Title
Transcriptional Regulatory Networks Controlling the Development of the Soybean Seed

Author
Jo, Leonardo

Publication Date
2021

Peer reviewed|Thesis/dissertation

Transcriptional Regulatory Networks Controlling the Development of the Soybean Seed

By

Leonardo Jo
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Plant Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Dr. John J. Harada, Chair

Dr. Siobhan Brady

Dr. Stacey Harmer

Committee in Charge
2021

Transcriptional Regulatory Networks Controlling the Development of the Soybean Seed

Abstract

Seed development in flowering plants is divided into two main contrasting developmental phases, morphogenesis and maturation. The morphogenesis phase is characterized by a series of cell division and differentiation events that establish the basic body plan of the embryo. Following morphogenesis, the seeds enter the maturation phase which is characterized by the accumulation of storage compounds and the embryo's acquisition of desiccation tolerance. Because seed development is a complex yet highly coordinated period of the plant life cycle, the temporal and spatial control of gene expression is crucial to ensure the proper development of the plant. The focus of my dissertation research was to investigate the complex network of transcription factors and their combinatorial efforts to control the onset of specific biological programs during the development of the soybean seed.

My first research topic investigated the function of four putative regulators of seed development in flowering plants: LEAFY COTYLEDON1 (LEC1), ABA-RESPONSIVE ELEMENT BINDING PROTEIN3 (AREB3), BASIC LEUCINE ZIPPER67 (bZIP67), and ABA INSENSITIVE3 (ABI3). I showed through genome-wide analyses of transcription factor binding sites that distinct transcription factor combinatorial interactions are controlling distinct biological programs in the soybean seed, such as embryo morphogenesis, photosynthesis, and seed storage protein accumulation. I also showed that these combinatorial interactions are assembled in *cis*-regulatory modules (CRMs) to control the expression of specific target genes. These distinct combinatorial interactions in CRMs are determined by the unique composition of DNA motifs in

the CRMs and the ability of transcription factors to physically interact with each other. I also explored the ability of CRMs to act as *cis*-acting enhancers.

The second topic of my dissertation described the role of the WRINKLED1 (WRI1) transcription factor in the regulation of lipid storage accumulation in soybean seeds. Lipid accumulation in seeds is a complex metabolic pathway that requires the action of many enzymes that act in processing carbohydrates into long chain fatty acids in addition to packaging triacyl-glycerol molecules into oil bodies. Genome-wide characterization of WRI1 binding sites revealed that this TF can bind to several genes that encode enzymes involved with every step of the fatty acid and triacylglycerol metabolic pathway. We also explored the collaboration between LEC1 and WRI1 in the regulation of genes involved with this metabolic process. Our results provided important and novel insights into the mechanisms of WRI1 in the control of lipid biosynthesis in soybean seeds.

Acknowledgement

I firmly believe that one's success is a direct result of the people that surrounds them. For that, I would like to dedicate this space to acknowledge all the amazing people I had the chance to interact and who made it possible for me to go through this amazing journey.

My family, Daniel Kum Suk Jo, Estela Yum Shim Im, Daniel Jo and Fernanda Jo and Emilyn Emy Matsumura for all their sacrifice and unconditional love and support.

My great mentor, John Harada who believed in me and for teaching me how to be a great scientist and a great teacher. I can't thank you enough for all the learning opportunities you have given me.

Julie Pelletier who taught and helped me in every single experiment. I could not have done it without your patience and support.

All the members of the Harada Lab. Rie Uzawa, Ssu-Wei Hsu, Russel Baden and every undergraduate student who worked in the lab for the great lab environment and for all scientific and emotional support whenever I needed it.

Bob Goldberg, for all valuable advice and for challenging and pushing me to become a better scientist.

My dissertation committee, Siobhan Brady and Stacey Harmer for all their career and scientific advice.

All the PBGG students. Your support was really important throughout the intense years of grad school. You also thought me the great value of community.

All the staff members in the Plant Biology Graduate Group and Plant Biology Department at UC Davis. Your assistance made my journey much easier.

Table of Contents

Chapter 1

Central role of the LEAFY COTYLEDON1 transcription factor in seed development

Jo, L., Pelletier, J.M. and Harada, J.J

This chapter was published as presented.
Jo, L., Pelletier, J.M. and Harada, J.J., 2019. Central role of the LEAFY COTYLEDON1 transcription factor in seed development. Journal of integrative plant biology, 61(5), pp.564-580.

Preface

The following attributes work within this chapter to the respective author(s)

Writing L.J., J.J.H.

Figures L.J. J.M.P.

JIPB Journal of Integrative Plant Biology

Central role of the LEAFY COTYLEDON1 transcription factor in seed development[FA]

Leonardo Jo, Julie M. Pelletier and John J. Harada*

Department of Plant Biology and Plant Biology Graduate Group, University of California, Davis, USA
doi: 10.1111/jipb.12806

Invited Expert Review

John J. Harada
*Correspondence:
jjharada@ucdavis.edu

Abstract Seed development is a complex period of the flowering plant life cycle. After fertilization, the three main regions of the seed, embryo, endosperm and seed coat, undergo a series of developmental processes that result in the production of a mature seed that is developmentally arrested, desiccated, and metabolically quiescent. These processes are highly coordinated, both temporally and spatially, to ensure the proper growth and development of the seed. The transcription factor, LEAFY COTYLEDON1 (LEC1), is a central regulator that controls several aspects of embryo and endosperm development, including embryo morphogenesis, photosynthesis, and storage reserve accumulation. Thus, LEC1 regulates distinct sets of genes at different stages of seed development. Despite its critical importance for seed development, an understanding of the mechanisms underlying LEC1's multifunctionality is only beginning to be obtained. Recent studies describe the roles of specific transcription factors and the hormones, gibberellic acid and abscisic acid, in controlling the activity and transcriptional specificity of LEC1 across seed development. Moreover, studies indicate that LEC1 acts as a pioneer transcription factor to promote epigenetic reprogramming during embryogenesis. In this review, we discuss the mechanisms that enable LEC1 to serve as a central regulator of seed development.

Edited by: Baocai Tan, Shandong University, China
Received Jan. 4, 2019; **Accepted** Mar. 16, 2019; **Online on** Mar. 27, 2019
FA: Free Access

Free Access

INTRICACIES OF SEED DEVELOPMENT

Overview of seed development

Seed development is a complex period of the flowering plant life cycle. As shown in Figure 1, the seed consists of three different regions, each with a distinct variation on a common genotype: diploid and filial embryo, triploid and filial endosperm, and diploid and maternal seed coat. Moreover, each region is comprised of distinct subregions, tissues, and cell types.

Seed development begins with the double fertilization of the egg and central cells of the embryo sac with two sperm cells that generate the embryo and endosperm, respectively (Goldberg et al. 1994). Fertilization also initiates seed coat development (Roszak and Kohler 2011).

Embryo and endosperm development can be divided temporally into two distinct phases: the morphogenesis phase, which is initiated immediately after fertilization, and the maturation phase, which partially overlaps and follows the morphogenesis phase (Figure 1). The morphogenesis phase is characterized by cell proliferation and differentiation that occur in both the embryo and endosperm. During this phase, the shoot and root apical meristems of the embryo are formed to set up the apical – basal plant axis, and the protoderm, ground meristem, and procambium develop as the tissue system progenitors that constitute the embryo's radial axis (reviewed by Lau et al.

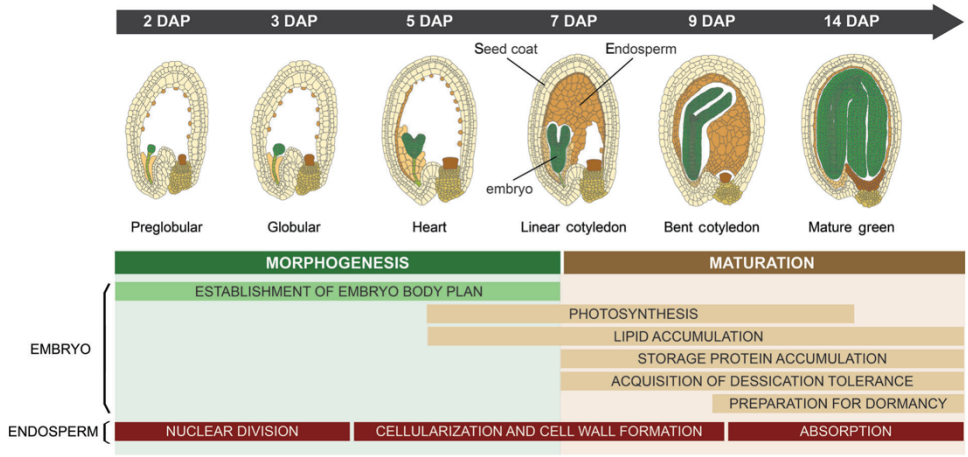


Figure 1. Overview of the major biological events that occur during seed development
Seed images diagram *Arabidopsis* seeds at the indicated stages and days after pollination (DAP). Bars indicate the morphogenesis and maturation phases and the major cellular processes that occur in embryos and endosperm.

2012; Palovaara et al. 2016). This basic body pattern which is established during embryogenesis is maintained throughout the sporophytic life cycle of the plant. The endosperm undergoes nuclear and cell proliferation, regionalization, and cell differentiation during the morphogenesis phase, and it develops into tissues that will provide nutrients for the developing embryo and/or postgerminative seedling (Li and Berger 2012).

By contrast, the maturation phase represents an interruption of the patterning, proliferation, and differentiation events that occur during the morphogenesis phase and that are reinitiated during seedling and vegetative development (Raz et al. 2001; Vicente-Carbajosa and Carbonero 2004). The maturation phase is characterized by the synthesis and massive accumulation of storage compounds, such as seed storage lipids and proteins (Harada 1997; Gutierrez et al. 2007; Baud et al. 2008). Storage compound accumulation results in cell expansion and a considerable increase in embryo cell size. It is also during the maturation phase that the embryo acquires the ability to survive desiccation that occurs at the latest stage of seed development through the accumulation of disaccharides, oligosaccharides, storage proteins, and late embryogenesis abundant proteins that preserve the integrity of membranes, proteins, and nucleic acids in the desiccated state (Angelovici et al. 2010; Leprince et al. 2017). Germination of the developing embryo is actively inhibited during the maturation phase, initially through accumulation of the hormone abscisic acid (ABA) and later through a reduction in water content (Kermode 1990). At the end of the maturation phase, the embryo and endosperm are developmentally arrested and metabolically quiescent, and they are typically maintained in this state until conditions favorable for germination are encountered.

Gene networks in seed development

The complexity of seed development suggests that the cellular processes that underlie specific seed functions must be highly coordinated both temporally and spatially. The onset and termination of these processes are controlled largely by changes in gene expression patterns. Therefore, understanding the mechanisms that control gene expression could aid in the development of strategies that can be used to modify the processes that occur during seed development and, potentially, improve seed quality in many important crop species.

The mRNA profiles of whole seeds and/or seed regions and subregions at different stages of development in several plant species have provided fundamental insights into the processes and regulatory

mechanisms that control seed development (Le et al. 2007; Benedito et al. 2008; Verdier and Thompson 2008; Xiang et al. 2011; Chen et al. 2012; Harada and Pelletier 2012; Belmonte et al. 2013; Terrasson et al. 2013; Becker et al. 2014; Chen et al. 2014; Khan et al. 2014; Li et al. 2014; Pradhan et al. 2014; Aghamirzaie et al. 2015; Gonzalez-Morales et al. 2016; Huang et al. 2017; Rangan et al. 2017). Gene expression patterns reflect spatial differences in seed regions and subregions and temporal differences in developmental stages. However, the most conspicuous change is a major reprogramming of gene expression that occurs in the embryo and endosperm during the transition between the morphogenesis and maturation phase of seed development (Verdier et al. 2008; Severin et al. 2010; Xiang et al. 2011; Chen et al. 2012; Belmonte et al. 2013; Chen et al. 2014). Many genes involved in patterning and morphological differentiation processes are preferentially expressed during the morphogenesis phase, whereas genes that are involved with seed storage macromolecule accumulation and desiccation tolerance are activated at the onset of the maturation phase. Although some aspects of gene expression are regulated posttranscriptionally in seeds (D'Ario et al. 2017), these findings suggest that transcriptional control mechanisms play major roles in regulating seed development.

An introduction to LEAFY COTYLEDON1

Many transcription factors have been shown to regulate biological processes during seed development (reviewed by Le et al. 2007; Verdier and Thompson 2008; Le et al. 2010; Jia et al. 2014; Pradhan et al. 2014; Baud et al. 2016; Devic and Roscoe 2016). Among these transcription factors, LEAFY COTYLEDON1 (LEC1) has been identified as a key, central regulator of seed development (Meinke 1992; Meinke et al. 1994; West et al. 1994; Lotan et al. 1998; Harada 2001; To et al. 2006; Braybrook and Harada 2008; Pelletier et al. 2017). LEC1 is a novel subunit of the nuclear factor Y (NF-Y) transcription factor that accumulates primarily in the embryo and endosperm, specifically during seed development (Figure 2A) (Lotan et al. 1998; Calvenzani et al. 2012; Gnesutta et al. 2017b). Although LEC1 has long been considered to be a central regulator of seed development, we are only beginning to understand the mechanisms by which LEC1 controls several aspects of seed development, including the biosynthesis of storage macromolecules, desiccation tolerance, photosynthesis, and hormone biosynthesis. In this review, we discuss the multifunctionality of LEC1 during seed development and recent findings that describe potential mechanisms by which LEC1 can regulate distinct biological processes across seed development.

LEC1 IS A KEY REGULATOR OF THE MATURATION PHASE

LEC1 is a central regulator of seed development that controls cellular processes that occur during the morphogenesis and maturation phases. Initial insights into LEC1 function were obtained through analyses of loss-of-function mutations of *Arabidopsis LEC1* that were identified in genetic screens for *embryo lethal* mutants (Harada 2001). Several characteristics of *lec1* mutants suggest that the transcription factor regulates several processes related to the maturation phase. First, LEC1 is required for embryos to acquire desiccation tolerance. Embryos with null mutations in *LEC1* die, because they do not survive maturation drying at the end of seed development (Meinke 1992; Meinke et al. 1994; West et al. 1994). Second, LEC1 is required for storage macromolecule accumulation. Storage protein and lipid accumulation are severely restricted in *lec1* mutants (Meinke 1992; Meinke et al. 1994; West et al. 1994). A genome-wide comparison of mRNA populations in wild type and *lec1* mutant seeds showed that the major difference in mRNA profiles is observed at the maturation phase of seed development (Pelletier et al. 2017). Genes involved with maturation processes, such as protein and lipid storage, desiccation tolerance, and seed dormancy, are downregulated in *lec1* mutant seeds. Third, postgerminative seedling development is suppressed during seed development by LEC1. The shoot apices of *lec1* mutant embryos are activated and possess leaf primordia, whereas wild type embryonic shoot apices are inactive and do not initiate leaf development (Meinke et al. 1994; West et al. 1994). One interpretation of these findings is that the maturation program prevents the precocious initiation of vegetative development during embryogenesis. Consistent with this interpretation, genes expressed seedling-specifically are prominently upregulated in *lec1* mutant embryos during the late stages of seed development (Pelletier et al. 2017). Thus, pleiotropic effects of the *lec1* mutation led to the conclusion that LEC1 is an essential

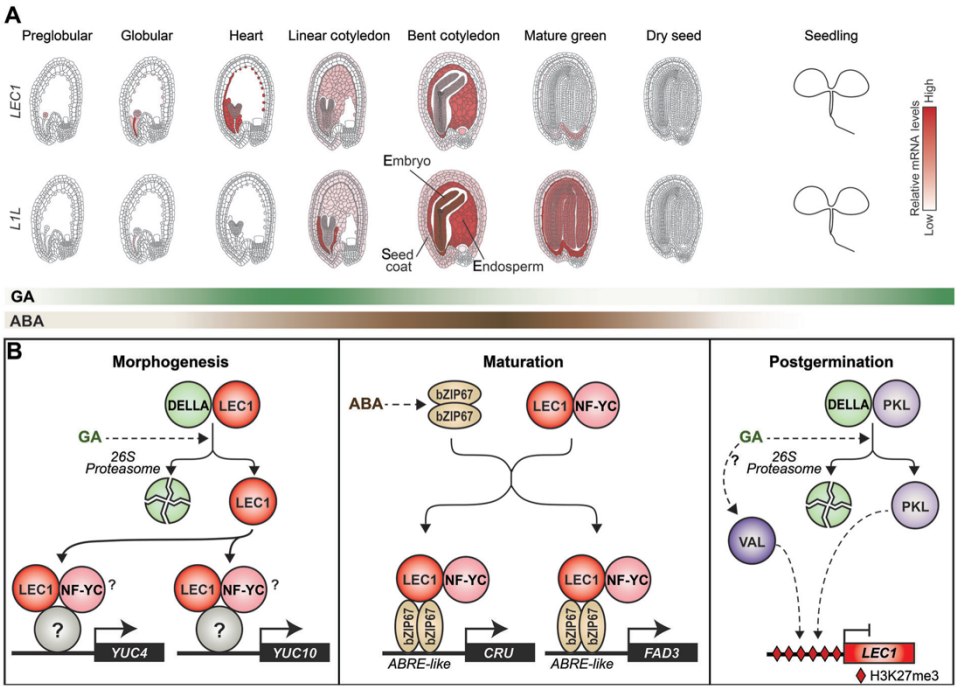


Figure 2. Modulation of LEC1 activity during seed development

(A) Heat map representations of *LEC1* and *L1L* mRNA levels in embryo, endosperm, and seed coat subregions during *Arabidopsis* seed development (top panel) and GA and ABA levels at the indicated stages of seed and postgerminative development, with darker colors indicating higher relative hormone levels (bottom panel). mRNA data are taken from Belmonte et al. (2013). **(B)** Mechanistic effects of GA and ABA on LEC1 activity. Morphogenesis panel. Because bioactive GA levels are high, DELLA is degraded, releasing LEC1 to activate gene encoding auxin biosynthetic enzymes, *YUC4* and *YUC10*, although the subunits with which LEC1 interacts is not known. Maturation panel. ABA levels are high, and the ABA-inducible transcription factor bZIP67 accumulates and forms a complex with a LEC1-NF-YC (or L1L-NF-YC) dimer. The complex binds *ABRE-like* DNA sequence motifs and activates maturation genes, such as *CRU* and *FAD3*. Postgermination panel. DELLA is degraded, because GA levels become high prior to and during germination and postgermination. PKL is released, resulting in an increase in H3K27me3 occupancy of the *LEC1* promoter and silencing of the *LEC1* gene. An increase in VAL activity, which is thought to be mediated by GA, also results in an increase in H3K27me3 occupancy.

regulator of the maturation phase (Meinke et al. 1994; West et al. 1994; Lotan et al. 1998; Harada 2001; To et al. 2006; Braybrook and Harada 2008; Lepiniec et al. 2018).

LEC1's role during the maturation phase was also demonstrated in gain-of-function genetic experiments. Ectopic expression of *LEC1* in *Arabidopsis* results in the upregulation of several genes involved in processes that occur during the maturation phase, such as seed storage proteins and lipid accumulation, desiccation tolerance, and seed dormancy (Lotan et al. 1998). For example, overexpression of *LEC1* in developing seeds results in the upregulation of key genes involved in fatty acid biosynthesis and storage and an increase in lipid content in a number of plant species (Kagaya et al. 2005; Mu et al. 2008; Tan et al. 2011; Elahi et al. 2016; Pelletier et al. 2017; Tang et al. 2018). These findings open the possibility that manipulating LEC1 expression might be useful to enhance the seed quality of crop plants.

The phenotypes induced by loss- and gain-of-function mutations suggest that LEC1 is a key regulator of the

maturation phase. Genome-wide characterization of LEC1 binding sites revealed that LEC1 can directly regulate several genes involved in processes that occur during the maturation phase of developing *Arabidopsis* and soybean seeds (Junker et al. 2012; Pelletier et al. 2017).

LEC1 has been implicated to have played a critical role in the evolution of the seed habit. In contrast to plant lineages that do not produce seeds, seed plant embryos undergo biochemical and physiological changes during the maturation phase that allow them to withstand maturation drying and metabolic quiescence and undergo the reinitiation of growth after germination. The processes that occur during the maturation phase account, in part, for the evolutionary success of seed plants (Steeves 1983; Harada 2001; Vicente-Carbajosa and Carbonero 2004). Thus, understanding the regulatory circuitry controlling seed maturation could provide insights into the mechanisms that underlie evolution of the seed habit. The requirement of LEC1 to regulate maturation processes opens the possibility that LEC1 may have played a critical role in the evolution of the maturation phase and the seed habit. Consistent with this possibility, phylogenetic analysis revealed that *LEC1*-type genes, which are shared among all spermatophytes, are first detected among basal land plant lineages in lycophytes (Xie et al. 2008; Kirkbride et al. 2013; Cagliari et al. 2014; Fang et al. 2017; Han et al. 2017), suggesting that LEC1 originated at least 30 million years before the appearance of seed plants in the fossil record. Based on their expression patterns, *LEC1*orthologs have been suggested to play roles in promoting desiccation tolerance and lipid accumulation in *Selaginella* (lycophyte) species and storage macromolecule accumulation in reproductive organs of the fern, *Adiantumcapillus-veneris* (Xie et al. 2008; Kirkbride et al. 2013; Fang et al. 2017; Han et al. 2017). Further studies of LEC1 function in basal plants could advance our understanding of seed plant evolution.

BEYOND MATURATION – ROLES FOR LEC1 IN OTHER ASPECTS OF SEED DEVELOPMENT

Importance of LEC1 for embryo morphogenesis

Although LEC1 is a key regulator of the maturation phase, several lines of evidence indicate that LEC1 also acts as a regulator during the morphogenesis phase of seed development. First, LEC1 is expressed within 24 h after fertilization, suggesting that it functions at the earliest stages of seed development (Figure 2A) (Lotan et al. 1998). Second, LEC1 is required to maintain embryonic suspensor identity early in seed development. The wild-type *Arabidopsis* suspensor is a transient structure comprised of a single file of six to eight cells. *lec1* mutant suspensors undergo abnormal cell divisions and often consist of more than eight cells (Lotan et al. 1998). Furthermore, combining the *lec1* mutation with mutations in *ABA INSENSITIVE3* (*ABI3*) or *FUSCA3* (*FUS3*) genes that encode other seed development regulators results in polyembryony, in which a second embryo proper forms from cells derived from proliferating suspensor cells (Vernon and Meinke 1994; Lotan et al. 1998). Thus, LEC1 is required to suppress the embryogenic potential of the suspensor early in embryo development. Third, LEC1 is required to specify cotyledon identity during embryogenesis (Meinke 1992; Meinke et al. 1994; West et al. 1994). *lec1* mutant embryo cotyledons, unlike wild type, undergo a heterochronic conversion in which they acquire leaf traits, such as trichomes on their adaxial surfaces and a cellular organization that is intermediate between cotyledons and leaves (Meinke et al. 1994; West et al. 1994). Consistent with this interpretation, trichome development is suppressed in plants overexpressing *LEC1* (Lotan et al. 1998; Huang et al. 2015a). Fourth, LEC1 regulates the expression of genes involved in embryo morphogenesis, including those encoding the transcription factors PHAVOLUTA and SCARECROW, and in auxin biosynthesis in *Arabidopsis* and soybean embryos early in seed development (Junker et al. 2012; Pelletier et al. 2017; Hu et al. 2018). Finally, a striking indication of LEC1's role in embryo morphogenesis is its ability to induce somatic embryo development in vegetative tissues of several plant species (Lotan et al. 1998; Lowe et al. 2003; Yang and Zhang 2010; Ledwon and Gaj 2011; Guo et al. 2013; Nic-Can et al. 2013; Orlowska et al. 2017). The mechanisms that underlie LEC1's ability to promote somatic embryogenesis are not fully understood, but it has been speculated that it acts to enhance embryogenic competence.

Involvement of LEC1 in photosynthesis and chloroplast development during seed development

Embryos of many angiosperm taxa possess chloroplasts that are highly shade adapted because of the light

quality to which they are exposed but that, nonetheless, photosynthesize during embryo development (reviewed by Puthur et al. 2013). In oilseeds, photosynthesis generates oxygen, which is limited in the internal tissues of the embryo, for mitochondria respiration, and it may aid in recycling carbon dioxide that is lost with each cycle of fatty acid elongation (Vigeolas et al. 2003; Rolletschek et al. 2005; Allen et al. 2009). LEC1 has been implicated to regulate photosynthesis and chloroplast biogenesis during seed development. *Arabidopsis lec1* mutants have a paler green coloration than wild-type embryos, suggesting that LEC1 promotes but is not absolutely required for proper chloroplast biogenesis during embryogenesis (Meinke 1992; West et al. 1994; Junker et al. 2012; Pelletier et al. 2017). LEC1 also transcriptionally activates the expression of representatives of most genes encoding the light-reaction components of photosystems I and II and of many other genes involved in chloroplast biogenesis in *Arabidopsis* and soybean embryos (Pelletier et al. 2017). These findings indicate a role for LEC1 in controlling photosynthesis and chloroplast biogenesis during seed development.

LEC1 plays a role in controlling endosperm development

mRNA profiles of *Arabidopsis* seeds revealed an extensive overlap in gene activity between embryo and endosperm subregions (Belmonte et al. 2013). Many of the same genes that are involved in processes that occur during the morphogenesis and maturation phases in the embryo are also expressed in the endosperm. The findings that chloroplasts and storage protein and oil bodies are present not only in the embryo but also in the endosperm support the functional significance of this overlap in gene expression programs (Belmonte et al. 2013).

LEC1 is expressed in the endosperm of many plant species, including *Arabidopsis*, maize, rapeseed, rice, and soybean (Figure 2A) (Lotan et al. 1998; Huang et al. 2009; Belmonte et al. 2013; Zhan et al. 2015; Pelletier et al. 2017; E et al. 2018). Moreover, *Arabidopsis* LEC1 directly activates genes that act both in the embryo and endosperm in processes related to the morphogenesis and maturation phases, suggesting the LEC1 regulates aspects of endosperm development, although the *lec1* mutant does not display obvious morphological defects in endosperm (Meinke 1992; Meinke et al. 1994; Lotan et al. 1998). Similarly, it was proposed that LEC1 can control endosperm development in rice through its interaction with AP2 transcription factors (Zhang and Xue 2013; Xu et al. 2016).

Thus, substantial evidence indicates that LEC1's role in seed development extends beyond simply control of the maturation phase. The ability of LEC1 to regulate cellular processes during both the morphogenesis and maturation phases and in distinct regions of the seed demonstrates that LEC1 is a central regulator of seed development.

TEMPORAL REGULATION OF LEC1 ACTIVITY BY HORMONES DURING SEED DEVELOPMENT

LEC1 regulates distinct processes at different stages of seed development, and its activity must be repressed after germination to promote vegetative development (Figure 2A, 2B). Thus, LEC1 activity must be highly temporally regulated during plant development.

Recent findings provide insight into the mechanisms by which LEC1 responds to the physiological cues that govern seed development. For example, gibberellic acid (GA) regulates LEC1 activity during seed development (Hu et al. 2018). As shown in Figure 2A, bioactive GA isoforms display a dynamic accumulation pattern, achieving highest levels during the early stages of seed development. In the absence of GAs, LEC1's ability to activate at least some of its target genes is repressed through its interaction with DELLA proteins, which are repressors of GA signaling pathways (Figure 2B). Bioactive GAs promote the degradation of DELLA proteins, releasing LEC1 to activate gene transcription. GAs have been shown to release LEC1 to activate the expression of *YUCCA* (*YUC*) genes involved in auxin biosynthesis (Hu et al. 2018).

Abscisic acid (ABA) accumulation during the late stages of seed development is at least partially responsible for the onset of the maturation phase and other developmental changes (Figure 2A) (Finkelstein et al. 2002; Gutierrez et al. 2007; Holdsworth et al. 2008; Nakashima and Yamaguchi-Shinozaki 2013). Given the importance of LEC1 and ABA in controlling the maturation phase, it is not surprising that ABA has been shown to augment LEC1's activation of genes involved in maturation. For example, ABA enhances the ability of LEC1 to activate the expression of the storage protein gene, *CRUCIFERIN* (*CRU*), and the lipid biosynthesis gene, *FATTY ACID DESATURASE3*

(*FAD3*), by promoting the activity of ABA RESPONSIVE ELEMENT BINDING (AREB) proteins, such as the transcription factor, BASIC LEUCINE ZIPPER67 (bZIP67) (Figure 2B) (Yamamoto et al. 2009; Mendes et al. 2013). It is not clear, however, if promotion results from enhanced *bZIP67* transcription or posttranslational phosphorylation of bZIP67, as has been shown to occur for another bZIP transcription factor, ABA INSENSITIVE5 (Lopez-Molina et al. 2001; Nakashima et al. 2009). The mechanistic relationship between LEC1 and bZIP transcription factors will be discussed, but it is likely that ABA modulates LEC1 function at least in part, by inducing AREB protein activity.

The central role of LEC1 in promoting seed development emphasizes a requirement to repress LEC1 activity during vegetative development. For example, ectopic *LEC1* expression in seedlings results in the repression of vegetative growth and the development of embryo-like seedlings (Lotan et al. 1998). Two lines of evidence indicate that chromatin conformation plays integral roles in regulating LEC1 expression postgermination (Jia et al. 2014; Pu and Sung 2015; Lepiniec et al. 2018). First, PICKLE (PKL), a CHD3 chromatin remodeling factor, negatively regulates *LEC1* expression and, therefore, embryonic programs during seedling development (Ogas et al. 1999; Dean Rider et al. 2003; Li et al. 2005). The seedling roots of *pkl* mutants display characteristics of embryos and accumulate storage lipids and proteins normally found in seeds. This phenotype results from the ectopic expression of LEC1 and other maturation regulators in *pkl* seedlings (Ogas et al. 1997; Henderson et al. 2004). Moreover, *pkl* mutants show spontaneous development of somatic embryos in postgerminative roots (Ogas et al. 1997). Second, the VIVIPAROUS ABI3-LIKE (VAL) proteins, also act to repress LEC1 activity during postgerminative development. *VAL1* and *VAL2* genes, also known as *HIGH-LEVEL EXPRESSION OF SUGAR-INDUCIBLE GENE2* (*HSI2*) and *HSI2-LIKE* genes, respectively, are B3 domain transcription factors that contain conserved CW and PHD domains frequently found in chromatin remodeling factors (Suzuki et al. 2007; Tsukagoshi et al. 2007). Monogenic *val* mutants do not display striking mutant phenotypes, however, *val1 val2* double mutants develop somatic embryos in shoot apical meristem regions of germinating seedlings (Suzuki et al. 2007). Although not normally active in wild-type seedlings, *LEC1* is expressed in *val1 val2* seedlings after germination, indicating that *VAL1* and *VAL2* inhibit embryonic development by repressing the expression of *LEC1* and other transcriptional regulators of maturation during seedling growth (Suzuki et al. 2007; Tsukagoshi et al. 2007).

Both PKL and VAL act epigenetically to repress *LEC1* expression (Figure 2B) (Jia et al. 2014; Pu and Sung 2015; Lepiniec et al. 2018). Repression of seed maturation genes by PKL is mediated through the trimethylation of the lysine 27 residue of histone H3 (H3K27me3), a repressive epigenetic mark, as indicated by the observation that *pkl* mutants display reduced H3K27me3 occupancy on *LEC1* postgermination (Zhang et al. 2008; Zhang et al. 2012). Similarly, *val1 val2* mutants show reduced accumulation of H3K27me3 and increased accumulation of active histone marks, such as histone H3 lysine 4 trimethylation, histone H3 acetylation, and histone H4 acetylation, in the promoter and coding regions of *LEC1* during seed germination (Zhou et al. 2013). VAL1 and VAL2 interact with HISTONE DEACETYLASE19 and 6, respectively, to inhibit *LEC1* activity (Zhou et al. 2013; Chhun et al. 2016). VAL2 binds with the promoter and coding/intron regions of *LEC1* to recruit HDA6 and suppress *LEC1* activity during seed germination.

The concerted actions of PKL and VAL1/VAL2 emphasize the importance of repressing the activities of LEC1 and other maturation regulators and, consequently, the embryonic program during vegetative development. GAs have been proposed to play an important role in controlling PKL and VAL activities (Figure 2B) (Ogas et al. 1997; Ogas et al. 1999; Suzuki et al. 2007; Zhang et al. 2014). An increase in GA levels prior to germination is responsible for breaking seed dormancy and promoting seed germination. In *pkl* mutant and *val1 val2* double mutant seedlings, the development of embryo-like structures is enhanced by GA biosynthesis inhibitors (Ogas et al. 1997; Suzuki et al. 2007). In addition, DELLA proteins interact with PKL to negatively regulate PKL activity (Figure 2B) (Zhang et al. 2014). Thus, GA induced degradation of DELLA proteins appears to activate PKL to repress embryonic gene expression after germination. The mechanism by which GAs influence VAL function remains to be determined. Nevertheless, the GA-mediated repression of LEC1 and other maturation regulators by PKL and VAL1 provides insight into the transition between seed and vegetative development in spermatophytes.

Together, these findings indicate that hormones play important roles in modulating LEC1 activity during seed development in response to physiological changes.

LEC1 REGULATES SEED DEVELOPMENT DIRECTLY AND INDIRECTLY THROUGH THE ACTIVATION OF OTHER KEY TRANSCRIPTION FACTORS

Genome-wide characterization of LEC1 occupancy coupled with gene expression analyses indicates that LEC1 can directly regulate many genes involved in the processes that occur during seed development (Junker et al. 2012; Pelletier et al. 2017). For instance, LEC1 directly regulates genes encoding enzymes involved in hormone biosynthesis and seed storage macromolecule accumulation. These studies also show that LEC1's involvement in controlling distinct processes during seed development may reflect, in part, its ability to regulate different sets of downstream transcription factors.

LEC1's function early in seed development is mediated, at least in part, through its direct activation of transcription factors involved in morphogenetic processes (Junker et al. 2012; Pelletier et al. 2017; Hu et al. 2018). For example, LEC1 directly regulates the transcription of the HD-ZIPIII transcription factors, *PHABULOSA* and *PHAVOLUTA*, that have been characterized as master regulators of apical fate early in embryogenesis and of *SCARECROW*, a key regulator of root architecture (Di Laurenzio et al. 1996; Smith and Long 2010; Pelletier et al. 2017). Moreover, LEC1 regulates genes involved in the biosynthesis of auxin, a hormone that plays key roles in embryonic pattern formation (Junker et al. 2012). Thus, LEC1 regulates the establishment of embryo body pattern by controlling the expression of genes involved in embryonic axis differentiation.

Among genes directly regulated by LEC1 are the "AFL" B3 domain transcription factors, ABI3, FUS3, and LEAFY COTYLEDON2 (LEC2), which are all key regulators of seed maturation (Braybrook and Harada 2008; Santos-Mendoza et al. 2008; Boulard et al. 2017). Single mutants for each gene display phenotypic similarities to *lec1* mutants and to each other (Finkelstein and Somerville 1990; Meinke 1992; Keith et al. 1994; Meinke et al. 1994; West et al. 1994; Harada 2001). The lack of redundancy among *AFL* genes indicates that they play similar though not identical roles during seed maturation. For example, *abi3* and *lec1* mutants but not *fus3* and *lec2* mutants have reduced sensitivity to exogenous ABA (To et al. 2006). *lec1, abi3* and *fus3* mutants are embryo lethal mutants, because they are desiccation intolerant, whereas *lec2* mutant embryos display only partial desiccation intolerance (Nambara et al. 1995; Harada 2001). LEC1 appears to act upstream of *ABI3, FUS3,* and *LEC2* in that *ABI3* and *FUS3* expression is reduced in *lec1* mutants, and overexpression of *LEC1* results in increased *ABI3* and *FUS3* expression in *Arabidopsis* seeds (Parcy et al. 1997; Kagaya et al. 2005; To et al. 2006; Mu et al. 2008; Pelletier et al. 2017). Moreover, *Arabidopsis ABI3, FUS3,* and *LEC2* are directly transcriptionally regulated by LEC1 (Pelletier et al. 2017).

Many maturation genes that are direct targets of LEC1 are also direct targets of ABI3 and FUS3 (Monke et al. 2012; Wang and Perry 2013; Pelletier et al. 2017). Thus, it appears that LEC1 activates both ABI3 and FUS3, and all three transcription factors act to promote maturation gene transcription during seed development. This type of network architecture is known as a feed-forward loop that can accelerate the response time of target gene expression following induction (Mangan and Alon 2003). Another potential example of a feed-forward loop is the relationship between LEC1 and WRINKLED1 (WRI1), another transcription factor that plays a key role in the maturation phase. WRI1 is a direct target of LEC1, and it is thought to directly regulate genes involved with fatty acid accumulation in *Arabidopsis* seeds that are also directly regulated by LEC1 (Baud et al. 2007; To et al. 2012; Pelletier et al. 2017). Thus, LEC1 works in concert with WRI1 to control fatty acid biosynthesis during seed development.

LEC1 indirectly controls seed development by regulating the expression of transcription factors that control independent developmental programs during seed development. However, LEC1's ability to directly regulate many of the structural genes in the regulatory network that are, in turn, regulated by its downstream transcription factor suggests a feed-forward mechanism of regulation that reinforces specific gene expression programs during seed development.

LEC1 FUNCTION IS MODULATED BY INTERACTIONS WITH OTHER TRANSCRIPTION FACTORS

The finding that LEC1 regulates distinct processes at different stages of development prompted the

question of how a single transcription factor can control different sets of genes. Genetic analyses suggested that LEC1 may interact synergistically with other transcription factors to regulate different processes during seed development (Parcy et al. 1997; To et al. 2006). Recent studies suggest that LEC1 acts sequentially during seed development to respond to different developmental signals by interacting with different combinations of transcription factors to alter the transcriptional specificity of LEC1 (Pelletier et al. 2017). In this section, we discuss LEC1's interactions with other transcription factors during seed development.

LEC1 as a subunit of a Nuclear Factor-Y transcription factor

LEC1 is a novel NF-YB subunit of the NF-Y complex, a transcription factor that is conserved among eukaryotes and binds the CCAAT DNA motif (Lotan et al. 1998; Calvenzani et al. 2012; Dolfini et al. 2012). In addition to NF-YB, the NF-Y complex is comprised of two other subunits, NF-YA and NY-YC (Petroni et al. 2012; Zhao et al. 2016). Different from other organisms, such as animals and yeast which contain only one gene for each subunit, plants possess NF-Y subunit gene families that consist of 8 to 14 members (Petroni et al. 2012; Zhao et al. 2016). This diversity of subunits offers the potential for the functional specialization of different combinations of NF-Y subunits (Siefers et al. 2009; Laloum et al. 2013). Seed plants possess two types of NF-YB subunits: the non-LEC1 type with B domains that are conserved across eukaryotes and the LEC1-type that, in *Arabidopsis*, consists of LEC1 (NF-YB9) and its paralog, LEC1-LIKE (L1L, NF-YB6), although LEC1 and L1L exhibit distinct accumulation patterns (Figure 2A) (Kwong et al. 2003b). LEC1-type subunits confer LEC1 activity whereas the non-LEC1 subunits do not (Kwong et al. 2003a; Lee et al. 2003). The B domains of LEC1-type subunits share sequence similarity with non-LEC1 type subunits, but they also possess unique amino acid residues. These unique residues are responsible for conferring LEC1 activity only to NF-Y complexes containing the LEC1-type subunits (Lee et al. 2003). The LEC1-type NF-YB subunits are found primarily in seed plants although they appear to have originated in land plant lineages in lycophytes (Xie et al. 2008; Kirkbride et al. 2013; Cagliari et al. 2014). Thus, non-LEC1 type and LEC1 type NF-YB subunits appear to have fundamentally different function.

The ability of NF-Y complexes containing non-LEC1-type NF-YB subunits to bind the CCAAT motif and to regulate gene transcription has been extensively studied in yeast, mammals and plants (Dolfini et al. 2012; Zhao et al. 2016; Myers and Holt 2018). The initial step in NF-Y complex formation involves dimerization between NF-YB and NF-YC through their histone-fold domains. NF-YC subunits possess nuclear localization sequences, whereas NF-YB subunits do not. Therefore, NF-YB/NF-YC dimers localize to the nucleus (Frontini et al. 2002; Kahle et al. 2005). The nuclear localized NF-YA subunit binds with the NF-YB/NF-YC dimer to form a functional transcription factor that binds the CCAAT DNA motif. All three subunits, particularly NF-YA, confer DNA binding specificity to the complex (Sinha et al. 1996; Zemzoumi et al. 1999).

Despite their difference from non-LEC1 subunits, both *Arabidopsis* LEC1 and L1L form functional NF-Y complexes, as diagramed in Figure 3 (Calvenzani et al. 2012; Gnesutta et al. 2017b). Assembly of the LEC1 NF-Y complex appears to occur similarly with non-LEC1 NF-Y complexes in that rice LEC1 preferentially localizes to tobacco epidermal cells nuclei only when a rice NF-YC subunit is coexpressed, suggesting that LEC1 lacks a nuclear localization sequence (E et al. 2018). Protein crystallography studies predict that the structure of the NF-Y complex containing L1L is very similar to NF-Y complexes from animals, and NF-Y complexes containing LEC1 or L1L bind CCAAT DNA motifs (Calvenzani et al. 2012; Nardini et al. 2013; Gnesutta et al. 2017b). Consistent with this finding, the CCAAT DNA motif is overrepresented in the promoter of several genes that are regulated by LEC1 during the early stages of embryo development in *Arabidopsis* and soybean (Pelletier et al. 2017). Thus, it is likely that LEC1 promotes transcription as a functional NF-Y complex during seed development.

LEC1 interactions with other transcription factors

Genome-wide analysis of LEC1 binding sites in the upstream regions of *Arabidopsis* and soybean genes that are transcriptionally regulated by LEC1 revealed a distinct set of DNA sequence motifs that were enriched in their promoter regions (Pelletier et al. 2017). The CCAAT DNA motif is enriched in genes that are LEC1 regulated early in seed development. By contrast, LEC1 regulated genes expressed at later stages of seed development were overrepresented for DNA motifs that resemble the G-Box (CACGTG), ABRE-like ((C/G/T)

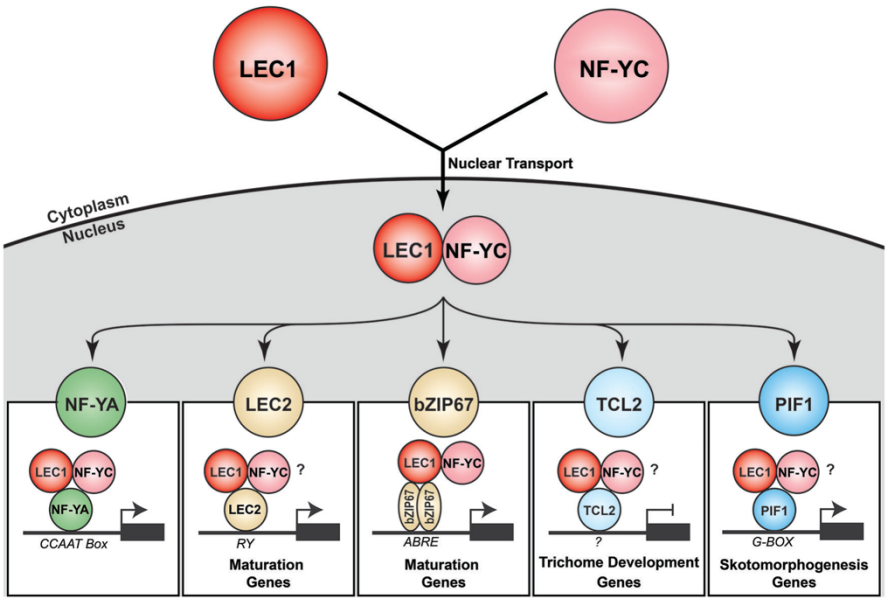


Figure 3. LEC1 regulates distinct processes during seed development through its interaction with other transcription factors

Binding of LEC1 with NF-YC enables transport of the dimer into the nucleus where it can interact with the indicated transcription factors, dependent on developmental stage. NF-YC subunits marked with a question mark indicate that NF-YC has not been shown to be required for the interaction of LEC1 with the transcription factor. In many cases, L1L may replace LEC1 in the complexes. The DNA sequence motif bound by TCL2 is not known.

ACGTG(G/T)(A/C)), RY (CATGCA) and BPC1 ((A/G)GA(A/G)AG(A/G)(A/G)A) *cis*-regulatory elements (Pelletier et al. 2017). Because NF-Y complexes bind CCAAT DNA motifs, it is hypothesized that LEC1 can interact with several other transcription factors and that these interactions specify which set of genes are regulated by LEC1 (Pelletier et al. 2017). Interactions between NF-Y subunits and other transcription factors have been reported extensively for plants and animals, and these interactions are important to specify the activity of these other transcription factors (Dolfini et al. 2012; Zhao et al. 2016; Myers and Holt 2018). Here, we discuss interactions between LEC1 and other transcription factors.

Studies of the B-BOX-type zinc finger transcription factor, CONSTANS (CO) that controls flowering in plants, provide insight into a potential mechanism by which the transcriptional specificity of LEC1 may be modulated. CO interacts with a NF-YB/NF-YC dimer to form a functional transcription factor by essentially replacing NF-YA in the NF-Y complex (Gnesutta et al. 2017a). Given that the NF-YA subunit participates in determining the DNA binding specificity of NF-Y complexes, the CO/NF-YB2/NF-YC3 complex does not bind the CCAAT DNA motif, but rather it binds the CORE element (CCACA) in the promoter regions of the CO target gene, *FLOWERING LOCUS T*. Interestingly, CO competes with NF-YA subunits for the NF-YB/YC dimer (Gnesutta et al. 2017a).

By analogy to the CO/NF-YB/NF-YC complex, the LEC1/NF-YC dimer appears to interact with other transcription factors to modulate LEC1 activity as shown in Figure 3. Consistent with the finding that G-box motifs are enriched in LEC1 binding regions in LEC1 target gene promoters, the basic leucine zipper transcription factor, bZIP67, has been shown to interact with the L1L/NF-YC2 dimer (Yamamoto et al. 2009). The LEC1/NF-YC2/bZIP67 complex binds the ABRE DNA motif, which has a G-box

core, but not the CCAAT DNA motif, in the promoters of genes involved in the maturation phase, such as *CRUCIFERIN C*, *FATTY ACID DESATURASE3*, and *SUCROSE SYNTHASE 2* (Yamamoto et al. 2009; Mendes et al. 2013). Similar to the CO/NF-YB/NF-YC complex, NF-YA strongly inhibits the activity of the LEC1 complex with *CRUCIFERIN C*, suggesting a competition between NF-YA and bZIP67 for the LEC1/NF-YC dimer (Yamamoto et al. 2009).

LEC1 also interacts with LEC2 (Figure 3) (Baud et al. 2016; Boulard et al. 2018). LEC2 is a B3 transcription factor that together with other B3 proteins, ABI3 and FUS3, regulates several processes during the maturation phase (Devic and Roscoe 2016; Lepiniec et al. 2018). LEC1, LEC2 and ABI3 synergistically promote the expression of the *OLEOSIN1* gene through RY and ABRE DNA motifs (Baud et al. 2016). Thus, LEC1's ability to control the maturation phase likely occurs through interactions with B3 and bZIP transcription factors that accumulate during the late stages of seed development.

LEC1 also interacts with other transcription factors to regulate diverse development processes (Figure 3). For example, LEC1 interacts with PHYTOCHROME INTERACTING FACTOR1 (PIF1) that is important for the expression of skotomorphogenesis genes through the G box element (Junker et al. 2012; Huang et al. 2015b). LEC1 also interacts with TRICHOMELESS2 (TCL2) to repress the expression of genes involved with trichome development during embryogenesis (Huang et al. 2015a).

Together, the ability of LEC1 to interact with many transcription factors provides potential mechanisms to explain how LEC1 can regulate distinct gene sets at different stages of seed development. Defining all of the transcription factors that interact with LEC1 during seed development and their impact on LEC1 activity could provide useful insights into the multifunctionality of LEC1 during seed development.

LEC1 AS A PIONEER TRANSCRIPTION FACTOR

The transition from the morphogenesis to the maturation phase represents a reprogramming of cellular identity. Cellular reprogramming in animals is often mediated, in part, by pioneer transcription factors that are involved in the initial steps that allow silenced genes to become competent for transcription (Guo and Morris 2017). Pioneer transcription factors have the capacity to bind compacted or "closed" chromatin and initiate chromatin remodeling, resulting in an increase in target site accessibility and facilitating the recruitment of other transcription factors to genes in the newly opened chromatin (reviewed by Zaret and Carroll 2011; Mayran and Drouin 2018; Sartorelli and Puri 2018).

LEC1 is the first pioneer transcription factor to be identified in plants based on its involvement in activating *FLOWERING LOCUS C* (*FLC*) (Figure 4) (Tao et al. 2017). FLC is a flowering repressor that undergoes epigenetic silencing during vernalization, resulting in the transition from vegetative to reproductive development (reviewed by Andres and Coupland 2012; Whittaker and Dean 2017). After plants flower, *FLC* remains silenced and in a repressed chromatin state, and it is maintained as such through gametogenesis (Sheldon et al. 2008). However, *FLC* expression must be

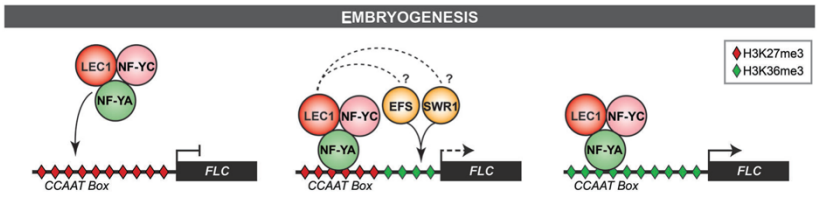


Figure 4. LEC1 is a pioneer transcription factor that promotes *FLC* transcription during embryogenesis
An NF-Y complex containing LEC1 binds the CCAAT DNA sequence motif in the *FLC* promoter in a closed chromatin conformation, as indicated by its occupancy by H3K27me3. The LEC1 NF-Y complex works through EFS and the SWR1 complex to initiate the establishment of an active chromatin state as indicated by occupancy of the active chromatin mark, H3K36me3.

reestablished to repress flowering prior to vernalization. As shown in Figure 4, LEC1 promotes the initial establishment of an active chromatin state at *FLC* in embryos (Tao et al. 2017). LEC1 binding at the *FLC* promoter is essential to engage EARLY FLOWERING IN SHORT DAYS (EFS) and the SWR1 complex to enhance chromatin accessibility and facilitate the recruitment of active histone marks on the *FLC* promoter, although the mechanistic relationship between LEC1 and the chromatin remodelers remains to be determined.

The characterization of LEC1 as a pioneer transcription factor opens the possibility that LEC1 may serve a similar function during seed development. Thus, LEC1 may bind compacted chromatin and promote chromatin conformational changes that allow other transcription factors to bind, in part, through their interactions with LEC1. Further analysis of the relationship between LEC1 and epigenetic changes that occur during seed development could provide insights into LEC1 role as a pioneer transcription factor.

CONCLUSION AND PERSPECTIVES

In this review, we have summarized recent findings that emphasize the role of LEC1 as a central regulator of seed development. LEC1 controls distinct processes at different stages of development. Therefore, its activity must sequentially regulate different sets of genes during seed development. The hormones GA and ABA may be involved in modulating LEC1 function in response to different physiological cues.

How does LEC1 regulate diverse sets of genes? First, LEC1 acts indirectly to regulate cellular processes during seed development by activating genes encoding transcription factors controlling structural genes that underlie these processes. In some cases, LEC1 also directly activates the same structural genes that are regulated by its downstream transcription factors, establishing a feed-forward loop that potentially promotes gene expression. LEC1 also interacts with different transcription factors at different stages of development, and the concerted actions of these transcription factor complexes may specify the particular set of genes that are activated. Moreover, the recent finding that LEC1 acts as a pioneer transcription factor provides potential insight into understanding LEC1 function to promote the activation of different gene sets during seed development.

ACKNOWLEDGEMENTS

We thank Rie Uzawa for reviewing the manuscript. Research from the Harada lab was supported by a grant from the National Science Foundation Plant Genome Research Program. L.J. was supported by a Coordination for the Improvement of Higher Education Personnel grant (CAPES, Brazil, No. 99999.013505/2013-00).

REFERENCES

Aghamirzaie D, Batra D, Heath LS, Schneider A, Grene R, Collakova E (2015) Transcriptome-wide functional characterization reveals novel relationships among differentially expressed transcripts in developing soybean embryos. **BMC Genomics** 16: 928

Allen DK, Ohlrogge JB, Shachar-Hill Y (2009) The role of light in soybean seed filling metabolism. **Plant J** 58: 220–234

Andres F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. **Nat Rev Genet** 13: 627–639

Angelovici R, Galili G, Fernie AR, Fait A (2010) Seed desiccation: A bridge between maturation and germination. **Trends Plant Sci** 15: 211–218

Baud S, Dubreucq B, Miquel M, Rochat C, Lepiniec L (2008) Storage reserve accumulation in *Arabidopsis*: Metabolic and developmental control of seed filling. **Arabidopsis Book** 6: e0113

Baud S, Kelemen Z, Thevenin J, Boulard C, Blanchet S, To A, Payre M, Berger N, Effroy-Cuzzi D, Franco-Zorrilla JM, Godoy M, Solano R, Thevenon E, Parcy F, Lepiniec L, Dubreucq B (2016) Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by master transcriptional regulators in *Arabidopsis* seed. **Plant Physiol** 171: 1099–1112

Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. **Plant J** 50: 825–838

Becker MG, Hsu SW, Harada JJ, Belmonte MF (2014) Genomic dissection of the seed. **Front Plant Sci** 5: 464

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB, Harada JJ (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. **Proc Natl Acad Sci USA** 110: E435–E444

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T, Moreau S, Niebel A, Frickey T, Weiller G, He J, Dai X, Zhao PX, Tang Y, Udvardi MK (2008) A gene expression atlas of the model legume *Medicago truncatula*. **Plant J** 55: 504–513

Boulard C, Fatihi A, Lepiniec L, Dubreucq B (2017) Regulation and evolution of the interaction of the seed B3 transcription factors with NF-Y subunits. **Biochim Biophys Acta Gene Regul Mech** 1860: 1069–1078

Boulard C, Thevenin J, Tranquet O, Laporte V, Lepiniec L, Dubreucq B (2018) LEC1 (NF-YB9) directly interacts with LEC2 to control gene expression in seed. **Biochim Biophys Acta** 1861: 443–450

Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. **Trends Plant Sci** 13: 624–630

Cagliari A, Turchetto-Zolet AC, Korbes AP, Maraschin Fdos S, Margis R, Margis-Pinheiro M (2014) New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants. **Genomics** 103: 380–387

Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, Petroni K, Mantovani R, Tonelli C (2012) Interactions and CCAAT-binding of *Arabidopsis thaliana* NF-Y subunits. **PLoS ONE** 7: e42902

Chen H, Wang FW, Dong YY, Wang N, Sun YP, Li XY, Liu L, Fan XD, Yin HL, Jing YY, Zhang XY, Li YL, Chen G, Li HY (2012) Sequence mining and transcript profiling to explore differentially expressed genes associated with lipid biosynthesis during soybean seed development. **BMC Plant Biol** 12: 122

Chen J, Zeng B, Zhang M, Xie S, Wang G, Hauck A, Lai J (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. **Plant Physiol** 166: 252–264

Chhun T, Chong SY, Park BS, Wong EC, Yin JL, Kim M, Chua NH (2016) HSI2 repressor recruits MED13 and HDA6 to downregulate seed maturation gene expression directly during *Arabidopsis* early seedling growth. **Plant Cell Physiol** 57: 1689–1706

D'Ario M, Griffiths-Jones S, Kim M (2017) Small RNAs: Big impact on plant development. **Trends Plant Sci** 22: 1056–1068

Dean Rider S Jr, Henderson JT, Jerome RE, Edenberg HJ, Romero-Severson J, Ogas J (2003) Coordinate repression of regulators of embryonic identity by PICKLE during germination in *Arabidopsis*. **Plant J** 35: 33–43

Devic M, Roscoe T (2016) Seed maturation: Simplification of control networks in plants. **Plant Sci** 252: 335–346

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN (1996) The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. **Cell** 86: 423–433

Dolfini D, Gatta R, Mantovani R (2012) NF-Y and the transcriptional activation of CCAAT promoters. **Crit Rev Biochem Mol Biol** 47: 29–49

E Z, Li T, Zhang H, Liu Z, Deng H, Sharma S, Wei X, Wang L, Niu B, Chen C (2018) A group of nuclear factor Y transcription factors are sub-functionalized during endosperm development in monocots. **J Exp Bot** 69: 2495–2510

Elahi N, Duncan RW, Stasolla C (2016) Modification of oil and glucosinolate content in canola seeds with altered expression of *Brassica napus* LEAFY COTYLEDON1. **Plant Physiol Biochem** 100: 52–63

Fang YH, Li X, Bai SN, Rao GY (2017) Sugar treatments can induce AcLEAFY COTYLEDON1 expression and trigger the accumulation of storage products during *Prothallus* development of *Adiantum capillus-veneris*. **Front Plant Sci** 8: 541

Finkelstein RR, Gampala SSL, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. **Plant Cell** 14: S15–S45

Finkelstein RR, Somerville CR (1990) Three classes of abscisic acid (ABA)-insensitive mutations of *Arabidopsis* define genes that control overlapping subsets of ABA responses. **Plant Physiol** 94: 1172–1179

Frontini M, Imbriano C, diSilvio A, Bell B, Bogni A, Romier C, Moras D, Tora L, Davidson I, Mantovani R (2002) NF-Y recruitment of TFIID, multiple interactions with histone fold TAF(II)s. **J Biol Chem** 277: 5841–5848

Gnesutta N, Kumimoto RW, Swain S, Chiara M, Siriwardana C, Horner DS, Holt BF 3rd, Mantovani R (2017a) CONSTANS imparts DNA sequence specificity to the histone fold NF-YB/NF-YC dimer. **Plant Cell** 29: 1516–1532

Gnesutta N, Saad D, Chaves-Sanjuan A, Mantovani R, Nardini M (2017b) Crystal structure of the *Arabidopsis thaliana* L1L/NF-YC3 histone-fold dimer reveals specificities of the LEC1 family of NF-Y subunits in plants. **Mol Plant** 10: 645–648

Goldberg RB, de Paiva G, Yadegari R (1994) Plant embryogenesis: Zygote to seed. **Science** 266: 605–614

Gonzalez-Morales SI, Chavez-Montes RA, Hayano-Kanashiro C, Alejo-Jacuinde G, Rico-Cambron TY, de Folter S, Herrera-Estrella L (2016) Regulatory network analysis reveals novel regulators of seed desiccation tolerance in *Arabidopsis thaliana*. **Proc Natl Acad Sci USA** 113: E5232–5241

Guo C, Morris SA (2017) Engineering cell identity: Establishing new gene regulatory and chromatin landscapes. **Curr Opin Genet Dev** 46: 50–57

Guo F, Liu C, Xia H, Bi Y, Zhao C, Zhao S, Hou L, Li F, Wang X (2013) Induced expression of AtLEC1 and AtLEC2 differentially promotes somatic embryogenesis in transgenic tobacco plants. **PLoS ONE** 8: e71714

Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C (2007) Combined networks regulating seed maturation. **Trends Plant Sci** 12: 294–300

Han JD, Li X, Jiang CK, Wong GK, Rothfels CJ, Rao GY (2017) Evolutionary analysis of the *LAFL* genes involved in the land plant seed maturation program. **Front Plant Sci** 8: 439

Harada J, Pelletier J (2012) Genome-wide analyses of gene activity during seed development. **Seed Sci Res** 22: S15–S22

Harada JJ (1997) Seed maturation and control of germination. In: Larkins BA, Vasi IK, eds. *Advances in Cellular and Molecular Biology of Plants, Volume 4, Cellular and Molecular Biology of Seed Development*. Kluwer Academic Publishers, Dordrecht. pp. 545–592

Harada JJ (2001) Role of *Arabidopsis LEAFY COTYLEDON* genes in seed development. **J Plant Physiol** 158: 405–409

Henderson JT, Li HC, Rider SD, Mordhorst AP, Romero-Severson J, Cheng JC, Robey J, Sung ZR, de Vries SC, Ogas J (2004) *PICKLE* acts throughout the plant to repress expression of embryonic traits and may play a role in gibberellin-dependent responses. **Plant Physiol** 134: 995–1005

Holdsworth MJ, Bentsink L, Soppe WJJ (2008) Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. **New Phytol** 179: 33–54

Hu YL, Zhou LM, Huang MK, He XM, Yang YH, Liu X, Li YG, Hou XL (2018) Gibberellins play an essential role in late embryogenesis of *Arabidopsis*. **Nat Plants** 4: 289–298

Huang J, Deng J, Shi T, Chen Q, Liang C, Meng Z, Zhu L, Wang Y, Zhao F, Yu S, Chen Q (2017) Global transcriptome analysis and identification of genes involved in nutrients accumulation during seed development of rice tartary buckwheat (*Fagopyrum tararicum*). **Sci Rep** 7: 11792

Huang M, Hu Y, Liu X, Li Y, Hou X (2015a) *Arabidopsis* LEAFY COTYLEDON1 controls cell fate determination during post-embryonic development. **Front Plant Sci** 6: 955

Huang M, Hu Y, Liu X, Li Y, Hou X (2015b) *Arabidopsis* LEAFY COTYLEDON1 mediates postembryonic development via interacting with PHYTOCHROME-INTERACTING FACTOR4. **Plant Cell** 27: 3099–3111

Huang Y, Chen L, Wang L, Vijayan K, Phan S, Liu Z, Wan L, Ross A, Xiang D, Datla R, Pan Y, Zou J (2009) Probing the endosperm gene expression landscape in *Brassica napus*. **BMC Genomics** 10: 256

Jia H, Suzuki M, McCarty DR (2014) Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. **Wiley Interdiscip Rev Dev Biol** 3: 135–145

Junker A, Monke G, Rutten T, Keilwagen J, Seifert M, Thi TM, Renou JP, Balzergue S, Viehover P, Hahnel U, Ludwig-Muller J, Altschmied L, Conrad U, Weisshaar B, Baumlein H (2012) Elongation-related functions of LEAFY COTYLEDON1 during the development of *Arabidopsis thaliana*. **Plant J** 71: 427–442

Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T (2005) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. **Plant Cell Physiol** 46: 399–406

Kahle J, Baake M, Doenecke D, Albig W (2005) Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13. **Mol Cell Biol** 25: 5339–5354

Keith K, Kraml M, Dengler NG, McCourt P (1994) fusca3: A heterochronic mutation affecting late embryo development in *Arabidopsis*. **Plant Cell** 6: 589–600

Kermode AR (1990) Regulatory mechanisms involved in the transition from seed development to germination. **CRC Crit Rev Plant Sci** 9: 155–195

Khan D, Millar JL, Girard IJ, Belmonte MF (2014) Transcriptional circuitry underlying seed coat development in *Arabidopsis*. **Plant Sci** 219–220: 51–60

Kirkbride RC, Fischer RL, Harada JJ (2013) LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of *Selaginella moellendorffii*. **PLoS ONE** 8: e67971

Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ (2003a) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. **Plant Cell** 15: 5–18

Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ (2003b) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. **Plant Cell** 15: 5–18

Laloum T, De Mita S, Gamas P, Baudin M, Niebel A (2013) CCAAT-box binding transcription factors in plants: Y so many? **Trends Plant Sci** 18: 157–166

Lau S, Slane D, Herud O, Kong J, Jurgens G (2012) Early embryogenesis in flowering plants: Setting up the basic body pattern. **Annu Rev Plant Biol** 63: 483–506

Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamuro JK, Harada JJ, Goldberg RB (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. **Proc Natl Acad Sci USA** 107: 8063–8070

Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB (2007) Using genomics to study legume seed development. **Plant Physiol** 144: 562–574

Ledwon A, Gaj MD (2011) LEAFY COTYLEDON1, FUSCA3 expression and auxin treatment in relation to somatic embryogenesis induction in *Arabidopsis*. **Plant Growth Regul** 65: 157–167

Lee H, Fischer RL, Goldberg RB, Harada JJ (2003) *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. **Proc Natl Acad Sci USA** 100: 2152–2156

Lepiniec L, Devic M, Roscoe TJ, Bouyer D, Zhou DX, Boulard C, Baud S, Dubreucq B (2018) Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. **Plant Reprod** 31: 291–307

Leprince O, Pellizzaro A, Berriri S, Buitink J (2017) Late seed maturation: Drying without dying. **J Exp Bot** 68: 827–841

Li G, Wang D, Yang R, Logan K, Chen H, Zhang S, Skaggs MI, Lloyd A, Burnett WJ, Laurie JD, Hunter BG, Dannenhoffer JM, Larkins BA, Drews GN, Wang X, Yadegari R (2014) Temporal patterns of gene expression in developing maize endosperm identified through transcriptome sequencing. **Proc Natl Acad Sci USA** 111: 7582–7587

Li HC, Chuang K, Henderson JT, Rider SD, Bai Y, Zhang H, Fountain M, Gerber J, Ogas J (2005) PICKLE acts during

germination to repress expression of embryonic traits. **Plant J 44**: 1010–1022

Li J, Berger F (2012) Endosperm: Food for humankind and fodder for scientific discoveries. **New Phytol** 195: 290–305

Lopez-Molina L, Mongrand S, Chua N-H (2001) A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in *Arabidopsis*. **Proc Natl Acad Sci USA** 98: 4782–4787

Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. **Cell** 93: 1195–1205

Lowe K, Hoerster G, Sun X, Rasco-Gaunt S, Lazerri P, Ellis S, Abbitt S, Glassman K, Gordon-Kamm B (2003) Maize LEC1 improves transformation in both maize and wheat. In: Vasil IK, ed. Plant Biotechnology 2002 and Beyond: Proceedings of the 10th IAPTC&B Congress. June 23–28, 2002 Orlando, Florida, USA. Springer Netherlands, Dordrecht. pp. 283–284

Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. **Proc Natl Acad Sci USA** 100: 11980–11985

Mayran A, Drouin J (2018) Pioneer transcription factors shape the epigenetic landscape. **J Biol Chem** 293: 13795–13804

Meinke DW (1992) A homoeotic mutant of *Arabidopsis thaliana* with leafy cotyledons. **Science** 258: 1647–1650

Meinke DW, Franzmann LH, Nickle TC, Yeung EC (1994) Leafy cotyledon mutants of *Arabidopsis*. **Plant Cell** 6: 1049–1064

Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, Eastmond PJ (2013) bZIP67 regulates the omega-3 fatty acid content of *Arabidopsis* seed oil by activating fatty acid desaturase3. **Plant Cell** 25: 3104–3116

Monke G, Seifert M, Keilwagen J, Mohr M, Grosse I, Hahnel U, Junker A, Weisshaar B, Conrad U, Baumlein H, Altschmied L (2012) Toward the identification and regulation of the *Arabidopsis thaliana* ABI3 regulon. **Nucleic Acids Res** 40: 8240–8254

Mu J, Tan H, Zheng Q, Fu F, Liang Y, Zhang J, Yang X, Wang T, Chong K, Wang XJ, Zuo J (2008) LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*. **Plant Physiol** 148: 1042–1054

Myers ZA, Holt BF, 3rd (2018) NUCLEAR FACTOR-Y: Still complex after all these years? **Curr Opin Plant Biol** 45: 96–102

Nakashima K, Fujita Y, Kanamori N, Katagiri T, Umezawa T, Kidokoro S, Maruyama K, Yoshida T, Ishiyama K, Kobayashi M, Shinozaki K, Yamaguchi-Shinozaki K (2009) Three *Arabidopsis* SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy. **Plant Cell Physiol** 50: 1345–1363

Nakashima K, Yamaguchi-Shinozaki K (2013) ABA signaling in stress-response and seed development. **Plant Cell Rep** 32: 959–970

Nambara E, Keith K, McCourt P, Naito S (1995) A regulatory role for the ABI3 gene in the establishment of embryo maturation in *Arabidopsis thaliana*. **Development** 121: 629–636

Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vonrhein C, Moras D, Romier C, Bolognesi M, Mantovani R (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. **Cell** 152: 132–143

Nic-Can GI, Lopez-Torres A, Barredo-Pool F, Wrobel K, Loyola-Vargas VM, Rojas-Herrera R, De-la-Pena C (2013) New insights into somatic embryogenesis: Leafy cotyledon1, baby boom1 and WUSCHEL-related homeobox4 are epigenetically regulated in *Coffea canephora*. **PLoS ONE** 8: e72160

Ogas J, Cheng JC, Sung ZR, Somerville C (1997) Cellular differentiation regulated by gibberellin in the *Arabidopsis thaliana* pickle mutant. **Science** 277: 91–94

Ogas J, Kaufmann S, Henderson J, Somerville C (1999) PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. **Proc Natl Acad Sci USA** 96: 13839–13844

Orlowska A, Igielski R, Lagowska K, Kepczynska E (2017) Identification of *LEC1*, *L1L* and *Polycomb Repressive Complex 2* genes and their expression during the induction phase of *Medicago truncatula* Gaertn. somatic embryogenesis. **Plant Cell Tiss Organ Cult** 129: 119–132

Palovaara J, de Zeeuw T, Weijers D (2016) Tissue and organ initiation in the plant embryo: A first time for everything. **Annu Rev Cell Dev Biol** 32: 47–75

Parcy F, Valon C, Kohara A, Misera S, Giraudat J (1997) The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of *Arabidopsis* seed development. **Plant Cell** 9: 1265–1277

Pelletier JM, Kwong RW, Park S, Le BH, Baden R, Cagliari A, Hashimoto M, Munoz MD, Fischer RL, Goldberg RB, Harada JJ (2017) LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. **Proc Natl Acad Sci USA** 114: E6710–E6719

Petroni K, Kumimoto RW, Gnesutta N, Calvenzani V, Fornari M, Tonelli C, Holt BF III, Mantovani R (2012) The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. **Plant Cell** 24: 4777–4792

Pradhan S, Bandhiwal N, Shah N, Kant C, Gaur R, Bhatia S (2014) Global transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds. **Front Plant Sci** 5: 698

Pu L, Sung ZR (2015) PcG and trxG in plants - friends or foes. **Trends Genet** 31: 252–262

Puthur JT, Shackira AM, Saradhi PP, Bartels D (2013) Chloroembryos: A unique photosynthesis system. **J Plant Physiol** 170: 1131–1138

Rangan P, Furtado A, Henry RJ (2017) The transcriptome of the developing grain: A resource for understanding seed development and the molecular control of the functional

and nutritional properties of wheat. **BMC Genomics** 18: 766

Raz V, Bergervoet JH, Koornneef M (2001) Sequential steps for developmental arrest in *Arabidopsis* seeds. **Development** 128: 243–252

Rolletschek H, Radchuk R, Klukas C, Schreiber F, Wobus U, Borisjuk L (2005) Evidence of a key role for photosynthetic oxygen release in oil storage in developing soybean seeds. **New Phytol** 167: 777–786

Roszak P, Kohler C (2011) Polycomb group proteins are required to couple seed coat initiation to fertilization. **Proc Natl Acad Sci USA** 108: 20826–20831

Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. **Plant J** 54: 608–620

Sartorelli V, Puri PL (2018) Shaping gene expression by landscaping chromatin architecture: Lessons from a master. **Mol Cell** 71: 375–388

Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA, Cannon SB, May GD, Vance CP, Shoemaker RC (2010) RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. **BMC Plant Biol** 10: 160–160

Sheldon CC, Hills MJ, Lister C, Dean C, Dennis ES, Peacock WJ (2008) Resetting of *FLOWERING LOCUS C* expression after epigenetic repression by vernalization. **Proc Natl Acad Sci USA** 105: 2214–2219

Siefers N, Dang KK, Kumimoto RW, Bynum WET, Tayrose G, Holt BF, 3rd (2009) Tissue-specific expression patterns of *Arabidopsis* NF-Y transcription factors suggest potential for extensive combinatorial complexity. **Plant Physiol** 149: 625–641

Sinha S, Kim IS, Sohn KY, de Crombrugghe B, Maity SN (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. **Mol Cell Biol** 16: 328–337

Smith ZR, Long JA (2010) Control of *Arabidopsis* apical-basal embryo polarity by antagonistic transcription factors. **Nature** 464: 423–426

Steeves TA (1983) The evolution and biological significance of seeds. **Can J Bot** 61: 3550–3560

Suzuki M, Wang HH, McCarty DR (2007) Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. **Plant Physiol** 143: 902–911

Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, Fu F, Li J, Guan R, Zhang H, Wang G, Zuo J (2011) Enhanced seed oil production in Canola by conditional expression of *Brassica napus LEAFY COTYLEDON1* and *LEC1-LIKE* in developing seeds. **Plant Physiol** 156: 1577–1588

Tang G, Xu P, Ma W, Wang F, Liu Z, Wan S, Shan L (2018) Seed-specific expression of AtLEC1 increased oil content and altered fatty acid composition in seeds of peanut (*Arachis hypogaea* L.). **Front Plant Sci** 9: 260

Tao Z, Shen L, Gu X, Wang Y, Yu H, He Y (2017) Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. **Nature** 551: 124–128

Terrasson E, Buitink J, Righetti K, Ly Vu B, Pelletier S, Zinsmeister J, Lalanne D, Leprince O (2013) An emerging picture of the seed desiccome: Confirmed regulators and newcomers identified using transcriptome comparison. **Front Plant Sci** 4: 497

To A, Joubes J, Barthole G, Lecureuil A, Scagnelli A, Jasinski S, Lepiniec L, Baud S (2012) WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in *Arabidopsis*. **Plant Cell** 24: 5007–5023

To A, Valon C, Savino G, Guilleminot J, Devic M, Giraudat J, Parcy F (2006) A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. **Plant Cell** 18: 1642–1651

Tsukagoshi H, Morikami A, Nakamura K (2007) Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in *Arabidopsis* seedlings. **Proc Natl Acad Sci USA** 104: 2543–2547

Verdier J, Kakar K, Gallardo K, Le Signor C, Aubert G, Schlereth A, Town CD, Udvardi MK, Thompson RD (2008) Gene expression profiling of M. truncatula transcription factors identifies putative regulators of grain legume seed filling. **Plant Mol Biol** 67: 567–580

Verdier J, Thompson RD (2008) Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. **Plant Cell Physiol** 49: 1263–1271

Vernon DM, Meinke DW (1994) Embryogenic transformation of the suspensor in twin, a polyembryonic mutant of *Arabidopsis*. **Dev Biol** 165: 566–573

Vicente-Carbajosa J, Carbonero P (2004) Seed maturation: Developing an intrusive phase to accomplish a quiescent state. **International J Dev Biol** 49: 645–651

Vigeolas H, van Dongen JT, Waldeck P, Huhn D, Geigenberger P (2003) Lipid storage metabolism is limited by the prevailing low oxygen concentrations within developing seeds of oilseed rape. **Plant Physiol** 133: 2048–2060

Wang F, Perry SE (2013) Identification of direct targets of FUSCA3, a key regulator of *Arabidopsis* seed development. **Plant Physiol** 161: 1251–1264

West M, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ (1994) LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. **Plant Cell** 6: 1731–1745

Whittaker C, Dean C (2017) The FLC locus: A platform for discoveries in epigenetics and adaptation. **Annu Rev Cell Dev Biol** 33: 555–575

Xiang D, Venglat P, Tibiche C, Yang H, Risseeuw E, Cao Y, Babic V, Cloutier M, Keller W, Wang E, Selvaraj G, Datla R (2011) Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*. **Plant Physiol** 156: 346–356

Xie Z, Li X, Glover BJ, Bai S, Rao GY, Luo J, Yang J (2008) Duplication and functional diversification of *HAP3* genes leading to the origin of the seed-developmental regulatory

gene, *LEAFY COTYLEDON1 (LEC1)*, in nonseed plant genomes. **Mol Biol Evol** 25: 1581–1592

Xu JJ, Zhang XF, Xue HW (2016) Rice aleurone layer specific OsNF-YB1 regulates grain filling and endosperm development by interacting with an ERF transcription factor. **J Exp Bot** 67: 6399–6411

Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T (2009) *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. **Plant J** 58: 843–856

Yang X, Zhang X (2010) Regulation of somatic embryogenesis in higher plants. **CRC Crit Rev Plant Sci** 29: 36–57

Zaret KS, Carroll JS (2011) Pioneer transcription factors: Establishing competence for gene expression. **Genes Dev** 25: 2227–2241

Zemzoumi K, Frontini M, Bellorini M, Mantovani R (1999) NF-Y histone fold alpha 1 helices help impart CCAAT specificity. **J Mol Biol** 286: 327–337

Zhan J, Thakare D, Ma C, Lloyd A, Nixon NM, Arakaki AM, Burnett WJ, Logan KO, Wang D, Wang X, Drews GN, Yadegari R (2015) RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. **Plant Cell** 27: 513–531

Zhang D, Jing Y, Jiang Z, Lin R (2014) The Chromatin-remodeling factor PICKLE Integrates brassinosteroid and gibberellin signaling during skotomorphogenic growth in *Arabidopsis*. **Plant Cell** 26: 2472–2485

Zhang H, Bishop B, Ringenberg W, Muir WM, Ogas J (2012) The CHD3 remodeler PICKLE associates with genes enriched for trimethylation of histone H3 lysine 27. **Plant Physiol** 159: 418–432

Zhang H, Rider SD, Henderson JT, Fountain M, Chuang K, Kandachar V, Simons A, Edenberg HJ, Romero-Severson J, Muir WM, Ogas J (2008) The CHD3 Remodeler *PICKLE* promotes trimethylation of histone H3 lysine 27. **J Biol Chem** 283: 22637–22648

Zhang JJ, Xue HW (2013) OsLEC1/OsHAP3E participates in the determination of meristem identity in both vegetative and reproductive developments of rice. **J Integr Plant Biol** 55: 232–249

Zhao H, Wu D, Kong F, Lin K, Zhang H, Li G (2016) The *Arabidopsis thaliana* nuclear factor Y transcription factors. **Front Plant Sci** 7: 2045

Zhou Y, Tan B, Luo M, Li Y, Liu C, Chen C, Yu CW, Yang S, Dong S, Ruan J, Yuan L, Zhang Z, Zhao L, Li C, Chen H, Cui Y, Wu K, Huang S (2013) HISTONE DEACETYLASE19 interacts with HSL1 and participates in the repression of seed maturation genes in *Arabidopsis* seedlings. **Plant Cell** 25: 134–148

Chapter 2

Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development

Jo, L., Pelletier, J.M., Hsu, S.W., Baden, R., Goldberg, R.B. and Harada, J.J.

This chapter was published as presented.
Jo, L., Pelletier, J.M., Hsu, S.W., Baden, R., Goldberg, R.B. and Harada, J.J., 2020. Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. Proceedings of the National Academy of Sciences, 117(2), pp.1223-1232.

Preface

The following attributes work within this chapter to the respective author(s)

Writing	L.J., J.J.H.
Figures	L.J. J.M.P.
Experiments	ChIP-Seq: L.J., R.B., J.M.P. Bioinformatics: L.J., J.M.P. Cloning: L.J. Protoplasts Transfections: L.J., S.W.H. Dual Luciferase Assays: L.J.
Intellectual Contributions	L.J., J.M.P.,R.B.G.,J.J.H.

Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development

Leonardo Jo[a], Julie M. Pelletier[a], Ssu-Wei Hsu[a,1], Russell Baden[a,2], Robert B. Goldberg[b,3], and John J. Harada[a,3]

[a]Department of Plant Biology, University of California, Davis, CA 95616; and [b]Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, CA 90095

Contributed by Robert B. Goldberg, November 27, 2019 (sent for review October 24, 2019; reviewed by Jerome Verdier and Lila O. Vodkin)

The LEAFY COTYLEDON1 (LEC1) transcription factor is a central regulator of seed development, because it controls diverse biological programs during seed development, such as embryo morphogenesis, photosynthesis, and seed maturation. To understand how LEC1 regulates different gene sets during development, we explored the possibility that LEC1 acts in combination with other transcription factors. We identified and compared genes that are directly transcriptionally regulated by ABA-RESPONSIVE ELEMENT BINDING PROTEIN3 (AREB3), BASIC LEUCINE ZIPPER67 (bZIP67), and ABA INSENSITIVE3 (ABI3) with those regulated by LEC1. We showed that LEC1 operates with specific sets of transcription factors to regulate different gene sets and, therefore, distinct developmental processes. Thus, LEC1 controls diverse processes through its combinatorial interactions with other transcription factors. DNA binding sites for the transcription factors are closely clustered in genomic regions upstream of target genes, defining *cis*-regulatory modules that are enriched for DNA sequence motifs that resemble sequences known to be bound by these transcription factors. Moreover, *cis*-regulatory modules for genes regulated by distinct transcription factor combinations are enriched for different sets of DNA motifs. Expression assays with embryo cells indicate that the enriched DNA motifs are functional *cis* elements that regulate transcription. Together, the results suggest that combinatorial interactions between LEC1 and other transcription factors are mediated by *cis*-regulatory modules containing clustered *cis* elements and by physical interactions that are documented to occur between the transcription factors.

cis-regulatory module | maturation | photosynthesis

The ability of plants to make seeds has conferred strong selective advantages to the angiosperms that, in part, explain their dominance within the plant kingdom (1). The seed habit requires that a novel, biphasic mode of development occurs at the earliest stage of the sporophytic life cycle. During the early, morphogenesis phase, the embryo and endosperm initially undergo regional specification into functional domains. The embryo develops further with the establishment of the shoot–root axis and differentiation of embryonic tissue and organ systems (2). Photosynthesis is initiated later during the morphogenesis phase, often in both the embryo and endosperm (3). During the maturation phase which follows morphogenesis, morphogenetic processes in the embryo are arrested; storage macromolecules, particularly proteins and lipids, accumulate and are stored; the embryo becomes desiccation tolerant; and seed germination is actively inhibited. The maturation phase is unique to seed plants, suggesting that this phase has been inserted into a continuous period of embryonic followed by postembryonic morphogenesis, characteristic of nonseed plants (4, 5). Relatively little is known of the gene regulatory networks that have enabled the maturation phase to be integrated into the angiosperm life cycle.

LEC1 is a central regulator of seed development that controls distinct developmental processes at different stages of seed development (reviewed in ref. 6). Analyses of loss- and gain-of-function mutants showed that LEC1 is a major regulator of the maturation phase that is required for storage macromolecule accumulation, the acquisition of desiccation tolerance, and germination inhibition during seed development (7, 8). However, LEC1 also appears to function during the morphogenesis phase. *LEC1* mRNA is detected in the zygote within 24 h after fertilization, loss-of-function mutations indicate that LEC1 is required to maintain embryonic suspensor and cotyledon identities, and LEC1 is also involved in regulating genes that underlie photosynthesis and chloroplast biogenesis (9, 10). It is not known how LEC1 is able to regulate the diverse developmental processes that occur during both the morphogenesis and maturation phases.

LEC1 is an atypical transcription factor (TF) subunit: a NF-YB subunit whose canonical role is to interact with NF-YC and NF-YA subunits to form a NF-Y TF that binds CCAAT DNA sequences (9, 11, 12). The LEC1-type NF-YB subunit is found only in plants, and it confers seed-specific functions (13). LEC1 also interacts physically with other TFs to regulate a variety of developmental processes (reviewed in ref. 6).

Significance

LEC1 is a central, transcriptional regulator of seed development, because it regulates diverse developmental processes at different stages, including embryo morphogenesis, photosynthesis, hormone biosynthesis and signaling, and the massive accumulation of seed storage macromolecules. We show that LEC1 acts in combination with the seed transcription factors (TFs), AREB3, bZIP67, and ABI3, and that different TF combinations regulate distinct gene sets. We show further that TF binding sites are closely clustered in the genome and contain enriched DNA sequence motifs that are bound by TFs and that distinct DNA motif sets recruit different TF combinations to binding site clusters. Our findings provide insights into the gene regulatory networks that govern seed development.

Author contributions: L.J., J.M.P., R.B.G., and J.J.H. designed research; L.J., J.M.P., and R.B. performed research; S.-W.H. contributed new reagents/analytic tools; L.J., J.M.P., R.B., and J.J.H. analyzed data; and L.J. and J.J.H. wrote the paper.

Reviewers: J.V., Institut de Recherche en Horticulture et Semences, Institut National de la Recherche Agronomique; and L.O.V., University of Illinois.

The authors declare no competing interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession nos. GSE101672, GSE101649, GSE140699, GSE140701, and GSE140700).

[1]Present address: Graduate Group in Integrative Pathobiology, University of California, Davis, CA 95616.

[2]Present address: Equine Analytical Chemistry Laboratory, University of California, Davis, CA 95616.

[3]To whom correspondence may be addressed. Email: jjharada@ucdavis.edu or bobg@g.ucla.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1918441117/-/DCSupplemental.

First published December 31, 2019.

We showed previously that LEC1 sequentially transcriptionally regulates distinct gene sets at different stages of seed development in *Arabidopsis* and soybean (10). As summarized in Fig. 1*A*, LEC1 regulates genes involved in growth and morphogenesis, photosynthesis, and maturation during the morphogenesis, transition from morphogenesis to maturation, and maturation phases, respectively. We showed further that LEC1 genomic binding sites are enriched for different DNA sequence motifs, the CCAAT, G box, RY, and BPC1 motifs. Different LEC1 target gene sets were enriched for distinct combinations of these DNA motifs, opening the possibility that LEC1 interacts with other TFs to regulate different gene sets.

In *Arabidopsis*, substantial information is available about the involvement of LEC1 and other TFs, including LEC1-LIKE, LEC2, ABA INSENSITIVE3 (ABI3), FUSCA3 (FUS3), WRINKLED1, MYB115/118, ABI4, ABI5, AGAMOUS-LIKE15, and a number of BASIC LEUCINE ZIPPER (bZIP) TFs, in regulating the maturation phase, and genetic studies generally place LEC1 atop the regulatory hierarchy (reviewed in refs. 14–17). The LEC1-NF-YC dimer interacts physically with the bZIP67 TF and binds with a G box-like but not a CCAAT DNA motif to activate maturation genes, such as the *CRUCIFERIN C*, *FATTY ACID DESATURASE3* (*FAD3*), and *DELAY OF GERMINATION1* (*DOG1*) (18–20). LEC1 also operates synergistically with LEC2 and ABI3, 2 B3 domain TFs that bind RY-like motifs, to promote maturation gene expression (21–24). LEC2 interacts physically with LEC1 through its B2 domain, but no direct physical interactions between LEC1 and ABI3 have been reported (22).

Here, we show that LEC1 regulates distinct developmental processes at different stages by acting combinatorially with other TFs, specifically bZIP67, ABA-RESPONSIVE ELEMENT BINDING PROTEIN3 (AREB3), a TF closely related to bZIP67, and ABI3. We showed that 1) LEC1 alone and LEC1 in combination with AREB3 primarily regulate genes involved in morphogenesis; 2) LEC1 and AREB3, LEC1, AREB3, and bZIP67, and LEC1, AREB3, bZIP67, and ABI3 regulate genes involved in photosynthesis; and 3) all 4 TFs regulate maturation genes. We also show that the binding sites for these TFs are closely clustered in the genome, and they are enriched for DNA motifs that correspond to annotated *cis* elements known to be bound by the 4 TFs. These results suggest that LEC1 functions combinatorially with AREB3, bZIP67, and ABI3 to regulate distinct gene sets and diverse developmental processes.

Results

Identification of AREB3, bZIP67, and ABI3 Target Genes in Developing Soybean Embryos. We hypothesized that LEC1 may act in combination with other TFs to regulate distinct gene sets at different stages of development, in part, because LEC1 has been shown to interact with a number of other TFs (reviewed in ref. 6). Based on their functions in *Arabidopsis*, we focused on 3 TFs: 1) bZIP67, a TF that interacts physically with the LEC1-NF-YC dimer to regulate maturation genes (19, 20); 2) AREB3, a TF closely related to and partially redundant functionally with bZIP67 that is expressed earlier in seed development than bZIP67 (25); and 3) the B3 domain TF, ABI3, a maturation regulator that interacts with bZIP TFs and, by extension, potentially with LEC1 (26–28).

We identified target genes directly regulated by AREB3, bZIP67, and ABI3 in soybean embryos at the early maturation (EM) stage (23 d after pollination) that corresponds to the transition from morphogenesis to maturation phases to compare the TFs' functions with LEC1. We used the chromatin immunoprecipitation–DNA sequencing (ChIP-Seq) strategy described by Pelletier et al. (10) to identify genes bound by AREB3-1 (Glyma.04G124200), AREB3-2 (Glyma.06G314400), bZIP67 (Glyma.13G317000), ABI3-1 (Glyma.08G357600), and ABI3-2 (Glyma.18G176100) (29, 30). The AREB3-1 and AREB3-2 homeologs and ABI3-1 and ABI3-2 homeologs are recognized by the AREB3 and ABI3 antibodies, respectively. Binding sites for these TFs were located primarily at the transcription start site, as we found previously for LEC1, and each TF bound the 1-kb upstream region of between 21,120 and 5,234 genes (*SI Appendix*, Fig. S1 and Dataset S1). Experiments with antibodies against 2 different peptides each from AREB3, bZIP67, and ABI3 confirmed the specificity of the ChIP experiments, and data analysis followed ENCODE guidelines (Dataset S2) (31–33). Our data analysis methods differed slightly from that reported previously; therefore, we also present results for the LEC1-1 (Glyma.07G268100) and

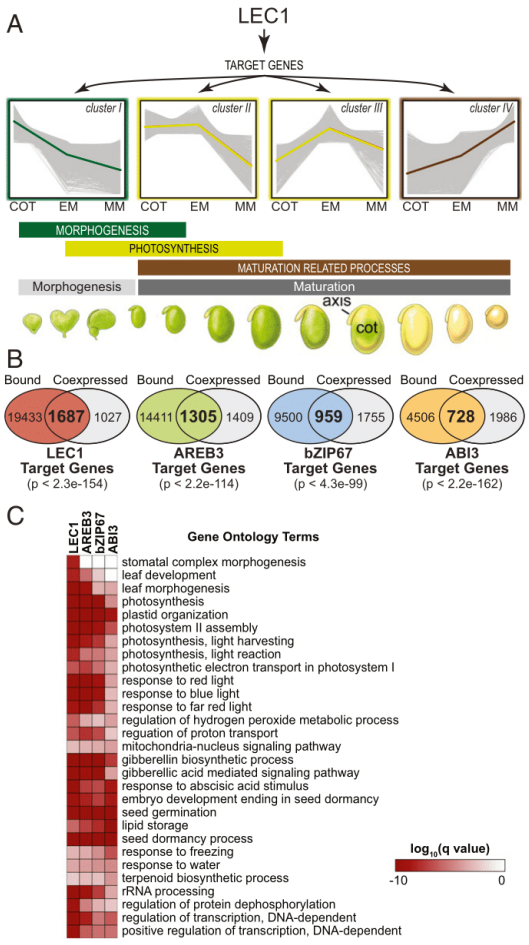


Fig. 1. Identification of LEC1, AREB3, bZIP67, and ABI3 target genes in soybean early maturation embryos. (*A*) Overview of LEC1's role in controlling distinct gene sets and developmental processes at different stages of seed development. (*B*) Target genes directly regulated by LEC1, AREB3, bZIP67, and ABI3 in soybean embryos at the EM stage. Venn diagrams show the overlap between bound genes (colored) and coexpressed genes (gray) for LEC1, AREB3, bZIP67, and ABI3. Statistical significance of the overlap between bound genes and coexpressed genes is indicated (hypergeometric distribution). (*C*) Heatmap showing the q value significance of GO terms for LEC1, AREB3, bZIP67, and ABI3 target genes. The GO terms listed are the top 5 enriched biological process GO terms for each TF. A comprehensive list of overrepresented GO terms is given in Dataset S1.

LEC1-2 (Glyma.17G005600) homeologs using previously reported primary data (10, 34).

Because only a fraction of the genes bound by a TF are transcriptionally regulated by that TF (35), we defined target genes regulated by these TFs as those that are both bound and coexpressed with the TF. All 4 of the TFs are expressed predominantly in embryos (*SI Appendix*, Fig. S2), and we used the Harada-Goldberg Soybean Seed Development Laser Capture Microdissection RNA-Seq Dataset (GEO accessions, GSE57606, GSE46096, and GSE99109) (36–38) to identify coexpressed genes as those whose mRNA levels accumulated at a 5-fold or higher level in embryo subregions compared with seed coat subregions ($q < 0.01$). As summarized in Fig. 1*B*, we identified 1,687, 1,305, 959, and 728 target genes, respectively, for LEC1, AREB3, bZIP67, and ABI3 and showed that the overrepresentation of bound and coexpressed genes was statistically significant ($P < 2.3 \times 10^{-154}$, $P < 2.2 \times 10^{-114}$, $P < 4.3 \times 10^{-99}$, and $P < 2.2 \times 10^{-162}$, respectively, Dataset S1). These TF target gene numbers are within the range reported for other plant TFs (39).

Gene Ontology (GO) representation analysis indicated that there was extensive overlap in the biological functions of the 4 TFs (Fig. 1*C* and Dataset S1), particularly processes related to morphogenesis, photosynthesis, GA biosynthesis and signaling, lipid storage, and seed dormancy. The results indicate that AREB3, bZIP67, and ABI3 TFs regulate developmental processes that are closely related to those controlled by LEC1.

LEC1 Operates in Combination with AREB3, bZIP67, and ABI3 to Regulate the Expression of Genes Involved in Distinct Developmental Processes in Soybean Embryos.

Deciphering combinatorial interactions among the 4 transcription factors. Because LEC1, AREB3, bZIP67, and ABI3 regulate genes involved in similar biological processes, we asked if they acted in coordination to regulate seed gene transcription by comparing their target genes. Fig. 2*A* shows that there was significant overlap in the target genes regulated by the 4 TFs. Of 1,687 LEC1 target genes, 1,243 (74%) were also targeted by at least 1 of the other TFs (Dataset S3). The vast majority of target genes were grouped into 4 categories: 1) those regulated by LEC1 alone (*L* genes); 2) LEC1 and AREB3 (*LA* genes); 3) LEC1, AREB3, and bZIP67 (*LAZ* genes); and 4) all 4 TFs (*LAZA* genes), with the largest number of target genes falling into the latter category. Thus, LEC1 appears to regulate gene transcription primarily in combination with AREB3, bZIP67, and ABI3.

Combinatorial control of developmental processes. We obtained insight into the biological processes regulated by different combinations of TFs by performing GO representation analysis on the different target gene sets. We were surprised to find that target genes regulated by different TF combinations were highly enriched for distinct GO term sets (Fig. 2*C* and Dataset S3). Specifically, 1) *L* and *LA* genes were most significantly overrepresented for GO terms related to morphogenesis, such as leaf morphogenesis, stomatal complex morphogenesis, polarity specification of adaxial/abaxial axis, and specification of organ position; 2) *LA*, *LAZ*, and *LAZA* genes were highly enriched for GO terms related to photosynthesis; 3) *LAZ* and *LAZA* genes were enriched for gibberellic acid (GA) biosynthesis and signaling; and 4) *LAZA* genes were overrepresented for GO terms related to maturation.

We asked if the accumulation of *AREB3*, *bZIP67*, and *ABI3* mRNAs could explain LEC1's ability to control the onset of the developmental processes temporally. Fig. 2*B* shows that *LEC1*, *AREB3*, and *ABI3* mRNAs accumulate early in embryo development, whereas *bZIP67* mRNA accumulates primarily at the midmaturation (MM, 40 to 45 d after pollination) stage. These results suggest that the TFs' mRNA accumulation patterns alone do not explain the temporal regulation of biological processes.

During embryo development, morphogenetic events are largely initiated before the onset of photosynthesis which, in turn, is followed by the maturation phase. To determine if the different TF combinations underlie the temporal regulation of these biological processes, we used clustering analysis to identify *L*, *LA*, *LAZ*, and *LAZA* mRNAs that accumulate at different stages of seed development. As shown in Fig. 2*D*, each target gene set exhibited 4 different expression patterns, with clusters I, II, III, and IV containing mRNAs that accumulated primarily at the 1) cotyledon (COT, 15 d after pollination) stage; 2) COT and EM stages; 3) EM stage; and 4) MM stage, respectively. *L* and *LA* genes were fairly evenly distributed among the 4 different clusters, whereas *LAZ* and *LAZA* genes were enriched in cluster III and cluster IV, respectively. We found that genes involved in morphogenesis, photosynthesis, and maturation were enriched in particular clusters: 1) *L* and *LA* genes involved in morphogenesis; 2) *L*, *LA*, *LAZ*, and *LAZA* genes involved in photosynthesis; and 3) *LAZA* genes involved in maturation were enriched in cluster I, cluster III, and cluster IV, respectively. These results emphasize that the genes that underlie specific developmental processes during seed development are precisely regulated temporally, regardless of which TF sets are involved in their regulation.

We determined which TF combinations regulate gene sets previously defined to be involved in either maturation or photosynthesis to validate the GO term enrichment analysis (10). Virtually all of the maturation genes bound by 1 of the TFs were bound by all 4 TFs, and orthologs of most of these genes were downregulated in *Arabidopsis lec1* and/or *abi3* mutants (*SI Appendix*, Fig. S3). Genes involved in the light reactions of photosynthesis were bound by between 1 and 4 of the TFs, and many were affected by the *Arabidopsis lec1* mutation (*SI Appendix*, Fig. S4). These results emphasize the importance of LEC1 and/or ABI3 in controlling maturation and photosynthesis genes. Additionally, *L* and *LA* genes involved in morphogenesis included known regulators of morphogenetic processes, *PHABULOSA*, *ASYMMETRIC LEAVES1*, *ARABIDOPSIS THALIANA HOMEOBOX PROTEIN13*, and *TCP1*. Among the genes related to GA biosynthesis and signaling, many of the *LAZ* genes encode proteins that promote GA synthesis, such as GA REQUIRING1 (GA1), GA3, GA4, and GA20 OXIDASE (GA20OX), and GA signaling, such as SLEEPY2 and GIBBERELLIN-INSENSITIVE DWARF1. By contrast, proteins encoded by the *LAZA* genes, *REPRESSOR OF GA1-3-LIKE2*, *PHYTOCHROME INTERACTING FACTOR3* (*PIF3*), and *PIF3-LIKE5*, negatively affect GA signaling, although others promote GA synthesis, such as *GA20OX*. Thus, LEC1 may act in both positive and negative feedback loops to control GA responses during embryo development.

Physical interactions between the 4 transcription factors. Combinatorial interactions among the TFs could indicate that they interact physically. In *Arabidopsis*, several of the 4 TFs have been shown to form complexes (18–20, 26–28, 40, 41). We obtained evidence indicating physical interactions between the soybean orthologs of LEC1 and bZIP67, LEC1 and AREB3, AREB3 and bZIP67, and bZIP67 and ABI3, as occurs in *Arabidopsis* (*SI Appendix*, Fig. S5). These results may indicate that the *LA*, *LAZ*, and *LAZA* genes are regulated by TF complexes.

LEC1, AREB3, bZIP67, and ABI3 Binding Sites Are Clustered and Contain Distinct Sets of DNA Sequence Motifs.

Identification of cis-regulatory module. We determined the organization of LEC1, AREB3, bZIP67, and ABI3 binding sites in the upstream regions of target genes to obtain insight into the mechanisms by which LEC1 works in combination with the other TFs to regulate different gene sets. We plotted the distance between the summit of the LEC1 ChIP-Seq peak, which approximates the TF binding site, and the ChIP-Seq peak summits of AREB3, bZIP67, and ABI3 (42). As shown in Fig. 3*A*, the TF binding sites in *LA*, *LAZ*, and *LAZA* target genes were in very close proximity to each other. Measurements showed that the median distance between peak summits for the different TFs was between 25 and 53 bp, indicating

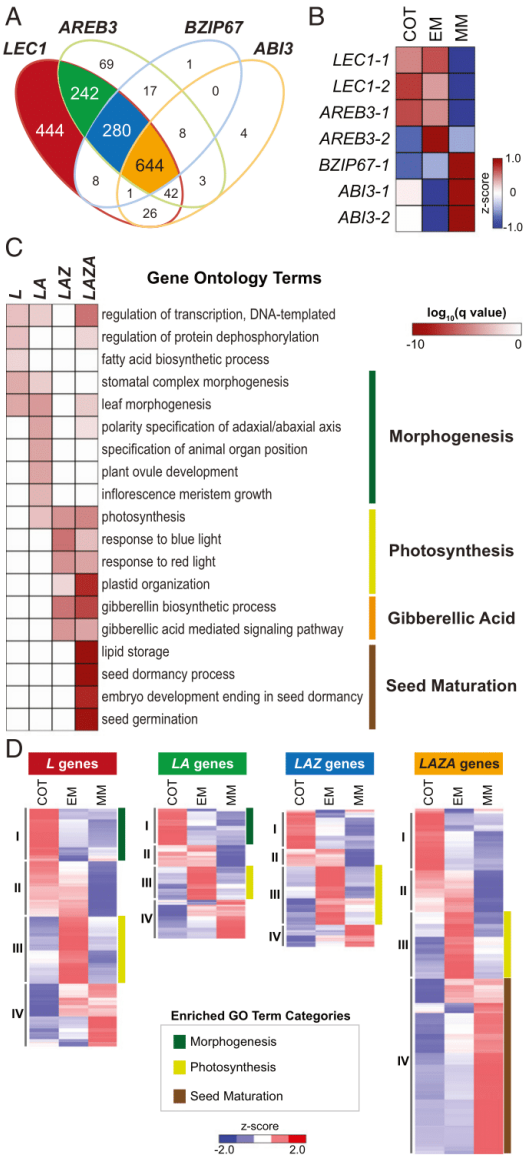


Fig. 2. Overlap between LEC1, AREB3, bZIP67, and ABI3 target genes indicate combinatorial interactions among transcription factors. (*A*) Venn diagram showing the overlap between LEC1, AREB3, bZIP67, and ABI3 target genes. Four major groups were identified: genes targeted by LEC1 (*L* genes, red); genes targeted by LEC1 and AREB3 (*LA* genes, green); genes targeted by LEC1, AREB3, and bZIP67 (*LAZ* genes, blue); and genes targeted by LEC1, AREB3, bZIP67, and ABI3 (*LAZA* genes, orange). Gene lists are given in Dataset S3. (*B*) *LEC1*, *AREB3*, *bZIP67*, and *ABI3* mRNA accumulation in soybean embryos at the COT, EM, and MM stages. mRNA levels were obtained from the Harada Embryo mRNA-Seq Dataset (GEO accession no. GSE99571, ref. 79). (*C*) Heatmap shows the q value significance of GO terms for *L*, *LA*, *LAZ*, and *LAZA* gene sets. GO terms listed are the top 5 most enriched biological process GO terms for each gene set. A comprehensive list of overrepresented GO terms is given in Dataset S1. (*D*) Hierarchical clustering of *L*, *LA*, *LAZ*, and *LAZA* gene sets. Heatmaps show relative embryo mRNA levels at the COT, EM, and MM stages. The major enriched developmental processes associated with each cluster are indicated. Gene lists and GO term enrichments for each cluster are given in Dataset S3.

that the binding sites are clustered. We hypothesized that the binding site clusters represent *cis*-regulatory modules (CRMs) or high occupancy target regions, genomic regions at which multiple, distinct TFs bind productively to regulate gene transcription (43–45). Therefore, we designated these binding site clusters as CRMs and used published criteria (46) to operationally define CRMs as genomic regions whose boundaries are extended by 100 bp on each side of the terminal ChIP peak summits within a cluster, although we also apply this term to *L* genes with single binding sites (Fig. 3*B*). Median CRM sizes for *L*, *LA*, *LAZ*, and *LAZA* genes were 201, 227, 235, and 240 bp, respectively.

Enriched DNA motifs in cis-regulatory modules. To investigate how different TF combinations are recruited to target genes, we identified overrepresented DNA sequence motifs within the CRMs that may serve as TF binding sites. We used de novo DNA motif discovery algorithms to identify the enriched DNA sequence motifs in *L*, *LA*, *LAZ*, and *LAZA* CRMs that are shown as position weight matrices in Fig. 3*C* and Dataset S4. *L* CRMs contained overrepresented CCAAT-like motifs, consistent with the observation that LEC1 is an atypical subunit of the NF-Y complex that binds CCAAT DNA sequences (11, 12). *LA* and *LAZ* CRMs were enriched for G box-like motifs, such as G box- and ABRE-like elements, although the specific position weight matrices identified in *LA* and *LAZ* CRMs differed. bZIP TFs, such as bZIP67 and AREB3, bind G-box motifs (47, 48). *LAZA* CRMs contained overrepresented G box-like motifs that were similar to those found in *LAZ* CRMs, and RY-like motifs. RY motifs are bound by B3 domain transcription factors, such as ABI3 (49). CCAAT-like motifs were not enriched in *LA*, *LAZ*, and *LAZA* CRMs even though LEC1 was bound at these CRMs. CRMs for the *L*, *LA*, and *LAZA* target gene sets were also enriched for BPC1 motifs that are bound by BASIC PENTACYSTEIN TFs that act as transcriptional activators and repressors (50, 51).

To validate the DNA motif discovery analyses, we conducted find individual motif occurrences-receiver operating characteristics-area under the curve (FIMO ROC-AUC) and Homer hypergeometric analyses. The former analysis measures the extent to which DNA motifs in CRMs exhibit similarities to the de novo discovered position weight matrices, whereas the latter measures the percent of CRMs that contain DNA motifs that are exact matches with annotated *cis* elements most closely related to the discovered position weight matrices. Both analyses provided independent evidence in support of the DNA motif discovery results (Fig. 3*D* and *SI Appendix*, Figs. S6 and S7). Together, these findings indicate that the binding sites of different TFs are clustered in the upstream regions of target genes, and they suggest that DNA sequence motifs may represent functional *cis* elements that recruit TFs to CRMs. The number and distribution of DNA motifs in CRMs varied greatly, even among those bound by the same set of TFs, indicating that there is no easily discernible arrangement of DNA motifs for *L*, *LA*, *LAZ*, and *LAZA* genes (*SI Appendix*, Fig. S8).

Binding Site Clusters Are Functional *cis*-Regulatory Modules.

Analysis of cis-regulatory module function. To test the hypothesis that the clustered TF binding sites represent functional CRMs, we determined whether 20 CRMs were sufficient to activate transcription. Functional *cis* elements are generally within 50 bp of ChIP peak summits and, therefore, they should be contained

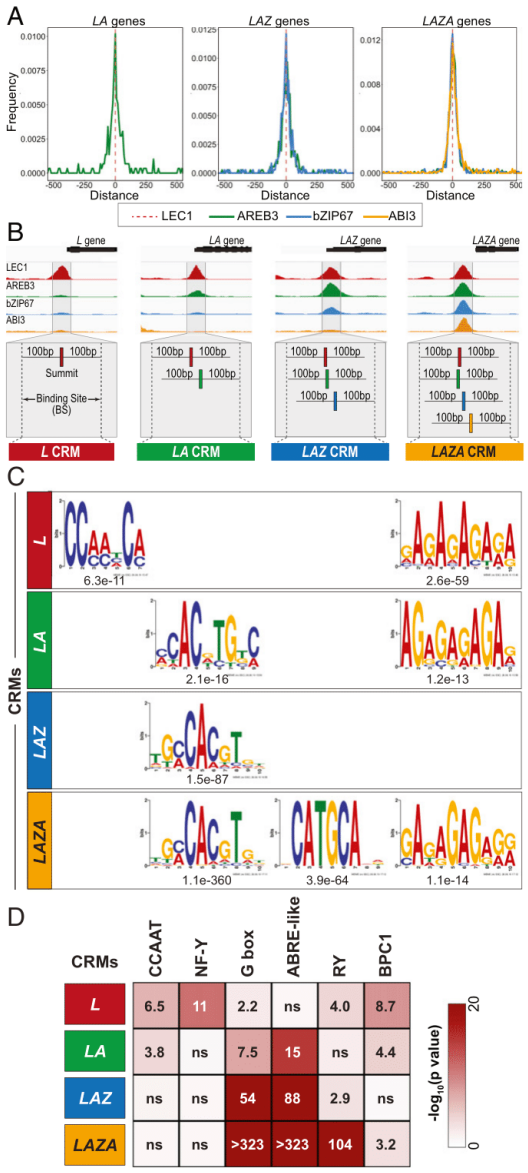


Fig. 3. Clustered binding sites for LEC1, AREB3, bZIP67, and ABI3 define *cis*-regulatory modules. (*A*) Distance between the positions of the ChIP peak summits of AREB3 (green), bZIP67 (blue), and ABI3 (yellow) and the LEC1 ChIP peak summit (dotted red line) for *LA*, *LAZ*, and *LAZA* genes. (*B*) Diagrammatic representation of the strategy used to define CRMs. (*B*, *Upper*) Genome browser representation of ChIP-Seq reads for LEC1 (red), AREB3 (green), bZIP67 (blue), and ABI3 (yellow) in the upstream genomic regions of *L* (Glyma.13G031500), *LA* (Glyma.08G106400), *LAZ* (Glyma.01G057700), and *LAZA* (Glyma.01G186200) genes. (*B*, *Lower*) CRMs are defined as the genomic region whose boundaries extend 100 bp beyond the terminal ChIP peak summits of a cluster. *LA*, *LAZ*, and *LAZA* CRMs were computed by merging the binding sites of 1) LEC1 and AREB3; 2) LEC1, AREB3, and bZIP67; and 3) LEC1, AREB3, bZIP67, and ABI3, respectively. *L* CRMs were defined as LEC1 binding sites of *L* genes. CRM genomic coordinates are listed in Dataset S4. (*C*) Position weight matrices of DNA sequence motifs discovered de novo in *L*, *LA*, *LAZ*, and *LAZA* CRMs and their relative enrichment as indicated by their associated E values. De novo discovered motifs and DNA motifs in the *Arabidopsis* DAP-Seq (67) or Human HOCOMOCOv11 (80) databases most closely related to the discovered motifs are listed in Dataset S4. (*D*) DNA motif enrichment in CRM regions. Heatmaps depict the statistical significance of the enrichment of annotated DNA motifs most closely related to de novo discovered motifs in *L*, *LA*, *LAZ*, and *LAZA* CRM regions, relative to the normal distribution of a population of randomly generated regions. Bonferroni-adjusted *P* values are listed, with a significance threshold of 0.01, with ns denoting no significant difference. Frequencies at which DNA motifs were identified in CRMs are shown in *SI Appendix*, Fig. S7.

within the CRMs (46). Each CRM was inserted upstream of a *35S* minimal promoter fused with the *FIREFLY LUCIFERASE* gene in a plasmid that also contained a *35S:RENILLA LUCIFERASE* gene, as diagrammed in Fig. 4*A*. The CRM activity of the constructs was assessed using transient assays with cotyledon protoplasts from EM-stage embryos. Protoplasts have been used extensively to investigate developmental gene expression (reviewed in ref. 52), consistent with our control experiments showing that seed-specific promoters were active in embryo cotyledon but not in leaf protoplasts (*SI Appendix*, Fig. S9). Transfection of the CRM constructs into embryo cotyledon protoplasts demonstrated that 16 of 20 CRMs were sufficient to induce reporter gene activity at a significantly higher level than the negative control lacking a CRM insert (Fig. 4*B*). Results of these gain-of-function experiments suggest that *L*, *LA*, *LAZ*, and *LAZA* CRMs are functional CRMs containing *cis* elements that are sufficient to activate the transcription of LEC1 target genes during seed development.

Defining* cis *elements. Because most of the tested CRMs activated transcription, we next asked if the overrepresented DNA sequence motifs are functional *cis* elements. We focused initially on 2 *LAZA* genes encoding the α′ subunit of the storage protein β-conglycinin (CG-1, Glyma.10G246300), and the lipid body protein oleosin1 (OLE1, Glyma.20G196600). A 5′-deletion series of the upstream region of each gene that was fused with the *GREEN FLUORESCENCE PROTEIN* (*GFP*) reporter gene in a plasmid that also contained *35S:mCHERRY* was generated (Fig. 4*A*). The *CG-1* CRM contained 4 G box-like and 5 RY-like motifs. As shown in Fig. 4*C*, deletion of the 2 5′-most G box-like and 1 RY-like motif caused a significant reduction in promoter activity relative to wild type, whereas deletion of all but 2 RY-like motifs eliminated detectable promoter activity. For the *OLE1* CRM which contains 8 G box-like and 7 RY-like motifs, deletion of all G box-like and RY-like motifs upstream of the CRM caused only a modest reduction in promoter activity, but deletion of 6 G box-like and all 7 RY-like motifs within the CRM essentially eliminated detectable promoter activity. These results indicate that the enriched DNA motifs are required to activate transcription of the *LAZA* genes.

To test more stringently the hypothesis that the enriched DNA motifs are involved in controlling *LAZA* gene transcription, we specifically mutagenized the G box-like and RY-like motifs in the *CG-1* and *OLE1* CRMs and assessed their ability to activate transcription in embryo protoplasts using the dual luciferase assay. Both of these CRMs were sufficient to activate the minimal promoter in transient assays in embryo cotyledon protoplasts (Fig. 4*A*). Fig. 4*D* shows that mutating all of the G box-like or RY-like motifs in the *CG1* CRM caused a reduction in promoter activity relative to wild type, with the RY-like motif mutations more severely compromising promoter activity. Mutating the *OLE1* CRM motifs also caused a reduction in promoter activity, although mutations of the G box-like motifs more strongly diminished promoter activity than did mutations in RY-like motifs.

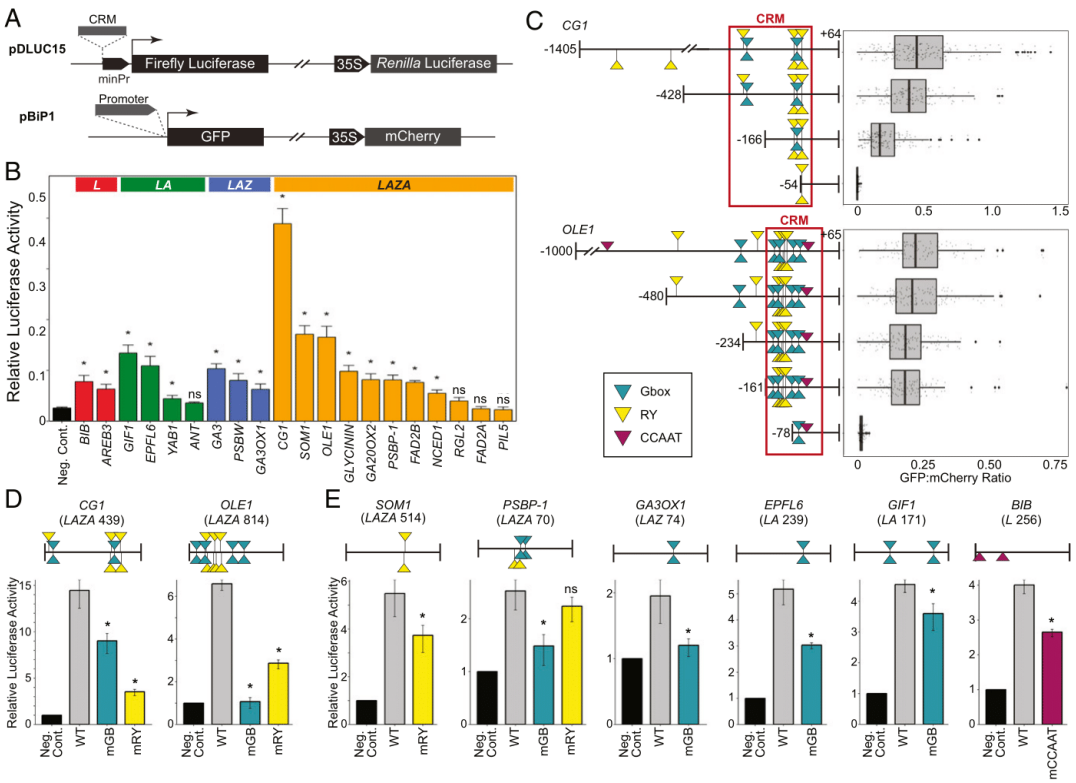


Fig. 4. Functional analysis of *cis*-regulatory modules and *cis* elements in soybean embryo cotyledon cells. (*A*, *Top*) Schematic diagram of the dual-luciferase vector used for the CRM assays which contains a *35S* minimal promoter fused with a *FIREFLY LUCIFERASE* gene and a *35S:RENILLA LUCIFERASE* gene. CRMs were inserted immediately upstream of the *35S* minimal promoter. (*A*, *Bottom*) The vector containing promoterless *GFP* and *35S:mCHERRY* genes were used for the 5′-deletion assays. The 5′-upstream regions were fused with the *GFP* gene. (*B*) CRMs effect on minimal promoter activity in soybean embryo cotyledon protoplasts at the EM stage as measured by the ratio of firefly to *Renilla* luciferase activities. Average value of 3 assays with SEs are plotted. Asterisks denote statistically significant differences relative to the negative control (Neg. Cont., no CRM insert), whereas ns indicates no significant difference ($P < 0.05$, paired, one-tailed *t* tests). (*C*) The 5′-deletion analyses of *CG1* and *OLE1* gene upstream regions. Diagrams of *CG1* and *OLE1* 5′ upstream regions that were fused with the *GFP* gene and used for embryo cotyledon cell transient assays. Teal, yellow, and magenta symbols indicate the positions of G box-like, RY-like, and CCAAT-like DNA motifs, respectively, with a FIMO score greater than 2.4. Positions relative to the transcription start site are indicated. Box plots show the ratios of GFP to mCherry activities for at least 150 transfected protoplasts. (*D*) Regulatory activities of *CG1* and *OLE1* CRMs containing mutations in all detectable G box-like (mGB) or RY-like (mRY) DNA motifs in embryo cotyledon transient assays. (*D*, *Upper*) Diagrams of *CG1* and *OLE1* CRMs with motif positions indicated. (*D*, *Lower*) The ratios of firefly to *Renilla* luciferase activities with SEs are shown ($n = 3$). Asterisks indicate significant activity ratio differences for mutant relative to wild-type (WT) CRMs ($P < 0.05$, one-tailed *t* tests). (*E*) Regulatory activities of *LAZA* (*SOM1* and *PSBP-1*), *LAZ* (*GA3OX1*), *LA* (*EPFL6* and *GIF1*), and *L* (*BIB*) CRMs with mutations in G box-like (mGB), RY-like (mRY), or CCAAT-like (mCCAAT) DNA motifs in embryo cotyledon cell transient assays. Data are presented as in *D*.

The results indicate that G box-like and RY-like motifs within the CRMs play key roles in controlling *LAZA* gene transcription.

We also determined if overrepresented motifs in the CRMs of 2 additional *LAZA* genes, 1 *LAZ* gene, 2 *LA* genes, and 1 *L* gene were involved in controlling promoter activity. As shown in Fig. 4*E*, mutations in RY-like motifs caused a reduction in promoter activity relative to wild type of 1 *LAZA* gene, *SOMNUS1* (*SOM1*, Glyma.12G205700), but RY-like motif mutations did not significantly alter the promoter activity mediated by the *PHOTOSYSTEM II SUBUNIT P-1* (*PSBP1*, Glyma.02G282500) CRM. Rather, mutation of the G box-like motifs in the *PSBP1* CRM caused promoter activity reduction. Promoter activity was reduced relative to wild type in constructs with mutations of G box-like motifs in the CRMs of the *LAZ* gene, *GA3 OXIDASE 1* (*GA3OX1*, *Glyma.04G071000*) and the *LA* genes, *EPIDERMAL PATTERNING FACTOR-LIKE 6* (*EPFL6*, Glyma.14G203100) and *GRF1-INTERACTING FACTOR 1* (*GIF1*, Glyma.10G164100), and of the CCAAT motif in the *L* gene, *BALDIBIS* (*BIB1*, Glyma.12G070300). Although the motif mutations resulted in a significant decrease in promoter activity relative to wild type, in most cases they did not reduce activity to the level of constructs lacking a CRM, suggesting that other *cis* elements are present in the CRMs. Together, our results suggest that CRMs contain enriched DNA sequence motifs that act as functional *cis* elements that are bound by specific TF combinations.

Discussion

LEC1 Regulates Distinct Gene Sets at Different Developmental Stages by Interacting with Specific Combinations of Transcription Factors. The rationale for this study is to determine how LEC1, a central

regulator of seed development, is able to regulate diverse developmental processes at different stages of seed development. We have shown that LEC1 interacts with different combinations of the TFs AREB3, bZIP67, and ABI3 to regulate distinct gene sets, and it is likely that other TFs also interact with these TFs during seed development (21). Gene expression is often dictated by combinatorial interactions among functionally active and distinct TFs in plant and animal cells (reviewed in refs. 53–55). For example, DNA binding experiments with 27 different *Arabidopsis* TFs showed that 63% of the target genes are bound by more than 1 TF, with 8% being bound by 8 or more TFs (39). LEC1 is a NF-Y TF subunit, and a comprehensive study of 154 TFs in human cells showed that 48 operate combinatorially with NF-Y. Thus, NF-Y TFs and their subunits may be particularly likely to coordinate their activities with other TFs.

A key finding, summarized in Fig. 5, is that LEC1 interacts with AREB3, bZIP67, and ABI3 in different combinations to regulate distinct developmental processes: 1) *L* and *LA* genes; 2) *LA*, *LAZ*, and *LAZA* genes; 3) *LAZ* and *LAZA* genes; and 4) *LAZA* genes are primarily involved in morphogenesis, photosynthesis, GA synthesis and signaling, and maturation, respectively. Our finding that LEC1, bZIP67, and ABI3 are involved in regulating maturation genes is consistent with other reports showing that LEC1 acts synergistically with bZIP67 to promote the expression of *CRUCIFERIN C*, *FAD3*, and *DOG1* and that LEC1 and ABI3 operate synergistically to regulate the *OLE1* gene in *Arabidopsis* (18–21).

Our studies establish that LEC1, AREB3, bZIP67, and ABI3 act combinatorially to globally regulate genes involved in maturation and other developmental processes (Fig. 2). Moreover, our studies also provide primary evidence that photosynthetic gene sets are regulated by LEC1 in combination with AREB3 and/or bZIP67 during seed development, as we suggested previously (10). These results indicate that transitions in biological processes during seed development are mediated by qualitative changes in the TF combinations in a cell, as shown for other developmental systems (reviewed in ref. 54).

Combinatorial interactions among LEC1, AREB3, bZIP67, and ABI3 are likely to result, in part, from their ability to interact physically and form a complex. In *Arabidopsis*, the LEC1-NF-YC2 dimer binds with bZIP67 (18–20). Because bZIP67 and AREB3 are closely related and they heterodimerize, the soybean homologs of both TFs are also likely to bind the LEC1-NF-YC dimer and form complexes (*SI Appendix*, Fig. S5; refs. 40 and 41). Although LEC1 and ABI3 operate synergistically to regulate maturation genes, no direct physical interaction between the TFs has been reported. Because *Arabidopsis* ABI3 binds bZIP TFs, AREB3 and/or bZIP67 are likely to serve as a bridge linking LEC1 and ABI3 in a complex (26–28). Given our findings indicating that soybean and *Arabidopsis* homologs of LEC1, AREB3, bZIP67, and ABI3 interact similarly (*SI Appendix*, Fig. S5), these results suggest that the TFs form complexes that regulate distinct gene sets during seed development.

Clustering of *L*, *LA*, *LAZ*, and *LAZA* mRNAs showed that genes involved in morphogenesis, photosynthesis, and maturation are expressed predominately at the COT, EM, and MM stages, respectively, regardless of which TF combination is involved in their regulation (Fig. 2*D*). This finding suggests strongly that combinatorial interactions among LEC1, AREB3, bZIP67, and ABI3 permit distinct developmental processes to be rigidly regulated temporally during seed development. Others have shown that a given combination of transcription factors can generate multiple and distinct spatial and temporal expression patterns (46, 56, 57).

The major temporal shift during seed development is the transition from the morphogenesis to maturation phase, and ABI3 and bZIP67 appears to be the key TFs associated with this transition. Among the 4 TFs, ABI3 is uniquely associated with the activation of maturation genes, and others have established the importance of ABI3 in controlling maturation in *Arabidopsis* (reviewed in refs. 14–17). Based on the *bZIP67* mRNA accumulation pattern (Fig. 2*B*), bZIP67 may trigger the onset of the maturation phase although it is difficult to predict bZIP TF activity, because it is regulated posttranslationally (41). Thus, developmental function reflects qualitative changes in the combination of TFs that are present in a cell.

LEC1's combinatorial interactions with AREB3, bZIP67, and ABI3 suggest that it may act as a pioneer TF. Pioneer TFs are able to bind DNA binding sequences associated with nucleosomes or compacted chromatin and increase chromatin accessibility,

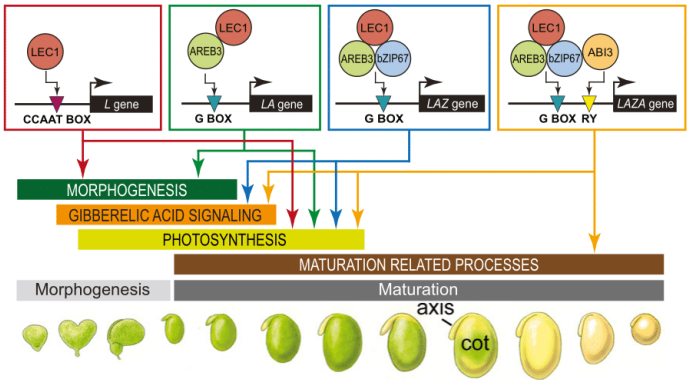


Fig. 5. Model for LEC1 combinatorial interactions with AREB3, bZIP67, and ABI3 in the control of soybean seed development. Combinatorial interactions of LEC1 with AREB3, bZIP67, and ABI3 TFs account for LEC1's ability to regulate different gene sets and diverse developmental processes during seed development. LEC1 interacts with NF-YC and NF-YA subunits to form a NF-Y TF that binds CCAAT motifs of genes involved in morphogenesis and photosynthesis. We propose that a LEC1-NF-YC dimer binds AREB3 or an AREB3-bZIP67 heterodimer to form complexes that bind G box-like motifs of genes involved in 1) morphogenesis and photosynthesis and 2) photosynthesis and GA signaling, respectively. We further propose that an AREB3-bZIP67 heterodimer binds both the LEC1-NF-YC dimer and ABI3 to form a complex that binds G box-like and RY-like motifs in genes involved with photosynthesis, GA signaling, and seed maturation.

thereby promoting the recruitment of other TFs to the target sites (58, 59). LEC1 was recently characterized as a pioneer TF that reprograms the negative regulator of flowering, *FLC*, from a silenced to an active state during embryogenesis (60), and NF-Ys act as pioneer TFs in human cells (61, 62). Because LEC1 in combination with AREB3, bZIP67, and ABI3 does not appear to act as a NF-Y TF and bind CCAAT, it will be important to determine if LEC1 in association with other TFs has pioneer TF activity.

LEC1, AREB3, bZIP67, and ABI3 Bind Functional *cis* Elements to Control Specific Developmental Programs during Seed Development. The striking arrangement of TF binding sites in the upstream regions of LEC1, AREB3, bZIP67, and ABI3 target genes explains, in part, the combinatorial interactions that occur between LEC1 and the other TFs (Fig. 3). We defined CRMs based on the close proximity of the binding sites of between 2 and 4 TFs, although we also used the term to describe genes bound only by LEC1.

A defining characteristic of CRMs is that they contain *cis* elements, short DNA sequences that, when recognized and bound by a TF, lead to transcriptional changes of the associated gene (43–45). Although not comprehensive, our results provide strong evidence that the CRMs contain functional *cis* elements. Gain-of-function experiments showed that 16 of 20 CRMs are sufficient to activate a minimal promoter, indicating that the CRMs contain functional *cis* elements (Fig. 4). Studies with human stem cells showed that between 9 and 25% of TF binding sites, defined as ChIP peaks, were sufficient to activate a minimal promoter (63). Consistent with these observations, others have shown that mutation of 4 soybean maturation genes, *CG1*, *GLYCININ* (Glyma.03G163500), *KUNITZ TRYPSIN INHIBITOR1* (Glyma.01G095000), and *LECTIN1* (Glyma.02G012600), in regions defined as CRMs in this study reduced transgene expression levels in developing seeds (64–66). These findings also demonstrate that results obtained with transgenic plants are reproduced in embryo protoplast transient assays.

The enriched DNA motifs in CRMs, CCAAT-like, G box-like, and RY-like motifs, correspond to annotated *cis* elements known to be bound by LEC1, AREB3, bZIP67, and ABI3 (Fig. 3 and *SI Appendix*, Figs. S6 and S7; refs. 47–49 and 67). Mutation of enriched motifs within CRMs compromised the CRM's ability to enhance minimal promoter activity (Fig. 4), indicating that many of the enriched DNA motifs represent functional *cis* elements. Consistent with this conclusion, others have shown the importance of the G box-like and RY-like motifs in controlling the *Arabidopsis OLE1* gene (21). Moreover, our results showing that *L*, *LA*, *LAZ*, and *LAZA* genes are overrepresented for 1) CCAAT-like, 2) G box-like, and 3) G box-like and RY-like DNA motifs, respectively, suggest that specific sets of DNA motifs are responsible for recruiting different TF combinations to the CRMs (Fig. 3). Together, these results provide insight into the basis of the combinatorial interactions between LEC1, AREB3, bZIP67, and ABI3 by showing that enriched DNA motifs in CRMs represent functional *cis* elements, and the recruitment of LEC1 along with AREB3, bZIP67, and/or ABI3 is determined, in part, by the specific set of *cis* elements present in the CRMs.

Although these results significantly advance our understanding of the TF networks that regulate seed development, several questions remain to be resolved. For example, although *L*, *LA*, *LAZ*, and *LAZA* gene sets are involved in diverse developmental processes, each contains genes that are expressed primarily at different seed development stages (Fig. 2). The temporal patterns are not explained by the DNA motifs in the CRMs. It is possible that other yet to be identified TFs bind the CRMs to dictate temporal expression patterns. Alternatively, "motif grammar," the specific arrangement and/or spacing of *cis* elements in CRMs, may account for temporal expression differences. Others have shown that motif grammar explains temporal and spatial gene expression patterns (46, 56, 68). Another question stems from the observation that several CRMs are able to enhance transcription, even when all of the discernible DNA motifs are mutagenized, suggesting the presence of cryptic *cis* elements in the CRMs (Fig. 4). Latent specificity, the modification of DNA recognition specificity through combinatorial interactions between TFs, offers a potential explanation for this observation (39, 69). Understanding the regulatory logic that controls seed gene expression requires further studies to identify distinct TFs that act combinatorially with LEC1, AREB3, bZIP67, and ABI3 and to decipher the motif grammar governing CRM activity.

Materials and Methods

ChIP-Seq. Soybean plants were grown and seeds were harvested for ChIP experiments as described by Pelletier et al. (10).

ChIP assays were performed, with modifications, as described previously (10) using peptide antibodies against AREB3, bZIP67, and ABI3 that were generated as described in *SI Appendix, SI Materials and Methods*. DNA sequencing libraries were prepared using the NuGEN Ovation Ultralow System V2, and DNA fragments were size selected by electrophoresis and sequenced at 50-bp single-end reads using an Illumina HiSEq. 4000 sequencing system.

ChIP-Seq data were analyzed as described previously (10), with the modifications described in *SI Appendix, SI Materials and Methods*. Briefly, sequencing reads were aligned using Bowtie v0.12.7 (70) and reproducible ChIP-Seq peaks were identified using MACS2 v2.1.0.20140616 (71) and the Irreproducible Discovery Rate pipeline (72) (https://github.com/nboley/idr). Antibody specificity was established by showing extensive overlap in genes bound by a given TF using antibodies generated against 2 separate peptides for each TF, and quality assessment of the ChIP-Seq data followed ENCODE guidelines (73) as summarized in Dataset S2. Because the data analysis pipeline was modified, we reanalyzed ChIP-Seq data for LEC1 ChIP-Seq experiments in EM-stage embryos (10). GO enrichment and hierarchical clustering analyses were performed as described (10). LEC1, AREB3, bZIP67, and ABI3 overlapping binding sites were used to define *L*, *LA*, *LAZ*, and *LAZA* CRMs using the bedtools merge function (74), as described in Fig. 3*B* and *SI Appendix, SI Materials and Methods*. De novo DNA motif discovery analysis of CRMs was performed using the MEME-ChIP tool from the MEME suite v5.0.5 (75), and DNA motif enrichment analysis was performed using the motifEnrich tool from HOMER (76) (homer.ucsd.edu/homer/motif/index.html) and the ROCR R package v1.0.2 (77), as detailed in *SI Appendix, SI Materials and Methods*.

Transient Assays in Embryo Cotyledon Protoplasts. Transient assays in protoplasts isolated from soybean embryo cotyledons and *Arabidopsis* leaves were performed as described by Yoo et al. (78) with the modifications described in *SI Appendix, SI Materials and Methods*. Plasmids used in the transient assays were constructed as detailed in *SI Appendix, SI Material and Methods*, and primers used for DNA manipulations are listed in Dataset S5. Activities of 5′-deletion constructs were evaluated by measuring GFP and mCherry activity in transfected soybean embryo protoplasts using fluorescence filters GFP-3035B and TXrRED-4040B (Semrock) with an Eclipse E600 microscope (Nikon) and calculating GFP to mCherry ratios as described in *SI Appendix, SI Material and Methods*.

Measurements of firefly and *Renilla* luciferase activities in CRM gain-of-function experiments with soybean embryo cotyledon protoplasts were made using the Dual-Luciferase Reporter Assay System (Promega) with a TriStar2 LB942 multiplate reader (Berthold) as described in *SI Appendix, SI Material and Methods*.

Constructs for bimolecular fluorescence complementation assays were created by fusing open reading frames for the TFs with the amino or carboxyl terminus of the citrine fluorescence protein. Constructs were transfected into *Arabidopsis* leaf protoplasts as described in *SI Appendix, SI Material and Methods*. Citrine fluorescence signal was detected using the GFP-3035B filter of an Eclipse E600 microscope (Nikon).

Data Availability. Data are available at Gene Expression Omnibus under the following accessions: AREB3-EM (GSE101672), bZIP67-EM (GSE101672), ABI3-EM (GSE101649), AREB3B-EM (GSE140699), bZIP67B-EM (GSE140701), and ABI3B-EM (GSE140700).

ACKNOWLEDGMENTS. We thank Bo Liu for advice about antibody production and testing, Dan Runcie for advice about statistical analyses, Savithramma Dinesh-Kumar and Samuel Hazen for providing plasmids containing the citrine and dual-luciferase constructs, respectively, and Sharon Belkin for the soybean seed diagrams. This work was supported by a grant from the National Science

Foundation Plant Genome Research Program (to R.B.G. and J.J.H.). L.J. was supported by a Coordination for the Improvement of Higher Level Personnel grant (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brazil, No. 99999.013505/2013-00).

1. T. A. Steeves, The evolution and biological significance of seeds. *Can. J. Bot.* **61**, 3550–3560 (1983).
2. J. Palovaara, T. de Zeeuw, D. Weijers, Tissue and organ initiation in the plant embryo: A first time for everything. *Annu. Rev. Cell Dev. Biol.* **32**, 47–75 (2016).
3. J. T. Puthur, A. M. Shackira, P. P. Saradhi, D. Bartels, Chloroembryos: A unique photosynthesis system. *J. Plant Physiol.* **170**, 1131–1138 (2013).
4. J. J. Harada, "Seed maturation and control of germination" *Advances in Cellular and Molecular Biology of Plants*, B. A. Larkins, I. K. Vasi, Eds. (Cellular and Molecular Biology of Seed Development, Kluwer Academic Publishers, Dordrecht, 1997), vol. 4, pp. 545–592.
5. J. Vicente-Carbajosa, P. Carbonero, Seed maturation: Developing an intrusive phase to accomplish a quiescent state. *Int. J. Dev. Biol.* **49**, 645–651 (2005).
6. L. Jo, J. M. Pelletier, J. J. Harada, Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *J. Integr. Plant Biol.* **61**, 564–580 (2019).
7. D. W. Meinke, A homoeotic mutant of Arabidopsis thaliana with leafy cotyledons. *Science* **258**, 1647–1650 (1992).
8. M. West *et al*., LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. *Plant Cell* **6**, 1731–1745 (1994).
9. T. Lotan *et al*., Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **93**, 1195–1205 (1998).
10. J. M. Pelletier *et al*., LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6710–E6719 (2017).
11. V. Calvenzani *et al*., Interactions and CCAAT-binding of Arabidopsis thaliana NF-Y subunits. *PLoS One* **7**, e42902 (2012).
12. N. Gnesutta, D. Saad, A. Chaves-Sanjuan, R. Mantovani, M. Nardini, Crystal structure of the Arabidopsis thaliana L1L/NF-YC3 histone-fold dimer reveals specificities of the LEC1 family of NF-Y subunits in plants. *Mol. Plant* **10**, 645–648 (2017).
13. R. W. Kwong *et al*., LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* **15**, 5–18 (2003).
14. S. A. Braybrook, J. J. Harada, LECs go crazy in embryo development. *Trends Plant Sci.* **13**, 624–630 (2008).
15. A. Fatihi *et al*., Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds. *Plant Sci.* **250**, 198–204 (2016).
16. L. Lepiniec *et al*., Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. *Plant Reprod.* **31**, 291–307 (2018).
17. M. Santos-Mendoza *et al*., Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. *Plant J.* **54**, 608–620 (2008).
18. F. M. Bryant, D. Hughes, K. Hassani-Pak, P. J. Eastmond, Basic LEUCINE ZIPPER TRANSCRIPTION FACTOR67 transactivates *DELAY OF GERMINATION1* to establish primary seed dormancy in Arabidopsis. *Plant Cell* **31**, 1276–1288 (2019).
19. A. Mendes *et al*., bZIP67 regulates the omega-3 fatty acid content of Arabidopsis seed oil by activating fatty acid desaturase3. *Plant Cell* **25**, 3104–3116 (2013).
20. A. Yamamoto *et al*., Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J.* **58**, 843–856 (2009).
21. S. Baud *et al*., Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by master transcriptional regulators in Arabidopsis seed. *Plant Physiol.* **171**, 1099–1112 (2016).
22. C. Boulard *et al*., LEC1 (NF-YB9) directly interacts with LEC2 to control gene expression in seed. *Biochim. Biophys. Acta. Gene Regul. Mech.* **1861**, 443–450 (2018).
23. S. A. Braybrook *et al*., Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3468–3473 (2006).
24. W. Reidt *et al*., Gene regulation during late embryogenesis: The RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant J.* **21**, 401–408 (2000).
25. S. Bensmihen, J. Giraudat, F. Parcy, Characterization of three homologous basic leucine zipper transcription factors (bZIP) of the ABI5 family during Arabidopsis thaliana embryo maturation. *J. Exp. Bot.* **56**, 597–603 (2005).
26. R. Alonso *et al*., A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. *Plant Cell* **21**, 1747–1761 (2009).
27. S. Nakamura, T. J. Lynch, R. R. Finkelstein, Physical interactions between ABA response loci of Arabidopsis. *Plant J.* **26**, 627–635 (2001).
28. P. Lara *et al*., Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. *J. Biol. Chem.* **278**, 21003–21011 (2003).
29. J. Pelletier *et al*., Identification of GLYMA.06G314400 and GLYMA.13G317000 binding sites in soybean early maturation embryos. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101672. Deposited 20 July 2017.
30. J. Pelletier *et al*., Identification of GLYMA.08G357600 binding sites in soybean early maturation embryos. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101649. Deposited 19 July 2017.
31. J. Pelletier *et al*., Identification of GLYMA.06G314400 binding sites in soybean early maturation embryos. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140699. Deposited 19 November 2019.
32. J. Pelletier *et al*., Identification of GLYMA.13G317000 binding sites in soybean early maturation embryos. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140701. Deposited 19 November 2019.
33. J. Pelletier *et al*., Identification of GLYMA.08G357600 binding sites in soybean mid-maturation embryos II. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140700. Deposited 19 November 2019.
34. J. Pelletier *et al*., Identification of LEC1 binding sites in soybean embryos at 3 developmental stages. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99882. Accessed 9 June 2017.
35. P. J. Farnham, Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.* **10**, 605–616 (2009).
36. J. J. Harada *et al*., Gene expression changes in the development of the soybean seed-cotyledon stage. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57606. Accessed 13 May 2014.
37. J. J. Harada *et al*., Gene expression changes in the development of the soybean seed-early maturation stage. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46096. Accessed 16 April 2013.
38. J. J. Harada *et al*., Gene expression changes in the development of the soybean seed mid-maturation (B1) stage. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99109. Accessed 19 May 2017.
39. K. S. Heyndrickx, J. Van de Velde, C. Wang, D. Weigel, K. Vandepoele, A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. *Plant Cell* **26**, 3894–3910 (2014).
40. Arabidopsis Interactome Mapping Consortium, Evidence for network evolution in an Arabidopsis interactome map. *Science* **333**, 601–607 (2011).
41. M. Jakoby *et al*.; bZIP Research Group, bZIP transcription factors in Arabidopsis. *Trends Plant Sci.* **7**, 106–111 (2002).
42. D. S. Johnson, A. Mortazavi, R. M. Myers, B. Wold, Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
43. E. H. Davidson *et al*., A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).
44. M. B. Gerstein *et al*.; modENCODE Consortium, Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
45. The modEncode Consortium *et al*., Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**, 1787–1797 (2010).
46. R. P. Zinzen, C. Girardot, J. Gagneur, M. Braun, E. E. Furlong, Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**, 65–70 (2009).
47. T. Izawa, R. Foster, N. H. Chua, Plant bZIP protein DNA binding specificity. *J. Mol. Biol.* **230**, 1131–1144 (1993).
48. S. Y. Kim, H.-J. Chung, T. L. Thomas, Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the Dc3 promoter using a modified yeast one-hybrid system. *Plant J.* **11**, 1237–1251 (1997).
49. G. Mönke *et al*., Seed-specific transcription factors ABI3 and FUS3: Molecular interaction with DNA. *Planta* **219**, 158–166 (2004).
50. R. J. Meister *et al*., Definition and interactions of a positive regulatory element of the Arabidopsis INNER NO OUTER promoter. *Plant J.* **37**, 426–438 (2004).
51. J. Xiao *et al*., Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat. Genet.* **49**, 1546–1552 (2017).
52. J. Sheen, Signal transduction in maize and Arabidopsis mesophyll protoplasts. *Plant Physiol.* **127**, 1466–1475 (2001).
53. M. Bemer, A. D. J. van Dijk, R. G. H. Immink, G. C. Angenent, Cross-Family transcription factor interactions: An additional layer of gene regulation. *Trends Plant Sci.* **22**, 66–80 (2017).
54. I. S. Peter, Regulatory states in the developmental control of gene expression. *Brief. Funct. Genomics* **16**, 281–287 (2017).
55. A. Reményi, H. R. Schöler, M. Wilmanns, Combinatorial control of gene expression. *Nat. Struct. Mol. Biol.* **11**, 812–815 (2004).
56. Z. Ouyang, Q. Zhou, W. H. Wong, ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21521–21526 (2009).
57. C. I. Swanson, N. C. Evans, S. Barolo, Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev. Cell* **18**, 359–370 (2010).
58. A. Mayran, J. Drouin, Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **293**, 13795–13804 (2018).
59. K. S. Zaret, J. S. Carroll, Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
60. Z. Tao *et al*., Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature* **551**, 124–128 (2017).
61. A. J. Oldfield *et al*., Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. *Mol. Cell* **55**, 708–722 (2014).
62. R. I. Sherwood *et al*., Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
63. T. S. Barakat *et al*., Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell* **23**, 276–288.e8 (2018).
64. P. A. Lessard, R. D. Allen, T. Fujiwara, R. N. Beachy, Upstream regulatory sequences from two beta-conglycinin genes. *Plant Mol. Biol.* **22**, 873–885 (1993).
65. G. R. De Paiva, "Transcriptional regulation of seed protein genes," PhD Dissertation, University of California, Los Angeles, CA (1994).
66. R. Yadegari, "Regional specification and cellular differentiation during early plant embryogenesis," PhD Dissertation, University of California, Los Angeles, CA (1996).
67. R. C. O'Malley *et al*., Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**, 1280–1292 (2016).
68. M. I. Arnone, E. H. Davidson, The hardwiring of development: Organization and function of genomic regulatory systems. *Development* **124**, 1851–1864 (1997).

69. M. Slattery *et al*., Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**, 1270–1282 (2011).
70. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
71. Y. Zhang *et al*., Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
72. Q. Li, J. B. Brown, H. Huang, P. J. Bickel, Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
73. S. G. Landt *et al*., ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
74. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
75. P. Machanick, T. L. Bailey, MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
76. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
77. T. Sing, O. Sander, N. Beerenwinkel, T. Lengauer, ROCR: Visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
78. S. D. Yoo, Y. H. Cho, J. Sheen, Arabidopsis mesophyll protoplasts: A versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).
79. J. J. Harada *et al*., Gene expression changes during embryo and seed maturation, quiescence and germination in soybean. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99571. Accessed 1 June 2017.
80. I. V. Kulakovskiy *et al*., HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).

PNAS
www.pnas.org

Supplementary Information for

Combinatorial Interactions of the LEC1 Transcription Factor Specify Diverse Developmental Programs during Soybean Seed Development

Leonardo Jo, Julie M. Pelletier, Ssu-Wei Hsu, Russell Baden, Robert B. Goldberg, and John J. Harada

John J. Harada
Email: jjharada@ucdavis.edu
Robert B. Goldberg
Email: bobg@g.ucla.edu

This PDF file includes:

SI Materials and Methods
Figures S1 to S9
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S5

www.pnas.org/cgi/doi/10.1073/pnas.1918441117

Supplementary Information Text

SI Materials and Methods

Chromatin Immunoprecipitation-DNA Sequencing

Antibodies

ChIP assays were performed according to Pelletier et al. (1). Antibodies to soybean AREB3 (Glyma.04G124200 and Glyma.06G314400), bZIP67 (Glyma.13G317000), and ABI3 (Glyma.08G357600 and Glyma.18G176100), respectively, were raised in rabbits against the following peptides: PEPRYQIRRTSSASF, MSLQQPNQEVPLQEP, and QNQGSDPHARMGGDNC. Peptides were conjugated to KLH, and antibodies were affinity purified (Eurogentec, AS-DOUB-LXP). Antibodies generated against different peptides from AREB3, bZIP67 and ABI3 (SDTRRPGRKRGTSED, TENLRAMRRPLSASW, and QQNPDPGLGGTVGEC, respectively) were used to validate TF specificity in the ChIP assays (Supplemental Dataset S2).

DNA Sequencing

AREB3, bZIP67 and ABI3 ChIP and input DNA libraries were prepared using the NuGEN Ovation Ultralow System V2 and fragments were enriched by 15 cycles of PCR. Libraries were size-selected (200 to 600 bp) by agarose gel electrophoresis and purified using the Qiagen MinElute Gel Extraction kit. Libraries were multiplexed and sequenced to obtain 50 bp single-end reads with the Illumina HiSeq 4000 sequencing system.

Data Analysis

ChIP Seq

ChIP-Seq data were analyzed as described previously (1), with modifications. ChIP-Seq reads were quality filtered and uniquely mapped to the Wm82.a2.v1 genome (Gmax 275) using Bowtie v0.12.7 (2), allowing up to two mismatches. Redundant reads were removed using samtools v.0.1.19 (3). Sequencing library complexity and quality were evaluated using the non-redundant fraction (NRF) and strand cross-correlation (CC) analysis (phantompeakqualtools: https://code.google.com/p/phantompeakqualtools), following ENCODE guidelines for ChIP-Seq quality standards (4). Quality assessments of the libraries are summarized in Dataset S2.

ChIP-Seq peaks were identified using MACS2 v2.1.0.20140616 (5) with two biological replicates, with a loose threshold of $P < 0.1$. The estimated ChIP fragment size was independently determined for each sample by MACS2 using default parameters. Genomic regions bound by a TF with statistical significance were determined by ChIP-Seq Peaks that were reproducible between the two independent biological replicates, as determined with the irreproducible discovery rate (IDR) pipeline (6)(https://github.com/nboley/idr) and the conservative IDR threshold of 0.01. Genes bound by a TF were defined as those with a reproducible peak within a 1 kb window upstream of the genes' transcription start site (TSS).

Target Genes

Target genes are defined as genes that are bound, as determined in ChIP-Seq experiments, and coexpressed by a TF as described previously (1). Because all of the TFs are expressed primarily in the embryo relative to the seed coat (Figs. 2B and S2), we used the Harada-Goldberg Soybean Seed Development LCM RNA-Seq Dataset (GEO accessions, GSE 57606, GSE46096, and GSE99109) to identify coexpressed genes that are five-fold upregulated in embryo cotyledon abaxial or adaxial parenchyma versus seed coat hilum and parenchyma subregions ($q < 0.01$).

GO Term Enrichment Analysis
GO term enrichment analysis was performed as described previously (1).

Hierarchical Clustering Analysis
L, LA, LAZ and *LAZA* gene sets were clustered as described previously (1), using the Harada Embryo mRNA-Seq Dataset, GEO accession no. GSE99571. TMM-normalized mRNA levels at the COT, EM and MM stages were averaged across biological replicates and clustered using dchip 2010_01 (7).

Cis-Regulatory Modules
LA, LAZ and *LAZA* CRMs were identified using the strategy shown in Figure 3B and the bedtools merge function (8) to combine overlapping 200 bps regions around the ChIP peak summits of: (i) LEC1 and AREB3 peaks in the upstream regions of *LA* genes, (ii) LEC1, AREB3, and bZIP67 peaks in the upstream regions of *LAZ* genes, and (iii) LEC1, AREB3, bZIP67 and ABI3 peaks in the upstream regions of *LAZA* genes (Supplemental Dataset S4). *L* CRMs were defined as 200 bp regions surrounding the LEC1 peak summit in the upstream regions of *L* genes.

DNA Motif Analysis
De novo DNA motif discovery in CRM regions was performed using the MEME-ChIP tool from the MEME suite v5.0.5 (9), with an E-value cutoff of 0.01. Default MEME discovery settings were used, except that the maximum discovered motif length was set to 10 nucleotides. Tomtom tool compared the *de novo* discovered DNA motifs to motifs found in the Arabidopsis DAP-Seq TF motifs database (10) and the Human HOCOMOCOv11 database (11). The top ranked motifs in the database determined by Tomtom are listed in the Supplemental Dataset S4.

Annotated motifs most similar to the *de novo* discovered motifs were screened for enrichment using HOMER (12)(homer.ucsd.edu/homer/motif/index.html) as described previously (1). We created "random CRMs", genomic regions for randomly selected genes that were of comparable number, length, and position to the relevant CRMs. Each set of CRMs were compared against 1,000 sets of "random CRMs" to calculate the P value significance of enrichment. P values were adjusted using the Bonferroni method, and a significance threshold of $P < 0.01$ was used to identify significantly enriched DNA motifs. *De novo* discovered motif enrichment was evaluated using Receiver Operating Characteristic - Area Under Curve (ROC-AUC) analyses similar to that reported by Siggers et al. (13). *De novo* discovered motifs and annotated motifs found in JASPAR were scanned in *L, LA, LAZ* and *LAZA* CRMs using the FIMO tool from the MEME Suite, with a P value threshold of 0.01. ROC curve analyses were performed to evaluate whether the motifs found in the CRMs (true positives) scored more highly than motifs found in genomic regions of comparable length and position relative to randomly selected genes (true negatives), and the AUCs were recorded using the ROCR package v1.0.4 (14) in R v3.5.1. Each CRM set was compared against 100 sets of "random CRMs".

Recombinant DNA Manipulations and Plasmid Construction
5' Deletions Constructs
The *GFP - 35S:mCHERRY* plasmid used to analyze 5' deletions of the *CG1* (Glyma.10G246300) and *OLE1* (Glyma.20G196600) upstream regions was constructed by replacing the *sGFP* open-reading frame (ORF) with the *mCHERRY* ORF in the *35S:sGFP(S65T):NOPALINE SYNTHASE terminator (NOSt)* plasmid (15). The *35S:mCherry:NOSt* gene was then cloned into the EcoRI site of the *35S:sGFP(S65T):NOSt* plasmid to generate the pBiP1 plasmid. DNA fragments with the

CG1 and *OLE1* upstream regions and their 5' deletion derivatives were generated by PCR amplification using the primers listed in Dataset S5 and cloned in place of the *35S* promoter upstream of the *sGFP* gene in the pBiP1 vector.

CRM Constructs
The backbone plasmid (pDLUC15) for the dual-luciferase assays was constructed by replacing the *35S:sGFP* and *35S:mCHERRY* genes in pBiP1 with the *35S MINIMAL PROMOTER:FIREFLY LUCIFERASE* and *35S:RENILLA LUCIFERASE* genes from the *pLAH-LARm* plasmid (16). CRM fragments were PCR amplified from genomic DNA using the primers listed in Dataset S5, and they were inserted into the pDLUC15 vector upstream of the *35S MINIMAL PROMOTER* using the In-Fusion® HD Cloning Kit, according the manufacturer's instructions. Mutant derivatives of CRMs were synthesized by Twist Bioscience and inserted into pDLUC15. The DNA sequences of wild type and mutant CRMs are listed in Dataset S5.

BiFC Constructs
Backbone plasmids for the bimolecular fluorescence complementation experiments, pBiP1cCitrine and pBiP1nCitrine, were made by replacing the *35S:sGFP* gene in pBiP1 with either the *cCitrine* or the *nCitrine* cassettes from the pSPDK919 and pSPDK920 plasmids, respectively, and fused with the *NOPALINE SYNTHASE* (*NOS*) promoter (17). ORFs lacking a stop codon for LEC1 (Glyma.07G268100), AREB3 (Glyma.06G314400), bZIP67 (Glyma.13G317000), ABI3 (Glyma.18G176100) and NF-YC (Glyma.08G17630) were amplified from EM embryo cDNA and cloned inframe into MluI and XbaI sites in the ORFs in pBiP1 *cCitrine* or *nCitrine* plasmids. *SPEECHLESS* ORF (Glyma.04G238400) was cloned and used as a negative control for BiFC experiments.

DNA sequences of all constructs were confirmed. DNA sequences and primers used for the DNA manipulation experiments are found in the Dataset S5.

Transient Assays with Soybean Embryo Cotyledon and Arabidopsis Leaf Mesophyll Protoplasts
Soybean embryo cotyledon and Arabidopsis leaf mesophyll protoplast isolations were performed according to Yoo et al (18), with modifications. Cotyledons from soybean embryos at the EM stage (6-7 mm seeds) and well-expanded leaves from 3-4 week-old Arabidopsis plants were cut into 0.5-1 mm strips, immersed in an enzyme solution containing 1% (w/v) Cellulase RS "Onozuka" and 0.25% (w/v) Macerozyme R-10 (Yakult Pharmaceutical Ind. Co. LTD.) and vacuum infiltrated for 15 minutes. Tissues were incubated in the enzyme solution in the dark with gentle agitation (50 rpm) for 2 hours at room temperature. Protoplasts were filtered through a 100 µm nylon mesh, washed two times in W5 buffer (154 mM NaCl, 125 mM CaCl2, 5 mM KCl, 2 mM MES pH 5.8, 5 mM Glucose), and incubated on ice for 30 minutes. Cell number was determined using a hemocytometer.

Protoplast transfection experiments were performed as described by Yoo et al. (18) using approximately 5 x 10^5 cells per transfection and 10 µg (dual-luciferase assay) or 20 µg (fluorescence assays) of plasmid DNA. Transfected protoplasts were incubated in W5 buffer in the light at 25°C for 18 hours before fluorescence or luciferase activity measurements were done. Transient assays with protoplasts have been used extensively to study developmental gene expression (19-23).

Fluorescence Microscopy

For the *CG1* and *OLE1* 5' deletion assays, soybean embryo cotyledon protoplasts were observed using an Eclipse 600 microscope equipped with 20X Plan-Apo Fluor objective (Nikon). GFP and mCherry fluorescence was imaged using fluorescence filters GFP-3035B and TXrRED-4040B from Semrock, respectively. Bright field and GFP and mCherry fluorescence images were acquired by an OptiMOS camera (Q Imaging) controlled by the μManager software package (24, 25). ImageJ (www.imagej.nih.gov/ij) software was used to measure the pixel integrated density of GFP and mCherry signal from individual protoplasts. Relative GFP activity was determined by averaging the GFP:mCherry signal ratio from at least 150 individual protoplasts.

For BiFC experiments, transfected protoplasts from Arabidopsis leaves were observed with the Eclipse 600 microscope as described above. The GFP-3035B filter was used to detect citrine fluorescence.

Dual Luciferase Assay

Firefly and *Renilla* luciferase activities were assayed according to (26), with modifications. Following an 18 hour incubation, transfected protoplasts were harvested. Protoplasts lysis and luciferase activity measurements were made using the Dual-Luciferase Reporter Assay System (Promega) following manufacturer's instructions, with the exception that the reactions were scaled in half. CRM activity was determined by averaging firefly:*Renilla* luciferase activity ratio of 10 plate measurements and three biological replicates. Wild type CRM activity was compared against the activity of a no insert, negative control, using the Student's t test (one-tailed, paired), with a significance threshold of 0.05. For mutant CRM analyses, the ratio of each biological replicate was normalized to the no insert control to determine relative CRM activity. Differences between mutant and wild type CRM activities were evaluated using the Student's t test (one-tailed, non-paired), with a significance threshold of 0.05.

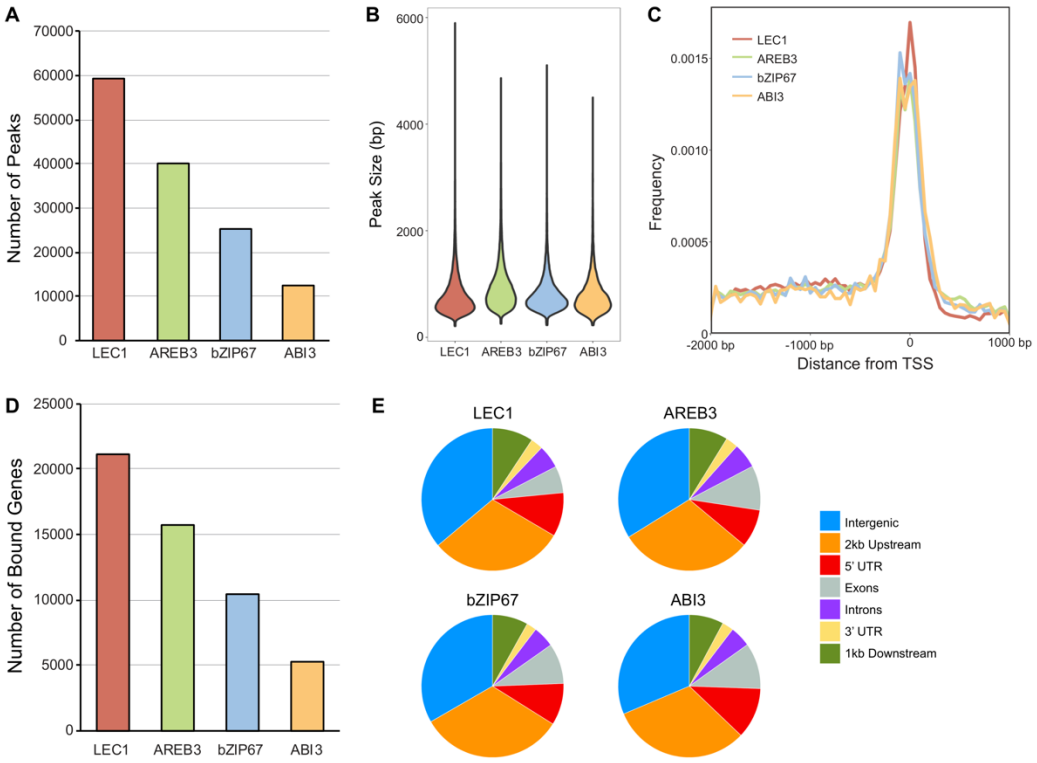


Figure S1. Profiling of LEC1, AREB3, bZIP67 and ABI3 ChIP-Seq peaks.
(*A*) Number of LEC1, AREB3, bZIP67 and ABI3 peaks in soybean embryos at the EM stage. (*B*) Distribution of LEC1, AREB3, bZIP67 and ABI3 peak sizes. (*C*) Frequency plot of the distance of LEC1, AREB3, bZIP67 and ABI3 ChIP-Seq peak summits relative to the transcription start sites (TSS) of the associated genes. (*D*) Numbers of genes bound by LEC1, AREB3, bZIP67 and ABI3 within 1kb of the TSS of the associated gene. (*E*) Association of LEC1, AREB3, bZIP67 and ABI3 ChIP-Seq peaks with genomic features.

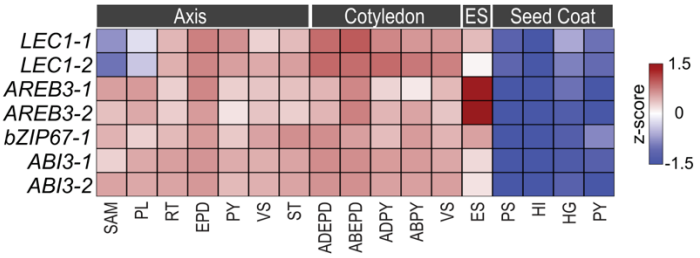


Figure S2. Spatial patterns of *LEC1*, *AREB3*, *bZIP67* and *ABI3* mRNA accumulation in soybean seed subregions at the early maturation stage.
mRNA accumulation data were obtained from the Harada-Goldberg Soybean Seed Development LCM RNA-Seq Dataset (GEO accessions GSE116036). Abbreviations: ABEPD, abaxial epidermis; ABPY, abaxial parenchyma; ADEPD, adaxial epidermis; ADPY, adaxial parenchyma; EPD, epidermis; ES, endosperm; HI, hilum; HG, hourglass; PL, plumule; PS, palisade; PY, parenchyma; RT, root tip; SAM, shoot apex; ST, stele; VS, vasculature.

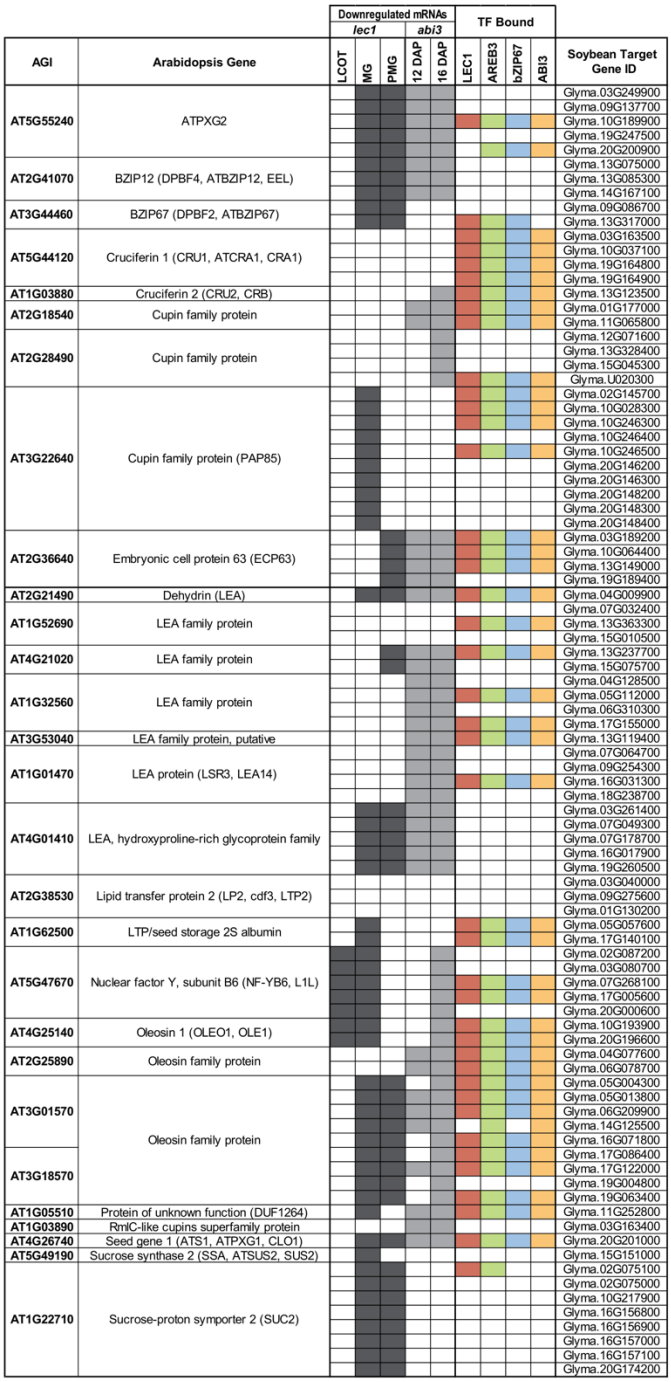


Figure S3. LEC1, AREB3, bZIP67 and ABI3 work in combination to regulate seed maturation genes.
Gray-scale filled squares indicate Arabidopsis orthologs of soybean genes that are downregulated in Arabidopsis *lec1* (dark gray) or *abi3* (light gray) mutants at the indicated seed development stage. Colored squares indicate that the soybean gene is bound by LEC1, AREB3, bZIP67, and/or ABI3 in soybean embryos at the EM stage. The maturation gene list is from Pelletier et al. (1). Abbreviations: LCOT, linear cotyledon; MG, mature green; PMG, post mature green.

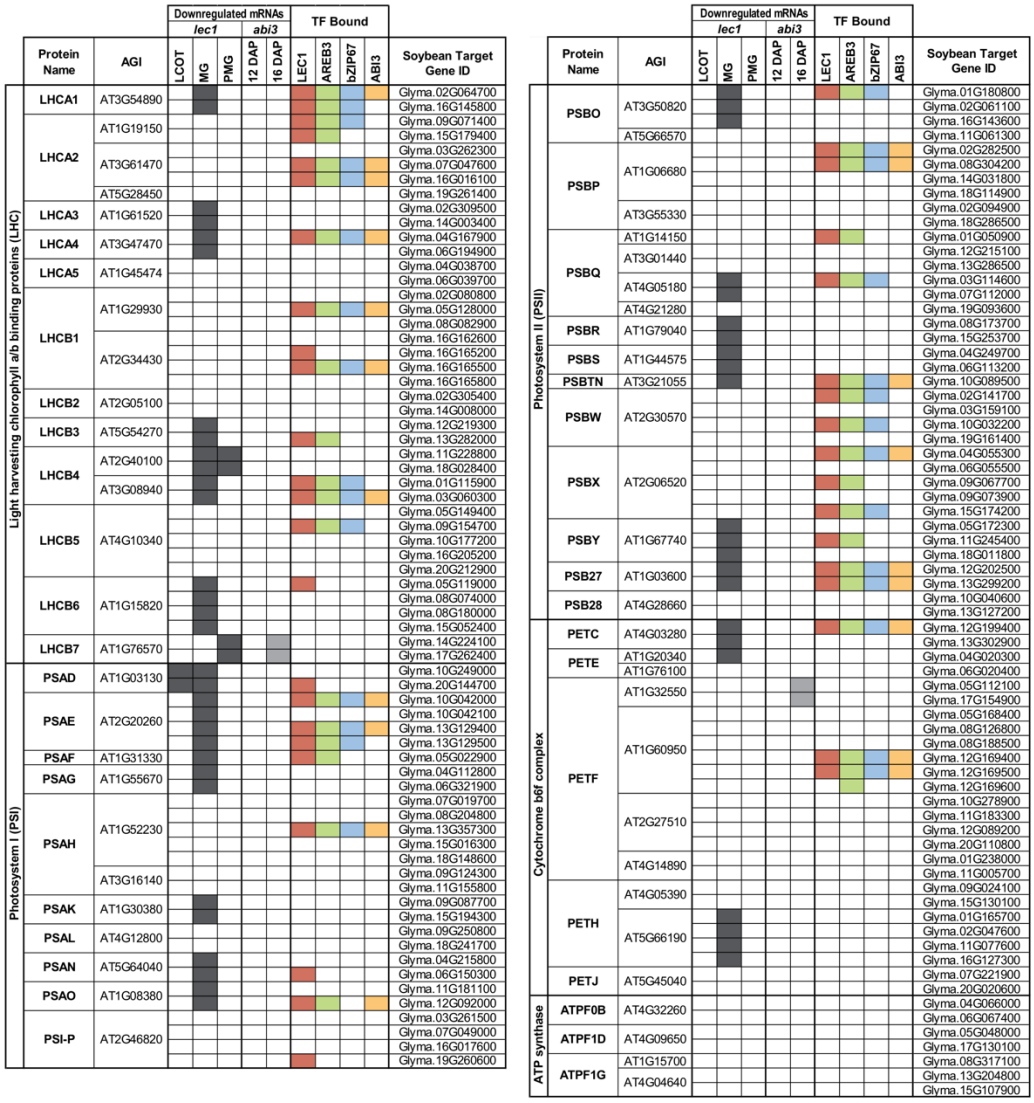


Figure S4. LEC1, AREB3, bZIP67 and ABI3 regulate photosynthesis genes.
Gray-scale filled squares indicate that Arabidopsis orthologs of soybean genes are downregulated in Arabidopsis *lec1* (dark gray) or *abi3* (light gray) mutants at the indicated seed development stage. Colored squares indicate that the soybean gene is bound by LEC1, AREB3, bZIP67, and/or ABI3 in soybean embryos at EM stage. The photosynthesis gene list is from Pelletier *et al.* (1). Abbreviations are as in Fig. S3.

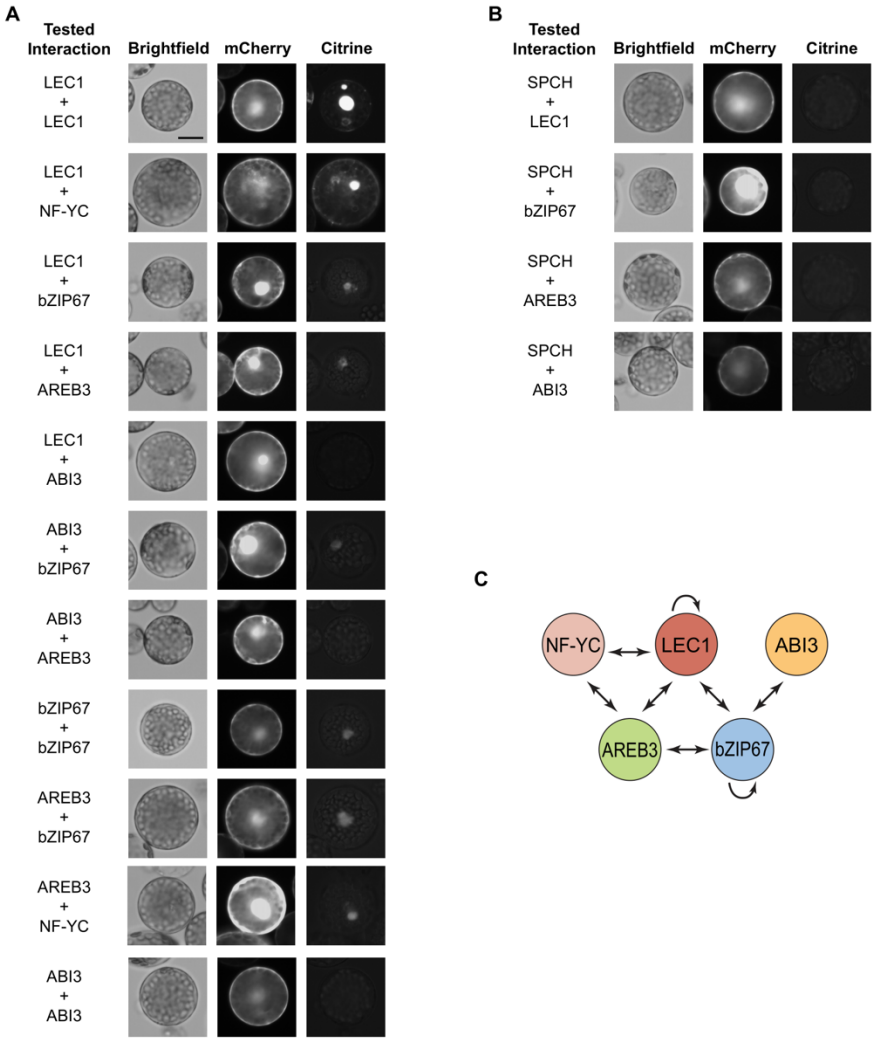


Figure S5. Interactions between LEC1, AREB3, bZIP67 and ABI3 transcription factors.
Bimolecular fluorescence complementation (BiFC) analyses were conducted in Arabidopsis leaf protoplasts. (*A*) Bright field and mCherry and citrine fluorescent images of protoplasts transfected with BiFC plasmids. mCherry signal indicates that a protoplast has been transfected, and citrine signal indicates interaction between the listed TFs. (*B*) The SPEECHLESS (SPCH) TF was used as a negative control. (*C*) Summary of LEC1, AREB3, bZIP67 and ABI3 interactions detected in BiFC experiments. Scale bar, 20 μm.

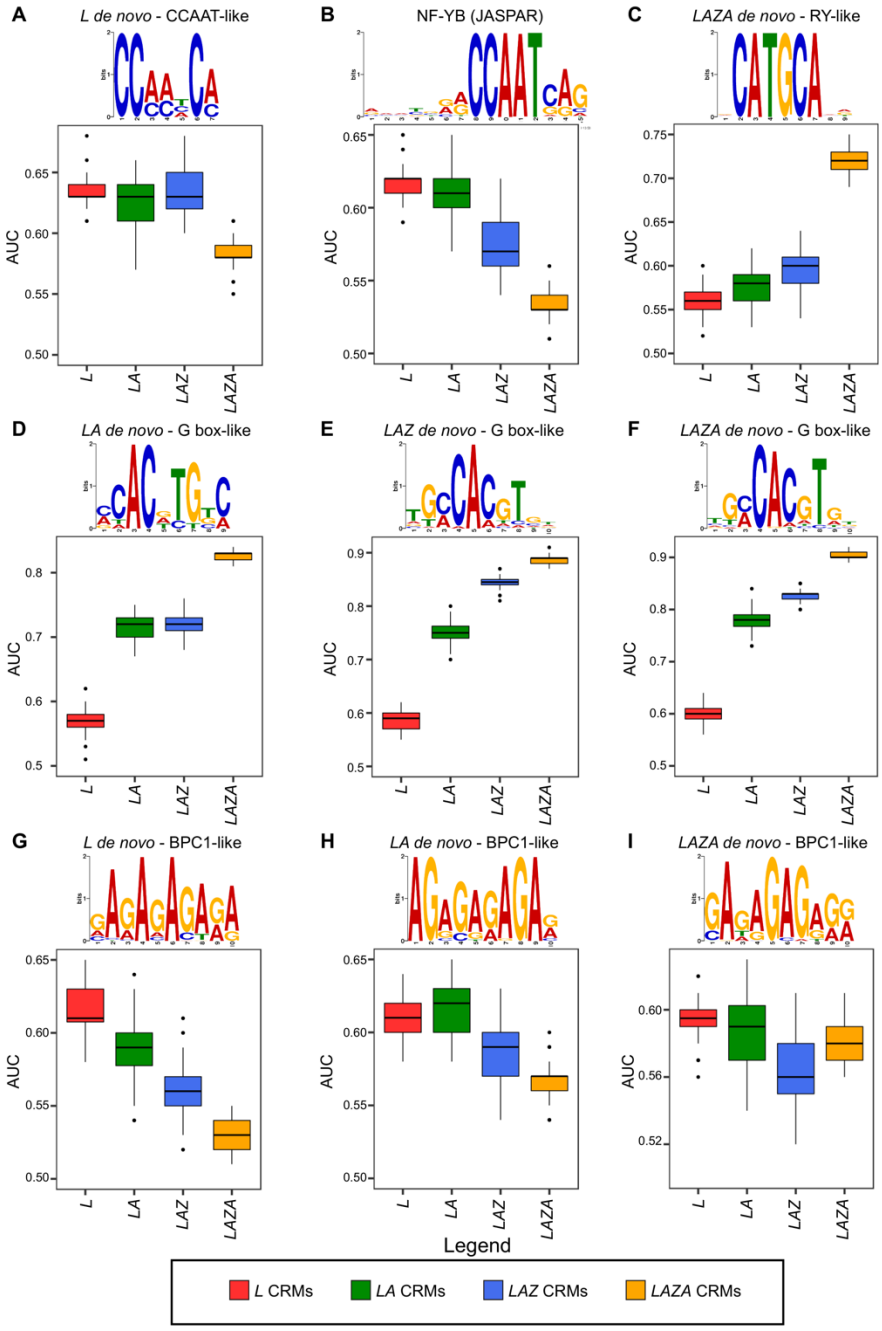


Figure S6. Enrichment of *de novo* discovered DNA motifs in CRM regions.
FIMO ROC-AUC analysis was conducted to determine the enrichment of *de novo* discovered DNA motifs in CRM sets and the JASPAR NF-YB motif (MA0502.1) were compared with regions of comparable length and position relative to randomly selected genes. Box plot graphs show the distribution of AUCs detected for each motif in *L*, *LA*, *LAZ* and *LAZA* CRM sets when compared with 1,000 random region sets. AUC scores of 1 and 0.5 indicate complete and no motif enrichment, respectively, in CRM sets relative to random DNA regions.

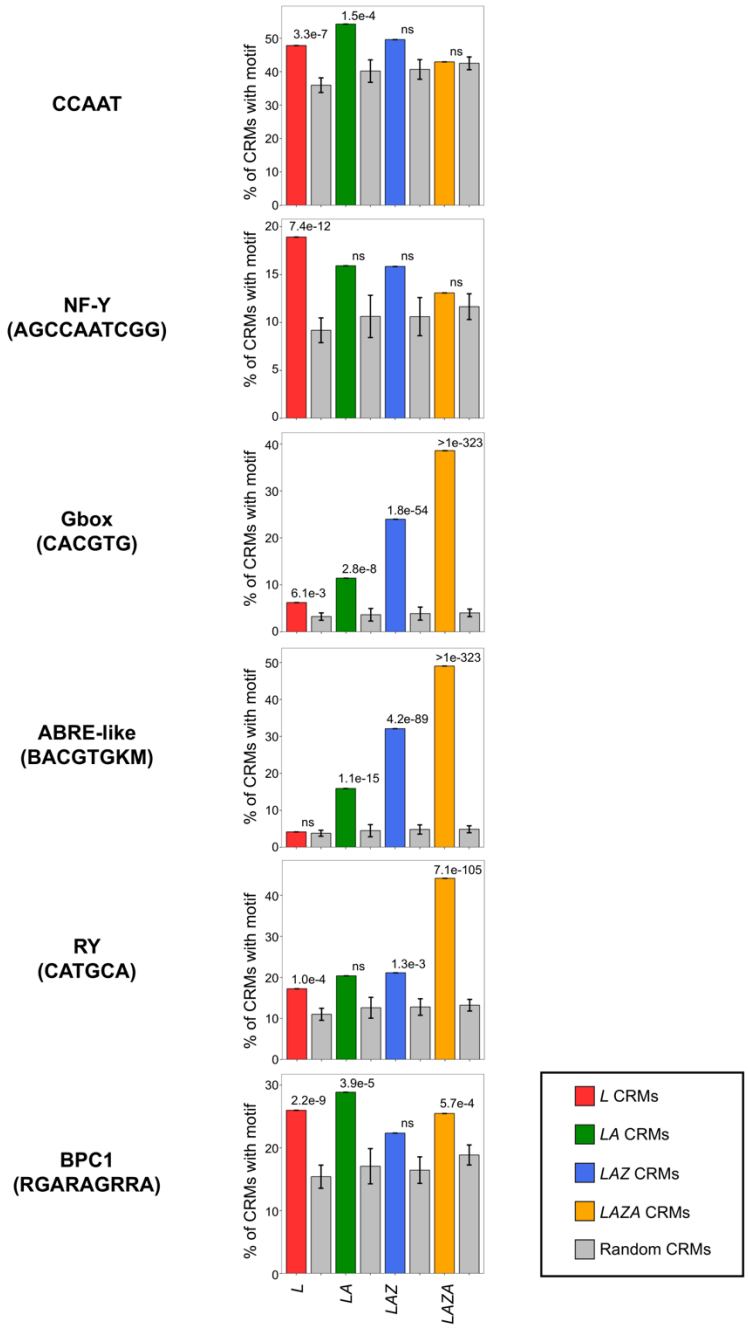


Figure S7. Enrichment of annotated DNA sequence motifs in CRM regions.
Graphs show the representation of DNA motifs in the indicated CRMs (colored bars) and in the average of 1,000 genomic regions of comparable length and position relative to randomly selected genes (gray). Adjusted P values (Bonferroni correction) are indicated.

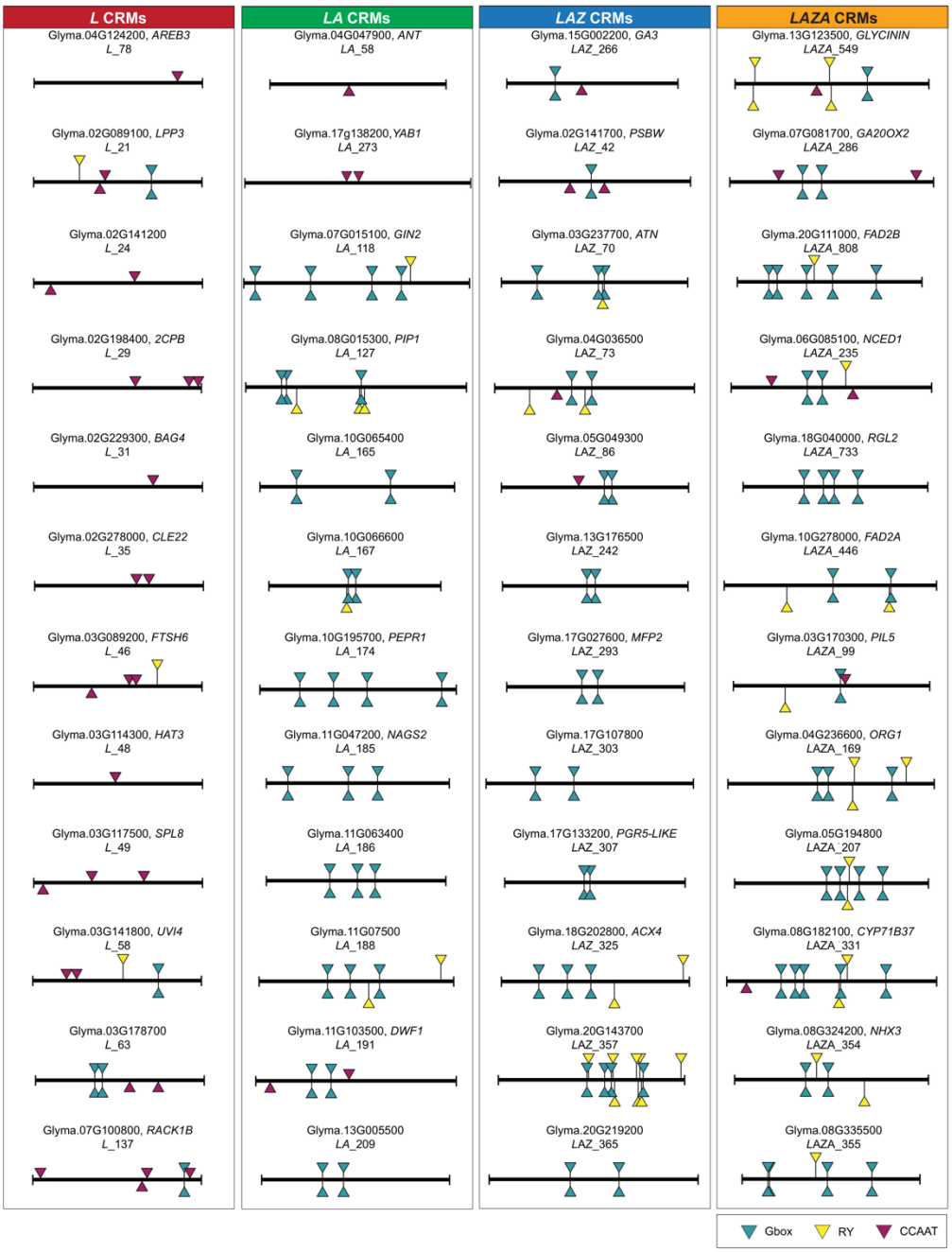


Figure S8. Schematic representation of DNA motif positions in selected *L*, *LA*, *LAZ*, and *LAZA* CRM regions.

The positions of CCAAT-like (magenta), G box-like (CACGTG, teal), and RY-like (CATGCA, yellow) motifs in the indicated CRM regions. Motifs were scanned using the FIMO tool, and only motifs with a P value of less than 0.02 and a FIMO score greater than 2.4 were identified. All CRMs are presented in the 5' to 3' orientation.

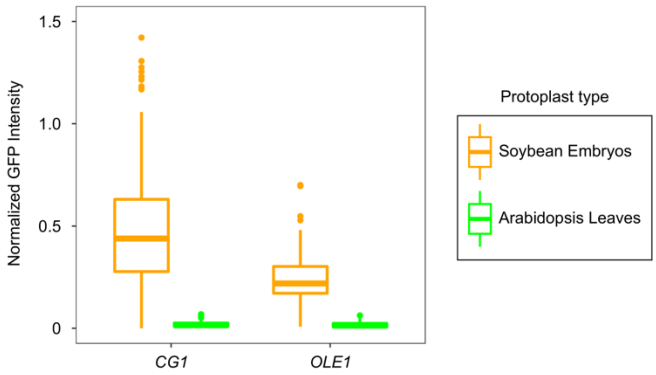


Figure S9. Developmental specificity of seed-specific gene activity in transient assays with protoplasts.
Relative activity of the *CG1* and *OLE1* chimeric genes shown in Fig. 4A in protoplasts isolated from soybean embryo cotyledons and Arabidopsis leaves. Box plots shows the ratios of GFP to mCherry fluorescence from at least 150 transfected protoplasts.

SI References

1. Pelletier JM, *et al.* (2017) LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proc Natl Acad Sci U S A* 114(32):E6710-E6719.
2. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
3. Li H, *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
4. Landt SG, *et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* 22(9):1813-1831.
5. Zhang Y, *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
6. Li Q, Brown JB, Huang H, & Bickel PJ (2011) Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics* 5(3):1752-1779.
7. Li C & Wong WH (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci USA* 98(1):31-36.
8. Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.
9. Machanick P & Bailey TL (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27(12):1696-1697.
10. O'Malley RC, *et al.* (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* 165(5):1280-1292.
11. Kulakovskiy IV, *et al.* (2017) HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* 46(D1):D252-D259.
12. Heinz S, *et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4):576-589.
13. Siggers T, *et al.* (2012) Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF-κB family DNA binding. *Nature immunology* 13(1):95.
14. Sing T, Sander O, Beerenwinkel N, & Lengauer T (2005) ROCR: visualizing classifier performance in R. *Bioinformatics* 21(20):3940-3941.
15. Fujii Y & Kodama Y (2015) In planta comparative analysis of improved green fluorescent proteins with reference to fluorescence intensity and bimolecular fluorescence complementation ability. *Plant Biotechnology* 32(1):81-87.
16. Taylor-Teeples M, *et al.* (2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 517(7536):571.
17. Zhang H, *et al.* (2018) Role of the BUB3 protein in phragmoplast microtubule reorganization during cytokinesis. *Nat Plants* 4(7):485.
18. Yoo SD, Cho YH, & Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc* 2(7):1565-1572.
19. Liu KH, *et al.* (2017) Discovery of nitrate-CPK-NLP signalling in central nutrient-growth networks. *Nature* 545(7654):311-316.
20. Dron M, Clouse SD, Dixon RA, Lawton MA, & Lamb CJ (1988) Glutathione and fungal elicitor regulation of a plant defense gene promoter in electroporated protoplasts. *Proc Natl Acad Sci U S A* 85(18):6738-6742.
21. Harkins KR, Jefferson RA, Kavanagh TA, Bevan MW, & Galbraith DW (1990) Expression of photosynthesis-related gene fusions is restricted by cell type in transgenic plants and in transfected protoplasts. *Proc Natl Acad Sci USA* 87(2):816-820.
22. Maeda K, Kimura S, Demura T, Takeda J, & Ozeki Y (2005) DcMYB1 acts as a transcriptional activator of the carrot phenylalanine ammonia-lyase gene (DcPAL1) in response to elicitor treatment, UV-B irradiation and the dilution effect. *Plant Mol Biol* 59(5):739-752.

23. Miyoshi H, Usami T, & Tanaka I (1995) High-Level of Gus Gene-Expression Driven by Pollen-Specific Promoters in Electroporated Lily Pollen Protoplasts. *Sexual Plant Reproduction* 8(4):205-209.
24. Edelstein A, Amodaj N, Hoover K, Vale R, & Stuurman N (2010) Computer control of microscopes using microManager. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 14:Unit14 20.
25. Edelstein AD*, et al.* (2014) Advanced methods of microscope control using muManager software. *J Biol Methods* 1(2).
26. Iwata Y, Lee MH, & Koizumi N (2011) Analysis of a Transcription Factor Using Transient Assay in Arabidopsis Protoplasts. *Plant Transcription Factors: Methods and Protocols* 754:107-117.

Chapter 3

Genome-wide characterization of the GmWRI1 binding profile provides novel insights into the mechanisms underlying fatty-acid biosynthesis during seed development

Jo, L., Pelletier, J.M., Melby, M., Mohammad, S., Goldberg, R.B. and Harada, J.J.

This chapter is *in preparation* for manuscript submission.

Preface

The following attributes work within this chapter to the respective author(s)

Writing	L.J., J.J.H.
Figures	L.J.
Experiments	GmWRI1 ChIP-Seq: L.J., J.M.P. ChIP-Seq library preparations: L.J., J.M.P. ChIP-Seq data analysis: L.J. DNA Motif Analysis: L.J. Cloning: L.J., M.M. Soybean Protoplasts Transfections: L.J., S.M. Arabidopsis Protoplasts Transfections: L.J. Dual Luciferase Assays: L.J. RNA-Seq library preparations: J.M.P. RNA-Seq data analysis: L.J.
Intellectual Contributions	L.J., J.M.P.,R.B.G.,J.J.H.

Abstract

Seeds are a major source of lipids. Understanding the regulatory mechanisms of lipid accumulation in seeds can provide opportunities for the development of new crops with increased oil levels. The transcription factor (TF) WRINKLED1 (WRI1) is as a central regulator of lipid biosynthesis in many plant species. To expand our knowledge into the mechanisms of WRI1 function, we characterized the binding profile of soybean WRI1 (GmWRI1) during the development of the soybean seed. Genome-wide characterization of GmWRI1 binding sites revealed that this TF can bind to several genes that encode enzymes involved with every step of the fatty acid (FA) and triacylglycerol (TAG) metabolic pathway. Additionally, we showed that the putative DNA element bound by GmWRI1 is enriched in the GmWRI1 binding sites and is partially necessary to specify GmWRI1 function. Our results also provided evidence that GmWRI1 and GmLEC1 act in a positive feedback subcircuit in the control of FA biosynthesis in soybean seeds. Interestingly, we identified the presence of a CTCCGCC-Box enriched in GmWRI1 binding sites that indicates that other TFs may collaborate with WRI1 in the induction of the FA biosynthetic network. Our results provided important and novel insights into the regulatory action of GmWRI1 to the control of lipid biosynthesis in soybean seeds.

Introduction

Seeds are a major source of plant oils in the world. In addition to being of major importance for food consumption, seed lipids have a broad application in the chemical industry and as biofuels (1). Due to its growing importance, it is expected that the demands for plant oils will double in the following decades (2). These demands can be addressed by developing crops with increased seed lipid content. In order to achieve this goal, it is crucial that we understand the regulatory mechanisms behind the accumulation of lipids in seeds. Understanding how the process of seed lipid accumulation is regulated in can provide important insights on the development of biotechnological tools to improve the quality and quantity of this important metabolite class in seeds.

Due to its importance, the metabolic pathway of fatty acid and lipid accumulation in seeds have been extensively studied in many crop species (Reviewed in (1) and (3)). Fatty acid (FA) biosynthesis initiates in plastids with the synthesis of malonyl-CoA by the enzymatic complex ACETYL-COA CARBOXYLASE (ACCase). BIOTIN CARBOXYL CARRIER PROTEIN2 (BCCP2) and ACETYL CO-ENZYME A CARBOXYLASE BIOTIN CARBOXYLASE SUBUNIT (CAC2) are important protein subunits present in ACCase complex. Malonyl-CoA is transferred to an ACYL-CARRIER PROTEIN (ACP) to form malonyl-ACP, which is the main carbon donor for all subsequent reactions of fatty acid elongation. The first 4-carbon long acyl-ACP is formed by the condensation reaction between a malonyl-ACP and an acetyl-CoA catalyzed by the *3-KETOACYL-ACYL CARRIER PROTEIN SYNTHASE* III (KASIII) enzyme. Consecutive and cyclic additions of 2 carbons from malonyl-ACP donors into the acyl-ACP is catalyzed by BETA-KETOACYL-ACP SYNTHASE I (KASI) to generate long 16 carbon long acyl-ACPs. Further elongation can be achieved by KASII to catalyze an extra condensation reaction to form

an 18-carbon long acyl-ACP. The release of FAs from ACPs is catalyzed by Acyl-ACPs FATTY ACYL THIOESTERASES (FAT) enzymes in the plastids. In addition to the enzymes that participate in the biosynthesis of FAs, several other enzymes are responsible for the elongation and desaturation of FAs, which account for the great diversity of FAs and TAGs found in the seeds. The constant activity of these enzymes in the plastids contribute to the increase in the acyl-CoA pool that is used for the *de novo* formation of membrane lipids as well of storage triacyl-glycerol (TAG). The acyl-CoA produced in the plastids, along with the glycerol-3-phosphate (G3P) are the main substrates for TAG assembly in the endoplasmic reticulum. The acylation of G3P in the endoplasmic reticulum requires the action of distinct enzymes such as GLYCEROL-3-PHOSPHATE ACYLTRANFERASE (GPAT) and DIACYLGLYCEROL ACYLTRANSFERASE (DGAT). Upon synthesis, TAGs are largely stored in discrete subcellular organelles called oil bodies (3).

As summarized above, the accumulation of lipids in seeds requires the coordinated activity of several enzymes in distinct group of organelles. Oilseed crops, such as soybean, are particularly efficient in the synthesis and storage of TAG in the seed. In some oilseed crops, lipid storage molecules can account for more than 40% of its dry mass (Taylor 2004). In soybean, lipids correspond to 20% of the seed dry mass (4). The great success of these crops to synthesize and store lipids suggests a precise regulation of FA and TAG biosynthesis during seed development. Some reports suggests that the regulation of FA and TAGs in seeds occurs mainly at the transcriptional level (5–8). Remarkably, the expression level of many genes involved with FA biosynthesis showed a similar bell-shaped pattern of expression that peaks during the onset of the maturation program and fades during the later stages of maturation, suggesting a precise

transcriptional coordination of gene expression for the FA biosynthesis during seed development (9).

Understanding the mechanisms of transcriptional regulation for genes involved with FA biosynthesis and lipid accumulation in seeds can represent an important step in the development of crops with increased oil content. Among well studied transcription factors (TFs), WRINKLED1 (WRI1), a member of the APETALA2 (AP2) family of TFs, has been characterized as a central regulator of FA biosynthesis in seeds of many plant species (10). This gene was initially identified in Arabidopsis mutant screens, where *wri1* loss-of-function mutants resulted in an 80% reduction in seed oil content (11). Since then, orthologues of WRI1 have been identified and its role in the regulation of FA biosynthesis in seeds has been shown in many oilseed crops (12,13,10). In addition, ectopic overexpression of WRI1 results in significant increases of FA related gene expression and consequently increased oil content in several studies (14,12,15,16,13). Such reports allowed WRI1 to be defined as a "master" regulator of seed lipid biosynthesis (10).

Despite its importance in FA biosynthesis, the detailed mechanisms by which WRI1 functions in in seeds is still not well understood. Characterization of WRI1 mechanisms in the regulation of FA related genes relied mainly in the analysis of loss-of-function mutants or ectopic overexpression of WRI1, where alteration of WRI1 expression almost always resulted in changes in the expression of FA related genes and consequently, altered the lipid accumulation program. Such approaches were important to show that WRI1 is involved in the active regulation of the expression of genes involved with FA biosynthesis, including the subunits of ACCase, ACP and KAS genes (17,18). Additionally, these approaches allowed for the identification of the AW-Box DNA motif (CNTNGNNNNNNNCG) to be the element responsible for WRI1 binding (19). However, because a genome wide characterization of WRI1 binding was never reported, the exact

mechanisms by which this important TF acts to regulate the expression of FA-related genes is still unclear. Such characterization is crucial if one wants to identify of all the genes that are regulated by this important TF in the control of lipid biosynthesis in seeds. In this work, we characterized the genome-wide binding profile of the soybean homologue for WRI1 (GmWRI1) during the development of the soybean seed. Such characterization contributed to reinforce many aspects of GmWRI1 function as a central regulator of lipid biosynthesis as well as providing new insights into WRI1 mechanisms to regulate FA biosynthesis and lipid accumulation in seeds.

Results

WRI1 ChIP-Seq binds to regions downstream the TSS

In order to identify potential genes that are regulated by GmWRI1 during the development of the soybean seed, we performed chromatin immunoprecipitation followed by DNA-sequencing (ChIP-Seq) using peptide antibodies against GmWRI1-A (Glyma.08G227700) and GmWRI1-B (Glyma.15G221600) in soybean embryos at cotyledon (cot) and early-maturation (em) stage that correspond to 15 and 23 days after pollination (dap), respectively. The selection of these stages relied on GmWRI1 expression, where its expression peaks at cot and em stages and it declines at later stages of embryo development (Figure 1A).

Upon DNA sequencing and data analysis, we identified genes that are potentially bound by GmWRI1 at cot and em stages of embryo development (Figure 1B). As per (20), we determined bound genes as those with a reproducible ChIP-seq peak (according to ENCODE metrics) that overlaps with the transcription start site (TSS) and 1000 bps windows upstream the TSS. A total of 1483 and 6547 genes were bound by GmWRI1 in cot and em stages, respectively. A major overlap between bound genes in cot and em stages was observed (Figure 1B). Data analyses were performed following ENCODE guidelines for ChIP-Seq (21). and unlike the ChIP-Seq data from em stage embryos, ChIP-Seq data from cot stage did not meet ENCODE minimum thresholds for normalized strand correlation (NSC) and relative strand correlation (RSC) (1.05 and 0.8, respectively) (Supplemental Figure 1), which can explain the lower number of bound genes detected in cot stage compared to the em stage (Figure 1B).

Surprisingly, we observed that the summit of ChIP-Seq peaks (the highest point of the peak) at both stages were located mostly downstream the transcription start site (TSS) (Figure 1C). Most peak summits were localized in the intragenic regions (5`UTR, exons and introns) (Figure

1D). These results are different from TF binding sites for GmLEC1, GmAREB3, GmbZIP67 and GmABI3, where ChIP-Seq peak summits were mostly located in regions upstream of the TSS (22,20).

GmWRI1 bound genes are related to fatty acid biosynthesis

Gene Ontology (GO) enrichment analysis revealed that in cot and em stages, GmWRI1 binds to several genes that are involved with FA biosynthesis and lipid accumulation in seeds (Figure 2A). Our results showed that that GmWRI1 can bind to regions of genes that encode for enzymes involved in the initiation process of FA biosynthesis in the chloroplast, such as the subunits of ACCase and KASIII (Figure 2B). We also showed that GmWRI1 can bind to genes that encode enzymes required for fatty-acid assembly into TAGs assembly in the endoplasmic reticulum, such as GPDHC and DGAT (Figure 2B). Because FA biosynthesis and TAG assembly are complex processes that involve many different enzymes in different organelles, it is remarkable that GmWRI1 can bind to genes required for all the steps required for lipid biosynthesis in soybean seeds.

Because not all TF binding events represent a transcriptional regulatory event (23), we evaluated whether GmWRI1 is responsible for the active transcription of FA related genes by evaluating the expression of GmWRI1 bound genes in soybean embryos. In support of GmWRI1 actively controlling the expression of those genes, we observed that these FA related genes that are bound by GmWRI1 are highly expressed in the soybean embryo at the EM stage (Figure 2 C and D). In fact, when compared to FA biosynthesis-related genes that are not bound to GmWRI1, GmWRI-bound genes were expressed at higher levels (Figure 2C). This pattern was not observed in a distinct developmental stage - soybean seedlings, where GmWRI1 expression is detected at a

low abundance(Figure 1A and 2D). This result suggests that that GmWRI1 is likely to be actively promoting the expression of FA biosynthesis-related genes in soybean embryos by binding to their regulatory regions and consequently, establishing the lipid accumulation program during the development of soybean seeds.

The putative DNA element bound by GmWRI1 is highly enriched in GmWRI1 bound regions

In order to investigate the potential mechanisms by which GmWRI1 controls lipid biosynthesis in soybean seeds, we analyzed the composition of *cis*-regulatory elements within the GmWRI1 binding sites (50 bps window around the ChIP-Seq peak summit). Motif *de novo* discovery analysis revealed the presence of consensus sequences in the GmWRI1 binding site in all bound genes and in the genes related to FA biosynthesis (Figure 3A and B). Interestingly, a GAGA rich DNA motif was de novo discovered in the GmWRI1 bound of both datasets (Figure 3A and B). We previously identified a GAGA rich DNA motif in the binding sites of GmLEC1, GmAREB3, GmBZIP67 and GmABI3 (22,20). The AW-BOX (CNTNGNNNNNNNCG) which was shown to be the DNA motif bound by GmWRI1 in vitro (13,24), was identified in the *de novo* discovery analysis (Figure 3A and B). These results suggests that GmWRI1 occupancy in many of those binding sites can be explained by the direct interaction with this DNA element. This result was confirmed in the motif enrichment test (Bonferroni-adjusted P values with a significance threshold of 0.001), where we found that 22.4% of the GmWRI1 binding sites for all bound genes and 40.6% of the GmWRI1 binding sites for genes involved with FA biosynthesis contained full AW-Boxes (Figure 3C). The percentage of regions with an AW-Box motif was greater when compared to background regions of equal number, length and genomic context (2.4% in the

background for all GmWRI1 bound genes and 2.7% in the background for GmWRI1 bound genes related to FA biosynthesis). Interestingly, the presence of CCAAT-Boxes (CCAAT), G-boxes (CACGTG) and RY (CATGCA) elements were not identified as enriched in the GmWRI1 bound sites for FA related genes. (Figure 3C) These elements are described as important for the control of the maturation program by LEC1, ABI3, FUS3 and LEC2 (LAFL) master regulators of seed maturation (25–27,20). This result suggests that GmWRI1 binding and transcriptional control of FA biosynthesis occur in distinct *cis*-regulatory modules to the ones identified by the LAFL group of transcription factors.

Because GmWRI1 peak position was often found in regions downstream of the TSS, we wanted to evaluate if the genomic position of AW-Boxes can explain this pattern. Interestingly, we observed that for genes that are bound by GmWRI1 as well as genes that are bound by GmWRI1 and are related to FA biosynthesis, AW-Boxes were predominately found in regions downstream of the TSS (Figure 3D). This is consistent with the observation that the GmWRI1 peak summit is located more often in regions downstream of the TSS, suggesting that GmWRI1 occupancy in these genes is likely to be directed by AW-Boxes downstream the TSS (Figure 4D). However, it is important to stress that enrichment of the AW-Box downstream the TSS could be an effect of the increased G/C content in region downstream the TSS. Nevertheless, the high enrichment of AW-Boxes in regions where ChIP-Seq Peak summits are located, reinforces that these elements are likely to play an important role in the specification of WRI1 occupancy in regions downstream of the TSS and correspondingly to control GmWRI1 regulatory function.

GmWRI1 and GmLEC1 control FA related genes through binding of distinct loci

The transcription factor LEAFY COTYLEDON1 (LEC1) is also reported to be a central regulator of lipid biosynthesis in seeds (28,27). Because we previously reported that GmLEC1 can

form transcriptional complexes with other TFs to control the expression of target genes (20), we wanted to investigate if GmWRI1 also works in concert with GmLEC1 to control FA biosynthesis in soybean seeds. We observed that 66 FA related genes were bound exclusively by GmWRI1, 266 genes were bound exclusively by GmLEC1, and 101 genes were bound by both GmWRI1 and GmLEC1 (Figure 4A). To evaluate if GmLEC1 binding position coincides with the GmWRI1 binding position for genes that are bound by both TFs, we examined the distance of the summit position of GmWRI1 in respect to the position of the GmLEC1 peak summit (Figure 4B). Interestingly, unlike GmAREB3, GmbZIP67 and GmABI3 and their relationship with GmLEC1 (20), GmWRI1 summit position does not appear to coincide with the position of GmLEC1 peak summit (Figure 4B). This result suggests that GmWRI1 and GmLEC1 bind to distinct loci in their shared bound genes. This result is corroborated by a motif enrichment analysis, where we observed that the GmWRI1 and GmLEC1 binding sites are enriched for distinct sets of DNA elements (Figure 4C). The binding sites for GmWRI1 were strongly enriched (Bonferroni-adjusted P values with a significance threshold of 0.001) for the presence of AW-Box for genes that were bound exclusively by GmWRI1, and genes bound by both GmWRI1 and GmLEC1 (Figure 4C). In the other hand, GmLEC1 were enriched for G-box type DNA elements and only slightly enriched for AW-Boxes for genes that are bound by both GmLEC1 and GmWRI1 (Figure 4C). Our results suggests that GmWRI1 and GmLEC1 bind to distinct locations within to which they coordinately bind and therefore regulate the expression of their target genes through distinct *cis*-regulation.

To evaluate how the relationship between these TFs affects their target gene expression, we looked at the expression of these F- related genes in soybean embryos (Figure 4D). We observed that genes that are bound exclusively by GmLEC1 and GmWRI1 to be expressed at higher levels when compared to genes that are not bound by either of these TFs (Figure 4D).

However, we observed that the group of genes that are bound by both TFs to be highly expressed when compared to genes bound by each TF as well for gene that are not bound by neither TF (Figure 4D). Our results suggests that although they bind independently from each other, GmWRI1 and GmLEC1 are important to regulate the expression of their FA related target genes.

Functional Analysis of AW-Boxes in GmWRI1 binding regions

Our results suggest that GmWRI1 binding and subsequent regulation of FA related gene expression is dependent on the AW-Box DNA motif. To determine whether the AW-Box is crucial to determine GmWRI1 function in controlling the expression of FA-related genes, we isolated and cloned the binding regions of GmWRI1 of FA related genes which contained a full AW-Box. The bound regions of GmWRI1 on the *ACP4* (*Glyma.15G098500*), *BCCP2* (*Glyma.13G057400*) and CAC2 (*Glyma.05G221100*) genes contained at least one full AW-Box sequence (Figure 5A, Supplemental Table 1). We also isolated the bound region on *ACP3* (*Glyma.19G240100*) which contained a partial AW-BOX (CNTNGNNNNNNCG) which was shown to be bound by WRI1 *in vitro* (29). First, we evaluated if the regions bound by GmWRI1 on those genes can function as enhancers in soybean embryo protoplasts (Figure 5B and C). Our results showed that these bound regions can function as enhancers in soybean embryo protoplasts, where GmWRI1 is expressed (Figure 5C). Interestingly, bound sites with peaks summits located downstream of the TSS (ACP4 and BCCP2) and upstream of the TSS (CAC2 and ACP3) showed enhancer-like capabilities (Figure 5C), suggesting that regardless of the binding position, these bound regions behave like *cis*-acting enhancers to promote the function of GmWRI1. In accordance with this, we found that GmWRI1-bound regions in the coding sequence of two seed storage proteins (Glyma.02G012600

(LECTIN1) and Glyma.10G028300 (PAP85)) to be sufficient for transactivation by the overexpression of GmWRI1 in Arabidopsis leaf protoplasts (Supplemental Figure 3).

In order to verify if the AW-Box contained in these bound regions is responsible for their enhancer capabilities, we generated mutant versions of these binding regions by altering the sequences of their AW-Boxes (Figure 6A, Supplemental Table 1). We tested whether the mutations in the AW-Box affected the ability of the binding sites to function as enhancers in protoplasts isolated from soybean embryo cotyledons, (Figure 6B). Additionally, we tested whether these mutations affected the ability of these bound sites to be transactivated by GmWRI1 in Arabidopsis leaf protoplasts (Figure 6C). As expected, we observed that mutations in AW-Boxes resulted in a reduction in the regulatory activity of the bound site for two genes, *ACP4* and *BCCP2* (Figure 6D). However, we observed that mutations in the full AW-Box of *CAC2* and the partial AW-Box of *ACP3* resulted in partial and no reduction of the binding site ability to work as an enhancer in soybean embryos, respectively (Figure 6D). This result suggests that for *ACP4* and *BCCP2*, GmWRI1 function is fully dependent of the AW-Box, while for *CAC2* and *ACP3*, GmWRI1 is partially or nondependent of a full AW-Box (Figure 6D). This interpretation is consistent with the Arabidopsis transactivation experiments (Figure 6E), where mutations in the AW-Box of *ACP4* and *BCCP2* completely removed their ability to be transactivated by GmWRI1, while mutations in *CAC2* and *ACP3* only partially or did not affect their ability to be transactivated by GmWRI1 (Figure 6E).

Because *ACP3* does not contain a full AW-Box but it showed the ability to be transactivated by GmWRI1, we made a series of 100 bps overlapping deletions in the GmWRI1 binding site of *ACP3* (Figure 7A) to identify potential elements, other than AW-Box, that may be responsible for GmWRI1 function. Interestingly, we observed that in soybean embryo protoplasts,

all deletions resulted in a reduction activity compared to the WT full region (Figure 7B). However, deletions in regions F and G resulted in the most dramatic effects (Figure 7B). Similarly, deletion in region F and G resulted in an inability of the binding site to be transactivated by GmWRI1 in Arabidopsis leaf protoplasts (Figure 7C). Interestingly, region F is where the GmWRI1 ChIP-Seq peak summit is located (Figure 7A and D), suggesting that this site responsible for GmWRI1 function. In order to identify other potential DNA binding elements in region F and G, we scanned these regions using Plant PAN3.0 promoter mapping program (http://plantpan.itps.ncku.edu.tw). We found two main DNA motifs in the region, a DELAY OF FLOWERING (DOF) binding motif (CTTT) and a C2H2 binding motif (CACACTT) (Supplemental Table 2). These motifs were not enriched in the WRI1 binding sites (Supplemental Figure 4), suggesting that these motifs are not responsible for GmWRI1 occupancy in bound regions.

CTCCGCC-Box is enriched in GmWRI1 bound regions on FA biosynthesis related genes

Our results showed that, in some instances, AW-Boxes are not critical to determine GmWRI1 function. In order to identify other DNA elements that are potentially involved with GmWRI1 function, we performed *de novo* discovery analysis only in WRI1 binding sites that are absent of AW-Boxes in FA related genes, with the expectation that we could find other DNA elements that can explain GmWRI1 occupancy. Motif *de novo* discovery revealed the presence of a CTCCGCC element (Figure 8A). Interestingly, this element was previously shown to be abundant in the 5'UTR region of FA related genes in many different plant species (30). This CTCCGCC-Box, as well as less stringent consensus sequences, CNCCNCC and CTCCGCC-Box

allowing 1 mismatch (CTCCGCC-Box 1mm), were highly enriched in the GmWRI1 binding sites for genes involved with FA biosynthesis (Figure 7B). These results suggest that GmWRI1's ability to regulate the expression of FA related genes is also determined by the CTCCGCC-Box. In accordance with this hypothesis, we found that this CTCCGCC-Box is present in the GmWRI1 binding sites of *ACP3* and *CAC2*, which were shown not to be fully dependent of the AW-BOX for GmWRI1 transactivation (Supplemental Table 1, Figure 6E and F). More specifically for ACP3, we identified the CTCCGCC-Box in region F, which is the region where the ChIP-Seq summit is located and the region that is responsible for GmWRI1 transactivation (Figure 7C and D). Altogether, these results suggests that GmWRI1 function is partially dependent on the CTCCGCC-Box.

Ectopic overexpression of GmWRI1 in Arabidopsis leaf protoplasts

In order to further investigate the role of GmWRI1 to control FA biosynthesis, we ectopically overexpressed GmWRI1 in Arabidopsis leaf protoplasts and analyzed the transcriptome of transfected cells (Figure 9A). Arabidopsis leaf protoplasts were transfected with a plasmid carrying GmWRI1 driven by the *CaMV35S* promoter (*p35S:GmWRI1*) and their transcriptome was compared to protoplasts transfected with a plasmid carrying the *MCHERRY* gene driven by the 35S promoter (*p35S:mCherry*) (Figure 9A). Upon data analysis, we identified 1264 differentially expressed genes (DEGs) that were upregulated (q value <= 0.001 and log2FC >= 1) and 126 DEGs that were downregulated (q value <= 0.001 and log2FC <= -1) in cells transfected by GmWRI1 when compared to the control (Figure 9B). GO enrichment analysis revealed that most genes found in the upregulated set were enriched (q value threshold of 0.05) for GO terms related to FA related processes, such as fatty acid biosynthetic process and acetyl-CoA

biosynthetic process from pyruvate (Figure 9C), supporting the central role of GmWRI1 to activate the FA biosynthesis program. Interestingly, genes in the downregulated set were identified to be enriched (q value threshold of 0.05) with genes involved with response to singlet oxygen (Figure 9D).

Our results suggest that like in soybean, GmWRI1 can actively bind to FA related genes in Arabidopsis and activate their expression. Because there is no data available for WRI1 ChIP-Seq in Arabidopsis, we asked if the soybean homologues for the Arabidopsis DEGs were bound by GmWRI1 in our ChIP-Seq data. We found that more than 35.9% and 25.4% of the upregulated and downregulated DEG, respectively, were found to have at least one soybean homolog that is bound by GmWRI1 in our ChIP-Seq data (Figure 9D).

In soybean, our results suggests that GmWRI1 is actively binding to regions downstream of the TSS due the position of AW-Boxes and CTCCGCC-Boxes. In order to investigate if the same pattern is found in Arabidopsis, we performed DNA motif enrichment analysis in downstream and upstream regions of the TSS for upregulated and downregulated DEGs (Figure 9E). Surprisingly, we didn't identify the enrichment of AW-Boxes and CTCCGCC-Boxes in the downstream and upstream regions of downregulated DEGs (Figure 9F). However, we observed the enrichment of AW-BOX and CTCCGCC-Boxes in upregulated DEGs (Figure 9F). We found that regions downstream the TSS were more predominantly enriched for those elements (Figure 9F). Around 65% and 20% of the regions downstream the TSS of the upregulated DEGs showed the presence of an AW-Box and CTCCGCC-Box, respectively (Figure 9F). This percentage was greater when compared to regions downstream the TSS of random sets of genes (40% and 9%, respectively). For regions upstream from the TSS for upregulated DEGs, we found only AW-Box to be enriched, with 28% of those regions showing the presence of AW-Boxes against 18% in the

background (Figure 8F). Our results suggest the role of GmWRI1 to act as a central regulator of lipid biosynthesis and that WRI1 pattern of binding downstream the TSS of its target genes to be shared between Arabidopsis and Soybean.

Discussion

GmWRI1 is a central regulator of lipid biosynthesis in seeds

Lipid accumulation in seeds requires the activity of several enzymes across many different organelles (1,3). In this sense, it is likely that the expression of all genes required for this complex process to be highly coordinated to ensure the rapid allocation of carbon into FAs and TAGs during embryogenesis. Here, we show that GmWRI1 is likely to be responsible for the coordination of expression of many FA biosynthesis related. Genome-wide characterization of GmWRI1 binding during the development of the soybean embryo revealed that this TF can bind to genes that encode for key enzymes required for FA biosynthesis and TAG assembly (Figure 2A and B). These genes encode for enzymes necessary for all the steps for FA biosynthesis (Figure 2 B and C), from the synthesis of Malonyl-CoA in the chloroplast to the assembly of triacyl-glycerol (TAG) in the endoplasmic reticulum (Figure 2B and C). However, because not every TF binding event equates to functional regulation, one can argue whether GmWRI1 binding in genes involved with FA biosynthesis is sufficient to cause transcriptional activation. The following evidence suggest that GmWRI1 acts to positively regulate the expression of FA related genes. First, we observed that FA related genes bound by GmWRI1 are highly expressed in the embryo, where GmWRI1 expression is higher, and their expression is reduced in later stages. Second, it was previously shown that overexpression of GmWRI1 in soybean roots and soybean seeds results in the upregulation of several genes involved with FA biosynthesis, which we showed are bound by GmWRI1 in soybean embryos (13,24) (Supplemental Figure 2). We observed that 15 out 17 genes reported as upregulated by ectopic expression of GmWRI1 in soybean roots (24) to be found in our GmWRI1 bound genes list (Supplemental Figure 2). In addition, we observed that 9 out of 18 upregulated genes in transgenic soybean plants carrying a chimeric GmWRI1 gene driven by the

Brassica napus NAPIN seed-specific promoter (*pNAPIN:GmWRI1*) (13) were identified as a bound gene by GmWRI1 (Supplemental Figure 2). Third, we showed that upon ectopic overexpression, GmWRI1 was able to upregulate several FA related genes in Arabidopsis (Figure 9B and C). Fourth, we showed that GmWRI1 can transactivate the expression of a reporter gene that is controlled by a minimal *35S* promoter fused with a GmWRI1 bound site (Figure 6E). Fifth, we showed that mutations in the canonical DNA element that is bound by GmWRI1 (AW-Box) in ACP4, BCCP2 and CAC2 bound regions results in significantly decreased the capabilities of these binding sites to work as functional enhancers (Figure 6D). Altogether, these results provide supporting evidence to suggest that GmWRI1 act as a positive regulator of FA accumulation by binding to FA biosynthesis related genes and positively acting on the transcription of binding genes, reinforcing the notion that GmWRI1 act as a central regulator of lipid accumulation in seeds.

GmWRI1 and GmLEC1 act in a positive feedback gene circuit to control lipid biosynthesis in seeds

In addition to controlling the expression of several structural genes involved with FA biosynthesis in seeds, our results also suggests that GmWRI1 acts to modify the regulatory state of the soybean embryo to promote lipid accumulation in soybean seeds. We showed that GmWRI1 binds to the promoter of several TFs, including GmLEC1, a well characterized regulator of lipid biosynthesis in seeds (28,27). We also observed that overexpression of GmWRI1 in Arabidopsis leaf protoplasts resulted in the up regulation *of AtLEC1 (AT1G21970)* (data not shown). Interestingly, it was previously showed that GmWRI1 is a target gene for GmLEC1 (22,20). Moreover, it was also shown that overexpression of AtLEC1 resulted in elevated expression of

AtWRI1 (28). These results suggests that WRI1 and LEC1 act in a positive feedback subcircuit in the control of FA biosynthesis in seeds by regulating the expression of each other as well as for the expression of several structural genes involved with FA biosynthesis and TAG assembly (Figure 10). In gene regulatory networks, positive feedback gene circuit are described to be important to trigger irreversible changes in transcriptional programs (31). The onset of the of the maturation program of the seed is hallmarked by the *de novo* initiation of many biological programs, such as the storage of lipids and proteins that were mostly inactive during the morphogenesis phase of the embryo (27). Initiation of the FA biosynthetic program in the embryo would require a robust and stable regulatory state to ensure proper expression of all the genes required for this complex metabolic program. In this sense, a positive feedback subcircuit between GmLEC1 and GmWRI1 is likely to involved for the efficient lipid accumulation program during the development of the soybean seed.

An important characteristic of LEC1 is that its ability to bind and regulate gene expression is modulated by the interactions with a distinct group of TFs (27). One could speculate that the involvement of LEC1 and WRI1 to regulate FA biosynthesis would occur in the same way. However, the binding profile of GmWRI1 strongly suggests that GmWRI1 and GmLEC1 are binding to distinct loci in FA related genes (Figure 4B). We showed that GmWRI1 occupancy in FA related genes is likely to be determined by AW-Boxes or CTCCGCC-Boxes, while occupancy of LEC1 in FA related genes is likely to be determined by an interaction with a bZIP type TF (Figure 4C, Figure 10). Interestingly, we also observed that FA related genes bound by both TFs to be expressed at higher levels when compared to genes bound by each individual TF (Figure 2C). This result suggests that GmWRI1 and GmLEC1 co-binding can synergistically affect the

transcriptional activation of their target genes. These results provide new insights into the relationship between these two important regulators of lipid biosynthesis in seeds.

Biological relevance of GmWRI1 preference for intragenic binding regions

Genome-wide characterization of GmWRI1 binding revealed that this TF binds to intragenic regions (Figure 1C and 3D). This unusual binding profile is likely to be caused by the enrichment of AW-Boxes and CTCCGCC-Boxes in regions downstream the TSS (Figure 3D). A similar binding pattern is observed in other plant TFs. A large-scale analysis of ChIP-Seq and ChIP-CHIP data of 27 TFs in Arabidopsis revealed a group of TFs with a substantial binding preference to intragenic regions (32). Even though this type of binding patter has been observed for other group of TFs, the modes by which intragenic binding transcriptionally regulates genes are still not clear. Here, we showed that when inserted upstream a minimal 35S promoter, intragenic GmWRI1 binding sites can be transactivated by GmWRI1 (Figure 5C, Figure 6E and Supplemental Figure 4), suggesting that these intragenic regions behave as *cis*-acting enhancers. A recent study in cucumber revealed that this intragenic binding events can also be important for chromatin remodeling events (33). Similar to GmWRI1, the *TENTRIL IDENTITY* gene (*TEN*), which belongs to the CYC/TB1 clade of the TCP gene family in cucumber, binds mostly on intragenic regions of its target genes (33). It was reported that intragenic binding of *TEN* to be important for transcription activation and chromatin disassembly during transcriptional activation (33). An intrinsically disordered region domain (IDR) in the N-terminal region of TEN was shown to have a histone acetyl transferase capability to facilitate chromatin accessibility during transcription of its target genes (33). Interestingly, WRI1 also has an IDR domain in its N-terminal region (34). The similarities in their binding pattern as well as the presence of the IDR region

might imply that the mechanisms of GmWRI1 to be like the ones described by TEN. However, further experiments are needed to show this association. Nevertheless, our results provided novel and exciting insights into the mechanisms of GmWRI1 function in the control of FA biosynthesis in soybean seeds.

GmWRI1 function requires the assistance of other TFs

Our results also suggests that GmWRI1 function to be partially dependent of other TFs. The following results support this hypothesis: First, we showed that even though the AW-Box is critical for GmWRI1 to bind to target sites in and regulate *ACP4* and *BCCP2* (Figure 6D and E), its presence was not critical for *CAC2* and *ACP3* binding sites to function as enhancers and to be transactivated by GmWRI1 (Figure 6D and E). Second, we showed that only ~40% of GmWRI1 binding sites for genes that are related to FA biosynthesis contain full AW-Boxes (Figure 8A). Third, we also showed the enrichment of a CTCCGCC-Box in the GmWRI1 binding sites (Figure 8A and B). Fourth, we found a CTCCGCC-Box in the GmWRI1 bound region on the ACP3 gene that is important for GmWRI1 transactivation (Figure 7E). Lastly, we showed that this CTCCGCC-Box is also enriched in regions downstream of the TSS for genes that are upregulated by ectopic overexpression of GmWRI1 in Arabidopsis leaf protoplasts (Figure 9F).

The enrichment of the CTCCGCC-Box in GmWRI1 binding sites suggests that GmWRI1 occupancy in this region is mediated through another TF that can bind to this element. To explore this model, it would be necessary to identify the TF that binds to this DNA-element. Unfortunately, this DNA element is not present in the Arabidopsis cistrome database (DAP-Seq) (35). However, a recent report identified a similar DNA element (CNCCNCC) in the bound regions of TEN (TCP-like TF in Cucumber) which is like the CTCCGCC-Box identified in GmWRI1 binding regions

(33). As mentioned previously, it was observed that the binding profile of TEN is like the pattern displayed by GmWRI1, in which, the binding positions was often located in intragenic positions (33). These observations would suggest that TCP-like TF could assist GmWRI1 and specify GmWRI1 function that is not totally dependent of an AW-Box. Recently, it was recently shown that AtWRI1 can physically bind to AtTCP4, AtTCP10 and AtTCP24 (36). The interaction between AtWRI1 and AtTCP4 was shown to be important to modulate the ability of AtWRI1 to activate the expression of FA related genes (36). One can propose a model where a TCP-like TF physically binds to GmWRI1 and regulates GmWRI1 occupancy and function in an AW-Box independent manner (Figure 10). In support of this model, we observed that in the overexpression of GmWRI1 in Arabidopsis protoplasts resulted in upregulation of AtTCP3 (Supplemental Table 3). Interestingly, we observed a similar pattern of expression between GmWRI1 and GmTCP3 paralogs during the development of the soybean (Figure 1A, Supplemental Figure 4). This result points to GmTCP3 as the TF that assists GmWRI1 to control gene expression. However, we didn't identify the enrichment of the canonical Arabidopsis TCP DNA binding element (GGGACCAC) (35) in the GmWRI1 binding sites (data not shown). In this sense, it would be necessary to evaluate if GmTCP3, or other GmTCP TFs, can physically bind to the CTCCGCC-Box identified in the GmWRI1-bound regions. Second, we need to know whether mutations in these CTCCGCC-Boxes will result in the loss of GmWRI1 to transactivate its binding sites. Moreover, it is important to evaluate the possibility that GmWRI1 can physically bind to the CTCCGCC-Box. Further experiments are necessary to complete the model by which GmWRI1 functions to control the expression of FA related genes (Figure 10).

Materials and Methods

Chromatin Immunoprecipitation and DNA Sequencing

Soybean plants were grown, and seeds were harvested for ChIP experiments as described by (22). ChIP assays were performed as described previously (22,20) using peptide antibodies against GmWRI1. DNA sequencing libraries were prepared using the NuGEN Ovation Ultralow System V2, and DNA fragments were size selected by electrophoresis and sequenced at 50-bp single-end reads using an Illumina HiSEq. 4000 sequencing system.

Recombinant DNA manipulation and plasmid construction.

p35S:GmWRI1 construct

The 35S:mCherry:NOSt plasmid was used for this construct (20). The cDNA for GmWRI1-A (Glyma.08G227700) was PCR amplified from cDNA samples of soybean embryos at cot stage. PCR amplified cDNA for GmWRI1-A was inserted into the 35S:mCherry:NOSt digested with SalI and NotI (to remove the mCherry cDNA fragment) using the In- Fusion® HD Cloning Kit, according to the manufacturer's instructions.

GmWRI1 binding sites constructs

The dual luciferase plasmid (pDLUC15) was used for these constructs (20). GmWRI1 binding sites were PCR amplified from genomic DNA they were inserted into the pDLUC15 vector upstream of the 35S MINIMAL PROMOTER using the In- Fusion® HD Cloning Kit, according to the manufacturer's instructions. Mutant derivatives of AW-Boxes were synthesized by Twist Bioscience and inserted into pDLUC15. The DNA sequences of WT and mutant versions of each binding site are listed in Supplemental Table 1.

Deletion series of ACP3 constructs

Deletion derivatives forms of GmWRI1 binding site for the ACP3 gene were synthesized by Twist Bioscience and inserted into the pDLUC15 vector upstream of the 35S MINIMAL PROMOTER using the In- Fusion® HD Cloning Kit, according to the manufacturer's instructions.

Transient Assays in Soybean embryo cotyledon protoplasts and Arabidopsis leaf protoplasts

Transient assays in protoplasts isolated from soybean embryo cotyledons and Arabidopsis leaves were performed as described by (37) and (20). Measurements of firefly and renilla luciferase activities in the loss-of-function analysis in soybean embryo protoplasts and the transactivation assays in Arabidopsis leaf protoplasts were made using the Dual-Luciferase Reporter Assay System (Promega) with a TriStar2 LB942 multiplate reader (Berthold) as described previously (20).

RNA isolation and library making for transcriptome analysis

Transfected cells were lysed, and the total RNA was isolated using the RNAqueous™-Micro Total RNA Isolation Kit, following manufacturer's instructions. RNA samples were then treated with DNAse for digestion of DNA using the TURBO DNA-free™ Kit, following manufacturer's instructions. RNA-Seq libraries were prepared using Tecan Ovation RNA-Seq System V2 using 50 ng of total RNA, and double-stranded cDNA was fragmented to ~200bp using a Covaris E220. End repair, A-tailing, adapter ligation and PCR enrichment were performed according to Methods 1 library preparation protocol (41), using the NEXTflex ChIP-Seq Barcodes (BiooScientific). Libraries were quantified using Quant-iT PicoGreen dsDNA Reagent (Grand Island, NY) and a Nanodrop ND-3300 instrument, and sequenced on a NovaSeq 6000 sequencer.

Data analysis

ChIP-Seq

ChIP-Seq data were analyzed as described previously (20,22) with the exception that reproducible ChIP-Seq peaks were identified using MACS3 (https://github.com/macs3-project/MACS).

RNA-Seq

Sequenced reads were demultiplexed and reads corresponding to rRNA sequences were removed. The resulting filtered reads were mapped to Arabidopsis primary transcripts (TAIR10) using bowtie v0.12.7 with parameters -v 2 -5 10 -3 10 -m 1 --best –strata. We used the EdgeR package (v3.10.5) to obtain normalized expression values using the Trimmed Mean of M-values (TMM) method and to identify differentially expressed genes between the different genotypes (FDR < 0.001) (38).

GO enrichment

GO enrichment were performed using the Bioconductor package GOSeq as described previously (22), with the soybase GO functional annotation, the hypergeometric method, and a q value threshold of 0.05.

Motif Analysis

DNA motif *de novo* discovery analysis was performed using the MEME-ChIP tool from the MEME suite v5.3.3 (39) and DNA motif enrichment analysis was performed using the motifEnrich tool from HOMER (40) as described previously (20).

References

1. Bates PD, Stymne S, Ohlrogge J. Biochemical pathways in seed oil synthesis. Current opinion in plant biology. 2013;16(3):358–64.

2. Bruinsma J. World agriculture: towards 2015/2030: an FAO study. Routledge; 2017.

3. Baud S. Seeds as oil factories. Plant reproduction. 2018;31(3):213–35.

4. Clemente TE, Cahoon EB. Soybean oil: genetic approaches for modification of functionality and total content. Plant Physiol. 2009/09/25 ed. 2009 Nov;151(3):1030–40.

5. O'Hara P, Slabas AR, Fawcett T. Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. Plant Physiology. 2002;129(1):310–20.

6. Baud S, Graham IA. A spatiotemporal analysis of enzymatic activities associated with carbon metabolism in wild-type and mutant embryos of Arabidopsis using in situ histochemistry. The Plant Journal. 2006;46(1):155–69.

7. Baud S, Lepiniec L. Physiological and developmental regulation of seed oil production. Progress in lipid research. 2010;49(3):235–49.

8. Troncoso-Ponce MA, Kilaru A, Cao X, Durrett TP, Fan J, Jensen JK, et al. Comparative deep transcriptional profiling of four developing oilseeds. The Plant Journal. 2011;68(6):1014–27.

9. Baud S, Lepiniec L. Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. Plant Physiology and Biochemistry. 2009;47(6):448–55.

10. Kong Q, Yuan L, Ma W. WRINKLED1, a "Master Regulator" in transcriptional control of plant oil biosynthesis. Plants. 2019;8(7):238.

11. Focks N, Benning C. wrinkled1: a novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant physiology. 1998;118(1):91–101.

12. Liu J, Hua W, Zhan G, Wei F, Wang X, Liu G, et al. Increasing seed mass and oil content in transgenic Arabidopsis by the overexpression of wri1-like gene from Brassica napus. Plant Physiology and Biochemistry. 2010;48(1):9–15.

13. Chen L, Zheng Y, Dong Z, Meng F, Sun X, Fan X, et al. Soybean (Glycine max) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation. Molecular genetics and genomics. 2018;293(2):401–15.

14. Cernac A, Benning C. WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. The Plant Journal. 2004;40(4):575–85.

15. Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, et al. Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. Plant physiology. 2010;153(3):980–7.

16. An D, Suh MC. Overexpression of Arabidopsis WRI1 enhanced seed mass and storage oil content in Camelina sativa. Plant Biotechnology Reports. 2015;9(3):137–48.

17. Baud S, Wuillème S, To A, Rochat C, Lepiniec L. Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in Arabidopsis. The Plant Journal. 2009;60(6):933–47.

18. To A, Joubès J, Barthole G, Lécureuil A, Scagnelli A, Jasinski S, et al. WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in Arabidopsis. The Plant Cell. 2012;24(12):5007–23.

19. Maeo K, Tokuda T, Ayame A, Mitsui N, Kawai T, Tsukagoshi H, et al. An AP2-type transcription factor, WRINKLED1, of Arabidopsis thaliana binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. The Plant Journal. 2009;60(3):476–87.

20. Jo L, Pelletier JM, Hsu S-W, Baden R, Goldberg RB, Harada JJ. Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. Proceedings of the National Academy of Sciences. 2020;117(2):1223–32.

21. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome research. 2012;22(9):1813–31.

22. Pelletier JM, Kwong RW, Park S, Le BH, Baden R, Cagliari A, et al. LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. Proceedings of the National Academy of Sciences. 2017;114(32):E6710–9.

23. Farnham PJ. Insights from genomic profiling of transcription factors. Nature Reviews Genetics. 2009;10(9):605–16.

24. Chen B, Zhang G, Li P, Yang J, Guo L, Benning C, et al. Multiple GmWRI1s are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (Glycine max). Plant biotechnology journal. 2020;18(1):155–71.

25. Fatihi A, Boulard C, Bouyer D, Baud S, Dubreucq B, Lepiniec L. Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds. Plant Science. 2016;250:198–204.

26. Lepiniec L, Devic M, Roscoe TJ, Bouyer D, Zhou D-X, Boulard C, et al. Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. Plant reproduction. 2018;31(3):291–307.

27. Jo L, Pelletier JM, Harada JJ. Central role of the LEAFY COTYLEDON1 transcription factor in seed development. Journal of integrative plant biology. 2019;61(5):564–80.

28. Mu J, Tan H, Zheng Q, Fu F, Liang Y, Zhang J, et al. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant physiology. 2008;148(2):1042–54.

29. Kong Q, Ma W, Yang H, Ma G, Mantyla JJ, Benning C. The Arabidopsis WRINKLED1 transcription factor affects auxin homeostasis in roots. Journal of experimental botany. 2017;68(16):4627–34.

30. Bonaventure G, Ohlrogge JB. Differential regulation of mRNA levels of acyl carrier protein isoforms in Arabidopsis. Plant physiology. 2002;128(1):223–35.

31. Xiong W, Ferrell JE. A positive-feedback-based bistable 'memory module'that governs a cell fate decision. Nature. 2003;426(6965):460–5.

32. Heyndrickx KS, de Velde JV, Wang C, Weigel D, Vandepoele K. A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. The Plant Cell. 2014;26(10):3894–910.

33. Yang X, Yan J, Zhang Z, Lin T, Xin T, Wang B, et al. Regulation of plant architecture by a new histone acetyltransferase targeting gene bodies. Nature Plants. 2020;6(7):809–22.

34. Ma W, Kong Q, Grix M, Mantyla JJ, Yang Y, Benning C, et al. Deletion of a C–terminal intrinsically disordered region of WRINKLED 1 affects its stability and enhances oil accumulation in Arabidopsis. The Plant Journal. 2015;83(5):864–74.

35. O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and epicistrome features shape the regulatory DNA landscape. Cell. 2016;165(5):1280–92.

36. Kong Q, Singh SK, Mantyla JJ, Pattanaik S, Guo L, Yuan L, et al. TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR4 interacts with WRINKLED1 to mediate seed oil biosynthesis. Plant physiology. 2020;184(2):658–65.

37. Yoo S-D, Cho Y-H, Sheen J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nature protocols. 2007;2(7):1565–72.

38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.

39. Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics. 2011;27(12):1696–7.

40. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010;38(4):576–89.

41. Kumar R, Ichihashi Y, Kimura S, Chitwood DH, Headland LR, Peng J, Maloof JN and Sinha NR (2012) A high-throughput method for Illumina RNA-Seq library preparation. Front. Plant Sci. 3:202. doi: 10.3389/fpls.2012.00202)

Figures and Table

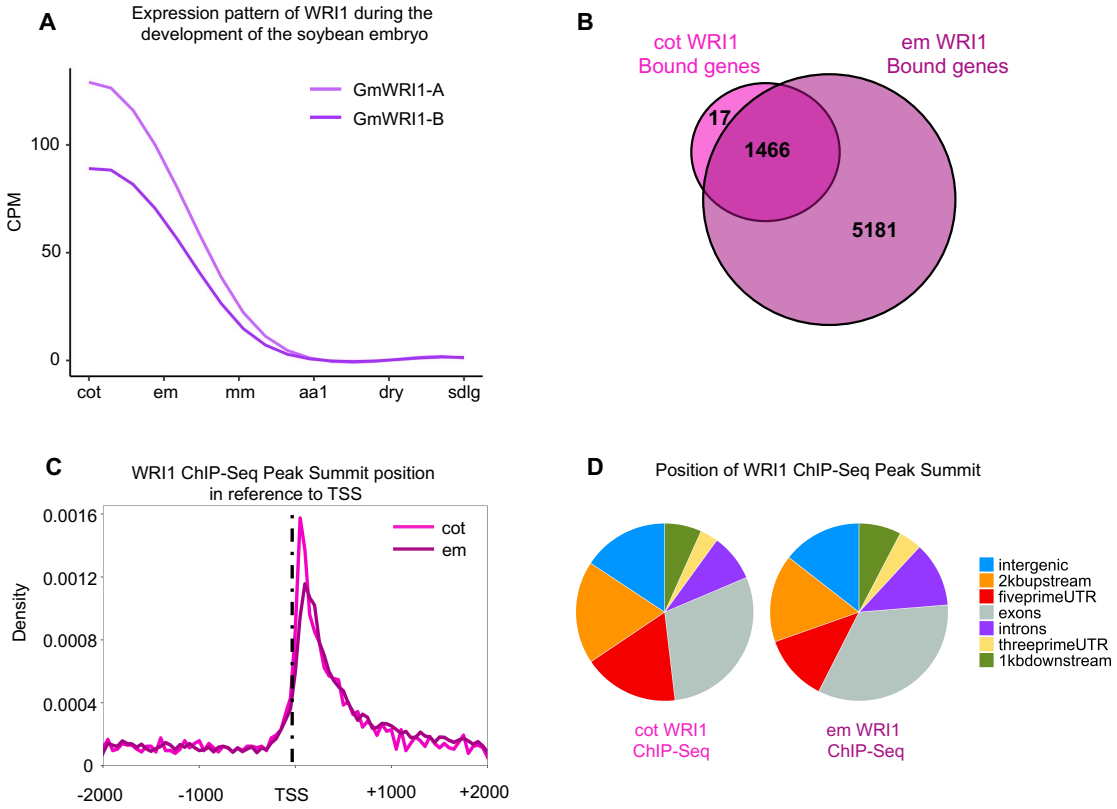


Figure 1. Identification of GmWRI1 bound genes in soybean embryos at cotyledon (cot) and early maturation (em stages). (A) Expression pattern of GmWRI1 during the development of the soybean embryo at the cot, em, mid-maturation (mm), double aa1 (aa1), dry seed stage (dry) and seedlings (sdlg). Transcriptome data was obtained from the Harada Embryo mRNA-Seq Dataset (GEO accession no. GSE99571). (B) Genes bound by GmWRI1 in soybean embryos at the cot and em developmental stages. Venn diagrams show the overlap between bound genes at cot and em stages. (C) Density plots of the positions of the GmWRI1 ChIP peak summits in reference to the transcriptional start site (TSS) of bound genes at cot and em stages. (D) Association of GmWRI1 ChIP-Seq peak summits with genomic features.

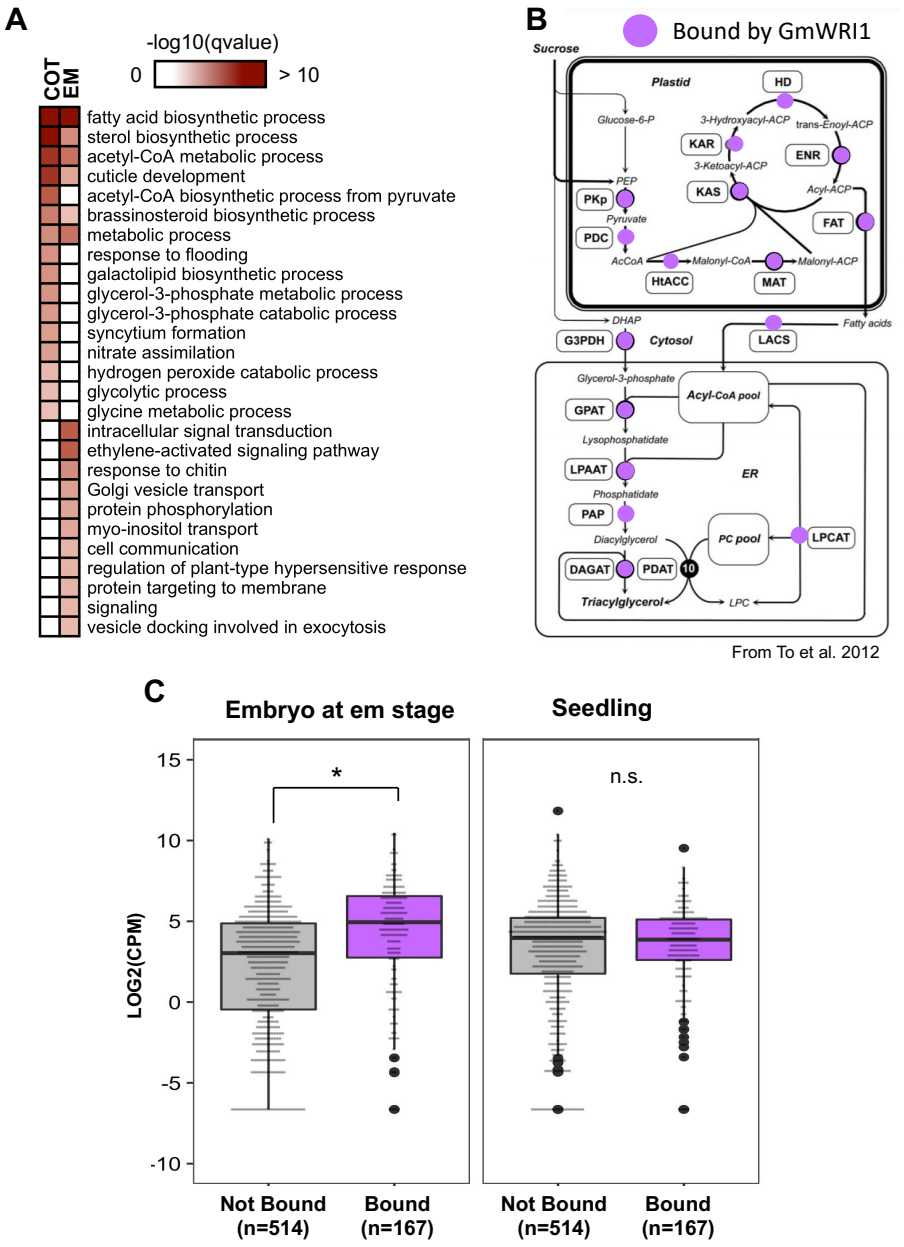


Figure 2. Analysis of GmWRI1 bound genes revealed that GmWRI1 bind to several genes involved with FA biosynthesis in soybean embryos. (A) Heatmap showing the q value significance of GO terms for GmWRI1 bound genes at cot and em stages. The GO terms listed are all the enriched biological process GO terms for each one of the developmental stages. (B) Simplified scheme of FA biosynthesis and TAG assembly from (18). Steps that are catalyzed by enzymes encoded by genes that are bound by GmWRI1 are highlighted with purple circles. (C) Box plots showing the expression levels of FA related genes that are bound by GmWRI1

(purple) and not bound by GmWRI1 (grey) in soybean embryos at em stage and in the seedling. Transcriptome data was obtained from the Harada Embryo mRNA-Seq Dataset (GEO accession no. GSE99571). Asterisk is shown to indicate significant differences between the expression of bound and not bound genes (Wilcoxon rank-sum test, P value < 0.001), with n.s. denoting non-significant differences between groups.

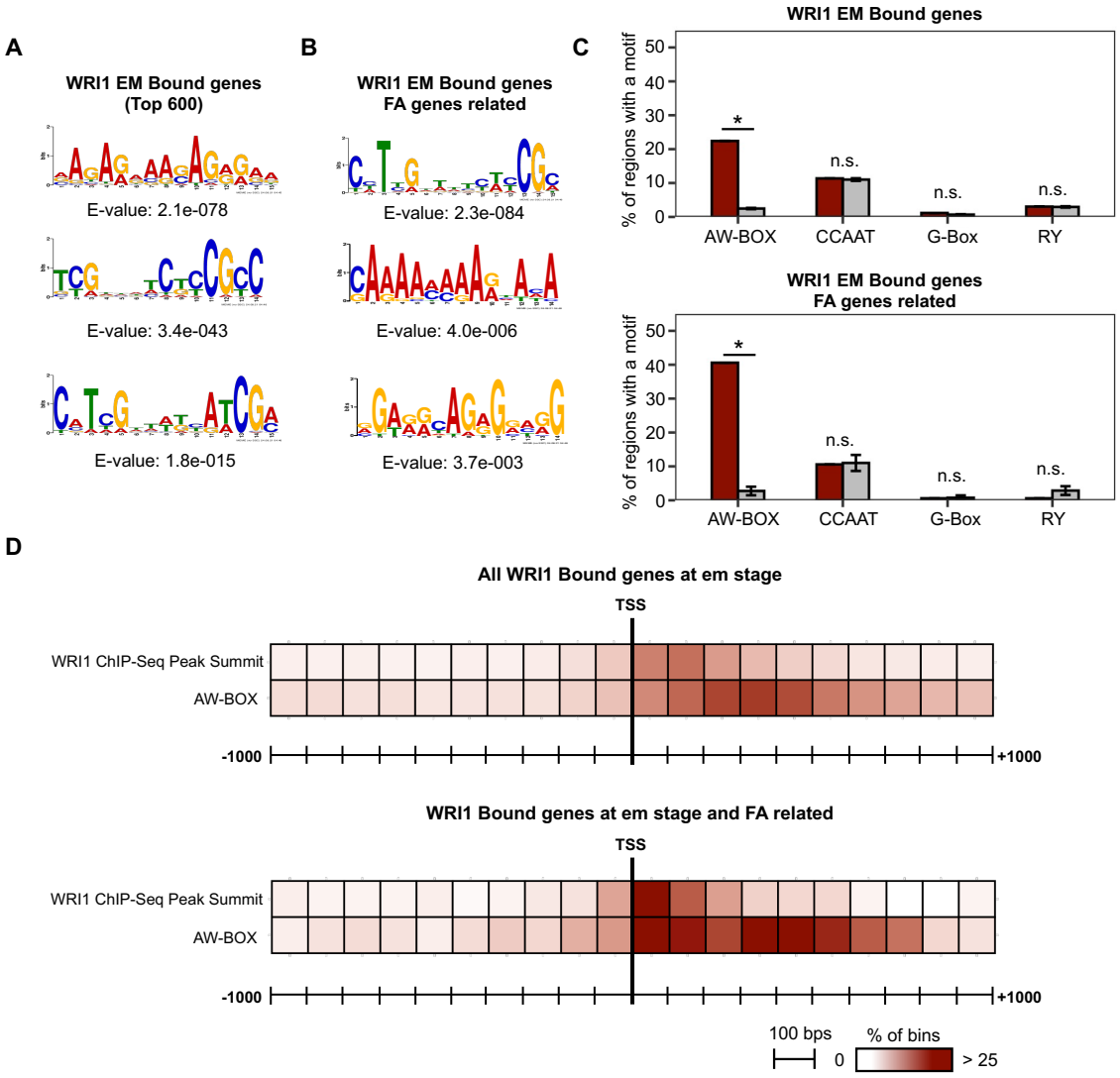


Figure 3. Motif analysis revealed enrichment of AW-Box elements (CNTNGNNNNNNCG) in GmWRI1 binding sites. (A) Position weight matrices of DNA sequence motifs discovered *de novo* in GmWRI1 bound regions (50 bps around the peak summit) for all genes bound by GmWRI1 (top 600 for ChIP-Seq signal) and (B) genes bound by GmWRI1 that are associated with FA biosynthesis. (C) DNA motif enrichment of in GmWRI1 binding sites regions. Graphs show the percentage of GmWRI1 binding sites with a DNA motif occurrence (red bars) and in the average of 1,000 population of genomic regions of comparable number, length and position relative to randomly selected genes (grey). Asterisks are to denote significant differences between groups (Bonferroni-adjusted P values with a significance threshold of 0.001), with n.s. denoting no significant difference. Motifs: AW-BOX (CNTNGNNNNNNCG), CCAAT, G-Box (CACGTG) and RY (CATGCA). (D) Heatmap representation of 2000 base pairs regions around the TSS of all GmWRI1 bound genes and GmWRI1 bound genes that are FA biosynthesis related. The 2000 bp region was divided into 100 base pair bins and the heatmap color intensity is used to depict the percentage of bins with a ChIP-Seq Peak summit or an AW-Box motif occurrence.

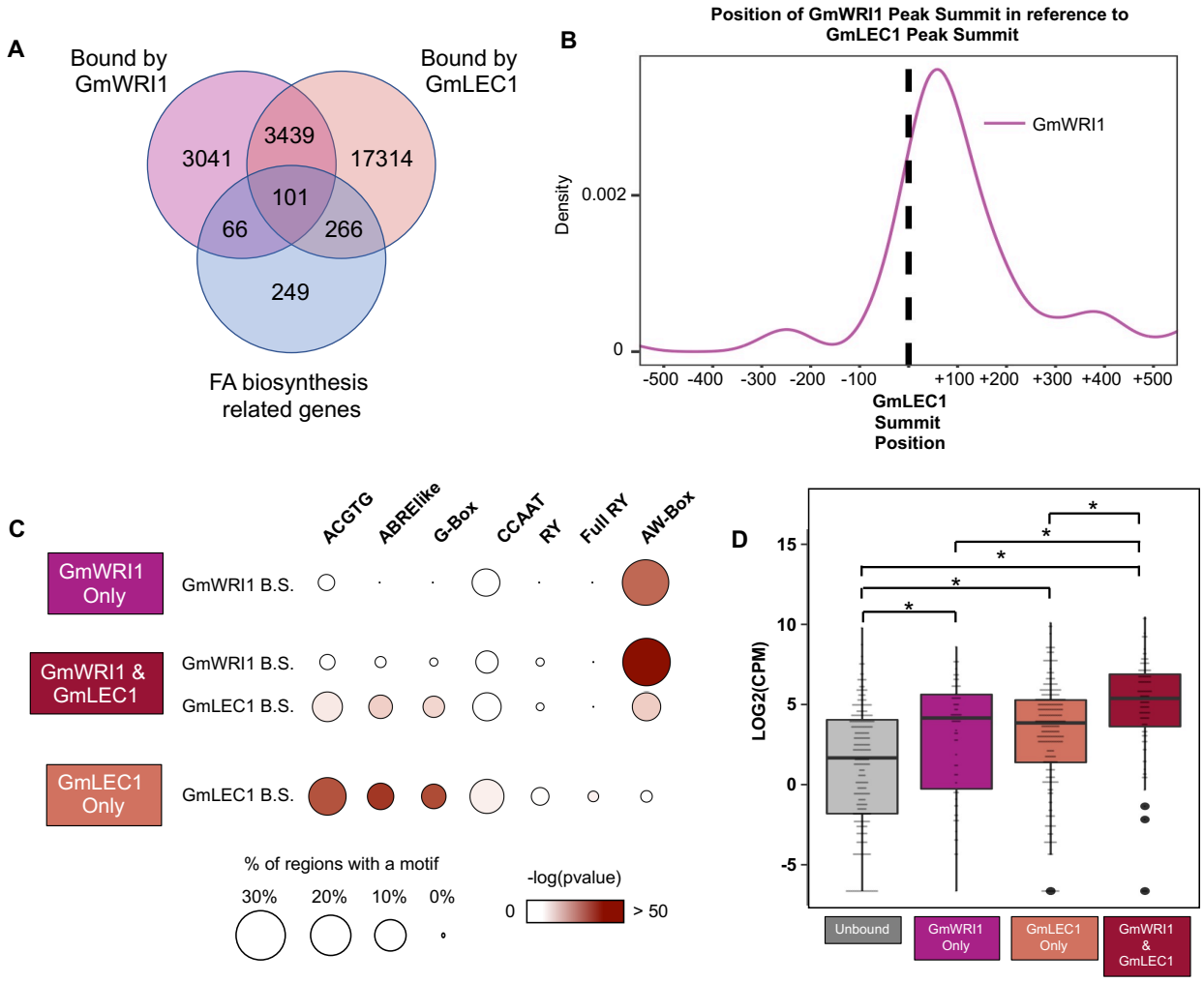


Figure 4. GmWRI1 and GmLEC1 act independently to control FA biosynthesis in soybean seeds. (A) Venn diagram showing the overlap between GmWRI1 bound genes (purple), GmLEC1 bound genes (salmon) and FA biosynthesis related genes (blue). (B) Density plot of the positions of the GmWRI1 ChIP peak summits in reference to the GmLEC1 peak summit position in genes that are bound by both TFs. (C) DNA motif enrichment in GmWRI1 and GmLEC1 bound sites (B.S.) (50 bps around the peak summit). Color intensity in the circles depict the statistical significance of the enrichment of annotated DNA motifs in bound regions relative to the normal distribution of a population of randomly generated regions. (Bonferroni-adjusted P values). Diameter of the circles depict the frequencies at which DNA motifs were identified in the bound sites. (D) Box plots showing the expression levels of FA related genes that are bound only by GmWRI1 (purple), only by GmLEC1 (salmon), both GmWRI1 and GmLEC1 (dark red) and not bound by both TFs (grey) in soybean embryos at em stage. Transcriptome data was obtained from the Harada Embryo mRNA-Seq Dataset (GEO accession no. GSE99571). Asterisk is to indicate significant differences between groups (Wilcoxon rank-sum test, with a significance threshold of $p < 0.01$).

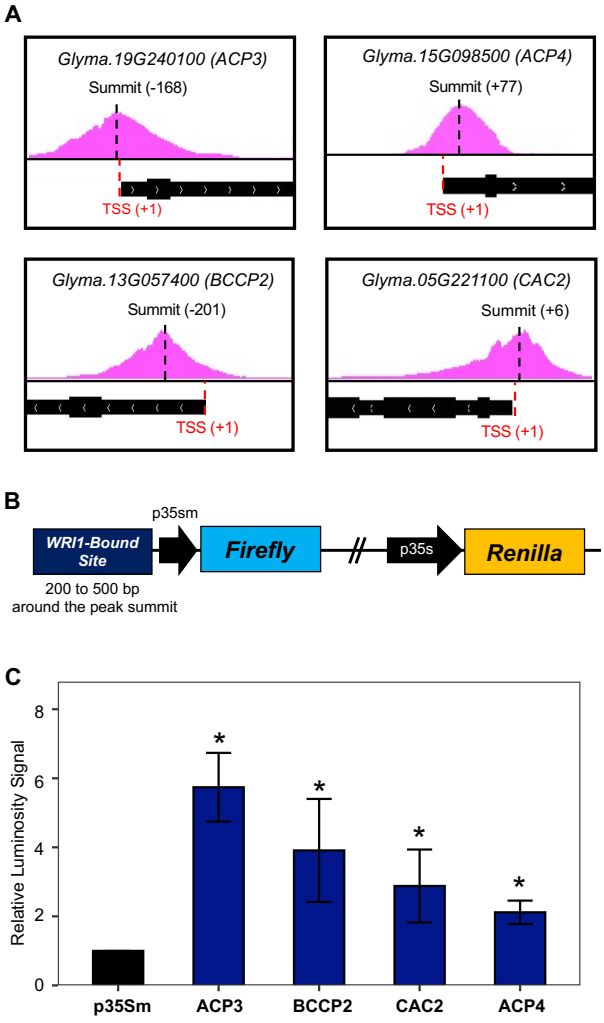


Figure 5. Functional analysis of GmWRI1 bound sites. (A) Genome browse view (Integrative Genome Viewer) of GmWRI1 (purple) ChIP-Seq peaks in bound genes *ACP3* (*Glyma.19G240100*), *ACP4* (*Glyma.15G098500*), *BCCP2* (*Glyma.13G057400*) and *CAC2* (*Glyma.05G221100*). Genes are not in the same size scale. Summits (black dashed lines) and TSS positions (red dashed lines) are shown. Numbers next the summit position depict relative distances of summit in reference to the TSS in base pairs. (B) Schematic representation of dual luciferase construct used to evaluate if bound sites function as transcriptional enhancers. GmWRI1 bound sites (fragments of 200 to 500 bps around peak summit) were inserted upstream a minimal p35Sm in the dual luciferase pDLUC15 plasmid. Sequences for bound sites can be found in Supplemental Table 1. (C) Values in barplots depict the relative luminosity signal, which corresponds to the ratio of Firefly to Renilla luciferase of protoplasts transfected by each construct normalized to the negative control (p35sm, no bound site inserted). Average value of 3 assays with SD are plotted. Asterisks denote statistically significant differences between bound sites and the negative control ($P < 0.05$, paired, one-tailed t tests).

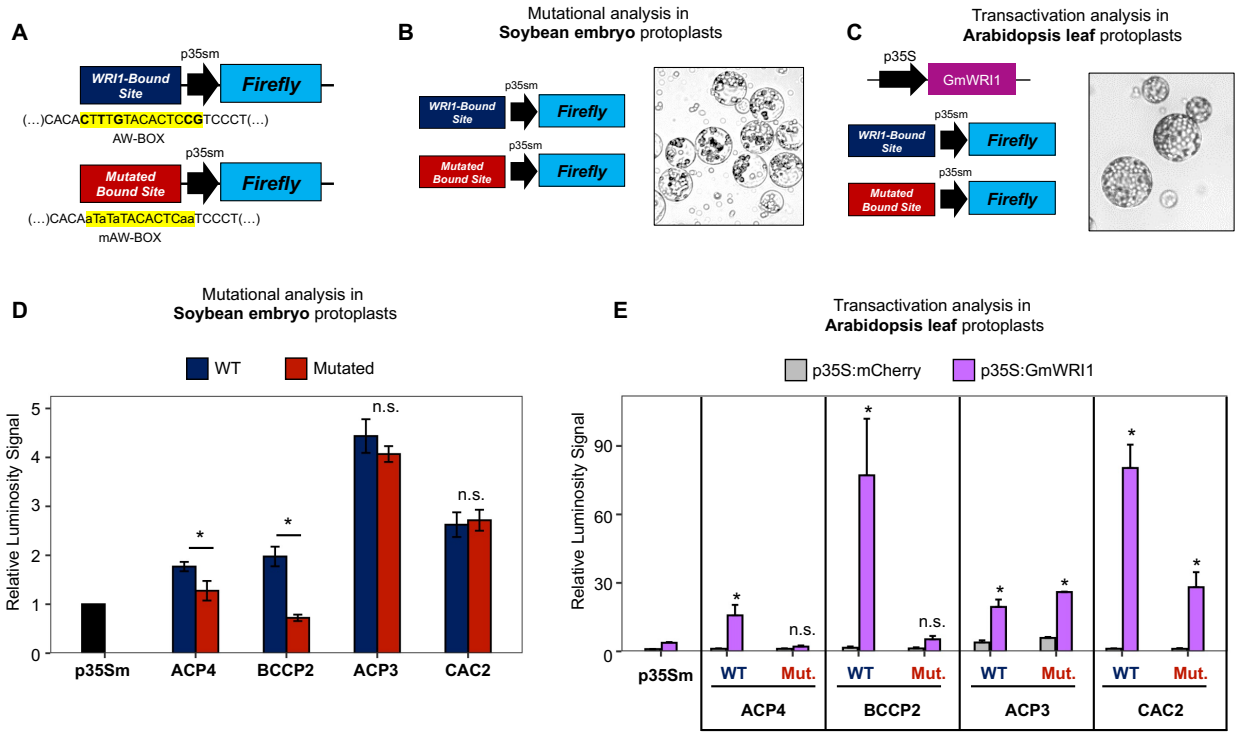


Figure 6. Functional analysis of AW-Boxes in GmWRI1 bound sites in FA related genes. (A) Schematic representation of constructs used for the analysis. WT and mutated bound sites were inserted immediately upstream of the 35S minimal promoter driving the expression of a Firefly luciferase gene. In the same plasmid a construct comprised of the p35S full promoter driving the expression of the Renilla luciferase gene was used as a transfection normalizing factor. (B) Schematic representation of constructs used in the loss-of-function analysis in soybean embryo protoplasts system. (C) Schematic representation of constructs used in transactivation analysis in Arabidopsis leaf protoplasts system. (D) Loss-of-function analysis in soybean embryo protoplasts at the em stage. Values depict the relative luminosity signal, which corresponds to the ratio of Firefly to Renilla luciferase activities normalized to the negative control (p35sm, no bound site inserted). Average value of 3 assays with SD are plotted. Asterisks denote statistically significant differences between WT (dark blue) and mutated (red) versions of the bound sites, whereas n.s. indicates no significant difference ($P < 0.05$, one-tailed t tests). (E) Transactivation analysis in Arabidopsis leaf protoplasts transfected with a p35S:mCherry (grey) or p35S:GmWRI1 (purple). For this assay, the relative luminosity signal is represented by the Firefly to Renilla ratio normalized to the negative control (p35Sm) co-transfected with a p35S:mCherry construct. Average value of 3 assays with SD are plotted. For this assay, asterisks denote statistically significant differences in the relative luminosity signal between protoplasts co-transfected with each construct and p35Sm:GmWRI1 against the negative control transfected with p35S:GmWRI1, whereas n.s. indicates no significant difference ($P < 0.05$, one-tailed t tests).

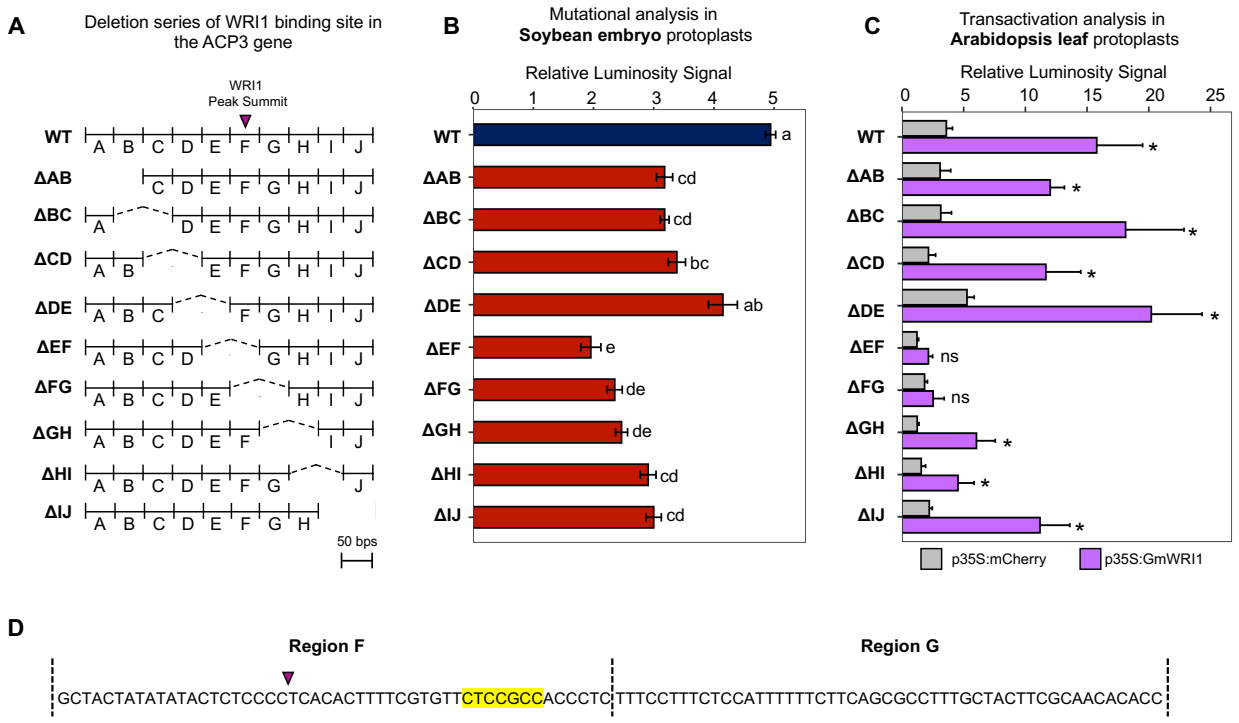


Figure 7. Deletion series of the GmWRI1 binding site of the ACP3 gene. A) Diagram representations of the WT form of GmWRI1 bound region in the ACP3 gene and the different variants used for the analysis. A series of overlapping 100 base pairs deletions were performed as indicated. Purple arrow indicates the position of the GmWRI1 Peak summit position in region F. (B) Loss-of-function analysis in Soybean embryo protoplasts at em stage. Similar to Figure 6D, values depict the relative luminosity signal of the WT form (dark blue) and deletion variants (red). Average value of 3 assays with SD are plotted. ANOVA and post hoc Tukey test results (depicted as lower-case letters) are showing the differences between the relative luminosity signal. (C) Transactivation analysis in Arabidopsis leaf protoplasts. Similar to Figure 6E, values depict the relative luminosity signal of protoplasts transfected with a p35S:mCherry (grey) or p35S:GmWRI1 (purple). Average value of 3 assays with SD are plotted. Asterisks denote statistically significant differences in the relative luminosity signal between protoplasts co-transfected with each construct and p35Sm:GmWRI1 against the negative control transfected with p35S:GmWRI1, whereas n.s. indicates no significant difference ($P < 0.05$, one-tailed t tests). (D) Nucleotide sequence of region F and G in the GmWRI1 bound region in the ACP3 gene. Purple arrow indicates the position of the GmWRI1 peak summit, and the CTCCGCC-Box is highlighted in yellow.

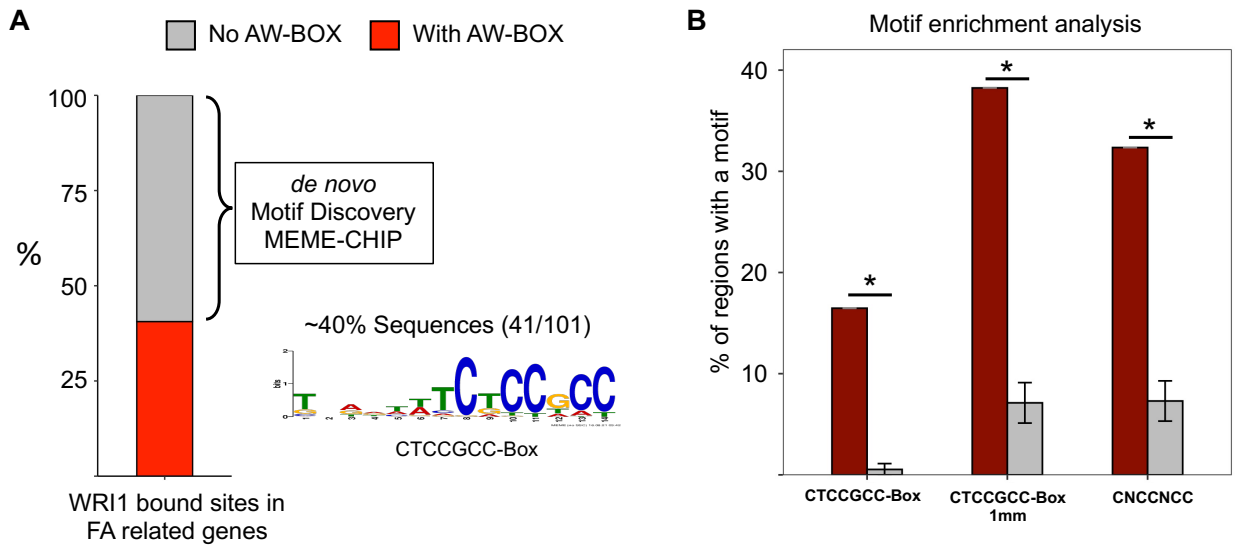


Figure 8. CTCCGCC-Box is enriched in GmWRI1 binding regions. (A, Left) Bar graph showing the percentage of GmWRI1 bound sites in genes that are FA related with (red) or without (grey) AW-Box. (A, right) Position weight matrices of the CTCCGCC-Box motif discovered *de novo* in the GmWRI1 bound regions that are absent of an AW-Box in genes that are associated with FA biosynthesis. (B) Motif enrichment analysis of CTCCGCC-Box, CTCCGCC-Box allowing 1 mismatch (CTCCGCC-Box-1mm) or CNCCNCC DNA elements in GmWRI1 bound regions in FA related genes. Graphs show the percentage of GmWRI1 binding sites with a DNA motif occurrence (red bars) and in the average of 1,000 population of genomic regions of comparable number, length and position relative to randomly selected genes (grey). Asterisks are to denote significant differences between groups (Bonferroni-adjusted P values with a significance threshold of 0.001).

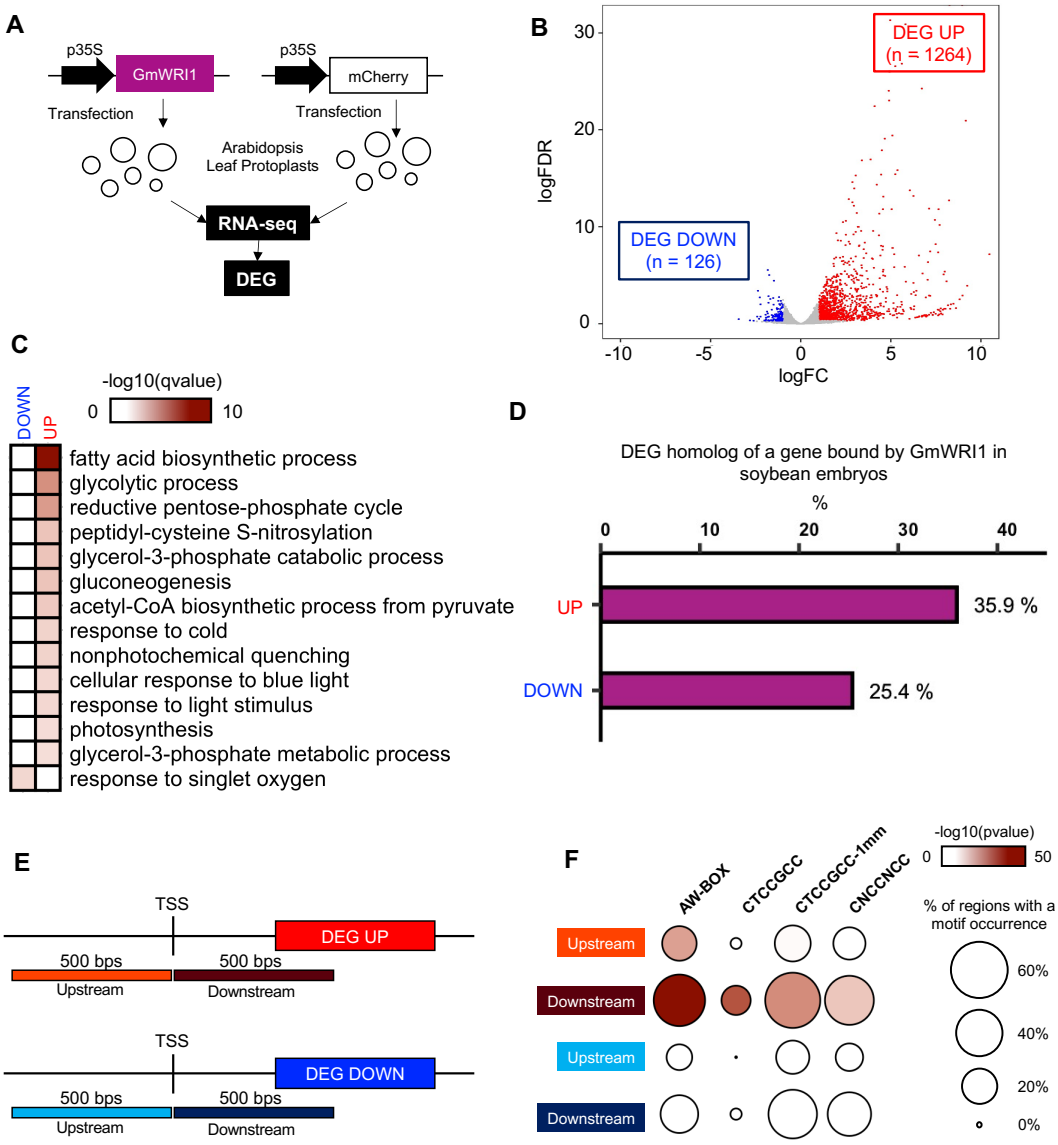


Figure 9. Ectopic overexpression of GmWRI1 in Arabidopsis leaf protoplasts. (A) Schematic representation of the gene consisting of the GmWRI1 cDNA sequence driven by the full 35S promoter (p35S:GmWRI1) that was used to transfect Arabidopsis leaf protoplasts. Protoplasts transfected by a gene consisting of MCHERRY cDNA driven by the full 35S promoter (p35S:mCherry) was used as control. RNA Seq experiments were conducted, and data analysis was performed to identify differentially expressed genes (DEGs). (B) Volcano plot of RNA-seq transcriptome data displaying the pattern of gene expression values for protoplasts transfected with p35S:GmWRI1 relative to protoplasts transfected with p35S:mCherry. Significant DEGs (FDR <= 0.001) are highlighted in red (upregulated, logFC >= 1) and blue (downregulated, logFC <= -1). (C) Heatmap showing the q value significance of GO terms for Upregulated and Downregulated DEGs. The GO terms listed are all the enriched biological process terms detected with a threshold of q value < 0.01. (D) Barplot showing the percentage of upregulated and downregulated genes with at least one soybean homolog gene that is bound by GmWRI1 at em stage. (E) Schematic representation of isolated regions used for motif

enrichment analysis. 500 bp regions upstream and downstream the annotated TSS were isolated for upregulated and downregulated genes. (F) DNA motif enrichment in the regions shown in (E). Intensity of the color in the circles depict the statistical significance of the enrichment of annotated DNA motifs in isolated regions relative to the normal distribution of a population of randomly generated regions. (Bonferroni-adjusted P values). Diameter of the circles depict the frequencies at which DNA motifs were identified in the isolated regions.

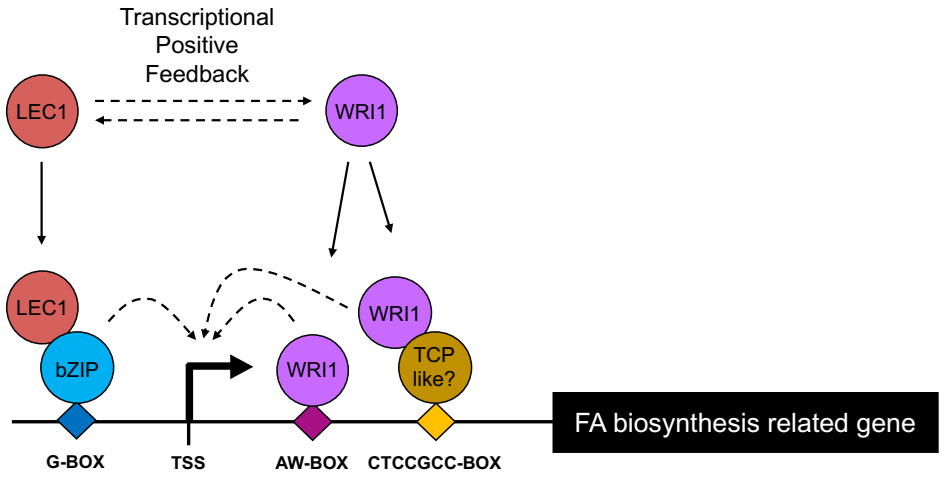
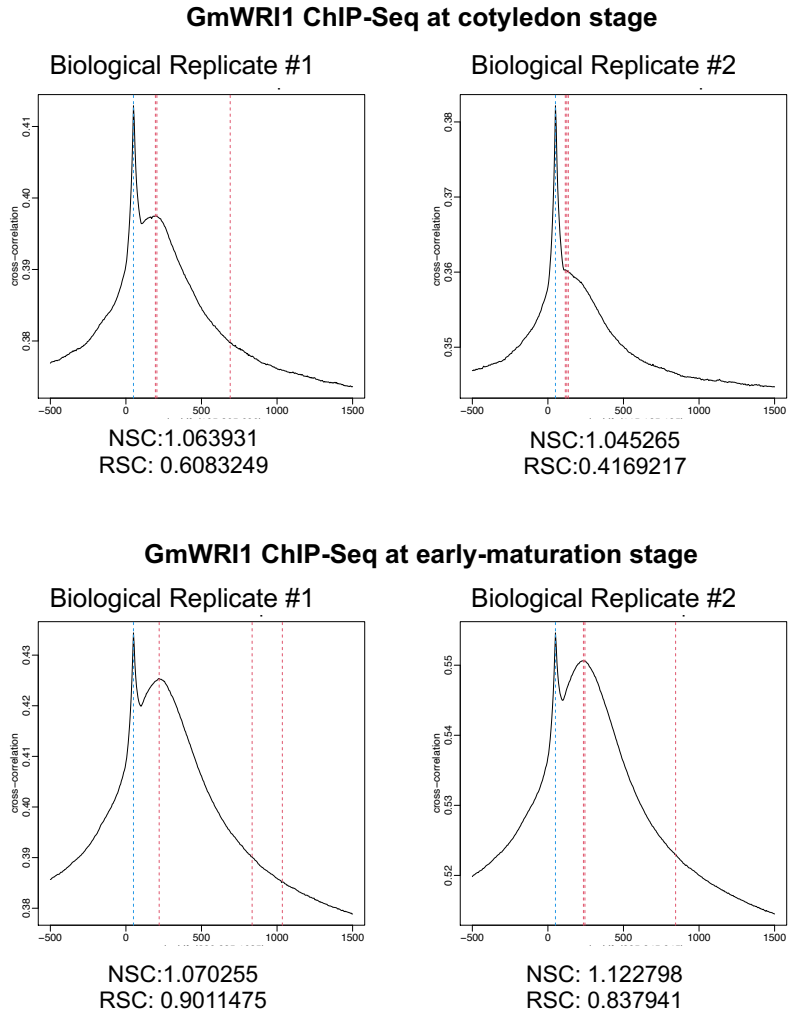
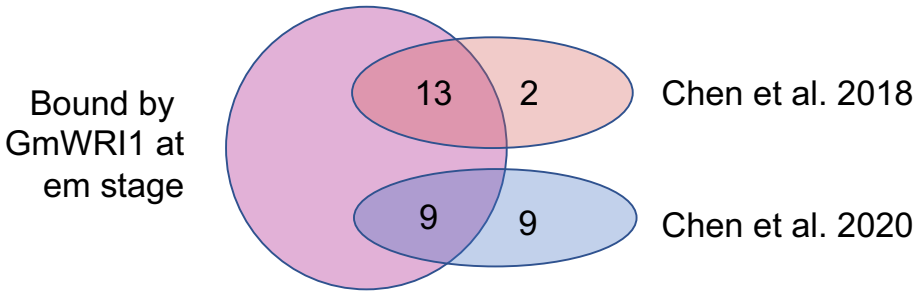


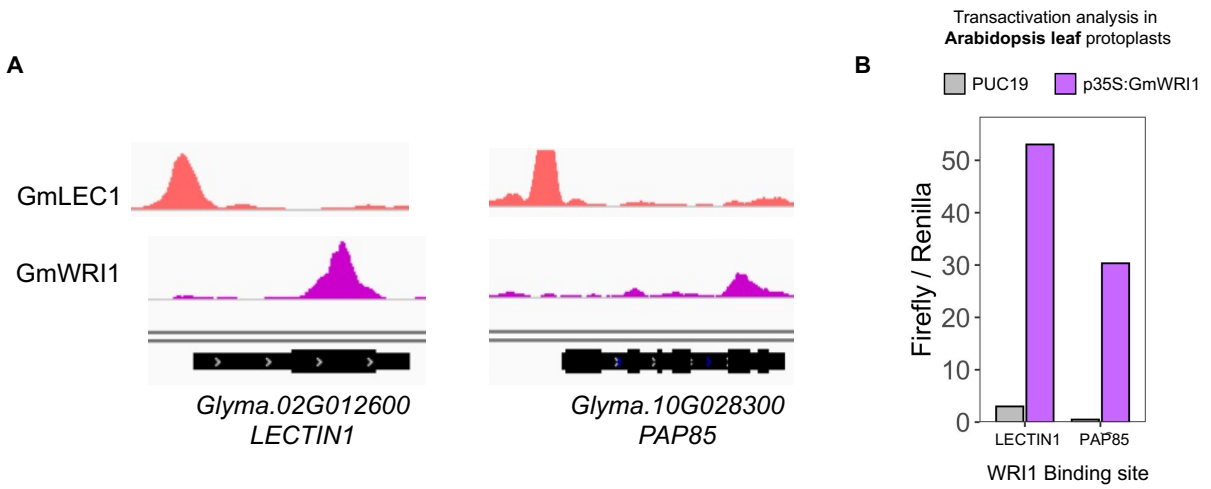
Figure 10. Model for the regulation of FA biosynthesis and lipid accumulation in seeds. GmWRI1 and GmLEC1 act in a transcriptional positive feedback subcircuit to regulate FA biosynthesis program in the seed. GmWRI1 regulate the expression of FA related genes by binding to AW-Boxes or a through binding to another TF (potentially a TCP-like TF) that binds to the CTCCGCC-Box located downstream the TSS. GmLEC1 another regulator of FA biosynthesis regulates the expression through the interaction with a bZIP TF that binds to G-Box DNA elements located mostly in regions upstream the TSS. Dashed arrows to indicate transcriptional activation.



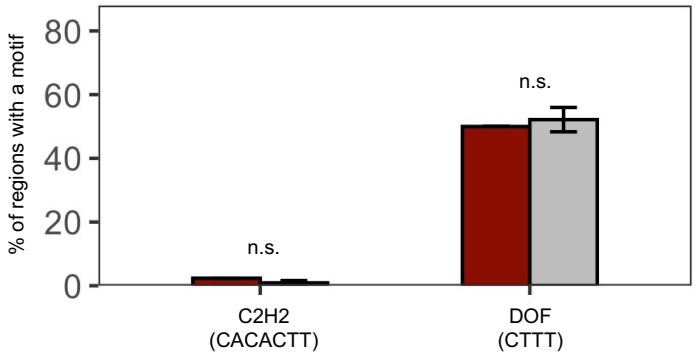
Supplemental Figure 1. Data quality metrics for the GmWRI1 ChIP-Seq datasets. Cross correlation plots of each biological replicate at cot and em developmental stages. Normalized strand correlation (NSC) and Relative strand correlation (RSC) values are reported for each replicate. The ENCODE minimum thresholds for NSC and RSC are 1.05 and 0.8, respectively.



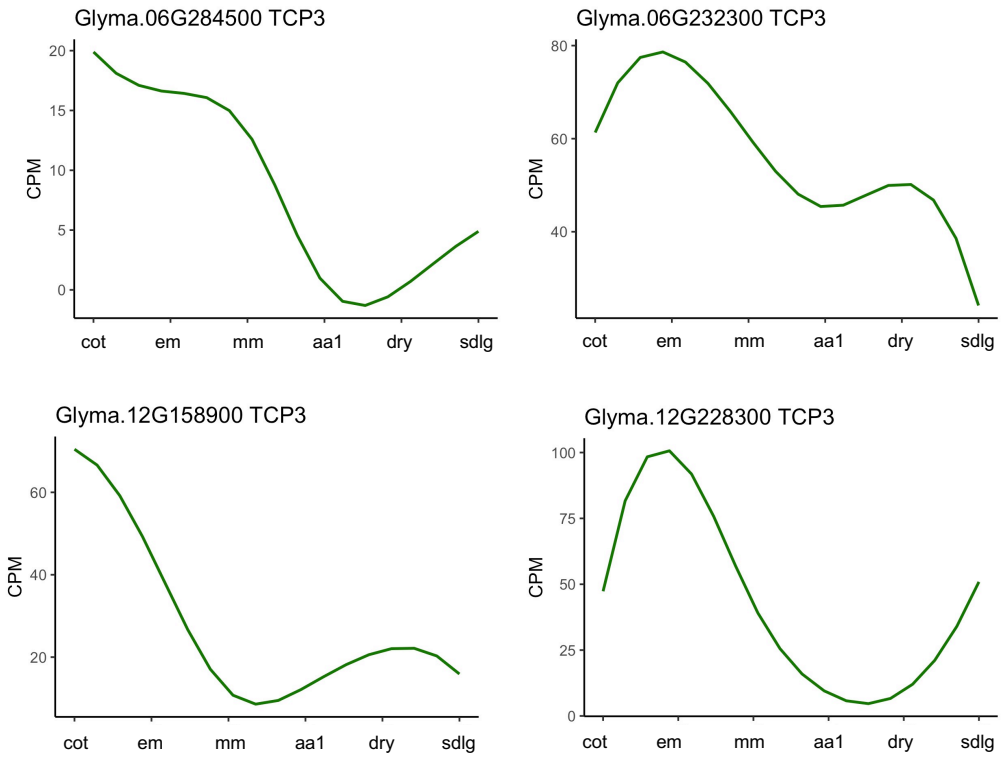
Supplemental Figure 2. Venn diagram to show the overlap between our reported GmWRI1 list of bound genes and previously described genes that are upregulated by the overexpression of GmWRI1 in hairy-roots (13) and soybean seeds (24).



Supplemental Figure 3. (A) Genome browser view of GmLEC1 (Salmon) and GmWRI1 (Purple) ChIP-Seq peaks in *Glyma.02G012600* (*LECTIN1*) and *Glyma.10G028300 (PAP85*). Genes are not in the same size scale. (B) Transactivation analysis in Arabidopsis leaf protoplasts. Values are depicting the relative luminosity signal of protoplasts transfected with a PUC19 (grey) or p35S:GmWRI1 (purple). Only one biological replicate was reported for this analysis.



Supplemental Figure 4. Motif enrichment analysis for the presence of C2H2 (CACACTT) and DOF (CTTT) DNA elements in GmWRI1 binding sites regions in genes related to FA biosynthesis. Graphs show the percentage of GmWRI1 binding sites with a DNA motif occurrence (red bars) and in the average of 1,000 population of genomic regions of comparable number, length and position relative to randomly selected genes (gray). No significant (n.s.) differences (Bonferroni-adjusted P value < 0.001) were detected between groups.



Supplemental Figure 5. Expression pattern of GmTCP3 genes during the development of the soybean embryo at the cot, em, mid-maturation (mm), double aa1 (aa1), dry seed stage (dry) and seedlings (sdlg). Transcriptome data was obtained from the Harada Embryo mRNA-Seq Dataset (GEO accession no. GSE99571)

Supplemental Table 1. Sequences of GmWRI1 bound regions used for functional analysis of AW-Boxes. Wild-type (WT) and mutated (mut) forms of AW-Boxes are highlighted in yellow and CTCCGCC-Box is highlighted in green.

Name	Sequence
CAC2-WT	GGTCGAAAGAAACCCAATTGTGGTGGGCCACGTAGCAGTCCAAATTGTCCGAGTACCTTCCAACTA GATTTAGATTTAGGCGGAGCTTTCAATTTCGCACCTCCCTCGTTTCCATCGACATCCGCATCAAAAG ACCTTCTCTTTCTCTTGCCATCATTTCATTCCCCATACGCTCCATCCATCACCCTTTCTTTCATTCTG
CAC2-mut	GGTCGAAAGAAACCCAATTGTGGTGGGCCACGTAGCAGTCCAAATTGTCCGAGTACCTTCCAACTA GATTTAGATTTAGGCGGAGCTTTCAATTTCGCACCTCaCaCaTTTCCATaaACATCCGCATCAAAAGAC CTTCTCTTTCTCTTGCCATCATTTCATTCCCCATACGCTCCATCCATCACCCTTTCTTTCATTCTG
ACP4-WT	CAGAACTCATATGTCAAATAAATTTTGTTGAAGTTATCATAATAATTTTATAATATAAATGTGTTAA CAAACTATATCATTAATAATATTTTAAGATAATTTTTAGTTATAATTCAGTAAGATCCTCGTATTATT TATTGTAAATTAATAAGCCCTTTACATATTAAATAACTAATGGTTGTCAGGTGCACCGAACAAAGA ATAAGAAGATCCAGATTCGAGGGACATGATCTCGGAGAGGCAAAGCGAACAAACGGCTACTCTCC TCCCACTTGCTATCCTCGAAATTAACGAAACGTAGTTGAAGAACCCACAGAATCAGAACCCCAACA CTGCCTTATGCTTTCCCTATAAATACACATTGTCCCTTCCTCTCGTCATTTCAAATACAAACTCACAA CACACTTTGTACACTCCGTCCCTCTTCCCTGTCTCTCAAATGGCTTCCCTGACGCAAACTTTCATGTC CCTCGCCTCTCTTCCCAATCTAGTTATGGTTTATAAAATCTCCTTTCTCGCTCTCTTCATTTCACTTGC TTCTGTTCC
ACP4-mut	CAGAACTCATATGTCAAATAAATTTTGTTGAAGTTATCATAATAATTTTATAATATAAATGTGTTAA CAAACTATATCATTAATAATATTTTAAGATAATTTTTAGTTATAATTCAGTAAGATCCTCGTATTATT TATTGTAAATTAATAAGCCCTTTACATATTAAATAACTAATGGTTGTCAGGTGCACCGAACAAAGA ATAAGAAGATCCAGATTCGAGGGACATGATCTCGGAGAGGCAAAGCGAACAAACGGCTACTCTCC TCCCACTTGCTATaCaCaAAATTAAaaAAACGTAGTTGAAGAACCCACAGAATCAGAACCCCAACACT GCCTTATGCTTTCCCTATAAATACACATTGTCCCTTCCTCTCGTCATTTCAAATACAAACTCACAACA CAaTaTaTACACTCaaTCCCTCTTCCCTGTCTCTCAAATGGCTTCCCTGACGCAAACTTTCATGTCCCTC GCCTCTCTTCCCAATCTAGTTATGGTTTATAAAATCTCCTTTCTCGCTCTCTTCATTTCACTTGCTTCT GTTCC
BCCP2-WT	GCACGTTATCGTTAAGTCAAAAGTGTACTCTTCTTTTGTACATACCAATGATGATACAATACCATAT TGGAAAAGTCAAAGACCTACAAAATCGGCCACGATAATAGCAAAAGCAGTGGCCATAGATCGATC AGTGCACTTGGAAGCTGTTTCGATCGTTGAAGACACCAACGTTGTGCACTTGGCTCTAGAGGGAAA CATCAAACATGTCTCCGAGACAAATGCAAAATCAGTGACTACAAGCCACGACATATGGTTTTTATT GTTGTTTGTTGCAAGTTAGTTAAAACTGTGTCACATTGATGGTTATTGCACCACTTGCATCGTAATC ATCGAAGCCACACTTGTAAAAATCTAAACTAAAACCAAACCAAACCAAACTAAAGTGTTCCAATTC TTCTTTGTATTCTCCATCCACCCTCATTGCACTTTTTAGTTTTTACCCTTTTTCGGTGTCTTCTCATATT CCTCACACGCACACCCTCTTCGATCTCAACGCCTCTTTGGCTTATATTCTCACCGACCCACATTGCCA TTCTCACCATTCTTTGCAGATCACGCACCCC
BCCP2-mut	GCACGTTATCGTTAAGTCAAAAGTGTACTCTTCTTTTGTACATACCAATGATGATACAATACCATAT TGGAAAAGTCAAAGACCTACAAAATCGGCCACGATAATAGCAAAAGCAGTGGCCATAGATCGATC AGTGCACTTGGAAGCTGTTTCGATCGTTGAAGACACCAACGTTGTGCACTTGGCTCTAGAGGGAAA CATCAAACATGTCTCCGAGACAAATGCAAAATCAGTGACTACAAGCCACGACATATGGTTTTTATT GTTGTTTGTTGCAAGTTAGTTAAAACTGTGTCACATTGATGGTTATTGCACCACTTGaAaaaTAATCAT aatAaCCACACTTGTAAAAATCTAAACTAAAACCAAACCAAACCAAACTAAAGTGTTCCAATTCTTCT TTGTATTCTCCATCCACCCTCATTGCACTTTTTAGTTTTTACCCTTTTTCGGTGTCTTCTCATATTCCT CACACGCACACCCTaTaCaATCTCAAaaCCTCTTTGGCTTATATTCTCACCGACCCACATTGCCATTCTC ACCATTaTaTaCAGATCAaaCACCCC

ACP3-WT	GATTGTGGATTTCATTATACTACTAATAAAAAATAACAATTAAAATTTACTGATAATAAAAAGTTGA TCAACATGCTGGTATTTTAAAATATTTTTATTTAAGAAAAAGTAGGGGTGTGAAGGAAAAATAAAT AAATAAGCAAACGTCTGTAAAAAGAAAGAAGCATGTGTAAAGCCCAGCCCAAAGGAAGGGGTACT TGAGATTCGGAATCCAAAGAGAGGACTCTTTCCTGCAGCCAACAATCACGGGGCTACTATATATAC TCTCCCCTCACACTTTTCGTGTTCTCCGCCACCCTCTTTCCTTTCTCCATTTTTTCTTCAGCGCCTTTG CTACTTCGCAACACACCACAATGGCCACCAATGCACTCGCTGGCACATCTCTCTCCATGCGATCTCT CCTTCCTCAAACACAGGTATTGTTTTCCCTTCCTCTTTTCATAAATCATAACCACTAGTACTATGTGC TTCGCACTGCCTCTTCTGTTCATTACTTCGCCGCACCTTTTACTCCTCACGACC
ACP3-mut	GATTGTGGATTTCATTATACTACTAATAAAAAATAACAATTAAAATTTACTGATAATAAAAAGTTGA TCAACATGCTGGTATTTTAAAATATTTTTATTTAAGAAAAAGTAGGGGTGTGAAGGAAAAATAAAT AAATAAGCAAACGTCTGTAAAAAGAAAGAAGCATGTGTAAAGCCCAGCCCAAAGGAAGGGGTACT TGAGATTCGGAATCCAAAGAGAGGACTCTTTCCTGCAGCCAACAATCACGGGGCTACTATATATAC TCTCCCCTCACACTTTTCGTGTTCTCCGCCACCCTCTTTCCTTTCTCCATTTTTTCTTCAGCGCaTaTaCT ACTTaaCAACACACCACAATGGCCACCAATGCACTCGCTGGCACATCTCTCTCCATGCGATCTCTCCT TCCTCAAACACAGGTATTGTTTTCCCTTCCTCTTTTCATAAATCATAACCACTAGTACTATGTGCTTC GCACTGCCTCTTCTGTTCATTACTTCGCCGCACCTTTTACTCCTCACGACC

Supplemental Table 2. Motifs identified through Plant Pan 3.0 promoter analysis tool (http://plantpan.itps.ncku.edu.tw/promoter.php). Only motifs with score 1 are shown.

Matrix ID	Family	Position	Strand	Score	Hit Sequence
TFmatrixID_0211	C2H2	23	+	1	cACACTt
TFmatrixID_0213	C2H2	23	+	1	cACACTt
TFmatrixID_0233	Dof	24	-	1	acaCTTTTcg
TFmatrixID_0235	Dof	26	-	1	aCTTTTcg
TFmatrixID_0243	Dof	26	-	1	aCTTTTcg
TF_motif_seq_0239	Dof	26	-	1	ACTTT
TF_motif_seq_0239	Dof	49	-	1	TCTTT
TF_motif_seq_0239	Dof	53	-	1	TCCTT
TF_motif_seq_0239	Dof	54	-	1	CCTTT
TF_motif_seq_0239	Dof	77	-	1	GCCTT
TF_motif_seq_0239	Dof	78	-	1	CCTTT
TF_motif_seq_0255	AP2; RAV; B3	92	+	1	CAACA
TF_motif_seq_0263	(Motif sequence only)	42	+	1	GCCAC
TF_motif_seq_0321	(Motif sequence only)	65	-	1	tTTTTC

Supplemental Table 3. Expression of TCP transcription factors in Arabidopsis leaf protoplasts transfected with p35S:mCherry or p35S:GmWRI1. Average CPMs of 3 biological replicates are shown. ND: non detected

TAIR10.agi	TAIR10.syn	p35S:mCherry	p35S:GmWRI1
AT3G02150	PTF1, TCP13, TFPD	87.43730924	74.06909425
AT1G53230	TCP3	19.44125937	69.01473446
AT1G58100	NA	22.21065863	19.96242487
AT2G45680	NA	14.2168882	18.14229547
AT3G15030	MEE35, TCP4	10.30070673	14.24161589
AT2G31070	TCP10	16.7535002	13.37044258
AT3G27010	AT-TCP20, ATTCP20, PCF1, TCP20	3.025938711	5.852662704
AT3G47620	AtTCP14, TCP14	5.252373434	5.550206667
AT4G18390	TCP2	4.597632577	4.646935927
AT1G72010	NA	2.534989871	3.660576306
AT1G69690	NA	3.42724129	2.987713313
AT1G30210	ATTCP24, TCP24	2.054892453	2.74637639
AT5G60970	TCP5	1.702020796	1.091442591
AT1G35560	NA	ND	ND
AT1G67260	TCP1	ND	ND
AT1G68800	BRC2, TCP12	ND	ND
AT2G37000	NA	ND	ND
AT3G18550	ATTCP18, BRC1, TCP18	ND	ND
AT3G45150	TCP16	ND	ND
AT5G08070	TCP17	ND	ND
AT5G08330	NA	ND	ND
AT5G23280	NA	ND	ND
AT5G41030	NA	ND	ND
AT5G51910	NA	ND	ND

Chapter 4

Transcriptional Regulatory Networks Controlling the Development of the Soybean Seed: Summary and Conclusions

Jo, L.

Introduction

The development of the seed is a complex yet elegant system, where distinct seed compartments undergo unique temporal and spatial developmental programs. The development of the main seed compartments (seed coat, endosperm and embryo) begins with the double fertilization of the egg and central cells of the embryo sac with two sperm cells that generate the embryo and endosperm, respectively (1). The development of the seed can be divided in two distinct developmental phases: morphogenesis and maturation. The morphogenesis phase is characterized by a series of cell proliferation, expansion and differentiation events which result in the development of the basic embryo body. This basic body pattern is maintained throughout the sporophytic life cycle of the plant. After the morphogenesis phase, the seed enters the maturation phase of development, which is characterized by the interruption of the patterning and proliferation events that occur during the morphogenesis phase (2,3). The maturation phase is also characterized by the synthesis and massive accumulation of storage compounds, such as seed storage lipids and proteins, that will be important for the establishment of the seedling after germination (4–6). It is also during the maturation phase that the embryo acquires the ability to survive desiccation (7). At the end of the maturation phase, the embryo is maintained in a developmentally arrested and quiescent state until conditions favorable for germination are encountered.

The development of the seed is hallmarked by the *de novo* initiation of developmental programs, particularly the development of the embryo. Thus, an important question in seed developmental studies is how novel developmental programs are initiated and controlled in the developing embryo. In a broad sense, inter- and intra-cellular signals promote shifts in the transcriptional landscape by regulating the localized expression of a specific set of regulatory genes (i.e., transcription factors (TFs)) that determine the specific developmental programs within

embryonic cells. These TFs can promote changes in the gene regulatory networks (GRNs) by interacting with specific DNA regulatory elements to regulate the expression of specific sets of genes (8,9). Because all cellular processes in differentiating cells depend on the genes they express, the GRN essentially dictates the temporal and spatial developmental programs (9). To fully understand the regulatory system that controls the development of the seed and the embryo, one needs to be able to characterize the GRNs associated with the commitment of individual developmental programs that occur in the developing seed. Such characterization includes understanding the expression of regulatory genes and the structure and function of DNA sequences that respond to these regulatory inputs.

My dissertation is aimed to understand how specific GRNs governs distinct biological programs that occur during the development of the soybean embryo. More specifically, my dissertation revolved around understanding the GRNs mechanisms of two putative regulators of seed development: the LEAFY COTYLEDON1 (LEC1) and WRINKLED1 (WRI1) TFs. These two TFs have been historically characterized as "master" regulator of seed development, due to their ability to initiate specific biological programs in the developing seed (10,11). While LEC1 has been characterized of a master regulator of seed maturation and embryo development (10), WRI1 was characterized as a master regulator of lipid biosynthesis program in seeds (11). Using functional genomic approaches, I was able to identify potential target genes of these TFs during the development of the soybean seed. The results presented in this dissertation confirmed that these TFs regulate many biological processes during the development of the soybean embryo. However, we also showed that often, these "master" regulators of seed development participate in a complex network where they associate with other TFs in the control of the expression of their target genes. Our work challenges the notion that these TFs act as isolated "master" regulators of seed

development and we provided important insights about the GRNs responsible for the onset of specific biological program in the developing soybean embryo.

Combinatorial interactions of LEC1 regulate diverse biological programs in soybean seeds

Our group previously identified all LEC1 target genes during the development of the soybean embryo (12). It was shown that LEC1 can regulate genes involved with many distinct biological programs, such as embryo morphogenesis, photosynthesis and seed maturation (12). Given that the onset of these biological programs occurs at distinct stages of seed development, one important question that remained was how LEC1 temporally regulates the expression of genes for these diverse biological programs. Previous studies showed that LEC1 can interact with several other TFs, and these interactions were important to specify LEC1's ability to regulate the expression of specific target genes (Reviewed in (10). We explored the possibility that LEC1 acts in combination with the TFs ABA-RESPONSIVE ELEMENT BINDING PROTEIN3 (AREB3), BASIC LEUCINE ZIPPER67 (bZIP67), and ABA INSENSITIVE3 (ABI3). We selected these TFs characterized because they are involved in similar biological process with LEC1, and they also showed the ability physically interact with LEC1 in other plant species (Reviewed in (10).

Using a combination of chromatin immunoprecipitation followed by DNA-sequencing experiments and the analysis of the seed transcriptome, we were able to identify potential target genes for these TFs (13). When compared to the target genes identified by LEC1, we found that LEC1 shares many of the target genes with AREB3, bZIP67 and ABI3 (13). Interestingly, we found that the distinct combinations of TFs with LEC1 were enriched for distinct gene ontology (GO) annotation categories. We found that LEC1 alone regulates the expression of genes involved with embryo morphogenesis while LEC1 associated with AREB3, bZIP67 and ABI3 often

regulates genes involved with seed maturation and photosynthesis (13). These results point to the fact that distinct combinations of LEC1 with AREB3, BZIP67 and ABI3 are responsible for the regulation of unique biological programs in soybean seeds. These results could indicate that distinct LEC1 associations are responsible for the temporal onset of specific biological programs during the development of the soybean embryo. However, when we evaluated the temporal expression pattern of genes targeted by distinct LEC1 combinatorial interactions, we didn't identify a clear relationship between temporal expression and distinct LEC1 combinatorial interactions. These results suggests that the genes that underlie specific developmental processes during seed development are regulated temporally, but the LEC1 associations alone can't explain entirely their temporal regulation. Perhaps other TFs also play an important role to fine tune the temporal expression of those genes.

Another interesting finding of our work is that LEC1, AREB3, bZIP67 and ABI3 bind to similar loci upstream of their target genes (13). We tested if these high occupancy regions represent *cis*-regulatory modules (CRMs). CRMs are genomic regions at which multiple, distinct TFs bind productively to regulate gene transcription (9). First, we demonstrated that the distinct LEC1 CRM regions function as enhancers in soybean embryo cells (13). Second, we found that this CRMs are often located in accessible regions in the chromatin (data not shown). These results suggests that these CRMs are likely to be important to determine the distinct LEC1 associations to regulate the expression of its target genes.

We then investigated the mechanism behind the distinct LEC1 associations with AREB3, bZIP67 and ABI3 in our characterized CRMs. First, we confirmed the physical interactions between LEC1 and bZIP67, LEC1 and AREB3, AREB3 and bZIP67, and bZIP67 and ABI3, as occurs in Arabidopsis (10,13). Second, we show that distinct LEC1 CRMs groups are enriched

for different sets of DNA motifs (13). Our results showed that the motif composition of CRMs can explain, at least partially, the distinct LEC1 associations with AREB3,bZIP67 and ABI3. Expression assays with embryo cells indicate that the enriched DNA motifs are functional cis elements that regulate transcription. Taken together, our results showed that the interactions between LEC1 and AREB3, bZIP67 and ABI3 in CRMs to be determined by their ability to physically interact with each other, as well as the DNA element composition in CRMs.

Finally, we also evaluated if the distinct DNA motifs found in CRMs are organized in a similar motif grammar (position and composition of DNA elements within the CRM). Even though we observed that genes regulated by distinct LEC1 combinatorial interactions are clearly enriched by distinct sets of DNA elements, we couldn't identify a common a motif grammar in the distinct group of CRMs (13). Perhaps not every DNA motif found in the CRMs can be considered as a functional DNA element, which makes it difficult to find a specific motif grammar in the distinct set of CRMs. It would be interesting to know how the motif grammar in CRMs can influence the steric hindrance or stability of LEC1 transcriptional complexes, and consequently, LEC1 function to regulate the expression of its target genes. The results presented here provided important insights into the mechanisms of LEC1 and other TFs to regulate the expression of distinct biological programs during the development of the soybean embryo.

Regulation of the lipid accumulation program during the development of the soybean seed

My dissertation identified important aspects of the regulatory mechanisms behind lipid accumulation in soybean seeds. We characterized the binding profile of the TF WRI1, a putative regulator of lipid biosynthesis in seeds (11). We show that this TF can bind to several genes that encode for enzymes involved with every step of the fatty acid (FA) and triacyl-glycerol (TAG)

metabolic pathway (Chapter 3). The WRI1 binding profile also provided important insights into the mechanisms by which WRI1 regulates the expression of FA biosynthesis related genes in soybean. We showed that the putative DNA element bound by WRI1 (AW-Box) is likely to be responsible for WRI1 regulation of many FA biosynthesis related genes. However, we observed that for some FA biosynthesis related genes, WRI1 function is not fully dependent on the AW-Box (Chapter 3). Interestingly, we identified the presence of a CTCCGCC-Box enriched in WRI1 binding sites that indicates that other TFs may collaborate with WRI1 in the induction of the FA biosynthetic network. For the next steps, it would be important to determine if this new DNA element is important for the function of WRI1 as well as to determine what other TFs could bind to this unique DNA motif.

I also presented data that suggest WRI1 and LEC1 act in a positive feedback subcircuit in the control of FA biosynthesis in seeds (Chapter 3). This is consistent with the notion that LEC1 is also an important regulator of lipid accumulation in seeds (14). We showed that WRI1 and LEC1 bind to each other and to many FA biosynthesis related genes, many of which are shared between these two TFs (Jo et al., 2020, Chapter 3). However, unlike the association between LEC1 and AREB3, bZIP67 and ABI3 (13), our data suggest that WRI1 and LEC1 bind to distinct loci in FA related genes (Chapter 3). Out results also suggests that WRI1 and LEC1 co-binding can synergistically affect the transcriptional activation of their target genes (Chapter 3). These results provide new insights into the relationship between these two important regulators of lipid biosynthesis in seeds.

In addition to the key enzymes required for the synthesis of FA and TAG, other cellular processes are crucial to ensure the proper accumulation of lipids in seeds. FA biosynthesis requires a constant source of carbon, energy and reducing power derived from reactions in plastids (15,16).

The metabolism of sucrose into hexose-phosphates to feed the FA biosynthesis pathway in seed plastids is crucial for the accumulation of lipids in seeds (16). Here, we showed that LEC1, AREB3, bZIP67 and ABI3 regulate several genes required for plastid development and the assembly of the light harvesting complex in plastids (12,13). Another important aspect of lipid accumulation in seeds is the ability to store lipids into distinct organelles called oil bodies (17). OLEOSINS (OLE) are important integral membrane proteins found in the phospholipid monolayer that surround the oil bodies (17). We also showed data indicating that many *OLE* genes are regulated by the LEC1, AREB3, bZIP67 and ABI3 transcriptional complex.

Taken together, we showed that the accumulation of lipid in seeds requires the coordination of several distinct TFs to regulate the expression of genes involved in all the distinct steps necessary for this complex metabolic pathway. Our results have provided important insights into the regulatory mechanisms of lipid accumulation in soybean seeds and can serve as a basis for the establishment of strategies to improve the accumulation of this important metabolite.

Summary and concluding remarks

The results presented in my dissertation have shown that the regulation of distinct biological programs is mediated by complex GRNs in the developing soybean seed. I have described the mechanisms of LEC1 association with AREB3, bZIP67 and ABI3 in the coordination of specific biological programs during the development of the soybean embryo. Moreover, we showed that high TF occupancy regions are likely to define the position of functional CRMs that behave as *cis*-acting enhancers in the regulation of specific genes in the developing seed. Finally, we discovered important aspects of the mechanisms involved in the regulation of lipid accumulation in soybean seeds. We showed that WRI1 and LEC1 act in a feedback positive

circuit in the regulation of the lipid accumulation program in the developing soybean seed. Additionally, I provided evidence to suggest that WRI1 function requires the assistance of other TFs to regulate the expression of FA biosynthesis related genes. Taken together, I hope my work contributes to an understanding of the complexity of the regulatory networks that governs seed development. For future developmental studies, more than understanding how unique TFs act to control the expression of specific genes, it is important that one understands how TFs that comprise the regulatory state of the seed are interacting and modulating the diversity of biological programs in the developing seed.

References

1. Goldberg RB, De Paiva G, Yadegari R. Plant embryogenesis: zygote to seed. Science. 1994;266(5185):605–14.

2. Raz V, Bergervoet JH, Koornneef M. Sequential steps for developmental arrest in Arabidopsis seeds. Development. 2001;128(2):243–52.

3. Vicente-Carbajosa J, Carbonero P. Seed maturation: developing an intrusive phase to accomplish a quiescent state. International Journal of Developmental Biology. 2004;49(5–6):645–51.

4. Harada JJ. Seed maturation and control of germination. In: Cellular and molecular biology of plant seed development. Springer; 1997. p. 545–92.

5. Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C. Combined networks regulating seed maturation. Trends in plant science. 2007;12(7):294–300.

6. Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L. Deciphering gene regulatory networks that control seed development and maturation in Arabidopsis. The Plant Journal. 2008;54(4):608–20.

7. Leprince O, Pellizzaro A, Berriri S, Buitink J. Late seed maturation: drying without dying. Journal of experimental botany. 2017;68(4):827–41.

8. Levine M, Davidson EH. Gene regulatory networks for development. Proceedings of the National Academy of Sciences. 2005;102(14):4936–42.

9. Peter IS, Davidson EH. Genomic control process: development and evolution. Academic Press; 2015.

10. Jo L, Pelletier JM, Harada JJ. Central role of the LEAFY COTYLEDON1 transcription factor in seed development. Journal of integrative plant biology. 2019;61(5):564–80.

11. Kong Q, Yuan L, Ma W. WRINKLED1, a "Master Regulator" in transcriptional control of plant oil biosynthesis. Plants. 2019;8(7):238.

12. Pelletier JM, Kwong RW, Park S, Le BH, Baden R, Cagliari A, et al. LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. Proceedings of the National Academy of Sciences. 2017;114(32):E6710–9.

13. Jo L, Pelletier JM, Hsu S-W, Baden R, Goldberg RB, Harada JJ. Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. Proceedings of the National Academy of Sciences. 2020;117(2):1223–32.

14. Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, et al. Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. Plant physiology. 2010;153(3):980–7.

15. Bates PD, Stymne S, Ohlrogge J. Biochemical pathways in seed oil synthesis. Current opinion in plant biology. 2013;16(3):358–64.

16. Baud S. Seeds as oil factories. Plant reproduction. 2018;31(3):213–35.

17. Buchanan BB, Gruissem W, Jones RL. Biochemistry and molecular biology of plants. John wiley & sons; 2015.