomplete nature of the dataset. In

both *E.coli* and yeast, FFL and DOR are enriched significantly in transcriptional gene networks, however, in the higher eukaryotes where still data is very limited; most network topologies are of SIM type. In plant field, only for a few TFs high quality ChIP-seq and Y1H data is available and therefore our view regarding biological networks still could be biased. In order to improve the quality of networks in plants, more interactome data is required to represent the real networks.

3.3.5.2 Co-expression analysis of TFs and their targets

Next, we asked a question whether the TFs and their downstream targets are co-expressed within the same cell types of the shoot. The available micro-array data for 10 different cell types of the *Arabidopsis thaliana* shoot, under normal and different experimental conditions, was used for co-expression study. To establish the co-expression between the bait and the prey TF, Pearson correlation coefficient was calculated between all the possible pairs of interacting baits and preys. The Pearson correlation coefficient values range from -1 to +1. Values close to +1 suggest perfect positive correlation while the -1 indicate opposite correlation for the given set of TFs. The value 0 means no correlation at all. Different cut-offs were applied and the number of interactions falling within different cut-off categories were calculated. P-value was also calculated for the observed number of interactions and were plotted along-side in the graph (Shivani, Jayesh and Ram Yadav unpublished data). Significant interactions were observed at co-expression cut-offs 0.65 and 0.45 with respective p-values of (0.01) and (0.03). However, no significant interactions were observed that were negatively correlated, suggesting that most of the interacting partners in the network are positively co-expressed.

Rank based scoring approach was also used to study co-expression between the interacting pairs. For this, three types of ranks were considered. First, the bait rank was assigned based on the co-expression value of the bait with all the preys. The highly co-expressed partners were ranked highest, followed by the next highly co-expressed partner. Second, the prey rank was assigned based on the co-expression value of the prey with all the baits, with the highly co-expressed partner having the highest rank, followed by next. Third, the mutual rank, which was calculated as the geometric mean of the bait and the prey rank. The mean of the mutual rank for the interaction network was calculated to be 20.6. To check the significance of average rank calculated from the network, 25000 randomizations were performed and a distribution graph was plotted (Shivani, Jayesh and Ram Yadav unpublished data). Mutual

rank was calculated for every interacting pair generated in the randomized network and the mean of all the mutual ranks was calculated. The distribution graph had a mean value of 22.015684326, with a standard deviation of 0.914535050. And the mean average rank for the interaction network was 19.9140566375161, with a P-value of (0.01). Therefore, it shows that the interacting partners in the network are expressed together and have good expression correlation.

3.3.5.3 Expression overlap between bait and prey TFs

Co-expression analysis was done for bait and prey interacting pairs to analyze if the two are significantly co-expressed either positively or negatively. Positively correlated pairs were found to be significant, however, the negatively correlated interacting pairs in the network were not found to be significant. Therefore, it was important to check the expression overlap between the bait and the prey TFs. The expression data for baits and preys for ten cell types of the shoot was considered for this analysis. Non-parametric Wilcoxon signed rank test was performed to find out if a gene was present in a particular cell type or not. And only if there was a 'present' call in all replicates, was a gene considered to be expressed in that cell type.

For every interacting pair, the percentage expression overlap of bait with the prey (bait score) and prey with the bait (prey score) was calculated. Bait score is defined as the number of cell types in which both bait and prey are expressed divided by the number of cell types in which only the bait is expressed. And prey score is defined as the number of cell types in which both prey and bait are expressed divided by the number of cell types in which only prey is expressed. Value of bait and prey score ranges from 0 to 1. Bait score of 1 means wherever the bait is expressed, prey is also expressed in those cell types. And a prey score of 1 means wherever the prey is expressed, bait is also expressed in those cell types. Both bait and prey score of 1 for a given pair means complete overlap in their expression domain. Bait score of less than 1 means bait is expressed in few cell types where prey is not present . And prey score of less than 1 means the prey is expressed in few cell types where bait is not expressed. The bait score was plotted against the prey score for all the interacting partners (Shivani, Jayesh and Ram Yadav unpublished data). And in the plot, one can observe a good number of interactions having both bait and prey score of 1, representing complete overlap of expression domain. However, there are as well a few interactions having a bait and prey score of zero. This signifies negative or inhibitory regulation between the two interacting pairs.

Based on expression degree of baits and preys, protein-DNA interaction density map was plotted (Shivani, Jayesh and Ram Yadav unpublished data). And genes having expression degree of more than 5 were classified as broadly expressed. The type of interactions that were over-represented in the plot were of a narrowly expressed protein prey regulating a narrowly expressed bait.

3.3.6 Motif analysis for predicted protein-DNA interactions

Specialized transcription factors require conserved cis-elements for binding to their target DNA. Based on the protein-DNA interactions concluded from the Y1H assay, we looked for binding site enrichment of upstream TFs in the promoter baits of their target genes. Plant transcription factor database (Jin et al., 2014, 2015, 2017) was used to collate the binding sites for all the available TFs. Binding sites from ChIP, ChIP-Seq, SELEX, DAP-Seq, ampDAP, PBM experiments were all taken together for performing the binding site analysis on DNA baits used in the Y1H study. Non-redundant and high quality binding motifs were available for a total of 674 TFs. Finding Individual Motif Occurrences (FIMO) software was then used for determining the TF motifs in the DNA bait promoters, with a p value cut-off of (10^{-4}). Out of 53 TFs that are a part of our interactome, binding sites for 29 have been identified. These 29 TF preys participate in a total of 101 interactions in the network. Out of these 101 interactions, binding site evidence in FIMO could be found for 73 interactions. An in-silico network was also constructed with 37 baits (the total number of baits in the interaction network) and 29 preys (total preys for which binding motifs available) and a total of 678 interactions were predicted. The Venn diagram below (Figure 3.8) shows the number of interactions having an overlap in the Y1H assay and the FIMO output.

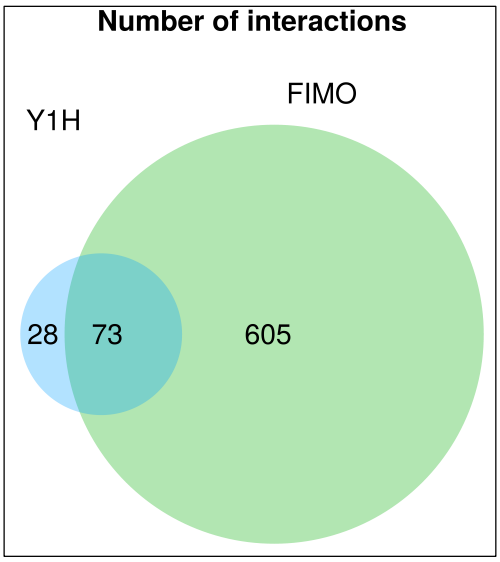


Figure 3.8: Motif prediction using FIMO. Venn diagram representing the overlap between the number of interactions found in Y1H and number of interactions predicted by FIMO, for 37 baits and 29 prey TFs.

To check whether the overlap between Y1H and FIMO was significant, network randomization was performed 25000 times, keeping the topology of the network intact and each time the overlap was calculated. The normal distribution plot was made for the number of overlapping interactions (Shivani, Jayesh and Ram Yadav unpublished data). The mean for the distribution plot was calculated to be 65.40876000, with a standard deviation of 3.57596354. And the number of overlapping interactions observed from the study was 73, with a P-value of 0.016885, which turns out to be significant.

3.3.7 Regulatory relationship between TFs and their targets

One of the key challenges in the Y1H based network studies is to find the biological significance of the PDIs. In other words, it is important to understand whether the prey (TF protein) of interest is involved in regulating the target gene (promoter) *in planta*. For this purpose, we obtained the T-DNA insertion lines from ABRC wherever it was available and I confirmed the loss of function (LOF) T-DNA mutant line by semi-quantitative RT-PCR (Refer to chapter 4). To determine the effect of putative upstream TF regulator on the target promoter nodes, total RNA was isolated from 8 TF mutant lines and 4 over-expression lines, and cDNA was prepared. Over-expression constructs were made for *DEWAX, WRKY54, AT2G28810* and *GRF3* (Harish and Ram Yadav unpublished data). For each of the TF, over-expression was determined either by qRT-PCR or by semi-quantitative PCR. The below figure shows the over-expression of the target gene in the lines used for *in- vivo* validation study.

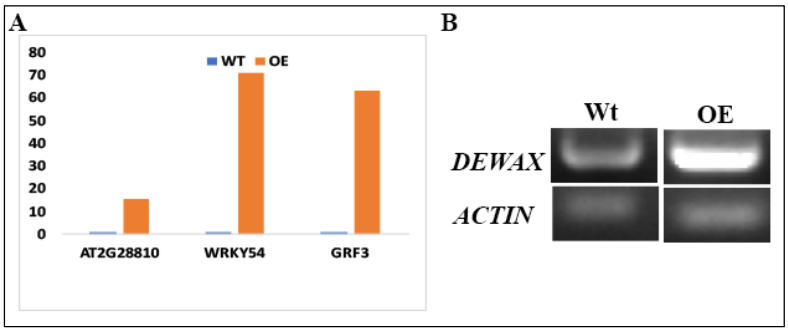


Figure 3.9: Over-expression quantification. (A) Histogram showing the over-expression of *AT2G28810, WRKY54, GRF3* in transgenic lines used for qRT-PCR experiments. (B) Over-expression of *DEWAX* shown using semi-quantitative PCR in *35S: DEWAX* transgenic line used for real time PCR.

To determine the effects of putative upstream TFs, qRT-PCR analysis was carried out in 5 day old seedlings of mutant or over-expression lines of the upstream regulators. Out of 40 regulators tested, molecular phenotypes were found for 4 out of 8 (50%) for DEWAX targets, 2 out of 3 (67%) for ARF9 targets, 2 out of 3 (67%) for WRKY54 targets, 2 out of 2 (100%) for ANAC082 targets, 2 out of 3 (67%) for ATHB-34 targets, 1 out of 2 (50%) for AZF2 targets, 1 out of 4 (25%) for ARF12 targets, 4 out of 11 (36%) for HMG targets and 1 out of 1 (100%) for GRF3 target. None of the targets could be reproduced for GATA8 and AT2G28810. A total of 6 activators and 13 repressors were identified from the qRT-PCR experiment. Out of the 40 interactions tested, 21 of them could not be validated. These interactions could either be false positives in the Y1H assay or the regulation might be happening in very few cells or narrow developmental intervals in the plant body. Hinting at the need of taking specific cells or tissues for capturing the undetected regulations.

For validating the downstream targets of DEWAX, qRT-PCRs were done in both *dewax* mutant and *35S: DEWAX* lines. Levels of *DREB19* were significantly reduced in the *dewax* mutant background in comparison to the WT, suggesting DEWAX to be an activator of *DREB19*. *DREB19* has been shown in literature, to be involved in drought responses. And DEWAX is an AP2/ERF type TF that gets induced by darkness and is a negative regulator of cuticular wax biosynthesis (Go et al., 2014b). The interaction between DEWAX and *DREB19* might be crucial for controlling wax biosynthesis in plants at the time of drought stress. However, the levels of *DREB19* did not significantly change in the DEWAX over-expression background in comparison to the WT. Suggesting that a particular threshold of DEWAX might be required for controlling the *DREB19* target. Other than this, DEWAX was found to act as a repressor for 3 target genes, namely, *ATHB-2, AT4G16610* and *NF-YA2*.

Three targets were validated for ARF9 namely, *WRKY25, AT2G31730* and *NF-YA2*. ARF9 acts as an activator for *WRKY25* and repressor for *NF-YA2*. A previous study has reported the role of *NF-YA2* in leaf development by controlling auxin biosynthesis and negatively regulating *ARF1/ARF2* (Zhang et al., 2017). However, in this study, we identified an auxin response factor that acts upstream of *NF-YA2* and negatively regulates it. Whether this regulation of *NF-YA2* by ARF9 also plays a crucial role in leaf development by affecting the levels of auxin biosynthesis, needs to be investigated in more detail in the future. Another auxin response factor, ARF12 binds to the promoter of 4 genes, *WRKY25, AT2G31730, ATHB-2* and *NF-YA2*. However, by qRT-PCR, the interaction was validated only for

AT2G31730, an ethylene responsive bHLH TF. The levels of *AT2G31730* were slightly reduced in *arf12* mutant background as compared to the WT, suggesting ARF12 to be an activator of *AT2G31730.* ARF12 and ARF9 both bind to the promoter of *AT2G31730* and a robust regulation of this downstream gene might get captured in higher order ARF mutants, due to redundancy among the family members. Also in future, to genetically test the interaction *in planta, AT2G31730* reporter can be crossed with *arf12* mutant or higher order *arf* mutants and YFP levels can be quantitated in comparison to the control.

An AP2/ERF TF, *DREB19* and a C2C2_Dof TF, *AT2G37590* were found to be negatively regulated by ANAC082. ANAC082 responds to disorders in ribosome biogenesis, leading to growth defects in *Arabidopsis* (Ohbayashi et al., 2017). What biological role ANAC082 plays by negatively regulating *DREB19* and *AT2G37590* is still unknown.

Targets of another TF prey, HMGB15 were validated by qRT-PCR. ATHMGB15 prey showed maximum out-degree in the network. Out of 11 targets validated for HMGB15, 7 could not be validated and 4 of them were negatively regulated. The four targets being repressed by HMGB15 were *ABF1, NF-YA2, ATHB-2* and *ATML1*. So far, not much is known about the role of HMGB15, except one report which suggests its role in regulating pollen tube growth by interacting with other TFs such as AGL66 and AGL104 (Xia et al., 2014). What role HMGB15 plays by repressing these downstream targets is not known. However, to test genetically the interaction *in vivo,* reporters of *ABF1* and *ATML1* (since they were made successfully using 3kb promoter, refer to chapter 2) can be crossed with *hmgb15* mutant and changes in the level of YFP can be studied in comparison to the control.

ATHB-34 was determined as an activator for two its downstream targets, *ABF1* and *AT2G31730.* And WRKY54 was determined as an activator for *DREB19* and *AT2G28810.* The biological significance of each of these regulations is an open question that needs to be addressed in future studies.

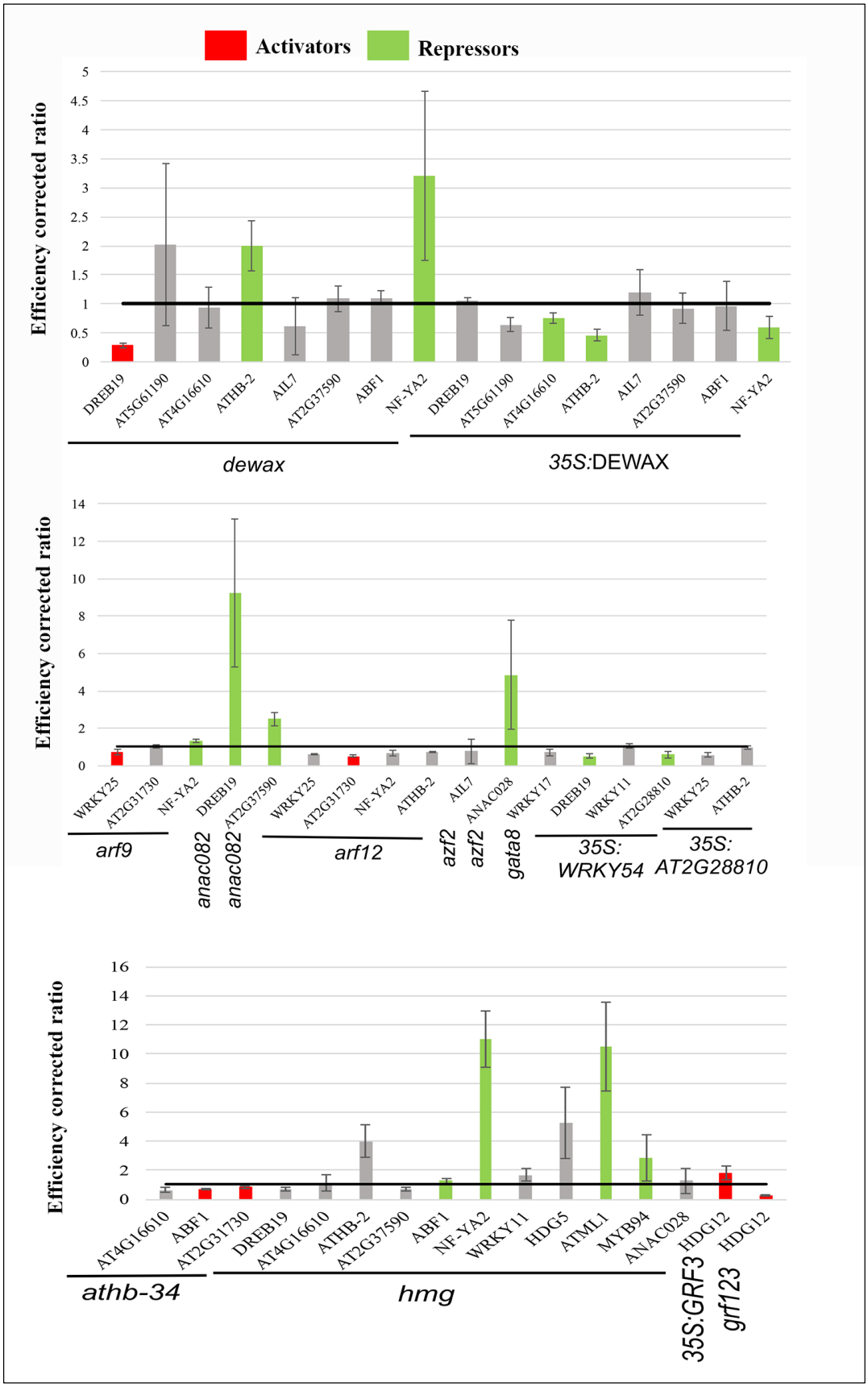


Figure 3.10: Network validation using qRT-PCRs. Sub-networks were validated by using mutant or over-expression lines of the upstream TFs and assaying in 5 day old seedlings for changes in downstream targets. Expression values were normalized by primer efficiency.

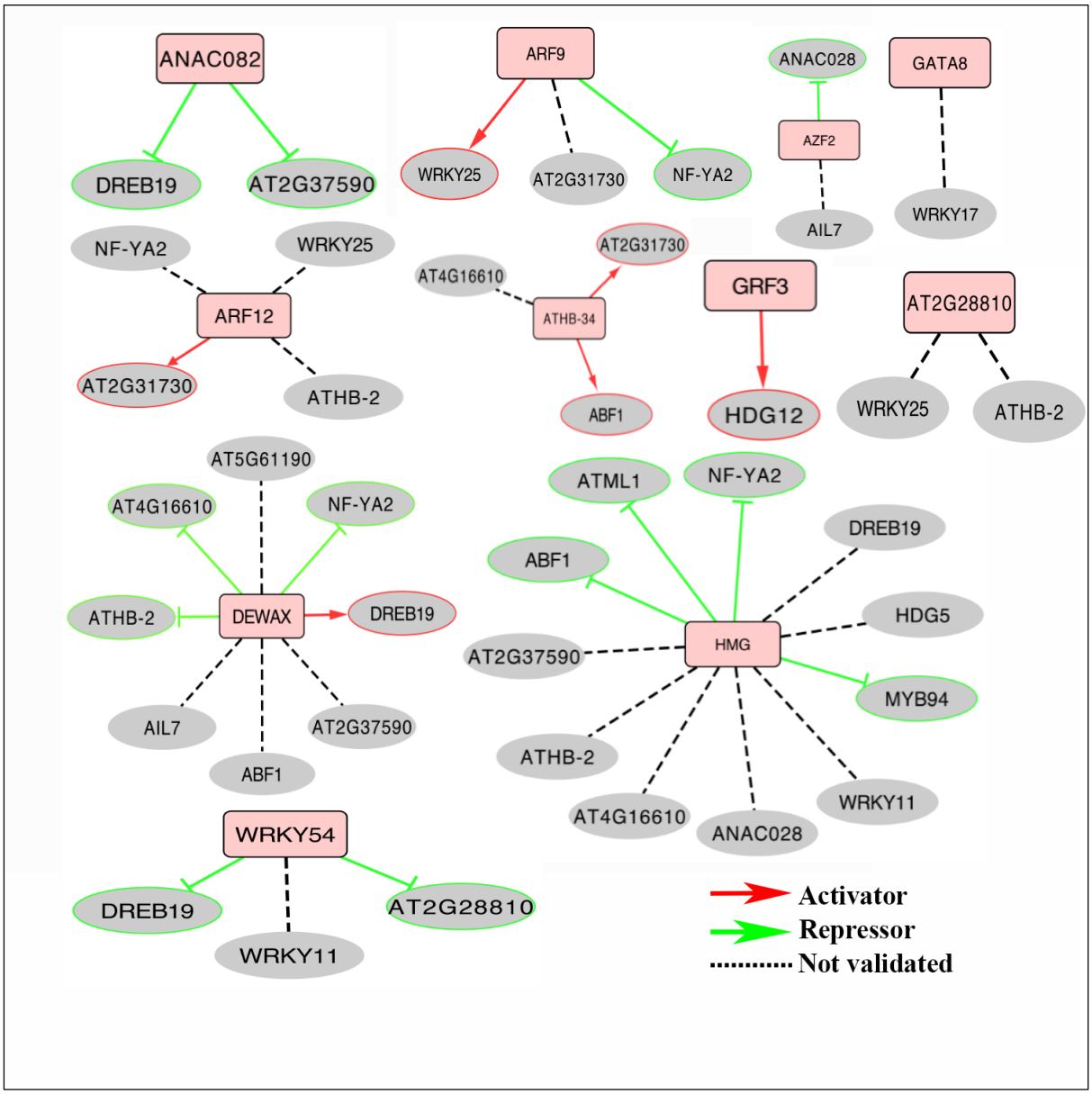


Figure 3.11: Validated sub-networks. Network validation by using mutant or over-expression lines of upstream TFs and assaying seedlings for RNA levels of downstream targets.

The table below lists the primers used for doing qRT-PCR.

Table 3.4: Primers used for qRT-PCR study

Gene ID	Forward primer	Reverse primer
AT2G38340	GAAAGTTCGGGAGGCTATTGG	TGGTGATAACTGGAGCTTTGAGG
AT2G37590	CGGTGGGAACTTGGAGGATTC	CAACGAGAAGAGCCGAAAGCA
AT1G49720	GTTCCTTACGTGTTTGGTCGG	TTGGCCAGCAATGGAGG
AT3G47600	GCTCTTTTGGGAAACAGGTGG	GGAGTTGCTGCTTTGATGTGAG
AT5G46880	GGACACCGTGAGGATCACGAC	TCAATGCAGCTTTCTTGAAGCATC
AT3G05690	GAAATTGAGTAGTAGATGCCGCAAG	CCGACACATTTAATCCGTTCG
AT4G16780	CTTCTTCAGTTCCAAACTCAGA	GATCTTTGAAGGTCTCTTCAAG
AT4G31550	GCAAGAAAAGCAGGAAAAATCG	TCAAGCCGAGGCAAACACTAA
AT4G16610	CGGGTCAGAGGAGATGTTTC	ATACAGCGACGGAGCTCAG
AT1G65910	GATTTACCTGGGAAGTCCTTGC	CTCGACAAAGTGCATAGGCATC
AT2G30250	GCTCCTCACACAGTTTGATA	CGTATTTTCTCCAACCATAAC
AT5G65510	GAGTCACCCGACATAGATG	GAGTCACCCGACATAGATG
AT2G28810	CAGCCGCCAGAGGGAG	GGGCTTGTAGCGTTGACCAT
AT1G17920	GAGAGAACTCGGCCTCAGTGG	GCGCTGCACTTGATGAGTCTATG
AT2G31730	CTTACGGAAAGACCGACACTG	GGAGTCTTCACAGGAGCCC
AT5G61190	CATCCCAGCAGGAGTGAAAGA	CTTGTAAGAGGAGGACCTCCCG
AT2G24570	GCAAGAAAAGCCGAAAAAATCG	CTGCATCGTGGACTGGTGATG
AT2G40750	AAAGAGGATGCTACACTAGAAAGACG	GGTAGCCAATGTATGTGATTTGG
AT4G21750	CAAGAAGCGAGTGGGATATAC	CATAGGACCCTGATGCG
PP2AA	GCATCTAAAGACAGAGTTCCA	AGACAACACACGACAAAGTATC

3.4 Discussion

3.4.1 Y1H assay is a powerful tool for mapping transcriptional regulatory networks

A major question in developmental biology is to understand the phenomenon of gene regulation in various tissues and understand the cascade of events happening in the process of gene regulation. Using Y1H assay, we generated a protein DNA network of transcription factors for the shoot enriched TFs as DNA elements and narrowly and broadly expressed TFs as protein preys. Each PDI determined was tested in quadruplicate and interactions concluded at least two times in three replicates were considered positive. Out of 15,042 interactions tested, 165 interactions were concluded, which constitutes approximately 1% of the total tested. Approximately similar amount of interactions have been concluded in a Y1H study on *Arabidopsis* root TFs (Sparks et al., 2016). Therefore, Y1H is a powerful tool for mapping gene interactions of *in vivo* significance. However, some *in-vivo* interactions may be missed in Y1H assay because of need for hetero-dimer formation or co-factor requirement for interaction. Also, due to high-level of activation in some baits, no interaction can be concluded for the same, limiting the amount of data that can be generated. Despite these limitations, Y1H assay provides a platform for high-throughput studies and capturing *in-vivo* relevant regulations as a lot of captured interactions could also be validated in planta. Also, the Y1H assay might generate false positives which might not be of significance. But the *in vivo* significance of interactions has been validated for a part of the network using qRT-PCR experiment.

3.4.2 Multiple upstream regulators are involved in regulation of downstream targets

Our network analysis data sheds light at the fact that a given gene is regulated by multiple upstream factors and also the upstream regulators are not solely dedicated for the regulation of one downstream target. Our results also indicate in some cases, how the narrowly expressed target genes are regulated by the broadly expressed upstream transcription factors. Using the Y1H data, one interaction was validated *in planta*, which possibly is part of the linear pathway affecting leaf growth and development. The interaction between GRF3 protein and *HDG12* promoter appeared to be specific and significant as it was the only interaction in which both these candidates were involved in the entire interactome. Also, the interaction between the two seems to be involved in controlling the leaf size in *Arabidopsis* by affecting the size of the pavement cells in the leaf epidermis (refer to chapter 5). Therefore, Y1H is a good approach; wherein the downstream targets involved in direct

interactions can be uncovered. And these transcriptional networks can be used for hypothesis generation about regulation of gene expression.

3.4.3 Converting gene networks into regulatory transcriptional modules

Network studies can capture interactions that are regulatory *in-vivo*. Approximately 50% of the interactions tested in TF mutant or over-expression lines, caused change in expression of its target gene. Similar results for PDIs identified through Y1H screens have been reported in the past in *Drosophila melanogaster, C.elegans, Arabidopsis thaliana* (Brady et al., 2011b; Fuxman Bass et al., 2016; Hens et al., 2011a). Not all interactions captured in Y1H show molecular phenotypes *in-vivo* because these may be false positives in the Y1H or the interaction may be occurring *in-vivo* but it might be neutral and may not have any regulatory effect (L.T. et al., 2015). And some regulations might have been missed or diluted because the regulation *in-vivo* might be cell type specific, whereas whole seedling was taken for validation experiments. Binding specificity is another important aspect of regulation of gene expression by transcription factors. Approaches like ChIP-seq, DAP-seq and Y1H assays are able to identify the binding of a TF on the promoter of downstream target gene, however none of them is able to provide insights into the functional significance of these interactions. Interactions discovered through Y1H assays may reflect functional significance if the binding sites for the interacting partner are enriched on the promoter of the downstream gene. However, the binding site enrichment alone may not reflect a regulatory relationship or binding because binding site prediction methods rely only on the presence of predicted DNA motifs (White et al., 2013).

Thus overall, our network shows the nature of complex gene regulation events happening *in vivo.* We showed phenotypic relationship between one of the interactions discovered from this network and validated it *in planta.* The remaining interactions captured might also have biological significance and need to be studied in the future. Also, in future, the scale of the study needs to be increased by including all the TFs from shoot. This can help us to have a complete overview of the shoot related events and our understanding at systems –level will accelerate further.

3.5 References

Arda, H.E., Taubert, S., MacNeil, L.T., Conine, C.C., Tsuda, B., Van Gilst, M., Sequerra, R., Doucette-Stamm, L., Yamamoto, K.R., and Walhout, A.J. (2010). Functional modularity of nuclear hormone receptors in a Caenorhabditis elegans metabolic gene regulatory network. Molecular Systems Biology *6.*

Brady, S.M., Zhang, L., Megraw, M., Martinez, N.J., Jiang, E., Yi, C.S., Liu, W., Zeng, A., Taylor-Teeples, M., Kim, D., et al. (2011a). A stele-enriched gene regulatory network in the Arabidopsis root. Molecular Systems Biology *7.*

Brady, S.M., Zhang, L., Megraw, M., Martinez, N.J., Jiang, E., Yi, C.S., Liu, W., Zeng, A., Taylor-Teeples, M., Kim, D., et al. (2011b). A stele-enriched gene regulatory network in the Arabidopsis root. Molecular Systems Biology *7.*

Brennan, M.B., and Struhl, K. (1980). Mechanisms of increasing expression of a yeast gene in Escherichia coli. Journal of Molecular Biology *136,* 333-338.

Burdo, B., Gray, J., Goetting-Minesky, M.P., Wittler, B., Hunt, M., Li, T., Velliquette, D., Thomas, J., Gentzel, I., dos Santos Brito, M., et al. (2014). The Maize TFome--development of a transcription factor open reading frame collection for functional genomics. Plant Journal *80,* 356-366.

Cui, F., Brosché, M., Lehtonen, M.T.T., Amiryousefi, A., Xu, E., Punkkinen, M., Valkonen, J.P.P.T., Fujii, H., and Overmyer, K. (2016). Dissecting Abscisic Acid Signaling Pathways Involved in Cuticle Formation. Molecular Plant *9,* 926-938.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A.M., Grove, C.A., Martinez, N.J., Sequerra, R., Doucette-Stamm, L., Reece-Hoyes, J.S., Hope, I.A., et al. (2006). A Gene-Centered C. elegans Protein-DNA Interaction Network. Cell *125,* 1193-1205

Deplancke, B., Vermeirssen, V., Arda, H.E., Martinez, N.J., and Walhout, A.J.M. (2006b). Gateway-Compatible Yeast One-Hybrid Screens. Cold Spring Harb. Protoc.

Dreze, M., Carvunis, A.-R., Charloteaux, B., Galli, M., Pevzner, S.J., Tasan, M., Ahn, Y.-Y., Balumuri, P., Barabasi, A.-L., Bautista, V., et al. (2011). Evidence for Network Evolution in an Arabidopsis Interactome Map. Science *333,* 601-607.

Fuxman Bass, J.I., Pons, C., Kozlowski, L., Reece-Hoyes, J.S., Shrestha, S., Holdorf, A.D., Mori, A., Myers, C.L., Walhout, A.J., Araya, C., et al. (2016). A gene-centered C. elegans protein-DNA interaction network provides a framework for functional predictions. Molecular Systems Biology *12,* 884.

Gaudinier, A., and Brady, S.M. (2016). Mapping Transcriptional Networks in Plants: Data-Driven Discovery of Novel Biological Mechanisms. Annu. Rev. Plant Biol.

Gaudinier, A., Zhang, L., Reece-Hoyes, J.S., Taylor-Teeples, M., Pu, L., Liu, Z., Breton, G., Pruneda-Paz, J.L., Kim, D., Kay, S.A., et al. (2011). Enhanced Y1H assays for Arabidopsis. Nature Methods *67,* 575-594.

Go, Y.S., Kim, H., Kim, H.J., and Suh, M.C. (2014a). Arabidopsis Cuticular Wax Biosynthesis Is Negatively Regulated by the DEWAX Gene Encoding an AP2/ERF-Type Transcription Factor. Plant Cell *26,* 1666-1680.

Go, Y.S., Kim, H., Kim, H.J., and Suh, M.C. (2014b). Arabidopsis Cuticular Wax Biosynthesis Is Negatively Regulated by the DEWAX Gene Encoding an AP2/ERF-Type Transcription Factor. Plant Cell *26*, 1666-1680.

Gubelmann, C., Waszak, S.M., Isakova, A., Holcombe, W., Hens, K., Iagovitina, A., Feuz, J.-D., Raghav, S.K., Simicevic, J., and Deplancke, B. (2013). A yeast one-hybrid and microfluidics-based pipeline to map mammalian gene regulatory networks. Molecular Systems Biology *9.*

Hens, K., Feuz, J.-D., Isakova, A., Iagovitina, A., Massouras, A., Bryois, J., Callaerts, P., Celniker, S.E., and Deplancke, B. (2011). Automated protein-DNA interaction screening of Drosophila regulatory elements. Nature Methods *8,* 1065-1073.

Jin, J., Zhang, H., Kong, L., Gao, G., and Luo, J. (2014). PlantTFDB 3.0: A portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Research *42.*

Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J., and Gao, G. (2015). An Arabidopsis Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors. Molecular Biology and Evolution *32,* 1767-1773.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., and Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research *45.*

Jones, A.M., Xuan, Y., Xu, M., Wang, R.-S., Ho, C.-H., Lalonde, S., You, C.H., Sardi, M.I., Parsa, S.A., Smith-Valle, E., et al. (2014). Border Control--A Membrane-Linked Interactome of Arabidopsis. Science *344,* 711-716.

Joung, J.K., Ramm, E.I., and Pabo, C.O. (2000). A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. Proceedings of the National Academy of Sciences *97,* 7382-7387.

Kim, Y., Park, S., Gilmour, S.J., and Thomashow, M.F. (2013). Roles of CAMTA transcription factors and salicylic acid in configuring the low-temperature transcriptome and freezing tolerance of Arabidopsis. Plant Journal *75,* 364-376.

L.T., M., C., P., H.E., A., G.E., G., and C.L., M. (2015). Transcription Factor Activity Mapping of a Tissue-Specific in Vivo Gene Regulatory Network. Cell Systems *1,* 152-162.

Lee, J.-Y., Colinas, J., Wang, J.Y., Mace, D., Ohler, U., and Benfey, P.N. (2006). Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. Proceedings of the National Academy of Sciences *103,* 6055-6060.

Leyser, H.M.O., and Furner, I.J. (1992). Characterisation of three shoot apical meristem mutants of Arabidopsis thaliana. Development *116,* 397.

Martinez, N.J., Ow, M.C., Barrasa, M.I., Hammell, M., Sequerra, R., Doucette-Stamm, L., Roth, F.P., Ambros, V.R., and Walhout, A.J.M. (2008). A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity. Genes and Development *22,* 2535-2549.

de Matos Simoes, R., Dehmer, M., and Emmert-Streib, F. (2013). Interfacing cellular networks of S. cerevisiae and E. coli: connecting dynamic and genetic information. BMC Genomics *14.*

Mukhtar, M.S., Carvunis, A.-R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M.T., et al. (2011). Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. Science *6042,* 596-601.

Ohbayashi, I., Lin, C.-Y., Shinohara, N., Matsumura, Y., Machida, Y., Horiguchi, G., Tsukaya, H., and Sugiyama, M. (2017). Evidence for a Role of ANAC082 as a Ribosomal Stress Response Mediator Leading to Growth Defects and Developmental Alterations in Arabidopsis. Plant Cell *29,* 2644-2660.

Petricka, J.J., Schauer, M.A., Megraw, M., Breakfield, N.W., Thompson, J.W., Georgiev, S., Soderblom, E.J., Ohler, U., Moseley, M.A., Grossniklaus, U., et al. (2012). The protein expression landscape of the Arabidopsis root. Proceedings of the National Academy of Sciences *109,* 6811-6818.

Sparks, E.E., Drapek, C., Gaudinier, A., Li, S., Ansariola, M., Shen, N., Hennacy, J.H., Zhang, J., Turco, G., Petricka, J.J., et al. (2016). Establishment of Expression in the SHORTROOT-SCARECROW Transcriptional Cascade through Opposing Activities of Both Activators and Repressors. Developmental Cell *39,* 585-596.

Sun, L., Yang, Z.T., Song, Z.T., Wang, M.J., Sun, L., Lu, S.J., and Liu, J.X. (2013). The plant-specific transcription factor gene NAC103 is induced by bZIP60 through a new cis-regulatory element to modulate the unfolded protein response in Arabidopsis. Plant Journal *76,* 274-286.

Thompson, W. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research *8,* 4321-4326.

Vermeirssen, V., Barrasa, M.I., Hidalgo, C.A., Babon, J.A.B., Sequerra, R., Doucette-Stamm, L., Barabási, A.L., and Walhout, A.J.M. (2007). Transcription factor modularity in a gene-centered C. elegans core neuronal protein-DNA interaction network. Genome Research *17,* 1061-1071.

White, M.A., Myers, C.A., Corbo, J.C., and Cohen, B.A. (2013). Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. Proceedings of the National Academy of Sciences *110,* 11952-11957.

Xia, C., Wang, Y.-J., Liang, Y., Niu, Q.-K., Tan, X.-Y., Chu, L.-C., Chen, L.-Q., Zhang, X.-Q., and Ye, D. (2014). The ARID-HMG DNA-binding protein AtHMGB15 is required for pollen tube growth in Arabidopsis thaliana. Plant Journal *79,* 741-756.

Yadav, R.K., Tavakkoli, M., Xie, M., Girke, T., Reddy, G.V., Abe, M., Katsumata, H.,

Komeda, Y., Takahashi, T., Aggarwal, P., et al. (2014). A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. Development *106,* 4941-4946.

Zhang, M., Hu, X., Zhu, M., Xu, M., and Wang, L. (2017). Transcription factors NF-YA2 and NF-YA10 regulate leaf growth via auxin signaling in Arabidopsis. Scientific Reports *7*.

CHAPTER 4

Characterizing the loss of function alleles of epidermal and sub-epidermal enriched transcription factors

4.1 Introduction

Studying gene function

Arabidopsis thaliana has got the distinction of being the first model plant, whose genetic blue print / genome sequence became available in the early 2000 (Bevan and Walsh, 2005). It harbours more than 30,000 genes. Despite the availability of high quality genome sequence data and gene annotation almost two decades ago, plant biologists made little progress in understanding the functions of majority of its genes. A quick solution to this problem is offered by numerous bio-informatics approaches that continue to develop in order to understand the function of genes. Many of these studies relied upon sequence-based similarity of the regions of genome to other model organism genomes where the functions were assigned by classical approaches. Bioinformatics based approaches showed great promise in understanding the evolutionary history of genomic loci, presence of cis-elements and their evolution in interacting with trans-acting factors (Petrey et al., 2015). *In-silico* analysis may provide insights into gene function, but *in-vivo* confirmation of the same using genetic tools is important to assign function to a gene (Bolle et al., 2011). In the post genome sequencing era reverse genetics tools combined with forward genetics information yielded great insights in the inner functioning of plant as an organism.

In reverse genetics, the nature of the gene product can be altered using several methods to infer its functions within the cell and tissues, where it is primarily expressed. A direct relationship between function of a gene product and its *in-vivo* role can be inferred using reverse genetics approaches, such as, a) by making a loss of function mutant, b) by over expressing the gene, or c) by knocking it down.

Loss of function mutants

In the past several techniques have been used to create complete loss function mutants at the genome scale to identify gene function, e.g. irradiation, ethyl methane sulfonate (EMS), transposon, and T-DNA insertion mutagenesis. The likelihood of getting a loss of function mutant in a desired gene was not very high, unlike the CRISPER/CAS9, because most mutations caused by EMS, transposon and T-DNA insertion were random in nature.

Chemical mutagenesis

a) EMS based mutagenesis

Ethyl methane sulfonate (EMS) based chemical mutagenesis brings change in DNA by inducing chemical modification in guanine nucleotide, in the process of DNA replication this base will be read by the DNA-polymerase as adenine, and in the newly synthesized strand it will insert a thymine instead of cytosine. Thus, often EMS mutagenesis causes point mutation by converting a G:C base pair into A:T (Greene et al., 2003). It is the most cost effective and easy method available to biologist to introduce mutations within a genome. One can design, a forward as well as a reverse genetics screen interrogating large population of plants to score the desired phenotype or mutation in the target gene using EMS mutagenesis. However, it is difficult to identify individuals carrying the mutation in the gene of interest from a large mutagenized population generated by methods causing point mutations (Gilchrist and Haughn, 2010).

b) TILLING

TILLING, used for creating polymorphism at nucleotide level in targeted genes, uses classical mutagenesis methods combined with high throughput sequencing. This technique takes advantage of the high frequency of mutations caused by chemical mutagenesis and sensitive techniques available for identifying single nucleotide polymorphisms. Mutations can be introduced by TILLING in almost any plant species, regardless of its ploidy, its genome size and the way it is propagated. TILLING makes use of chemical mutagens that have a high rate of causing point mutations in a random nature in the genome. Saturation mutagenesis can be reached in a much smaller population using TILLING (Østergaard and Yanofsky, 2004). Using TILLING, one can study small genes, as well as larger genomes, which are difficult targets for insertional mutagenesis techniques. Since TILLING uses chemical mutagens to cause point mutations, it can lead to a variety of mutations, such as missense mutations and mutations in splice junctions. Also, using TILLING, one can create a series of alleles in a given locus. Therefore, other than generating loss of function mutants, chemical mutagens can also give rise to hypomorphic and gain of function alleles, that can give a range of phenotypes (Alonso and Ecker, 2006). Another advantage of TILLING is that it does not require transformation and therefore can be applicable to species that are difficult to transform or are recalcitrant. Lately, TILLING has become an important tool in crop breeding, as opposed to transgenic approach and

it has already been used for crops such as rice, wheat, barley, rapeseed, soybean and tomato (Kurowska et al., 2011).

Insertional mutagenesis

a) Transposon mutagenesis

An alternative approach to EMS, is based on insertional mutagenesis, where a foreign DNA is inserted into the genome by transformation. Insertional mutants are generated either by using transposable elements or transfer DNA (T-DNA) (Krysan, 1999). Genome wide insertional mutagenesis screen offers advantage over EMS in terms of quick identification of mutation in the DNA without genetic mapping. PCR is used to identify the flanking regions of inserted DNA elements. Using T-DNA mediated transformation method; transposons can also be introduced into the genome. Transposons can jump from one chromosomal location to another, as long as an active transposase enzyme is present. This leads to possibility of having a mutation both in the excision as well as in landing site of the genome (Smith et al., 1996). Usually, transposons tend to jump to linked loci, there are strategies designed that can make the transposon land at unlinked loci (Sundaresan et al., 1995). Positive as well as negative selectable markers are harboured both by T-DNA integration site launch pad and the mobile element. Plants that are able to select on both the markers simultaneously are likely to be enriched for unlinked events of transposition. However, to generate stable mutants, it is desirable to have the transposon fixed after one jump. Transposon activity can be regulated by introducing both transposase and transposon element in the same plant. Subsequently, crossing them to WT would stabilize the mutation and pure mutant lines carrying only transposon will be established by further follow up of the segregating mutant allele of the gene. Plants with insertions that do not carry the active transposase are then selected Although, it is easy to generate transgenic lines carrying transposon insertions, however, selecting plants with stable insertion requires a huge amount of time and efforts.

b) T-DNA mutagenesis

T-DNA insertional mutagens have advantage over transposons because they do not transpose from one position to another following insertion, and are physically stable over the generations (Krysan et al, 1999). An inserted T-DNA element not only knocks out the gene of interest but also acts as a marker subsequently for the identification of mutation. Since the intergenic region

in *Arabidopsis* genome is relatively smaller, therefore the probability of disrupting a gene with large T-DNA element (5-20kb length) is quite high (Krysan, 1999; O'Malley and Ecker, 2010). Independent research groups conducted genome wide T-DNA insertion mutagenesis screens in the past. They identified a large number of T-DNA insertion lines for *Arabidopsis* genes. The likelihood of having T-DNA mediated disruption in a gene of interest is high in the T-DNA lines given the sheer number of lines generated by different groups. Most T-DNA lines are screened for gene specific insertions and are deposited in seed stock centres after initial confirmation of T-DNA for distribution to interested research group around the globe. Usually, T-DNA insertions occur at a frequency of one insertion per plant, with a little chance to have more than one per plant. Thus, T-DNA insertions offer the advantage of random insertions with low copy number. Also, homozygous lethal mutations can be maintained in the population as heterozygous individuals.

In advanced plant transformation vector, which is also called binary vector, a segment of DNA is flanked by 25bp imperfect direct repeats, called right and left border of the T-DNA. The T-DNA segment is sandwiched between the two border boundaries, and gets transferred into host cell along with additional genes present in it. The helper plasmid or Ti plasmid harbours the *vir* genes, which are required to mobilize the T-DNA into host cells by *Agrobacterium.* In modern plant biology, T-DNA based transformation of plant cells has become a routine for studying gene functions, and it is an indispensable tool for many research labs around the globe. One can make loss of function, gain of function and knockdown constructs for desired genes to reveal their biological functions in different cellular contexts. In the late 1990s, large T-DNA insertion screens were carried out with an aim of identifying at least one knockout mutation in every gene of a genome. To map the insertion site of T-DNA within the genome, the forward primer is targeted against the left border, while the reverse primer binds to the flanking genic regions. For analysis of phenotype, it is important to identify homozygous plants from the population. Polymerase chain reaction (PCR) based methods are developed for easy isolation of homozygous and heterozygous individuals, carrying the T-DNA in gene of interest, from a given population of segregating individuals (May et al, 2002).

Gain of function studies

Sometimes, knock-down and knock-out approaches are not sufficient for inferring the complete function of a gene. Often, additional approaches, such as gain of function or over-expression are needed. Over-expressing a gene product can be unsettling for the organism and may lead to modification in the phenotype as compared to the WT. Gene expression levels can be increased *in-vivo* either by randomly activating endogenous genes through enhancers or over-expression constructs for individual genes can be created and transformed *in-planta.* The loss-of-function and gain-of function phenotypes are often complimentary to each other (Bolle et al., 2011). Over-expressing a gene might lead to changes in phenotype by inhibiting another protein involved in a similar biological function (Prelich, 2012).

Objectives of the present study

To understand the inner functioning of SAM, a cell type specific gene expression study was conducted by isolating the fluorescently labelled cells (Yadav et al., 2014). This study revealed expression pattern of hundreds of gene across the different cell layers and domains of shoot at a very high resolution. Despite the impressive array of transcripts showing cell type specific gene expression, very little is known, how the cell type specific gene expression is achieved at transcription level. By analysing this microarray data, we identified dozens of transcription factors (TFs), which are expressed in epidermal and sub-epidermal cell layers of shoot apex. To understand the role of these TFs in shoot development and growth, I analysed the T-DNA insertion mutant lines and characterized their phenotype in detail.

4.2 Materials and Methods

4.2.1 Molecular Biology Techniques

4.2.1.1 Genomic DNA extraction

A modified CTAB buffer method was used for genomic DNA isolation (Murray, M.G., Thompson, 1980). Plant tissue was first collected in aluminium foil and frozen immediately in liquid nitrogen. Tissue was then grinded into fine power in liquid nitrogen, and 600 µl CTAB buffer (2% (w/v) CTAB, 1.42M NaCl, 20mM EDTA, 100mM Tris, pH 8.0) was added to the grinded tissue, and it was incubated at 65^{o}C for 20 mins. 6µl β-mercaptoethanol was also added

to the CTAB buffer prior to use. Phenol: Chloroform: Isoamyl alcohol (600µl) was added, and the mixture was put on shaker for 5 minutes. To precipitate the proteins, the mixture was centrifuged at 14,000 rpm for 10 minutes. The upper aqueous layer was transferred into fresh tube and 600µl chloroform: Isoamylalcohol was added to it. Turning the tube up and down several times mixed the contents of tube. The mix was then centrifuged at 14,000 rpm for 10 mins. The upper aqueous layer was again transferred to a fresh tube and DNA was precipitated with isopropanol. Finally, the pellet was washed with 70% ethanol and DNA was resuspended in 50µl of dH2O.

4.2.2 Imaging

Nikon D-5100 camera was used for taking whole plant pictures. Adobe Photoshop software (Adobe systems Inc.) was used for image manipulations such as brightness and contrast.

4.3 Results

The dicot SAM consists of three cell layers, namely, epidermal / L1-cell layer, sub-epidermal / L2 cell layer and the corpus / L3 cell layer. To understand the epidermal and sub-epidermal cell type fate specification and maintenance in shoot apex, I analysed the cell type specific microarray data to identify TFs that are enriched in these cell types. Gene expression profiles have been generated for these cell types as well as for central zone / stem cells, rib meristem / niche cell, peripheral zone including vasculature cell types, resulting in a total of 10 transcriptome profiles. Of the estimated 1538 *Arabidopsis* TFs in this study, 1225 are expressed in shoot apex cell types. MAS5 analysis revealed a present call for 817 TFs, representing at least in one cell type. To identify the TFs enriched in cell layer data, a three-way analysis of epidermal, sub-epidermal and L3 cell layer was carried out. This analysis revealed, 44 and 21 TFs enriched in epidermal and sub-epidermal cell types, respectively.

To gain insight into the functions of epidermal and sub-epidermal cell type enriched TFs, we identified at least one T-DNA insertion line in them. The plants were grown in the plant growth chambers to collect the tissue for DNA isolation and to determine the insertion. Most of the lines were in Col-0 background, and were obtained from Arabidopsis Biological Resource Center (Ohio, USA). Most of the lines were deposited by T-DNA mutagenesis led by SALK,

WiscDsLox and GABI-Kat consortium (Alonso et al., 2003; Rosso et al., 2003; Yamada et al., 2003) . To study the gain of function phenotype wherever possible, ectopic expression lines of TFs were either ordered from stock Center or made in house. A few lines were requested from the laboratories that have already reported them in the literature.

4.3.1 T-DNA genotyping and identification of null alleles for selected TFs

To identify loss of function mutant in the epidermal and sub-epidermal cell type enriched TFs, T-DNA lines were identified with the help of genome browser at "www.arabidopsis.org" and www.gabi-kat.de. The lines were ordered from ABRC Ohio. Individual T-DNA lines were put on the MS-agar plate (for 7-days), and then transferred on soil. Two-three young leaves (100mg tissue) were collected for DNA isolation. The isolated DNA was used for setting T-DNA PCR. PCR primers for individual T-DNA lines were designed using SIGnAL Salk tool. Individual plants of the above mentioned lines were genotyped to identify the homozygous plants. For identifying homozygous plants, both alleles need to be analysed independently. PCR was set up using three different sets of primers, 1) left primer (LP) and right primer (RP) pair, 2) LP and LBb3 (T-DNA primer) pair, and 3) RP and LBb3 pair (Figure 4.1). Eight individual plants were genotyped independently to identify the homozygous recessive allele for a given TF. WT DNA was used as a control (Figure 4.2).

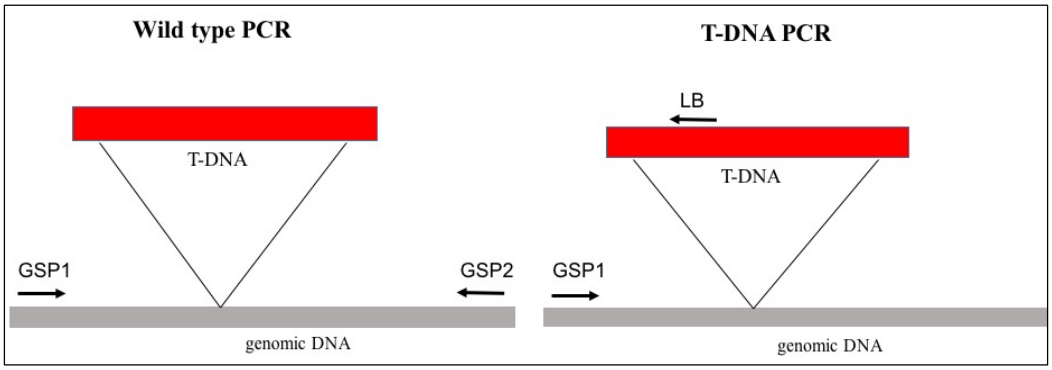


Figure 4.1: T-DNA confirmation strategy. Schematic representation of T-DNA insertion within the genome and strategy used for confirming the insertion and homozygous and heterozygous lines. GSP1, GSP2 and LB refers to Genome Specific Primer1, Genome Specific Primer 2 and Left Border primer respectively.

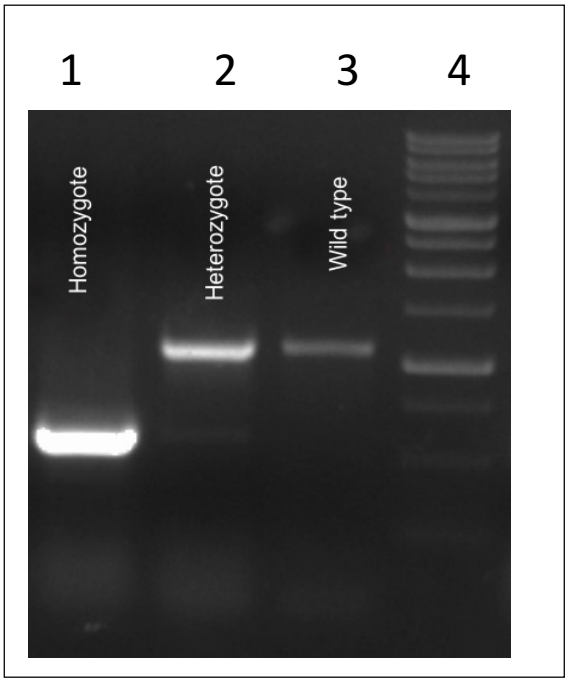


Figure 4.2: Homo and heterozygote allele identification. A representative gel image showing the T-DNA insertion confirmed by PCR using genomic DNA. First lane shows the homozygote plant which has only T-DNA specific band. Second lane shows the heterozygote plant that carries both T-DNA specific as well as a WT specific band. The third lane shows the WT plant, which carries no T-DNA insertion in the genome. And the fourth lane represents the 1kb ladder.

The table below lists the primers used for genotyping each of the T-DNA lines screened (Table 4.1) and the common primers (Table 4.2) used for each collection of T-DNA mutants.

Table 4.1: List of T-DNA lines screened and primers used for genotyping

S.No.	Gene Name	T-DNA	Primer1 (GSP1)	Primer2 (GSP2)
1	***AT1G12610***	SALK_114390	TTTCATATCCTTCCATCGTGC	TAAACTGGGTGACGTGTCTCC
2	***AT1G22190***	SALK_139727	GTGTATCGGTGAGGCTGAGAG	GTCCTCCTCCGGTAGTTTCAC
3	***AT2G38340***	SALK_144950	TTGGGTGACAATCCACCTTAC	TTGTCGAACACCTCTAAACCG
4	***AT3G61630***	SALK_063548	GATCGCTTCAATCTCATGCTC	CTCAATCTGAGATGCGGAAAG
5	***AT5G44210***	SALK_091532	AGGAAATCAACGAAATCTCCC	ACATGCAAACGGAAATCTCTG
6	***AT5G61590***	SALK_015182	CAGCTTAGGATTCGAACCATG	GAAAACGCAGAAGTTCCATTG
7	***AT1G04880***	SALK_100002	TGCTTCAGGGGTAACACAAAC	TGACCTATTTTATAGCCACCCC
8	***AT4G01460***	SALK_027604	TTGATCTGGAGGACAGGACAC	CTGCTCCGTTTTCTGTTCTTG
9	***AT2G20180***	SALK_072677	AGATCGTCGAAGACCTTGTTG	GGGTGAAGATGATGATCTTATGG
10	***AT2G31730***	SALK_115465	CGCTGATCAATTGGACATTTG	ACCAAAAATGGGGAACAAAAC

11	***AT1G06850***	SALK_033320	TTCTCTCTTTGGATCGAGCTG	CAGACGTGGGGAGATGTAATG
12	***AT1G07640***	SALK_139462	AAGTTGAGTGTTTGGTGGTGG	GCGGATTAGAAAAGAAGATGG
13	***AT2G28810***	SALK_056801	TAACCAAAACAGAAAAGTGCG	GCGGTAAAAATCACACAAAAAG
14	***AT1G29160***	SALK_045465	TCCCATTATTGTGTGATTGTCAG	GAGCTACAAGCAACAACAGCC
15	***AT1G64620***	SALK_130584	TGTACACGTTCATCAAATGGG	AAAGCTCCAAAGACGAGGAAG
16	***AT1G75710***	SALK_048268	CCGCCTTTAAGAAGGTCAAAC	CGTTAGATCTCGTTGAGCCAG
17	***AT5G54630***	SALK_098531	CCGTGCATTAAATCAAGTTTACC	TCCTCCTCCTCGATCTCTCTC
18	***AT5G61190***	SALK_010662	TTTCACTCCTGTTCCAACCAC	GTTGTAGTTTTGTTGCTCCGG
19	***AT3G16940***	SALK_078900	CTGCAAGAGCATCCTTGAGAC	TGAACCCATGGTTATGCCTAG
20	***AT1G54160***	SALK_042760	TGACGTTTCAGCACATTGAAG	TAGCCAATACCCAATTTGCAG
21	***AT4G14770***	SALK_021952	AGATTGCAGACAAAGCAAAGC	TGGAGAATCCTGCATTTTCAG
22	***AT5G14960***	SALK_093190	TAACGTTTGTGAGTTCGGTCC	CATCTCCCAGACCTGGTAATG
23	***AT2G27050***	SALK_049679	GCTCCATACGCTAAACGATTG	ACAAAATGCGTTTGAAACGTC
24	***AT2G36400***	SALK_026786	TCCATCCATGTTCAACTAGCC	GAAAAGACTCACTGGGGAACC
25	***AT1G05230***	SALK_138646	AAGATGTGAGTTCTCATGCCG	ATGTTCGAGCCAAATATGCTG
26	***AT1G17920***	SALK_127261	CTATTCCCCGAGATCTTTTGG	TGGCTGAGATGGTAAACTTGG
27	***AT4G04890***	CS304455	GTCAGTTTCGATTCCTTCTGAGAC	GAATTGATCAGAATAGCCCC
28	***AT4G04890***	SALK_109425	CTCGTTCGCTTGATCTTGAAG	AGATCACCAGGAATGTTGCTG
29	***AT4G16780***	SALK_106790	CTCACGACTCAACGATCTAACC	CGTCACTGATTCCTCTTGAGC
30	***AT4G17710***	CS303999	ACATGAATGGTCTTAGCCAGAGTT	AATGGAAACGAAAGACAAGAAAGA
31	***AT4G21750***	SALK_033408	GACCAATATTTTGCTTTTCGG	CTCGGAGCTTGTCATTTTCTG
32	***AT5G52170***	SALK_132114	ATACAAGACATAGGCAACGGC	ACAGCAGCACGAAGAAGAGAG
33	***AT3G28910***	SALK_027644	AAGATATGACGCAATTGCAGC	CTTTGGAGGCTTTACCTCCAC
34	***AT3G47600***	CS859284	AAAGTCAACGAGTTGGGTGTG	ATACCCATCTGGGGTCTATGC
35	***AT5G49330***	CS9979	CCAACAAGCTACTACAAAACCACA	
36	***AT5G62470***	CS333785	GACAAGGCCTCACAAAGAGC	
37	***AT1G28470***	SALK_000287	TGGTTATCGCGATTTCATTTC	CTCGAGGTTAAAGTTACGCCC
38	***AT1G65910***	CS868895	GAGCTTCTCTGACGCACCTAG	TGTTTGTCACAACTCGTCAGG
39	***AT5G64060***	CS850179	ACTCTGACAGACCCACCATTG	GCAATGACCAAGTCGTAGAGC
40	***AT2G24570***	SALK_076337	TGGATTTTGGTTAAAGACCTTC	AGCAAGAAAGATCGAAGAGCC
41	***AT2G40750***	SALK_111964	GCTGGTGTTGTTCTCTTGCTC	GGGTTGGTAAGGGTAAAAGAGG
42	***AT4G31550***	SALK_141511	TGTCGTATTGATGAATCGCTG	GTCAGTGATCTCGGAGCAGTC
43	***AT5G56270***	SALK_020399	CCAAGAATTTGGCTGAATCTC	TGTTAGAACACGAATCACCCC
44	***AT4G01250***	SALK_047120	TACTGCTGACGGATTATTCCG	CCTTTACCAAAAATGTAACGCAG

45	***AT1G04880***	CS2102953	TACAATACATTACCCCCAAGAAGC	GATCTGATGTGTGGAGCCTTAAAGTC

Table 4.2: List of common primers used for T-DNA genotyping

S.No.	Source of line	Primer name	Primer1
1	SALK	LBb3	ATTTTGCCGATTTCGGAAC
2	GABI-KAT	08474	ATAATAACGCTGCGGACATCTACATTTT
3	Wisconsin	P745	AACGTCCGCAATGTGTTATTAAGTTGTC

4.3.2 TFs families represented in the epidermal and sub-epidermal cell type data

A protein is annotated as true TF based on its ability to bind DNA and activate the transcription of target genes. Most TFs have a DNA binding domain and an activation domain. The DNA binding domain helps the TF protein in binding to the specific DNA sequence within the chromatin called as *cis*-element, while the activation domain assists in assembling the transcription initiation complex on the promoter of target genes to drive transcription efficiently. Based on the DNA binding domain, TFs are classified in to different families. According to the Plant TF database (http://planttfdb.cbi.pku.edu.cn), plant TFs are classified in to 64 families. Of the 64 TF families reported, 50 TF families are present in *Arabidopsis*. Of the 50 TF families reported in *Arabidopsis*, TFs belonging to 19 families are expressed in epidermal and sub-epidermal cell types. This includes, nine AP2-EREBP, one ARID, five bHLH, two bZIP, one C2C2-CO, four C2C2-DOF, seven C2H2, one CAMTA, three CCAAT-HAP2, one CPP, one E2F-DP, one EIL, 9 HB, three MADS, six MYB, five NAC, one TUB, and six WRKY family TFs.

4.3.2.1 APETALA2/ETHYLENE RESPONSIVE ELEMENT BINDING PROTEIN (AP2/EREBP) Transcription Factor Family

AP2/EREBP-like protein family of TFs is one of the largest TF family and also unique to the plants. The AP2/EREBP members contain an AP2 DNA-binding domain of about 60 amino acids. There are 169 (as per the PTFDB) members in this family (represented in phylogenetic tree in figure 4.4). The AP2 superfamily is sub-divided into three families, the ERF, AP2 and RAV families. Proteins having single AP2 domain and less introns belong to the ERF family

(Figure 4.3). Proteins having tandem repetition of two AP2 domains e.g. APETALA2, AINTEGUMENTA, and RAP2.7 and a few proteins with a single AP2 domain having high similarity to the AP2 present in double-AP2 proteins are classified into AP2 family. Proteins that possess ERF domain along with a B3 domain belong to the RAV family e.g. RAV1 and RAV2. ERF proteins are further sub-divided into DREB and ERF subfamily. DREB1, DREB2, ABI4, TINY, RAP2.1 and RAP2.4 belongs to DREB subfamily, while the AtERFs, RAP2.2, RAP2.6 and RAP2.11 belongs to ERF subfamily.

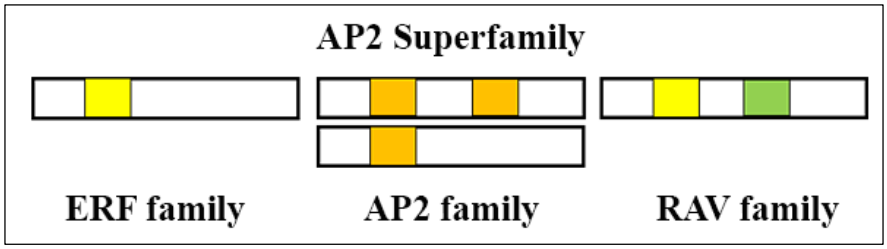


Figure 4.3: AP2 Superfamily. Domain structure of the APETALA2/Ethylene Responsive Factor (AP2/ERF) superfamily (Licausi et al., 2013a). Single AP2 domain proteins (ERF family), single or double ERF domain proteins (AP2 family), Proteins carrying one AP2 domain plus a B3 DNA binding domain (RAV family).

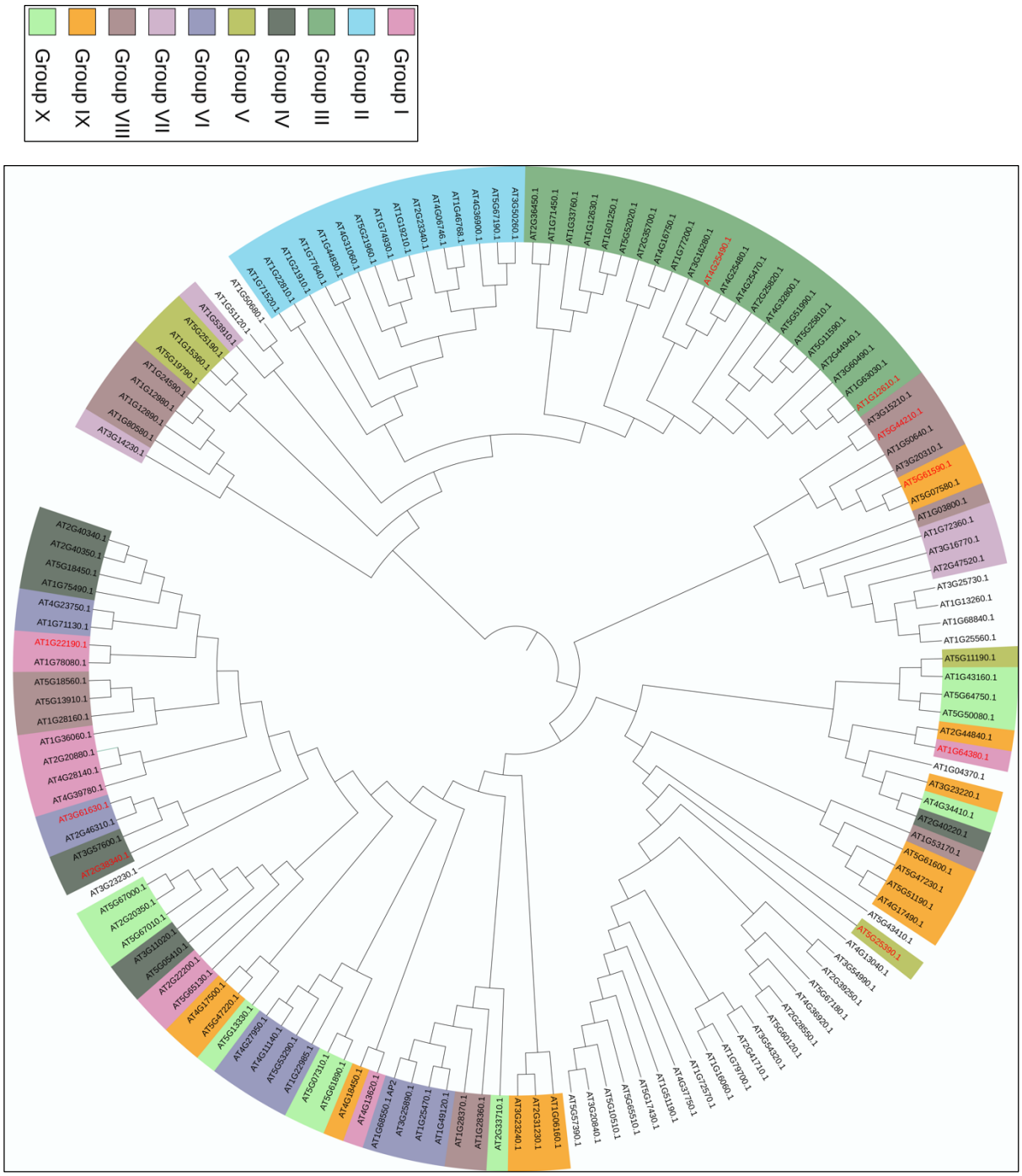


Figure 4.4: Phylogenetic tree for AP2/EREBP family. A circular phylogenetic tree showing all the members of the AP2/EREBP family of transcription factors in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). Members belonging to different groups (Licausi et al., 2013b) have been highlighted in different colours. TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

AP2/ERF

AP2/ERF members *DEWAX*/AT5G61590, *AT3G61630/CRF6,* and *AT5G25390/SHN2* are expressed in the epidermal and sub-epidermal layers of SAM. Among the AP2/ERF members, *CRF6* expression is induced in response to cytokinin signalling, and its role in inhibition of senescence of leaves has been shown previously. Recently, the role of *DEWAX* was shown in inhibition of cuticle biosynthesis in the dark. A role *AT5G25390/SHN2* has been reported in cuticle biosynthesis. Given the expression of these genes in epidermal layer, it is easy to speculate such a function. To analyse the developmental role of ***DEWAX***, T-DNA line (SALK_015182) was characterized by T-DNA PCR. The T-DNA insertion was found within the annotated exon of the *DEWAX*. Semi-quantitative-RT PCR revealed this line was null for *DEWAX* transcript (Figure 4.5). The mutant plants have slightly whitish green stem as compared to the WT. To study the gain of function, *35S: DEWAX* plants were characterized for ectopic expression and were analysed further. Developmental defect was not observed in any of these lines screened under normal growth condition so far, although gene was found to be over-expressed at the level of RNA (refer to chapter 3).

T-DNA insertion line (SALK_063548) of ***CRF6*** was analysed for insertion as well as for mRNA expression. Sequencing of T-DNA PCR product revealed insertion of the T-DNA 17bps upstream of ATG, within the 5'UTR of *CRF6*. Furthermore, this line was used to check the *CRF6* transcript level. Despite having insertion in the 5'UTR, expression level of *CRF6* was observed in T-DNA lines comparable to that in WT (Figure 4.5). Although, the transcript does get detected in semi-quantitative PCR, but due to the presence of a long T-DNA fragment just 17bps upstream of ATG, the functional protein might still not be made. However, plants carrying insertion in *CRF6* gene did not show any morphological phenotype and resembled to the WT. However, analysis of loss of function mutants have revealed that CRFs function redundantly in controlling embryo development and also development of leaves and cotyledons (Rashotte et al., 2006).

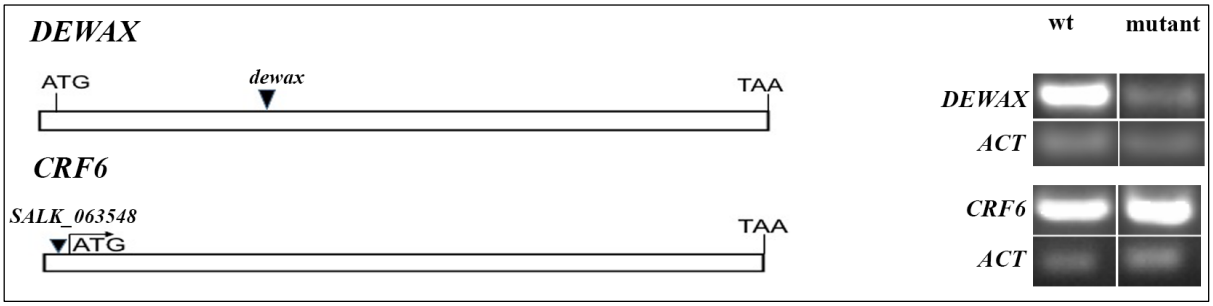


Figure 4.5: T-DNA characterisation of AP2/ERF family members. Genomic structure of *DEWAX and CRF6* genes and locations of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT-PCR experiments.

DREBs

DREB proteins are involved in regulating plant responses to various abiotic stresses. DREB proteins bind to an A/GCCGAC element. This *cis*-element is present in the promoters that responds to dehydration and therefore also termed as dehydration responsive element (DRE). DREB1A and DREB2A were the first proteins identified for this family (Sakuma et al., 2002). The genes encoding *DREB19/AT2G38340, RAP2.4/AT1G22190, DWARF AND DELAYED FLOWERING1* (*DDF1*)*/AT1G12610*, *CBF1/AT4G25490* and *AT1G64380* are enriched in epidermal and sub-epidermal cell layer.

DDF1 is a member of the DREB sub-family A1 of ERF/AP2 transcription factor family. A mutant line (SALK_114390) was screened for insertion in *DDF1*. The mutant line did not display any visible phenotype. T-DNA insertion was deduced within 300bps of the 5'end of *AT1G12610*, in the promoter region of the gene (Figure 4.6). RT PCR analysis does not show any down regulation in the level of transcript as compared to that in WT. However, studies in literature have shown that over-expression of this gene results in dwarf plants with delayed flowering phenotype, mainly due to reduced gibberellic acid biosynthesis (Magome et al., 2004).

DREB19/AT2G38340 is a member of the DREB subfamily A2 of AP2/ERF transcription factor family. SALK_144950C T-DNA line was screened for *DREB19*. The T-DNA was present in the 5' promoter region and lies within 300bps of the 5' end of *DREB19* (Figure 4.6). RT-PCR analysis revealed presence of significant *DREB19* transcript, comparable to that in WT.

RAP2.4/WIND2 belongs to the subgroup A6 of DREB sub-family. T-DNA line (SALK_139727) was ordered for *RAP2.4*. T-DNA insertion was found within the exon. There was a significant decrease in the level of *RAP2.4* transcript in comparison to WT control (Figure 4.6). I did not see any phenotype in the mutant line of *RAP2.4*.

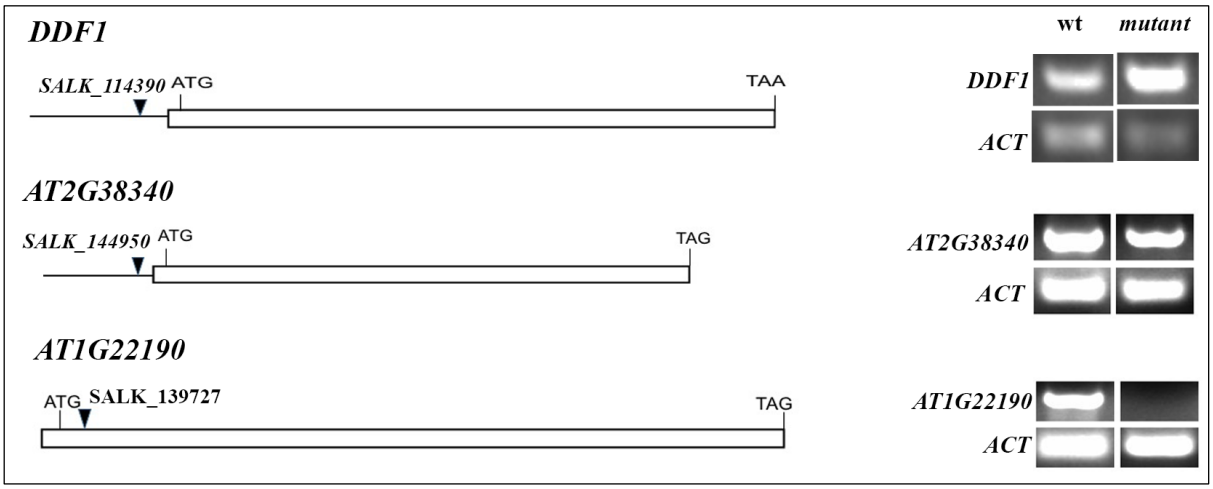


Figure 4.6: T-DNA characterisation of DREB family members. Genomic structure of *Arabidopsis* DREB genes and locations of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

ERF sub-family

The ERF subfamily is subdivided into twelve sub-groups I, II, III, IV, V, VI, VII, VIII, IX, X, VI-L, Xb-L (Licausi et al., 2013a). Members of the ERF sub-family bind to the GCC-box found in the promoter of genes involved in response to pathogens, ethylene and wounding (Ohme-Takagi and Shinshi, 1995; Solano et al., 1998). ERF family members are also involved in responding to biotic stresses to the plant. A member of ERF family, ***ERF9*** is also present in the epidermal layer of SAM and is a part of this study.

ERF9/AT5G44210, an ethylene response factor belongs to the B1 sub-family of the AP2/ERF family. SALK_091532 mutant line was used for *AT5G44210*. The mutant plant does not have any developmental or growth related defects. This line, has insertion within 300 bps of the 5' end of the genes, in the 5' UTR region. Equal transcript as that of WT was detected in mutant in RT PCR analysis (Figure 4.7).

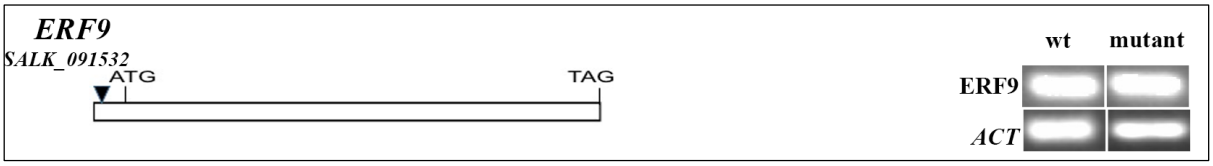


Figure 4.7: T-DNA characterisation of ERF family member, *ERF9*. Genomic structure of *Arabidopsis* ERF family gene, *ERF9* and location of T-DNA insertion. Exon is denoted by box. Arrow head indicates the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.2 ARID Transcription Factor Family

There are 11 members in this family of transcription factors in *Arabidopsis* (represented in the form of phylogenetic tree in figure 4.8).

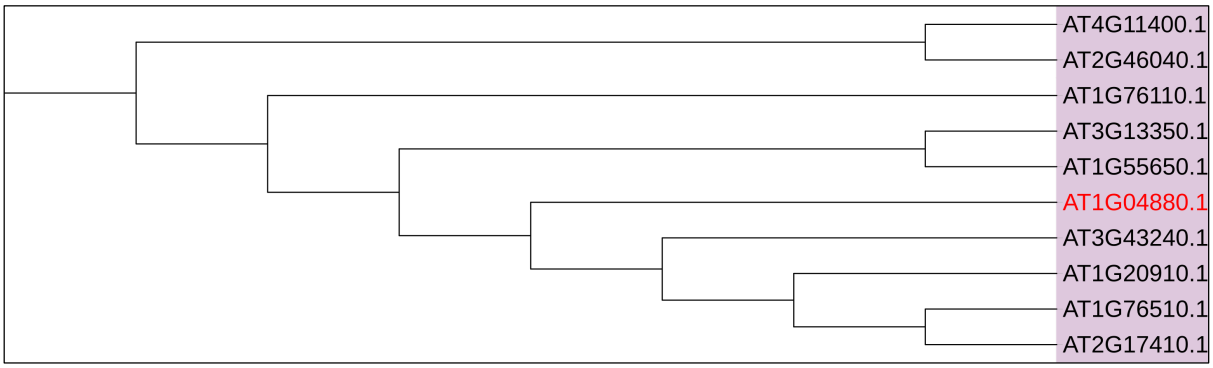


Figure 4.8: Phylogenetic tree for ARID TF family. Phylogenetic tree representing all the members of the ARID transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016).TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

ARID is a DNA binding module that was first identified in mouse B-cell transcription factor BRIGHT and the Drosophila DEAD RINGER protein. The ARID (A-T rich interaction domain) is a helix-turn-helix motif based DNA binding domain which is conserved in all eukaryotes. Another class of proteins is there, which in addition to an ARID domain also possess an HMG-box. The HMG proteins are most abundant and ubiquitous non-histone proteins found in nuclei. The HMG proteins have been divided into 3 families based on their structural features- HMG1/2, HMG 14/17 & HMG I/Y. A HMG1/2 family member, *AT1G04880*, is enriched in the epidermal layer of shoot apical meristem and the role of *AT1G04880* is known in pollen tube development via interaction with AGL66 and AGL104 (Xia et al., 2014). For deeper understanding of the role

of HMG1/2 TF in development, SALK_100002 mutant was analysed. *hmgb15-x* mutation is present in the annotated first intron of the gene. No downregulation in the transcript level of *HMG* was found in RT PCR experiment (Figure 4.9). The single mutant plants were indistinguishable from the WT plants.

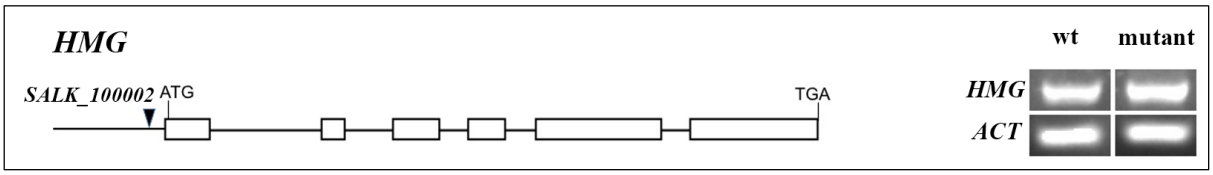


Figure 4.9: T-DNA characterisation of ARID TF family member, *HMG.* Genomic structure of *HMG* gene and location of T-DNA insertion. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.3 bHLH Transcription Factor Family

This is one of the largest transcription factor families in plants & is highly conserved in all the kingdoms (Toledo-Ortiz et al., 2003). bHLH domain comprises of a basic region required for DNA binding & a HLH region responsible for dimerization. This family has 225 members in *Arabidopsis* (as per the PTFDB) and phylogenetic tree in figure 4.10 represents all the bHLH proteins of *Arabidopsis.* The bHLH family is sub-divided into 12 major groups based on structural similarities (Heim et al., 2003a). Members of different groups are known to be involved in different functions. Such as group III members are known to regulate genes of flavonoid metabolism, group XII members act as positive regulators in early brassinosteroid signalling pathways. Group VII members are known to be important in controlling development of floral structures. For example, mutants of *spatula* have defective style and stigma and in stronger alleles, pollen transmitting tract tissue doesn't form (Heim et al., 2003a). In mutants of a member closely related to *SPATULA, ALCATRAZ,* siliques fail to dehisce because this gene is important for development of a non-lignified cell layer important for cell separation during dehiscence.

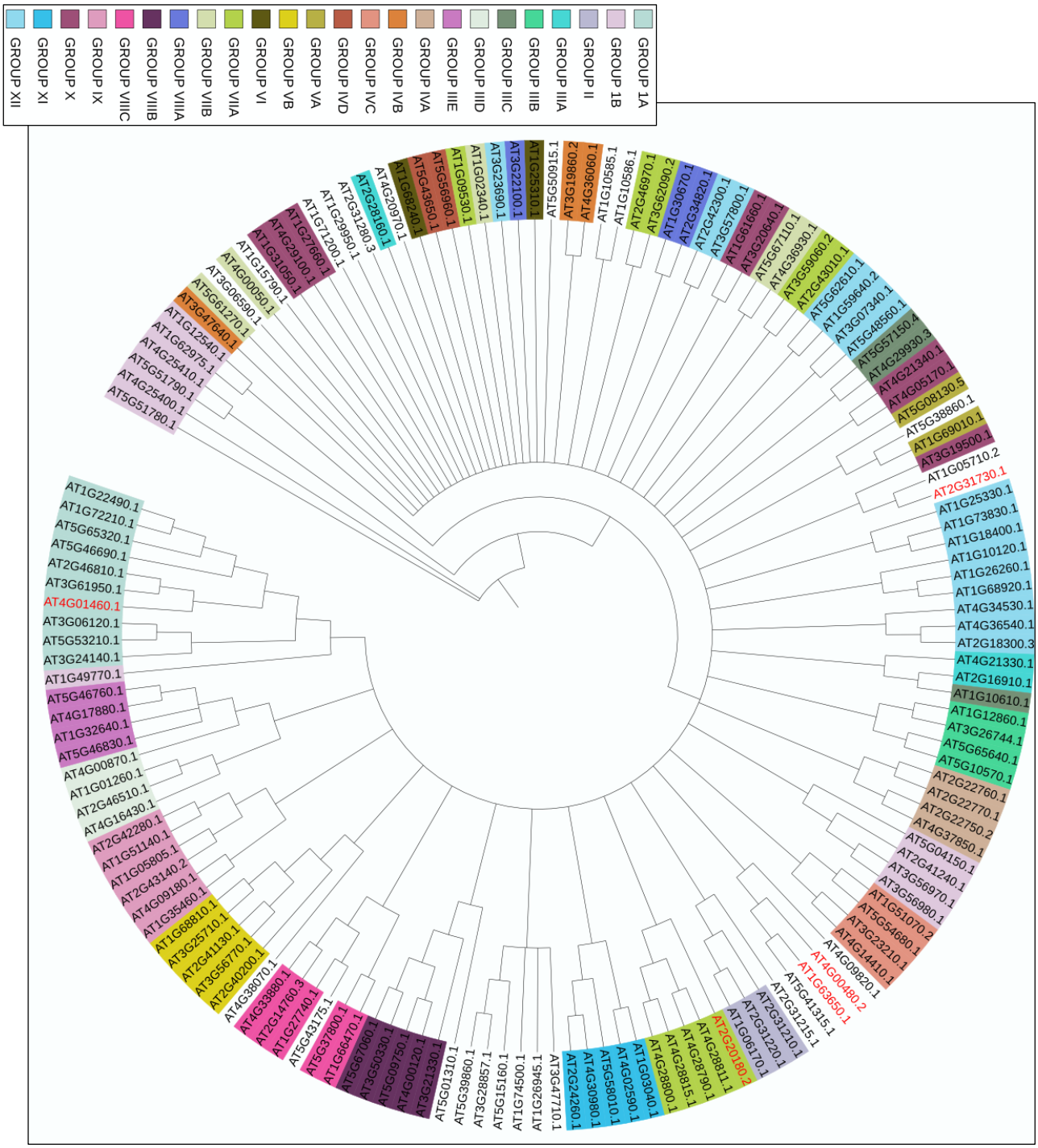


Figure 4.10: Phylogenetic tree for bHLH TF family. A circular phylogenetic tree representing all the members of the bHLH transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). Group I to group XII members (Heim et al., 2003b) have been highlighted in different colours. TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

Five of these bHLH TFs are expressed in L1 and L2 layers of the shoot. bHLH proteins recognize the E-box motifs (CANNTG), with CACGTG motif (G-box) being mostly exploited by the plant bHLHs. Most of the animal bHLH proteins function in developmental pathways such as myogenesis, neurogenesis or sex-determination. However, in plants, some of the bHLH family members in maize, such as LC, B-Peru & R-S are involved in regulating the expression of genes involved in anthocyanin biosynthesis pathway along with the C1/MYB family members. Genes encoding 5 bHLH TFs, ***EGL3/AT1G63650, PIL5/AT2G20180, AT2G31730***, ***AT4G00480, AT4G01460*** are expressed in the epidermal and sub-epidermal layers of the *Arabidopsis* shoot. *Phytochrome interacting factor 3-like 5/PIL5/AT2G20180* shows transcriptional activity in the dark and is known to be a negative regulator of seed germination. *PIL5* is known to be involved in photomorphogenic development of *Arabidopsis* via GA signalling. To further understand the role of ***PIL5*** in development of shoot, T-DNA (SALK_072677) mutant line was characterized by PCR. Sequencing analysis revealed the presence of T-DNA insert within the second exon of the gene. However, in RT-PCR analysis, a similar amount of transcript was detected as that in WT (Figure 4.11). Also, the plants did not show any developmental defect. T-DNA line was also characterized for another bHLH TF, ***AT2G31730.*** SALK_115465 mutant line carries the T-DNA insertion in the promoter of *AT2G31730* gene, approximately 1kb upstream of the translation start site (Figure 4.11). But no downregulation of transcript was found in the mutant plants as compared to the WT, suggesting it not to be true null mutant.

AT4G01460 is also a member of the bHLH TF family and is enriched in the sub-epidermal layer of shoot**.** T-DNA mutant line, SALK_027604 was screened for insertion. Sequencing analysis revealed the presence of T-DNA insertion within the intron of the gene. However, this mutant was confirmed null at the level of RNA as compared to its wild counterpart (Figure 4.11). But the mutant plant did not display any phenotypic changes as compared to the wild-type.

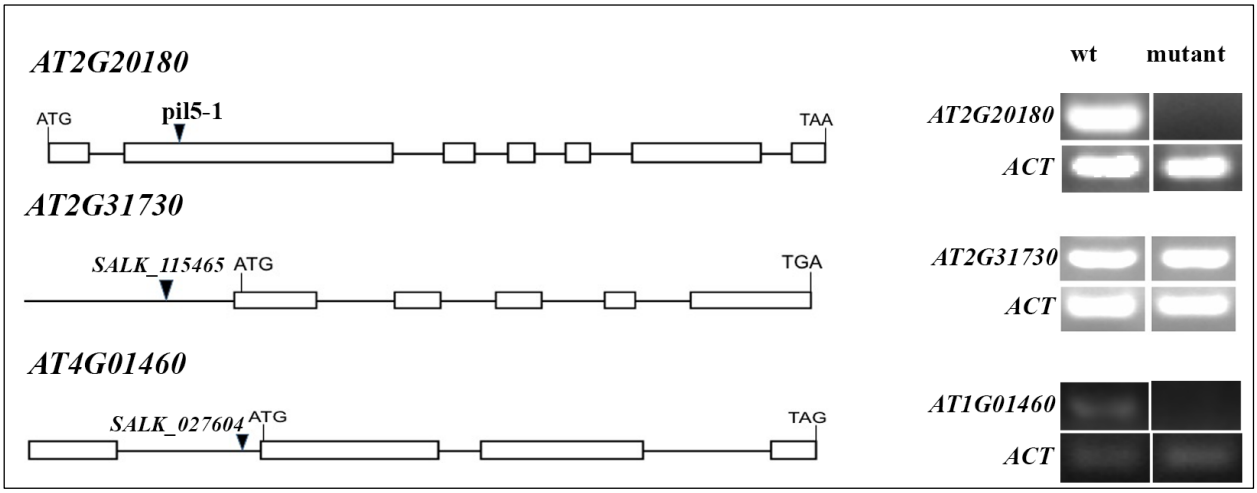


Figure 4.11: T-DNA characterisation of bHLH TF family members. Genomic structure of *Arabidopsis* bHLH family genes and locations of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.4 bZIP Transcription Factor Family

This family of transcription factors are characterized by the presence of a DNA binding and a dimerization bZIP domain. A stretch of 25 basic amino acid residues i.e. the basic region is present adjacent to the leucine zipper, where Leu residue appears every seventh residue over three to six repeat units. In *Arabidopsis,* a total of 127 members, represented in a phylogenetic tree in figure 4.12, are present within the bZIP TF family. Based on similar basic region & additional conserved motifs, bZIPs can be categorized into ten groups. Group A members are involved in ABA or stress signalling pathways (Jakoby et al., 2002). Group C members such as *OPAQUE2* and other related members from monocot suggest their role in regulating seed storage protein production. Group D members are involved in defense against pathogens as well as developmental processes, such as *bZIP46/PERIANTHA. PERIANTHIA (PAN)*, plays an important role in floral meristem control. *PAN* interacts with various developmental pathways, such as light, hormone and controls stem cell fate in flowers by controlling the expression of floral homeotic gene *AGAMOUS* (Das et al., 2009). *pan* mutants display alterations in number of floral organs and meristem size (Maier et al., 2011).Group G genes are involved in regulation of light responsive promoters. Group H members are known to have a role in promoting photomorphogenesis, such as *HY5. HY5*, a bZIP TF, which has a role in photo morphogenesis, is also involved in stimulus dependent regulation of root and hypocotyl growth. *hy5* mutant is

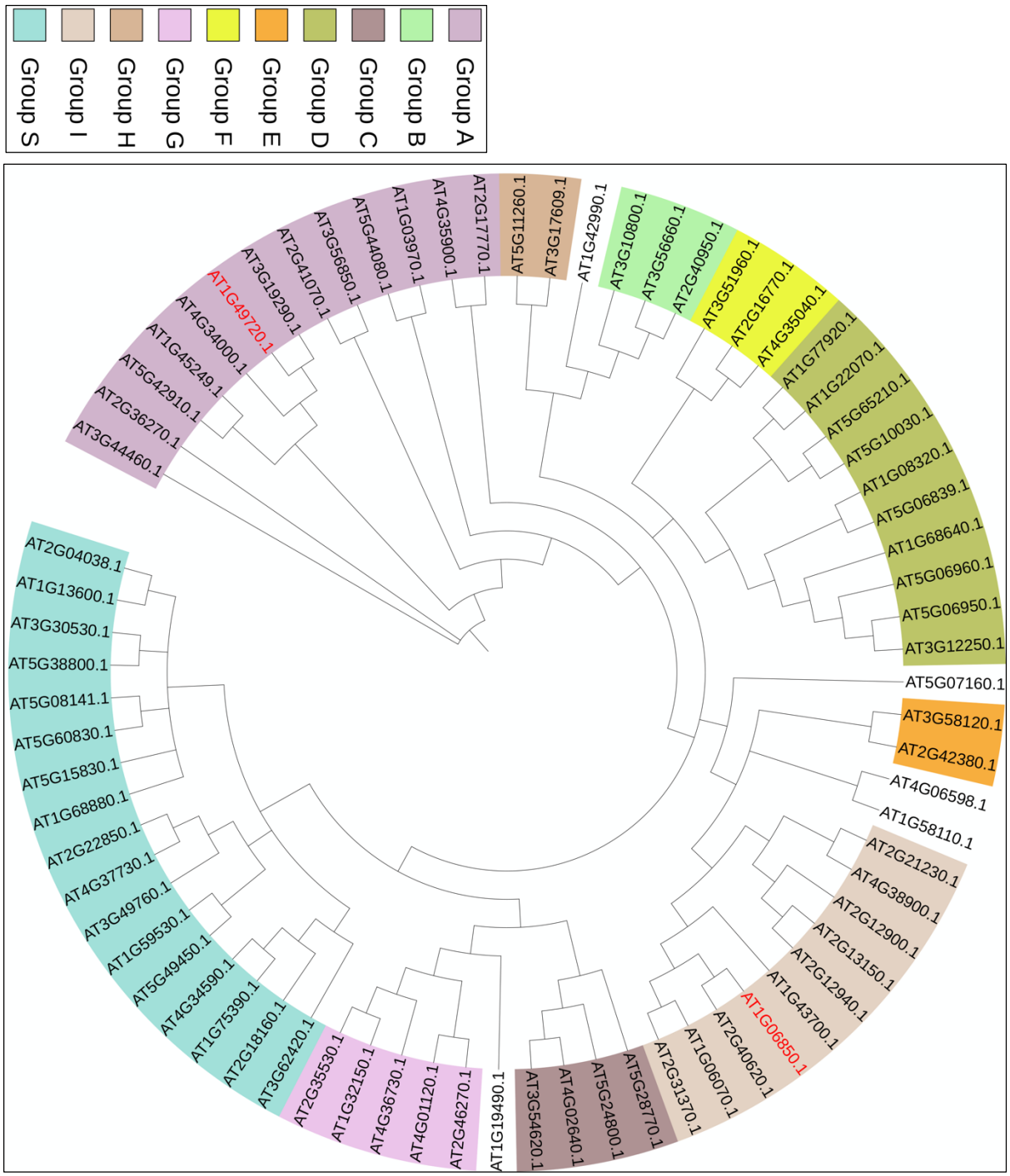


Figure 4.12: Phylogenetic tree for bZIP TF family. A circular phylogenetic tree representing all the members of the bZIP transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). Members belonging to different groups (Jakoby and Vicente-carbajosa, 2002) have been highlighted in different colours. TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

known to be deficient in a variety of stimulus responses, such as light dependent elongation of hypocotyl and gravitropic root growth (Oyama et al., 1997). Group I members play a role in vascular development. Group S members play a role in sucrose signalling and stress responses (Jakoby et al., 2002).

bZIP52/AT1G06850 is a member of the bZIP TF family that is enriched in the epidermal layer of shoot apical meristem. T-DNA insertion line, SALK_033320 was characterized for *bZIP52*. The T-DNA was found to be present within 300bps of the 5' end of the gene, in the 5'UTR annotated region (Figure 4.13). RT PCR data shows no downregulation in the transcript level of *bZIP52* as compared to WT, reflecting that it is not a true null mutant. Also, the mutant was morphologically indistinguishable from its WT counterpart.

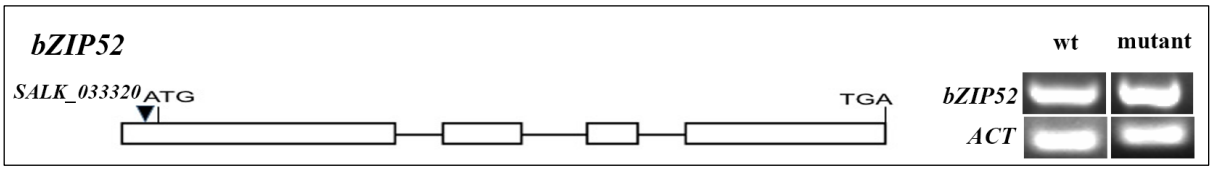


Figure 4.13: T-DNA characterisation of bZIP TF family members, *bZIP52* Genomic structure of *Arabidopsis bZIP52* gene and location of T-DNA insertion. Exons are represented by boxes and introns are indicated by lines. Arrow head indicates the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for the same. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.5 C2C2_CO-like Transcription Factor Family

Constans like gene family has recently been identified in *Arabidopsis* and other plant species. The first member identified was *CONSTANS*, which is a zinc finger protein involved in induction of flowering in long photoperiods. 17 members (as per PTFDB) of this family in *Arabidopsis* are shown in the phylogenetic tree in figure 4.14, with members of different groups highlighted in different colours.

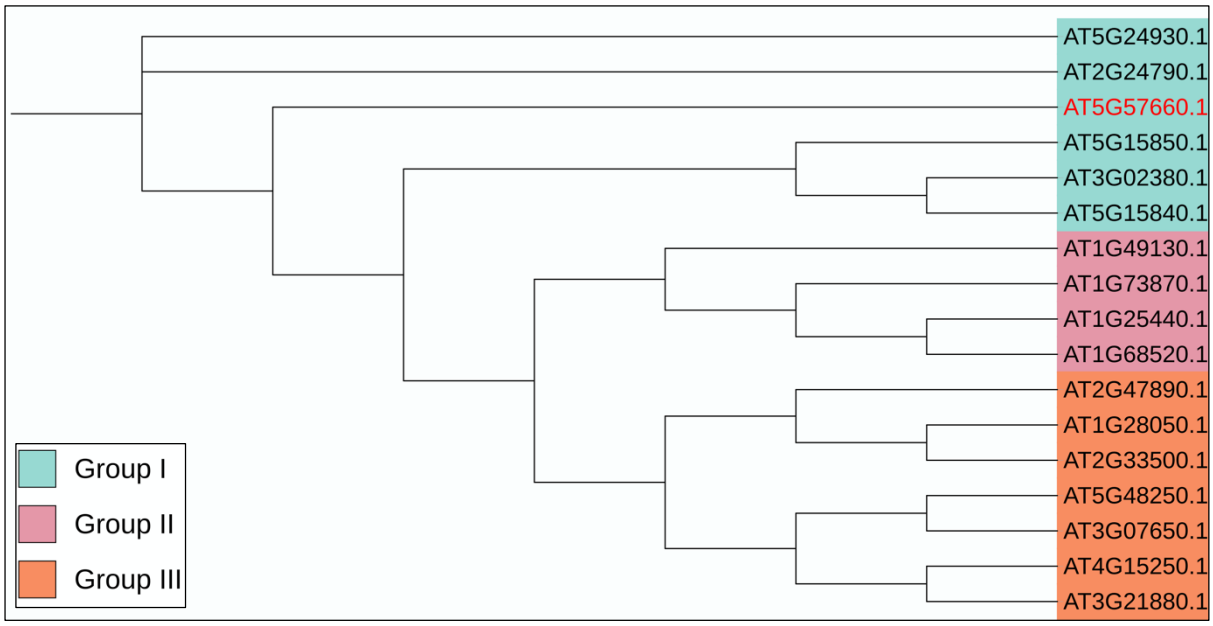


Figure 4.14: Phylogenetic tree for C2C2_CO TF family. A phylogenetic tree representing all the members of the C2C2_CO-like transcription factor family in *Arabidopsis.* Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). Members belonging to different groups (Griffiths et al., 2003) have been highlighted in different colours. *COL5*, a TFs enriched in the epidermal layer of the shoot is highlighted in red.

CO-like gene family has been sub-divided into three sub categories, CO and COL1 to COL5 (two B-box genes), COL6 to COL8 and COL16 (one B-box gene) and COL9 to COL15 (Robson et al., 2001). ***COL5/AT5G57660*** is a member of the C2C2_CO family that is enriched in the epidermal layer of the shoot and is represented in the phylogenetic tree, along with other members of this family in *Arabidopsis*.

4.3.2.6 C2C2_DOF Transcription Factor Family

This family of transcription factors is specific to the plants and contain a Dof domain (DNA-binding with one finger), which includes a single C_2-C_2 zinc finger. There are 36 members in this family in *Arabidopsis,* shown in figure 4.15. The first Dof mutant identified was for *DAG1*. The seeds of the *dag1* mutant were non-dormant and did not need light for germination (Papi et al., 2000). Another Dof mutant, *cog1* was identified from activation tagging pool screen. *COG1* acts as a negative regulator in both PhyA and PhyB signalling pathways. Plants over

expressing another DOF, *OBP3*, show severe growth defects, with yellow leaves and altered root development (Papi et al., 2000).

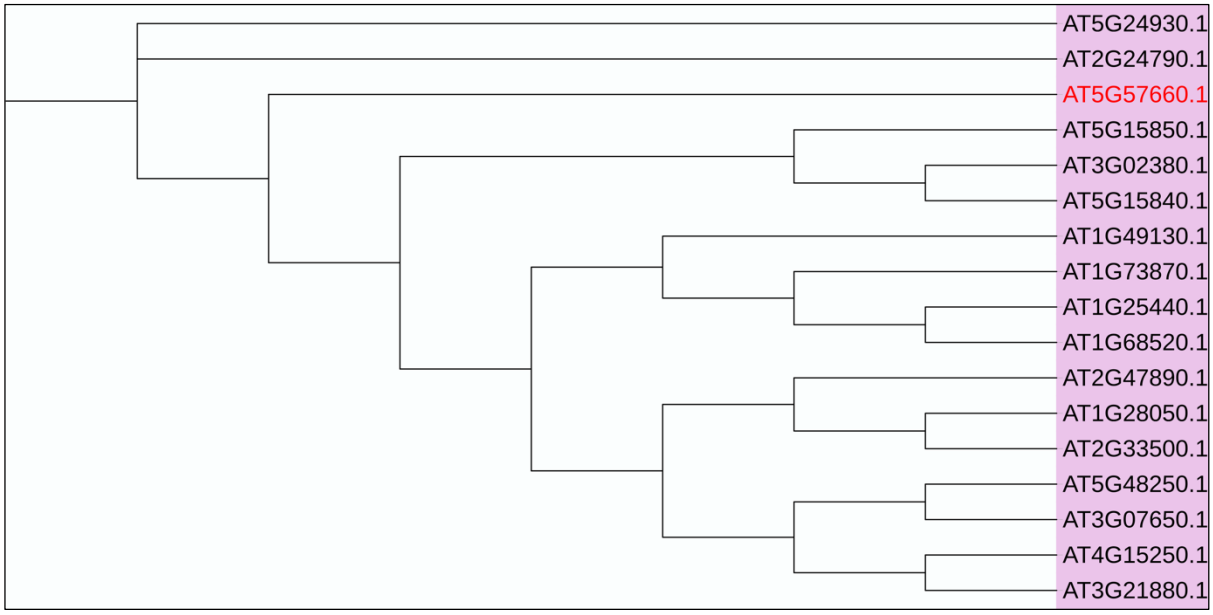


Figure 4.15: Phylogenetic tree for C2C2_Dof TF family. Phylogenetic tree representing all the members of the C2C2_Dof transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

Dof domain proteins are involved in various plant specific biological processes and can act as both activators and repressors. This family members contain a Dof domain which is usually present in the N-terminal region of the protein and outside of the Dof domain, amino acid sequences are very diverse (Shuichi, 2002). Dof proteins are known to bind to the AAAG core sequence in the promoters of the genes they control. Dof domain was initially recognized to be involved in DNA binding, but now is known to be a bifunctional domain involved in DNA binding as well as protein-protein interactions. Four TFs, ***COG1/AT1G29160*, *DOF2.4/AT2G37590*, *OBP2/AT1G07640* and *AT2G28810*** are enriched in the epidermal and sub-epidermal layers of the shoot. Recently, the role of *COG1* has been discovered in light perception and seed coat development. For understanding its role in shoot development, SALK_045465 mutant line of ***COG1*** was studied. T-DNA insertion was found in the 3' end of the gene, downstream of the stop codon. No downregulation in the expression of this gene was

found in mutant, reflecting it not to be a null mutant. ***OBP2*** is another member of the Dof TF family found in shoot. A T-DNA line, SALK_139462 was studied to identify insertion in *OBP2* gene. A detailed analysis of the mutant suggested the presence of T-DNA insertion within 500bps of the 5' end of the gene, in the annotated promoter region (Figure 4.16).

AT2G28810 is also a Dof type Zn finger DNA binding protein family member. SALK_056801 mutant line was analysed for the same. The mutant carried a T-DNA insertion within 500bps of the 5' end of the gene, in the annotated promoter region of the gene *AT2G28810.* In RT PCR data no downregulation of transcript was found in mutant plants (Figure 4.16). Mutant plants also morphologically resembled that of the WT.

Dof1.8/AT1G64620 is a C2C2-Dof family enriched in the CLV3 domain of the shoot. A T-DNA insertion line, SALK_130584 was characterized for presence of T-DNA insert as well as at the level of mRNA. T-DNA insertion was found in the coding sequence of the gene. RT PCR data showed downregulation in the transcript of *ATIG64620* in comparison to the WT (Figure 4.16). Mutant lines however displayed no phenotype.

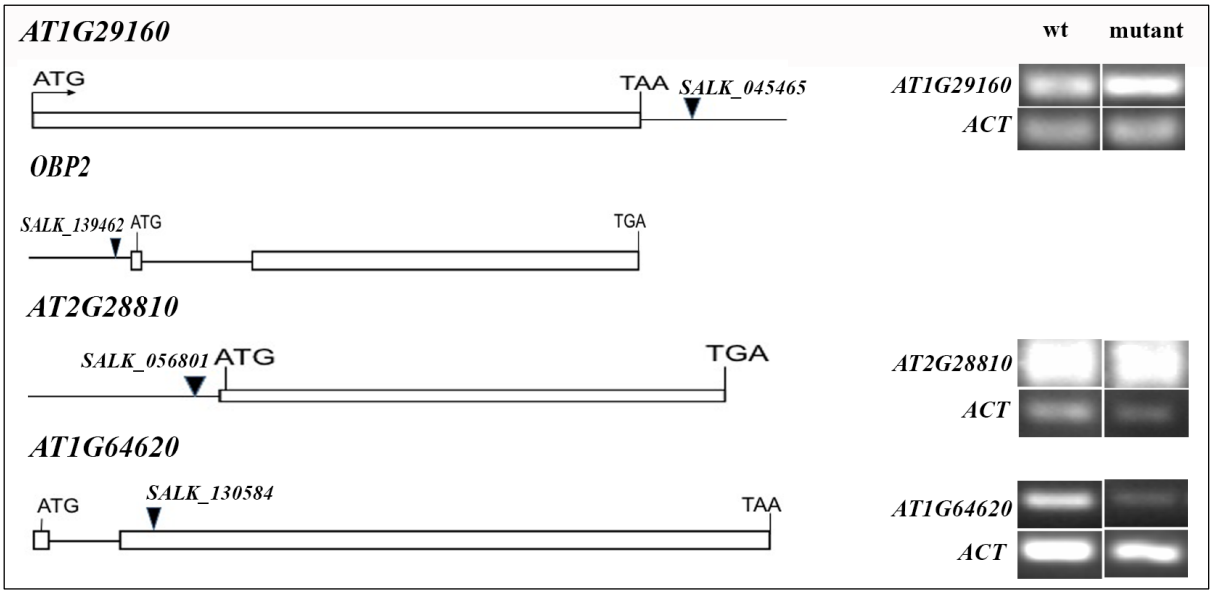


Figure 4.16: T-DNA characterisation of C2C2_Dof TF family members. Genomic structure of *Arabidopsis* C2C2_Dof genes and locations of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.7 C2H2 Transcription Factor Family

C2H2 TF family, shown in figure 4.17 in phylogenetic tree, is represented by 211 members in *Arabidopsis*.

Figure 4.17: Phylogenetic tree for C2H2 TF family. A circular phylogenetic tree representing all the members of the C2H2 transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

C2H2 proteins are involved in a variety of functions, such as DNA or RNA binding and protein-protein interactions. Therefore, this class of proteins are involved in transcriptional regulation, by either modifying the chromatin or site specific modification, in RNA metabolism and also other functions of the cell which require specific contacts of zinc finger domains. Seven members of this family, ***AZF3/AT5G43170, HDA3/AT3G44750, STOP1/AT1G34370, AT1G75710, AT4G16610***, ***AT5G54630*** *and* ***AT5G61190*** are enriched in the epidermal and sub-epidermal layers of the shoot. For understanding the function of these TFs, T-DNA lines were characterized for the same. T-DNA line, SALK_048268 was studied for ***AT1G75710***. Sequencing analysis revealed the presence of T-DNA insertion within the annotated exon of the gene. Also, significant downregulation was found in the transcript of *AT1G75710* gene in mutant versus WT plants (Figure 4.19). A multitude of phenotypes were observed in the mutant of this gene. Firstly, as compared to the WT plants, the vegetative phase in the mutant plants was quite prolonged. The mutant plants transited from the vegetative to the reproductive phase much later than their WT counterparts (Figure 4.18). Also, mutant was much more sturdy and tall growing. Spiral pattern of phyllotaxy was also not maintained in the mutant and flower primordia were not positioned at 137.50 angle in the shoot apical meristem, as evident from the arrangement of siliques on the adult plant. The siliques produced were sterile i.e. viable seeds were not present within the siliques. To investigate this further, male and female organs of the flower were looked in detail. While the female organ, the carpel, looked morphologically correct, the male organ was found to be defective. Pollen fertility was low as compared to the WT plants. Also, as compared to 6 anthers mostly in the WT plants, there were only 5 or less anthers consistently present in the mutant flowers and anthers were sometimes fused (Figure 4.18). Size of one of the sepals was also unusually large, which covered the remaining organs of the flower, thereby interfering with their normal development. The petals of the mutant flowers showed aberrant morphology with unusual twisting (Figure 4.18). ***AT5G54630*** is another member of the C2H2 TF family and its function was studied by analysing the T-DNA insertion line, SALK_098531. The insertion was found to be present in the first exon of the gene. Full length transcript was significantly reduced in the mutant plant as compared to WT (Figure 4.19). This T-DNA represents a true null for *AT5G54630*. But, no growth or development related defect was found in the single mutant plants. This may be due large redundancy among transcription factors in *Arabidopsis.*

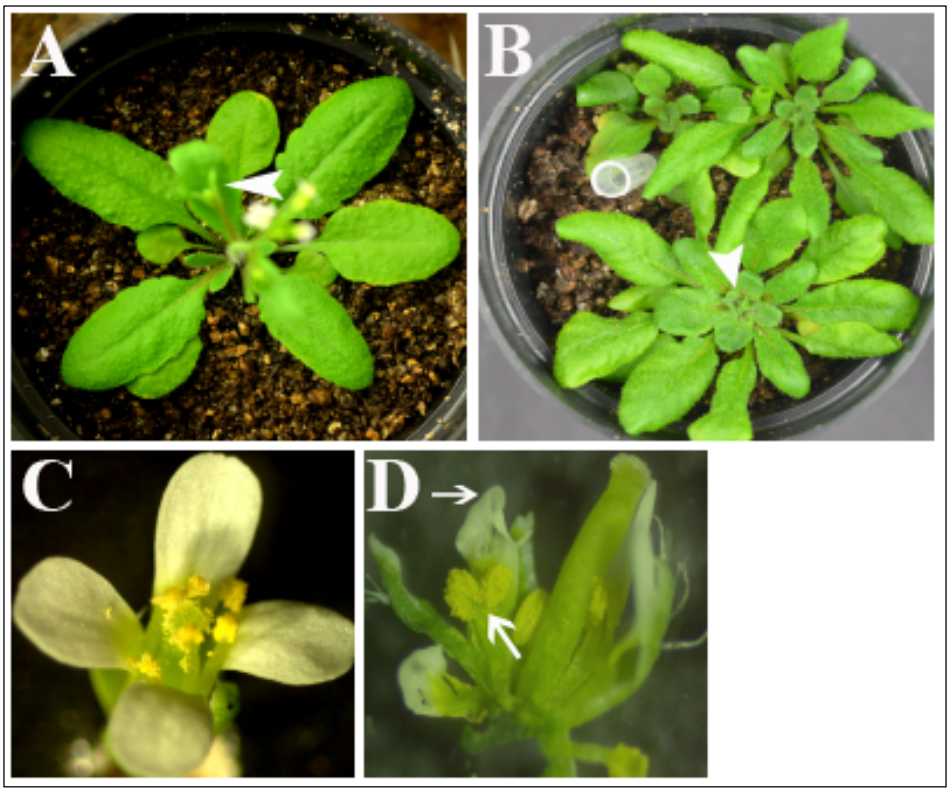


Figure 4.18: Mutant phenotype of *at1g75710.* (A) *Arabidopsis thaliana* WT Col-0 plant. The white arrowhead denotes the onset of reproductive phase in WT plant. (B) Mutant plant of *at1g75710.* The mutant has more number of rosette leaves and white arrow head indicates delayed onset of reproductive phase. (C) Flower of WT Col-0 plant. (D) Aberrant flower of *at1g75710* mutant. The petals are deformed and anthers are fused, as indicated by white arrows.

Another C2H2 family member, ***AT5G61190*** is enriched in the shoot. SALK_010662 mutant line was screened for the same. Sequencing analysis revealed the presence of T-DNA within the 5' UTR region of the gene present upstream of ATG. However, no downregulation at the level of transcript was detected in mutant background in comparison to the wild (Figure 4.19). Also, the mutant plants morphologically resembled the WT counterpart.

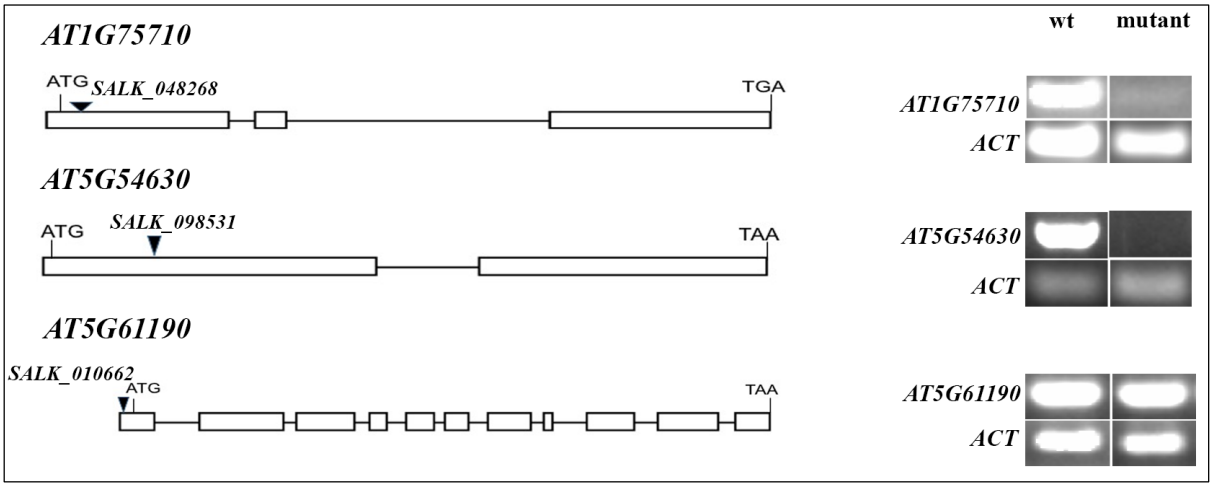


Figure 4.19: T-DNA characterisation of C2H2 TF family members. Genomic structure of *Arabidopsis* C2H2 genes and locations of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.8 CAMTA Transcription Factor Family

cDNA expression libraries derived from plants exposed to stress conditions were screened with 35S labelled recombinant calmodulin as probe. This revealed a new family of proteins that contained a transcription activation domain and DNA binding domain of two kinds, the CG-1 domain and the TF immunoglobulin domain, ankyrin repeats and IQ calmodulin binding motifs. Based on amino acid similarities and domain organization, similar proteins with the same domain organization were identified in other multicellular organisms including *C. elegans*, humans and Drosophila. This family of proteins was designated as calmodulin binding transcription activators (CAMTAs) (Bouché et al., 2002). Members of this family should possess a CG-1 IQ domain. There are six members in *Arabidopsis* that belong to CAMTA family and are shown in figure 4.20 in the form of a phylogenetic tree constructed using maximum likelihood method. CAMTA genes in *Arabidopsis* respond rapidly to various environmental cues such as drought, high salinity, heat, cold, UV and signalling intermediates such as H2O2, methyl jasmonate and phytohormones, ethylene, ABA and salicylic acid (Finkler et al., 2007). These rapid responses of CAMTA genes, suggest their role in cross-talk among multiple signalling pathways involved in response to stress.

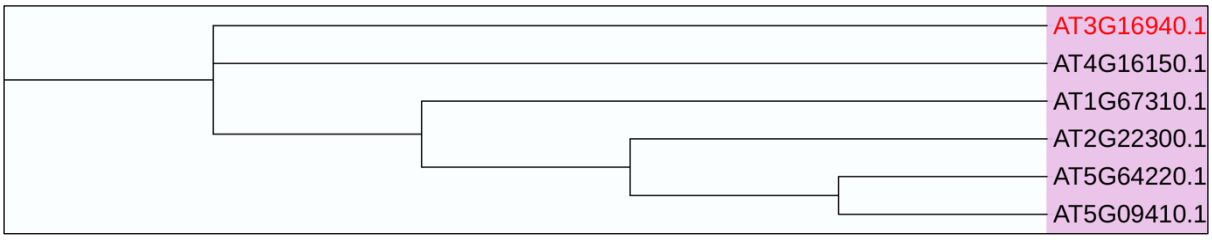


Figure 4.20: Phylogenetic tree for CAMTA TF family. A phylogenetic tree representing all the members of the CAMTA transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). *CAMTA6*, a TF enriched in the epidermal layer of the shoot is highlighted in red.

AT3G16940, a member of the CAMTA family is expressed in the epidermal layer of the shoot. A T-DNA mutant line, SALK_078900 was studied for *AT3G16940*. T-DNA insertion was found to be present in the 8th exon of the gene (Figure 4.21). No transcript was detected in RT PCR analysis of this gene, suggesting it to be null mutant. However, mutant plants were phenotypically similar as the WT, possibly due to redundancy among TFs. And since, CAMTA TFs are involved in responding to cold stress, possibly the mutant might be more susceptible to cold conditions and its developmental trajectory might also need to be studied under cold conditions.

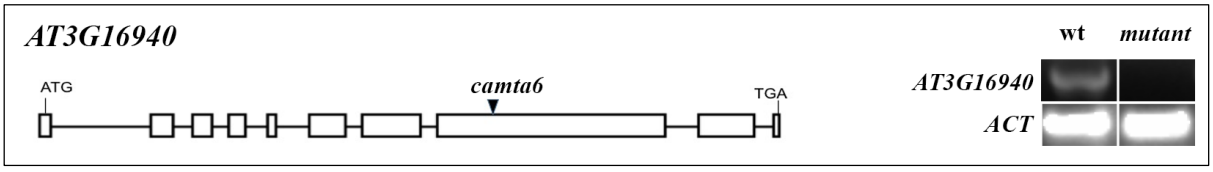


Figure 4.21: T-DNA characterisation of CAMTA TF family members, *CAMTA6*. Genomic structure of *Arabidopsis* CAMTA gene, *AT3G16940* and location of T-DNA insertion. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.9 CCAAT_HAP2 Transcription Factor Family

CCAAT box binding factors (CBF) or Heme associated proteins (HAPs) or Nuclear Factory Y (NF-Y) transcription factor family is found in all eukaryotes. NF-Ys bind to the CCAAT site on the DNA as heterotrimeric complexes composed of single subunits from each of the three protein families, NF-YA, NF-YB and NF-YC. Because of the heterotrimeric nature of NF-Y family, it provides flexibility and allows for the formation of a large number of combinations of TFs, that

may be effective under various environmental conditions and may be exploited by the plants for growth and development in adverse conditions. NF-YB and NF-YC initially form a dimer in the cytoplasm and then translocate to the nucleus, where they interact with NF-YA and then bind to the CCAAT box (Frontini et al., 2004; Steidl et al., 2004). This family of TFs are known to play important roles in embryogenesis. The first identified member involved in embryogenesis was *NF-YB9*, which was later identified as *LEAFY COTYLEDON1(LEC1)* (Lee et al., 2003; Lotan et al., 1998; West, 1994). *NF-YB9/LEC1* helps in maintaining the embryonic cells and prevents immature germination of seeds. NF-Y genes also play role in photoperiod dependent control of flowering. NF-YB2 and NF-YB3 act as activators of *FLOWERING LOCUS T (FT),* a gene required for vegetative to floral meristem transition. *nf-yb2 nf-yb3* mutants also show the late flowering phenotype. NF-YA subunits are also known to play redundant roles in seed germination and embryo development. NF-YB and NF-YC subunits also physically interact with CO, a major regulator of photoperiod induced flowering time (Petroni et al., 2012). NF-YC subunits are also required for seed germination through ABA responses. There are a total of 43 members in this family in *Arabidopsis*, out of which three are enriched in the shoot, namely ***NF-YA2/AT3G05690, NF-YA5/AT1G54160, NF-YA10/AT5G06510***. *NF-YA5* is involved in drought resistance, as its expression is strongly induced by osmotic stress, salt stress and drought stress (Petroni et al., 2012). *NF-YA2* and *NF-YA10* are known to be involved in regulation of leaf growth via auxin signalling (Zhang et al., 2017). All the members of this family are represented in a phylogenetic tree in figure 4.22.

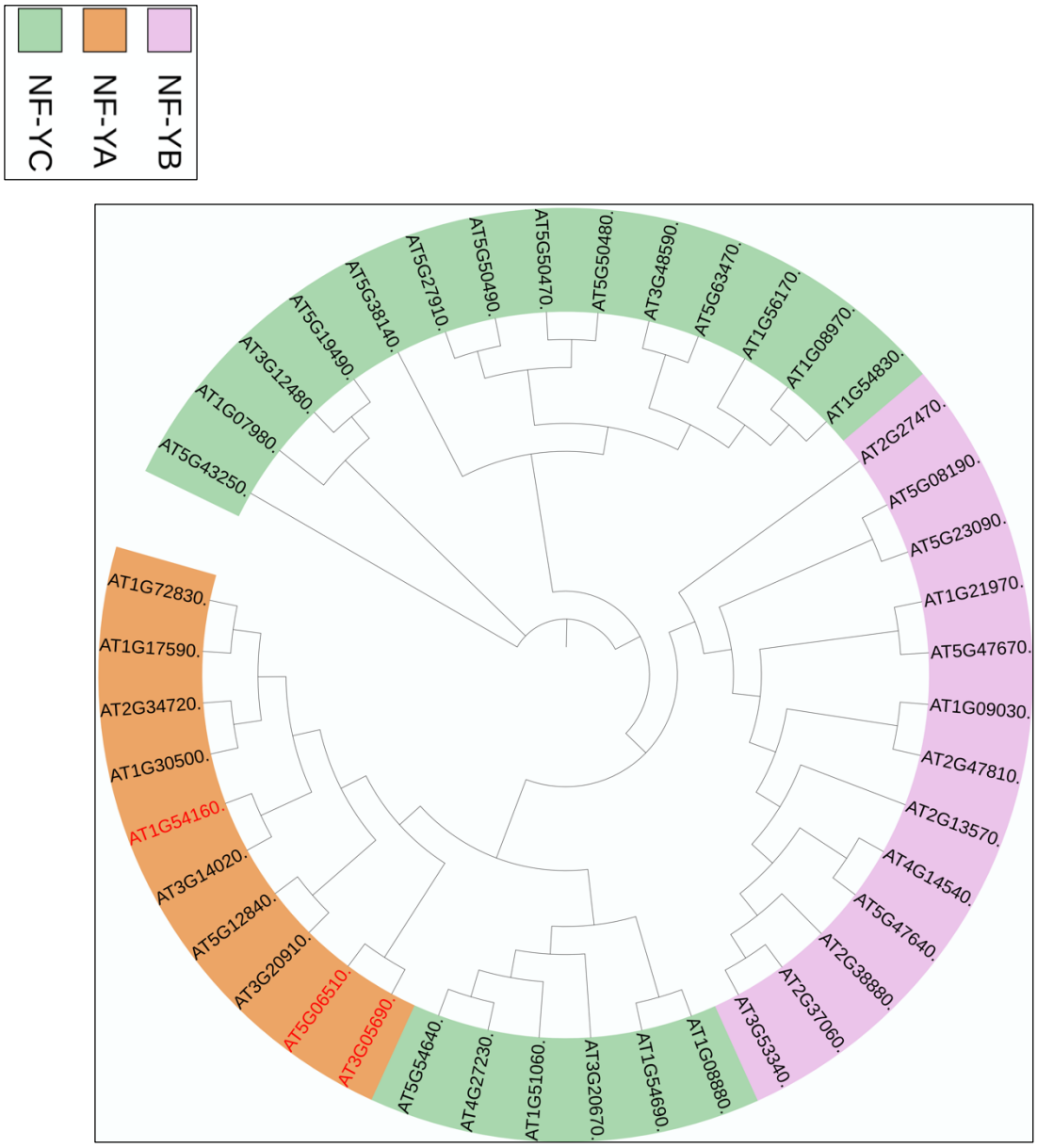


Figure 4.22: Phylogenetic tree for CCAAT_HAP2 TF family. A circular phylogenetic tree representing all the members of the CCAAT_HAP2 transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). The three different protein families, NF-YA, NF-YB, NF-YC have been highlighted in different colours. TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

NF-YA5 is a member of the CBF-B/NF-YA family and for understanding its function, its T-DNA mutant, SALK_042760 was analysed. Sequencing results revealed the presence of T-DNA insertion within the intron of the gene (Figure 4.23). However, RT-PCR reflected presence of no

transcript in the mutant background, reflecting it to be true null for this gene. No phenotype was recorded for the single mutant.

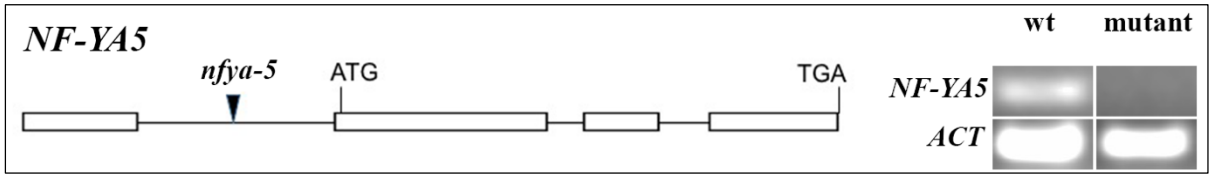


Figure 4.23: T-DNA characterisation of CCAAT_HAP2 TF family member, *NF-YA5*. Genomic structure of *Arabidopsis* CCAAT_HAP2 gene, *NF-YA5* and location of T-DNA insertion. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.10 CPP (Cysteine rich polycomb like proteins) Transcription Factor Family

Nodulin genes are found to be expressed specifically in nitrogen fixing root nodules. A novel DNA binding protein was identified (CPP1), which was found to interact with the leghaemoglobin gene, *Gmlbc3* of the Soybean. CPP1 DNA binding domain contains two similar Cys rich domains with 9 and 10 Cys, respectively (Cvitanich et al., 2000). Proteins with similar domains were also identified in other organisms such as *C. elegans*, humans, mouse and *Arabidopsis*. Fig 4.24 represents the phylogenetic tree of all the members of the CPP family in *Arabidopsis*.

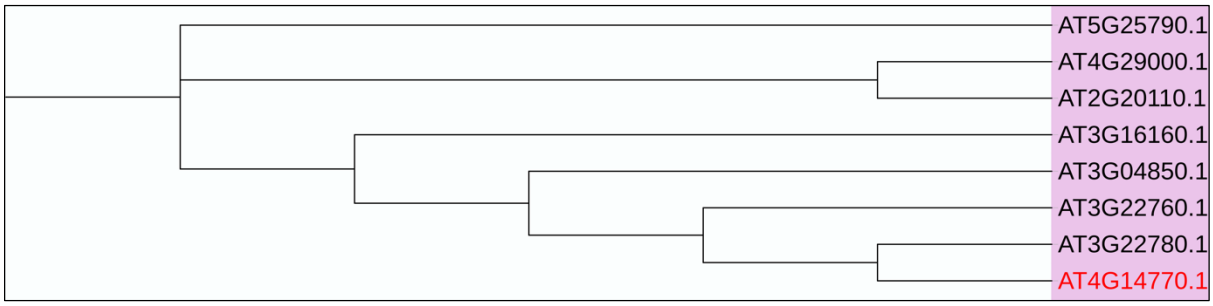


Figure 4.24: Phylogenetic tree for CPP TF family. A phylogenetic tree representing all the members of the CPP transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). *TCX2*, a TF enriched in the sub-epidermal layer of the shoot is highlighted in red.

CPP domain also shares homology with Cys rich region present in some of the polycomb proteins. Proteins with such domains are a part of the CPP family of TFs. There are 8 members in the CPP family in *Arabidopsis thaliana,* of which one CPP protein, ***TCX2/AT4G14770*** is enriched in the sub-epidermal layer of the shoot. T-DNA mutant, SALK_021952 was screened for *TCX2*. Insertion was found in the first intron of the gene (Figure 4.25). A significant downregulation in the level of the mRNA was also detected in the mutant as compared to the WT.

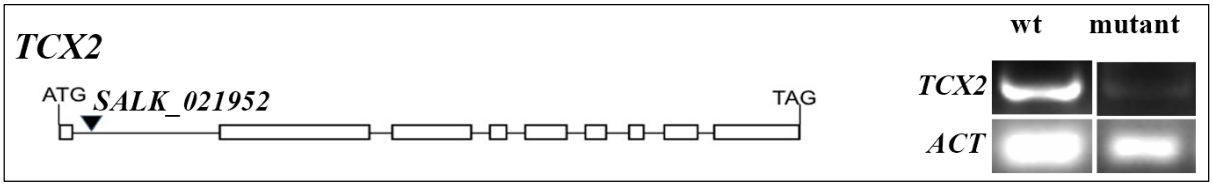


Figure 4.25: T-DNA characterisation of CPP TF family member, *TCX2.* Genomic structure of *Arabidopsis* CPP family gene, *TCX2* and location of T-DNA insertion. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.11 E2F-DP Transcription Factor Family

E2F-DP transcription factors are important members of the cyclin D/retinoblastoma/E2F pathway. These TFs regulate the expression of genes involved in G1/S transition and S-phase progression. Structurally and functionally, the AtE2Fs can be divided into 2 groups. Three members of the first group, E2Fa, E2Fb, E2Fc possess all the domains which are found in other animal and plant E2Fs. The other AtE2Fs are novel proteins with duplicated DNA binding domain but no other conserved region (Mariconti et al., 2002). E2Fs of the first group, interact specifically with AtDP proteins and bind to the E2F *cis* elements. However, members of the group II can bind directly to the DNA without interacting with the DP proteins (Mariconti et al., 2002). This family has 7 members in *Arabidopsis.* One member of this family, ***DEL2/AT5G14960*****,** is enriched in the sub-epidermal layer of the shoot and is represented in the phylogenetic tree along with other members of the family in figure 4.26.

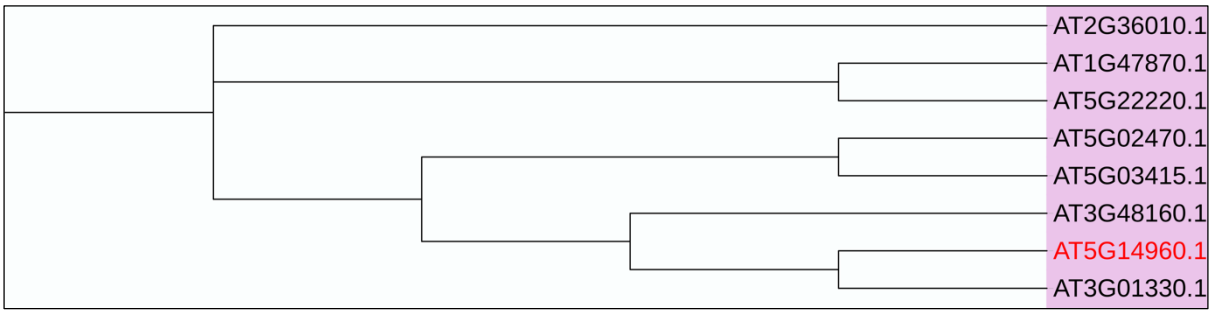


Figure 4.26: Phylogenetic tree for E2F-DP TF family. A phylogenetic tree representing all the members of the E2F-DP transcription factor family in *Arabidopsis.* Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). *DEL2*, a TF enriched in the sub-epidermal layer of the shoot is highlighted in red.

A T-DNA mutant, SALK_093190 of *DEL2* was studied in detail and sequencing revealed the presence of T-DNA insert in the 7th exon of the gene (Figure 4.27). However, the transcript levels for this gene couldn't be determined by semi-quantitative RT PCR.

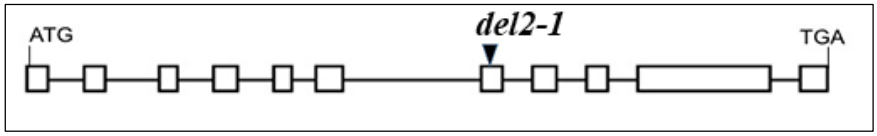


Figure 4.27: T-DNA characterisation of E2F-DP TF family members, *DEL2.* Genomic structure of *Arabidopsis* E2F-DP gene, *DEL2* and location of T-DNA insertion. Exons are denoted by boxes and introns are indicated by lines. Arrow head indicates the position of T-DNA insertion in the 7th exon of the gene.

4.3.2.12 EIL Transcription Factor Family

Ethylene-insensitive3 (EIN3) and EIN3-like (EIL) proteins are transcription factors involved in ethylene signalling in higher plants. *ein3* mutants are insensitive to ethylene and show loss of ethylene mediated effects, such as gene expression, cell growth inhibition and accelerated senescence (Chao et al., 1997). *EIN3* acts downstream of the histidine kinase ethylene receptor, *ETR1*. In *Arabidopsis thaliana* , there are six members in the EIL family and one EIL TF, ***EIL1/AT2G27050***, is expressed in the epidermal layer of the shoot. Complete EIL family phylogenetic tree has been shown in figure 4.28.

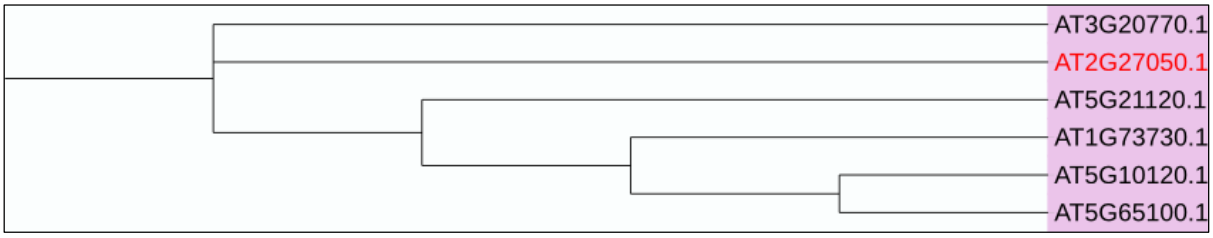


Figure 4.28: Phylogenetic tree for EIL TF family. A phylogenetic tree representing all the members of the EIL transcription factor family in *Arabidopsis.* Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). *EIL1*, a TF enriched in the epidermal layer of the shoot is highlighted in red.

To study the function of ***EIL1,*** SALK_049679 mutant line was characterized in detail. Gene was found to be disrupted by T-DNA insertion in the only exon of the gene. In *eil1-3,* no transcript was detected in comparison with the WT plants (Figure 4.29). This possibly represents a null mutant of this gene. However, the plant does not show any defects in growth and development.

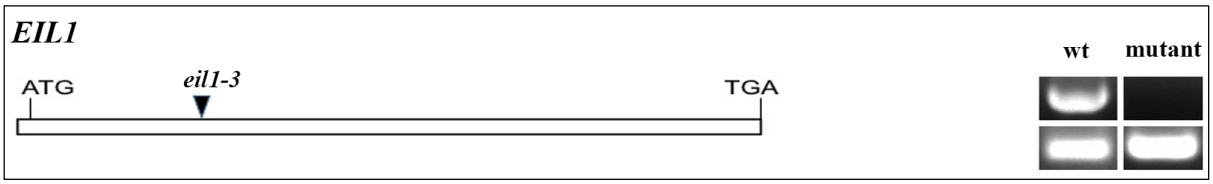


Figure 4.29: T-DNA characterisation of EIL TF family members, *EIL1.* Genomic structure of *Arabidopsis* EIL family gene, *EIL1* and location of T-DNA insertion. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.13 GRF Transcription factor family

GRF proteins contain a highly conserved WRC and QLQ domains in their N-terminal regions. Similar domains are also conserved in the Growth regulating factor of rice, *OsGRF1*. Based on C-terminal amino acid sequence similarity, phylogenetic analysis showed that AtGRF proteins can be divided into 5 categories. AtGRF1 and AtGRF2 (44% amino acid identity in their N-terminal region); AtGRF3 and AtGRF4 having 54% amino acid identity; AtGRF5 and AtGRF6

having 16% identity; and AtGRF7 and AtGRF8 having 17% identity. AtGRF9 stands out as an isolated branch on the phylogenetic tree and also has a WRC domain. The phylogenetic tree in figure 4.30 shows all the 9 members of this family present in *Arabidopsis* .

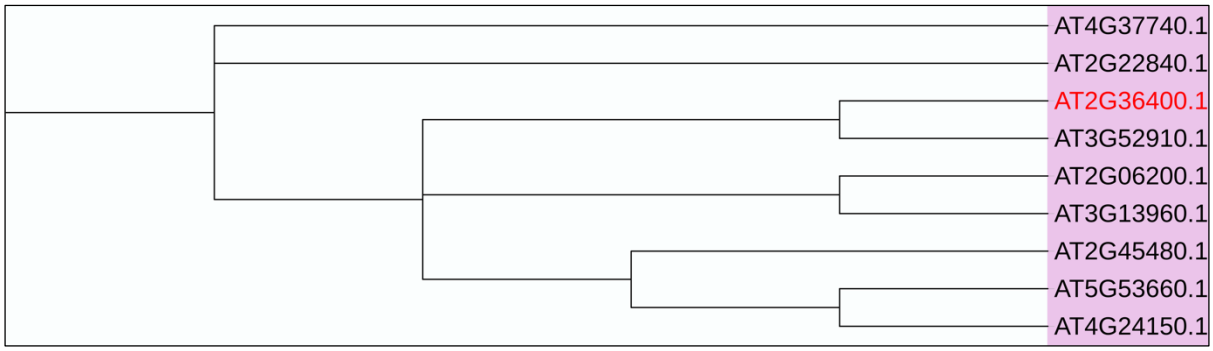


Figure 4.30: Phylogenetic tree for GRF TF family. A phylogenetic tree representing all the members of the GRF transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). *GRF3*, that is enriched in the CLV3 domain of the shoot is highlighted in red.

GRF3/AT2G36400 is a GRF family member enriched in the CLV domain of the shoot. Role of GRF TFs has been shown in regulating cell size in the leaf and root development. To understand the function of *GRF3* in shoot development, T-DNA mutant, SALK_026786 was analysed. *grf3-1* insertion is present within the promoter of the gene *AT2G36400*. Single mutant of *GRF3* doesn't show any growth defects. Slight downregulation of transcript was found in the mutant plant as compared to WT (Figure 4.31). Analysis of the transcript by real time PCR revealed a fivefold reduction in the mutant in comparison to the WT. Suggesting, that sometimes real time PCRs are important to reveal the actual downregulation in the level of transcripts.

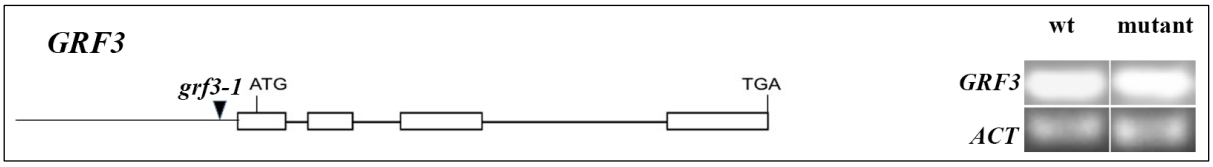


Figure 4.31: T-DNA characterisation of GRF TF family member, *GRF3*. Genomic structure of *Arabidopsis* GRF family gene, *GRF3* and location of T-DNA insertion. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.14 Homeobox Transcription Factor Family

This class of proteins is unique to the plants because of the closely linked leucine zipper motif to the homeodomain (Ariel et al., 2007). There are 109 (TFDB) members in this family in *Arabidopsis,* all members represented in phylogenetic tree in figure 4.32. Based on the sequence homology of the homeodomain, these proteins have been divided into four families, from HD-ZIP I to IV. HD-ZIP proteins bind in dimeric form to the palindromic DNA sequences containing the motif AATNATT. Plant homeodomain proteins also possess an additional domain apart from HD DBD. These are homeodomain leucine zipper (HD-ZIP), plant homeodomain with a finger domain (PHD-HD), bell domain (Bell), Zn-finger with homeodomain (ZF-HD), Knotted homeobox (KNOX) and Wuschel homeobox (WOX) (Ariel et al., 2007). Homeodomain is a specific DNA binding domain and homeodomain proteins are involved in various biological functions. Homeodomain is composed of 60 amino acids that folds into a stable globular structure, allowing it to bind to the DNA (Gehring et al., 1994).

The first homeodomain protein discovered in plants was *KNOTTED1(KN1)* in maize (Vollbrecht et al., 1991). *KNOTTED1* is involved in regulating developmental program, determining cell fate. Mis-expression of *KN1* and related genes, such as *OsH1* in rice, *HvKnox3* in barley and *KNAT* in Arabidopsis have shown to result in highly abnormal leaf morphology (Lincoln et al., 1994; Matsuoka et al., 1995; Müller et al., 1995; Sinha et al., 1993). In maize, *rough sheath 1* mutation causes unregulated cell division and expansion in leaves (Chan et al., 1998). *stm* mutation in *Arabidopsis* produces seedlings that fail to form a shoot apical meristem during embryogenesis. A PHD finger protein, *GL2* is required for normal trichome and root development. *gl2* mutants have aborted trichomes that grow laterally on the surface of the leaf. Another gene known to affect trichome development is *TTG1*. Just like *GL2*, *TTG1* is also involved in root hair development and seed coat mucilage production. *bell1* mutant plants are female sterile due to defects in ovule development. *ath1* mutants show constitutive photo morphogenesis just like *cop1* mutants, because *ATH1* is positively regulated by light in *Arabidopsis* seedlings (Chan et al., 1998). *atml1 pdf2* double mutants are embryonic lethal and the embryo fails to survive due to the defects in the development of epidermal layer (Abe, 2003a). Few HD-ZIP proteins in Arabidopsis are involved in growth control of embryos and cotyledons, such as *HAT4*, *ATHB-2*, *ATHB-4* (Schena et al., 1993; Sessa et al., 1998). While

others have role in regulating growth in water deficit conditions, such as *ATHB-5*, *ATHB-6*, *ATHB-7* (Soderman et al., 1994).

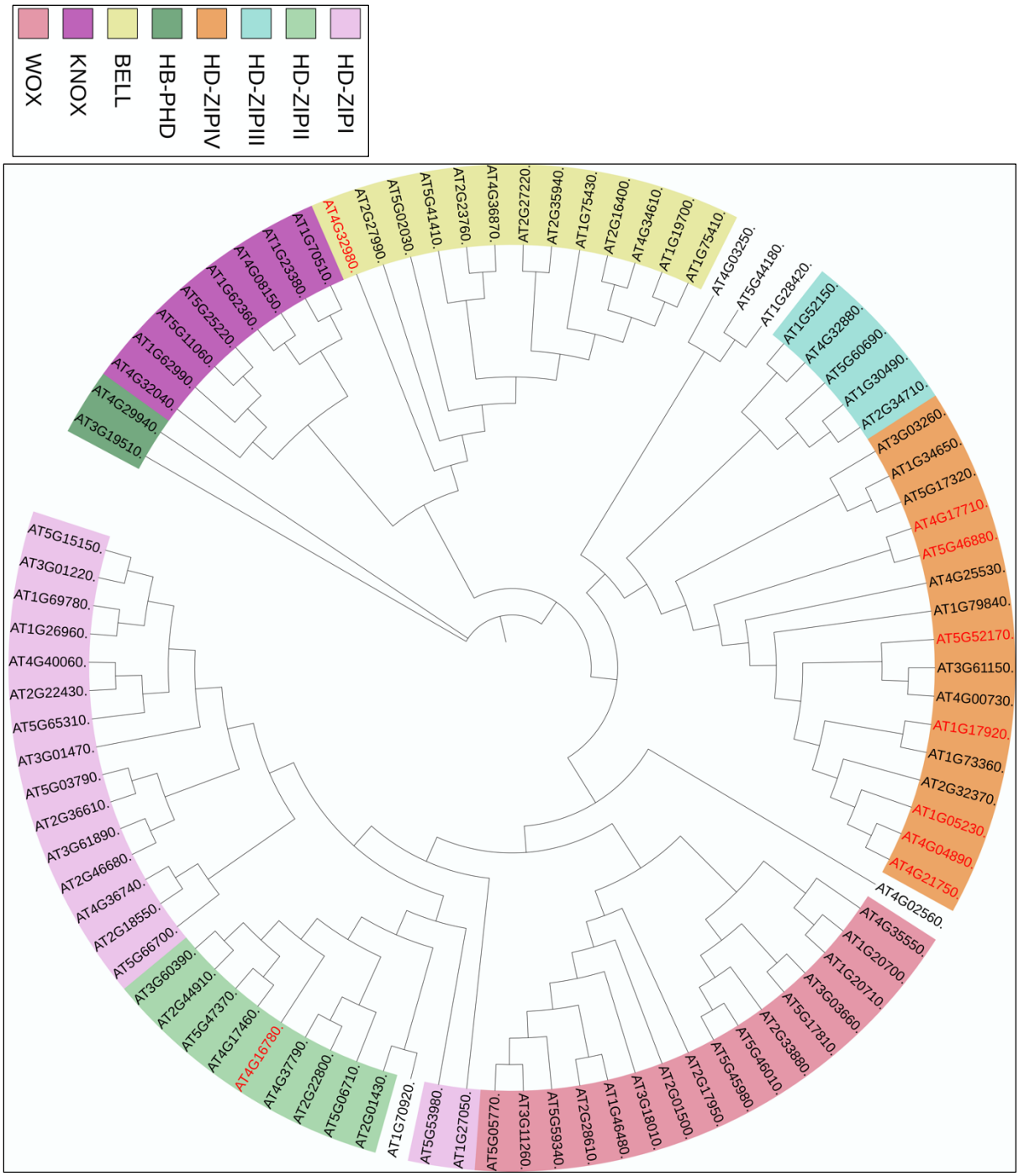


Figure 4.32: Phylogenetic tree for homeobox TF family. A circular phylogenetic tree representing all the members of the HB transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum

likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). Members belonging to different sub-families have been highlighted in different colours. TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

Nine members of the homeobox family are expressed in the epidermal and sub-epidermal layers of the shoot, ***ATH1/AT4G32890***, ***ATHB-2/AT4G16780, ATML1/AT4G21750, HDG2/AT1G05230, HDG4/AT4G17710, HDG5/AT5G46880, HDG7/AT5G52170, HDG12/AT1G17920,*** and ***PDF2/AT4G04890.*** *ATH1* belongs to the BELL sub-family of homeobox proteins and the remaining members belong to the HD-ZIP sub-family. ***ATHB-2/AT4G16780*** is a member of the HD-ZIP II family and has been involved in sensing changes in R/FR ratio and inducing the shade avoidance response.(Carabelli et al., 1993, 2005; Steindler et al., 1997). Plants over-expressing *ATHB-2* have longer petioles and smaller and fewer leaves. To understand the impact of this gene on shoot development, its mutant, SALK_106790 was analysed. *athb2-1* allele exhibited T-DNA insertion within the first exon of the gene (Figure 4.33). However, transcript was detected in comparable amount in the mutant and the WT plants. No phenotypic defects were identified for this allele.

ATML1/AT4G21750 is a member of the HD-ZIP IV family and has been known to play an important role in specifying the epidermal layer, along with other genes, very early during embryonic development (Takada et al., 2013). To study the function of ***ATML1*** in plant development, mutant line, SALK_033408 was studied. *atml1-3* allele carries a T-DNA insertion within the coding sequence of the gene (Figure 4.33). No full length transcript was detected in the mutant line, representing a possible true knock-out. Though null, single mutant doesn't display any morphological defect possibly due to redundancy with other family members.

HDG2/AT1G05230 is a member of the HD-ZIP IV TF family and is an important player in stomatal differentiation pathway. Stomatal differentiation gets disturbed in the *hdg2* loss of function mutants, thereby resulting in aberrant stomata (Peterson et al., 2013). To understand the effect of *hdg2* mutation on the overall growth and development of the plant, its mutant line, SALK_138646 was screened. *hdg2-3* allele carries an insertion within the first exon of the gene. RT PCR did not amplify full length transcript in the mutant as opposed to the WT (Figure 4.33).

This allele might represent a true null at the level of RNA. However, homozygous plants display no phenotypic defects.

HDG4/AT4G17710 belongs to the HD-ZIP IV family and is a part of the sub-epidermal layer of SAM. Nothing is known about the role of *HDG4* in plant growth and development so far. To understand the role of this gene, a GABI-KAT mutant line CS303999 was analysed. Insertion is present within the annotated exon region (Figure 4.33). Transcript levels in WT and mutant background could not be established by semi-quantitative PCR. Although highly expressed in the L2 layer of shoot apical meristem, the single mutant doesn't exhibit any abnormality in development. This in part may be due to high redundancy among genes within the plants.

HDG7/AT5G52170 is a member of the HD-ZIP IV family. This gene is enriched in the sub-epidermal layer of the shoot and nothing so far is known about the role of this gene in *Arabidopsis*. A mutant line, SALK_132114 was studied to understand the effect of loss of this gene on plant growth. Sequencing results indicated the insertion of T-DNA within the annotated intron of this gene. Semi-quantitative RT PCR analysis revealed that there was no down regulation in the level of RNA in the mutant versus WT (Figure 4.33). This possibly indicates that the above-mentioned line does not represent a true knock-out or a null mutant for the *HDG7* gene. The homozygous plants display no altered phenotype as compared to the wild counterpart.

HDG12/AT1G17920 belongs to HD-ZIP IV family and is expressed in the L1 layer of the shoot. *hdg12* loss of function in the plants along with *pdf2*, show weak phenotype with altered flower morphology, where petals and stamens get converted into sepals and carpels respectively (Kamata et al., 2013). However, nothing is known about the role of this gene in shoot development in the plant. To study the function of *HDG12*, its mutant, SALK_127261 was analysed. *hdg12-2* allele carries an insertion within the coding sequence of the gene. No full-length transcript was detected in the homozygous mutant plant in RT PCR analysis (Figure 4.33). This allele represents a true knockout for *HDG12* gene. No discernible phenotype of growth and development was observed in *hdg12-2* mutant allele.

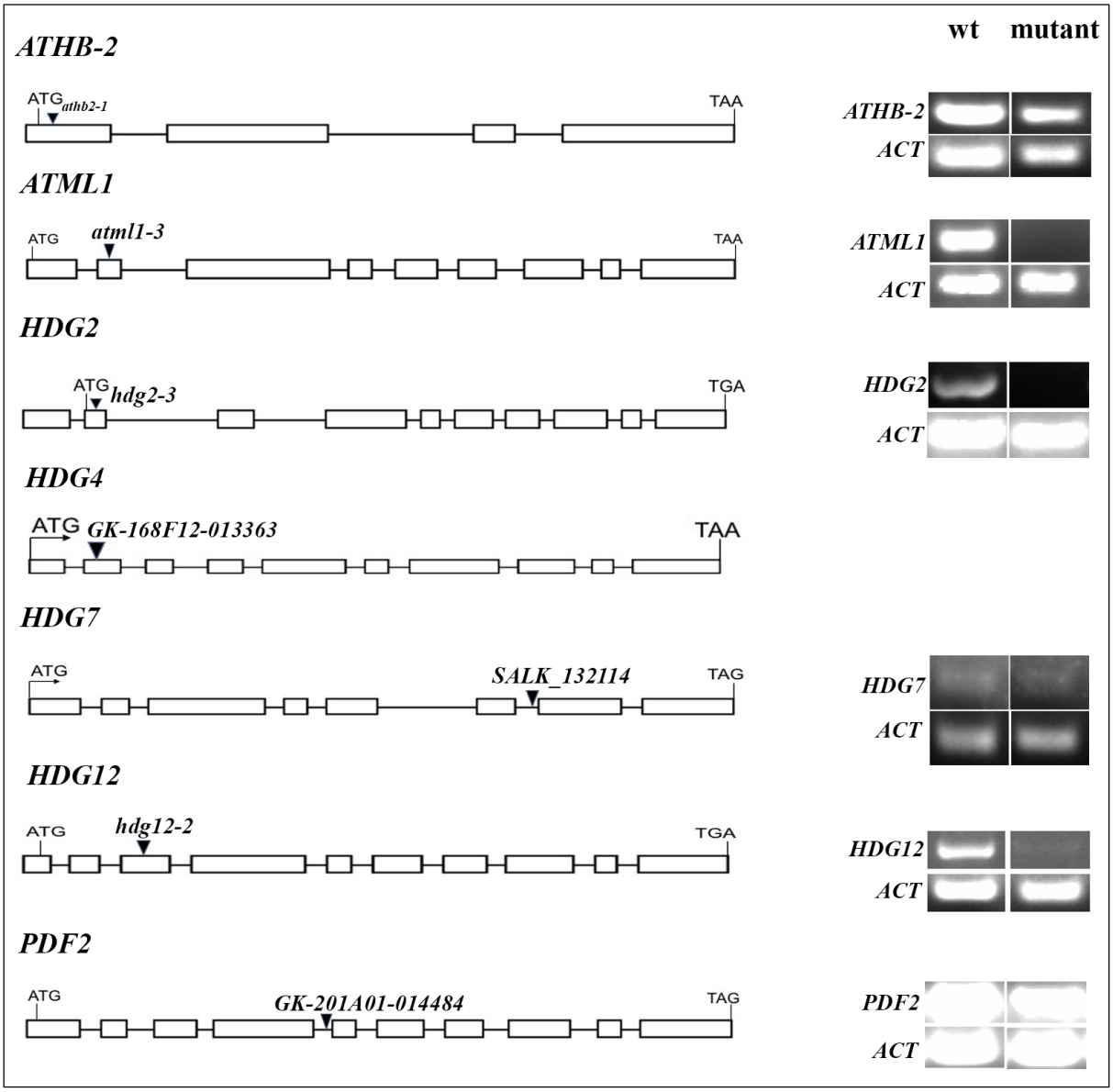


Figure 4.33: T-DNA characterisation of homeobox TF family members. Genomic structure of *Arabidopsis* homeobox genes and location of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

***PDF2/AT4G04890*,** a member of the HD-ZIP IV TF family, along with *ATML1* has been known to be involved in specification of epidermal layer in *Arabidopsis*. *PDF2* expresses very early during embryogenesis and controls epidermal cell differentiation (Abe, 2003b). For deeper understanding of role of *PDF2* in shoot development, two independent insertion mutant lines were analysed, CS304455 and SALK_109425. CS304455 displayed insertion of the T-DNA

within the exon of the gene. Also, *pdf2-2 allele,* SALK_109425, exhibited insertion within the annotated exon. RT PCR analysis showed no downregulation in the level of RNA in the mutant lines (Figure 4.33). No abnormal phenotype related to growth and development was exhibited by homozygous mutant lines.

4.3.2.15 MADS Transcription Factor Family

This family has 103 (as per PTFDB) members. MADS box genes are present in two major monophyletic lineages, type I and type II. More than half of the *Arabidopsis* MADS box genes are type I. Type I genes usually have one or two exons, while the type II genes have six to eight exons. Figure 4.34 shows all the members of the MADS family in *Arabidopsis* in the form of a phylogenetic tree, with type I and type II members highlighted in different colours. The name of the family was derived from the four initially identified members (Schwarz-Sommer et al., 1990) yeast *MCM1* involved in cell specific transcription and pheromone response (Dolan and Fields, 1991), *Arabidopsis AGAMOUS (AG)* involved in flower development (Yanofsky et al., 1990), Antirrhinum *DEFICIENS (DEF)* involved in flower development (Schwarz-Sommer et al., 1990) and the *human serum response factor (SRF)* involved in the transcriptional regulation of immediate early genes . MADS box motif appears to be structurally conserved and is composed of 56 amino acids (Shore and Sharrocks, 1995) (Davies and Schwarz-Sommer 1994).

Two sub-regions are present in the MADS box: A N-terminal region rich in basic and hydrophilic residues involved in the contact with DNA and a C-terminal region rich in hydrophobic residues.

Plant MADS box genes are involved in regulation of floral development (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994). Plant MADS box proteins have been classified by parsimony analysis into several distinct groups: AG group, AP3/PI group, AP1/AGL4 group and others (Purugganan et al., 1995). The AG group contains *AGAMOUS* and Antirrhinum *PLENA* genes required for stamen and carpel development (Bradley et al., 1993; Yanofsky et al., 1990)The AP3/PI group includes *Arabidopsis APETALA3 (AP3)*, *PISTILLATA (PI)* necessary for stamen and petal development (Goto and Meyerowitz, 1994; Jack et al., 1992; Sommer et al., 1990; Tröbner et al., 1992). The AP1/ AGL9 group contains *Arabidopsis APETALA1 (AP1)* and

CAULIFLOWER (CAL) genes required for floral meristem identity and development of Perianth organs (Huijser et al., 1992; Kempin et al., 1995; Mandel et al., 1992).

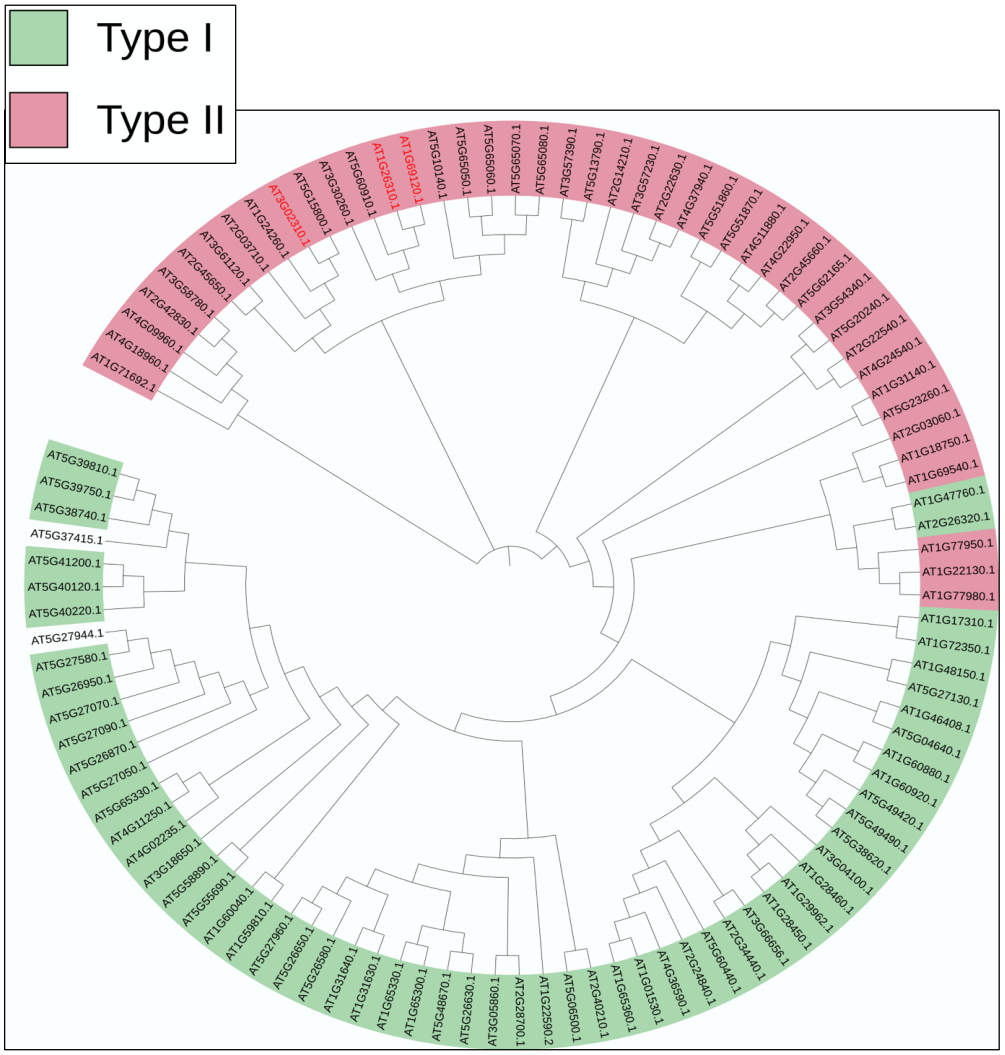


Figure 4.34: Phylogenetic tree for MADS TF family. Phylogenetic tree representing all the members of the MADS transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). Members belonging to different sub-families (Yanhui et al., 2006) have been highlighted in different colours. TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

Extensive studies have been done on genetic interactions, hierarchical relationships and cross regulatory networks between MADS box and other homeotic genes (Weigel and Meyerowitz, 1994). For example, *AP1* and *CAL1* are negatively regulated by AG (Kempin, Savidge et al.

1995) (Gustafson-Brown, Savidge et al. 1994). *AGL5* is a direct target of AG (Savidge, Rounsley et al. 1995). Three members of this family are enriched in the epidermal and sub-epidermal cell types of the shoot, namely ***AP1/AT1G69120, CAL/AT1G26310***, **SEP2/AT3G02310**. *APETALA1* is a floral homeotic gene that specifies floral meristem and sepal identity. In single mutant of *ap1*, there is a homeotic conversion of sepals to bracts. Secondary and tertiary flowers start forming in the axils of transformed sepals. Petals are usually absent. *CAL* is also a floral homeotic gene that shares high homology with *AP1*, but the single mutant of *cal* does not show any phenotype. However, it enhances the flower to shoot transformation in *ap1* mutants. *sep2* single mutant shows a very subtle phenotype, however, the triple mutant of *sep1 sep2 sep3* shows a reduction in the number of flowers, with otherwise normal looking flowers (Ditta et al., 2004).

4.3.2.16 MYB Transcription Factor Family

MYB superfamily consists of one-eighth of the TFs in *Arabidopsis* and is divided into three families, R1R2R3, R2R3 and MYB-related families. The MYB domain usually consists of one to three imperfect repeats, each of about 52 amino acid residues. R2R3 members consist of two adjacent repeats and R1R2R3 members consist of three adjacent repeats. MYB related protein usually contain single MYB repeat (Yanhui et al., 2006). The DNA binding domain of the protooncogene *Myb* defines the MYB domain in proteins of this family. AACNG is the binding sequence preferred by the MYB family members. This family has 168 (as per PTFDB) members in *Arabidopsis,* as shown in figure 4.35..

Contrary to the animals, there are a large number of MYB proteins in plants, with highly diverse functions (Avila et al., 1993; Jackson et al., 1991; Oppenheimer et al., 1991; Paz-Ares et al., 1987). MYB proteins such as C1, P, PL, Zm1, Zm38 of maize, MYB305 of Antirrhinum and MYB.Ph3 of Petunia have role in regulation of phenyl-propanoid biosynthetic gene(Avila et al., 1993; Grotewold et al., 1991; Paz-Ares et al., 1987; Sablowski et al., 1994; Solano et al., 1995). The *GL1* gene encodes for a MYB domain transcription factor that is involved in trichome initiation process. Another *Arabidopsis* MYB protein, AtMYB2 is involved in transcriptional regulation in response to salt stress or dehydration and to ABA (Urao, 1993). *MIXTA* gene of Antirrhinum also encodes for a MYB TF, that controls cell shape by activating directional

synthesis of specialized epidermal cell wall material (Noda et al., 1994). In *Arabidopsis,* mutation in *FOUR LIPS,* a MYB family TF, results in abnormal stomatal patterning due to the failure of guard mother cells to adopt guard cell fate (Lai, 2005). Double mutant of *flp-1 myb88* further enhances the phenotype and results in clusters of cells consisting of mature stomata (Lee et al., 2013). Mutations in another MYB family gene, *LOF1,* causes defects in organ separation. *lof1*mutants have cauline leaves fused to the inflorescence and loss of accessory shoot formation (Lee et al., 2009). A MYB domain protein, *AS1* is involved in specification of proximal-distal axis of the leaves, with *as1* mutants having smaller rosette with lobed and asymmetric leaves (Sun et al., 2002).

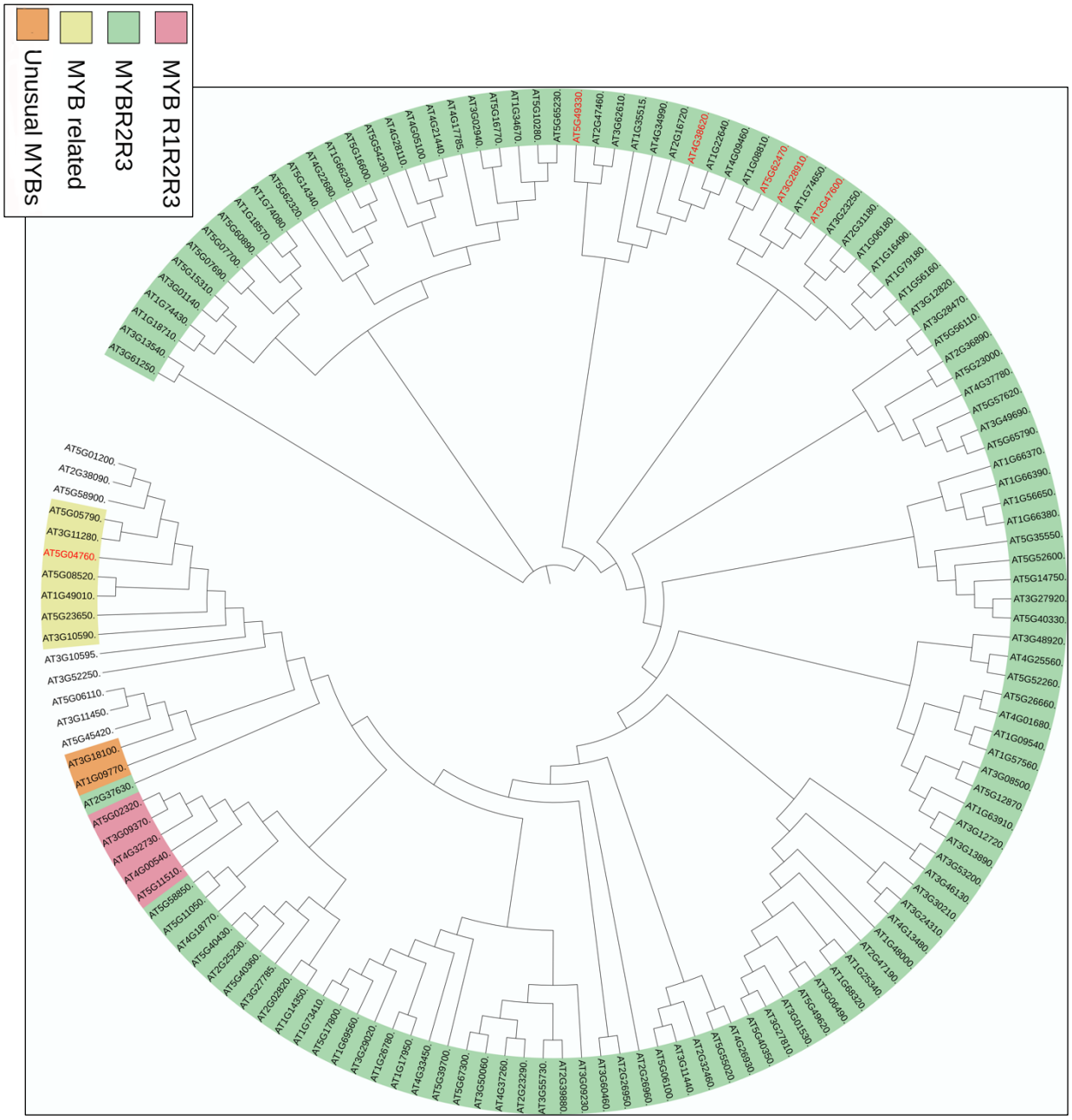


Figure 4.35: Phylogenetic tree for MYB TF family. A circular phylogenetic tree representing all the members of the MYB transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). Members belonging to different sub-families (Yanhui et al., 2006) have been highlighted in different colours. TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

Six members of the MYB family, ***MYB4/AT4G38620, MYB30/AT3G28910, MYB94/AT3G47600, MYB96/AT5G62470, MYB111/AT5G49330 and AT5G04760,*** are enriched in the epidermal layer of the shoot. Recently, *MYB30* was shown to be involved in oxidative and heat stress responses in *Arabidopsis.* To further investigate the role of ***MYB30/AT3G28910*** in development, its T-DNA line, SALK_027644 was screened. *myb30-1* insertion is present within 300bps of the 5' end of the gene, in the annotated exon of the gene. Equal *MYB30* transcript was amplified from mutant plants in RT PCR experiment, in comparison to the WT (Figure 4.36). Also, *myb30* plants did not vary phenotypically from the WT plants.

MYB94/AT3G47600 is another member of the MYB TF family, enriched in the shoot. Wisconsin T-DNA insertion line CS859284 was studied for *MYB94*. Sequencing analysis revealed the presence of T-DNA insertion in the 3' UTR of the gene (Figure 4.36). Although downregulation in the level of transcript was detected in mutant as compared to WT, the mutant plants did not exhibit any phenotype.

MYB96/AT5G62470, is a R2R3 type MYB transcription factor, and T-DNA line, CS333785 from GABI-KAT collection was used for studying the function of *MYB96*. Sequencing analysis revealed the presence of the T-DNA within the intron of the gene (Figure 4.36). This might not be a represent a true knock-out for *MYB96*. Also, phenotypically the mutant plants resembled the WT. RT PCR analysis shows no downregulation in the level of *MYB96* transcript in mutant versus WT.

MYB111/AT5G49330 belongs to the R2R3 family of the MYB superfamily. A GABI-KAT line CS9979 was used for confirming insertion in *myb111*. This is primarily a double mutant of *myb11 myb111.* T-DNA insertion is present in the third exon of the annotated gene sequence for *MYB111* (Figure 4.36). The double mutant plants also did not show any growth related defects, probably due to other redundant factors present in the system. The transcript level in the mutant was not much reduced compared to the WT.

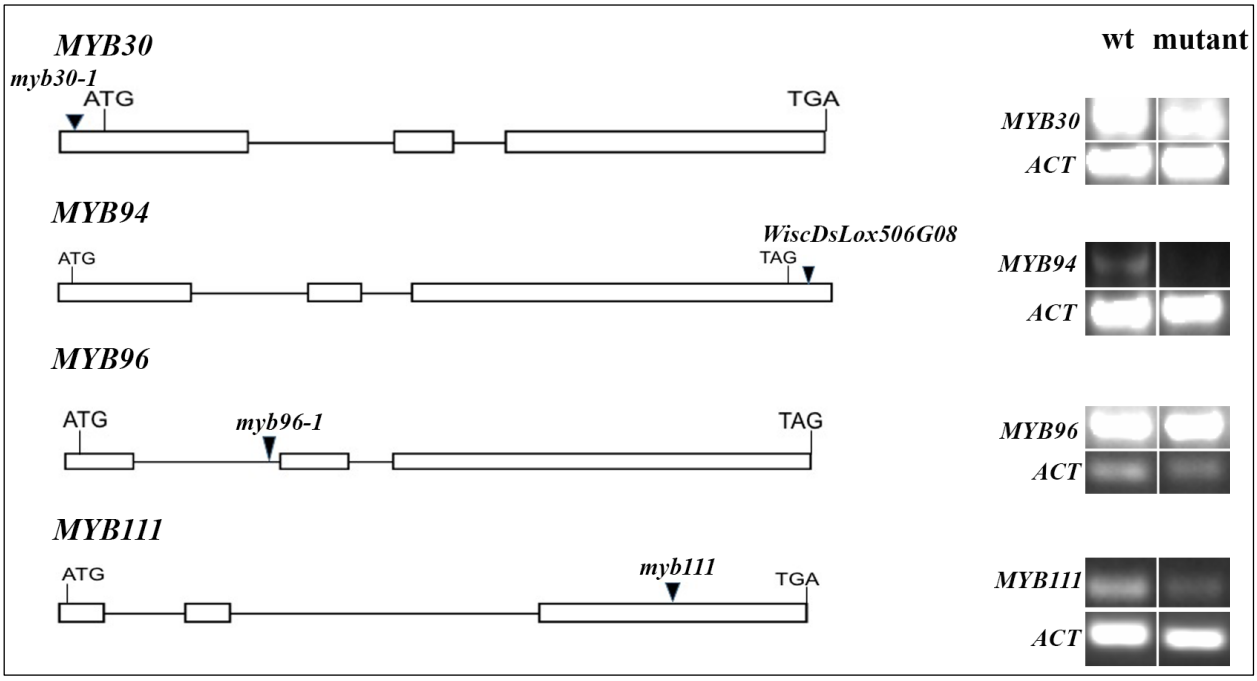


Figure 4.36: T-DNA characterisation of MYB TF family members. Genomic structure of *Arabidopsis* MYB genes and location of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.17 NAC Transcription Factor Family

There are 101 members in this family in *Arabidopsis thaliana,* represented in figure 4.37 in the form of a phylogenetic tree. This family is characterized by the presence of a NAC domain (petunia NAM and *Arabidopsis* ATAF1, ATAF2 and CUC2) which is present at the N terminal region and is divided into subdomains from A to E (Apweiler, 2001; Duval et al., 2002; Kikuchi et al., 2000). Mutations in the *NAM (No Apical Meristem)* genes in petunia plants results in absence of shoot apical meristem, suggesting the role of this gene in positioning of shoot and primordia in this plant (Souer et al., 1996). Mutations in *CUP-SHAPED COTYLEDON (cuc1* and *cuc2)* leads to defects in separation of cotyledons, stamens and sepals, also defective shoot apical meristem formation (Aida, 1997; Aida et al., 1999; Ishida et al., 2000). Therefore, the *CUC2* gene is thought to be involved in development of embryos and flowers (Aida, 1997). *ANAC043* is involved in regulating secondary wall biosynthesis in anthers and siliques. *anac043* mutants have reduced secondary cell walls in stems and siliques, resulting in indehiscent siliques (Mitsuda and Ohme-Takagi, 2008). ANAC033/URP7 is a NAC domain protein involved in root

cap development and *anac033* mutants experience delay in maturation and shedding of root cap cells, thereby causing the root cap to extend into differentiation zone (Willemsen et al., 2008). One member of this family, *NAC1* gets induced by auxin and promotes lateral root development by mediating auxin signalling (Xie et al., 2000). Members of this family are known to be involved in various processes of the plant such as development and morphogenic systems.

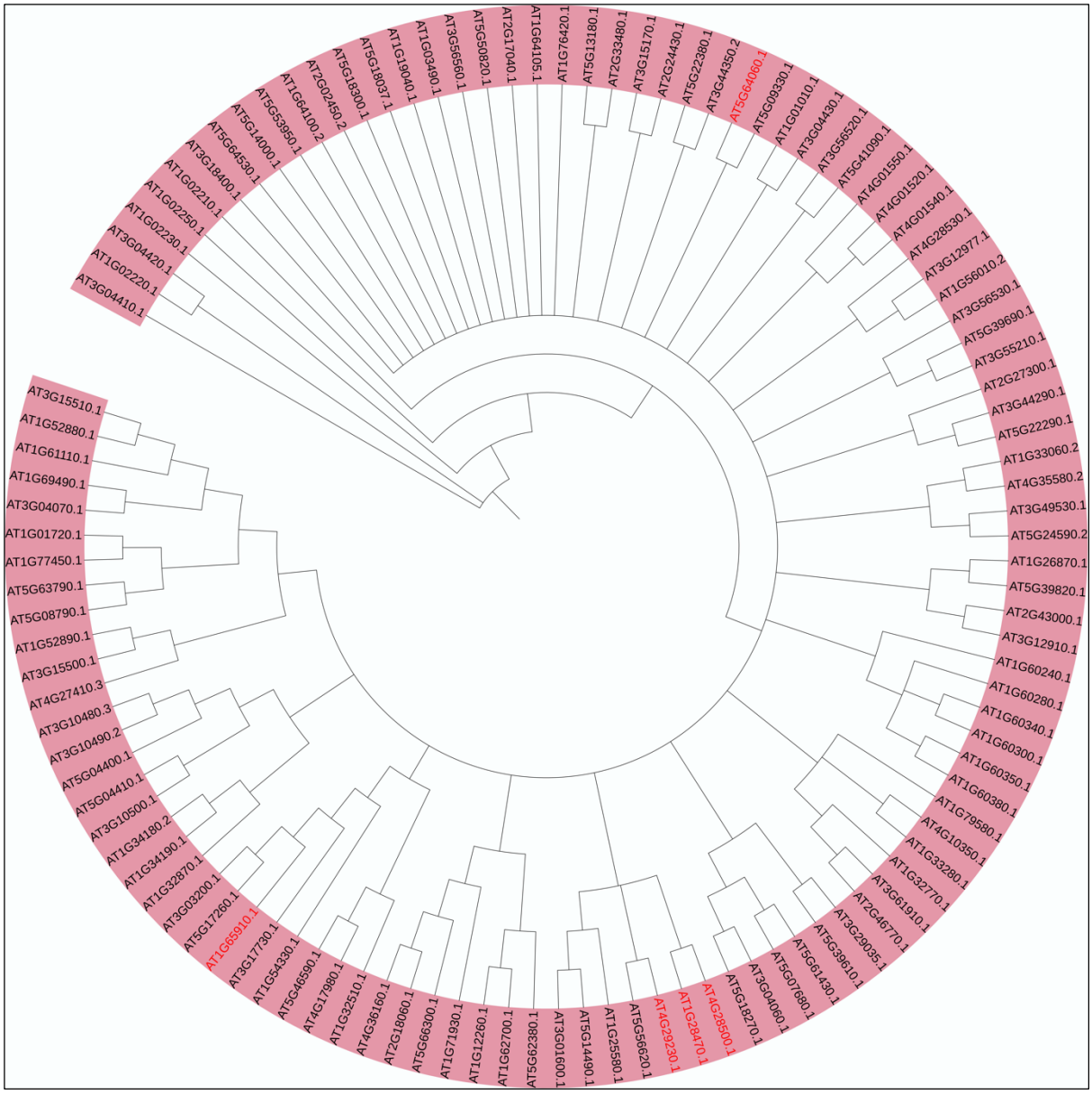


Figure 4.37: Phylogenetic tree for NAC TF family. A circular phylogenetic tree representing all the members of the NAC transcription factor family in *Arabidopsis.* Full length amino acid

sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

A total of 5 NAC family members are expressed in the epidermal and sub-epidermal layers of the *Arabidopsis* shoot apical meristem, ***ANAC010/SND3/AT1G28470, ANAC028/AT1G65910 ANAC073/AT4G28500, ANAC075/AT4G29230*** **and** ***ANAC103/AT5G64060***. In the past, role of *SND3* has been shown in secondary wall biosynthesis in vessels. To study the role of ***ANAC010/SND3*** in shoot development, T-DNA line, SALK_000287 was screened. Sequencing data revealed the presence of T-DNA element in the predicted exon of the gene. However, no developmental defects were found in the mutant line. Also, the transcript levels in mutant and WT background could not be established by PCR.

ANAC028/AT1G65910 is another member of the NAC family, enriched in the shoot. T-DNA line, CS868895 was characterized for insertion in *ANAC028* and also for level of transcript in mutant versus WT. In *anac028-1,* the T-DNA is present within the coding sequence of the gene (Figure 4.38). No morphological defects related to growth and development were associated with this mutant line and resembled the WT counterpart. However, no transcript for *ANAC028* was detected in the mutant as compared to WT.

ANAC103/AT5G64060 is a ANAC family member whose role has been shown in modulating the unfolded protein response via bZIP60. Wisconsin line CS850179 was used for studying the function of *ANAC103* in context of shoot. In *anac103-1,* 3' end of the gene was found to be disrupted by the T-DNA insertion (Figure 4.38). This line possibly doesn't represent a true knock-out for *ANAC103*. Also, no downregulation in the level of transcript was seen in mutant versus WT.

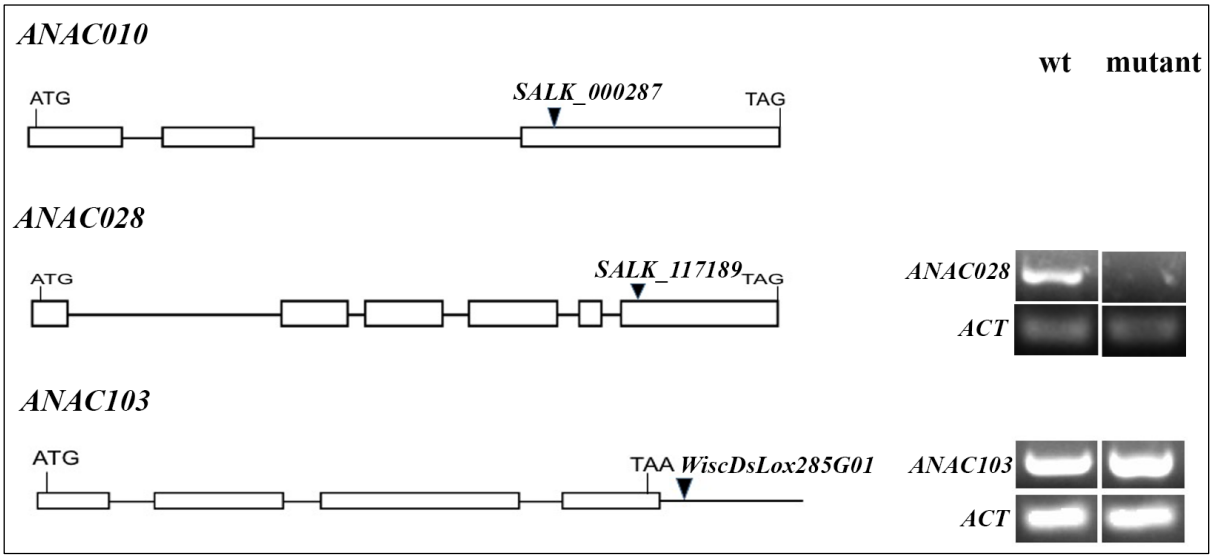


Figure 4.38: T-DNA characterisation of NAC TF family members. Genomic structure of *Arabidopsis* NAC genes and location of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.2.18 TUBBY Transcription Factor Family

A total of 10 members of tubby like protein family are present in *Arabidopsis,* of which one is a part of the sub-epidermal cell type. There is 30% to 80% amino acid similarity, among the 10 family members, across their C terminal tubby domains. All members of the AtTLP except *TLP8* contain F-box domain that is conserved (51-57 residues). A member of this family, *AtTLP9* is known to be involved in ABA signalling pathway (Lai et al., 2004). The phylogenetic tree showing evolutionary relationship between all members of TUB family is presented in figure 4.39. ***AtTLP8/AT1G16070*** is a member of the TUB family which is enriched in the sub-epidermal layer of the shoot.

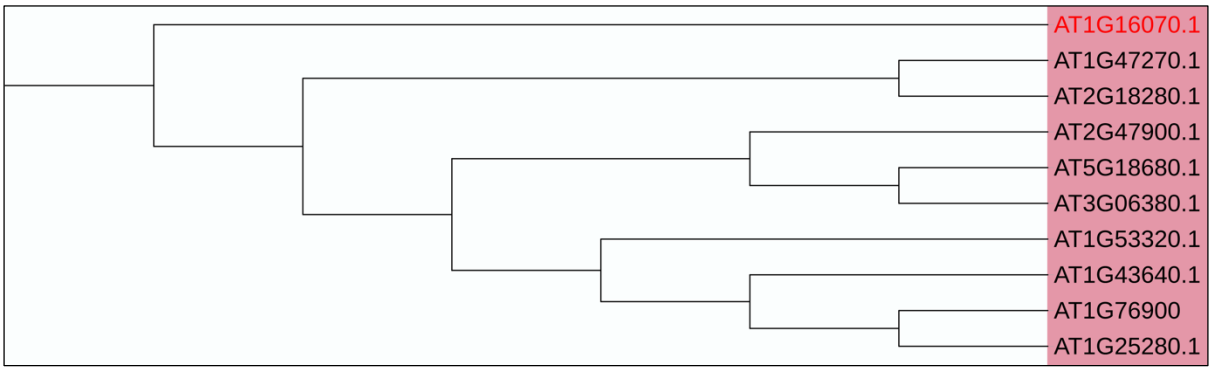


Figure 4.39: Phylogenetic tree for Tubby TF family. A phylogenetic tree representing all the members of the TUBBY transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). *TLP8,* enriched in the sub-epidermal cell type is highlighted in red.

4.3.2.19 WRKY Transcription Factor Family

The WRKY family members are characterized by the presence of conserved one or two WRKY domains, about 60 amino residues with conserved WRKYGQK at the N-terminal followed by a zinc finger motif (Rushton et al., 1995). WRKY domain is a DNA binding domain and all the characterized WRKY proteins prefer the same DNA motif. WRKY proteins can be classified based on the number of WRKY domains and the features of their zinc finger like motifs. Proteins with two WRKY domains fall in group I, whereas protein with one WRKY domain fall in group II. Similar type of finger motif (C-X_{4-5} -C-X_{22-23} -H-X_{13}-H) is present in group I and group II family members. However, a distinct finger motif is associated with a subset of WRKY proteins, C_2-HC, which have been classified into group III. The WRKY proteins bind to the DNA via a W-box motif, (T)(T)TGAC(C/T). Members of this family are involved in responses to pathogens (Fukuda and Shinshi, 1994; Rushton et al., 1995). Also, the W-box is present in the promoters of a large number of genes involved in pathogen defense (Fukuda and Shinshi, 1994). WRKY proteins also play role in biotic and abiotic stress responses in plants. A total of 71 members are a part of the WRKY family in *Arabidopsis* and are shown in the form of phylogenetic tree in figure 4.40, with members of different groups highlighted in different colours.

A large number of WRKY TFs are involved in defense related processes by regulating the expression of pathogenesis related (PR) genes. A large scale expression profiling study also revealed the involvement of WRKY TFs in defense program. Mutations in the *WRKY72* gene makes the plants more susceptible to fungal infections (Bhattarai et al., 2010). *wrky8* mutants are also more susceptible to *Botrytis cinerea* (Chen et al., 2010). WRKY2 is involved in proper embryo development by regulating its downstream targets *WOX8/9*. *wrky2* mutants fail to establish polar organelle positioning in the zygote, causing equal cell division and distorted embryo development (Ueda et al., 2011a).

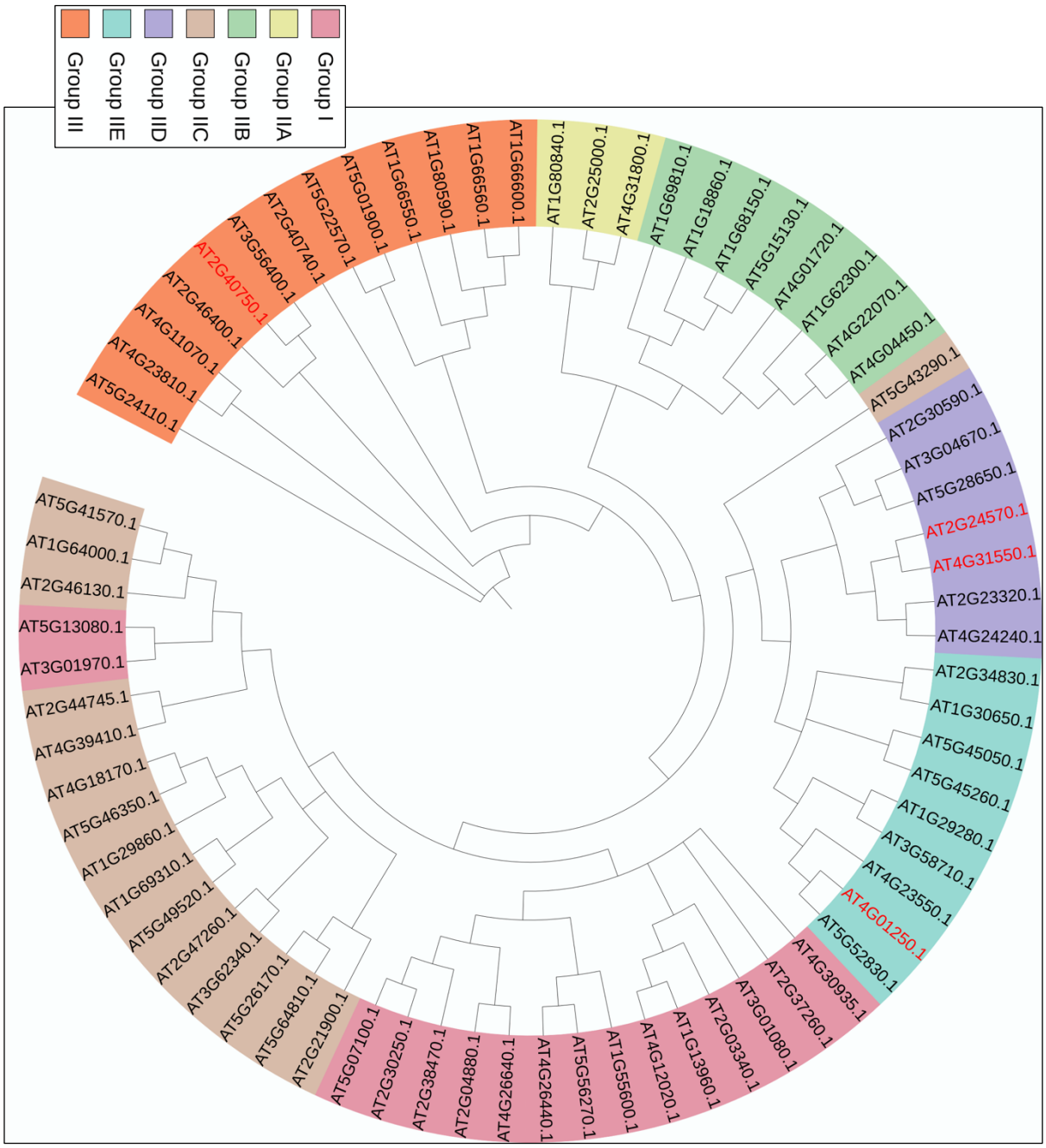


Figure 4.40: Phylogenetic tree for WRKY TF family. A circular phylogenetic tree representing all the members of the WRKY transcription factor family in *Arabidopsis*. Full length amino acid sequence of each protein was aligned. The tree was constructed using Maximum likelihood method. iTOL v3 online tool was used for tree visualisation (Letunic and Bork, 2016). Group I, II, III members (Eulgem et al., 2000) have been highlighted in different colours. TFs that are enriched in the epidermal or sub-epidermal cell types are highlighted in red.

A total of six members of this family are enriched in the epidermal and sub-epidermal layers of shoot, namely ***WRKY3/AT2G03340, WRKY11/AT4G31550, WRKY17/AT2G24570,***

WRKY22/AT4G01250, WRKY25/AT2G30250, WRKY54/AT2G40750. Another WRKY family member, ***WRKY2/AT5G56270,*** is enriched in the shoot and has been characterized in this study. Role of *WRKY2* has been previously reported in embryo patterning (Ueda et al., 2011b). In order to investigate the function of this gene in adult shoot development, a knock-out mutant line of the same, SALK_020399 line was characterised. *wrky2-1* allele carries a T-DNA insertion in the annotated exon of the *WRKY2* gene (Figure 4.41). However, *WRKY2* transcript was being detected in the mutant line also in the RT-PCR analysis. The mutant showed no morphological defects or defects related to growth, development and seed size.

WRKY11/AT4G31550 is another member of the WRKY family expressed in the epidermal layer of the shoot. *WRKY11* along with *WRKY17* has been known to act as negative regulators of basal resistance against *Pseudomonas syringae* pv *tomato* (Journot-Catalino et al., 2006). To further investigate its role in plant development, its insertion mutant line, SALK_141511 line was studied. T-DNA insertion was present within 300bps of the 5' end of the gene, in the annotated promoter region of the gene (Figure 4.41). RT-PCR analysis show no downregulation in the transcript level of the mutant line as compared to the WT. No discernible phenotype observed, possibly because it doesn't represent a null allele.

Another WRKY family member, ***WRKY17/AT2G24570*****,** with its role known in providing immunity to the plant, is enriched in the epidermal layer of the shoot. To understand the function of this gene, I screened a T-DNA insertion line, SALK_076337. T-DNA insertion was found to be present in the 5' UTR of the gene, 39bps upstream of the start codon (Figure 4.41). A comparable full length transcript was being amplified from the mutant as well as WT seedlings, but due to presence of the T-DNA very close to the ATG, it is most likely to disrupt the protein translation. However, the plants displayed no developmental defects in comparison to the WT. ***WRKY22/AT4G01250*** is another transcription factor from this family that is enriched in the epidermal layer. Role of *WRKY22* is known in providing innate immunity to the plant under conditions of submergence during flooding, where plants are at high risk of infection. For understanding developmental significance of *WRKY22,* SALK_047120 T-DNA insertion line was screened. Sequencing analysis confirmed the presence of T-DNA insertion within the first exon the gene and the T-DNA was also confirmed a null mutant at the level of RNA (Figure

4.41). Although null for RNA, but *wrky22-1* plants were phenotypically similar to WT. SALK_111964 T-DNA insertion line was screened for insertion as well as presence of RNA for ***WRKY54/AT2G40750***. *wrky54* mutant allele carries an insertion within the first annotated intron of the gene. Equal amount of *WRKY54* transcript was amplified from the mutant (Figure 4.41). No phenotype could be observed in the mutant line probably because it was not a null mutant and presence of other redundant members.

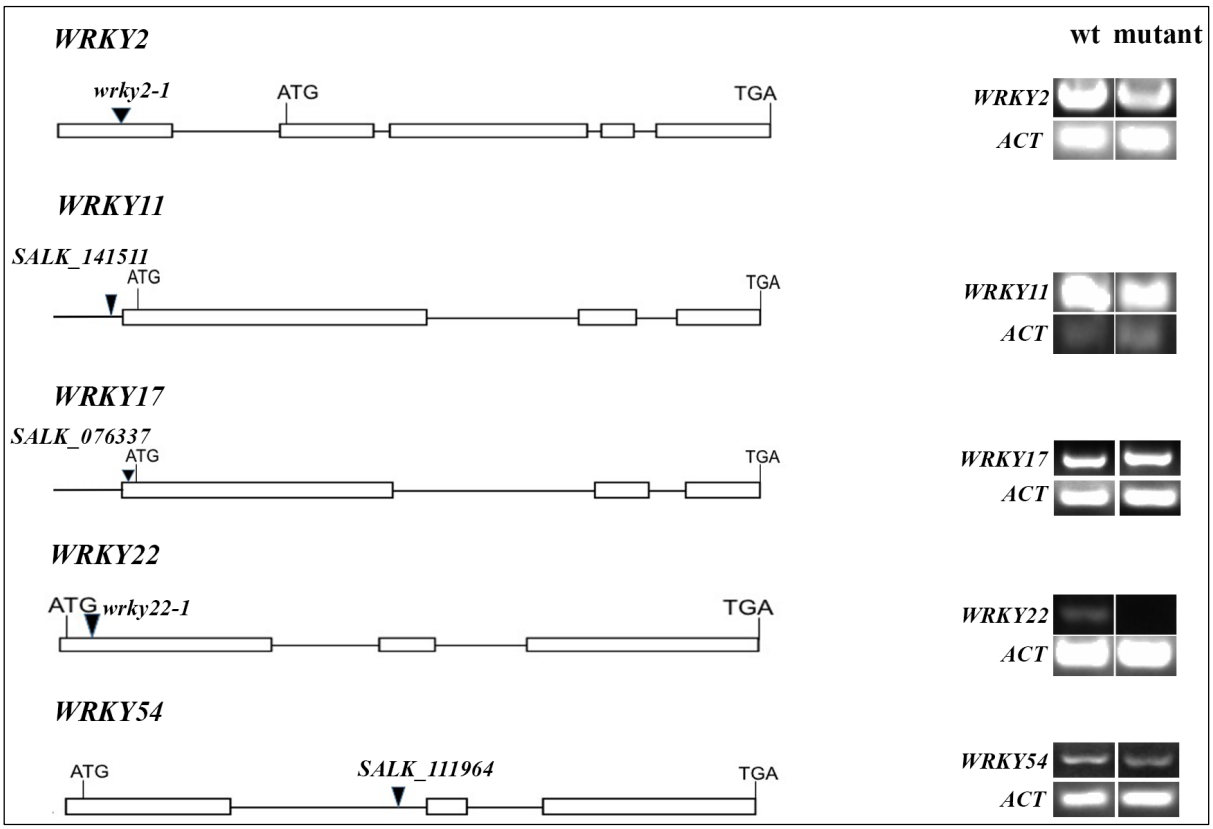


Figure 4.41: T-DNA characterisation of WRKY TF family members. Genomic structure of *Arabidopsis* WRKY genes and location of T-DNA insertions. Exons are denoted by boxes and introns are indicated by lines. Arrow heads indicate the position of T-DNA insertion. Alongside is also shown the RT PCR analysis of transcript levels in the WT versus mutant for each gene studied. *ACTIN* was used as internal control in RT PCR experiments.

4.3.3 Higher order mutants (Double and triple mutants)

Double and triple mutants were made for *hdg4 hdg7, hdg7 pdf2, hdg4hdg7pdf2.* 50 to 70 plants were genotyped for each combination. Double mutant for *HDG4* and *HDG7* was made because expression studies and micro-array data analysis indicate that both the TFs are expressed in the

L2 layer of the shoot and they share 65% identity in their nucleotide sequences. Also, the two transcription factors belong to HD-ZIP IV family. However, the double mutant plants did not display any phenotypic defects and resembled their WT counterparts. *HDG7* and *PDF2* also belong to the HD-ZIP IV family and share 65% identity at the nucleotide level. Based on this, their double mutant was attempted. *hdg7 pdf2* double mutant plants also did not show any developmental defects. Furthermore, a triple mutant of *hdg7 hdg4 pdf2* was made. All three genes belong to the HD-ZIP family and share sequence similarity. To our surprise, even the triple mutant plants were phenotypically indistinguishable from the WT. SALK_132114, CS304455 and CS303999 alleles of *HDG7, PDF2 and HDG4* respectively were used for making double and triple mutants.

4.4 Discussion

This study focusses on large scale comprehensive characterization of shoot enriched transcription factor genes in *Arabidopsis.* T-DNA insertional mutants were analysed for 43 transcription factors. Homozygous mutants were isolated for all of the above-mentioned genes. However, most of the homozygous mutants studied did not show any abnormalities in growth and development and phenotypically resembled the WT under normal conditions. A rough estimate suggests that roughly 8-10% gene knockouts result in phenotype. This is due to high redundancy in their function. In the current study, except the homozygote T-DNA allele of *AT1G75710* gene, none have shown a phenotype, suggesting the presence of other redundant factors within the system. I found a lot of candidate genes encoding TFs in the epidermal and sub-epidermal cell type dataset that were previously uncharacterized. However, when the T-DNA insertion lines were analysed for these TFs by T-DNA PCR, I found insertion outside the coding region of the genes. Further annotation of these genes revealed in few, a STOP codon. Furthermore, I analysed these lines for mRNA transcript of particular TF. Interestingly in a few T-DNA lines, despite confirming the insertion closer to start codon of a TF gene, the transcripts level of a gene of interest either got unchanged or elevated a little, and in a few instances it got diminished a little bit. Indicating a little or no consequence of T-DNA insertion. RT-PCR experiments carried out on homozygous alleles, reflected downregulation of full length transcript

for 18 TF mutants out of 38 validated by semi-quantitative RT-PCR. Therefore, these mutant lines used in the study are necessarily null alleles.

For most of the mutant alleles used in the study, no phenotype was recorded, however, similar alleles used in the past for creating higher order mutants did display phenotypic defects. *atml1-3 pdf2* were found to cause arrest of the embryo at globular stage (Ogawa et al., 2014). *hdg12-2 hdg11-1* double mutants caused excess branching of the trichome (Nakamura et al., 2006), however, *hdg12-2* allele used in this study alone did not display any trichome branching defect. *grf1-1 grf2 grf3* displayed smaller rosette leaves and cotyledons as compared to the WT (Kim et al., 2003). Also, triple mutant plants showed fused cotyledons in some cases. These redundant members present within the plants, possibly explain why the single mutant *grf3-1* allele used in the study does not display any defect. *dewax* mutant plants had slightly whitish green stems. *DEWAX*, an AP2/ERF type transcription factor is involved in regulating cuticular wax biosynthesis in *Arabidopsis*. Therefore, disrupting the *DEWAX* gene, causes an increased deposition of epicuticular wax crystals on the surface of the stem, which was related with upregulation of genes involved in wax biosynthesis (Go et al., 2014). A double mutant of *wrky54 wrky70* (SALK_111964 line of *WRKY54*, also used in this study) is reported in literature to display premature senescence of leaves (Besseau et al., 2012). But, *wrky54* single mutant had no signs of premature senescence. *wrky2* single mutant plants looked very similar to the WT phenotypically. But there are reports suggesting that *wrky2* mutants are hypersensitive to ABA responses during seed germination and early seedling growth (Jiang and Yu, 2009). *anac010, anac028* and *anac103,* members of the *Arabidopsis* NAC family of transcription factors also did not display any growth defects, possibly due to redundantly acting other members of the NAC family present in the plant. However, there are evidences in the literature suggesting role of *SND1/ANAC010* in activating genes involved in secondary cell wall biosynthesis (Zhong et al., 2008). Suggesting that the mutants should analysed in more depth for secondary cell wall defects in various organs and higher order mutants with other homologs be made, which may display a much stronger defect as compared to the single mutant.

at1g75710, a member of the Zn-finger family displays a multitude of phenotypes such as phyllotaxy defects, abnormal flowers and a prolonged vegetative phase. The promoter reporter fusion created for *AT1G75710* using a 3kb promoter region, shows expression of this gene in the

L1 layer of the shoot and in the emerging and fully developed organ primordia. Expression of this gene was also found in the mature flowers of various stages. Based on the expression studies and mutant phenotype, a possible role of this gene might be in regulating organ development, especially flower formation. However, for further understanding of the function of this gene, over expression lines can be created by over-expressing the gene under CaMV 35S promoter. Over expression work is in progress. Also, double mutant of *AT1G75710* needs to be made with its closest homolog expressed in the shoot, *AT5G54630*.

Double mutants *hdg7 hdg4* and *hdg7 pdf2* and triple mutant *hdg7 hdg4 pdf2* did not display any phenotypic defect despite knocking out three genes of the same family. This is possibly because *HDG7* carries a T-DNA insertion in the annotated intron and also no downregulation was seen in the transcript level in mutant versus WT. *HDG7* allele SALK_132114 does not appear to be null. In future, the double and triple mutants for the same should be analysed with different mutant alleles which are true null.

Over expression of *HDG7* resulted in plants with more vigour, prolonged vegetative phase, more number of axillary meristems and siliques and tremendous amount of anthocyanins (Shivani Bhatia and Ram Yadav unpublished data). This suggests that *HDG7* might be involved in the production of anthocyanin in the plant by regulating the key genes involved in the pathway. Also, presence of more number of axillary meristems indicate the role of *HDG7* in organ formation. *HDG7* might be involved in cross-talk with some of the important organ boundary genes such as, *CUC1, CUC2* or *LAS*, which function as an important regulators of axillary meristem initiation (Greb et al., 2003). However, these possibilities need to be explored in much detail and validated experimentally.

Lack of phenotypes in most of the single mutant genes, raises the possibility of evaluating higher order mutants for in-depth functional analysis of the genes under study. Another method by which the function of these genes can be understood is by creating transgenic plants with chimeric fusions of the genes coding sequence with either VP16 activation domain of herpes simplex virus or the Engrailed repressor domain of Drosophila (Khaled et al., 2005).

4.5 References

Abe, M. (2003a). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development *130,* 635-643.

Abe, M. (2003b). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development *130*, 635–643.

Aida, M. (1997). Genes Involved in Organ Separation in Arabidopsis: An Analysis of the cup-shaped cotyledon Mutant. PLANT CELL ONLINE *9,* 841-857.

Aida, M., Ishida, T., and Tasaka, M. (1999). Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. Development *126,* 1563-1570.

Alonso, J.M., and Ecker, J.R. (2006). Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. Nature Review Genetics *7,* 524-536.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science *301,* 653-657.

Apweiler, R. (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Research *43,* D213-D221.

Ariel, F.D., Manavella, P.A., Dezar, C.A., and Chan, R.L. (2007). The true story of the HD-Zip family. Trends in Plant Science *12,* 419-426.

Avila, J., Nieto, C., Cañas, L., Benito, M.J., and Paz-Ares, J. (1993). Petunia hybrida genes related to the maize regulatory C1 gene and to animal myb proto-oncogenes. Plant Journal *3,* 553-562.

Besseau, S., Li, J., and Palva, E.T. (2012). WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in Arabidopsis thaliana. Journal of Experimental Botany *63*, 2667–2679.

Bevan, M., and Walsh, S. (2005). The Arabidopsis genome : A foundation for plant research The Arabidopsis genome : A foundation for plant research. Genome Research *15,* 1632-1642.

Bhattarai, K.K., Atamian, H.S., Kaloshian, I., and Eulgem, T. (2010). WRKY72-type transcription factors contribute to basal immunity in tomato and Arabidopsis as well as gene-for-gene resistance mediated by the tomato R gene Mi-1. Plant Journal *63,* 229-240.

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

Bouché, N., Scharlat, A., Snedden, W., Bouchez, D., and Fromm, H. (2002). A novel family of calmodulin-binding transcription activators in multicellular organisms. Journal of Biological Chemistry *277,* 21851-21861.

Bradley, D., Carpenter, R., Sommer, H., Hartley, N., and Coen, E. (1993). Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of antirrhinum. Cell *72,* 85-95.

Carabelli, M., Sessa, G., Baima, S., Morelli, G., and Ruberti, I. (1993). The Arabidopsis Athb-2 and -4 genes are strongly induced by far-red-rich light. Plant Journal *4,* 469-479.

Carabelli, M., Morelli, G., Whitelam, G., and Ruberti, I. (2005). Twilight-zone and canopy shade induction of the Athb-2 homeobox gene in green plants. Proceedings of the National Academy of Sciences *93,* 3530-3535.

Chan, R.L., Gago, G.M., Palena, C.M., and Gonzalez, D.H. (1998). Homeoboxes in plant development. Biochimica et Biophysica Acta - Gene Structure and Expression *1442,* 1-19.

Chao, Q., Rothenberg, M., Solano, R., Roman, G., Terzaghi, W., and Ecker, J.R. (1997). Activation of the ethylene gas response pathway in arabidopsis by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. Cell *89,* 1133-1144.

Chen, L., Zhang, L., and Yu, D. (2010). Wounding-induced WRKY8 is involved in basal defense in Arabidopsis. Molecular Plant Microbe Interaction *23,* 558-565.

Coen, E.S., and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. Nature *353,* 31-37.

Cvitanich, C., Pallisgaard, N., Nielsen, K.A., Hansen, A.C., Larsen, K., Pihakaski-Maunsbach, K., Marcker, K.A., and Jensen, E.O. (2000). CPP1, a DNA-binding protein involved in the expression of a soybean leghemoglobin c3 gene. Proceedings of the National Academy of Sciences *97,* 8163-8168.

Das, P., Ito, T., Wellmer, F., Vernoux, T., Dedieu, A., Traas, J., and Meyerowitz, E.M. (2009). Floral stem cell termination involves the direct regulation of AGAMOUS by PERIANTHIA. Development *136,* 1605-1611.

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M.F. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Current Biology *14,* 1935-1940.

Dolan, J.W., and Fields, S. (1991). Cell-type-specific transcription in yeast. BBA - Gene Structure and Expression *1088,* 155-169.

Duval, M., Hsieh, T.F., Kim, S.Y., and Thomas, T.L. (2002). Molecular characterization of

AtNAM: A member of the Arabidopsis NAC domain superfamily. Plant Molecular Biology *50,* 237-248.

Eulgem, T., Rushton, P.J., Robatzek, S., and Somssich, I.E. (2000). The WRKY superfamily of plant transcription factors. Trends in Plant Science *5,* 199-206.

Finkler, A., Ashery-Padan, R., and Fromm, H. (2007). CAMTAs: Calmodulin-binding transcription activators from plants to human. FEBS Letter *581,* 3893-3898.

Frontini, M., Imbriano, C., Manni, I., and Mantovani, R. (2004). Cell cycle regulation of NF-YC nuclear localization. Cell Cycle *3,* 217-222.

Fukuda, Y., and Shinshi, H. (1994). Characterization of a novel cis-acting element that is responsive to a fungal elicitor in the promoter of a tobacco class I chitinase gene. Plant Molecular Biology *24, 485-493.*

Gehring, W.J., Affolter, M., and Burglin, T. (1994). Homeodomain Proteins - Annual Review of Biochemistry *63*, 487-526.

Gilchrist, E., and Haughn, G. (2010). Reverse genetics techniques: Engineering loss and gain of gene function in plants. Briefings in Functional Genomics and Proteomics *9,* 103-110.

Go, Y.S., Kim, H., Kim, H.J., and Suh, M.C. (2014). Arabidopsis Cuticular Wax Biosynthesis Is Negatively Regulated by the DEWAX Gene Encoding an AP2/ERF-Type Transcription Factor. Plant Cell *26*, 1666–1680.

Goto, K., and Meyerowitz, E.M. (1994). Function and regulation of the Arabidopsis floral homeotic gene PISTILLATA. Genes and Development *8,* 1548-1560.

Greb, T., Clarenz, O., Schäfer, E., Müller, D., Herrero, R., Schmitz, G., and Theres, K. (2003). Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. Genes and Development *17,* 1175-1187.

Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., Odden, A.R., et al. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics *164,* 731-740.

Griffiths, S., Dunford, R.P., Coupland, G., and Laurie, D.A. (2003). The evolution of CONSTANS-like gene families in barley, rice and Arabidopsis. Plant Physiology *131,* 1855-1867.

Grotewold, E., Athma, P., and Peterson, T. (1991). Alternatively spliced products of the maize P gene encode proteins with homology to the DNA-binding domain of myb-like transcription factors. Proceedings of the National Academy of Sciences *88,* 4587-4591.

Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P.C. (2003a). The

basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Molecular Biology and Evolution *20,* 735-747.

Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P.C. (2003b). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol. Biol. Evol. *20*, 735–747.

Huijser, P., Klein, J., Lönnig, W.E., Meijer, H., Saedler, H., and Sommer, H. (1992). Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene squamosa in Antirrhinum majus. EMBO Journal *11,* 1239-1249.

Ishida, T., Aida, M., Takada, S., and Tasaka, M. (2000). Involvement of CUP-SHAPED COTYLEDON genes in gynoecium and ovule development in Arabidopsis thaliana. Plant and Cell Physiology *41,* 60-67.

Jack, T., Brockman, L.L., and Meyerowitz, E.M. (1992). The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. Cell *68,* 683-697.

Jackson, D., Culianez-Macia, F., Prescott, a G., Roberts, K., and Martin, C. (1991). Expression patterns of myb genes from Antirrhinum flowers. Plant Cell *3,* 115-125.

Jakoby, M., and Vicente-carbajosa, J. (2002). bZIP transcription factors in Arabidopsis. *7*, 106–111.

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., Parcy, F., and Group, bZIP R. (2002). bZIP transcription factors in Arabidopsis. Trends in Plant Science *7,* 106-111.

Jiang, W., and Yu, D. (2009). Arabidopsis WRKY2 transcription factor mediates seed germination and postgermination arrest of development by abscisic acid. BMC Plant Biology *9*, 96.

Journot-Catalino, N., Somssich, I.E., Roby, D., and Kroj, T. (2006). The Transcription Factors WRKY11 and WRKY17 Act as Negative Regulators of Basal Resistance in Arabidopsis thaliana. Plant Cell Online *18*, 3289–3302.

Kamata, N., Okada, H., Komeda, Y., and Takahashi, T. (2013). Mutations in epidermis-specific HD-ZIP IV genes affect floral organ identity in Arabidopsis thaliana. Plant Journal *75*, 430–440.

Kempin, S. a, Savidge, B., and Yanofsky, M.F. (1995). Molecular basis of the cauliflower phenotype in Arabidopsis. Science *267,* 522-525.

Khaled, A.S., Vernoud, V., Ingram, G.C., Perez, P., Sarda, X., and Rogowsky, P.M. (2005). Engrailed-ZmOCL1 fusions cause a transient reduction of kernel size in maize. Plant Mol. Biol. *58*, 123–139.

Kikuchi, K., Ueguchi-Tanaka, M., Yoshida, K.T., Nagato, Y., Matsusoka, M., and Hirano, H.Y. (2000). Molecular analysis of the NAC gene family in rice. Molecular and General Genetics *262,* 1047-1051.

Kim, J.H., Choi, D., and Kende, H. (2003). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. Plant Journal *36,* 94-104

Krysan, P.J. (1999). T-DNA as an Insertional Mutagen in Arabidopsis. PLANT CELL ONLINE *11,* 2283.

Kurowska, M., Daszkowska-Golec, A., Gruszka, D., Marzec, M., Szurman, M., Szarejko, I., and Maluszynski, M. (2011). TILLING - a shortcut in functional genomics. J. Appl. Genet *42,* 293-308.

Lai, L.B. (2005). The Arabidopsis R2R3 MYB Proteins FOUR LIPS and MYB88 Restrict Divisions Late in the Stomatal Cell Lineage. PLANT CELL ONLINE *17,* 2754-2767.

Lai, C.-P., Lee, C.-L., Chen, P.-H., Wu, S.-H., Yang, C.-C., and Shaw, J.-F. (2004). Molecular analyses of the Arabidopsis TUBBY-like protein gene family. Plant Physiology *134,* 1586-1597.

Lee, D.-K., Geisler, M., and Springer, P.S. (2009). LATERAL ORGAN FUSION1 and LATERAL ORGAN FUSION2 function in lateral organ separation and axillary meristem formation in Arabidopsis. Development *136,* 2423-2432.

Lee, E., Liu, X., Eglit, Y., and Sack, F. (2013). FOUR LIPS and MYB88 conditionally restrict the G1/S transition during stomatal formation. Journal of Experimental Botany *64,* 5207-5219.

Lee, H., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (2003). Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proceedings of the National Academy of Sciences *100,* 2152-2156.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research *44,* W242-W245.

Licausi, F., Ohme-Takagi, M., and Perata, P. (2013a). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. New Phytologist *199,* 639–649.

Licausi, F., Ohme-Takagi, M., and Perata, P. (2013b). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytologist *199,* 639-649.

Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K., and Hake, S. (1994). A knotted1-like homeobox gene in Arabidopsis is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. Plant Cell *6,* 1859-1876.

Lotan, T., Ohto, M., Yee, K.M., West, M.A.., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B., and Harada, J.J. (1998). Arabidopsis LEAFY COTYLEDON1 Is Sufficient to Induce Embryo Development in Vegetative Cells. Cell *93,* 1195-1205.

Magome, H., Yamaguchi, S., Hanada, A., Kamiya, Y., and Oda, K. (2004). Dwarf and delayed-flowering 1, a novel Arabidopsis mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. Plant Journal *37,* 720-729.

Maier, A.T., Stehling-Sun, S., Offenburger, S.-L., and Lohmann, J.U. (2011). The bZIP Transcription Factor PERIANTHIA: A Multifunctional Hub for Meristem Control. Frontiers in Plant Science *2.*

Mandel, M.A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1992). Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. Nature *360,* 273-277.

Mariconti, L., Pellegrini, B., Cantoni, R., Stevens, R., Bergounioux, C., Cella, R., and Albani, D. (2002). The E2F family of transcription factors from Arabidopsis thaliana. Novel and conserved components of the retinoblastoma/E2F pathway in plants. Journal of Biological Chemistry *277,* 9911-9919.

Matsuoka, M., Tamaoki, M., Tada, Y., Fuyjimura, T., Tagiri, A., Yamamoto, N., and Kano-Murakami, Y. (1995). Expression of rice OSH1 gene is localized in developing vascular strands and its ectopic expression in transgenic rice causes altered morphology of leaf. Plant Cell Reports *14,* 555-559.

Mitsuda, N., and Ohme-Takagi, M. (2008). NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. Plant Journal *56,* 768-778.

Müller, K.J., Romano, N., Gerstner, O., Garcia-Maroto, F., Pozzi, C., Salamini, F., Rohde, W., Muller, K.J., Romano, N., Gerstner, O., et al. (1995). The barley Hooded mutation caused by a duplication in a homeobox gene intron. Nature *374,* 727-730.

Murray, M.G., Thompson, W.F. (1980). Rapid isolation of high molecular weight plant ONA. Nucleic Acids Research, *8* 432-4326.

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., and Takahashi, T. (2006). Characterization of the Class IV Homeodomain-Leucine Zipper Gene Family in Arabidopsis1[W]. Plant Physiology *141*, 1363–1375.

Noda, K., Glover, B.J., Linstead, P., and Martin, C. (1994). Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. Nature *369,* 661-664.

O'Malley, R.C., and Ecker, J.R. (2010). Linking genotype to phenotype using the Arabidopsis unimutant collection. Plant Journal *61,* 928-940.

Ogawa, E., Yamada, Y., Sezaki, N., Kosaka, S., Kondo, H., Kamata, N., Abe, M., Komeda, Y., and Takahashi, T. (2014). ATML1 and PDF2 play a redundant and essential role in arabidopsis embryo development. Plant Cell Physiology *56*, 1183–1192.

Ohme-Takagi, M., and Shinshi, H. (1995). Ethylene-Inducible DNA Binding Proteins that interact with an ethylene responsive element. Plant Cell *7,* 173.

Oppenheimer, D.G., Herman, P.L., Sivakumaran, S., Esch, J., and Marks, M.D. (1991). A myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. Cell *67,* 483-493.

Østergaard, L., and Yanofsky, M.F. (2004). Establishing gene function by mutagenesis in Arabidopsis thaliana. Plant Journal *39,* 682-696.

Oyama, T., Shimura, Y., and Okada, K. (1997). The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus- induced development of root and hypocotyl. Genes and Development *11,* 2983-2995.

Papi, M., Sabatini, S., Bouchez, D., Camilleri, C., Costantino, P., and Vittorioso, P. (2000). Identification and disruption of an Arabidopsis zinc finger gene controlling seed germination. Genes and Development *14,* 28-33.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A., and Saedler, H. (1987). The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. EMBO Journal *6,* 3553-3558.

Peterson, K.M., Shyu, C., Burr, C.A., Horst, R.J., Kanaoka, M.M., Omae, M., Sato, Y., and Torii, K.U. (2013). Arabidopsis homeodomain-leucine zipper IV proteins promote stomatal development and ectopically induce stomata beyond the epidermis. Development *140*, 1924–1935.

Petrey, D., Chen, T.S., Deng, L., Garzon, J.I., Hwang, H., Lasso, G., Lee, H., Silkov, A., and Honig, B. (2015). Template-based prediction of protein function. Current Opinion in Structural Biology *12,* 431-440.

Petroni, K., Kumimoto, R.W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B.F., and Mantovani, R. (2012). The Promiscuous Life of Plant NUCLEAR FACTOR Y Transcription Factors. Plant Cell *24,* 4777-4792.

Prelich, G. (2012). Gene overexpression: Uses, mechanisms, and interpretation. Genetics.
Purugganan, M.D., Rounsley, S.D., Schmidt, R.J., and Yanofsky, M.F. (1995). Molecular evolution of flower development: Diversification of the plant MADS-box regulatory gene family. Nature Reviews Genetics *2,* 186-195.

Rashotte, A.M., Mason, M.G., Hutchison, C.E., Ferreira, F.J., Schaller, G.E., and Kieber, J.J.

(2006). A subset of Arabidopsis AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. Proceedings of the National Academy of Sciences *103,* 11081-11085.

Robson, F., Costa, M.M.R., Hepworth, S.R., Vizir, I., Piñeiro, M., Reeves, P.H., Putterill, J., and Coupland, G. (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. Plant Journal *28,* 619-631.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Molecular Biology *53,* 247-259.

Rushton, P.J., Macdonald, H., Huttly, A.K., Lazarus, C.M., and Hooley, R. (1995). Members of a new family of DNA-binding proteins bind to a conserved cis- element in the promoters of alpha-Amy2 genes. Plant Molecular Biology *29,* 691-702.

Sablowski, R.W., Moyano, E., Culianez-Macia, F. a, Schuch, W., Martin, C., and Bevan, M. (1994). A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. EMBO Journal *13,* 128-137.

Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). DNA-Binding Specificity of the ERF/AP2 Domain of Arabidopsis DREBs, Transcription Factors Involved in Dehydration- and Cold-Inducible Gene Expression. Biochemical and Biophysical Research Communications *25,* 998-1009.

Schena, M., Lloyd, A.M., and Davis, R.W. (1993). The HAT4 gene of Arabidopsis encodes a developmental regulator. Genes and Development *7,* 367-379.

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus. Science *250,* 931-936.

Sessa, G., Steindler, C., Morelli, G., and Ruberti, I. (1998). The Arabidopsis Athb-8, -9 and -14 genes are members of a small gene family coding for highly related HD-ZIP proteins. Plant Molecular Biology *38,* 609-622.

Shore, P., and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. European Journal of Biochemistry *229,* 1-13.

Shuichi, Y. (2002). The Dof family of plant transcription factors. Trends in Plant Science *7,* 555-560.

Sinha, N.R., Williams, R.E., and Hake, S. (1993). Overexpression of the maize homeo box gene, KNOTTED-1, causes a switch from determinate to indeterminate cell fates. Genes and Development *7,* 787-795.

Smith, D., Yanai, Y., Liu, Y.G., Ishiguro, S., Okada, K., Shibata, D., Whittier, R.F., and Fedoroff, N. V. (1996). Characterization and mapping of Ds-GUS-T-DNA lines for targeted insertional mutagenesis. Plant Journal *10,* 721-732.

Soderman, E., Mattsson, J., Svenson, M., Borkird, C., and Engstrom, P. (1994). Expression patterns of novel genes encoding homeodomain leucine-zipper proteins in Arabidopsis thaliana. Plant Molecular Biology *26,* 145-154.

Solano, R., Nieto, C., Avila, J., Cañas, L., Diaz, I., and Paz-Ares, J. (1995). Dual DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB.Ph3) from Petunia hybrida. EMBO Journal *14,* 1773-1784.

Solano, R., Stepanova, A., Chao, Q., and Ecker, J.R. (1998). Nuclear events in ethylene signaling: A transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. Genes and Development *12,* 3703-3714.

Sommer, H., Beltrán, J.P., Huijser, P., Pape, H., Lönnig, W.E., Saedler, H., and Schwarz-Sommer, Z. (1990). Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors. EMBO Journal *9,* 605-613.

Souer, E., Van Houwelingen, A., Kloos, D., Mol, J., and Koes, R. (1996). The no apical Meristem gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell *85,* 159-170.

Steidl, S., Tuncher, Goda, H., Guder, C., Papadopoulou, N., Kobayashi, T., Tsukagoshi, N., Kato, M., and Brakhage, A.A. (2004). A single subunit of a heterotrimeric CCAAT-binding complex carries a nuclear localization signal: Piggy back transport of the pre-assembled complex to the nucleus. Journal of Molecular Biology *342,* 515-524.

Steindler, C., Carabelli, M., Borello, U., Morelli, G., and Ruberti, I. (1997). Phytochrome A, phytochrome B and other phytochrome(s) regulate ATHB-2 gene expression in etiolated and green Arabidopsis plants. Plant, Cell and Environment *20,* 759-763.

Sun, Y., Zhou, Q., Zhang, W., Fu, Y., and Huang, H. (2002). ASYMMETRIC LEAVES1, an Arabidopsis gene that is involved in the control of cell differentiation in leaves. Planta *214,* 694-702.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D.G., Dean, C., Ma, H., and Martienssen, R. (1995). Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. Genes and Development *9,* 1797-1810.

Takada, S., Takada, N., and Yoshida, A. (2013). ATML1 promotes epidermal cell differentiation in Arabidopsis shoots. Development *140*, 1919–1923.

Toledo-Ortiz, G., Huq, E., and Quail, P.H. (2003). The Arabidopsis Basic / Helix-Loop-Helix

Transcription Factor Family. Plant Cell *15,* 1749-1770.

Tröbner, W., Ramirez, L., Motte, P., and Hue, I. (1992). GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis. EMBO Journal *1,* 4693-4704.

Ueda, M., Zhang, Z., and Laux, T. (2011a). Transcriptional Activation of Arabidopsis Axis Patterning Genes WOX8/9 Links Zygote Polarity to Embryo Development. Developmental Cell *20,* 264-270.

Ueda, M., Zhang, Z., and Laux, T. (2011b). Short Article Transcriptional Activation of Arabidopsis Axis Patterning Genes WOX8 / 9 Links Zygote Polarity to Embryo Development. Developmental Cell *20*, 264–270.

Urao, T. (1993). An Arabidopsis myb Homolog Is Induced by Dehydration Stress and Its Gene Product Binds to the Conserved MYB Recognition Sequence. PLANT CELL ONLINE *5,* 1529.

Vollbrecht, E., Veit, B., Sinha, N., and Hake, S. (1991). The developmental gene Knotted-1 is a member of a maize homeobox gene family. Nature *350,* 241-243.

Weigel, D., and Meyerowitz, E.M. (1994). The ABCs of floral homeotic genes. Cell *78,* 203-209.

West, M. (1994). LEAFY COTYLEDON1 Is an Essential Regulator of Late Embryogenesis and Cotyledon Identity in Arabidopsis. THE PLANT CELL *6,* 1731-1745.

Willemsen, V., Bauch, M., Bennett, T., Campilho, A., Wolkenfelt, H., Xu, J., Haseloff, J., and Scheres, B. (2008). The NAC Domain Transcription Factors FEZ and SOMBRERO Control the Orientation of Cell Division Plane in Arabidopsis Root Stem Cells. Developmental Cell *15,* 913-922.

Xia, C., Wang, Y.-J., Liang, Y., Niu, Q.-K., Tan, X.-Y., Chu, L.-C., Chen, L.-Q., Zhang, X.-Q., and Ye, D. (2014). The ARID-HMG DNA-binding protein AtHMGB15 is required for pollen tube growth in Arabidopsis thaliana. Plant Journal *79,* 741-756.

Xie, Q., Frugis, G., Colgan, D., and Chua, N.H. (2000). Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes and Development *14,* 3024-3036.

Yadav, R.K., Tavakkoli, M., Xie, M., Girke, T., and Reddy, G.V. (2014). A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. Development *141*, 2735–2744.

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., et al. (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome. Science *302,* 842-846.

Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei, Z., Xiaoxiao, W., Xiaoming, Q., et al. (2006). The MYB transcription factor superfamily of Arabidopsis: Expression analysis and phylogenetic comparison with the rice MYB family. Plant Molecular Biology *60,* 107-124.

Yanofsky, M., Ma, H., Bowman, J., Drews, G., Feldmann, K., and Meyerowitz, E. (1990). The protein encoded by the Arabidopsis homeotic gene AGAMOUS resembles transcription factors. Nature *346,* 35-39.

Zhang, M., Hu, X., Zhu, M., Xu, M., and Wang, L. (2017). Transcription factors NF-YA2 and NF-YA10 regulate leaf growth via auxin signaling in Arabidopsis. Scentific Reports *7.*

Zhong, R., Lee, C., Zhou, J., McCarthy, R.L., and Ye, Z.-H. (2008). A Battery of Transcription Factors Involved in the Regulation of Secondary Cell Wall Biosynthesis in Arabidopsis. Plant Cell Online *20*, 2763–2782.

CHAPTER 5

GROWTH REGULATING FACTOR 3, a positive regulator of *HOMEODOMAIN GLABROUS 12 (HDG12)*

5.1 Introduction

Growth regulating factors (GRFs) belongs to a small family of transcription factors (TFs), which is found exclusively in plants. GRF family of TFs has been identified in a number of plant species, including *Arabidopsis thaliana, Solanum tuberosum, Zea mays, Glycine max, Brassica napus* and land plants whose genome has been sequenced to date. The first member of this family was identified in rice and named as *Oryza sativa GROWTH REGULATING FACTOR 1 (OsGRF1)* (van der Knaap et al., 2000). Later on, 12 GRF TFs were identified in rice and nine in *Arabidopsis*. In *Arabidopsis,* GRF family is comprises of *AtGRF1* to *AtGRF9*.

The structure of GRF proteins is conserved among different species. GRF proteins contain a QLQ domain (glutamine, lysine, glutamine), followed by presence of bulky aromatic or hydrophobic amino acids. These conserved regions are important for mediating protein-protein interactions. In addition to GRFs, the QLQ domain is also present in yeast SWI2/SNF2 proteins, where it facilitates interaction of SWI2/SNF2 with other proteins to form chromatin-remodeling complex. Other than QLQ domain, a second conserved region is also present in GRF proteins and known as WRC (tryptophan, arginine, cysteine). The WRC domain in these proteins, contains a putative nuclear localization signal and three cysteine and one histidine residue in the sequence $CX_9CX_{10}CX_2H$, which functions in metal binding. Therefore, this conserved region helps in targeting the GRF proteins to the nucleus and binding to the DNA (Choi et al., 2004; Kim et al., 2003a; van der Knaap et al., 2000; Raventós et al., 1998). Proteins with QLQ domain are found in all eukaryotes; however, WRC domain is specific only to the plants. The QLQ and WRC domains are present at the N-terminal region of the GRF proteins. All the nine members of GRF family in *Arabidopsis,* possess the QLQ and WRC domains. However, AtGRF9 possesses a second WRC domain at the C-terminus of the protein.

Although the N-terminal region of the GRFs is quite conserved, their C-terminal parts are quite variable. In *Arabidopsis,* AtGRF3 and AtGRF4 are the most similar members with 54% identity in their C-terminal region. AtGRF1 and AtGRF2 share next highest similarity, with 44% identical amino acids at C-terminus of the protein. AtGRF7 and AtGRF8, AtGRF5 and AtGRF6 show very less sequence similarity of 16% and 17%, respectively at their C-terminus. AtGRF9 represents a distinct member due to its unique structure and presence of an extra WRC domain.

All GRF family of TFs are regulated post-transcriptionally by miRNA396, except *AtGRF5* and *AtGRF6*. miRNA396 shares complementarity with the WRC coding region of the GRFs and thereby regulates their expression. miRNA396 is known to play an important role in leaf growth and development by post-transcriptionally regulating the GRF expression. Expression of miRNA396 increases with the maturity of leaf and thereby it reduces the transcript levels of target GRFs. Therefore, expression of GRFs always remains very high during early leaf development and with maturity their expression goes down and thus allowing the transition from cell proliferation to cell differentiation.

In the past, various studies have shown the role of GRFs in regulating the size and growth of leaf in *Arabidopsis*. Kim *et al*. (2003) showed that over-expression of *AtGRF1* and *AtGRF2* results in larger leaves and cotyledons size compared to WT. The length and the width of leaf blade were increased in GRF over-expression lines. The analysis of *grf123* mutant phenotype also supports their role in leaf development. *grf123* triple mutant plants developed smaller leaves and cotyledons. Scanning electron microscopy revealed the differences in the size of epidermal cells in GRF over-expression and mutant plant leaves. In *35S:AtGRF2* over-expression plant lines the leaf size is bigger compared to the WT. In contrast, *grf123* triple mutant plant leaves are smaller. The cellular basis of alterations in the size of the leaf were determined by scanning electron microscopy of epidermal cells. The epidermal cells were bigger in the over-expression lines and smaller in the mutant. Cells of the petiole were also longer in comparison to the WT. However, the numbers of cells were not significantly different in over-expression and mutant as compared to the WT. Thus, showing that changes in the size of the leaves were due to corresponding changes in the cell size. Also, the plants over-expressing miRNA396, *35S:miR396* display a stronger reduction in leaf size due to reduced expression of several GRFs with redundant functions.

Other than being regulated by miRNA, GRFs work in concert with cofactor, growth interacting factor or GIF. Co-expressing *rGRF3* and *GIF1* causes an additive increase in the size of the leaf than *rGRF3* alone. GIF1 increases the activity of GRF3 by causing an increase in the number of palisade cells in the leaf. GIF1 stimulates the activity of GRF3 by controlling the expression of cell proliferation markers *CYCLINB1*, *CYCLIND3* and *KNOLLE* (Debernardi et al., 2014a)..

Increase in the expression of the proliferation markers, causes the cells to divide, thereby increasing the number of cells and thereby the size of the leaf.

In this study, *HOMEODOMAIN GLABROUS 12 (HDG12)* was identified as a downstream target of GROWTH REGULATING FACTOR 2 / 3 (GRF 2 / 3), and shown that *HDG12* functions downstream of GRFs in promoting leaf growth by regulating cell expansion and cell growth (Harish and Ram Yadav unpublished data). The evidence presented in this chapter of my thesis clearly establishes GRF3 as positive regulator of *HDG12*.

5.2 Materials and methods

5.2.1 Vector construction and transformation

For ectopic expression of *GRF3* and *HDG12,* the open reading frames were amplified from *Arabidopsis* Col-0 cDNA and cloned into pENTR/D/TOPO vector and sequence verified. Forward primer CACCATGGAGTTTCTCGGCGACAG and reverse primer TCAAGCAGTTGAAGGACAATTCAA was used to amplify *HDG12* ORF. Forward primer CACCCATATGGATTTGCAACTGAAACA and reverse primer TCAATGAAAGGCTTGTGTCGAGAC was used for amplifying *GRF3* ORF. The pENTR/D/TOPO clones were then used to set up LR clonase recombination reaction with gateway compatible over-expression vector pMDC32 carrying the 35S promoter. Over-expression constructs were then dipped into *Arabidopsis* Col-0 plants by floral dip method using *Agrobacterium* GV3101 strain. All the over-expression vectors were made and followed *in planta* (Harish and Ram Yadav unpublished data).

5.2.2 Confocal Imaging

Imaging of leaf epidermis was done using Leica SP8 upright confocal microscope. Cell outlines were marked using Propidium iodide (100 μg/ml, P21493, Invitrogen) dye in dH_2O containing silwett. Propidium iodide dye was excited using argon laser and emission spectra was collected through a long-pass filter (BP 610-672nm). Raw confocal images were then used for measuring cell size using ImageJ software.

5.3 Results

5.3.1 GRF2 and GRF3 bind to the *HDG12* promoter

A 3 kb promoter region of *HDG12* gene was cloned into yeast vector pMW2 upstream of the *HIS3* reporter gene. The pMW2 construct was then integrated into the yeast genome and transformants were selected on selective media lacking histidine. *HDG12* bait was screened against a library of 327 preys in the high-throughput Y1H screen. Only GRF3 prey protein was found to interact with *HDG12* promoter even at very high concentrations of 3-AT, a competitive inhibitor of histidine biosynthesis enzyme HIS3. Protein-DNA interaction of GRF3 on *HDG12* promoter obtained in the high-throughput screen was also tested manually and was reproduced. Since the *grf123* triple mutant display smaller leaf size in comparison to WT, I investigated the interaction of GRF1 and GRF2 on the promoter DNA of *HDG12* manually because prey vector of both these TFs were missing from the library. Y1H assay was conducted using GRF1, GRF2 and GRF3 as preys against the *HDG12* promoter bait. My results show GRF2 interacts with *HDG12* promoter DNA. Surprisingly, GRF1 did not bind to the full length *HDG12* promoter (Figure 5.1).

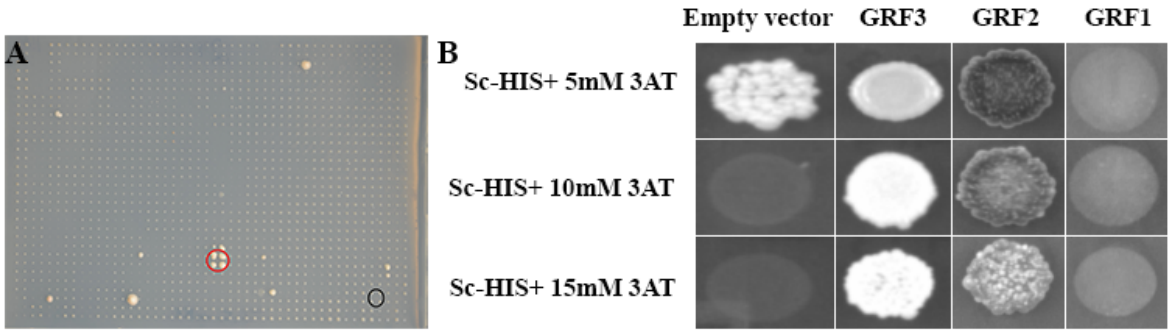


Figure 5.1: Y1H assay for *HDG12* promoter. (A) High-throughput Y1H assay plate showing the interaction of GRF3 (in red) on *HDG12* promoter bait, on selective media containing 3-AT. The negative control or the empty vector spot is marked with black circle. (B) Y1H analysis of *HDG12* bait with GRF1, GRF2 and GRF3 TF protein.

5.3.2 GRF1, GRF2 and GRF3 show strong binding within the first 500bp of *HDG12* promoter

In order to figure out the region of the promoter that is preferred by GRFs for binding on *HDG12* promoter, the *HDG12* promoter was chopped into 4 smaller fragments of 500bps. The fragments ranged from -1 to -500, -501 to -1000, -1001 to -1500, -1501 to -2000. Within the 2kb region, 4 GTCAG cores were predicted at -350, -855, -1356, -1937 position. The smaller promoter fragments were cloned into yeast vector pMW2 and transformed into yeast.

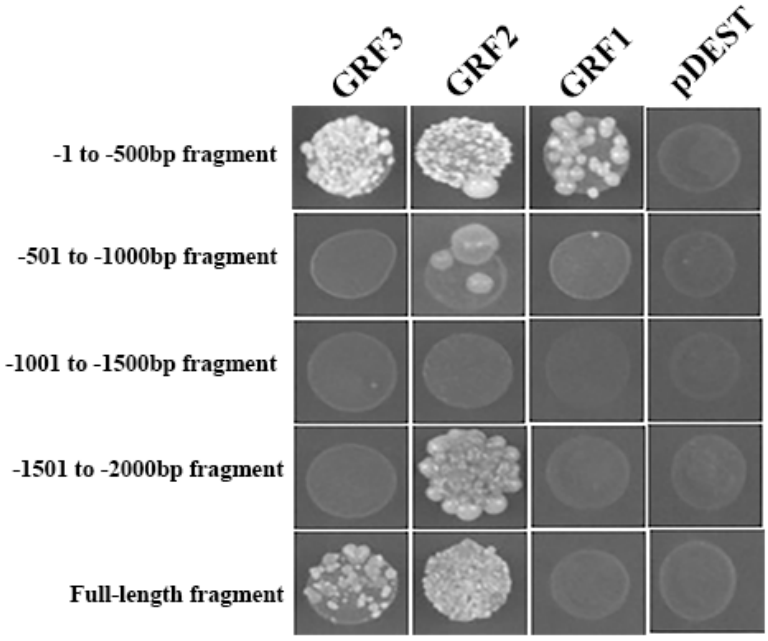


Figure 5.2: Y1H assay for *HDG12* promoter fragments. Y1H analysis of chopped and full length fragments of *HDG12* promoter bait with GRF1, GRF2 and GRF3 TF protein.

Interaction of these smaller fragments of *HDG12* promoter was then tested against GRF1, GRF2 and GRF3. From the assay, I concluded that GRF1, GRF2 and GRF3 bind to *HDG12* within the first 500bp promoter region. Furthermore, a strong interaction was found on the 4th fragment whereas and a weak interaction on the 2nd fragment for GRF2 suggesting that GRF2 must be binding to other motifs that are present within the 2nd and the 4th fragment of HDG12 promoter region. GRF1 did not show any binding on *HDG12* promoter in yeast, when full-length fragment was used. However, GRF1 was bound to *HDG12* promoter when the smaller promoter fragment was used (Figure 5.2), suggesting that smaller fragments might work better in yeast and give

positive interactions which I probably missed while using bigger promoter fragment. Based on Y1H data of shorter fragments (500bp), it appeared that GRF3 might be using the GTCAG core present at -350 position on the *HDG12* promoter. To validate this further, the 500bp fragment was further chopped into five distinct fragments of 100bp each. Interestingly, GRF2 and GRF3 were found to bind within the -100bp upstream region (Figure 5.3). GRF1 also bound within the -100bp upstream region, though its binding was not as strong as GRF2 and GRF3. This suggests that GRFs prefer a novel binding site in the core promoter of *HDG12*. However, the exact binding site of GRFs present in the core promoter still needs to be determined.

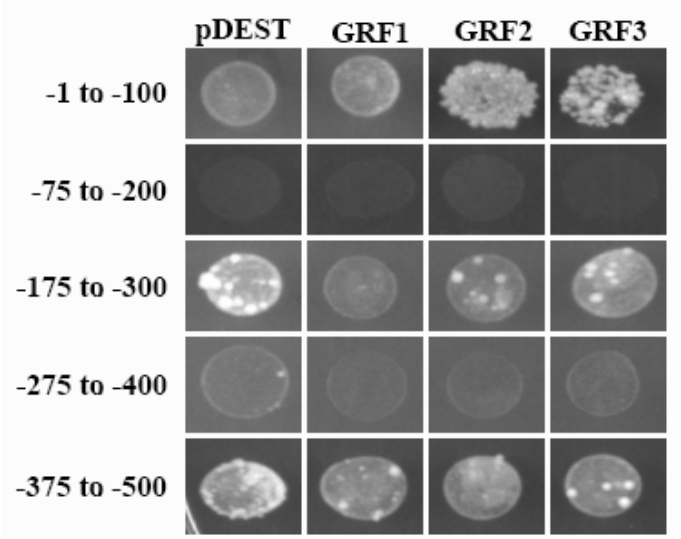


Figure 5.3: Y1H assay for *HDG12* promoter fragments. Y1H analysis of 100bp fragments of *HDG12* promoter bait with GRF1, GRF2 and GRF3 TF protein.

5.3.3 GRF2 and GRF3 expression overlaps with *HDG12*

HDG12 transcriptional and translational expression patterns were studied by generating stable transgenic lines, respectively. The *pHDG12::H2B-YFP* and *pHDG12::HDG12-EGFP* constructs were introduced into WT L*er* plants. The transgenic lines were then screened for YFP and EGFP fluorescence. *HDG12* expression is restricted to the epidermal layer of shoot apex and floral primordia (Prince, Harish and Ram Yadav unpublished data). Previous studies have also reported the expression of *HDG12* in the shoot epidermis as well as in protodermal layer of early embryos (Horstman et al., 2015). In the mature leaf tissue, its expression was confined to the pavement cells of the epidermis (Harish and Ram Yadav unpublished data). Using GUS reporter

also in the past, *HDG12* expression has been reported in leaf primordia and developing leaves (Nakamura et al., 2006). *HDG12* reporter construct displays an overlap in its expression with *GRF2* and *GRF3* expression. In the shoot apical meristem, both GRF2 and GRF3 are expressed more broadly (Figure 5.4). However, unlike *GRFs,* expression of *HDG12* is restricted to the epidermis. Also, using GUS reporter, GRF3 expression has been reported in the SAM tissue where cells are dividing frequently and in the early stages of leaf development (Kim et al., 2003b). Based on the binding evidence, I speculated overlap of *HDG12* expression pattern with that of GRF2 and GRF3 (Fig 5.4).

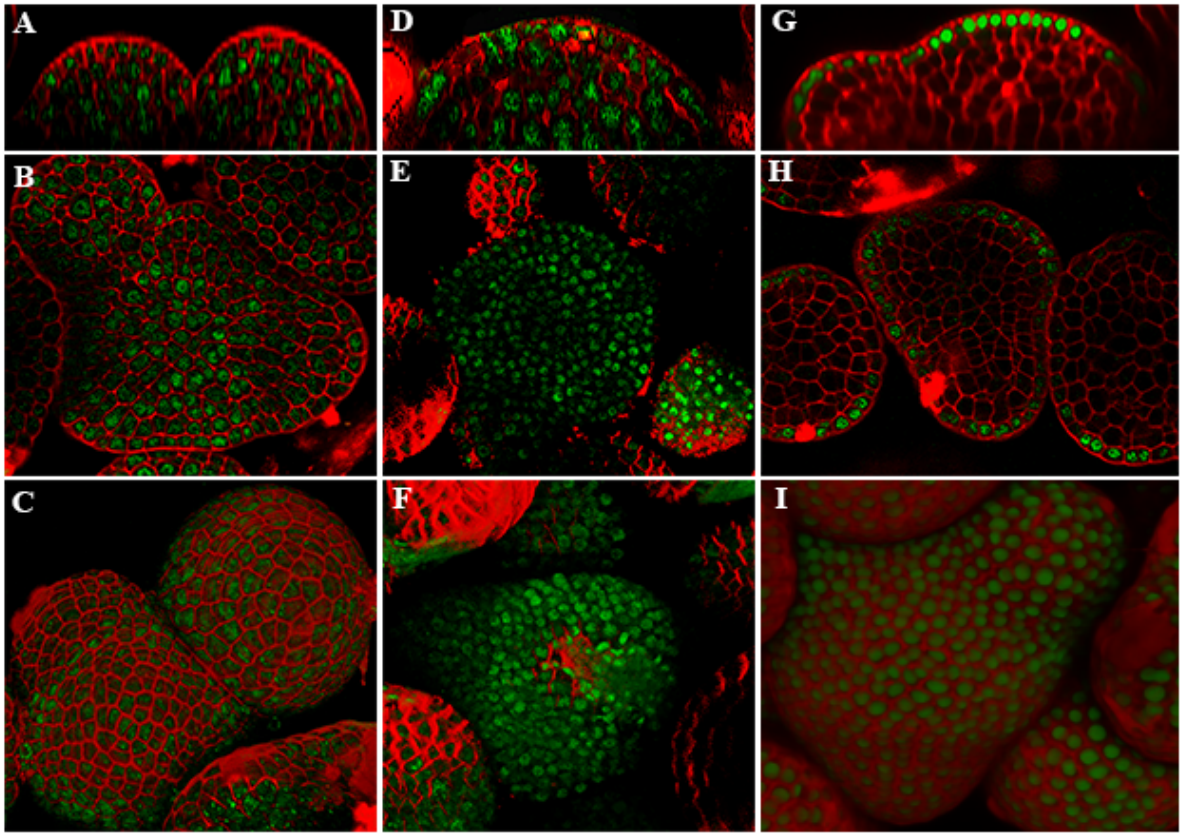


Figure 5.4: Expression pattern of GRF3, GRF2 and *HDG12* in the shoot. (A, B, C) Side view, transverse section and 3D reconstructed top view, respectively, of the WT L*er* shoot apex, showing the expression pattern of GRF3 protein. (D, E, F) Side view, transverse section and 3D reconstructed top view, respectively, of the WT L*er* shoot apex, showing the expression pattern of GRF2 protein. (G, H, I) Confocal images showing the epidermis specific expression of *pHDG12:H2B*-YFP in side view, transverse section and 3D reconstructed top view respectively. Cell outlines highlighted with Propidium iodide (red).

5.3.4 *hdg12* single mutant plants display larger leaf size

To study the function of *HDG12,* T-DNA insertion knockout for *HDG12* was obtained from the stock center, and insertion was determined. SALK_127261 line carries a T-DNA insertion within the 3rd exon of the gene and was found to be null for RNA (Refer chapter 4). *hdg12* mutant plants showed larger leaf size as compared to the WT plants. Confocal imaging of epidermis of the first leaf of *hdg12* single mutant displayed a reduction in the size of the pavement cells (data not shown). Pavement cell size of *hdg12* single mutant was rescued by the *35S:HDG12* over-expression (Harish and Ram Yadav unpublished data). The above results suggest that *HDG12* plays a role in regulating the size and growth of leaf during development.

5.3.5 GRF3 positively regulates the *HDG12* expression

HDG12 expression was detected in epidermal cell layer by confocal microscopy in shoot apex and leaf pavement cells. HDG12-EGFP fusion protein stays in the same cells where it is synthesized at the first place (Harish and Ram Yadav unpublished data). However, upstream regulators (GRFs) are expressed broadly. Initially, I investigated the expression of *HDG12* in GRF3 overexpression lines (# 4) to see its regulation. Despite doing several independent qRT-PCR experiments to confirm the finding in gain of function *GRF3* lines. I did not find convincing evidence for *HDG12* regulation. Therefore, in order to study *in-vivo,* the cell type specific regulation of *HDG12* by GRF3, *pHDG12:HDG12-EGFP* transgenic line was crossed with *35S:GRF3* transgenic line. Plant line showing both high *GRF3* expression levels and bigger leaf size phenotype was crossed with *pHDG12:HDG12-EGFP* line. The *pHDG12:HDG12-EGFP* line served as a control. Eighth rosette leaf was taken from sample and control plants, respectively, for cell sorting (Shivani, Monika, Harish and Ram Yadav unpublished data).

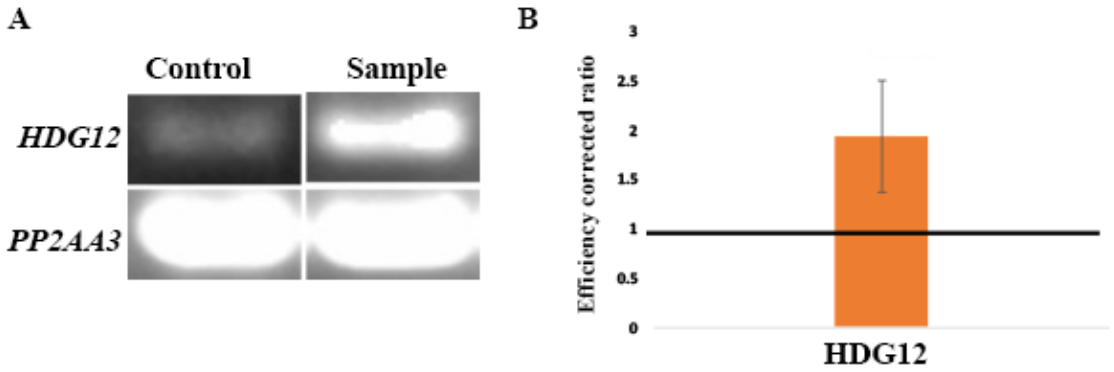


Figure 5.5: Quantification of *HDG12* levels. (A) Semi-quantitative PCR showing the over-expression of *HDG12* transcript in sorted cells of *35S:GRF3 X HDG12:HDG12-EGFP* plants in comparison to control *HDG12:HDG12-EGFP* plants. (B) Real time qRT-PCR results showing the up regulation in *HDG12* transcript levels in sample as compared to control (shown in black line).

Total RNA was isolated from approximately 80,000 sorted cells. mRNA was converted into cDNA for both sample and control. *HDG12* transcript levels were checked both semi-quantitatively and by real time qRT-PCR. *HDG12* transcript levels was found to be approximately two times up regulated in *GRF3* over-expression lines in comparison to control. Clearly indicating a positive regulation of *HDG12* by GRF3.

5.4 Discussion

One of the major questions in network biology is how broadly expressed upstream TFs regulate the narrowly expressed target genes. Often the transient perturbation experiments coupled with microarray studies or RNAseq experiments fail to capture the regulation of narrowly expressed cell type specific genes. Similarly, the promoter fragments of narrowly expressed target genes are also under represented in the chromatin immunoprecipitation assay due to their restricted expression in few cells, and even if they pulled efficiently they further gets diluted when entire tissue is harvested for immunoprecipitation. To overcome such technical difficulties, Y1H based approaches can be used to resolve the regulatory network. Often developmental biologists studying phenotypes come across genes whose phenotypes are closely similar. However, it is challenging to deduce the regulatory hierarchy among them. The Y1H strategy employed in this

study clearly indicated linear relationship among the genes involved in closely related phenotypes.

The role of *GRF3* is well studied in the past in the context of leaf development. *grf3* single mutant display reduction in the size of the leaves by 15% as compared to the WT (Debernardi et al., 2014b). *grf123* triple mutant displays a significant reduction in the size of leaf and cotyledons. However, single mutants of *grf1* and *grf2* do not exhibit much difference in the size of the leaf in comparison to WT, possibly due to redundancy among members of this family. But plants over expressing miR396 transgene under 35S promoter display significantly reduced leaf and cotyledon size in Arabidopsis. This is due to the miR396 mediated down regulation of GRFs in the transgenic lines. In contrast, plants over expressing miR396 resistant version of *GRF3* transcript have bigger leaf size in comparison to WT control (Debernardi et al., 2014b). Previous studies also suggested that the increase in the leaf size is both due to increase in cell proliferation and cell expansion in leaf epidermis (Kim et al., 2003). However, later studies dissected the function of GRFs in cell proliferation by showing their direct role in regulating Cyclins and CDKs expression in the early leaf development. But, it is still not clear how GRFs coordinate both cell proliferation and cell expansion and cell growth at the same time across the epidermal and sub-epidermal cell layer.

Function of GRFs in cell growth and cell expansion in leaf is still not resolved. In this study, I discovered GRF3 binds to the *HDG12* promoter in the high-throughput Y1H screen. Further, I studied the transcriptional regulation of *HDG12* in *grf123* mutant background and confirmed it's down regulation in the absence of functional GRFs. Analysis of T-DNA insertion mutant line of *hdg12* revealed reduction in the size of the leaf and cotyledon in the mutant plant. The leaf and cotyledon phenotype linked to *hdg12* loss of function was rescued by over-expressing of *HDG12* WT transgene. Also, over-expressing *HDG12* in *grf123* triple mutant background, rescues the leaf size (Harish and Ram Yadav unpublished data). Taken together our results suggest that GRF3 might be controlling the leaf development by regulating a key downstream transcription factor, *HDG12,* in the epidermal cell layer.

Confocal imaging of pavement cells of the *hdg12* single mutant plants revealed a smaller cell size in comparison to the WT. And the size of the pavement cells was rescued when *HDG12* was expressed in the single mutant background. Suggesting, that *HDG12* plays a role in controlling cell expansion, thereby affecting the growth and development of the leaf. The past studies have also shown that HDG proteins play a role in cell differentiation (Horstman et al., 2015). The cells grow in size up to a certain extent, before they can undergo differentiation. During leaf morphogenesis, cell proliferation and cell differentiation events occur concurrently to allow proper leaf development. Wherein, *GRF3* plays a role regulating cell proliferation in the leaf, *HDG12* seems to affect cell expansion and cell differentiation, thereby affecting the leaf size and growth during the development. Our findings indicate that GRF3 positively regulates the expression of *HDG12* in the leaf epidermis.

Recently, studies in *Arabidopsis* root has revealed the role of various GRFs in repressing the stem cell-promoting genes in cells that are actively proliferating and repression of GRF expression in the quiescent center of the root to maintain the stem cell niche in the roots. However, there is no knowledge about what role the GRFs play in the shoot. Future studies would explore downstream targets of GRFs in the shoot apex to understand their role in cell proliferation and cell growth.

5.5 References

Choi, D., Jeong, H.K., and Kende, H. (2004). Whole genome analysis of the OsGRF gene family encoding plant-specific putative transcription activators in rice (Oryza sativa L.). Plant and Cell Physiology *45,* 897-904.

Debernardi, J.M., Mecchia, M.A., Vercruyssen, L., Smaczniak, C., Kaufmann, K., Inze, D., Rodriguez, R.E., and Palatnik, J.F. (2014a). Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. Plant Journal *79,* 413-426.

Debernardi, J.M., Mecchia, M.A., Vercruyssen, L., Smaczniak, C., Kaufmann, K., Inze, D., Rodriguez, R.E., and Palatnik, J.F. (2014b). Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. Plant Journal *79,* 413-426.

Horstman, A., Fukuoka, H., Muino, J.M., Nitsch, L., Guo, C., Passarinho, P., Sanchez-Perez, G., Immink, R., Angenent, G., and Boutilier, K. (2015). AIL and HDG proteins act antagonistically to control cell proliferation. Development *142,* 454-464.

Kim, J.H., Choi, D., and Kende, H. (2003a). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. Plant Journal *36,* 94-104.

Kim, J.H., Choi, D., and Kende, H. (2003b). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. Plant Journal *36,* 94-104.

Van der Knaap, E., Kim, J.H., and Kende, H. (2000). A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth. Plant Physiology *122,* 695-704.

Nakamura, M., Katsumata, H., Abe, M., Yabe, N., Komeda, Y., Yamamoto, K.T., and Takahashi, T. (2006). Characterization of the class IV homeodomain-Leucine Zipper gene family in Arabidopsis. Plant Physiology *141,* 1363-1375.

Raventós, D., Skriver, K., Schiein, M., Karnahl, K., Rogers, S.W., Rogers, J.C., and Mundy, J. (1998). HRT, a novel zinc finger, transcriptional repressor from barley. Journal of Biological Chemistry *273,* 23313-23320.