ROLE OF A GATA-TYPE ZINC FINGER PROTEIN IN REGULATING SEED DORMANCY IN *ARABIDOPSIS THALIANA*

PRATIBHA RAVINDRAN

B.Tech.(Hons.) Biotechnology, SASTRA, Thanjavur

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

FACULTY OF SCIENCE

NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

_____

Pratibha Ravindran

20 January 2014

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank each and everyone who has been a part of this thesis, either in small or big way.

First of all, I am very grateful to my supervisor, Professor Prakash Kumar, for having given me the opportunity and platform to work on this project. I thank him for his timely guidance and support throughout the years of my research. His constant motivation and trust in me has helped me to design my own experiments with confidence and evolve as an independent researcher.

My next big 'thank you' would be to my lab mate and mentor, Dr. Petra Stamm. Having joined right after my undergraduate course, with very little knowledge in field of research, Petra helped to nurture me into a scientist. She introduced me to every experiment and protocol in lab and taught me with great patience and passion. She also played a huge part in helping to initialize this project. I owe every bit of my growth as a scientist to her. It was always a pleasure talking to her, be it about science or outside of it. Thank you Petra!

I would also like to thank all my other lab mates who have helped me throughout my years of research. Special thanks to Vivek and Vijay for all the discussions and fun talk that kept me cheerful and for having made me feel more comfortable in the lab. I thank Vivek for his constant motivation and support in times of both triumph and failures that helped me move forward. Thank you Pavithra for being a good friend and colleague. Working on and optimizing some of the protocols could not have been easier without your help. Thank you Ram for always being ready to help whenever I faced some technical issues. I appreciate the friendship of all my other former and current lab mates and feel lucky to have met them all.

I owe a special thanks to Mrs. Ang-Lim Swee Eng, our lab officer, who always made sure that the lab was well-equipped and safe. Working in lab could not have been as smooth as it was without her around. Thank you for always being ready to help wholeheartedly and for making the lab environment more pleasant with your friendly talk and cheerful smile.

I would also like to thank all my friends and former housemates in Singapore. These four and a half years could not have been as much as fun as it was without you all. Your presence always made me feel secure and happy and kept me moving through all the difficult phases. Special thanks to my dear

friends, Tanujaa and Sindhuja. The times spent with you both, be it at home, university canteen, fitness classes or at some vacation destination, will remain as some of the most cherished and unforgettable moments of my life.

It is with a deep sense of gratitude that I thank the Department of Biological Sciences, NUS, for having supported me financially throughout my postgraduate course.

I wholeheartedly thank my parents for having given me the freedom to pursue my interests and for always unconditionally supporting me in whatever I chose to do. Their unwavering love and care has been instrumental in my success. I would also like to thank my other family members for all the kindness and love they have showered upon me.

Last but not the least, I would like to thank my life partner, Arvind. He has always been very loving, understanding and a constant source of encouragement. He has been my punch bag on bad days, and has been ever patient and caring. His unconditional love and concern has helped me to come this far.

PUBLICATIONS

P. Stamm, **P. Ravindran**, B. Mohanty, E. Tan, H. Yu and P. Kumar (2012). "Insights into the molecular mechanism of RGL2-mediated inhibition of seed germination in Arabidopsis thaliana." BMC plant biology **12**(1): 179.

El-Sharkawy, S. Sherif, W. El Kayal, A. Mahboob, K. Abubaker**, P. Ravindran**, P. A. Jyothi-Prakash, P. P. Kumar and S. Jayasankar (2013). "Characterization of gibberellin-signalling elements during plum fruit ontogeny defines the essentiality of gibberellin in fruit development."Plant molecular biology: 1-15.

Conference contributions:

- 24th International Conference on Arabidopsis Research (ICAR), 24 to 28 June, 2013, Sydney, Australia.

Poster: Role of a GATA-type transcription factor in regulating seed dormancy downstream from DELLA protein, RGL2. **P. Ravindran**, P. Stamm and P. Kumar.

- 17th Biological Sciences Graduate Congress, December 2012, Thailand.

Poster: Molecular mechanisms underlying seed germination downstream of the DELLA protein, RGL2. **P. Ravindran**, P. Stamm and P. Kumar.

- 16th Biological Sciences Graduate Congress, December 2011, Singapore.

Oral: Molecular mechanisms underlying seed germination downstream of the DELLA protein, RGL2. **P. Ravindran**, P. Stamm and P. Kumar.

- 15th Biological Sciences Graduate Congress, December 2010, Malaysia.

Poster: Characterization of the downstream targets of *Arabidopsis* DELLA protein RGL2. **P. Ravindran**, P. Stamm and P. Kumar.

TABLE OF CONTENTS

SUMMARY

Seed biology is one of the most widely studied fields in plant physiology because of its immense significance in agriculture. However, most mechanisms regulating seed germination and dormancy are not fully understood. Pre-harvest sprouting and asynchronous germination are significant problems in agriculture and understanding the mechanism regulating seed dormancy will help to device strategies to overcome these problems.

Seed germination is under the tight control of phytohormones, gibberellic acid (GA) and abscisic acid (ABA), the levels of which determine the switch from dormancy to germination under favourable environmental conditions. One of the key transcriptional repressors of GA signalling in seed germination is the DELLA protein, RGA-Like 2 (RGL2). However, the molecular mechanism of repression downstream of RGL2 is unknown. Based on an earlier microarray analysis done in our laboratory, we identified a gene-encoding a GATA-type zinc finger transcription factor (GATA12) as one of the downstream targets of RGL2 protein. We found that freshly harvested seeds of *GATA12* antisense suppression lines have reduced dormancy than WT, while overexpression lines show enhanced dormancy. Our data shows that *GATA12* transcripts are negatively regulated by GA. Also, its transcript levels reduce dramatically under dormancy-breaking conditions like dry storage and cold stratification of seeds. Using chromatin immunoprecipitation studies we showed that a protein complex containing RGL2 can bind to the promoter of

GATA12 and thereby regulate the gene. A motif analysis conducted showed that *GATA12* promoter has several GAMYB and Dof-associated motifs that are known to be GA and RGL2 responsive, respectively. This reinforces the GA-mediated regulation of *GATA12*. The findings described in this thesis have contributed more to our knowledge of gibberellin signalling downstream of RGL2 and thus have shed more light on the mechanism of dormancy release and seed germination.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Gene / protein names and descriptions	
ABI5	Abscisic acid-insensitive 5
BME3	*BLUE MICROPYLAR END3*
BSA	bovine serum albumin
CPS	copalyl diphosphate synthase
Dof	DNA-binding with one finger
FUS3	*FUSCA3*
GA20ox	*Gibberellin 20-oxidase*
GA3ox	*Gibberellin 3-oxidase*
GAI	Gibberellic Acid Insensitive
GAMT1	*Gibberellin Methyltransferase1*
GAMT2	*Gibberellin Methyltransferase2*
GFP	Green Fluorescent Protein
GID1	*Gibberellin Insensitive Dwarf1*
GR	glucocorticoid receptor
GUS	β-glucuronidase
HRP	horseradish peroxidase
JAZ1	JA ZIM-domain 1
KO	ent-kaurene oxidase
KAO	ent-kaurenoic acid oxidase
KS	*ent-kaurene synthase*
LEC1	*LEAFY COTYLEDON1*
LEC2	*LEAFY COTYLEDON2*
MFT	*MOTHER OF FT AND TFL1*
NCED	nine-*cis* epoxycarotenoid dioxygenase
PIF4	PHYTOCHROME INTERACTING FACTOR 5
PIL5	PIF3-like 5
RGA	Repressor of ga1-3
RGL2	RGA-like2
RGL3	RGA-like3
SCF	Skp1-Cullin-F-box
SLY	*SLEEPY*
SPT	*SPATULA*
SPY	*SPINDLY*

Chemicals and compounds	
ABA	abscisic acid
Amp	Ampicillin
CaCl2	calcium chloride
DAPI	diamidino phenylindole
DEPC	diethylpyrocarbonate
DEX	dexamethasone
DIG	Digoxigenin
DSG	disuccinimidyl glutarate
DTT	Dithiothreitol
EDTA	ethylene diamine tetraacetic acid
EGTA	ethylene glycol tetraacetic acid
GA	gibberellic acid
HA	hemagglutinin
HCl	hydrochloric acid
IAA	indole acetic acid
JA	jasmonic acid
Kan	Kanamycin
MgCl2	magnesium chloride
NaCl	sodium chloride
Na2HPO4	disodium phosphate
NaH2PO4	monosodium phosphate
PBS	phosphate-buffered saline
PEG	polyethylene glycol
PVA	polyvinyl alcohol
PVDF	polyvinylidene difluoride
SDS	sodium dodecyl sulfate
SSC	saline-sodium citrate
TAE	Tris/acetic acid/EDTA
Tris	Tris (hydroxymethyl) aminomethane
Tween20	polyoxyethylene sorbitan monolaurate
dH2O	deionized water

Units and Measurements	
%	Percent
°C	degree Celsius
µg	microgram
µl	microliter
µM	micromolar
bp	base pairs
cm	centimeter
dpa	days post anthesis
g	acceleration of gravity
g/l	gram per litre
g/ml	gram per millilitre
h	hour
kb	kilo base pairs
kDa	kilo Dalton
kV	Kilo Volt
l	liter
M	molar (mol per liter)
mg	milligram
min	minute
ml	millilitre
mm	millimetre
mM	millimolar
ng	Nanogram
nm	nanometer
OD595	optical density at 595 nm
pH	power of hydrogen
RH	relative humidity
rpm	revolutions per minute
s	second
V	Volt
v/v	percent by volume
w/v	percent by weight

Others	
ABRC	Arabidopsis Biological Resource Center
bHLH	basic helix-loop-helix
cDNA	complementary deoxyribonucleic acid
ChIP	chromatin immunoprecipitation
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
et al.	et alii (Latin for 'and others')
LB	Luria Bertani
LB	Left border
LC	liquid chromatography
LP	Left (forward) primer
mRNA	messenger RNA
MS	Murashige & Skoog
PCR	Polymerase Chain Reaction
RB	Right Border
RNA	ribonucleic acid
RNase	ribonuclease
RP	Right (reverse) primer
RT	room temperature
SD	standard deviation
Taq	Thermophilus aquaticus
T-DNA	Transfer-deoxyribonucleic acid
WT	wild type
qRT-PCR	quantitative real time - Polymerase Chain Reaction

1. INTRODUCTION

1.1 Plant hormones

During the 19th century, pioneering experiments demonstrated that certain 'chemical substances' played an integral role in controlling the growth and development of plants. These substances were later coined as 'plant hormones' or 'phytohormones' (Salisbury and Ross 1992). A plant hormone is generally described as an organic compound synthesized in one part of the plant and translocated to another part where it elicits a physiological response.

The five 'classical' plant hormones are auxins, gibberellins (GAs), abscisic acid (ABA), cytokinin (CK) and ethylene. Other plant growth regulators identified later include: salicylic acid (SA), jasmonic acid (JA) and brassinosteroids (BR). Recently, strigolactones, nitric oxide (NO) and karrikins have also been known to regulate certain aspects of plant growth (Flematti et al. 2004, Umehara et al. 2008, Mur et al. 2013). Each class of hormones has positive as well as inhibitory functions, and most often work in tandem with each other, with varying ratios of one or more interplaying to affect growth regulation (Weier et al. 1979).

In the last decade, study of phytohormone signalling using genetic and molecular tools has gained importance. GAs form an excellent study model to understand how plants modulate their growth and development in response to environmental stimuli. Understanding the signalling pathway downstream

of gibberellins in *Arabidopsis thaliana* (hereafter referred to as *Arabidopsis*) forms the core focus of this thesis.

1.2 Gibberellins

The discovery of gibberellins originated from an investigation of a soil borne disease of rice caused by the fungus *Gibberella fujikuroi.* In the early stages of disease (called 'bakanae' in Japan or 'foolish seedling disease'), the leaves and stem of the infected seedlings elongated more rapidly than those of healthy plants and eventually they wilted and died. In 1938, Teijiro Yabuta and his collaborators (Yabuta and Sumiki. 1938) succeeded in isolating a crystalline active material from the fungal culture which they named as gibberellin A. To date 136 gibberellins have been identified from plants, fungi and bacteria (http://www.plant-hormones.info/gainfo.asp) and have been named GA_1 to GA_{136} in the order of their discovery (MacMillan 2002). But only a small fraction of them are known to be biologically active in plants (e.g. GA1, GA3, GA4, GA7) (Letham et al. 1978, Yamaguchi 2008, Peter and Stephen 2012). They are derived from a basic diterpenoid carboxylic acid skeleton, and commonly have a C3- hydroxyl group (Figure 1.1).

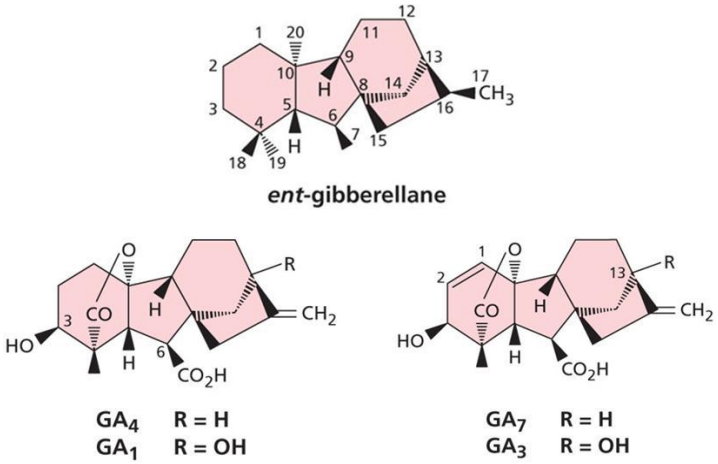


Figure 1.1: Structure of bioactive gibberellins
All GAs possess a tetracyclic (four-ringed) *ent*-gibberellane skeleton containing either 20 carbon atoms (C-20) or a 20-nor-*ent*-gibberellane skeleton, which has only 19 carbon atoms (C-19) because carbon-20 has been lost. The most biologically active GAs namely GA_1, GA_3, GA_4, and GA_7 are all C_{19}-GAs. (structures from Plant Physiology Online, Fifth Edition, Chapter 20)

In addition to stem elongation in plants, there are numerous developmental processes in which GAs participate, including seed development and seed germination, seedling growth, root extension, determination of leaf size and shape, flower induction and development, pollination and fruit expansion (Garcia-Martinez et al. 1997, Peng and Harberd 1997, Lee et al. 2002, Yu et al. 2004). But, the function of GAs in bacteria and fungi still remains unknown.

The function of GA in promoting seed germination which is the first step in a plant's life cycle, and its downstream signalling components will be elaborated in this thesis.

1.3 Gibberellin metabolism

1.3.1 Biosynthesis

GA biosynthesis in plants has been studied for a long time, and numerous genes encoding relevant enzymes have been identified through genetic and molecular approaches. All GAs are tetracyclic diterpenoid carboxylic acids possessing either a *ent*-gibberellane skeleton containing 20 carbon atoms (C20-GA) or a 20-nor *ent*-gibberellane skeleton containing 19 carbon atoms (C19-GA) with a carboxyl group on C-6, a lactone function between C-4 and C-10, and a hydroxyl or other functionality at C-3β (Schomburg et al. 2003, Rieu et al. 2008).

Biosynthesis of GAs in plants occurs in three stages via three classes of enzymes namely: terpene synthases (TPSs), cytochrome P450 monooxygenases (P450s) and 2-oxoglutarate-dependent dioxygenases (2ODDs) (Peter and Stephen 2012).

Geranylgeranyl diphosphate (GGDP) is a common precursor for GAs, carotenoids and chlorophylls. It is converted to tetracyclic hydrocarbon ent-kaurene via a two-step cyclization reaction catalyzed by two TPSs: ent-copalyl diphosphate synthase (CPS) and ent-kaurene synthase (KS) in the plastid (Sun and Kamiya 1997). This is followed by sequential oxidation by two cytochrome P450 monooxygenases in the endoplastic reticulum (ER) namely ent-kaurene oxidase (KO) and ent-kaurenoic acid oxidase (KAO) to produce GA_{12}. In the third stage of GA biosynthesis, GA12 is further converted to GA53 by 13-hydroxylation. GA12 and GA53 are then converted to various GA

intermediates and bioactive GAs in the cytoplasm through a series of oxidation steps catalyzed by 2ODDs, GA 20-oxidases (GA20ox) and 3-oxidases (GA3ox) (Appleford et al. 2006) (Figure 1.2). It has been shown that methylerythritol phosphate pathway in the plastid is the major contributor of isoprene units to GAs in *Arabidopsis* seedlings whereas there is only a minor contribution from the cytosolic mevalonate pathway (Kasahara et al. 2002).

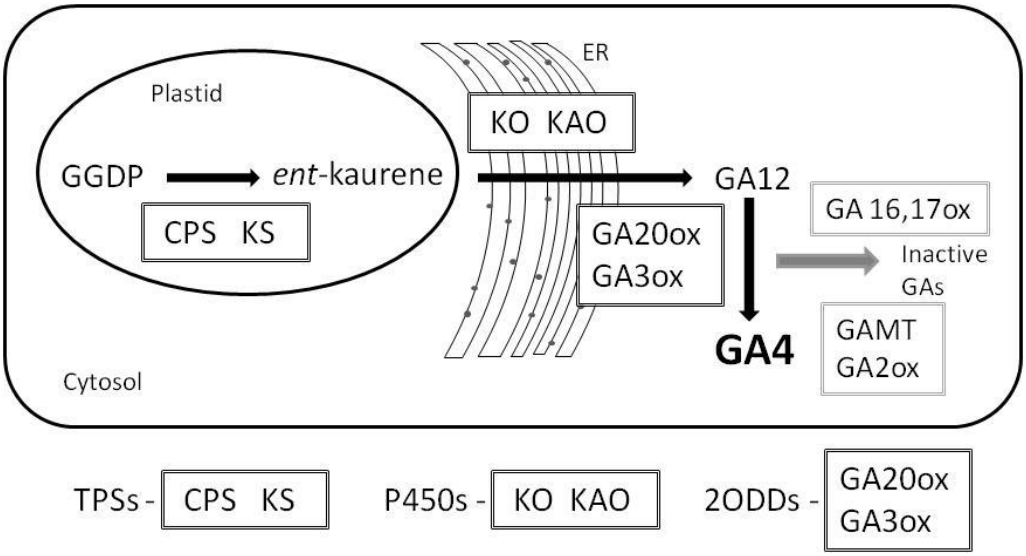


Figure 1.2: Biosynthesis of gibberellins

Biosynthesis of GAs begins from its precusor, geranylgeranyl diphosphate which is converted to ent-kaurene by two terpene synthases : ent-copalyl diphosphate synthase (CPS) and ent-kaurene synthase (KS) in the plastid. This is followed by sequential oxidation by two cytochrome P450 monooxygenases in the endoplastic reticulum (ER) namely ent-kaurene oxidase (KO) and ent-kaurenoic acid oxidase (KAO) to form GA12. Finally, GA12 is further converted to bioactive GAs in the cytoplasm through a series of oxidation steps catalyzed by 2ODDs, GA 20-oxidases (GA20ox) and 3-oxidases (GA3ox). Adapted from Yamaguchi (2008).

1.3.2 Catabolism

Deactivation of GAs is essential to ensure a proper regulation of the bioactive forms. There are several ways by which GAs can be deactivated. A well known deactivation reaction is the 2β-hydroxylation by a class of ODDs, GA 2-oxidases (GA2oxs) (Thomas et al. 1999). Another known mechanism is through the 16α,17-epoxidation of GAs carried out by EUI/CYP714D1, a cytochrome P450 monooxygenase. This was discovered through the study of *eui* mutants in rice that accumulated high levels of bioactive GAs in the upper internodes (Zhu et al. 2006). More recently, it has been shown that *Arabidopsis GAMT1* and *GAMT2,* that encode gibberellin methyltransferases, catalyze the deactivation of GAs by methylating its C-6 carboxyl group using *S* adenosine-L-methionine as a methyl donor. *gamt1gamt2* double mutant seeds germinated better under the influence of a GA biosynthesis inhibitor relative to wild type seeds (Varbanova et al. 2007).

Deactivation may also result from GA conjugation, which involves mainly the formation of glucosyl ethers and esters (Schneider and Schliemann 1994). But, it still remains to be ascertained if these conjugates are indeed formed as a result of feedback loop, to maintain levels of bioactive GAs. In addition, the gibberellin levels are also regulated by external cues like light, temperature, biotic and abiotic stresses (reviewed in (Yamaguchi 2008)).

1.4 Gibberellin signalling

Several GA signalling components have been identified by genetic studies (Davies 1995, Hedden and Phillips 2000, Stamm et al. 2012). The importance

of gibberellins was greatly manifested by its role in generating high yielding varieties of rice and wheat during the 'Green Revolution'. The dwarf varieties had greater resistance to lodging that occurs with high use of chemical fertilizers and thus a higher harvest index (Peng et al. 1999). This significant role in agriculture has further boosted studies on the molecular mechanism of GA action in the last two decades. Several positive and negative regulators of GA signalling have been identified.

1.4.1 Positive regulators

Positive regulators of GA signalling were identified through mutant plants showing impaired responses to GA like dwarfed stature with dark green compact leaves, delayed flowering, reduced fertility, and no or poor seed germination The GA-signalling pathway is partially impaired in the rice dwarf mutant, d1, which is defective in the α-subunit of heterotrimeric G proteins. The d1 mutant is severely affected in GA responsive events like internode elongation and expression of GA-inducible genes in aleurone. Thus, D1 was identified as a positive regulator of GA signalling (Ueguchi-Tanaka et al. 2000).

A major breakthrough was the identification of a soluble GA receptor in rice by the studies of a GA-insensitive dwarf mutant (GID1) (Ueguchi-Tanaka et al. 2005) (ref). One of the well-studied GA responses, the induction of α-amylase gene expression in the aleurone layer of cereal seeds, was absent in the aleurone of *gid1-1* rice seeds even after an external supply of very high amounts of GA. Also, these mutant plants accumulated high levels of

bioactive GAs since they lacked the feedback control of GA signalling. Kinetic analysis of GID1 binding to GAs showed that bioactive GAs had 10-fold higher affinity for GID1 than the inactive forms (Ueguchi-Tanaka et al. 2005). These experiments led to the discovery that GID1 functions as a GA receptor in rice. Following its discovery in rice, three *Arabidopsis* homologs namely GIDa, GIDb and GIDc were found to function analogously to GID1 (Griffiths et al. 2006).

The rice GID1 and the structurally similar *Arabidopsis* SLEEPY1 (SLY1) encode putative F-box proteins and function to degrade the negative regulators of GA signalling through the ubiquitin/proteasome pathway. (McGinnis et al. 2003, Sasaki et al. 2003, Ariizumi et al. 2011).

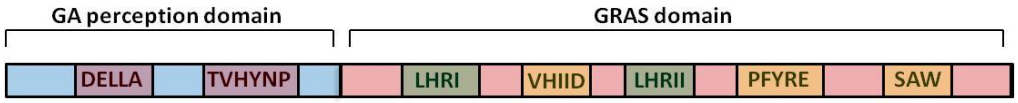
1.4.2 Negative regulators

The GA signalling events are known to be negatively regulated by a class of repressors called DELLA proteins, belonging to the GRAS (GAI, RGA and SCARECROW) family of transcription factors. The DELLA proteins are the most widely researched GA signalling components. They are named after their highly conserved N-terminal DELLA motif, which mediates gibberellin-responsiveness (Peng et al. 1997, Dill et al. 2001, Willige et al. 2007). Like all GRAS proteins, DELLAs share a conserved C-terminal GRAS domain that is involved in transcriptional regulation and is characterized by two leucine heptad repeats (LHRI and LHRII) and three conserved motifs, VHIID, PFYRE and SAW (Bolle 2004) (Figure 1.3A). A single DELLA protein is present in rice (SLENDER RICE1 [SLR1]) and barley (SLENDER1 [SLN1]) and functions to

repress every aspect of GA responses in these species (Peng et al. 1999). Surprisingly, there are five DELLA proteins in *Arabidopsis*: GA-INSENSITIVE (GAI), REPRESSOR OF GA1-3 (RGA), RGA-LIKE1 (RGL1), RGL2 and RGL3 (Koornneef et al. 1985, Peng et al. 1997, Dill and Sun 2001, Peng 2002).

The five DELLA proteins have partially redundant or overlapping functions (Gallego-Bartolome et al. 2010) in repressing GA responses. RGA and GAI are the major repressors of stem elongation (Dill and Sun 2001, King et al. 2001), RGA, RGL1 and RGL2 impair flower development (Wen and Chang 2002, Cheng et al. 2004, Tyler et al. 2004, Yu et al. 2004) whereas RGL2 is the major repressor of seed germination and its function is enhanced by GAI, RGA and RGL1, and RGL3 (Cao et al. 2005, Piskurewicz and Lopez-Molina 2009) (Figure 1.3B). RGL3 also contributes to plant fitness during environmental stress (Wild et al. 2012).

A



B

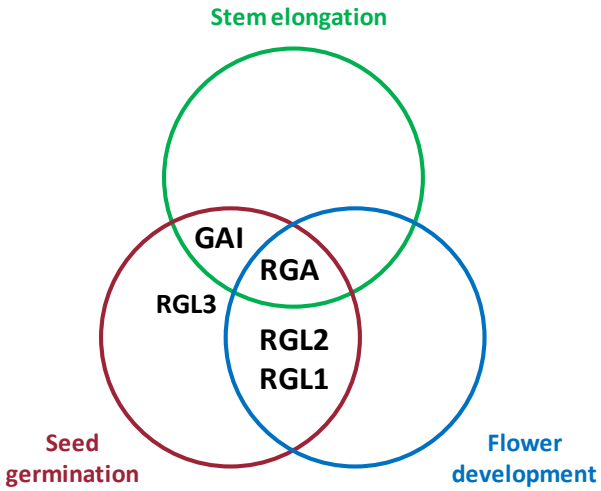


Figure 1.3: DELLA protein structure and function

(A) Schematic representation of domain structure of DELLA proteins showing its conserved N-terminal DELLA (that provides it the name) and TVHYNP domains. The C-terminal region that is responsible for the suppressive function of DELLAs also has conserved motifs like VHIID, PFYRE and SAW and two leucine heptad repeats, LHRI and LHRII. Domain structure representation is adapted from Davière and Achard (2013) **(B)** The five DELLA proteins in *Arabidopsis* have distinct but overlapping functions. GAI and RGA repress stem elongation; RGA, RGL1 and RGL2 repress flower development; RGL2 is the major repressor of seed germination, and this repression is enhanced by RGA, RGL1 and RGL3.

So how do gibberellins suppress the inhibitory effect of these DELLA proteins? A major breakthrough came from the discovery that GA stimulates the disappearance of DELLAs (Silverstone et al. 2001, Gao et al. 2011). The current model for destruction of DELLAs is that upon perception of GA by the soluble GID1-like receptors, the latter undergoes structural changes promoting its interaction with DELLAs (Wang et al. 2009). GID1 contains a GA-binding pocket and a flexible N-terminal extension. Crystal data from Murase et al. (2008) has shown that the C3-hydroxyl group of the GA molecule and the Tyr31 residue of GID1 form a hydrogen bond, inducing a conformational change in the N-terminal extension and thereby causing the closure of the pocket. The upper surface of the lid then binds to the DELLA protein to form the GA-GID1-DELLA complex (Hedden 2008, Murase et al. 2008, Shimada et al. 2008). This, in turn, leads to the recruitment of an SCF E3 ubiquitin ligase, resulting in the destruction of DELLA via the 26S proteasome pathway (Jiang and Fu 2007, Achard and Genschik 2009) (Figure 1.4). The N-terminal DELLA domain (and the adjacent VHYNP domain) is sufficient to mediate this degradation, suggesting that the DELLA domain (and the VHYNP domain) is receiver domain for GA receptors (Griffiths et al. 2006, Willige et al. 2007).

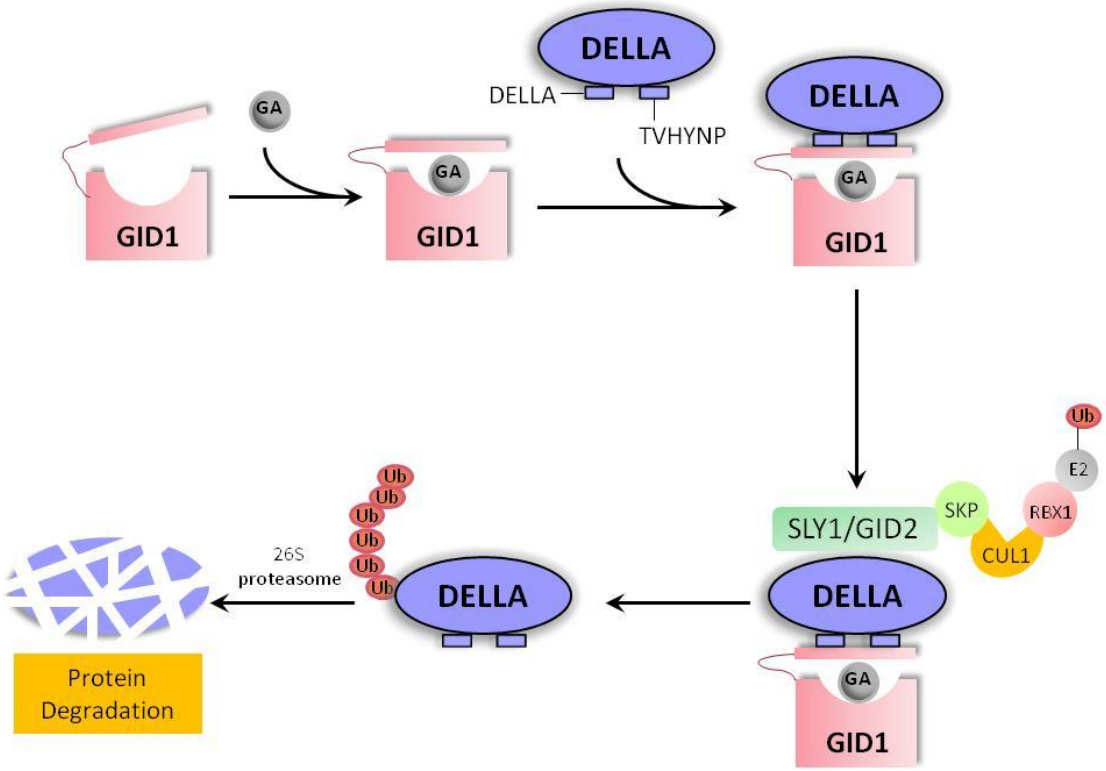


Figure 1.4: Destruction of DELLAs by gibberellins

The current model for destruction of DELLAs under GA-rich conditions is as follows: When GA binds to its receptor GID1, the N-terminal tail of GID1 closes over the pocket as a result of a conformational change. This allows DELLAs to bind to the surface of the lid thereby forming the GA-GID1-DELLA complex. Formation of this complex results in the recruitment of an SCF E3 ubiquitin ligase. Ubiquitin is a 76-amino-acid protein that is covalently linked to proteins in order to flag them for proteolytic processing or destruction. The E3 ubiquitin ligase catalyzes transfer of ubiquitin from the E2 ubiquitin conjugating enzyme onto the DELLA protein, thereby targeting it to the 26 proteasome for degradation. Model re-drawn from Davière and Achard (2013) based on the information from Murase et al. (2008).

1.5 Role of GA mediated signalling in control of seed germination

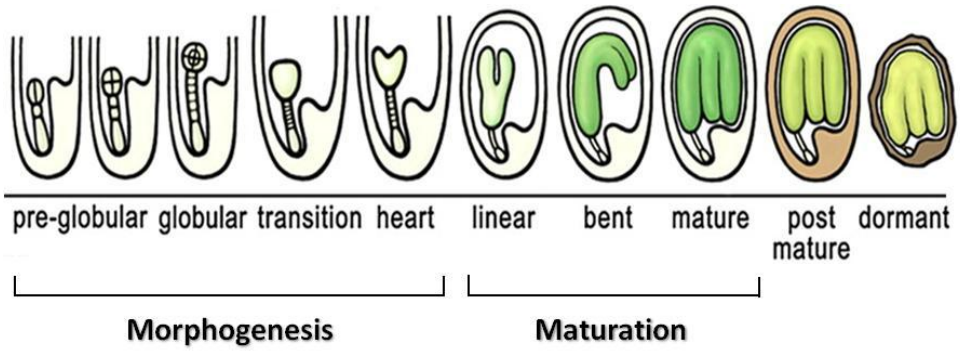
1.5.1 Embryo development and seed structure of *Arabidopsis*

Seed germination is the first crucial step in a plant's life cycle. The *Arabidopsis* mature seed consists of an embryo which is composed of cotyledons (seed leaves), hypocotyl (embryonic axis), shoot apex and radicle (embryonic root) (Figure 1.5B). The embryo is surrounded by an endosperm, a storage tissue, that provides nutrition to the embryo. The outermost testa (seed coat) provides protection to the embryo. It is a dead tissue as it undergoes programmed cell death at maturity. Seed germination begins with the uptake of water by the embryo and ends with the emergence of the embryonic radicle through the testa (Bewley 1997, Finch-Savage and Leubner-Metzger 2006).

Embryogenesis in *Arabidopsis* occurs in two phases: pattern formation or morphogenesis followed by maturation (Figure 1.5A). Morphogenesis starts with the formation of a single-cell zygote and ends at the heart stage, when all embryo structures are patterned and the axes of the plant body plan are defined. At this stage, cell division is arrested and the embryo begins to grow in size. This is the moment at which the seed enters the maturation phase during which processes related to the embryo's entry into metabolic quiescence and subsequent germination occur (Yang and Zhang 2010). Maturation begins with a switch from maternal control to filial control of seed growth (Weber et al. 2005). Dormancy is initiated early during seed maturation and increases until the seed is fully developed (Raz et al. 2001,

Kanno et al. 2010). At the end of maturation, seeds are rich in storage compounds, water content has decreased, ABA levels have increased thereby inducing desiccation tolerance and primary dormancy. It has been demonstrated that nine-*cis* epoxycarotenoid dioxygenases (NCEDs) in *Arabidopsis*, are the key regulatory enzymes in the ABA biosynthetic pathway (Nambara and Marion-Poll 2005). Studies have shown that in whole siliques, *AtNCED6* and *AtNCED9* were highly expressed at mid-embryo development, whereas the expression of *AtNCED2* and *AtNCED3* was high at later stages of embryo development, corresponding to the two peaks of ABA accumulation (Nambara and Marion-Poll 2005, Kanno et al. 2010). *Arabidopsis* embryogenesis is rapid, with the production of desiccated mature seeds within 14 d post fertilization (Lindsey and Topping 1993). Seed maturation and dormancy induction are genetically controlled by at least four major regulators, namely FUSCA3 (FUS3), ABSCISIC ACID INSENSITIVE 3 (ABI3), LEAFY COTYLEDON 1 (LEC1) and LEC2, whose mutants are severely affected in seed maturation and thus show decreased dormancy at maturation (Raz et al. 2001) and reduced expression of seed storage proteins (Gutierrez et al. 2007).

A



B

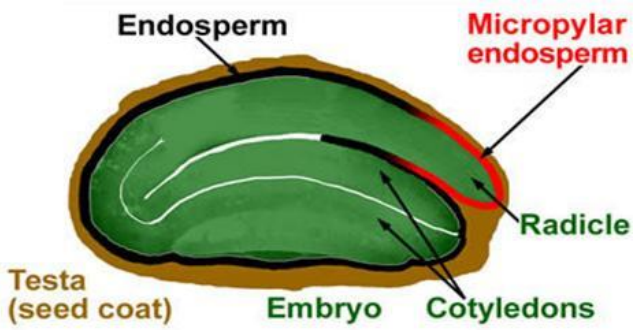


Figure 1.5: *Arabidopsis* embryogenesis and structure of the mature dormant seed

(A) Embryogenesis begins a single-celled zygote that follows a defined pattern of cell division and differentiation to form the mature embryo. The process of embryogenesis is divided into two phases: morphogenesis and maturation. Morphogenesis is characterized by the formation of the basic body plan of the plant. It ends at the heart stage where all the structure are formed and cell division is arrested. This is followed by the maturation phase which is characterized by cell expansion, endosperm absorption, storage reserve accumulation, induction of desiccation tolerance and primary dormancy. The mature seed is thus highly dormant. This figure is taken from the Le et al. (2010) (B) Structure of mature seed : *Arabidopsis* seeds consist of a mature embryo protected by the surrounding endosperm and the seed coat. The mature embryo consists of cotyledons (seed leaves), hypocotyl (embryonic axis) and radicle (embryonic root). The endosperm, being a storage tissue provides nutrition to the embryo. The outermost testa (seed coat) provides protection to the embryo. It is a dead tissue as it undergoes programmed cell death at maturity. This figure is taken from Müller et al. (2006).

1.5.2 Gibberellins release ABA-induced seed dormancy

Freshly harvested mature *Arabidopsis* seeds exhibit primary dormancy, which means that seeds cannot germinate without the aid of dormancy-breaking exogenous factors like moist-chilling (stratification), exposure to light, dry storage (after-ripening) and most importantly endogenous cues like gibberellins (Finkelstein et al. 2008). Seed dormancy is however an adaptive trait the delays germination during harsh environmental conditions under which the probability of survival for seedlings is very low. In agriculture, primary dormancy in seeds is desirable as it helps to prevent precocious germination of seeds while it is still on the plant, a phenomenon called as pre-harvest sprouting that is common in cereal crops like wheat and barley. Early sprouting damage reduces yield and quality of grains. This results in major losses to the farmers and the agriculture industry. On the other hand, some crops are bred to exhibit reduced dormancy so that they have higher germination rates on soil (Burson et al. 2009). Thus, the use of biotechnological tools to alter seed dormancy has captured the interest of many researchers.

Dormancy is maintained by the phytohormone ABA whose levels rise during embryogenesis and are high in fully mature seeds (Karssen et al. 1983, Ogawa et al. 2003). It has been suggested that ABA inhibits water uptake by preventing cell wall loosening of the embryo and thereby reduces embryo growth potential (Schopfer and Plachy 1985). ABA also causes the accumulation of ABSCISIC ACID INSENSITIVE 5 (ABI5), a basic leucine zipper

transcription factor that causes growth arrest by recruitment of the Late Embryogenesis Abundant (LEA) genes, whose products confer osmotolerance to the embryo under harsh environmental conditions (Finkelstein and Lynch 2000, Lopez-Molina and Chua 2000). The ABA responsive element (ABRE), a *cis*-acting element triggering ABA-mediated transcription, and its RY enhancer element, are one of the best-studied components of ABA signalling (Dickinson et al. 1988, Shen et al. 1996, Busk and Pages 1998). It has also been reported that ABI3, which activates RY element-mediated transcription, interacts with ABI5 physically (Nakamura et al. 2001, Park et al. 2011). This physical interaction appears to be important for the synergistic repression of germination through ABRE and RY elements. Thus, ABA is antagonistic in function to GA; ABA promotes dormancy, GA promotes germination.

The repressive effects of ABA are overcome by gibberellins under favourable conditions of light, temperature and moisture. GA biosynthesis and response pathways are activated in response to cold stratification and light resulting in an increase in bioactive GAs via transcription factors like Blue Micropylar End3 (BME3), SPATULA (SPT) and PIF3-like 5 (PIL5) (Liu et al. 2005, Penfield et al. 2005, Oh et al. 2006). Cold treatment not only enhances biosynthesis of GA in seeds but also increases tissue sensitivity to GAs (Ogawa et al. 2003, Yamauchi et al. 2004). The synthesized GAs induce genes encoding enzymes such as endo-β-1,3 glucanase (Leubner-Metzger et al. 1995), β-1,4 mannan endohydrolase (Bewley 1997, Nonogaki et al. 2000) that hydrolyze the endosperm and release the inhibitory effects of ABA on embryo growth

potential (Koornneef et al. 2002, Iglesias-Fernández et al. 2011). This means that ABA and GA have an antagonistic relationship; favourable conditions lead to high GA and low ABA levels in seeds whereas unfavourable conditions cause the vice-versa (Figure 1.7). Thus, it is important to strike a balance between ABA and GA levels in seeds and this is best understood by studying the ABA response mutants and GA-deficient mutants. ABA-insensitive mutants like *abi1*, *abi2* and *abi3* show reduced dormancy and allow germination at ABA concentrations that inhibit wild-type germination (Finkelstein and Somerville 1990, Leung et al. 1997). On the other hand, GA-deficient mutants like *ga1-3* and *ga2* are highly dormant and fail to germinate in the absence of exogenous GA (Koornneef and Van der Veen 1980, Ogawa et al. 2003).

1.6 RGL2 is a major repressor of GA-induced seed germination

RGL2 has been proposed as the main DELLA protein that needs to be inactivated during GA-induced breaking of seed dormancy (Lee et al. 2002, Tyler et al. 2004, Cao et al. 2005). This is shown from Figure 1.6 wherein an additional knockout of *RGL2,* but not *RGA,* is able to rescue the germination defect of the GA biosynthetic mutant *ga1-3,* thereby allowing germination even in the absence of exogenous GA and thus mimicking WT germination. Cao et al. (2005) showed that the triple mutants ga1-*3 rga-t2 rgl2-1, ga1-3 gai-t6 rgl2-1* and *ga1-3 rgl1-1 rgl2-1* all germinated in the absence of exogenous GA without chilling, in contrast to *ga1-3 rgl2-1* that needed prolonged chilling to allow germination. This suggests that the DELLA proteins

RGA, GAI and *RGL1* all encode negative regulators that help to enhance the function of RGL2 in repressing seed germination.

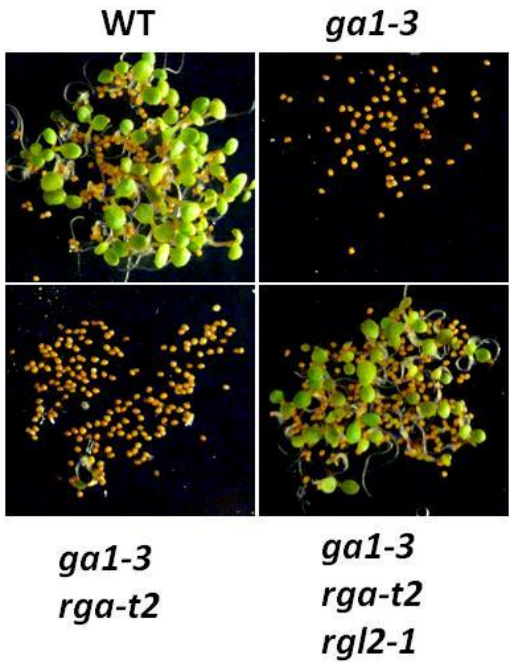


Figure 1.6: RGL2 is the major repressor of seed germination

ga1-3 seeds are gibberellin-deficient and thus are unable to germinate in the absence of exogenous GA. A knock out of RGA is unable to restore the germination phenotype. However, knockout of RGL2 is able to completely rescue the germination defect of the GA biosynthetic mutant *ga1-3,* even in the absence of exogenous GA. Thus, RGL2 is a major repressor of seed germination in the gibberellin signalling pathway.

At the transcriptional level, *RGL2* transcripts are known to rise following the onset of imbibition in both wild-type and *ga1-3* seeds. But once the seeds begin to germinate, *RGL2* transcripts are known to drop rapidly in wild-type seeds with increase in bioactive GAs, but remain at high levels in *ga1-3* non-germinating seeds. At the protein level, it is known that GA can relieve the RGL2-induced repression by degrading it via the 26S proteasome pathway. In addition to being controlled by GA, studies have shown the involvement of RGL2 in ABA signalling as well. RGL2 induces the expression of *XERICO*, which encodes a RING-H2 factor, which, in turn, elevates endogenous ABA levels and thus ABI5 activity in an unknown fashion (Ko et al. 2006, Zentella et al. 2007). Lee et al. (2010) have shown that high ABA levels in imbibed dormant seeds requires the permanent expression of the DELLA gene, RGL2. Endosperm of dormant seeds releases ABA in an RGL2-dependent manner, thus sustaining its dormancy. ABA, in turn, also enhances the *RGL2* expression. On the other hand, upon imbibition, non-dormant seeds expressed RGL2 only transiently and germinated. Therefore, RGL2 helps to integrate the GA and ABA signalling pathways with environmental cues in regulating seed dormancy.

RGL2 up-regulates *MOTHER OF FT AND TFL1 (MFT)* transcription, which encodes a phosphatidylethanolamine-binding protein, by binding to its promoter region in complex with an unknown protein partner (Xi et al. 2010). *MFT* expression is also directly repressed by ABA-INSENSITIVE3 (ABI3) and promoted by ABI5. MFT, in turn, directly represses *ABI5,* thereby providing a

negative feedback regulation of ABA signalling. Thus, MFT functions downstream of RGL2 and serves as an intersection point of ABA and GA signalling pathways during seed germination (Xi et al. 2010).

Although DELLA proteins are known to function as nuclear-localized transcription factors, they lack a DNA binding domain. Thus, it is hypothesized that their downstream function is mediated by interaction with other transcription factors (Peng et al. 1997, Silverstone et al. 1997, Hong et al. 2012). DELLA proteins were demonstrated to sequester transcription factors causing a transcriptional silencing of downstream GA responses. Therefore, the binding of these transcription factors to DELLA proteins and to their gene promoters would be mutually exclusive. An example is the binding to members of the PIF family of phytochrome interacting bHLH transcription factors and also brassinosteroid activated-BZR1 to sequester them as inactive complexes and thereby inhibit their downstream functions (de Lucas et al. 2008, Bai et al. 2012). Hou et al. (2010) reported that DELLA proteins could also function as transcriptional activators by sequestering inhibitors of transcription. The authors showed that the DELLA protein RGA interacts with JA ZIM-domain 1 (JAZ1) protein, the key repressor of jasmonic acid signalling, thus preventing JAZ1-mediated repression of transcription. More recently, Park et al. (2013) have shown that not only do DELLA proteins sequester transcription factors, but they also form protein complexes for a synergistic downstream repressive action. The authors show that BOTRYTIS SUSCEPTIBLE1 INTERACTOR (BOI) and its homologs interact with DELLAs to

form a complex that represses downstream GA responses like seed germination, juvenile-to-adult phase transition and flowering. Another study shows that DELLA proteins RGL2 and RGL3 interact with SWI/SNF chromatin remodelling complex and affect transcriptional activation of GID1 and GA3ox genes thereby controlling GA perception and biosynthesis, respectively (Sarnowska et al. 2013).

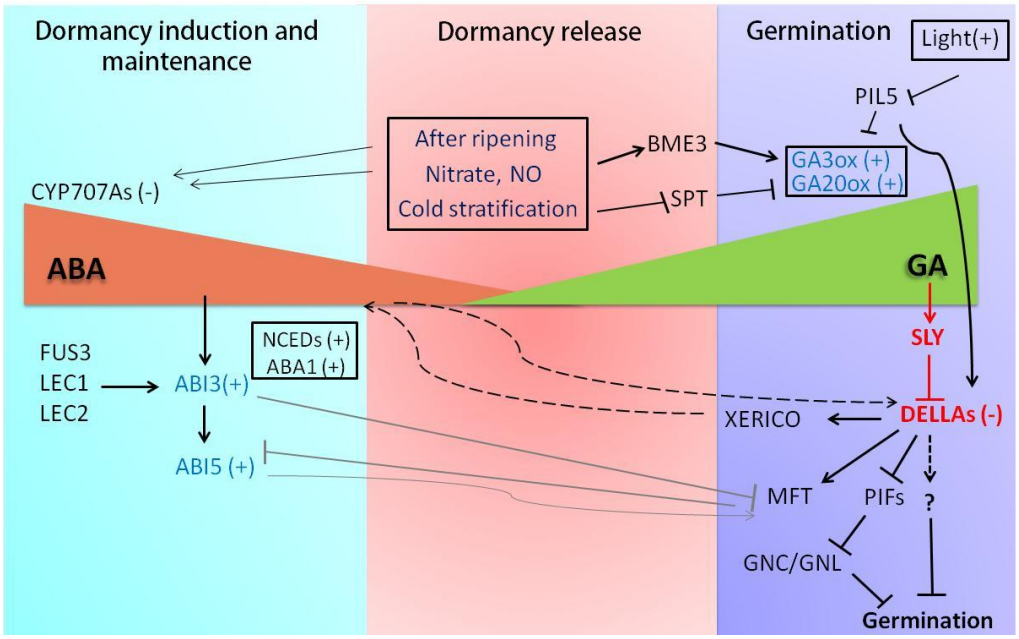


Figure 1.7: Interplay of ABA and GA signalling in the regulation of seed dormancy and germination.

The switch from dormancy to germination is controlled by the balance between ABA and GA levels. Mature seeds are highly dormant due to high levels of ABA. ABA1 and 9-cis-epoxycarotenoid dioxygenase (NCED) are key enzymes in ABA biosynthesis. Dormancy induction in embryo is genetically controlled by at least four major regulators, namely FUSCA3 (FUS3), ABSCISIC ACID INSENSITIVE 3 (ABI3), LEAFY COTYLEDON 1 (LEC1) and LEC2. GA biosynthesis and response pathways are activated in response to germination-inducing conditions like cold stratification, after-ripening, nitrate and light resulting in an increase in bioactive GAs. Cold-activated transcription factors like BME3 and SPT and light cause a rise in GA biosynthetic genes GA3ox and GA20ox. After-ripening and nitrate cause a rise in *CYP707As* which encode encodes ABA 8'-hydroxylases, key enzymes in ABA catabolism. RGL2 is the major repressor of seed germination, with minor roles played by RGA, RGL1 and RGL3. RGL2 is known to induce levels of ABA via XERICO. Its function is mediated by other transcription factors like MFT, GNC/GNL and several other unknown proteins. The GA-mediated destruction of DELLAs is thus, the most important step in the interplay of ABA and GA signalling in regulation of seed germination. Adapted from Finkelstein et al. (2008) and re-drawn based on information from several review papers (Liu et al. 2005, Penfield et al. 2005, Oh et al. 2006).

To further probe the molecular mechanism downstream of RGL2 in maintaining seed dormancy, a microarray was conducted in our laboratory by Stamm et al. (2012), wherein we identified a wide set of genes regulated by RGL2. We show that RGL2 positively regulates the expression of a number of transcription factors of MYB-, AP2/ERF-, bHLH- and HD-Zip-types, which are involved in various responses to phytohormones as well as developmental stages. On the other hand it downregulates the expression of genes encoding cell wall modifying enzymes, viz., *CP1*, *EXPA3* and *EXPA8.* Thus, it is proposed that RGL2 causes seeds to enter a state of secondary dormancy by directly arresting embryonic growth.

Furthermore, a strong enrichment of motifs associated with Dof-type transcription factors was observed in the promoters of up-regulated transcripts. Dof proteins are a family of plant-specific transcription factors in *Arabidopsis*, many of which have roles in regulation of seed germination. For example, Dof Affecting Germination1 (DAG1) and DAG2 have been shown to possess antagonistic roles in the regulation of germination, with DAG1 inhibiting germination by mediating PIL5 activity as well as directly affecting gibberellin biosynthesis (Gualberti et al. 2002, Gabriele et al. 2010). Another Dof transcription factor, Dof6, was shown to negatively regulate germination by affecting abscisic acid signalling in seeds (Rueda-Romero et al. 2012). This led to the hypothesis that either Dof transcription factors themselves or their inhibitors could be binding partners of the RGL2-complex that regulates downstream gene expression to maintain seed dormancy.

One of the transcription factors that appeared in the microarray from Stamm et al. (2012) was a GATA-type transcription factor, GATA12 (AT5G25830). The GATA gene family comprises 30 members and very few of them have been characterized. Recently, a few of them have been characterized and were found to play roles in regulating embryo development and seed germination (Liu et al. 2005, Nawy et al. 2010, Richter et al. 2010). Therefore, we decided to investigate the function of GATA12 in regulating seed germination and dormancy downstream of RGL2. This study forms the core of this thesis.

1.7 Major aims of this project

The DELLA protein, RGL2, is the major repressor of seed germination and functions downstream of GA signalling. Although a lot is known about how GA de-represses the negative effect of RGL2 on seed germination, not much is known about the mechanism downstream of RGL2. This project was initiated to understand the downstream mechanism of RGL2 in negatively regulating seed germination by studying its downstream target(s). From an earlier microarray data done in our laboratory, several downstream targets of RGL2 were identified. This project began with confirmation of the differential regulation of several selected genes from the microarray. We selected a GATA-type transcription factor for further molecular and functional characterization to understand its role downstream of RGL2. Our main objectives were:

- **Molecular characterization of GATA12**

  - Expression pattern of *GATA12*

  We analysed expression pattern of *GATA12* in various tissues of WT using quantitative real-time PCR (qRT-PCR). Histochemical GUS staining studies using *GATA12pro:GUS* lines were also done. We also analysed the levels of *GATA12* transcripts at different stages of embryogenesis.

  - Sub-cellular localization of GATA12

  Sub-cellular localization of GATA12 protein was studied by transient expression in *Arabidopsis* mesophyll protoplasts.

  - Phenotypic analysis of *GATA12* transgenic plants

  We generated transgenic lines with suppression and ectopic expression of *GATA12* and analyzed the phenotypic differences, paying particular attention to all aspects of seed dormancy and germination.

- **Study of regulation of *GATA12* by gibberellin signalling**

  - Regulation by GA

  We analysed the regulation of *GATA12* by GA at the transcriptional level. Analysis was done using GA and mock-treated gibberellin-deficient seeds.

  - Regulation by RGL2

  GATA12, being picked up from the microarray as differentially regulated, we analysed its immediate response to a functional RGL2 protein. Also, chromatin immunoprecipitation (ChIP) assay was used

to analyse if the differential regulation by the RGL2-complex was through direct binding to *GATA12* promoter.

- Promoter motif analysis

We analysed the promoter of *GATA12* and scanned for GA or RGL2-related cis-elements.

- **Preliminary identification of protein partners in RGL2 complex**

  We initiated a yeast library screening to identify proteins that interact with RGL2.

From this study, we aimed to gain a better understanding of gibberellin signalling in regulation of seed dormancy and germination. We hope that the observations can make a valuable contribution to future crop improvement strategies.

2. MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Plant Materials and Growth Conditions

Arabidopsis thaliana accessions used in this study were either in Columbia (Col-0) or Landsberg *erecta* (L*er*) background. Seeds were sown on compost soil and placed in a cold room for 4 days at 4°C to synchronize germination. Trays with cold-stratified seeds were then transferred into environmentally controlled growth chambers. Plants were grown at 23°C and 75% RH under long days (16 h of light/8 h of dark).

T-DNA insertion mutants used in this study were obtained from the *Arabidopsis* Biological Resource Centre (ABRC) seed stock. All ABRC mutant lines used here are listed in Table 2.1 Plants homozygous for the T-DNA insertion were screened by genotyping with primers designed by the T-DNA primer design tool (http://signal.salk.edu/tdnaprimers.2.html; see also Table 2.3).

All **gibberellin-related mutants** are described in Table 2.2. All other transgenic plants described here were generated by transforming wild-type plants cv. Col-0 with a binary vector harbouring the respective gene of interest.

Table 2. 1: List of *Arabidopsis thaliana* mutants obtained from Arabidopsis Biological Resource Centre (ABRC)

Locus	Gene Name	ABRC stock	Type of mutant	Background
AT5G25830.1	GATA12	SALK_052546	T-DNA insertion	Col-0
AT5G25830.1	GATA12	SALK_012051C	T-DNA insertion	Col-0
AT4G32890.1	GATA9	SALK_152156C	T-DNA insertion	Col-0

Table 2. 2: List of gibberellin-related mutant plants used in this study

Mutants	Background	Source
ga1	Col-0	SALK T-DNA
ga1-3	L*er*	SALK T-DNA
rgl2-1	L*er*	(Yu et al. 2004)
ga1-3 rga-t2	L*er*	(Yu et al. 2004)
ga1-3 rga-t2 rgl2-1	L*er*	(Yu et al. 2004)
ga1-3 rga-t2 rgl2-1 35S::RGL2-GR	L*er*	(Stamm et al. 2012)
ga1-3 rgl2-1 35S::RGL2-6XHA	L*er*	(Xi et al. 2010)

2.1.2 Bacterial Strains

Species	Strain(s)	Purpose
Escherichia coli	DH5α, MAC	cloning, propagation of plasmids
Agrobacterium tumefaciens	GV3101	*Agrobacterium*-mediated plant transformation

2.1.3 Yeast Strains

Species	Strain(s)	Purpose
Saccharomyces cerevisiae	Y2HGold	Yeast library screening
Saccharomyces cerevisiae	AH109	Yeast library screening

2.1.4 Vectors and Plasmids

Plasmid name	Purpose
pGEM®-T Easy vector system (Promega)	TA-cloning
pJET (fermentas)	blunt end cloning
pGreen binary vectors HY105 backbone (Liu *et al.*, 2007) pGreen 0229 backbone (Yu *et al.*, 2004a)	*Agrobacterium*-mediated plant transformation
pGADT7 and pGBKT7	Yeast library screening

2.1.5 Primers and Oligonucleotides

Sequence information for full length genes and promoters were obtained from The *Arabidopsis* Information Resource (TAIR) database (www.arabidopsis.org) and the Plant Promoter Database 2.1 (http://ppdb.agr.gifu-u.ac.jp), respectively. Forward primer was designed using the first 20-24 nucleotides of the sequence of interest and reverse primers were designed by reverse complementing its last 20-24 nucleotides. For genes that need to be cloned into vectors with a C-terminal epitope tag, the stop codon was omitted to allow fusion proteins to be formed. Restriction sites were added to the 5' end of both primers as per the multiple cloning sites (MCS) of their respective vectors. Primers for quantitative real time PCR (qRT-PCR) analysis were designed using the NCBI primer designing tool. All primers were synthesized by 1st Base (Singapore) to a final concentration of 100µM. For working solutions, primers were further diluted 1:10 with ddH2O (final concentration of 10µM) . The complete set of primers used for this work are listed in Table 2.3. Primers are sorted according to purpose and associated gene locus.

Table 2. 2: List of primers used during the course of this work.

Primer name	Sequence (5'→ 3')
ChIP-qRT-PCR	
GATA12_ChIP_P1_FW	GTCGGTACTGCCGATATCCAC
GATA12_ChIP_P1_RV	GGATCGAGTCACCGGTTTTGG
GATA12_ChIP_P2_FW	GTGGATCGGTTTCAACTTTCG
GATA12_ChIP_P2_RV	GGTATAGTGGTGGGGACAAC
GATA12_ChIP_P3_FW	GATGATAATCAACGGCGTGATC
GATA12_ChIP_P3_RV	GTGTAGTCACGTTTGTGGGGAG
GATA12_ChIP_P4_FW	CTCCCCACAAACGTGACTACAC
GATA12_ChIP_P4_RV	CCGGTAAACTCTCGTCGGAGGT
GATA12_ChIP_P5_FW	GACCTCTTGTCCCCTTCAAG
GATA12_ChIP_P5_RV	CCGGTAAACTCTCGTCGGAGGT

TUB2_ChIP_FW	ATCCGTGAAGAGTACCCAGAT
TUB2_ChIP_RV	AAGAACCATGCACTCATCAGC
ACTIN_ChIP_FW	CGTTTCGCTTTCCTTAGTGTTAGCT
ACTIN_ChIP_RV	AGCGAACGGATCTAGAGACTCACCTTG
MFT-2_ChIP_FW	ATGATCCTCCTAACCGACGGC
MFT-2_ChIP_RV	AACCTATAGGCCATGCATAGG
Vector-specific	
pGEMT_insitu_ FW	CTTAATGCGCCGCTAC
pGEMT_insitu_ RV	GAAGAGCGCCCAATACG
pGREEN_35S_ FW	GACCCTTCCTCTATATAAGGAAGTTC
pGREEN_terminator_ RV	CCTTATCGGGAAACTACTCACAC
pGREEN_HAtag_RV	AGCGTAATCTGGAACATCGTATGGGTA
pJET1.2_FW	CGACTCACTATAGGGAGAGCGGC
pJET1.2_RV	AAGAACATCGATTTTCCATGGCAG
GUS_RV	CTGCCAGTTCAGTTCGTTGT
qRT-PCR	
GATA12_qRT_FW	CCACCACTACCACCACCATAACCG
GATA12_qRT_RV	CACCGGAGAAGCTAGTGCCGTC
GATA9_qRT_FW	TATGGACGTGGATTGCGGAG
GATA9_qRT_RV	TACAAAGCGTCTTCGGACCC
GATA2_qRT_FW	GTGGAGGACAGGACCACTTG
GATA2_qRT_RV	AGGACTCGAAGCCGGTCTAT
GATA4_qRT_FW	AACTTGGGCTCCGATGTCTG
GATA4_qRT_RV	ATCCGCCGTAACCGATTCAG
GA3ox1_qRT_FW	ACCGTGCCACCGTTTCCTGG
GA3ox1_qRT_RV	TAACCCGAGCGCGGTGCAAC
GA20ox3_qRT_FW	TCGTATGGCCCGACCACGAGA
GA20ox3_qRT_RV	ATGCCAAGCACGAGTCGCCG
TUB2_qRT_FW	CCAGCTTTGGTGATTTGAAC
TUB2_qRT_RV	GGAATGGGATGAGATTCACAG
microarray confirmation	
AtEXPA1_qRT_FW	CGCCGGAGGAGGTTGGGTCA
AtEXPA1_qRT_RV	TTAGCGCCGCCGTGTTGGTT
AtEXPA3_qRT_FW	GTCGGCTTGTGGTTGGCCGT
AtEXPA3_qRT_RV	GCCGCCGTGTTCACACCGTA
AtEXPA8_qRT_FW	TGCCCACCTAACCCTGGCCT
AtEXPA8_qRT_RV	TGCCGGCACGATACTGAGCG
AtEXPA9_qRT_FW	TGCCTTTACGGCCAACGCCA
AtEXPA9_qRT_RV	TCCCGTACCCACACGCACCA
AtEXPA11_qRT_FW	TGGCGGTTTTGGCCGCTCTT
AtEXPA11_qRT_RV	ATTGTCCCGTACCCCGCCGA
AtEXPA14_qRT_FW	CGGCGTCGTCCCTGTCCAAT
AtEXPA14_qRT_RV	GCGCCGGCGACGTTGGTTAT
AtEXPA15_qRT_FW	TTACGGGACCAACACGGCGG
AtEXPA15_qRT_RV	GCGCCGTCGCTCTGACACTT
CYP707A2_qRT_FW	TCGGGGACAAAGAGGAGCCCA
CYP707A2_qRT_RV	GCCCGGTAGGTCGAGAGGCA
CYP707A3_qRT_FW	TCCGCCGTAGCTCCTCCACG
CYP707A3_qRT_RV	CGCTGCTTCAGGGCTCGAGAT
CYP716A1_qRT_FW	ACCTCTTTGGGTCTCCCACCGC
CYP716A1_qRT_RV	TCTGGCCACCACGAGACCACA
CYP81F4_qRT_FW	GTGGGCGATGGCGAGTTTGTTGA
CYP81F4_qRT_RV	GAACCGGCGCTGCTGGATGT
COR47_qRT_FW	TGGTGGAGCATGACCATCCCGA

COR47_qRT_RV	CCAACGGCGTGGACGTGACA
PIF4_qRT_FW	CACATGCAGGCCGCGACTCA
PIF4_qRT_RV	CGGTGGTCTTCGTCGGCACA
bHLH_AT1G74500_qRT_FW	GGGACAGTCGTCGTTCCGACA
bHLH_AT1G74500_qRT_RV	AGCAGCTTGTGCAGTGTCTGAGT
Dof_AT5G66940_qRT_FW	TCTGCCGCTACCACCTCCGT
Dof_AT5G66940_qRT_RV	GCACTTCCCTTCGCCGTCGT
C2H2_AT5G03510_qRT_FW	CCAAGCGCTAGGAGGGCACA
C2H2_AT5G03510_qRT_RV	CCGTAGACGCAGACGCCGTT
C2H2_AT2G18490_qRT_FW	TGCGAGATTTTAAACCCTGAGCCTCT
C2H2_ AT2G18490_qRT_RV	GGGGATTTTTGTCGAAGATCGGCCT
C3HC4_AT2G01275_qRT_FW	TTGCAGCGGCAGCGTCAAGT
C3HC4_ AT2G01275_qRT_RV	CTCTAGCAACGGAGGCGGCG
C3HC4_AT4G13100_qRT_FW	TGAACTCAGCGCGCGAGACG
C3HC4_ AT4G13100_qRT_RV	CAACCCAACTCCGCTGCCCC
C3HC4_AT4G30370_qRT_FW	TCTCCGCCGCAAGTCGATGT
C3HC4_ AT4G30370_qRT_RV	GTGTTCGCTTTTGGCGGCGG
GATA_AT3G51080_qRT_FW	GGCTCGCCAGCGGTCAGTTT
GATA_AT3G51080_qRT_RV	GCGTCTGCGTCTGCGTTTGC
Dof_AT2G28510_qRT_FW	CGGTGGTGGTTGCCGTCGAA
Dof_AT2G28510_qRT_RV	TGGAGAGTAGCGAAAGCGAGGTTG
MIF1_AT1G74660_qRT_FW	GCTAACATCGGAGGCTACGCCG
MIF1_AT1G74660_qRT_RV	AACCACAAGCCGCGCATCTCA
Genotyping	
GATA12_ SALK _012501C _LP	TAAAGGTGGTGTCGGTACTGC
GATA12_ SALK _012501C_ RP	AGTGGTGGAATCAGCAACAAC
GATA12_ SALK _052546C_LP	GTTTTAAGAGTCGACCGGACC
GATA12_ SALK _052546C _RP	CCAACACTGAGACACTCTCCC
GATA9_ SALK _152156C _LP	CCCCTTGTTGTTAAGCAAATG
GATA9_ SALK _152156C _RP	CAAGCGTTACAAAGCGTCTTC
Cloning and Sequencing	
gene cloning	
GATA12_HindIII_FW	*AAGCTT*ATGGAAGATGAAGCACATGA
GATA12_SpeI_no terminator_RV	TAAT*ACTAGT*AATAAGCTGTCTGAAATC
GATA12_internal_FW	CATATGACCGGAAGCCCGGAAAACC
promoter cloning	
GATA12pro_HindIII_FW	*AAGCTT*TGATCAAATCTTTGATACGTTG
GATA12pro_SmaI_FW	*CCCGGG*AAGTTTCGTTGATTAAAACT
antisense construct and insitu	
GATA12_AS_insitu_SpeI_FW	*ACTAGT*ACCACCATAACCGACAG
GATA12_AS_insitu_HindIII_RV	*AAGCTT*AGACAAGTGTTGTTGG

2.2 PLASMID CONSTRUCTION

2.2.1 RNA extraction and cDNA synthesis

Before extraction, 1.5 and 2ml eppendorf tubes, pestles and pipette tips were autoclaved at 121°C for 1h and then kept in 60°C for drying. DEPC-treated

water was prepared by treating ultrapure water with 0.1% diethylpyrocarbonate (DEPC) at 37°C for 2 hours and then autoclaved. All buffers were prepared using DEPC-treated water.

For total RNA from leaves or seedlings, extraction was done using TRIzol® (Invitrogen) following the manufacturer's recommendations. For the extraction of total RNA from seeds (dry or imbibed) or siliques, we followed the protocol from Vicient and Delseny (1999), with many modifications. 20-50mg of seeds (dry weight), or 4 to 6 siliques were ground in liquid nitrogen to a very fine powder and transferred immediately into 0.5ml of pre-cooled precipitation buffer (8M lithium chloride + 2% β-mercaptoethanol). Precipitation was allowed to occur overnight at 4°C. The next day, samples were centrifuged at 14,000 rpm at 4°C for 20min. Following centrifugation, the crude RNA pellet was washed with 1ml ice-cold 75% ethanol (prepared in DEPC-treated water) and centrifuged at 14,000 rpm at 4°C for 5min. The pellet was then resuspended thoroughly in 0.5ml RNA solubilisation buffer (10mM Tris/HCl pH 7.6, 25mM EDTA, 100mM NaCl, 0.5% SDS, 2% β-mercaptoethanol). 0.5ml cholorofom was then added to this resuspension and shaken vigorously and centrifuged at 14,000 rpm at 4°C for 15min. The upper aqueous phase was then extracted again with an equal volume of TRIzol® (Invitrogen) and then precipitated in 2.5 times (v/v) ethanol and 1/10 (v/v) 3M sodium acetate (pH 5.2) in -20°C for 1.5h. Following centrifugation, the RNA pellet was dissolved in 80µl DEPC-treated water, precipitated again with 1x volume high-salt buffer (0.8M sodium citrate, 1.2M NaCl) and 1x

volume isopropanol for 30min at -20°C. The RNA pellet was washed with 1ml ice-cold 75% ethanol, air dried and dissolved in 30 to 50µl DEPC-treated water.

RNA concentration and quality was checked using a micro-volume spectrophotometer and a minimum of 500-600ng of RNA for each sample was used for cDNA synthesis. The cDNA was synthesized using MAXIMA® First Strand cDNA Synthesis kit (Thermo SCIENTIFIC, USA) as per manufacturer's protocol in a total reaction volume of 20µl. For PCR applications, cDNA was finally diluted to a final volume of 100 µl (1:5).

2.2.2 PCR Amplification

Coding regions or promoters of the gene of interest were amplified using cDNA or gDNA as template and their respective primers (refer Table .2). A 20µl PCR reaction was set up with 1µl template, 0.5µl of each primer (10µM), 0.5µl dNTP mix (10mM), 2µl 10x KAPATaq buffer A, 0.2µl KAPATaq DNA Polymerase (5U/µl) (KAPA Biosystems) and sterile dH_2O. PCR was performed with an initial denaturation at 95°C for 5min, and 27 cycles of denaturation at 95°C for 45s, annealing for 25s, and extension at 72°C, followed by final extension at 72°C for 5min. Annealing temperature and extension time were adjusted according to the melting temperature (Tm) of primers used (typically 55 to 60°C) and length of the fragment to be amplified (approximately 1min per 1kb fragment length).

The PCR product was mixed with 6X DNA loading dye and loaded onto 1% TAE agarose gel containing nucleic acid fluorescent dye (dil. 1:20,000). The

gel was run in 1X TAE (4.84g/l Tris Base, 50mM EDTA pH 8.0, 0.11% (v/v) acetic acid) tank buffer and viewed/imaged under ultraviolet light using a gel documentation system.

2.2.3 Gel extraction and ligation with vector

The PCR band of expected size was excised with a scalpel blade and DNA was extracted from it using Wizard® SV Gel and PCR Clean-Up System (Promega). The gel-extracted DNA was ligated into pGEM®-T Easy Vector System (Promega) according to the manufacturer's protocol.

2.2.4 Transformation, confirmation and plasmid extraction

The ligated mixture was used to transform *E.coli* chemically-competent cells that were prepared as per the protocol mentioned in chapter 2.2.7 and then plated on selective medium.

The antibiotic-resistant colonies were subjected to colony-PCR to confirm the presence of the desired plasmid. For this, the bacterial colony was resuspended in 10ul of dH2O and 2 ul of this was used as the template for the PCR reaction. PCR components and conditions are the same as that mentioned in chapter 2.2.2. PCR products were size-separated on 1% agarose gels, and positive colonies were detected based on the presence of the band of expected size.

Positive colonies were inoculated in 4ml liquid LB medium containing the appropriate antibiotic and incubated overnight at 37°C in a shaker incubator. Plasmid extraction was then carried out using the Wizard® Plus SV Minipreps

DNA Purification System (Promega) according to the manufacturer's instructions. Plasmids were eluted in 30µl TE buffer, and concentration was determined using NanoDrop 2000c (Thermo Scientific).

2.2.5 Sequence analysis

The generated constructs were sequenced to make sure there are no mutations in the gene and it is in reading frame. Sequencing reactions were set up using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), with 1µl template plasmid, 0.5µl primer (10µM), 2µl BigDye reaction mix, and sterile dH2O to a total volume of 5µl. PCR was performed with an initial denaturation at 96°C for 1min, followed by 25 cycles of denaturation at 96°C for 10s, annealing at 50°C for 5s, and extension at 60°C for 4min. DNA was precipitated by adding 3µl sodium acetate (3M, pH5.2), 69.4µl ethanol (100%) and 17.6µl dH2O to the reaction. After 30min incubation on ice, samples were centrifuged at 15,400*g* at 4°C for 20min, the supernatant was removed and the pellet was washed once with 250µl ice-cold 75% ethanol. The air-dried pellet was submitted for sequencing analysis to the department's DNA sequencing laboratory (DSL) which uses the ABI3130xl sequencer (Applied Biosystems). The output data was analysed using Chromas LITE (v2.01) and NPS@MULTIALIN available at the Pole Bioinformatique Lyonnais (PBIL) Network Protein Sequence Analysis. (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_multalinan.html).

2.2.6 Sub cloning into respective vectors

Digestion and Ligation: The sequenced inserts in the MCS of pGEM®-T Easy were digested with their respective restriction sites and ligated into either pGREEN binary vector or pGBKT7. Digestion of both, the clones in TA vector and the subsequent empty vectors, was carried out using FastDigest® restriction enzymes and buffer (Fermentas) as follows: reactions containing approximately 1µg plasmid, 1µl of each respective restriction enzyme, 3µl 10x FastDigest® Green buffer, and dH2O to a total volume of 30µl were mixed and incubated at 37°C for 15 to 30min. The digested reaction was directly run on 1% TAE agarose gel, required fragments were excised and DNA was gel extracted using Wizard® SV Gel and PCR Clean-Up System (Promega). The concentration of the eluted products was determined and ligation reaction was setup in 3:1 molar ratio of insert : vector. Rapid ligation was done using Rapid DNA Ligation kit (Thermo SCIENTIFIC), according to which 2 µl of 5X ligation buffer, 1 µl of T4 DNA Ligase and calculated amounts of vector and insert were added along with dH2O to a final volume of 10 µl and incubated at room temperature for 30min.

Transformation & confirmation of clones: Transformation of *E.coli* with the ligated product and confirmation of the new construct using colony PCR were carried out as mentioned in chapter 2.2.4. The construct was sequenced as in chapter 2.2.5 to double confirm the clone.

2.2.7 Preparation of *E.coli* Competent Cells

To prepare chemically competent *E. coli* cells, 200ml LB broth was inoculated with 5ml of overnight culture and incubated at 37°C in a shaker-incubator set to 200 rpm, until an optical density (OD_{595}) of 0.4 to 0.5 was reached. The culture was poured into four 50ml falcon tubes and cells were harvested by centrifugation at 5000 rpm for 10min at 4°C. They were resuspended in 25ml (1/4 vol.) ice cold 100mM MgCl2 and again pelleted at 5000rpm for 10min at 4°C. The cell pellet was then resuspended in 50ml (1/2 vol.) ice cold 100mM CaCl2, and incubated on ice for 30min. Cells were again harvested by centrifugation at 5000rpm for 10min at 4°C, and finally resuspended in 1ml ice cold 100mM CaCl2 containing 15% glycerol. Aliquots of this cell suspension (typically 100µl) were snap-frozen in liquid nitrogen, and stored at -80°C for future use.

2.3 Plant Transformation

2.3.1 Preparation of *A. tumefaciens* Competent Cells

The *Agrobacteria tumefaciens* GV3101 strain was used in our laboratory for making electroporation-competent cells. *A. tumefaciens* was streaked on LB agar plate with 25µg/ml of rifampicin and gentamycin and allowed to grow at $28^{O}C$ (1-2 days). A single colony was innoculated in 10mL LB media containing selection antibiotics and grown at $28^{O}C$ in shaker-incubator for 2 days. 100ml LB media was innoculated with the 10ml culture and cells were allowed to grow at $28^{O}C$ (200 rpm) until an OD_{595} of 1-1.5 (~4-5h). The culture was transferred to two cold falcon tubes (50mL) and placed on ice for 20min. Cells

were then harvested at 4^{o}C for 15min at 4,000xg. The pellet was then gently resupended in 30mL of cold sterile water. Cells were again harvested at 4^{o}C for 15min at 4,000 xg. This wash step were repeated 3 more times. The cells were finally re-suspended again in 4ml (per falcon tube) of fresh ice-cold sterile H2O and placed on ice for 10min. 100 µl of cell suspension was aliquoted into a pre-chilled 1.5ml tubes and snap-frozen in liquid nitrogen. The tubes were then kept at -80°C for long term storage up to several months without dramatic decrease of competency.

2.3.2 Plasmid Transformation of *A. tumefaciens* Competent Cells

A tube of frozen GV3101 competent cells was thawed on ice and 0.5-1 µl of purified plasmid was pipetted to the tube and mixed by gentle tapping of the tube. The mixture was then transferred into a 1 mm Gene Pulser® cuvette (Bio-Rad) and subject to electroporation at 25 µF, 2.5 kV, 200Ω using a MicroPulser electrporator (Bio-Rad). The electroporated bacteria was immediately placed in ice and then cultured in 1ml of LB medium for 3h with shaking at 28°C. 15-20 µl of the culture was spread evenly onto an LB agar plate containing 25µg/ml gentamycin, 10 µg/ml tetracycline and 25 µg/ml rifampicin for the selection of the GV3101 strain and a specific antibiotic for the selection of the transfected plasmid. For pGreen-based plasmids, 50 µg/ml kanamycin was used. The plate was incubated at 28°C for 2-3 days. The colonies were verified by PCR and the confirmed colonies with the transgene were grown in 4ml LB liquid media containing the four selection antibiotics and used for subsequent floral dip.

2.3.3. Floral Dip and Selection of Transgenic Plants

Agrobacterium-mediated floral dip method was established previously (Clough and Bent, 1998). GV3101 cells transformed by a desired construct were cultured at 28°C for 2 days in 4ml of LB liquid medium containing the desired selection antibiotics. This culture was then used to innoculate a 200ml LB liqiud media with selection antibiotics and incubated at 28°C for 2 days with shaking at 200 rpm. After 2 days, the cells were pelleted at 4000 rpm for 15min at room temperature. The cell pellet was re-suspended completely in a solution containing 0.25 g/l MES (pH 5.7), 5% sucrose and 0.02% surfactant Silwet L-77. Fully opened flowers and siliques of healthy *Arabidopsis* WT Col-0 plants were removed and all flower buds were submerged in this *Agrobacterium* cell suspension for 1min. Dipped plants were incubated overnight in the dark (covered with black plastic) under high humidity before being placed back into normal growth conditions. Seeds from the dipped plants were later collected as T1 generation. All seeds obtained were sown on soil and stratified at 4°C in cold room for 4-5 days before transferring to the growth chamber .The plants were screened by 3% BASTA® (glufosinate-ammonium) at the four-leaf stage for the selection of transgenics.

2.3.4. Genotyping of selected transgenic plants

BASTA application selected for successful integration of the pGREEN construct with the plant genome. However, the survivors had to be further confirmed for the presence of the transgene construct using a PCR-based

genotyping method. A PCR was done with primers specific to the vector backbone using genomic DNA extracted from each survivor plant as the template. Plants that gave a PCR band of the right size were selected and rest discarded.

Genomic DNA extraction: Genomic DNA (gDNA) was extracted using the protocol given by *Dellaporta et al.* (1983) with slight modifications. A tiny leaf sample was directly crushed in 300µl extraction buffer [100 mM Tris (pH 8.0), 50 mM EDTA (pH 8.0), 500 mM NaCl and 1.5% β-ME freshly added] with no use of liquid nitrogen. The sample was then incubated at 60°C for 10min with vortexing in between. To this, 125 µl of potassium acetate (KOAc) was added, mixed by vortexing and incubated on ice for 5-10min. The samples were centrifuged at 14000 rpm for 10min at room temperature and the supernatant was transferred to a fresh eppendorf containing an equal volume of chloroform: isoamyl alcohol (24:1). The tubes were shaken vigorously and then centrifuged at 14000 rpm for 10min. The upper aqueous phase was transferred into 800 µl of 100% ethanol and again incubated on ice for 30min. The samples were again centrifuged at 14000 rpm for 10min at room temperature. The DNA pellets obtained were air dried and resuspended in 50 µl water.

The gDNA extracted was then used as template and a PCR set up with appropriate primers to check for presence of transgene. We used GoTaq® DNA Polymerase (PROMEGA) for the genotyping PCR. The positive plants

were retained and their seeds were collected to be sown into the next generation.

2.3.5. Quantitative real-time PCR analysis

Comparative analysis of relative gene expression was performed by qRT-PCR. Transformed plants were checked for significant suppression or ectopic expression of the transgene. Reactions were performed on cDNA, prepared from RNA of various *Arabidopsis* tissues. The cDNA prepared was diluted 5X before use. Reaction mixtures were prepared using 'KAPA SYBR® FAST qPCR Kit' (KAPA Biosystems), containing 1µl template cDNA, 5µl 2x KAPA SYBR® Master Mix, 0.2 µl of each primer (10 µM) and sterile ddH2O to a final volume of 10µl. Reactions were performed in the 'StepOne™ Real-Time PCR Systems' and analysed using the 'StepOne™ Software' (v2.1; Applied Biosystems). The run setup was as follows: denaturation at 95°C for 2s, followed by 40 cycles of denaturation at 95°C for 3s and annealing/extension at 60°C for 30s. The amplification of *TUB2* gene was used as the internal control to normalize data. Primers used for qRT-PCR were designed using the NCBI primer BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and are listed in Table 2.3.

2.4 Chromatin Immunoprecipitation (ChIP) assay

2.4.1 Sample Fixation

Chromatin Immunoprecipitation (ChIP) was performed using *ga1-3 rgl2-1 35S::RGL2-6HA* seeds that were cold stratified for 2 days. WT seeds were used as negative control since they lack a HA-tagged RGL2 protein. ChIP was

performed according to *Kaufmann et al*. (2010), with minor changes. We included an additional step of protein-protein cross-linking using 2mM disuccinimidyl glutarate (DSG; Nowak *et al.*, 2005), prior to cross-linking of DNA-protein complexes by formaldehyde. All centrifugation steps were performed at maximum speed of 15,400*g.*

100mg seeds that were cold-stratified for 2 days were used as starting material. Samples were submerged and vacuum-infiltrated in ice-cold MC buffer (10mM potassium phosphate pH 7.0, 50mM NaCl, 100mM sucrose) + 2mM DSG for 30min. This was followed by cross-linking in ice-cold MC buffer + 1% formaldehyde and tissues were vacuum-infiltrated for 45min. Fixation was stopped by adding glycine to a final concentration of 150mM, and a brief vacuum-infiltration. Samples were then washed three times with fresh ice-cold MC buffer, and snap-frozen in liquid nitrogen.

2.4.2 Homogenization and sonication

To isolate nuclear protein complexes, samples were ground to a fine powder in liquid nitrogen using a mortar and pestle and immediately transferred into 250µl of freshly prepared ice-cold M1 buffer (10mM sodium phosphate pH 7.0, 100mM NaCl, 10mM β-mercaptoethanol, 1M hexylene glycol, complete protease inhibitor cocktail tablet (Roche)). After centrifugation at 15,400*g* at 4°C for 5min, the pellet was washed four to five times by thoroughly resuspending in an equal volume ice-cold M2 buffer (10mM sodium phosphate pH 7.0, 100mM NaCl, 10mM β-mercaptoethanol, 1M hexylene glycol, 10mM MgCl2, 0.5% Triton X-100, complete protease inhibitor cocktail

tablet (Roche), centrifuging at 14,000 rpm at 4°C for 5min each time. Pellets were then washed twice with M3 buffer (10mM sodium phosphate pH 7.0, 100mM NaCl, 10mM β-mercaptoethanol, complete protease inhibitor cocktail tablet (Roche)) as above. Nuclear pellets were finally resuspended in 500µl sonication buffer (10mM sodium phosphate pH 7.0, 100mM NaCl, 10mM EDTA pH 8.0, 0.5% sarkosyl), and chromatin was sheared by sonication on ice for 7 pulses of 10s each, at 10 to 14 micron amplitude (SONIPREP 150). After centrifugation at 15,400*g* at 4°C for 5min, the supernatant was transferred into a new tube, and the pellet was resuspended in another 250µl sonication buffer. This was again centrifuged for 5min as before, and the supernatant was combined with the previous. From this, 75µl was saved as 'INPUT DNA'.

2.4.3 Immunoprecipitation and DNA purification

To pull down protein-DNA complexes, an equal volume of IP buffer (50mM HEPES, 150mM KCl, 5mM MgCl2, 10µM ZnSO4, 1% Triton X-100, 0.05% SDS) was added, and samples were pre-cleared by two rounds of centrifugation at 15,400*g* at 4°C for 5min, transferring the supernatant into fresh tubes each time, followed by 90min incubation with 40µl protein A-agarose (50% slurry conditioned in IP buffer; Sigma-Aldrich) on ice on a rocker. Agarose beads were removed by centrifugation at 3,800*g* at 4°C for 5min, followed by centrifugation of the supernatant at 14,000 rpm at 4°C for 5min. The supernatant was equally distributed into two tubes, and 1µg of antibody was added to each. After 1h incubation on ice on a rocker, samples were

centrifuged at 15,400*g* at 4°C for 5min, and the supernatant was transferred into a new tube containing 20µl protein A-agarose. After 50min incubation on ice on a rocker, samples were centrifuged at3,800*g* at 4°C for 5min, and the supernatant was kept as 'POST BINDING'. Agarose beads were washed five times, thrice with 1ml IP buffer, once with IP buffer + 0.35M NaCl, and once with 1ml TE buffer (10mM Tris/HCl pH 7.5, 1mM EDTA pH 8.0), incubating at RT on a rocker for 3min before centrifuging at 3,800*g* at RT for 2min each time.

To elute, 200µl elution buffer (50mM Tris/HCl pH8.0, 10mM EDTA pH 8.0, 1% SDS) were added to the agarose beads, and samples were incubated at 65°C for 15min. After centrifugation at 15,400g at RT for 2min, supernatant was collected, and the elution step was repeated thrice with 100µl elution buffer. The eluate (total volume of 500µl) was centrifuged at 15,400g at RT for 2min, and the supernatant was transferred into a new tube. 50µl were kept aside as 'ELUATE'.

To reverse cross-link, 11.25µl of proteinase K (20mg/ml; Sigma-Aldrich) was added, and samples were incubated at 37°C over night. A second aliquot of proteinase K was added, and incubation was continued at 65°C for 6h. Chromatin was finally purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. Bound chromatin was eluted in 35µl MilliQ H_2O. Purification of chromatin was later modified (to get better yield of DNA) to extraction with equal volume phenol: cholorofom (1:1) and precipitation of the upper aqueous phase with 1/10

volume of 3M sodium acetate (pH 5.2) and 2 volumes ethanol. The precipitated pellet was then washed in 75% ethanol and dissolved in 35µl MilliQ H_2O.

2.4.4 ChIP qRT-PCR

qRT-PCR analysis was done to check for promoter fragment enrichment in the final eluted chromatin from the ChIP experiment. Reaction mixtures were prepared using 'KAPA SYBR® FAST qPCR Kit' (KAPA Biosystems), containing 1µl template gDNA (chromatin), 5µl 2x KAPA SYBR® Master Mix, 0.2 µl of each primer (10 µM) and sterile ddH2O to a final volume of 10µl. Reactions were performed in the 'StepOne™ Real-Time PCR Systems' and analysed using the 'StepOne™ Software' (v2.1; Applied Biosystems). The PCR conditions were modified from the default as follows: denaturation at 95°C for 2 mins, followed by 42 cycles of denaturation at 95°C for 4s and annealing/extension at 60°C for 30s. The amplification of *ACTIN* gene was used as the internal control to normalize data. Enrichment of promoter fragments over that of *TUB2* fragment was calculated and a difference of 2.5 fold and above was considered as significant. Primers used for ChIP qRT-PCR were designed using the NCBI primer BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and are listed in Table 2.3.

2.5 Western Blotting

2.5.1 Whole protein extraction

For extraction of whole protein, samples were weighed (about 6-8 seedlings, 10d old), pulverised with a pestle in an eppendorf tube and dissolved in 5

times (v/v) Extraction Buffer (EB). The samples were then centrifuged for 20min at top speed at 4°C. The supernatant was transferred to a fresh tube and centrifuged again to remove the remaining debris. Again the supernatant was collected in a fresh tube. Samples can be stored in -80°C until further use.

Composition of whole protein EB (prepared fresh):

100mM Tris-Hcl pH 7.0

5mM EDTA pH 8.0

5mM EGTA pH 8.0

150mM NaCl

10mM DTT

Complete EDTA-free protease inhibitor tablet (Roche)

0.5% Triton-X

2.5.2 Immunoblotting

For the detection of RGL2-6XHA and GATA12-6XHA, whole protein extracts were resolved under reducing conditions on a 12% SDS/polyacrylamide gel. Proteins were transferred onto polyvinylidene difluoride (PVDF) membranes (Bio-Rad Laboratories) by tank-blotting at 100V for 1.5h. PVDF membranes were blocked by incubation with 5% BSA in Phosphate-buffered saline + 0.1% Tween 20 (PBS-T) at room temperature with mild shaking for 2h, followed by incubation with monoclonal mouse anti-HA antibody (1:2,000; Santa Cruz Biotechnology, Inc) overnight at 4°C, with mild shaking. After washing with PBS-T thrice, membranes were incubated with secondary antibody, horseradish peroxidase (HRP)-conjugated goat anti-mouse antibody (1:10,000; Santa Cruz Biotechnology, Inc) for 1.5h at room temperature, with

mild shaking. Membranes were then washed with PBS-T for 2h with frequent buffer changes, and then immune complexes were detected on X-ray film (Fuji medical x-ray film) using the SuperSignal West Femto Chemiluminescent Substrate according to the manufacturer's instructions (Thermo Scientific).

2.6 Isolation and Transfection of *Arabidopsis* Mesophyll Protoplasts

Mesophyll protoplasts were isolated from leaves of 3- to 4-week-old wild-type *Arabidopsis* (Col-0) plants and transfected according to the protocol described in Yoo et al. (2007), with minor modifications. 200 x g was used for 3min for pelleting protoplasts instead of the recommended speed of 100 x g for 1-2min. Similarly, the transfection time was increased to 20 to 30min. For each transfection, 8 to 15µg of plasmid DNA was used. PEG-calcium transfection solution used was as follows: 25% PEG, 0.4M mannitol, 150mM CaCl2. The transfected protoplasts were incubated overnight (12-16h) in 1ml MMG buffer in the dark instead of the recommended WI buffer.

Protoplasts were observed the next day using a Zeiss LSM 510 META laser scanning confocal microscope (ZEN 2009 software) using EC Plan-Neofluar 40×/0.75 in multi-track channel mode. Excitation wavelengths and emission filters were 488 nm/band-pass 505-530 nm for GFP, 405/ band pass 425-480 nm for DAPI and 488 nm/band-pass 650-710 nm for chloroplast auto-fluorescence. Image analysis was done using LSM image browser 4.2 and converted to JPEG format.

2.7 Physiological Methods

2.7.1 Seed mucilage staining

To stain acidic polysaccharides in *Arabidopsis* mucilage, seeds were imbibed for 5min in water and then in 0.03% (w/v) Ruthenium Red (Sigma-Aldrich) for 15min, without shaking. Staining solution was carefully removed, and seeds were rinsed once with water. The seeds were then transferred to a glass slide with a cover slip on them and imaged using a stereomicroscope.

2.7.2 Seed Sterilization

For assays on MS medium, *Arabidopsis* seeds were sterilised as follows: Seeds were imbibed in MilliQ H_2O for 5min (in 2ml eppendorf tubes) and then the water was pipetted out. This was followed by a 70% (v/v) ethanol wash for 2min and 10% (v/v) bleach treatment for 5min. The seeds were then washed in water four to five times to remover residual traces of ethanol and bleach.

2.7.3 Germination Assays

Germination assays were done on 1X MS medium containing 1% sucrose and 0.5% gelrite. The sterilised seeds were placed on the MS medium containing plates under sterile conditions. For cold-stratification treatments, the plates were covered with aluminium foil and placed in 4°C for three days and then transferred to tissue culture room set at 24°C with 16h light/ 8h dark cycle. For assays without cold-stratification, plates were directly placed in the tissue culture room after sowing**.** All assays were done with freshly harvested seeds that were stored for 5 days with silica beads.

2.7.4 Histochemical GUS staining

For GUS staining assays, seeds were imbibed for 1 hr (dry seed) or 24 h (imbibed seed). The testa and embryo were then separated under the stereomicroscope using a thin forceps and scalpel blade. The separated testa/endosperm and embryo were then immersed into the GUS staining solution (0.1M sodium phosphate buffer pH 7.0, 10mM EDTA, 0.1% Triton-X, 2mM 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc)). Other plants parts were directly immersed in the staining solution. The tissues were vacuum-infiltrated for 5min and then incubated overnight in the dark at 37°C without shaking.

2.8 Quantification of Hormones

2.8.1 Extraction of multiple hormones from seeds and siliques

Fresh samples were weighed (20-50 mg) and snap frozen in liquid nitrogen before extraction to calculate hormone levels per FW. Extraction was done with 5 times volume of 80% ice-cold HPLC-grade methanol for 30min at 4°C, sonicated for 10min and then centrifuged at 4°C for 20min at 14,000 rpm. The supernatants were collected and centrifuged again to remove any remaining debris.

2.8.2 LC-ESI-MS/MS analysis

The supernatant were completely dried with SpeedVac after extraction and resuspended in 80 µl of 5mM Ammonium formate containing 0.1 % (v/v) formic acid (HPLC solvent A) to inject into the LC-ESI-MS/MS system. Quantification of GA1, GA3, GA4 and ABA by LC-ESI-MS/MS was performed as

described in Mitsunori Seo *et al*. (2011).The data obtained are from three biological replicates with three technical replicates in each to ensure reliability. Conditions and parameters for the Agilent G4790A equipped with a ZORBAX Eclipse Plus C18 column (1.8 µm, 2.1 × 50 mm) for analysis of the metabolites are summarized below.

Compound	Retention time on LC(min)	MRM transition monitored for quantification (m/z)	Collision energy(V)	Fragmentor voltage (V)	Mode of Ionization (ESI)
GA1	2.79	347.4/229	35	380	Negative
GA3	2.75	345.1/143.2	40	380	Negative
GA4	4.94	331.4/213	26	380	Negative
ABA	3.75	263/153.1	10	380	Negative

Solvent A (Aqueous)	Solvent B (Organic)	Gradient (composition of solventB)
Water with 0.1% acetic acid	Acetonitrile with 0.1% acetic acid	5-50% over 5min

2.8.3 LC conditions

Flow rate: 0.4 ml/min

Time (min)	0	0.5	5.5	5.6	7.6	7.7	12
% B	5	5	50	95	95	5	5

Column temperature: 35 °C

2.8.4 MS/MS conditions

Instrument model: G4790A; Firmware version: A.00.06.39; MS1 and MS2 heater: 101 °C; Gas temp: 200 °C; Gas flow: 14 l/min; Nebulizer: 35psi; Sheath gas temp: 300 °C; Sheath gas flow: 11 l/min

2.9 Non-radioactive RNA *In Situ* Hybridization

In situ hybridization technique was derived from several protocols.

2.9.1 Preparation of RNA probe:

A fragment of DNA that is unique to *GATA12* was amplified using forward primer 5'-ACCACCATAACCGACAG-3' and reverse primer 5'-AGACAAGTGTTGTTGG-3' and cloned into pGEM®-T Easy vector system (Promega). Primers were designed outside the T7 and SP6 RNA polymerase promoters of pGEM®-T Easy vector, and the fragment containing the cloned insert amplified. This amplified product was then used as a template DNA in the transcription reaction as follows:

Template DNA 1µg

10X Transcription buffer (Roche) 2µl

10x DIG labeling Mix (Roche) 2µl

RNase inhibitor (Promega) 1µl

RNA polymerase (Promega) 2µl

RNase free H_2O to a total of 20µl

The reaction was incubated at 37°C for 2h, following which 2µl of RNase-free DNase I (10000U, Promega) was added and incubated for an additional 30min at 37°C. In order to verifiy RNA production, 1 µl of the reaction was run on 1% agarose gel. The transcription reaction was topped up to 100µl with DEPC-

treated H_2O and the synthesized RNA was chopped into pieces of 150bp long by adding 100µl of 2X CO_3- buffer (80mM $NaHCO_3$, 120mM Na_2CO_3) to the reaction and incubated at 60°C for a period of time calculated using the following formula:

$$\text{Incubation time (min)} = \frac{\text{initial length of probe (in kb)} - 0.15}{0.11 * \text{initial length of probe(in kb)} * 0.15}$$

After hydrolysis, 10µl 10% acetic acid was added to neutralize the reaction. RNA probes were precipitated with 1/10 volume 3M NaAc (pH 5.2), 2.5 volume ethanol and 2µl of 10mg/ml tRNA (Roche). Pellets were rinsed with 70% ethanol (prepared in DEPC-water) and re-suspended in 50% formamide. RNA probes were used at a final concentration of 0.5ng/µl/kb.

2.9.2 Tissue Preparation:

WT siliques of stages 6DAF and 9DAF were used as tissues for *in situ* RNA hybridization analysis. Fresh tissues were fixed in ice-cold fixative: 4% paraformaldehyde prepared in 1X PBS (10X PBS stock: 1.3M NaCl, 70mM Na_2HPO_4, 30mM NaH_2PO_4). Vacuum was applied to samples for 15min or more until the tissue sank. Fixative was replaced with fresh solution and incubated in 4°C for 2 days.

Tissues were dehydrated at 4°C as follows: 2×30min 1×PBS, 60min 30% ethanol, 60min 40% ethanol, 60min 50% ethanol, 60min 60% ethanol, 60min 70% ethanol, 60min 85% ethanol, and finally 95% ethanol with 0.05 g/ml of eosin for overnight.

After dehydration, tissue was cleared at room temperature as follows: 2×30 min 100% ethanol with eosin, 2×60min 100% ethanol with eosin, 60min 25% Histoclear in ethanol, 60min 50% Histoclear in ethanol, 60min 75% Histoclear in ethanol, 2×60min 100% Histoclear. Finally, the tissue was immersed in 100% Histoclear with 1/4 volume of paraplast chips and incubated overnight without shaking.

The next day the tissue was transferred to a 42°C incubator until chips melted. 1/4 volume paraplast chips were added again and tissues were moved to 55°C for several hours. The wax/Histoclear mixture was replaced with freshly melted wax and incubated overnight at 55°C. For the next four days, fresh wax changes were done twice a day. The tissue was then placed in molds and allowed to solidify at room temperature. They were stored at 4°C for later use.

2.9.3 Sectioning:

Sectioning was done using an ultra microtome. 10µm thick sections were cut and placed on warm DEPC-treated water on top of ProbeOn Plus™ glass slide (Fisher Biotechnology). The slide was then placed on the top of a 42°C heated slide warmer to allow evaporation of H_2O and allow the ribbon to flatten out on the slide. Excessive H_2O was then removed and the slide was kept on the slide warmer at 42°C overnight to achieve complete drying and tight adherence of tissues onto slides. Sectioned tissues can be stored for several weeks at 4°C.

2.9.4 *In situ* section pre-treatment:

The prepared sections were deparaffinized and rehydrated as follows: 2X12min Histoclear, 2X2min 100% ethanol, 1min 95% ethanol, 1min 90% ethanol, 1min 80% ethanol, 1min 60% ethanol, 1min 30% ethanol, 1min H_2O. After rehydration, slides were incubated in 2X SSC (from 20X SSC stock: 3M NaCl, 300mM sodium citrate, pH 7.0) for 20min. This was followed by 30min proteinase K (2µg/ml) in 100mM Tris pH8.0, 50mM EDTA at 37°C. Incubation in 2mg/ml glycine in PBS for 2min to quench the proteinase K. The slides were then washed with 1XPBS for 2X2min at room temperature. The slides were then fixed with fresh 4% (w/v) paraformaldehyde in 1X PBS pH7 at room temperature. The slides were then washed in PBS for 2X5min. Then, slides were incubated for 10min in triethanolamine solution (freshly-prepared by adding 2.68ml of triethanolamine into 200ml RNase-free H2O containing 0.8ml of HCl and 1ml of acetic anhydride, mixed vigorously) and followed by 2× 5min washes with 1×PBS. Dehydration of slides was done with the following washes: 30s 30%ethanol, 30s 60% ethanol, 30s 80% ethanol, 30s 90% ethanol, 30s 95% ethanol and 2×30s 100% ethanol.

2.9.5 In situ hybridization

Hybe solution contains 100 µl 10× in situ salts, 400 µl deionised formamide, 200 µl 50% dextran sulphate, 20 µl 50× Denhardt's solution (warm to 50°C before pipetting), 10 µl tRNA (10 mg/ml) and 70 µl H2O (DEPC-treated). The total volume of 800 µl Hybe solution was enough for 3 pairs of slides. The RNA probes were denatured as follows: 1-2 µl of RNA probe from stock at -

80°C was topped up to 60 µl with 50% formamide, followed by heating at 80°C for 2 min and immediate cooling with ice. The slides were sandwiched together and Hybe solution was pulled up by capillary action. Slides were elevated on a rack in a sealed plastic container containing sterile H2O and incubated at 50°C in a hybridization oven overnight.

2.9.6 Post-hybridization

Each pair of slides was separated by dipping in 55°C pre-warmed 0.2× SSC buffer before placing in the glass slide holder. The slides were washed three times with 0.2× SSC for 60min at 55°C in a shaker incubator. Thereafter, the slides were washed with 1× PBS buffer at room temperature for 5min. Subsequently, the slides were blocked for 45min with 1% Boehringer blocking reagent dissolved in 100 mM Tris pH 7.5, 150 mM NaCl, followed by another 45min incubation with a second blocking solution (BSA/Tris/NaCl/Triton) containing 1.0% Bovine Serum Albumin (BSA) dissolved in 100 mM Tris pH 7.5, 150 mM NaCl, 0.3% Triton X-100. All incubation steps were performed at room temperature on a platform rocker. After blocking, anti-DIG antibody (Roche) was diluted 1:500 in 4ml of new BSA/Tris/NaCl/Triton solution described in previous washing step, and the antibody solution was poured in a plastic weighing dish. Slides were sandwiched together and antibody solution was pulled up by capillary action. Solution was drained on Kimwipes and the dipping process was repeated. Care was taken to avoid bubbles between two slides. Slides were elevated on a rack in a sealed plastic container containing sterile H2O and allowed to sit at room temperature for 2

h. After antibody incubation, slides were drained on Kimwipes and separated in the glass slide holder containing BSA/Tris/NaCl/Triton solution. This washing step was repeated 4 times for 15min each at room temperature on a platform rocker, followed by washing with 100 mM Tris pH 9.5, 100 mM NaCl, 50mM MgCl2 solution (Tris pH 9.5/NaCl/MgCl2) for 10min to remove detergent. The substrate solution Tris-NaCl-PVA for color development was prepared as follows: 10% (w/v) 40kDa polyvinyl alcohol (PVA) (Sigma) was dissolved in Tris pH 9.5/NaCl/MgCl2 solution, which was then heated to 50°C, mixed vigorously, and cooled down to room temperature. 60 µl of NBT/BCIP stock solution (Roche) was mixed with 3ml of Tris-NaCl-PVA substrate solution. 200 µl of prepared substrate solution was added to each pair of slides which were then sandwiched together. Slides were elevated on a rack in the plastic container containing sterile H2O in total darkness for 24h at room temperature. On the next day, slides were separated and placed in the glass slide holder. The slides were rinsed with tap water for three times to stop the reaction. The slides were dehydrated by washing with 70% ethanol for 5s and 2 times 100% ethanol for 2s each. Slides were air dried before being mounted with 50% glycerol for maintenance of signals for at least three months and observation under microscopes.

2.10 *Ab initio* Promoter Analysis

Motifs/putative transcription factor binding sites on GATA12 promoter region (1 kb upstream of the CDS of AT5G25830, http://ppdb.agr.gifu-u.ac.jp/ppdb/cgi-bin/index.cgi) were predicted by using MATCH program version 2013.3 (TRANSFAC: https://portal.biobase-international.com/cgi-

bin/build_t/idb/1.0/Match/index.cgi). Prediction was done by using only high-quality matrices by setting two different cut-offs i.e. by minimizing false positives and by minimizing the sum of both positive and negative error rates.

2.11 Statistical Analysis

Statistical differences between samples were determined by two-tailed, unpaired t-test with equal variance (computed with Microsoft Excel) and samples differences were considered to be statistically significant if $P<0.05$, indicated in figure legends as $*P<0.05$ or $**P<0.01$.

3. RESULTS

3.1. UNDERSTANDING THE ROLE OF GATA12 (AT5G25830) IN REGULATING SEED GERMINATION AND DORMANCY

3.1.1 Selection of *GATA12* from a microarray

RGL2 being the major DELLA protein that represses seed germination, a microarray was done in our laboratory to identify the genes that are differentially regulated by this protein. For this the seeds of *ga1-3 rga-t2 rgl2-1* and *ga1-3 rga-t2* were stratified at 4°C for 5 days and their transcriptomes compared. This study helped to shed light on some of the genetic events downstream of RGL2. Several genes were found to be differentially regulated by a functional RGL2 protein, like genes encoding cell wall modifying enzymes, viz., *CP1* and *EXPA8*, with at least *EXPA8* being directly regulated (Stamm et al. 2012). In addition, many genes encoding for transcription factors were also differentially regulated (Figure 3.1). Of these, a GATA-type transcription factor (*GATA12*) was chosen for further characterization in our study.

There are 30 genes within the GATA family of zinc finger transcription factors in *Arabidopsis* (Reyes et al. 2004). Several characterized members of the GATA family have roles in regulating embryo development and seed germination (Table 3.1). Also, public microarray data indicate that *GATA12* is highly expressed in mature dormant seeds (Figure 3.2) and thus it likely regulates seed germination or dormancy downstream of RGL2. Thus, we

decided to characterize the role of GATA12 in regulation of *Arabidopsis* seed germination/dormancy downstream of RGL2.

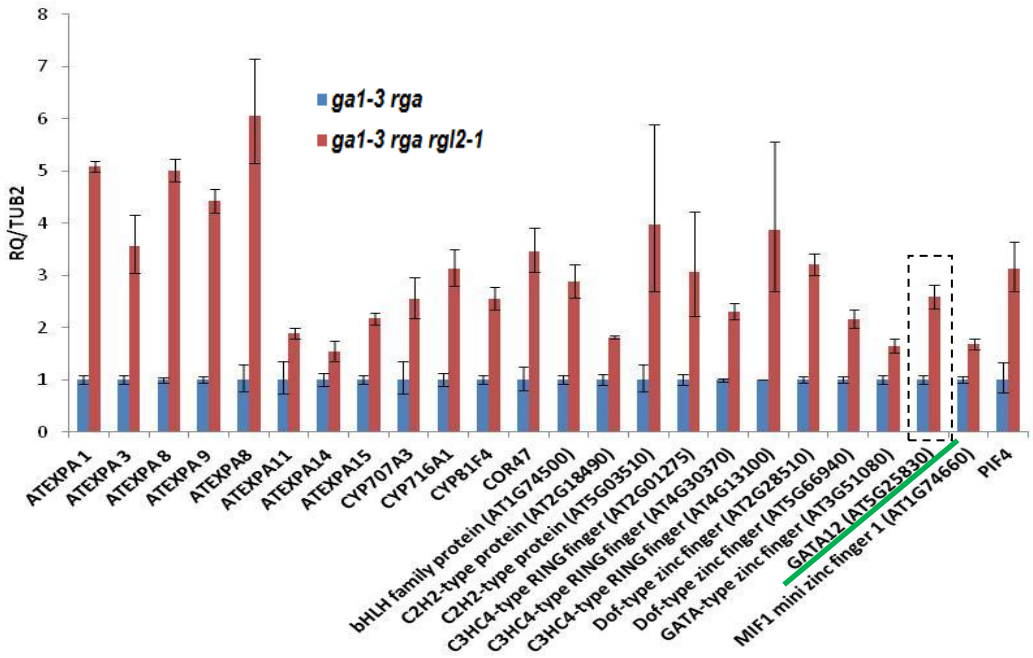


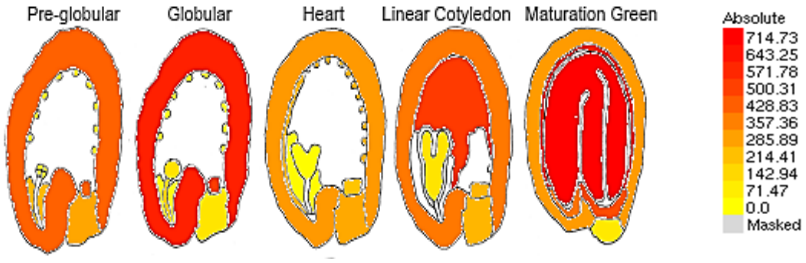
Figure 3.1: Differential regulation of genes by a functional RGL2 protein

Several genes were identified from the microarray as differentially regulated by a functional RGL2 protein. This graph represents the qRT-PCR analysis of selected down-regulated genes, several of them encoding transcription factors. Error bars represent SD. Among others, *GATA12* transcript levels are significantly different (t-test, $P<0.05$, n=2) in the transcriptomes of *ga1-3 rga-t2 rgl2-1* and *ga1-3 rga-t2* seeds, suggesting it is regulated by a functional RGL2. Thus, GATA12 (AT5G25830) may play a role in GA-mediated regulation of seed germination. Error bars denote standard deviation.

Table 3. 1: List of characterized members of the GATA-type zinc finger family in *Arabidopsis*.

Of the 30 GATA factors, only 8 have been functionally characterized. Most of them play roles either in embryo development or seed germination.

GATA-type zinc finger	Function	Reference
GATA 8 (BME3)	Positive regulator of seed germination	(Liu et al. 2005)
GATA 21 (GNC) GATA22 (GNL/CGA1)	GNC and GNL repress germination downstream from DELLA proteins and PHYTOCHROMEINTERACTING FACTORS. *gnc* and *gnl* mutations suppress the *ga1* phenotype	(Richter et al. 2010)
GATA18 (HANABA TARANU)	Positions the inductive proembryo boundary Affects flower and shoot apical meristem (SAM) development	(Zhao et al. 2004, Nawy et al. 2010)
GATA2	Mediates the crosstalk between BR- and light-signalling pathways	(Luo et al. 2010)
GATA25 (ZIM) GATA24 (ZML1) GATA28 (ZML2)	Elongation of hypocotyls and petioles under all wavelengths of light Essential components of the cry1-mediated photoprotective response.	(Shikata et al. 2004) (Shaikhali et al. 2012)



Data provided by Goldberg and Harada labs. Images adapted from images drawn by Meryl Hashimoto. Data published in Le et al. (2010)

Figure 3.2: Public microarray data showing expression of *GATA12* during *Arabidopsis* embryogenesis.

A public microarray data from *Arabidopsis* eFP browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) showing high expression of *GATA12* in the mature dormant embryo.

3.1.2 Cloning and molecular characterization of GATA12 and its closest homolog

GATA factors are a class of C2-C2 zinc finger transcriptional regulators present in plants, fungi and metazoans. The name 'GATA' comes from the fact that these proteins generally recognize the consensus sequence WGATAR (W=T or A; R=G or A) in their downstream targets. They are constituted by a type IV zinc finger in the form $CX_2CX_{17-20}CX_2C$. The four cysteine residues in the zinc finger domain are highly conserved throughout the GATA family and thus they are classified as C2-C2 zinc finger proteins (Figure 3.3A). In *Arabidopsis*, so far 30 GATA factors have been identified and they have been classified into four subfamilies based on the phylogenetic analysis of their full-length protein sequences (Figure 3.3B) (Reyes et al.

2004).GATA12 belongs to the subfamily I and contains an N-terminal acidic transactivation domain followed by a C-terminal zinc finger domain of the form $CX_2CX_{18}CX_2C$ (Figure 3.3C).

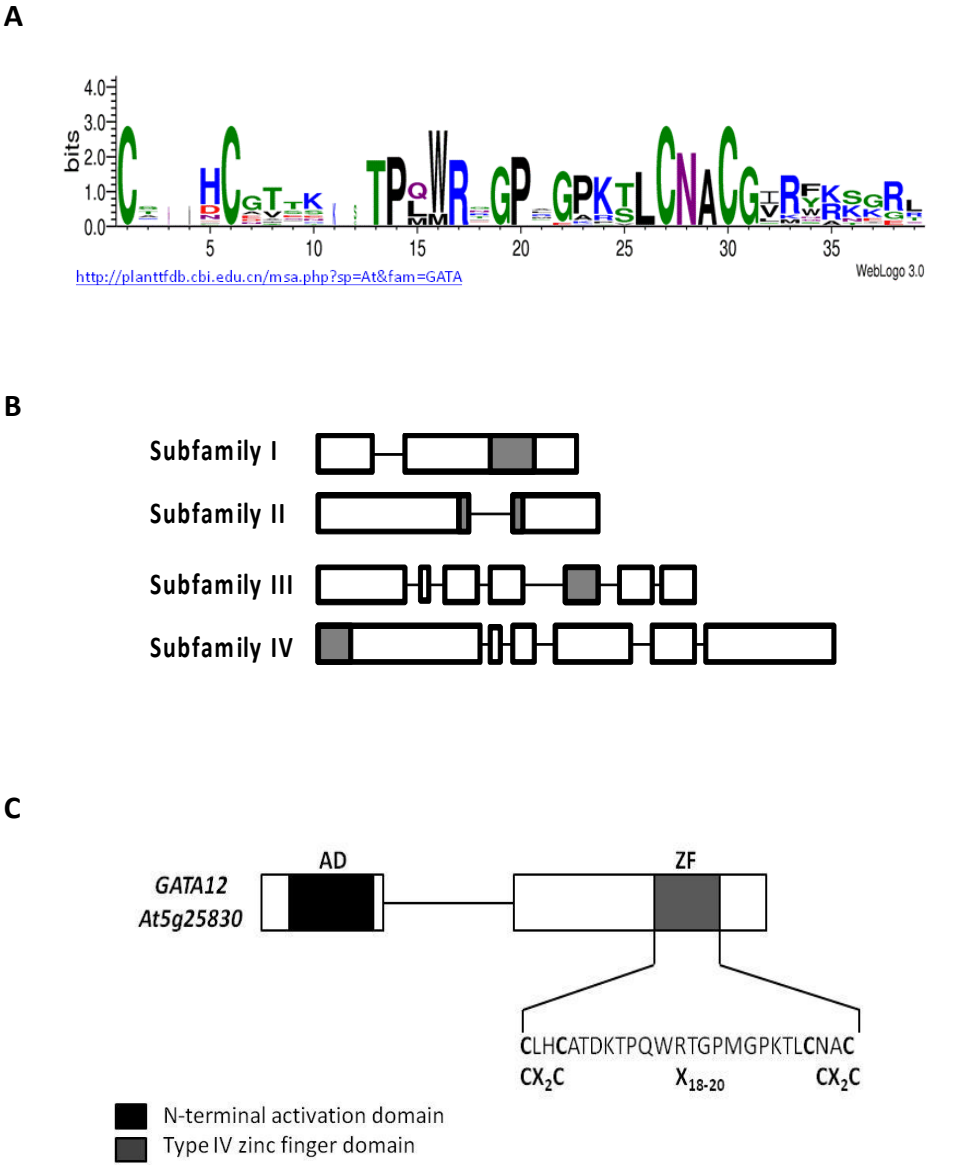


Figure 3.3: GATA12 belongs to sub-family I of the C2-C2 GATA-type zinc finger family of transcription factors

(A) Sequence logo of all 30 members of the GATA-type family showing the four conserved cysteine residues that give the name C2-C2 to this family. **(B)** The 30 members of C2-C2 zinc finger GATA-type family of transcription factors are classified into four sub-families. The grey shaded boxes represent the position of the zinc finger domain which differs across the sub-families. **(C)** GATA12 is a member of sub-family I containing an N-terminal activation domain and a C-terminal zinc finger domain of the form $CX_2CX_{18}CX_2C$.

GATA12 is located on chromosome 5 of *Arabidopsis*. Its gene model is comprised of 2 exons with a single intron in between them (Figure 3.4A). Cloning of the *GATA12* full length gene was done by using primers designed to amplify the coding sequence (CDS) obtained from The *Arabidopsis* Information Resource (TAIR). The full length CDS comprises of 996bp (Figure 3.4B). The promoter region of *GATA12* was cloned by amplifying the region that is 1kb upstream of the CDS (Figure 3.4C). The amplified PCR products were first cloned into pGEMT-easy vector and subsequently into pGREEN vectors for further downstream studies.

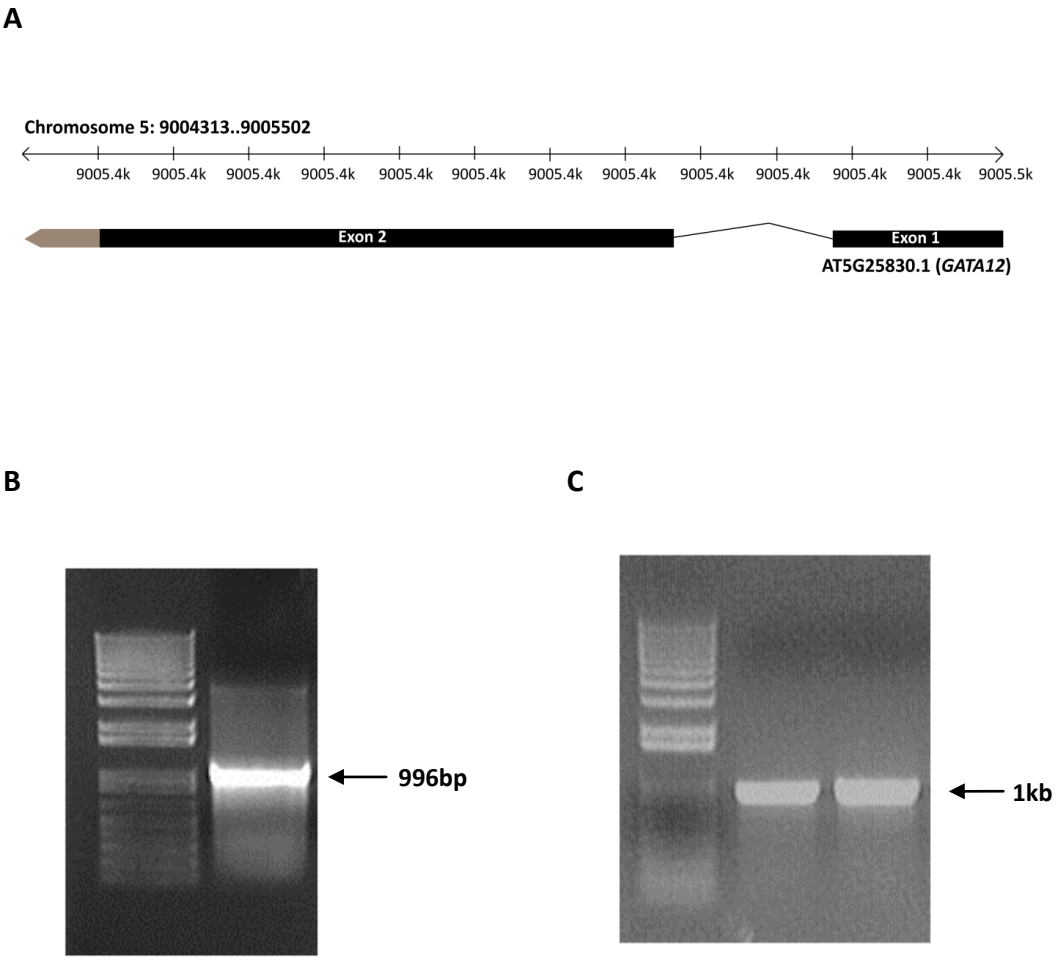
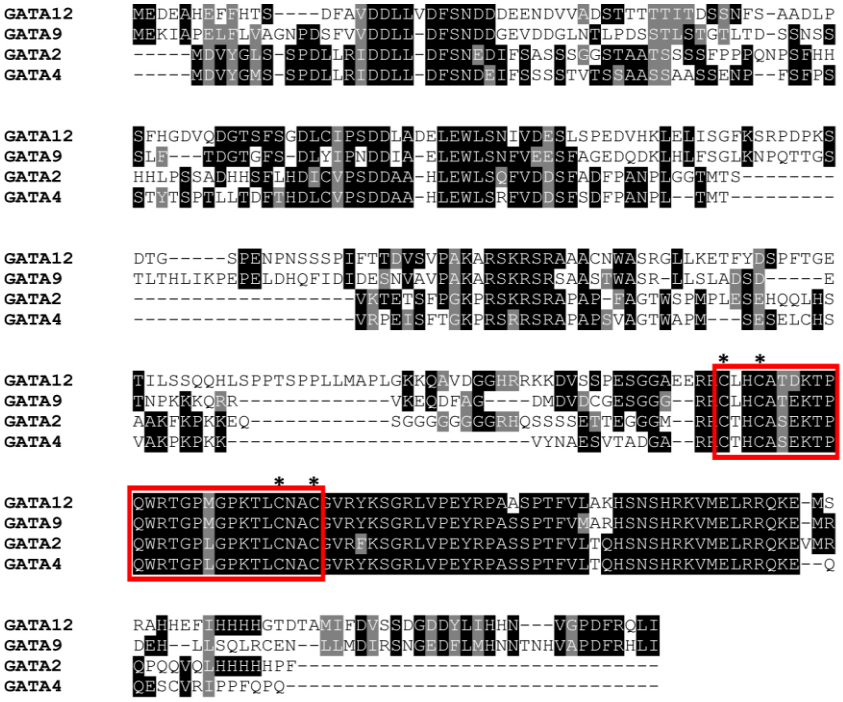


Figure 3.4: Cloning of *GATA12* CDS and *GATA12* promoter sequences

(A) Protein coding gene model of GATA12 showing two exons and one intron. It is located in chromosome 5 (AT5G25830) **(B)** PCR amplification of full-length CDS of *GATA12* using seed cDNA as template. **(C)** PCR amplification of 1kb upstream promoter region of *GATA12* gene using genomic DNA from *Arabidopsis* seedlings as template.

Among the members of sub-family I, the closest homologs of GATA12 were identified as GATA9, GATA2 and GATA4 based on phylogenetic analysis of their protein sequences (Reyes et al. 2004). A ClustalW2 alignment of the four homologous proteins shows that C-terminal zinc finger domains and the adjoining basic regions are highly similar in all of them. A phylogenetic tree generated using ClustalW2 showed that amongst these homologs, GATA9 was the closest relative to GATA12 (Figure 3.5A and 3.5B). Thus, we decided to include GATA9 along with GATA12 in some of our molecular characterization studies to see if they play redundant roles.

A



B

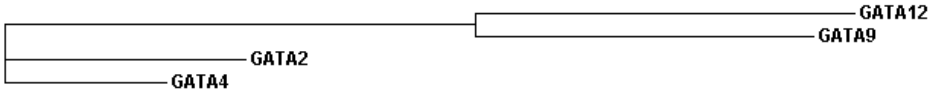


Figure 3.5: Phylogeny and multiple sequence alignment of GATA12 and homologs

(a) Phylogenetic relationship of the members of subfamily I based on protein sequence shows GATA9, GATA2 and GATA4 as close homologs of GATA12**.** ClustalW2 multiple sequence alignment of protein sequences of the four close homologs show that their C-terminal zinc finger domain and the subsequent residues are highly conserved. **(B)** ClustalW2 phylogenetic tree of the four close homologs shows that GATA9 is the closest relative of GATA12.

3.1.2.1 Tissue-specific expression analysis of GATA12 and GATA9

Tissue-specific expression of *GATA12* and its closest homolog *GATA9* was studied by extracting RNA from different tissues of WT plants and analysing the transcript abundance using quantitative real time-PCR (qRT-PCR). *GATA12* showed a high level of expression in freshly harvested (FH) mature seeds. The next highest level was in the stem (inflorescence axis, 10-fold lower than in the seeds), followed by similar levels in roots, rosette leaf, cauline leaf and flowers (Figure 3.6A). A similar pattern was seen with tissue-specific expression of *GATA9*. The levels were the highest in freshly harvested mature seeds, followed by rosette leaf, cauline leaf, flowers, root and stem (Figure 3.6B).

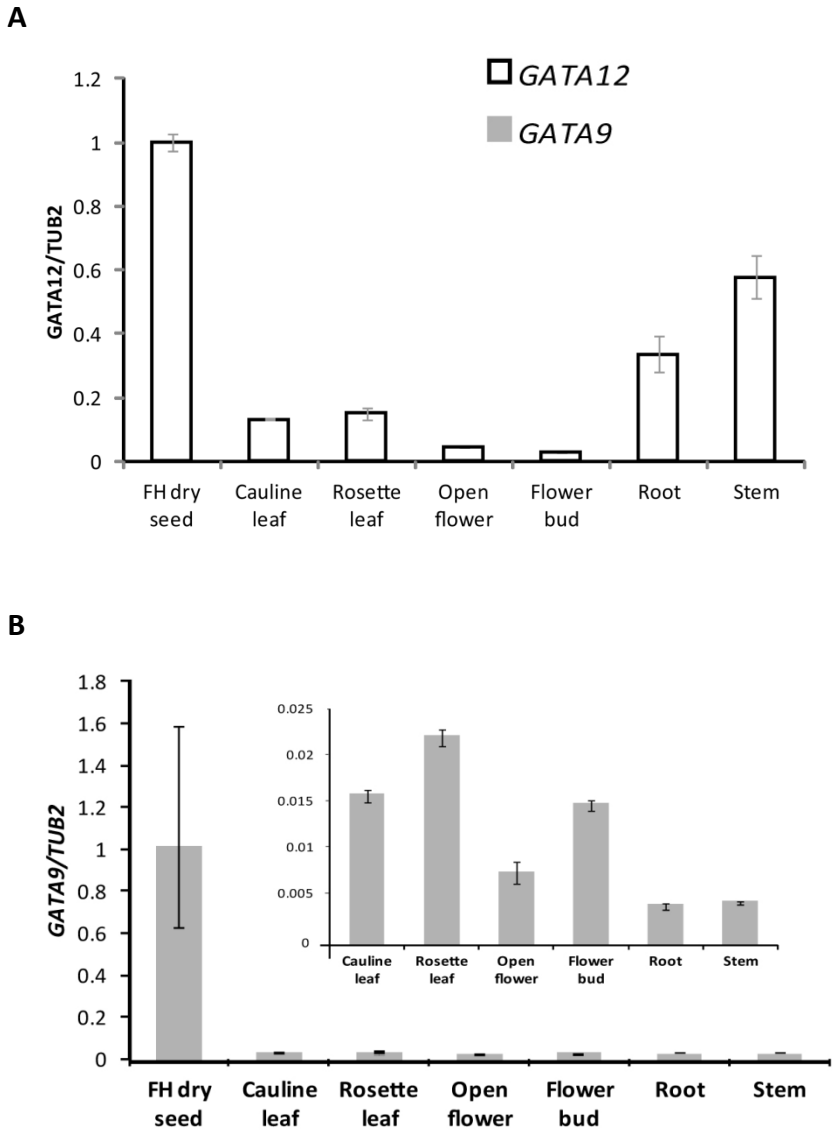


Figure 3.6: Tissue-specific expression levels of *GATA12* and *GATA9* transcripts

(A) *GATA12* expression pattern in wild-type tissues as determined by qRT-PCR. *GATA12* transcripts are highly abundant in freshly harvested dry seeds (t-test, P<0.01, n=3) followed by inflorescence axis (stem) and roots. Error bars represent standard deviation. **(B)** A similar expression pattern seen for *GATA9* transcripts, with predominantly high amounts in freshly harvested dry seeds (t-test, P<0.01, n=2) compared to other tissues, followed by rosette leaves. Error bars represent standard deviation.

Since GATA-type transcription factors have been implicated to have roles in light-mediated and circadian-regulated gene expression, we decided to check the expression levels of *GATA12* and *GATA9* in response to light. RNA was extracted from light-grown (constant light) and dark-grown etiolated seedlings of the same age and a qRT-PCR was done to check the transcripts levels of both genes. We found that both genes showed a ~2-fold increase in transcript abundance in etiolated seedlings compared to light grown seedlings (Figure 3.7).

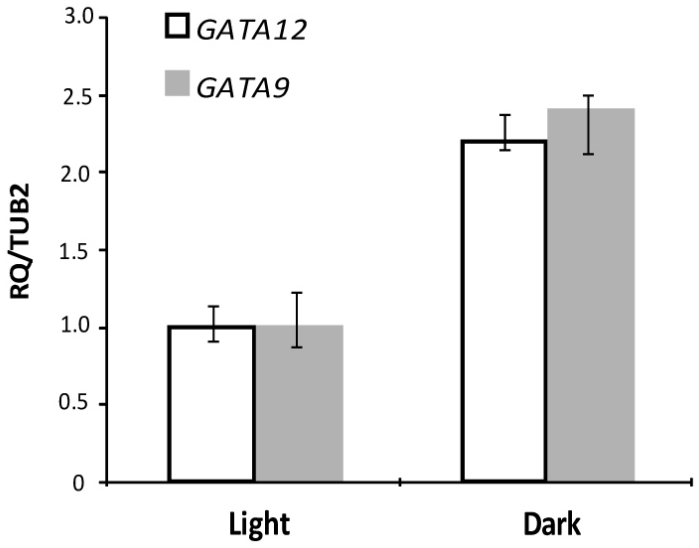


Figure 3.7: Expression of *GATA12* and *GATA9* transcripts in light- and dark-grown seedlings

Transcript abundance *GATA12* and *GATA9* was measured by qRT-PCR analysis using cDNA from both light-grown (photomorphogenetic) and dark-grown (skotomorphogenetic) seedlings (t-test, $P<0.05$, n=2). Error bars denote standard deviation.

3.1.2.2 Expression analysis in GATA12pro::GUS plants.

In order to further probe the expression of *GATA12* in various tissues and under different developmental conditions, wild type plants were transformed with constructs constituting the 1 kb upstream promoter region of *GATA12* driving the expression of *GUS* gene. Histochemical GUS staining of *GATA12pro*::GUS plants showed high expression in dry seeds, especially in the embryonic radicle. Germinated seeds showed lesser expression in the radicle. *GATA12pro* driven *GUS* expression was also seen in the hydathodes of rosette leaves and was quite high in the etiolated seedling, stamens and inflorescence axis. As the staining was quite high in the stamens, we further investigated the expression in pollen grains. The levels were high in mature pollen grains but diminished in the germinated pollen grains and their pollen tube (Figure 3.8).

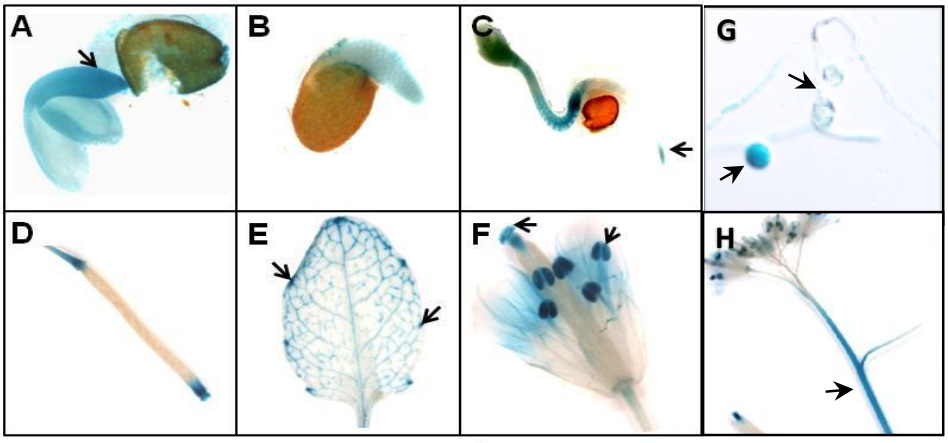


Figure 3.8: Histochemical GUS staining of *GATA12pro*::GUS transgenic plants

GUS staining of various tissues **(A)** Freshly harvested seed imbibed for 1h, embryo and seed coat seperated **(B)** Germinated seed **(C)** etiolated seedling **(D)** Silique **(E)** Rosette leaf **(F)** Inflorescence **(G)** Mature pollen and germinated pollen with pollen tube **(H)** Inflorescence axis (stem)

3.1.2.3 Sub-cellular localization of GATA12-GFP

Sub-cellular localization studies were done in *Arabidopsis* mesophyll protoplasts by transfecting them with *35S: GATA12-GFP* plasmid construct. Confocal microscopy images show that the green fluorescence signal was localised to nucleus (Figure 3.9). Thus, like RGL2, GATA12 is also a nuclear-localised transcription factor.

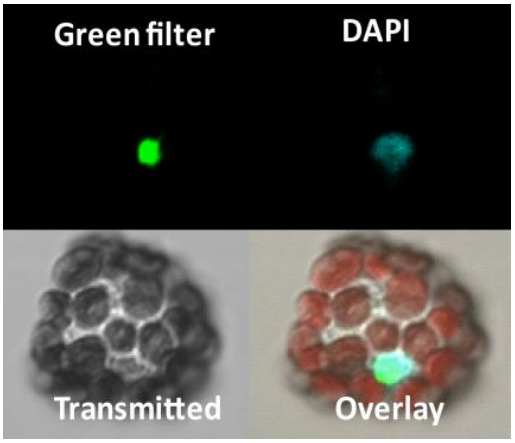


Figure 3.9: Sub-cellular localization studies in *Arabidopsis* mesophyll protoplasts

Mesophyll protoplasts transfected with *35S:GATA12-GFP* construct emit a fluorescence signal that overlays with the signal from the DNA binding stain DAPI, showing it is a nuclear-localized transcription factor (n=3).

3.1.2.4 Expression of *GATA12* and *GATA9* during embryogenesis.

Since, the two GATA factors are highly expressed in the freshly harvested seeds, we went on to detect their expression during various stages of embryogenesis. RNA was extracted from siliques collected at 3 days post anthesis (dpa), 6dpa, 9dpa, 12dpa and freshly harvested dry seeds. qRT-PCR analysis showed that the *GATA12* and *GATA9* transcript levels rose through embryogenesis and peaked in the fully mature 'dormant' seeds (Figure 3.10A and 3.10B). This data was consistent with public microarray data of *GATA12* as shown before (Figure 3.2). *In situ* hybridization studies were done using a RNA probe that hybridizes with GATA12 transcripts (Figure 3.10C). This study further confirmed the high expression of *GATA12* in mature embryos. Thus, GATA12 and GATA9 are likely to play an important role during embryogenesis.

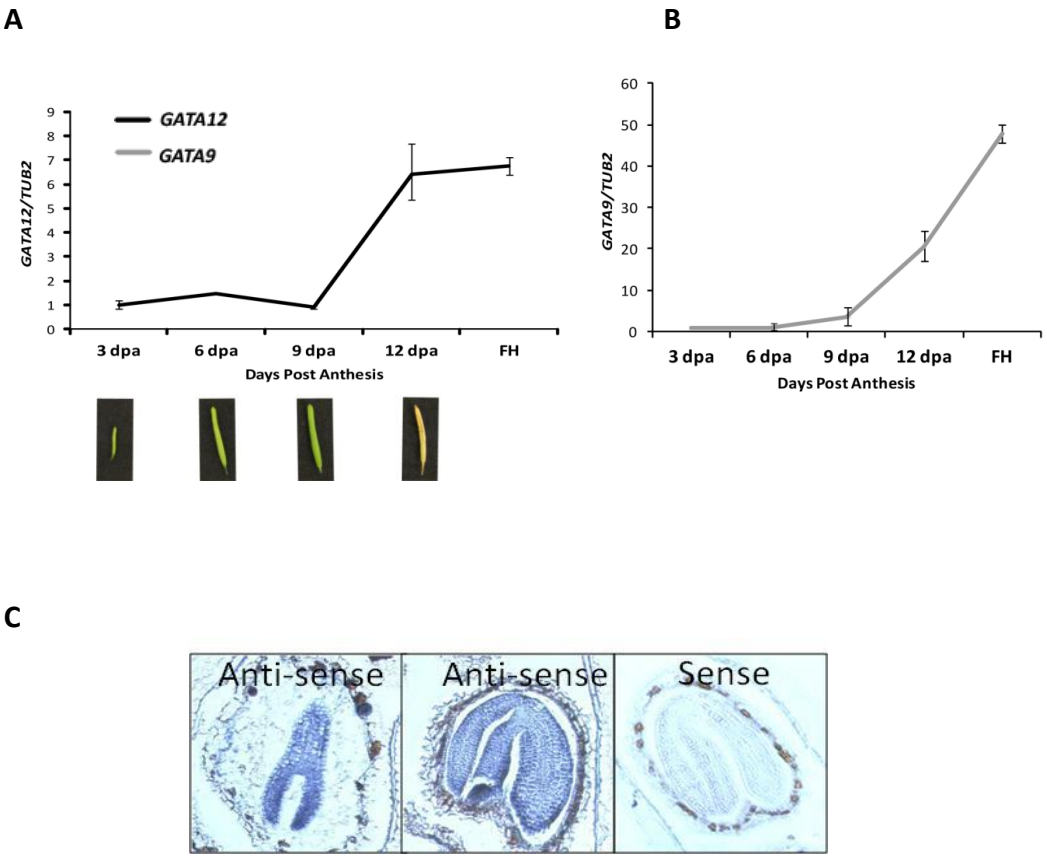


Figure 3.10: *GATA12* and *GATA9* transcript levels rise during embryogenesis

During silique development, **(A)** *GATA12* and **(B)** *GATA9* transcripts start to increase from 9 days post anthesis (dpa) and accumulate to high levels in siliques at 12dpa and freshly harvested mature seeds that are highly dormant (t-test, $P<0.01$ at 12dpa and FH from 3dpa, n=3 for GATA12 and n=2 for GATA9). Error bars denote standard deviation (SD). **(C)** *In-situ* hybridization studies using *GATA12* anti-sense RNA probe show transcript abundance in mature embryos. The sense probe serves as a negative control.

3.1.3 Phenotypic characterization of *GATA12* and *GATA9* transgenic and mutant plants

To further probe the role of GATA12 and its closest homolog in regulation of seed germination and dormancy, we generated transgenic plants showing suppression and overexpression of *GATA12* transcripts and phenotypically characterized these lines.

3.1.3.1 Generation of Transgenic Plants

T-DNA insertion lines were ordered from the ABRC seed resource. For GATA12, unfortunately there were no T-DNA insertions in its exon region. We obtained two T-DNA insertions lines, one in the 3'UTR (SALK_052546C) and other in the upstream (-328) promoter region (SALK_012501C) (Figure 3.11A). However, for GATA9, a T-DNA insertion in exon2 (SALK_152156C) was available (Figure 3.11B). When tested for expression of transcripts, the levels of *GATA9* were significantly suppressed in this line (referred to as *gata9*) (Figure 3.13D).

Figure 3.11: T-DNA insertion lines obtained from ABRC seed resource

(A) Schematic representation of *GATA12* gene structure showing the positions of T-DNA insertion, one in 3'UTR (SALK_052546C) and other within the promoter region at 328bp upstream of CDS (SALK_012501C). **(B)** Schematic representation of *GATA9* gene structure showing the position of T-DNA insertion in exon2 (SALK_152156C).

Since the suppression of the *GATA12* was not in significant levels in the T-DNA insertion lines (Figure 3.12E) , we also generated *GATA12*-antisense (*GATA12*-AS) lines. Generation of *GATA12*-AS and *GATA12* overexpression (*GATA12*-OE) lines was done by transforming WT plants by floral dip method with 35S:*GATA12* and 35S:*GATA12*-AS constructs respectively. The transgenic plants obtained were genotyped to confirm the presence of construct (Figure 3.12D) and expression analysis was done in T2 plants to check for significant suppression or overexpression of *GATA12*. All the *GATA12*-OE plants tested showed significant overexpression (50-200 fold) of *GATA12* transcripts (Figure 3.13A). For expression analysis of 35S:*GATA12*-AS plants, in addition to checking GATA12 expression, we also checked the expression of GATA9, GATA2 and GATA4 (Figure 3.13C). This was done to make sure that any phenotypic differences seen were due to suppression of *GATA12* only. These plants showed 2-5 fold decrease in *GATA12* transcripts (Figure 3.13B). The confirmed antisense lines were renamed as AS1, AS2, AS3 and AS4. These confirmed transgenic and mutant lines were then used for phenotypic assays in the T3 generation.

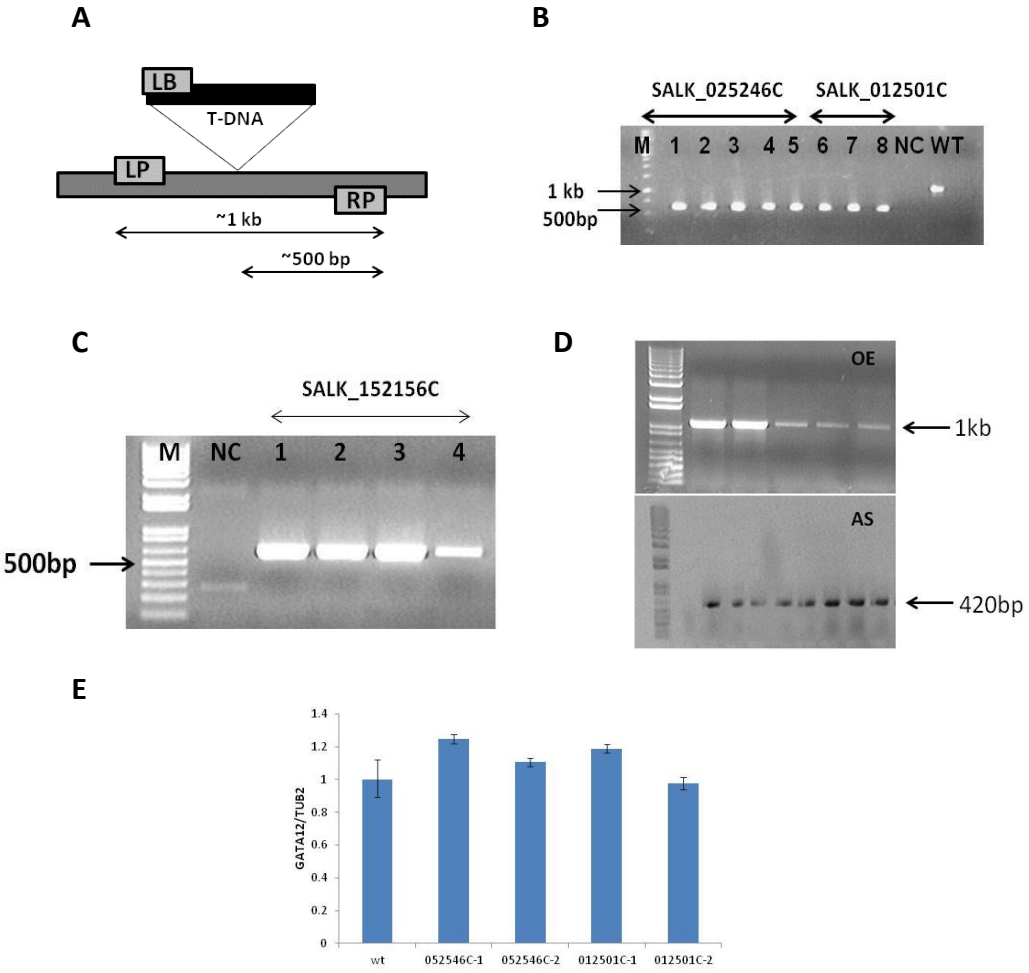


Figure 3.12: Genotyping of *GATA12*-AS, *GATA12*-OE and *gata9* lines

(A) A schematic representation of the basic design of primers obtained from SALK T-DNA express **(B)** Genotyping of T-DNA insertion lines of GATA12 using SALK T-DNA primers (LP+LB+RP). A single band at ~ 500 bp and no band corresponding to 1kb confirms that plants are homozygous for the T-DNA insertion. **(C)** Gel image showing genotyping results for *gata9* mutant plants using SALK T-DNA primers (LP+LB+RP). **(D)** Genotyping of T1 plants to confirm the presence of 35S:*GATA12* and 35S:*GATA12*-AS constructs in *GATA12*-OE and *GATA12*-AS plants respectively. **(E)** Expression of *GATA12* in SALK T-DNA insertion lines shows no significant suppression. Error bars denote SD.

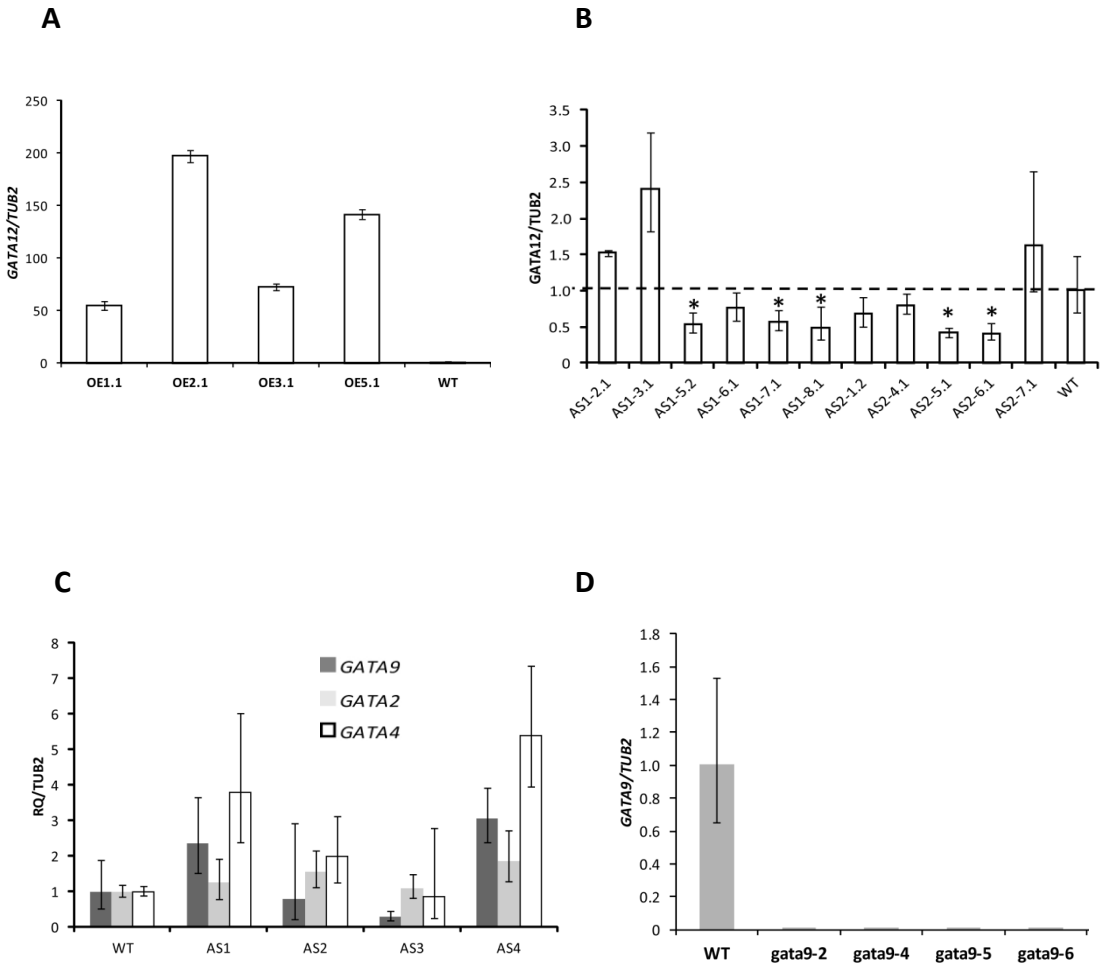


Figure 3.13: Expression analysis of *GATA12*-AS, *GATA12*-OE and *gata9* lines

(A) Expression analysis of 35S:GATA12 T2 lines show significant overexpression (t-test, P<0.01, n=2) (ranging from 50-200 fold) of *GATA12* transcripts. Error bars denote SD. **(B)** qRT-PCR analysis to check suppression of *GATA12* transcripts in *GATA12*-AS T2 plants. The lines marked with asterisk show significant suppression (t-test, *P<0.05, n=2) and were used for further phenotypic assays. Error bars denote SD. **(C)** qRT-PCR analysis to check if expression of GATA12 homologs: *GATA9, GATA2* and *GATA4* are affected in *GATA12*-AS plants. Error bars denote SD. AS1, AS2, AS3, AS4 are selected *GATA12*-antisense lines that show significant suppression of *GATA12* transcripts. **(D)** Analysis of *GATA9* transcript expression in *gata9* mutant plants using qRT-PCR shows significant suppression of *GATA9* transcripts (t-test, P<0.01, n=2). Error bars denote SD

3.1.3.2 Phenotypic Analysis

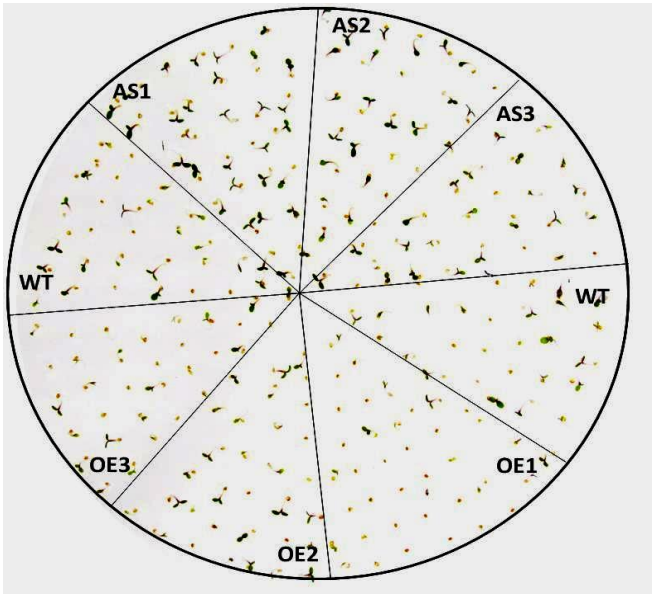
Phenotypic analysis of GATA12 transgenic plants showed no differences at maturity in comparison to WT plants. The height, flower structure, time of bolting were all normal (Figure 3.14). Since GATA12 was picked up from a microarray as one of downstream targets of RGL2, we decided to check for differences at the germination level. Freshly harvested seeds from transgenic and WT plants were dried for a week and then sown on MS plates. We found that, in the absence of cold stratification treatment, the *GATA12*-AS seeds germinated a little earlier that WT. On the other hand, *GATA12*-OE seeds showed a poor germination percentage. Thus, *GATA12*-AS seeds were less dormant and GATA12-OE seeds were more dormant that WT (Figure 3.15A and 3.16A). The *gata9* unstratified seeds however showed no significant differences even at the germination level in comparison to WT (Figure 3.15C). However, the phenotypic differences observed in GATA12 transgenic seeds were observed only when unstratified seeds were used and were abolished when subjected to cold stratification treatment at 4°C. Also, these phenotypic differences observed with freshly harvested (FH) seeds diminished with dry-storage (after-ripening) of seeds (Figure 3.15B). The rate of germination and percentage germination of 'stored' WT seeds became more closer to that of antisense transgenic seeds.

A

B

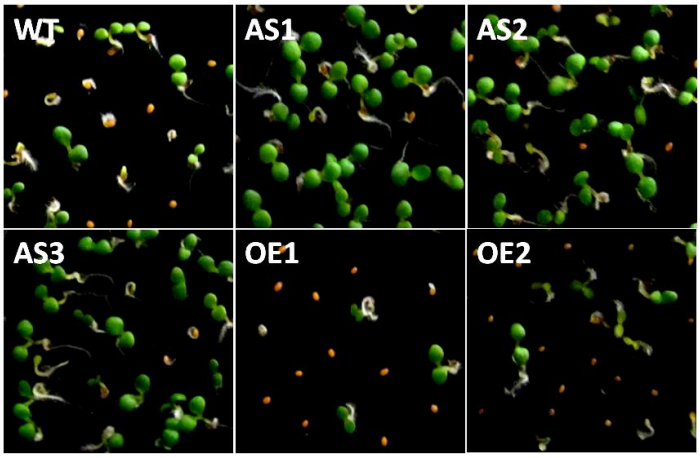
Figure 3.14: *GATA12* transgenic plants show no phenotypic differences in comparison to WT plants at maturity.
The flower development **(A)** and time of bolting and height of transgenic plants **(B)** were no different from that of WT.

A



WT - wild-type
AS - GATA12 antisense
OE - GATA12 over expression

B



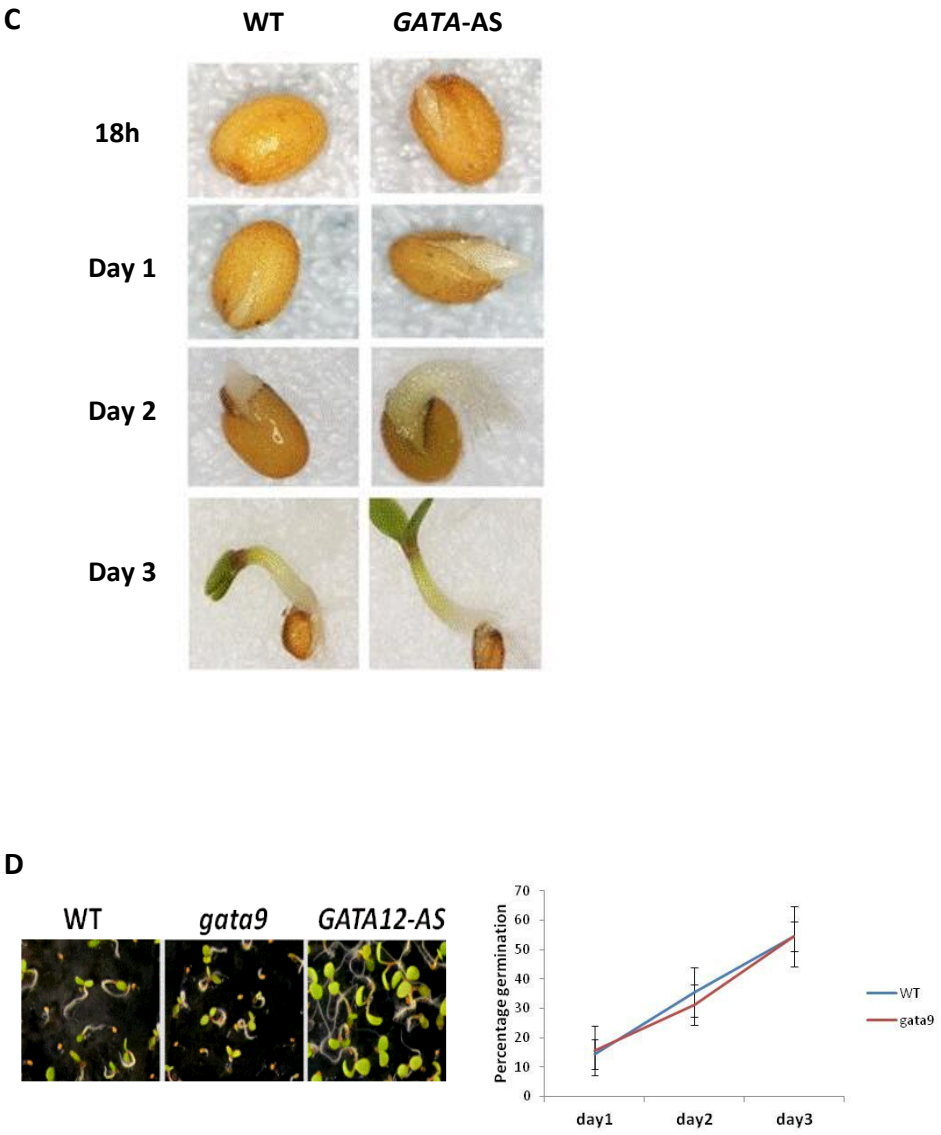


Figure 3.15: Phenotypic analysis of *GATA12* transgenic seeds and *gata9* mutant seeds

(A) GATA12-OE, *GATA12*-AS and WT seeds were sown on MS plates with 1% sucrose and 0.5% gelrite. The GATA12-OE (OE1, OE2, OE3) seeds showed a poor germination percentage whereas *GATA12*-AS (AS1, AS2, AS3) seeds show a better germination percentage and germination rate in comparison to WT seeds. **(B)** A close-up shot of the GATA12 transgenic seeds showing phenotypic differences. **(C)** A close-up shot of one representative individual seed at different time points of sowing from the anti-sense transgenic line that show phenotypic differences in comparison to WT. The shot represent the stage of seed germination shown by not all but majority of the seeds of the two lines, at different days of sowing.
(D) *gata9* mutant seeds show no significant differences in germination from WT, even under unstratified conditions. Error bars denote SD. n=3.

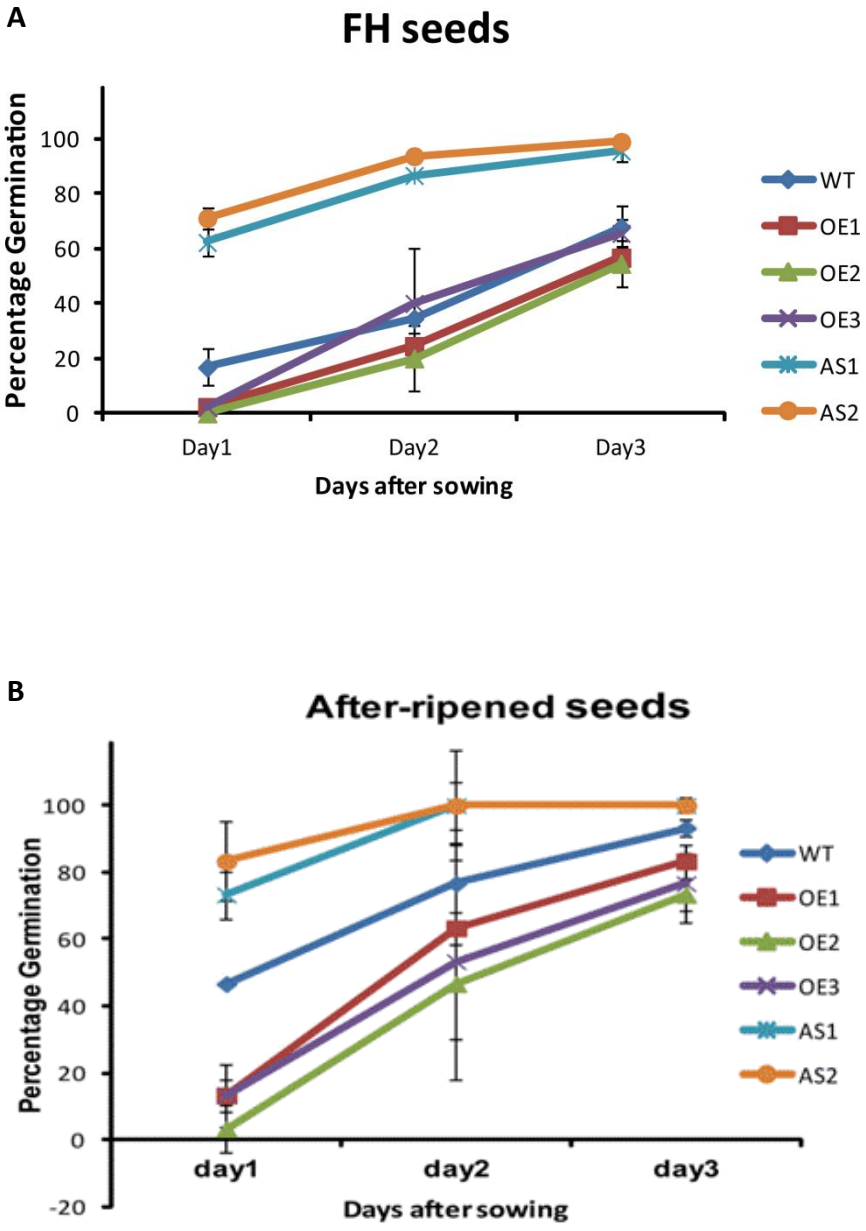


Figure 3.16: Germination percentages of GATA12 transgenic seeds: FH and after-ripened

Percentage germination of **(A)** freshly harvested (FH) and **(B)** One month after-ripened seeds (AR) of GATA12- OE and AS lines in comparison to WT seeds (t-test, n=3, *P<0.05 for all transgenics in comparison to WT on day1 and 2 for FH and on day1 for AR seeds). Error bars denote SD. Percentage germination here refers to percentage of seeds showing 'radicle protrusion'.

Since phenotypic differences were significantly reduced with cold stratification and dry storage of seeds, we decided to check if the level of *GATA12* transcripts were affected under these conditions. RNA was extracted from FH wild-type Col-0 seeds that were cold stratified for different time points. qRT-PCR analysis showed a decrease in *GATA12* transcripts with cold stratification (Figure 3.17A). Similarly, qRT-PCR analysis was done for RNA extracted from FH wild-type seeds stored for different time-points at 37°C (to allow accelerated after-ripening). Again, there was drastic reduction in *GATA12* transcripts with dry-storage (Figure 3.17B). This trend was also observed for the homolog, *GATA9*, transcripts.

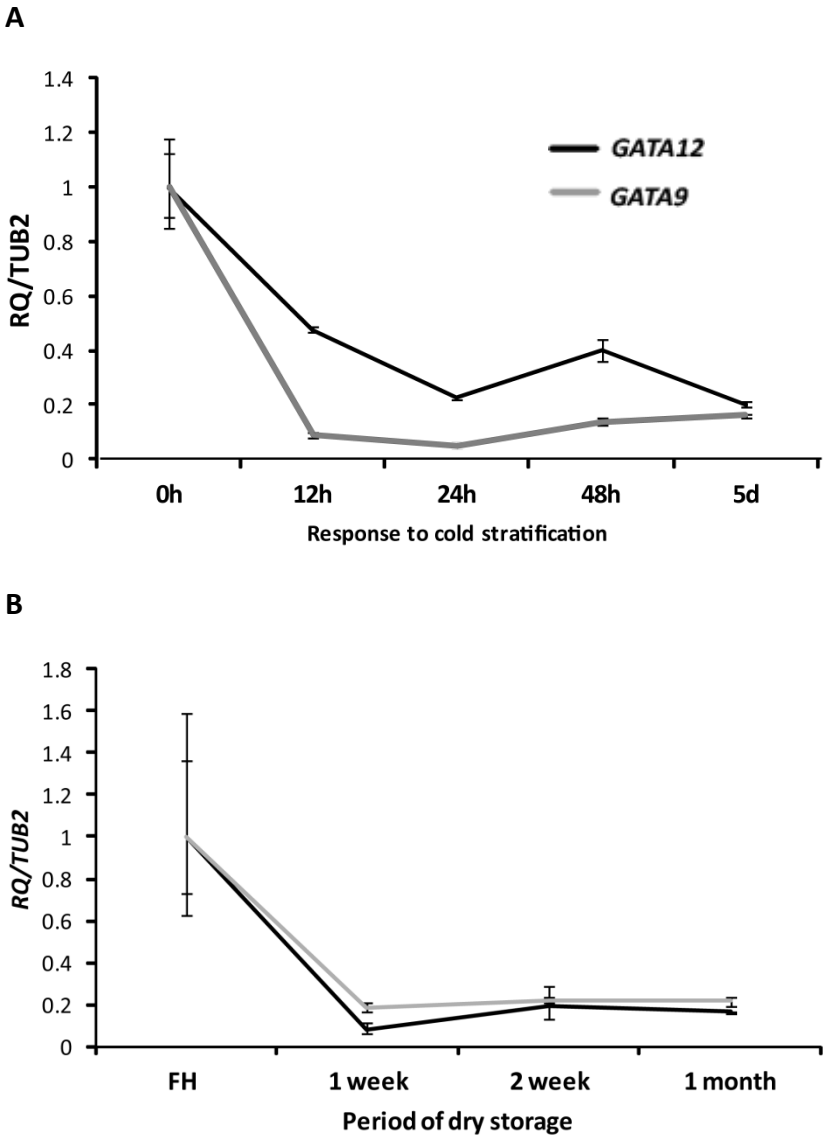


Figure 3.17: Effect of cold-stratification and dry-storage of seeds on levels of *GATA12* and *GATA9* transcripts

(A) qRT-PCR analysis of cDNA from FH wild-type seeds cold stratified for different time-points shows a drastic reduction in *GATA12* and *GATA9* transcripts (t-test, $P<0.05$ at 12h compared to FH, n=3). Error bars denote SD. **(B)** qRT-PCR analysis of cDNA from FH WT seeds stored at 37°C for different time-points (to allow accelerated after-ripening) shows a drastic reduction in *GATA12* and *GATA9* transcripts (t-test, $P<0.05$ at 1 month compared to FH, n=2). Error bars denote SD.

We then went on to check for any differences in the seed morphology that could be attributed to the differences in germination/dormancy of the transgenic seeds. Average seed size and seed mucilage extrusion were check. Images of dry and imbibed seeds were taken under the microscope and average seed size was measured using the ImageJ software. For checking mucilage extrusion, a ruthenium red staining of seeds was performed. Both seed size and mucilage extrusion upon imbibition were normal in the transgenic seeds in comparison to that of WT Col-0 seeds (Figure 3.18).

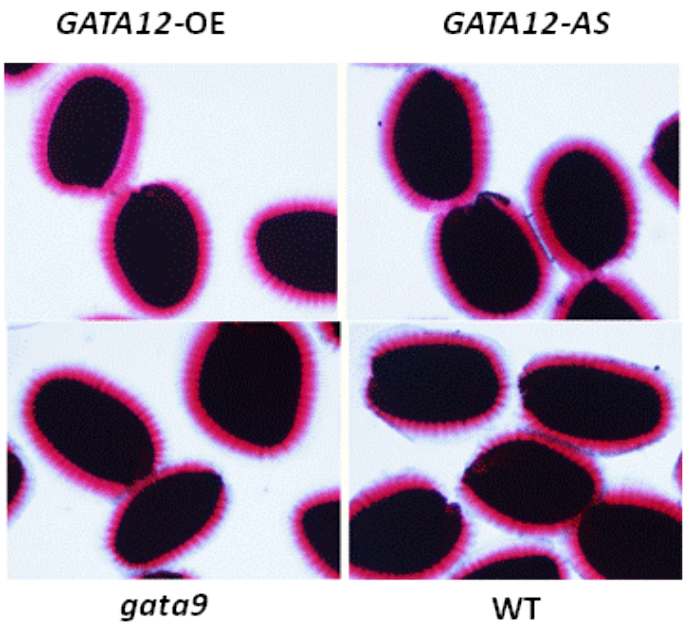


Figure 3.18: Comparison of seed size and mucilage extrusion between GATA12 transgenics, *gata9* mutant and WT

Ruthenium red staining was done to check seed size and mucilage extrusion for *GATA12* transgenics and *gata9* mutants in comparison to WT Col-0 seeds (n=2).

We then wanted to check if GATA9 is redundant in function with GATA12. So went on to analyze the double mutant *gata9 GATA12-AS* by crossing *GATA12-AS* and *gata9* plants. The FH T3 seeds of genotyping confirmed plants (Figure 3.19A) were used in germination assays on 1X MS plates without cold stratification. However, there was no significant enhancement of the GATA12-AS phenotype with additional gata9 knockdown (*GATA12-AS gata9*) (Figure 3.19B). Thus, GATA9 is not redundant in function to GATA12, consistent with the prediction by Manfield et al. (2007).

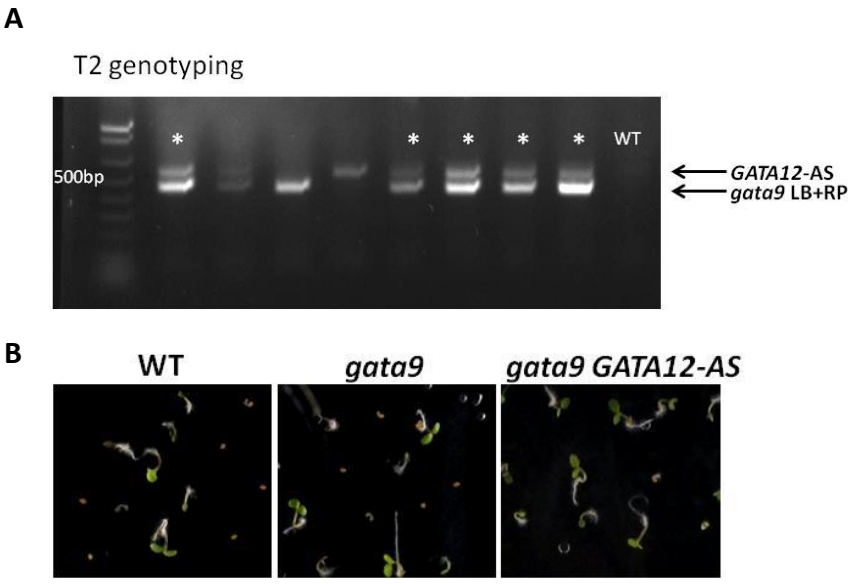


Figure 3.19: GATA9 may not be redundant in function to GATA12
(A) Genotyping of *gata9 GATA12-AS* T2 plants. **(B)** T3 seeds were sown on MS plates with 1% sucrose and 0.5% gelrite. The *gata9 GATA12-AS* shows a better germination percentage in comparison to *gata9* seeds, but no significant enhancement from *GATA12-AS* seed germination. Thus, GATA9 may not be redundant in function to GATA12.

3.2 STUDY OF THE REGULATION OF *GATA12* BY GIBBERELLINS AND RGL2

Since the levels of *GATA12* and *GATA9* were quite high in dormant seeds, we decided to check if the phytohormones GA or ABA, that have prominent roles in seed germination and dormancy respectively, have any role in regulating the GATA expression levels.

3.2.1 Regulation by GA

This was studied using *ga1* (Col-0) mutant seeds, where GA biosynthesis is blocked. RNA was extracted from *ga1* seeds cold stratified with exogenous GA_3 and mock solutions for 12 h and also from WT seeds mock treated for 12h. qRT-PCR analysis showed that GATA12 levels were very high in *ga1*+mock, but reduced drastically with *ga1*+GA_3 showing similar levels as in WT+ mock treatments (Figure 3.20A). Also, *GATA12* transcript levels were measured in WT-Ler and *ga1-3* FH dry seeds and it was found that the transcripts were highly abundant in dry seeds of *ga1-3* mutant in comparison to that in WT seeds (Figure 3.20B). Additional DELLA mutations reduced this drastic difference. This experiment was conducted on freshly harvested (no storage) seeds. Also, motif analysis of promoter region (2kb upstream) of *GATA12* revealed binding sites for GAMYB transcription factor, a GA signalling component (Figure 3.20C).

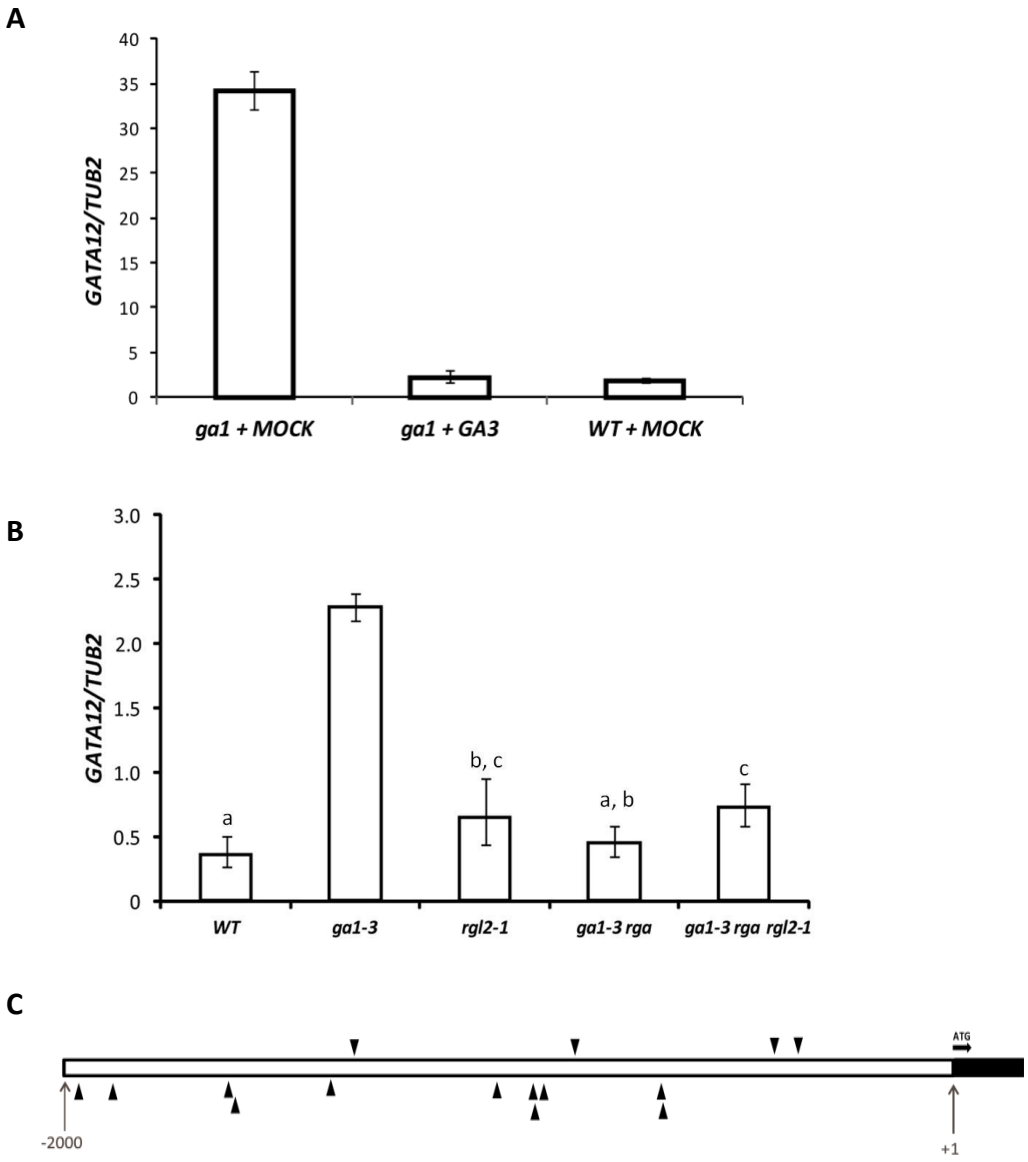


Figure 3.20 *GATA12* is negatively regulated by GA at the transcriptional level in DELLA-dependant fashion

(A) qRT-PCR analysis of cDNA from *ga1* mutant seeds treated with GA_3 and mock solutions for 12h. *GATA12* transcripts are present in high levels in *ga1* non-germinating (mock-treated) seeds when compared to ga1+GA3 seeds (t-test, $P<0.05$, n=3). WT seeds treated with mock solution served as a control. **(B)** *GATA12* transcript levels are abundant in FH dry seeds of *ga1-3* than in WT Ler seeds. The bars marked with the same alphabets are statistically insignificant. The rest of the samples show statistically significant difference (t-test, $P<0.05$, n=2). **(C)** Promoter motif analysis of *GATA12pro* (2kb upstream of CDS) shows sites of binding for GAMYB transcription factors.

There is a drop in *GATA12* and *GATA9* transcripts during cold stratification as shown earlier (Figure 3.17A). This was attributed to the rise in endogenous GA levels under this condition. Thus, we tested the levels of GA biosynthetic genes *GA3ox1* and *GA20ox* in response to cold treatment and imbibition of seeds and found that they rise with cold stratification as shown by Yamauchi et al. (2004) (Figure 3.21).

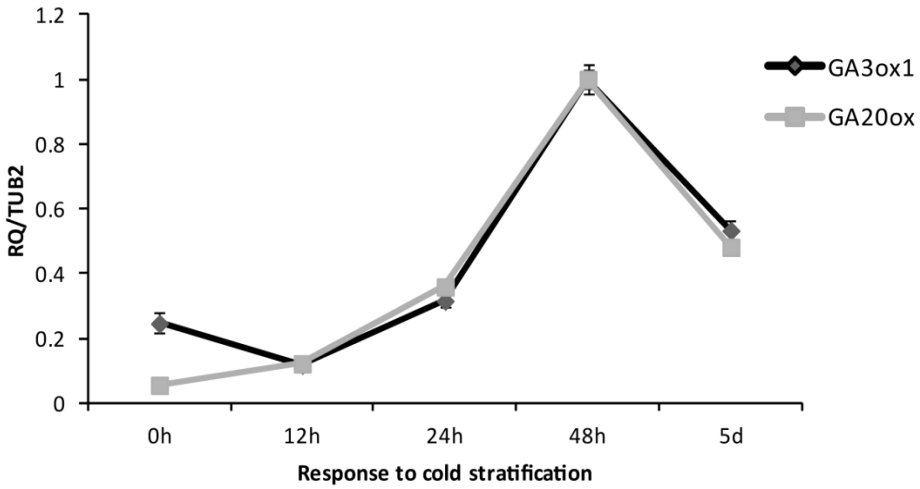


Figure 3.21: *GATA12* transcript levels drop during cold stratification due to rise in transcription of GA biosynthetic genes

Transcript abundance of GA biosynthetic genes, *GA3ox1* and *GA20ox* during different time points of cold stratification in the dark (n=1). Error bars denote SD.

Thus, *GATA12* was demonstrated to be negatively regulated by gibberellins at the transcriptional level. To further understand the regulation of GATA12 by gibberellins, a genetic cross of *ga1* mutant and *GATA12-AS* line was done to obtain a *ga1 GATA12-AS* transgenic lines. Mature pollens from *GATA12-AS* line were transferred onto the stigma of an emasculated bud on *ga1* plants. T1 plants were selected using BASTA as the antisense lines were BASTA-resistant. Surviving plants when taken to T2 generation were still heterozygous for the *ga1* insertion mutant (Figure 3.22A). T3 plants segregated to give to transgenic plants that were homozygous for *ga1* insertion and also contained the 35S:*GATA12*-AS construct (Figure 3.22B). Seeds from these plants were used in germination assays to check if *GATA12* suppression can revert the non-germination phenotype of *ga1* seeds. Although there was no reversion of *ga1* phenotype, we found that germination of *ga1 GATA12-AS* seeds was more sensitive to lower concentrations of exogenous GA3 (2.5µM and 5µM) than *ga1* seeds (Figure 3.23).

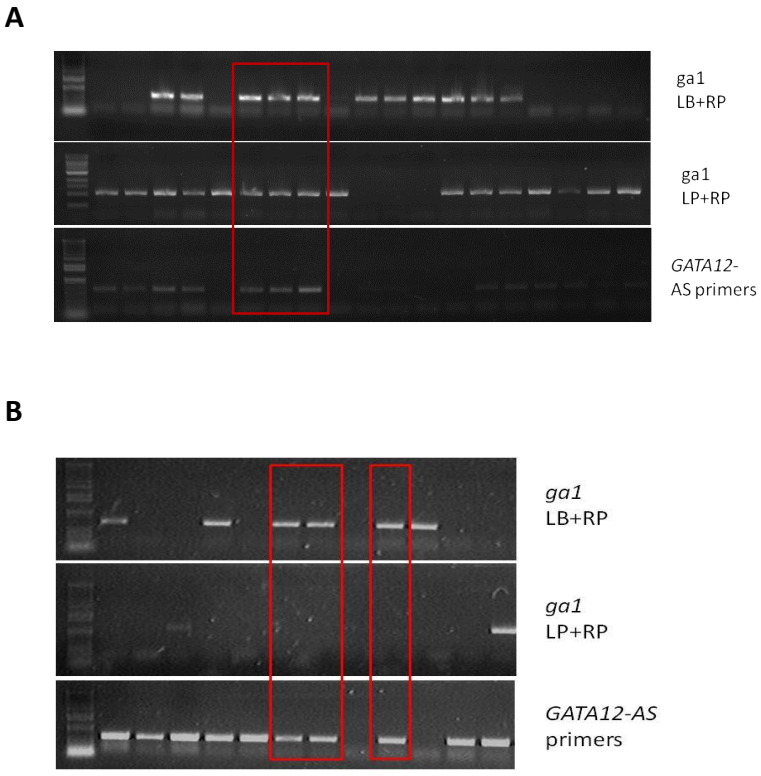


Figure 3.22: Genotyping of *ga1 GATA12-AS* lines

(A) Genotyping of T2 plants; plants containing the antisense construct were heterozygous for the *ga1* insertion **(B)** Genotyping of T3 plants; plants containing the antisense construct were homozygous for the *ga1* insertion (*ga1* LP+RP primers show no amplification). Seeds from these confirmed (highlighted in red box) plants will be used for future germination assays.

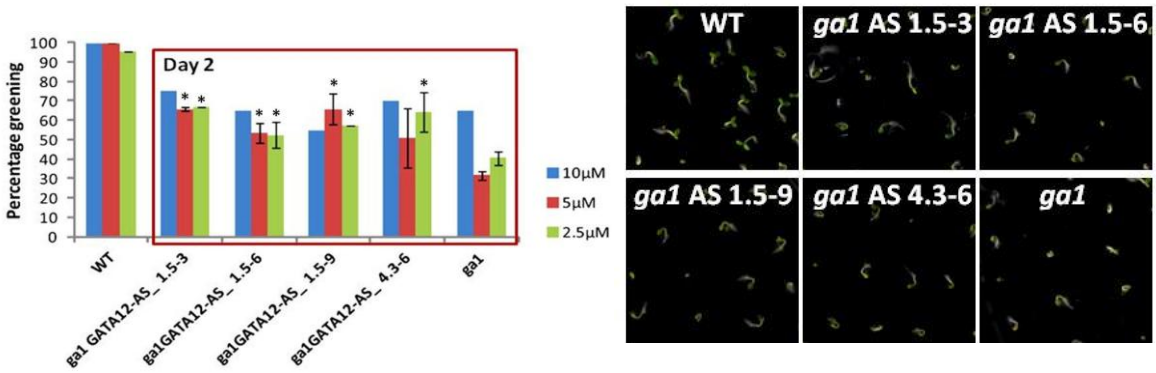


Figure 3.23: *ga1 GATA12-AS* seeds are more sensitive to exogenous GA treatment than *ga1* seeds.

Germination assay was conducted using *ga1 GATA12-AS* **T3** seeds with *ga1* seeds used as control. Although the *ga1 GATA12-AS* mutants were not able to revert back the non-germination phenotype of *ga1* seeds, we found that the double mutants were more sensitive to low concentrations (2.5µM and 5µM) of exogenous GA3 than *ga1* seeds (t-test, *P<0.05 in comparison to *ga1*, n=2 for 2.5µM and 5µM; for 10µM, n=1). Error bars denote SD.

Since GATA12 seems to play a role in release of seed dormancy, and is regulated by gibberellins, we decided to check if GATA12 has any role in regulating ABA or GA biosynthesis. This was studied by analysing the levels of endogenous ABA and GA_3 in transgenic seeds (*GATA12*-AS and *GATA12*-OE) in comparison to WT seeds using LC-ESI-MS approach. The levels of GA_3 were negligible in the seeds and could be detected well. However, we found no significant differences in the levels of ABA as well (Figure 3.24), thereby concluding that GATA12 does not affect ABA biosynthesis in freshly harvested mature seeds.

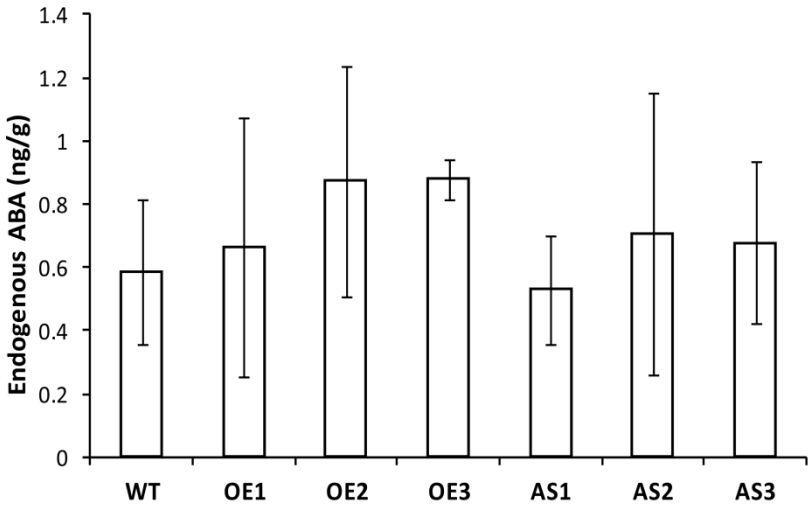


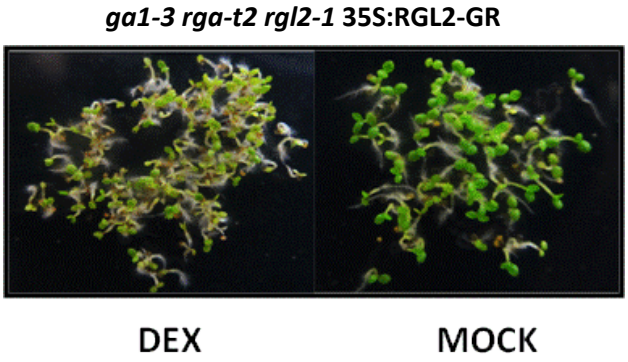
Figure 3.24: Measurement of endogenous ABA levels in GATA transgenic seeds using LC-ESI-MS.

LC-ESI-MS approach was used to measure levels of endogenous ABA in wild type, *GATA12*-OE and *GATA12*-AS seeds. No significant differences were observed (t-test, P>0.05, n=2). Error bars represent SD.

3.2.2 Regulation by RGL2

Since *GATA12* was picked up from the microarray as being differentially regulated by RGL2, we decided to check the immediate response of *GATA12* and *GATA9* transcripts to RGL2 protein. We used *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR seeds to study the immediate effect on *GATA12* transcripts on induction of a functional RGL2. Before carrying out the experiment, we first analysed if the *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR transgenic seeds were functional i.e. whether RGL2 protein is translocated to the nucleus upon dexamethasone (DEX) treatment. The transgenic seeds were cold-stratified in mock and DEX solutions for 3 days and then sown on 1X MS plates. It was observed the germination rate of DEX treated seeds were delayed/slowed due to the inhibitory effect of a functional RGL2, thereby confirming its functionality (Figure 3.25A). In addition, *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR seeds were imbibed in mock and DEX solutions for different time points and RNA was extracted from them. qRT-PCR analysis was done to check the *RGL2* transcript levels, and it was found they increased with DEX treatment, most likely as a result of positive feedback through ABA (Figure 3.25B). Similar treatment was done to *ga1-3 rga-t2 rgl2-1* seeds that showed no rise in *RGL2* levels thus serving as a negative control. These experiments validated the use of *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR seeds for study of immediate response of *GATA12* transcripts to a functional RGL2 protein.

A



B

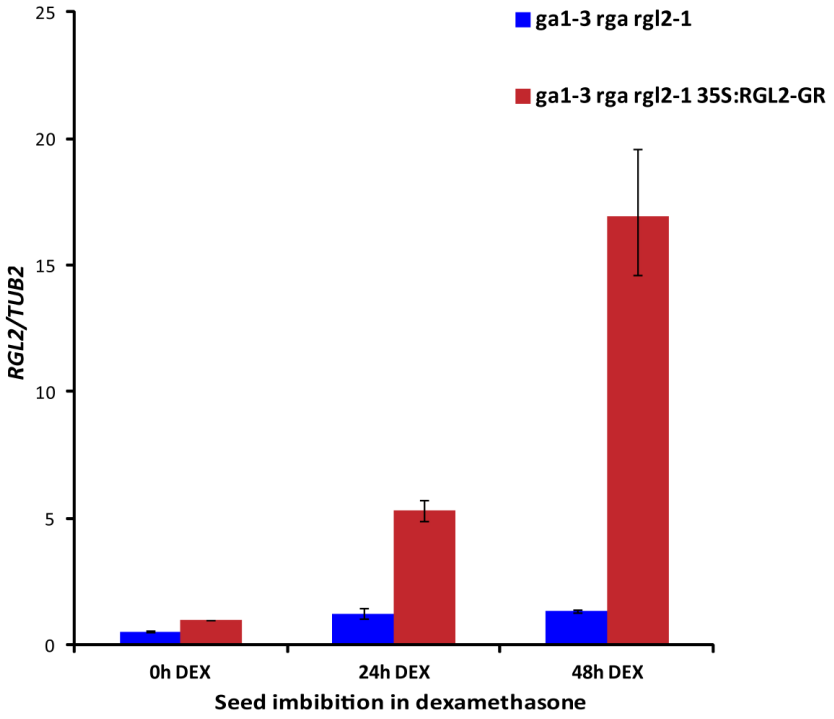


Figure 3.25: DEX treatment of *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR seeds allows RGL2 translocation into nucleus that causes a rise in *RGL2* transcript levels by positive feedback.

(A) Seeds of ga1-3 rga-t2 rgl2-1 35S:RGL2-GR plants were cold stratified in mock and DEX solutions and sown on 1XMS plate. DEX treated seeds, owing to the translocation of RGL2 to nucleus, showed a poor germination rate than mock treated seeds (n=2) **(B)** ga1-3 rga-t2 rgl2-1 35S:RGL2-GR seeds imbibed in DEX show a rise in *RGL2* transcript levels with time, by virtue of a positive feedback loop by ABA (n=1). Error bars denote SD.

Freshly harvested *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR seeds were imbibed in mock and DEX solutions for 4h, 8h and 12h and the *GATA12* transcript regulation was studied by qRT-PCR analysis. Although the microarray was done using *ga1-3 rga-t2 rgl2-1* and *ga1-3 rga-t2* seeds imbibed for 5 days, and *GATA12* transcripts were found to be down-regulated in response to a functional RGL2 (Figure 3.1), the above experiment showed that they are up-regulated as an immediate response (4h,8h) to nuclear-localized (functional) RGL2 (Figure 3.26). However, stratification for a longer time point (12h) showed a mild repression of *GATA12* transcripts by RGL2. This could be a compensatory/feedback mechanism exhibited by RGL2, most likely to avoid excessive growth suppression in seeds.

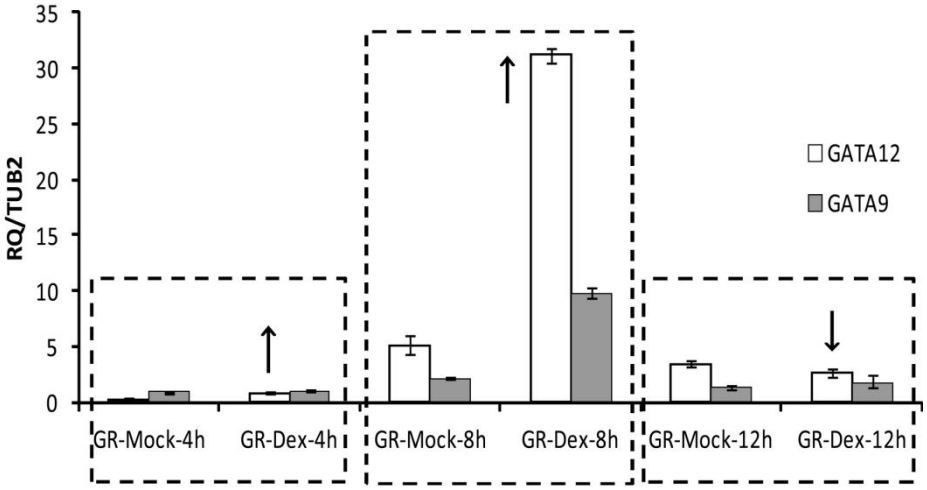


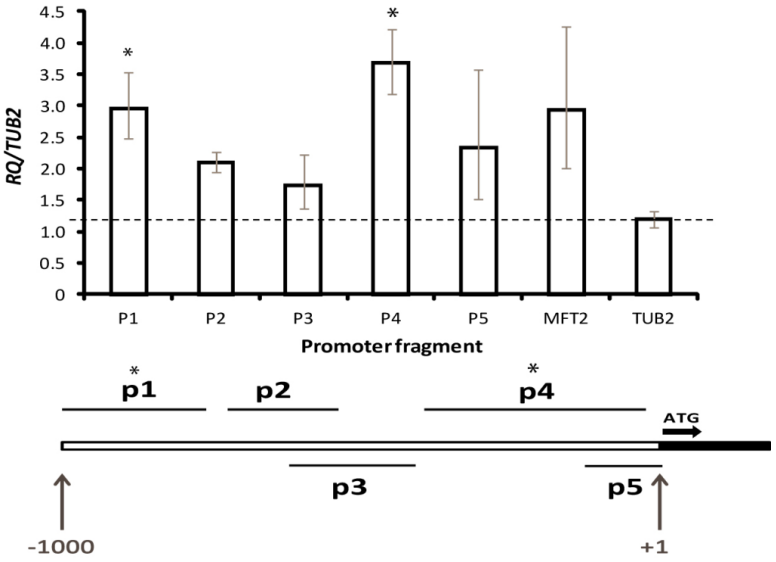
Figure 3.26: Regulation of *GATA12* and *GATA9* transcription by RGL2 at different time points.

Regulation of *GATA12* and *GATA9* transcripts by RGL2 at different time points was studied by treating *ga1-3 rga-t2 rgl2*-1 35S::RGL2-GR seeds with DEX and mock solutions for 4h, 8h and 12h. The graph below shows the positive regulation of *GATA12* by RGL2 in the earlier time points of 4h and 8h (t-test, P>0.05 for DEX treatment in comparison to mock, n=3). However, stratification for longer time point (12h) shows a mild repression of *GATA12* transcripts by RGL2. This could be a compensatory/feedback mechanism exhibited by RGL2, most likely to avoid excessive growth suppression in seeds. Error bars denote SD.

Since *GATA12* is regulated by RGL2, we decided to check if this regulation was direct. A chromatin immunoprecipitation (ChIP) assay was done using *ga1-3 rgl2-1* 35S:RGL2-6HA seeds. Single knockout of RGL2 in *ga1-3* background can restore germination after prolonged chilling , like *ga1-3 rga-t2 rgl2-1* seeds (Cao et al. 2005). The reason for choosing these seeds as the starting material instead of *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR seeds was due to the specificity of anti-HA antibody in pull down of protein over anti-GR antibodies.

WT seeds were used as negative control as they lack a HA tag. *In vivo* DNA-protein cross linking was done using formaldehyde. In addition, we also included a DSG treatment prior to formaldehyde treatment to allow protein-protein cross linking since RGL2 cannot directly bind to DNA and does so through protein complexes. Chromatin was fragmented by sonication and the efficiency of sonication was checked by running the sheared chromatin on a 1% agarose gel. The gel picture in Figure 3.27B shows that the size of sheared fragments varied between 200bp and 1kb, with maximum fragments in the 250-300bp range. The RGL2-protein complex bound DNA that was pulled down was then analysed by qRT-PCR, with ACTIN serving as endogenous control. A fragment was considered to be enriched if its levels were ~3 fold higher than that of TUB2 fragment. Fragments P1 and P4 were found to be enriched and thus hypothesized to contain sites for RGL2 complex binding (Figure 3.27A).

A



B

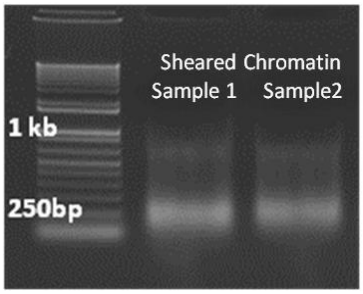


Figure 3.27: ChIP assay showing direct regulation of *GATA12* by RGL2 protein complex

(A) ChIP enrichment test showing the binding of RGL2-6HA to the *GATA12* promoter. The 1kb upstream region (+1 to -1000) is represented by a white box, while the first exon is represented by a black box in the lower panel. The solid lines represent the position of the amplified fragments. The upper panel shows the qRT-PCR results, with asterisks representing promoter fragments that show ~**3 fold** or more significant enrichment over *TUB2* fragment (t-test, *P<0.05, n=3). Amplification of *MFT2* promoter fragment was used as positive control. Error bars denote SD. **(B)** Gel picture showing sonication efficiency of two technical replicates. Chromatin was sonicated at 7cycle 10s pulses into fragments of size ranging from 200bp to 1kb, with most fragments in 250-300bp size.

Earlier study in our laboratory by Stamm et al. (2012) showed that promoters of genes regulated by RGL2 protein were rich in GARE and Dof motifs. Thus, we did a promoter motif analysis of *GATA12* promoter region to identify the various motifs in them. And indeed we found that the fragments 1 and 4 are enriched in Dof motifs, which could be the likely binding regions for the RGL2 complex (Figure 3.28).

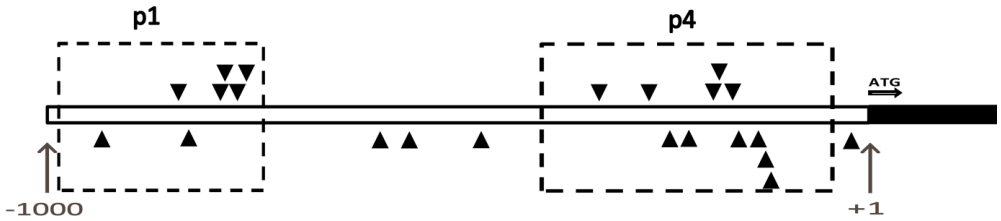


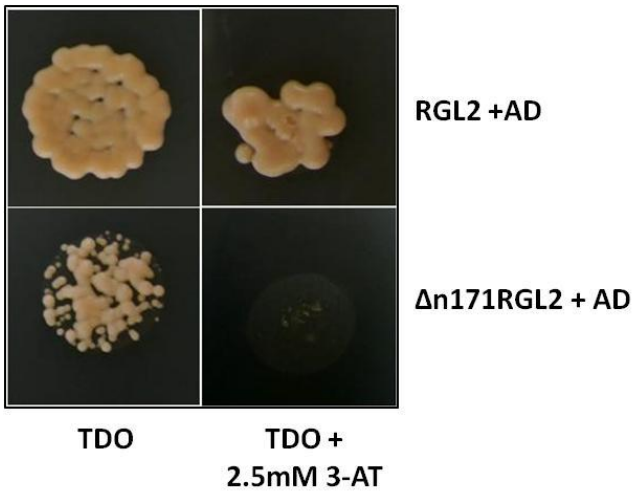
Figure 3.28: Dof motifs are enriched in *GATA12* promoter region

An analysis of various motifs present in *GATA12* promoter (1 kb upstream of CDS) showed that the fragments P1 and P4, that were enriched in the ChIP assay, were highly clustered with Dof motifs.

3.3 YEAST LIBRARY SCREENING TO IDENTIFY PROTEIN PARTNERS IN RGL2 COMPLEX

Since RGL2 is known to have no DNA binding domain, it is highly likely that it regulates it downstream targets through other proteins or transcription factors. Thus, we decided to identify the protein(s) present in the RGL2 complex through an yeast library screening. The screening was performed using Matchmaker® Gold Yeast Two-Hybrid Library screen. We first cloned RGL2 protein into pGBKT7 DNA-BD cloning vector. But it showed self activation with empty pGADT7 AD vector (Figure 3.29A). So we went on to make a deletion construct of RGL2 protein that lacks the first 171 residues in the N-terminal (referred to as RGL2Δn171). This deletion of the N-terminal DELLA domain reduced self-activation dramatically when plated on TDO+2.5mM 3-AT (SD/‑Ade/‑Leu/‑Trp + 2.5mM 3-AT) plates (Figure 3.29A). This construct was then used for further screening assays (construct was cloned by Petra Stamm). Y2HGold Yeast Strain was transformed with RGL2Δn171-pGBKT7 and allowed to mate with Mate&Plate Library-Universal *Arabidopsis*. The diploids were then plated and selected on TDO+10mM 3-AT plates. The selected colonies were then re-plated/selected on QDO + X-α-GAL plates (Figure 3.29B). Surviving blue colonies were sequenced to identify the putative interacting proteins. Around 30 putative interacting proteins were identified (Table 3.2). However, these interactions require further confirmation through yeast-two hybrid assay.

A



B

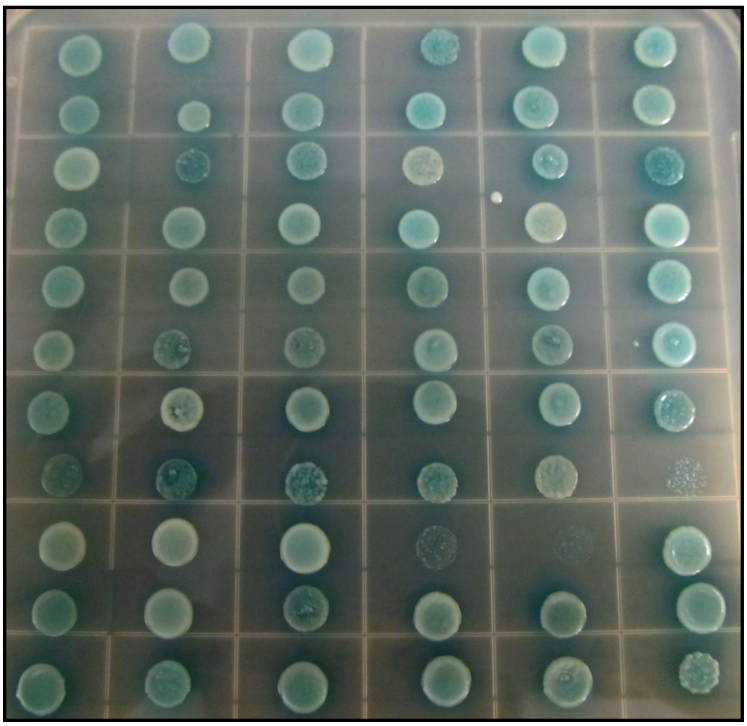


Figure 3.29: Yeast library screening to identify protein partners of RGL2.

(A) Check of self-activation: RGL2 shows self-activation with empty AD vector on TDO+2.5mM 3-AT. A deletion of 171 residues in N-terminal abolished this self-activation. **(B)** Colonies selected on QDO + X-α-GAL were sequenced to identify putative protein partners.

Table 3.2: List of protein partners of RGL2 identified by yeast library screening

The tissue-specific expression of the putative interacting proteins was determined from the *Arabidopsis* eFP browser ('Developmental map' and 'seed' data source).

Name of protein	**TAIR accession number**	**No. of hits**	**Putative expression region**
PLATZ transcription factor family protein, **DNA-binding**	AT2G27930	2	
F-box family protein-like protein	AT3G20620	1	mature pollens, siliques, during embryogenesis
hypothetical protein	AT2G39950	1	mature pollen, secondary dormant seeds
protein phosphatase 2A B55 alpha	AT1G51690	1	
Alba **DNA**/RNA-**binding** protein	AT1G20220	1	
hypothetical protein	AT1G23970	1	dry seeds
auxin transport protein BIG, **DNA-binding**	AT3G02260	2	mature embryos
RING/U-box domain-containing protein, **DNA binding**	AT1G63840	2	
shewenella-like protein phosphatase 1	AT1G07010	1	mature embryos
homeobox-leucine zipper protein HAT2, **DNA-binding**	AT5G47370	1	stamens, chalazal seed coat
octicosapeptide/Phox/Bem1p domain-containing protein	AT5G49920	1	
global transcription factor group A2	AT4G08350	1	Shoot apex, endosperm
leucine-rich repeat-containing protein, **DNA-binding**	AT5G21090	1	internode, seeds
serine-arginine-aspartate-rich protein SHW1, **DNA-binding**	AT1G69935	1	dry seeds, leaves
chaperone protein dnaJ 11	AT4G36040	1	Mature embryos, dry seeds, senescent leaves
protein TIFY 10A (JAZ1), **DNA binding**	AT1G19180	1	mature embryos, stamens, petals
SKU5 similar 6	AT1G41830	1	Imbibed seeds
tetraspanin family protein	AT5G46700	2	Shoot apex

Rae1-like protein	AT1G80670	1	Shoot apex, seeds
Lipin-like protein	AT3G09560	1	Dry seeds
cysteine proteinase RD21a	AT1G47128	2	water-Imbibed seeds
translation initiation factor eIF-4E	AT4G18040	1	Germinated embryo, shoot apex
30S ribosomal protein S9	AT1G74970	1	
aluminum induced protein with YGL and LRDR motif	AT3G22850	2	stamens, endosperm
selenium-binding protein 2	AT4G14040	1	Mature embryos
succinate dehydrogenase subunit 4	AT2G46505	1	Imbibed seeds
basic region/leucine zipper motif 53 protein (BZIP53), DNA-binding	AT3G62420	1	Mature embryos
NADH dehydrogenase (ubiquinone) flavoprotein 1	AT5G08530	1	Imbibed seeds, young siliques, shoot apex, reproductive organs
putative NDP-L-rhamnose synthase	AT1G53500	1	Mature pollen, young siliques, seed coat
Cyclin/Brf1-like TBP-binding protein	AT3G09360	1	embryogenesis

Identification of specific protein partners that interact/complex with RGL2 will help to further widen our knowledge about the downstream mechanism of RGL2. Also, identifying which of these protein partners is able to regulate *GATA12* directly will help us to understand how RGL2 complex is able to bind to *GATA12* promoter motif and regulate it.

Our study has identified a novel downstream target of RGL2 that helps to maintain dormancy in seeds. We have shown that this gene is downstream of GA signalling and directly regulated by the RGL2 complex. *GATA12* is highly expressed in freshly harvested dormant seeds. qRT-PCR studies have shown

that its expression drops under favourable conditions of germination, like during stratification and after-ripening in response to gibberellins, thereby releasing dormancy and allowing seeds to germinate. Histochemical GUS staining of *GATA12pro:GUS* seeds have confirmed that its expression is high in embryonic radicle of dry seeds and drops in that of germinated seeds. Also, *GATA12* levels have been shown to rise during embryogenesis and peak in fully mature embryos. Phenotypic analysis of *GATA12* transgenic seeds has further confirmed their role in maintaining dormancy. Thus, GATA12 controls the release of primary dormancy in seeds downstream of GA and RGL2.

4. DISCUSSION

4.1 GATA12 functions to maintain primary dormancy in seeds

This study was initiated to identify downstream targets of RGL2 and thereby understand a key step of GA signalling pathway in regulating seed dormancy. RGL2 has been shown to be the most important of the DELLA proteins (RGA, GAI, RGL1, RGL2, RGL3) in repressing GA-mediated seed germination (Lee et al. 2002). An earlier microarray analysis of the transcriptomes of *ga1-3 rga-t2 rgl2-1* and *ga1-3 rga-t2* seeds (stratified for 5 days) revealed several genes to be differentially regulated by a functional RGL2 protein (Stamm et al. 2012). A gibberellin-deficient (*ga1-3*) background was selected since active gibberellins degrade RGL2 protein (Tyler et al. 2004, Murase et al. 2008). RGA is known to have minor roles in suppressing germination similar to RGL2; the knockout of RGA in both genotypes served to ensure that the results obtained were specific to the function of RGL2. qRT-PCR analysis was done to verify the differential regulation of several genes (Figure 3.1). Some of the genes verified to be down-regulated include members of the Expansin family (*AtEXPA1,3,8,9,11,14,15*), *Cytochrome P450's* like *CYP707A*s that encode (+)-Abscisic Acid 8′-Hydroxylases (*CYP707A2*, *CYP707A3*) and several other transcription factors. RGL2, although considered a transcriptional regulator, is known to have no DNA binding domain (Peng and Harberd 1997, Silverstone et al. 1998). Hence, any downstream regulation must be through other proteins, most likely transcription factors, that complex with RGL2. Of the several transcription factors verified for differential regulation as part of our

study, we chose a gene encoding a GATA-type zinc finger transcription factor (GATA12) for further molecular and phenotypic assays.

Why did we choose to understand the molecular function of this GATA-type zinc finger protein? GATA factors are type-IV zinc finger proteins named so due to their affinity for promoter elements containing this sequence. The *Arabidopsis* GATA-type family is a large family, consisting of 30 members. Most of the members of this family that have been characterized (8 out of 30 family members) have roles in embryogenesis or seed germination (Table 3.1). Also, since public microarray data showed high expression of *GATA12* in mature dormant embryos (Figure 3.2), we decided it would be a good candidate to understand the downstream mechanism of RGL2 in negatively regulating seed germination.

Reyes et al. (2004) performed a phylogenetic analysis of GATA factors and classified them into four sub-families based on the position of the zinc finger domains (Figure 3.3B). The neighbour-joining tree they generated revealed GATA12 to be a member of sub-family I with GATA9 as its closest homolog (Figure 3.5B). Thus, GATA9 was included in some of the molecular characterization studies. The other two closest homologs are GATA2 and GATA4 (Figure 3.5A and B).

The high expression of *GATA12* and *GATA9* transcripts in freshly harvested dry seeds (Figure 3.6 A and B) in comparison with other tissues suggests that these genes are likely to play roles in maintaining seed dormancy since freshly harvested mature seeds are known to be highly dormant (Karssen et

al. 1983, Holdsworth et al. 2008, Nakabayashi et al. 2012). Some of the plant GATA transcription factors have been implicated to have roles in light-regulated gene expression (Manfield et al. 2007). Although *GATA12* and *GATA9* transcripts were shown to be expressed at higher levels in the dark-grown (etiolated) seedlings when compared to seedlings drown in constant light (Figure 3.7), Manfield et al. (2007) have suggested that these two genes do not show significant co-expression with any of the light-signalling transcription factors unlike *GATA2* and *GATA4.* Although the four gene sequences are closely related, it is likely that *GATA12* and *GATA9* have diverged from *GATA2* and *GATA4* in the function of light-regulation. Also, the authors show that *GATA12* and *GATA9* have also diverged sufficiently in function that they potentially regulate different sets of genes. Thus, they predicted that *GATA12* and *GATA9* may not show functional redundancy.

Stamm et al. (2012) did a motif analysis of promoters (−1,000 to +200 nt relative to TSS) of genes regulated by RGL2 and found that they were enriched in Dof motifs. Using in vivo assay, the authors have shown that transcriptional activation by RGL2 involves promoter motifs bound by Dof-type transcription factors. An analysis of the promoter (-1000 to +1) of *GATA12* revealed that it was rich in Dof-associated motifs as seen in Figure 3.28. The -1000 to -2000 region contained very few Dof motifs. Thus, we decided to use the highly enriched 1kb upstream region for driving the *GUS* reporter gene expression. Histochemical GUS staining analysis of *GATA12pro*: GUS plants further confirmed the high expression of *GATA12* in embryonic

radicle of dry seeds in comparison to that of germinated seeds (Figure 3.8A and B). A similar GUS expression pattern was also observed in pollen grains. The mature pollen grains were highly stained whereas the germinated pollen grains with pollen tubes showed mild staining (Figure 3.8G). In *Arabidopsis*, mature pollen grains are released from anthers in a dormant state. Once they fall onto stigma rehydration of pollen grains breaks dormancy and turns on pollen germination (McCormick 2004). This observation further strengthened our claim that *GATA12* is predominantly expressed in dormant conditions.

The protein encoding the gene *GATA12* was shown to be nuclear localised (Figure 3.9) by the transient expression of *35S:GATA12-GFP* in *Arabidopsis* mesophyll protoplasts (TEAMP). This supports the view that GATA12 is a transcription factor.

Dormancy is instilled in the middle stage of embryogenesis (between 6dpa to 9dpa) and peaks in the fully mature dry seeds (Kanno et al. 2010). Siliques of different developmental stages should possess embryos at different developmental stages corresponding to progressively increasing levels of dormancy. Hence, our analysis of the expression pattern in various stages of siliques that revealed the rise in *GATA12* and *GATA9* transcripts during embryogenesis with its levels peaking in fully mature dormant embryos (Figure 3.10A and B) helped to further corroborate our hypothesis. This was also confirmed by detection of *GATA12* transcripts using RNA probes in mature siliques by RNA *in situ* hybridization studies (Figure 3.10C). From all the above results put together, it was thus concluded that *GATA12* levels are

high under dormant conditions but drop under germination conditions in seeds to allow the release of dormancy.

To understand the biological function of *GATA12*, we generated suppression and ectopic expression transgenic lines. T-DNA insertion mutants from ABRC (SALK_052546C and SALK_012501C) lacked an insertion in the exon (Figure 3.11A) and showed poor suppression of *GATA12* (Figure 3.12E). To address this shortcoming, we generated anti-sense lines showing significant suppression, to study the effects of suppression of *GATA12.* Transgenic lines showing significant overexpression and suppression of *GATA12* transcripts were chosen for phenotypic analysis (Figure 3.13 A and B). However for *GATA9,* SALK_152156C line having a T-DNA insertion in its exon was retained for phenotypic analysis (Figure 3.11B) as it showed severe suppression of *GATA9* transcripts (Figure 3.13D). Subsequently, these lines were used for seed germination assays. Although the mature transgenic plants showed no phenotypic differences in comparison to WT plants (Figure 3.14), they did show differences at the germination level. Consistent with our earlier conclusion that GATA12 may play roles in maintaining seed dormancy, *GATA12*-OE lines showed a lesser germination percentage that WT and *GATA12*-AS lines showed a higher germination rate and germination percentage than that of WT seeds (Figure 3.15A and B). As observed in Figure 3.16A, the percentage of seeds showing radicle protrusion is higher on day1 for seeds of *GATA12*-AS lines, while almost negligible for *GATA12*-OE lines. Also, *GATA12*-AS reach almost 100% germination on day3 whereas *GATA12*-

OE seeds show only 30-60% overall germination. However, this phenotypic difference was observed only with freshly harvested seeds that were directly sown on MS plates without being subjected to cold stratification. Dry after-ripening of seeds for a month brought WT germination levels closer to that of *GATA12*-AS seeds, especially on day2 and day3 (Figure 3.16B). Long-term dry storage of seeds abolished such phenotypic differences. The down regulation of *GATA12* during cold stratification and dry-after ripening are plausible explanations for the loss/reduction in phenotypic differences under these conditions (Figure 3.17A and B). Like *GATA12*, the *GATA9* transcripts also drop with cold stratification and dry storage of freshly harvested seeds. The *gata9* mutant however showed no differences in phenotype at germination level (Figure 3.15C). This could be due to one of two reasons: redundant roles played by these GATA factors, with one being dominant over the other or the two GATA factors show no functional redundancy and thus do not show similar phenotypes (as predicted by Manfield et al. (2007)).

Thus, the phenotypic analysis of GATA12 has demonstrated that it plays a role in regulating the release of primary dormancy in seeds.

Although dormancy is largely controlled by hormonal signals, several seed parts (e.g. endosperm, seed coat) can contribute to appearance of dormancy (Kucera et al. 2005). During early seed imbibition, the mucilage is extruded from the testa epidermis and forms a pectin hydrogel. Lack of mucilage extrusion is associated with reduced germination under dry conditions, suggesting that mucilage is an important factor for seedling

establishment under adverse conditions (Penfield et al. 2001). Ruthenium red staining of *GATA12* transgenic seeds showed no defects in mucilage extrusion (Figure 3.18), thereby leading us to the hypothesis that these germination differences observed under unstratified conditions cannot be due to defects in seed coat. Instead, it may be due to hormonal control of GATA12 or its downstream targets.

The crossing of *GATA12-AS* and *gata9* was done to check if GATA9 is redundant in function to its homolog, GATA12. The FH T3 seeds of genotyping confirmed plants were used in germination assays on 1X MS plates without cold stratification. However, there was no significant enhancement of the *GATA12-AS* phenotype with additional gata9 knockdown (Figure 3.19B). Thus, GATA9 is not redundant in function to GATA12, consistent with the prediction by Manfield et al. (2007). The authors have shown through an ACT scatter plot analysis for *GATA9* and *GATA12* that there is significant divergence of regulation for these two genes. They predict that *GATA9* and *GATA12* have not only diverged from *GATA2* and *GATA4*, but they are also diverging in expression from each other and would not be predicted to show functional redundancy.

4.2 Hormonal control of *GATA12*: Regulation by GA

Since *GATA12* was identified as a downstream target of RGL2, experiments were conducted to understand its regulation by GA signalling pathway. High expression of *GATA12* in *ga1*+mock seeds in comparison with *ga1*+GA3 or WT+mock seeds suggests that *GATA12* transcripts are negatively regulated by GA signalling in seeds (Figure 3.20A). Also, accumulation of *GATA12* transcripts to higher levels in *ga1-3* dry seeds than WT dry seeds (Figure 3.20B) suggests that GATA12 controls the release of dormancy in mature seeds in a GA-dependant manner.

We tested if DELLA proteins mediate GA regulation of *GATA12* by checking *GATA12* expression in *rga-t2* and *rgl2-1* mutants in the *ga1-3* background. Loss of activity of two DELLA proteins, RGA and RGL2, reduced *GATA12* expression to a level comparable to that in the WT (Figure 3.20B). Thus, *GATA12* is regulated by GA in a DELLA-dependant manner. Interestingly, we detected gibberellin response element (GARE)-like motifs in the promoter of *GATA12*. GARE motifs are typically bound by GAMYB-type transcription factors, which respond to GA signalling, and regulate transcription of GA-responsive genes (Gubler et al. 1995, Gubler et al. 1999). The presence of these motifs further supports the hypothesis that *GATA12* is regulated by GA.

GA biosynthesis begins under favourable conditions of light and temperature. Cold stratification causes an increased expression of the GA biosynthesis genes *GA20ox1* (*GIBBERELLIN 20 OXIDASE*), *GA20ox2* and *GA3ox1* and

decreased expression of the GA catabolic gene *GA2ox2* (Yamauchi et al. 2004). The observed increase in *GA20ox2* and *GA3ox1* expression could explain for the decrease in *GATA12* levels upon cold stratification (Figure 3.21 and 3.17A).

T3 seeds of *ga1 GATA12-AS* plants were used in germination assays to check if *GATA12* suppression can revert the non-germination phenotype of *ga1* seeds. Although there was no reversion of *ga1* non-germinating phenotype, the germination of *ga1 GATA12-AS* seeds was more sensitive to lower concentrations of exogenous GA3 (2.5µM and 5µM) than *ga1* seeds (Figure 3.23). This helped to further corroborate that GATA12 is regulated by GA signalling pathway.

In addition to GA, ABA also plays an equally important but antagonistic role in regulating seed germination (Koornneef et al. 2002). However, *GATA12* showed no regulation by ABA, neither do transgenic seeds show any differences in accumulation of endogenous ABA levels (Figure 3.24). Thus, the differences in the levels of dormancy in the transgenic seeds may be attributed to differences in sensitivity to GA. It should be noted that ABA is not the only signal in establishment and dormancy release (Koornneef et al. 2002). Dormancy release is also affected by NCEDs and CYP707As (Cadman et al. 2006), crosstalk with ethylene (Kucera et al. 2005), reactive oxygen species (ROS) (Finkelstein et al. 2008), sugars such as glucose (Finkelstein et al. 2008, Zhu et al. 2009) and also nitrate (Alboresi et al. 2005).

Although *GATA12* was shown to be negatively regulated by RGL2 in the microarray using *ga1-3 rga-t2* and *ga1-3 rga-t2 rgl2-1* seeds stratified for 5 days (Figure 3.1), analysis using dexamethasone (DEX)-treated *ga1-3 rga-t2 rgl2-1* and *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR seeds showed an immediate positive regulation. The glucocorticoid receptor (GR)-tagged RGL2 protein is retained in the cytosol as it complexes with heat-shock proteins (Pratt et al. 2006). It moves into the nucleus when treated with DEX, which is an agonist of the GR. DEX treatment causes the release of heat-shock proteins thereby resulting in an RGL2-receptor-DEX complex that translocates via active transport into the nucleus (Hayashi et al. 2004).

RGL2 is the major DELLA repressor of seed germination (Lee et al. 2002, Tyler et al. 2004, Cao et al. 2005). The translocation of RGL2 into nucleus will thus induce repressive effects on the germination potential of seeds. Thus, the slower germination rate of DEX-treated *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR seeds in comparison to that of *ga1-3 rga-t2 rgl2-1* seeds (Figure 3.25A) helped to confirm the functionality of these seeds. Also, the rise in *RGL2* transcripts with DEX treatment (Figure 3.25B) can be attributed to a positive feedback loop. Among the five DELLA factors present in *Arabidopsis*, RGL2 plays a unique and central role to promote ABA synthesis in seeds. In turn, ABA has been shown to strongly stimulate *RGL2* mRNA levels, but not those of the other DELLA genes, in a positive feedback loop (Piskurewicz et al. 2008, Piskurewicz and Lopez-Molina 2009). Thus, the spike in the levels of *GATA12* and *GATA9* transcripts, on DEX treatment of *ga1-3 rga-t2 rgl2-1* 35S:RGL2-GR

seeds for 4h and 8h, is the result of the presence of a functional RGL2 protein in the nucleus (Figure 3.26). This immediate rise in *GATA12* transcripts in response to RGL2 will help to enforce primary dormancy in seeds and the negative regulation of *GATA12* observed at later time point of 12h (Figure 3.26) and in 5d stratified seeds (Figure 3.17A) may be attributed to a feedback mechanism or a compensatory mechanism exhibited by RGL2, in order to avoid excessive growth suppression.

The pull-down of *GATA12* promoter fragments in a ChIP experiment conducted using *ga1-3 rgl2-1* 35S:RGL2-6XHA seeds revealed that this regulation by RGL2-complex was direct. The fragment 1 and fragment 4 of *GATA12* promoter were enriched in comparison to the *TUB2* fragment, with *ACTIN* used as endogenous control (Figure 3.27A). Fragment 4 showed the highest enrichment of all. Motif analysis of the *GATA12* promoter (1kb upstream) region revealed that these two fragments were highly enriched in Dof (DNA-binding with one finger)-associated motifs (Figure 3.28).

Dof (DNA-binding with one finger) domain proteins are plant-specific transcription factors with a highly conserved DNA-binding domain, with a single C2-C2 zinc finger (Yanagisawa 2002). Dof proteins have diverse roles in gene expression associated with plant-specific phenomena including light, phytohormones and defence responses, seed development and germination. Recent studies revealed two closely related *Arabidopsis Dof* genes, *Dof AFFECTING GERMINATION1 (DAG1)* and *DAG2* to be involved in the control of seed germination, although their actions are antagonistic (Gualberti et al.

2002, Papi et al. 2002). Another Dof protein, Dof6, negatively regulates seed germination by increasing the expression of ABA-related genes (Rueda-Romero et al. 2012). Thus, Dof family of proteins likely play important roles in regulation of GA-mediated germination, which may include *GATA12* regulation implied in the observation presented above.

Interestingly, a bioinformatics analysis of the promoters of genes identified to be up-regulated by RGL2 has revealed an enrichment in Dof-associated cis-elements (Stamm et al. 2012). Two of the Dof motifs, an AAAAG element and a TAAAG element, occurred in 42 and 37%, respectively, of the promoters analysed by them. The authors have thus speculated that Dof-associated motifs could indeed be binding sites for RGL2-mediated regulation of transcription. Their suggested mechanism is that RGL2 either interacts with Dof-type transcription factors to regulate downstream genes or that RGL2 binds to sequester the competitors or inhibitors of Dof proteins, thus allowing Dof proteins to bind to their target promoters and induce transcription. Everything considered, the enrichment of Dof-associated motifs in our ChIP-enriched *GATA12pro* fragments is of biological significance.

All the above data put together, we can conclude that GATA12 plays a role in controlling the release of primary dormancy in seeds and functions downstream of GA and RGL2. The proposed model of seed germination/dormancy mediated by GATA12 is shown in Figure 4.1.

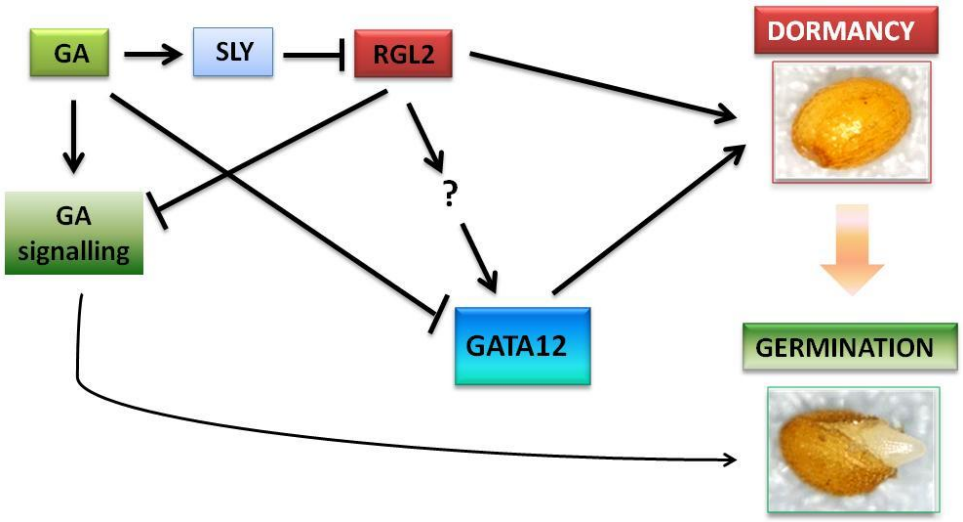


Figure 4.1: Proposed model of dormancy release mediated by GATA12

GATA12 functions downstream of GA signalling pathway. It is negatively regulated by GA at the transcriptional level. Also, it is directly regulated by an RGL2-containing protein complex that can bind to its promoter. RGL2 up-regulates *GATA12* transcripts as an immediate response in order to retain primary dormancy in seeds, until favorable conditions for germination set it. Thus, GATA12 controls the release of seed dormancy and negatively regulates seed germination

4.3 Investigating the constituents of the RGL2 complex

RGL2, although referred to as a transcription factor, has no known DNA binding domain (Peng and Harberd 1997, Silverstone et al. 1998). So far, we have only been able to hypothesize that it functions by binding to other transcription factors, one example being the above mentioned Dof proteins. Identification of the protein partners in the RGL2 complex will aid to widen our knowledge of how RGL2 mediates the downstream regulation of its target genes. The yeast library screening was initiated for this purpose. Auto-activation of full-length RGL2 with empty PGADT7 AD vector led us to design a deletion construct of RGL2 (Figure 3.29A). The RGL2Δn171 deletion protein lacks the N-terminal conserved motifs DELLA and VHYNP. The DELLA motif functions mainly for gibberellin-induced recognition and degradation of the protein (Dill et al. 2001). The suppressive function of DELLA proteins depends on the C-terminal region, which contains the VHIID and RVER domains. This is supported by the fact that the deleted proteins missing the N-terminal regions (ΔDELLA, ΔVHYNP), as shown for rice SLR1, have a constitutive suppressive function (Itoh et al. 2002) The importance of the C-terminal half to the suppression activity has also been described for GAI/RGA (Peng and Harberd 1993, Peng et al. 1997, Silverstone et al. 1998). Thus, the library screening was performed with RGL2Δn171 deletion construct. We identified several putative interactors (Table 3.2), like transcription factors, enzymatic protein and proteins that function in hormonal pathways. The reliability of this screen is shown by the identification of JA ZIM-domain 1 (JAZ1) protein, a key repressor of JA

signalling, as one of the putative interacting proteins. Hou et al. (2010) have shown that JAZ1 can interact in vivo with DELLA proteins. Further verification of the interaction of the other identified proteins can help to discover novel pathways and hormonal integrations downstream of RGL2. Also, this screen can be used to identify the protein(s) in the RGL2 complex that help to directly regulate *GATA12.*

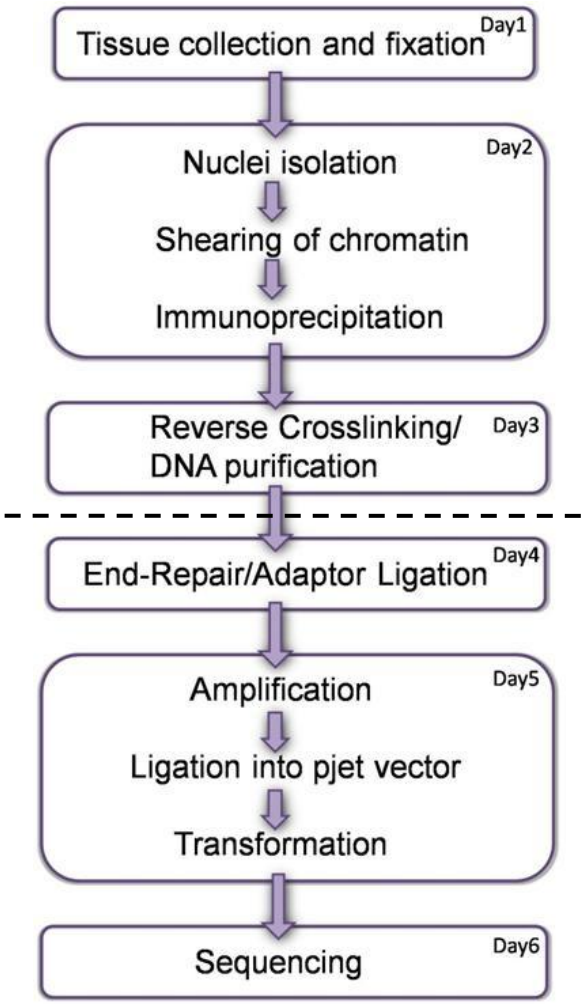
5. FUTURE PERSPECTIVES

Our study was initiated in an attempt to understand the molecular mechanism of GA signalling in seeds and thereby contribute towards advancement of basic scientific knowledge on GA-mediated germination.

Short term goals

In order to corroborate our hypothesis that GATA12 functions to maintain primary dormancy in seeds, it will be desirable to identify the global downstream targets of GATA12 protein by ChIP-seq analysis, which has been initiated. ChIP assay was performed using 35S:GATA12-6HA seeds and the DNA pulled down along with GATA12 protein was purified and sent for next-generation sequencing to BGI-Hong Kong. Unfortunately, the sample was rejected as it did not meet the concentration requirements. DNA purification using the PROMEGA PCR purification kit causes a loss in yield. To improve yield, more starting material will be used and the purification will be done by phenol/chloroform extraction method. Following these modifications, fresh sample will be sent to BGI. Global analysis of the DNA fragments will help us to better understand the downstream action of GATA12.

A



B

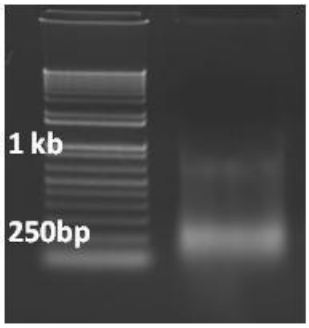


Figure 5.1: Work flow of ChIP-seq analysis to identify global downstream targets of GATA12

(A) The workflow above the dotted line was performed in laboratory. The work flow below the dotted line will be handled by BGI-Hong Kong **(B)** Sonication efficiency was checked by running the sheared chromatin on 1% agarose gel. The fragments were in the range of 200bp to 1kb, with most fragments in the 250-300bp range, as per BGI-Hong Kong requirements.

Since it has been shown that *GATA12* levels rise during embryogenesis, it would be interesting to check if GATA12 has any effect on embryo development as such, by analysing the different stages of embryos in *GATA12* transgenic siliques.

A genetic cross between *ga1 rgl2-1* and *GATA12-AS* to get triple mutant '*ga1 rgl2-1 GATA12-AS*' (Col-0) will help to further strengthen the genetic link between RGL2 and GATA12.

Although we found that GATA12 transgenic seeds showed no differences in endogenous ABA levels, it would be interesting to check if ABA signalling events are affected in these seeds or if GATA12 interacts with any of the ABA-related transcription factors to maintain dormancy.

Long term goals

One of the major issues faced by cereal crops is pre-harvest sprouting. Reduced dormancy in seeds causes seeds to germinate when on the crop, thereby severely affecting yield. Producers suffer loss when the sprouting damaged wheat is purchased at a discount, millers are faced with reduced flour yields and bakers encounter problems in processing and quality due to starch damage. Sprouting resistance can be enhanced by prolonging seed dormancy by genetic selection and breeding. On the other hand, increased dormancy in seeds is undesirable in direct seeding technology that requires uniformity in seed germination. Thus, control of seed dormancy is an extremely important agronomic trait. Identification of genes that control the time of release of dormancy will therefore help to reduce such huge

agronomic losses. On a global perspective, our attempt to understand the mechanism underlying maintenance and release of seed dormancy is a small contribution in addressing such an important agricultural issue seen in all crops.

6. GENERAL CONCLUSIONS

Seed germination is the first step in the growth cycle of plants and thus has been the focus of study for many researchers. Regulatory roles of GA in seed germination have been extensively explored. RGL2, a major repressor of seed germination, is an important component of GA signalling in seeds. However, the downstream mechanism of RGL2 is only poorly understood. In order to examine this, downstream targets of RGL2 were identified through microarray studies in our laboratory. One of the several interesting candidates, a GATA-type transcription factor, was chosen for further characterization as several members of the GATA family are known to mediate roles in seed germination or embryo development.

Although public microarray data show that *GATA12* is expressed in embryos, not much is known about the function of this protein. Our expression analysis of *GATA12* reveals that it is expressed in high amounts in dormant seeds. Indeed, levels of GATA12 rose during embryogenesis and peaked in the mature and highly dormant embryos. Of all the *Arabidopsis* tissues tested, *GATA12* levels peaked in freshly harvested seeds that are highly dormant. Also, histochemical GUS staining analysis revealed that the high levels in embryonic radicle of dry seeds drops upon germination. In addition, activity of *GATA12* promoter is higher in mature pollen grains than in germinated pollen grain/ tube. Phenotypic analyses of *GATA12* transgenic seeds further substantiated our results that GATA12 helps to maintain primary dormancy in

seeds. The *GATA12* overexpression seeds (freshly harvested) exhibited enhanced dormancy whereas suppression lines were less dormant that WT, under unstratified conditions.

We then investigated if this control of dormancy release exhibited by GATA12 was under hormonal regulation. Since *GATA12* was picked up from the microarray of RGL2 regulated genes, it most likely lies downstream of GA signalling. Indeed, *GATA12* expression was down-regulated by gibberellins in DELLA-dependant manner. Its transcripts accumulated in higher amounts in freshly harvested seeds of gibberellin-deficient background in comparison to WT seeds. However, analysis at the translational level will help to better understand the hormonal control over GATA12 function. Also, *GATA12* seemed to be up-regulated as an immediate response to translocation of functional RGL2 into nucleus in freshly harvested seeds. This regulation occurs by direct binding of RGL2-complex to the promoter of *GATA12.* We have discovered that the promoter regions to which this complex binds are enriched in Dof-associated motifs, which could likely be the recognition sites for binding.

To date, only a few interacting partners of RGL2 have been identified. Thus, we attempted an yeast library screen and identified several putative interacting proteins. Further confirmation of these interactions will help to identify the proteins that complex with RGL2 and help to regulate its downstream target.

Our results have helped to contribute more to our knowledge of gibberellin signalling in seeds. The findings described in this thesis have shed more light on the mechanism of RGL2 function and have contributed to our understanding of dormancy release and seed germination. This will help in future crop improvement strategies either via biotechnology or by marker-assisted breeding to allow seeds to germinate in uniformity upon sowing and at the same time prevent pre-harvest sprouting problems, thus improving the crop yield index.

7. REFERENCES

Achard, P. and P. Genschik (2009). "Releasing the brakes of plant growth: how GAs shutdown DELLA proteins." J Exp Bot **60**(4): 1085-1092.

Alboresi, A., C. Gestin, M. T. LEYDECKER, M. Bedu, C. Meyer and H. N. TRUONG (2005). "Nitrate, a signal relieving seed dormancy in Arabidopsis." Plant, Cell & Environment **28**(4): 500-512.

Appleford, N. E. J., D. J. Evans, J. R. Lenton, P. Gaskin, S. J. Croker, K. M. Devos, A. L. Phillips and P. Hedden (2006). "Function and transcript analysis of gibberellin-biosynthetic enzymes in wheat." Planta **223**(3): 568-582.

Ariizumi, T., P. K. Lawrence and C. M. Steber (2011). "The role of two F-box proteins, SLEEPY1 and SNEEZY, in Arabidopsis gibberellin signaling." Plant Physiology **155**(2): 765-775.

Bai, M. Y., J. X. Shang, E. Oh, M. Fan, Y. Bai, R. Zentella, T. P. Sun and Z. Y. Wang (2012). "Brassinosteroid, gibberellin and phytochrome impinge on a common transcription module in Arabidopsis." Nat Cell Biol **14**(8): 810-817.

Bewley, J. D. (1997). "Seed Germination and Dormancy." Plant Cell **9**(7): 1055-1066.

Bolle, C. (2004). "The role of GRAS proteins in plant signal transduction and development." Planta **218**(5): 683-692.

Burson, B. L., C. Tischler and W. Ocumpaugh (2009). Breeding for Reduced Post-Harvest Seed Dormancy in Switchgrass. American Society of Agronomy Branch Meeting..

Busk, P. K. and M. Pages (1998). "Regulation of abscisic acid-induced transcription." Plant Molecular Biology **37**(3): 425-435.

Cadman, C. S., P. E. Toorop, H. W. Hilhorst and W. E. Finch-Savage (2006). "Gene expression profiles of Arabidopsis Cvi seeds during dormancy cycling indicate a common underlying dormancy control mechanism." The Plant Journal **46**(5): 805-822.

Cao, D., A. Hussain, H. Cheng and J. Peng (2005). "Loss of function of four DELLA genes leads to light- and gibberellin-independent seed germination in Arabidopsis." Planta **223**(1): 105-113.

Cheng, H., L. Qin, S. Lee, X. Fu, D. E. Richards, D. Cao, D. Luo, N. P. Harberd and J. Peng (2004). "Gibberellin regulates Arabidopsis floral development via suppression of DELLA protein function." Development **131**(5): 1055-1064.

Davière, J.-M. and P. Achard (2013). "Gibberellin signaling in plants." Development **140**(6): 1147-1151.

Davies , P. J. (1995). Plant Hormones: Physiology, Biochemistry and Molecular Biology. Dordrecht, The Netherlands, Kluwer Academic Publishers..

de Lucas, M., J.-M. Davière, M. Rodríguez-Falcón, M. Pontin, J. M. Iglesias-Pedraz, S. Lorrain, C. Fankhauser, M. A. Blázquez, E. Titarenko and S. Prat (2008). "A molecular framework for light and gibberellin control of cell elongation." Nature **451**(7177): 480-484.

Dickinson, C. D., R. P. Evans and N. C. Nielsen (1988). "RY repeats are conserved in the 5'-flanking regions of legume seed-protein genes." Nucleic Acids Research **16**(1): 371.

Dill, A., H.-S. Jung and T.-p. Sun (2001). "The DELLA motif is essential for gibberellin-induced degradation of RGA." Proceedings of the National Academy of Sciences **98**(24): 14162-14167.

Dill, A. and T. P. Sun (2001). "Synergistic Derepression of Gibberellin Signaling by RGA and GAI Function in Arabidopsis thaliana." Genetics **159**: 777–785..

Finch-Savage, W. E. and G. Leubner-Metzger (2006). "Seed dormancy and the control of germination." New Phytologist **171**(3): 501-523.

Finkelstein, R., W. Reeves, T. Ariizumi and C. Steber (2008). "Molecular aspects of seed dormancy." Annual Review of Plant Biology **59**: 387-415.

Finkelstein, R. R. and T. J. Lynch (2000). "The Arabidopsis abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor." Plant Cell **12**(4): 599-609.

Finkelstein, R. R. and C. R. Somerville (1990). "Three classes of abscisic acid (ABA)-insensitive mutations of Arabidopsis define genes that control overlapping subsets of ABA responses." Plant Physiology **94**(3): 1172-1179.

Flematti, G. R., E. L. Ghisalberti, K. W. Dixon and R. D. Trengove (2004). "A compound from smoke that promotes seed germination." Science **305**(5686): 977-977.

Gabriele, S., A. Rizza, J. Martone, P. Circelli, P. Costantino and P. Vittorioso (2010). "The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene AtGA3ox1." Plant J **61**(2): 312-323.

Gallego-Bartolome, J., E. G. Minguet, J. A. Marin, S. Prat, M. A. Blazquez and D. Alabadi (2010). "Transcriptional Diversification and Functional Conservation between DELLA Proteins in Arabidopsis." Molecular Biology and Evolution **27**(6): 1247-1256.

Gao, X.-H., S.-L. Xiao, Q.-F. Yao, Y.-J. Wang and X.-D. Fu (2011). "An updated GA signaling 'relief of repression'regulatory model." Molecular Plant **4**(4): 601-606.

Garcia-Martinez, J. L., I. Lopez-Diaz, M. J. Sanchez-Beltran, A. L. Phillips, D. A. Ward, P. Gaskin and P. Hedden (1997). "Isolation and transcript analysis of gibberellin 20-oxidase genes in pea and bean in relation to fruit development." Plant Molecular Biology **33**(6): 1073-1084.

Griffiths, J., K. Murase, I. Rieu, R. Zentella, Z. L. Zhang, S. J. Powers, F. Gong, A. L. Phillips, P. Hedden, T. P. Sun and S. G. Thomas (2006). "Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis." Plant Cell **18**(12): 3399-3414.

Gualberti, G., M. Papi, L. Bellucci, I. Ricci, D. Bouchez, C. Camilleri, P. Costantino and P. Vittorioso (2002). "Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of Arabidopsis seeds." The Plant Cell Online **14**(6): 1253-1263.

Gubler, F., R. Kalla, J. K. Roberts and J. V. Jacobsen (1995). "Gibberellin-regulated expression of a myb gene in barley aleurone cells: evidence for Myb transactivation of a high-pI alpha-amylase gene promoter." The Plant Cell Online **7**(11): 1879-1891.

Gubler, F., D. Raventos, M. Keys, R. Watts, J. Mundy and J. V. Jacobsen (1999). "Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone." The Plant Journal **17**(1): 1-9.

Gutierrez, L., O. Van Wuytswinkel, M. Castelain and C. Bellini (2007). "Combined networks regulating seed maturation." Trends in Plant Science **12**(7): 294-300.

Hayashi, R., H. Wada, K. Ito and I. M. Adcock (2004). "Effects of glucocorticoids on gene transcription." European Journal of Pharmacology **500**(1): 51-62.

Hedden, P. (2008). "Plant biology: Gibberellins close the lid." Nature **456**(7221): 455-456.

Hedden, P. and A. L. Phillips (2000). "Gibberellin metabolism: new insights revealed by the genes." Trends in Plant Science **5**(12): 523-530.

Holdsworth, M. J., L. Bentsink and W. J. Soppe (2008). "Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination." New Phytologist **179**(1): 33-54.

Hong, G.-J., X.-Y. Xue, Y.-B. Mao, L.-J. Wang and X.-Y. Chen (2012). "Arabidopsis MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression." The Plant Cell Online **24**(6): 2635-2648.

Hou, X., L. Y. Lee, K. Xia, Y. Yan and H. Yu (2010). "DELLAs modulate jasmonate signaling via competitive binding to JAZs." Developmental Cell **19**(6): 884-894.

Iglesias-Fernández, R., M. C. Rodríguez-Gacio, C. Barrero-Sicilia, P. Carbonero and A. Matilla (2011). "Three endo-β-mannanase genes expressed in the micropylar endosperm and in the radicle influence germination of Arabidopsis thaliana seeds." Planta **233**(1): 25-36.

Itoh, H., M. Ueguchi-Tanaka, Y. Sato, M. Ashikari and M. Matsuoka (2002). "The gibberellin signaling pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei." The Plant Cell Online **14**(1): 57-70.

Jiang, C. and X. Fu (2007). "GA action: turning on de-DELLA repressing signaling." Curr Opin Plant Biol **10**(5): 461-465.

Kanno, Y., Y. Jikumaru, A. Hanada, E. Nambara, S. R. Abrams, Y. Kamiya and M. Seo (2010). "Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA biosynthesis, ABA transport and hormone interactions." Plant and Cell Physiology **51**(12): 1988-2001.

Karssen, C., D. Brinkhorst-Van der Swan, A. Breekland and M. Koornneef (1983). "Induction of dormancy during seed development by endogenous abscisic acid: studies on abscisic acid deficient genotypes of Arabidopsis thaliana (L.) Heynh." Planta **157**(2): 158-165.

Kasahara, H., A. Hanada, T. Kuzuyama, M. Takagi, Y. Kamiya and S. Yamaguchi (2002). "Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in Arabidopsis." Journal of Biological Chemistry **277**(47): 45188-45194.

King, K. E., T. Moritz and N. P. Harberd (2001). "Gibberellins are not required for normal stem growth in Arabidopsis thaliana in the absence of GAI and RGA." Genetics **159**(2): 767-776.

Ko, J. H., S. H. Yang and K. H. Han (2006). "Upregulation of an Arabidopsis RING-H2 gene, XERICO, confers drought tolerance through increased abscisic acid biosynthesis." plant journal **47**(3): 343-355.

Koornneef, M., L. Bentsink and H. Hilhorst (2002). "Seed dormancy and germination." Current Opinion in Plant Biology **5**(1): 33-36.

Koornneef, M., A. Elgersma, C. J. Hanhart, E. P. van Loenen-Martinet, L. van Rijn and J. A. D. Zeevaart (1985). "A gibberellin insensitive mutant of Arabidopsis. ." Plant Physiol **65**(33-39).

Koornneef, M. and J. Van der Veen (1980). "Induction and analysis of gibberellin sensitive mutants in Arabidopsis thaliana (L.) Heynh." Theoretical and Applied Genetics **58**(6): 257-263.

Kucera, B., M. A. Cohn and G. Leubner-Metzger (2005). "Plant hormone interactions during seed dormancy release and germination." Seed Science Research **15**(04): 281-307.

Le, B. H., C. Cheng, A. Q. Bui, J. A. Wagmaister, K. F. Henry, J. Pelletier, L. Kwong, M. Belmonte, R. Kirkbride, S. Horvath, G. N. Drews, R. L. Fischer, J. K. Okamuro, J. J. Harada and R. B. Goldberg (2010). "Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors." Proceedings of the National Academy of Sciences **107**(18): 8063-8070.

Lee, K. P., U. Piskurewicz, V. Turečková, M. Strnad and L. Lopez-Molina (2010). "A seed coat bedding assay shows that RGL2-dependent release of abscisic acid by the endosperm controls embryo growth in Arabidopsis dormant seeds." Proceedings of the National Academy of Sciences **107**(44): 19108-19113.

Lee, S., H. Cheng, K. E. King, W. Wang, Y. He, A. Hussain, J. Lo, N. P. Harberd and J. Peng (2002). "Gibberellin regulates Arabidopsis seed germination via RGL2, a GAI/RGA-like gene whose expression is up-regulated following imbibition." Genes Dev **16**(5): 646-658.

Letham, D., P. Goodwin and T. Higgins (1978). Phytohormones and Related Compounds: A Comprehensive Treatise. The Biochemistry of Phytohormones and Related Compounds, Elsevier..

Leubner-Metzger, G., C. Frundt, R. Vogeli-Lange and F. Meins, Jr. (1995). "Class I [beta]-1,3-Glucanases in the Endosperm of Tobacco during Germination." Plant Physiol **109**(3): 751-759.

Leung, J., S. Merlot and J. Giraudat (1997). "The Arabidopsis ABSCISIC ACID-INSENSITIVE2 (ABI2) and ABI1 genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction." The Plant Cell Online **9**(5): 759-771.

Lindsey, K. and J. F. Topping (1993). "Embryogenesis: a question of pattern." Journal of Experimental Botany **44**: 359–374.

Liu, P.-P., N. Koizuka, R. C. Martin and H. Nonogaki (2005). "The BME3 (Blue Micropylar End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination." The Plant Journal **44**(6): 960-971.

Lopez-Molina, L. and N. H. Chua (2000). "A null mutation in a bZIP factor confers ABA-insensitivity in Arabidopsis thaliana." Plant Cell Physiol **41**(5): 541-547.

Luo, X. M., W. H. Lin, S. Zhu, J. Y. Zhu, Y. Sun, X. Y. Fan, M. Cheng, Y. Hao, E. Oh, M. Tian, L. Liu, M. Zhang, Q. Xie, K. Chong and Z. Y. Wang (2010). "Integration of light- and brassinosteroid-signaling pathways by a GATA transcription factor in Arabidopsis." Developmental Cell **19**(6): 872-883.

MacMillan, J. (2002). "Occurrence of Gibberellins in Vascular Plants, Fungi, and Bacteria." Journal of Plant Growth Regulation **20**(4): 387-442.

Manfield, I. W., P. F. Devlin, C.-H. Jen, D. R. Westhead and P. M. Gilmartin (2007). "Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the Arabidopsis GATA gene family." Plant Physiology **143**(2): 941-958.

McCormick, S. (2004). "Control of male gametophyte development." The Plant Cell Online **16**(suppl 1): S142-S153.

McGinnis, K. M., S. G. Thomas, J. D. Soule, L. C. Strader, J. M. Zale, T.-p. Sun and C. M. Steber (2003). "The Arabidopsis SLEEPY1 Gene Encodes a Putative F-Box Subunit of an SCF E3 Ubiquitin Ligase." The Plant Cell Online **15**(5): 1120-1130.

Müller, K., S. Tintelnot and G. Leubner-Metzger (2006). "Endosperm-limited Brassicaceae seed germination: abscisic acid inhibits embryo-induced endosperm weakening of Lepidium sativum (cress) and endosperm rupture of cress and Arabidopsis thaliana." Plant and Cell Physiology **47**(7): 864-877.

Mur, L. A., J. Mandon, S. Persijn, S. M. Cristescu, I. E. Moshkov, G. V. Novikova, M. A. Hall, F. J. Harren, K. H. Hebelstrup and K. J. Gupta (2013). "Nitric oxide in plants: an assessment of the current state of knowledge." AoB Plants **5**.

Murase, K., Y. Hirano, T.-p. Sun and T. Hakoshima (2008). "Gibberellin-induced DELLA recognition by the gibberellin receptor GID1." Nature **456**(7221): 459-463.

Nakabayashi, K., M. Bartsch, Y. Xiang, E. Miatton, S. Pellengahr, R. Yano, M. Seo and W. J. Soppe (2012). "The time required for dormancy release in Arabidopsis is determined by DELAY OF GERMINATION1 protein levels in freshly harvested seeds." The Plant Cell Online **24**(7): 2826-2838.

Nakamura, S., T. J. Lynch and R. R. Finkelstein (2001). "Physical interactions between ABA response loci of Arabidopsis." The Plant Journal **26**(6): 627-635.

Nambara, E. and A. Marion-Poll (2005). "Abscisic acid biosynthesis and catabolism." Annual Review of Plant Biology **56**: 165-185.

Nawy, T., M. Bayer, J. Mravec, J. Friml, K. D. Birnbaum and W. Lukowitz (2010). "The GATA Factor HANABA TARANU Is Required to Position the

Proembryo Boundary in the Early Arabidopsis Embryo." Developmental Cell **19**(1): 103-113.

Nonogaki, H., O. H. Gee and K. J. Bradford (2000). "A germination-specific endo-β-mannanase gene is expressed in the micropylar endosperm cap of tomato seeds." Plant Physiology **123**(4): 1235-1246.

Ogawa, M., A. Hanada, Y. Yamauchi, A. Kuwahara, Y. Kamiya and S. Yamaguchi (2003). "Gibberellin biosynthesis and response during Arabidopsis seed germination." Plant Cell **15**(7): 1591-1604.

Oh, E., S. Yamaguchi, Y. Kamiya, G. Bae, W. I. Chung and G. Choi (2006). "Light activates the degradation of PIL5 protein to promote seed germination through gibberellin in Arabidopsis." The Plant Journal **47**(1): 124-139.

Papi, M., S. Sabatini, M. M. Altamura, L. Hennig, E. Schäfer, P. Costantino and P. Vittorioso (2002). "Inactivation of the Phloem-Specific Dof Zinc Finger GeneDAG1 Affects Response to Light and Integrity of the Testa of Arabidopsis Seeds." Plant Physiology **128**(2): 411-417.

Park, J., N. Lee, W. Kim, S. Lim and G. Choi (2011). "ABI3 and PIL5 collaboratively activate the expression of SOMNUS by directly binding to its promoter in imbibed Arabidopsis seeds." The Plant Cell Online **23**(4): 1404-1415.

Park, J., K. T. Nguyen, E. Park, J. S. Jeon and G. Choi (2013). "DELLA proteins and their interacting RING Finger proteins repress gibberellin responses by binding to the promoters of a subset of gibberellin-responsive genes in Arabidopsis." Plant Cell **25**: 927-943.

Penfield, S., E.-M. Josse, R. Kannangara, A. D. Gilday, K. J. Halliday and I. A. Graham (2005). "Cold and light control seed germination through the bHLH transcription factor SPATULA." Current Biology **15**(22): 1998-2006.

Penfield, S., R. C. Meissner, D. A. Shoue, N. C. Carpita and M. W. Bevan (2001). "MYB61 is required for mucilage deposition and extrusion in the Arabidopsis seed coat." The Plant Cell Online **13**(12): 2777-2791.

Peng, J. (2002). "The role of GA-mediated signalling in the control of seed germination." Curr Opin Plant Biol **5**(5): 376-381.

Peng, J., P. Carol, D. E. Richards, K. E. King, R. J. Cowling, G. P. Murphy and N. P. Harberd (1997). "The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses." Genes & Development **11**(23): 3194-3205.

Peng, J. and N. P. Harberd (1993). "Derivative alleles of the Arabidopsis gibberellin-insensitive (gai) mutation confer a wild-type phenotype." The Plant Cell Online **5**(3): 351-360.

Peng, J. and N. P. Harberd (1997). "Gibberellin deficiency and response mutations suppress the stem elongation phenotype of phytochrome-deficient mutants of Arabidopsis." Plant Physiology **113**(4): 1051-1058.

Peng, J. R., D. E. Richards, N. M. Hartley, G. P. Murphy, K. M. Devos, J. E. Flintham, J. Beales, L. J. Fish, A. J. Worland and F. Pelica (1999). "''Green revolution'' genes encode mutant gibberellin response modulators." Nature **400**: 256–261.

Peter, H. and G. T. Stephen (2012). "Gibberellin biosynthesis and its regulation." Biochemical Journal **444**(1): 11-25.

Piskurewicz, U., Y. Jikumaru, N. Kinoshita, E. Nambara, Y. Kamiya and L. Lopez-Molina (2008). "The gibberellic acid signaling repressor RGL2 inhibits Arabidopsis seed germination by stimulating abscisic acid synthesis and ABI5 activity." The Plant Cell Online **20**(10): 2729-2745.

Piskurewicz, U. and L. Lopez-Molina (2009). "The GA-signaling repressor RGL3 represses testa rupture in response to changes in GA and ABA levels." Plant Signaling & Behavior **4**(1): 63-65.

Pratt, W., Y. Morishima, M. Murphy and M. Harrell (2006). Chaperoning of glucocorticoid receptors. Molecular Chaperones in Health and Disease, Springer**:** 111-138.

Raz, V., J. H. Bergervoet and M. Koornneef (2001). "Sequential steps for developmental arrest in Arabidopsis seeds." Development **128**(2): 243-252.

Reyes, J. C., M. I. Muro-Pastor and F. J. Florencio (2004). "The GATA family of transcription factors in Arabidopsis and rice." Plant Physiology **134**(4): 1718-1732.

Richter, R., C. Behringer, I. K. Müller and C. Schwechheimer (2010). "The GATA-type transcription factors GNC and GNL/CGA1 repress gibberellin signaling downstream from DELLA proteins and PHYTOCHROME-INTERACTING FACTORS." Genes & Development **24**(18): 2093-2104.

Rieu, I., S. Eriksson, S. J. Powers, F. Gong, J. Griffiths, L. Woolley, R. Benlloch, O. Nilsson, S. G. Thomas and P. Hedden (2008). "Genetic analysis reveals that C19-GA 2-oxidation is a major gibberellin inactivation pathway in Arabidopsis." The Plant Cell Online **20**(9): 2420-2436.

Rueda-Romero, P., C. Barrero-Sicilia, A. Gómez-Cadenas, P. Carbonero and L. Oñate-Sánchez (2012). "Arabidopsis thaliana DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14." Journal of Experimental Botany **63**(5): 1937-1949.

Salisbury, F. B. and C. W. Ross (1992). Plant Physiology. Belmont, CA, Wadsworth..

Sarnowska, E. A., A. T. Rolicka, E. Bucior, P. Cwiek, T. Tohge, A. R. Fernie, Y. Jikumaru, Y. Kamiya, R. Franzen and E. Schmelzer (2013). "DELLA-interacting SWI3C core subunit of SWI/SNF chromatin remodeling complex modulates gibberellin responses and hormonal crosstalk in Arabidopsis." Plant Physiology: pp. 113.223933.

Sasaki, A., H. Itoh, K. Gomi, M. Ueguchi-Tanaka, K. Ishiyama, M. Kobayashi, D.-H. Jeong, G. An, H. Kitano, M. Ashikari and M. Matsuoka (2003). "Accumulation of Phosphorylated Repressor for Gibberellin Signaling in an F-box Mutant." Science **299**(5614): 1896-1898.

Schneider, G. and W. Schliemann (1994). "Gibberellin conjugates–an overview." Plant Growth Regulation **15:247–60**.

Schomburg, F. M., C. M. Bizzell, D. J. Lee, J. A. Zeevaart and R. M. Amasino (2003). "Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants." The Plant Cell Online **15**(1): 151-163.

Schopfer, P. and C. Plachy (1985). "Control of Seed-Germination by Abscisic-Acid " Plant Physiol **77**(3): 676-686.

Shaikhali, J., J. de Dios Barajas-Lopéz, K. Ötvös, D. Kremnev, A. S. Garcia, V. Srivastava, G. Wingsle, L. Bako and Å. Strand (2012). "The CRYPTOCHROME1-dependent response to excess light is mediated through the transcriptional activators ZINC FINGER PROTEIN EXPRESSED IN INFLORESCENCE MERISTEM LIKE1 and ZML2 in Arabidopsis." The Plant Cell Online **24**(7): 3009-3025.

Shen, Q., P. Zhang and T. Ho (1996). "Modular nature of abscisic acid (ABA) response complexes: composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley." The Plant Cell Online **8**(7): 1107-1119.

Shikata, M., Y. Matsuda, K. Ando, A. Nishii, M. Takemura, A. Yokota and T. Kohchi (2004). "Characterization of Arabidopsis ZIM, a member of a novel plant-specific GATA factor gene family." Journal of Experimental Botany **55**(397): 631-639.

Shimada, A., M. Ueguchi-Tanaka, T. Nakatsu, M. Nakajima, Y. Naoe, H. Ohmiya, H. Kato and M. Matsuoka (2008). "Structural basis for gibberellin recognition by its receptor GID1." Nature **456**(7221): 520-523.

Silverstone, A. L., C. N. Ciampaglio and T. Sun (1998). "The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway." Plant Cell **10**(2): 155-169.

Silverstone, A. L., H.-S. Jung, A. Dill, H. Kawaide, Y. Kamiya and T.-p. Sun (2001). "Repressing a repressor gibberellin-induced rapid reduction of the RGA protein in Arabidopsis." The Plant Cell Online **13**(7): 1555-1566.

Silverstone, A. L., P. Y. A. Mak, E. C. Martinez and T. P. Sun (1997). "The New RGA Locus Encodes a Negative Regulator of Gibberellin response in Arabidopsis thaliana." Genetics **146**: 1087-1099.

Stamm, P., P. Ravindran, B. Mohanty, E. Tan, H. Yu and P. Kumar (2012). "Insights into the molecular mechanism of RGL2-mediated inhibition of seed germination in Arabidopsis thaliana." BMC Plant Biology **12**(1): 179.

Sun, T. P. and Y. Kamiya (1997). "Regulation and cellular localization of ent-kaurene synthesis." Physiologia Plantarum **101**(4): 701-708.

Thomas, S. G., A. L. Phillips and P. Hedden (1999). "Molecular cloning and functional expression of gibberellin 2- oxidases, multifunctional enzymes involved in gibberellin deactivation." Proc Natl Acad Sci U S A **96**(8): 4698-4703.

Tyler, L., S. G. Thomas, J. Hu, A. Dill, J. M. Alonso, J. R. Ecker and T.-p. Sun (2004). "DELLA proteins and gibberellin-regulated seed germination and floral development in Arabidopsis." Plant Physiology **135**(2): 1008-1019.

Ueguchi-Tanaka, M., M. Ashikari, M. Nakajima, H. Itoh, E. Katoh, M. Kobayashi, T. Y. Chow, Y. I. Hsing, H. Kitano, I. Yamaguchi and M. Matsuoka (2005). "GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin." Nature **437**(7059): 693-698.

Ueguchi-Tanaka, M., Y. Fujisawa, M. Kobayashi, M. Ashikari, Y. Iwasaki, H. Kitano and M. Matsuoka (2000). "Rice dwarf mutant d1, which is defective in the α subunit of the heterotrimeric G protein, affects gibberellin signal transduction." Proceedings of the National Academy of Sciences **97**(21): 11638-11643.

Umehara, M., A. Hanada, S. Yoshida, K. Akiyama, T. Arite, N. Takeda-Kamiya, H. Magome, Y. Kamiya, K. Shirasu and K. Yoneyama (2008). "Inhibition of shoot branching by new terpenoid plant hormones." Nature **455**(7210): 195-200.

Varbanova, M., S. Yamaguchi, Y. Yang, K. McKelvey, A. Hanada, R. Borochov, F. Yu, Y. Jikumaru, J. Ross, D. Cortes, C. J. Ma, J. P. Noel, L. Mander, V. Shulaev, Y. Kamiya, S. Rodermel, D. Weiss and E. Pichersky (2007). "Methylation of gibberellins by Arabidopsis GAMT1 and GAMT2." Plant Cell **19**(1): 32-45.

Wang, F., D. Zhu, X. Huang, S. Li, Y. Gong, Q. Yao, X. Fu, L.-M. Fan and X. W. Deng (2009). "Biochemical insights on degradation of Arabidopsis DELLA proteins gained from a cell-free assay system." The Plant Cell Online **21**(8): 2378-2390.

Weber, H., L. Borisjuk and U. Wobus (2005). "Molecular physiology of legume seed development." Annual Review of Plant Biology **56**: 253-279.

Weier, T. E., T. L. Rost and T. E. Weier (1979). "Botany: a brief introduction to plant biology." New York: Wiley: 155–170.

Wen, C.-K. and C. Chang (2002). "Arabidopsis RGL1 encodes a negative regulator of gibberellin responses." The Plant Cell Online **14**(1): 87-100.

Wild, M., J.-M. Davière, S. Cheminant, T. Regnault, N. Baumberger, D. Heintz, R. Baltz, P. Genschik and P. Achard (2012). "The Arabidopsis DELLA RGA-LIKE3 is a direct target of MYC2 and modulates jasmonate signaling responses." The Plant Cell Online **24**(8): 3307-3319.

Willige, B. C., S. Ghosh, C. Nill, M. Zourelidou, E. M. Dohmann, A. Maier and C. Schwechheimer (2007). "The DELLA domain of GA INSENSITIVE mediates the interaction with the GA INSENSITIVE DWARF1A gibberellin receptor of Arabidopsis." The Plant Cell Online **19**(4): 1209-1220.

Xi, W., C. Liu, X. Hou and H. Yu (2010). "MOTHER OF FT AND TFL1 regulates seed germination through a negative feedback loop modulating ABA signaling in Arabidopsis." Plant Cell **22**(6): 1733-1748.

Yabuta, T. and Y. Sumiki. (1938). "On the Crystal of Gibberellin, a Substance to Promote Plant Growth." Journal of the Agricultural Chemical Society Japan **14**: 1526.

Yamaguchi, S. (2008). "Gibberellin metabolism and its regulation." Annual Review of Plant Biology **59**: 225-251.

Yamauchi, Y., M. Ogawa, A. Kuwahara, A. Hanada, Y. Kamiya and S. Yamaguchi (2004). "Activation of gibberellin biosynthesis and response pathways by low temperature during imbibition of Arabidopsis thaliana seeds." The Plant Cell Online **16**(2): 367-378.

Yanagisawa, S. (2002). "The Dof family of plant transcription factors." Trends in Plant Science **7**(12): 555-560.

Yang, X. and X. Zhang (2010). "Regulation of somatic embryogenesis in higher plants." Critical Reviews in Plant Science **29**(1): 36-57.

Yu, H., T. Ito, Y. Zhao, J. Peng, P. Kumar and E. M. Meyerowitz (2004). "Floral homeotic genes are targets of gibberellin signaling in flower development." Proc Natl Acad Sci U S A **101**(20): 7827-7832.

Zentella, R., Z. P. Zhang, M., S. G. Thomas, A. Endo, K. Murase, C. M. Fleet, Y. Jikumaru, E. Nambara and Y. Kamiya (2007). "Global analysis of DELLA direct targets in early gibberellin signaling in Arabidopsis." Plant Cell Physiol **19**: 3037–3057.

Zhao, Y., L. Medrano, K. Ohashi, J. C. Fletcher, H. Yu, H. Sakai and E. M. Meyerowitz (2004). "HANABA TARANU is a GATA transcription factor that

regulates shoot apical meristem and flower development in Arabidopsis." The Plant Cell Online **16**(10): 2586-2600.

Zhu, G., N. Ye and J. Zhang (2009). "Glucose-induced delay of seed germination in rice is mediated by the suppression of ABA catabolism rather than an enhancement of ABA biosynthesis." Plant and Cell Physiology **50**(3): 644-651.

Zhu, Y. Y., T. Nomura, Y. H. Xu, Y. Y. Zhang, Y. Peng, B. Z. Mao, A. Hanada, H. C. Zhou, R. X. Wang, P. J. Li, X. D. Zhu, L. N. Mander, Y. Kamiya, S. Yamaguchi and Z. H. He (2006). "ELONGATED UPPERMOST INTERNODE encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice." Plant Cell **18**(2): 442-456.

8. APPENDICES

8.1 GATA12 gene and protein sequence:

> *GATA12* CDS

ATGGAAGATGAAGCACATGAATTCTTCCACACATCTGATTTTGCCGTTGATGACCTTTTAGTTGATTTC
TCTAACGATGATGACGAAGAAAACGATGTTGTTGCTGATTCCACCACTACCACCACCATAACCGACAG
CTCTAACTTCTCCGCTGCTGATCTTCCCAGTTTCCACGGTGATGTTCAAGACGGCACTAGCTTCTCCGG
TGACCTTTGTATACCTTCTGATGATTTGGCTGATGAGTTAGAGTGGCTTTCGAACATTGTGGATGAAT
CATTGTCGCCTGAAGATGTACACAAGCTCGAGCTAATATCCGGTTTTAAGAGTCGACCGGACCCGAA
ATCCGATACCGGAAGCCCGGAAAACCCGAATAGCAGCAGTCCGATTTTTACTACCGACGTTTCTGTAC
CGGCCAAAGCTAGAAGCAAACGCTCACGCGCCGCTGCGTGTAATTGGGCCTCACGTGGGCTTCTCAA
GGAAACGTTTTACGACAGTCCTTTCACCGGAGAAACCATTCTCTCTAGCCAACAACACTTGTCTCCGCC
AACCTCGCCGCCTTTGTTGATGGCTCCGCTAGGGAAAAAGCAAGCCGTTGATGGAGGACACCGACGG
AAGAAGGATGTTTCTTCACCGGAGTCTGGTGGCGCAGAGGAGAGACGGTGTCTCCACTGCGCCACG
GATAAGACTCCGCAATGGCGGACAGGCCCAATGGGCCCGAAGACGTTGTGCAACGCTTGCGGTGTT
AGGTACAAATCGGGACGTTTAGTGCCGGAGTATCGGCCCGCGGCGAGTCCGACGTTTGTGCTGGCG
AAACACTCAAATTCTCATCGGAAAGTTATGGAGCTCCGGCGACAGAAGGAGATGAGTAGGGCCCATC
ATGAGTTCATACATCACCATCACGGTACGGACACTGCCATGATTTTCGACGTTTCATCGGACGGTGAT
GATTACTTGATCCACCACAACGTTGGCCCAGATTTCAGACAGCTTATTTGA

> GATA12 protein

MEDEAHEFFHTSDFAVDDLLVDFSNDDDEENDVVADSTTTTTIT
DSSNFSAADLPSFHGDVQDGTSFSGDLCIPSDDLADELEWLSNIVDESLSPEDVHKLE
LISGFKSRPDPKSDTGSPENPNSSSPIFTTDVSVPAKARSKRSRAAACNWASRGLLKE
TFYDSPFTGETILSSQQHLSPPTSPPLLMAPLGKKQAVDGGHRRKKDVSSPESGGAEE
RRCLHCATDKTPQWRTGPMGPKTLCNACGVRYKSGRLVPEYRPAASPTFVLAKHSNSH
RKVMELRRQKEMSRAHHEFIHHHHGTDTAMIFDVSSDGDDYLIHHNVGPDFRQLI

8.2 *Ab Initio* Promoter Analysis

TRANSFAC

>At5g25830 (GATA12)- 2kb upstream promoter

TAGCTACAGTTATTACATTAGTAATTCTCATTTCGTCTTTTGAAGTGATTACCGTGTCTGAATTTTAGGA
AACGGGATTGGATTCTAATTATGACTGATATTTCATATTTGTAACGTGTAGTGGTTACTAGGTTAATA
GTAAGACTAAAACTTAGGGGGTAACTAAAGAAGAGAATGTTTTGACCGGTCATACTATTTCTGATTGC
CTGAGCTGGAAAATTGTCGTGGCTAAAAGTGGGAATCGGAAAAAGGAATAAAGGGATTTAACATGG
GATTAGGAAGAAGGGTTAACGTGTTTATAAGATAACAATAGTTAAAAGATAAATAACTCTTAATATAT
TCTTATACAAGTCCCGAGATATATTGTTCGATGTGGTAGATAATGGAACTCGGCCAATAGAAATATAG
AATGTCTTGATATTAAGAATTAAGTCGCTTTCCATTTTCTGGTTGTATAGTTAGCTAATTTAGTTGCCTA
TTTAGGGTTATTAATGATTTCTTAGTTTAGTTTTCTTGAAATGGGTTATAATTAATTACTTATATCTAAC
GTGAAAAGTTTTTTTTTTTTTTTTGTCTTCTTATTTGCATTGTTGAGAGGGTTTGTTCTTGTCGCATCCAT

TACTGATAAAGTTTACGACAAACTGCTAAAAGATAAAATCACTTAAGACTTAAGAGCAACAGCTATTC
AATTTTTTTTTATCCCATTTGGATAGCCTAAAAGTCTAAAACTATATGCAAAGCAATAAACTTAACTAT
AATCTAGCTCTACCAAAAATGTGAAGATAAGAAAATATTAAAACATATTCAAGTTCGAGTGTCGCTAG
TAATTACTACCAACATGAATAATCACATACTCGATCAGCCATTTTATGTCATATAAATATATTTTTGTTT
TATACTTTCTATAATTAATTCTCTAGATCCGAATATACCTCAAATGTACATGATTTGATACACTATCATT
TTATCAAAAAGGTCTATTGTTATCGCCTAAGCTTAAAGGTGGTGTCGGTACTGCCGATATCCACCCTG
ATCAAATCTTTGATACGTTGTTGTTTCAGTTTTTACTTTTTTTCCTTATTTTGTGGACCGATTGGCCCGTA
TAATACATACTATTGTTAACCTATTTGGGTTCAGTTAATATATGCCGTTTAAATCAATAACAAAGAGAA
AAACTTTATTCTTATAGGTATTACATTCTAATTAAAAAGCGGATGAACAAAGAAAAAGGAAAAAGAA
GGTATTATACCAAAACCGGTGACTCGATCCAGTTTACAATACATATAAATAATTAATGTGGATCGGTT
TCAACTTTCGCGTAATGGTGTCGACGCTTATCGAGTGTTGGTTGATGATAATCAACGGCGTGATCATC
ATCCATCACATGCACGTGCGTGGTACTTTCTGTAATTTGAAGCAATCTTATGACTCTTTTGGTAAAGTT
GTCCCCACCACTATACCATAATTGCACATGTTTGATTCAATTAATCCTTTAATGAAGTCAAAAACATTTC
CCACCTTTGTGTTTGTCAACTAACCAAATTCTAATTTTCCCAACAAATTTATGGCTTAGTAACTAGGAA
AGTGTAAAAATAAATAACTCCCCACAAACGTGACTACACATGTGCTTGTACAAGCATGTAAACAAACT
AAAATAACACTAATAATTACAATTTAACAAAGGAAAAAAAACTAGTAAAAATAGGAATAATAGTAAG
TAGTGAGTGTAAAAAGAAATAATTCCCATAAAAACTCCTTTGTCTTCGTCAAACTTTTCCGAATTAGAC
CTCTTGTCCCCTTCAAGCGCAATAAAAGTGAAAGTAAAATAAACAAAAGTTTATCTTTTCTACAGAGA
AACCATAACTTTCTCTAACTTTACTCCTTTTTAAACTCACATTTTCCCCTCACTTCACTCTCTCCTCTCTCC
AAAGACCTCCGACGAGAGTTTACCGGAAAAAAAAAAAAAAAAAACAACTCTTTAGTTTTAATCAACG
AAACTATG

XXX These are your search results from Fri, 20.4.2012 - 11:25 MEZ
XXX for the following search: default27314

Search for sites by WeightMatrix library: matrixTFP20083.lib
Sequence file: default.seq
Matrix groups: plants

Cut-offs: to minimize false positive matches

Scanning sequence ID: At5g25830

FFF sequence: 1, searchname: default27314.out,

matrix identifier (always the strand is shown)	position factor name (strand)	core match	matrix match	sequence (+)-
P$ARR10_01 ARR10	35 (-)	0.996	0.975	caAATCT
P$DOF_Q2 ttaCTTTTttt	65 (-) Dof	1.000	1.000	
P$ASR1_01 ASR-1	129 (-)	1.000	1.000	TGGGT

```
P$TGA1A_Q2_01                 157 (+)  0.822  0.828
tcaATAACaaagagaaaaactt           TGA1a
P$PBF_Q2                      226 (+)  1.000  1.000  aAAAGG
PBF
P$BHLH66_01                   387 (+)  1.000  1.000  gCACGTgc
bHLH66
P$BHLH66_01                   387 (-)  1.000  1.000  gcACGTGc
bHLH66
P$OCSBF1_01                   388 (+)  1.000  1.000  CACGT
OCSBF-1
P$OCSBF1_01                   389 (-)  1.000  1.000  ACGTG
OCSBF-1
P$BZR1_01                     390 (+)  1.000  1.000  CGTGCg
BZR1
P$BPC1_Q2                     401 (-)  1.000  1.000  cTTTCT
BPC1
P$TGA1A_Q2_01                 487 (-)  0.905  0.892
aatcctttaatgaaGTCAAaaa           TGA1a
P$TGA1A_Q2_01                 493 (+)  0.793  0.863
ttaATGAAgtcaaaaacatttc           TGA1a
P$OCSBF1_01                   611 (-)  1.000  1.000  ACGTG
OCSBF-1
P$PBF_01                      808 (+)  1.000  0.995
aatAAAAGtga                      PBF
P$TGA1A_Q2_01                 857 (+)  0.822  0.799
accATAACtttctctaacttta           TGA1a
P$BPC1_Q2                     864 (-)  1.000  1.000  cTTTCT
BPC1
P$PBF_Q2                      881 (-)  1.000  1.000  CCTTTt
PBF


 Total sequences length=999
 Total number of sites found=18
 Frequency of sites per nucleotide=0.018018



Search for sites by WeightMatrix library:
file:///usr/local/apache/cgi-
bin/build_t/idb/1.0/Match/data/2014.1/matrix.dat
Sequence file: file:///usr/local/apache/cgi-
bin/build_t/idb/1.0/Match/usr/btibioinfo1/seq/13554.seq
Profile:
file:///tmp/cfa5d57bc418384b9d9eb86603f3f276_9839/taxon.prf,
plants.prf
Configuration file:
/tmp/cfa5d57bc418384b9d9eb86603f3f276_9839/project.xml

Inspecting sequence ID   tmp.seq_0

 P$ATHSFA1D_01         |        3 (+) |  1.000 |  0.940 |
gCTACA                      | ATHSFA1D
 P$GAMYB_Q2            |        6 (-) |  0.982 |  0.925 |
acaGTTATtacat               | GAMYB
 P$NAC6_01             |       13 (-) |  0.861 |  0.884 |
TTACAt                      | NAC6
 P$HSFA4A_01           |       17 (-) |  0.947 |  0.952 |
ATTAGt                      | HSFA4A
```

P$NAC6_01	19 (+)	0.894	0.765	tAGTAA	NAC6
P$O2_03	28 (-)	0.800	0.818	tcattTCGTC	Opaque-2
P$NAC6_01	31 (-)	0.782	0.818	TTTCGt	NAC6
P$AT2G41690_01	31 (-)	0.970	0.949	tTTCGT	At2g41690
P$TGA1A_01	32 (+)	0.860	0.763	tTCGTCt	TGA1A
P$ATHSFA1D_01	41 (-)	0.950	0.953	TGAAGt	ATHSFA1D
P$TGA1A_01	41 (-)	0.833	0.848	tGAAGTg	TGA1A
P$GT1_Q6	45 (+)	0.888	0.923	GTGATta	GT-1
P$ATHSFA1D_01	64 (-)	0.918	0.908	TTTAGg	ATHSFA1D
P$GBP_Q6	110 (-)	1.000	0.949	tgtaACGTGtag	GBP
P$O2_02	111 (-)	1.000	0.946	gtaACGTGta	Opaque-2
P$NAC6_01	112 (-)	0.864	0.886	TAACGt	NAC6
P$TGA1A_01	112 (-)	0.731	0.762	tAACGTg	TGA1A
P$LIM1_01	113 (-)	0.963	0.904	aacgtgtAGTGG	LIM1
P$OCSBF1_01	114 (-)	1.000	1.000	ACGTG	OCSBF-1
P$ABI5_Q2	114 (+)	1.000	0.979	ACGTGt	ABI5
P$ATHSFA1D_01	117 (-)	1.000	1.000	TGTAGt	ATHSFA1D
P$GAMYB_Q2	121 (-)	0.969	0.925	gtgGTTACtaggt	GAMYB
P$ATMYB15_Q2	122 (-)	0.989	0.991	tGGTTA	Myb-15
P$NAC6_01	125 (-)	0.894	0.765	TTACTa	NAC6
P$GT1_Q6	132 (+)	0.984	0.989	GTTAAta	GT-1
P$HSFA4A_01	135 (-)	1.000	1.000	AATAGt	HSFA4A
P$NAC6_01	137 (+)	0.894	0.765	tAGTAA	NAC6
P$HMG1_01	143 (-)	0.848	0.857	gactAAAAC	HMG-1
P$ATHSFA1D_01	144 (+)	0.918	0.923	aCTAAA	ATHSFA1D
P$ASR1_01	154 (-)	0.962	0.962	AGGGG	ASR-1
P$GT1_Q6_01	158 (-)	0.994	0.935	ggtaactAAAGA	GT-1
P$NAC6_01	161 (+)	0.723	0.768	aACTAA	NAC6
P$DOF1_01	161 (+)	1.000	0.994	aacTAAAGaag	Dof1

P$ATHSFA1D_01	162 (+)	0.918	0.923	aCTAAA	ATHSFA1D
P$TGA1A_Q2_01	173 (-)	1.000	0.877	gaatgttttgaccgGTCATact	TGA1a
P$TGA1A_Q2_01	177 (+)	0.905	0.845	gttTTGACcggtcatactattt	TGA1a
P$GT1_Q6	178 (-)	0.980	0.957	ttTTGAC	GT-1
P$ZAP1_01	180 (+)	1.000	0.933	TTGACcggtc	ZAP1
P$SPF1_Q2	189 (-)	1.000	0.949	catACTATtt	SPF1
P$HSFA4A_01	192 (+)	1.000	1.000	aCTATT	HSFA4A
P$SQUA_01	193 (+)	0.917	0.917	ctaTTTCTga	SQUA
P$AGL1_02	202 (+)	1.000	0.941	aTTGCCtgagctggaaaa	AGL1
P$AG_01	202 (+)	0.869	0.873	attgCCTGAgctggaaaa	AG
P$AGL1_01	202 (+)	0.968	0.944	attgCCTGAgctggaaaa	AGL1
P$AG_03	202 (+)	0.869	0.873	attgCCTGAgctggaaaa	AG
P$AGL1_02	202 (-)	0.997	0.904	attgcctgagctGGAAAa	AGL1
P$AG_02	203 (+)	0.994	0.910	TTGCCtgagctggaaa	AG
P$AG_02	203 (-)	1.000	0.910	ttgcctgagctGGAAA	AG
P$LIM1_01	217 (-)	0.968	0.919	aaattgtCGTGG	LIM1
P$HBPA1_Q6_01	220 (-)	0.869	0.876	tTGTCGtggc	HBP-1a
P$TGA1A_01	221 (-)	0.808	0.827	tGTCGTg	TGA1A
P$HY5_01	221 (+)	0.837	0.867	tgTCGTGgct	HY5
P$AGL3_01	225 (-)	0.902	0.882	gtggctaaaAGTGGgaat	AGL3
P$AGL3_02	225 (-)	0.883	0.881	gtggctaaaAGTGGgaat	AGL3
P$AGL3_03	225 (-)	0.883	0.881	gtggctaaaAGTGGgaat	AGL3
P$RIN_Q2	228 (-)	0.978	0.912	gcTAAAAgtgg	RIN
P$DOF3_01	228 (+)	0.994	0.986	gctaAAAGTgg	Dof3
P$HSFA4A_01	231 (-)	0.926	0.933	AAAAGt	HSFA4A
P$PBF_Q2	231 (+)	0.963	0.964	aAAAGT	PBF
P$PBF_Q2_01	231 (-)	0.998	0.998	aAAAGT	PBF
P$MADSA_Q2	240 (+)	0.800	0.840	aatcggAAAAAggaat	MADS-A
P$AT3G51080_01	244 (+)	1.000	1.000	GGAAAaa	At3g51080

P$DOF_Q2	244 (+)	1.000	0.992	ggaAAAAGgaa	Dof
P$HSFA4A_01	247 (-)	0.926	0.897	AAAAGg	HSFA4A
P$PBF_Q2	247 (+)	1.000	1.000	aAAAGG	PBF
P$PBF_Q2_01	247 (-)	1.000	1.000	aAAAGG	PBF
P$DOF3_01	252 (+)	0.992	0.991	gaatAAAGGga	Dof3
P$PBF_Q2	255 (+)	1.000	0.986	tAAAGG	PBF
P$PBF_Q2_01	255 (-)	1.000	0.998	tAAAGG	PBF
P$NAC6_01	265 (-)	0.725	0.770	TAACAt	NAC6
P$HSFA4A_01	274 (-)	0.947	0.916	ATTAGg	HSFA4A
P$AT3G51080_01	278 (+)	0.934	0.954	GGAAGaa	At3g51080
P$ASR1_01	284 (-)	0.973	0.973	AGGGT	ASR-1
P$O2_02	288 (-)	1.000	0.940	ttaACGTGtt	Opaque-2
P$NAC6_01	289 (-)	0.864	0.886	TAACGt	NAC6
P$TGA1A_01	289 (-)	0.731	0.762	tAACGTg	TGA1A
P$OCSBF1_01	291 (-)	1.000	1.000	ACGTG	OCSBF-1
P$ABI5_Q2	291 (+)	1.000	0.979	ACGTGt	ABI5
P$GT1_Q6	295 (+)	0.932	0.953	GTTTAta	GT-1
P$HMG1_01	303 (-)	0.861	0.876	gataACAAT	HMG-1
P$ATMYB15_Q2	305 (+)	1.000	1.000	TAACAa	Myb-15
P$HSFA4A_01	309 (-)	1.000	1.000	AATAGt	HSFA4A
P$GT1_Q6_01	312 (-)	0.953	0.899	agttaaaAGATA	GT-1
P$GT1_Q6	313 (+)	0.984	0.959	GTTAAaa	GT-1
P$GT1_Q6_01	336 (+)	0.970	0.915	TATATtcttata	GT-1
P$PEND_02	338 (+)	1.000	0.909	taTTCTTata	PEND
P$AT2G41690_01	366 (-)	0.946	0.948	gTTCGA	At2g41690
P$LIM1_01	368 (-)	0.813	0.817	tcgatgtGGTAG	LIM1
P$AT5G54070_01	369 (-)	0.984	0.944	CGATGt	At5g54070
P$GT1_Q6	376 (+)	0.948	0.964	GTAGAta	GT-1
P$TEIL_01	383 (+)	0.969	0.915	ATGGAact	TEIL

P$CBF1_01	387 (-)	0.923	0.749	aactCGGCCa	CBF1
P$ESR1_01	387 (-)	1.000	0.663	aactCGGCCa	ESR1
P$HSFA4A_01	396 (-)	1.000	0.914	AATAGa	HSFA4A
P$SQUA_01	399 (-)	0.887	0.844	agAAATAtag	SQUA
P$SBF1_01	417 (-)	1.000	0.941	gatATTAAgaatta	SBF-1
P$PEND_02	420 (-)	1.000	0.888	attAAGAAtt	PEND
P$PEND_01	422 (+)	1.000	0.958	tAAGAAtt	PEND
P$NAC6_01	428 (-)	0.805	0.837	TTAAGt	NAC6
P$MADSA_Q2	437 (-)	1.000	0.925	tttccATTTTctggtt	MADS-A
P$RIN_Q2	440 (+)	0.883	0.921	ccatTTTCTgg	RIN
P$SQUA_01	441 (+)	0.917	0.859	catTTTCTgg	SQUA
P$LIM1_01	442 (-)	0.813	0.816	attttctGGTTG	LIM1
P$HSFA2_01	445 (-)	0.979	0.980	tTCTGG	HSFA2
P$GT1_Q6_01	452 (+)	0.968	0.908	TGTATagttagc	GT-1
P$C1_Q2	452 (-)	1.000	0.962	tgtaTAGTTag	C1
P$ATHSFA1D_01	467 (-)	0.918	0.923	TTTAGt	ATHSFA1D
P$NAC6_01	468 (-)	0.723	0.768	TTAGTt	NAC6
P$GAMYB_Q2	468 (-)	0.987	0.908	ttaGTTGCctatt	GAMYB
P$AG_01	471 (+)	0.872	0.873	gttgCCTATttagggtta	AG
P$AG_03	471 (+)	0.872	0.873	gttgCCTATttagggtta	AG
P$AG_01	471 (-)	0.994	0.835	gttgcctatTTAGGgtta	AG
P$AG_03	471 (-)	0.994	0.835	gttgcctatTTAGGgtta	AG
P$AGL15_02	472 (-)	0.913	0.916	TTGCCtatttagg	AGL15
P$HSFA4A_01	475 (+)	1.000	0.964	cCTATT	HSFA4A
P$RIN_Q2	475 (+)	0.925	0.933	cctaTTTAGgg	RIN
P$RIN_01	476 (-)	1.000	0.888	CTATTtaggg	RIN
P$SQUA_01	476 (+)	0.748	0.847	ctaTTTAGgg	SQUA
P$ATHSFA1D_01	479 (-)	0.918	0.908	TTTAGg	ATHSFA1D
P$ASR1_01	482 (-)	0.973	0.973	AGGGT	ASR-1

P$GAMYB_Q2	482 (-)	0.982	0.907	aggGTTATtaatg	GAMYB
P$SBF1_01	485 (-)	1.000	0.915	gttATTAAtgattt	SBF-1
P$PEND_01	495 (-)	1.000	0.972	atTTCTTa	PEND
P$PEND_02	495 (+)	1.000	0.876	atTTCTTagt	PEND
P$NAC6_01	500 (-)	0.723	0.768	TTAGTt	NAC6
P$ATHSFA1D_01	504 (-)	0.918	0.923	TTTAGt	ATHSFA1D
P$NAC6_01	505 (-)	0.723	0.768	TTAGTt	NAC6
P$AT2G15660_01	509 (+)	0.935	0.922	TTTTCttgaaa	At2g15660
P$RIN_Q2	512 (-)	0.879	0.931	tcTTGAAatgg	RIN
P$ASR1_01	520 (-)	1.000	1.000	TGGGT	ASR-1
P$SBF1_01	524 (+)	1.000	0.915	ttataaTTAATtac	SBF-1
P$GT1_Q6	531 (-)	0.888	0.923	taATTAC	GT-1
P$PEND_02	533 (+)	0.863	0.878	atTACTTata	PEND
P$NAC6_01	534 (-)	0.894	0.911	TTACTt	NAC6
P$O2_Q2	541 (+)	0.906	0.892	tatctAACGTgaa	Opaque-2
P$O2_02	544 (-)	1.000	0.953	ctaACGTGaa	Opaque-2
P$NAC6_01	545 (-)	0.864	0.886	TAACGt	NAC6
P$TGA1A_01	545 (-)	0.731	0.762	tAACGTg	TGA1A
P$O2_Q4	545 (-)	1.000	0.921	taACGTGaaaa	Opaque-2
P$TGA1A_01	546 (+)	0.731	0.762	aACGTGa	TGA1A
P$OCSBF1_01	547 (-)	1.000	1.000	ACGTG	OCSBF-1
P$NAC6_01	547 (+)	0.811	0.842	aCGTGA	NAC6
P$GT1_Q6	549 (+)	1.000	0.971	GTGAAaa	GT-1
P$HSFA4A_01	552 (-)	0.926	0.933	AAAAGt	HSFA4A
P$PBF_Q2	552 (+)	0.963	0.964	aAAAGT	PBF
P$PBF_Q2_01	552 (-)	0.998	0.998	aAAAGT	PBF
P$AZF3_01	554 (+)	0.937	0.947	aAGTTTttttt	AZF3
P$HMG1_01	570 (+)	0.861	0.881	TTTGTcttc	HMG-1
P$ID1_01	570 (+)	1.000	0.944	tTTGTCttctt	ID1

P$GT1_Q6_01	572 (+)	0.951	0.908	TGTCTtcttatt	GT-1
P$PEND_02	574 (+)	1.000	0.937	tcTTCTTatt	PEND
P$C1_Q2	593 (-)	0.978	0.961	gagaGGGTTtg	C1
P$ASR1_01	596 (-)	0.973	0.973	AGGGT	ASR-1
P$HMG1_01	600 (+)	0.861	0.849	TTTGTtctt	HMG-1
P$GAMYB_Q2	600 (-)	0.955	0.927	tttGTTCTtgtcg	GAMYB
P$HMG1_01	603 (+)	0.865	0.862	GTTCTtgtc	HMG-1
P$PBF_Q2_01	626 (-)	0.998	0.995	tAAAGT	PBF
P$ID1_01	631 (-)	1.000	0.913	tttacGACAAa	ID1
P$NAC6_01	632 (-)	1.000	0.854	TTACGa	NAC6
P$TGA1A_01	633 (+)	0.808	0.838	tACGACa	TGA1A
P$NAC6_01	660 (+)	0.805	0.837	aCTTAA	NAC6
P$NAC6_01	667 (+)	0.805	0.837	aCTTAA	NAC6
P$GAMYB_Q2	670 (+)	0.987	0.903	taagaGCAACagc	GAMYB
P$MYBAS1_01	673 (+)	0.999	0.993	gaGCAACagct	MYBAS1
P$RAV1_01	673 (+)	1.000	0.957	gagCAACAgcta	RAV1
P$HSFA4A_01	681 (+)	1.000	0.910	gCTATT	HSFA4A
P$MADSA_Q2	684 (-)	1.000	0.827	attcaATTTTttttta	MADS-A
P$SQUA_01	687 (+)	1.000	0.867	caaTTTTTtt	SQUA
P$GT1_Q6_01	690 (+)	1.000	0.982	TTTTTttttatc	GT-1
P$AT3G51080_01	696 (-)	0.999	0.999	ttTATCC	At3g51080
P$AG_01	696 (-)	1.000	0.829	tttatcccaTTTGGatag	AG
P$AG_03	696 (-)	1.000	0.829	tttatcccaTTTGGatag	AG
P$ATHSFA1D_01	714 (+)	0.918	0.908	cCTAAA	ATHSFA1D
P$HSFA4A_01	717 (-)	0.926	0.933	AAAAGt	HSFA4A
P$PBF_Q2	717 (+)	0.963	0.964	aAAAGT	PBF
P$PBF_Q2_01	717 (-)	0.998	0.998	aAAAGT	PBF
P$CBNAC_01	738 (-)	1.000	1.000	AAGCAa	CBNAC
P$NAC6_01	738 (+)	0.724	0.770	aAGCAA	NAC6

P$NAC6_01	747 (+)	0.805	0.837	aCTTAA	NAC6
P$AT1G18860_01	747 (+)	1.000	0.960	acTTAACta	At1g18860
P$AT5G15130_01	747 (+)	1.000	0.961	acTTAACta	At5g15130
P$AG_01	765 (+)	1.000	0.841	tctaCCAAAaatgtgaag	AG
P$AG_03	765 (+)	1.000	0.841	tctaCCAAAaatgtgaag	AG
P$MADSA_Q2	766 (+)	1.000	0.831	ctaccaAAAATgtgaa	MADS-A
P$MADSB_Q2	767 (+)	1.000	0.932	taccAAAAAtgtgaa	MADS-B
P$HSFA2_01	769 (+)	1.000	1.000	CCAAAa	HSFA2
P$PEND_02	782 (-)	1.000	0.914	gatAAGAAaa	PEND
P$GT1_Q6_01	782 (-)	0.996	0.937	gataagaAAATA	GT-1
P$GT1_Q6_01	792 (-)	0.996	0.909	tattaaaACATA	GT-1
P$AZF3_01	796 (-)	0.934	0.947	aaaacATATTc	AZF3
P$AT2G41690_01	809 (-)	0.946	0.948	gTTCGA	At2g41690
P$BBM_01	814 (-)	0.622	0.691	agtgTCGCTa	BBM
P$PLT1_01	814 (-)	0.616	0.671	agtgTCGCTa	PLT1
P$BBM_01	816 (+)	0.790	0.700	tGTCGCtagt	BBM
P$PLT1_01	816 (+)	0.791	0.692	tGTCGCtagt	PLT1
P$NAC6_01	822 (+)	0.894	0.765	tAGTAA	NAC6
P$GT1_Q6	824 (+)	0.888	0.923	GTAATta	GT-1
P$GT1_Q6	825 (-)	0.888	0.923	taATTAC	GT-1
P$NAC6_01	828 (-)	0.894	0.765	TTACTa	NAC6
P$ATHSFA1D_01	830 (+)	0.902	0.908	aCTACC	ATHSFA1D
P$RAV1_01	832 (+)	1.000	0.937	tacCAACAtgaa	RAV1
P$HSFA2_01	834 (+)	0.948	0.952	CCAACa	HSFA2
P$GT1_Q6_01	842 (-)	0.996	0.923	aataatcACATA	GT-1
P$GT1_Q6	844 (-)	0.888	0.923	taATCAC	GT-1
P$RIN_Q2	863 (+)	0.950	0.940	ccatTTTATgt	RIN
P$SQUA_01	864 (+)	0.961	0.869	catTTTATgt	SQUA
P$WRKY_Q2	865 (+)	0.919	0.783	attttaTGTCA	WRKY

P$O2_Q2	867 (-)	0.883	0.888	tttATGTCatata	Opaque-2
P$NAC6_01	868 (-)	0.810	0.841	TTATGt	NAC6
P$TGA1A_01	869 (+)	0.808	0.838	tATGTCa	TGA1A
P$AZF3_01	879 (+)	0.934	0.955	aAATATatttt	AZF3
P$GT1_Q6	893 (-)	0.948	0.935	ttTATAC	GT-1
P$BPC1_Q2	899 (-)	1.000	1.000	cTTTCT	BPC1
P$GT1_Q6_01	902 (+)	0.968	0.931	TCTATaattaat	GT-1
P$ARR10_01	919 (+)	1.000	1.000	AGATCcg	ARR10
P$AT2G41690_01	923 (+)	1.000	0.974	CCGAAt	At2g41690
P$TGA1A_01	930 (+)	0.833	0.859	tACCTCa	TGA1A
P$LIM1_01	932 (+)	0.845	0.818	CCTCAaatgtac	LIM1
P$TEIL_01	938 (+)	1.000	0.940	ATGTAcat	TEIL
P$TEIL_01	938 (-)	1.000	0.940	atgTACAT	TEIL
P$O2_03	943 (+)	0.800	0.799	CATGAtttga	Opaque-2
P$AZF3_01	955 (+)	0.937	0.921	cACTATcattt	AZF3
P$HSFA4A_01	956 (+)	0.899	0.908	aCTATC	HSFA4A
P$O2_03	960 (-)	0.800	0.818	tcattTTATC	Opaque-2
P$HSFA4A_01	971 (-)	0.926	0.897	AAAAGg	HSFA4A
P$PBF_Q2	971 (+)	1.000	1.000	aAAAGG	PBF
P$PBF_Q2_01	971 (-)	1.000	1.000	aAAAGG	PBF
P$GT1_Q6_01	977 (+)	0.968	0.933	TCTATtgttatc	GT-1
P$HSFA4A_01	977 (+)	1.000	0.914	tCTATT	HSFA4A
P$HMG1_01	980 (+)	0.861	0.876	ATTGTtatc	HMG-1
P$GAMYB_Q2	980 (-)	0.982	0.933	attGTTATcgcct	GAMYB
P$ATMYB15_Q2	981 (-)	1.000	1.000	tTGTTA	Myb-15
P$ALFIN1_Q2	997 (+)	0.970	0.947	ttaaagGTGGTgtcg	Alfin1
P$PBF_Q2	998 (+)	1.000	0.986	tAAAGG	PBF
P$PBF_Q2_01	998 (-)	1.000	0.998	tAAAGG	PBF
P$CBF1_01	1006 (-)	0.777	0.712	gtgtCGGTAc	CBF1

P$DREB1B_01	1008 (-)	1.000	1.000	GTCGG	DREB1B
P$CBF1_01	1015 (+)	1.000	0.822	cTGCCGatat	CBF1
P$GT1_Q6	1022 (-)	0.948	0.964	taTCCAC	GT-1
P$ASR1_01	1027 (+)	0.973	0.973	ACCCT	ASR-1
P$ARR10_01	1035 (-)	0.996	0.975	caAATCT	ARR10
P$NAC6_01	1045 (-)	0.751	0.792	ATACGt	NAC6
P$GAMYB_Q2	1046 (-)	1.000	0.931	tacGTTGTtgttt	GAMYB
P$AT5G54070_01	1048 (-)	1.000	0.958	CGTTGt	At5g54070
P$HMG1_01	1049 (+)	1.000	0.925	GTTGTtgtt	HMG-1
P$GAMYB_Q2	1049 (-)	1.000	0.953	gttGTTGTttcag	GAMYB
P$GAMYB_Q2	1058 (-)	0.947	0.907	tcaGTTTTtactt	GAMYB
P$GT1_Q6	1062 (-)	1.000	0.971	ttTTTAC	GT-1
P$NAC6_01	1065 (-)	0.894	0.911	TTACTt	NAC6
P$DOF3_01	1065 (-)	0.994	0.988	ttACTTTtttt	Dof3
P$DOF_Q2	1065 (-)	1.000	1.000	ttaCTTTTttt	Dof
P$PBF_Q2	1067 (-)	0.963	0.964	ACTTTt	PBF
P$PBF_Q2_01	1067 (+)	0.998	0.998	ACTTTt	PBF
P$HSFA4A_01	1067 (+)	0.926	0.933	aCTTTT	HSFA4A
P$SQUA_01	1068 (+)	1.000	0.855	cttTTTTTcc	SQUA
P$GT1_Q6_01	1071 (+)	1.000	0.895	TTTTTccttatt	GT-1
P$AT3G51080_01	1071 (-)	1.000	1.000	ttTTTCC	At3g51080
P$LIM1_01	1077 (-)	0.977	0.934	cttatttTGTGG	LIM1
P$HSFA2_01	1083 (-)	0.998	0.998	tTGTGG	HSFA2
P$BBM_01	1093 (-)	0.637	0.701	attgGCCCGt	BBM
P$PLT1_01	1093 (-)	0.640	0.689	attgGCCCGt	PLT1
P$AZF3_01	1105 (-)	1.000	0.957	aatacATACTa	AZF3
P$HSFA4A_01	1112 (+)	1.000	1.000	aCTATT	HSFA4A
P$ATMYB15_Q2	1116 (-)	1.000	1.000	tTGTTA	Myb-15
P$HSFA4A_01	1123 (+)	1.000	0.964	cCTATT	HSFA4A

P$ASR1_01	1129 (-)	1.000	1.000	TGGGT	ASR-1
P$SBF1_01	1132 (+)	1.000	0.932	gttcagTTAATata	SBF-1
P$GT1_Q6_02	1136 (-)	1.000	0.994	aGTTAAtatatg	GT-1
P$GT1_Q6	1137 (+)	0.984	0.989	GTTAAta	GT-1
P$CBF1_01	1145 (+)	1.000	0.828	aTGCCGttta	CBF1
P$TGA1A_01	1146 (-)	0.794	0.814	tGCCGTt	TGA1A
P$GT1_Q6	1150 (+)	0.932	0.924	GTTTAaa	GT-1
P$GAMYB_Q2	1155 (+)	0.982	0.953	aatcaATAACaaa	GAMYB
P$ATMYB15_Q2	1161 (+)	1.000	1.000	TAACAa	Myb-15
P$AT5G54070_01	1163 (+)	0.942	0.905	aCAAAG	At5g54070
P$HMG1_01	1168 (-)	0.848	0.877	gagaAAAAC	HMG-1
P$DOF3_01	1173 (-)	0.994	0.987	aaACTTTattc	Dof3
P$PBF_Q2_01	1175 (+)	0.998	0.995	ACTTTa	PBF
P$GT1_Q6_01	1177 (+)	0.974	0.918	TTTATtcttata	GT-1
P$PEND_02	1179 (+)	1.000	0.909	taTTCTTata	PEND
P$NAC6_01	1193 (-)	0.861	0.884	TTACAt	NAC6
P$SBF1_01	1198 (+)	1.000	0.934	ttctaaTTAAAaag	SBF-1
P$GT1_Q6_01	1199 (-)	1.000	0.927	tctaattAAAAA	GT-1
P$DOF2_01	1204 (+)	1.000	0.991	ttaaAAAGCgg	Dof2
P$SED_Q2	1207 (-)	0.985	0.985	aAAAGCggat	SED
P$PBF_Q2	1207 (+)	0.963	0.964	aAAAGC	PBF
P$AT3G51080_01	1213 (+)	0.933	0.953	GGATGaa	At3g51080
P$GT1_Q6_01	1214 (-)	0.994	0.908	gatgaacAAAGA	GT-1
P$DOF3_01	1217 (+)	0.992	0.987	gaacAAAGAaa	Dof3
P$AT5G54070_01	1219 (+)	0.942	0.905	aCAAAG	At5g54070
P$MADSA_Q2	1219 (+)	0.800	0.835	acaaagAAAAAggaaa	MADS-A
P$DOF_Q2	1223 (+)	1.000	0.996	agaAAAAGgaa	Dof
P$HSFA4A_01	1226 (-)	0.926	0.897	AAAAGg	HSFA4A
P$PBF_Q2	1226 (+)	1.000	1.000	aAAAGG	PBF

P$PBF_Q2_01	1226 (-)	1.000	1.000	aAAAGG	PBF
P$ID1_01	1226 (-)	0.855	0.888	aaaagGAAAAa	ID1
P$AT3G51080_01	1230 (+)	1.000	1.000	GGAAAaa	At3g51080
P$DOF_Q2	1230 (+)	1.000	0.994	ggaAAAAGaag	Dof
P$AGL1_02	1245 (+)	0.967	0.905	tATACCaaaaccggtgac	AGL1
P$AG_01	1245 (+)	1.000	0.887	tataCCAAAaccggtgac	AG
P$AGL1_01	1245 (+)	1.000	0.918	tataCCAAAaccggtgac	AGL1
P$AG_03	1245 (+)	1.000	0.887	tataCCAAAaccggtgac	AG
P$HSFA2_01	1249 (+)	1.000	1.000	CCAAAa	HSFA2
P$HSFA1E_01	1255 (+)	0.910	0.926	cCGGTG	HSFA1E
P$AT2G41690_01	1300 (-)	0.945	0.946	gATCGG	At2g41690
P$AT4G12750_01	1300 (-)	1.000	0.943	gaTCGGTttc	At4g12750
P$GT1_Q6	1306 (-)	0.932	0.924	ttTCAAC	GT-1
P$NAC6_01	1317 (+)	1.000	0.835	gCGTAA	NAC6
P$PLT1_01	1321 (-)	0.575	0.668	aatgGTGTCg	PLT1
P$CBF1_01	1325 (+)	0.775	0.761	gTGTCGacgc	CBF1
P$C1_Q2	1342 (-)	0.996	0.961	gtgtTGGTTga	C1
P$HSFA2_01	1343 (-)	0.948	0.952	tGTTGG	HSFA2
P$GAMYB_01	1344 (-)	1.000	0.953	gttGGTTG	GAMYB
P$GAMYB_Q2	1345 (-)	0.921	0.904	ttgGTTGAtgata	GAMYB
P$AT4G36620_01	1346 (-)	1.000	1.000	TGGTTgat	At4g36620
P$AT5G54070_01	1359 (+)	1.000	0.910	tCAACG	At5g54070
P$CBF1_01	1359 (-)	0.903	0.801	tcaaCGGCGt	CBF1
P$PLT1_01	1363 (+)	0.789	0.656	cGGCGTgatc	PLT1
P$HSFA2_01	1378 (+)	0.941	0.945	CCATCa	HSFA2
P$TEIL_01	1385 (+)	0.976	0.930	ATGCAcgt	TEIL
P$O2_02	1386 (+)	1.000	0.940	tgCACGTgcg	Opaque-2
P$TGA1A_Q2	1386 (+)	1.000	0.978	tgCACGTgcg	TGA1a
P$PIF1_01	1386 (-)	1.000	0.971	tgCACGTgcg	PIF1

P$O2_02	1386 (-)	1.000	0.936	tgcACGTGcg	Opaque-2
P$TGA1A_Q2	1386 (-)	1.000	0.978	tgcACGTGcg	TGA1a
P$PIF1_01	1386 (+)	1.000	0.971	tgcACGTGcg	PIF1
P$BHLH66_01	1387 (+)	1.000	1.000	gCACGTgc	bHLH66
P$BHLH66_01	1387 (-)	1.000	1.000	gcACGTGc	bHLH66
P$OCSBF1_01	1388 (+)	1.000	1.000	CACGT	OCSBF-1
P$OCSBF1_01	1389 (-)	1.000	1.000	ACGTG	OCSBF-1
P$BZR1_01	1390 (+)	1.000	1.000	CGTGCg	BZR1
P$BPC1_Q2	1401 (-)	1.000	1.000	cTTTCT	BPC1
P$RIN_Q2	1404 (-)	0.881	0.916	tcTGTAAtttg	RIN
P$CBNAC_01	1415 (-)	1.000	1.000	AAGCAa	CBNAC
P$NAC6_01	1415 (+)	0.724	0.770	aAGCAA	NAC6
P$AG_01	1424 (-)	1.000	0.887	tatgactctTTTGGtaaa	AG
P$AG_03	1424 (-)	1.000	0.887	tatgactctTTTGGtaaa	AG
P$LIM1_01	1426 (-)	0.796	0.830	tgactctTTTGG	LIM1
P$HSFA2_01	1432 (-)	1.000	1.000	tTTTGG	HSFA2
P$PBF_Q2_01	1438 (-)	0.998	0.995	tAAAGT	PBF
P$HMG1_01	1442 (+)	1.000	0.872	GTTGTcccc	HMG-1
P$ALFIN1_Q2	1444 (-)	1.000	0.939	tgtcCCCACcactat	Alfin1
P$ASR1_01	1447 (+)	0.989	0.989	CCCCA	ASR-1
P$LIM1_01	1449 (+)	1.000	0.922	CCACCactatac	LIM1
P$HSFA2_01	1449 (+)	0.946	0.950	CCACCa	HSFA2
P$LIM1_01	1452 (+)	0.963	0.943	CCACTataccat	LIM1
P$HSFA2_01	1460 (+)	0.993	0.994	CCATAa	HSFA2
P$TGA1A_Q2_01	1487 (-)	0.905	0.892	aatcctttaatgaaGTCAAaaa	TGA1a
P$PBF_Q2	1490 (-)	1.000	0.986	CCTTTa	PBF
P$PBF_Q2_01	1490 (+)	1.000	0.998	CCTTTa	PBF
P$TGA1A_Q2_01	1493 (+)	0.793	0.863	ttaATGAAgtcaaaaacatttc	TGA1a
P$ATHSFA1D_01	1497 (-)	0.950	0.953	TGAAGt	ATHSFA1D

P$TGA1A_01	1497 (-)	0.833	0.834	tGAAGTc	TGA1A
P$TGA1A_01	1498 (+)	0.794	0.800	gAAGTCa	TGA1A
P$GT1_Q6	1501 (+)	0.980	0.957	GTCAAaa	GT-1
P$AT5G54070_01	1519 (-)	0.942	0.905	CTTTGt	At5g54070
P$HMG1_01	1520 (+)	0.861	0.849	TTTGTgttt	HMG-1
P$WRKY_Q2	1522 (+)	0.919	0.787	tgtgttTGTCA	WRKY
P$C1_Q2	1530 (+)	1.000	0.958	tcAACTAacca	C1
P$GAMYB_01	1531 (+)	0.978	0.935	CAACTaac	GAMYB
P$NAC6_01	1532 (+)	0.723	0.768	aACTAA	NAC6
P$WEREWOLF_Q2	1532 (-)	1.000	0.928	aaCTAACca	WEREWOLF
P$ATMYB46_Q6	1534 (+)	0.952	0.980	CTAACcaaat	Myb-46
P$ATMYB15_Q2	1535 (+)	0.989	0.991	TAACCa	Myb-15
P$RAV1_01	1552 (+)	1.000	0.958	tccCAACAaatt	RAV1
P$LIM1_01	1554 (+)	0.819	0.814	CCAACaaattta	LIM1
P$HSFA2_01	1554 (+)	0.948	0.952	CCAACa	HSFA2
P$RIN_Q2	1558 (+)	0.950	0.943	caaaTTTATgg	RIN
P$HSFA2_01	1563 (-)	0.993	0.994	tTATGG	HSFA2
P$WRKY_Q2	1566 (-)	0.900	0.886	TGGCTtagtaa	WRKY
P$NAC6_01	1571 (+)	0.894	0.765	tAGTAA	NAC6
P$GT1_Q6	1585 (+)	0.948	0.935	GTGTAaa	GT-1
P$GT1_Q6_01	1586 (-)	0.974	0.971	tgtaaaaATAAA	GT-1
P$GT1_Q6	1587 (+)	1.000	0.971	GTAAAaa	GT-1
P$GAMYB_Q2	1592 (+)	0.982	0.902	aataaATAACtcc	GAMYB
P$MYBAS1_01	1595 (+)	0.999	0.986	aaATAACtccc	MYBAS1
P$PLT1_01	1596 (-)	0.599	0.661	aataACTCCc	PLT1
P$ASR1_01	1603 (+)	0.989	0.989	CCCCA	ASR-1
P$LIM1_01	1605 (+)	0.977	0.905	CCACAaacgtga	LIM1
P$HSFA2_01	1605 (+)	0.998	0.998	CCACAa	HSFA2
P$O2_02	1608 (-)	1.000	0.939	caaACGTGac	Opaque-2

P$TGA1A_01	1610 (+)	0.731	0.762	aACGTGa	TGA1A
P$OCSBF1_01	1611 (-)	1.000	1.000	ACGTG	OCSBF-1
P$NAC6_01	1611 (+)	0.811	0.842	aCGTGA	NAC6
P$ATHSFA1D_01	1616 (+)	1.000	1.000	aCTACA	ATHSFA1D
P$NAC6_01	1639 (+)	0.861	0.884	aTGTAA	NAC6
P$NAC6_01	1648 (+)	0.723	0.768	aACTAA	NAC6
P$ATHSFA1D_01	1649 (+)	0.918	0.923	aCTAAA	ATHSFA1D
P$GAMYB_Q2	1650 (+)	0.982	0.903	ctaaaATAACact	GAMYB
P$AZF3_01	1653 (-)	0.916	0.924	aaataACACTa	AZF3
P$HSFA4A_01	1660 (+)	0.947	0.952	aCTAAT	HSFA4A
P$GT1_Q6	1665 (-)	0.888	0.923	taATTAC	GT-1
P$HMG1_01	1666 (-)	0.861	0.846	aattACAAT	HMG-1
P$ATMYB15_Q2	1676 (+)	1.000	1.000	TAACAa	Myb-15
P$AT5G54070_01	1678 (+)	0.942	0.905	aCAAAG	At5g54070
P$AT3G51080_01	1683 (+)	1.000	1.000	GGAAAaa	At3g51080
P$AZF3_01	1685 (-)	0.937	0.953	aaaaaAAACTa	AZF3
P$HSFA4A_01	1692 (-)	0.921	0.928	ACTAGt	HSFA4A
P$HSFA4A_01	1692 (+)	0.921	0.928	aCTAGT	HSFA4A
P$NAC6_01	1694 (+)	0.894	0.765	tAGTAA	NAC6
P$GT1_Q6	1696 (+)	1.000	0.971	GTAAAaa	GT-1
P$SQUA_01	1696 (-)	1.000	0.933	gtAAAAAtag	SQUA
P$HSFA4A_01	1701 (-)	1.000	0.964	AATAGg	HSFA4A
P$HSFA4A_01	1710 (-)	1.000	1.000	AATAGt	HSFA4A
P$NAC6_01	1712 (+)	0.894	0.765	tAGTAA	NAC6
P$HSFA4A_01	1717 (-)	0.932	0.938	AGTAGt	HSFA4A
P$GT1_Q6	1725 (+)	0.948	0.935	GTGTAaa	GT-1
P$GT1_Q6_01	1726 (-)	0.957	0.961	tgtaaaaAGAAA	GT-1
P$GT1_Q6	1727 (+)	1.000	0.971	GTAAAaa	GT-1
P$DOF_Q2	1727 (+)	1.000	0.995	gtaAAAAGaaa	Dof

P$PEND_02	1729 (-)	1.000	0.912	aaaAAGAAat	PEND
P$HSFA2_01	1744 (+)	0.993	0.994	CCATAa	HSFA2
P$PEND_02	1745 (-)	0.890	0.909	catAAAAAct	PEND
P$AT5G54070_01	1756 (-)	0.942	0.905	CTTTGt	At5g54070
P$HMG1_01	1757 (+)	0.861	0.881	TTTGTcttc	HMG-1
P$ID1_01	1757 (+)	1.000	0.912	tTTGTCttcgt	ID1
P$WRKY_Q2	1759 (+)	0.919	0.816	tgtcttCGTCA	WRKY
P$AT2G41690_01	1762 (-)	0.970	0.959	cTTCGT	At2g41690
P$TGA1A_01	1763 (+)	0.860	0.882	tTCGTCa	TGA1A
P$PBF_Q2	1771 (-)	0.963	0.964	ACTTTt	PBF
P$PBF_Q2_01	1771 (+)	0.998	0.998	ACTTTt	PBF
P$HSFA4A_01	1771 (+)	0.926	0.933	aCTTTT	HSFA4A
P$AG_01	1773 (+)	0.971	0.838	ttttCCGAAttagacctc	AG
P$AG_03	1773 (+)	0.971	0.838	ttttCCGAAttagacctc	AG
P$AT2G41690_01	1777 (+)	1.000	0.974	CCGAAt	At2g41690
P$ASR1_01	1795 (+)	0.962	0.962	CCCCT	ASR-1
P$ATHSFA1D_01	1797 (+)	0.950	0.938	cCTTCA	ATHSFA1D
P$PLT1_01	1802 (+)	0.828	0.662	aAGCGCaata	PLT1
P$PBF_01	1808 (+)	1.000	0.995	aatAAAAGtga	PBF
P$DOF3_01	1808 (+)	0.994	0.989	aataAAAGTga	Dof3
P$HSFA4A_01	1811 (-)	0.926	0.933	AAAAGt	HSFA4A
P$PBF_Q2	1811 (+)	0.963	0.964	aAAAGT	PBF
P$PBF_Q2_01	1811 (-)	0.998	0.998	aAAAGT	PBF
P$NAC6_01	1819 (+)	0.894	0.911	aAGTAA	NAC6
P$HSFA4A_01	1832 (-)	0.926	0.933	AAAAGt	HSFA4A
P$SED_Q2	1832 (-)	0.985	0.985	aAAAGTttat	SED
P$PBF_Q2	1832 (+)	0.963	0.964	aAAAGT	PBF
P$PBF_Q2_01	1832 (-)	0.998	0.998	aAAAGT	PBF
P$GT1_Q6	1844 (-)	0.948	0.935	ttTCTAC	GT-1

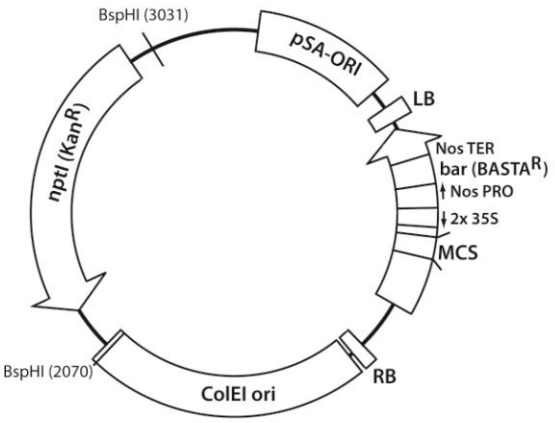
P$ATHSFA1D_01	1846 (+)	1.000	0.941	tCTACA	ATHSFA1D
P$HSFA2_01	1858 (+)	0.993	0.994	CCATAa	HSFA2
P$MYBAS1_01	1858 (+)	0.999	0.985	ccATAACtttc	MYBAS1
P$NAC6_01	1861 (-)	0.758	0.798	TAACTt	NAC6
P$BPC1_Q2	1864 (-)	1.000	1.000	cTTTCT	BPC1
P$NAC6_01	1871 (-)	0.758	0.798	TAACTt	NAC6
P$DOF3_01	1871 (-)	0.994	0.990	taACTTTactc	Dof3
P$PBF_Q2_01	1873 (+)	0.998	0.995	ACTTTa	PBF
P$DOF_Q2	1879 (-)	1.000	0.992	ctcCTTTTtaa	Dof
P$PBF_Q2	1881 (-)	1.000	1.000	CCTTTt	PBF
P$PBF_Q2_01	1881 (+)	1.000	1.000	CCTTTt	PBF
P$HSFA4A_01	1881 (+)	0.926	0.897	cCTTTT	HSFA4A
P$GT1_Q6	1885 (-)	0.932	0.924	ttTAAAC	GT-1
P$ASR1_01	1901 (+)	0.962	0.962	CCCCT	ASR-1
P$TGA1A_01	1906 (+)	0.833	0.848	cACTTCa	TGA1A
P$ATHSFA1D_01	1907 (+)	0.950	0.953	aCTTCA	ATHSFA1D
P$AT5G54070_01	1926 (+)	0.942	0.947	cCAAAG	At5g54070
P$DREB1B_01	1936 (+)	1.000	1.000	CCGAC	DREB1B
P$AT2G41690_01	1936 (+)	0.950	0.939	CCGACg	At2g41690
P$HSFA1E_01	1936 (+)	0.996	0.997	cCGACG	HSFA1E
P$AT5G54070_01	1936 (+)	0.933	0.939	cCGACG	At5g54070
P$AT4G01720_01	1943 (-)	1.000	1.000	gagTTTACc	At4g01720
P$AT3G51080_01	1952 (+)	1.000	1.000	GGAAAaa	At3g51080
P$HMG1_01	1967 (-)	1.000	0.953	aaaaACAAC	HMG-1
P$DOF1_01	1974 (-)	1.000	0.994	actCTTTAgtt	Dof1
P$ATHSFA1D_01	1978 (-)	0.918	0.923	TTTAGt	ATHSFA1D
P$NAC6_01	1979 (-)	0.723	0.768	TTAGTt	NAC6
P$HMG1_01	1982 (+)	0.848	0.874	GTTTTaatc	HMG-1
P$AT5G54070_01	1989 (+)	1.000	0.910	tCAACG	At5g54070

```
 P$AT2G41690_01        |     1992 (+) |  0.970 |  0.949 |
ACGAAa                      | At2g41690
 P$NAC6_01             |     1992 (+) |  0.782 |  0.818 |
aCGAAA                      | NAC6

Total sequences length = 1999

Total number of found sites = 470

Frequency of sites per nucleotide = 0.23512
```

8.3 Vector maps

The following vector map illustrations were made by Dr. Petra Stamm.

pGreenII 0229 backbone (5425bp)



MCS sequence - '35S::MCS'

5′ ... CGT CTC GAG GTC GAC GGT ATC GAT AAG CTT GAT ATC GAA TTC CTG CAG CCC GGG GGA TCC ACT AGT TCT AGA TAA ...3′
3′ ... GCA GAG CTC CAG CTG CCA TAG CTA TTC GAA CTA TAG CTT AAG GAC GTC GGG CCC CCT AGG TGA TCA AGA TCT ATT ... 5′

XhoI | SalI | ClaI | HindIII | EcoRV | EcoRI | PstI | SmaI/XmaI | BamHI | SpeI | XbaI | STOP

MCS sequence - '35S::GFP-MCS'

contains the ~800bp mGFP, cloned with XhoI and HindIII

5′ ... AAG CTT GAT ATC GAA TTC CTG CAG CCC GGG GGA TCC ACT AGT TCT AGA TAA ...3′
3′ ... TTC GAA CTA TAG CTT AAG GAC GTC GGG CCC CCT AGG TGA TCA AGA TCT ATT ... 5′

HindIII | EcoRV | EcoRI | PstI | SmaI/XmaI | BamHI | SpeI | XbaI | STOP

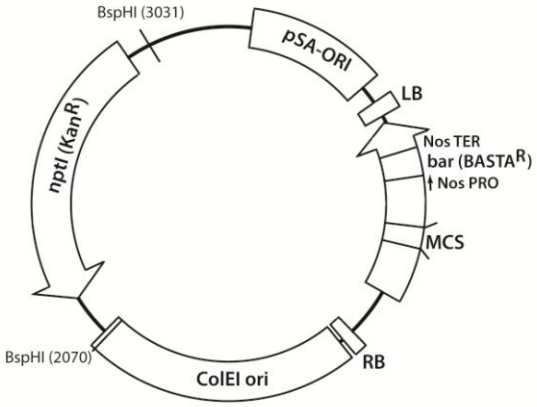
MCS sequence - '35S::MCS-GFP'

contains the 850bp mGFP-His$_{(6)}$ cloned with SpeI and XbaI

5′ ... CGT CTC GAG GTC GAC GGT ATC GAT AAG CTT GAT ATC GAA TTC CTG CAG CCC GGG GGA TCC ACT AGT ATG ...3′
3′ ... GCA GAG CTC CAG CTG CCA TAG CTA TTC GAA CTA TAG CTT AAG GAC GTC GGG CCC CCT AGG TGA TCA TAC ... 5′

XhoI | SalI | ClaI | HindIII | EcoRV | EcoRI | PstI | SmaI/XmaI | BamHI | SpeI | mGFP

pGreen HY105 backbone (~5500bp)



MCS sequence - 'MCS::GUS'

contains the 1921bp *uidA* gene, cloned with XbaI

5' ... GTC GAC GGT ATC GAT AAG CTT GAT ATC GAA TTC CTG CAG CCC GGG GGA TCC ACT AGT TCT AGA CAG TCC CTT ATG TTA ...3'
3' ... CAG CTG CCA TAG CTA TTC GAA CTA TAG CTT AAG GAC GTC GGG CCC CCT AGG TGA TCA AGA TCT GTC AGG GAA TAC AAT... 5'

SalI **ClaI** **HindIII** *EcoRV* **EcoRI** **PstI** **SmaI/XmaI** **BamHI** **SpeI** **XbaI** **GUS**

MCS sequence - 'MCS::GFP'

contains the 850bp mGFP-His$_{(6)}$ cloned with SpeI and XbaI

5' ... CGT CTC GAG GTC GAC GGT ATC GAT AAG CTT GAT ATC GAA TTC CTG CAG CCC GGG GGA TCC ACT AGT ATG ...3'
3' ... GCA GAG CTC CAG CTG CCA TAG CTA TTC GAA CTA TAG CTT AAG GAC GTC GGG CCC CCT AGG TGA TCA TAC ... 5'

XhoI *SalI* *ClaI* **HindIII** *EcoRV* **EcoRI** **PstI** **SmaI/XmaI** **BamHI** **SpeI** **mGFP**