UC Davis
UC Davis Electronic Theses and Dissertations

Title
Regulation of Gene Expression in Soybean Seed Development

Permalink
https://escholarship.org/uc/item/2zw029wf

Author
Uzawa, Rie

Publication Date
2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Regulation of Gene Expression in Soybean Seed Development

By

Rie Uzawa

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. Andrew Groover

Dr. Robert L. Fischer

Committee in Charge

2021

Regulation of Gene Expression in Soybean Seed Development

Abstract

Soybean (*Glycine max*) has been one of the most important crops to feed humans and animals for over 3000 years. Soybean seeds are a major source of proteins and oils. The goal of my dissertation is to understand how genes are regulated during soybean seed development. My dissertation is focused on two gene regulatory mechanisms for soybean seed development: gene regulation by micro RNAs (miRNAs) and by transcription factors.

I first conducted studies on soybean seed development by profiling the miRNA populations in each seed subregion at different developmental stages. miRNAs have been shown to play important roles in plant development by regulating target mRNAs. Plant miRNAs have been profiled in different tissues of various plant species including soybean seeds. However, no studies have been comprehensively conducted to profile miRNAs at the single subregion level in soybean seeds. In order to get insight into the roles of miRNAs for soybean seed development, I have demonstrated that many miRNAs predominately accumulate in one subregion or stage. In particular, the endosperm subregion is enriched with miRNAs that have been identified only in soybean, whose functions have not yet been studied. Some of these non-conserved miRNAs specifically accumulate in the endosperm at a high level. I also explored the function of the target mRNAs of miRNAs. Based on the analysis of the target mRNAs, the endosperm-specific miRNAs may play important roles in biotic and abiotic stress responses. These results shed light on the role of miRNAs in soybean seed development.

The second focus of my dissertation is to elucidate the gene regulatory network for soybean seed maturation. The seed maturation program is controlled by four transcription factors: LEAFY COTYLEDON1 (LEC1), ABA-RESPONSIVE ELEMENT BINDING PROTEIN3 (AREB3), BASIC LEUCINE ZIPPER67 (bZIP67) and ABA INSENSITIVE3 (ABI3). Two transcription factors, GROWTH-REGULATING FACTOR5 (GRF5) and HOMEOBOX22 (HB22) were identified as direct targets of *LEC1, AREB3, bZIP67* and *ABI3*. I have investigated the genes regulated by GRF5 and HB22 to expand the maturation gene regulatory network. I have demonstrated that GRF5 and HB22 shared many of the same target genes that are jointly regulated by *LEC1, AREB3, bZIP67* and *ABI3.* Furthermore, GRF5 binding sites are located close to those of LEC1, AREB3, bZIP67 and ABI3 on the common targets. By contrast, HB22 binding sites are more distantly located from those of LEC1, AREB3, bZIP67 and ABI3. My results suggest that GRF5 may regulate target genes involved in the soybean seed maturation program in combination with LEC1, AREB3, bZIP67 and ABI3 whereas HB22 may regulate the same genes independently. These results elucidate the roles of GRF5 and HB22 in soybean seed maturation program.

Acknowledgement

I would like to thank my mentor, John J. Harada who have guided me throughout my enormous journey for Ph. D., which seemed to me impossible to achieve. Without his support, guidance and patience, I could not have been accomplished. He also taught me how to be a compassionate educator.

I would like to thank the present and past members of Harada Lab, especially Julie Pelletier and Leonardo Jo. You not only helped me but also taught me how to be a great scientist with your patience.

I would like to thank Bob Goldberg who showed me great professionalism and passion toward science.

I would like to thank Andrew Groover to serve both my qualifying exam and dissertation committees. His advices are always thoughtful.

I would like to thank Bob Fischer to be my "science father". He has always believed in me and encouraged me.

I want to thank all my friends. Too many to name, but you all helped me go through my journey.

I want to thank my kitty babies, Serenity (aka Nyan-Nyan), Tama and Zoey. Your cuteness always comforts me.

I want to express my gratitude to my family outside the US (namely my parents, Norihiko and Misae Inaoka, to my sister, Aki Inaoka, and to my brother, Masahiro Inaoka). You always give me a big laugh.

Finally, I want to thank my husband, daughter and son (namely Satoru, Karin and Daichi). You all are the joy and treasure of my life.

Table of Contents

Chapter 1

Gene regulation during seed development

Introduction

Seed development is initiated by double fertilization. The haploid egg cell and the homodiploid central cell are each fertilized with a sperm cell to generate the diploid zygote/embryo and the triploid endosperm, respectively. The embryo and endosperm are surrounded by the maternally-derived seed coat which is originated from the ovule integuments (1). These seed regions give rise to more specialized subregions through cell division and differentiation.

The first cell division of the zygote generates the apical and basal cells, which develop into the embryo proper and the suspensor with the hypophysis, respectively. The suspensor is a terminally differentiated tissue that anchors the embryo to the embryo sac and serves as a conduit between the seed coat and embryo proper. Cells derived from the upper-most part of the basal cell differentiate into the hypophysis, which later becomes part of the root apical meristem (RAM). The apical cell becomes the embryo proper which gives rise to the cotyledons and the axis. The axis contains the RAM and shoot apical meristem (SAM) that generate all the tissues required for the mature plants after germination. The cotyledons are terminally differentiated tissues that synthesize and store food reserves for the germinating seedling (2). This early phase of embryo development is called the morphogenesis phase. Once the morphogenesis phase is over, the embryo enters the maturation phase when seeds accumulate storage molecules to prepare for seedling growth (3). At the late maturation stage, seeds acquire desiccation tolerance to withstand the reduction of water content. In order to maintain the viability of seeds for a long time before germination, seeds undergo further metabolic changes by accumulating lipids and sugars (4–8).

In dicots such as Arabidopsis and soybean, the endosperm initially undergoes nuclear division without cell wall formation after fertilization. These syncytial nuclei migrate from the embryo-surrounding micropylar end to the chalazal end. Subsequently, cell walls form around the nuclei to generate endosperm cells, and the endosperm is fully cellularized in the seed. As the embryo grows, the endosperm is depleted gradually. By the time the embryo enters the maturation phase, most endosperm tissues have degenerated, with only a single cell layer persisting below the seed coat at maturation (9–12).

The seed coat starts differentiating and expanding from the ovule integuments upon fertilization. The seed coat consists of multiple cell layers that protect the embryo. The seed coat also provides nutrition for the embryo at an early developmental stage before photosynthesis begins. As the embryo grows larger, some of the seed coat layers are crushed and degenerate (7,13,14).

In order to produce a viable seed, it is crucial that spatial and temporal development are coordinated at the single subregion level. The coordination of spatial and temporal developmental processes depends on the precise control of gene expression.

In this chapter, I review gene regulation during seed development in each seed region. Next, I will discuss genome-wide gene expression profiles in the seeds in eudicot and monocot to compare and contrast the gene expression profiles with underlying functions between the two groups. Finally, I will discuss the importance of studying soybean seed development.

Development and gene regulation in seed regions

In this section, I will discuss the overall seed developmental processes in each region and the gene regulation that governs the spatio-temporal developmental processes.

Molecular mechanisms of gene regulation

Gene expression is controlled by a variety of molecular mechanisms, including transcriptional activation and repression by transcription factors (TFs) (15), post-transcriptional gene regulation (16,17) and through changes in chromatin conformation (18,19).

TFs are proteins that play an important role in the regulation of tissue-specific gene expression as well as gene expression in response to specific stimuli by binding to *cis*-regulatory elements with distinct DNA sequences (DNA motifs) at gene promoters or enhancers (15,20). Coexpressed genes often share the same set of *cis*-regulatory elements, suggesting that these genes are regulated by the same set of TFs (21–23). Furthermore, a single TF can regulate different sets of gene by forming complexes with different TFs or co-regulators (24,25). Thus, TFs are major participants in the regulation of gene expression.

Regulation of transcription is often mediated by chromatin structure. In eukaryotes, DNA is packaged into chromatin. In order for TFs to bind to specific *cis*-regulatory elements, the *cis*-regulatory elements must be accessible. Thus, chromatin remodeling proteins play important roles to achieve the accessibility of *cis*-regulatory elements to TFs (26). There are three major classes of proteins that are involved in chromatin remodeling: histone chaperones, histone modification enzymes and ATP-dependent chromatin remodeling enzymes. The first class, histone chaperones interact with core histones to assemble or disassemble nucleosomes in an

ATP-independent manner. Histone chaperones are often associated with other chromatin modifiers and the DNA replication machinery to remodel chromatin architecture (27,28). Histone modification enzymes are the second class of chromatin remodeling proteins. Post-translational covalent histone modifications, such as acetylation/deacetylation, phosphorylation/dephosphorylation, methylation/demehtylation and ubiquitination/deubiquination, affect the degree of chromatin compaction (28,29). The third class is ATP-dependent chromatin remodeling enzymes that are brought to specific areas of chromatin by other proteins and use energy from ATP to separate or bring together nucleosomes, which controls accessibility of DNA sequences to DNA binding proteins (19,28).

Micro RNAs (miRNAs) are a class of small RNAs 21-22 nt in length. miRNAs are a major posttranscriptional regulator of gene expression. miRNAs guide a RNA endonucleolytic enzyme within a protein complex, the RNA-induced silencing complex (RISC), to the target mRNAs with complementary nucleotide sequences. Binding of the RISC complex either cleaves the target mRNA transcript or inhibits its translational (16,17).

These independent gene regulatory mechanisms also work together to precisely control gene expression. For example, some TFs (i.e., Pioneer TFs (30)) can interact with DNA sequences in the packaged chromatin state. This interaction can change chromatin architecture by recruiting chromatin remodeling factors to initiate gene expression (31,32). Also, many TF genes are regulated by specific miRNAs in various plant developmental processes. The regulatory relationships between TF and miRNAs are highly conserved in land plants, suggesting the importance of miRNA regulation of TF gene expression (17). Furthermore, plants with a mutation in a subunit of the ATP-dependent chromatin remodeler SWR-1, decrease accumulation of some miRNAs, suggesting that chromatin state may affect miRNA gene transcription (33).

Taken together, genes are regulated at multiple levels to coordinate the processes required to make viable seeds. In the following sections, I will review how genes are regulated during seed development.

Early embryo development and gene regulation

Embryo development starts from a single fertilized cell called the zygote. After the egg cell is fertilized, the zygote starts to elongate asymmetrically in many angiosperms. The elongated zygote then undergoes a transverse cell division to produce the apical cell with a small cytoplasm and the basal cell with a large vacuole (34). The apical cell undergoes two rounds of longitudinal and one round of transverse cell divisions to generate the 8-cell proembryo. The transverse cell division gives rise to the two separate domains of the proembryo. The cells in the upper tier divide and differentiate into the shoot apical meristem (SAM) and cotyledons, whereas the cells in the lower tier become the hypocotyl, radicle and part of the cotyledons (35,36). The 8-cell proembryo further undergoes periclinal cell divisions, which gives rise to the protoderm, and ground tissues to form the dermatogen stage globular embryo. The protoderm is the outer layer of the embryo proper, which serves as the epidermis layer of the embryo later (35).

The basal cell of the zygote gives rise to the hypophysis and suspensor. The hypophysis is derived from the upper most cell, which differentiates into the root quiescent center of the embryo. The suspensor is derived from the lower part of the basal cell which anchors the embryo proper to the surrounding maternal tissue and serves as a nutrient conduit (5,35). The morphology of the suspensor varies from species to species. Some species consist of a single file of cells whereas others have more elaborate shapes (34). The suspensor eventually degenerates at later stages of embryo development.

WOX (*WUSCHEL HOMEOBOX*) transcription factors encode a plant-specific subclade of the eukaryotic homeobox superfamily (37). Some WOX family members regulate apical-basal body axis establishment. *WOX2* and *WOX8* are directly induced by the *WRKY DNA-BINDING PROTEIN2 (WRKY2)* TF (38) and are co-expressed in the egg cell and zygote (39). After the zygote undergoes the first cell divisions, *WOX2* and *WOX8* are expressed predominantly in the apical and basal cell, respectively (39). Weak loss-of-function *wrky2* mutants have an altered asymmetric cell division and generate another proembryo-like structure whereas strong *wrky2* mutants arrest embryo development at an early stage. The *wrky2* mutant phenotype is restored by introducing the *WOX8* transgene, suggesting *WOX8* play an important role for the initial apical-basal embryo axis establishment (38).

WOX9*,* another WOX family member, also plays an important role in establishing the apical-basal embryo axis. WOX9, closely related to WOX8, is first expressed in the same domain as *WOX8* after the first cell division (39,40). At the four-cell stage of the embryo proper, *WOX9* expression is restricted to the hypophysis. The results of *in situ* hybridization showed *WOX9* expression was observed in the lower tier of the proembryo at the eight-cell stage. After the eight-cell stage, *WOX9* expression was not observed in the hypophysis and was restricted to epidermal cells in the lower part of the embryo proper (39). Wu et al. (40) also observed similar expression patterns with WOX9:GFP fusion protein in *wox9* loss-of-function mutant background, except GFP signals were detected more broadly in the embryo and suspensor at the same stages. Embryos with different mutant *wox9* alleles show variable degrees of severities in morphological phenotypes, including embryo lethality, reduction in embryo length, horizontal growth, and abnormal cell divisions in the suspensor. These phenotypes were likely caused by the cell division arrest based on observations of marker gene expression (40). Together, the WRKY2-WOX2-WOX8/9 gene regulatory network is required for establishment of the embryonic apical-basal axis.

Mutations in additional genes affect apical-basal organization (41). Two of those mutants, gain-of-function *bondelos (bdl)* and loss-of-function *monopteros (mp)*, have similar phenotypes. Both mutants affect embryo structures, including the hypophysis. Map-based analyses identified *BDL* and *MP* as encoding the IAA12 transcriptional repressor and the ARF5 TF, respectively (42,43), which regulate auxin responses. An auxin gradient is established in the very early stages of embryogenesis. Auxin is first detected in the apical cells. By the preglobular stage, auxin has predominately accumulated in the hypophysis, through the actions of PIN auxin efflux carriers (44). This auxin flux induces RAM formation. An auxin response was not detected in *bdl* or *mp* mutants bearing the synthetic auxin-responsive gene construct, DR5 fused with GFP (DR5rev::GFP) (44), suggesting BDL and MP are required for auxin activity to establish the basal embryonic structure. Moreover, plants with a *pin7* mutation show apical-basal pattern defects in Arabidopsis (44). These results suggest that auxin plays an important role in establishing apical-basal patterning in the embryo. Interestingly, *WOX9* expression depends on BDL and MP*,* suggesting there is a crosstalk between *WOX* and *BDL/MP* gene pathways (39).

After the initiation of the apical-basal body axis, the embryo starts to grow radially. The first step for radial patterning is the specification of the protoderm by periclinal cell divisions in the embryo proper which occurs during the transition from the 8-cell to 16-cell stage. A class IV homeodomain-leucine zipper TF, *ARABIDOPSIS THALIANA MERISTEM LAYER 1 (ATML1)* mRNAs are first detected in the apical cell after the first transverse division of zygote. *ATML1* expression expanded to the entire proembryo by the 8-cell proembryo stage. Strikingly, *ATMl1* expression was restricted to the protoderm at the 16-cell stage (45). Although *atml1* mutants do not display any phenotypes, Arabidopsis plants with mutations in *ATML1* and another class IV homeodomain-leucine zipper TF family member gene, *PROTODERMAL FACTOR 2 (PDF2),* caused structural defects on the embryo surface as well as growth arrest at the globular stage, suggesting the functional redundancy of two genes required for epidermal development during early embryo development.

Ground and vascular tissues are derived from inner cells of the proembryo. The *ZWILLE/ARGONAUTE10 (ZLL/AGO10)* gene is initially expressed at the 4-cell stage, and by the dermatogen stage, its expression is restricted to the inner cells of the proembryo (46). Although *zll/ago10* mutant early embryos did not show any mutant phenotypes, the shoot apical meristem (SAM) was not maintained in the embryo at the later stages. The centrally localized stem cells in a wild-type SAM are undifferentiated, and as they proliferate and move outside the center, they begin to differentiate and generate aerial plant organs, leaves and stems (47). However, the central cells in the *zll/ago10* SAM precociously differentiate based on histological analyses (48). ZLL/AGO10 is closely related to AROGONAUTE1 (AGO1) which is a subunit of the effector protein complex required for miRNA-mediated gene regulation (16,49). AGO10 is predominately associated with miR165/166 in *Arabidopsis* based on AGO10 immunoprecipitation experiments. Interaction with AGO10 sequesters miR165/166 to prevent cleavage of their target mRNAs encoding class III HD-ZIP TFs. Therefore AGO10 promotes the activity of class III HD-ZIP TFs in the SAM (49). *PHABULOSA (PHB)* encodes a class III HD-ZIP TF. *PHB* expression is detected as early as the 8-cell stage embryo. By the dermatogen stage, *PHB* expression is restricted at the future SAM location that maintains the undifferentiated state of stem cells in the SAM (50). Thus, SAM formation is regulated by class III HD-ZIP TFs, mainly *PHB*, under the control of ZLL/AGO10 regulatory network.

The stem cells in the SAM are maintained by multiple pathways other than the ZLL/AGO10-miR165/166- class III HD-ZIP TF network. At the time when *PHB* expression is confined to the SAM, *WUSCHEL (WUS)* expression is also detected in the inner cells of the dermatogen embryo. *WUS* expressions is restricted in the organizing center at the later stage. WUS is a member of the WOX TF family. SAM size and the position of stem cell populations are regulated by the CLAVATA/WUS feedback-loop network (36).

The cotyledons and the coleoptile emerge at the region surrounding the SAM in dicots and monocots (47,51). At the time of cotyledon and coleoptile emergence, the SAM becomes more visible. Two major developmental domains of the embryo proper, the central and peripheral domains, are defined. The central domain is under the control of the ZLL/AGO10-miR165/166- class III HD-ZIP TF network, whereas the peripheral domain is controlled by the KANADI (KAN) and YABBY (YAB) genes. The expression of *KAN* and *YAB* are complementary to the expression of class III HD-ZIP TFs. Mutations in either *KAN* or *YAB* causes morphological change in the embryo (52,53) due to the expanded expression of class III HD-ZIP TFs in the embryo, suggesting that KAN or YAB antagonize HD-ZIP TFs to maintain the central and peripheral domains (53). In addition to the antagonistic action between KAN/YAB and class III HD-ZIP TFs, KAN genes are involved in auxin transport. Auxin transport was monitored using a PIN1-GFP reporter line in *kan* mutant plants. Compared to plants with functional KAN, PIN1-GFP expression expands to the peripheral domain and the hypocotyl in *kan* mutants, suggesting morphological changes in *kan* mutants may result from abnormal auxin transport (54).

In addition to KAN and YAB, Miyashima et al. (55) showed that miR165/166 primary transcripts from MIR165/166 microRNA genes were detected in the peripheral domain of the embryo proper. They demonstrated that the mature processed miRNAs, miR165/166, regulated the class III HD-ZIP TFs, PHABULOSA(PHB) gene expression in the peripheral domain of the embryo. To elucidate the regulation of PHB by miRNAs, they generated transgenic plants with reporter lines for *PHB* expression: wild-type PHB (PHB-GFP) and mutant PHB (PHBmu-GFP) containing silent mutations in the miR165/166 binding site. The embryos with PHB-GFP exhibited GFP signals only in the central domain without any morphological changes. By contrast, PHBmu-GFP exhibited expanded GFP signals in the entire embryo proper. Plants bearing the PHBmu-GFP transgene also showed morphological alterations. Furthermore, the same GFP patterns were recovered when they co-transformed PHBmu-GFP and an altered miR165 with mutations that are complementary to the PHBmu-GFP construct. These results confirmed that miR165/166 regulate class III HD-ZIP TFs to maintain the central and peripheral domain.

Taken together, embryo morphogenesis phase is controlled by various gene regulatory networks.

Embryo development in the maturation phase and gene regulations

After establishing the basic body plan, the embryo ceases cell divisions. The embryo starts expanding by accumulating storage proteins and lipids to prepare for desiccation and germination. The embryo undergoes hormonal and metabolic changes. Abscisic acid (ABA) is produced and accumulated in the embryo, which peaks at the maturation phases to promote seed maturation processes (3). ABA acts antagonistically against gibberellins (GA) to prevent early germination (56,57).

As the embryo undergoes maturation process, the carbohydrate profile in the embryo changes. During the morphogenesis phase in legumes, sucrose imported from the maternal tissue is enzymatically converted to hexoses by invertase to promote cell division and growth. Once the seed enters the maturation phase, the cotyledon can take up sucrose directly via a sucrose transporter of epidermis transfer cells. As a result, the sucrose:hexose ratio increases in the embryo (3,58). The physiological, hormonal and metabolic changes act as signals for maturation processes, which are highly integrated by the regulation of gene expression.

LEAFY COTYLEDON1 (LEC1), LEAFY COTYLEDON2 (LEC2), ABSCISIC ACID INSENTITIVE3 (ABI3) and FUSCA3 (FUS3) are the core maturation phase regulators. *LEC1* encodes a nuclear transcription factor-YB (NF-YB) subunit of the NF-Y CCAAT-binding TF (59). *LEC2, ABI3,* and *FUS3* are B3 DNA binding domain containing TFs (60). The *lec1, lec2* or *fus3* mutations affect storage protein and lipid accumulation in seeds (60,61) although the *lec2* mutant show somewhat different storage molecule accumulation patterns (60–62). The *abi3* mutation affects desiccation tolerance and storage protein accumulations. The *abi3* mutation induces the germination program prematurely, bypassing the late maturation program, leading to seed shrinkage with green, mature embryos (63). Although *lec1, lec2*, *fus3* and *abi3* share similar phenotypes, all four TFs have unique gene expression domains at different developmental stages (64). Moreover, the phenotypes of the double mutants of *lec1, lec2, fus3*, and *abi3* in different combinations showed synergistic effects on maturation processes with some overlaps in function (61). In order to dissect the gene regulatory network for *LEC1, LEC2, FUS3* and *ABI3*, To et al (60) monitored gene expression of *LEC2* and *ABI3* in different combinations of *lec1, lec2, fus3*, and *abi3* double or triple mutant plants. They found that the embryo maturation gene regulatory network of *LEC1, LEC2, FUS3* and *ABI3* acted locally and redundantly. For example, expression of the gene for the storage protein, At2S3, was not detected in part of the *lec2* mutant embryo cotyledons, but it was detected in the embryo axis. *At2S3* gene expression pattern matched with *ABI3* expression in *lec2* mutant plants, suggesting *LEC2* regulates *ABI3* for storage protein expression in a spatially restricted manner. Furthermore, overexpression of *FUS3* partially restores *At2S3* expression in *lec2* mutant plants. These results confirm the involvement of LEC1, LEC2, FUS3 and ABI3 in the embryo maturation program. It also reflects the complexity of the gene regulatory network of the four maturation TFs.

Based on analyses of the *lec1* mutant (61) and inducible overexpression of *LEC1* (65), *LEC1* is placed upstream of *LEC2, ABI3* and *FUS3* in the maturation regulatory program. Yamamoto et al (66) shed light into the molecular mechanism of LEC1 regulation of storage protein genes. LEC1-LIKE (L1L) is most closely related to LEC1 among the members of the NF-YB subunit gene family in Arabidopsis. The NF-Y CCAAT-binding TF is a complex consisting of NF-YA, NF-YB and NF-YC subunits, which binds to the CCAAT DNA motif to regulate target gene expression. Each subunit is highly conserved in eukaryotes (67). NF-Y subunits can interact with other TFs as a trimer of all NF-Y subunits (67). Furthermore, NF-YB and NF-YC heterodimers can interact with proteins other than NF-YA to bind different DNA motifs (68,69). Yamamoto et al (66) showed, using an Arabidopsis leaf protoplast transient gene expression assay, that *L1L* activated the expression of the seed storage protein *cruciferin C (CRC)* gene with NF-YC2 and the seed-specific ABA-response element (ABRE) binding TF, *BASIC LEUCINE ZIPPER TRANSCRIPTION FACTOR 67 (bZIP67)* in an ABA-dependent manner. They also showed that L1L physically interacted with bZIP67 by immunoprecipitation experiments. Although the promoter of *CRC* carries both CCAAT and ABRE motifs, only the ABRE motif is required for activation of CRC. Together, these results showed that L1L, NF-YC2 and bZIP67 form a TF complex to regulate *CRC* via bZIP67 binding at the ABRE motif.

lec1 and *fus3* mutant embryos in the maturation phase have decreased lipid content compared with wild type (61). Plants ectopically expressing *LEC2* induce the formation of somatic embryos with a high level of lipid body protein mRNAs (70). These studies indicated the involvement of LEC1, FUS3 and LEC2 in lipid synthesis during embryo maturation. Baud et al (71) elucidated the transcriptional regulation for oleosin, an oil body protein that is required for triacylglycerol (TAG) storage. Transactivation studies using a GFP reporter in moss protoplasts showed that the *OLE1* promoter becomes most active when *LEC1, LEC2* and *ABI3* are co-expressed, suggesting LEC1, LEC2 and ABI3 synergistically activate *OLE1*.

The *WRINKED1* gene was isolated by mutant screening for embryos with low TAG contentment (72,73). *WRI1* encodes a AP2/EREB family TF. Overexpression of *WRI1* causes increases in the total fatty acid amount in seeds (73). One of *wri1* mutant phenotypes is higher carbohydrate accumulation, suggesting defects in the incorporation of carbohydrates into TAGs (72). Furthermore, *wri1* phenotypes are observed only in the maturation phase. Interestingly, overexpressed *WRI1* seedlings grown in sucrose media causes TAG accumulation, which normally only occurs in embryos, further suggesting that WRI1 is a key regulator for the seed maturation program (73). Expression of *WRI1* is induced by LEC2 in cotyledons of embryos, not in the axis, suggesting LEC2-mediated *WRI1* gene regulation is localized and *WRI1* is regulated by different TFs in a tissue-specific manner (74).

Inducible *LEC1* overexpression in Arabidopsis LEC1-OX seedlings showed an increase in various types of fatty acids (75). mRNA profiles using microarray experiments with LEC1-

OX seedlings showed upregulation of genes related to fatty acid and lipid metabolism. In LEC1-OX seedlings, both *ABI3* and *FUS3* are upregulated. Seedlings with LEC1-OX in *abi3* and *fus3* backgrounds show reductions of certain fatty acids, suggesting that LEC1, ABI3 and FUS3 work together to regulate fatty acid biosynthesis. Furthermore, *WRI1* expression is upregulated in LEC1-OX seedlings. LEC1-OX seedlings in a *wri1* mutant background show a reduction of certain fatty acids. These results suggest that LEC1 is a master regulator for fatty acid biosynthesis via ABI3, FUS3 and WRI1. WRI1 is likely regulated by LEC2 as well.

In addition to the core seed maturation transcriptional regulators, miRNAs play an important role in seed maturation. Loss-of-function mutants of the miRNA biogenesis gene, *dcl1*, arrest embryo development at an early stage in *Arabidopsis.* Plants with a *dcl1* mutation upregulated many miRNA target genes. One of the most significantly upregulated target mRNAs encode the *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 10/11 (SPL10 and SPL11)* transcription factors, which are regulated by miR156. Early stage embryos with the *dcl1* mutation also upregulated several other seed maturation genes. Expression of these seed maturation genes was suppressed in *dcl-1* and *spl11* double-mutant embryos. Furthermore, SPL10 and SPL11 with miR156-resistant mutations upregulated the seed maturation genes at the early stage of embryo development. These observations suggest that miR156 prevents the early onset of maturation programs by regulating SPL10 and SPL11 (76). It would be interesting to know how the miR156 gene regulatory network may interact with LEC1, ABI3, FUS3 and LEC2 gene regulatory module for the seed maturation process.

Taken together, LEC1, ABI3, FUS3 and LEC2 govern various biological processes in the maturation phase, including storage protein and fatty acid biosynthesis. The gene regulatory network for LEC1, ABI3, FUS3 and LEC2 is intricate and highly redundant.

Endosperm development and gene regulations

The endosperm is a terminal tissue that is only present in the angiosperms The main function of the endosperm is to sustain embryo development and germination (9). Endosperm development starts with fertilization of the homodiploid central cell in the megagametophyte with one of the two sperm cells in the pollen. The first several divisions after fertilization of the central cell are nuclear divisions without cell wall formation to form a structure called the coenocyte. At this stage, the nuclei migrate from the embryo-surrounding area to the area away from the embryo (6,77). After the coenocytic phase, the endosperm undergoes cellularization by forming cell walls around the nuclei (9,10). In cereals there are four endosperm domains: the outermost aleurone layer, the central starchy endosperm which is the largest domain, the thin layer around the embryo called the Embryo Surrounding Region (ESR), and the transfer cells domain at the chalazal end called the Basal Endosperm Transfer Layer. In *Arabidopsis*, nuclei migrate from the micropylar region to the chalazal region. Three different domains are established along the micropylar-chalazal axis. Micropylar endosperm (MCE) is located around the embryo. Peripheral endosperm (PEN) is located in the central chamber. Chalazal endosperm (CZE) is located at the opposite end of the endosperm from the MCE. Cereal (monocot) endosperm store storage molecules such as starch and storage protein in the starchy endosperm and endosperm is retained in the mature seed. These storage molecules are mobilized by enzymes released from the aleurone layer during germination (9,10). By contrast, in *Arabidopsis* and other eudicot species the endosperm is largely degraded when the seed maturation starts (9,10) and its storage molecules are stored in cotyledons. Overall, the core processes for early endosperm development are similar between maize and *Arabidopsis*. However, as described above, there are significant differences between monocots and dicots later in endosperm development.

The endosperm shares the space with the embryo confined within the seed coat. Thus, proper endosperm development is critical for the embryo growth as well. The endosperm is the unique tissue where genomic imprinting occurs. Genomic imprinting refers to parent-of-origin gene expression. Imprinted genes are either downregulated or silenced at either the paternal or maternal allele. Certain paternal-expressed and maternal-expressed imprinted genes promote or restrict endosperm cell proliferation and development, respectively, and mutations in expressed alleles can result in seed abortion (78). The mechanisms governing genomic imprinting involve DNA methylation and histone modification to downregulate or silence imprinted genes at one of their alleles (79,80). *DEMETER (DME)* encodes a DNA glycosylase domain protein that demethylates previously methylated loci by removing methylated cytosine in CG, CHG, and CHH sequence contexts where H indicates A, C, or T (81). *DME* is primarily expressed the central cell of the female gametophyte (78). *MEDEA (MEA)* and *FERTILIZATION-INDEPENDENT SEEDS2 (FIS2)* are the members of the Polycomb Repressing Complex 2 (PRC2) proteins that repress target genes via chromatin remodeling by forming the complex with *FERTILIZATION INDEPENDENT ENDOSPERM (FIE)* and *MULTIPLE SUPPRESOSOR OF IRA1 (MSI1)*. DME activity is required for maternal expression of *MEA* and *FIS2* imprinted genes. In the central cell, maternal allele expression of *MEA* and *FIS2 is* induced by DME-mediated DNA demethylation. Maternal-derived MEA polypeptide, a repressive H3K27 methylation enzyme in the PRC2 complex, silences the paternal *MEA* allele in the endosperm. Hence, *MEA* is a self-imprinted gene. *FIS2* gene imprinting is solely regulated by DNA methylation – DME demethylates the maternal *FIS2* allele in the central cell, and the MET1 DNA methyltransferase methylates and silences the paternal allele in the sperm.

mea mutants show developmental embryo arrest and overproliferated endosperm. Fourquin et al. (82) investigated the underlying mechanisms for the developmental relationship between the embryo and endosperm. *ZHOUPI (ZOU)* encodes a basic helix-loop-helix (bHLH) TF that is expressed exclusively in the endosperm. In *zou* mutants, the endosperm could not be broken down due to the increase in stiffness of the endosperm cell. As a result, embryo development was arrested early. They found that cell wall modification genes were significantly downregulated in *zou* mutant siliques seeds. In wild type plants, the expression of several cell wall modification genes gradually increased and peaked at the young torpedo stage when the endosperm was completely cellularized. However, these genes were not detected in *zou* mutants. They hypothesized that the lack of the endosperm breakdown in *zou* mutants was due to the absence of cell wall modifications. In order to answer this question, they measured the physical stiffness of the cell walls by using a nano-indenter. They found that the endosperm in *zou* mutants became much stiffer than the endosperm in the wild type plants, suggesting that ZOU is required for embryo growth by modifying the endosperm cell walls, which leads to the endosperm breakdown.

Taken together, proper gene regulation in the endosperm is critical for embryo growth and for seed development as a whole.

Seed coat development and gene regulations

The seed coat is the protective layer of the seed. It is derived from the ovule integuments. The major role for the seed coat is not limited to embryo protection. The seed coat can produce and transport the metabolites necessary for early embryo development. The seed coat is also the place where molecules for seed defense are synthesized and deposited. The seed coat consists of different tissue types – the epidermis, outer integument, inner integument and endothelium from the outside to inside. The outer and inner integuments are divided into several layers. The numbers of the layers vary from species to species (13).

After fertilization, the seed coat undergoes cell divisions and differentiates. As the seed coat develops, different chemical compounds are produced in the different parts of the seed coat. In *Arabidopsis,* mucilage, pectic polysaccharide compounds that hydrate the seed surface when the seed is imbibed, are produced and secreted from the epidermis. By contrast, the epidermis of legumes such as soybean and *Medicago* do not produce mucilage. The Arabidopsis seed coat synthesizes flavonoid compounds in the endothelium, the inner most seed coat layer, which gives the brown color to the seeds. Plants use flavonoids for various purposes, including defense against microbial activities, UV-B light, and the regulation of auxin transport (13). In soybean, chitinase and peroxidase were isolated from soybean seed coat, which are thought to promote plant defense (83). Some of the seed coat layers are broken down and crushed as the embryo grows inside the seeds, and the eventually a few layers remain in mature seeds (7,13,83).

The seed coat is not involved in plant fertilization, but its development is initiated by fertilization (7). The study of Arabidopsis plants heterozygous for a loss-of-function *FIE* mutation showed that seed coat initiation depends on endosperm development. The central cell only has the *fie* mutant allele and initiates endosperm development. The heterozygous seed coat, with the dominant wild-type *FIE* allele, undergoes develops, suggesting that it responds to a signal from the endosperm (84,85). Furthermore, the MADS box TF, *AGL62,* is expressed in the central cell and the endosperm. The *agl62* mutation causes premature cellularization of the endosperm and arrest of the seed coat development (86,87), further suggesting the dependency of seed coat development on endosperm development.

Figueiredo et al (87) investigated the endosperm-derived “switch” that initiates seed coat development after fertilization. Previously, genetic analyses showed that seed coat initiation was actively silenced by the Polycomb Repressive Complex 2 (PRC2) in the ovule integuments before fertilization (88). Mutations in PCR2 genes cause precocious seed coat initiation in the ovule by upregulating GA and auxin responsive genes. Exogenous applications of auxin and GA were sufficient to trigger seed coat development. While auxin application induced GA responses, GA did not induce auxin responses, suggesting auxin acts upstream of GA. Mutants deficient in auxin signaling did not affect seed coat development whereas mutants deficient in auxin biosynthesis showed the arrest of seed coat growth. In their previous study, they observed that auxin is produced in the endosperm after the fertilization (89). They also showed that the unfertilized wildtype Arabidopsis pistil was able to initiate endosperm development in the seeds with exogenous auxin application. In order to see if the auxin production in the endosperm was required to induce the seed coat development, they generated transgenic plants that ectopically expressed auxin biosynthesis genes in the central cell. Seed coat development was initiated in the ovule containing the transgene without fertilization, suggesting that auxin production is sufficient to start the seed coat development, not the presence of the endosperm itself. Taken together, these studies suggest that seed coat development in Arabidopsis is driven by Auxin production in the endosperm.

They further investigated the role of AGL62 and auxin in the seed coat development. They observed that the seed coat of an *agl62* mutant did not develop. Using the DR5::VENUS auxin reporter, they showed that VENUS signal, and therefore the presence of auxin, was detected in both endosperm and seed coat in the wild type plants, whereas the VENUS signal was only detected in the endosperm in *agl62* mutant, suggesting that auxin flow into the seed coat is required for the proper seed coat development (87).

The ovules from double-mutant plants in two PCR2 subunits, *VERNALIZTION2(VRN2)* and *EMBRYONIC FLOWER2(EMF2)*, were able to generate the seed coat without fertilization. Ectopic seed coat formation was further enhanced by introducing the dominant overexpressing auxin gene, *YUCCA6(yuc6-2d)*, into the *vern2/emf2* double mutant genetic background. The number of the ovules with ectopic seed coat formation was increased in the triple mutant of *yuc6-2d/ vern2/emf2* compared to the double *vern2/emf2* mutants. Furthermore, the expression of PRC2 subunits, VRN2, EMF2, MSI1 and SWINGER(SWN) were downregulaed after fertilization, releasing the development of the seed coat. In summary, auxin transport from the endosperm to the seed coat requires AGL62, and auxin influx into the seed coat represses PRC2-mediated silencing to initiate seed coat development.

Most soybean cultivars have yellow seeds because of *CHALCONE SYNTHASE (CHS)* gene silencing. CHS is the enzyme in the flavonoid pathway that leads to the biosynthesis of anthocyanins and proanthocyanidins. These compounds give wild soybean seeds the black pigmentation in the seed coat (90). Tuteja et al (91) investigated the regulation of seed coat pigmentation by using different soybean cultivars with different alleles at the *I* locus that regulates CHS genes. The dominant form of *I* gives the yellow seed coat whereas soybean with the recessive form has a dark-colored seed coat. The *I* locus was identified by deletions at this locus that has two separate inverted cluster cassettes consisting of three *CHS* genes, *CHS1, CHS3, and CHS4. CHS* are multi-gene families with nine members. A study suggested that *CHS* is regulated post-transcriptionally by CHS-derived small interference RNAs (siRNAs) (92). In their study, *CHS*-derived sRNAs were detected only in the seed coat with the dominant *I* allele and not in the seed coat of soybean with the recessive *i* allele. Furthermore, siRNAs were not detected in any other tissues other than the seed coat of soybean with either alleles, suggesting that CHS sRNAs accumulate only in soybean seed coat with the dominant *I* allele in a tissue-specific manner. sRNA-seq further confirmed that sRNA sequences were mapped to all the *CHS* gene family genes, particularly to *CHS7* and *CHS8* genes of the dominant *I* allele, not with the recessive *i* allele, suggesting the yellow seed coat color was caused by siRNA silencing of CHS genes. On the other hand, the recessive *i* allele carries a deletion in one of the inverted *CHS1/CHS3/CH4* cluster cassettes. They speculated that the inverted *CHS1/CHS3/CH4* cluster cassettes may generate the double-stranded RNAs into sRNAs. Thus, the deletions of this cassettes may prevent sRNA production so that *CHS* genes are not silenced. However, this speculation cannot explain the complete absence of siRNAs in the seeds with the recessive *i* allele.

The same group later identified another locus *K* that is epistatic to *I* locus (93). A mutation in the *k* locus causes partial pigmentation (the "saddle" pattern) in the dominant *I* alleles. They profiled *CHS*-derived sRNAs at all *CHS* genes with sRNA-seq data from the seed coat tissues at the pigmented and non-pigment sections. All of the *CHS* loci, particularly at *CHS7* and *CHS8* accumulate high levels of sRNAs in the non-pigmented seed coat section, but not in the pigmented section. The mRNA profiles in the same tissues show the opposite profiles from sRNA profiles, suggesting *CHS* gene repression is caused by *CHS*-derived sRNAs. Whole genome sequencing data identified the recessive *k* allele as a deletion of the *ARGONAUTE 5 (AGO5)* gene. *AGO5* is closely related to *AGO1*, which is the effector gene that mediate miRNA regulation (16). AGO5 is mainly associate with repeat-associated siRNAs as well as sRNAs generated from protein-coding genes (94). In order to further confirm deletions in *AGO5* are responsible for the pigmentation, they compared different soybean varieties with the saddle pattern in the seed coats. All of the varieties with the saddle patterns have different deletions in AGO5. These results showed the sRNA pathway play an important role for the seed coat color in a tissue-specific manner.

Genome-wide examination of gene expression in seed development

In the early 2000's, genome-wide technologies to study gene expression using the GeneChip (95) and Next Generation Sequencing (96) became widely available to monitor gene expression in different tissues and developmental stages. Biological processes are influenced by expressed genes and their level of expressions. Thus, genome-wide studies give a comprehensive view of gene expressions.

Schmid et al (97) analyzed global gene expression datasets from 79 *Arabidopsis thaliana* samples with Affymetrix ATH1 arrays, including whole seeds samples collected at different developmental stages. In seed samples, they observed strong temporal opposing gene expression modules between earlier and later seed stages, which coincided with morphogenesis and maturation phases.

In order to get more insight into comprehensive gene expression profiles during the seed development, Belmonte et al. (98) profiled mRNA populations at the single subregion level from five different developmental stages of *Arabidopsis thaliana* seeds with Laser-capture microdissection (LCM) and Affymetrix ATH1 GeneChip hybridization analysis. They used the principal component analysis (PCA) to compare the global mRNA profiles from 31 combinations of subregions and stages. There were the four major clusters: a) a cluster for all of the seed coat subregions at different stages, b) a cluster for the embryo proper (EP), micropylar endosperm (MCE) and peripheral endosperm (PEN) subregions at the earlier stages, c) a cluster for the embryo proper (EP), micropylar endosperm (MCE) and peripheral endosperm (PEN) subregions at the later stages and d) a cluster for the chalazal endosperm (CZE) subregion at different stages. The differences in mRNA profiles were further supported by GO term enrichment analyses of the clusters. The functions of mRNAs that accumulated specifically in the EP, MCE, and PEN, at the mature green stage are related to the seed maturation (e.g., nutrient reservoir activity, lipid storage and seed oil biogenesis). In general, eudicots accumulate storage molecules in the cotyledons of the embryo since the endosperm largely degenerates by the maturation phase. It was surprising to observe lipid storage functions in the endosperm subregions at the later stage. In addition, the set of TFs known to be involved in maturation processes accumulated specifically in the EP, MCE, and PEN, suggesting that the similar maturation processes occur in these subregions. As noted above, another endosperm subregion, the CZE, has a unique profile in the PCA analysis.

The number of genes specifically expressed in the CZE subregion was the largest, compared with the number of genes specifically expressed in other subregions. The genes that were only detected exclusively in the seeds, not other tissues, were highly enriched in the CZE. These observations suggest that the CZE is a developmentally unique subregion. They performed the GO representation analysis to gain insights into the function of CZE subregion-specific genes. The overrepresented GO terms were ubiquitin-dependent protein catabolism, suggesting CZE subregion-specific genes may have a posttranscriptional regulatory role in the CZE. The

CZE was also enriched with the rate-limiting enzymes for the biosynthesis of plant hormones (i.e., gibberellic acid, abscisic acid and cytokinin), further suggesting the CZE may play a regulatory role in seed development. Taken together, mRNA profiling at a single tissue level provides significant insights into the developmental processes of Arabidopsis seed development.

In addition to Arabidopsis, studies of genome-wide mRNA profiles in seeds were conducted in *Medicago* (99), soybean (100), common beans (101) and maize seeds (102) . As seen in Arabidopsis genome-wide mRNA profiling studies (97,98), the approximate numbers of expressed genes do not greatly differ among the tissues compared. Moreover, the pattern of two temporal modules for morphogenesis and maturation phases was found in all species, suggesting that the overall temporal pattern of gene expression during seed development is common among these flowering plant species, including monocots.

Chen et al. (102) investigated mRNA profiles from maize whole seed, embryo and endosperm samples at 21 different time points using RNA-seq. They used PCA analysis to compare the overall mRNA profiles. The clusters were separated by tissue identities; all the embryo samples clustered together, and all the endosperm samples clustered together, which was different from Arabidopsis mRNA PCA profiles (98). Within each tissue cluster, the subclusters were delineated based on developmental stage. The unbiased hierarchical clustering showed the distinct temporal patterns between the embryo and endosperm. These data suggest that mRNA profiles demonstrated the distinct developmental processes in maize. In order to get insight into the cellular function based on the patterns of gene expression, they assigned the gene functions by using MapMan annotations. The functional categories for each cluster reflect the biological processes occurring in a tissue at the corresponding to the temporal phases, which is mostly common between maize and *Arabidopsis*. The major difference is storage molecule accumulation in the endosperm. The timing for storage molecule accumulation is similar to eudicots, but eudicots accumulate those molecules in the embryo, not the endosperm.

Taken together, the global mRNA profiles highlight the universal seed developmental processes that are fundamental for seed development.

Importance of elucidating gene regulatory network in soybean seeds.

Soybean is one of the most important agricultural crops in the world. Soybean, *Glycine max*, has been grown by humans for over four thousand years (103). High oil and protein content in the seeds make them suitable as a source of human and animal nutrition, biofuels, and various other commercial products such as adhesives, solvents, and candles (104) (https://ncsoy.org/media-resources/uses-of-soybeans/). Soybean seeds are also rich in other molecules that are beneficial to human health. For example, some clinical research indicated that isoflavone has anticancer and cholesterol lowering effects (105). Elucidating gene regulatory network in soybean seed development may help find the way to improve the yield and quality of soybean seeds.

My main research goal is to understand how genes are regulated in a spatio-temporal manner through different molecular mechanisms, specifically a) post-transcriptional gene regulations by miRNAs during soybean seed development, and b) TF regulations during the seed maturation phase. These studies may shed light on the complex gene regulatory network in soybean seed development.

Reference:

1. Bleckmann A, Alter S, Dresselhaus T. The beginning of a seed: Regulatory mechanisms of double fertilization. Vol 5, Frontiers in Plant Science. 2014.
2. Goldberg RB, Barker SJ, Perez-Grau L. Regulation of gene expression during plant embryogenesis. Vol 56, Cell. 1989. p 149–60.
3. Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C. Combined networks regulating seed maturation. Vol 12, Trends in Plant Science. 2007. p 294–300.
4. Angelovici R, Galili G, Fernie AR, Fait A. Seed desiccation: a bridge between maturation and germination. Vol 15, Trends in Plant Science. 2010. p 211–8.
5. Goldberg RB, De Paiva G, Yadegari R. Plant embryogenesis: Zygote to seed. Science (80-). 1994;266(5185):605–14.
6. Olsen OA. Endosperm development: Cellularization and cell fate specification. Annu Rev Plant Biol. 2001;52:233–67.
7. Haughn G, Chaudhury A. Genetic analysis of seed coat development in Arabidopsis. Vol 10, Trends in Plant Science. 2005. p 472–7.
8. Baud S, Boutin JP, Miquel M, Lepiniec L, Rochat C. An integrated overview of seed development in Arabidopsis thaliana ecotype WS. Plant Physiol Biochem. 2002;40(2):151–60.
9. Berger F. Endosperm: The crossroad of seed development. Vol 6, Current Opinion in Plant Biology. Elsevier Ltd; 2003. p 42–50.
10. Olsen OA. Nuclear endosperm development in cereals and Arabidopsis thaliana. Vol 16, Plant Cell. 2004.
11. Dute RR, Peterson CM. Early endosperm development in ovules of soybean, Glycine max (L) Merr. (Fabaceae). Ann Bot. 1992;69(3):263–71.
12. Schmidt MA, Herman EM. Characterization and functional biology of the soybean aleurone layer. BMC Plant Biol. 2018;18(1).
13. Moïse JA, Han S, Gudynaitę-Savitch L, Johnson DA, Miki BLA. Seed coats: Structure, development, composition, and biotechnology. Vol 41, In Vitro Cellular and Developmental Biology - Plant. 2005. p 620–44.
14. Weber H, Borisjuk L, Wobus U. Molecular physiology of legume seed development. Annu Rev Plant Biol. 2005;56:253–79.
15. Latchman DS. Transcription factors: An overview. Int J Biochem Cell Biol. 1997;29(12):1305–12.
16. Chen X. Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol. 2009;25:21–44.
17. Voinnet O. Origin, Biogenesis, and Activity of Plant MicroRNAs. Vol 136, Cell. 2009. p 669–87.
18. Goodrich J, Tweedie S. Remembrance of things past: Chromatin remodeling in plant development. Vol 18, Annual Review of Cell and Developmental Biology. 2002. p 707–46.
19. Ojolo SP, Cao S, Priyadarshani SVGN, Li W, Yan M, Aslam M, et al. Regulation of plant growth and development: a review from a chromatin remodeling perspective. Vol 9, Frontiers in Plant Science. 2018.
20. Boeva V. Analysis of genomic sequence motifs for deciphering transcription factor binding and transcriptional regulation in Eukaryotic cells. Vol 7, Frontiers in Genetics.

2016.
21. Biłas R, Szafran K, Hnatuszko-Konka K, Kononowicz AK. Cis-regulatory elements used to control gene expression in plants. Vol 127, Plant Cell, Tissue and Organ Culture. 2016. p 269–87.
22. Hobert O. Common logic of transcription factor and microRNA action. Trends Biochem Sci. 2004;29(9):462–8.
23. Wittkopp PJ, Kalay G. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Vol 13, Nature Reviews Genetics. 2012. p 59–69.
24. Farnham PJ. Insights from genomic profiling of transcription factors. Vol 10, Nature Reviews Genetics. 2009. p 605–16.
25. Hobert O. Gene regulation by transcription factors and MicroRNAs. Vol 319, Science. 2008. p 1785–6.
26. Wagner D. Chromatin regulation of plant development. Vol 6, Current Opinion in Plant Biology. 2003. p 20–8.
27. Grasser KD. The FACT Histone Chaperone: Tuning Gene Transcription in the Chromatin Context to Modulate Plant Growth and Development. Vol 11, Frontiers in Plant Science. 2020.
28. Jarillo JA, Piñeiro M, Cubas P, Martínez-Zapater JM. Chromatin remodeling in plant development. Vol 53, International Journal of Developmental Biology. 2009. p 1581–96.
29. Zhang P, Torres K, Liu X, Liu C, E. Pollock R. An Overview of Chromatin-Regulating Proteins in Cells. Curr Protein Pept Sci. 2016;17(5):401–10.
30. Zaret KS. Pioneer Transcription Factors Initiating Gene Network Changes. Annual Review of Genetics. 2020.
31. Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. Vol 37, Current Opinion in Genetics and Development. 2016. p 76–81.
32. Tao Z, Shen L, Gu X, Wang Y, Yu H, He Y. Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. Nature. 2017;551(7678):124–8.
33. Choi K, Kim J, Müller SY, Oh M, Underwood C, Henderson I, et al. Regulation of microRNA-mediated developmental changes by the SWR1 chromatin remodeling complex. Plant Physiol. 2016;171(2):1128–43.
34. Ueda M, Laux T. The origin of the plant body axis. Vol 15, Current Opinion in Plant Biology. 2012. p 578–84.
35. Park S, Harada JJ. Arabidopsis embryogenesis. In: Methods in molecular biology (Clifton, NJ). 2008. p 3–16.
36. Jenik PD, Gillmor CS, Lukowitz W. Embryonic patterning in Arabidopsis thaliana. Vol 23, Annual Review of Cell and Developmental Biology. 2007. p 207–36.
37. van der Graaff E, Laux T, Rensing SA. The WUS homeobox-containing (WOX) protein family. Vol 10, Genome Biology. 2009.
38. Ueda M, Zhang Z, Laux T. Transcriptional Activation of Arabidopsis Axis Patterning Genes WOX8/9 Links Zygote Polarity to Embryo Development. Dev Cell. 2011;20(2):264–70.
39. Haecker A, Groß-Hardt R, Geiges B, Sarkar A, Breuninger H, Herrmann M, et al. Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. Development. 2004;131(3):657–68.
40. Wu X, Chory J, Weigel D. Combinations of WOX activities regulate tissue proliferation during Arabidopsis embryonic development. Dev Biol. 2007;309(2):306–16.

41. Mayer U, Ruiz RAT, Berleth T, Miseéra S, Jürgens G. Mutations affecting body organization in the Arabidopsis embryo. Nature. 1991;353(6343):402–7.
42. Hamann T, Benkova E, Bäurle I, Kientz M, Jürgens G. The Arabidopsis BODENLOS gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning. Genes Dev. 2002;16(13):1610–5.
43. Hardtke CS, Berleth T. The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. EMBO J. 1998;17(5):1405–11.
44. Friml J, Vieten A, Sauer M, Weijers D, Schwarz H, Hamann T, et al. Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. Nature. 2003;426(6963):147–53.
45. Lu P, Porat R, Nadeau JA, O'Neill SD. Identification of a Meristem L1 Layer-Specific Gene in Arabidopsis That Is Expressed during Embryonic Pattern Formation and Defines a New Class of Homeobox Genes. Plant Cell. 1996;8(12):2155–68.
46. Lynn K, Fernandez A, Aida M, Sedbrook J, Tasaka M, Masson P, et al. The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis development and has overlapping functions with the ARGONAUTE1 gene. Development. 1999;126(3):469–81.
47. Bowman JL, Eshed Y. Formation and maintenance of the shoot apical meristem. Vol 5, Trends in Plant Science. 2000. p 110–5.
48. Moussian B, Schoof H, Haecker A, Jürgens G, Laux T. Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis. EMBO J. 1998;17(6):1799–809.
49. Zhu H, Hu F, Wang R, Zhou X, Sze SH, Liou LW, et al. Arabidopsis argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. Cell. 2011;145(2):242–56.
50. McConnell JR, Emery J, Eshed Y, Bao N, Bowman J, Barton MK. Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature. 2001;411(6838):709–13.
51. Vollbrecht E, Reiser L, Hake S. Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, knotted1. Development. 2000;127(14):3161–72.
52. Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS. KANADI regulates organ polarity in Arabidopsis. Nature. 2001;411(6838):706–9.
53. Siegfried KR, Eshed Y, Baum SF, Otsuga D, Drews GN, Bowman JL. Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. Development. 1999;126(18):4117–28.
54. Izhaki A, Bowman JL. KANADI and class III HD-Zip gene families regulate embryo patterning and modulate auxin flow during embryogenesis in Arabidopsis. Plant Cell. 2007;19(2):495–508.
55. Miyashima S, Honda M, Hashimoto K, Tatematsu K, Hashimoto T, Sato-Nara K, et al. A comprehensive expression analysis of the arabidopsis MICRORNA165/6 gene family during embryogenesis reveals a conserved role in meristem specification and a non-cell-autonomous function. Plant Cell Physiol. 2013;54(3):375–84.
56. White CN, Proebsting WM, Hedden P, Rivin CJ. Gibberellins and seed development in maize. I. Evidence that gibberellin/abscisic acid balance governs germination versus maturation pathways. Plant Physiol. 2000;122(4):1081–8.

57. Braybrook SA, Harada JJ. LECs go crazy in embryo development. Vol 13, Trends in Plant Science. 2008. p 624–30.

58. Weber H, Borisjuk L, Wobus U. Sugar import and metabolism during seed development. Trends Plant Sci. 1997;2(5):169–74.

59. 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. Proc Natl Acad Sci U S A. 2017;114(32):E6710–9.

60. To A, Valon C, Savino G, Guilleminot J, Devic M, Giraudat J, et al. A network of local and redundant gene regulation governs Arabidopsis seed maturation. Plant Cell. 2006;18(7):1642–51.

61. Meinke DW, Franzmann LH, Nickle TC, Yeung EC. Leafy cotyledon mutants of Arabidopsis. Plant Cell. 1994;6(8):1049–64.

62. Parcy F, Valon C, Kohara A, Miséra S, Giraudat J. The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of arabidopsis seed development. Plant Cell. 1997;9(8):1265–77.

63. Nambara E, Keith K, McCourt P, Naito S. A regulatory role for the ABI3 gene in the establishment of embryo maturation in Arabidopsis thaliana. Development. 1995;121(3):629–36.

64. 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. Vol 54, Plant Journal. 2008. p 608–20.

65. Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T. LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. Plant Cell Physiol. 2005;46(3):399–406.

66. Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J. 2009;58(5):843–56.

67. Mantovani R. The molecular biology of the CCAAT-binding factor NF-Y. Vol 239, Gene. 1999. p 15–27.

68. Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, et al. bZIP67 regulates the omega-3 fatty acid content of arabidopsis seed oil by activating fatty acid DESATURASE3. Plant Cell. 2013;25(8):3104–16.

69. Gnesutta N, Kumimoto RW, Swain S, Chiara M, Siriwardana C, Horner DS, et al. CONSTANS imparts DNA sequence specificity to the histone fold NF-YB/NF-YC Dimer. Plant Cell. 2017;29(6):1516–32.

70. Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, et al. LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proc Natl Acad Sci U S A. 2001;98(20):11806–11.

71. Baud S, Kelemen Z, Thévenin J, Boulard C, Blanchet S, To A, et al. Deciphering the molecular mechanisms underpinning the transcriptional control of gene expression by master transcriptional regulators in arabidopsis seed. Plant Physiol. 2016;171(2):1099–112.

72. Focks N, Benning C. Wrinkled 1: A novel, low-seed-oil mutant of arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant Physiol. 1998;118(1):91–101.

73. Cernac A, Benning C. WRINKLED1 encodes an AP2/EREB domain protein involved in

the control of storage compound biosynthesis in Arabidopsis. Plant J. 2004;40(4):575–85.
74. Baud S, Mendoza MS, To A, Harscoët E, Lepiniec L, Dubreucq B. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. Plant J. 2007;50(5):825–38.
75. Mu J, Tan H, Zheng Q, Fu Y, Liang Y, Zhang J, et al. LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiol. 2008;148(2):1042–54.
76. Nodine MD, Bartel DP. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev. 2010;24(23):2678–92.
77. Berger F. Endosperm: The crossroad of seed development. Vol 6, Current Opinion in Plant Biology. 2003. p 42–50.
78. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, et al. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell. 2002;110(1):33–42.
79. Huh JH, Bauer MJ, Hsieh TF, Fischer R. Endosperm gene imprinting and seed development. Vol 17, Current Opinion in Genetics and Development. 2007. p 480–5.
80. Bauer MJ, Fischer RL. Genome demethylation and imprinting in the endosperm. Vol 14, Current Opinion in Plant Biology. 2011. p 162–7.
81. Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, et al. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell. 2006;124(3):495–506.
82. Fourquin C, Beauzamy L, Chamot S, Creff A, Goodrich J, Boudaoud A, et al. Mechanical stress mediated by both endosperm softening and embryo growth underlies endosperm elimination in Arabidopsis seeds. Dev. 2016;143(18):3300–5.
83. Smýkal P, Vernoud V, Blair MW, Soukup A, Thompson RD. The role of the testa during development and in establishment of dormancy of the legume seed. Vol 5, Frontiers in Plant Science. 2014. p 1–19.
84. Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, et al. Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell. 1999;11(3):407–15.
85. Ohad N, Margossian L, Hsu YC, Williams C, Repetti P, Fischer RL. A mutation that allows endosperm development without fertilization. Proc Natl Acad Sci U S A. 1996;93(11):5319–24.
86. Kang IH, Steffen JG, Portereiko MF, Lloyd A, Drews GN. The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. Plant Cell. 2008;20(3):635–47.
87. Figueiredo DD, Batista RA, Roszak PJ, Hennig L, Köhler C. Auxin production in the endosperm drives seed coat development in Arabidopsis. Elife. 2016;5(NOVEMBER2016).
88. Roszak P, Köhler C. Polycomb group proteins are required to couple seed coat initiation to fertilization. Proc Natl Acad Sci U S A. 2011;108(51):20826–31.
89. Figueiredo DD, Batista RA, Roszak PJ, Köhler C. Auxin production couples endosperm development to fertilization. Nat Plants. 2015;1.
90. Senda M, Kurauchi T, Kasai A, Ohnishi S. Suppressive mechanism of seed coat pigmentation in yellow soybean. Vol 61, Breeding Science. 2011. p 523–30.
91. Tuteja JH, Zabala G, Varala K, Hudson M, Vodkin LO. Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in glycine max seed

coatsWOA. Plant Cell. 2009;21(10):3063–77.

92. Senda M, Masuta C, Ohnishi S, Goto K, Kasai A, Sano T, et al. Patterning of virus-infected Glycine max seed coat is associated with suppression of endogenous silencing of chalcone synthase genes. Plant Cell. 2004;16(4):807–18.

93. Cho YB, Jones SI, Vodkin LO. Mutations in Argonaute5 illuminate epistatic interactions of the K1 and I loci leading to saddle seed color patterns in glycine max. Plant Cell. 2017;29(4):708–25.

94. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, et al. Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5′ Terminal Nucleotide. Cell. 2008;133(1):116–27.

95. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo M V., Chee MS, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol. 1996;14(13):1675–80.

96. Thermes C. Ten years of next-generation sequencing technology. Vol 30, Trends in genetics : TIG. 2014. p 418–26.

97. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, et al. A gene expression map of Arabidopsis thaliana development. Nat Genet. 2005;37(5):501–6.

98. Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, et al. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc Natl Acad Sci U S A. 2013;110(5).

99. Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, et al. A gene expression atlas of the model legume Medicago truncatula. Plant J. 2008;55(3):504–13.

100. Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, et al. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biol. 2010;10.

101. O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, et al. An RNA-Seq based gene expression atlas of the common bean. BMC Genomics. 2014;15(1).

102. Chen J, Zeng B, Zhang M, Xie S, Wang G, Hauck A, et al. Dynamic transcriptome landscape of maize embryo and endosperm development. Plant Physiol. 2014;166(1):252–64.

103. Mujić A, Salkić B, Mešinović E, Buljubašić B. Agroeconomic impact indicators of various nourishment methods on the soybean crop during 2017. 2019;09(May):27428–32.

104. Pagano MC, Miransari M. The importance of soybean production worldwide. In: Abiotic and Biotic Stresses in Soybean Production. 2016. p 1–26.

105. Mateos-Aparicio I, Redondo Cuenca A, Villanueva-Suárez MJ, Zapata-Revilla MA. Soybean, a promising health source. Vol 23, Nutricion Hospitalaria. 2008. p 305–12.

Chapter 2

Profiling soybean miRNAs during soybean seed development.

Rie Uzawa[1], Julie Marie Pelletier[1], Tashinga Tsokodyi[2], Leonardo Jo[1], Ssu-Wei Hsu[1], Robert B. Goldberg[3] and John J. Harada[1]

[1]Department of Plant Biology, University of California, Davis, CA 95616 and Graduate Program in Plant Biology, University of California, Davis, CA 95616

Department of Agriculture and Natural Sciences, University of Maryland East Shore, MD21853

[2]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095

Abstract

Seeds consist of three regions, embryo, endosperm and seed coat, and each seed region is further differentiated into subregions. Each subregion undergoes unique temporal and spatial developmental programs that are mediated by changes in gene expressions controlled at the transcriptional and post-transcriptional level. Micro RNAs (miRNAs) are key regulators for gene expression. Although plant miRNAs have been extensively studied in recent years, little is known about miRNA accumulations and their functions at the single-tissue level during seed development. We identified 113 miRNA families originated from 174 miRNAs that are present in 37 soybean seed subregions at four different developmental stages by using Laser Capture Microdissection (LCM) coupled with small RNA sequencing (sRNA-Seq). About the half of miRNA families are conserved in different plant species while the rest are non-conserved miRNA found only in soybean. The majority of miRNAs accumulate primarily in one subregion or stage, suggesting that miRNAs may play important roles in regulating developmental processes in soybean seeds. I found that the endosperm subregions were enriched with non-conserved miRNAs. In order to understand miRNA functions, I identified 596 target mRNAs of soybean miRNAs using publicly available soybean parallel analysis of RNA ends (PARE) Sequencing (PARE-Seq) databases. The target analysis suggested that non-conserved miRNAs present in the endosperm including the novel miRNA may be involved in abiotic and biotic stress responses. miRNA profiling at the single tissue level gives insights into the functions of miRNAs, including non-conserved miRNAs.

Introduction

The beginning of seed development occurs at double fertilization. Prior to fertilization, the megagametophyte, surrounded by the ovule integuments, contains the egg cell and the central cell. The egg and the central cells are each fertilized by a sperm cell giving rise to the diploid embryo and triploid endosperm, respectively (1). Fertilization also initiates seed coat development from the ovule integuments (2). Each seed region, the embryo, endosperm and seed coat, undergoes a series of cell divisions and differentiations to establish organs, tissues, and cell types, called subregions. The first division of the zygote generates the embryo proper and the suspensor, the terminally differentiated tissue that provides nutrients for the embryo proper. The embryo proper further establishes the basic plant body plan during the earlier stages of seed development, called the morphogenesis phase. The morphogenesis phase is followed by the maturation phase when the embryo expands by accumulating storage molecules such as storage proteins and lipids before entering the seed maturation and desiccation (3). The central cell nucleus undergoes a series of mitosis without cell division shortly after double fertilization. The daughter nuclei migrate from the micropylar to charazal ends of the endosperm. Once the syncytial divisions are over, the endosperm is cellularized and fills up the entire endosperm cavity. As the embryo proper grows it absorbs the endosperm, and only a few layers of endosperm tissues remain at the beginning of the maturation phase (4,5). The seed coat starts differentiating from the ovule integuments upon double fertilization. The seed coat consists of five different subregions: the endothelium, the inner integuments, outer integuments, hilum and the epidermis. Before the embryo enters maturation phase, the layers in the endothelium and the inner integument are crushed and degraded (6). In summary, each seed region undergoes distinct spatial and temporal developmental processes, which are highly coordinated.

The coordination of developmental processes is often controlled by the regulation of gene expression. An important posttranscriptional mechanism for gene regulation involves micro RNAs (miRNAs). miRNAs are a class of small non-coding RNA, 20-24 nt in length, that negatively regulate target gene expression.

miRNAs guide the RNA-induced Silencing Complex (RISC) that cleaves the target mRNA or inhibits its translation in a sequence-specific manner (7,8). miRNAs are key regulators of seed development. For example, in Arabidopsis, loss of function mutations in *Dicer-like RNase III endonucleases 1(DCL1*), one of the key genes for miRNA biogenesis that process miRNAs from primary transcripts, arrest embryo growth and promote the precocious onset of the photosynthesis and embryo maturation (9,10). Members of the SQUAMOSA BINDING PROTEIN-LIKE (SPL) transcription factor family, were identified as miR156 targets (11–13). Nodine et al. (9) showed that upregulation of storage protein genes was caused by overaccumulation of SPL10 and SPL11 due to mutations in the miR156 binding sites. Another miRNA, miR160 is known to target *ARF10, ARF16* and *ARF17*, which are the members of AUXIN RESPONSE FACTOR (ARFs) transcription factors required for auxin signaling (14). miR160-resistant ARF10 that contained mutations at the miR160 binding sites showed poor germination rates due to the increase in sensitivity to the seed maturation-promoting hormone, abscisic acid (ABA). In contrast, overexpression of miR160 caused decrease in ABA sensitivity (15). Thus, miRNAs are involved in various biological processes in seed development.

miRNAs are classified as conserved and non-conserved miRNAs. Conserved miRNAs are detected in a wide range of plant species. Some conserved miRNAs are present in most or all land plant lineages. Those highly conserved miRNAs are abundant but low in sequence variation across plant species (16,17). The highly conserved miRNAs families are expanded with several family members, which often accumulate in different tissues or domains of the tissues (18–21) . In many cases, conserved miRNAs regulate target mRNAs encoding regulatory protein such as transcription factors and F box proteins that are involved in various developmental processes, signal transductions and stress responses (22–24). Furthermore, conserved miRNAs have mRNA targets that are themselves conserved among plant species. That is, the sequences of target mRNAs where miRNAs binds are conserved as well (22). Extensive conserved miRNAs and their mRNA targets indicates that gene regulation by these miRNAs play important roles in plants.

Non-conserved miRNAs are, by contrast, observed in a single plant lineage or species. These miRNAs are usually single copy miRNAs. The number of non-conserved miRNAs outnumbers the number of the conserved miRNAs. Non-conserved miRNAs are usually low in their abundance (17,25). Target mRNAs of non-conserved miRNAs are often difficult to identify by *in silico* or experimental approaches (22,26). Moreover, many non-conserved miRNAs have been shown to be non-functional based on comparative analyses using miRNA biogenesis mutants in *Arabidopsis*. That is, the level of the target mRNAs of the conserved miRNAs generally increased in the mutants whereas no changes were observed in many target mRNAs of the non-conserved miRNAs (27). Together, these observations suggest a high birth and death rate of non-conserved miRNAs (16,22,23,25). However, some targets of non-conserved miRNAs have been validated and shown to be functional (26,27). The functions of target mRNAs for non-conserved miRNAs are proteins with diverse functions involved in a broad range of biological processes (16,23,27).

miRNAs have been extensively studied mainly in Arabidopsis tissues. However, little is known on miRNA functions in soybean seeds despite the fact that the soybean is an important crop in the world. Our group successfully generated sRNA libraries from each seed subregion at four different soybean developmental stages. These libraries were then used to profile miRNA populations. My goal for this miRNA study is to shed light on miRNA functions at a single tissue level to understand how miRNAs are involved in the gene regulatory network for spatial and temporal development of soybean seeds.

Results

Evaluation of miRNA candidates

In order to understand how miRNAs are involved in gene regulation for soybean seed development, we profiled small RNA populations by small RNA sequencing (sRNA-Seq) from 6-17 subregions at the four different developmental stages using Laser Capture Microdissection (LCM). We collected tissues from the embryo proper and suspensor subregions of the embryo region, the endosperm subregion, and the endothelium, inner and outer integuments, hilum and epidermis subregions of the seed coat region at the globular, heart, cotyledon and early maturation stages (Figure 2.1-A, Table 2.1). By the time the seed enters the early maturation stage, the embryo proper has established the axis and cotyledons, and completed the pattern formation program. Because the embryo has grown larger, more specialized subregions can be isolated. Thus, we were able to construct libraries from seven subregions and five subregions from the axis and cotyledon, respectively. The early maturation seed coat palisade is differentiated from the seed coat epidermis, whereas the early maturation seed coat hourglass and parenchyma are derived from the outer integuments. The inner integument tissue is completely compressed by the end of the cotyledon stage (28). The suspensor and the endothelium subregion samples were only obtained at the heart stage due to the technical difficulties. Small RNA (sRNA) populations in every subregion of the soybean seed were sequenced in at least one developmental stage. In total, 74 libraries from six to seventeen subregions at four stages with two biological replicates was constructed and sequenced. (GEO accessions: GSE57845, GSE57874, GSE57883, GSE57906, sequencing summary in Table S2.1). After removing the reads matching rRNA and tRNA sequences, sRNA reads of 18 to 26 nucleotides in length that were perfectly aligned to the soybean genome (Wm82.a2.v1/Gmax275) were retained. We chose 18 to 26 nucleotides as the range of miRNA length because this range matches plant miRNA lengths listed in miRBase v22 (http://www.mirbase.org) (29). The total number of miRNA sequences in each biological replicate ranged from 2.4 to18 million (Table S2.1). The abundance of each sequence in a biological replicate was normalized to sRNA counts per million (CPM) adjusted by the depths of sequencing from all the biological replicates (30).

Reads from sRNA-Seq contain different classes of sRNAs such as small interfering RNAs (siRNAs) other than miRNA (31). In order to filter non-miRNA populations, I evaluated miRNA candidates by adapting published miRNA criteria (32) and methods (33) based on the characteristics of miRNA biogenesis (Figure S2.1, Material and Methods).

miRNAs are generated by the precise excision of the miRNA and complementary miRNA star (miRNA*) duplex from a stemloop precursor RNA (32). Both miRNA and miRNA* can become a functional miRNA by being incorporated into a RNA-induced Silencing Complex (RISC) that cleaves its target mRNAs or inhibits their translation. However, miRNA* is less frequently incorporated into the RISC. This unique miRNA biogenesis process can be utilized to sort out miRNAs from other sRNA populations. First, the formation of a stemloop structure can be inferred by the presence and close proximity of the miRNA and its complementary sequence. Second, we examined the degree of the sequence complementarity between each miRNA and its miRNA*. Third, the dominance of miRNA and miRNA* sequence enrichment was analyzed at the stemloop genomic loci. Fourth, we determined whether both miRNA and miRNA* are derived from the same strand at the stemloop genomic loci. The strand bias is not observed at loci where other sRNAs are generated.

The miRNA candidates include miRNAs listed in miRBase v22 (29), miRNAs identified and published after the miRBase v22 release (Table S2.2), and novel miRNA candidates that our group identified using a miRNA discovery pipeline (34). As shown in Figure S2.1, a total of 174 individual miRNAs were identified from over 230,000 miRNA candidates from soybean seeds. Among the 174 miRNAs, 109 miRNAs are identified in two or more species and are designated as conserved miRNAs, whereas 65 miRNAs are uniquely identified in soybean and are designated as non-conserved miRNAs. Three non-conserved miRNAs were discovered previously by our group (34).

In order to assess the quality of the data, I compared the read counts of each miRNA between two biological replicates in each subregion (Figure S2.2-A, B and Table S2.3). The correlation coefficients between two biological replicates showed the strong positive correlation (more than 0.8) between two biological replicates, verifying the integrity and reproducibility of our sRNA-Seq libraries.

miRNAs can be grouped into families based on sequence similarities of their stemloops and miRNA sequences (35). The expansion in the number of miRNA family members was derived from a process of genome-wide duplication events (36). Most of the miRNA family members target the same target mRNAs (21,37). Thus, I merged the miRNA read counts within the miRNA families based on the sequence similarities of miRNA. As a result, 174 individual miRNAs were grouped into 113 miRNA families (Table 2.2). Out of 113 miRNA families, 54 miRNA families are conserved and 59 miRNA families are non-conserved.

Little variation is observed in the number of miRNA families in seed subregions

In Arabidopsis, some mature miRNAs have been shown to accumulate in specific tissues at certain developmental stages to maintain proper embryo developmental processes (38–40). Thus, the number of miRNAs may be different from one subregion to another in different stages. I compared the numbers of miRNA family detected in each subregion (Figure 2.1.-B). Overall, the numbers of miRNA families did not show large differences among the subregions although some variation was observed. Differences in the numbers of miRNA family may be influenced by spatial or/and temporal changes. Alternatively, differences in the numbers may depend on the depth of sequencing. To address these questions, I sampled the same numbers of the reads from all the libraries to compare the numbers of miRNA families detected in all the subregions. If there were no spatial or/and temporal effects on the numbers, all the subregions would have the similar numbers of miRNA family. The profiles for their numbers were similar to the original samples (Figure S2.2-C), suggesting spatial or/and temporal effects may contribute to some of the differences in numbers. In order to examine the effect of the sequence depth, I compared the number of miRNAs with total sRNA numbers in each sample. As shown in Figure S2.2-D, the number of miRNAs has a modest positive correlation with the depth of sequencing. Taken together, differences in the numbers of miRNA families may depend on spatial and temporal variation as well as the depth of sequencing.

Distinct spatial and temporal patterns of miRNA accumulation

miRNAs have been shown to accumulate at different levels in different tissues and developmental stages in plants (41). To observe the dynamics of global miRNA accumulation during soybean seed development, I examined the distributions of miRNA family accumulation in each subregion (Figure 2.2-A). The overall distributions of miRNA family accumulation did not show large differences. The distributions for the outer integument at the cotyledon stage and the suspensor at heart stage were significantly higher and lower than the total miRNA distribution (Pairwise Wilcoxon Rank Test, p values < 0.05), respectively, suggesting at least some spatial effect on global miRNA family accumulation.

The accumulation of individual miRNA families changes during soybean seed development. In order to get more insight into the global spatial and temporal changes in accumulation patterns of miRNA families, I used the unbiased hierarchical clustering method (Figure 2.2-B and C). It has been reported that non-conserved miRNAs tend to accumulate in limited tissues or organs whereas many conserved miRNAs accumulated more broadly (42,43). The categories of soybean miRNA conservation are indicated alongside the heatmaps for the hierarchical clustering (Figure 2.2-B and C, the bars with green and blue colored boxes) as are the levels of miRNA families (Figure 2.2-B and C, the bars with orange and red colored boxes). miRNA families with a low level of accumulation (cpm <1) in all the subregions of each heatmap were not included.

Many miRNA families primarily accumulated in one subregion or stage, indicating that the accumulation of miRNA family may be influenced by spatial and temporal cues. The numbers of conserved miRNA families either exceed or are similar to the numbers of non-conserved miRNA families in most of the clusters. However, miRNA families that predominately accumulated in the endosperm subregion at the early maturation stage contained mostly non-conserved miRNA families (Figure 2.2-B). Similarly, miRNA families that predominately accumulated in the endosperm subregion at all four stages contained mostly non-conserved miRNA families (Figure 2.2 C). Thus, non-conserved miRNA families account for much of the miRNA families that accumulate in the endosperm subregion, particularly at later stages. These results suggested that non-conserved miRNA family may play an important role for spatial and temporal development of the endosperm.

Those miRNA families specifically accumulated in one subregion or stage may be functionally important for soybean seed development. In order to identify subregion- or stage-specific miRNA families that accumulated at statistically significant levels, I adapted the method described in Belmonte et al (44). Briefly, subregion specific miRNA families are defined as the ones that accumulated fivefold or higher in only one subregion at one stage with a statistically significance q value <0.001 (Anova model with edgeR R package (45), whereas stage specific miRNA families are defined as the ones that accumulated at one stage relative to any other stages in one subregion with the same criteria described above. As a result, 17 subregion-specific and 7 stage-specific miRNA families were identified, respectively. (Table 2.3). The same miRNA families were identified as subregion- or stage-specific miRNAs for multiple subregions or stages. For example, miR171-5p was identified as a subregion-specific miRNA family in epidermal tissues in the early maturation embryo subregions (SAM, plumule and epidermis of axil, adaxial and abaxial epidermises of cotyledon) as well as a stage-specific miRNA family that accumulates in the cotyledon stage embryo proper, suggesting miR171-5p may play an important role for the epidermal tissues of the embryo at the later seed developmental stages. miR4414-5p may be a key spatial and temporal regulator for the embryo axis stele at the early maturation stage because it was identified as both a stage- and subregion-specific miRNA family for the subregion. Furthermore, four miRNAs: miR5380, miR5770, miR9740, and miR9753 were identified as endosperm-specific miRNAs at two or three developmental stages, suggesting that these miRNAs may contribute to the endosperm developmental process. Interestingly, all the subregion-specific miRNA families identified in the endosperms were non-conserved miRNAs, further supporting non-conserved miRNA family enrichment in the endosperm subregions observed in the hierarchical clustering analysis.

Taken together, many miRNA families accumulate specifically in a single cell type or tissue or at a specific developmental stage.

Elucidating the functions of miRNAs by studying target mRNAs in soybean seed development.

The accumulation patterns of miRNA families suggests that miRNAs may regulate spatial and temporal soybean seed development. In order to understand more clearly the functions of these miRNAs, it is important to identify potential target mRNAs with complementary sequences. I analyzed the publicly available datasets of parallel analysis of RNA ends (PARE) for high-throughput identification of miRNA targets (PARE-Seq) (46). I analyzed 20 different PARE-Seq datasets (Table S2.4) from 14 different soybean samples using the sPARTA target identifying tool (47), I identified 596 mRNAs as the targets for 84 miRNA families (603 miRNA-target pairs in total). Several studies showed that the major target mRNAs for the conserved miRNAs encode transcription factors, whereas non-conserved miRNAs were known to target the mRNAs encoding the proteins involving a wide range of biological processes (reviewed in (8,23,48)). In order to get insights into the target mRNA functions for soybean seed miRNAs, I examined the annotated functions for the target mRNAs. Out of 596 target mRNAs identified in my analysis, 493 and 106 mRNAs were cleaved by conserved and non-conserved miRNA families, respectively. Among 596 target mRNAs, seven mRNAs were identified as targets of two different miRNA families. Of these seven target mRNAs, three mRNAs were identified as both conserved and non-conserved miRNA targets. The percent of target mRNAs encoding transcription factors relative to the total target mRNAs for conserved miRNAs was much higher (26.6%) than the percent for non-conserved miRNAs (1%), agreeing with the previous reports (Table 2.4).

PARE-Seq can identify potential target mRNAs. However, more evidence is needed to show that the miRNA cleaves the mRNA. In order to examine the effects of miRNA regulation on its target mRNA, the relationship between the levels of a miRNA family and its target mRNA was globally compared. If the miRNA cleaves the target mRNA, there should be a negative correlation between accumulation of the miRNA and accumulation of its target mRNA. The relationships were indicated by using Spearman's rank correlation (Figure 2.3-A as a representation). Target mRNA expression data were previously generated in our group by LCM followed by RNA-Seq (GEO accessions: GSE57349, GSE57350, GSE57606, GSE46906). miRNAs and target mRNAs with no or extremely low accumulations were excluded (cpm < 1). As a result, 518 miRNA family-target mRNA pairs were examined. I classified miRNA-target mRNA pairs into the five categories based on correlation coefficients with arbitrary thresholds: a) strong negative correlation (less than -0.5), b) weak negative correlation (between -0.2 and -0.5), c) no correlation (between -0.2 and 0.2), d) weak positive correlation (between 0.2 and 0.5), and e) strong positive correlation (more than 0.5). (Table 2.5) Although a fraction of the pairs (27 %) showed strong or weak negative correlations, the majority (48.6%) showed no correlations, and some pairs (24.3%) showed either weak or strong positive correlations. The small proportion of the pairs that exhibited negative correlation values suggests that an inversed relationship between miRNA and their target mRNA levels was not very common.

Next, I asked if the relationship between the levels of a miRNA family and its target mRNA may be different for conserved and non-conserved miRNA families. The correlation coefficient distributions were statistically similar between conserved and non-conserved miRNA families (Student t-test P = 0.27, Figure 2.3-B). Interestingly, the pairs with strong negative correlation were all for conserved miRNAs (Figure 2.3-B). The majority of the pairs with strong negative correlation were for miR156, miR167, miR169 and miR396 (Figure 2.3-C), suggesting a small subset of conserved miRNAs, particularly these four miRNAs, may play a critical role to determine the expression of target mRNAs in the soybean seeds.

In order to confirm if a miRNA cleaves its target mRNA *in vivo*, I generated miRNA target "sensor" constructs to detect endogenous miRNA activities in early maturation embryo protoplasts (49). Briefly, the construct contains two fluorescent proteins - GFP with the 3' end of the ORF fused to the 21nt target mRNA sequence in frame and mCHERRY with no target sequences that serves as a transfection control. GFP fused to a miRNA target sequence and mCHERRY were both overexpressed in protoplasts by the 35S promoter. The construct was transfected into protoplasts (Figure 2.4-A). The expression levels of GFP were measured relative to the control expression level of mCHERRY.

I tested three different classes of target mRNAs sequences for the sensor constructs based on my target analysis: a) two target mRNA sequences for previously identified miRNA family/target mRNA pairs whose miRNAs are accumulated at the high level in the early maturation embryo (*SPL2/9* sequence for miR156 family (50,51), and *MYB33/65* sequence for miR159a-3p,miR319 family (52)), b) two target mRNA sequences for the previously identified family/target mRNA pairs whose miRNAs are accumulated at a relatively low level in the early maturation embryo (*NAC100* sequence for miR164 family (53), and *HAIRY MERISTEM3 (HAM3)* sequence for miR171 (b-3p,c-3p,d) family (54)) and c) two target mRNA sequences for the novel miRNA family/target mRNA pairs whose miRNAs are accumulated at a relatively high level in the early maturation embryo (*BRAP2 RING ZNF UBP DOMAIN-CONTAINING PROTEIN (BRIZ2)* sequence for miR165-5p family, and *REPRESSOR OF SILENCING1(ROS1)* sequence for miR398/ miRPubNew170 families family) (Figure 2.4-B and Table S2.5). As controls, I used two constructs for each target sequence in which the target sequence was mutated and a construct without the miRNA target sequences. Each transfection was repeated twice for the reproducibility.

I observed reduction of GFP signal for the protoplasts containing the constructs for *SPL2/9* sequence (miR156 family) and *MYB33/65* sequence (miR159a-3p,miR319 family) relative to the protoplasts with the negative control constructs (Figure 2.4-C). Interestingly, the miR159a-3p,miR319 family and *MYB33/65* pair showed no correlation in the analysis for the global comparison of miRNA-target mRNA levels mentioned in the previous section (Table S2.5). This may indicate that the role of miR159a-3p,miR319 family may be to maintain the level of *MYB33/65* gene expression at a constant level.

The results for the other constructs were either inconclusive or no change was observed (Figure 2.4-D and E). The protoplasts with the constructs with *HAM3, BRIZ2* and *ROS1* sequences showed inconsistent results between biological replicates (Table 2.6). The protoplasts containing construct with *NAC100* sequence showed no reduction of GFP signals relative to the protoplasts with the negative control constructs. No reduction of GFP signal of protoplasts with *NAC100* sequence containing construct remain the same with the protoplasts when I replace the

35S promoter for GFP to less robust promoter OLE1. These results suggest that downregulating target mRNAs may require for very high level of miRNAs. It is also possible that the overexpression of target sequence may overwhelm the endogenous miRNAs to repress GFP. It would be informative to perform the same experiments with the protoplasts co-transfected by the sensor constructs and the miRNAs driven by a 35S promoter to see if reduction of GFP signals can be detected.

The subregion-specific miRNA families for the endosperm may be involved in biotic and abiotic stress responses.

As I described in the previous section, I identified 21 miRNA families as stage- or/and subregion-specific miRNA families (17 subregion-specific and 7 stage-specific miRNA families). Interestingly, the endosperm-specific miRNA families at the heart, cotyledon and early maturation stages were all non-conserved miRNAs. Non-conserved miRNAs generally accumulate at a low level and are often non-functional (25,27,31). Since the five endosperm-specific miRNA families accumulated at the moderate or high levels in the endosperm subregions (Figure S2.3), I further examined the underlying roles of these miRNA families for the endosperm subregions at the heart, cotyledon and early maturation stages.

Due to the challenges of transforming soybean plants (55), I utilized the LCM RNA-seq datasets described in the above. The hierarchical clustering analysis was performed on the 15,000 most varying mRNAs (56) with all miRNA families at the heart, cotyledon and early maturation stages (Figure 2.5). I performed GO term representation analysis of the gene clusters containing the five endosperm-specific miRNA families: cluster 6 at the heart stage, the cluster 3 at the cotyledon stage and the cluster 11 at the early maturation stages. The overrepresented GO terms for these clusters were related to abiotic and biotic stress response and defense (Table 2.7). Of the five endosperm-specific miRNA families, target mRNAs for miR5770, miR9740 and miRC14 were identified in my target mRNA analysis as described in the previous section. No target mRNAs were identified for gma-miR5380 and miR9753 in the same analysis (Table 2.8).

Surprisingly, the annotated gene functions for target mRNAs of the endosperm-specific miRNA family were also related to abiotic and biotic stress responses and defense. As shown in Table 2.8, the target mRNAs for the miR5770 family include copper amine oxidase family proteins (Glyma.01G062400, Glyma.17G019300), and ClpC1 homologs (Glyma.04G200400, Glyma.06G165200). miR5770 was identified as the endosperm-specific miRNAs at the cotyledon and early maturation stages. The Arabidopsis Copper amine oxidase family protein is induced by methyl jasmonate (57), suggesting its involvement in plant defense. ClpC1 is required for iron homeostasis in Arabidopsis leaves. ClpC1 was upregulated under the iron-depleted condition. (58). Another target mRNA of miR5770, *Arabidopsis Epsin-like chlathrin adaptor1(AtECA1)* encodes a cargo protein that is involved in clathrin-coated vesicle formation (59). *AtECA1* can interact with phosphatidic acid under salt stress-induced condition in *Arabidopsis* (60). miRC14 was identified as the endosperm-specific miRNA family at the cotyledon stage. It was also identified as one of the novel miRNAs based on our group's novel miRNA discovery pipeline (34). One of target mRNAs for miRC14 is MIN7 (Glyma.17G044000), which encodes ADP ribosylation factor guanine nucleotide exchange factor protein (ARF-GEFs). MIN7 is required for the normal cuticle formation for Arabidopsis leaves to protect the plants from the bacterial pathogen, *Pseudomonas syringae* (61).

Interestingly, one of the enriched GO terms for the cluster 6 at the heart stage, the cluster 3 at the cotyledon stage and the cluster 11 at the early maturation stages was the promotion in miRNA production, suggesting the importance of miRNA regulation in the endosperm (Table 2.7). Mosher et al. (62) showed a subset of siRNAs accumulated from maternal chromosomes in Arabidopsis endosperm. Many siRNAs are generated from the transposable elements. These siRNAs are required to silence the same the transposable elements where siRNAs are originally generated (63). Hsieh et al. (64) suggested that siRNAs generated in the endosperm initiate non-cell autonomous silencing of homologous transposable elements in the embryo to reinforce the genome integrity. Thus, endosperm development may require two different classes of sRNAs to regulate different biological processes for the endosperm development.

Taken together, the endosperm-specific miRNA families at the heart, cotyledon and early maturation stages may play important roles for biotic and abiotic stress responses by maintain the gene expression levels of target mRNA.

Discussion

Previous studies showed that miRNAs are required for spatio-temporal development in seeds. Loss-of-function mutations in miRNA biogenesis mutant, *dcl1,* caused the defects in embryo morphology as well as the precocious onsets of photosynthesis and seed maturation processes in *Arabidopsis* seeds (9,10). The *dcl1* loss of function mutant of soybean also displayed the abnormal seed shape (65). Furthermore, accumulation of miRNAs is often restricted in certain tissues or to certain times ((66) reviewed in (7,8,48)). In order to comprehensively understand the roles of miRNAs in soybean seed development, we profiled miRNA populations in each subregion at the different developmental stages using LCM coupled with sRNA-Seq. Using a rigorous evaluation method, we identified 113 miRNA families from 174 individual miRNAs in 37 subregions at four developmental stages.

Many miRNA families accumulate predominantly in one subregion or stage.

Based on the hierarchical clustering analysis, the change of miRNA family accumulations is influenced by spatial and temporal cues. Some miRNA families predominately accumulated in a specific subregion or stage at the statistically significant level. This may suggest the importance of gene regulation by miRNAs in a specific subregion or stage. I identified 17 and 7 miRNA families as subregion or stage specific miRNA families, respectively. Some specific miRNAs are worth noting. For instance, miR171-5p was identified as a subregion-specific miRNA family in the subregions containing epidermal tissues in the early maturation subregions (SAM, plumule and epidermis of axil, adaxial and abaxial epidermises of cotyledon) as well as the cotyledon stage embryo proper. miR4414-5p was identified as both stage- and subregion-specific miRNA family for the embryo axis stele at the early maturation stage. The targets of miR171-5p and miR4414-5p are *CYTOCHROME P450, FAMILY 710 (CYP710A1*, Glyma.15G095000*)* and *tRNA METHYLTRANSFERASE 2A (TRM2A* Glyma.13G352500*)* were identified as the target mRNAs, respectively. However, the functions for these target mRNAs were not well-characterized in soybean and other plants. The *Arabidopsis* homologs of these target mRNAs are highly conserved in plants. It will be interesting to perform functional analyses in *Arabidopsis* by disturbing gene expressions such as target mimicry analysis (67) to observe the effects on the seeds since manipulating gene expressions in soybean is technically challenging.

A small number of target mRNAs may be important to control the expressions of target mRNAs.

PARE-Seq datasets were a great resource to identify target mRNAs of miRNA families that were cleaved by miRNAs. However, PARE-Seq cannot show the extent of miRNA involvement in gene regulations. Several modes of miRNA regulations have been reported. In *Arabidopsis*, miR166 and closely related miR165 were shown to regulate the target mRNAs by a dose-dependent manner to generate the expression threshold for the spatial development of the leaf and root, respectively (68,69). In rice, miR812q inversely regulated its target mRNA, *CBL-interacting protein kinase (CIPK10)* under the cold stress (33). Decrease in the level of miR156 caused increase in the level of target mRNAs*, SQUAMOSA PROMOTING BINDING PROTEIN-LIKE 3*, which promoted the transition to adult vegetative traits in leaves (12). Furthermore, miRNAs serve as buffering regulator to maintain the level of target mRNAs in *Arabidopsis* leaves and flowers (20,70,71). To explore how miRNAs regulate their target mRNAs, I globally compared the levels between a miRNA family and its target mRNAs in different subregions and stages.

A small subset of miRNA family-target mRNA pairs showed a strong negative correlation, suggesting those miRNAs are a major regulator controlling target mRNA accumulation. The majority of miRNA family-target mRNA pairs showed little or no correlation with their mRNA target. Some pairs even showed a strong positive correlation. The positive correlation seems counterintuitive to the general consensus that miRNAs are negative regulators of gene expression. These observations likely reflect the different modes of miRNA actions, including keeping target mRNA levels constant (Reviewed in (7,8,48)). Furthermore, target mRNAs may be regulated by translational inhibition, which might have little effect on target mRNA levels (72). Although miRNA-mediated target mRNA cleavage is a dominant form of miRNA regulation, translational inhibition has been shown to be more widespread in plant than previously thought (73).

A "gene expression buffer" mechanism for maintaining a target mRNA level constant or translational inhibition may explain the miRNA target sensor analysis results. MYB33/65 homologs were identified as miR159a-3p/miR319 family in our miRNA target identification analysis. The pairs of miR159a-3p/miR319-MYB33/65 homologues showed no correlation in the comparison of miRNA family-target mRNA levels (Table S2.5). However, I observed a strong GFP signal reduction in miRNA target sensor.

On the other hand, comparison of miR164 and NAC100 levels showed a strong negative correlation. However, no reduction of GFP signal was observed in the protoplast with miR164 sensor constructs. miR164 was shown to regulate *CUP-SHAPED COTYLEDON2 (CUC2),* one of NAC family member genes, in a tissue-specific manner in *Arabidopsis* flower (20) . In the same study, *CUC2* expression was upregulated in the carpel margin of miR164abc triple-mutant plants, but not in the stamens although miR164c expression was detected in the stamen margins, suggesting miR164 did not regulate *CUC2* expression in the stamens. Similarly, miR164 may be able to downregulate *NAC100* in the other soybean seed subregions, not embryo at the early maturation stage. The downregulation of target mRNAs by miR164 may be specific to a certain subregion.

Taken together, miRNAs may not be a major regulator for gene expressions of most target mRNAs. Rather, miRNAs may be required to maintain the level of target mRNAs.

Non-conserved miRNAs may play important roles to regulate stress responses in soybean endosperm subregions.

In the Results section, I showed that some miRNA families accumulated specifically in one subregion or stage at a significant level. Understanding the functions of the specific miRNAs are key to understand miRNA-mediate gene regulation in soybean seeds. The challenge to study miRNA functions in soybean plants is the availability of mutant and transgenic lines. In order to overcome this challenge, I combined 15,000 highly varying mRNAs and all miRNA families to perform the hierarchical clustering analysis to obtain clues about miRNA regulations by comparing the function of coexpressed genes and annotated functions for target genes of miRNAs.

I was particularly intrigued to find the five endosperm-specific miRNAs at the heart, cotyledon or early maturation stages. All of these miRNA families were only found in soybean. In general, non-conserved miRNAs accumulated at a low level without conferring functions (25,27). However, these endosperm-specific miRNAs accumulated at a high level in the endosperm (Figure S2.3), suggesting they may play important roles for the endosperm development.

GO terms related to biotic and abiotic stress responses were overrepresented for the genes that were coexpressed with these endosperm-specific miRNAs. The upregulations of the genes related to the defense against oxidative stress and fungi in the endosperm has been reported in germination Arabidopsis endosperm (74). Similarly, the endosperm in germinating tomato seed were enriched with the protein required for the virus defense (75). Jerkovic et al. (76) was able to micro-dissect the aleurone layer of the wheat endosperm to perform proteomic analysis. They found fungal defense protein in the aleurone layer specifically. Furthermore, Vendel et al. (77)

showed the shift of the protein profiles in wheat endosperm from the early to late filling stages. The early endosperm proteins were related to biosynthesis for nitrogen and amino acids whereas the late endosperm proteins are related to storage molecules as well as defense and stress responses.

miRNAs negatively regulate target mRNAs. Why did the enriched GO terms for the clusters agree with the annotated functions of target mRNAs? One possible answer may be that the miRNAs may serve a buffering role (7,8,48) to maintain a certain level of target mRNAs in the endosperm. Nikovics et al. (70) showed that both miR164 and miR164 target, *CUP-SHAPED COTYLEDON2 (CUC2),* one of NAC family member gene, were present in leaf margin of *Arabidopsis.* They also showed that the degree of leaf serration at the leaf margin were determined by the balance between miR164 and *CUC2* accumulation levels, suggesting the leaf phenotype depends on the level of *CUC2* that was maintained by miR164. In addition, the relationships of the accumulation level between a miRNA family and its target mRNA for the endosperm-specific miRNA families showed weak or no correlations (Table2.8), suggesting the endosperm-specific miRNAs accumulate in the same subregions as target mRNAs as well. Thus, the endosperm-specific miRNAs may control the level of target mRNAs in a similar way as miR164.

The endosperm fills the seed by the heart stage, then gradually degenerates as the embryo grows. (5,28). Finding GO terms related to biotic and abiotic stress responses in the clusters for the endosperm subregion was surprising because the biological functions of the endosperm is to nourish the embryo until the embryo becomes self-sustained. The endosperm of another legume, fenugreek, serves as a protective layer to prevent desiccation of the germinating embryo with galactomannan in the cell wall (78). Although soybean endosperm lacks galactomannan, it may have a function as a protective layer in addition to the seed coat since the endosperm is located adjacent to the embryo. Our miRNA study shed the lights on the underlying miRNA functions at a single tissue level. Further functional analyses will be needed to expand our understanding of miRNA-mediated gene regulations.

Materials and Methods

Plant Materials and Growth

Soybean cv William 82 plants were grown as described in Pelletier et al. (79)

Laser-Capture Microdissection

LCM processed samples were prepared and modified described in Belomonte et al (44). Whole seeds were collected and immediately fixed in 3:1 100% (vol/vol) ethanol:acetic acid fixative under RNAse-free condition. The material was vacuum-infiltrated for up to 1 hour, depending on size of tissue. The material was fixed overnight at 4 °C. The plant material was rinsed three times with 75% (vol/vol) ethanol, dehydrated in a graded ethanol series (75%, 85%, 95%, 100%, 100%, 100% ethanol), and infiltrated with xylenes (1:3, 1:1, 3:1 xylenes:ethanol, followed by 100% xylenes twice). Samples were incubated with paraffin chips overnight at room temperature, at 42 °C for 3 to 4 hours. The xylenes-paraffin mixture was replaced with 100% paraffin, and samples were infiltrated with change in paraffin for 2-3 days at 60 °C.

Seeds were sectioned and microdissected described as described in Belmonte et al.(44). Two biological replicates were captured for each subregion at each developmental stage. All subregions were captured within 1 moth of fixation to maximize RNA quality.

RNA extraction and library preparation

Total RNA was extracted as described in Pelletier et al (79). Sequencing libraries were prepared as described in the methods GEO accessions: GSE57845, GSE57874, GSE57883, and GSE57906. Briefly, the libraries were prepared using 250 nanogram of total RNA for each biological replicate except early maturation stage seed coat palisade biological replicate 1, both heat stage endothelium biological replicates and both heart stage suspenser biological replicates. The extracted RNA was quantified with Quant-IT^{TM} RiboGreen ® RNA reagent (Grand Island, NY) on a Nanodrop ND-3300 instrument (Thermo Scientific, Waltham MA).

Sequencing libraries were prepared using total RNA, using the illumine TruSeq Small RNA Sample Preparation Kit (cat no RS-200-0012 & RS-200-0024). Briefly, total RNA was ligated with 3' and 5' RNA adapter using T4 RNA ligase. Adapter-ligated small RNA was converted to complementary DNA and amplified by PCR for 15 cycles. Libraries were size-selected 147-160 bp and purified. Libraries were quantified with Quant-IT^{TM} PicoGreen dsDNA reagent (Grand Island, NY) on a Nanodrop ND-3300 instrument. Phi-Z174 was spiked in and to the library by the sequencing facility before cluster formation and sequencing.

Data processing

sRNA-Seq data were quality filtered. Quality-filtered reads were trimmed to remove adapter sequences and mapped to the Wm82.a2.v1 genome (Gmax275) using Bowtie v0.12.7 (80). Sequences that mapped perfectly to the Soybean genome were kept and filtered further by mapping against a reference containing soybean ribosomal RNA (rRNA) and transfer RNA (tRNA) using Bowtie. Sequences that did not map to rRNA and tRNA sequence were kept. The remaining filtered sequences with 18 to 26nt in length were counted (GSE57349, GSE57350, GSE57606, and GSE46906). The total number of miRNA sequences in each biological replicate ranged was normalized to sRNA counts per million (CPM) adjusted by the depths of sequencing from all the biological replicates (30). The normalized counts were averaged between biological replicates.

The names of miRNAs are based on the following categories. First, the name of miRNAs from miRBase v22 are followed miRBase convention as described in Griffin-Jones et al. (81). The names of miRNA with the exact same short sequences are pooled together in the parentheses. “NV” indicates “new variants” which have one or two bases offset from the original miRNAs on miRBase. “New” indicates the new miRNA sequences originated from the miRBase miRNA stemloops. Second, the name of miRNAs from the publications listed in Table S2.2 are indicated as “gma-PubNew”. These miRNAs have been published but not registered in miRBase v22. Third, we identified three novel miRNAs by our novel miRNA discovery pipeline: miRC14, novel-miR81 and nove-miR84, respectively.

miRNA evaluation

The total of 238830 miRNA candidates (569 miRNAs from miRBase V22 (29), 805 published miRNAs not registered in miRBase V22(Table S2.2), and 237634 miRNAs from our group’s novel miRNA identifying pipeline (34)) were screened to evaluate for miRNAs based on the criteria described in Axtell and Meyers (32). The procedure was adopted and modified from Jeong et al. (33).

The process for evaluating miRNA candidates involved six steps. The first step was to examine the stemloop structures of miRNA candidate. The miRNAs that contained the following structures were excluded; a) larger than 300 nts in size, b) any secondary stems, and c) any large internal loops with more than 5 bases at the stem. The structures of stemloop were screened either manually with Mfold web tool (82) or through the novel miRNA discovery pipeline (34).

The second step for evaluating miRNA candidates was detection of miRNA* in at least two biological replicates. miRNA* sequence was either adopted from miRBase v22 or manually curated. The presence of miRNA* sequences was then searched in our sRNA-seq datasets.

The third step for evaluating miRNA candidates was to screen the base mismatches in miRNA/miRNA* duplexes. The miRNA/miRNA* duplexes that contained more than five mismatches were excluded. The miRNA/miRNA* duplex that contained no more than five mismatches were also excluded if the duplex contains asymmetric bulges with three or more bases. The mismatches were manually screened.

The fourth step for evaluating miRNA candidates was to screen the enrichment of miRNA and miRNA* sequences at the corresponding stemloop genomic loci. miRNA and miRNA* sequences should be dominant and occupy more than 75% of total sequence counts at the stemloop genomic loci in least two biological replicates. These miRNA and mRNA* sequence must be derived from the same strand as the stemloop loci. The ratio calculation was executed with R. One-nucleotide positional variants of miRNA and miRNA* were also included for the ratio calculations as described in Axtell and Meyers (32).

The fifth step for evaluating miRNA candidates was the length of miRNAs. The miRNA length must be the range of 20 and 24 nt.

The sixth step for evaluating miRNA candidates was the minimum sequence counts for miRNAs. miRNAs must be detected more than 3 row counts for both biological replicates.

Data analysis

Unbiased hierarchial clustering

The hierarchical clustering of miRNA counts was performed using the ComplexHeatmap R Package (version 2.1.0) (83). The normalized and average miRNA counts were represented by z-score.

Identification of subregion and stage specific miRNA families

The method to identify subregion- and stage-specific miRNA family specific was adopted from Belmonte et al. (44). A miRNA family specific to a seed subregion was defined as one whose relative level is at least fivefold higher and significantly different (q <0.001*,* mixed-model ANOVA) than those detected in all other subregions at a given developmental stage. A miRNA family specific to a seed developmental stage was defined as one whose relative level is at least fivefold higher and significantly different (q <0.001*,* mixed-model ANOVA) than those detected in all other stages at a given seed subregion.

The subregions derived from the embryo proper and outer integuments at early maturation stage are further differentiated into 17 and 2 separate subregions, respectively. To identify subregion and stage-specific miRNAs for those 19 subregions, the subregions that are derived from the same embryo proper or outer integument subregion in the early stages were not compared with each other. For example, to identify the axis SAM-specific miRNAs at early maturation stage, the axis SAM subregion was only compared with endosperm and four seed coat subregions.

Identification of target mRNA of miRNA families

Genome-wide target of 113 miRNA families were identified by using publicly available soybean parallel analysis of RNA ends (PARE) Sequencing (PARE-Seq) databases listed in Table S2.4. Cleaved target mRNA fragments were identified by using sPARTA target identifying pipeline (47). sPARTA's built-in target prediction module *miRferno* was used to match the input miRNAs to the cleaved target mRNAs as described in Arikit et al (42). The standard scoring setting for miRNA/target mismatches was used and the cutoff were set less than or equal to 7. A cutoff threshold of a corrected P value that indicates the significant enrichment of fragment accumulation was set at 0.05. A read abundance threshold for cleaved fragment of target mRNAs was set more than or equal to 5.

Recombinant DNA manipulation and plasmid construction for miRNA sensor construct.

The 35S:GFP -35S:mCHERRY plasmid used for miRNA target sensor was constructed by inserting the 21nt target mRNA sequences at 3' end of GFP ORF in frame. For miR164 family and gma-miR171(b-3p,c-3p,d) target sensors, 35S promoter for GFP was replaced by OLE1(Glyma.20G196600) upstream region described in Jo et al. (84). For negative controls, the constructs with The 35S:GFP -35S:mCHERRY without the 21nt target mRNA sequences and with the mutated target mRNA sequences were used. 35S DNA sequence of the construct was confirmed by the Sanger method.

Transient assays with soybean embryo cotyledon protoplasts for mRNA target sensor

Plasmids with the miRNA target sensor construct were transfected into soybean embryo cotyledon at the early maturation stage (6-7 mm seeds in length) as described in Jo et al (84).

Briefly, cotyledons from early maturation soybean embryos were cut into 0.5-1mm strips, immersed in an enzyme solution containing 1% (w/v) Cellulase RS “Onozuka" and 0.25% (w/v) Macerozyme R-10 (Yakult Pharmaceutical Industry CO., Ltd). The cotyledon tissues were vacuum infiltrated for 15 min and incubated in the dark with gentle agitation (50 rpm) for 2 hours at room temperature. Protoplasts were filtered to remove unnecessary tissues, washed twice with W5 buffer (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2mM MES pH5.8, 5 mM glucose) and incubated on ice for 30 min. After the incubation, protoplasts (approximately 5 x10^5 cells per 200 ul) were transfected with 10ug of plasmid DNA. Transfected protoplasts were washed with W5 buffer twice and incubated in W5 buffer in the light at 25 C for 16 hours.

Fluorescence microscopy

GFP and mCHERRY fluorescence images from the transfected cotyledon protoplasts were acquired as described in Jo et al. (84). Relative GFP activity was determined by averaging the GFP:mCHERRY signal ratio.

GO representation analysis

Gene Ontology (GO) term enrichment analysis were performed using Bioconductor package GOseq, the soybean GO functional annotation, the hypergeometric method and a *q* value threshold of 0.05 (85).

Reference:

1. Berger F, Hamamura Y, Ingouff M, Higashiyama T. Double fertilization - caught in the act. Vol 13, Trends in Plant Science. 2008. p 437–43.
2. Figueiredo DD, Köhler C. Auxin: A molecular trigger of seed development. Vol 32, Genes and Development. 2018. p 479–90.
3. Goldberg RB, De Paiva G, Yadegari R. Plant embryogenesis: Zygote to seed. Science (80-). 1994;266(5185):605–14.
4. Berger F. Endosperm: The crossroad of seed development. Vol 6, Current Opinion in Plant Biology. Elsevier Ltd; 2003. p 42–50.
5. Olsen OA. Nuclear endosperm development in cereals and Arabidopsis thaliana. Vol 16, Plant Cell. 2004.
6. Moïse JA, Han S, Gudynaitę-Savitch L, Johnson DA, Miki BLA. Seed coats: Structure, development, composition, and biotechnology. Vol 41, In Vitro Cellular and Developmental Biology - Plant. 2005. p 620–44.
7. Chen X. Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol. 2009;25:21–44.
8. Voinnet O. Origin, Biogenesis, and Activity of Plant MicroRNAs. Vol 136, Cell. 2009. p 669–87.
9. Nodine MD, Bartel DP. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev. 2010;24(23):2678–92.
10. Willmann MR, Mehalick AJ, Packer RL, Jenik PD. MicroRNAs regulate the timing of embryo maturation in Arabidopsis. Plant Physiol. 2011;155(4):1871–84.
11. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. Endogenous siRNA and miRNA Targets Identified by Sequencing of the Arabidopsis Degradome. Curr Biol. 2008;18(10):758–62.
12. Wu G, Poethig RS. Temporal regulation of shoot development in Arabidopsis thaliana by miRr156 and its target SPL3. Development. 2006;133(18):3539–47.
13. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D. Specific effects of microRNAs on the plant transcriptome. Dev Cell. 2005;8(4):517–27.
14. Mallory AC, Bartel DP, Bartel B. MicroRNA-directed regulation of Arabidopsis Auxin Response Factor17 is essential for proper development and modulates expression of early auxin response genes. Plant Cell. 2005;17(5):1360–75.
15. Liu PP, Montgomery TA, Fahlgren N, Kasschau KD, Nonogaki H, Carrington JC. Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. Plant J. 2007;52(1):133–46.
16. Baldrich P, Beric A, Meyers BC. Despacito: the slow evolutionary changes in plant microRNAs. Vol 42, Current Opinion in Plant Biology. 2018. p 16–22.
17. Chávez Montes RA, De Fátima Rosas-Cárdenas F, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, et al. Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun. 2014;5.
18. Miyashima S, Honda M, Hashimoto K, Tatematsu K, Hashimoto T, Sato-Nara K, et al. A comprehensive expression analysis of the arabidopsis MICRORNA165/6 gene family during embryogenesis reveals a conserved role in meristem specification and a non-cell-autonomous function. Plant Cell Physiol. 2013;54(3):375–84.
19. Nogueira FTS, Chitwood DH, Madi S, Ohtsu K, Schnable PS, Scanlon MJ, et al.

Regulation of small RNA accumulation in the maize shoot apex. PLoS Genet. 2009;5(1).
20. Sieber P, Wellmer F, Gheyselinck J, Riechmann JL, Meyerowitz EM. Redundancy and specialization among plant microRNAs: Role of the MIR164 family in developmental robustness. Development. 2007;134(6):1051–60.
21. Yu N, Niu QW, Ng KH, Chua NH. The role of miR156/SPLs modules in Arabidopsis lateral root development. Plant J. 2015;83(4):673–85.
22. Jones-Rhoades MW. Conservation and divergence in plant microRNAs. Vol 80, Plant Molecular Biology. 2012. p 3–16.
23. Axtell MJ, Bowman JL. Evolution of plant microRNAs and their targets. Vol 13, Trends in Plant Science. 2008. p 343–9.
24. Khraiwesh B, Zhu JK, Zhu J. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim Biophys Acta - Gene Regul Mech. 2012;1819(2):137–48.
25. Cuperus JT, Fahlgren N, Carrington JC. Evolution and functional diversification of MIRNA genes. Vol 23, Plant Cell. 2011. p 431–42.
26. Lu C, Kulkarni K, Souret FF, MuthuValliappan R, Tej SS, Poethig RS, et al. MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. Genome Res. 2006;16(10):1276–88.
27. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, et al. High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. PLoS One. 2007;2(2).
28. Miller SS, Bowman LAA, Gijzen M, Miki BLA. Early development of the seed coat of soybean (Glycine max). Ann Bot. 1999;84(3):297–304.
29. Kozomara A, Birgaoanu M, Griffiths-Jones S. MiRBase: From microRNA sequences to function. Nucleic Acids Res. 2019;47(D1):D155–62.
30. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3).
31. Axtell MJ. Classification and comparison of small RNAs from plants. Annu Rev Plant Biol. 2013;64:137–59.
32. Axtell MJ, Meyers BC. Revisiting criteria for plant microRNA annotation in the Era of big data. Vol 30, Plant Cell. 2018. p 272–84.
33. Jeong DH, Park S, Zhai J, Gurazada SGR, de Paoli E, Meyers BC, et al. Massive analysis of rice small RNAs: Mechanistic implications of regulated MicroRNAs and variants for differential target RNA cleavage. Plant Cell. 2011;23(12):4185–207.
34. Douglass S, Hsu SW, Cokus S, Goldberg RB, Harada JJ, Pellegrini M. A naïve Bayesian classifier for identifying plant microRNAs. Plant J. 2016;86(6):481–92.
35. Li A, Mao L. Evolution of plant microRNA gene families. Cell Res. 2007;17(3):212–8.
36. Maher C, Stein L, Ware D. Evolution of Arabidopsis microRNA families through duplication events. Genome Res. 2006;16(4):510–9.
37. Yang L, Conway SR, Poethig RS. Vegetative phase change is mediated by a leaf-derived signal that represses the transcription of miR156. Development. 2011;138(2):245–9.
38. Kidner CA, Martienssen RA. Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature. 2004;428(6978):81–4.
39. Knauer S, Holt AL, Rubio-Somoza I, Tucker EJ, Hinze A, Pisch M, et al. A Protodermal miR394 Signal Defines a Region of Stem Cell Competence in the Arabidopsis Shoot Meristem. Dev Cell. 2013;24(2):125–32.
40. Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. The Sequential Action

of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. Cell. 2009;138(4):750–9.

41. Válóczi A, Várallyay É, Kauppinen S, Burgyán J, Havelda Z. Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. Plant J. 2006;47(1):140–51.
42. Arikit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, et al. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. Plant Cell. 2014;26(12):4584–601.
43. Talmor-Neiman M, Stav R, Frank W, Voss B, Arazi T. Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. Plant J. 2006;47(1):25–37.
44. Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, et al. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc Natl Acad Sci U S A. 2013;110(5).
45. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139–40.
46. German MA, Luo S, Schroth G, Meyers BC, Green PJ. Construction of parallel analysis of rna ends (Pare) libraries for the study of cleaved mirna targets and the rna degradome. Nat Protoc. 2009;4(3):356–62.
47. Kakrana A, Hammond R, Patel P, Nakano M, Meyers BC. SPARTA: A parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. Nucleic Acids Res. 2014;42(18).
48. Garcia D. A miRacle in plant development: Role of microRNAs in cell differentiation and patterning. Vol 19, Seminars in Cell and Developmental Biology. 2008. p 586–95.
49. Parizotto EA, Dunoyer P, Rahm N, Himber C, Voinnet O. In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. Genes Dev. 2004;18(18):2237–42.
50. Liu Q, Wang F, Axtell MJ. Analysis of complementarity requirements for plant MicroRNA targeting using a Nicotiana benthamiana quantitative transient assay. Plant Cell [Internet]. 2014;26(2):741–53. Available at: http://www.plantcell.org/cgi/doi/10.1105/tpc.113.120972
51. Xu M, Hu T, Zhao J, Park MY, Earley KW, Wu G, et al. Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in Arabidopsis thaliana. PLoS Genet. 2016;12(8).
52. Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, Rodriguez R, et al. Sequence and Expression Differences Underlie Functional Specialization of Arabidopsis MicroRNAs miR159 and miR319. Dev Cell. 2007;13(1):115–25.
53. Mallory AC, Dugas D V., Bartel DP, Bartel B. MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. Curr Biol. 2004;14(12):1035–46.
54. Larue CT, Wen J, Walker JC. A microRNA-transcription factor module regulates lateral organ size and patterning in Arabidopsis. Plant J. 2009;58(3):450–63.
55. Jia Y, Yao X, Zhao M, Zhao Q, Du Y, Yu C, et al. Comparison of soybean transformation efficiency and plant factors affecting transformation during the agrobacterium infection process. Int J Mol Sci. 2015;16(8):18522–43.
56. Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, et al. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific

transcription factors. Proc Natl Acad Sci U S A. 4 May 2010;107(18):8063–70.

57. Fraudentali I, Ghuge SA, Carucci A, Tavladoraki P, Angelini R, Rodrigues-Pousada RA, et al. Developmental, hormone- and stress-modulated expression profiles of four members of the Arabidopsis copper-amine oxidase gene family. Plant Physiol Biochem. 2020;147:141–60.

58. Wu H, Ji Y, Du J, Kong D, Liang H, Ling HQ. ClpC1, an ATP-dependent Clp protease in plastids, is involved in iron homeostasis in Arabidopsis leaves. Ann Bot. 2010;105(5):823–33.

59. Song K, Jang M, Kim SY, Lee G, Lee GJ, Kim DH, et al. An A/ENTH domain-containing protein functions as an adaptor for clathrin-coated vesicles on the growing cell plate in Arabidopsis root cells. Plant Physiol. 2012;159(3):1013–25.

60. McLoughlin F, Arisz SA, Dekker HL, Kramer G, De Koster CG, Haring MA, et al. Identification of novel candidate phosphatidic acid-binding proteins involved in the salt-stress response of Arabidopsis thaliana roots. Biochem J. 2013;450(3):573–81.

61. Zhao Z, Yang X, Lü S, Fan J, Opiyo S, Yang P, et al. Deciphering the novel role of atmin7 in cuticle formation and defense against the bacterial pathogen infection. Int J Mol Sci. 2020;21(15):1–21.

62. Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC. Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature. 2009;460(7252):283–6.

63. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Vol 11, Nature Reviews Genetics. 2010. p 204–20.

64. Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, et al. Genome-wide demethylation of Arabidopsis endosperm. Science (80-). 2009;324(5933):1451–4.

65. Curtin SJ, Michno JM, Campbell BW, Gil-Humanes J, Mathioni SM, Hammond R, et al. MicroRNA maturation and microRNA target gene expression regulation are severely disrupted in soybean dicer-like1 double mutants. G3 Genes, Genomes, Genet. 2016;6(2):423–33.

66. Plotnikova A, Kellner MJ, Schon MA, Mosiolek M, Nodine MD. MicroRNA dynamics and functions during arabidopsis embryogenesis[CC-BY]. Plant Cell. 2019;31(12):2929–46.

67. Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, et al. Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet. 2007;39(8):1033–7.

68. Skopelitis DS, Benkovics AH, Husbands AY, Timmermans MCP. Boundary Formation through a Direct Threshold-Based Readout of Mobile Small RNA Gradients. Dev Cell. 2017;43(3):265-273.e6.

69. Miyashima S, Koi S, Hashimoto T, Nakajima K. Non-cell-autonomous microRNA 165 acts in a dose-dependent manner to regulate multiple differentiation status in the Arabidopsis root. Development. 2011;138(11):2303–13.

70. Nikovics K, Blein T, Peaucelle A, Ishida T, Morin H, Aida M, et al. The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. Plant Cell. 2006;18(11):2929–45.

71. Baker CC, Sieber P, Wellmer F, Meyerowitz EM. The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in Arabidopsis. Curr Biol. 2005;15(4):303–15.

72. Chen X. A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. Science (80-). 2004;303(5666):2022–5.

73. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, et al. Widespread translational inhibition by plant miRNAs and siRNAs. Science (80-). 2008;320(5880):1185–90.

74. Endo A, Tatematsu K, Hanada K, Duermeyer L, Okamoto M, Yonekura-Sakakibara K, et al. Tissue-specific transcriptome analysis reveals cell wall metabolism, flavonol biosynthesis and defense responses are activated in the endosperm of germinating arabidopsis thaliana seeds. Plant Cell Physiol. 2012;53(1):16–27.

75. Sheoran IS, Olson DJH, Ross ARS, Sawhney VK. Proteome analysis of embryo and endosperm from germinating tomato seeds. Proteomics. 2005;5(14):3752–64.

76. Jerkovic A, Kriegel AM, Bradner JR, Atwell BJ, Roberts TH, Willows RD. Strategic distribution of protective proteins within bran layers of wheat protects the nutrient-rich endosperm. Plant Physiol. 2010;152(3):1459–70.

77. Vensel WH, Janaka CK, Cai N, Wong JH, Buchanan BB, Hurkman WJ. Developmental changes in the metabolic protein profiles of wheat endosperm. Proteomics. 2005;5(6):1594–611.

78. Grant Reid JS, Derek Bewley J. A dual rôle for the endosperm and its galactomannan reserves in the germinative physiology of fenugreek (Trigonella foenum-graecum L.), an endospermic leguminous seed. Planta. 1979;147(2):145–50.

79. 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. Proc Natl Acad Sci U S A. 2017;114(32):E6710–9.

80. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3).

81. Griffiths-Jones S, Saini HK, Van Dongen S, Enright AJ. miRBase: Tools for microRNA genomics. Nucleic Acids Res. 2008;36(SUPPL. 1).

82. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31(13):3406–15.

83. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9.

84. Jo L, Pelletier JM, Hsu SW, Baden R, Goldberg RB, Harada JJ. Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. Proc Natl Acad Sci U S A. 2020;117(2):1223–32.

85. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2).

86. Diener AC, Gaxiola RA, Fink GR. Arabidopsis ALF5, a multidrug efflux transporter gene family member, confers resistance to toxins. Plant Cell. 2001;13(7):1625–37.

87. Solís-Guzmán MG, Argüello-Astorga G, López-Bucio J, Ruiz-Herrera LF, López-Meza J, Sánchez-Calderón L, et al. Expression analysis of the Arabidopsis thaliana AtSpen2 gene, and its relationship with other plant genes encoding spen proteins. Genet Mol Biol. 2017;40(3):643–55.

88. Nomura K, Mecey C, Lee YN, Imboden LA, Chang JH, He SY. Effector-triggered immunity blocks pathogen degradation of an immunity-associated vesicle traffic regulator in Arabidopsis. Proc Natl Acad Sci U S A. 2011;108(26):10774–9.

89. Guo N, Ye W, Yan Q, Huang J, Wu Y, Shen D, et al. Computational identification of

novel microRNAs and targets in Glycine max. Mol Biol Rep. 2014;41(8):4965–75.

90. Li Y, Varala K, Hudson ME. A survey of the small RNA population during far-red light-induced apical hook opening. Front Plant Sci. 2014;5(APR).

91. Yan Z, Hossain MS, Arikit S, Valdés-López O, Zhai J, Wang J, et al. Identification of microRNAs and their mRNA targets during soybean nodule development: Functional analysis of the role of miR393j-3p in soybean nodulation. New Phytol. 2015;207(3):748–59.

92. Li W, Wang P, Li Y, Zhang K, Ding F, Nie T, et al. Identification of MicroRNAs in response to different day lengths in soybean using high-throughput sequencing and qRT-PCR. PLoS One. 2015;10(7).

93. Kulcheski FR, Molina LG, da Fonseca GC, de Morais GL, de Oliveira LFV, Margis R. Novel and conserved microRNAs in soybean floral whorls. Gene. 2016;575(2):213–23.

94. Ding X, Li J, Zhang H, He T, Han S, Li Y, et al. Identification of miRNAs and their targets by high-throughput sequencing and degradome analysis in cytoplasmic male-sterile line NJCMS1A and its maintainer NJCMS1B of soybean. BMC Genomics. 2016;17(1).

95. Sun Z, Wang Y, Mou F, Tian Y, Chen L, Zhang S, et al. Genome-wide small RNA analysis of soybean reveals auxin-responsive microRNAs that are differentially expressed in response to salt stress in root apex. Front Plant Sci. 2016;6(JAN2016).

96. Wang Y, Lan Q, Zhao X, Xu W, Li F, Wang Q, et al. Comparative profiling of microRNA expression in soybean seeds from genetically modified plants and their near-isogenic parental lines. PLoS One. 2016;11(5).

97. Xu S, Liu N, Mao W, Hu Q, Wang G, Gong Y. Identification of chilling-responsive microRNAs and their targets in vegetable soybean (Glycine max L.). Sci Rep. 2016;6.

98. Zhao M, Cai C, Zhai J, Lin F, Li L, Shreve J, et al. Coordination of MicroRNAs, PhasiRNAs, and NB-LRR Genes in Response to a Plant Pathogen: Insights from Analyses of a Set of Soybean Rps Gene Near-Isogenic Lines. Plant Genome. 2015;8(1).

Figures and Tables

A.

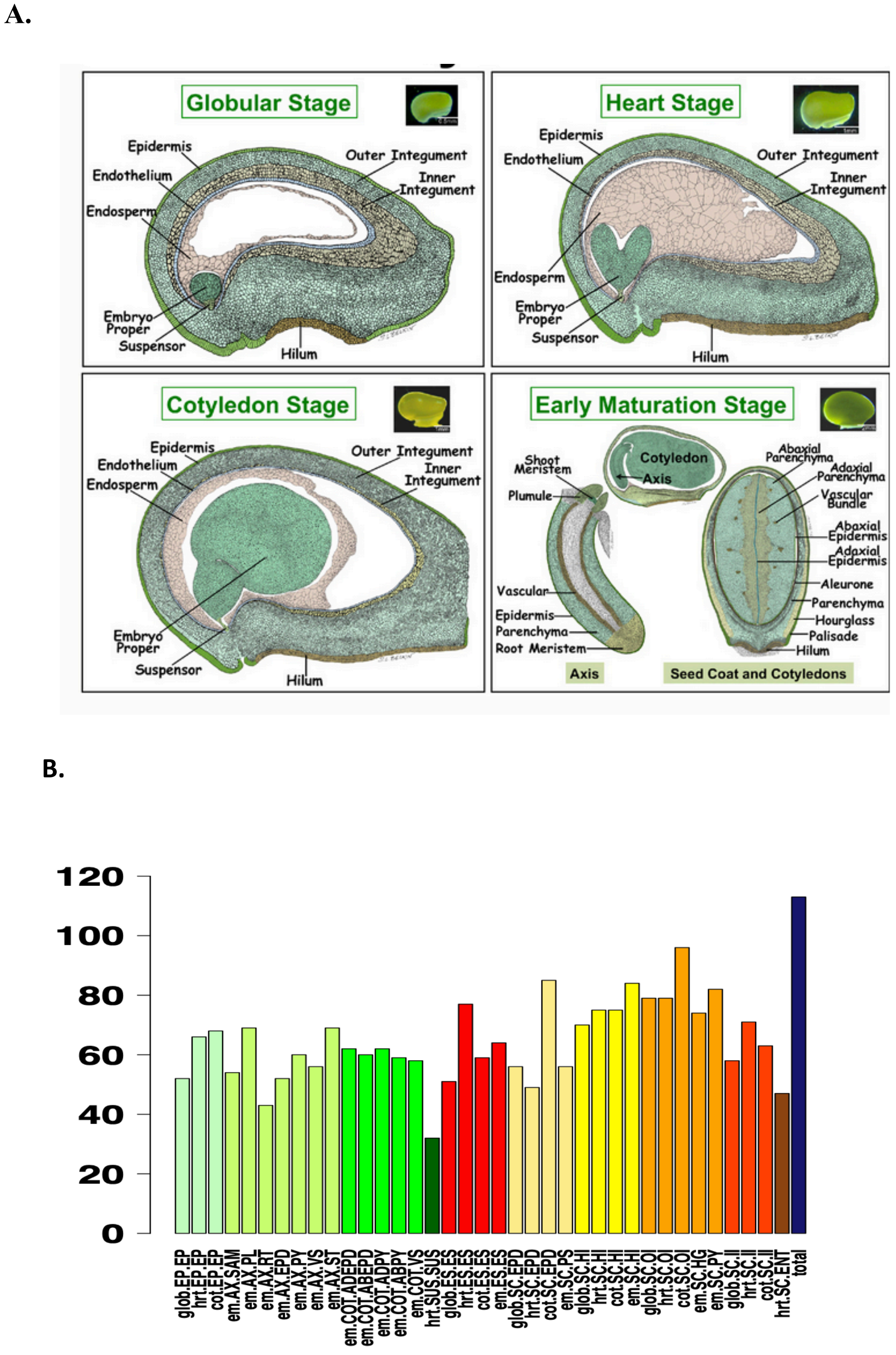
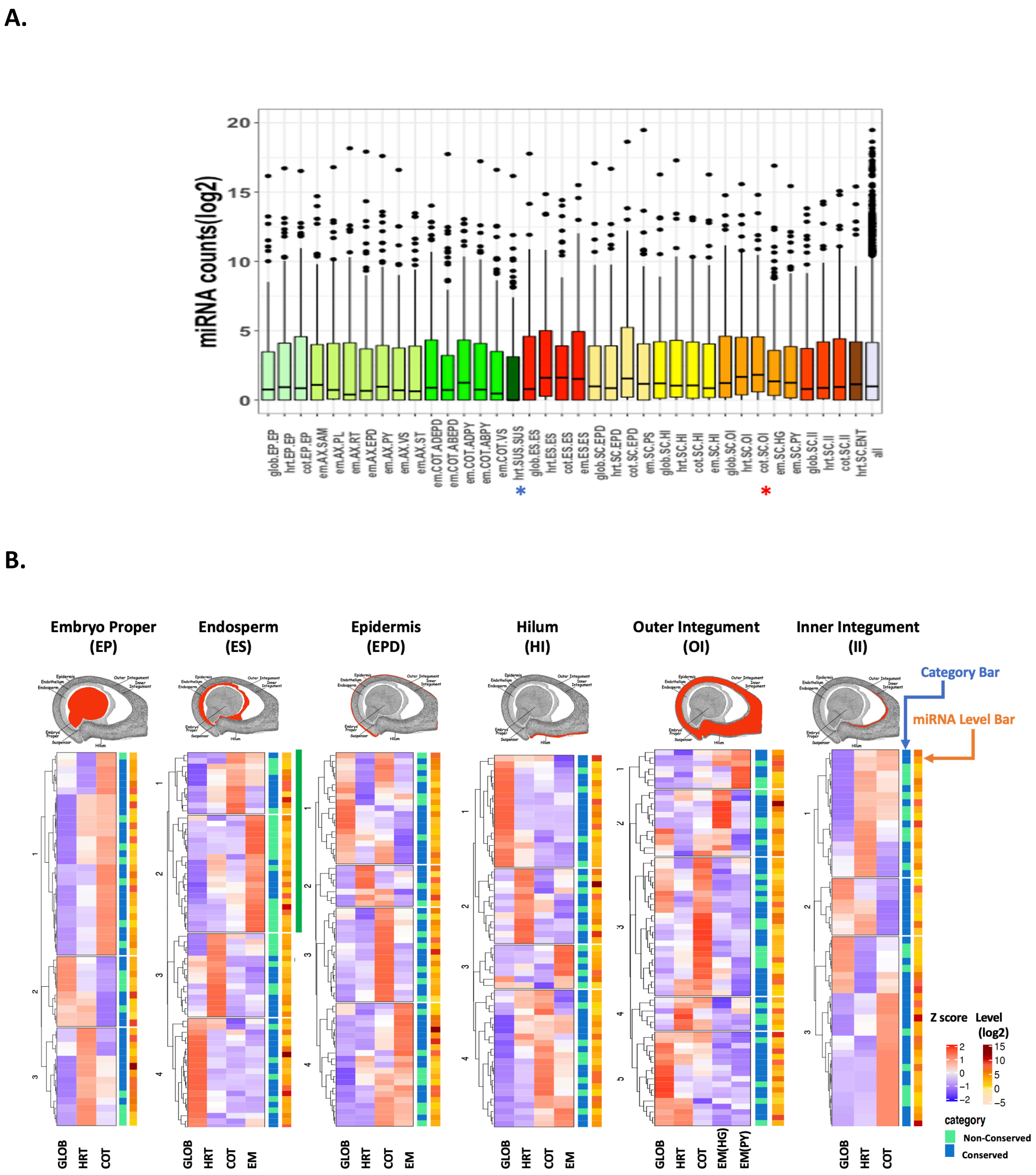


Figure 2.1. Number of miRNAs in each subregion at different developmental stages. **(**A) Soybean seed subregions and stages. The figures were taken from Gene Networks in Seed Development website (http://seedgenenetwork.net). (B) Number of miRNA families present in subregions at different stages. The total number of miRNA families (113 families) detected in all the subregions and stages is indicated on the furthest right side in navy blue for the comparison. Abbreviations for stages: glob (globular), hrt (heart), cot (cotyledon) and em (early maturation). Abbreviations for subregion: ABEPD (abaxial epidermis), ABPY (abaxial parenchyma), ADEPD (adaxial epidermis), ADPY (adaxial parenchyma), AX (axis), COT (cotyledon), ENT (endothelium), EP (embryo proper), EPD (epidermis), ES (endosperm), HG (hourglass), HI (hilum), II (inner integument), OI (outer integument), PL (plumule), PS (palisade), PY (Parenchyma), RT (root tip), SAM (shoot apical meristem), SC (seed coat), ST (stele), SUS (suspensor), VS (vasculature).


A.
miRNA counts(log2)
20
15
10
5
0
glob.EP.EP
hrt.EP.EP
cot.EP.EP
em.AX.SAM
em.AX.PL
em.AX.RT
em.AX.EPD
em.AX.PY
em.AX.VS
em.AX.ST
em.COT.ADEPD
em.COT.ABEPD
em.COT.ADPY
em.COT.ABPY
em.COT.VS
hrt.SUS.SUS
glob.ES.ES
hrt.ES.ES
cot.ES.ES
em.ES.ES
glob.SC.EPD
hrt.SC.EPD
cot.SC.EPD
em.SC.PS
glob.SC.HI
hrt.SC.HI
cot.SC.HI
em.SC.HI
glob.SC.OI
hrt.SC.OI
cot.SC.OI
em.SC.HG
em.SC.PY
glob.SC.II
hrt.SC.II
cot.SC.II
hrt.SC.ENT
all
*
*
B.
Embryo Proper (EP)
Endosperm (ES)
Epidermis (EPD)
Hilum (HI)
Outer Integument (OI)
Inner Integument (II)
Category Bar
miRNA Level Bar
Epidermis
Endothelium
Endosperm
Outer Integument
Inner Integument
Embryo Proper
Suspensor
Hilum
1
2
3
4
5
GLOB
HRT
COT
EM
EM(HG)
EM(PY)
Z score
Level (log2)
2
1
0
−1
−2
15
10
5
0
−5
category
Non-Conserved
Conserved

C.

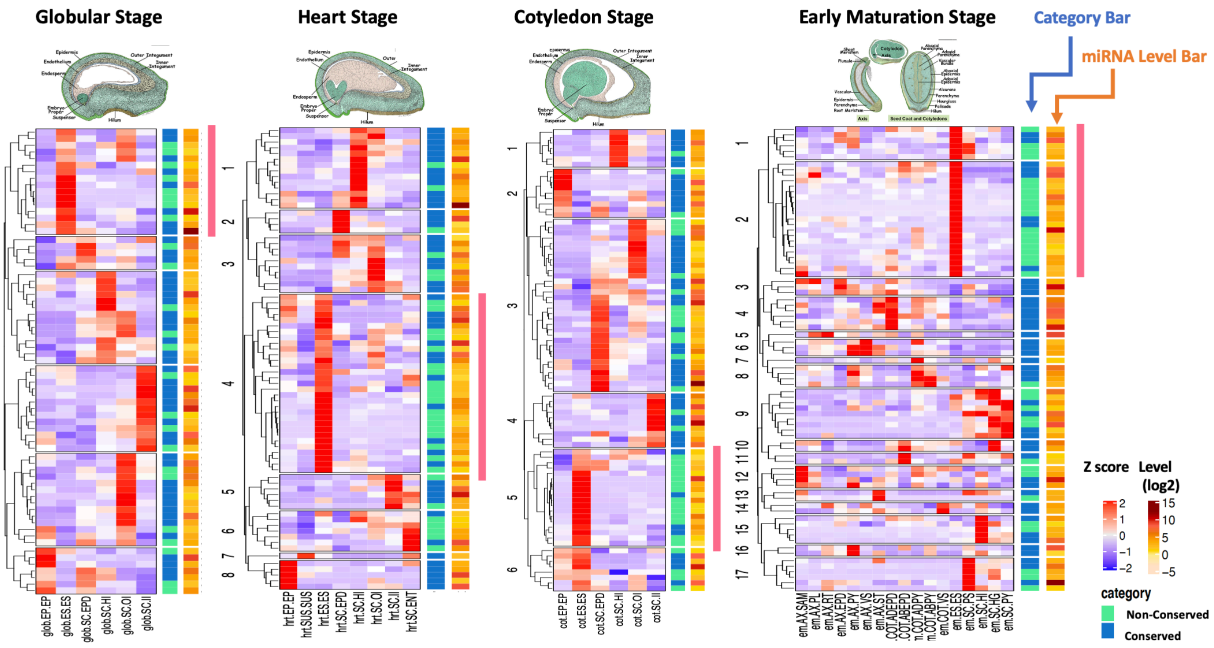


Figure 2.2. miRNA family accumulation in subregions and stages. (A) Global Distributions of miRNA family accumulation in each subregion. The furthest right bar in lavender shows the level of miRNA families present in all subregions. Red and blue stars mark the subregions indicate higher or lower than the total miRNA family distributions with a statistical significance, respectively. (B) Hierarchical clustering of miRNAs temporally by subregions. Subregions are highlighted in red in the seed cross section sketches. (C) Hierarchical clustering of miRNAs in space by stages. miRNA families with a maximum count of less than 1cpm in a given subregions or stages are excluded. The middle thin bar indicates miRNA family conservations. The right thin bar shows the highest miRNA family level in the subregions in each heatmap. The values of the miRNA family levels are log2 transformed. The vertical green bars in (B) indicate the clusters where miRNA families predominately accumulated in the endosperm at the cotyledon and early maturation stages. The vertical pink bars in (C) indicate the clusters where miRNA

families predominately accumulated in the endosperm subregions. The heart, cotyledon and early maturation endosperm subregions are particularly enriched with non-conserved miRNAs. The scale bars used for the heatmaps are shown on the furthest right on each figure. Abbreviations for stages: glob (globular stage), hrt (heart stage), cot (cotyledon stage), em (early maturation). Abbreviations for subregions: ABEPD (abaxial epidermis), ABPY (abaxial parenchyma), ADEPD (adaxial epidermis), ADPY (adaxial parenchyma), AX (axis), COT (cotyledon), ENT (endothelium), EP (embryo proper), EPD (epidermis), ES (endosperm), HG (hourglass), HI (hilum), II (inner integument), OI (outer integument), PL (plumule), PS (palisade), PY (Parenchyma), RT (root tip), SAM (shoot apical meristem), SC (seed coat), ST (stele), SUS (suspensor), VS (vasculature).

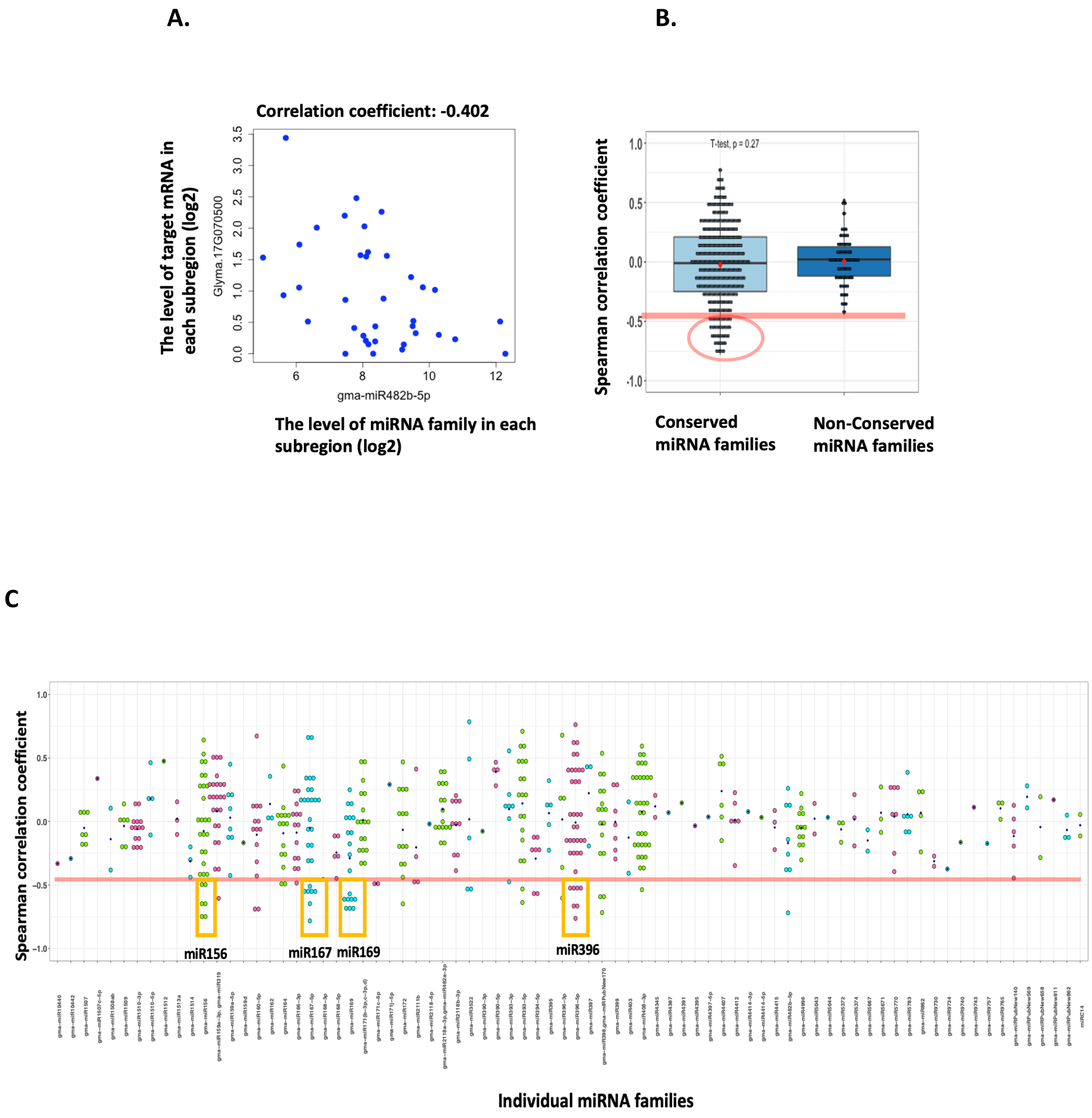
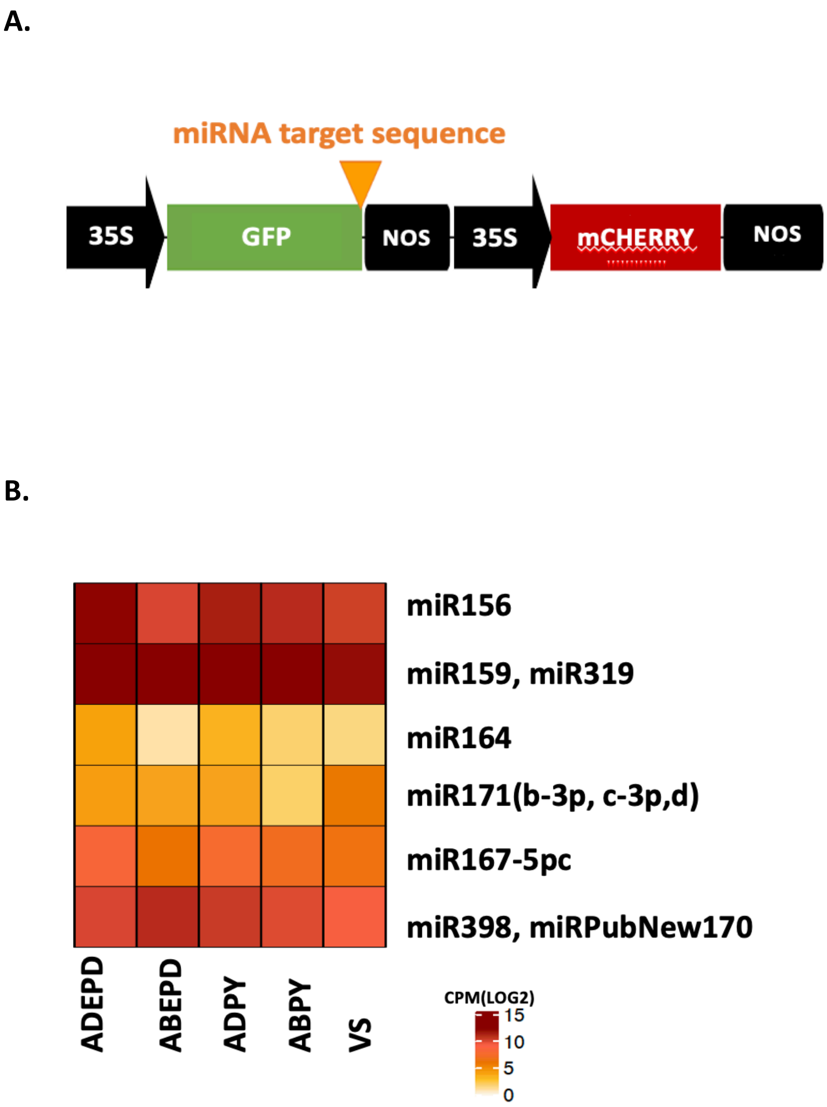


Figure 2.3. Global pairwise comparison of the levels of a miRNA family and its target mRNA. (A) A representation of the scattered plot used to compare the level of a miRNA family and its target mRNA. X axis shows the level of a miRNA family. Y axis shows the level of an

associated target mRNA. Spearman correlation coefficient is indicated on the top. miRNA family and target mRNA pairs with less than 1 cpm read counts for miRNA family or target mRNAs in all subregions were not included in this assay. (B) Distributions of Spearman correlation coefficient for conserved miRNAs and non-conserved miRNAs. Red line indicates -0.5 for correlation coefficient, corresponding to the cutoff value that was used as strong negative correlation coefficient for this assay. Red circle indicates the miRNA family-target mRNA pairs with the correlation coefficient value less than -0.5. (C) Distributions of Spearman correlation coefficient for the individual miRNA families. Red line indicates the correlation coefficient value of -0.5. Orange squares show the miRNA family-target mRNA pairs with the correlation coefficient value less than -0.5.


A.
miRNA target sequence
35S
GFP
NOS
35S
mCHERRY
NOS
B.
miR156
miR159, miR319
miR164
miR171(b-3p, c-3p,d)
miR167-5pc
miR398, miRPubNew170
ADEPD
ABEPD
ADPY
ABPY
VS
CPM(LOG2)
15
10
5
0

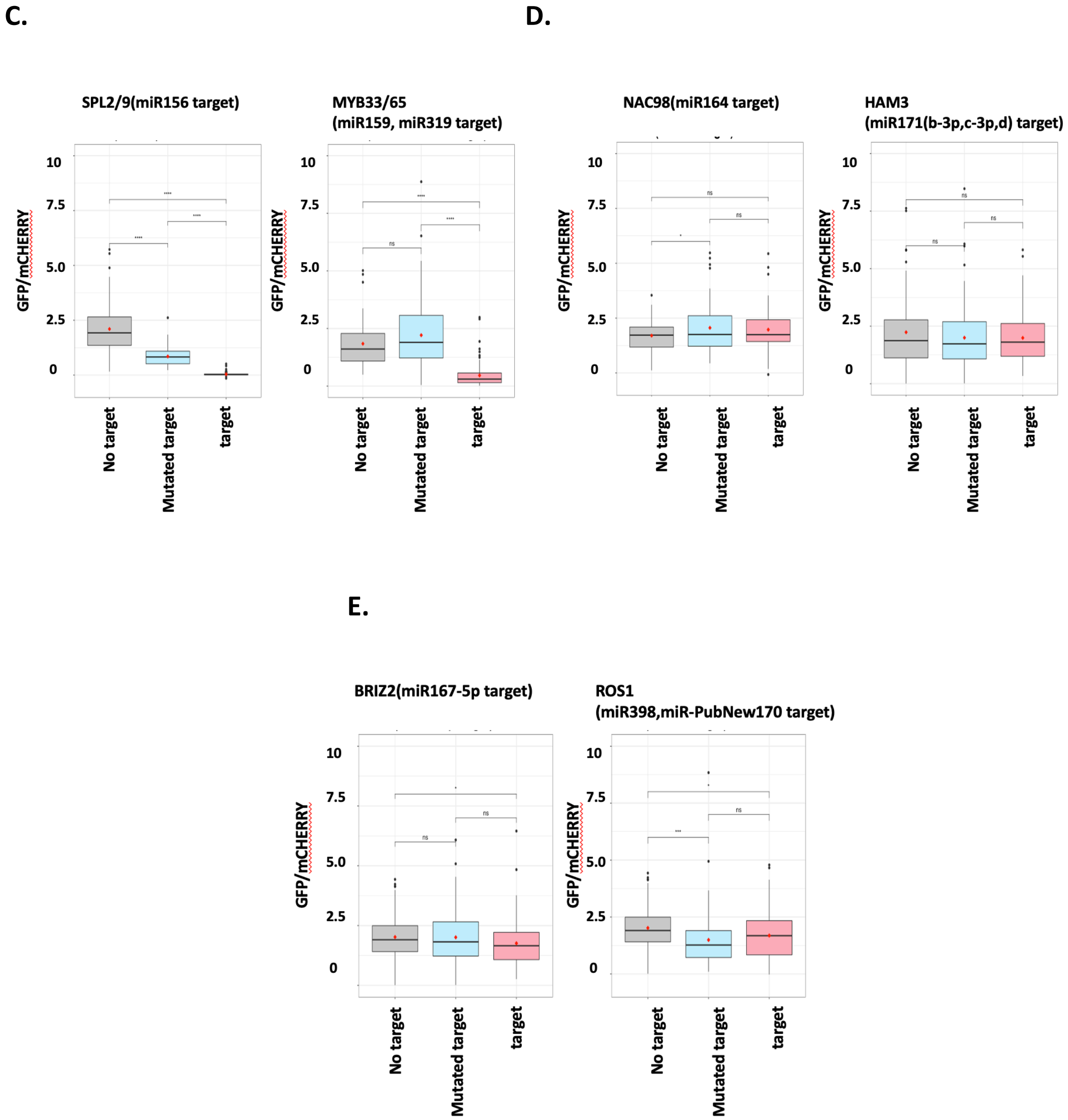


Figure 2.4. miRNA target sensor to detect endogenous miRNA family activities. (A) Schematic picture of the sensor construct. Target sequences were inserted at the 3' end of GFP in frame. (B) Level of miRNA families corresponding to the target sensors in the embryo cotyledon subregions at the early maturation stage. Protoplasts were harvested from embryo cotyledons.

Abbreviations: ADEPD (adaxial epidermis), ABEPD (abaxial epidermis), ADPY (adaxial parenchyma), ABPY (abaxial parenchyma) and VS (vasculature). ADPY and ABPY occupy most of the cotyledon tissues. The read counts are log2 transformed. (C-E) Detection of endogenous miRNA activities by transfecting protoplasts with the sensors. GFP levels were measured as the ratio of GFP to mCHERRY signal. Gray, blue, and pink bars represent sensors without any additional sequences, with mutated target sequences, and with target sequences, respectively. (C) GFP levels of protoplasts with the construct containing the target mRNA sequences for miRNA family/target mRNA pairs whose miRNAs are accumulated at the high level. These miRNA family-target mRNA pairs are identified in other plant species. (D) GFP levels of protoplasts with the construct containing the target mRNA sequences for miRNA family/target mRNA pairs whose miRNAs are accumulated at the low level. These miRNA family-target mRNA pairs are identified in other plant species. (E) GFP levels of protoplasts with the construct containing the target mRNA sequences for miRNA family/target mRNA pairs whose miRNAs are accumulated at the relatively high level. These miRNA family- target mRNA pairs are novel.

.

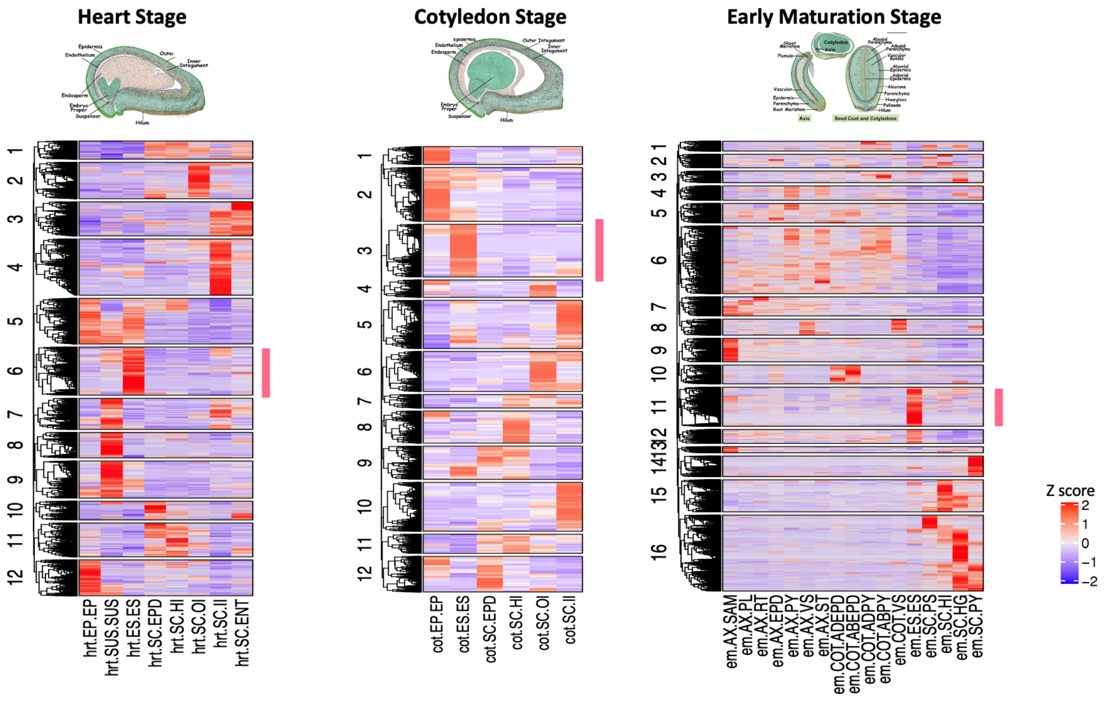


Figure 2.5. Hierarchical clustering of the 15,000 most variable mRNA and all miRNA families combined. The clusters with a pink bar in each heatmap contain endosperm-specific miRNA families in each stage. The scale bar shows on the right. Abbreviations for stages: hrt (heart stage), cot (cotyledon stage), em (early maturation). Abbreviations for subregions: ABEPD (abaxial epidermis), ABPY (abaxial parenchyma), ADEPD (adaxial epidermis), ADPY (adaxial parenchyma), AX (axis), COT (cotyledon), ENT (endothelium), EP (embryo proper), EPD (epidermis), ES (endosperm), HG (hourglass), HI (hilum), II (inner integument), OI (outer integument), PL (plumule), PS (palisade), PY (Parenchyma), RT (root tip), SAM (shoot apical meristem), SC (seed coat), ST (stele), SUS (suspensor), VS (vasculature).

Table 2.1. Abbreviations for soybean developmental stages and subregions used for this study on the left column in orange.

Abbreviation[*]	Stage[§]	Seed subregion category [†]	Subregion [‡]
glob-EP-EP	globular	embryo proper	embryo proper
hrt-EP-EP	heart	embryo proper	embryo proper
cot-EP-EP	cotyledon	embryo proper	embryo proper
em-AX-SAM	early maturation	axis (embryo proper)	shoot apical meristem
em-AX-PL	early maturation	axis (embryo proper)	plumule
em-AX-RT	early maturation	axis (embryo proper)	root tip
em-AX-EPD	early maturation	axis (embryo proper)	epidermis
em-AX-PY	early maturation	axis (embryo proper)	parenchyma
em-AX-VS	early maturation	axis (embryo proper)	vascular
em-AX-ST	early maturation	axis (embryo proper)	stele
em-COT-ADEPD	early maturation	cotyledon (embryo proper)	adaxial epidermis
em-COT-ABEPD	early maturation	cotyledon (embryo proper)	abaxial epidermis
em-COT-ADPY	early maturation	cotyledon (embryo proper)	adaxial parenchyma
em-COT-ABPY	early maturation	cotyledon (embryo proper)	abaxial parenchyma
em-COT-VS	early maturation	cotyledon (embryo proper)	vascular
hrt-SUS-SUS	heart	suspensor	suspensor
glob-ES-ES	globular	endosperm	endosperm
hrt-ES-ES	heart	endosperm	endosperm
cot-ES-ES	cotyledon	endosperm	endosperm
em-ES-ES	early maturation	endosperm	endosperm
glob-SC-EPD	globular	seed coat	epidermis
hrt-SC-EPD	heart	seed coat	epidermis
cot-SC-EPD	cotyledon	seed coat	epidermis
em-SC-PS	early maturation	seed coat	palisade
glob-SC-HI	globular	seed coat	hilum
hrt-SC-HI	heart	seed coat	hilum
cot-SC-HI	cotyledon	seed coat	hilum
em-SC-HI	early maturation	seed coat	hilum

Table2.1 continued

Abbreviation*	Stage§	Seed subregion category†	Subregion‡
glob-SC-OI	globular	seed coat	outer integument
hrt-SC-OI	heart	seed coat	outer integument
cot-SC-OI	cotyledon	seed coat	outer integument
em-SC-HG	early maturation	seed coat	hourglass
em-SC-PY	early maturation	seed coat	parenchyma
glob-SC-II	globular	seed coat	inner integument
hrt-SC-II	heart	seed coat	inner integument
cot-SC-II	cotyledon	seed coat	inner integument
hrt-SC-ENT	heart	seed coat	endothilium

*The abbreviations are indicated the following order used the names in each column: stage§- seed subregion category†-subregion‡.

Table 2.2. List of miRNAs identified in this study.

miRNA name	miRNA Family	Sequence	miRNA Conservation
gma-miR10440	gma-miR10440	TTGGGACAATACTTTAGATAT	Non-Conserved
gma-miR10442	gma-miR10442	CTACATGCTGCACTTGGATCC	Non-Conserved
gma-miR1507a	gma-miR1507	TCTCATTCCATACATCGTCTGA	Conserved
gma-miR1507b	gma-miR1507	TCTCATTCCATACATCGTCTG	Conserved
gma-miR1507c-5p	gma-miR1507c-5p	GAGGTGTTTGGGATGAGAGAA	Conserved
gma-MIR1508a-5pNew	gma-miR1508ab	ACTGCTATTCCCATTTCTAAA	Conserved
gma-miR1509a	gma-miR1509	TTAATCAAGGAAATCACGGTCG	Conserved
gma-miR1509b	gma-miR1509	TTAATCAAGGAAATCACGGTT	Conserved
gma-miR1510b-3p	gma-miR1510-3p	TGTTGTTTTACCTATTCCACC	Conserved
gma-miR1510b-5p	gma-miR1510-5p	AGGGATAGGTAAAACAACTACT	Conserved
gma-miR1511-NV	gma-miR1511	AACCAGGCTCTGATACCATGG	Conserved
gma-miR1512b	gma-miR1512	TAACTGGAAATTCTTAAAGCAT	Non-Conserved
gma-miR1512c	gma-miR1512	TAACTGAACATTCTTAGAGCAT	Non-Conserved
gma-miR1513(a-5p,b)	gma-miR1513a	TGAGAGAAAGCCATGACTTAC	Non-Conserved
gma-miR1513b-5pNew	gma-miR1513b	AAATCATGACTTTCTCTTGTA	Non-Conserved
gma-miR1514a-5p-NV	gma-miR1514	TTCATTTTTAAAATAGGCATTG	Conserved
gma-miR1531-5p	gma-miR1531	ATATGGACGAAGAGATAGGTAAAT	Non-Conserved
gma-miR156(a,h,u,v,w,x,y)	gma-miR156	TGACAGAAGAGAGTGAGCAC	Conserved
gma-miR156(b,f)-NV	gma-miR156	TTGACAGAAGAGAGAGAGCAC	Conserved
gma-miR156(c,d,i,j,l,m)	gma-miR156	TTGACAGAAGATAGAGAGCAC	Conserved
gma-miR156(k,n,o)	gma-miR156	TTGACAGAAGAGAGTGAGCAC	Conserved
gma-miR156(p,t)	gma-miR156	TTGACAGAAGAAAGGGAGCAC	Conserved
gma-miR156(q,s)	gma-miR156	TGACAGAAGAGAGTGAGCACT	Conserved
gma-miR156e	gma-miR156	CTGACAGAAGATAGAGAGCAC	Conserved
gma-miR156f	gma-miR156	TTGACAGAAGAGAGAGAGCACA	Conserved
gma-miR159(a-3p,e-3p)	gma-miR159a-3p,gma-miR319	TTTGGATTGAAGGGAGCTCTA	Conserved

Table2.2 continuedd

miRNA name	miRNA Family	Sequence	miRNA Conservation
gma-miR319(a,b,e)	gma-miR159a-3p,gma-miR319	TTGGACTGAAGGGAGCTCCC	Conserved
gma-miR319(g,l)	gma-miR159a-3p,gma-miR320	TTGGACTGAAGGGAGCTCCTTC	Conserved
gma-miR319(h,j,k,m)	gma-miR159a-3p,gma-miR321	TTGGACTGAAGGGAGCTCCCT	Conserved
gma-miR319c	gma-miR159a-3p,gma-miR322	TTGGACTGAAGGGAGCTCCT	Conserved
gma-miR319n	gma-miR159a-3p,gma-miR323	TTTGGACCGAAGGGAGCCCCT	Conserved
gma-miR319p	gma-miR159a-3p,gma-miR324	TTTTGGACTGAAGGGAGCTCC	Conserved
gma-miR319q	gma-miR159a-3p,gma-miR325	TGGACTGAAGGGAGCTCCTTC	Conserved
gma-miR159a-5p	gma-miR159a-5p	GAGCTCCTTGAAGTCCAATTG	Conserved
gma-miR159e-5p	gma-miR159a-5p	GAGCTCCTTGAAGTCCAATT	Conserved
gma-miR159d	gma-miR159d	AGCTGCTTAGCTATGGATCCC	Conserved
gma-miR160a-3p	gma-miR160-3p	GCGTATGAGGAGCCAAGCATA	Conserved
gma-miR160(a-5p,f)	gma-miR160-5p	TGCCTGGCTCCCTGTATGCCA	Conserved
gma-miR160(b,c,d,e)	gma-miR160-5p	TGCCTGGCTCCCTGTATGCC	Conserved
gma-miR162(b,c)	gma-miR162	TCGATAAACCTCTGCATCCAG	Conserved
gma-miR162a	gma-miR162	TCGATAAACCTCTGCATCCA	Conserved
gma-miR164(a,e,f,g,h,i,j,k)	gma-miR164	TGGAGAAGCAGGGCACGTGCA	Conserved
gma-miR164(b,c,d)	gma-miR164	TGGAGAAGCAGGGCACGTGC	Conserved
gma-miR166(a-3p,b,c-3p,d,e,f,g,i-3p,n,o)	gma-miR166-3p	TCGGACCAGGCTTCATTCCCC	Conserved
gma-miR166(h-3p,k)	gma-miR166-3p	TCTCGGACCAGGCTTCATTCC	Conserved
gma-miR166(p,q,r,s,t)	gma-miR166-3p	TCGGACCAGGCTTCATTCCC	Conserved
gma-miR166j-3p	gma-miR166-3p	TCGGACCAGGCTTCATTCCCG	Conserved
gma-miR166u	gma-miR166-3p	TCTCGGACCAGGCTTCATTC	Conserved
gma-miR166(a-5p,c-5p,l)	gma-miR166-5p	GGAATGTTGTCTGGCTCGAGG	Conserved
gma-miR167(a,b,d)	gma-miR167-5p	TGAAGCTGCCAGCATGATCTA	Conserved
gma-miR167(c,j)	gma-miR167-5p	TGAAGCTGCCAGCATGATCTG	Conserved
gma-miR167(e,f)	gma-miR167-5p	TGAAGCTGCCAGCATGATCTT	Conserved
gma-miR167g	gma-miR167-5p	TGAAGCTGCCAGCATGATCTGA	Conserved
gma-miR168(a-3p,b-3p)New	gma-miR168-3p	CCCGCCTTGCATCAACTGAAT	Conserved
gma-miR168a	gma-miR168-5p	TCGCTTGGTGCAGGTCGGGAA	Conserved

Table2.2 continued

miRNA name	miRNA Family	Sequence	miRNA Conservation
gma-miR169(a,f,g,m)	gma-miR169	CAGCCAAGGATGACTTGCCGG	Conserved
gma-miR169(k,l-5p)	gma-miR169	CAGCCAAGAATGACTTGCCGG	Conserved
gma-miR169d	gma-miR169	TGAGCCAAGGATGACTTGCCGGT	Conserved
gma-miR169j-5p	gma-miR169	TAGCCAAGAATGACTTGCCGG	Conserved
gma-miR169n-3p	gma-miR169-3p	TGCCGGCAAGTTTCTCTTGGC	Conserved
gma-miR169j-5p-NV	gma-miR169jw	AAGAATGACTTGCCGGAATGCA	Conserved
gma-miR171(c-3p,o-3p,q)	gma-miR171(b-3p,c-3p,d)	TTGAGCCGTGCCAATATCACA	Conserved
gma-miR171(e,f,g,j-3p,u)	gma-miR171(b-3p,c-3p,d)	TGATTGAGCCGTGCCAATATC	Conserved
gma-miR171(m,t)	gma-miR171(b-3p,c-3p,d)	TTGAGCCGCGTCAATATCTCA	Conserved
gma-miR171(n,p)	gma-miR171(b-3p,c-3p,d)	TTGAGCCGCGTCAATATCTTA	Conserved
gma-miR171b-3p	gma-miR171(b-3p,c-3p,d)	CGAGCCGAATCAATATCACTC	Conserved
gma-miR171d	gma-miR171(b-3p,c-3p,d)	TGATTGAGTCGTGTCAATATC	Conserved
gma-miR171k-3p	gma-miR171(b-3p,c-3p,d)	TTGAGCCGCGCCAATATCACT	Conserved
gma-miR171(k-5p,l)	gma-miR171c-5p	CGATGTTGGTGAGGTTCAATC	Conserved
gma-miR171c-5p	gma-miR171c-5p	AGATATTGGTGCGGTTCAATC	Conserved
gma-miR171o-5p	gma-miR171c-5p	AGATATTGGTACGGTTCAATC	Conserved
gma-miR171(a-5pNew,i-5p-NV)	gma-miR171j-5p	ATATTGGTCCGGTTCAATAAG	Conserved
gma-miR171j-5p	gma-miR171j-5p	TATTGGCCTGGTTCACTCAGA	Conserved
gma-miR172c	gma-miR172	GGAATCTTGATGATGCTGCAG	Conserved
gma-miR2109-3p	gma-miR2109-3p	GGAGGCGTAGATACTCACACC	Conserved
gma-miR2109-5p	gma-miR2109-5p	TGCGAGTGTCTTCGCCTCTG	Conserved
gma-miR2111(a,d)	gma-miR2111a	GTCCTTGGGATGCAGATTACG	Conserved
gma-miR2111(b,c,e,f)	gma-miR2111b	TAATCTGCATCCTGAGGTTTA	Conserved
gma-miR2118(a-5p,b-5p)	gma-miR2118-5p	GGAGATGGGAGGGTCGGTAAAG	Conserved
gma-miR482(b-3p,d-3p)	gma-miR2118a-3p,gma-miR482a-3p	TCTTCCCTACACCTCCCATACC	Conserved
gma-miR482c-3p	gma-miR2118a-3p,gma-miR482a-3p	TTCCCAATTCCGCCCATTCCT	Conserved

Table2.2 continued

miRNA name	miRNA Family	Sequence	miRNA Conservation
gma-miR2118(a-3p,b-3p)	gma-miR2118b-3p	TTGCCGATTCCACCCATTCCT	Conserved
gma-miR3522-NV2	gma-miR3522	TGAGACCAAATGAGCAGCTGA	Conserved
gma-miR390(a-3p,c)	gma-miR390-3p	CGCTATCCATCCTGAGTTTC	Conserved
gma-miR390(a-5p,f,g)	gma-miR390-5p	AAGCTCAGGAGGGATAGCGCC	Conserved
gma-miR393c-3p	gma-miR393-3p	ATCATGCTATCCCTTTGGATT	Conserved
gma-miR393(c-5p,d,e,f,g)	gma-miR393-5p	TCCAAAGGGATCGCATTGATCC	Conserved
gma-miR393(h,i,j,k)	gma-miR393-5p	TTCCAAAGGGATCGCATTGATC	Conserved
gma-miR393(h,i,j,k)-NV	gma-miR393-5p	TCCAAAGGGATCGCATTGATCT	Conserved
gma-miR393a	gma-miR393-5p	TCCAAAGGGATCGCATTGATC	Conserved
gma-miR394(a-5p,b-5p,c-5p,d,e,f,g)	gma-miR394-5p	TTGGCATTCTGTCCACCTCC	Conserved
gma-miR394b-3p	gma-miR394ab-3p	AGGTGGGCATACTGTCAACT	Conserved
gma-miR395(a,b,c)	gma-miR395	CTGAAGTGTTTGGGGGAACTC	Conserved
gma-miR396(b-3p,k-3p)	gma-miR396-3p	GCTCAAGAAAGCTGTGGGAGA	Conserved
gma-miR396a-3p	gma-miR396-3p	TTCAATAAAGCTGTGGGAAG	Conserved
gma-miR396i-3p	gma-miR396-3p	GTTCAATAAAGCTGTGGGAAG	Conserved
gma-miR396(a-5p,i-5p)	gma-miR396-5p	TTCCACAGCTTTCTTGAACTG	Conserved
gma-miR396(b-5p,c,k-5p)	gma-miR396-5p	TTCCACAGCTTTCTTGAACTT	Conserved
gma-miR397(a,b-5p)	gma-miR397	TCATTGAGTGCAGCGTTGATG	Conserved
gma-miRPubNew519	gma-miR397	TCATTGAGTGTAGCATTGATG	Non-Conserved
gma-miR398(a,b-3p)	gma-miR398,gma-miRPubNew170	TGTGTTCTCAGGTCACCCCTT	Conserved
gma-miR398(c,d)	gma-miR398,gma-miRPubNew170	TGTGTTCTCAGGTCGCCCCTG	Conserved
gma-miRPubNew170	gma-miR398,gma-miRPubNew170	ATGTGTTTTCAGGTCACCCATG	Non-Conserved
gma-miR399(a,b,c,h)	gma-miR399	TGCCAAAGGAGAGTTGCCCTG	Conserved
gma-miR399(d,e,f,g)	gma-miR399	TGCCAAAGGAGATTTGCCCAG	Conserved
gma-miR399i	gma-miR399	TGCCAAAGGAGAATTGCCCTG	Conserved
gma-miR403(a,b)	gma-miR403	TTAGATTCACGCACAAACTTG	Conserved
gma-miR408(a-3p,b-3p,c-3p)	gma-miR408-3p	ATGCACTGCCTCTTCCCTGGC	Conserved
gma-miR408d	gma-miR408-3p	TGCACTGCCTCTTCCCTGGC	Conserved
gma-miR408(a-5p,c-5p)	gma-miR408-5p	CAGGGGAACAGGCAGAGCATG	Conserved
gma-miR408b-5p	gma-miR408-5p	CTGGGAACAGGCAGGGCACG	Conserved

Table2.2 continued

miRNA name	miRNA Family	Sequence	miRNA Conservation
gma-MIR4345-3pNew	gma-miR4345	CAATCTTTTTAAGTTTCGTCT	Non-Conserved
gma-miR4349-3p	gma-miR4349	TCACTTATATACTCTTTCTTGGCC	Non-Conserved
gma-miR4387e	gma-miR4387	TGTTAGTGATAAGGCGTGATG	Non-Conserved
gma-miR4391	gma-miR4391	TCTCGGCAAAGAACTAAGAAGAAG	Non-Conserved
gma-MIR4395-3pNew	gma-miR4395	CAGCAGCTTTCTCGGACCCATACT	Non-Conserved
gma-miR4397-3p	gma-miR4397-3p	TGTCAAAGATGTGGCGAATACT	Non-Conserved
gma-miR4397-5p	gma-miR4397-5p	CATCGTTGACGCTGACTGTACG	Non-Conserved
gma-miR4407	gma-miR4407	CAGAGGAAGCAGCACTTGTACC	Non-Conserved
gma-miR4412-5p	gma-miR4412	TGTTGCGGGTATCTTTGCCTC	Non-Conserved
gma-miR4414a-3p	gma-miR4414-3p	TCCAACGATGCGGGAGCTGC	Non-Conserved
gma-miR4414a-5p	gma-miR4414-5p	AGCTGCTGACTCGTTGGCTC	Non-Conserved
gma-miR4415(a-3p,b-3p)	gma-miR4415	TTGATTCTCATCACAACATGG	Non-Conserved
gma-miR4416(a-5pNew,b-NV)	gma-miR4416	TGGGTGAGAGAAACGCGTATCG	Non-Conserved
gma-miR4416b	gma-miR4416	TGGGTGAGAGAAACGCGTATC	Non-Conserved
gma-miR482(a-5p-NV,c-5p-NV)	gma-miR482b-5p	GGAATGGGCTGATTGGGAAG	Conserved
gma-miR482(b-5p,d-5p)	gma-miR482b-5p	TATGGGGGGATTGGGAAGGAAT	Conserved
gma-MIR482c-5p-NV2	gma-miR482b-5p	GGAATGGGCTGATTGGGAAGT	Conserved
gma-miR482e	gma-miR482b-5p	TATGGGGGGATTGGGAAGGAA	Conserved
gma-miR4996	gma-miR4996	TAGAAGCTCCCCATGTTCTC	Non-Conserved
gma-miR5035-NV	gma-miR5035	TGAGGGAAAAAATGTTTAGAAGCT	Non-Conserved
gma-miR5036	gma-miR5036	AGAGGCCCTTGGGGAGGAGTAA	Non-Conserved
gma-miR5043	gma-miR5043	TGTCCCCTTCTCTGCACCACC	Non-Conserved
gma-miR5037a-5p-NV_5044-5pNew	gma-miR5044	CCTCAAAGGCTTCCACTACTG	Non-Conserved
gma-miR5372	gma-miR5372	TTGTTCGATAAAACTGTTGTG	Non-Conserved
gma-miR5374-5p	gma-miR5374	TTATAGTCTGACATCTGGAAT	Non-Conserved

Table2.2 continued

miRNA name	miRNA Family	Sequence	miRNA Conservation
gma-miR5380(a,b)-NV	gma-miR5380	GAATGATGAGGATGGGGAGTAACT	Non-Conserved
gma-miR5667-3p	gma-miR5667	AAACAGATCTAAATGGATTCC	Non-Conserved
gma-miR5671b	gma-miR5671	TACCCGAATTTGCTTCCATGAT	Non-Conserved
gma-miR5770a	gma-miR5770	TTAGGACTATGGTTTGGACGA	Non-Conserved
gma-miR5770b-NV	gma-miR5770	TAGGACTATGGTTTGGACAAG	Non-Conserved
gma-miR5780a	gma-miR5780	ATCACTTAGCTGACGGTAGGGAC	Non-Conserved
gma-miR5783	gma-miR5783	GACGACGACGGGGAGGACGCGC	Non-Conserved
gma-miR828(a,b)	gma-miR828	TCTTGCTCAAATGAGTATTCCA	Conserved
gma-miR862a	gma-miR862	TGCTGGATGTCTTTGAAGGAAT	Conserved
gma-miR862b	gma-miR862	GCTGGATGTCTTTGAAGGA	Conserved
gma-miR9727	gma-miR9727	TGAAGTTACTCTGAGCACTGAG	Non-Conserved
gma-miR9730	gma-miR9730	CGATTGCTGTCATAACTGCTGC	Non-Conserved
gma-miR9731	gma-miR9731	ATACATATCGTGTTGCCAAGC	Non-Conserved
gma-miR9734	gma-miR9734	TCGTGAATGAGATTTGTGTTGCTT	Non-Conserved
gma-miR9735	gma-miR9735	TACGGCTTAAGTTCAACTTTGGAG	Non-Conserved
gma-miR9740	gma-miR9740	TGTAGGTTCCAGTGAGGGAAA	Non-Conserved
gma-miR9743-NV	gma-miR9743	TTGAGAGAGTGCTTTGAAGAAAAT	Non-Conserved
gma-miR9745	gma-miR9745	AGAATTAAATTTGGACCGTATAAC	Non-Conserved
gma-miR9749	gma-miR9749	TTAGCTTCTTTCACCTTTCCC	Non-Conserved
gma-miR9753	gma-miR9753	ACAATTTGGGACTTAGGGCTACAA	Non-Conserved
gma-miR9757	gma-miR9757	CAACCCTCCTCAGTTAGATCTC	Non-Conserved
gma-miR9763	gma-miR9763	ACCCATCCAACTCTGAAGATA	Non-Conserved
gma-miR9765	gma-miR9765	TAATACAGAATTCGGAGACAAC	Non-Conserved
gma-miR9765-NV	gma-miR9765	AATACAGAATTCGGAGACAAC	Non-Conserved

Table2.2 continued

miRNA name	miRNA Family	Sequence	miRNA Conservation
gma-miRPubNew140	gma-miRPubNew140	AGTTCCTCTGAATGCTTCATA	Non-Conserved
gma-miRPubNew169	gma-miRPubNew169	ATGTACAACTACCTTGAATGT	Non-Conserved
gma-miRPubNew187	gma-miRPubNew187	ATTTTGCGAATTTGAGGGACT	Non-Conserved
gma-miRPubNew192	gma-miRPubNew192	CAAATATAATTTGCAAGGACTA	Non-Conserved
gma-miRPubNew497	gma-miRPubNew497	TATTGGATGGAATCAGGATGGA	Non-Conserved
gma-miRPubNew569	gma-miRPubNew569	TCTTGACTTTGGACTTTTGGGT	Non-Conserved
gma-miRPubNew572	gma-miRPubNew572	TCTTGGGACAGAGGTATCAATAC	Non-Conserved
gma-miRPubNew603	gma-miRPubNew603	TGACCTGTGAAACCACCCTAT	Non-Conserved
gma-miRPubNew658	gma-miRPubNew658	TGCGCTACAGGTATAGGGACC	Non-Conserved
gma-miRPubNew811	gma-miRPubNew811	TTGGCGGAAGTAATACTAGGTA	Non-Conserved
gma-miRPubNew862	gma-miRPubNew862	TTTGTTCTGGATCCCTGTCGTC	Non-Conserved
łmiRC14	miRC14	AGATGGTTGAGGAGCGTGAGAAGG	Non-Conserved
łnovel-miRNA-81	novel-miRNA-81	AATCCGTATTGTTCTGTAAAAGGC	Non-Conserved
łnovel-miRNA-84	novel-miRNA-84	AACCGTCGTAGTTTTCATTCTCTT	Non-Conserved

Ɨ Novel miRNAs identified by our lab’s pipeline.

Table 2.3. List of stage-and subregion-specific miRNAs. that accumulated fivefold or higher in one stage or subregion level with a statistically significance (q value<0.001, Anova model).

miRNA family	Conservation of miRNA family	Stage- or subregion-specific miRNA family	Subregion where miRNA family was identified as specific*
gma-miR164	Conserved	stage	glob-ES-ES
gma-miR396-5p	Conserved	stage	glob-ES-ES
gma-miR160-5p	Conserved	stage	glob-SC-HI
gma-miR398,gma-miRPubNew170	Conserved	stage	glob-SC-HI
gma-miR5770	Non-conserved	stage	glob-SC-HI
gma-miR160-5p	Conserved	stage	glob-SC-OI
gma-miR4414-5p	Non-conserved	stage	em-AX-ST
miRC14	Non-conserved	stage	em-COT-ADEPD
miRC14	Non-conserved	stage	em-COT-ADPY
miRC14	Non-conserved	stage	em-SCPY No SCHG

Table2.3 continued

miRNA family	Conservation of miRNA family	Stage or subregion specific miRNA family	Subregion where miRNA family was identified as specific*
gma-miR408-3p	Conserved	subregion	glob-SC-II
gma-miR408-5p	Conserved	subregion	glob-SC-II
gma-miR9740	Non-conserved	subregion	hrt-ES-ES
gma-miR9753	Non-conserved	subregion	hrt-ES-ES
gma-miR393-5p	Conserved	subregion	hrt-SC-HI
gma-miR156	Conserved	subregion	cot-EP-EP
gma-miR160-3p	Conserved	subregion	cot-EP-EP
gma-miR171j-5p	Conserved	subregion	cot-EP-EP
gma-miR5380	Non-conserved	subregion	cot-ES-ES
gma-miR5770	Non-conserved	subregion	cot-ES-ES
gma-miR9740	Non-conserved	subregion	cot-ES-ES
gma-miR9753	Non-conserved	subregion	cot-ES-ES
miRC14	Non-conserved	subregion	cot-ES-ES
gma-miRPubNew140	Non-conserved	subregion	cot-SC-EPD
gma-miR2118a-3p,gma-miR482a-3p	Conserved	subregion	cot-SC-II
gma-miR9765	Non-conserved	subregion	cot-SC-II
gma-miR171j-5p	Conserved	subregion	em-AX-SAM
gma-miR171j-5p	Conserved	subregion	em-AX-PL
gma-miR171j-5p	Conserved	subregion	em-AX-EPD
gma-miR160-3p	Conserved	subregion	em-AX-VS
gma-miR156	Conserved	subregion	em-AX-ST
gma-miR4414-5p	Non-conserved	subregion	em-AX-ST
gma-miR171j-5p	Conserved	subregion	em-COT-ADEPD
gma-miR171j-5p	Conserved	subregion	em-COT-ABEPD
gma-miR5380	Non-conserved	subregion	em-ES-ES
gma-miR5770	Non-conserved	subregion	em-ES-ES
gma-miR9753	Non-conserved	subregion	em-ES-ES
gma-miR164	Conserved	subregion	em-SC-PS
gma-miR4407	Non-conserved	subregion	em-SC-HI

*The abbreviations used this column are same as Table2.1.

Table 2.4. Summary of miRNA family target mRNA annotations.

miRNA family	Target Annotation	Number of target mRNAs	Total number of target mRNAs
Conserved (45 families)	transcription factor	131	493
	non-transcription factor	362	
Non-conserved (39 families)	transcription factor	1	106
	non-transcription factor	105	

Table 2.5. Classifications of miRNA family-target mRNAs pair based on Spearman correlation coefficients values.

Category	Number of miRNA family-target mRNA pairs
Strong negative correlation (less than -0.5)	46
Weak negative correlation (-0.5 to -0.2)	94
No correlation (-0.2 to 0.2)	252
Weak positive correlation (0.2 to 0.5)	103
Strong positive correlation (more than 0.5)	23

Table 2.6. Summary of miRNA sensor results

Target mRNA sequence inserted into the constructs.	miRNA family targeting mRNAs	Promoter used to drive GFP	Number of protoplasts with no target construct screened	Number of protoplasts with mutated target construct screened	Number of protoplasts with target construct screened	Reduction of GFP in protoplasts with target constructs relative to no target control*	Reduction of GFP in protoplasts with target constructs relative to mutated target control*
SPL2/9-1	gma-miR156	35S	111	77	125	TRUE	TRUE
SPL2/9-2	gma-miR156	35S	57	62	51	TRUE	TRUE
MYB33/65-1	gma-miR159a-3p,gma-miR319	35S	39	55	101	TRUE	TRUE
MYB33/65-2	gma-miR159a-3p,gma-miR319	35S	57	106	104	TRUE	TRUE
NAC100-1	gma-miR164	35S	57	59	48	FALSE	FALSE
NAC100-2	gma-miR164	35S	57	57	43	FALSE	FALSE
NAC100-1	gma-miR164	OLE	51	32	44	TRUE	FALSE
NAC100-2	gma-miR164	OLE	109	104	109	FALSE	FALSE
HAM3-1	gma-miR171(b-3p,c-3p,d)	35S	57	59	105	FALSE	TRUE
HAM3-2	gma-miR171(b-3p,c-3p,d)	35S	85	100	102	FALSE	FALSE
HAM3-1	gma-miR171(b-3p,c-3p,d)	OLE	109	106	101	TRUE	TRUE
HAM3-2	gma-miR171(b-3p,c-3p,d)	OLE	107	101	100	FALSE	FALSE
BRIZ-1	gma-miR167-5p	35S	82	91	96	TRUE	FALSE
BRIZ-2	gma-miR167-5p	35S	103	109	108	TRUE	FALSE
ROS1-1	gma-miR398,gma-miRPubNew170	35S	82	101	102	TRUE	TRUE
ROS1-2	gma-miR398,gma-miRPubNew170	35S	103	106	101	TRUE	FALSE

*TRUE: showing the reduction of GFP signal with a statistical significance (Student t-test, P <0.05). FALSE: showing the reduction of GFP signal with a statistical significance (Student t-test, P≥0.05).

Table 2.7. List of enriched GO terms for the cluster.

Stage	Stage	category	term	q value
Heart	Cluster6	GO:0009651	response to salt stress	1.46E-05
		GO:0034976	response to endoplasmic reticulum stress	2.57E-05
		GO:0051788	response to misfolded protein	2.63E-05
		GO:0046686	response to cadmium ion	2.03E-04
		GO:0009408	response to heat	1.72E-03
		GO:0071472	cellular response to salt stress	1.64E-02
		GO:0050826	response to freezing	2.51E-02
		GO:0035196	production of miRNAs involved in gene silencing by miRNA	3.77E-02
		GO:0042542	response to hydrogen peroxide	4.44E-02
Cotyledon	Cluster3	GO:0034976	response to endoplasmic reticulum stress	1.91E-11
		GO:0046686	response to cadmium ion	4.29E-11
		GO:0051788	response to misfolded protein	5.75E-11
		GO:0009651	response to salt stress	1.99E-07
		GO:0009408	response to heat	2.47E-05
		GO:0042542	response to hydrogen peroxide	1.62E-04
		GO:0009615	response to virus	4.62E-03
		GO:0001666	response to hypoxia	9.99E-03
		GO:0009644	response to high light intensity	1.51E-02
		GO:0035196	production of miRNAs involved in gene silencing by miRNA	3.70E-02
		GO:0050687	negative regulation of defense response to virus	4.74E-02
Early Maturation	Cluster11	GO:0031349	positive regulation of defense response	3.36E-04
		GO:0010332	response to gamma radiation	1.32E-03
		GO:0035196	production of miRNAs involved in gene silencing by miRNA	2.36E-03
		GO:1900150	regulation of defense response to fungus	3.07E-03
		GO:0009742	brassinosteroid mediated signaling pathway	1.06E-02
		GO:0050826	response to freezing	1.07E-02
		GO:0009410	response to xenobiotic stimulus	2.27E-02

Table 2.8. List of the target mRNAs for the subregion-specific miRNA families for endosperm.

miRNA family	Target ID	Spearman Correlation Coefficient*	Arabidopsis homolog ID	Arabidopsis homolog gene name	Arabidopsis homolog gene annotataion	Function summary
gma-miR5770	Glyma.04G200400	-0.3951	AT5G50920	CLPC, ATHSP93-V, HSP93-V, DCA1, CLPC1	CLPC homologue 1	Required for iron homeostasis (58)
	Glyma.06G165200	-0.2507				
	Glyma.01G062400	0.0649	AT1G31710	(CUAOα3, AtCuAO2)	Copper amine oxidase family protein	Induced by methyl jasmonate (57)
	Glyma.17G019300	NA				
	Glyma.05G139700	0.2495	AT2G01600	*AtECA1, PICALM1A*	ENTH/ANTH/VHS superfamily protein	salt stress signal (59)
	Glyma.08G095000	0.0477				
	Glyma.03G005400	0.2857	AT1G33110		MATE efflux family protein	A homologue involves detoxication in Arabidopsis.(86)
	Glyma.20G049300	0.2637	AT4G12640	AtSpen2	RNA recognition motif (RRM)-containing protein	Expressed in mature rice endosperm(87)
gma-miR9740	Glyma.02G146500	-0.1636	AT5G36740		Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein	NA
miRC14	Glyma.15G107900	0.055	AT4G04640	ATPC1	ATPase, F1 complex, gamma subunit protein	NA
	Glyma.17G044000	-0.1143	AT3G43300	AtMIN7, BEN1	HOPM interactor 7	Defense against bacterium infection (88)

* The values on this column indicate that Spearman correlation coefficient that was used for global comparisons between the levels of miRNA family-target mRNA pair shown in Figure 2.3.-A.

Supplemental Figures and Tables

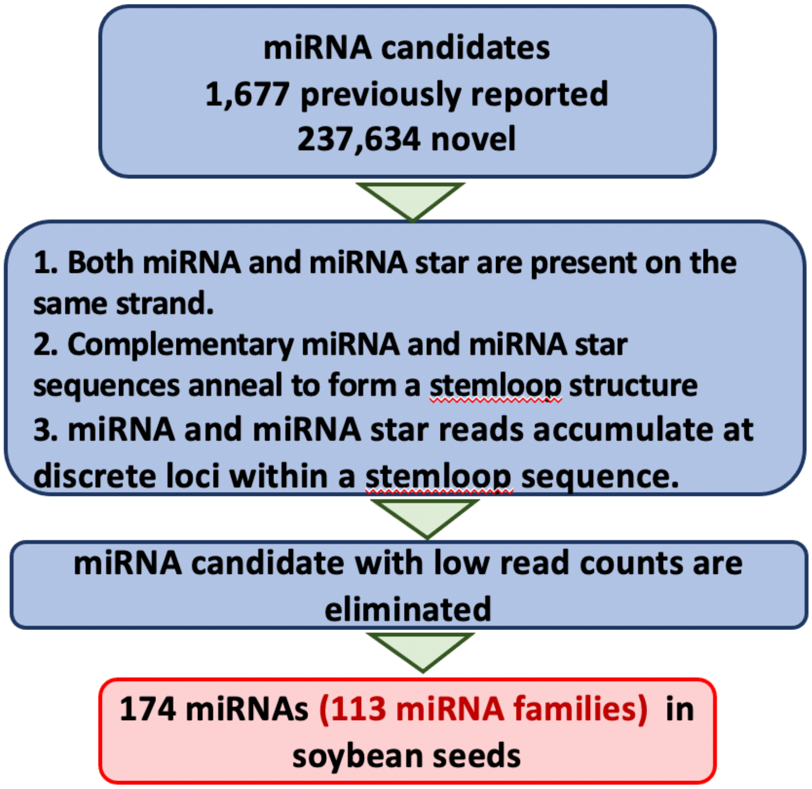
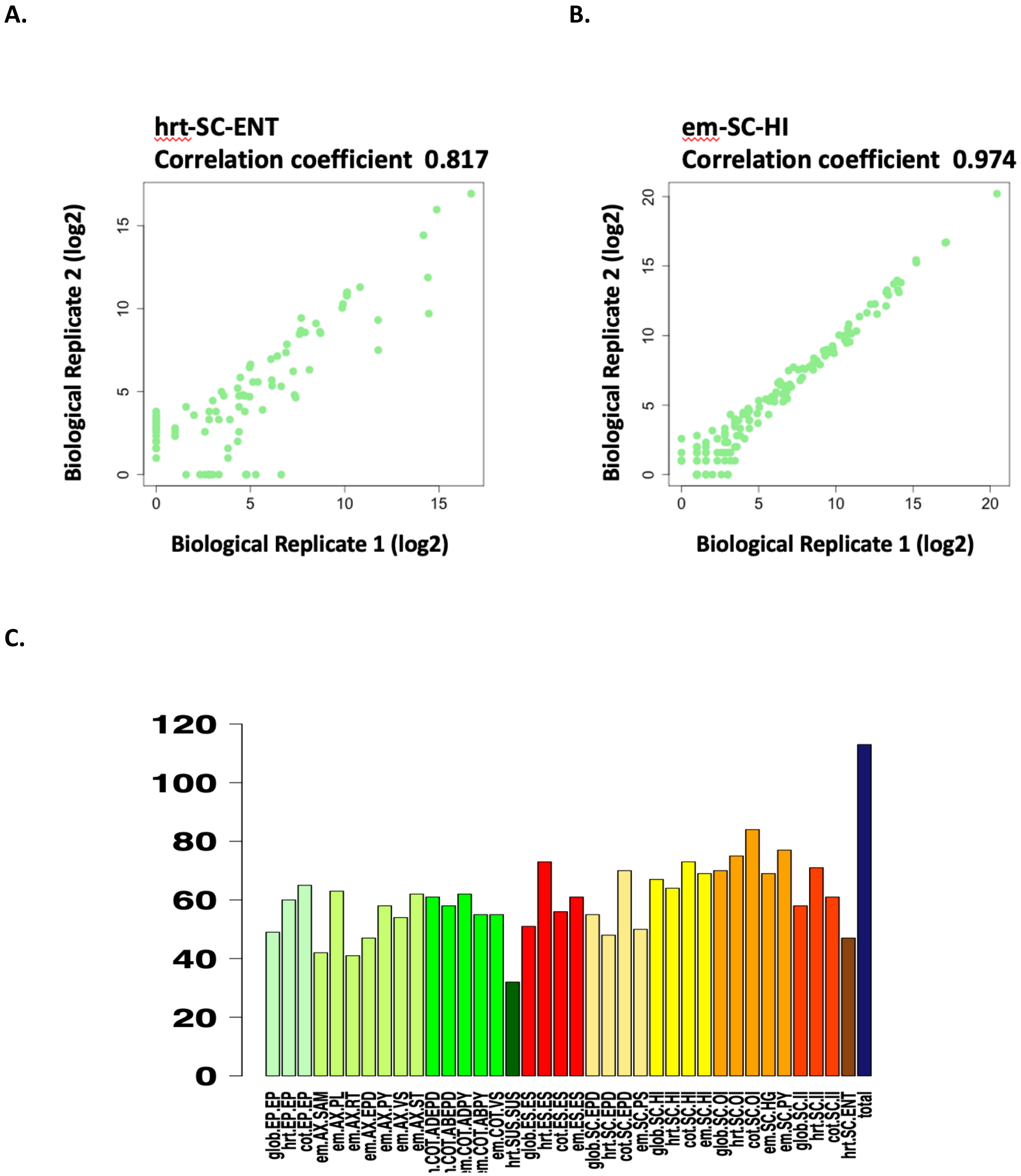

miRNA candidates
1,677 previously reported
237,634 novel
1. Both miRNA and miRNA star are present on the same strand.
2. Complementary miRNA and miRNA star sequences anneal to form a stemloop structure
3. miRNA and miRNA star reads accumulate at discrete loci within a stemloop sequence.
miRNA candidate with low read counts are eliminated
174 miRNAs (113 miRNA families) in soybean seeds

Figure S2.1. Flow chart for miRNA evaluation. The detailed descriptions are in Materials and Method section.


A.
B.
hrt-SC-ENT
Correlation coefficient 0.817
Biological Replicate 2 (log2)
Biological Replicate 1 (log2)
0
5
10
15
em-SC-HI
Correlation coefficient 0.974
Biological Replicate 2 (log2)
Biological Replicate 1 (log2)
0
5
10
15
20
C.
120
100
80
60
40
20
0
glob.EP.EP
hrt.EP.EP
cot.EP.EP
em.AX.SAM
em.AX.PL
em.AX.RT
em.AX.EPD
em.AX.PY
em.AX.VS
em.AX.ST
n.COT.ADEPD
n.COT.ABEPD
em.COT.ADPY
em.COT.ABPY
em.COT.VS
hrt.SUS.SUS
glob.ES.ES
hrt.ES.ES
cot.ES.ES
em.ES.ES
glob.SC.EPD
hrt.SC.EPD
cot.SC.EPD
em.SC.PS
glob.SC.HI
hrt.SC.HI
cot.SC.HI
em.SC.HI
glob.SC.OI
hrt.SC.OI
cot.SC.OI
em.SC.HG
em.SC.PY
glob.SC.II
hrt.SC.II
cot.SC.II
hrt.SC.ENT
total

D.

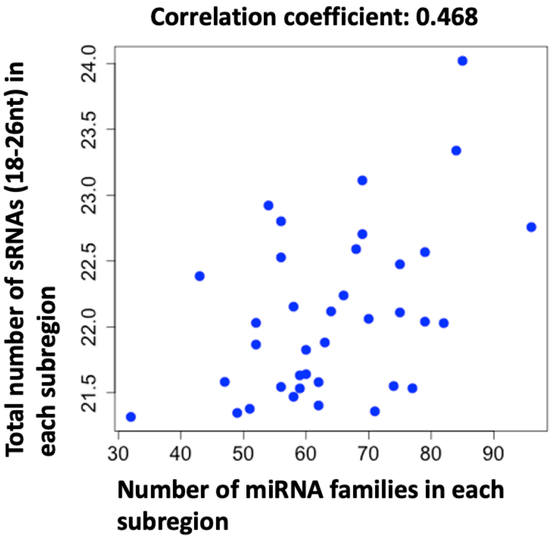
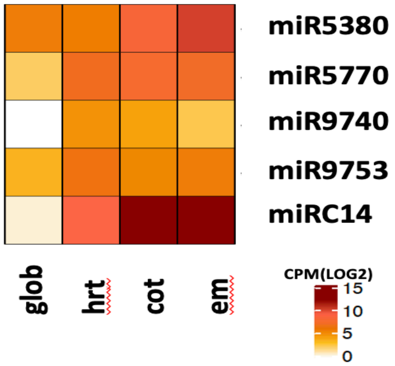


Figure S.2.2. In order to evaluate the quality of sRNA-Seq libraries, Pearson correlation coefficients for miRNA levels are calculated between two biological replicates in each subregion. Correlation coefficient values for all the subregions are listed in Table S2.3. (A) Scatter plot for the seed coat endothelium at the heart stage. The correlation coefficient for this subregion shows the lowest value among all the samples. (B) Scatter plot for the seed coat hilum at the early maturation stage. The correlation coefficient for this subregion shows the highest value among all the samples. (C) Number of miRNA families after sampling the same number of reads in subregions and stages. (D) Scatter plot showing the relationship between miRNA family numbers and the sequence depth in subregions. A value of Pearson correlation coefficient is indicated on the top. The sequence depth is the average of the total sRNA reads in each

subregion. Abbreviations for stages: glob (globular stage), hrt (heart stage), cot (cotyledon stage), em (early maturation). Abbreviations for subregions: ABEPD (abaxial epidermis), ABPY (abaxial parenchyma), ADEPD (adaxial epidermis), ADPY (adaxial parenchyma), AX (axis), COT (cotyledon), ENT (endothelium), EP (embryo proper), EPD (epidermis), ES (endosperm), HG (hourglass), HI (hilum), II (inner integument), OI (outer integument), PL (plumule), PS (palisade), PY (Parenchyma), RT (root tip), SAM (shoot apical meristem), SC (seed coat), ST (stele), SUS (suspensor), VS (vasculature).



S Figure 2.3. Temporal accumulations of the subregion-specific miRNA families for the endosperm subregions**.** gma-miR5380 and gma-miR5770 were identified for the cotyledon and early maturation stages. gma-miR9740 was identified for the heart and cotyledon stages. gma-miR9753 was identified for the heart, cotyledon and early maturation stages. miRC14was identified for the cotyledon stage. Abbreviations: glob (globular stage), hrt (heart stage), cot (cotyledon stage) and em (early maturation stage).

Table S2.1. Summary of sRNA-Seq libraries.

Libraries	raw reads	Reads aligned to soybean Genome	18-26 nt reads
glob-EP-EP-1	33,809,048	6,197,716	3,910,070
glob-EP-EP-4	30,947,029	6,720,147	4,656,752
hrt-EP-EP-1	32,562,435	8,424,975	6,313,088
hrt-EP-EP-2	21,895,927	5,395,011	3,583,430
cot-EP-EP-1	39,033,085	7,787,714	5,978,712
cot-EP-EP-2	52,203,194	9,699,238	6,650,810
em-AX-SAM-1	36,251,224	9,593,591	6,658,413
em-AX-SAM-2	43,934,142	12,569,916	9,229,866
em-AX-PL-1	25,240,527	9,427,005	8,105,594
em-AX-PL-2	29,650,323	11,782,314	10,021,928
em-AX-RT-1	71,506,553	7,706,516	4,115,683
em-AX-RT-2	34,046,061	9,137,633	6,838,660
em-AX-EPD-1	33,098,989	6,064,116	4,473,300
em-AX-EPD-2	41,828,172	5,280,024	3,169,567
em-AX-PY-1	18,533,353	4,632,219	3,124,279
em-AX-PY-2	22,735,051	4,651,503	3,415,387
em-AX-VS-1	81,136,158	17,258,373	8,252,150
em-AX-VS-2	23,576,902	5,709,334	3,836,215
em-AX-ST-1	32,022,729	7,658,744	5,831,750
em-AX-ST-2	30,374,688	9,209,549	7,832,771
em-COT-ADEPD-1	40,203,295	5,642,938	2,883,171
em-COT-ADEPD-2	36,060,623	4,859,196	2,663,350
em-COT-ABEPD-1	43,542,568	7,126,289	3,750,962
em-COT-ABEPD-2	43,713,317	7,517,803	3,682,904
em-COT-ADPY-1	36,723,837	5,235,123	3,095,305
em-COT-ADPY-2	31,292,074	5,295,243	3,175,838
em-COT-ABPY-1	29,178,960	4,200,999	2,569,099
em-COT-ABPY-2	28,383,731	5,412,437	3,498,050
em-COT-VS-1	65,151,405	11,641,760	5,741,399
em-COT-VS-2	35,881,251	6,213,511	3,587,093
hrt-SUS-SUS-1	50,082,718	5,076,265	2,468,467
hrt-SUS-SUS-2	39,357,178	4,987,107	2,753,682

Table S2.1 continued

Libraries	raw reads	Reads aligned to soybean Genome	18-26nt reads
glob-ES-ES-3	29,800,929	4,694,536	2,796,355
glob-ES-ES-4	40,860,327	3,876,915	2,654,131
hrt-ES-ES-1	29,643,210	4,589,774	2,912,265
hrt-ES-ES-2	25,901,982	5,117,520	3,159,328
cot-ES-ES-1	58,448,412	6,905,352	3,314,369
cot-ES-ES-2	55,851,509	5,644,706	3,184,904
em-ES-ES-1	65,554,874	8,507,767	3,138,900
em-ES-ES-2	94,887,032	12,992,179	5,961,793
glob-SC-EPD-1	29,014,106	5,870,127	3,473,082
glob-SC-EPD-2	25,990,039	5,724,559	2,641,895
hrt-SC-EPD-1	36,285,367	6,085,783	2,689,344
hrt-SC-EPD-2	38,536,196	6,095,732	2,643,988
cot-SC-EPD-1	53,481,753	22,325,514	17,980,773
cot-SC-EPD-2	34,226,799	18,750,777	16,035,200
em-SC-PS-1	53,096,793	8,196,507	5,876,221
em-SC-PS-2	53,929,933	11,076,021	8,743,689
glob-SC-HI-1	49,434,908	9,045,934	5,332,445
glob-SC-HI-2	29,819,053	5,877,844	3,418,295
hrt-SC-HI-1	47,668,015	10,153,056	6,074,067
hrt-SC-HI-2	62,682,869	7,776,572	5,585,407
cot-SC-HI-1	37,927,396	7,334,256	5,881,590
cot-SC-HI-2	14,538,451	4,091,697	3,167,686
em-SC-HI-1	55,767,203	14,704,996	11,672,587
em-SC-HI-2	74,217,440	13,469,244	9,539,307
glob-SC-OI-1	38,505,384	10,321,340	7,481,351
glob-SC-OI-2	25,739,527	7,075,191	4,947,614
hrt-SC-OI-1	30,365,421	5,532,801	3,711,433
hrt-SC-OI-2	36,691,634	7,756,218	4,908,675
cot-SC-OI-1	27,471,536	6,453,318	5,008,284
cot-SC-OI-2	34,190,867	10,950,755	9,171,769
em-SC-HG-1	39,458,456	5,148,320	3,114,260
em-SC-HG-2	43,612,753	5,375,758	3,028,084
em-SC-PY-1	16,658,745	3,491,174	2,602,031
em-SC-PY-2	25,491,755	7,289,754	5,953,552

Table S2.1 continued

Libraries	raw reads	Reads aligned to soybean Genome	18-26nt reads
glob-SC-II-1	37,108,008	7,435,329	3,146,035
glob-SC-II-2	47,338,497	7,204,403	2,661,117
hrt-SC-II-1	34,347,639	5,267,597	2,810,804
hrt-SC-II-2	32,577,522	5,094,761	2,567,401
cot-SC-II-1	74,693,726	8,775,188	3,914,322
cot-SC-II-2	48,431,945	7,420,363	3,812,250
hrt-SC-ENT-1	45,610,170	6,121,525	3,084,602
hrt-SC-ENT-2	59,909,346	7,296,431	3,194,949

Table S2.2 List of the publications used for miRNA evaluations after miRBase v22 in 2016

title	Year	journal
Computational identification of novel microRNAs and targets in Glycine max (89).	2014	Molecular Biology Report
A survey of the small RNA population during far-red light-induced apical hook opening (90)	2014	Frontiers Plant Science
An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes (42)	2014	Plant Cell
Identification of microRNAs and their mRNA targets during soybean nodule development: functional analysis of the role of miR393j-3p in soybean nodulation (91).	2015	New Phytologist
Identification of MicroRNAs in Response to Different Day Lengths in Soybean Using High-Throughput Sequencing and qRT-PCR (92).	2015	PLoS One
Novel and conserved microRNAs in soybean floral whorls (93)	2016	Gene
Identification of miRNAs and their targets by high-throughput sequencing and degradome analysis in cytoplasmic male-sterile line NJCMS1A and its maintainer NJCMS1B of soybean (94).	2016	BMC Genomics
Genome-Wide Small RNA Analysis of Soybean Reveals Auxin-Responsive microRNAs that are Differentially Expressed in Response to Salt Stress in Root Apex (95).	2016	Frontiers Plant Science
Comparative Profiling of microRNA Expression in Soybean Seeds from Genetically Modified Plants and their Near-Isogenic Parental Lines (96).	2016	PLOS One
Identification of chilling-responsive microRNAs and their targets in vegetable soybean (Glycine max L.) (97).	2016	Scientific Report
Coordination of MicroRNAs, PhasiRNAs, and NB-LRR Genes in Response to a Plant Pathogen: Insights from Analyses of a Set of Soybean Rps Gene Near-Isogenic Lines(98).	2014	The Plant Genome

Table S.2.3. Summary of correlation coefficient between two biological replicates in subregions.

Subregion	Pearson correction coefficient
glob-EP-EP	0.94
hrt-EP-EP	0.95
cot-EP-EP	0.95
em-AX-SAM	0.95
em-AX-PL	0.94
em-AX-RT	0.94
em-AX-EPD	0.96
em-AX-PY	0.88
em-AX-VS	0.91
em-AX-ST	0.97
em-COT-ADEPD	0.96
em-COT-ABEPD	0.95
em-COT-ADPY	0.94
em-COT-ABPY	0.93
em-COT-VS	0.93
hrt-SUS-SUS	0.91
glob-ES-ES	0.87
hrt-ES-ES	0.95
cot-ES-ES	0.88
em-ES-ES	0.93
glob-SC-EPD	0.92
hrt-SC-EPD	0.88
cot-SC-EPD	0.96
em-SC-PS	0.94
glob-SC-HI	0.90
hrt-SC-HI	0.94
cot-SC-HI	0.96
em-SC-HI	0.97
glob-SC-OI	0.97
hrt-SC-OI	0.95
cot-SC-OI	0.97
em-SC-HG	0.95
em-SC-PY	0.96
glob-SC-II	0.95
hrt-SC-II	0.96
cot-SC-II	0.96
hrt-SC-ENT	0.82

Table S.2.4. Soybean PARE libraries used for miRNA family target analysis.

GEO ID	Library Description	Library Name
GSM1419377	10 days nodule	10Dnod-GSM1419377_PARE
GSM1419378	15 days nodule	15Dnod-GSM1419378_PARE
GSM1419379	20 days nodule	20DNod-GSM1419379_PARE
GSM1419380	25 days nodule	25DNod-GSM1419380_PARE
GSM1419381	30 days nodule	30DNod-GSM1419381_PARE
GSM1419382	Anthers, rep 1	Ant-GSM1419382_PARE-1
GSM1419383	Anthers, rep 2	Ant-GSM1419383_PARE-2
GSM647200	Seed developmental stage: 15 days after flowering	cotWS-GSM647200_PARE
GSM848963	Embryo cotyledon in early maturation (green 25-50 mg)	emCOT-GSM848963_PARE
GSM848964	Seed coat in early maturation (green 25-50 mg)	emSC-GSM848964_PARE
GSM1419388	Opened flowers, rep 1	OF-GSM1419388_PARE-1
GSM1419389	Opened flowers, rep 2	OF-GSM1419389_PARE-2
GSM1419384	Ovaries, rep 1	Ova-GSM1419384_PARE-1
GSM1419385	Ovaries, rep 2	Ova-GSM1419385_PARE-2
GSM1419386	Flower buds, rep 1	UOF-GSM1419386_PARE-1
GSM1419387	Flower buds, rep 2	UOF-GSM1419387_PARE-2
GSM1419391	Well-watered soybean leaves; line IA3023	WS1-GSM1419391_PARE
GSM1419393	Well-watered soybean leaves; line LD003309	WS2-GSM1419393_PARE
GSM1419390	Drought stressed soybean leaves; line IA3023	WW1-GSM1419390_PARE
GSM1419392	Drought stressed soybean leaves; line LD00330	WW2-GSM1419392_PARE

Table S.2.5. List of genes used for miRNA sensor construct with miRNA family-target mRNA pair Spearman correlation coefficient.

miRNA	Target	Spearman Correlation Coefficient*	Target annotation	miR-target relationship
miR156	Glyma.02G177500	0.26	SPL9	Previously identified
	Glyma.03G143100	-0.41		
	Glyma.09G113800	0.64		
	Glyma.19G146000	0.02		
	Glyma.11G251500	0.30	SPL2	
	Glyma.18G005600	-0.50		
gma-miR159a-3p,gma-miR319	Glyma.04G125700	0.18	MYB33	Previously identified
	Glyma.06G312900	0.08	MYB65	
gma-miR164	Glyma.05G025500	0.02	NAC100	Previously identified
	Glyma.06G195500	-0.48		
	Glyma.17G101500	0.05		
gma-miR171(b-3p,c-3p,d)	Glyma.01G136300	-0.24	HAM3	Previously identified
	Glyma.03G031800	-0.20		
gma-miR167-5p	Glyma.15G005300	-0.56	BRIZ2	Novel
gma-miR398,gma-miRPubNew170	Glyma.10G065900	-0.02	DML1	Novel

* The values on this column indicate that Spearman correlation coefficient that was used for global comparisons between the levels of miRNA family-target mRNA pair shown in Figure 2.3-A.

Chapter 3

Expanding the LEAFY COTYLEDON 1-mediated gene regulatory network that controls seed maturation program in soybean.

Rie Uzawa[1], Julie Marie Pelletier[1], Leonardo Jo[1], Robert B. Goldberg[2] and John J. Harada[1]

[1]Department of Plant Biology, University of California, Davis, CA 95616 and Graduate Program in Plant Biology, University of California, Davis, CA 95616

[2]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095

Abstract

Soybean seed is a major source of oils and proteins to feed humans and animals. These important molecules are produced during the seed maturation phase. The LEC1 transcription factor is a master regulator of the seed maturation program. Previously, LEC1 was shown to regulate maturation processes in combination with other transcription factors, AREB3, bZIP67 and ABI3, collectively abbreviated "LAZA". GRF5 and HB22 are transcription factors that are directly regulated targets of LAZA. GRF5 and HB22 have been shown to be involved in chloroplast proliferation and seed longevity in Arabidopsis, respectively. However, the function of these genes in the seed maturation program is not known. To expand the maturation gene regulatory network, we identified the target genes of GRF5 and HB22 by chromatin immunoprecipitation-DNA sequencing experiments (ChIP-Seq). We found that there is a major overlap of the target genes among LAZA, GRF5 and HB22.The functions of the common targets are related to the seed maturation program. Furthermore, we found that GRF5 binding sites are located close to those of the LAZA, whereas HB22 binding sites are distal from LAZA binding sites. Our results suggest that the LAZA, GRF5 and HB22 act combinatorially to regulate maturation processes, possibly through their direct interactions.

Introduction

Seed development consists of two temporal phases: morphogenesis and maturation. During the morphogenesis phase, the embryo undergoes cell divisions and differentiations to establish the basic body plan of the plant. The morphogenesis phase is followed by the maturation phase when the embryo undergoes cell expansion by accumulating storage molecules. During the maturation phase, the seed acquires desiccation tolerance to withstand seed desiccation, followed by a period of the developmental arrest before germination (1,2).

Soybean seeds store high levels of protein and oil which are required for seed viability. Oil from soybean seeds constitutes more than 50% of the world's vegetable oil production. It is one of the most important crops for food supplies and commercial applications (3). Storage proteins and oils are produced during the seed maturation phase under the tight control of genes.

LEAFY COTYLEDON 1 (LEC1) encodes a seed-specific transcription factor (TF) that regulates various biological processes in seed development, including the maturation program. Arabidopsis plants homozygous for a loss-of-function *lec1* mutation showed desiccation intolerance (4) and a reduction of storage proteins and lipids (5,6). Overexpressed *LEC1* induced storage protein gene expression in seedlings, suggesting LEC1 is a key regulator of the seed maturation program. Loss-of-function *lec1* mutants in *Arabidopsis* also showed morphological defects at earlier stages, including altered morphology of the embryo, and trichome growth on the cotyledons which is not normally observed in the wild type embryo (7), suggesting the embryonic identity was affected. Thus, LEC1 is involved in diverse biological processes during seed development.

LEC1 can form a protein complex with other TFs to regulate various biological processes. *LEC1* is the B subunit (NF-YB) of the CCAAT binding NF-Y TF complex that is widely conserved in eukaryotes, including animals and plants. The NF-Ys complex consists of three subunits, NF-YA, NF-YB and NF-YC (8). Each subunit comprises a family of up to ten genes. NF-Y TFs are involved in controlling diverse biological processes throughout the plant lifecycle through the interaction of different combinations of NF-Y subunits. (9). Furthermore, NF-Y TFs can interact with non-NF-Y TF subunits to regulate diverse biological processes (9,10). For example, NF-YB and NF-YC interact with flowering regulator, CONSTANS (CO) to promote flowering (11,12). LEC1 and the LEC1 homolog (13), LEC1-LIKE (L1L), were shown to interact with bZIP67 along with NF-YC2 to activate the promoter of the storage protein gene, *CRUCIFFERIN C (CRC)* by a transactivation assay in Arabidopsis mesophyll protoplasts (14). L1L and NF-YC2 were also shown to interact with bZIP67 to activate the promoter of the fatty acid synthesis gene, *FATTY ACID DESATURASE3 (FAD3)* in the same study (15). Although endogenous LEC1 and L1L do not function redundantly due to the temporal expression differences, ectopically expressed L1L was able to replace LEC1 functions (13). Thus, LEC1 is likely able to activate the promoter of *FAD3* as well.

Previously, LEC1was shown to regulate distinct gene sets for different biological processes sequentially in both the Arabidopsis and soybean seed development, suggesting evolutional conservation in gene regulations for seed development processes (16). Furthermore, LEC1-mediated temporal gene regulation is governed by interactions between LEC1 and three other transcription factors, ABA-RESPONSIVE ELEMENT BINDING PROTEIN 3 (AREB3), BASIC LEUCINE ZIPPER67 (bZIP67) and ABA INSENTIVE3 (ABI3). The genes involved in the morphogenesis and photosynthesis program are regulated either by LEC1 only or LEC1 with AREB3 (abbreviated LA), whereas the genes involved in photosynthesis and GA signaling are regulated by LEC1 with AREB3, and bZIP67 (abbreviated LAZ) as well as the LEC1 with AREB3, bZIP67, and ABI3 (abbreviated LAZA). Finally, genes involved in the maturation programs are regulated by LAZA as well (17).

GROWTH-REGULATING FACTOR 5 (GRF5) and HOMEOBOX22 (HB22) were identified as LAZA complex target genes in chromatin immunoprecipitation-DNA sequencing

(ChIP-Seq) analysis in early maturation soybean seeds (17). *GRF5* encodes a plant-specific TF. Arabidopsis *grf5* mutants show reduction of leaf size, especially in width, as well as in the number of the cells in leaf parenchyma cells. Arabidopsis plants that overexpress GRF5 show an increase in leaf size caused by increased cell expansion (18). Vercruyssen et al. (19) showed that overexpressed GRF5 induced cell proliferation in a cytokinin-depending manner. Additionally, they showed that Arabidopsis with overexpressed GRF5 increased the number of chloroplasts per cell, photosynthesis efficiency and leaf longevity.

HOMEOBOX22 (HB22) is a member of the Class I HD-Zip (HD-Zip I) TF family (20). The expression of many Class I HD-Zip TFs is induced by abiotic stress and light in *Arabidopsis thaliana* (21). Although loss-of-function homozygous mutant *hb22* Arabidopsis plants do not show any detectable mutant phenotypes, double loss-of-function mutations in the *HB22* and the closely related *HB25* HD-Zip TFs improve seed longevity by positively regulating GA biosynthesis, likely in the seed coat. Furthermore, *hb22* and *hb25* double mutants generated smaller seeds relative to wild type (22).

The functions of both GRF5 and HB22 have been reported in *Arabidopsis.* However, the involvement of GRF5 and HB22 in the seed maturation program has not been well-studied in plants, including soybean seeds. In order to expand the LEC1-mediated seed maturation gene regulatory network in soybean seeds, I used ChIP-Seq to identify GRF5 and HB22 target genes. I showed that GRF5 and HB22 share target genes with LAZA. Furthermore, GRF5 regulates these target genes by binding closely to LAZA binding sites, whereas HB22 binds away from LAZA binding sites. These results suggest that GRF5 may regulate the same maturation processes with the LAZA.

Results

Identification of GRF5 and HB22 target genes in early maturation soybean embryos.

Approximately 75% of the genes in the soybean genome are present in multiple copies due to two whole genome duplications that took place 13 and 59 million years ago (23). Both GRF5 and HB22 have several gene copies. Based on our transcriptomic data of soybean seeds from the different developmental stages (GEO accession: GSE99571), we focused our study on four GRF5 (Glyma.07G038400, Glyma.16G007600, Glyma.01G23440, Glyma.11G008500) and two HB22 (Glyma.04G093300, Glyma.06G095200) homologs based on the expression patterns of these genes. The mRNAs of *GRF5* and *HB22* homologs accumulate during the maturation phase (Figure 3.1).

To expand the LAZA gene regulatory network for the seed maturation program, we identified genes bound by GRF5 and HB22. To this end we performed ChIP-Seq with embryos at the early maturation stage as described by Pelletier et al. (16). We generated two polyclonal antibodies against GRF5 and HB22 peptides listed in Table S1. We generated two antibodies targeting separate peptides of the same TF to ensure that each antibody identify the same bound genes and verify the specificities of the GRF5 and HB22 antibodies. ChIP-seq peaks were identified from two biological replicates with a statistical significance as described in Jo et al. (17). We followed ENCODE guidelines for quality control as described in Figure S3.1 and Table S3.2 (24). All but the ChIP-seq data with GRF5-b antibody exceeded the quality guidelines.

Because the GRF5-b antibody did not yield interpretable data, I used a different approach to confirm the bound gene data using the GRF5-a antibody. A plasmid with a MYC-tagged GRF5-2 (Glyma.16G007600) construct under the control of 35S promoter (p35S:4xMYC-GRF5) was transfected into early maturation soybean embryo protoplasts, and chromatin bound by GRF5 was pulled down with an anti-MYC antibody for ChIP-Seq analysis. Preliminary ChIP-seq data with the MYC antibody did not meet ENCODE quality guidelines (Figure S3.1-B and Table S3.2). Thus, ChIP-qPCR assays were performed to quantify the enrichment of the genes bound by the GRF5. We used the ChIP peak information for the bound genes from the ChIP-seq data with GRF5-a antibody mentioned above. As a negative control, I used ChIP DNA from early maturation embryo protoplast transfected with overexpressed LEC1 plasmids (35S:LEC1) and immunoprecipitated with the MYC antibody. I tested a total of 28 bound loci, including 22 loci upstream of the genes and 6 intergenic loci with two biological replicates for each locus (Figure S3.2, primer information in Table S.3.3). MYB-GRF5 lines show enrichment for all 28 loci, whereas no significant enrichments were observed in the negative control except only one of the two biological replicates from intergenic region 1. It is possible that the result for the intergenic region 1 may be an artifact of either ChIP or qPCR quantification. The qPCR quantification results provided support for the specificity of the GRF5-a antibody.

The binding by a TF does not always result in transcriptional regulation by the TF (25). In order to identify the transcriptionally regulated genes by GRF5 and HB22, we defined a target gene as being both bound by and coexpressed with the TF. We designate "bound" genes as those with a ChIP-seq peak within 1kb upstream of the transcriptional start site (TSS), whereas "coexpressed" genes have mRNA levels accumulated at a 5-fold or higher level in embryo subregions than in the seed coat subregions ($q<0.01$) as described in Jo et al. (17). We used the gene coexpresson as one of the criteria to define target genes because all the members of LAZA as well as GRF5 and HB22 are predominately expressed in the embryo at a 5-fold or higher level than in the seed coat (Figure S3.3). The numbers and the overlap of bound genes and coexpressed genes of HB22-1, HB22-2, and GRF5-1 are shown in Figure 3.2-A. For HB22, bound and target genes identified from both ChIP-seq samples show significant overlaps (Figure 3.2-B).

I used *de novo* DNA motif discovery algorithms to identify the enriched DNA sequence motifs in the promoter region of 600 bound genes for GRF5-1 and HB22-1 with the highest peak enrichments. Figure 3.3 shows three DNA sequence motifs each for GRF5 and HB22 with highest significance. These motifs include ones for GRF6 and HB34 binding, which are the TFs in the same gene families of GRF5 and HB22, respectively. These results further validated GRF5 and HB22 binding specificities.

Taken together, we were able to identify and validate GRF5 and HB22 target gene using stringent criteria.

Underlying functions of GRF5 and HB22 target genes in soybean seed development.

GRF5 and HB22 were identified as the target genes of LAZA TFs, which are involved in seed maturation process as well as photosynthesis and gibberellic acid (GA) signaling. GRF5 and HB22 may play important roles for these biological processes. In order to obtain insight into GRF5 and HB22 functions, I performed Gene Ontology (GO) term representation analysis (26) on GRF5 and HB22 target genes. Overrepresented GO terms with the top 10 highest significance included seed maturation processes, whereas photosynthesis and GA signaling were also overrepresented GO terms of LAZA target genes (Figure 3.4). This suggests that GRF5 and HB22 regulate similar biological processes as LAZA.

I asked if GRF5 and HB22 targeted the same set of the genes regulated by LAZA. As Figure 3.5 shows, GRF5 and HB22 shared significant numbers of target genes with LAZA.

Furthermore, I conducted GO term representation analysis for the common targets of LAZA TF, GRF5 and HB22 (LAZA-GH target gene). The overrepresented GO terms were related to maturation process, photosynthesis and GA signaling, further suggesting GRF5 and HB22 regulate the same biological processes as LAZA TFs (Table 3.1).

Interestingly, overrepresented GO terms for 471 target genes which were uniquely targeted by HB22, and not targeted by either LAZA or GRF5 (Figure 3.5), were related to photosynthesis and light responses, suggesting HB22 is a key regulator for these biological processes during the maturation phase (Table 3.2). I asked if these genes may be regulated by HB22 with LEC1, LA or LAZ as well. First, I identified 33 genes whose genes are associated with the GO term IDs for photosynthesis and light responses found in Table3.2 (GO:0009637, GO:0009657, GO:0010114, GO:0010207, GO:0015979, GO:00196840, highlighted in green) in 471 target genes mentioned above. The majority of these targets (21 genes) were also identified as the target genes regulated by HB22 with LEC1, LA, or LAZ (Table S3.4). Thus, HB22 may be responsible to regulate the photosynthetic processes with different combinations of LEC1, AREB3, bZIP67 and ABI3 in the early maturation soybean embryo.

GRF5 and HB22 binds to different genomic loci at LAZA-GH target genes.

Next, I asked where GRF5 and HB22 bound at the LAZA-GH target genes relative to LAZA binding sites since GRF5 and HB22 regulate a large subset of target genes with LAZA. First, I examined the summit of the GRF5 and HB22 ChIP-seq peak, which are presumably the GRF5 and HB22 binding sites. Previously, Jo et al. (17) showed that the binding sites for the members of LAZA are closely clustered in the promoter region of the target genes. These clusters are designated as *cis*-regulatory modules (CRM) to indicate a high occupancy of multiple TF binding sites. CRMs were operationally defined as promoter regions whose boundaries extended 100 bp on each side of the outermost ChIP peak summits within a cluster. The average size of LAZA CRMs was 240 bp, suggesting LAZA TFs generally bind very closely in CRMs.

I asked how many GRF5 and HB22 peak summits overlapped with LAZA CRMs at LAZA-GH target genes (Figure 3.6-A for the schematic figure). A total of 253 LAZA CRMs were found at 300 LAZA-GH target genes. The number of LAZA CRMs were lower than the number of LAZA-GH target genes because not all the peak summits for the members of LAZA overlap with each other. As shown in Table 3.3, 87% of the total GRF5 summits at LAZA-GH target genes were observed within LAZA CRMs whereas only 16.2% of the total HB22 summits were found in LAZA CRMs, suggesting most GRF5 TFs bound within LAZA CRM, in contrast to HB22.

To investigate how close to or distant from GRF5 or HB22 binding sites are located relative to the peak summits of the LAZA TF members, I examined the distances of GRF5 and HB22 from LEC1 binding sites (Figure 3.6-B). Figure3.6.-C shows that the distance between the LEC1 and AREB3, bZIP67 and ABI3 summits were close to each other as expected. Interestingly GRF5 summits were located close with LAZA TF summits, whereas HB22 summits were located further away from those of LAZA TFs. These results suggest that GRF5 binds closely to LAZA TFs whereas HB22 binds away from LAZA TFs.

To get more insights into GRF5 and HB22 binding sites at LAZA-GH target genes, I examined the specific DNA sequence motif enrichments of the GRF5 and HB22 summit regions at LAZA-GH target gene loci by using HOMER hypergeometric analysis (27). I chose the following five DNA sequence motifs to examine the LAZA-GH target gene loci: CCAAT box motif for LEC1 (28,29), G box motif for AREB3 and bZIP67 (30,31), RY motif for ABI3 (32), and the GRF5 and HB22 motifs that we identified in *de novo* motif discovery analysis as described above. I queried regions 100 bp upstream and downstream of GRF5 and HB22 ChIP-seq peak summits at LAZA-GH target genes to see if the DNA sequence motifs listed above were enriched relative to randomly chosen sequences of the same size. GRF5 summit regions were enriched for G box, RY and GRF5 motifs whereas HB22 summit regions were enriched primarily for the HB22 motif. This result further suggests that GRF5 binds closely with LAZA TFs whereas HB22 binds away from LAZA TFs (Figure 3.7).

Discussion

GRF5 and HB22 regulate the same set of genes as LAZA TF complex.

The major aim for identifying GRF5 and HB22 target genes was to expand the LEC1-mediated seed maturation gene regulatory network. GRF5 and HB22 were identified as LAZA targets in a previous study (17). LAZA were shown to regulate the seed maturation program as well as photosynthesis and GA signaling in the soybean early maturation embryos. Based on studies in Arabidopsis, my expectation was that GRF5 and HB22 would be specifically involved in photosynthesis and GA signaling, respectively. Surprisingly, I found that both TFs shared a significant number of target genes with LAZA: 43.7% (420 out of 962) and 33.5% (415 out of 1240) for GRF5 and HB22 targets, respectively, were in common with LAZA targets (Figure 3.5). In addition, 300 targets were shared with LAZA, GRF5 and HB22 (LAZA-GH). The LAZA-GH target genes were involved in the same biological processes as previously described for LAZA TFs (16), suggesting that GRF5 and HB22, together with LAZA play, important roles for seed maturation, photosynthesis and GA signaling.

I observed 471 target genes (38% of total HB22 target genes) were targeted by HB22, not by either LAZA or GRF5 (Figure 3.5). GO term representation analysis showed that overrepresented GO terms for these 471 target genes were related to photosynthetic processes and light responses. I identified the genes associated with these GO terms. Many of these genes were targeted by HB22 with LEC1, LA, or LAZ (Table S3.4), not with LAZA and/or GRF5, suggesting some photosynthesis gene regulatory networks may be regulated by HB22 with the members of LAZ without GRF5.

Taken together, our study of GRF5 and HB22 target genes shed light into their functions in soybean seed development, which had not been previously described.

GRF5 and HB22 binds differently at LAZA-GH target gene relative to LAZA TF complex.

Although I observed a significant overlap among LAZA, GRF5 and HB22 target genes, GRF5 and HB22 binding sites were different at LAZA-GH target loci. I showed that many more GRF5 summits were overlapped with LAZA CRMs at LAZA-GH target loci than HB22 summits (Table 3.3). This is because GRF5 binds closer to LAZA TFs whereas HB22 binds away from LAZA TFs (Figure 3.6, 3.7 and 3.8). It is tempting to speculate that GRF5 may be directly interacting with LAZA. However, I only found that GRF TFs interact with GRF-INTERACTING FACTOR1 (GIF1)/ ANGUSTIFOLIA3 (AN3), but not with other TFs or transcription activators (18). GIF1 is a member of GIF family of transcriptional coactivators that interacts with SWITCH/SUCROSE NONFERMENTATING (SWI/SNF) chromatin remodeling ATPase complex (33). Based on inducible GIF1 and tandem chromatin affinity purification followed by sequencing (TChAP-Seq), a variant of ChIP-seq analyses, GIF1 was able to bind and activate GRF5 in *Arabidopsis*. Together, this study suggests that GRF5, with GIF1 and SWI/SNF, can self-activate gene transcription by chromatin remodeling. Motif discovery analysis showed that GIF1 binding sites were enriched for G-box and GAGA motif sequences. In our soybean analyses, GRF5 homologs are identified as LAZA-GH targets. My *de novo* discovery analysis for GRF5 target sites at LAZA-GH targets enrich the G-box and GAGA motif sequences as well (S. Figure 3.4). Furthermore, LEC1 was shown to establish the active chromatin state and reactivate the expression of *FLOWERING LOCUS C (FLC)*, a negative regulator for flowering. LEC1 reactivates *FLC*, which has been silenced during the previous generation in Arabidopsis, at an early stage of embryo development (34). FLC activation was also shown to depend on the subunit of the SWR1 complex, a member of the SWITCH/SUCROSE NONFERMENTATING (SWI/SNF) chromatin remodeling ATPase complex (34,35). One possibility is that GRF5 interacts indirectly with LEC1 via chromatin remodeling proteins. Another possibility is GRF5 can bind where LEC1-mediated chromatin opening occurs. Further studies using biochemical assay for interaction of GRF5 and LEC1 will help to understand the LAZA-GH target gene regulatory mechanisms.

HD-Zip TFs can form homo-dimers or hetero-dimers within the HD-ZIP family (36), but not with other TFs (37) I showed HB22 binds away from the LAZA TFs and GRF5 at LAZA-GH target genes. HB22 may not directly interact with LAZA or GRF5 to regulate the LAZA-GH target genes. It is possible that GRF5 and LAZA interaction may allow HB22 binding at LAZA-GH target sites where LEC1 establishes open chromatin conformation. Another possibility is HB22 may interact with LAZA and GRF5 through the chromatin looping.

The remaining important question is if GRF5 and HB22 are required to control LAZA-GH target genes. One way we can test our hypothesis is to use targeted gene knockdown or knockout methods such as RNA interference or CRISPR/CAS technologies followed by ChIP-Seq analysis by using the protoplast transfection approach. (17). A further study will be necessary to answer this question.

Materials and Methods

Chromatin Immunoprecipitation-DNA Sequencing

Soybean plants were grown and seeds were harvested for ChIP experiments as described by Pelletier et al (16), and Jo et al (17). Briefly, early maturation embryos were collected and cross-linked with formaldehyde. Embryo tissues were ground for nuclei isolation. Isolated nuclei were sonicated to fragment chromatins to 100-500bp. Chromatins were incubated with the GRF5 or HB22 antibodies. Immune complexes were captured with Dynabeads-Protein A beads (Thermo Fischer Scientific). DNA was eluted, cross-links were reversed, and proteins were digested. DNA was extracted with phenol:chloroform mixture and precipitated. ChIP DNA was quantified using the Invitrogen Quant-iT PicoGreen dsDNA assay kit on a NanoDrop 3300 fluorospectrometer (Thermo Fischer Scientific).

ChIP assay was performed using peptide antibodies against soybean GRF5 and HB22 as listed in Table S3.1 with the peptide sequences. The antibodies were raised in rabbits against the peptides. The antibody specificities were tested with the peptide sequences expressed in E. coli by the western blot assay.

ChIP-seq libraries were prepared using the NuGEN Ovation Ultralow DR Multiplex System. Libraries were size-selected by electrophoresis, purified and sequenced at 50-bp single-end reads using Illumina HiSeq 2000 sequencing system. qPCR validation experiments were done in triplicate with 30 pg unamplified chromatin or 200 pg of amplified DNA.

Recombinant DNA manipulation and plasmid construction for MYC-GRF5 construct.

The coding region of Glyma.16G007600 (GRF5-2) was amplified from the early maturation embryo cDNA and cloned in-frame into the binary vector, pGWB18 (38), by using the GATEWAY technology (Invitrogen). Glyma.16G007600(GRF5-2) was fused with the 35S promoter with 4xMYC genes (35S:4xMYC:GRF5). DNA sequence of the construct was confirmed by the Sanger method.

ChIP with soybean embryo cotyledon protoplast using MYB antibody

Plasmids with the 35S:4xMYC:GRF5 construct were transfected into soybean embryo cotyledon at the early maturation stage (6-7 mm seeds in length) as described in Jo et al (17). Briefly, cotyledons from early maturation soybean embryos were cut into 0.5-1mm strips, immersed in an enzyme solution containing 1% (w/v) Cellulase RS “Onozuka" and 0.25% (w/v) Macerozyme R-10 (Yakult Pharmaceutical Industry CO., Ltd). The cotyledon tissues were vacuum infiltrated for 15 min and incubated in the dark with gentle agitation (50 rpm) for 2 hours at room temperature. Protoplasts were filtered to remove unnecessary tissues, washed twice with W5 buffer (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2mM MES pH5.8, 5 mM glucose) and incubated on ice for 30 min. After the incubation, protoplasts (approximately 5 x10^5

cells per 200 ul) were transfected with 10ug of plasmid DNA. Transfected protoplasts were washed with W5 buffer twice and incubated in W5 buffer in the light at 25 C for 16 hours.

ChIP assay was performed according to Pelletier et al. (16) and Jo et al (17) with modifications. Protoplasts were collected and cross-linked with formaldehyde. Cross-linked protoplasts were sonicated to fragment chromatin to 100-500 bp. Chromatin was incubated with the c-MYC antibody (9E10). Immune complexes were captured with Dynabeads-Protein A beads (Thermo Fischer Scientific). The subsequent procedures were the same as described above.

qPCR validation experiments were done in triplicate with 200 pg of amplified DNA. The primer information is listed in Table S3.3.

Data analyses

The data analyses were done according to Jo et al. (17). ChIP-Seq data were quality filtered and uniquely mapped to the Wm82.a2.v1 genome (Gmax275) using Bowtie v0.12.7 (39) with two mismatches allowed. Redundant reads were removed using samtools v.0.1.19 (40). Sequencing library complexity and quality were evaluated using the non-redundant fraction and strand cross-correlation analysis (phantompeakqualtools:https://code.google.com/p/phantompeakqualtools), following ENCODE guideline for ChIP-Seq quality standards (24). Quality assessment of the libraries are summarized in Table S3.2.

ChIP-Seq peaks were identified using MACS2 v2.1.0.2014616 (41) with two biological replicates with P <0.1 threshold. The estimated ChIP fragment size was independently determined for each sample by MACS2 with 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 (42) (https://github.com/nboley/idr) and the IDR threshold of 0.01. Genes bound by a TF were defined as those with a reproducible peak within a 1kb window upstream of the genes transcriptional start site (TSS).

Target gene are defined as genes that are bound as determined in ChIP-seq experiments and coexpressed with LAZA TFs as described in Pelletier et al. (16) and Jo et al (17). Briefly, mRNA of the coexpresed genes accumulated at a 5-fold or higher level in embryo subregions than in the seed coat subregions (q<0.01) based on the Harada-Goldberg Soybean Seed Development LCM RNA-Seq Datasets (GEO accessions, GSE 57606, GSE46096, and GSE99109).

Gene Ontology (GO) term representation analysis were performed using Bioconductor package GOseq, the soybean GO functional annotation, the hypergeometric method and a *q* value threshold of 0.05(43).

De novo DNA motif discovery analysis was performed with MEME-CHIP tool from the MEME suite v5.0.5 (44) with an E-value cutoff of 0.01. Default MEME discovery setting were used, except that the maximum discovered motif length was set to 10 nucleotides. Tomtom tool compared the *de novo* discovery DNA motifs to motifs found in the Arabidopsis DAP-Seq TF motif database (45) and the Human HOCOMOCOv11 database (46).

Distance between the position of the ChIP peak summits of GRF5 and HB22 were calculated at the LAZA-GH target genes using the R script. AREB3, bZIP67 and ABI3 ChIP peak summits were used as the controls.

In order to screen the annotated motifs at LAZA-GH target sites, annotated motifs for LEC1, AREB3, bZIP67 and ABI3 used in Jo et al. (17) were screened for enrichment using

HOMER (27) (homer.ucsd.edu/homer/motif/index.html), in addition to annotated motifs most similar to the *de novo* discovered motif of GRF5 and HB22, as described in Pelletier et al (16). "Random expanded summits" were generated from the randomly selected genes that were of comparable number length and position to the expanded GRF5 and HB22 summits. A significance threshold of *p* value <0.01 was used to identify significantly enriched DNA motifs.

Reference:

1. West MAL, Harada JJ. Embryogenesis in higher plants: An overview. Plant Cell. 1993;5(10):1361–9.
2. Goldberg RB, De Paiva G, Yadegari R. Plant embryogenesis: Zygote to seed. Science (80-). 1994;266(5185):605–14.
3. Chaudhary J, Patil GB, Sonah H, Deshmukh RK, Vuong TD, Valliyodan B, et al. Expanding omics resources for improvement of soybean seed composition traits. Vol 6, Frontiers in Plant Science. 2015. p 1–16.
4. Lotan T, Ohto MA, Matsudaira Yee K, West MAL, Lo R, Kwong RW, et al. Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell. 1998;93(7):1195–205.
5. Vicient CM, Bies-Etheve N, Delseny M. Changes in gene expression in the leafy cotyledon1 (lec1) and fusca3 (fus3) mutants of Arabidopsis thaliana L. J Exp Bot. 2000;51(347):995–1003.
6. Meinke DW, Franzmann LH, Nickle TC, Yeung EC. Leafy cotyledon mutants of Arabidopsis. Plant Cell. 1994;6(8):1049–64.
7. West MAL, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, et al. LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in Arabidopsis. Plant Cell. 1994;6(12):1731–45.
8. Lee H, Fischer RL, Goldberg RB, Harada JJ. Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proc Natl Acad Sci U S A. 2003;100(4):2152–6.
9. Petroni K, Kumimoto RW, Gnesutta N, Calvenzani V, Fornari M, Tonelli C, et al. The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. Vol 24, Plant Cell. 2013. p 4777–92.
10. Jo L, Pelletier JM, Harada JJ. Central role of the LEAFY COTYLEDON1 transcription factor in seed development. Vol 61, Journal of Integrative Plant Biology. 2019. p 564–80.
11. Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, et al. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. Plant Cell. 2006;18(11):2971–84.
12. Kumimoto RW, Zhang Y, Siefers N, Holt BF. NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana. Plant J. 2010;63(3):379–91.
13. Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, et al. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. Plant Cell. 2003;15(1):5–18.
14. Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J. 2009;58(5):843–56.
15. Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, et al. bZIP67 regulates the omega-3 fatty acid content of arabidopsis seed oil by activating fatty acid DESATURASE3. Plant Cell. 2013;25(8):3104–16.
16. 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. Proc Natl Acad Sci U S A. 2017;114(32):E6710–9.

17. Jo L, Pelletier JM, Hsu SW, Baden R, Goldberg RB, Harada JJ. Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. Proc Natl Acad Sci U S A. 2020;117(2):1223–32.
18. Horiguchi G, Kim GT, Tsukaya H. The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana. Plant J. 2005;43(1):68–78.
19. Vercruyssen L, Tognetti VB, Gonzalez N, Van Dingenen J, De Milde L, Bielach A, et al. Growth regulating factor5 stimulates arabidopsis chloroplast division, photosynthesis, and leaf longevity. Plant Physiol. 2015;167(3):817–32.
20. Ariel FD, Manavella PA, Dezar CA, Chan RL. The true story of the HD-Zip family. Vol 12, Trends in Plant Science. 2007. p 419–26.
21. Henriksson E, Olsson ASB, Johannesson H, Johansson H, Hanson J, Engström P, et al. Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns and phylogenetic relationships. Vol 139, Plant Physiology. 2005. p 509–18.
22. Bueso E, Muñoz-Bertomeu J, Campos F, Brunaud V, Martínez L, Sayas E, et al. ARABIDOPSIS THALIANA HOMEOBOX25 uncovers a role for gibberellins in seed longevity. Plant Physiol. 2014;164(2):999–1010.
23. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010;463(7278):178–83.
24. 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 Res. 2012;22(9):1813–31.
25. Farnham PJ. Insights from genomic profiling of transcription factors. Vol 10, Nature Reviews Genetics. 2009. p 605–16.
26. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47(D1):D330–8.
27. 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. Mol Cell. 2010;38(4):576–89.
28. Calvenzani V, Testoni B, Gusmaroli G, Lorenzo M, Gnesutta N, Petroni K, et al. Interactions and CCAAT-binding of Arabidopsis thaliana NF-Y subunits. PLoS One. 2012;7(8).
29. Gnesutta N, Saad D, Chaves-Sanjuan A, Mantovani R, Nardini M. Crystal Structure of the Arabidopsis thaliana L1L/NF-YC3 Histone-fold Dimer Reveals Specificities of the LEC1 Family of NF-Y Subunits in Plants. Vol 10, Molecular Plant. 2017. p 645–8.
30. Izawa T, Foster R, Chua NH. Plant bZIP protein DNA binding specificity. J Mol Biol. 1993;230(4):1131–44.
31. Kim SY, Chung HJ, Thomas TL. 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. 1997;11(6):1237–51.
32. Mönke G, Altschmied L, Tewes A, Reidt W, Mock HP, Bäumlein H, et al. Seed-specific transcription factors ABI3 and FUS3: Molecular interaction with DNA. Planta. 2004;219(1):158–66.
33. Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han SK, et al. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during Arabidopsis leaf development. Plant Cell. 2014;26(1):210–29.

34. Tao Z, Shen L, Gu X, Wang Y, Yu H, He Y. Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. Nature. 2017;551(7678):124–8.

35. Choi K, Park C, Lee J, Oh M, Noh B, Lee I. Arabidopsis homologs of components of the SWR1 complex regulate flowering and plant development. Development. 2007;134(10):1931–41.

36. Meijer AH, De Kam RJ, D'Erfurth I, Shen W, Hoge JHC. HD-Zip proteins of families I and II from rice: Interactions and functional properties. Mol Gen Genet. 2000;263(1):12–21.

37. Sessa G, Morelli G, Ruberti I. The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. EMBO J. 1993;12(9):3507–17.

38. Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, et al. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng. 2007;104(1):34–41.

39. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3).

40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

41. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9).

42. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. Ann Appl Stat. 2011;5(3):1752–79.

43. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2).

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

45. O'Malley RC, Huang SSC, 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.

46. Kulakovskiy I V., Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, 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. 2018;46(D1):D252–9.

Figures and Tables

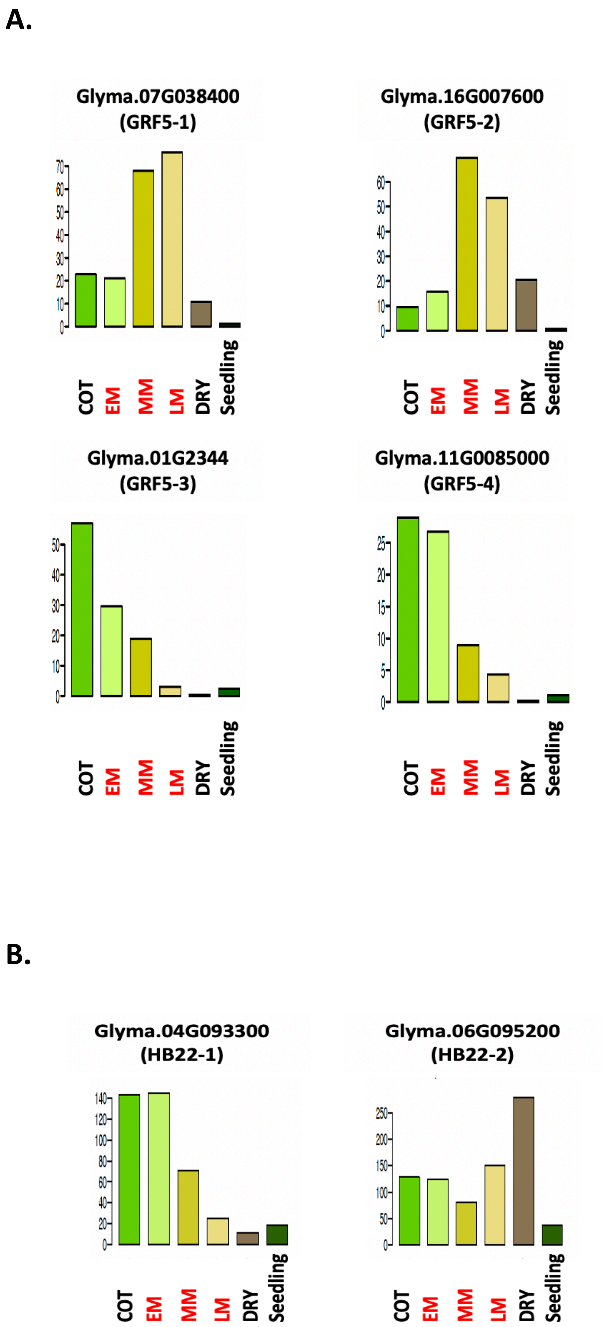


Figure 3.1. Temporal mRNA accumulation of GRF5 and HB22 homologs in soybean seeds and seedlings. RNA-seq data were obtained from GEO accession: GSE99571. (A) *GRF5* homolog expression. (B) *HB22* homolog expression. COT (cotyledon stage), EM (early maturation stage), MM (mid maturation stage), DRY (desiccation stage), and Seedling (seedling after germination).

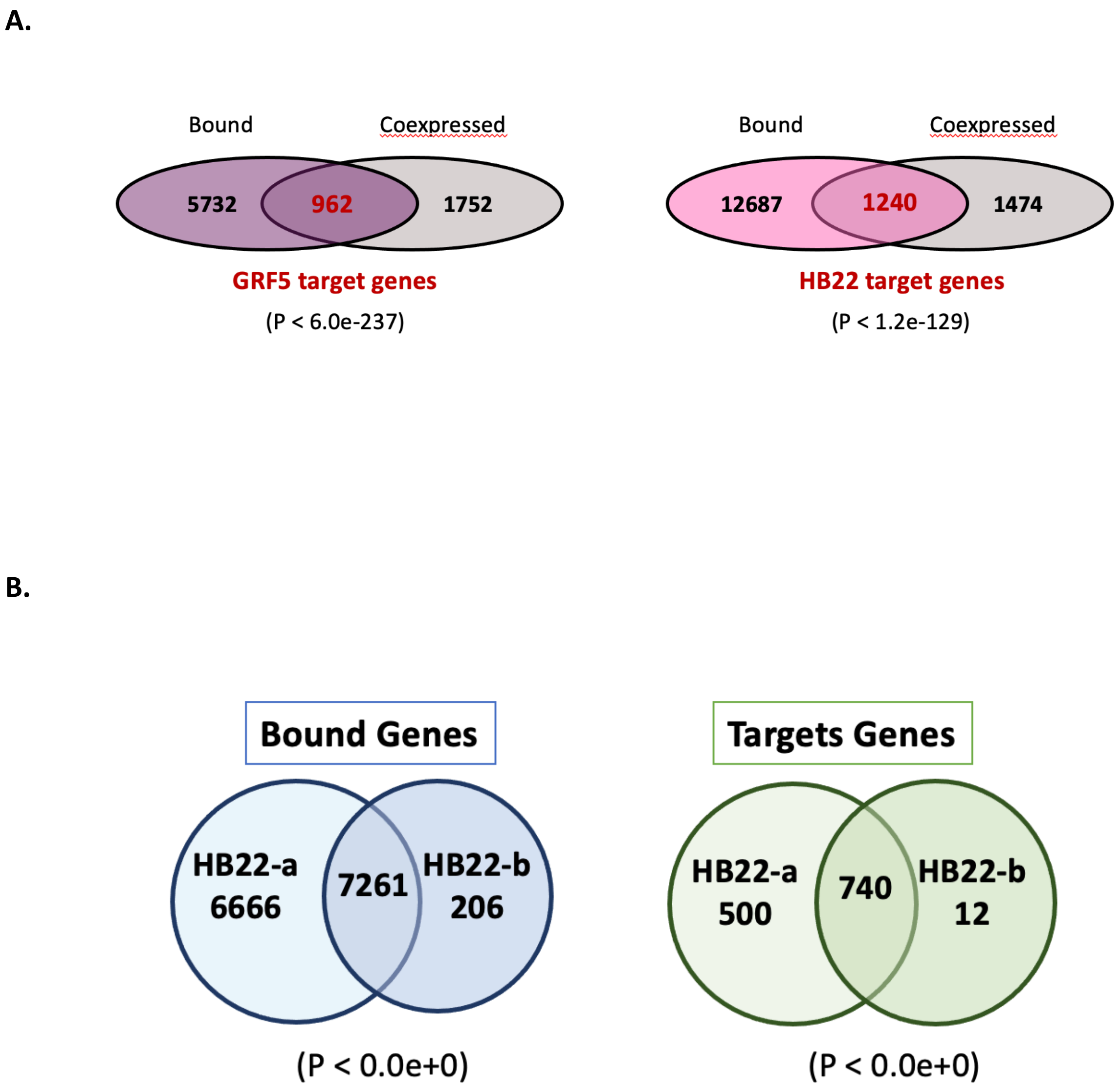
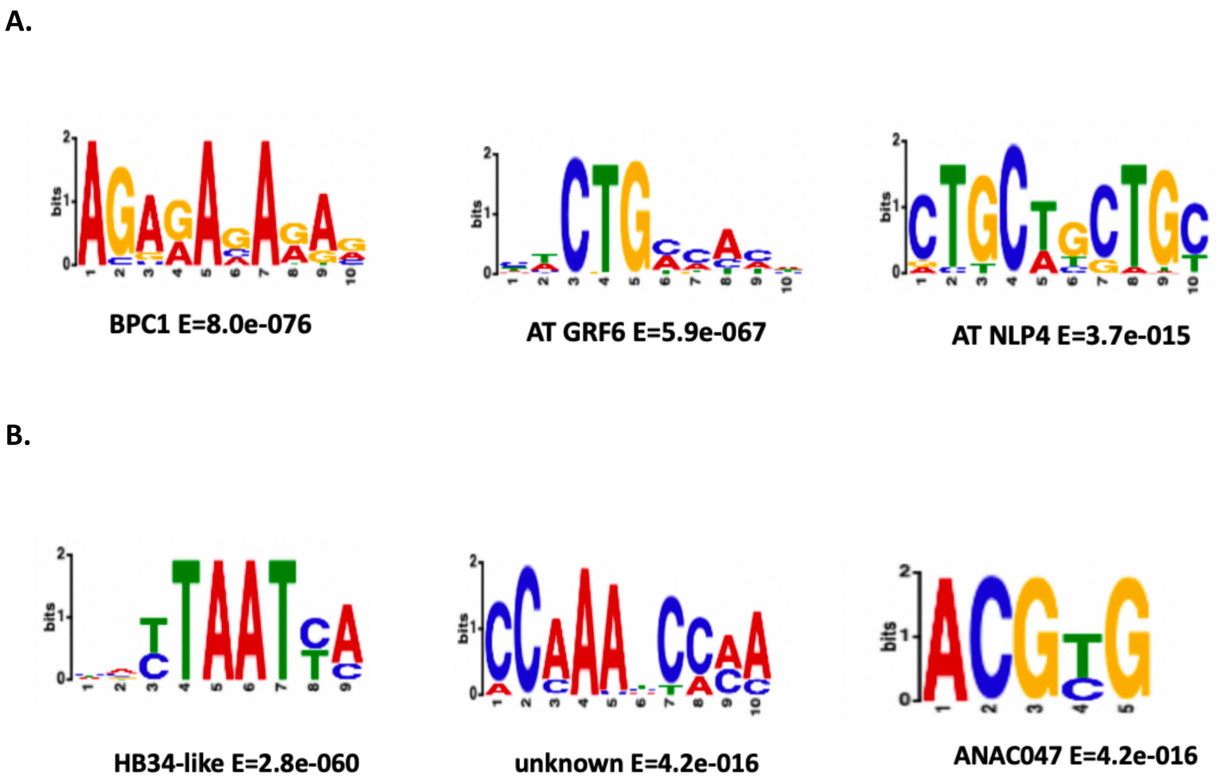


Figure 3.2. Identification of GRF5 and HB22 target genes in soybean early maturation embryos. (A) Target genes directly regulated by GRF5 and HB22 in soybean early maturation embryos. Venn diagrams show the overlap between bound genes (colored) and coexpressed genes (grey). (B) HB22-a and HB22-b antibodies bound the same genes, which verify the HB22 binding specificities. Venn diagrams show the major overlaps in the bound and target genes identified with HB22-a and HB22-b antibodies.



BPC1 E=8.0e-076

AT GRF6 E=5.9e-067

AT NLP4 E=3.7e-015

HB34-like E=2.8e-060

unknown E=4.2e-016

ANAC047 E=4.2e-016

Figure 3.3. *De novo* DNA motif discovery analysis for 600 GRF5 and HB22 bound genes with the highest peak signals. The motifs with three highest E-values are shown. (A) GRF5 bound gene *de novo* DNA motifs. (B) HB22 bound gene *de novo* DNA motifs

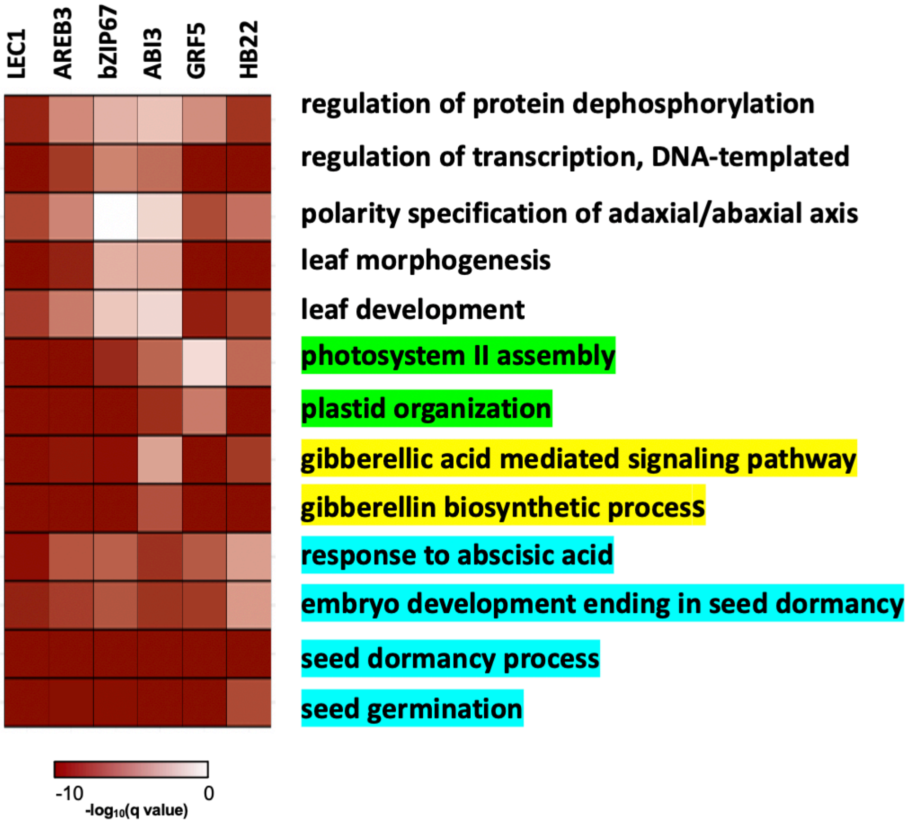


Figure 3.4. Direct target gene overrepresented GO terms with top 10 highest significances for GRF5 and HB22 are shown in the heatmap. LEC1, AREB3, bZIP67 and ABI3 *q* values are indicated in the heatmap for comparison. GO terms related to seed maturation process, GA signaling, and photosynthesis are highlighted in blue, yellow and green, respectively.

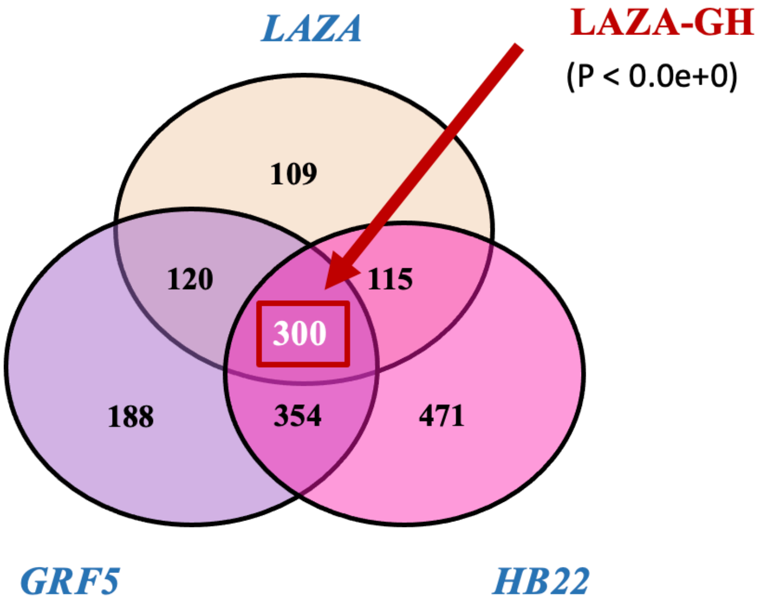


Figure 3.5. LAZA, GRF5 and HB22 share many target genes. Venn diagram shows the overlap of target genes for LAZA, GRF5 and HB22 TFs.

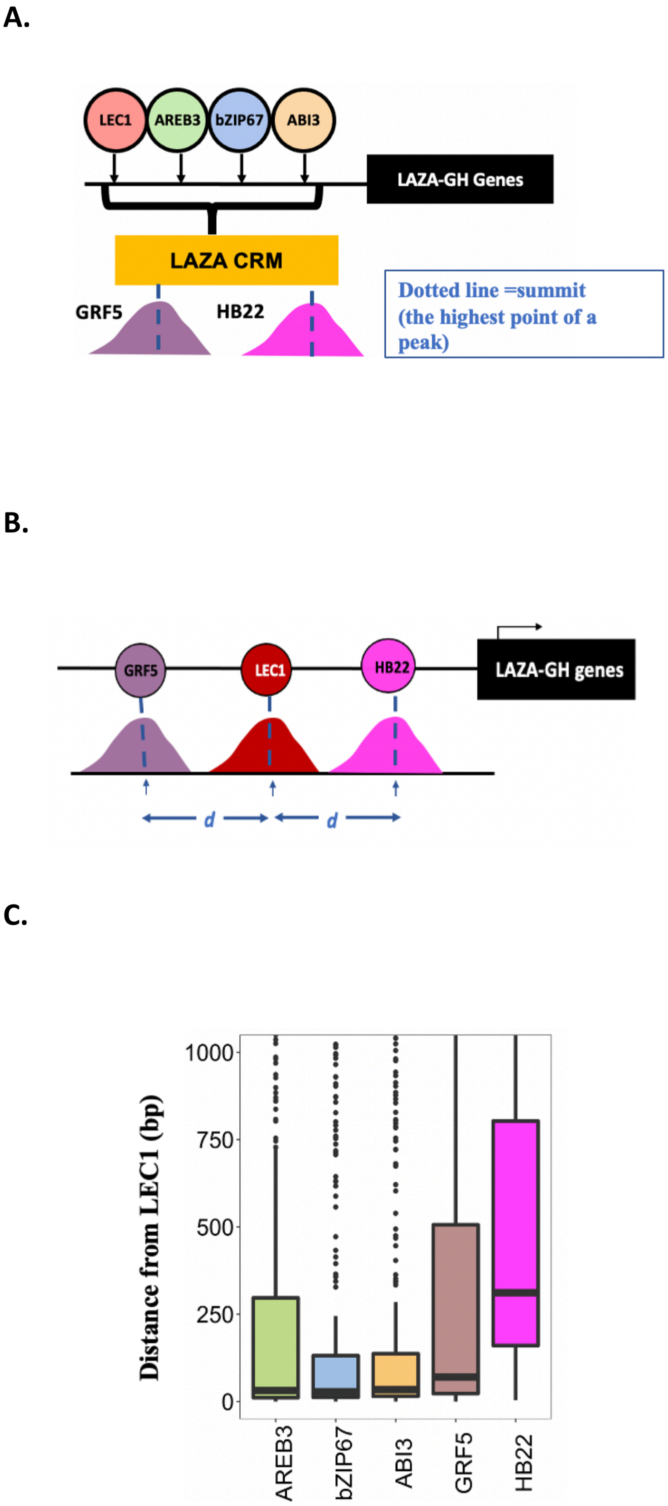

A.
LEC1
AREB3
bZIP67
ABI3
LAZA-GH Genes
LAZA CRM
GRF5
HB22
Dotted line =summit (the highest point of a peak)
B.
GRF5
LEC1
HB22
LAZA-GH genes
d
d
C.
Distance from LEC1 (bp)
1000
750
500
250
0
AREB3
bZIP67
ABI3
GRF5
HB22

Figure 3.6. GRF5 and HB22 bind different loci at LAZA-GH target genes. (A) The schematic figure to show the overlap with LAZA *cis* regulatory module (CRM) and the summits of GRF5 and HB22 at LAZA-GH target genes. CRMs were operationally defined as promoter regions whose boundaries extended 100 bp on each side of the outermost ChIP peak summits within a cluster. The total number of LAZA CRMs at LAZA-GH target genes is 253. The number of overlapped GRF5 and HB22 summits with LAZA CRM is summarized in Table 3.3. (B) Schematic figure showing how the distance between LEC1 and GRF5/HB22 summits are measured. (C) The box plot shows the distributions of the distances between LEC1 and AREB3/bZIP67/ABI3/GRF5/HB22 summits at LAZA-GH target loci. The bar in the box indicates the median distances of the summits.

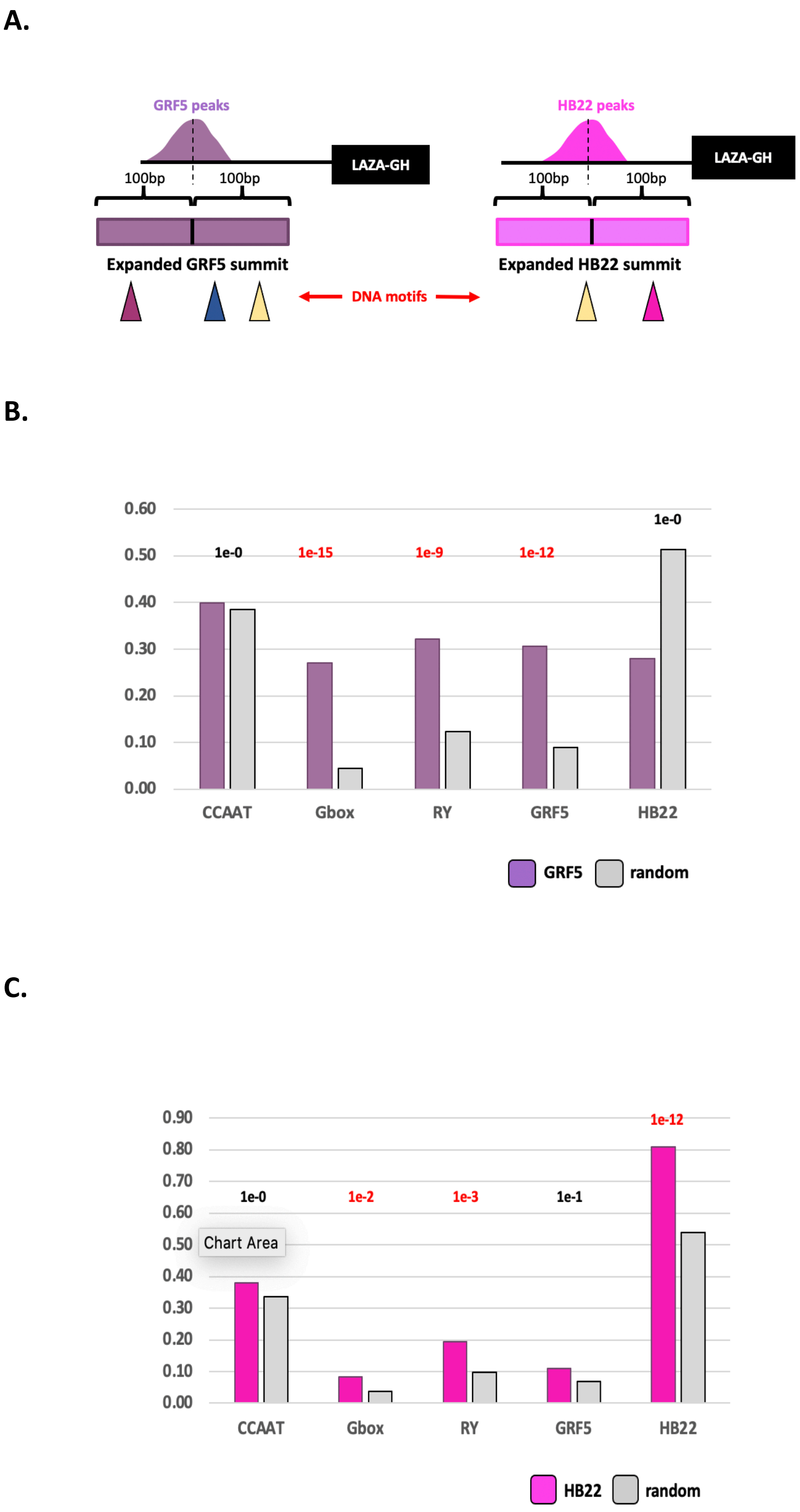

A.
GRF5 peaks
LAZA-GH
100bp
100bp
Expanded GRF5 summit
HB22 peaks
LAZA-GH
100bp
100bp
Expanded HB22 summit
DNA motifs
B.
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1e-0
1e-15
1e-9
1e-12
1e-0
CCAAT
Gbox
RY
GRF5
HB22
GRF5
random
C.
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Chart Area
1e-0
1e-2
1e-3
1e-1
1e-12
CCAAT
Gbox
RY
GRF5
HB22
HB22
random

Figure 3.7. DNA motif enrichment analyses for expanded GRF5 and HB22 summits. (A) The schematic picture to show how the peak summits were queried regions 100 bp upstream and downstream of GRF5 and HB22 ChIP-seq peak summits at the LAZA-GH target sites. (B) Motif enrichment for the queried GRF5 peak summits. (C) Motif enrichment for the queried HB22 peak summits. The numbers above each bar indicate the *p* values for significance. The purple, pink and gray bars show queried GRF5, queried HB22 and randomly sequence, respectively. CCAAT box for LEC1 binding motif, G-box for AREB3 and bZIP67 binding motif, RY for ABI3 binding motif, GRF5 and HB22 motifs are based on my *de novo* discovery motif analysis shown in Figure 3.3.

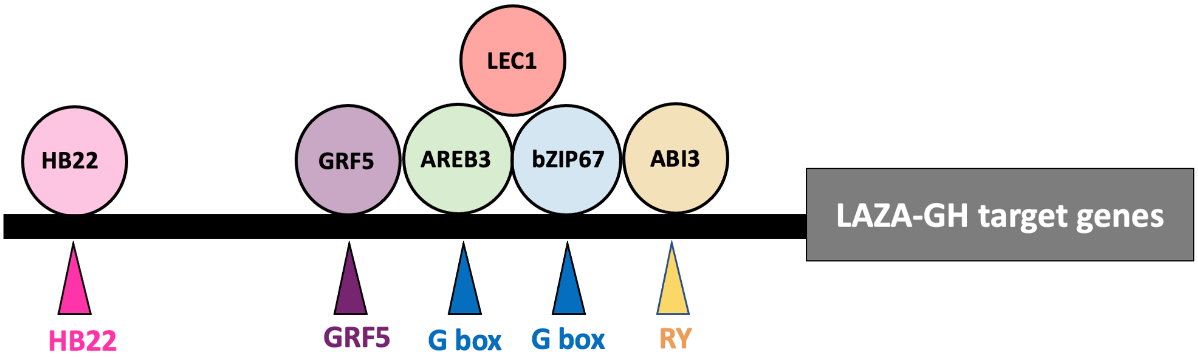


Figure 3.8. The model for the regulatory circuitry controlling maturation gene targeted by the members of LAZA, GRF5 and HB22. GRF5 binds closely to LAZA TFs whereas HB22 binds away from LAZA TFs at the LAZA-GH target genes.

Table 3.1. Top 15 most overrepresented GO terms (biological function) for LAZA-GH are shown. Yellow for GO terms for the seed maturation, blue for GO terms for GA signaling/biosynthesis and green for GO terms for photosynthesis.

category	term	qval
GO:0010162	seed dormancy process	3.05E-08
GO:0006355	regulation of transcription, DNA-templated	5.54E-06
GO:0009737	response to abscisic acid	1.82E-05
GO:0031930	mitochondria-nucleus signaling pathway	1.82E-05
GO:0009845	seed germination	3.63E-05
GO:0009686	gibberellin biosynthetic process	5.14E-05
GO:0010075	regulation of meristem growth	0.000447037
GO:0009740	gibberellic acid mediated signaling pathway	0.000816138
GO:0009657	plastid organization	0.004208537
GO:0009793	embryo development ending in seed dormancy	0.005937621
GO:0010373	negative regulation of gibberellin biosynthetic process	0.005937621
GO:0010896	regulation of triglyceride catabolic process	0.005937621
GO:0019915	lipid storage	0.005937621
GO:0045893	positive regulation of transcription, DNA-templated	0.016102637
GO:0009765	photosynthesis, light harvesting	0.016102637
GO:0009062	fatty acid catabolic process	0.016102637
GO:0009944	polarity specification of adaxial/abaxial axis	0.018829303
GO:0010262	somatic embryogenesis	0.023497384
GO:0010080	regulation of floral meristem growth	0.023497384
GO:0016570	histone modification	0.034499152

Table 3.2. Overrepresented GO terms for the genes targeted by HB22, not by LAZA or GRF5. Green for GO terms for photosynthesis.

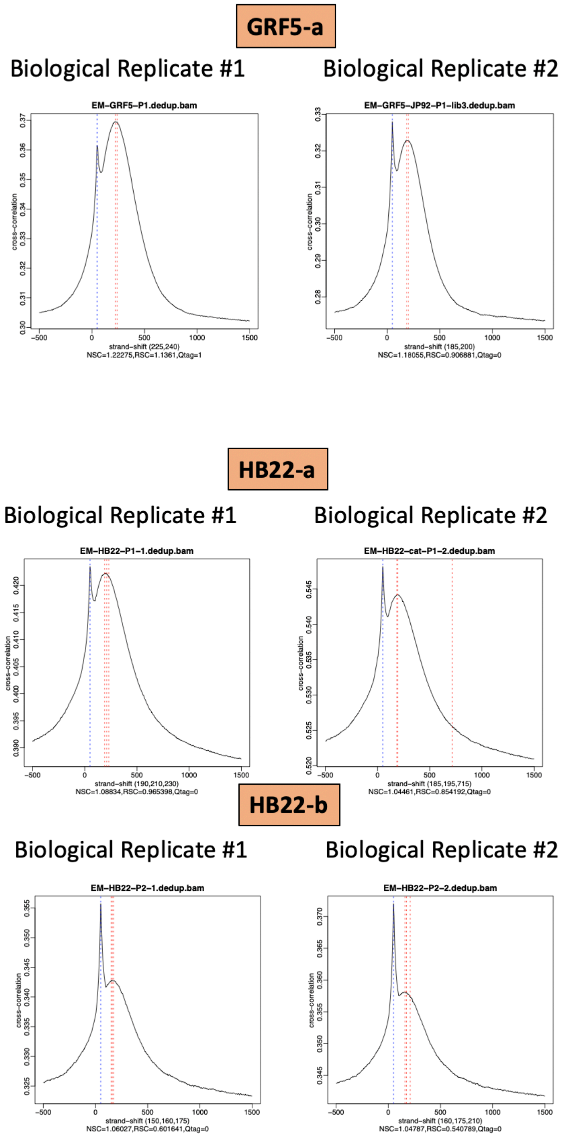
category	term	qval
GO:0010207	photosystem II assembly	0.00015952
GO:0009657	plastid organization	0.00015952
GO:0006364	rRNA processing	0.00015952
GO:0019684	photosynthesis, light reaction	0.00046764
GO:0006633	fatty acid biosynthetic process	0.00051021
GO:0035304	regulation of protein dephosphorylation	0.00304066
GO:0015979	photosynthesis	0.00319767
GO:0006098	pentose-phosphate shunt	0.00542997
GO:0043085	positive regulation of catalytic activity	0.00908421
GO:0010114	response to red light	0.01637397
GO:0000226	microtubule cytoskeleton organization	0.02338138
GO:0009637	response to blue light	0.02338138
GO:0008610	lipid biosynthetic process	0.02420695
GO:0019464	glycine decarboxylation via glycine cleavage system	0.04665568
GO:0010103	stomatal complex morphogenesis	0.04977705

Table 3.3. Summary of GRF5 and H22 summits overlapping with LAZA CRMs at LAZA-GH target genes. Total number of LAZA CRM is 253.

	Overlapping Summits #	% of the total LAZA CRM #
GRF5	220	87.0%
HB22	41	16.2%

Supplemental Figures and Tables

A.



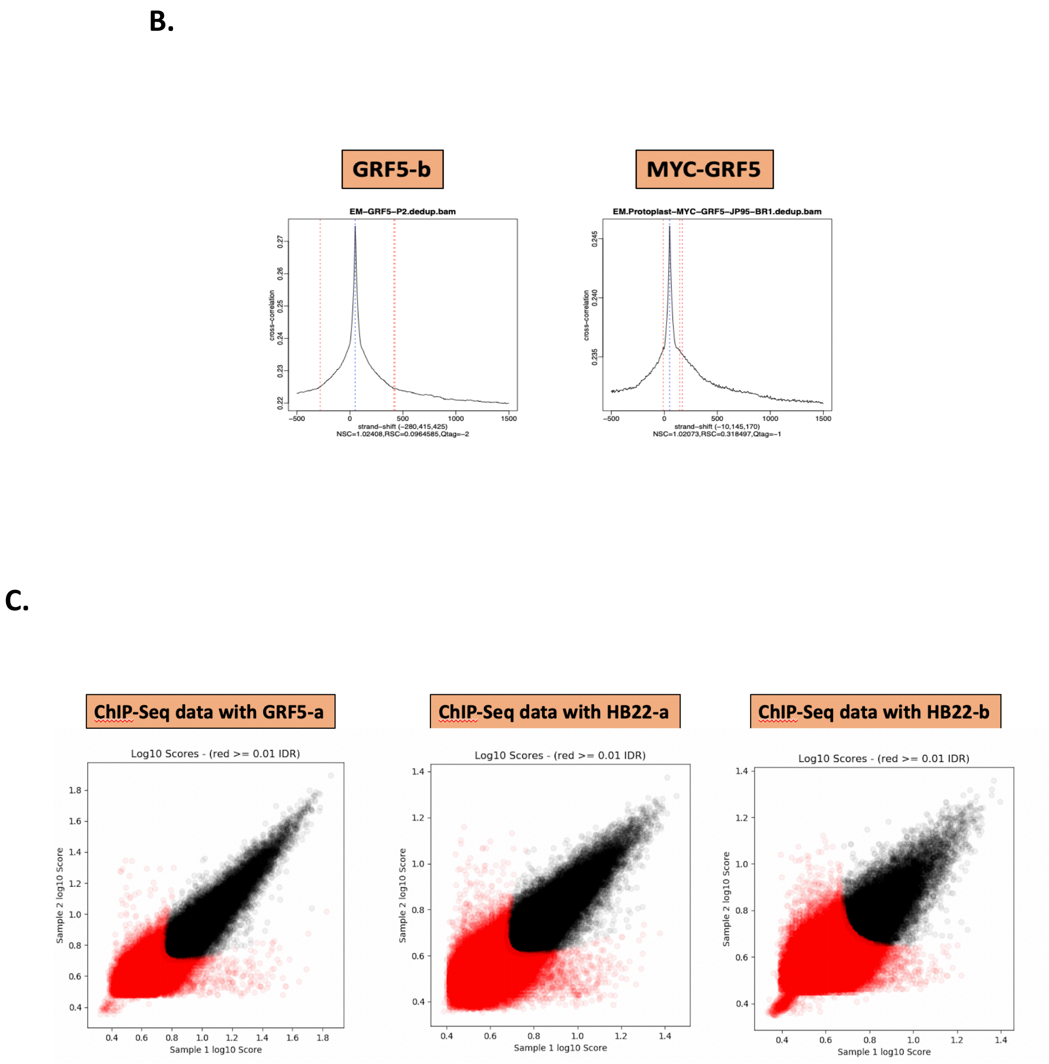


Figure S3.1. Data quality metrics for the ChIP-Seq datasets. The details are listed in Table S3.2. (A) Cross correlation of each biological replicate for the data with GRF5-a, HB22-a, and HB22-b antibodies. (B) Cross correlation of each biological replicate for the GRF5 data with GRF5-b and MYC antibodies. The second libraries were not generated due to the insufficient qualities. (C) Plots show correlation between the peak scores in biological replicates to show reproducibility of ChIP-seq experiments. Each axis shows two biological replicates. Black dots show the peaks that passed IDR threshold (0.01) whereas red dots show the peaks did not pass IDR threshold.

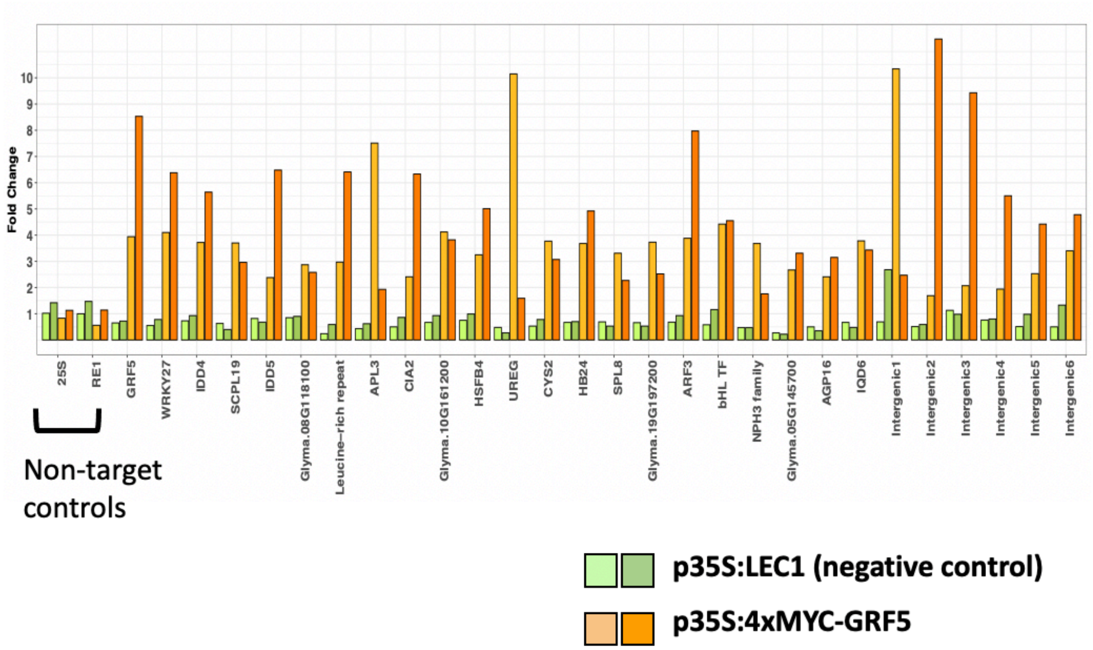


Figure S3.2. ChIP-qPCR validation for 35S:4xMYC-GRF5. The first two samples are for non-target control as marked. Primers were designed based on peaks from ChIP-seq data with GRF5-a antibody. Each bar indicates the individual biological replicates. P35S:LEC1 is used as a negative control for MYC pull-down. X axis shows GRF5 bind loci whereas Y axis shows the fold change relative to the input samples. Primer information is listed on Table S3.3.

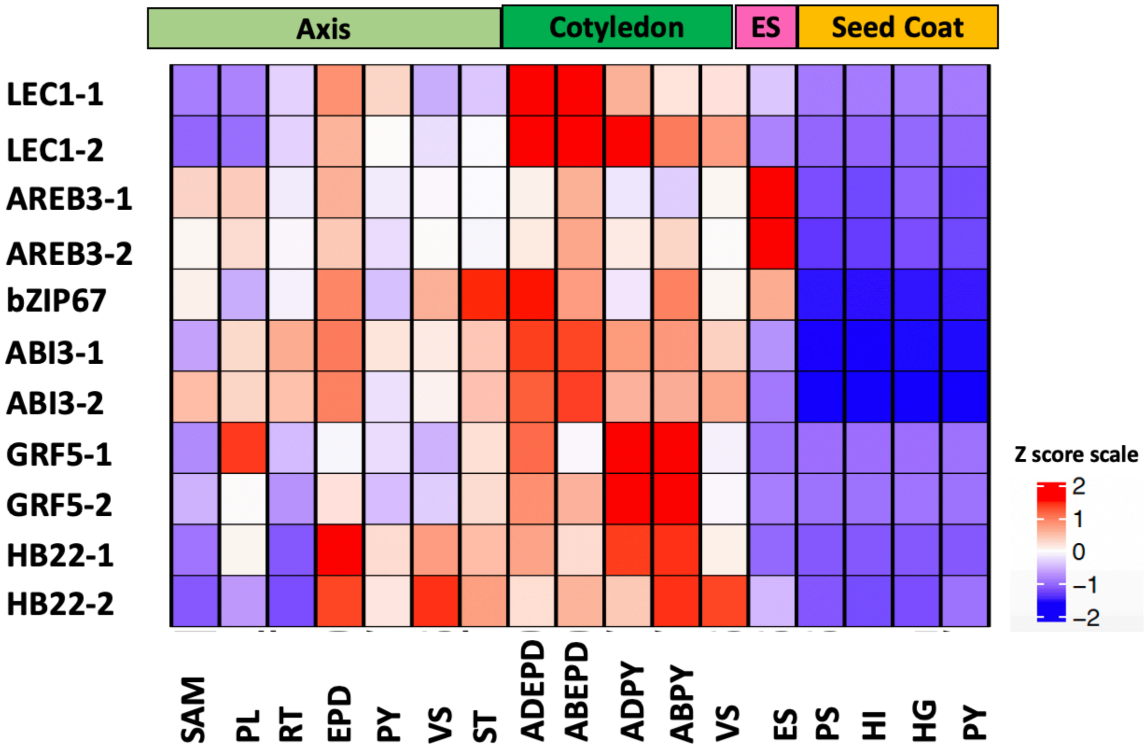
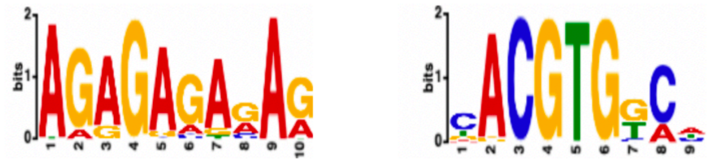


Figure S3.3. Spatial patterns of GRF5 and HB22 mRNA accumulation were compared with ones of LEC1(Glyma.07G268100 and Glyma.17G005600), AREB3 (Glyma.04G124200, Glyma.06G314400), bZIP67 (Glyma.13G317000) and ABI3(Glyma.08G357600 and Glyma.18G176100) in soybean seed subregions at the early maturation stage. mRNA accumulation data were obtained from the Harada-Goldberg Soybean Seed Development LCM RNA-Seq Dataset (GEO accession: GSE 116036). Antibodies for each transcription factor were designed for the genes listed in this figure. 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 apical meristem), ST (stele), VS (vasculature).



BPC1 E=1.4e-040

G box-like E=4.5e-032

Figure S3.4. *De novo* motif discovery analysis for GRF5 target genes at LAZA-GH. The similar motifs were observed in GIF1 binding sites mentioned in Vercruyssen et al. (33).

Table S3.1. Antibodies used for ChIP-seq for early maturation embryos.

Antibody Name	Genes Name	Target Protein	Peptide Sequence with Linker	QC
GRF5-a	GROWTH-REGULATING FACTOR 5	Glyma.07G038400 Glyma.16G007600	C+STSSRPPDADFPPQD	Passed
GRF5-b	GROWTH-REGULATING FACTOR 5	Glyma.01G234400 Glyma.11G008500	C+LRSNNNSMLQGDYLQ	Not Passed
HB22-a	HOMEOBOX PROTEIN 22	Glyma.04G093300 Glyma.06G095200	RSQPQPQPLHPQYHH+C	Passed
HB22-b	HOMEOBOX PROTEIN 22	Glyma.04G093300 Glyma.06G095200	C+ASGGVFSREE	Passed

Table S3.2. Metrics of quality control.

	Number of ChIP-Seq Reads			Library Complexity		Strand Cross-Correlation Coefficient ł		Number of Peaks Identified	
Library	Raw	Uniquely Mapped	Non-Redundant & Uniquely Mapped	Non-Redundant Fraction (NRF) *	Complexity	Normalized (NSC) ǂ	Relative (RNC) §	MACS (FDR 0.01)	Reproducible Peaks (IDR <0.01)
EM-GRF5-a ChIP #1	16,594,018	12,723,923	11,045,639	0.87	Compliant	1.22	1.14	94,063	20,553
EM-GRF5-a ChIP #2	17,244,218	11,237,781	9,515,074	0.85	Compliant	1.18	0.91	126,583	
EM-GRF5-a Input #1	18,987,589	14,149,299	12,549,137	0.89	Compliant				
EM-GRF5-a Input #2	33,304,458	24,264,497	20,936,601	0.86	Compliant				
EM-GRF5-b ChIP #1	11,723,728	7,948,089	6,924,423	0.87	Compliant	1.02	0.10	NA	NA
EM-GRF5-b Input #1	20,731,577	14,849,907	13,139,729	0.88	Compliant				
EM-HB22-a ChIP #1	22,631,112	16,874,007	15,449,002	0.92	Ideal	1.09	0.97	165,668	35,851
EM-HB22-a ChIP #2	42,235,469	31,634,417	26,807,748	0.85	Compliant	1.04	0.85	163,560	
EM-HB22-a Inpt #1 & #2	36,128,465	25,628,670	23,620,450	0.92	Ideal				
EM-HB22-b ChIP #1	18,272,525	13,216,102	11,645,948	0.88	Compliant	1.06	0.60	201,159	18,733
EM-HB22-b ChIP #2	19,512,629	14,078,300	12,689,345	0.90	Compliant	1.05	0.54	189,020	
EM-HB22-b Inpt #1 & #2	35,945,163	25,353,424	23,629,912	0.93	Ideal				
MYC-GRF5-1 ChIP #1	16,963,339	8,553,924	7,280,159	0.85	Compliant	1.02	0.32	NA	NA
MYC-GRF5-1 Input #2	38,007,709	26,153,931	22,616,732	0.86	Compliant				

*** Non-redundant fraction (NRF):** measure of the library complexity. Ratio between the number of genomic positions that reads map uniquely to and the total number of uniquely mappable reads: Ideal (NRF>0.9), Compliant (0.8 ≤ NRF ≤ 0.9), Acceptable (0.5 ≤ NRF < 0.8) and Concerning (NRF < 0.5)
ł Strand Cross-Correlation Coefficient: measure of the degree of clustering of the immunoprecipitated fragments at the protein binding sites. Computation of the Pearson correlation between the Watson and the Crick strands, as a function of the shift (k) applied to one of the strands.
ǂ Normalized strand cross-correlation coefficient (NSC): ratio between the cross-correlation at the fragment length and the background cross-correlation. Pass (NSC > 1.05)
§ Relative strand cross-correlation coefficient: ratio between cross-correlation at the fragment length and the cross-correlation at the read length. Pass (RSC > 0.8)

Table S3.3. qPCR primer list for 35S:4xMYC-GRF5 protoplast ChIP-qPCR validation.

Gene	Target	Forward	Reverse
NA	25S	GATGTCGGCTCTTCCTATCATTGTG	CTTAGAGGCGTTCAGTCATAATCCAG
NA	RE1	CAATGGGCCTGTGATCTTCT	GAAGGCCACCTTGTTCTTCA
Glyma.09G068700	GRF5	TTTGAGGGACCATTGCGTTACAG	GTGTGGCAGAGAGTAGAGAGAGAAA
Glyma.09G129100	WRKY27	AGGGTAACAATGCAGAAATATGGCAGTGG	TGGATAGTGCACGTAATCGAGGAAACG
Glyma.07G158200	IDD4	ATTGCATGCAACCCAAGGCAGCTAAACA	TGAAATTCAGCAGTGGCAGAGTGTACTGTG
Glyma.06G047400	SCPL19	CTGCCAGCAGTTCTTGTCACTA	GCCTTCTTCACGACTCTGGTTTAT
Glyma.10G280000	IDD5	ACCTCACTCCTCAACCCAAGT	GGGATGAATGACACTGTTTGTCTGAA
Glyma.08G118100	Glyma.08G118100	GCGTCAGTACATGTGATACAACTGTG	GCCTTTGTAATTGTCTCTCTCTACTCTCTA
Glyma.03G141600	Leucine-rich repeat	CGTGGTGTCCTTGAGGAATGATTT	CAGTGCAGTGTAGTGTGTTGGTAAT
Glyma.06G011700	APL3	AAATACGACTAAAGGGTTCAGAC	GCTTGATCTCTACCGTTAAAGTG
Glyma.09G059800	CIA2	CTCATCTCCAGAGTCAGTGAGCATAC	CTTTAGCTATAGCAGAGAGAAGGTCACG
Glyma.10G161200	Glyma.10G161200	GCACTGTGGTGAAGCTGTGTACTA	AGGATGGCAGGACACGTGAAA
Glyma.14G083700	HSFB4	TTTGCTCTGCACTGCACCTT	CAAGACAAACCCGTGACTCTCTTTC
Glyma.08G084800	UREG	AACAATGAATACGACAGCCTGACA	GCTGCATTTCTTTGCTTGTTTAATTGAG
Glyma.11G253300	CYS2	GCCATCATCACACTCACACCATA	GGGCTGACAGATAAGCAAGTATCG
Glyma.01G174200	HB24	AGGAAAGAGTGGGTTGCAGATAATG	GGGCCACTACACTTGTCTTGTTATC
Glyma.03G117500	SPL8	GACGAAATGGTAAACTCCGTCCAAG	TCTCTCTCTCTTGCTATTGGTGAGTG
Glyma.19G197200	Glyma.19G197200	GGGCACAATCTGTCAAGTGTC	CCTATTCTATTAGCATCCTTTCATGAATCTC
Glyma.07G202200	ARF3	CTCCTTCTACTGCACTACAGTCTTTCT	TTTGCTTGTCTGCTGGCTGAG
Glyma.08G271000	bHL TF	TCGAGAAGGGTAGCTTACAGTCA	GCAAACAACAGAACCTTTAGCCATATC
Glyma.09G281100	NPH3 family	TGCATACGTGACACACAACTGAAA	GGGCAACGAATGGCAATGTAGTA
Glyma.05G145700	Glyma.05G145700	ATCGGAAGCACTGCTCTCTCT	GTGGGATGCTTGCAAGAACATAGA
Glyma.11G018900	AGP16	AAAGTTAGACAAAGCATATGGG	CTGAAGCTGTTGTAACTTGTATG
Glyma.06G052800	IQD6	CAGATCTCTCTCTACTGCACTGACTAC	GTGCCTTTGAGGGAGGGTAATAAAG
Chr19-48892131	Intergenic1	CACCCTAATGCACCAAACAAATGATAAC	AAAGCTTCTGGAGGCTATTGACAC
Chr09-1568808	Intergenic2	AGTAGTAGAGTCTGAATCTGACACATGAT	GTGGTTCTGAAGCGAAGAAGCTAAG
Chr18-999247	Intergenic3	TCTTCAGGCAACCTCAAACCAAA	GTCCCAATGGAACAACTAGAACATGATA
Chr20-35093113	Intergenic4	GACACACTTCAGAATCAGAGAGAGAG	TTGTGGCGTGGTAATGGTAATGG
Chr18-55544952	Intergenic5	GTGTGGTGGTCCATACCTGATTT	AGCTGAATGTGCATGGAAGTAAGAG
Chr02-1963948	Intergenic6	CGTCTGCATGCTGTTGGTTAGT	AGCCACCACCGTCAAAGTTTC

Table S3.4. HB22 only target genes associated with photosynthesis or light response-related GO terms. Those GO terms are photosystem II assembly (GO:0010207), plastid organization (GO:0009657), photosynthesis/light reaction (GO:0019684), photosynthesis (GO:0015979), response to red light (GO:0010114), response to blue light (GO:0009637). Those genes were also regulated by LAZA TFs in the different combinations indicated in the "category" column.

Gene	Category
Glyma.01G089300	HB22_LEC1
Glyma.01G199700	HB22_LEC1
Glyma.13G155300	HB22_LEC1
Glyma.18G049600	HB22_LEC1
Glyma.19G260600	HB22_LEC1
Glyma.14G177200	HB22_LEC1
Glyma.10G153100	HB22_LEC1
Glyma.02G012500	HB22_LEC1
Glyma.17G241000	HB22_LEC1
Glyma.11G245400	HB22_LA
Glyma.18G193600	HB22_LA
Glyma.13G282000	HB22_LA
Glyma.17G174500	HB22_LA
Glyma.02G101100	HB22_LAZ
Glyma.10G032200	HB22_LAZ
Glyma.17G133200	HB22_LAZ
Glyma.13G129500	HB22_LAZ
Glyma.03G114600	HB22_LAZ
Glyma.13G046200	HB22_LAZ
Glyma.16G145800	HB22_LAZ
Glyma.15G100200	HB22_LAZ
Glyma.12G092000	HB22 _ LEC1 _ AREB3 _ ABI3
Glyma.20G214000	HB22 _ AREB3
Glyma.15G043600	HB22_AREB3_ABI3

Table S3.4 continued

Gene	Category
Glyma.11G042200	HB22
Glyma.14G094500	HB22
Glyma.16G172700	HB22
Glyma.19G227700	HB22
Glyma.01G198700	HB22
Glyma.11G043200	HB22
Glyma.19G152600	HB22
Glyma.15G105900	HB22
Glyma.13G063700	HB22

Chapter 4

Summary and Conclusion

Introduction

The angiosperms are the most diverse group in plants. The sexual reproduction system, including production of seeds, has contributed to the success of the angiosperm diversification. The embryo is embedded inside of the seed, surrounded by the endosperm and seed coat in most angiosperms where the embryo is nourished and protected.

Seeds are composed of three regions with different genetic architectures. The embryo, zygotic and diploid, establishes the basic body plan of the plant (1). The endosperm, zygotic and triploid, maintains embryo growth and reserve storage (2). The seed coat is maternal and diploid, and it serves as a protective tissue for the embryo (3). Seed development is biphasic, consisting of morphogenesis and maturation phases. During the morphogenesis phase, the embryo undergoes a series of cell divisions and specifications to generate specialized tissues and organs, leading to the establishment of the shoot-root axis (1). When the seed enters the maturation phase, the embryo cells start expanding and accumulate storage molecules such as oils, seed storage proteins and starch. During the late maturation phase, the seed acquires desiccation tolerance and becomes metabolically quiescent to prepare for dormancy and subsequent germination (4).

Seed development is comprised of highly regulated processes. Each developmental process must occur in the right tissue at the right time. Seed development is controlled largely by various gene regulatory networks. My work focuses on how genes are regulated during soybean (*Glycine* max) seed development. Soybean is one of the most important agricultural crops in the world due to its high oil and protein content (5). In order to dissect the gene regulatory networks controlling soybean seed development, we employed genomic approaches to comprehensively understand the underlying gene regulatory mechanisms. I have focused on gene regulation by micro RNAs (miRNAs) and transcription factors (TFs) to obtain insight into the mechanisms regulating soybean seed development.

miRNA profiling

miRNAs are a class of small RNAs that negatively regulate their target genes. miRNAs play important roles for plant development including seed (6–8), and stress responses (9). Arabidopsis embryos carrying the loss-of-function mutant allele of *DICER-LIKE1(DCL1)*, the enzyme responsible for the miRNA biosynthesis, showed various morphological defects. miRNAs accumulate in specific tissues at certain developmental stage to maintain proper developmental processes (10–12), including Arabidopsis embryo (13,14), suggesting miRNA may control spatial and temporal developmental processes in plants.

Various studies have showed that many miRNAs are induced or suppressed under the abiotic or biotic stress conditions (reviewed in (9,15)), suggesting the involvement of miRNAs in stress response gene regulatory networks. The miRNAs that play important roles for plant development are also involved in stress responses (9). Thus, it has been suggested that miRNAs may have a role to adjust the growth and development under the stresses (16). These findings are important, because miRNA-based genetic modification technology has been explored to improve the crop yield and quality (17).

Information about which miRNAs accumulate in seed subregions at different developmental stages, and the biological processes in which miRNA may be involved, advances an integrated understanding of seed development. The use of Laser Capture Microdissection allowed us to profile miRNA populations in specific seed subregions at different developmental stages. My research focused on studying miRNAs accumulation during seed development and how the patterns of miRNA accumulation provide insights into the roles of miRNA in soybean seed development (Chapter 2).

One of the most surprising observations of miRNA accumulation was the enrichment of non-conserved miRNA families in endosperm subregions, particularly at the heart, cotyledon and early maturation stages (Chapter 2). Most non-conserved miRNA are known to accumulate at a low level (18,19), and they are not functional (20). Interestingly, several miRNA families, including one novel miRNA discovered in our laboratory, accumulate at high levels in soybean seeds. I was able to identify target mRNAs for the endosperm subregion-specific miRNA families using publicly available datasets (Chapter2). The annotated functions for these target mRNAs are related to biotic and abiotic stress responses. To my best knowledge, miRNA-mediated gene regulations for abiotic and biotic stresses in the soybean endosperm has not been previously reported.

A similar conclusion was reached using unbiased hierarchical clustering analysis of mRNA and miRNA families, which was used to obtain insights into miRNA family functions. The miRNA families that accumulate specifically in the endosperm subregion with statistical significance cluster together with the mRNAs whose annotated functions are related to abiotic and biotic stress responses (Chapter 2). The genes expressions and proteins related to abiotic and biotic stress responses have been detected in the endosperm in other plant species (21–23).

The fact that the annotated functions for target mRNAs for the endosperm subregion-specific miRNA families were related to the abiotic and biotic stress responses seems counterintuitive because miRNAs are the negative regulator for the target mRNAs. One possible explanation is that miRNAs may keep target mRNA levels constant as “gene expression buffer” (reviewed in (24–26)). In fact, I found that most miRNAs did not show an inverse correlation of expressions to their target mRNA levels. This result may support the idea of miRNA as a “buffering regulator”.

However, this result does not rule out the possibility that the miRNAs are simply not functional at least in the seed subregions because the target identification analysis was based on datasets from other soybean tissues in addition to the seeds. Additional functional assays may be able to reveal the exact functions of miRNAs in a specific subregion.

Expanding the gene regulatory network for soybean seed maturation program

Soybean seeds are rich in oils and proteins which make them one of the most important crops in the world for food and commercial use (5). The proteins and oils accumulate in the embryo during the maturation phase. *LEAFY COTYLEDON 1 (LEC1)* is a master regulator for the seed maturation program (27). *LEC1* is a B subunit of the CCAAT binding NF-Y transcription factor (TF) that can form a oligomeric complex consisting of NF-YA, NF-YB and NF-YC (28). LEC1 can also form a protein complex with other TFs (29–32). Previously, it was shown that photosynthesis, GA signaling, and the seed maturation program were sequentially and combinatorially regulated by a group of the transcription factors, called LAZA that consists of LEC1, ABA-RESPONSIVE ELEMENT BINDING PROTEIN3 (AREB3), BASIC LEUCINE ZIPPER67 (bZIP67) and ABA INSENTIVE3 (ABI3) in soybean (33). In the same study, GROWTH-REGULATING FACTOR 5 (GRF5) and HOMEOBOX22 (HB22) were identified as the target genes of LAZA. *GRF5* encodes a plant-specific TF (34). HB22 encodes a member of Class I HD-Zip TFs (HD-Zip I) (35). The functions of these TFs have not been elucidated in the seed development.

In order to expand the gene regulatory network for soybean seed maturation, we identified target genes of GRF5 and HB22 by using chromatin immunoprecipitation-DNA sequencing experiments (ChIP-Seq). My dissertation research is focused on understanding how GRF5 and HB22 are involved in the soybean seed maturation program. First, I performed GO term representation analysis for target genes of GRF5 and HB22 to get insights into the functions for GRF5 and HB22 in the seed maturation program. The overrepresented GO terms for both GRF5 and HB22 were similar to those of LAZA (33). Furthermore, I found major overlaps of the target genes among LAZA, GRF5 and HB22 (Chapter 3). I hypothesized that GRF5 and HB22 might bind to the same genomic loci as LAZA at the common target genes of LAZA, GRF5 and HB22. To test this hypothesis, I looked at the distances of the genomic binding sites between LAZA and GRF5 or HB22 in the upstream regions of common target gene. I found that GRF5 binding sites were very close to LAZA binding sites whereas HB22 genomic binding sites were located further away from LAZA binding sites.

Based on the results above, I speculate GRF5 may regulate genes involved in the seed maturation program with LAZA, possibly by direct interaction with LAZA. Further investigation is required to test this hypothesis, possibly by using assays such as bimolecular fluorescence complementation (BiFC). So far, GRF5 has been shown to interact with transcriptional coactivators, GRF-INTERACTING FACTOR1 (GIF1) (36). GIF1 is known to interact with SWITCH/SUCROSE NONFERMENTATING (SWI/SNF) chromatin remodeling ATPase complex (37). LEC1 was shown to establish the active chromatin state for reactivation of *FLOWERING LOCUS C (FLC)* gene expressions in Arabidopsis seeds. Thus, LEC1 is speculated to play a role as a pioneer TF (38). Given the role of GRF5 and GIF1, it is possible that LEC1 may function as a pioneer TF with the help of GRF5 and GIF5 to control soybean seed maturation program. Thus, interactions among GRF5, GIF5 and LEC1 can set up an open chromatin structure that is necessary to activate the common target gene transcription. The open chromatin structures may allow HB22 to bind its DNA sequence motif.

The remaining important question is if GRF5 and HB22 are required to control common target genes with LAZA. Applying genetic modification technology to the soybean system is challenging because soybean gene transformation is time- and labor-intensive. Our laboratory has established a transient gene expression method by using soybean embryo protoplasts (33). One way we can test our hypothesis is to use targeted gene knockdown or knockout methods such as RNA interference, or CRISPR/CAS technologies, followed by ChIP-Seq analysis. This approach could address our question to elucidate the roles of GRF5 and HB22 in the regulation of gene transcription during the soybean seed maturation program.

Summary and concluding remarks

My dissertation research has focused on gene regulation by miRNAs and TFs. For the miRNA study, my goal was to examine how miRNA profiles change spatially and temporally, and how changes in miRNA profiles in each subregion can provide insight into their functions. Previous studies of miRNA profiles at the whole seed level provided some insights into the roles of miRNAs (13,18,39,40), and our extremely high resolution data provides additional information about miRNA-mediated gene regulations in soybean seeds.

For the transcription factor study, my goal was to expand the soybean seed maturation program that is controlled by GRF5 and HB22. Identifying target genes of GRF5 and HB22 shed light on the roles of these TFs in soybean seed maturation gene regulatory network.

My dissertation allowed me to study two different mechanisms for gene regulation in soybean seeds by using multiple genomic approaches. The opportunity to engage in this dynamic study made me realize the complexity and sophistication of soybean seed development. My work is just a small step to understand the gene regulation in soybean seed development. I hope my work can contribute to next step for soybean seed research.

References

1. Goldberg RB, De Paiva G, Yadegari R. Plant embryogenesis: Zygote to seed. Science (80-). 1994;266(5185):605–14.
2. Berger F. Endosperm: The crossroad of seed development. Vol 6, Current Opinion in Plant Biology. Elsevier Ltd; 2003. p 42–50.
3. Moïse JA, Han S, Gudynaitę-Savitch L, Johnson DA, Miki BLA. Seed coats: Structure, development, composition, and biotechnology. Vol 41, In Vitro Cellular and Developmental Biology - Plant. 2005. p 620–44.
4. 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. Vol 54, Plant Journal. 2008. p 608–20.
5. Pagano MC, Miransari M. The importance of soybean production worldwide. In: Abiotic and Biotic Stresses in Soybean Production. 2016. p 1–26.
6. Nodine MD, Bartel DP. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev. 2010;24(23):2678–92.
7. Willmann MR, Mehalick AJ, Packer RL, Jenik PD. MicroRNAs regulate the timing of embryo maturation in Arabidopsis. Plant Physiol. 2011;155(4):1871–84.
8. Li C, Zhang B. MicroRNAs in Control of Plant Development. Vol 231, Journal of Cellular Physiology. 2016. p 303–13.
9. Khraiwesh B, Zhu JK, Zhu J. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim Biophys Acta - Gene Regul Mech. 2012;1819(2):137–48.
10. Kidner CA, Martienssen RA. Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature. 2004;428(6978):81–4.
11. Knauer S, Holt AL, Rubio-Somoza I, Tucker EJ, Hinze A, Pisch M, et al. A Protodermal miR394 Signal Defines a Region of Stem Cell Competence in the Arabidopsis Shoot Meristem. Dev Cell. 2013;24(2):125–32.
12. Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. Cell. 2009;138(4):750–9.
13. Plotnikova A, Kellner MJ, Schon MA, Mosiolek M, Nodine MD. MicroRNA dynamics and functions during arabidopsis embryogenesis[CC-BY]. Plant Cell. 2019;31(12):2929–46.
14. Miyashima S, Honda M, Hashimoto K, Tatematsu K, Hashimoto T, Sato-Nara K, et al. A comprehensive expression analysis of the arabidopsis MICRORNA165/6 gene family during embryogenesis reveals a conserved role in meristem specification and a non-cell-autonomous function. Plant Cell Physiol. 2013;54(3):375–84.
15. Guleria P, Mahajan M, Bhardwaj J, Yadav SK. Plant Small RNAs: Biogenesis, Mode of Action and Their Roles in Abiotic Stresses. Vol 9, Genomics, Proteomics and Bioinformatics. 2011. p 183–99.
16. Sunkar R, Li YF, Jagadeeswaran G. Functions of microRNAs in plant stress responses. Vol 17, Trends in Plant Science. 2012. p 196–203.
17. Zhou M, Luo H. MicroRNA-mediated gene regulation: Potential applications for plant genetic engineering. Vol 83, Plant Molecular Biology. 2013. p 59–75.
18. Arikit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, et al. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. Plant Cell.

2014;26(12):4584–601.

19. Talmor-Neiman M, Stav R, Frank W, Voss B, Arazi T. Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. Plant J. 2006;47(1):25–37.
20. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, et al. High-throughput sequencing of Arabidopsis microRNAs: Evidence for frequent birth and death of MIRNA genes. PLoS One. 2007;2(2).
21. Endo A, Tatematsu K, Hanada K, Duermeyer L, Okamoto M, Yonekura-Sakakibara K, et al. Tissue-specific transcriptome analysis reveals cell wall metabolism, flavonol biosynthesis and defense responses are activated in the endosperm of germinating arabidopsis thaliana seeds. Plant Cell Physiol. 2012;53(1):16–27.
22. Sheoran IS, Olson DJH, Ross ARS, Sawhney VK. Proteome analysis of embryo and endosperm from germinating tomato seeds. Proteomics. 2005;5(14):3752–64.
23. Jerkovic A, Kriegel AM, Bradner JR, Atwell BJ, Roberts TH, Willows RD. Strategic distribution of protective proteins within bran layers of wheat protects the nutrient-rich endosperm. Plant Physiol. 2010;152(3):1459–70.
24. Chen X. Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol. 2009;25:21–44.
25. Voinnet O. Origin, Biogenesis, and Activity of Plant MicroRNAs. Vol 136, Cell. 2009. p 669–87.
26. Garcia D. A miRacle in plant development: Role of microRNAs in cell differentiation and patterning. Vol 19, Seminars in Cell and Developmental Biology. 2008. p 586–95.
27. 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. Proc Natl Acad Sci U S A. 2017;114(32):E6710–9.
28. Lee H, Fischer RL, Goldberg RB, Harada JJ. Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proc Natl Acad Sci U S A. 2003;100(4):2152–6.
29. Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, et al. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. Plant Cell. 2006;18(11):2971–84.
30. Kumimoto RW, Zhang Y, Siefers N, Holt BF. NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana. Plant J. 2010;63(3):379–91.
31. Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T. Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J. 2009;58(5):843–56.
32. Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, et al. bZIP67 regulates the omega-3 fatty acid content of arabidopsis seed oil by activating fatty acid DESATURASE3. Plant Cell. 2013;25(8):3104–16.
33. Jo L, Pelletier JM, Hsu SW, Baden R, Goldberg RB, Harada JJ. Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. Proc Natl Acad Sci U S A. 2020;117(2):1223–32.
34. Omidbakhshfard MA, Proost S, Fujikura U, Mueller-Roeber B. Growth-Regulating Factors (GRFs): A Small Transcription Factor Family with Important Functions in Plant Biology. Vol 8, Molecular Plant. 2015. p 998–1010.
35. Ariel FD, Manavella PA, Dezar CA, Chan RL. The true story of the HD-Zip family. Vol

12, Trends in Plant Science. 2007. p 419–26.

36. Horiguchi G, Kim GT, Tsukaya H. The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana. Plant J. 2005;43(1):68–78.

37. Vercruyssen L, Verkest A, Gonzalez N, Heyndrickx KS, Eeckhout D, Han SK, et al. ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during Arabidopsis leaf development. Plant Cell. 2014;26(1):210–29.

38. Tao Z, Shen L, Gu X, Wang Y, Yu H, He Y. Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. Nature. 2017;551(7678):124–8.

39. Song QX, Liu YF, Hu XY, Zhang WK, Ma B, Chen SY, et al. Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol. 2011;11.

40. Shamimuzzaman M, Vodkin L. Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. BMC Genomics. 2012;13(1).