HOKKAIDO UNIVERSITY

Title	Molecular-Genetic Study on a Novel Photoperiod-Insensitive Gene in Soybean
Author(s)	朱, 江慧
Citation	北海道大学. 博士(農学) 甲第13593号
Issue Date	2019-03-25
DOI	10.14943/doctoral.k13593
Doc URL	http://hdl.handle.net/2115/91548
Type	theses (doctoral)
File Information	Zhu_Jianghui.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Molecular-Genetic Study on a Novel Photoperiod-Insensitive Gene in Soybean

（ダイズにおける新規の非感光性遺伝子に関する分子遺伝学的研究）

Hokkaido University Graduate School of Agriculture

Division of Agrobiology and Bioresources Doctoral Course

Jianghui Zhu

Contents

Chapter 1 Introduction

Photoperiod response of flowering determines the adaptation of crops to a wide range of latitudes with different daylengths during growing seasons. Its regulatory mechanisms vary with plant species, and may rely on both evolutionally conserved and species-specific gene systems. In *Arabidopsis thaliana*, a long-day (LD) plant, *CONSTANS* (*CO*) plays a key role in regulation of photoperiodic flowering; transcriptional and post-translational regulations of *CO* result in accumulation of the CO protein in the late afternoon under LD conditions, which in turn activates *FLOWERING LOCUS T* (*FT*), florigen, gene expression (reviewed by Andrés and Coupland, 2012; Song et al., 2013). Similarly, in rice, a short-day (SD) plant, a *CO* orthologue, *Heading date 1* (*Hd1*) (Yano et al., 2000), regulates the *FT* orthologues *Heading date 3a* (*Hd3a*) and *Rice FT-like 1* (*RFT1*) (Kojima et al., 2002; Tamaki et al., 2007).

Soybean (*Glycine max* (L.) Merr.) is a typical short-day plant. Eleven major genes for flowering have been reported so far (*E1–E9* and *J*, reviewed by Cao et al., 2017; *E10*, Samanfar et al., 2017). Among them, four maturity genes, *E1* to *E4*, are the main contributors to soybean adaptation to a wide range of latitudes (Liu et al., 2011; Jia et al., 2014; Jiang et al., 2014; Langewisch et al., 2014; Tsubokura et al., 2014; Zhai et al., 2014; Lu et al., 2015; Kurasch et al., 2017; Li et al., 2017).

The floral repressor *E1* encodes a protein that contains a bipartite nuclear localization signal (NLS) and a region distantly related to the B3-like DNA-binding domain, and is a possible transcription factor that represses the expression of major soybean *FT* orthologues *FT2a* and *FT5a* (Xia et al., 2012). *E1* expression is up-regulated under long day (LD) conditions under the control of the phytochrome A (phyA) proteins E3 and E4

(Liu et al., 2008; Watanabe et al., 2009; Xia et al., 2012). *E2*, a soybean orthologue of *Arabidopsis GIGANTEA* (*GI*) (Watanabe et al., 2011), inhibits flowering under LD conditions through a pathway distinct from the phyA-regulated *E1* pathway (Xu et al., 2015; reviewed by Cao et al., 2017). *E1* has two homologues, *E1-Like-a* (*E1La*) and *E1Lb*, located 10,640 kb apart from each other in the homoeologous region of chromosome 4 (Xia et al., 2012; Xu et al., 2015). Down-regulation of the *E1L* genes by virus-induced gene silencing (VIGS) in a cultivar deficient in the *E1* gene leads to early flowering and abolishes the night-break response, suggesting that the two *E1L* genes are also involved in the photoperiod responses of soybean (Xu et al., 2015).

Photoperiod-insensitivity in soybean is conditioned by combinations of various alleles at *E1*, *E3*, and *E4* (Tsubokura et al., 2013, 2014; Xu et al., 2013; Zhai et al., 2015). *E3* and *E4* were originally identified as major genes for different responses of flowering to artificially induced LD conditions, where natural daylength was extended to 20 h with red light (R)-enriched cool white fluorescent lamps (fluorescent-long daylength; FLD) or far red light (FR)-enriched incandescent lamps (incandescent-long daylength; ILD) (Buzzell 1971; Buzzell and Voldeng, 1980; Saindon et al., 1989). *e3* conditions flowering under the FLD condition (Buzzell 1971), whereas *e4* does so under the ILD condition in the *e3* background (Saindon et al., 1989), suggesting that E3 and E4 are functionally diverged and have an epistatic relationship. On the basis of the functions of alleles at the three loci, Xu et al. (2013) classified ILD-insensitive cultivars into three genotypic groups: (group 1) the dysfunction of both *E3* and *E4*; (group 2) the dysfunction of *E1* in combination with that of either *E3* or *E4*; and (group 3) a combination of *e1-as* (hypomorphic allele), *e3*, and *E4*. Because *E4* inhibits flowering under ILD conditions (Saindon et al., 1989; Cober et al., 1996; Abe et al., 2003; Liu and Abe, 2010), the group

3 cultivars are predicted to have novel genes that abolish or reduce ILD-sensitivity. One such gene is an early-flowering allele (*FT5a-ef*) at *qDTF-J*, a QTL for days to flowering in linkage group J, which encodes FT5a; early flowering is caused by its increased transcriptional activity or mRNA stability associated with an insertion in the promoter and/or deletions in the 3′-untranslated region (UTR) (Takeshima et al., 2016).

In the thesis, I studied a molecular-genetic mechanism of photoperiod-insensitivity in group 3 soybean cultivars introduced from Far-Eastern Russia. In Chapter 2, I identified a novel loss-of-function allele at the *E1Lb* locus as a most likely causal factor for the photoperiod-insensitivity, by the genetic analysis including the association test, fine-mapping and sequence analysis. In Chapter 3, I characterized the function and roles of the loss-of-function allele on flowering under LD conditions, by using near-isogenic lines. In Chapter 4, I determined the interaction between the E1 family proteins, E1, E1La and E1Lb, by yeast two-hybrid assay. In Chapter 5, I surveyed the molecular diversity for the *E1La* and *E1Lb* genes. Finally, I discussed 1) the molecular-genetic basis for the photoperiod-insensitivity of Far-Eastern Russian cultivars, and 2) the regulatory mechanisms and functions of *E1* family genes in the control of flowering, and 3) the molecular diversity of *E1La* and *E1Lb* genes and its usefulness in soybean breeding in high latitudes.

Chapter 2 Identification of a novel gene for photoperiod-insensitivity

2.1 Background and Purposes

Soybean is basically a short day plant. Its flowering is generally repressed under long day conditions. High latitude areas are characterized by long day lengths in early summer and relatively short duration of warm climate suitable for the soybean cultivation. Thus soybean cultivars cultivated in high latitudes need to lose or lower the sensitivity of flowering and maturation to longer daylengths. In particular, the photoperiod-insensitivity has been known to be mainly controlled by loss-of-function alleles at *E1* or at both *E3* and *E4* (Watanabe et al., 2012; Buzzell et al., 1971; Buzzell et al., 1980; Saindon et al., 1989; Cober et al., 1996; Abe et al., 2003). In addition to these major genotypes, Xu et al (2014) found that some photoperiod-insensitive cultivars introduced from North-Eastern China and Far-Eastern Russia possessed the same genotype at the *E1*, *E3* and *E4* loci as Harosoy isoline for *e3* (H-*e3*), a photoperiod-sensitive line whose genotype is *e1-as*/*e3*/*E4*. Thus, these cultivars may possess a novel gene(s) involved in the photoperiod-insensitivity. Takeshima et al (2016) identified an upregulated transcript abundance of *FT5a* caused by DNA polymorphisms in the promoter and/or 3′-UTR regions as one of the factors for the photoperiod-insensitivity. In this chapter, I performed the genetic analyses for the photoperiod-insensitivity of three Russian cultivars, Zeika, Yubileinaya and Sonata.

2.2 Materials and Methods

2.2.1 Plant materials and segregation analysis

The indeterminate Far Eastern Russian soybean cultivars Zeika (ZE), Yubileinaya (YU), and Sonata were crossed with the Canadian indeterminate cultivar Harosoy (L58-266; HA); ZE and YU were also crossed with a Harosoy near-isogenic line for *e3* (PI547716; H-*e3*). The three Russian cultivars have the same genotype as H-*e3* at five maturity loci, *E1*, *E2*, *E3*, *E4,* and *E9* (*e1-as*/*e2*/*e3*/*E4*/*E9*), but unlike H-*e3* they flower without any marked delay under ILD conditions in comparison with natural daylength (ND) conditions (maximum daylength, 15.2 h) in Sapporo, Japan (43°07′N, 141°35′E) (Xu et al., 2013). The ILD condition was set at an experimental farm of Hokkaido University by extending the ND to 20 h by supplemental lighting from 2:00 to 7:00 and from 18:00 to 22:00 with incandescent lamps with a red-to-far-red (R:FR) quantum ratio of 0.72 (Abe et al., 2003). Seeds of F_2 populations and parents were sown in paper pots (Paperpots No. 2, Nippon Beet Sugar Manufacturing Co., Tokyo, Japan) on 28 May 2013 for the crosses with HA and 26 May 2014 for crosses with H-*e3*. The pots were put under the ILD condition, and 12 days later seedlings were transplanted into soil. The progeny test was carried out for 48 F_2 plants randomly selected from the H-*e3* × ZE cross and recombinant plants used in fine mapping. Seeds of these plants were sown in paper pots on late May in 2015 to 2017 (25 May, 2015, 28 May, 2016, 26 May, 2017). After 12 days under the ILD condition, 15 seedlings per plant were transplanted into the same field. The number of days from sowing to the first flower opening (R1) (Fehr et al., 1971) of each plant was recorded.

2.2.2 DNA extraction and SSR marker analysis

Total DNA was extracted from trifoliate leaves of each of 150 H-*e3* × ZE and 156 H-*e3* × YU F_2 plants as described by Doyle and Doyle (1990), and from each of 492 seeds from two F_2 plants from the H-*e3* × ZE cross, as described by Xia et al. (2012). Each PCR mixture for SSR marker analysis contained 30 ng of total genomic DNA as a template, 0.2 μL of each primer (10 μM), 0.8 μL of dNTPs (2.5 mM), 0.1 μL of Taq DNA polymerase (Ampliqon, Denmark), and 1 μL of 10× ammonium buffer (Ampliqon, Denmark) in a total volume of 10 μL; amplification conditions were 35 cycles at 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. PCR products were separated by electrophoresis in 10.5% (w/v) polyacrylamide gels, stained with ethidium bromide, and visualized under UV light.

2.2.3 Association test, linkage map construction, and fine mapping

A total of 16 F_2 plants from the H-*e3* × ZE cross were used to test the association of ILD sensitivity with simple sequence repeat (SSR) marker genotypes. They were selected based on the segregation pattern in their progeny, and included 8 plants fixed for ILD insensitivity and 8 plants fixed for ILD sensitivity. SSR markers were chosen from those located in genomic regions that harbored the soybean orthologues of *Arabidopsis* flowering genes (Song et al., 2004; Watanabe et al., 2012). The SSR markers significantly associated with ILD sensitivity were genotyped for a total of 306 F_2 plants from the H-*e3* × ZE and H-*e3* × YU crosses to confirm the detected association. Plants recombined in the targeted region were subjected to fine mapping; the genotypes for the target gene were

estimated based on the segregation of flowering under the ILD condition in the progeny and were compared with the graphical genotypes constructed by using additional 11 BARCSOY SSR markers (Song et al., 2010) (Supplemental Table 1).

2.2.4 Sequencing analysis

The coding sequences of the three gene models, Glyma.04G143300, Glyma.04G143400 and Glyma.04G143500, were analyzed for H-*e3* and ZE. The coding sequences were amplified from the cDNAs by using primers listed in Supplemental Table 3. The amplified fragments were cloned into a pGEM-T Easy vector (Promega, Japan). Plasmid DNAs were extracted with Wizard® Plus SV Minipreps DNA Purification Systems (Promega, Japan) following the manufacturer's instructions and sequenced with a BigDye Terminator v3.1 Cycle Sequencing kit and an ABI PRISM 3100 Avant Genetic Analyzer (both from Applied Biosystems, Japan) according to the manufacturer's instructions.

2.2.5 DNA marker analysis of *E1Lb*

A derived cleaved amplified polymorphic sequence (dCAPS) marker targeting the single-base deletion observed in ZE was developed to discriminate the functional *E1Lb* allele from H-*e3* and the loss-of-function *e1lb* allele from ZE. The 275-bp DNA fragment amplified from ZE by PCR with the forward primer 5′-GTGTAAACACTCAAAGTCCTT-3′ and the reverse primer 5′-CGTCTTCTTGATCTTCCAACG-3′ was digested with HpyCH4IV (New England Biolabs, Japan) into two fragments, 254 bp and 21 bp, but the 276-bp fragment amplified

from H-*e3* was resistant to HpyCH4IV digestion. The PCR products were treated with HpyCH4IV for 1 h and then separated by electrophoresis in 2.5% NuSieve 3:1 gel (Lonza, Japan), stained with ethidium bromide, and visualized under UV light.

2.3 Results and Discussion

2.3.1 Segregation of flowering time in F_2 and F_3 populations

The three Russian cultivars are photoperiod insensitive (Xu et al., 2013). They flowered 45 to 47 days after sowing (DAS) under the ND condition of Sapporo, whereas H-*e3* and HA flowered approximately 5 and 10 days later, respectively. Under the ILD condition, the three cultivars and H-*e3* flowered 2–4 days and around 20 days later than under ND, respectively, whereas HA continued vegetative growth and did not develop any flower buds until the end of light supplementation (10 August, 76 DAS).

Flowering time under the ILD condition in F_2 populations of the H-*e3* × ZE and H-*e3* × YU crosses varied continuously from that of ILD-insensitive parents (45 DAS for ZE and 46 DAS for YU) to the end of light supplementation; 10 out of 150 and 12 out of 156 plants had no flower bud in the H-*e3* × ZE and H-*e3* × YU F_2 populations, respectively (Figure 1). In both populations, the distribution of flowering time tended to be bi-modal; plants which flowered at 56 DAS and later or remained vegetative segregated more than those flowered earlier. 48 H-*e3* × ZE F_2 plants were randomly selected and the progeny for flowering time segregation under the ILD condition were tested. Based on the segregation pattern, the 48 F_2 plants could be classified into three groups: (1) plants fixed for ILD insensitivity (all F_3 plants tested flowered as ZE did; *e*/*e*); (2) those segregating for flowering time (*E*/*e*) and (3) those fixed for ILD sensitivity (all F_3 plants tested showed delayed or no flowering; *E*/*E*) (Figure 1A). The number of plants was 8 in *e*/*e*, 23 in *E*/*e*, and 17 in *E*/*E*, in consistence with a monogenic 1:2:1 ratio (χ^2 = 3.81, df = 2, p = 0.18), suggesting the involvement of a single recessive gene for ILD insensitivity. Based on the results of the progeny test, the 306 F_2 plants were classified

into early-flowering ILD-insensitive plants, which flowered before 56 DAS, and late- or non-flowering ILD-sensitive plants (Figure 1). The segregation ratios of the two classes fit the expected 3:1 ratio ($\chi^2 = 0.33$, $df = 1$, $p = 0.56$ for H-*e3* × ZE, $\chi^2 = 3.28$, $df = 1$, $p = 0.07$ for H-*e3* × YU), confirming that ILD insensitivity is controlled mainly by a single recessive gene.

The segregation of flowering time under the ILD condition for the crosses between HA and the three Russian cultivars were also examined. Because HA has the *E3* allele while the three cultivars have the *e3* allele, it is considered that, in addition to the gene for ILD insensitivity segregated in the crosses with H-*e3*, the *E3* locus would also segregate in the F_2 populations. In the three crosses, however, ILD-insensitive plants segregated at frequencies of 21.1% to 33.9 %; the remaining plants remained vegetative until the end of light supplementation (Table 1). These segregation frequencies were thus inconsistent with those of a two-gene model, but were close to those expected from monogenic inheritance, as in the crosses with H-*e3* (Table 1).

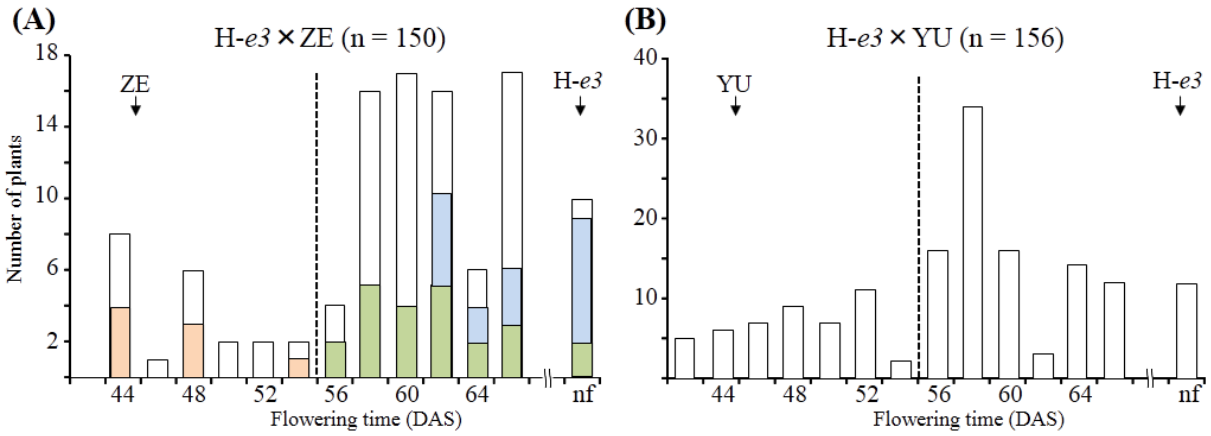


Figure 1. Segregation of flowering time in F_2 populations of crosses between a Harosoy NIL for *e3* (H-*e3*) and the incandescent-long daylength (ILD)-insensitive cultivars Zeika (ZE) and Yubileinaya (YU) under far red light-enriched ILD conditions. **(A)** H-*e3* × ZE; **(B)** H-*e3* × YU. In a cross between H-*e3* and ZE, 48 F_2 plants were selected for the progeny test; ILD-sensitivity genotypes were estimated based on the segregation in the progeny. Pink bars, homozygotes for ILD-insensitivity (*e*/*e*); yellow–green bars, heterozygotes (*E*/*e*); light-blue bars, homozygotes for ILD-sensitivity (*E*/*E*). Arrows indicate mean values of flowering time in parents. Dotted vertical lines indicate the threshold for classification of F_2 plants into early-flowering ILD-insensitive and late- or non-flowering ILD-sensitive. nf, no flower buds by the end of light supplementation. DAS, days after sowing.

Table 1. Segregation of ILD-insensitivity in F_2 of crosses of an ILD-sensitive cultivar Harosoy with ILD-insensitive Russian cultivars

Cross combination	No. of plants			χ^2 value for 1:3	*P* value
	ILD-insensitive	ILD-sensitive	TOTAL		
Harosoy × Zeika	19	37	56	3.57	0.059
Harosoy × Yubileinaya	28	105	133	1.66	0.198
Harosoy × Sonata	19	54	73	0.06	0.803

2.3.2 Association test, linkage map construction, and fine mapping

To determine the genomic position of the gene for ILD insensitivity from ZE, I tested the association between ILD sensitivity and SSR marker genotypes. Based on the results of the progeny test, 16 F_2 plants from the H-*e3* × ZE cross were selected, 8 homozygous for ILD insensitivity (*e*/*e*), and 8 homozygous for ILD sensitivity (*E*/*E*). Among the SSR markers tested, Satt190 and Sat_085 in linkage group C1 (chromosome 4; Chr04) showed genotypic variation in complete accordance with the ILD sensitivity (Figure 2A). Then the genotypes of the two markers in the whole F_2 plants of H-*e3* × ZE and H-*e3* × YU populations (Figure 2B and 2C) were determined. The two markers were tightly linked to each other with a recombination value of 2.1, and were closely associated with ILD sensitivity. All of the plants homozygous for the allele from ILD-insensitive parents at Sat_085 (I/I) flowered before 56 DAS (H-*e3* × ZE) or 52 DAS (H-*e3* × YU), whereas those homozygous for the allele from ILD-sensitive H-*e3* (S/S) flowered at ≥60 DAS or did not flower in both crosses. Heterozygous plants (I/S) mostly flowered at ≥58 DAS (H-*e3* × ZE) or ≥54 DAS (H-*e3* × YU), which partly overlapped with the flowering date ranges of the S/S plants; only a few plants flowered as early as the I/I plants. These results strongly suggested that a gene for ILD insensitivity is located near the two SSR markers.

Satt190 and Sat_085 are located 17.3 Mb from each other in the pericentromeric region of chromosome 4 (Schmutz et al., 2010) (Phytozome v12.1/*Glycine max Wm82.a2.v1*; https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax). To delimit the genomic region of the gene for ILD-insensitivity more precisely, plants with recombination between the two markers (7 from 306 F_2 plants from the H-*e3* × ZE and H-*e3* × YU crosses and 3 from 492 F_3 plants from the H-*e3* × ZE cross) were selected and

constructed their graphical genotypes with 11 SSR markers. A comparison of the graphical genotypes with the genotype of ILD-insensitivity estimated by the progeny test revealed that the gene for ILD-insensitivity was located between SSR markers M5 (BARC-18g-0889) and M6 (BARC-18g-0895) (Figure 3). The physical distance between the two markers was 842 kb, and the delimited region contained only 6 annotated genes (Phytozome v12.1/*Glycine max Wm82.a2. v1*) (Figure 3 and Table 2). RNA-sequencing Atlas in Phytozome v12.1/*Glycine max Wm82.a2. v1* indicates that Glyma.04G143000, Glyma.04G143100 and Glyma.04G143200 are expressed only in flower or root tissues, whereas Glyma.04G143300, Glyma.04G143400 and Glyma.04G143500 are expressed in leaves (Severin et al., 2010). Because ZE exhibited significantly higher expressions for *FT2a* and *FT5a* in leaves in the 2nd and 3rd trifoliate leaf stages than H-*e3* under R-enriched LD condition (Figure 4), I focused on the three genes expressed in leaves as a possible candidate of the gene for ILD-insensitivity that upregulates the two *FT* genes.

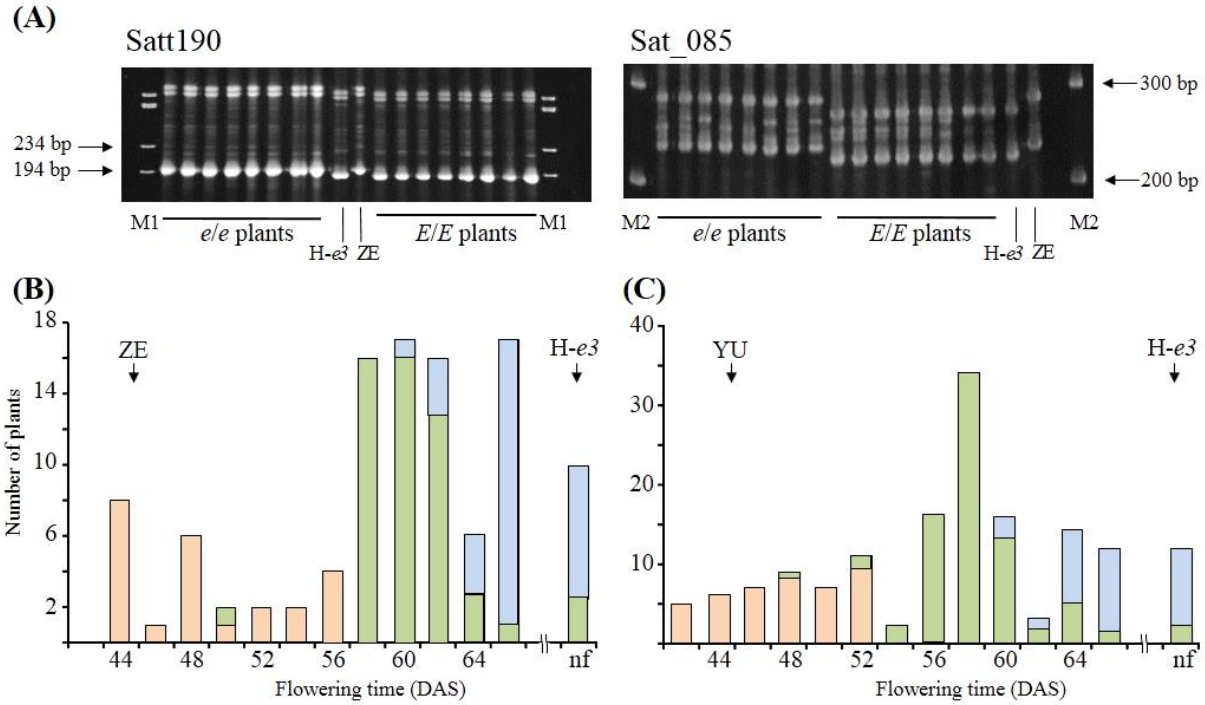


Figure 2. Simple sequence repeat (SSR) marker analyses in F_2 plants from a Harosoy isoline for *e3* (H-*e3*) × Zeika (ZE) cross. **(A)** Gel electrophoresis for the analysis of Satt190 and Sat_085. Eight plants homozygous for the ILD-insensitive allele (*e*/*e*) and 8 plants homozygous for the ILD-sensitive allele (*E*/*E*) were selected on the basis of the results of the progeny test. M1, *φ*X174/HaeIII digest; M2, 100 bp DNA ladder. **(B, C)** Association between Sat_085 and flowering time in H-*e3* × ZE **(B)** and H-*e3* × YU **(C)**. F_2 plants were classified based on the genotype at Sat_085. Pink bars, homozygotes for the allele from ILD-insensitive parents (I/I); yellow-green bars, heterozygotes (I/S); light-blue bars, homozygotes for the allele from ILD-sensitive H-*e3* (S/S). Arrows indicate mean values of flowering time in parents. DAS, days after sowing.

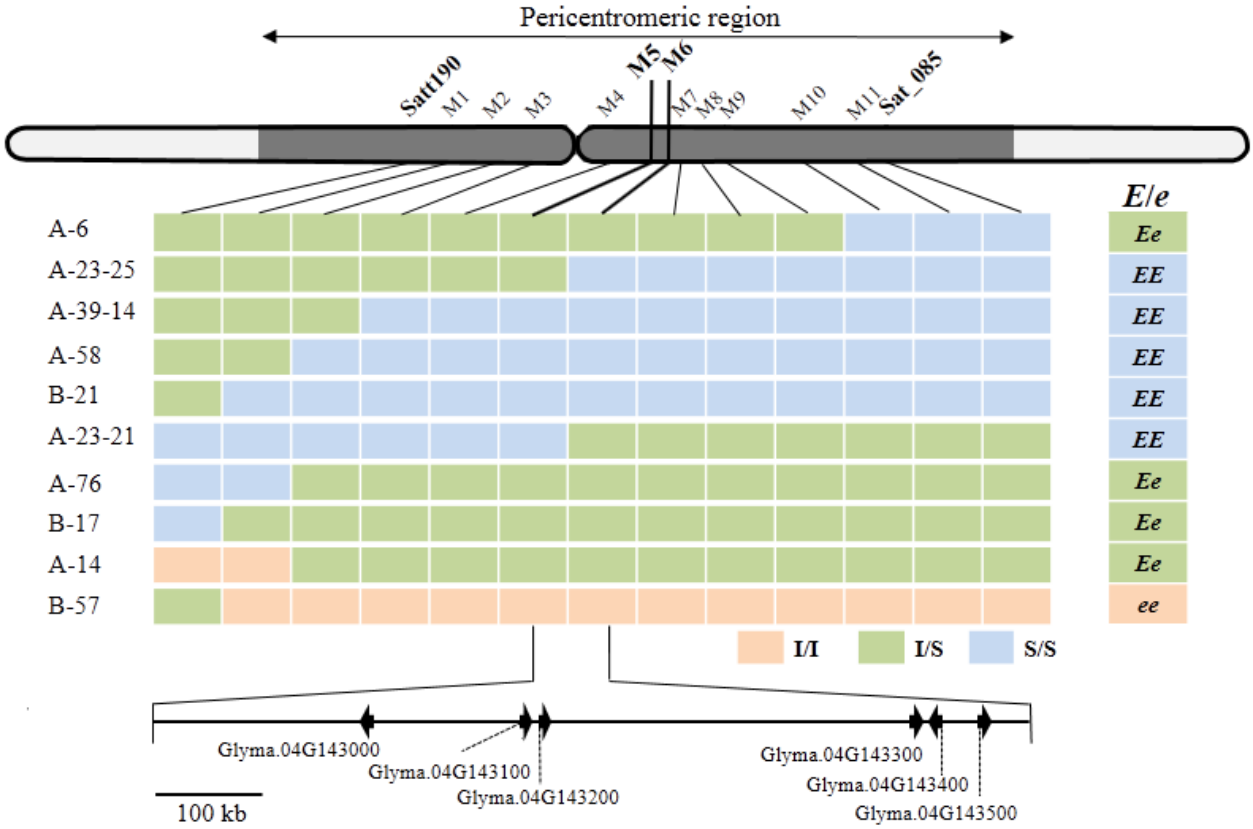


Figure 3. Fine mapping of a locus for ILD insensitivity (*E*/*e*) and annotated genes in the delimited region of Williams 82 chromosome 4 (*Glycine max Wm82.a2. v1*). M1 to M11, BARCSOY-SSR markers; Pink bars, homozygotes for the allele from ILD-insensitive parents (I/I); yellow–green bars, heterozygotes (I/S); light blue bars, homozygotes for the allele from ILD-sensitive H-*e3* (S/S). The *E* genotype for ILD insensitivity was estimated from the segregation patterns in the progeny. Six genes annotated in the region between M5 and M6 are shown at the bottom.

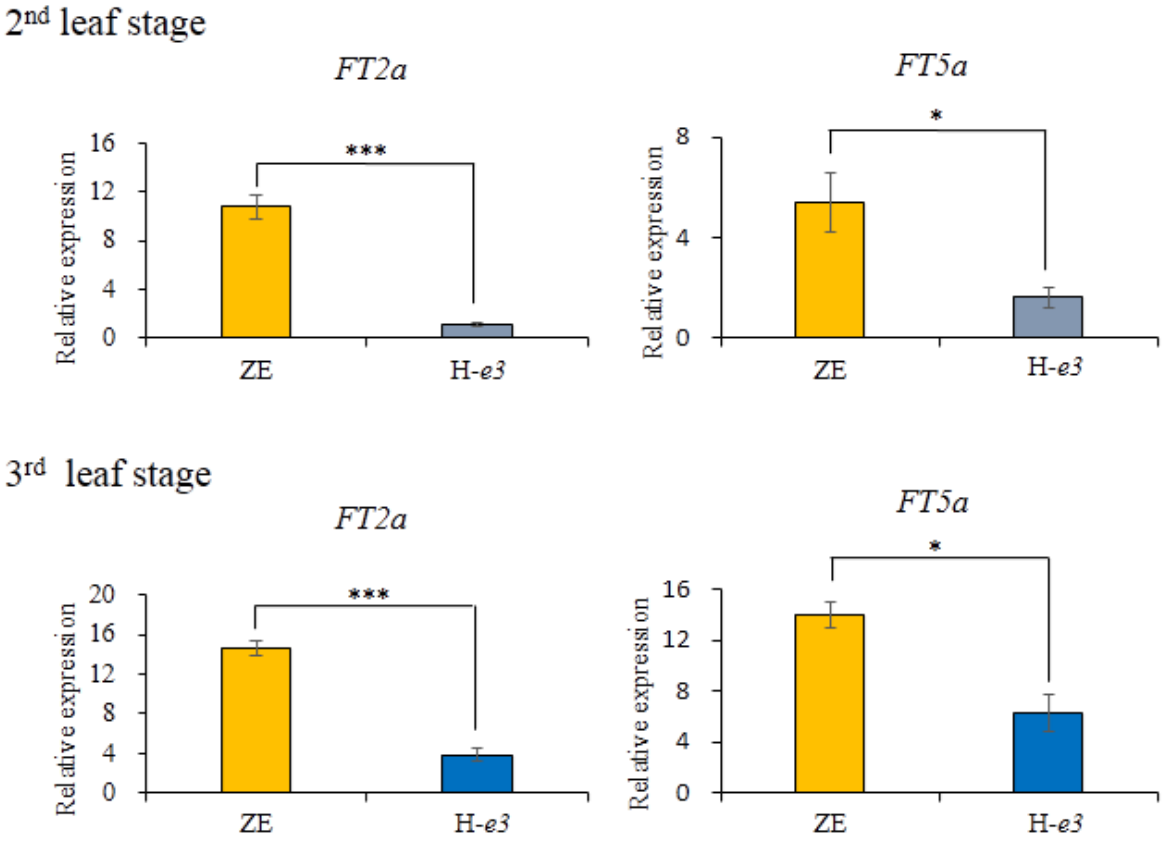


Figure 4. Expression levels of *FT2a* and *FT5a* at the second and third leaf stages in ILD-insensitive Zeika (ZE) and ILD-sensitive Harosoy NIL for *e3* (H-*e3*) under R-enriched LD (20 h) conditions. Relative mRNA levels (mean and standard error, n = 3) are expressed as the ratios to *β-tubulin* transcript levels. *, $p < 0.05$; ***, $p < 0.005$.

Table 2. Genes annotated in an 842-kb genomic region in chromosome 4 delimited by fine-mapping

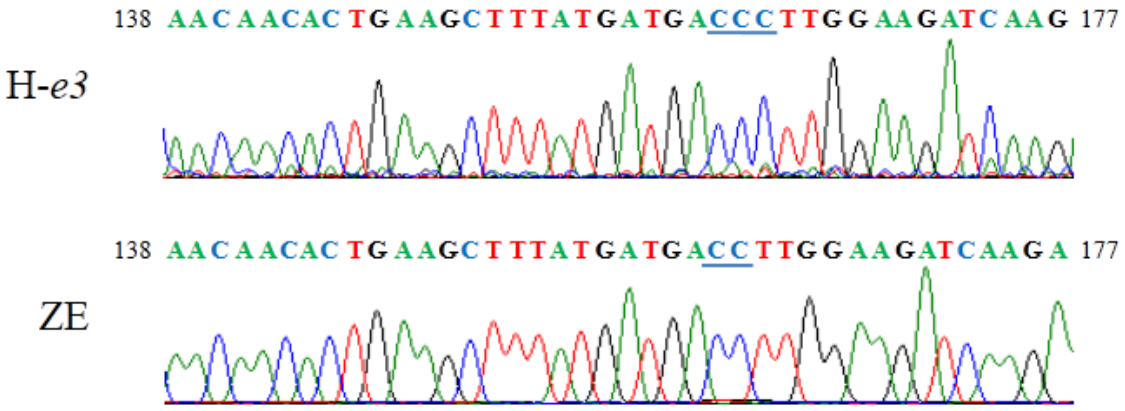
No.	Gene	Annotation (Phytozome V12.1/ *Glycine max Wm82.a2. v1*)	Expressed tissues
1	Glyma.04G143000	Diacylglycerol kinase 7	flower
2	Glyma.04G143100	RNA-binding (RRM/RBD/RNP motifs) family protein	root
3	Glyma.04G143200	Pectin lyase-like superfamily protein	flower
4	Glyma.04G143300	AP2/B3-like transcriptional factor family protein, *E1Lb*	leaf
5	Glyma.04G143400	Cytidine/deoxycytidylate deaminase family protein	leaf, root
6	Glyma.04G143500	Mitochondrial substrate carrier family protein	flower, leaf

Data on expressed tissues are referred from *Glycine max Wm82.a2. v1*. (Severin et al., 2010)

2.3.3 Sequence analysis

The coding sequences of the three gene candidates, Glyma.04G143300, Glyma.04G143400 and Glyma.04G143500 for H-*e3* and ZE were analyzed. Sequence analysis revealed that ZE and H-*e3* possessed identical sequences for Glyma.04G143400 and Glyma.04G143500, whereas one of cytosines at the 162th nucleotide to 164th nucleotide from the adenine of the start codon was deleted in the Glyma.04G143300 from ZE. This deletion generated a premature stop codon, and the Glyma.04G143300 from ZE was predicted to encode a truncated protein of 61 amino acids (Figure 5). Glyma.04G143300 is reported as *E1Lb*, one of two homoeologues (*E1La* and *E1Lb*) of floral repressor *E1* (Xia et al., 2012). Because the down-regulation of *E1La* and *E1Lb* expressions by VIGS promotes flowering under non-inductive conditions such as LD and night break (Xu et al., 2015), it is considered the loss-of-function allele of *E1Lb* (designated *e1lb* hereafter) as the most probable causal factor for the ILD-insensitivity.

(A)



(B)

```
W82    MSNHSDEKEQCQKKRKSTICEASNFRTSRRRFCSNKNEEEMNKGVSTTLKLYDDPWKIKK 60
H-e3   MSNHSDEKEQCQKKRKSTICEASNFRTSRRRFCSNKNEEEMNKGVSTTLKLYDDPWKIKK 60
ZE     MSNHSDEKEQCQKKRKSTICEASNFRTSRRRFCSNKNEEEMNKGVSTTLKLYDDLGRSRR 60
       ************************************************************

W82    TLTDSDLGILSRLSLATDLVKKQILPMLGADHARAAETEEGTPVRVWDMDTKSMHQLVLK 120
H-e3   TLTDSDLGILSRLSLATDLVKKQILPMLGADHARAAETEEGTPVRVWDMDTKSMHQLVLK 120
ZE     R*------------------------------------------------------------ 61

W82    RWSSSKSYVLIAKWNQDFVRRRDLKKGDEIGFHWDPYNCVFNFCVLKRAMPEN* 173
H-e3   RWSSSKSYVLIAKWNQDFVRRRDLKKGDEIGFHWDPYNCVFNFCVLKRAMPEN* 173
ZE     ---------------------------------------------------
```

Figure 5. DNA and predicted amino acid sequences of Glyma.04G143300 (*E1Lb*) in Williams 82 (W82), Harosoy isoline for *e3* (H-*e3*), and Zeika (ZE). **(A)** DNA sequences of the 138th nucleotide to 177th nucleotide from the adenine of the stat codon. One of cytosines at the 162th nucleotide to 164th nucleotide underlined was deleted in ZE. **(B)** Predicted amino acid sequences. The asterisks under the amino acid sequences mean identities of amino acids, the asterisk at the end of each amino acid sequence means stop codon.

2.3.4 DNA marker analysis of *E1Lb*

I developed a dCAPS marker to discriminate *e1lb* from *E1Lb* (Figure 6). The PCR-amplified fragment of 275 bp from ZE produced a shorter fragment of 254 bp when digested with HpyCH4IV, whereas that from H-*e3* (276 bp) was not digested. The digestion of the PCR products from YU and Sonata (Russian cultivar) produced 254-bp fragments, indicating that these two cultivars had the same deletion as ZE (Figure 6B). Therefore, the segregation of ILD-insensitive plants in the crosses of these cultivars with HA and H-*e3* were most likely caused by *e1lb*.

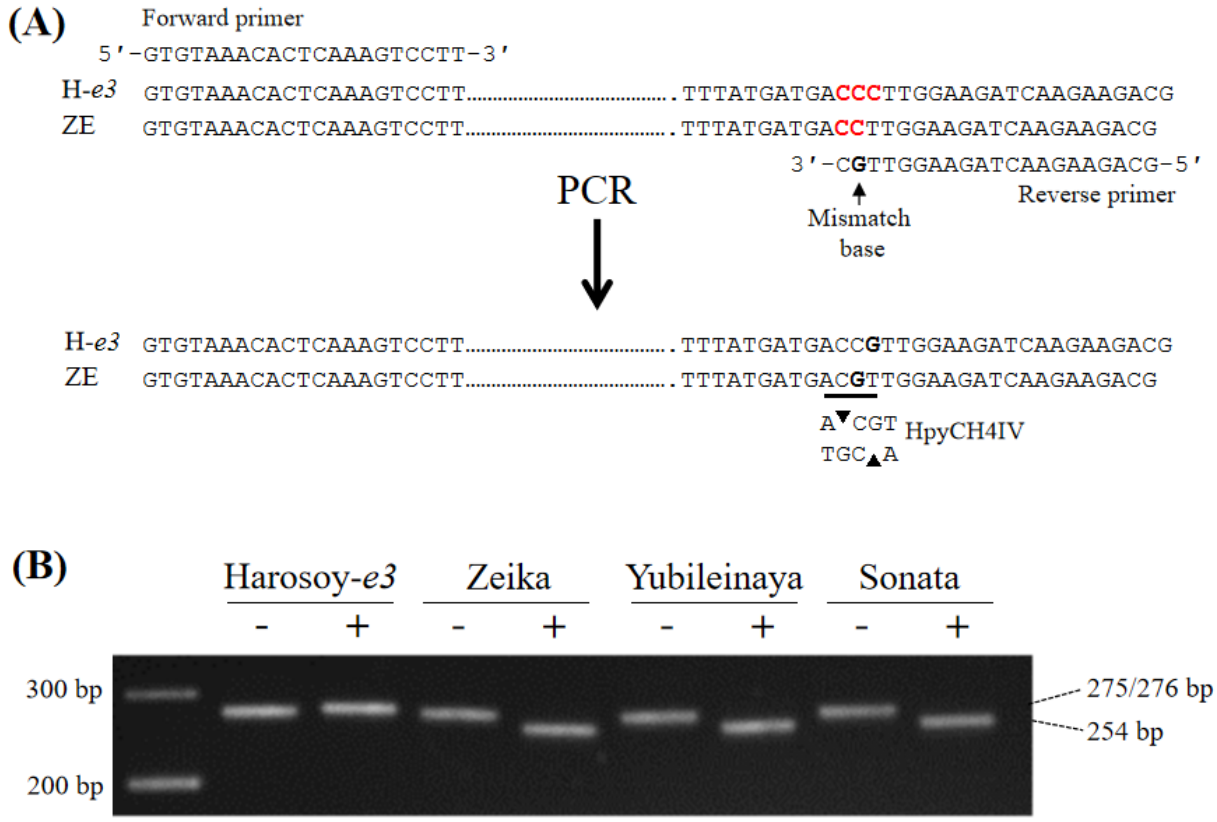


Figure 6. Derived cleaved amplified polymorphic sequence (dCAPS) marker analysis to detect the single-base deletion in the *e1lb* allele. **(A)** Primers designed and the generation of an HpyCH4IV restriction site. One of three cytosines marked by red was deleted in ZE. **(B)** Gel electrophoresis of PCR products without (–) or with (+) HpyCH4IV digestion. H-*e3*, Harosoy NIL for *e3*; ZE, Zeika.

Chapter 3 Characterization of the *e1lb* allele in the control of flowering

3.1 Background and Purposes

The *e1lb* allele was considered as the most probable causal factor for the ILD-insensitivity of the three Russian cultivars. To determine the effect of *e1lb* on flowering, four sets of NILs for *e1lb* and *E1Lb* were developed from the crosses of ZE with Harosoy and H-*e3*. The two sets of NILs possessed the *e3*/*E4* genotype, and the other two did the *E3*/*E4* genotype. I compared the effect of *e1lb* on flowering time and *FT2a*/*FT5a* expressions against *E1Lb* under both *e3*/*E4* and *E3*/*E4* genetic backgrounds.

3.2 Materials and Methods

3.2.1 Development of NILs

Four sets of NILs, each including one NIL for ILD insensitivity and another for sensitivity, were developed from heterozygous inbred F_5 plants derived from different F_2 plants (#4 and #21) from the H-*e3* × ZE cross and those (#11 and #20) from the HA × ZE cross. The former two sets of NILs had the recessive *e3* allele, whereas the latter two had the dominant *E3* allele. These lines, together with parents and an ILD-insensitive NIL of HA for *e3* and *e4* (PI546043; H-*e3e4*), were cultivated in a growth chamber (25°C, 20-h daylength) with an average photon flux of 120 μmol m^{-2} s^{-1} and an R:FR ratio of 2.2 at 1 m below light sources, or in the field under the ILD condition (sowing date, 26 May 2018), as described above. For comparison, indeterminate NILs for alleles, *e1-nl* and *e1-as*, at *E1* (NIL-*E1*; *e2*/*E3*/*E4*/*E9*), which were developed from a heterozygous inbred F_5

plant derived from a cross between the Japanese determinate cultivar Toyomusume (*e1-nl*/*e2*/*E3*/*E4*/*e9*) and HA, were included in the evaluation of flowering under the ILD condition.

3.2.2 RNA extraction and reverse transcription

A new fully expanded leaflet was sampled from each of four plants per parent and NIL at Zeitgeber time 3 in two different growing stages, the 2nd and 3rd leaf stages. The sampled leaves were bulked, immediately frozen in liquid N_2, and stored at −80°C. Total RNA was isolated from frozen tissues with TRIzol Reagent (Thermo Fisher Scientific, Japan). DNase I (TaKaRa, Japan) was used to remove genomic DNA. The complementary DNAs (cDNAs) were synthesized from 0.8 μg of total RNA by using the M-MLV reverse transcriptase system (Invitrogen, Japan) with an oligo (dT) 20 primer in a volume of 20 μL.

3.2.3 Expression analysis

Transcript levels of *E1*, *E1La*, *E1Lb*, *FT2a*, and *FT5a* were determined by quantitative real-time PCR. The PCR mixture (20 μL) contained 0.1 μL of the cDNA synthesis reaction, 5 μL of 1.2 μM primer premix, and 10 μL SYBR Premix Ex Taq II (TaKaRa, Japan). A CFX96 Real-Time System (Bio-Rad, Japan) was used. The PCR cycling conditions were 95°C for 3 min followed by 40 cycles of 95°C for 10 s, 59°C for 30 s, 72°C for 20 s, and 78°C for 2 s. Fluorescence was quantified before and after the incubation at 78°C to monitor the formation of primer dimers. The mRNA for *β-tubulin*

(Glyma.08G014200) was used for normalization. The expression levels were expressed as the tratios to β-tubulin transcript levels. A reaction mixture without reverse transcriptase was also used as a control to confirm the absence of genomic DNA contamination. Amplification of a single DNA fragment was confirmed by melting curve analysis and gel electrophoresis of the PCR products. Averages and standard errors of relative expression levels were calculated from PCR results for three independently synthesized cDNAs. Primer sequences used in expression analyses are listed in Supplemental Table 2.

3.3 Results and Discussion

3.3.1 Comparison of flowering time and gene expression among NILs

The allelic effects of *E1Lb* and *e1lb* on flowering under the R-enriched LD condition (daylength, 20 h) were evaluated in four sets of NILs, each for *E1Lb* and *e1lb*, developed from different F_2 plants from the H-*e3* × ZE cross (#4 and #21; *e3*/*E4*) and the HA × ZE cross (#11 and #20; *E3*/*E4*). In the two sets of the *e3*/*E4* NILs, each NIL for *e1lb* flowered at the same or almost the same time (#4, 31.7 DAS; #21, 30.3 DAS) as ZE (30.3 DAS); this was on average 6.7 to 7.6 days earlier than the respective NILs for *E1Lb*, which flowered at almost the same time as H-*e3* (Figure 7A). Flowering times of the *E3*/*E4* NILs were around 20 days or more later than those of the *e3*/*E4* NILs. *e1lb* also promoted flowering in the *E3*/*E4* background: each NIL for *e1lb* flowered around 10 days earlier than the respective NIL for *E1Lb* and HA (Figure 7A). This flowering-promoting effect of *e1lb* versus *E1Lb* under the R-enriched LD condition was smaller than that of *e4* vs. *E4* and that of *e3* vs. *E3*, because H-*e3e4* and H-*e3* flowered, on average, 13 and 25 days earlier than H-*e3* and HA (*E3E4*), respectively.

The effect of *e1lb* vs. *E1Lb* on flowering under FR-enriched ILD condition (Figure 7B) was also evaluated. *e1lb* induced flowering at 58 DAS (NILs #4) or 49 DAS (NILs #21) in the *e3*/*E4* genetic background and at 56 DAS (NILs #11 and #20) in the *E3*/*E4* genetic background. All these NILs produced pods of up to 3 cm in length at the end of light supplementation, similar to those of ZE and H-*e3e4*. In contrast, the *e3*/*E4* NILs for *E1Lb* and H-*e3* flowered around 20 days later, and *E3*/*E4* NILs for *E1Lb* and HA continued vegetative growth and did not produce any flower buds until the end of light supplementation. Therefore, *e1lb* was sufficient to induce flowering under the ILD condition, irrespective of the *E3* genotype (Figure 7B). Interestingly, a similar flowering-

promoting effect was observed in the NIL-*E1* for a loss-of-function allele *e1-nl* (e1); it initiated flowering under the ILD condition, as the *E3*/*E4* NILs for *e1lb*, whereas the NIL for *e1-as* (E1) did not (Figure 7B).

The expression levels of *E1*, two *E1L* genes, and two *FT* orthologues were tested in two different growing stages (the 2nd and 3rd leaf stages) in the *e3*/*E4* NILs grown under the R-enriched LD condition (Figure 8). The expression levels of *E1* and *E1La* were similar between the NILs for *E1lb* and *e1lb* at both stages in NILs #4 or at the 3rd stage in NILs #21; both *E1* and *E1La* were significantly up-regulated in the 2nd leaf stage in NILs #21 for *e1lb* relative to those for *E1Lb*. On the other hand, the expression of *E1Lb* was significantly down-regulated in the NILs for *e1lb* at both stages (#4) or at the 3rd leaf stage (#21). The extremely low expressions of *E1Lb* gene in NILs for *e1lb* were considered to be due to nonsense-mediated mRNA decay or different promoter activity. In contrast, the expressions of both *FT2a* and *FT5a* were up-regulated at both stages in the NILs for *e1lb* relative to those for *E1Lb* in both NIL sets. The similar effect of *e1lb* vs. *E1Lb* on the expression of *FT2a* and *FT5a* was also observed at the 3rd leaf stage in both sets of *E3*/*E4* NILs (#11 and #20; Figure 9). As observed in the *e3*/*E4* NILs for *e1lb*, the expression levels of *FT2a* and *FT5a* were significantly upregulated in the *E3*/*E4* NIL for *e1lb*.

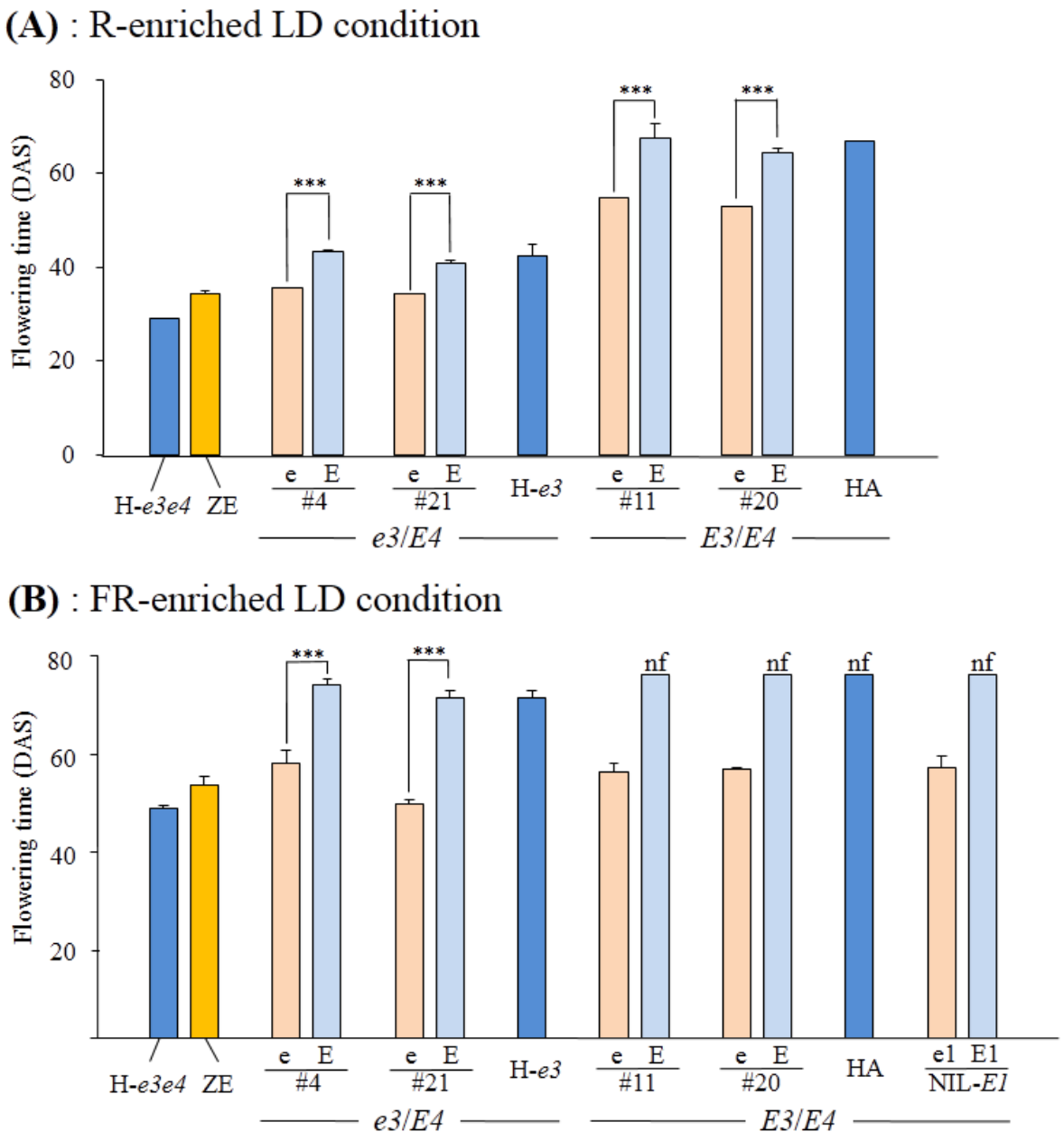


Figure 7. Flowering time in NILs for the *e1lb* (e) and *E1Lb* (E) alleles under R-enriched and FR-enriched LD conditions. Two sets of NILs (#4 and #21) have the *e3*/*E4* genotype, similar to H-*e3*, whereas the other two (#11 and #20) have the *E3*/*E4* genotype, similar to HA. A set of *E3*/*E4* NILs for *e1-nl* (e1) and *e1-as* (E1) at *E1* locus (NIL-*E1*) were also evaluated for the flowering under the FR-enriched LD condition. Plants were grown under 20-h **(A)** R-enriched LD or **(B)** FR-enriched ILD conditions. Data are presented as mean and standard error (n = 5). nf, no flower buds by the end of light supplementation, DAS, days after sowing, ***$p < 0.001$.

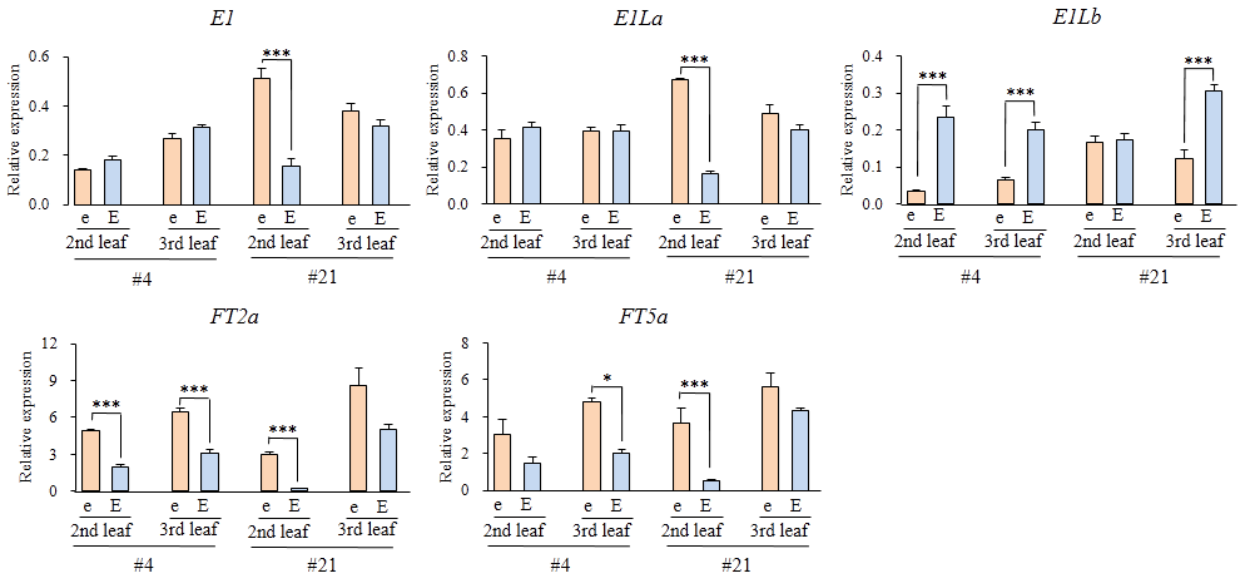


Figure 8. Expression levels of *E1*, *E1-Like*, and *FT* genes at the second and third leaf stages in two sets of *e3*/*E4* NILs for the *e1lb* (e) and *E1Lb* (E) alleles under R-enriched LD (20 h) conditions. Relative mRNA levels (mean and standard error, n = 3) are expressed as the ratios to *β-tubulin* transcript levels. **p* < 0.05, ****p* < 0.005.

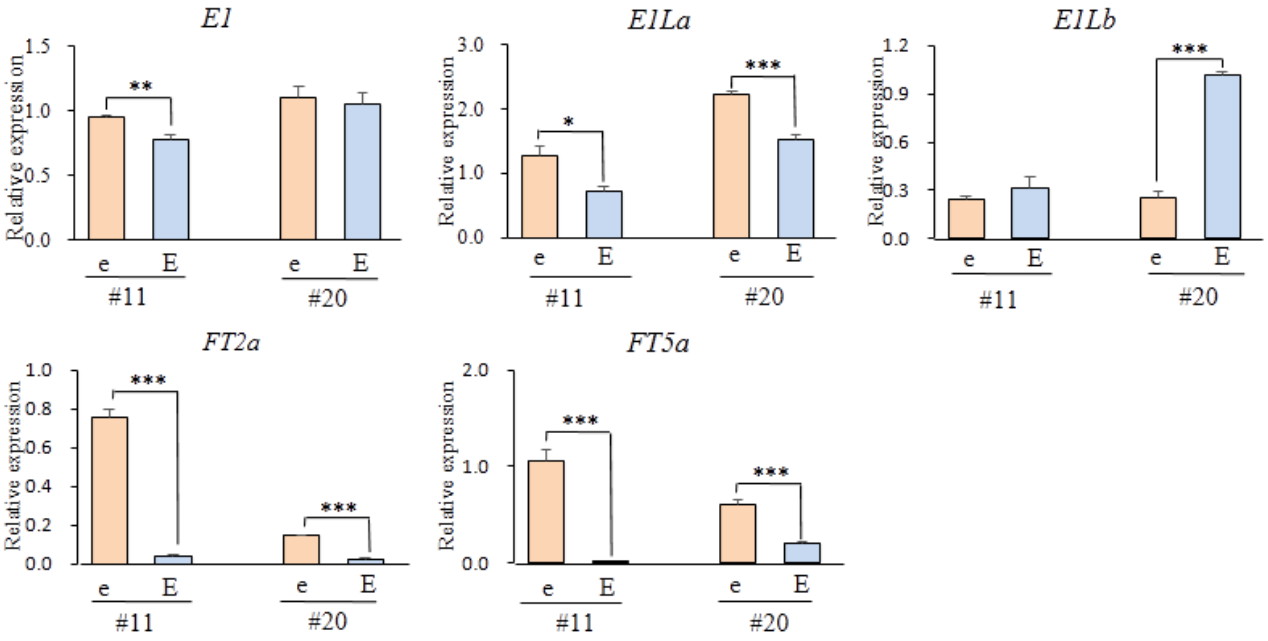


Figure 9. Expression levels of *E1*, *E1-Like*, and *FT* genes at the third leaf stage in two sets of *E3*/*E4* NILs for the *e1lb* (e) and *E1Lb* (E) alleles under R-enriched LD (20 h) conditions. Relative mRNA levels (mean and standard error, n = 3) are expressed as the ratios to *β-tubulin* transcript levels. **p* < 0.05, ***p* < 0.01, ****p* < 0.005.

Chapter 4 Interaction between the E1 family proteins

4.1 Background and Purposes

The results obtained in the previous chapters strongly suggested that *E1Lb* represses the *FT* expressions and as a consequence controls flowering, as *E1* does. However, it still remains undetermined how *E1Lb* and *E1* regulate the *FT* expressions. It is thus meaningful to check if the E1Lb and E1 proteins physically interact with each other. In this chapter, yeast two-hybrid assays were carried out to determine the interactions between E1 family proteins.

4.2 Materials and Methods

4.2.1 Yeast two-hybrid assays

Yeast two-hybrid (Y2H) assays were performed according to the method described by Nan et al. (2014). The yeast cloning vectors pGBKT7 and pGADT7, the control vectors pGADT7-T and pGBKT7-53, and the yeast strain Y2HGold used in the yeast two-hybrid assays were obtained from Matchmaker™ Gold Yeast Two-Hybrid System (TaKaRa, Japan). The full-length coding sequences of *E1*, *E1La* and *E1Lb* were inserted into pGBKT7 vectors to generate fused GAL4 DNA-binding domains (BD) as baits. These sequences were also cloned into pGADT7 to generate fused GAL4 DNA activation domains (AD) as preys. The bait and prey plasmids were cotransformed into the yeast strain Y2H using the lithium acetate method. The cotransformed yeasts were selected on SD medium lacking leucine (Leu) and tryptophan (Trp) (SD/–Leu/–Trp). After 3 days of

incubation at 30°C, the yeast cells were replated on selection plates with SD medium lacking Leu, Trp, histidine (His) and adenine (Ade) but including the X-α-gal substrate (SD/–Ade/–His/–Leu/–Trp/+X-α-gal) for the interaction test. The primers and restriction sites used to generate the yeast bait and prey constructs are presented in Supplemental Table 5.

4.3 Results and Discussion

4.3.1 Yeast two-hybrid assays

Tocixity of AD vectors in the yeast strain Y187 on SD/-Leu mdedium (Figure 10) and BD vectors in the yeast strain Y2HGold on SD/-Trp mdedium (Figure 11) were tested. The BD-E1 on SD/-Trp medium did not grow any clony suggested that the *E1* gene inserted in the pGBKT7 vector may have toxicity in the yeast strain Y2HGold. So the combinations of BD-E1with other vectors (AD-E1/BD-E1, AD-E1La/BD-E1 and AD-E1Lb/BD-E1) were not performed in following experiments. The results of Y2H assays were shown in Figure 12. A total of six combinations between the E1 family proteins were tested. The transformed yeasts alive in the SD/–Leu/–Trp medium (Figure 12A) were then replated on the selection medium (SD/–Ade/–His/–Leu/–Trp/+X-*α*-gal). All of the yeasts except positive control died in the selection medium (Figure 12B), indicating that the three E1 family proteins did not interact physically with one another and themselves. These results suggest that the *E1* family genes control the *FT2a* and *FT5a* expressions independently of one another.

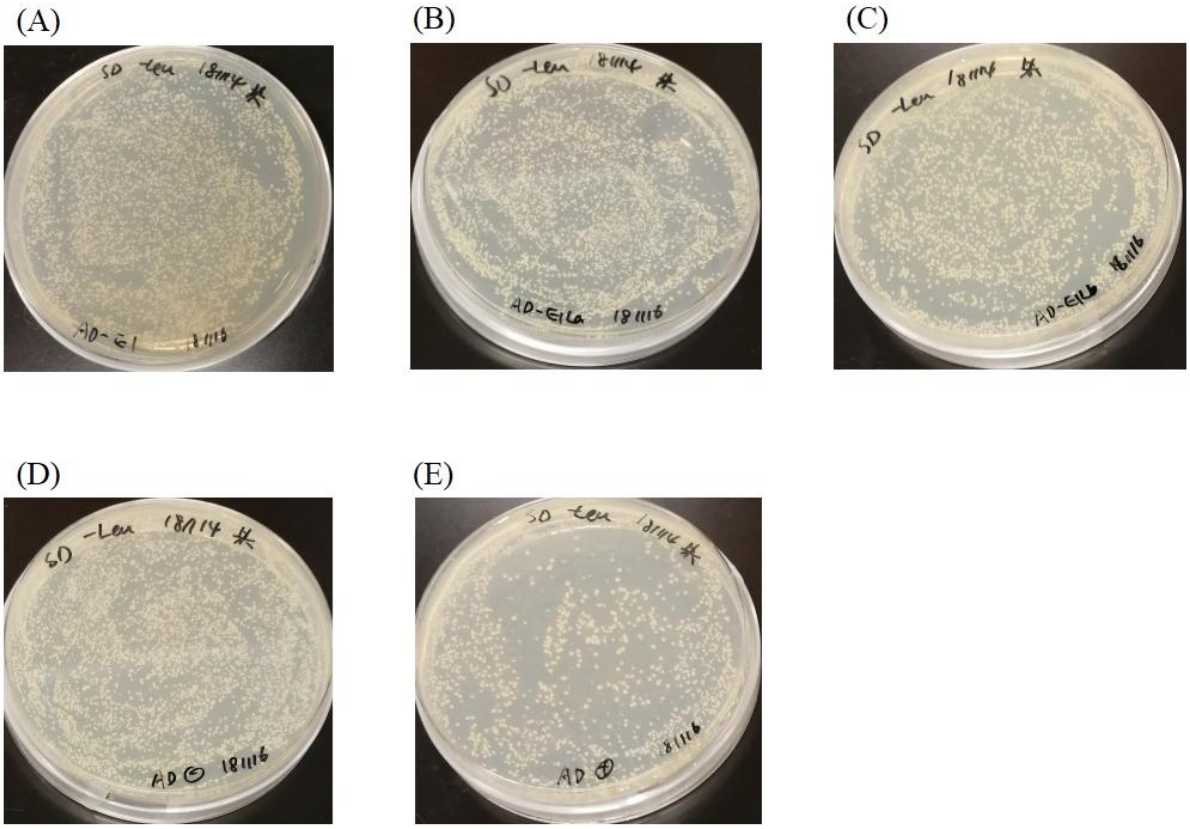


Figure 10. Toxicity test of AD vectors in the yeast strain Y187 on SD/-Leu mdedium. **(A)** E1; **(B)** E1La; **(C)** E1Lb; **(D)** negetive control, empty pGADT7 vector; **(E)** positive control, pGADT7-T. All five AD vectors can grow on SD/-Leu medium.

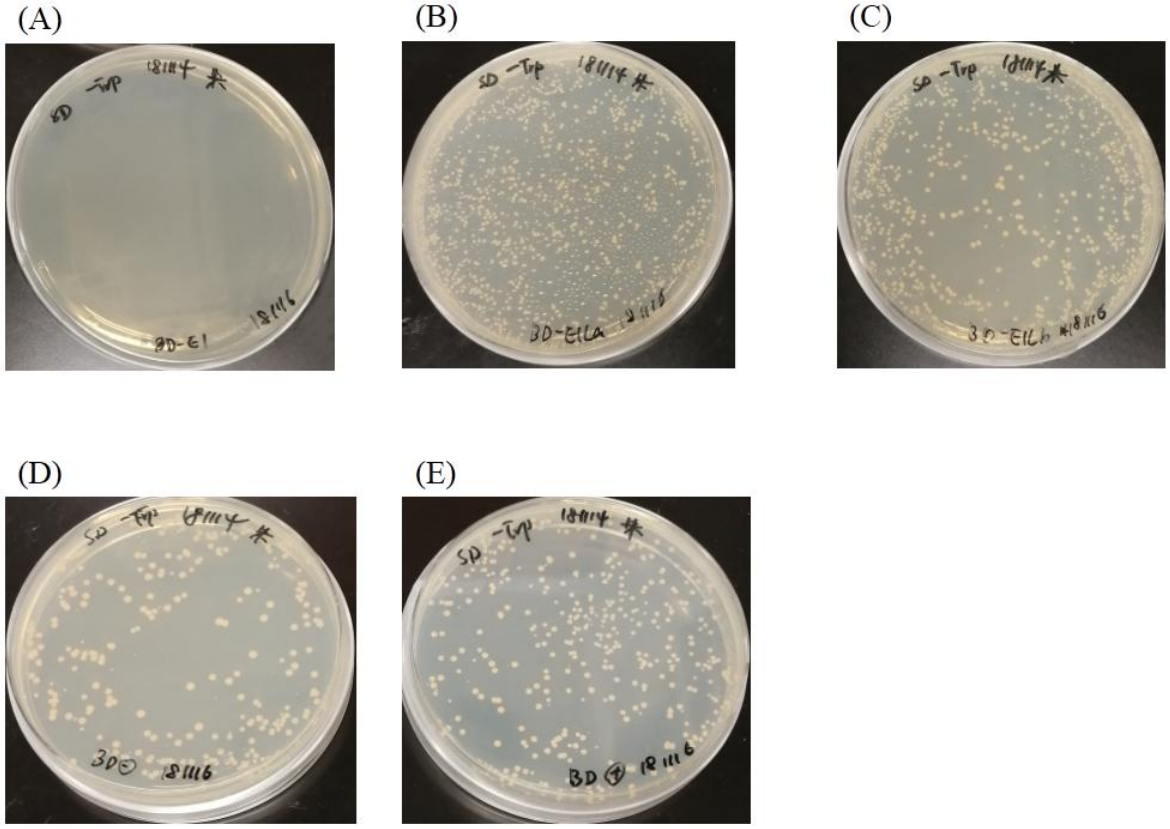


Figure 11. Toxicity test of BD vectors in the yeast strain Y2HGold on SD/-Trp medium. **(A)** E1; **(B)** E1La; **(C)** E1Lb; **(D)** negetive control, empty pGBKT7 vector; **(E)** positive control, pGBKT7-53. BD-E1 showed toxicity in the yeast strain Y2HGold, all other four vectors can grow on SD/-Trp medium.

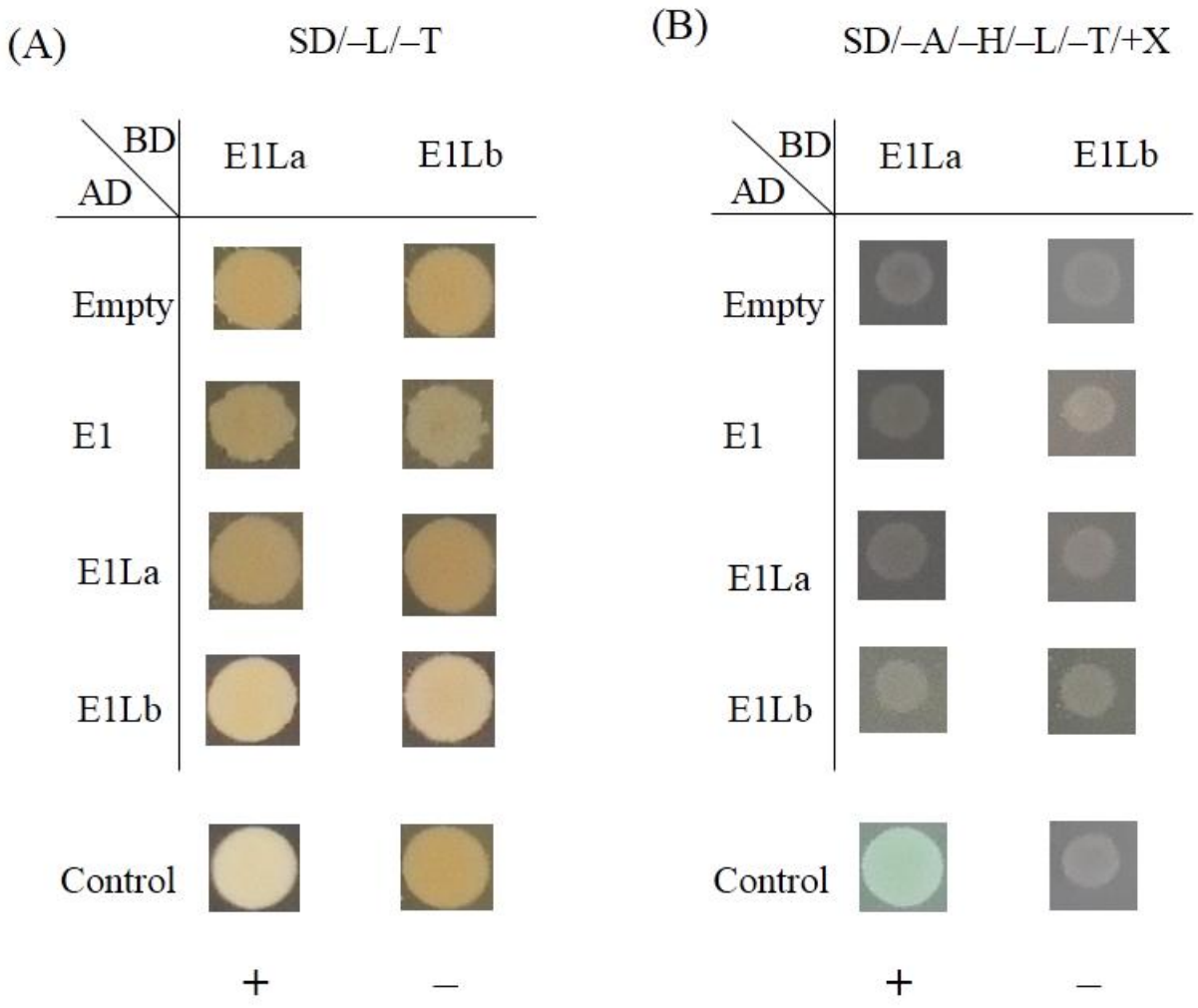


Figure 12. Yeast two-hybrid assays of the E1 family proteins**.** After cotransformation of the baits (BD) and preys (AD), an equal amount of yeast clones was plated on **(A)** SD/–Leu/–Trp and **(B)** SD/–Ade/–His/–Leu/–Trp/+X-α-gal selective plates. The plates were incubated at 30˚C until the emergence of the yeast clonies. AD, GAL4 DNA activation domain, prey; BD, GAL4 DNA-binding domain; Control +, positive control, in which the yeasts were cotransformed with pGADT7-T and pGBKT7-53 vectors; Control –, negative control, in which the yeasts were cotransformed with the empty vectors of pGBKT7 and pGADT7. **(A)** All six combinations together with positive control and negative control could grow on SD/–Leu/–Trp medium. **(B)** Only positive control could grow on SD/–Ade/–His/–Leu/–Trp/+X-α-gal medium and showed blue; all six combinations and negative control could not grow on SD/–Ade/–His/–Leu/–Trp/+X-α-gal medium.

Chapter 5 Molecular diversity of *E1Lb* and its homologue *E1La*

5.1 Background and Purposes

Four maturity genes, *E1* to *E4*, are major *FT* repressors in soybean (Liu et al., 2011; Jia et al., 2014; Jiang et al., 2014; Langewisch et al., 2014; Tsubokura et al., 2014; Zhai et al., 2014; Lu et al., 2015; Kurasch et al., 2017; Li et al., 2017). Many complete or partial loss-of-function natural variants at these loci have been reported so far (Xia et al., 2012; Tsubokura et al., 2013, 2014; Xu et al., 2014, Zhang et al., 2016; reviewed by Cao et al., 2016). Allelic variation in the *E1* gene includes a leaky allele *e1-as* in which a SNP at 44th nucleotide results in a nonsynonymous substitution in nuclear localization signal, and null alleles such as *e1-fs* (frame-shift allele), *e1-nl* and *e1-b3a* (deletion alleles), *and e1-re* (retroelement-inserted allele) (Xia et al., 2012; Tsubokura et al., 2014; Zhang et al., 2016). In the *E2* gene, there is a nonsense allele *e2-ns* (Watanabe et al., 2011). In the *E3* gene, there are three null alleles, deletion, flame-shift and nonsense alleles (*e3-tr*, *e3-fs* and *e3-ns)* and a missense allele *e3-mo* (Watanabe et al., 2009; Xu et al., 2014). The *E4* gene has five loss-of-function alleles, four single-base-deletion alleles (*e4-tsu*, *e4-oto*, *e4-kam*, and *e4-kes*), and a retrotransposon-inserted allele *e4-SORE-1* (Liu et al., 2008; Tsubokura et al., 2013). These results suggest that the variation in flowering time among soybean cultivars has been created by classification of multiple loss-of-function alleles at a number of key flowering loci. In this chapter, I surveyed the sequence variations for *E1Lb* and its homologue *E1La* in cultivated and wild soybean accessions by the dCAPS marker developed in Chapter 2 and direct PCR sequencing analyses.

5.2 Materials and Methods

5.2.1 Survey of the *e1lb* allele by the allele-specific DNA marker

A total of 59 ILD-insensitive accessions were surveyed for the *E1Lb* genotype using the allele-specific DNA marker, according to the method described in Chapter 2. They included 9 accessions from northern Japan, 26 from North-Eastern China, 13 from Far Eastern Russia, 8 from Ukraine, and 3 from Poland (Supplemental Table 6). The maturity genotypes at *E1* to *E4* of 47 accessions were determined previously by Xu et al. (2013), and those of the remaining 12 accessions were assayed according to Xu et al. (2013) and Tsubokura et al. (2014).

5.2.2 Sequence analysis for *E1La* and *E1Lb*

Fourteen ILD-insensitive accessions, together with H-*e3* and Zeika, and eight wild soybean (*Glycine soja*) accessions introduced from the Far Eastern Russia were surveyed for sequence variation of *E1La* and *E1Lb*. The accessions tested are listed in Tables 3 and 4. The sequence was determined by direct PCR sequencing analysis. Each PCR mixture contained 30 ng of total genomic DNA as a template, 0.5 μL of each primer (10 μM), 1.6 μL of dNTPs (2.5 mM), 0.1 μL of Ex Taq DNA polymerase (TaKaRa, Japan), and 2 μL of 10× Ex Taq buffer (TaKaRa, Japan) in a total volume of 20 μL; amplification conditions were 35 cycles at 94°C for 30 s, 62°C for 30 s, and 72°C for 60 s. 5 μL PCR products were then treated with 2μL ExoSAP-IT (Thermo Fisher Scientific, Japan) to remove excess primers and dNTPs, following the conditions 37°C for 15 min, 80°C for 15 min. The treated products were then diluted 5 times with distilled water and used for direct

sequencing analysis. The sequencing analysis were performed with a BigDye Terminator v3.1 Cycle Sequencing kit and an ABI PRISM 3100 Avant Genetic Analyzer (both from Applied Biosystems, Japan) according to the manufacturer's instructions. The primers for PCR amplification of *E1La* and *E1Lb* are presented in Supplemental Table 4.

5.2.3 Analysis of plant *cis*-acting regulatory DNA elements

New PLACE program (Higo et al., 1999; http://www.dna.affrc.go.jp/htdocs/PLACE/) was used to determine whether the nucleotide polymorphisms among accessions created or destroyed known *cis*-elements.

5.2.4 Putative 3D protein structure analysis

The putative 3D protein structures of E1La and its amino-acid substitution variant detected in the wild soybean were constructed with I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The structure analysis was performed with UCSF Chimera 1.13.1 soft (http://www.rbvi.ucsf.edu/chimera).

5.3 Results and Discussion

5.3.1 Survey of the *e1lb* allele in ILD-insensitive soybean accessions

To determine whether or not the deletion in the *E1Lb* gene is region-specific, I surveyed the polymorphism in the ILD-insensitive soybean accessions by using the dCAPS marker. These accessions were introduced from Northern Japan, North-Eastern China, Far-Eastern Russia, Ukraine and Poland. In addition to the three Russian cultivars used in the genetic analysis, another two Russian cultivars, Salyut 216 and DYA-1, were found to have the *e1lb* allele, whereas all the other accessions had the functional *E1Lb* allele (Supplemental Table 6). All of the Russian cultivars with *e1lb* possessed the maturity genotype of *e1-as*/*e3*/*E4* (Supplemental Table 6). There was no cultivar which had loss-of-function alleles at both *E1* and *E1Lb* loci. I also examined the *E1Lb* genotype for 7 wild accessions introduced from the Far Eastern Russia (Table 4), because the Russian cultivar Zeika was developed from the crossing between cultivated and wild soybeans. I expected that the *e1lb* allele had been introduced from the wild soybean in Russia. However, there was no wild accession with the *e1lb* allele.

5.3.2 Sequence polymorphisms in *E1La* and *E1Lb*

As mentioned above, all of the wild accessions did not possess the single-nucleotide deletion in *E1Lb*, unlike in ZE. To confirm if these wild accessions possessed any other polymorphism in *E1Lb*, their sequences were determined. All of the 7 accessions tested showed a new single nucleotide polymorphism (SNP) in 3'-UTR; it changed the 530th nucleotide from the adenine of start codon from guanine (G) to thymine (T) (Figure 13). I designated the *E1Lb* sequence of Williams 82 reference sequence (*Glycine max*

Wm82.a2.v1) as haplotype 1 (Hap 1), *e1lb* as haplotype 2 (Hap 2), and the sequence variant observed in the wild accessions as haplotype 3 (Hap 3). As a result of a survey of *cis*-elements with New PLACE (Higo et al., 1999), the SNP in the wild soybean destroyed two *cis*-elements, CACTFTPPCA1 with the recognition site 'YACT' (Y = T/C), and CANBNNAPA with the recognition site 'CNAACAC' (N = A/C/G/T), in the reverse strand of Haplotypes 1 and 2. Further studies such as those with yeast one-hybrid assay are needed to determine whether the SNP influences the regulation of *E1Lb* expression.

Fourteen ILD-insensitive accessions, together with H-*e3* and Zeika, and 8 wild soybean accessions were surveyed for sequence diversity of *E1La*. Besides the reference sequence of Williams 82 (designated haplotype 1; Hap 1), one haplotype (haplotype 2; Hap 2) in wild soybean accessions, two haplotypes (haplotype 3 and 4; Hap 3 and Hap 4) in cultivated accessions were detected (Figure 14). Compared with Hap 1, Hap 2 contained three SNPs, synonymous and non-synonymous nucleotide substitutions at the 222nd and the 244th nucleotide from the adenine of start codon in the exon, respectively, and a SNP in 3'-UTR at the 563rd nucleotide. The non-synonymous substitution converted the amino acid from lysine (K) to glutamic acid (E). Hap 3 contained the same SNPs as Hap 2 at the 222nd and the 563rd nucleotide. Hap 4 contained the SNP at the 563rd nucleotide as Haplotypes 2 and 3 did. The Russian wild accessions possessed two haplotypes, Hap 1 and Hap 2 (Table 4).

I also sequenced the coding region of *E1* gene for the 8 wild soybean accessions. All of the 8 accessions showed the same sequence as the functional *E1* allele (Table 4). Accordingly, the genotypes at *E1* family genes for the Russian wild soybean accessions were *E1*, *E1La*-Hap 1 or *E1La*-Hap 2, and *E1Lb*-Hap 3. The wild accessions, B00148 (*E1La*-Hap2) and B00161 (*E1La*-Hap1), flowered without any marked delay in the LD

of 16h, compared with the photoperiod-insensitive soybean cultivars, such as ZE and Harosoy NIL for *e3* and *e4* (Figure 14). Since the two accessions may possess the functional *E1* and *E1Lb* genes (Table 4), an unknown genetic factor(s) may be responsible for the photoperiod-insensitivity for these wild soybeans.

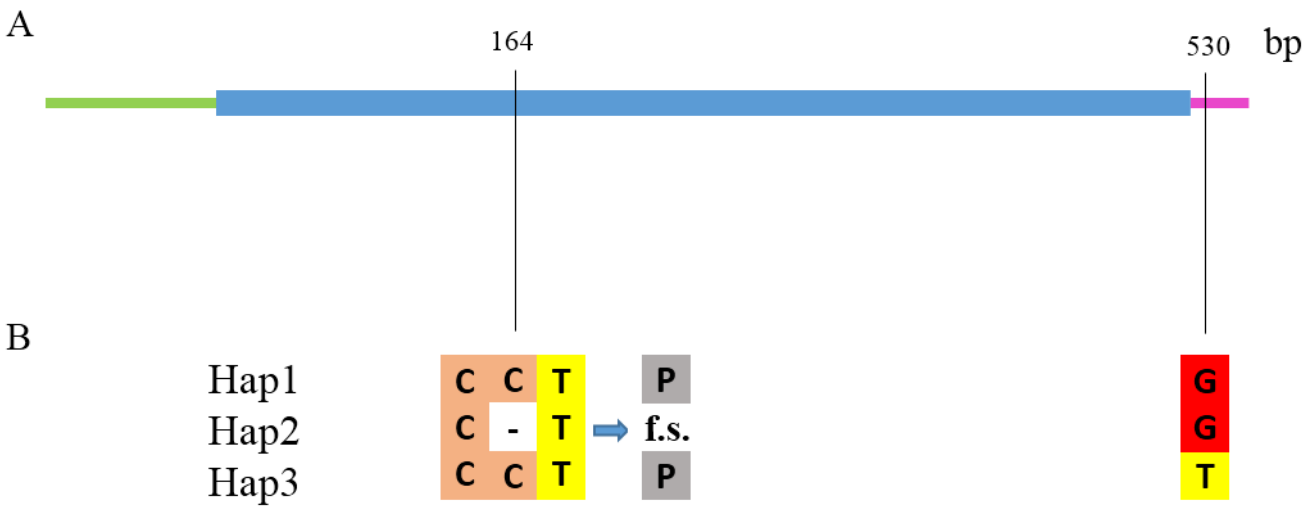


Figure 13. Sequence polymorphism of *E1Lb* gene. **(A)** Gene structure of *E1Lb*; green bar, 5'-UTR; blue bar, coding region; purple bar, 3'-UTR. **(B)** Haplotypes of *E1Lb*. Hap 1, Haplotype 1; Hap 2, Haplotype 2; Hap 3, Haplotype 3; P, proline; f.s., frame shift.

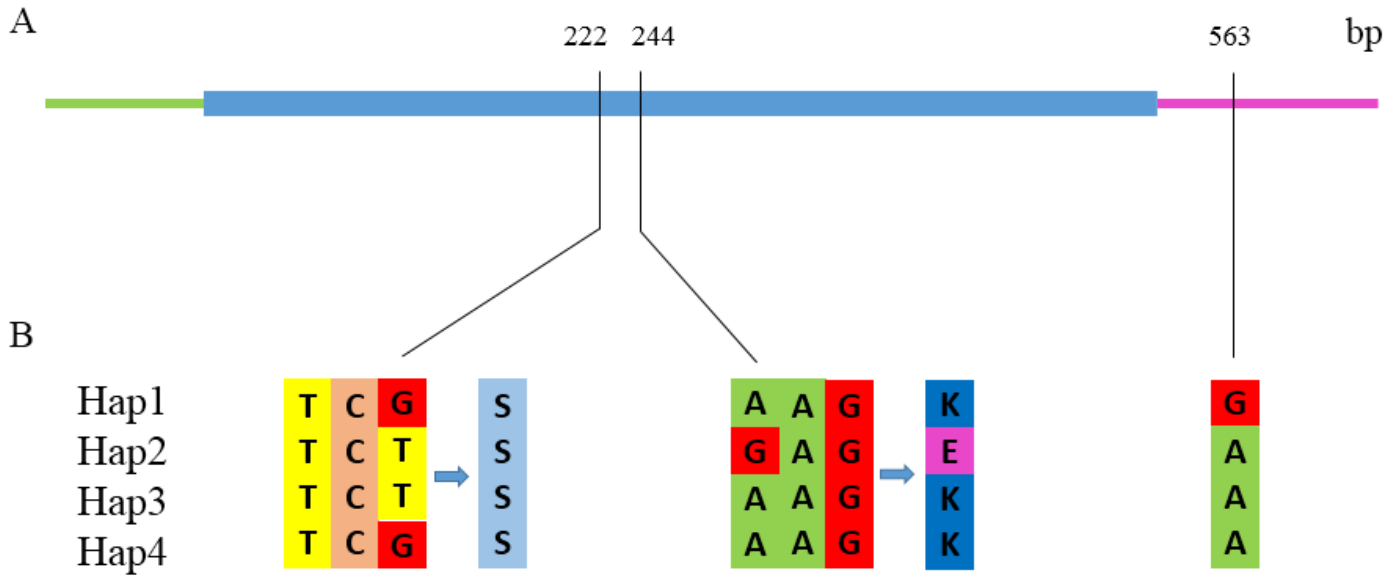


Figure 14. Sequence polymorphism of *E1La* gene. **(A)** Gene structure of *E1La*; green bar, 5'-UTR; blue bar, coding region; purple bar, 3'-UTR. **(B)** Haplotypes of *E1La*. Hap 1, Haplotype 1; Hap 2, Haplotype 2; Hap 3, Haplotype 3; Hap 4, Haplotype 4; S, serine; K, Lysine; E, glutamic acid.

Table 3. Haplotypes of the *E1* family genes of cultivated soybean accessions used for sequencing analysis

Cultivated accession	Collection/cultivation area	*E1*	*E1La*	*E1Lb*
Harosoy	Canada	*e1-as*	Hap 1	Hap 1
Ohyachi 2	Hokkaido, Japan	*E1*	Hap 1	Hap 1
Sakamoto wase	Hokkaido, Japan	*e1-fs*	Hap 3	Hap 1
Kamaishi 17	Tohoku, Japan	*E1*	Hap 1	Hap 1
Otome wase	Tohoku, Japan	*E1*	Hap 1	Hap 1
Gai	Poland	*e1-nl*	Hap 3	Hap 1
Nawiko	Poland	*e1-as*	Hap 1	Hap 1
Yug 30	Ukraine	*e1-as*	Hap 1	Hap 1
Ustya	Ukraine	*e1-nl*	Hap 1	Hap 1
Zeya 2	Russia	*e1-as*	Hap 1	Hap 1
DYA-1	Russia	*e1-as*	Hap 4	Hap 2 (*e1lb*)
Zeika	Russia	*e1-as*	Hap 1	Hap 2 (*e1lb*)
Yubileinaya	Russia	*e1-as*	Hap 4	Hap 2 (*e1lb*)
Keshuang	Northeast China	*E1*	Hap 1	Hap 1
Dongnong 50	Northeast China	*e1-as*	Hap 1	Hap 1
Heihe 33	Northeast China	*e1-as*	Hap 1	Hap 1

Genotype at the *E1* locus for the cultivated accessions are cited from Xu et al. (2013) and Kong et al. (2018).

Table 4. Haplotypes for *E1* family genes of wild soybean accessions used for sequencing analysis

Wild accession	Collection/cultivation area	*E1*	*E1La*	*E1Lb*
B00139	Ivanovka, Amrskaya, Russia (A2)	*E1*	Hap 2 (*e1la-as*)	Hap 3
B00143	Ivanovka, Amrskaya, Russia (A6)	*E1*	n.t.	n.t.
B00148	Sadovoe, Amrskaya, Russia (A11)	*E1*	Hap 2 (*e1la-as*)	Hap 3
B00155	Garovka, Khabarovsk,Russia (K2)	n.t.	Hap 2 (*e1la-as*)	n.t.
B00159	Anastafievka, Khabarovsk, Russia (K6)	*E1*	Hap 2 (*e1la-as*)	Hap 3
B00161	Elabug, Khabarovsk, Russia (K8)	*E1*	Hap 1	Hap 3
B00171	Vladivostok, Primorsky, Russia (P1)	*E1*	Hap 1	Hap 3
B00177	Novoseliische, Primorsky, Russia (P7)	*E1*	Hap 2 (*e1la-as*)	Hap 3
B00188	Erolovka, Primorsky, Russia (P18)	n.t.	Hap 1	n.t.
T106	Primorsky, Russia (*G. ussuriensis*)	*E1*	n.t.	Hap 3

n.t., not tested; *e1la-as*, putative missense allele by non-synonymous mutation

5.3.3 Putative 3D protein structures of wild and variant types of E1La protein

The nonsynonymous substitution at the 244th nucleotide of *E1La* converted the 82nd amino acid from lysine (lys; K) to glutamic acid (glu; E) (Figure 14). To confirm whether this substitution affects the protein structure of E1La, the putative 3D protein structures were constructed with the I-TASSER program (Zhang, 2008; Roy et al., 2010; Yang et al., 2015) and the UCSF Chimera 1.13.1 software (Pettersen et al., 2004). The structure of B3 domain of E1 resembled that of RAV1 (related to ABI3/VP1) (Zhang et al., 2016). The primary structure of the B3 domain of RAV1 possesses seven β-strands (β1–β7) and two short α-helices (α1 and α2); each of the two α-helices are present between β-strands 2 and 3 and between β-strands 5 and 6, respectively (Yamasaki et al., 2004). According to the protein structure estimated in this study, the B3-like DNA-binding domain of E1La possessed seven β-strands (β1, I58-T61; β2, L73-L75; β3, G101-W107; β4, M114-L119; β5, Y128-I131; β6, E149-D155; β7, V160-L166) as well, but did three α-helices (α1, D64-L67; α2, L79-Q83; α3, K133-R142.) positioning between β1 and β2, between β2 and β3, and between β5 and β6, respectively. The K to E substitution was located in the second α-helix (α-helix 2) (Figure 15). The amino acid sequence of α-helix 2 (L79–Q83) is highly conserved in legume species; three soybean E1 family proteins, and one common bean (*Phaseolus vulgaris*), one medicago (*Medicago truncatula*), one *Lotus japonicus*, two pigeon pea (*Cajanus cajan*), and one chickpea (*Cicer arietinum*) E1-like proteins shared an identical amino acid sequence of LVKKQ (Zhang et al., 2016). The structure analysis for the wild type and variant type E1La proteins indicated that the nonsynonymous substitution changed the pattern of hydrogen bonds between neighboring amino acids (Figure 16). In the wild type E1La protein, the amino group of K82 residue

formed hydrogen bonds with the carboxylic acid group of Q83 residue, whereas the amino group of K81 residue formed hydrogen bonds with the carboxylic acid group of D78 residue (Figure 16A). In contrast, in the variant E1La protein, the carboxylic acid group of E82 residue formed hydrogen bonds with the amino group of E81 residue, in place of the hydrogen bonds between K81 and D78 (Figure 16B). Therefore, the change of hydrogen bonding pattern caused by the K to E substitution was predicted to result in the structure change of α-helix 2 in the variant protein. Additionally, lysine is basic amino acid while glutamic acid is acidic. This different electrical charge could also affect the stability of α-helix 2. These results strongly suggested that the amino acid substitution detected in E1La Hap2 produced an adverse effect on the DNA-binding ability of E1La to the *cis*-elements in *FT2a* and *FT5a*. Here, I tentatively designated the Hap 2 sequence as a recessive dysfunctional allele *e1la-as*. A further study is needed to determine the effect of *e1la-as* on flowering in segregating populations.

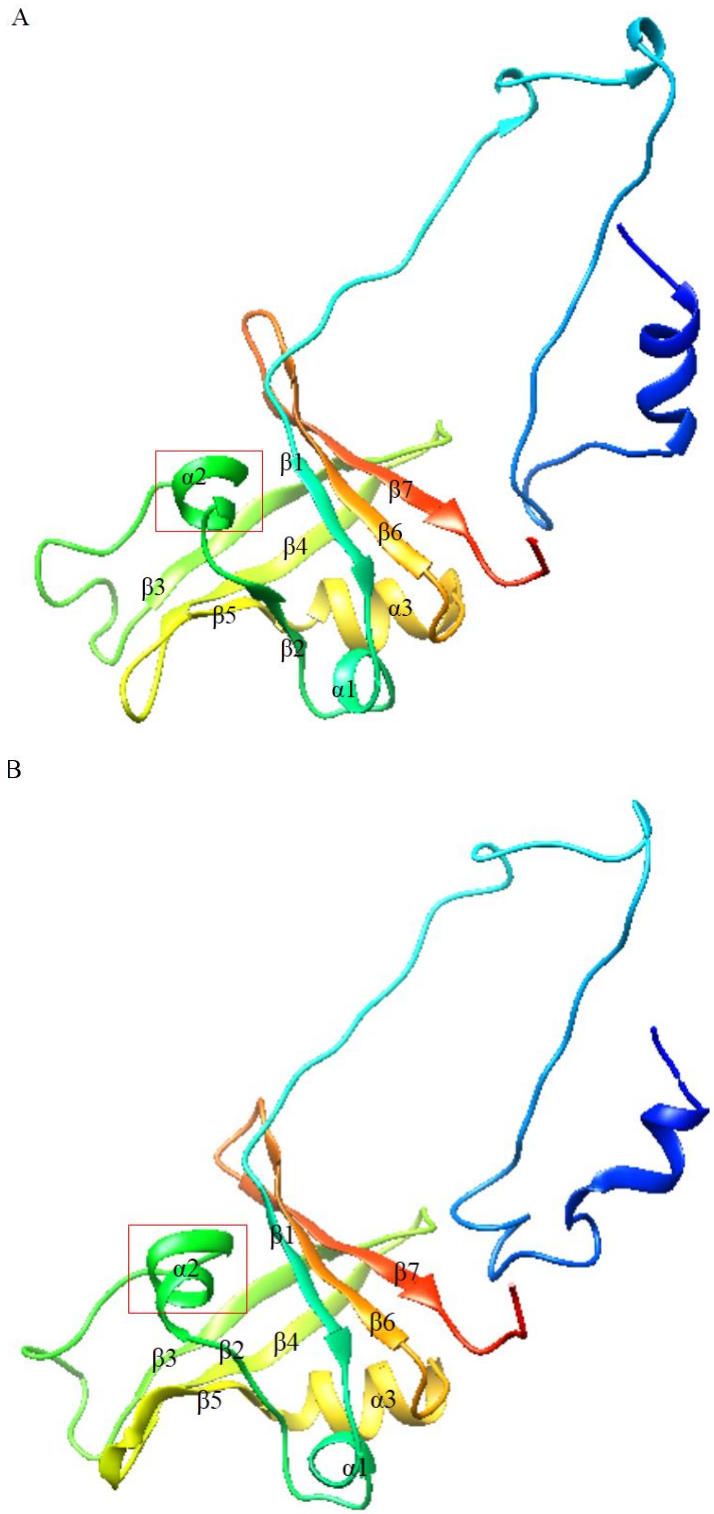


Figure 15. Putative 3D protein structures of wild and variant types E1La proteins. **(A)** Wild type E1La; **(B)** Variant type E1La. Red boxes show the second α-helices where the amino acid substitution occurred.

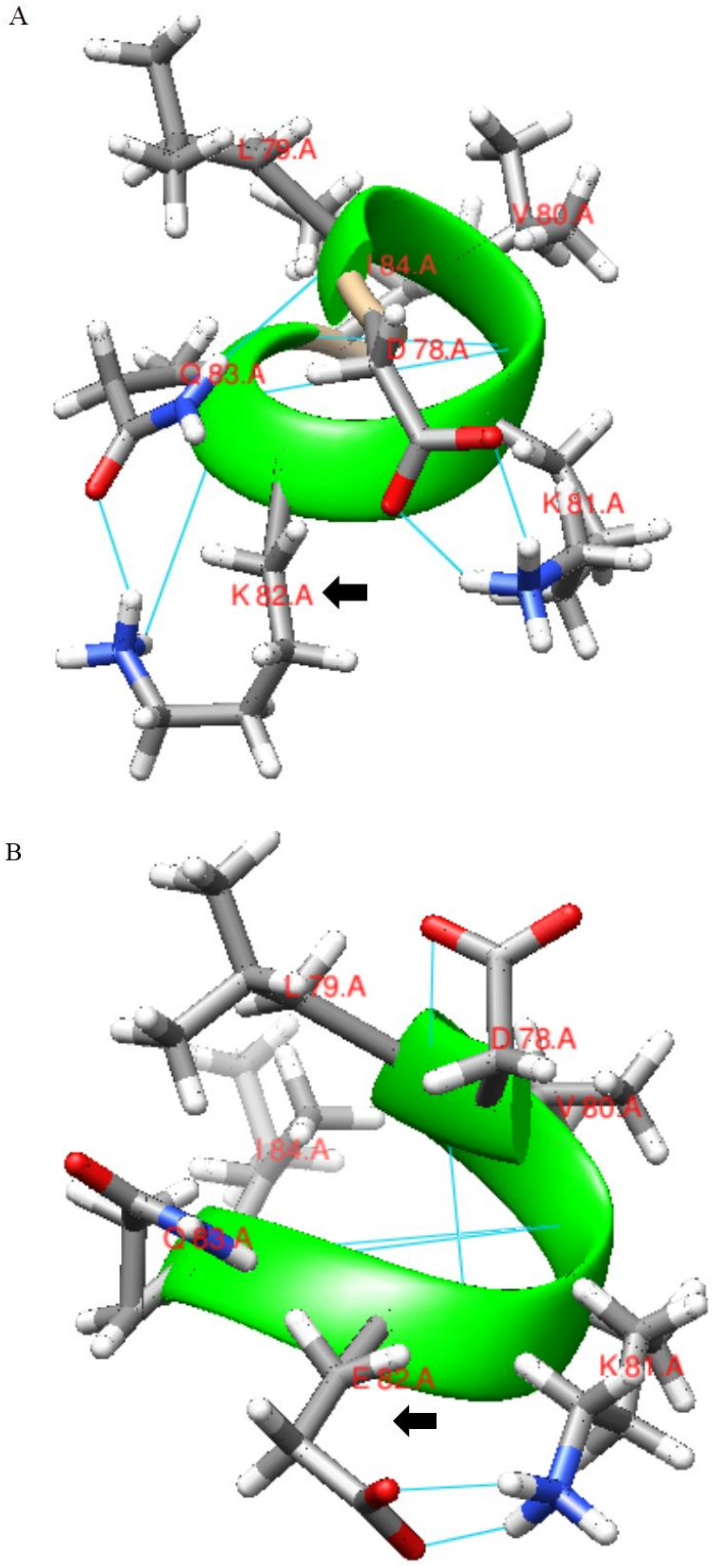


Figure 16. Details of the second helices of wild and variant types E1La proteins. **(A)** Wild type E1La; **(B)** Variant type E1La. Black arrows showed the amino acid which substitutes from lysine (K) in wild type E1La to glutamic acid (E) in the E1La variant. Light blue lines, hydrogen bonds between amino acids.

Chapter 6 General Discussion

The soybean maturity loci, *E1* to *E4*, are major flowering loci that determine the ability of cultivars to adapt to different latitudes. Diverse allelic combinations at the *E1*, *E3*, and *E4* loci, each of which has multiple loss-of-function alleles (Tsubokura et al., 2013, 2014; Xu et al., 2013; Jiang et al., 2014; reviewed by Cao et al., 2017), have resulted in cultivars with various degrees of sensitivity to photoperiod. Photoperiod-insensitivity is an adaptive trait for cultivars at high latitudes; such cultivars are classified into three genotypic groups according to the allelic functions at each of the three loci (Xu et al., 2013). Among the ILD-insensitive accessions tested, the predominant group has the loss-of-function alleles of the *phyA* genes *E3* and *E4* (*e3*/*e4*) (Figure 17B), followed by a group which has the loss-of-function of the *E1* repressor for *FT2a* and *FT5a* in combination with a dominant *E3* or *E4* allele (Figure 17C) (Xu et al., 2014). Cultivars of the third group, which were introduced from North-Eastern China and Far-Eastern Russia, have a novel genetic mechanism(s) that abolishes or reduces the sensitivity to daylength, because they have the same genotype at the three loci (*e1-as*/*e3*/*E4*) as an Haroosy NIL for *e3*, which is sensitive to FR-enriched ILD conditions (Saindon et al., 1989; Cober et al., 1996; Abe et al., 2003; Liu and Abe, 2010; Xu et al., 2013). Takeshima et al. (2016) carried out QTL analysis of ILD-insensitivity by a testcross of a Chinese cultivar of group 3 with the Harosoy NIL for *e3* and demonstrated that an early-flowering allele (*FT5a-ef*) at *qDTF-J*, which encodes the FT5a protein, up-regulates *FT5a* expression by *cis*-activation to induce flowering under ILD conditions (Figure 17D). However, the Far-Eastern Russian cultivars of group 3 did not possess *FT5a-ef*, so the molecular-genetic basis for the photoperiod insensitivity of these cultivars remains undetermined. In this study, I tried to

identify a novel gene(s) that is responsible for the photoperiod-insensitivity in these Far-Eastern Russian cultivars.

A novel loss-of-function allele *e1lb* conditions the photoperiod-insensitivity

Based on the results obtained from the association analysis followed with fine-mapping and sequence analyses, I detected a novel loss-of-function allele that resulted from a frameshift mutation at the *E1Lb* locus in the Far-Eastern Russian photoperiod-insensitive cultivars. *E1Lb* and its tandemly linked homologue, *E1La*, have high sequence similarity to *E1* (Xia et al., 2012). *E1* has the most marked effect on time to flowering in the flowering genes of soybean (McBlain et al., 1987; Upadhyay et al., 1994; Tsubokura et al., 2014). The polymorphism of *E1* (or its flanking genomic region) largely accounts for the variation in flowering time and related agronomic traits in segregating populations of different genetic backgrounds (Yamanaka et al., 2000; Wang et al., 2004; Zhang et al., 2004; Funatsuki et al., 2005; Liu et al., 2007; Zhai et al., 2015). In contrast to the *E1* gene, only a few reports have demonstrated the presence of major genes or QTLs for flowering and maturing associated with the genomic region of chromosome 4 harboring *E1La* and *E1Lb* (Cober et al., 2010; Cheng et al., 2011; Watanabe et al., 2017; Kong et al., 2018). Down-regulation by VIGS revealed that the *E1L* genes inhibit flowering under LD and night-break conditions in the *e1-nl* background, suggesting that the *E1L* genes also have similar effects on flowering as *E1* does (Xu et al., 2015). However, the functions of each of *E1La* and *E1Lb* have remained undetermined. The results obtained in this study revealed that *E1Lb* also has an inhibitory effect on flowering through the repression of *FT2a* and *FT5a* expressions. In particular, the allelic effect of *e1lb* to *E1Lb* was almost

the same as that of *e1-nl* to *e1-as* under FR-enriched ILD condition; the *E3*/*E4* NILs for these loss-of-function alleles flowered similarly as a Harosoy NIL for *e3*/*e4*, whereas those for *E1Lb* and *e1-as* did not flower until the end of lighting treatment (early August) as Harosoy (*E3*/*E4*). Therefore, the loss-of-function allele *e1lb* has an ability to cancel the inhibitory effect of FR-enriched LD on flowering, which is mediated by *E4*, and singly confer the photoperiod-insensitivity, independently of the *E1* genotypes. The photoperiod-insensitivity in soybean is thus brought about by at least four genetic mechanisms, loss-of-function of *E1* or *E1Lb*, loss-of-function of *phyA* genes *E3* and *E4*, and elevated *FT5a* expression (Figure 17). Two North-Eastern Chinese cultivars, Heihe 34 and Jiagedaqi 20, also belong to the group 3 cultivar (Xu et al., 2014). The two cultivars have neither the *e1lb* allele nor the *FT5a-ef* allele. There are still some undetermined genetic variants left to lower or abolish the photoperiod-sensitivity.

Functions of *E1Lb* and *E1* in the control the *FT2a* and *FT5a* expressions

Comparison of NILs for *E1Lb* and *e1lb* further provides an interesting finding on the regulatory mechanism of the *E1* family genes by the *E3* and *E4* genes. As discussed above, the loss-of-function in *E1* and *E1Lb* can cancel the inhibitory effect on flowering under FR-enriched LD mediated by *E4*. In particular, the *e1lb* allele can promote flowering under the FR-enriched LD condition, irrespective of the *E3* genotype. The *e1lb* allele could also promote flowering under R-enriched LD conditions; the NILs for *e1lb* flowered earlier than those for *E1Lb* in both the *e3*/*E4* and *E3*/*E4* genetic background. However, the effect of *e1lb* on flowering was small under R-enriched LD; *e1lb* could not cancel flowering inhibition by *E3* as efficiently as *e3* did. Accordingly, the function of

E1Lb may depend on light quality of LD. *E1La* might be involved in the control of flowering under R-enriched LD conditions, although the loss-of-function for either of *E1* or *E1Lb* is sufficient for the photoperiod-insensitivity in FR-enriched LD conditions. It may be tempting in a further study to develop double recessive lines for the loss-of-function alleles at the *E1* and *E1Lb* loci not only to elucidate the interaction between the two genes and the function of another *E1* homologue, *E1La*, but also to explore the regulatory mechanisms of these *E1* family genes by E3 and E4 under different light conditions.

The flowering-promoting effects of *e1lb* involve the up-regulation of *FT2a* and *FT5a* expression. Intriguingly, their expression levels were not associated with those of *E1* and *E1La*; the expressions of the two genes were maintained at relatively high levels under the LD conditions, irrespective of the *E1Lb* genotypes. One likely explanation for this observation is that the total expression level and/or activity of *E1*, *E1La* and *E1lb* may be important for the repression of *FT2a* and *FT5a* expression. The yeast two-hybrid assays did not detect any interaction between the E1 family proteins. The E1 family proteins may therefore regulate the *FT2a* and *FT5a* expressions, independently of one another. A further study is needed to determine why the loss-of-function allele at *E1Lb* can singly upregulate the *FT2a* and *FT5a* expression under LD condition, even though the remaining *E1* family genes are expressed normally.

Molecular diversity of *E1Lb* and *E1La* and their use in soybean breeding

In Chapter 5, I surveyed the sequence variation for *E1Lb* and *E1La* in photoperiod-insensitive cultivars and wild soybeans introduced from Far-Eastern Russia. Genotyping with an allele-specific DNA marker revealed that *e1lb* is a rare and region-specific allele even in early maturing photoperiod-insensitive cultivars. Only Russian cultivars possessed *e1lb*; this allele was not detected in the other cultivated and wild soybean accessions used in the DNA marker and sequencing analysis. The *e1lb* allele most likely has neither largely contributed to the diversity of flowering behaviors nor been used widely in the current soybean breeding. The *e1lb* allele may thus be useful as a new resource to broaden the genetic variability of soybean cultivars for flowering under LD conditions at high latitudes.

I also discovered a nonsynonymous substitution in the B3-like DNA-binding domain of E1La in the wild soybean accessions. The E1 family proteins contain a putative bipartite nuclear localizing signal and a domain distantly related to the plant-specific B3-like DNA-binding domain, which shares 21–27% of amino acid sequence identity to those of other proteins (Xia et al., 2012). The B3 DNA-binding domain is exclusively found in higher plants and can be classified into four families, as the auxin response factor (ARF) family, the *LEAFY COTYLEDON2-ABI3-VAL* (LAV) family, the related to ABI3/VP1 (RAV) family, and the reproductive meristem (REM) family (reviewed by Swaminathan et al., 2008). The ARF family and the LAV family are reported to response to auxin and abscisic acid, respectively, while RAV and REM families are related to stress responses and vernalization, respectively. Some of the RAV family members have been reported to be related to flowering. Overexpression of *RAV1* in *Arabidopsis* delayed flowering (Hu

et al., 2004). Castillejo et al (2008) indicated that *TEMPRANILLO* genes (*TEM1* and *TEM2*) can repress *FT* by directly binding to its 5′-UTR in *Arabidopsis*. *VERNALIZATION1* (*VRN1*) is a member of REM family and functions in stable repression of the floral repressor *FLC* of the vernalization pathway in *Arabidopsis* (Levy et al. 2002). The B3-like DNA-binding domains of E1 family proteins may thus have a critical role to regulate *FT2a* and *FT5a* genes in soybean. The structure analysis revealed that the amino acid substitution in the B3-like DNA-binding domain changes hydrogen bonding patterns between neighboring amino acids in α-helix 2. This change most likely reduces or loses the function of DNA-binding. Further study will be needed to confirm the effect of this nonsynonymous substitution of *E1La* on flowering by the genetic analysis. The loss-of-function allele at *E1Lb* and the putative missense variant of *E1La* detected in this study may be useful to better comprehend the network among *E1* family genes in the control of soybean flowering. The missense *E1La* variant can also be used to breed the cultivars toward unexplored regions of higher latitudes for soybean cultivation.

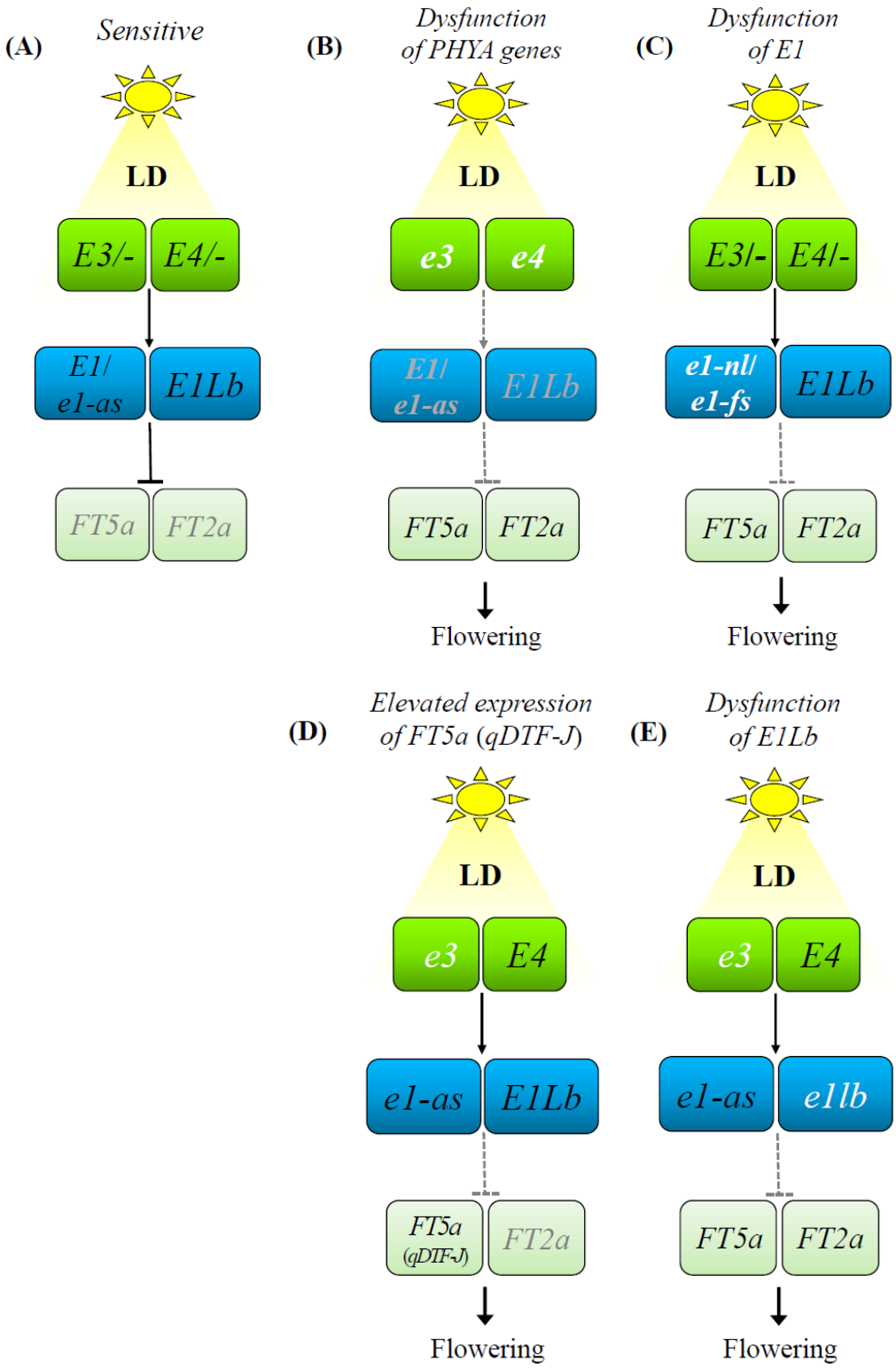


Figure 17. Genetic models of the photoperiod-insensitivity in soybean. **(A)** Photoperiod-sensitive genotypes. **(B)** Dysfunction of *PHYA* genes *E3* and *E4* (group 1 cultivar). **(C)** Dysfunction of *E1* gene (repressor for *FT2a* and *FT5a*) (group 2 cultivar). **(D)** Elevated expression of *FT5a* (group 3 cultivar). **(E)** Dysfunction of *E1Lb* (repressor for *FT2a* and *FT5a*) (group 3 cultivar).

Summary

Photoperiod response of flowering determines plant adaptation to different latitudes. Soybean is basically a short-day crop, and its flowering is inhibited under long day (LD) conditions. However, soybean has gained the ability to flower under LD conditions in early summer to produce seeds in the limited growing season of high latitudes. The lack of or lowered photoperiod-sensitivity in soybean cultivars adapted in high latitudes is mainly conditioned by dysfunction of phytochrome A genes (*E3* and *E4*) or the floral repressor *E1*. However, some cultivars from North-Eastern China and Far-Eastern Russia possess the photoperiod-insensitivity, despite that they have the identical maturity genotype as photoperiod-sensitive ones at the above three loci. The aims of my thesis were to identify a molecular-genetic basis for the photoperiod-insensitivity of Far-Eastern Russian soybean cultivars, and to determine the molecular diversity of the responsible gene and its homologue. The results obtained are summarized as follows.

In Chapter 2, I performed the genetic analyses for the insensitivity in the test-crossings of three Russian cultivars (Zeika, Yublienaya and Sonata) with Harosoy and its near-isogenic line (NIL) for *e3* (H-*e3*). The Russian cultivars and H-*e3* possessed the same maturity genotype of *e1-as*/*e2*/*e3*/*E4*. The F_2 populations produced a segregation of flowering time close to a bi-modal distribution under artificially-induced LD conditions by incandescent lamps of low R-to-FR quantum ratio. Association tests with SSR markers followed with fine mapping revealed that the insensitivity was inherited as a single recessive gene located in an 842-kb interval in the pericentromeric region of chromosome 4. Sequencing analysis for three possible candidate genes annotated in the region detected a single-nucleotide deletion in the coding sequence of *E1Lb*, a homoeologue of *E1* gene,

in the insensitive cultivars, which generated a premature stop codon. A previous study with virus-induced gene silencing approach has revealed that the *E1L* genes including *E1Lb* and *E1La* inhibit flowering in LD and night-break conditions. These results strongly suggested that the loss-of-function in the *E1Lb* gene is the most likely causal factor for the photoperiod-insensitivity of Russian cultivars. I designated this variant as *e1lb*.

In Chapter 3, I developed four sets of NILs for *e1lb* and *E1Lb* from the crossings of Zeika with Harosoy and H-*e3*. In all of the four sets of NILs, the NILs for *e1lb* flowered earlier than their respective NILs for *E1Lb* under LD conditions in both the *e3*/*E4* and *E3*/*E4* genetic backgrounds. The NILs for *e1lb* could cancel the inhibitory effect on flowering by FR–enriched LD conditions, which was mediated by *E4*. However, it could not cancel the inhibitory effect of R-enriched LD conditions by cool fluorescent lamps, which was mediated by *E3*. The NILs for *e1lb* further exhibited upregulated expression of soybean *FLOWERING LOCUS T* (*FT*) orthologues, *FT2a* and *FT5a*. Accordingly, the loss-of-function allele *e1lb* reduced the inhibitory effect of *E1Lb* on *FT2a* and *FT5a* expressions and in turn induced flowering under LD conditions.

In Chapter 4, I carried out yeast two-hybrid (Y2H) assays to determine the interaction among E1 family proteins, E1, E1La and E1Lb. The Y2H assays showed no interactions between any members of E1 family proteins or with themselves. These findings suggest that *E1Lb* repressed the expressions of *FT2a* and *FT5a* independently of *E1* and possibly *E1La* as well.

In Chapter 5, I studied the geographical distribution of *e1lb*, and evaluated the molecular diversity for *E1Lb* and its homologue *E1La* for cultivated and wild soybeans. Genotyping with an allele-specific DNA marker revealed that *e1lb* is a rare and region-specific allele even in early maturing photoperiod-insensitive cultivars. Only Russian

cultivars possessed *e1lb*. The *e1lb* allele therefore has neither largely contributed to the diversity of flowering behaviors nor been used widely in the current soybean breeding. The *e1lb* allele may be useful as a new resource to broaden the genetic variability of soybean cultivars for flowering under LD conditions at high latitudes. I also found a nonsynonymous substitution in the *E1La* gene of wild soybean accessions from Russia. The 3D protein structure analysis demonstrated that the substitution was located at the second α-helix of the B3-like DNA-binding domain, and changed the hydrogen bonding patterns between neighboring amino acids. This structural change most likely influences the DNA-binding ability of the variant E1La protein. Further research is needed to confirm this possibility.

In Chapter 6, I discussed 1) the molecular-genetic basis for the photoperiod-insensitivity of Far-Eastern Russian cultivars, and 2) the regulatory mechanisms and functions of *E1* family genes in the control of flowering, and 3) the molecular diversity of *E1La* and *E1Lb* genes. The loss-of-function *e1lb* allele and a missense *e1la-as* allele detected in this study can be used as new gene resources in breeding of photoperiod-insensitive cultivars in higher latitudes, and may also be useful to improve our understanding of the function of the *E1* family genes in the photoperiod responses of flowering in soybean.

References

Abe, J., Xu, D., Miyano, A., Komatsu, K., Kanazawa, A. and Shimamoto, Y. (2003). Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. *Crop Sci.* 43, 1300–1304. doi: 10.2135/cropsci2003.1300

Andrés, F. and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* 13, 627–639. doi: 10.1038/nrg3291

Buzzell, R.I. (1971). Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Can. J. Genet. Cytol.* 13, 703–707. doi: 10.1139/g71-100

Buzzell, R.I. and Voldeng, H.D. (1980). Inheritance of insensitivity to long daylength. *Soyb. Genet. Newsl.* 7, 26–29

Cao, D., Li, Y., Lu, S., Wang, J., Nan, H., Li, X., et al. (2015). *GmCOL1a* and *GmCOL1b* function as flowering repressors in soybean under long-day conditions. *Plant Cell Physiol.* 56, 2409–2422. doi: 10.1093/pcp/pcv152

Cao, D., Takeshima, R., Zhao, C., Liu, B., Abe, J. and Kong, F. (2017). Molecular mechanisms of flowering under long days and stem growth habit in soybean. *J. Exp. Bot.* 68, 1873–1884. doi: 10.1093/jxb/erw394

Castillejo, C. and Pelaz, S. (2008). The balance between CONSTANS and TEMPRANILLO activities determines *FT* expression to trigger flowering. *Curr. Biol.* 18, 1338–1343, doi: 10.1016/j.cub.2008.07.075

Cheng, L., Wang, Y., Zhang, C., Wu, C., Xu, J., Zhu, H., et al. (2011). Genetic analysis and QTL detection of reproductive period and post-flowering photoperiod responses in soybean. *Theor. Appl. Genet.* 123, 421–429. doi: 10.1007/s00122-011-1594-8

Cober, E.R., Tanner, J.M. and Voldeng, H.D. (1996). Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. *Crop Sci.* 36, 601–605. doi: 10.2135/cropsci1996.0011183X003600030013x

Cober, E.R., Molnar, S.J., Charette, M. and Voldeng, H.D. (2010). A new locus for early maturity in soybean. *Crop Sci.* 50, 524–527. doi: 10.2135/cropsci2009.04.0174

Doi, K., Izawa, T., Fuse, T., Yamanouchi, U., Kubo, T., Shimatani, Z., et al. (2004). *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression independently of *Hd1*. *Genes Dev.* 18, 926–936. doi: 10.1101/gad.1189604

Doyle, J.J. and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13–15.

Fan, C., Hu, R., Zhang, X., Wang, X., Zhang, W., Zhang, Q., et al. (2014). Conserved

CO-FT regulons contribute to the photoperiod flowering control in soybean. *BMC Plant Biol*.14, 9. doi: 10.1186/1471-2229-14-9

Fehr, W.R., Caviness, C.E., Burmood, D.T. and Pennington, J.S. (1971). Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183X001100060051x

Funatsuki, H., Kawaguchi, K., Matsuba, S., Sato, Y. and Ishimoto, M. (2005). Mapping of QTL associated with chilling tolerance during reproductive growth in soybean. *Theor. Appl. Genet.* 111, 851–861. doi: 10.1007/s00122-005-0007-2

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T. (1999). Plant *cis*-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.* 2, 297–300. doi: 10.1093/nar/26.1.358

Hu, Y., Wang, Y., Liu, X. and Li, J. (2004). *Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. *Cell Research* 14, 8–15. doi: 10.1038/sj.cr.7290197

Izawa, T., Oikawa, T., Sugiyama, N., Tanisaka, T., Yano, M. and Shimamoto, K. (2002). Phytochrome mediates the external light signal to repress *FT* orthologs in photoperiodic flowering of rice. *Genes Dev.* 16, 2006–2020. doi: 10.1101/gad.999202

Jia, H., Jiang, B., Wu, C., Lu, W., Hou, W., Sun, S., et al. (2014). Maturity group classification and maturity locus genotyping of early-maturing soybean varieties from high-latitude cold regions. *PLoS One* 9, e94139. doi: 10.1371/journal.pone.0094139

Jiang, B., Nan, H., Gao, Y., Tang, L., Yue, Y., Lu, S., et al. (2014). Allelic combinations of soybean maturity Loci *E1*, *E2*, *E3* and *E4* result in diversity of maturity and adaptation to different latitudes. *PLoS One* 9, e106042. doi: 10.1371/journal.pone.0106042

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., et al. (2002). *Hd3a*, a rice ortholog of the Arabidopsis *FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant Cell Physiol.* 43, 1096–1105. doi: 10.1093/pcp/pcf156

Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B. M., Watanabe, S., et al. (2010). Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. *Plant Physiol.* 154, 1220–1231. doi: 10.1104/pp.110.160796

Kong, L., Lu, S., Wang, Y., Fang, C., Wang, F., Nan, H., et al. (2018). Quantitative trait locus mapping of flowering time and maturity in soybean using next-generation sequencing-based analysis. *Front. Plant Sci.* doi: 10.3389/fpls.2018.00995

Kurasch, A.K., Hahn, V., Leiser, W.L., Vollmann, J., Schori, A., Bétrix, C.A., et al. (2017). Identification of mega-environments in Europe and effect of allelic variation at maturity *E* loci on adaptation of European soybean. *Plant Cell Environ.* 40, 765–778. doi: 10.1111/pce.12896

Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D. and Bilyeu, K. (2014). Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. *PLoS One* 9, e94150. doi: 10.1371/journal.pone.0094150

Levy, Y., Mesnage, S., Mylne, J., Gendall, A. and Dean, C. (2002). Multiple roles of *arabidopsis VRN1* in vernalization and flowering time control. *Science* 297, 243–246. doi: 10.1126/science.1072147

Li, J., Wang, X., Song, W., Huang, X., Zhou, J., Zeng, H., et al. (2017). Genetic variation of maturity groups and four *E* genes in the Chinese soybean mini core collection. *PLoS One* 12, e0172106. doi: 10.1371/journal.pone.0172106

Liu, B. and Abe, J. (2010). QTL mapping for photoperiod insensitivity of a Japanese soybean landrace Sakamotowase. *J. Hered.* 101, 251–256. doi: 10.1093/jhered/esp113

Liu, B., Fujita, T., Yan, Z. H., Sakamoto, S., Xu, D. and Abe, J. (2007). QTL mapping of domestication-related traits in soybean (*Glycine max*). *Ann. Bot.* 100, 1027–1038. doi: 10.1093/aob/mcm149

Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K. and Abe, J. (2008). Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics* 180, 995–1007. doi: 10.1534/genetics.108.092742

Liu, W., Kim, M.Y., Kang, Y.J., Van, K., Lee, Y.H., Srinives, P., et al. (2011). QTL identification of flowering time at three different latitudes reveals homeologous genomic regions that control flowering in soybean. *Theor. Appl. Genet.* 123, 545–553. doi: 10.1007/s00122-011-1606-8

Lu, S., Li, Y., Wang, J., Srinives, P., Nan, H., Cao, D., et al. (2015). QTL mapping for flowering time in different latitude in soybean. *Euphytica* 206, 725–736. doi: 10.1007/s10681-015-1501-5

McBlain, B., Hesketh, J. and Bernard, R. (1987). Genetic effect on reproductive phenology in soybean isolines differing in maturity genes. *Can. J. Plant Sci*. 67, 105–116. doi: 10.4141/cjps87-012

Nan, H., Cao, D., Zhang, D., Li, Y., Lu, S., Tang, L., and et al., (2014). GmFT2a and GmFT5a redundantly and differentially regulate flowering through interaction with and upregulation of the bZIP transcription factor GmFDL19 in soybean. *PLoS One* 9, e97669. doi: 10.1371/journal.pone.0097669

Nemoto, Y., Nonoue, Y., Yano, M. and Izawa, T. (2016). *Hd1a*, *CONSTANS* ortholog in rice, functions as an *Ehd1* repressor through interaction with monocot−specific CCT−domain protein Ghd7. *Plant J*. 86, 221–223. doi: 10.1111/tpj.13168

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., et al. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem*. 25, 1605–1612. doi: 10.1002/jcc.20084

Roy, A., Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 5, 725–738. doi: 10.1038/nprot.2010.5

Saindon, G., Voldeng, H.D., Beversdorf, W.D. and Buzzell, R.I. (1989). Genetic control of long daylength response in soybean. *Crop Sci.* 29, 1436–1439. doi: 10.2135/cropsci1989.0011183X002900060021x

Samanfar, B., Molnar, S.J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., et al. (2017). Mapping and identification of a potential candidate gene for a novel maturity locus, *E10*, in soybean. *Theor. Appl. Genet.* 130, 377. doi: 10.1007/s00122-016-2819-7

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Severin A.J., Woody J.L., Bolon Y.T., Joseph B., Diers B.W., Farmer A.D., et al. (2010) RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.* 10, 160. doi: 10.1186/1471-2229-10-160.

Swaminathan, K., Peterson, K. and Jack, T. (2008). The plant B3 superfamily. *Trends Plant Sci*. 13, 647–655. doi: 10.1016/j.tplants.2008.09.006

Song, Q., Jia, G., Zhu, Y., Grant, D., Nelson, R.T., Hwang, E.Y., et al. (2010). Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. *Crop Sci*. 50, 1950–1960. doi: 10.2135/cropsci2009.10.0607

Song, Q., Marek, L.F., Shoemaker, R.C., Lark, K.G., Concibido, V.C., Delannay, X., et al. (2004). A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* 109, 122–128. doi: 10.1007/s00122-004-1602-3

Song, Y., Ito, S. and Imaizumi, T. (2013). Flowering time regulation: photoperiod- and temperature-sensing in leaves. *Trends Plant Sci.* 18, 575–583. doi: 10.1016/j.tplants.2013.05.003

Takeshima, R., Hayashi, T., Zhu, J., Zhao, C., Xu, M., Yamaguchi, N., et al. (2016). A soybean quantitative trait locus that promotes flowering under long days is identified

as *FT5a*, a *FLOWERING LOCUS T* ortholog. *J. Exp. Bot.* 67, 5247–5258. doi: 10.1093/jxb/erw283

Tamaki, S., Matsuo, S., Wong, H.L., Yokoi, S. and Shimamoto, K. (2007). Hd3a protein is a mobile flowering signal in rice. *Science* 316, 1033–1036. doi: 10.1126/science.1141753

Tsubokura, Y., Matsumura, H., Xu, M., Liu, B., Nakshima, H., Anai, T., et al. (2013). Genetic variation in soybean at the maturity locus *E4* is involved in adaptation to long days at high latitudes. *Agronomy* 3, 117–134. doi: 10.3390/agronomy3010117

Tsubokura, Y., Watanabe, S., Xia, Z., Kanamori, H., Yamagata, H., Kaga, A., et al. (2014). Natural variation in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. *Ann. Bot.* 113, 429–441. doi: 10.1093/aob/mct269

Upadhyay, A.P., Ellis, R.H., Summerfield, R.J., Roberts, E.H. and Qi, A. (1994). Characterization of photothermal flowering responses in maturity isolines of soyabean [*Glycine max* (L.) Merrill] cv. Clark. *Ann. Bot.* 74, 87–96. doi: 10.1093/aob/74.1.87

Wang, D., Graef, G.L., Procopiuk, A.M. and Diers, B.W. (2004). Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theor. Appl. Genet.* 108, 458–467. doi: 10.1007/s00122-003-1449-z

Watanabe, S., Harada, K. and Abe, J. (2012). Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breed. Sci*. 61, 531–543. doi: 10.1270/jsbbs.61.531

Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., Sato, S., Nakamoto, Y., et al. (2009). Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics* 182, 1251–1262. doi: 10.1534/genetics.108.098772.

Watanabe, S., Tsukamoto, C., Oshita, T., Yamada, T., Anai, T. and Kaga, A. (2017). Identification of quantitative trait loci for flowering time by a combination of restriction site-associated DNA sequencing and bulked segregant analysis in soybean. *Breed. Sci.* 67, 277–285. doi: 10.1270/jsbbs.17013

Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., et al. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics* 188, 395–407. doi: 10.1534/genetics.110.125062

Wu, F., Price, B.W., Haider, W., Seufferheld, G., Nelson, R. and Hanzawa, Y. (2014). Functional and evolutionary characterization of the *CONSTANS* gene family in short–day photoperiodic flowering in soybean. *PLoS One* 9, e85754. doi: 10.1371/journal.pone.0085754

Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., et al., (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2155–2164. doi: 10.1073/pnas.1117982109

Xu, M., Xu, Z., Liu, B., Kong, F., Tsubokura, Y., Watanabe, S., et al. (2013). Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. *BMC Plant Biol.* 13, 91. doi: 10.1186/1471-2229-13-91

Xu, M., Yamagishi, N., Zhao, C., Takeshima, R., Kasai, M., Watanabe, S., et al. (2015). The soybean-specific maturity gene *E1* family of floral repressors controls night-break responses through down-regulation of *FLOWERING LOCUS T* orthologs. *Plant Physiol.* 168, 1735–1746. doi: 10.1104/pp.15.00763

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., et al. (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nature Genetics* 40, 761–767. doi: 10.1038/ng.143

Yamada, T., Hajika, M., Yamada, N., Hirata, K., Okabe. A., Oki, N., et al. (2012). Effects on flowering and seed yield of dominant alleles at maturity loci *E2* and *E3* in a Japanese cultivar, Enrei. *Breed. Sci.* 61, 653–660. doi: 10.1270/jsbbs.61.653

Yamanaka, N., Nagamura, Y., Tsubokura, Y., Yamamoto, K., Takahashi, R., Kouchi H., et al. (2000). Quantitative trait locus analysis of flowering time in soybean using a RFLP linkage map. *Breed. Sci.* 50, 109–115. doi: 10.1270/jsbbs.50.109

Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., et al. (2004). Solution structure of the B3 DNA binding domain of the arabidopsis cold-responsive transcription factor RAV1. *Plant Cell* 16, 3448–3459. doi: 10.1105/tpc.104.026112

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods* 12, 7–8. doi: 10.1038/nmeth.3213

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., et al. (2000). *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS. Plant Cell* 12, 2473–2483. doi: 10.1105/tpc.12.12.2473

Zhai, H., Lü, S., Wang, Y., Chen, X., Ren, H., Yang, J., et al. (2014). Allelic variations at four major maturity *E* genes and transcriptional abundance of the *E1* gene are associated with flowering time and maturity of soybean cultivars. *PLoS One* 9, e97636. doi: 10.1371/journal.pone.0097636

Zhai, H., Lü, S., Wu, H., Zhang, Y., Zhang, X., and Yang, J., (2015). Diurnal expression pattern, allelic variation, and association analysis reveal functional features of the *E1*

gene in control of photoperiodic flowering in soybean. *PLoS One* 10, e0135909. doi: 10.1371/journal.pone.0135909

Zhang, W., Wang, Y., Luo, G., Zhang, J., He, C., Wu, X., et al. (2004). QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. *Theor. Appl. Genet.* 108, 1131–1139. doi: 10.1007/s00122-003-1527-2

Zhang, X., Zhai, H., Wang, Y., Tian, X., Zhang, Y., Wu, H., et al. (2016). Functional conservation and diversification of the soybean maturity gene *E1* and its homologs in legumes. *Scientific Reports* 6, 29548. doi: 10.1038/srep29548

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9, 40. doi: 10.1186/1471-2105-9-40

Appendix

Supplemental Table 1. Primers used for fine-mapping in this study

SSR marker		Primer sequences (5′ to 3′)
Satt190	F	GGGAGTGTGAACTTACATTGTCT
	R	GGGCCTTGAATTTTGTGCTAT
Sat_085	F	GGTTTTAGATCCTTAAATTTGT
	R	GGGGAAGCAAGTAGCT
BARC-04g-0685; M1	F	TGCATTCACATCAGTAGAGGCT
	R	CAATCTCAACCCAATATCTCACC
BARC-04g-0725; M2	F	CACAACCCTAGCACCTACACC
	R	ATGGGTCAAACCCAACTCAA
BARC-04g-0881; M3	F	AATGAGAGAGCTTGCAGGAA
	R	TGGGGTAAGTTGTTACATCAAA
BARC-04g-0859; M4	F	AAGCCAACCTTATAATTCTTTCAT
	R	ATATGGGCTTACTTACCCATCATAGA
BARC-18g-0889; M5	F	GACAATTTGATATGTCTTCCCCA
	R	GAGGCAGAATGCAATGGTTT
BARC-18g-0895; M6	F	TACGTCATCCCCAAATGCTT
	R	TGAAAATCGAATCATAAATAGCAAA
BARC-18g-0905; M7	F	AACGCAGTACCACACCTTCC
	R	ACCCAACTTGTGAACCCGTA
BARC-18g-0908; M8	F	AGAAAGGGGGTTGAATCGTG
	R	AGGCAGCAAATGATTTTGGA
BARC-18g-0911; M9	F	TGCACAACCACCAAAGTCAT
	R	GTCGTCCTGGAAATTGCATC
BARC-04g-0751; M10	F	CACTTCCATTTAAAACCTGCAA
	R	CCTCGCTCACTTTAGTTGTGA
BARC-04g-0781; M11	F	TTGCTTCTGCATCTCCTTTTT
	R	CCAAATCTGAAAAAGAACCGA

Supplemental Table 2. Primers used for expression analysis in this study

Gene		Primer sequences (5′ to 3′)
β-Tubulin	F	GACCCGATAACTTCGTGTTC
	R	CAGTTCCACCTCCCAGTGAGTGAC
E1	F	CACTCAAATTAAGCCCTTTCA
	R	TTCATCTCCTCTTCATTTTTGTTG
E1La	F	AAACACTCAAAGCCCGATCA
	R	ATCCTCTTCATTTTTGTTGCTGA
E1Lb	F	GTGTAAACACTCAAAGTCCTT
	R	CTCCTCTTCATTTTTGTTGCTGC
FT2a	F	GGATTGCCAGTTGCTGCTGT
	R	GAGTGTGGGAGATTGCCAAT
FT5a	F	GCCTTACTCCAGCTTATACT
	R	GGCATGCTCTAGCATTGCAA

Supplemental Table 3. Primers used for sequencing analysis in this study

Gene		Primer sequences (5′ to 3′)
Glyma.04G143300	F	GTGTAAACACTCAAAGTCCTT
	R	GATTTGAAAGTAAAATAAAGCTAACACTT
Glyma.04G143400	F	CGTTGTATGACTCTGTAGGCCTTAC
	R	GTTTCAAACTCATACACACAATCG
Glyma.04G143500	F	GTGCATATTAGGAATCTATAGGACCAAT
	R	AGGGCAGCTAAAATAAAGGCCCT

Supplemental Table 4. Primers used for direct PCR sequencing analysis in this study

Gene		Primer sequences (5′ to 3′)
E1La	F	TTGAAAGAAAGGGGAAATGAGT
	R	TTAAACAACCTCATCAAACAAATCCAC
E1Lb	F	TTGAAAGAAAGGGGAAATGAGC
	R	ATAGAAAGGAAATAAACAATGACAA

Supplemental Table 5. Primers used for yeast two-hybrid in this study

Primer name	Restriction enzyme		Primer sequences (5′ to 3′)
E1 for pGADT7	EcoRI	F	CGAGGCCGAATTCATGAGCAACCCTTCAGATGAAAGG
	XhoI	R	GAGCCGCTCGAGCTTATTCTCTGGCATAGCTTGTTT
E1 for pGBKT7	NcoI	F	CATGCCATGGAGATGAGCAACCCTTCAGATGAAAGG
	EcoRI	R	CGAGGCCGAATTCTTAATTCTCTGGCATAGCTTGTTT
E1L for pGADT7	EcoRI	F	CGAGGCCGAATTCATGAGTTTACAATCACAAATCCCA
	XhoI	R	GAGCCGCTCGAGCTTAATTCTCTGGCATAGCTCGTTT
E1L for pGBKT7	NcoI	F	CATGCCATGGAGATGAGTTTACAATCACAAATCCCA
	EcoRI	R	CGAGGCCACCTTCTTAATTCTCTGGCATAGCTCGTTT

E1La and *E1Lb* share the same primers, products were amplified from plasmid DNAs containing full length sequence of *E1La* and *E1Lb,* respectively.

Supplemental Table 6. Accessions used in genpotyping of *E1Lb* by dCAPS marker

	Accession	Region	Genotype at				
			E1	*E2*	*E3*	*E4*	*E1Lb*
1	Kamaishi 17	Tohoku, Japan	*E1*	*e2*	*e3*	*e4*	*E1Lb*
2	Ohfunato 45	Tohoku, Japan	*E1*	*e2*	*e3*	*e4*	*E1Lb*
3	Otome wase	Tohoku, Japan	*E1*	*e2*	*e3*	*e4*	*E1Lb*
4	Gokuwase Kamishunbetsu	Hokkaido, Japan	*e1-re*	*e2*	*e3*	*e4*	*E1Lb*
5	Karafuto 1	Hokkaido, Japan	*e1-nl*	*e2*	*e3*	*e4*	*E1Lb*
6	Miharudaizu	Hokkaido, Japan	*E1*	*e2*	*e3*	*e4*	*E1Lb*
7	Ohyachi 2	Hokkaido, Japan	*E1*	*e2*	*e3*	*e4*	*E1Lb*
8	Okuhara 1	Hokkaido, Japan	*E1*	*e2*	*e3*	*e4*	*E1Lb*
9	Sakamoto wase	Hokkaido, Japan	*e1-fs*	*e2*	*e3*	*E4*	*E1Lb*
10	Dongda 2	North-East China	*e1-nl*	*e2*	*e3-tr*	*E4*	*E1Lb*
11	Heihe 12	North-East China	*e1-as*	*e2*	*e3-tr*	*e4-SORE1*	*E1Lb*
12	Heihe 13	North-East China	*e1-as*	*e2*	*e3-fs*	*e4-kes*	*E1Lb*
13	Heihe 21	North-East China	*e1-as*	*e2*	*e3-tr*	*e4-SORE1*	*E1Lb*
14	Heihe 28	North-East China	*e1-nl*	*e2*	*e3-tr*	*E4*	*E1Lb*
15	Heihe 33	North-East China	*e1-as*	*e2*	*e3-tr*	*e4-kes*	*E1Lb*
16	Heihe 34	North-East China	*e1-as*	*e2*	*e3-tr*	*E4*	*E1Lb*
17	Heihe 35	North-East China	*e1-as*	*e2*	*e3-fs*	*e4-kes*	*E1Lb*
18	Heihe 40	North-East China	*e1-as*	*e2*	*e3-fs*	*e4-kes*	*E1Lb*
19	Jiagedaqi 01	North-East China	*e1-nl*	*e2*	*e3*	*E4*	*E1Lb*
20	Jiagedaqi 02	North-East China	*e1-as*	*e2*	*e3*	*E4*	*E1Lb*
21	Jiagedaqi 03	North-East China	*e1-nl*	*e2*	*e3*	*E4*	*E1Lb*
22	Jiagedaqi 04	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
23	Jiagedaqi 05	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
24	Jiagedaqi 08	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
25	Jiagedaqi 09	North-East China	*e1-nl*	*e2*	*e3*	*E4*	*E1Lb*

26	Jiagedaqi 10	North-East China	*e1-nl*	*e2*	*e3*	*E4*	*E1Lb*
27	Jiagedaqi 11	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
28	Jiagedaqi 12	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
29	Jiagedaqi 13	North-East China	*e1-nl*	*e2*	*e3*	*E4*	*E1Lb*
30	Jiagedaqi 14	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
31	Jiagedaqi 16	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
32	Jiagedaqi 17	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
33	Jiagedaqi 18	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
34	Jiagedaqi 19	North-East China	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
35	Jiagedaqi 20	North-East China	*e1-as*	*e2*	*e3*	*E4*	*E1Lb*
36	Darta	Poland	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
37	Gai	Poland	*e1-nl*	*e2*	*e3*	*e4*	*E1Lb*
38	Nawiko	Poland	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
39	Amurskaya 41	Russia	*e1-as*	*e2*	*E3*	*E4*	*E1Lb*
40	DB no. 23	Russia	*e1-as*	*e2*	*e3*	*E4*	*E1Lb*
41	DYA-1	Russia	*e1-as*	*e2*	*e3*	*E4*	*e1lb*
42	Krasnoarmeiskaya	Russia	*e1-as*	*E2*	*e3*	*E4*	*E1Lb*
43	Mesinaya	Russia	*e1-as*	*e2*	*e3*	*E4*	*E1Lb*
44	Mesinaya Salatnaya	Russia	*E1*	*e2*	*E3*	*E4*	*E1Lb*
45	Oktyabr-70	Russia	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
46	Sadovy	Russia	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
47	Salyut 216	Russia	*e1-as*	*E2*	*e3*	*E4*	*e1lb*
48	Sonata	Russia	*e1-as*	*e2*	*e3*	*E4*	*e1lb*
49	Vega	Russia	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
50	Venera	Russia	*e1-as*	*e2*	*E3*	*E4*	*E1Lb*
51	VIR-29	Russia	*e1-as*	*e2*	*e3*	*E4*	*E1Lb*
52	Yubileinaya	Russia	*e1-as*	*e2*	*e3*	*E4*	*e1lb*

53	Zeika	Russia	*e1-as*	*e2*	*e3*	*E4*	*e1lb*
54	Zeya 2	Russia	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
55	Arcadia	Ukline	*e1-as*	*e2*	*e3*	*E4*	*E1Lb*
56	Hadjibey	Ukraine	*e1-as*	*e2*	*E3*	*E4*	*E1Lb*
57	Kiev 242 BH	Ukraine	*e1-nl*	*e2*	*E3*	*e4*	*E1Lb*
58	Kiev 242 WH	Ukraine	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*
59	Odesskaya 150	Ukraine	*e1-as*	*e2*	*e3*	*E4*	*E1Lb*
60	Palmira	Ukraine	*e1-as*	*E2*	*e3*	*E4*	*E1Lb*
61	Ustya	Ukraine	*e1-nl*	*e2*	*E3*	*e4*	*E1Lb*
62	Yug 30	Ukraine	*e1-as*	*e2*	*e3*	*e4*	*E1Lb*

Genetypic information on *E1* to *E4* is cited from Xu et al. (2013)

NOTE:

Part of the results presented in this thesis has been published in *Frontiers in Plant Science.*

Jianghui Zhu, Ryoma Takeshima, Kohei Harigai, Meilan Xu, Fanjiang Kong, Baohui Liu4, Akira Kanazawa, Tetsuya Yamada and Jun Abe. (2018). Loss of function of the *E1-Like-b* gene associates with early flowering under long-day conditions in soybean. doi: 10.3389/fpls.2018.01867

Acknowledgements

First of all, I would like to express my sincere appreciation to my supervisor Professor Jun Abe for the continuous support of my doctoral study and thesis throughout my doctoral study life. My doctoral study and thesis would not have accomplished without his great and kind support.

I acknowledge Professor Yuji Kishima, Associate Professor Akira Kanazawa and Lecturer Tetsuya Yamada for their kind, critical comments and suggestions, which were indispensable for the accomplishment of my doctoral study and thesis.

I also appreciate Professor Baohui Liu and Professor Fanjiang Kong (School of Life Science, Guangzhou University; The Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences) for their constructive suggestions and cooperation to my study. I would like to thank all members of The Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, especially Dr. Sijia Lu for instructing me the yeast two-hybrid assay.

I am really grateful to all members of Plant Genetics and Evolution laboratory. Special thanks for Dr. Meilan Xu, Mr. Kohei Harigai and Dr. Ryoma Takeshima (Institude of Crop Science, National Agriculture and Food Research Organization) for their advices and experimental help.

Finally, I would like to thank my parents for their mental and financial support and encouragement during my study life.