PLOS | ONE

RESEARCH ARTICLE

Cloning and Characterization of 5′ Flanking Regulatory Sequences of *AhLEC1B* Gene from *Arachis Hypogaea* L.

Guiying Tang[1], Pingli Xu[1], Wei Liu[2], Zhanji Liu[1], Lei Shan[1,2]*

1 Bio-Tech Research Centre, Shandong Academy of Agricultural Sciences / Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan, 250100, China, **2** College of Agriculture, Shandong University, Jinan, 250100, China

* Shlei1025@sina.com

OPEN ACCESS

Citation: Tang G, Xu P, Liu W, Liu Z, Shan L (2015) Cloning and Characterization of 5′ Flanking Regulatory Sequences of *AhLEC1B* Gene from *Arachis Hypogaea* L. PLoS ONE 10(10): e0139213. doi:10.1371/journal.pone.0139213

Editor: Martina Stromvik, McGill University, CANADA

Received: May 9, 2015

Accepted: September 9, 2015

Published: October 1, 2015

Copyright: © 2015 Tang et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by National Natural Science Foundation of China, grant No. 30971546 and No. 31470349, to LS, and No. 31171621 to ZJL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

LEAFY COTYLEDON1 (LEC1) is a B subunit of Nuclear Factor Y (NF-YB) transcription factor that mainly accumulates during embryo development. We cloned the 5′ flanking regulatory sequence of *AhLEC1B* gene, a homolog of *Arabidopsis LEC1*, and analyzed its regulatory elements using online software. To identify the crucial regulatory region, we generated a series of GUS expression frameworks driven by different length promoters with 5′ terminal and/or 3′ terminal deletion. We further characterized the GUS expression patterns in the transgenic *Arabidopsis* lines. Our results show that both the 65bp proximal promoter region and the 52bp 5′ UTR of *AhLEC1B* contain the key motifs required for the essential promoting activity. Moreover, *AhLEC1B* is preferentially expressed in the embryo and is co-regulated by binding of its upstream genes with both positive and negative corresponding *cis*-regulatory elements.

Introduction

NF-Y (Nuclear Factor Y) transcription factor is ubiquitous in eukaryotic organisms. The three subunits of NF-Y, NF-YA, NF-YB, and NF-YC, play an important role in regulating the expression of multiple genes (both positively and negatively) by recognizing and binding to the CCAAT promoter sequence [1, 2]. In the *Arabidopsis* genome, there are 36 NF-Y subunits, including 10 NF-YA, 13 NF-YB and 13 NF-YC. These subunits are differentially expressed in a tissue- or organ-specific pattern, or in the distinctive profile of developmental stages, and participate in regulating of many genes in a wide range of biological processes [3–5].

The NF-Y transcription factor genes such as *LEAFY COTYLEDON1* (*LEC1* or *NF-YB9*) and *LEC1-LIKE* (*L1L* or *NF-YB6*)–first identified in *Arabidopsis*–are genes related to embryonic development. *AtLEC1* and *AtL1L* mRNA accumulate in different spatial and temporal patterns. Higher levels of *AtLEC1* mRNA are present in the early-stage embryo at the proembryo stage, globular stage, transition stage, heart stage, torpedo stage, and curled cotyledon stage than in the late maturation embryo, but is not detectable in leaves, stems, roots, and flowers [6], while *AtL1L* mRNA levels are higher in seeds than in vegetative tissues. *AtL1L* RNA levels peak at a

later stage of embryogenesis (mainly from the torpedo stage to the bent-cotyledon stage) as compared with *LEC1* levels. Warpeha et al. (2007) [7] showed *NF-YB6* and *NF-YB9* expression in the 6-d-old etiolated seedlings of *Arabidopsis*. Siefers et al. (2009) [3] identified 36 nuclear factor transcription subunits that can combine to govern tissue-specific expression patterns of flowering time, embryo maturation, meristem development, etc. in *Arabidopsis*. The *turnip* (*tnp*) mutant represents a gain-of-function mutant of Arabidopsis LEC1. In *tnp* mutant, the elements required for the repression of *LEC1* in vegetative tissue are deleted in the distal upstream promoter region causing a higher constitutive expression of *LEC1* [8].

Here, we analyze the phylogenetic relationship among the peanut transcription factors *AhLEC1A*, *AhLEC1B*, and the *Arabidopsis* NF-YB transcription factors. We also cloned the 5′ flanking regulatory sequence of the *AhLEC1B* gene and analyzed the *cis*-regulatory elements existing in this region by computational analyses. We further constructed a set of GUS expression frameworks driven by different length promoters with 5′ terminal and 3′ terminal deletion to identify the crucial regulatory regions and characterize the GUS expression patterns in their transgenic *Arabidopsis* lines.

Materials and Methods

Plant materials and growth conditions

Peanut (*Arachis hypogaea* L.) cv. 'Luhua 14' seeds were grown in the experimental field of Shandong Academy of Agricultural Sciences. Seeds at different developmental stages were collected at 10~70 days after pegging (DAP) and kept in -80°C refrigerator for isolation of total RNA and construction of a cDNA library.

Cloning of 5' flanking region of *AhLEC1*

Peanut genomic DNA was isolated from Luhua 14 leaves using CTAB method [9]. For each DNA library construction, 2.5μg genomic DNA was digested with four blunt-end restriction enzyme *DraI*, *EcoRV*, *PvuII*, and *StuI* respectively. The digested samples were purified with phenol and chloroform; and then 4μl digested DNA was connected with the BD Genome-Walker adaptor (Table 1) provided by BD GenomeWalker Universal Kit (Clontech, USA), resulting in the library containing digestions by *DraI*, *EcoRV*, *PvuII*, and *StuI* (LD, LE, LP, and LS). Based on the sequence of *AhLEC1B* genomic DNA (S1 Fig), two nested gene-specific primers (GSP), LEC1BGSP1-2 and LEC1BGSP2-2 (Table 1), were designed. The first round of PCR reaction was done as per the manufacturer's instructions in a 25μl reaction system using an AP1 (Table 1) provided by Kit and LEC1BGSP1-2 as 5' terminal and 3' terminus primer, and 1μl DNA of each library as template. The nested PCR reaction was also performed using the same volume and conditions with primers AP2 (Table 1) and LEC1BGSP2-2, and 1μl of the 10-fold diluted primary PCR products as template. The specific PCR fragments from the second round reaction were isolated and inserted into the vector pEASY-T3. The recombinants harboring the target gene were validated by *EcoRI* digestion and two-way sequencing using ABI3730 model DNA sequencer.

Localization of transcriptional start site

High-quality total RNA was isolated from Luhua14 mixed seeds at different developmental stages (ranging from 10 to 70 DAP) using the improved CTAB method [10]. According to the GeneRacer^{TM} Kit's recommendation, 5μg total RNA was dephosphorylated by Calf Alkaline Phosphatase (CIAP or CIP), and the full-length mRNA among them was removed using the 5' Cap structure, which was then ligated to the RNA adaptor (GeneRacer RNA Oligo, Table 1).

Table 1. The primers used in this study.

Serial No.	Primer name	Sequences	Purposes
1	BD GenomeWalker adaptor	5' GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCG GGCTGGT 3'	No.1~5 used for the amplification of 5′ flanking sequence of *AhLEC1B* genome DNA
2	LEC1BGSP1-2	5' CCTTGTTCCCATGTAAAACCATGAAAGCA 3'	
3	LEC1BGSP2-2	5' AGGTAAAGCAGCCGCTAATCTAGTTAGT 3'	
4	AP1	5' GTAATACGACTCACTATAGGGC 3'	
5	AP2	5' ACTATAGGGCACGCGTGGT 3'	
6	GeneRacer RNA Oligo	5' CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAG UAGAAA 3'	No.6~10 used for the localization of the transcriptional start site of *AhLEC1B* gene
7	TSS LEC1BGSP1-1	5' TCTTTTGCGTCGTCGGAGATTTTAGC 3'	
8	TSS LEC1BGSP2	same as LEC1BGSP2-2	
9	5' GeneRacer Primer	5' CGACTGGAGCACGAGGACACTGA 3'	
10	5' Nested Primer	5' GGACACTGACATGGACTGAAGGAGTA 3'	
11	BF1	5′ AAGCTTTCGTGAATAAAGGAACAC 3′	No.11~17 used for the deletion analysis of *AhLEC1B* promoter
12	BF2	5′ AAGCTTACTCTATGATATTCCGAAGG 3′	
13	BF3	5′ AAGCTTCCTCGGTTGCATCGCCCT 3′	
14	BF4	5′ AAGCTTGCATTGCTTGCAGCTCTTTG 3′	
15	BF5	5′ AAGCTTGTTACTCCGTTTCTTCATAC 3′	
16	BR1	5′ CCATGGGTAAAGCAGCCGCCAATCTA 3′	
17	BR2	5′ CCATGGCTCGCCCTTCGGAATATCAT 3′	

doi:10.1371/journal.pone.0139213.t001

The ds-cDNA was synthesized based on the manufacturer's instruction using the above decapped, full-length mRNA with RNA Oligo as template, and oligo dT provided by SuperScript^{TM} III RT kit as a primer. The ds-cDNA was cloned into vector pCR4-TOPO to establish the full-length cDNA library.

For amplifying the transcription start site (TSS) of the target gene, two 3′ terminus gene-specific primers for each gene, TSS LEC1BGSP1-1 and TSS LEC1BGSP2 (Table 1), were designed, for use in the nested PCR reaction. The 5' terminus general primers for two rounds of PCR were 5' GeneRacer^{TM} Primer and 5' Nested Primer (Table 1). According to the recommended system of BD Advantage™ 2 PCR Kit, the primary PCR was performed as per the following conditions: 94°C denatured for 2 min, and 5 cycles of 94°C for 30 sec and 72°C for 30sec, and then 5 cycles of 94°C for 30 sec and 70°C for 30 sec, and 20 cycles of 94°C for 30 sec, 63°C for 30sec and 68°C for 30sec, and finally extension for 10 min at 68°C. The nested PCR was performed using a 50-fold dilution of the primary PCR product as template. The PCR condition were: denaturation at 94°C for 2 min; 35 cycles of 94°C for 30 sec, 65°C for 30 sec and 68°C for 10 sec; and finally 68°C for 10 min.

The nested PCR products were collected and sequenced by ABI3730 model DNA sequencer.

Computational *cis*-regulatory motif analysis of the promoter of *AhLEC1B* gene

Two different online software PLACE (http://www.dna.affrc.go.jp/PLACE/) and PlantCARE (http://bioinformatics.psb.ugebp.be/webtools/plantcare/html/) were used to predict the *cis*-regulatory elements in the 5' flanking region of *AhLEC1B* gene, including the 5' untranslated region (5' UTR) and the upstream regulatory region.

Constructs of GUS expressing system, *Arabidopsis* transformation, and GUS staining

The different length promoters with 5' or 3' terminal deletion were obtained by PCR. All primers are listed in Table 1. BR1 and BR2 are reverse primers localized in 5' UTR of *AhLEC1B*. BF1-BF5 are the forward primers situated in the different sites of the *AhLEC1B* promoter (Table 1). For cloning purposes, a *HindIII* site (AAGCTT) and a *NcoI* site (CCATGG) was added to the 5' border and 3' border of each fragment by PCR amplification with an appropriately designed oligonucleotide. The six fragment-deleted promoters replacing the CaMV 35S promoter were cloned into pCAMBIA3301 digested with *HindIII* and *NcoI*.

The binary vectors constructed above were transferred into *Agrobacterium tumefaciens* strain GV3101 and then transformed into *Arabidopsis* Col-0 plants using the floral dip method [11]. Seeds were harvested and stored at room temperature. For screening, seeds were sterilized in 95% (v/v) ethanol for 1 min and 0.1% (v/v) HgCl for 20 min, followed by several washes with sterile water. Herbicide-resistant plants were selected by incubating plants for 14d on MS [12] basal medium supplemented with10 mg/L Basta.

GUS staining was performed using a standard protocol [13]. The roots and leaves at the 4-leaf stage, stems at the bolting stage, flowers, and seeds of 6–10 days after pollination in transgenic T_2 lines were incubated with the staining buffer (0.1% TritonX-100 and 2mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc), cyclohexyl ammonium salt in 100mM sodium phosphate buffer, pH7.0) at 37°C overnight or 24h and then decolorized with 70% ethanol. The analyses were performed using at least six independent transgenic lines for analysis.

Results

Phylogenetic analysis of AhLEC1A and AhLEC1B

In the *Arabidopsis* genome 13 *NF-YB* genes with distinctive expression patterns were found [3–5]. To predict the evolutionary relationship of AhLEC1A and AhLEC1B, a sequence comparison of AhLEC1A, AhLEC1B, and Arabidopsis NF-YB transcription factors was performed using MAGE 4.0. AhLEC1A, AhLEC1B, *Arabidopsis* NF-YB6 (L1L) and NF-YB9 (LEC1) have higher sequence similarity and group together (Fig 1). AhLEC1A and AhLEC1B share 95% sequence identity and diverge at only 12 amino acid sites. However, the expression profile of *AhLEC1A* was substantially different from that of *AhLEC1B*. *AhLEC1A* is expressed specifically in seeds during different developmental stages while *AhLEC1B* mRNA accumulates at higher levels in seeds as compared with roots, stems, rosettes, and flowers [14].

Cloning and sequence analysis of 5' flanking region of *AhLEC1B* and localization of TSS

To investigate the major regulatory regions or elements of *AhLEC1B*, we isolated the promoter using chromosomal walking. As a result, the 5' flanking fragment of 1289 bp in length including the promoter region (1235bp) and 5'UTR sequences (54bp) was obtained from the peanut DNA library LP (Fig 2)

Based on the cDNA sequence of *AhLEC1B*, we further amplified the 5'UTR of the gene from the full-length cDNA library of Luhua14 developing seeds using nested 5' RACE. As a result, we obtained PCR products of about 400bp and 60bp (Fig 3). The transcription of *AhLEC1B* gene starts at the first 'A' within the sequence of CCAAACT. This sequence is located 83 bp upstream to the translation start codon ATG, and is consistent with the general feature in most eukaryotes (Fig 4).

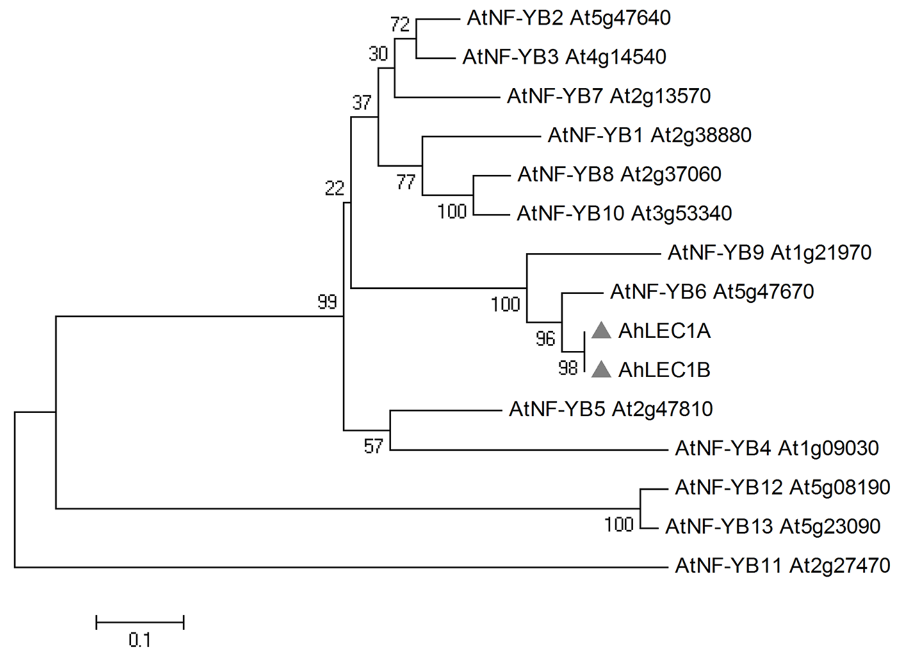

Fig 1. Phylogenetic tree for peanut AhLEC1A and AhLEC1B, and the *Arabidopsis* NF-YB family.

doi:10.1371/journal.pone.0139213.g001

Cis-elements prediction of *AhLEC1B* promoter

To predict *cis*-regulatory elements in 5' flanking fragment of *AhLEC1B*, we submitted the 1318 bp sequences containing 1235 bp promoter region and 83 bp 5'UTR to PLACE and PlantCARE online to detect *cis*-regulatory elements.

The putative TATA box (TATATAT) in the core region of promoter was located –36 from TSS. The other *cis*-regulatory elements were classified into two groups (Fig 4). The first group

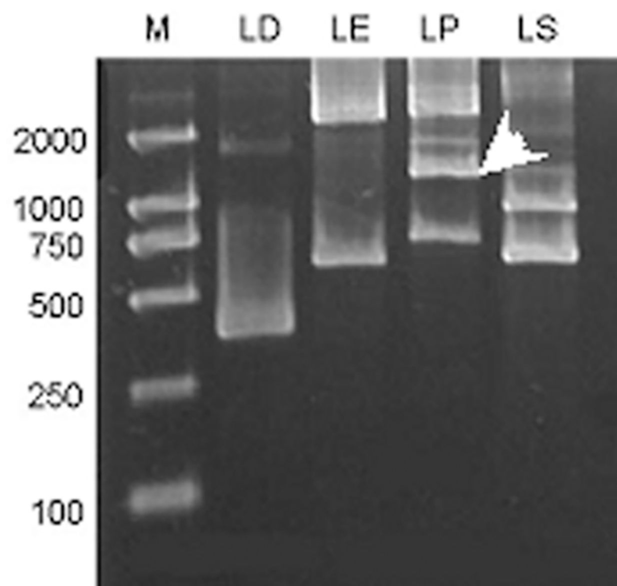

Fig 2. The second round of PCR amplification products of 5′ flanking regulation regions of peanut *AhLEC1B* gene by chromosome walking. The arrow indicates the target band.

doi:10.1371/journal.pone.0139213.g002

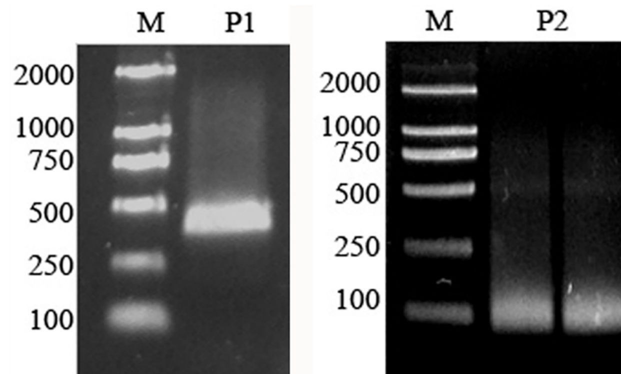

Fig 3. Localization of transcription start sites of the peanut *AhLEC1B* gene using 5′ RACE. P1 –Product of the first round PCR; P2 –Product of the second round PCR

doi:10.1371/journal.pone.0139213.g003

contains the multiple-copied elements, such as E BOX (CANNTG, 5 copies), CARGCW8GAT (CWWWWWWWWG, 4 copies), SEF4 MOTIF (RTTTTTR, 7 copies) and the DOF CORE (AAAG, 17 copies). All of these elements exist in the regulatory regions of many genes that are

```
 -1235  gtataagcattgcttgcagctctttgCAACAtattggtctgttacttggtgttcttttCT
                                  RAV1A AT
 -1175  TGTGTaaaaatgagaatcaagaaCATGCAaacGTCActCCATGTGTctagagaggaaaca
        DPBFCORE              RY REPEAT  WRKY71OS DPBFCORE/E BOX
 -1115  gaggaCATTTGagtttcttcacaagaGAAAATgaaagattgggtttcagcatagaggaag
              E BOX                 GT1CONSENSUS
 -1055  agagaaagtgtgatgagtaagttgtagtaatgagaacatggaagagagtgatcctcgccc
  -995  gaacttgatgtcccaaccttcctccttTGACtgcaaaacatagactaaatcatccatcca
                                 WRKY71OS
  -935  catcgCAACTCatGCCACttctccatcctcacttcttttgccAATTCAAAacatCAACTC
              CARE  SORLIP1AT                         ERE MOTIF   CARE
  -875  atcattattctaaaaatataaatattctatatacttattctttgctaatttatataatgt
  -815  tcaataattaatactttggtttcagacaaatctataagttactccgtttcttcatacaca
  -755  tttttattaatattctCTAATAaaataaaaggacaaattaatcgttaatattttctaaaa
                         CPB
  -695  taactaattagtcttgaatttttttaagaagaatatgattgagagattaaATTATCtga
                                                         GT1CONSENSUS
  -635  aatttttaaaaataaaaaaattaTTTATCttCTTTATTATGgaCAAATTTTTGtttcaga
                            GT1CONSENSUS ARGCW8GAT   CARGCW8GAT
  -575  tcaaataaTTTATCttttTTTTTCgtgaataaaggaacacactcttgattttatttagtt
              GT1CONSENSUS GT1CONSENSUS
  -515  aactttttaataaagttaaccgaaattatccaaatcatcaaattttggtttgatagtTAA

  -455  TCATTAaaaaaattacatacgcatagtaaaattttaatataataacatagttaattgtta
    HDZIP2ATATHB2
  -395  aaataaccaaaccttgattttatggttgatatttgtatggcatcCTTTTTTTGtatttt
                                                    CARGCW8GAT
  -335  aaGATAATtctactctatgatattccgaagggcgaGATAtatttaatttcttgattatt
        GT1CONSENSUS                     GATA BOX
  -275  tattcatggttgaagagatTTTGAATTttTGTTGgatGATAAAgggcagagtCTTATATT
                          ERE MOTIF  RAV1AAT GT1CONSENSUS CARGCW8GAT/ CPB
  -215  AGgaacattGAGTTGagaggtgaatataatttaaaataataaCATTTGATAgacatacag
                  CARE                            E BOX/GATA BOX
  -155  taaaataatttggagtacttgtGAAAAAaaaaacgtaCAAGTGTgggtgcattaggGTCA
                            GT1CONSENSUS  E BOX/DPBF CORE     WRKY71OS
   -95  taagagtgtaaTGACtggttattggccgtttcctcggttgcatcgccctCTTTgttcacT
                 WRKY71OS                                DOFCORE
   -35  ATATATCttgcCACTTGccaactacttcattgcccAaacttcacttcatttcatcatcac
   TATA BOX/GATA BOX  E BOX
   +26  cactaactagattggcggctgCTTTacctgccctCTTTcactgcatttttcttcacac
                            DOFCORE      DOFCORE
```

Fig 4. The sequence of 5′ flanking regulation region of peanut *AhLEC1B* gene and some major elements harbored in this region. The bold capital letter “A” represents the transcription start site (TSS), and other capital letters show different regulatory elements.

doi:10.1371/journal.pone.0139213.g004

preferentially expressed in the seed or embryo [15–18]. Moreover, the CACTFT (YACT, 22 copies), TAAAG MOTIF (5 copies), ROOT MOTIF (ATATT, 15 copies), OSE2 ROOT NODULE (CTCTT, 8copies) and POLLEN1 LELAT52 (AGAAA, 7 copies) are expressed in leaf, root and flower, respectively [19–22]. Some motifs required for light regulation (twelve copies of GATA BOX and ten copies of GT1 CONSENSUS) are dispersed in the promoter region of *AhLEC1B* [23, 24]. Four copies of TGAC core sequences (WRKY71OS) were also scattered in the promoter region. Zhang et al. (2004) [25] found that the TGAC core motif could bind with rice WRKY71 transcriptional repressor to participate in the regulation of the gibberellin signaling pathway. The second group of *cis*-regulatory elements included a large number of elements with lower copies (less than three copies) or a single copy. These include several phytohormone-regulated elements such as CPB Sequence (TATTAG, cytokinin response), ERE Motif (AWTTCAAA, ethylene-induced transcription), GARE Motif (TAACAGA, Gibberellin-responsive element) [26–28], and some elements (TGACGT Sequence and PROLAMIN BOX) related to gene expression levels [29, 30], and some tissue- or organ-preferential regulatory elements DPBF CORE (ACACNNG, associated with embryo- or seed-preferential expression), and RAV1A AT (CAACA) which expresses in relatively higher level in rosette leaves and roots, and etc (Fig 4) [31, 32]. A copy of RY REPEAT sequence (CATGCA) was present in the upstream region of *AhLEC1B* promoter. RY REPEAT sequence is present in the promoters of many genes regulating seed development [33] and is also found in the promoter and intron regions of *AtLEC1*. Other specific *cis*-regulatory elements, such as CELL CYCLE BOX (CACGAAAA) and HEXAMER MOTIF (ACGTCA) were present in *AhLEC1B* promoter. The CELL CYCLE BOX (CACGAAAA) is involved in cell-cycle-specific activation of transcription [34] while HEXAMER MOTIF (ACGTCA) functions in the regulation of replication-dependent expression of the histone H3 gene [35, 36]. They all exist mainly by the style of a single copy in the promoter region of this gene.

GUS expression driven by *AhLEC1B* promoter fragments

To identify the crucial regulatory regions that are essential for gene expression, we generated a series of constructs containing different length *AhLEC1B* promoter with 5' terminal deletion and 52bp 5' UTR or 3' terminal deletion fused with GUS reporter gene (Fig 5). All constructs were introduced into the *Arabidopsis* genome by *Agrobacterium*-mediated transformation. The resulting transgenic T_2 lines containing a single copy homologous gene were screened for use in GUS histochemical staining studies. The results of staining in diverse tissues or organs showed that the longest fragment (Q1, 1281bp) containing 1229bp promoter region and 52bp 5' UTR, mainly regulates the GUS expression in the developing embryo. Moreover, three fragments (Q4, Q5, and Q6) with a 5' terminal deletion could drive the GUS expression in all

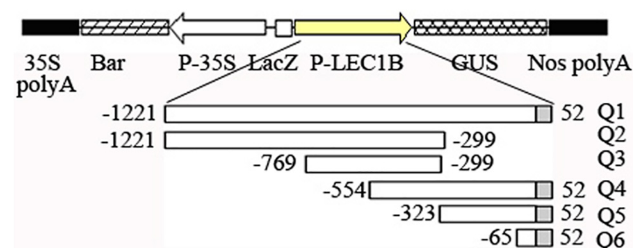

Fig 5. The constructs of GUS expression driven by *AhLEC1B* promoter and schematic representation of the different length promoters with 5′ or 3′ terminal deletion. Q1-Q6 indicates the different promoters with 5′ or 3′ terminal deletion. The white and gray rectangles show the upstream promoter region from TSS and 5′ UTR region respectively.

doi:10.1371/journal.pone.0139213.g005

tissues detected (Fig 6). However, the promoter fragment (Q2 and Q3) with 351bp deletion from 3' terminus lost the promoter function that had crucial activity responsive elements (Table 2). The shortest fragment (Q6, 118bp) including 66bp promoter region and 52bp 5' UTR contains the main elements that control the constitutive expression of the downstream gene (Fig 6).

Discussion

Arabidopsis LEC1 and *L1L* genes regulate embryogenesis, but they have distinct function during embryo development [6, 47]. *LEC1* expression in the embryos peaks at the early stage of seed development and declines thereafter, up to the green premature seed stage [6, 48]. The loss–of–function mutation in *LEC1* results in desiccation intolerance of embryos and defective in the production of storage proteins and lipids. However, as compared with *LEC1* levels, the *L1L* mRNA levels peak at the later stage of embryogenesis. The suppression of *L1L* in RNAi transgenic lines results in abnormal embryos and the embryo lethal phenotype [47], but its mutants *l1l-1* and *l1l-2* have no apparent altered phenotypes during seed development [49]. *AhLEC1A* and *AhLEC1B* from peanut are homologous genes of *Arabidopsis NF-YB6 (L1L)* and *NF-YB9 (LEC1)* and have differential expression patterns in vegetative tissues. Our RT-PCR data shows that *AhLEC1B* mRNA, as similar as *AtL1L* does, accumulates at a higher level in seeds but at a lower level in vegetative tissues [14]. Thus, our expression data and phylogenetic analysis together shows that *AhLEC1B* is an ortholog of *AtL1L*.

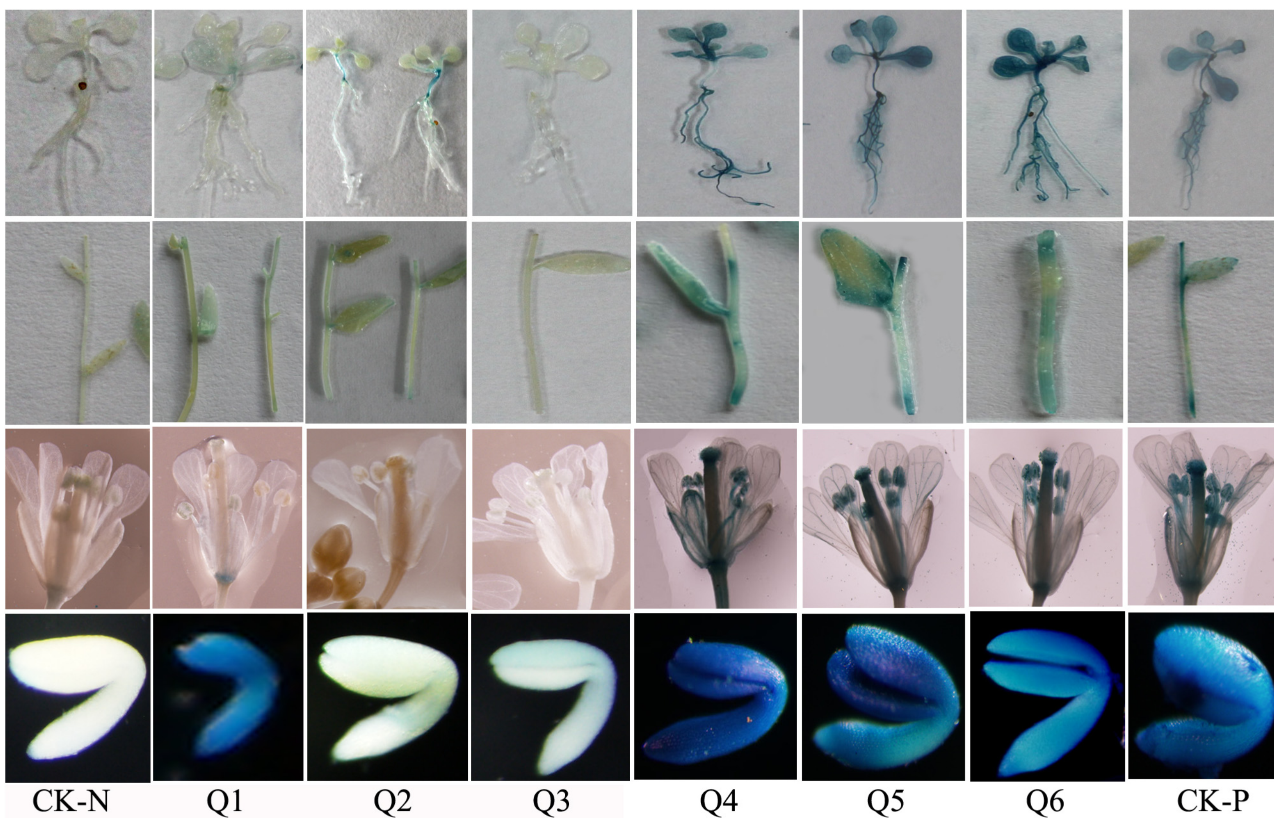

Fig 6. Effects of *AhLEC1B* promoter deletions on the expression profile of *GUS* gene in transgenic Arabidopsis lines. Q1-Q6 indicate the GUS expression patterns in different transgenic Arabidopsis lines containing 5′ or 3′ terminal deletion promoters, and the CK-N and CK-P showed the GUS expression profiles in non-transformed negative control and in positive control harboring 35S:GUS constructs, respectively.

doi:10.1371/journal.pone.0139213.g006

Table 2. Major elements in 5′UTR and 300bp promoter region.

Elements	Sequence[a]	Location[b,c]	Putative Function
ACGT Sequence	ACGT	-120(+,-)	ACGT sequence (from -155 to -152) required for etiolation-induced expression of erd1 (early responsive to dehydration) in Arabidopsis [37].
ARR1AT	NGATT	-257(+), 36(+)	"ARR1-binding element" found in Arabidopsis; Required for transcriptional activation in response to cytokinin [38].
CACTFT PPCA1	YACT	-36(+), -21(+), -137(+), -10(+), -154(-), -139(-), -113(-), -88(-), 9(+), 28(+), 66 (+)	Tetranucleotide (CACT) is a key component of Mem1 (mesophyll expression module 1, which direct mesophyll-specific expression of gene) [19].
CARE element	CAACTC	-203(-)	CAREs, CAACTC regulatory elements, are required for GA-inducible expression of hydrolase genes in the germinating seeds [39].
CARGCW8GAT	CWWWWWWWWG	-220(+,-)	A variant of CArG motif with a longer A/T-rich core is a preferential binding site for the transcriptional regulator AGL15 that accumulates during embryo development [17].
CCAAT BOX1	CCAAT	-72(-), 38(-)	Common sequence found in the 5'-non-coding regions of eukaryotic genes, which involved in increasing the promoter activity [40].
CPB Sequence	TATTAG	-216(+)	The sequence is critical for Cytokinin-enhanced Protein Binding in vitro [27].
CURE CORE	GTAC	-138(+,-), -118(+,-)	Copper-response element, also involved in oxygen-response of some genes [41].
DOF CORE	AAAG	-43(-), 49(-), 62(-)	Core site is required for binding of Dof proteins, which may be associated with the plant-specific pathway for carbon metabolism in maize [42].
DPBF CORE	ACACNNG	-115(-)	The binding core sequence of bZIP transcription factor DPBF-1 and 2 (Dc3 promoter-binding factor-1 and 2); Involved in embryo-specific expression, and responding to ABA [31].
E2F CONSENSUS	WTTSSCSS	-72(+)	E2F consensus sequence of all different E2F-DP-binding motifs that were involved in cell cycle regulation, DNA replication, and chromatin dynamics [43].
E BOX	CANNTG	-170(+,-), -115(+,-), -21(+,-)	The *cis*-elements in the promoter regions of most genes encoding the storage protein [18].
ERE Motif	AWTTCAAA	-253(-)	The ethylene responsive element mediate ethylene-induced activity of transcription [28].
GATA BOX	GATA	-297(+), -235(+), -165(+), -29(-)	Required for high level, light regulated, and tissue specific expression [23].
GT1 CONSENSUS	GRWAAW	-235(+), -130(+), 50(-), 73(-)	Consensus GT-1 binding site in the promoter regions of many light-regulated genes [24].
GTGA Motif	GTGA	-193(+), -132(+)	"GTGA motif" found in the promoter of the tobacco late pollen gene g10 and the tomato gene lat56, required for the gene expression in pollen [44]
I BOX CORE	GATAA	-235(+)	Conserved sequence upstream of light-regulated genes of both monocots and dicot.
POLLEN1 LELAT52	AGAAA	-285(-), 75(-)	One of two co-dependent regulatory elements (AGAAA and TCCACCATA) responsible for pollen specific activation of gene [21].
RAV1A AT	CAACA	-243(-)	Binding consensus sequence of Arabidopsis transcription factor RAV1, which expresses in relatively higher level in rosette leaves and roots [32].
ROOT MOTIF	ATATT	-294(+), -217(+), -190(-)	Motif found both in promoters of rolD, which expresses strongly in roots [20].
SEF4 MOTIF	RTTTTTR	-248(+)	Binding with SEF4, one of soybean embryo factor (SEF) [15].
SORLIP1 AT	GCCAC	-23(+)	One of "Sequences Over-Represented in Light-Induced Promoters (SORLIPs) in Arabidopsis; Involved in phyA-regulated gene expression [45].
TAAAG Motif	TAAAG	-233(+), 49(-)	TAAAG motif controls guard cell-specific gene expression [46].

(*Continued*)

Table 2. (*Continued*)

Elements	Sequence[a]	Location[b,c]	Putative Function
WRKY71 OS	TGAC	-81(+), -96(-)	A core of TGAC-containing W-box; Binding site of rice WRKY71, a transcriptional repressor of the gibberellin signaling pathway or the regulation of the pathogenesis-related genes [25].

[a]N = G/A/C/T; R = A/G; S = C/G; W = A/T; Y = T/C

[b]The symbol '+' or '-' in the bracket represents the DNA strand in which the element is situated.

[c]The positive number indicates the location of element in 5′UTR, while the negative represents that in promoter.

doi:10.1371/journal.pone.0139213.t002

In this study, we cloned and analyzed the 5′ flanking regulatory sequence of *AhLEC1B*, and found that *GUS* gene, driven by the whole-length Q1 construct, preferentially expressed in embryos of the transgenic *Arabidopsis*. On the other hand, the transgenic lines with 452bp-1156bp deletion constructs of Q1 from 5′ terminal showed higher GUS expression in roots, rosettes, stems, flowers, and seeds. Previous studies showed that the upstream region of *AtLEC1* promoter contains elements that repress its function in vegetative tissues [8]. Moreover, the seed-specific expression of the *AtLEC1* gene is controlled by combinatorial properties of negative and positive *cis*-regulatory elements in its promoter [8]. *PICKLE* (*PKL*)–a putative chromatin-remodeling factor–forms part of a NuRD histone deacetylase complex, which as a negative regulator of *AtLEC1* expression, represses embryonic identity and contributes to the transition from embryonic to postembryonic development in vegetative tissues [50, 51]. We hypothesize that the expression of *AhLEC1B* gene has a similar regulatory mode in peanut. The VP1/ABI3-LIKE (VAL) B3 proteins (as another repressor) in *Arabidopsis*, specific binding to the canonical sequence of Sph/RY *cis*-elemnets (CATGCA), are required for repression of the LEC1/B3 transcription factor network during gemination and vegetative development [33, 52]. Our results showed that an RY REPEAT element (CATGCA) localized at -1149bp of the *AhLEC1B* promoter region from TSS may be the binding site for VAL. The binding of VAL to the RY REPEAT element probably inhibits *AhLEC1B* expression in vegetative tissues. Moreover, the distal region of the *AhLEC1B* promoter consists of several other negative regulatory elements such as WRKY71OS (a transcriptional repressor of the gibberellin signaling pathway) and SRE (sugar-repressive element), which may be associated with upstream genes to decline its expression in particular way. Many elements required for the expression in embryo or endosperm, such as E BOX, CARGCW8GAT, and DPBF CORE, and so on, disperse in the Q1 construct. The E BOX elements are concentrated in the region from -250 to -50 in the promoters of some genes involved in fatty acid biosynthesis, triacylglycerol synthesis, and reserve including *SeFAD2*, *Cs-ACP1* and *Cs-4PAD*, *acyl-CoA-diacylglycerol acyltransferase* (At2g19450), *phosphatidylcholine*: *diacylglycerol acyltransferase* (At3g44830), several oil-body *oleoresins* (At3g01570, At3g18570, At3g27660, At5g40420, and At5g51210), and two *caleosins* (At4g26740 and At5g55240) [18]. The non-canonical CArG motif—CARGCW8GAT, which is an AGL15 (AT5G13790) transcription factor (TF) binding site is present in many endosperm-specific TF gene promoters [17, 53]. AGL15 might act upstream of the chalazal endosperm-specific TF genes and functions in activating at least one chalazal endosperm gene regulatory network [54].

The 300bp proximal region and 52bp 5′ UTR of the *AhLEC1B* promoter have crucial regulatory elements that are required for its basic activity and function. Deletion of these regulatory elements causes loss of reporter expression in Q2 transgenic lines (Fig 6). In this region, with the exception of TATA BOX, many tissue- or organ-specific elements, including phytohormone-responsive elements, light-regulated elements, elements associated with biotic stress and

abiotic stress response, etc., were found (Table 2). Furthermore, the 5′-end deletion analysis of Q1 construct indicated that the 65bp promoter fragment with the 52bp 5′ UTR, where TATA BOX, CACTFTPPCA1, DOF CORE, GATA BOX, SORLIP1AT, and the like exist, could satisfy its basic driving function. The promoter also drives the GUS activity in a manner similar to that of the CaMV 35S promoter in all detected tissues. In general, plant promoters have a distal region (upstream activation sequence) and a proximal region (core region of the promoter) located at about 30-40bp upstream of TSS. In our study, we found that *AhLEC1B* promoter harbored within -65~+52bp region has those crucial elements such as DOF CORE and GATA BOX and the like. Morton et al. (2014) [55] found that ROEs (regions of enrichment) of transcription factor binding site (TFBS) in the proximal promoter region within 40 nucleotides from the TSS are present either in Narrow Peak promoters or in those of Broad with Peak, where these crucial elements helpfully determine the profiles and levels of gene expression.

In conclusion, the *AhLEC1B* gene–with transcripts preferentially in the embryo–is co-regulated by the binding of upstream genes and the corresponding *cis*-regulatory elements in its promoter. The promoter elements may negatively and positively regulate the gene, and its 65bp promoter region plus 52bp 5′ UTR contain the key motifs required for the essential promoter activity.

Supporting Information

S1 Fig. The sequence of *AhLEC1B* genomic DNA. The sequences underlined indicated the exon1 and exon2, and the italics showed the 5′UTR sequence. The start and stop codon were shown using the bold letters.
(TIF)

Author Contributions

Conceived and designed the experiments: LS. Performed the experiments: GYT PLX WL. Analyzed the data: LS ZJL. Contributed reagents/materials/analysis tools: GYT PLX WL ZJL. Wrote the paper: LS.

References

1. Peng Y, Jahroudi N. The NFY transcription factor functions as a repressor and activator of the von Willebrand factor promoter. Blood 2002; 99: 2408–2417. PMID: 11895773
2. Ceribelli M, Dolfini D, Merico D, Gatta R, Vigano AM, Pavesi G, et al. The histone-like NF-Y is a bifunctional transcription factor. Mol Cell Biol. 2008; 28: 2047–2058. doi: 10.1128/MCB.01861-07 PMID: 18212061
3. Siefers N, Dang KK, Kumimoto RW, Bynum WE IV, Tayrose G, Holt BF III. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant Physiol. 2009; 149: 625–641. doi: 10.1104/pp.108.130591 PMID: 19019982
4. Petroni K, Kumimoto RW, Gnesutta N, Calvenzani V, Fornari M, Tonelli C, et al. The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. Plant Cell 2012; 24(12): 4777–4792. doi: 10.1105/tpc.112.105734 PMID: 23275578
5. Laloum T, De Mita S, Gamas P, Baudin M, Niebel A. CCAAT-box binding transcription factors in plants: Y so many? Trends Plant Sci. 2013; 18(3): 157–166. doi: 10.1016/j.tplants.2012.07.004 PMID: 22939172
6. Lotan T, Ohto M, 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: 1195–1205. PMID: 9657152
7. Warpeha KM, Upadhyay S, Yeh J, Adamiak J, Hawkins SI, Lapik YR, et al. The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in *Arabidopsis*. Plant Physiol. 2007; 143: 1590–1600. PMID: 17322342

8. Casson SA, Lindsey K. The turnip mutant of *Arabidopsis* reveals that *LEAFY COTYLEDON1* expression mediates the effects of auxin and sugars to promote embryonic cell identity. Plant Physiol. 2006; 142: 526–541. PMID: 16935993
9. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 1980; 8: 4321–4325. PMID: 7433111
10. Chen G, Shan L, Zhou LX, Tang GY, Bi YP. The comparison of different methods for isolating total RNA from peanuts. Chinese Agricultural Science Bulletin 2011; 27(1): 214–218. (In Chinese)
11. Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. 1998; 16: 735–743. PMID: 10069079
12. Murashige T, Skoog F. Arevised medium for rapid growth and bioassay with tobacco tissue cultures. Physiology Plant 1962; 15: 473–497.
13. Jefferson RA, Kavanagh, Bevan MW. GUS fusion: *beta*- glucuronidase as a sensitive and versatile gene fusion marker in higher plants. The European Molecular Biology Organization Journal 1987; 6: 3901–3907.
14. Li AQ, Xia H, Wang XJ, Li CS, Zhao CZ, Bi YP, et al. Cloning and expression analysis of peanut (*Arachis hypogaea* L.) *LEC1*. Acta Bot. Boreal. Occident. Sin. 2009; 29(9): 1730–1735. (In Chinese)
15. Allen RD, Bernier F, Lessard PA, Beachy RN. Nuclear factors interact with a soybean beta-conglycinin enhancer. Plant Cell 1989; 1: 623–631. PMID: 2535514
16. Yanagisawa S, Schmidt RJ. Diversity and similarity among recognition sequences of Dof transcription factors. Plant J 1999; 17: 209–214. PMID: 10074718
17. Tang W, Perry SE. Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. J Biol. Chem. 2003; 278: 28154–28159. PMID: 12743119
18. Kim MJ, Kim H, Shin JS, Chung CH, Ohlrogge JB, Suh MC. Seed-specific expression of sesame microsomal oleic acid desaturase is controlled by combinatorial properties between negative *cis*-regulatory elements in the SeFAD2 promoter and enhancers in the 5′UTR intron. Mol.Genet.Genomics 2006; 276: 351–368. PMID: 16862401
19. Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, et al. *cis*-Regulatory elements for mesophyll-specific gene expression in the C4 plant Flaveria trinervia, the promoter of the C4 phosphoenolpyruvate carboxylase gene. Plant Cell 2004; 16: 1077–1090. PMID: 15100398
20. Elmayan T, Tepfer M. Evaluation in tobacco of the organ specificity and strength of the rol D promoter, domain A of the 35S promoter and the 35S2 promoter. Transgenic Res. 1995; 4: 388–396. PMID: 7581519
21. Filichkin SA, Leonard JM, Monteros A, Liu PP, Nonogaki H. A novel endo-beta-mannanase gene in tomato *LeMAN5* is associated with anther and pollen development. Plant Physiol. 2004; 134: 1080–1087. PMID: 14976239
22. Sahoo DK, Sarkar S, Raha S, Maiti IB, Dey N. Comparative analysis of synthetic DNA promoters for high-level gene expression in plants. Planta 2014; 240: 855–875. doi: 10.1007/s00425-014-2135-x PMID: 25092118
23. Teakle GR, Manfield IW, Graham JF, Gilmartin PM. Arabidopsis thaliana GATA factors: organisation, expression and DNA-binding characteristics. Plant Mol. Biol. 2002; 50: 43–57. PMID: 12139008
24. Terzaghi WB, Cashmore AR. Light-regulated transcription. Annu Rev Plant Physiol Plant Mol. Biol. 1995; 46: 445–474.
25. Zhang ZL, Xie Z, Zou X, Casaretto J, Ho TH, Shen QJ. A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. Plant Physiol. 2004; 134: 1500–1513. PMID: 15047897
26. Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S. Gibberellin biosynthesis and response during *Arabidopsis* seed germination. Plant Cell 2003; 15: 1591–1604. PMID: 12837949
27. Fusada N, Masuda T, Kuroda H, Shimada H, Ohta H, Takamiya K. Identification of a Novel *Cis*-Element Exhibiting Cytokinin-Dependent Protein Binding in Vitro in the 5'-region of NADPH-Protochlorophyllide Oxidoreductase Gene in Cucumber. Plant Mol. Biol. 2005; 59: 631–645. PMID: 16244912
28. Rawat R, Xu ZF, Yao KM, Chye ML. Identification of *cis*-elements for ethylene and circadian regulation of the Solanum melongena gene encoding cysteine proteinase. Plant Mol. Biol. 2005; 57: 629–643. PMID: 15988560
29. Yamauchi D. A TGACGT motif in the 5'-upstream region of *alpha*-Amylase gene from *Vigna mungo* is a *cis*-element for expression in cotyledons of germinated seeds. Plant Cell Physiol. 2001; 42: 635–641. PMID: 11427683

30. Wu C, Washida H, Onodera Y, Harada K, Takaiwa F. Quantitative nature of the Prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal *cis*-element requirements for endosperm-specific gene expression. Plant J. 2000; 23: 415–421. PMID: 10929134
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: 1237–1251. PMID: 9225465
32. Kagaya Y, Ohmiya K, Hattori T. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. Nucleic Acids Res. 1999; 27: 470–478. PMID: 9862967
33. Suzuki M, Wang HHY, McCarty DR. Repression of the LEAFY COTYLEDON1/B3 regulatory network in plant embryo development by *VP1/ABSCISIC ACID INSENSITIVE3-LIKE B3* genes. Plant Physiol. 2007; 143: 902–911. PMID: 17158584
34. Breeden L, Nasmyth K. Cell cycle control of the yeast HO gene: *cis*- and *trans*-acting regulators. Cell 1987; 48: 389–397. PMID: 3542227
35. Mikami K, Tabata T, Kawata T, Nakayama T, Iwabuchi M. Nuclear protein(s) binding to the conserved DNA hexameric sequence postulated to regulate transcription of wheat histone genes. FEBS Lett. 1987; 223: 273–278. PMID: 2822486
36. Terada R, Nakayama T, Iwabuchi M, Shimamoto K. A type I element composed of the hexamer (ACGTCA) and octamer (CGCGGATC) motifs plays a role(s) in meristematic expression of a wheat histone H3 gene in transgenic rice plants. Plant Mol. Biol. 1995; 27: 17–26. PMID: 7865787
37. Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Two different novel *cis*-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. Plant J. 2003; 33: 259–270. PMID: 12535340
38. Ross EJ, Stone JM, Elowsky CG, Arredondo-Peter R, Klucas RV, Sarath G. Activation of the *Oryza sativa* non-symbiotic haemoglobin-2 promoter by the cytokinin-regulated transcription factor, ARR1. J Exp Bot. 2004; 55: 1721–1731. PMID: 15258171
39. Sutoh K, Yamauchi D. Two *cis*-acting elements necessary and sufficient for gibberellin-upregulated proteinase expression in rice seeds. Plant J. 2003; 34: 635–645. PMID: 12787245
40. Haralampidis K, Milioni D, Rigas S, Hatzopoulos P. Combinatorial interaction of *cis*- elements specifies the expression of the Arabidopsis AtHsp90-1 gene. Plant Physiol. 2002; 129: 1138–1149. PMID: 12114568
41. Quinn JM, Merchant S. Two copper-responsive elements associated with the Chlamydomonas Cyc6 gene function as targets for transcriptional activators. Plant Cell 1995; 7: 623–628. PMID: 7780310
42. Yanagisawa S. Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. Plant J. 2000; 21: 281–288. PMID: 10758479
43. Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GT, Gruissem W, et al. Genome-wide identification of potential plant E2F target genes. Plant Physiol. 2005; 139: 316–328. PMID: 16126853
44. Rogers HJ, Bate N, Combe J, Sullivan J, Sweetman J, Swan C, et al. Functional analysis of *cis*-regulatory elements within the promoter of the tobacco late pollen gene g10. Plant Mol. Biol. 2001; 45: 577–585. PMID: 11414616
45. Hudson ME, Quail PH. Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. Plant Physiol. 2003; 133: 1605–1616. PMID: 14681527
46. Plesch G, Ehrhardt T, Mueller-Roeber B. Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression. Plant J. 2001; 28: 455–464. PMID: 11737782
47. 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: 5–18. PMID: 12509518
48. Lee HS, Fischer RL, Goldberg RB, Harada JJ. *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. Proc Natl Acad Sci USA 2003; 100: 2152–2156. PMID: 12578989
49. Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori. *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. Plant J. 2009; 58(5):843–856. doi: 10.1111/j.1365-313X.2009.03817.x PMID: 19207209
50. Ogas J, Kaufmann S, Henderson J, Somerville C. PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. Proc Natl Acad Sci USA 1999; 96(24): 13839–13844. PMID: 10570159

51. Dean Rider S Jr, Henderson JT, Jerome RE, Edenberg HJ, Romero-Severson J, Ogas J. Coordinate repression of regulators of embryonic identity by PICKLE during germination in Arabidopsis. Plant J. 2003; 35: 33–43. PMID: 12834400

52. Jia H, McCarty DR, Suzuki M. Distinct roles of LAFL network genes in promoting the embryonic seedling fate in the absence of VAL repression. Plant Physiol. 2013; 163(3):1293–1305. doi: 10.1104/pp.113.220988 PMID: 24043445

53. Zhu C, Perry SE. Control of expression and autoregulation of AGL15, a member of the MADS-box family. Plant J. 2004; 41: 583–594.

54. 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 USA 2010; 107(18): 8063–8070. doi: 10.1073/pnas.1003530107 PMID: 20385809

55. Morton T, Petricka J, Corcoran DL, Li S, Winter CM, Carda A, et al. Paired-end analysis of transcription start sites in *Arabidopsis* reveals plant-specific promoter signatures. Plant Cell 2014; 26(7): 2746–2760. doi: 10.1105/tpc.114.125617 PMID: 25035402