# Molecular-Genetic Study on Soybean Maturity Gene *E9* and its Role on Flowering

# （ダイズの感光性遺伝子 *E9* とその開花に及ぼす役割に関する分子遺伝学的研究）

**Hokkaido University　Graduate School of Agriculture**

**Division of Agrobiology and Bioresources　Doctor Course**

**Zhao Chen**

# CONTENTS


Chapter Ⅰ Introduction ..... 1
Chapter Ⅱ Review on molecular mechanism on flowering in plant ..... 5
2.1 Photoperiod ..... 5
2.1.1 Photoreceptors ..... 5
2.1.2 Photoperiodism and the circadian clock ..... 6
2.2 Photoperiodic flowering in long-day plant *Arabidopsis* and short-day plant rice ..... 8
2.2.1 Photoperiodic flowering in *Arabidopsis* ..... 8
2.2.2 Photoperiodic flowering in rice ..... 9
2.3 Photoperiodic flowering in soybean ..... 10
2.3.1 *E1*, which is the largest effect gene on flowering, and *E1-like* genes ..... 11
2.3.2 *E2*, which is a *GIGANTEA* ortholog, and CO-FT pathways in soybean ..... 12
2.3.3 *E3*, *E4*, and other photoreceptor-related genes ..... 14
2.3.4 GmFT, a flowering integrator ..... 15
Chapter Ⅲ *E9* gene identified in a cross between Harosoy and Toyomusume ..... 19
3.1 Background and purpose ..... 19
3.2 Materials and methods ..... 20
3.2.1 Plant materials and segregation analysis ..... 20
3.2.2 DNA marker analysis in association tests ..... 20
3.3 Results and discussion ..... 21
3.3.1 Segregation of flowering time in $F_2$ and $F_3$ populations ..... 21
3.3.2 Test for association between flowering time and SSR markers ..... 22
Chapter Ⅳ Fine-mapping and identification of responsible gene for the *E9* gene ..... 24
4.1 Background and purpose ..... 24
4.2 Materials and methods ..... 24
4.2.1 Plant materials for fine-mapping expression and sequencing analyses ..... 24
4.2.2 Genotyping for DNA markers and maturity loci ..... 25
4.2.3 qRT-PCR ..... 26
4.2.4 Sequencing analysis of *FT2a* and *SORE-1* ..... 27
4.4 Results and discussion ..... 28
4.4.1 Fine-mapping and association analysis ..... 28
4.4.2 cDNA sequencing and expression analysis ..... 31
4.4.3 Sequence analysis of the *FT2a* genomic region ..... 33
4.4.4 Expression of different *FT2a* alleles in near-isogenic lines and photoperiod insensitive accessions ..... 35
Chapter Ⅴ Molecular mechanism of delayed flowering mediated by the *e9* allele ..... 38
5.1 Background and purpose ..... 38
5.2 Materials and methods ..... 39
5.2.1 Plant materials ..... 39
5.2.2 Analysis of RNA processing ..... 39

5.2.3 Methylation analysis ..... 40
5.3 Results and Discussion ..... 40
5.3.1 RNA processing of the *E9* gene ..... 40
5.3.2 DNA methylation at the *E9* locus ..... 44
Chapter Ⅵ Discussion ..... 46
6.1 Flowering gene *E9* is *FT2a* ..... 46
6.2 Factors responsible for attenuation of *FT2a* expression ..... 47
6.3 Origin and adaptive role of the *e9* allele ..... 48
Summary ..... 51
Lists for figures ..... 75
List for tables ..... 76
List for abbreviations ..... 77
Acknowledgements ..... 84

# Chapter Ⅰ Introduction

Soybean (*Glycine max* (L.) Merr.) is an important crop as the primary supplier of vegetable proteins and oil. It is cultivated widely in the world, and geographically distributed from an equatorial region to regions with high latitude such as 50 degree north or 35 degree south. However, the cultivation area of a soybean cultivar is relatively limited to a narrow range of latitudes (Watanabe *et al.* 2012). The wide adaptability of soybean has thus been attributed to a genetic variation in photoperiod sensitivity, which is generated by natural variation in the major genes and/or quantitative trait loci (QTLs) controlling flowering.

At present, ten major genes, *E1* to *E9* and *J*, and many QTLs controlling time to flowering are reported in soybean (Watanabe *et al.* 2012). However, only four genes (*E1* to *E4*) are identified for the underlying molecular bases. *E1* encodes a possible transcription factor down-regulating the expressions of soybean *FLOWERING LOCUS T* orthologs, *FT2a* and *FT5a* (Xia *et al.* 2012). *E2* is an ortholog of *Arabidopsis GIGANTEA* (*GI*) (Watanabe *et al.* 2011). *E3* and *E4* encode the phytochrome A isoforms, GmPHYA3 and GmPHYA2, respectively (Liu *et al.* 2008a; Watanabe *et al.* 2009). Identification of responsible genes for the four loci has further facilitated the detection of allelic variations at each locus and novel genes (Xu *et al.* 2013; Tsubokura *et al.* 2014). Natural variations at these loci have conferred diverse photoperiod responses in flowering among soybean cultivars, although the genetic variation still remains to be accounted for only by the four loci among cultivars (Xu *et al.* 2013; Tsubokura *et al.* 2014). However, the molecular basis of natural variation in time to flowering and maturity is still poorly understood. Understanding of molecular mechanisms for flowering has become more and more important the significance in soybean breeding, because times to flowering and maturity determine the adaptability and yield of cultivars in a region.

Genetic signaling pathways of flowering have been well characterized in a model

long day (LD) plant *Arabidopsis thaliana* (Wilczek *et al*. 2010). The findings that have been accumulated in *Arabidopsis* are so important to provide information about genes and genetic pathways controlling the flowering process for other flowering plants including important crops. Natural variation in flowering time in major crops such as rice, wheat, and pea has been reported to result from the variation in orthologs of *Arabidopsis* flowering genes, such as *Hd1* (*CO*) and *Hd3a* (*FT*) in rice (Yano *et al.* 2000; Kojima *et al.* 2002; Yan *et al.* 2003, 2006), *Vrn1* (*APETALA1*) and *Vrn3* (*FT*) in wheat (Foucher *et al.* 2003), and *LATE FLOWERING* (*TERMINAL FLOWER 1*; *TFL1*), *LATE BLOOMER1* (*GI*) and *GIGAS* (*FT*) in pea (Hecht *et al.* 2007, 2011; Watanabe *et al.* 2012). In soybean, identification and characterization of orthologs of *Arabidopsis* flowering genes has been also an important research approach to detect novel genes involved in photoperiodic flowering response and understand the diverse regulation mechanisms underlying this process. For example, Kong *et al.* (2010) identified 10 *FLOWERING LOCUS T* (*FT*) homologs in soybean, using the soybean EST and genome sequence databases. Transcription analyses for six of the *FT* homologs revealed that two homologs, *GmFT2a* and *GmFT5a*, were strongly expressed in trifoliate leaves and the expression levels of *GmFT2a* and *GmFT5a* were lower under LD conditions than short day (SD) conditions (Kong *et al.* 2010). Ectopic expression analyses in *Arabidopsis* further confirmed that both *GmFT2a* and *GmFT5a* had the same function as *Arabidopsis FT* (Kong *et al.* 2010). A number of other orthologs of *Arabidopsis* flowering genes have been also characterized: *AP1* (*APETALA1*) (Nan *et al*. 2014), *COL* (*CO-like*) (Fan *et al.* 2014; Wu *et al.* 2014), *CPDs* (Wang *et al*. 2015), *CRY* (*CRYPTOCHROME*) (Zhang *et al.* 2008; Matsumura *et al.* 2009), *FKF1* (Li *et al*. 2013), *LFY* (*LEAFY*) (Meng *et al*. 2007; Nan *et al*. 2014), *FLD* (*FLOWERING LOCUS D*) (Hu *et al*. 2014), *FUL* (*FRUITFULL*) (Jia *et al*. 2015), *MADS28* (Huang *et al*. 2014), *RAV-like* (*RELATED TO ABI3/VP1-like*) (Lu *et al*. 2014), *SOC1/AGL20* (*SUPPRESSOR OF OVEREXPRESSION OF COL1/AGAMOUS-LIKE 20*) (Zhong *et al*. 2012; Na *et al*. 2013), *TOE* (*TARGET OF EAT1*) (Zhao *et al*. 2015), and *ZTL* (*ZEITLUPE*) (Xue *et al*. 2012). A genome-wide association study also revealed a number of SNPs that were significantly associated

with flowering time; some of these SNPs implied an involvement of orthologs to *Arabidopsis* flowering genes, such as *EARLY FLOWERING 8* and *SOC1* or *AGAMOUS-LIKE 6*, in the control of flowering time in soybean (Zhang *et al.* 2015). In addition to these genes, a total of 333 orthologs of 92 *Arabidopsis* genes from among a total of 46,367 annotated genes was detected in the soybean genome (Watanabe *et al.* 2012). However, it remains to be determined whether these orthologs are involved in genetic variation of flowering in soybean.

Another approach in identifying genes responsible for flowering is to utilize QTL analysis and subsequent fine-mapping with a help of genome sequence information. The information on the physical position of orthologs to known *Arabidopsis* genes may be also greatly useful for identifying candidate genes for targeted major loci and QTLs. For example, Watanabe *et al.* (2004) detected three QTLs, *qFT1*, *qFT2* and *qFT3* for flowering time in a recombinant inbred lines (RILs) population derived from a cross between a Japanese cultivar Misuzudaizu and a Chinese cultivar Moshido Gong 503 (Yamanaka *et al.* 2001). Using fine mapping approaches, they could successfully narrow the genomic region of *qFT2* to a region of approximately 100 kb. Based on the annotated genes in the region in the Williams 82 genome sequence (Schmutz *et al*. 2010), Watanabe *et al.* (2011) detected a soybean otholog of the *Arabidopsis GIGANTEA* (*GI*) in nine annotated genes as a candidate gene for *qFT2* (*E2*). They finally concluded by sequence analysis in *E2* and *e2* genotypes, together with the experiment using a mutant line for *E2*, that *qFT2* (*E2*) is a soybean *GI* ortholog.

Liu *et al.* (2007) detected a QTL for flowering time in linkage group (LG) J (chromosome 16), *FT04*, in a RIL population derived from a cross between cultivated soybean line TK780 (TK) and a wild soybean Hidaka 4 (H4). Kong *et al.* (2014) made test crosses between RILs which possessed the same maturity genotype at *E1* to *E4*, but differed in the alleles at *FT04*, and found that the QTL segregated as a single locus designated as *E9*, at which a dominant allele *E9* conferred early-flowering, whereas the recessive allele *e9* conditioned the delayed flowering. A further

fine-mapping delimited *E9* to a 245-kb region where two soybean orthologs of *FT*, *FT2a* and *FT2b*, were located. However, it remains undetermined whether either of the two genes is a responsible gene and what mechanisms are involved in different controls of flowering observed between the two alleles.

In this thesis, I studied molecular bases of *E9* to answer the above questions. First of all, I reviewed recent findings on molecular bases of flowering which have been accumulated in model plant species, *Arabidopsis* and rice. After that, I reviewed findings of previous reports on flowering in soybean. In Chapters III, I examined segregations of flowering time in the progeny of a cross between early-maturing cultivars Toyomusume and Harosoy, which were associated with those at the *E1* and *E9* loci. In Chapter IV, I performed fine mapping of *E9* and detected a candidate of responsible gene with the use of the genome sequence information of Williams 82, and confirmed that it was *FT2a* by using sequencing and expression analyses. I found that *SORE-1*, a *Ty1/copia*-like retrotransposon, was inserted in the first intron of *FT2a* in the *e9* allele. In Chapter V, I analyzed RNA processing and DNA methylation at the *e9* allele to determine the function of inserted retrotransposon in the *FT2a* expression. In Chapter VI, I discussed the molecular bases of *E9* and its function on soybean flowering. The findings obtained in this study may be useful not only in understanding of molecular mechanism of flowering, but also in breeding of cultivars of both early-flowering and late-flowering genetic backgrounds in soybean.

# Chapter Ⅱ Review on molecular mechanism on flowering in plant

## 2.1 Photoperiod

### 2.1.1 Photoreceptors

Light is a critical factor for plant growth, and has three principal characteristics to affect plant growth and productivity: quantity (intensity), duration of light (photoperiod), and quality (color or wavelength) (Kumar and Rani 1999). Light quantity to plant is the amount of light particles exposed around the plants, and varies within a year and can be changed by greenhouse or reflective objects (Macinnes *et al*. 1969). Duration of light, referred to the length of the light period, is also crucial for initiating flowering (Roden *et al*. 2002). In seven kinds of wavelengths, blue light and red/far-red (R/FR) light are most important for plant vegetable growth and flowering (Tibbitts *et al*. 1983; Okamoto *et al*. 1996; Schuerger *et al*. 1997). For the indoor gardening, incandescent light lamps and fluorescent cool white lamps are used to simulate light conditions by creating red and blue lights, respectively (Stasinopoulos *et al*. 1990; Liscum *et al*. 1991).

Light is absorbed by plant photoreceptors, which transmit its signals to control most aspects in plant development. It is well known that cryptochromes (CRYs) and phytochromes (PHYs) are two main photoreceptor families in plant (Lariguet *et al*. 2005).

PHYs, firstly discovered in 1950s, can perceive R/FR light signals (Kendrick *et al*. 1994; Butler *et al*. 1959; Somers *et al*. 1998). Inactive (Pr) and active (Pfr) forms are different conformations of phytochrome, and absorb R light and FR light, respectively, to transform into each other (Sharrock *et al*. 2008). In *Arabidopsis*, phytochromes are encoded by a small gene family, designated PHYA to PHYE, among which PHYA, PHYB, and PHYC are conservative in most monocot species, and PHYD to PHYF genes are derived from PHYB in eudicot plant species (Sharrock *et al*. 2008; Mathews *et al*. 2010). It was shown that all the PHYs are involved in the

regulation of flowering time. PHYA, a light-labile type I phytochrome which is highly abundant in etiolated seedlings, plays a key role in flowering initiating as a daylength sensor in *Arabidopsis*, while PHYB members are light-stable type I phytochromes, which are mainly required for shade-avoidance responses (Li *et al*. 2011). FR light and R light can respectively induce nuclear translocation of PHYA-Pfr and PHYB-Pfr via distinct nuclear import mechanisms to trigger most light responses (Kircher *et al*. 1999; Hisada *et al*. 2000; Monte *et al*. 2003; Nagatani *et al*. 2004). PHYC appears to function as a flowering repressor under non-inductive photoperiods in both *Arabidopsis* and rice (Takano *et al*. 2005; Kevei *et al*. 2007).

In *Arabidopsis*, cryptochromes CRY1 and CRY2 absorb blue light to inhibit de-etiolation and regulate floral initiation controlled by photoperiod, respectively (Guo *et al*. 1998; El-Din El-Assal S *et al*. 2001). Both CRY1 and CRY2 stimulate the expression of miRNA172 to regulate flowering in a CONSTANS (CO)-independent manner (Jung *et al*. 2007). CRY2 physically interacts with a basic-helix-loop-helix 1 (bHLH) transcription factor, cryptochrome-interacting bHLH1 (CIB1) to regulate *FT* expression in a blue light-dependent manner (Liu *et al*. 2008b). Even though the protein expression level of CRY1 is not regulated by light, CRY1 can prevent COP1-SPA1 (CONSTITUTIVE PHOTOMORPHOGENIC 1-SUPPRESSOR OF PHYTOCHROME A) complex formation by interaction with SPA1 proteins via a blue light-dependent manner (Lian *et al*. 2011; Liu *et al*. 2011).

The interaction between PHYs and CRYs may also mediate cross-talk between the R/FR and blue/UV-sensing pathways, enabling fine-tuning of light responses to different spectral inputs (Hughes *et al*. 2012).

### 2.1.2 Photoperiodism and the circadian clock

As the earth is endless revolution and rotation, the daylength and darkness length vary with the season and latitude (Jackson 2009). The daylength is a major determinant factor rather than the light quantity in plant development (Kobayashi and Weigel 2007).

In 1920s, Garner and Allard firstly pointed out the relationship between the daylength and the flowering time (Garner and Allard 1920). A SD cultivar ‘Merryland Mammoth’ of tobacco and a SD species, soybean (cv. Biloxi) were used in their study, and it was found that the relative length of the day was an important factor in the growth and development of the plants, particularly flowering, and in some species it was found that the normal plant could attain flowering and fruiting stages only when the length of the day falls within certain limits (Garner and Allard 1920).

In 1936, Bünning proposed the hypothesis that the physiological clock recognized the cyclic change of light phase and dark phase, and these endogenous circadian rhythms were a polygenic trait to control photoperiodic responses such as flowering (Bünning 1936; Bünning 1960). This hypothesis was further supported by Hamner in later (Hamner 1960).

According to different photoperiod responses, there are three types of flowering plants: short-day plants in which flowering is induced when the darkness length is longer than the critical day length, such as *Glycine max*, *Oryza sativa*, *Chrysanthemum sp.*, *Pharbitis nil*; long-day plants in which flowering is induced when the darkness length is shorter than the critical day length, such as *Arabidopsis thaliana*, *Secale cereale*, *Triticum aestivum*; and day-neutral plants in which flowering is not related to photoperiod, such as *Heliantbus annuus*, *Pisum sativum*, *Zea mays* (Salisbury 1985; Hopkins 1999; Jackson 2009). Both short-day plants and long-day plants are photoperiod sensitive.

In 1938, Hamner and Bonner clarified that the uninterrupted darkness length is a key factor to induce flowering, which from the experiments with light break into dark period and night break into light period in *Xanthium strumarium* (Hamner and Bonner 1938). However, *X. strumarium* is a qualitative short-day species, in which flowering only happens under SD conditions (Hopkins 1999). For soybean, which is a quantitative short-day plant, it flower under both SD and LD conditions, but flowering occurs earlier under SD conditions than LD conditions (Roberts and Summerfield 1987; Hopkins 1999). Long-day plants also include qualitative long-day plant, such as

*Lolium temulentum* (Perilleux *et al*. 1994) and quantitative long day plant, such as wheat (*Triticum sp.*) (Hopkins 1999).

## 2.2 Photoperiodic flowering in long-day plant *Arabidopsis* and short-day plant rice

### 2.2.1 Photoperiodic flowering in *Arabidopsis*

Photoperiodic pathway in *Arabidopsis* includes a number of genes involved in circadian clock and light-signal transduction, among which four genes, *GIGANTEA* (*GI*), *FLAVIN KELCHF BOX 1* (*FKF1*), *CYCLING DOF FACTOR* (*CDF*) and *CONSTANS* (*CO*), have main function to regulate *FT* expression (Putterill *et al*. 1995; Fowler *et al*. 1999; Imaizumi *et al*. 2003; Sawa *et al*. 2007). A zinc finger transcriptional regulator CO, which contains two B boxes and a CCT domain, integrates circadian rhythm and light signal to directly interact with the *FT* promoter for activating *FT* transcription (Putterill *et al*. 1995; Imaizumi *et al*. 2003). NUCLEAR FACTOR Y (NF-Y) complex, bound at the *FT* distal enhancer element, helps recruit CO to proximal *cis*-regulatory elements in the *FT* promoter (Cao *et al*. 2014). CO-binding elements (CORE1 and CORE2) with the consensus sequence TGTG(NN or NNN)ATG also have been identified in the *FT* promoter (Tiwari *et al*. 2010). It was reported that the transcription of *FT* is regulated by both transcriptional and post-translational regulation of *CO* gene (Andrés and Coupland 2012). On the level of transcriptional regulation of *CO* gene, two circadian clock components *GI* and *FKF1* interactively increase *CO* transcription through triggering the degradation of CDF protein under LD conditions (Knott 1934; Sawa *et al*. 2007; Fornara *et al*. 2009; Andrés and Coupland 2012). CO protein is degraded by the 26S proteasome in the dark, which are mainly promoted by activity of SPA1-COP1 complex and by other CO-degradation genes *HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 1* (*HOS1*) and *PHYB* in the morning (Valverde *et al*. 2004; Laubinger *et al*. 2006; Sawa *et al*. 2007; Jang *et al*. 2008; Liu *et al*. 2008c; Fornara *et al*. 2009; Lazaro *et al*. 2012). At the end of day under LD conditions, PHYA and CRY2 can maintain

stabilization of CO protein. The combination of transcriptional and post-translational regulation causes a transient peak of CO protein to induce high *FT* expression at the end of the day under LD conditions (Andrés and Coupland 2012).

The *FT* promoter is comprised of a 5.7 kb region upstream of the start codon ATG, which contains all of the *cis*-elements to regulate *FT* expression (Adrian *et al*. 2010). GI, FKF1 and CDF1 were reported to affect expression of *FT* by regulating *CO* transcription, but can also directly bind to the *FT* locus (Sawa *et al*. 2007; Sawa *et al*. 2011; Song *et al*. 2012). In addition to the three genes, there are several genes which affected *FT* expression in responses to environmental stimulus, such as *PHYTOCHROME-INTERACTING FACTOR 4* (*PIF4*) in response to high temperature, *CIB1* in response to blue light, *SHORT VEGETATIVE PHASE* (*SVP*) and *FLOWERING LOCUS C* (*FLC*) in response to vernalization, *TEMPRANILLO* (*TEM*) genes binding directly to the 5′ UTR of *FT*, and SCHLAFMUTZE (SMZ) directly binding to a ~1.5 kb region downstream of the *FT* stop codon (Helliwell *et al*. 2006; Searle *et al*. 2006; Lee *et al*. 2007; Castillejo *et al*. 2008; Liu *et al*. 2008b; Mathieu *et al*. 2009; Kumar *et al*. 2012). LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), as a plant-specific component of polycomb repressive complex (PRC), together with histone H3 modified by trimethylation on lysine 27 (H3K27me3) widely cover the *FT* locus to make chromatin prone to access regulatory factors (Turck *et al*. 2007; Adrian *et al*. 2010).

### 2.2.2 Photoperiodic flowering in rice

In rice, photoperiodic flowering is regulated by two distinct pathways: the Hd1-Hd3a module, similar to the *Arabidopsis* CO-FT module and the unique Ghd7-Ehd1-Hd3a/RFT1 pathway (Izawa *et al*. 2002; Xue *et al*. 2008; Itoh *et al*. 2010; Shrestha *et al*. 2014).

In the Hd1-Hd3a module, light signals and the circadian clock regulate diurnal expression of an *Arabidopsis CO* ortholog, *Heading date 1* (*Hd1*), through the function of *OsGI*, which is an ortholog of *Arabidopsis GI* (Hayama *et al*. 2003). Hd1,

interacted with NF-Y, activates the expression of *Heading date 3a* (*Hd3a*), an ortholog of *Arabidopsis FT* gene in SD conditions but suppresses it under LD conditions. Under LD conditions, the activity of Hd1 can converts from activating to repressing *Hd3a* expression via R light signals absorbed by PHYB. *Heading date 6* is also involved in enhancing this repressive activity, which encodes the α-subunit of CK2 (Izawa *et al*. 2002; Ogiso *et al*. 2010; Ishikawa *et al*. 2011).

In the rice-specific Ghd7-Ehd1-Hd3a/RFT1 pathway, expression of *Grain yield and heading date* 7 (*Ghd7*) and *Early heading date 1* (*Ehd1*) is controlled by the circadian clock and light signaling. *Ehd1* activates expression of *Hd3a* and *RFT1*, independently of *Hd1* (Doi *et al*. 2004). *Ehd1* mediates *Hd3a* expression under both SD and LD conditions, but only activates *RFT1* expression in LD conditions (Song *et al*. 2014). Circadian-related *OsGI* regulates *Ehd1* expression as an activator and *Ghd7* negatively regulates *Ehd1* expression (Xue *et al*. 2008; Song *et al*. 2014). Phytochrome is also required for R light-dependent induction of *Ghd7* in this progress (Itoh *et al*. 2010).

### 2.3 Photoperiodic flowering in soybean

Soybean is a paleopolyploid plant species with a much greater proportion of gene duplicates (Force *et al*. 1999; Shoemaker *et al*. 2006). In soybean, about 75% genes are present in multiple copies, and about 50% of paralogs have differential expression patterns, which causes a more complex mechanism in gene expression and flowering regulation compared with *Arabidopsis* (Schmutz *et al*. 2010; Fan *et al*. 2014).

Studies on flowering in soybean have begun as early as in 1927, when Owen detected a major pair of genes controlling maturity and designated them as *E* and *e* (Owen 1927). So far ten major loci, *E1* to *E9* and *J*, are identified to control flowering time and maturity in soybean (Bernard 1971; Buzzell 1971; Buzzell and Voldeng 1980; McBlain and Bernard 1987; Ray *et al*. 1995; Bonato and Vello 1999; Cober and Voldeng 2001; Cober *et al*. 2010; Kong *et al*. 2014). The time to initiate flowering in

soybean is determined by interaction between these major genes and quantitative trait loci (QTLs), and between them and environments. After extensive research over the last decades, the molecular bases of maturity loci *E1-E4* were well understood, however, the detail of *E5* to *E9* and *J* is still unknown.

### 2.3.1 *E1*, which is the largest effect gene on flowering, and *E1-like* genes

The *E1* locus is the most important gene responsible for the variation in flowering time among soybean cultivars (Bernard 1971; Abe *et al*. 2003; Xia *et al*. 2012). Positional cloning suggested that E1 is a soybean-specific transcription factor as there is no distinct homolog of the *E1* family in the *Arabidopsis* and other plant species, and it contains a putative bipartite nuclear localization signal (NLS) and a region distantly related to B3 domain (Xia *et al*. 2012). *E1* gene plays a critical role in repressing flowering and delaying maturity in soybean under LD, and early-flowering phenotype was observed in cultivars with null alleles, such as *e1-fs* and *e1-nl* (Xia *et al*. 2012).

So far, seven different *E1* alleles have been identified: *E1*, *e1-as*, *e1-fs*, *e1-nl*, *e1-re*, *e1-p*, and *e1-b3a* (Xia *et al*. 2012; Tsubokura *et al*. 2014; Zhai *et al*. 2015). *e1-nl* and *e1-fs* are dysfunctional alleles; the former lacks the entire *E1* gene; and the latter produces a nonfunctional truncated protein due to a 1-bp deletion that results in a premature stop codon (Xia *et al*. 2012). The *e1-as* allele is a leaky allele and retains partial *E1* function (Xia *et al*. 2012). Mutations of both *e1-re* and *e1-p* occur outside the *E1* coding region: the *e1-re* allele has a long interspersed nuclear element (LINE) that is inserted in the *E1* promoter region; while *e1-p* has an altered 5′ upstream region, and both the effects of *e1-re* and *e1-p* on flowering time have not been clarified (Langewisch *et al*. 2014; Tsubokura *et al*. 2014); *e1-b3a* retains the intact bipartite NLS, but with only approximately half of the B3-like domain (Zhai *et al*. 2015).

The expression of the *E1* gene, which is promoted by LD and suppressed by SD,

negatively correlates with the expression of *GmFT2a* (*FT2a*) and *GmFT5a* (*FT5a*). These findings revealed that *E1* is a main factor in the phytochrome A signaling pathway and controls two functionally coordinated *GmFT* genes (*FT2a* and *FT5a*) (Xia *et al*. 2012). The diurnal expression pattern showed that *E1* is associated with the circadian clock to a small extent (Xia *et al*. 2012; Xu *et al*. 2015; Zhai *et al*. 2015).

Two *E1* homologs *E1La* (*Glyma.04G156400.1*) and *E1Lb* (*Glyma.04G143300.1*) showed an similar expression pattern as *E1*, and may also inhibit flowering by down-regulating the *FT2a* and *FT5a* genes, despite of their weaker effect on flowering compared with *E1* (Xu *et al*. 2015). The expression of *E1* and *E1Ls* is induced by the proper appearance time of the light signal, especially by a period of light before dusk of the previous day (Xu *et al*. 2015). Taken together, regulation of *E1* and *E1L* genes by light are crucial for the photoperiodic response mediated by PHYA in soybean (Xia *et al*. 2012; Xu *et al*. 2015).

### 2.3.2 *E2*, which is a *GIGANTEA* ortholog, and CO-FT pathways in soybean

*E2*, a soybean ortholog of *Arabidopsis GIGANTEA* (*GI*), has two alleles: a functional dominant allele (*E2*) and a recessive loss-of-function allele (*e2*) (Watanabe *et al*. 2011).

Using a residual heterozygous line-derived map-based cloning method, Watanabe *et al*. (2011) indentitied the candidate gene for *E2*, named also as *GmGIa* (*Glyma10g36600*). The loss of function of *GmGIa* caused early flowering probably through the up-regulation of *GmFT2a* expression; no difference was observed in the expression of *GmFT5a* between the *E2* and *e2* alleles (Watanabe *et al*. 2011). Another *E2* homolog, *GIb* has also been identified in the soybean genome (Watanabe *et al*. 2011). Li *et al*. (2013) found that, among the three homologs (*GmGI1*, *GmGI2* and *GmGI3*), *GmGI1* (*GmGIb*) has two alternative splicing forms, *GmGI1α* and *GmGI1β*. GmFKFs (GmFKF1 and GmFKF2) can interact with GmGI1α and GmGI2, but not with GmGI1β and GmGI3 (E2) (Li *et al*. 2013). The transgenic *Arabidopsis* plants over-expressing *GmGI1α* and *GmFKF2* produced more rosette leaves and exhibited

more vigorous growth and senescence retardation compared to the wild type plants in LD, suggesting their functions on vegetative growth (Li *et al*. 2013). The expression pattern of *GmGI2*, which is similar with *GmFKFs*, showed circadian rhythms (Li *et al*. 2013). They indicated that *GmGI2* may have a role in flowering via CO/FT model, which is different from the PHYA-E1 pathway, and *E2* (*GmGI3*) may activate *FT* expression directly binding with *FT* promoter region.

Fan *et al*. (2014) proposed two homologous pairs *GmCOL1* (*GmCOL1a*) / *GmCOL2* (*GmCOL1b*) and *GmCOL5* (*GmCOL2a*) / *GmCOL13* (*GmCOL2b*) were the good candidates of *CO* orthologs in the soybean, which was consistent with Jung *et al* (2007). Until now, 26 *CO* homologs were identified in soybean genome, in which *GmCOL1a* (*Glyma08g28370*) / *GmCOL1b* (*Glyma18g51320*) and *GmCOL2a* (*Glyma13g07030*) / *GmCOL2b* (*Glyma19g05170*) show the highest sequence homology to *Arabidopsis CO* gene (Wu *et al*. 2014). All of the four *CO* homologs fully complemented the late-flowering phenotype of the *Arabidopsis* co-1 mutant (Wu *et al*. 2014). *GmCOL1a* and *GmCOL1b* display strong photoperiod responsiveness and the expression level of *GmCOL1a* and *GmCOL1b* were clearly higher than *GmCOL2a* and *GmCOL2b* in soybean. For the circadian patterns, *GmCOL1a* and *GmCOL1b* have the highest expression level in the morning, while *GmCOL2a* and *GmCOL2b* appeared peak in the afternoon, and these patterns were consistent with peaks of *GmFTs* (Wu *et al*. 2014; Guo *et al*. 2015).

NF-Ys, composed of NF-YA, NF-YB and NF-YC proteins, have multiple functions including mediating floral promotion in CO/FT pathway to regulate the expression of *CO* in *Arabidopsis* (Calvenzani *et al*. 2012). The soybean *NF-Y* homologs *GmNF-YB02* (*Glyma02g46970.1*), *GmNF-YB13* (*Glyma08g44140.2*), *GmNF-YC04* (*Glyma06g17780.1*), *GmNF-YC06* (*Glyma08g17630.1*), and *GmNF-YC14* (*Glyma19g42460.1*) are closely similar to *Arabidopsis* and rice flowering genes *AtNF-YB2*, *AtNF-YB3*, *OsHAP3H*, *AtNF-YC3* and *AtNF-YC9*, respectively (Miyoshi *et al*. 2003; Kumimoto *et al*. 2008; Kumimoto *et al*. 2010). It is believed that these soybean *NF-Y* homologs have similar functions as *NF-Y* in

*Arabidopsis* and rice (Quach *et al*. 2015).

### 2.3.3 *E3, E4,* and other photoreceptor-related genes

Dominant alleles *E3* and *E4* encode GmPHYA3 and GmPHYA2, respectively, which are homologs of the photoreceptor PHYA in *Arabidopsis*, together with GmPHYA1 and GmPHYA4 (Liu *et al*. 2008a; Watanabe *et al*. 2009; Franklin and Quail 2010). GmPHYA1 most likely functions on photo-morphogenesis such as de-etiolation response and flowering under FR-enriched environments (Liu *et al*. 2008a), whereas GmPHYA4 is dysfunctional in the Williams 82 genome, although a recent pan-genome analysis suggests that the wild soybean possessed functional GmPHYA4 (Li *et al*. 2014).

*E3* locus includes functional dominant allele *E3* and three recessive alleles: *e3-ns* (nonsense mutation), *e3-tr* (deletion of a 13.33 kb region including the exon 4), and *e3-fs* (frameshift due to insertion of "T") (Xu *et al*. 2013). *E4* locus includes functional dominant allele *E4* and four dysfunctional alleles: *e4-SORE1*, *e4-kam*, *e4-oto*, and *e4-kes* (Liu *et al*. 2008a; Tsubokura *et al*. 2013).

*E3* and *E4* control the expression of *E1* and *E1L* genes to mediate different flowering responses to LD and SD conditions; *E3* is involved in the response to light with a high or low R:FR ratio, while *E4* is to light with a low R:FR ratio (Buzzell 1971; Buzzell and Voldeng 1980; Saindon *et al*. 1989; Cober *et al*. 1996; Xu *et al*. 2015). The recessive allele *e3* and *e4* have been reported to be responsible for photoperiod insensitivity in soybean (Cober *et al*. 1996). Further studies are needed to clarify how PHYAs mediate the induction of *E1* and *E1L*.

As blue light receptors, cryptochromes also contribute to regulate photoperiodic light responses in plants (Cashmore 2003; Sancar 2003; Chaves *et al*. 2011). In the soybean genome, there are at least six cryptochromes, including four CRY1s (CRY1a, CRY1b, CRY1c, and CRY1d) and two CRY2s (CRY2a and CRY2b) proteins (Zhang *et al*. 2008). The expression of the *GmCRY1a* was shown to exhibit

photoperiod-dependent circadian rhythm and correlated with latitudinal distribution of soybean cultivars (Zhang *et al*. 2008). *CRY1a* also can activate flowering in transgenic *Arabidopsis* (Zhang *et al*. 2008). CRY2a interacts with the bHLH transcription factor GmCIB1 (Glyma11g12450) to suppress its DNA binding activity and mainly mediate leaf senescence in soybean, and it is reported that CRY2a can be degraded by the ubiquitin-26S proteasome in response to blue light (Meng *et al*. 2013).

### 2.3.4 GmFT, a flowering integrator

So far, 10 *FT* orthologs have been identified in soybean: *GmFT1a*, *Glyma18g53680*), *GmFT1b* (*Glyma18g53690*), *GmFT2a* (*Glyma16g26660*), *GmFT2b* (*Glyma16g26690*), *GmFT3a* (*Glyma16g04840*), *GmFT3b* (*Glyma19g28390*), *GmFT5a* (*Glyma16g04830*), *GmFT5b* (*Glyma19g28400*), *GmFT4* (*Glyma08g47810*), and *GmFT6* (*Glyma08g47820*). *GmFT2a* and *GmFT5a* were shown to exhibit a high expression and circadian rhythm under floral inductive conditions, which is similar as *FT* in *Arabidopsis* (Kong *et al*. 2010). Fan *et al.* (2014) also gained 11 *FT-like* genes in the soybean genome, and divided them into two clades, one composing of *GmFTL1* to *GmFTL7* and the other including *GmTSF1* to *GmTSF4*. Compared with the previous results of Kong *et al.* (2010), *GmFTL1* to *GmFTL6* correspond to *GmFT3a*, *GmFT3b*, *GmFT2a*, *GmFT5a*, *GmFT2b* and *GmFT5b* respectively, while *GmFT7* did to *GmFT2c*, a paralogous gene of *GmFT2a* and *FT2b*, which possess only a part of phosphatidylethanolamine binding protein domain (PEBP) domain and lacks the N-terminal segment (Kong *et al*. 2010; Fan *et al.* 2014). *GmTSF1* to *GmTSF4* correspond to *GmFT1b*, *GmFT1a*, *GmFT6* and *GmFT4*, respectively, among them, *GmTSF1* and *GmTSF2* showed similar amino acid sequences with *Arabidopsis TSF*, whereas *GmTSF3* and *GmTSF4* displayed much similar sequences with *FT* rather than *TSF*, but they should belong to the *TSF* homologs according to the collinear relationship and ectopic expression in *Arabidopsis* which did not have flower-promoting phenotype under LD conditions, like as *TSF* in *Arabidopsis* (Fan *et*

*al*. 2014). Thus, now *GmFTL1* to *GmFTL6* (*GmFT3a*, *GmFT3b*, *GmFT2a*, *GmFT5a*, *GmFT2b* and *GmFT5b*) were considered as *FT* orthologs in soybean.

Expression analysis of *GmFTs* in different leaves of different cultivars with different maturities showed that all the *GmFTs* increased in pace with the growth days and, under SD, reached expression peaks at 20 days after emergence or the growing stage with the 3rd trifoliolates (Kong *et al*. 2010; Guo *et al*. 2015). Among them, *GmFT2a* exhibited the highest abundance at all stages in all cultivars. *In situ* hybridization using the *GmFT2a*- 3′ UTR specific probe also revealed the presence of *GmFT2a* transcripts in vascular tissues as well as meristems and primordial, which is different from the expression pattern of *Arabidopsis FT* (Sun *et al*. 2011). Ectopic expression confirmed both *GmFT2a* and *GmFT5a* could induce premature flowering, while *GmFT5a* had a more prominent effect than *GmFT2a* in *Arabidopsis* (Kong *et al*. 2010).

The sequence polymorphisms in the coding sequence and promoter of the *GmFT2a* were analyzed in cultivars of China and North America (Jiang *et al*. 2013). The results indicated that the *GmFT2a* coding sequence is highly conserved, only with one A/T SNP site located at position 48 in the coding region, and the diversity in flowering time and maturation time in soybeans is not affected by this polymorphism (Jiang *et al*. 2013). Different from the coding sequence, the promoter region of *GmFT2a* is highly polymorphic and exhibits 17 haplotypes with ten SNP sites and six indels, in which a SNP (S17) showed a relationship with the day to first flowering under SD and ND conditions while two SNPs (S162 and S1849) showed such a relationship only under SD conditions (Jiang *et al*. 2013). A CIACADIANLELHC element (CAANNNNATC) and an IBOXCORENT element (GATAAGR) were identified near SNP-S17 (G/A) and SNP-S1849 (T/C), respectively. Both of elements may be related to photoperiod responses in soybean because the CIACADIANLELHC element is associated with circadian expression and the IBOXCORENT element is associated with light-responsive regulation (Jiang *et al*. 2013). Further studies on the function of these SNPs, especially SNP-S17, SNPs-S162

and S1849 will aid to understand the regulation of *FT2a* gene in the photoperiod pathway of soybean, despite of polymorphism of *GmFT2a* seems not to be responsible for maturity diversity (Jiang *et al*. 2013).

*GmFT2a* and *GmFT5a* genes exhibited a diurnal circadian rhythm under SD conditions, suggesting that their expression was partly regulated by circadian clock genes (Kong *et al*. 2010; Guo *et al.* 2015). The expression patterns of *GmFT2a* and *GmFT5a* in Harosoy, and its near isogenic lines (NILs) for *e3* and *e4*, *He3e4*, showed that these two genes was mainly controlled by the PHYA genes (*GmPHYA2* and *GmPHYA3*) under LD conditions (Kong *et al*. 2010). The SD-to-LD transfer experiment further demonstrated that the expression of *GmFT2a* is strictly regulated by photoperiod; while the expression of *GmFT5a* could also be detected albeit at low levels after the plants were transferred to LD conditions (Kong *et al*. 2010). Therefore, the expression of *GmFT2a* and *GmFT5a* were regulated by the PHYA-mediated photoperiodic regulation pathway, and the *GmFT5a* expression was also regulated by a photoperiod-independent system in LD conditions. *GmFT2a* and *GmFT5a* may coordinately control flowering and enable the adaptation of soybean to a wide range growth and productivity adaptively in soybean (Kong *et al*. 2010).

Different researchers use different naming systems for orthologs of *Arabidopsis* flowering genes, particularly for *GmFTs*, *GmCOs*, and *GmGIs*. Table 2.1 shows the correspondence among different naming systems. In this thesis, I used the *GmFTs* names consistent with Kong *et al*. (2010).

**Table 2.1 Correspondences between different naming systems for *GmFTs*, *GmCOs*, and *GmGIs*.**

| Gene name | Locus name | Names of homologs used in different references | |
|---|---|---|---|
| | | Watanabe *et al* . (2011) | Li *et al* . (2013) |
| *GmGIs* | Glyma20g30980 | *GmGIb* | *GmGI1* |
| | Glyma09g07240 | | *GmGI2* |
| | Glyma10g36600 | *GmGIa*（*E2*） | *GmGI3* |
| | | | |
| | | Fan *et al* . (2014) | Wu *et al* . (2014) |
| *GmCOs* | Glyma08g28370 | *GmCOL1* | *GmCOL1a* |
| | Glyma18g51320 | *GmCOL2* | *GmCOL1b* |
| | Glyma13g07030 | *GmCOL5* | *GmCOL2a* |
| | Glyma19g05170 | *GmCOL13* | *GmCOL2b* |
| | | | |
| | | Kong *et al* . (2010) | Fan *et al* . (2014) |
| | Glyma18g53680 | *GmFT1a* | *GmTSF2* |
| | Glyma18g53690 | *GmFT1b* | *GmTSF1* |
| | Glyma16g26660 | *GmFT2a* | *GmFTL3* |
| | Glyma16g26690 | *GmFT2b* | *GmFTL5* |
| *GmFTs* | Glyma16g04840 | *GmFT3a* | *GmFTL1* |
| | Glyma19g28390 | *GmFT3b* | *GmFTL2* |
| | Glyma16g04830 | *GmFT5a* | *GmFTL4* |
| | Glyma19g28400 | *GmFT5b* | *GmFTL6* |
| | Glyma08g47810 | *GmFT4* | *GmTSF4* |
| | Glyma08g47820 | *GmFT6* | *GmTSF3* |
| | | | *GmFT7* |

# Chapter Ⅲ *E9* gene identified in a cross between Harosoy and Toyomusume

## 3.1 Background and purpose

Understanding the molecular mechanisms of flowering and maturity is important for improving the adaptability and yield of seed crops in different environments. In soybean, a facultative short-day plant, genetic variation at four maturity genes, *E1* to *E4*, play an important role in adaptation to environments with different photoperiods. Among the four genes, *E1* has the largest effect on flowering times in LD conditions; it is a repressor for *FT* orthologs, and the locus consists of multiple alleles with different functions (Xia *et al*. 2012). The *e1-as* allele, which has been named as *e1*, is a leaky allele, which retains the E1 function partially, whereas the *e1-fs* and *e1-nl* alleles are dysfunctional due to a frame-shift by a deletion for the former and a deletion of the entire *E1* gene for the latter (Xia *et al*. 2012). Accordingly, the *e1-nl* and *e1-fs* alleles are expected to confer an early-flowering phenotype, when compared with the *e1-as* allele (Xia *et al*. 2012). However, Tsubokura *et al*. (2014) found that a Japanese cultivar Toyomusume having *e1-nl* flowered almost the same time as or slightly later than a Canadian cultivar Harosoy having *e1-as* in each of three different sowing conditions tested. The two cultivars have the same maturity genotype at *E2*, *E3* and *E4* (*e2*/*e2 E3*/*E3 E4*/*E4*), suggesting that an unknown genetic factor(s) different from the four maturity genes may counteract the early-flowering that was conditioned by the *e1-nl* allele in Toyomusume.

In this chapter, I carried out a genetic study of flowering genes segregated in the cross between Harosoy and Toyomusume. The purpose was to identify a flowering gene (s) which segregated independently of the *E1* locus.

## 3.2 Materials and methods

### 3.2.1 Plant materials and segregation analysis

$F_2$ and $F_3$ populations were developed from the cross between Harosoy (L58-266; HA) and Toyomusume (TO). Seeds of the $F_2$ population (n = 82) and both parents (n = 10) were sown in paper pots on 25 May 2012, and 10 days later seedlings were transplanted into soil at an experimental farm of Hokkaido University, Sapporo (43°07′N, 141°35′E). The 82 $F_2$ plants were genotyped with a DNA marker at the *E1* locus (Tsubokura *et al*. 2014) and its flanking SSR marker (Xia *et al*. 2012), and 16 plants homozygous for *e1-nl* and 16 plants homozygous for *e1-as* were selected for the progeny test. Seeds of each $F_2$ plant were sown on 25 May 2013, and 10 days later 20 seedlings 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.

### 3.2.2 DNA marker analysis in association tests

Total DNA was extracted from trifoliate leaves as described (Michaels *et al*. 2003) and from seeds as described (Xia *et al*. 2012). DNA marker analysis for the *E1* genotype followed with the method developed by Xu *et al.* (2013). Sixty-one SSR markers mapped on the consensus map (SOYBASE; http://soybase.org/sbt/) (Cregan *et al*. 1999; Song *et al*. 2010) and located in the genomic regions where orthologs to *Arabidopsis* flowering genes are clustered (Watanabe *et al*. 2012) were chosen for tests of their association with flowering time. Each PCR for SSR marker analyses contained 30 ng of total genomic DNA as template, 1 μl of each primer (10 μM) and dNTP (2.5 mM), 0.5 μl of ExTaq polymerase, and 2.5 μl of 10 × ExTaq buffer (Takara, Otsu, Japan) in a total volume of 25 μl; amplification conditions were 35 cycles at 94 ℃ for 30 s, 56 ℃ to 60 ℃ (depending on the primers used) for 30 s, and 72 ℃ for 30 s to 90 s. PCR products were separated by electrophoresis in 10.5% (w/v) polyacrylamide gels, stained with ethidium bromide, and visualized under UV light. Association tests were performed with one-way analysis of variance.

## 3.3 Results and discussion

### 3.3.1 Segregation of flowering time in $F_2$ and $F_3$ populations

Flowering times of Toyomusume (TO) and Harosoy (HA) were 54.4 and 56.7 days after sowing (DAS), respectively, in 2012, and 54.5 and 58.4 DAS, respectively, in 2013. HA thus flowered 2 to 4 days later than TO. Despite of similar flowering times for parents, flowering times in the $F_2$ population varied widely (46–67 DAS; Figure 3.1A). Since the allelic variation at *E1* has a large effect on flowering time, the effects of *E1* alleles on flowering time were first evaluated in the population. The *E1* genotypes of $F_2$ plants were determined with an allele-specific DNA marker (Xu *et al*. 2013) and flanking simple sequence repeat (SSR) markers (Xia *et al*. 2012). As expected, plants homozygous for *e1-nl* (from TO) flowered, on average, 11 days earlier than those homozygous for *e1-as* (from HA) (Figure 3.1A). Since plants homozygous for each allele still varied considerably in flowering time, the progeny test for 16 plants homozygous was carried out for each allele. Flowering times of $F_2$ individuals were closely correlated with the average flowering times of their progeny (Figure 3.1B). Parent–offspring correlation coefficients were 0.676 for the *e1-nl* homozygote and 0.823 for the *e1-as* homozygote, suggesting that a genetic factor(s) other than *E1* segregated in each of the two genotypic classes.

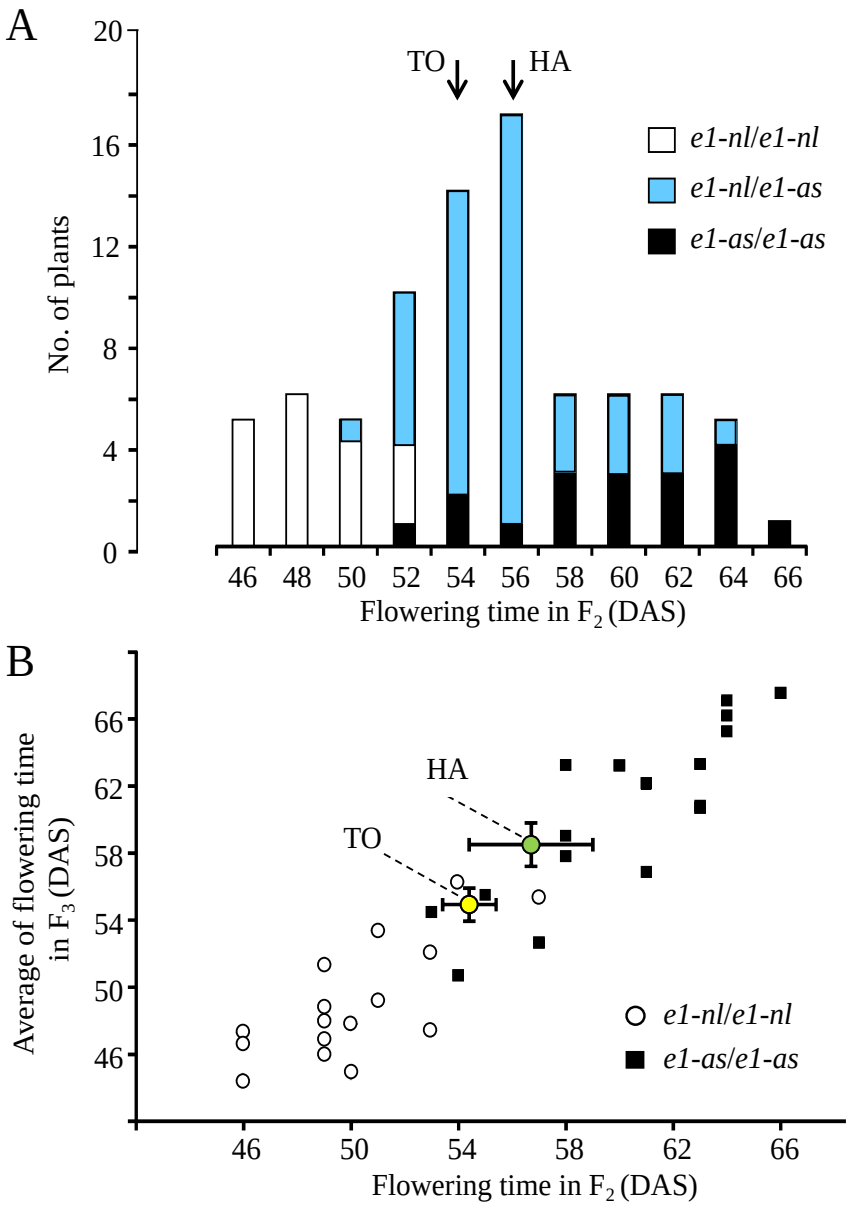



**Figure 3.1 Flowering time in the progeny of the cross between Toyomusume and Harosoy.** (A) Frequency distribution of flowering time in $F_2$. (B) Scatter diagram of flowering time in $F_2$ and $F_3$ progeny. Averages and standard deviations of flowering time for Toyomusume (TO) and Harosoy (HA) are shown. Parental data are means ± SD of ten independent plants.

### 3.3.2 Test for association between flowering time and SSR markers

To detect flowering genes that segregated independently of *E1*, flowering time and SSR marker association was tested in each of the *e1-nl* and *e1-as* genotypic classes; 61 SSR markers were also used, which located in the genomic regions where orthologs to *Arabidopsis* flowering genes are clustered (Watanabe *et al*. 2012). Two

markers were significantly associated with flowering time in *e1-nl* homozygotes and five in *e1-as* homozygotes (Table 3.1). Plants homozygous for the TO alleles (A) at all loci except Sat235 flowered later than those homozygous for the HA alleles (B). Only Sat_350 showed significant associations in both *e1-nl* and *e1-as* genotypic classes. Sat_350 was located near the SSR marker Satt686 on LG J, which is a tagging marker for the *E9* gene identified in a cross between cultivated (TK780) and wild (Hidaka 4) soybeans (Kong *et al*. 2014). Because TO is a parent of TK780 (Tanaka *et al*. 2003), which carries the recessive *e9* allele (Kong *et al*. 2014), it is plausible that the gene tagged by Sat_350 is identical to *E9* and that TO has the same recessive allele for late flowering as TK780.

**Table 3.1 Association tests of SSR marker genotypes with flowering time.** 16 plants homozygous for *e1-nl* and 16 plants homozygous for *e1-as* were used in the association tests. A and B indicate the alleles from Toyomusume and Harosoy, respectively. LG, linkage group

| | | Average | | | One-way ANOVA | |
|---|---|---|---|---|---|---|
| Marker | LG | A | H | B | F | *p* |
| | | | | | | |
| *e1-nl* | | | | | | |
| Satt681 | C2 | 16.7 | 11.4 | 12.3 | 5.65 | 0.017 |
| Sat_036 | D1a | 18.5 | 13.3 | 10.7 | 4.97 | 0.025 |
| Sat_350 | J | 20.1 | 12.7 | 8.6 | 14.84 | < 0.001 |
| | | | | | | |
| *e1-as* | | | | | | |
| Satt519 | B1 | 27.4 | 25.9 | 18.8 | 7.40 | 0.007 |
| Satt583 | B1 | 28.2 | 23.2 | 20.0 | 4.52 | 0.032 |
| Satt195 | C1 | 21.7 | 22.9 | 29.0 | 3.88 | 0.048 |
| Sat_235 | C1 | 14.7 | 23.0 | 28.9 | 6.99 | 0.009 |
| Sat_336 | C2 | 19.7 | 22.1 | 28.0 | 4.94 | 0.025 |
| Satt031 | D2 | 27.5 | 25.3 | 18.3 | 6.04 | 0.014 |
| Satt146 | F | 27.5 | 27.7 | 19.2 | 19.41 | < 0.001 |
| Sat_350 | J | 27.6 | 24.9 | 18.7 | 9.26 | 0.003 |
| | | | | | | |

# Chapter Ⅳ Fine-mapping and identification of responsible gene for the *E9* gene

## 4.1 Background and purpose

Results obtained in the previous chapter revealed that $F_2$ and $F_3$ populations of the cross between TO and HA segregated for the *E1* and *E9* loci. The allelic effects at the *E9* locus on flowering times were detected in both *e1-nl* and *e1-as* genotypic classes. Different allelic combinations at the two loci most likely made TO and HA flowered almost the same time, and produced a transgressive segregation in the progeny. The *E9* locus was fine mapped to a 245-kb region on LG J (Gm16), where two soybean orthologs of *FT*, *GmFT2a* and *GmFT2b*, located (Kong *et al*. 2014). Kong *et al*. (2014) detected the allelic effects on flowering time at the *E9* locus only under the *e1-nl* RILs of the cross between a breeding line TK780 and a wild accession Hidaka 4, but not under the *E1* RILs. Accordingly, the genetic background appears to be important to identify the hidden genetic factors for flowering, such as the *E9*/*e9* gene, the effects of which are generally abolished by the *E1* gene with marked effects on flowering (Kong *et al*. 2014).

In this chapter, I performed fine-mapping, expression and sequencing analyses to identify a molecular basis of *E9* and the potential factor(s) that results in the recessive *e9* allele.

## 4.2 Materials and methods

### 4.2.1 Plant materials for fine-mapping, expression and sequencing analyses

In order to fine-map the *E9* locus, I selected two heterozygous $F_3$ plants derived from the same $F_2$ family (#41) of the cross between TO and HA. A total of 300 seeds were genotyped for SSR and indel markers flanking the *E9* locus. Three indel markers were used; these markers were developed by Kong *et al*. (2014), and mapped in a

flanking region of *E9*. Eight recombinants between markers were detected; these were cultivated in a glasshouse during winter, and the seeds produced were used for the progeny test during summer (sowing date: 15 May) in 2014.

Four sets of $F_6$ near-isogenic lines (NILs) were developed for the *E9* gene through repetitive heterozygous selections: two sets from the cross between TO and HA and two from the cross between TO and a Japanese cultivar Hayahikari (HY). HY is a photoperiod-insensitive cultivar bred in Hokkaido, whose maturity genotype is *E1*/*E1*, *e2*/*e2*, *e3*/*e3*, and *e4*/*e4*. All NILs developed had the same genotype (*e1-nl*/*e1-nl*, *e2*/*e2*, *E3*/*E3*, and *E4*/*E4*) as TO. The three cultivars (TO, HA and HY), and the parents of RILs were used for identification of the *E9* gene (Liu *et al*. 2007; Kong *et al*. 2014), TK780 and Hidaka 4 were used for sequencing analyses. Five photoperiod-insensitive accessions (Karafuto 1, Gokuwase-Kamishunbetsu, Nawiko, Heihe 13, and Kitamusume) were also used for expression analyses.

### 4.2.2 Genotyping for DNA markers and maturity loci

The procedure of SSR analyses followed the method in the previous chapter. Briefly, each PCR contained 30 ng of total genomic DNA as template, 1 μl of each primer (10 μM) and dNTP (2.5 mM), 0.5 μl of ExTaq polymerase, and 2.5 μl of 10× ExTaq buffer in a total volume of 25 μl; amplification conditions were 35 cycles at 94 ℃ for 30 s, 48 ℃ to 60 ℃ (depending on the primers used) for 30 s, and 72 ℃ for 30–90 s. PCR products were separated by electrophoresis in 10.5% (w/v) polyacrylamide gels, stained with ethidium bromide, and visualized under UV light.

DNA markers were developed to detect a 10-bp deletion in the 5′ UTR and the insertion of SNP #17 and *SORE-1* in the first intron. For the 10-bp deletion, the primers 5′-GGAATCGAGGCTATTGACTA-3′ and 5′-CTTCCACTAGGCATGGGATA-3′ were used. For *SORE-1*, two forward primers, 5′-GCTCTCTCTCTTCCACTCTCTAGATGG-3′ (in the long terminal repeat [LTR] of *SORE-1*) and 5′-ACCCTCTCAAGTGGACATGT-3′ (in the first *FT2a* intron), and the common reverse primer 5′-CTAGGTGCATCGGGATCAAC-3′ (in the second

FT2a exon) were used. To identify the SNP, a dCAPS marker was developed: PCR was performed with the primers 5′-TTCAAACAATCTCATAATTATGAGT-3′ and 5′-TAATAGTAGTATGGATGGTCAAA-3′, and the amplified products were digested with *Hin*fI. The PCR and detection of amplified fragments were performed as described above. The genotyping for the *E1*, *E3*, and *E4* loci was performed using allele-specific DNA markers as described (Xu *et al*. 2013; Tsubokura *et al*. 2014). Primers, PCR conditions and expected fragment sizes are presented in Additional table 3.

### 4.2.3 qRT-PCR

Expression analyses were carried out in growth cabinets in which the daylength was set 12 h (SD) or 18 h (LD) at an air temperature of 24 ℃. Fully developed trifoliate leaves of four 20-day-old plants were sampled as a bulk at Zeitgeber time 3, immediately frozen in liquid $N_2$, and stored at −80°C. Total RNA was isolated from frozen tissues by lithium chloride precipitation according to Napoli *et al*. (1990), except that DNase I was used to remove genomic DNA. cDNA was synthesized from 1 μg of total RNA using an oligo (dT) 18 primer or random primer cocktail according to Dwiyanti *et al*. (2011). *FT2a* transcript levels were determined by semi-quantitative RT-PCR or quantitative real-time PCR (qRT-PCR). The qRT-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 ExTaq Perfect Real Time. A CFX96 Real-Time System was used. The PCR cycling conditions were 95 ℃ for 3 min followed by 35 cycles of 95 ℃ for 10 s, 60 ℃ for 30 s, 72 ℃ for 20 s and 78 ℃ for 2 s. Fluorescence was quantified before and after the incubation at 78 ℃ to monitor the formation of primer dimers. The mRNA for *β-tubulin* was used as a control. 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 Additional table 2.

### 4.2.4 Sequencing analysis of *FT2a* and *SORE-1*

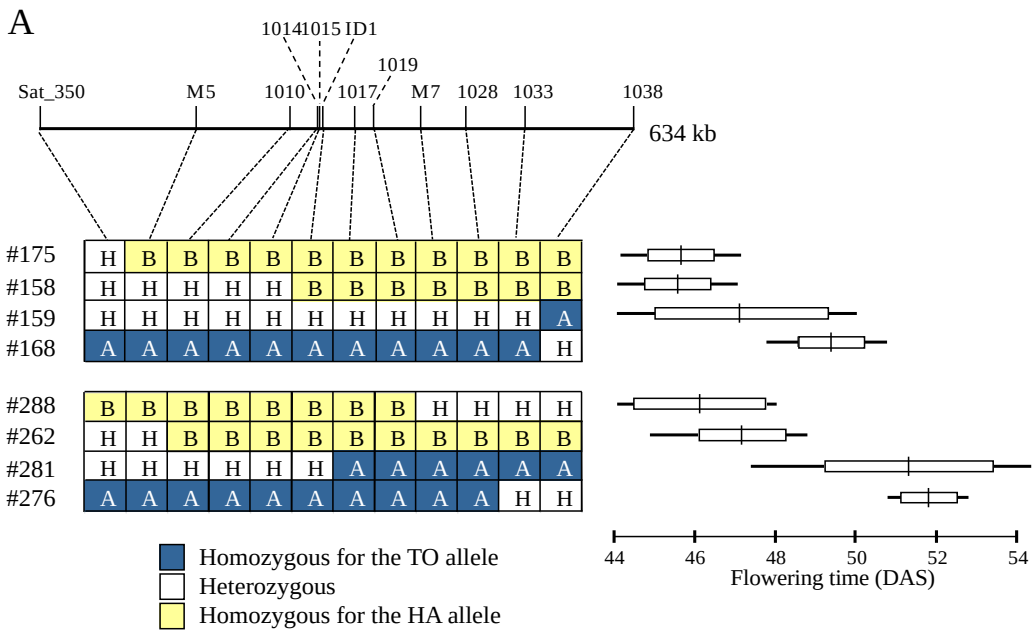
cDNAs from the three cultivars, TK780 and Hidaka 4 were used to sequence the *FT2a* coding regions. Each of the two *FT2a* genomic regions (the 5′ upstream region and the region to 3′ downstream region) was divided into three parts, which were amplified from total DNA with KOD FX polymerase (Toyobo Life Science, Osaka, Japan) and sequenced. Genome walking with a BD Genome-Walker Universal kit (Takara Clontech, Otsu, Japan) was used to sequence the first intron of TO, in which *SORE-1* was inserted. According to the manufacturer's instructions, we constructed four kinds of genomic libraries by digesting total DNA from TO in separate reactions with four blunt-end endonucleases (*Dra*I, *Eco*RV, *Puv*II, and *Stu*I) and ligating the ends of the digested DNA to an adaptor sequence. Nested PCR was performed for each library using adaptor primers and gene-specific primers. The inserted *SORE-1* was then amplified with the forward primer in intron 1 and the reverse primer in intron 2, and the resultant amplicon was used for PCR amplification of each of five divided regions of *SORE-1* to obtain the whole sequence (Additional table 2). The amplified fragments were sequenced directly or were first cloned into a pGEM-T Easy vector (Promega, Madison, WI, USA) and then sequenced. Sequence analysis was performed by using a BigDye Terminator v. 3.1 Cycle Sequencing kit and an ABI PRISM 3100 Avant Genetic Analyzer (Applied Biosystems Japan, Tokyo, Japan) according to the manufacturer's instructions. A BLAST search of the NCBI genome database and PLACE (Higo *et al.*1999) analysis were carried out to detect sequences homologous to the fragment identified by genome-walking and possible *cis*-elements in the first intron of *FT2a*. Primer sequences used in genome sequencing are listed in Additional table 2.

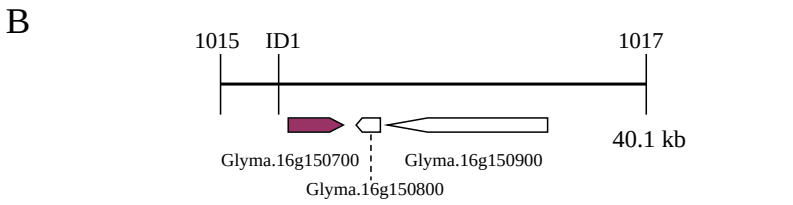
## 4.4 Results and discussion

### 4.4.1 Fine-mapping and association analysis

For fine-mapping of the *E9* gene, a total of 300 seeds from two heterozygous $F_3$ plants derived from the same $F_2$ family (#41) were genotyped for the SSR markers Sat_350 and BARCSOYSSR_16_1038. Eight recombinants (four progenies from each of two heterozygous $F_3$ plants) were detected in the flanking region, which were genotyped for seven additional SSR markers and three insertion/deletion (indel) markers (ID1, M5, and M7) used in the identification of *E9* (Kong *et al*. 2014). The genotype at *E9* was estimated from the segregation pattern in the progeny test (Figure 4.1A). Among the four plants derived from one $F_3$ parent, two plants (#158 and #175) flowered early and one (#168) flowered late, whereas plant #159 segregated for flowering time. Among the four plants derived from the other $F_3$ parent, two plants (#262 and #288) flowered early and one (#276) flowered late, whereas one plant (#281) segregated. By comparing the graphical genotypes and estimated *E9* genotypes, the QTL to a 40.1-kb region was delimited between markers BARCSOYSSR_16_1015 and BARCSOYSSR_16_1017, in which only the ID1 marker completely co-segregated with the genotype at *E9*.

To confirm co-segregation between flowering time and ID1 genotype, 14 $F_2$ families were examined homozygous for *e1-nl* and 14 homozygous for *e1-as* (Table 4.1). Among the *e1-nl* families, plants of two families homozygous for the TO allele flowered late, whereas plants of two families homozygous for the HA allele flowered early. A highly significant association between flowering time and marker genotypes was observed in the 10 heterozygous families. Similarly, a highly significant association was detected between flowering time and marker genotypes in the 5 heterozygous families with the *e1-as* genotype. Therefore, the variation in flowering time in each $F_2$ family could be mostly accounted for by the genotypes at the ID1 marker.




Glyma.16g150700; *FT2a*
Glyma.16g150800; EXOCYST COMPLEX PROTEIN EXO70
Glyma.16g150900; TATD FAMILY DEOXYRIBONUCLEASE

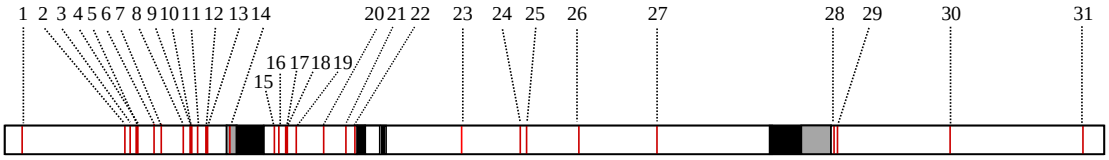
**Figure 4.1 Fine mapping of the *E9* locus and annotated genes in the delimited genomic region.** (A) Eight recombinants (four from each of two $F_3$ heterozygous plants) in the region between Sat_350 and BARCSOYSSR_16_1038 were genotyped at 7 BARCSOYSSR (1010 to 1033) and 3 indel markers (bold). The genotype at *E9* was estimated by progeny testing. The ranges (horizontal lines), averages (vertical lines), and standard deviations (open boxes) of flowering time (DAS: days after sowing) are indicated. (B) Three annotated genes in a delimited genomic region.

**Table 4.1 Association tests of ID1, a tagging marker of *E9*, with flowering.** The progeny of 14 plants homozygous for *e1-nl* and 14 plants homozygous for *e1-as* were used in the association tests. A and B indicate the alleles from Toyomusume and Harosoy, respectively.

| $F_2$ | Average (SD) of flowering time (DAS) in $F_3$ | | | One-way ANOVA | |
|---|---|---|---|---|---|
| Plant number | AA | AB | BB | F value | Probability ($10^{-3}$) |
| | | | | | |
| $F_2$ families with *e1-nl*/*e1-nl* | | | | | |
| | | | | | |
| #34 | | | 44.5 (1.1) | | |
| #66 | | | 44.9 (0.8) | | |
| | | | | | |
| #02 | 51.8 (2.2) | 44.6 (1.4) | 44.3 (0.5) | 27.2 | 0.005 |
| #05 | 53.0 (1.7) | 46.3 (1.4) | 43.0 (0.8) | 45.6 | 0.000 |
| #25 | 52.8 (2.8) | 47.7 (1.9) | 44.8 (1.2) | 22.1 | 0.018 |
| #27 | 52.6 (2.1) | 45.5 (0.8) | 44.0 (0.0) | 73.2 | 0.000 |
| #28 | 56.0 (1.7) | 52.1 (1.2) | 49.5 (2.4) | 14.4 | 0.223 |
| #41 | 54.0 (2.0) | 46.2 (1.7) | 44.2 (1.6) | 40.7 | 0.000 |
| #50 | 56.0 (1.0) | 53.6 (2.1) | 50.4 (2.9) | 8.8 | 2.324 |
| #79 | 56.0 (1.4) | 47.8 (1.8) | 45.4 (1.3) | 61.3 | 0.000 |
| #81 | 55.3 (1.6) | 46.3 (1.0) | 44.7 (0.5) | 156.8 | 0.000 |
| #82 | 56.4 (0.5) | 52.1 (1.0) | 46.9 (3.7) | 26.2 | 0.006 |
| | | | | | |
| #18 | 55.4 (1.5) | | | | |
| #46 | 56.4 (1.5) | | | | |
| | | | | | |
| $F_2$ families with *e1-as*/*e1-as* | | | | | |
| | | | | | |
| #12 | | | 60.7 (1.9) | | |
| #29 | | | 50.7 (1.5) | | |
| #30 | | | 52.7 (2.9) | | |
| #43 | | | 54.5 (2.4) | | |
| | | | | | |
| #22 | 64.8 (1.3) | 57.5 (1.6) | 52.0 (2.8) | 92.0 | 0.000 |
| #36 | 65.7 (0.6) | 57.3 (3.2) | 50.4 (3.8) | 57.9 | 0.020 |
| #48 | 65.7 (1.6) | 59.5 (2.7) | 55.0 (1.7) | 63.7 | 0.005 |
| #69 | 66.7 (1.4) | 62.5 (3.8) | 59.3 (2.1) | 42.6 | 7.615 |
| #73 | 65.4 (1.5) | 55.6 (2.7) | 55.0 (4.0) | 58.9 | 0.015 |
| | | | | | |
| #13 | 66.2 (0.8) | | | | |
| #16 | 67.4 (0.6) | | | | |
| #33 | 63.2 (3.2) | | | | |
| #76 | 63.3 (1.9) | | | | |
| #78 | 65.9 (0.8) | | | | |
| | | | | | |

### 4.4.2 cDNA sequencing and expression analysis

According to the Williams 82 reference genome sequence (Schmutz *et al*. 2010), the region delimited by fine mapping contained three genes: *Glyma16g150700* (*FT2a*), *Glyma16g150800* (*EXOCYST COMPLEX PROTEIN EXO70*), and *Glyma16g150900* (*TATD FAMILY DEOXYRIBONUCLEASE*) (Figure 4.1B). *FT2a* was focused on as a candidate for *E9* because of its importance in floral induction in soybean (Kong *et al*. 2010; Sun *et al*. 2011; Thakare *et al*. 2011; Jiang *et al*. 2013; Fan *et al*. 2014; Nan *et al*. 2014; Wang *et al*. 2015; Xu *et al*. 2015). cDNA sequence analysis was carried out for HA and TO, the Japanese cultivar Hayahikari (HY), and the parents (TK780 and Hidaka 4) of the RIL population used for the identification of *E9* (Kong *et al*. 2014). There were no nucleotide substitutions in their coding regions, which were identical to that of Williams 82; a SNP (#28; Figure 4.2) after the stop codon was identified between HA and TO or HY. Then the expression profiles of *FT2a* were compared under SD and LD conditions in plants homozygous for the TO allele and those homozygous for the HA allele at ID1 in the progeny of 10 $F_2$ families with the *e1-nl*/*e1-nl* genotype that segregated for *E9*. In all tested families, plants with the HA allele had higher *FT2a* expression than plants with the TO allele, regardless of daylength, although the expression was much higher in SD than LD in both homozygotes (Figure 4.3). Thus, late flowering in plants homozygous for the TO allele at ID1 was tightly associated with reduced *FT2a* expression.




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Promoter | | | | | | | | | | | | | 5' UTR |
| | -1544 | -772 | -731 | -698 | -694 | -571 | -544 | -402 | -379 | -375 | -349 | -279 | -274 | -47 |
| Toyomusume | A | T | 43 bp | A | T | G | AT | G | G | G | T | - | - | - |
| TK780 | A | T | 43 bp | A | T | G | AT | G | G | G | T | - | - | - |
| Hayahikari | T | T | 43 bp | A | T | G | AT | G | G | G | T | - | GATC | 10 bp |
| Hidaka 4 | A | A | - | T | - | T | AT | A | G | G | C | - | - | 10 bp |
| Harosoy | A | A | - | T | T | G | - | G | A | A | T | G | - | 10 bp |

| | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|
| | Intron I | | | | | | | |
| | 294 | 315 | 398 | 403 | 475 | 710 | 965 | 988 |
| Toyomusume | G | C | C | T | T | T | SORE-1 | T |
| Hayahikari | G | C | A | T | C | C | - | C |
| Harosoy | T | T | A | A | T | T | - | T |

| | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
|---|---|---|---|---|---|---|---|---|---|
| | Intron II | | | | | 3' downstream region | | | |
| | 2150 | 2604 | 2648 | 3022 | 3783 | 4606 | 4628 | 5300 | 6630 |
| Toyomusume | $(C)_{10}$ A $(C)_{8}$ $(A)_{6}$ | $(T)_{11}$ T | $(AT)_{8-11}$ | $(A)_{10}$ | - | A | $(A)_{6}$ | G | A |
| Hayahikari | $(C)_{10}$ A $(C)_{8}$ $(A)_{6}$ | $(T)_{10}$ A | $(AT)_{18}$ | $(A)_{10}$ | 9 bp | A | $(A)_{7}$ | G | G |
| Harosoy | $(C)_{16}$ $(A)_{9}$ | $(T)_{11}$ A | $(AT)_{27-30}$ | $(A)_{9}$ | - | G | $(A)_{6}$ | T | A |

**Figure 4.2 DNA polymorphisms detected in the *FT2a* genomic region for four soybean cultivars or breeding line and a wild soybean accession.**

43 bp-indel CCCCTGCTAATCTTCTTCCTCCTCACACATTGCTTCTCACCCC

10 bp-indel GAAAGCATAA

*SORE-1* Ty1/copia-like retrotransposon (6,224 bp)

9 bp-indel AAAGAAAAA

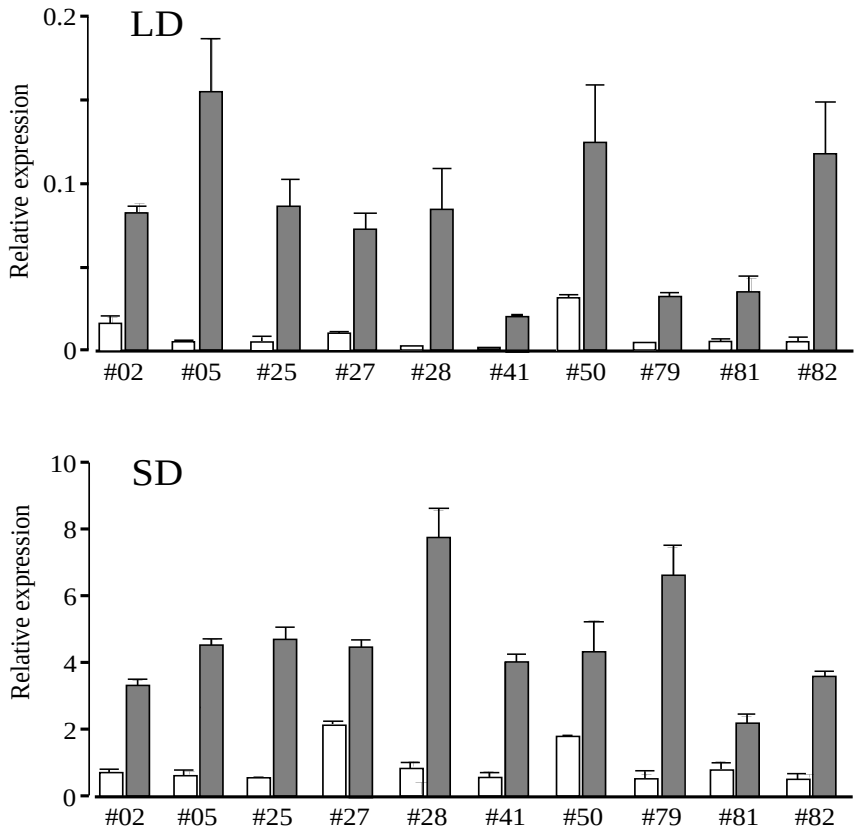



**Figure 4.3 *FT2a* expression in the progeny of $F_2$ plants from a cross between Toyomusume and Harosoy.** Four plants from the progeny of each $F_2$ plant, which were homozygous for the Toyomusume allele (white bars) or the Harosoy allele (gray bars) at the ID1 tagging marker for *FT2a*, were used. Relative mRNA levels are expressed as the ratios to *β-tubulin* transcript levels.

### 4.4.3 Sequence analysis of the *FT2a* genomic region

In *Arabidopsis*, *FT* is regulated by various transcription factors, which bind to the promoter or to the first intron and 3′ downstream region (Andrés *et al*. 2012; Pin and Nilsson 2012; Itoh *et al*. 2013). To detect the cause of the reduced *FT2a* expression, the 5′-upstream region of *FT2a* was first sequenced in the three cultivars and in TK780 and Hidaka 4. Eight SNPs and six indels were also detected (Figure 4.2). The sequences of TO and TK780 were identical to each other, but differed from those of HA and Hidaka 4 in a 43-bp indel in the promoter and a 10-bp indel in the 5′ UTR, which were located 731 and 47 bp upstream of the start codon, respectively, and in two SNPs (#2 and #4) (Figure 4.2). The sequence of HY was similar to those of TO and TK780 (including the 43-bp segment), but differed from them in one SNP (#1), a

4-bp indel 274 bp upstream of the start codon, and the 10-bp indel in the 5′UTR.

The introns and the 3′-downstream region in TO, HA, and HY were also sequenced to test whether the polymorphism(s) observed in the promoter and 5′ UTR could be responsible for late flowering in TO. The primers based on the gene model *Glyma16g150700* worked well for PCR amplification of these regions except for the first intron of TO. To sequence the first intron in TO, genome walking was used in this studies. Nested PCR analysis of genomic libraries produced an amplicon of 370-bp from the library constructed by using *Eco*RV. Sequencing revealed that it consisted of an unknown sequence of 137-bp fused with a 233-bp segment of the first intron of *FT2a* proximal to the second exon. A BLAST search of the NCBI genome database showed that the unknown sequence was identical to a part of an LTR of *SORE-1* (AB370254), which has been previously detected in a recessive allele at the *E4* locus (Liu *et al*. 2008a; Kanazawa *et al*. 2009). The inserted retrotransposon and its flanking regions were then amplified by nested PCR and sequenced. The retrotransposon was 6,224 bp long; its sequence was 100% identical to the LTRs of *SORE-1* and 99.7% identical to its coding region. Using a DNA marker for the *SORE-1* detection, it confirmed that TK780 also had *SORE-1* in the first intron, but Hidaka 4, HA, and HY did not. A total of 17 polymorphisms (10 SNPs, 2 indels, and 5 SSRs) were detected from the first intron to 3′ downstream regions among the three cultivars (Figure 4.2).

Thus, three early-maturing cultivars—TO, HA, and HY—had different FT2a sequences, which were designated as the *FT2a-TO*, *FT2a-HA*, and *FT2a-HY* haplotypes. *FT2a-TO* differed from both *FT2a-HA* and *FT2a-HY* in the 10-bp deletion in the 5′ UTR, and in SNP #17 and the *SORE-1* insertion in intron 1 (Figure 4.2 and Figure 4.4A). By using the database of plant *cis*-acting regulatory DNA elements (PLACE) (Higo *et al*. 1999), a W-box element (AGTCAAA) was detected that was created by SNP #17 in TO, and two *cis*-elements, RBCSCONSENSUS (AATCCAA) and ARR1AT (NGATT), in the genomic region flanking the *SORE-1* integration site.

A


DNA polymorphisms discriminating *FT2a-TO* from *FT2a-HA* and *FT2a-HY*

| | 10bp-indel<br>-47 | SNP#17<br>+398 | Insertion of *SORE-1*<br>+965 |
|---|---|---|---|
| Toyomusume (TO) | – | C | + |
| Hayahikari (HY) | + | A | – |
| Harosoy (HA) | + | A | – |

B

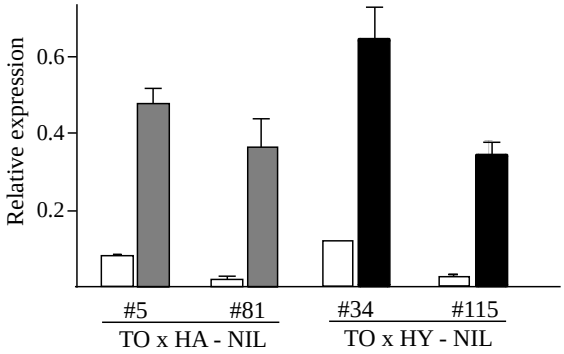



**Figure 4.4 DNA polymorphisms that discriminate between the *FT2a* alleles and *FT2a* transcript abundance in their NILs.** (A) Genomic positions and types of three DNA polymorphisms between Toyomusume (TO) and both Harosoy (HA) and Hayahikari (HY). (B) *FT2a* expression in 20-DAS-old plants of NILs for *FT2a-TO* (white) and *FT2a-HA* (gray) or *FT2a-HY* (black) under SD conditions. Relative mRNA levels are expressed as the ratios to *β-tubulin* transcript levels.

### 4.4.4 Expression of different *FT2a* alleles in near-isogenic lines and photoperiod insensitive accessions

Four sets of NILs for the above three *FT2a* haplotypes were developed from the progeny of $F_5$ heterozygous plants: two from the cross between TO and HA (#5 and #81) and two from the cross between TO and HY (#34 and #115). It was found that, under SD conditions, *FT2a* expression in *TO* was much lower than that of *FT2a* in *HA* and *HY* (Figure 4.4B).

Using 3 markers, five photoperiod-insensitive *e3e4* cultivars were selected, all of

which had the 10-bp deletion in 5′ UTR, but differed in SNP #17 and in the presence or absence of *SORE-1* (Figure 4.5A). *FT2a* expression was analyzed in fully-expanded trifoliate leaves at different leaf stages (first, second, and third true leaves) (Figure 4.5B). *FT2a* expression was markedly low in all stages in Karafuto 1, but was relatively high in the other four. Because Karafuto 1 differed from the other cultivars only in the presence of *SORE-1*, low expression of *FT2a* in *TO* was caused by the insertion of *SORE-1*, not by the 10-bp deletion or by SNP #17.

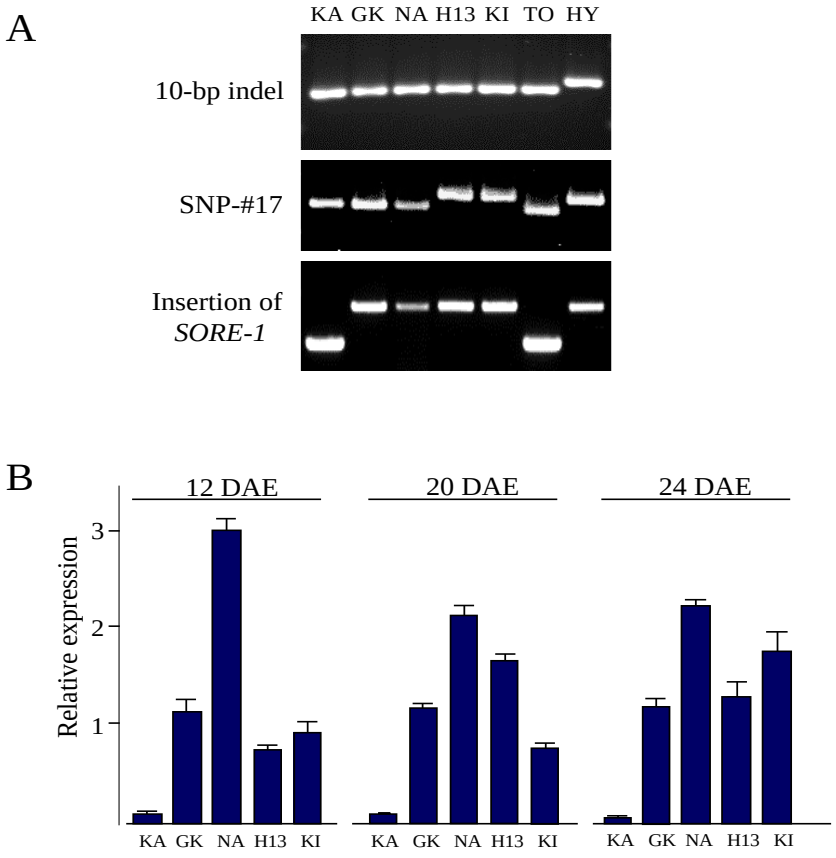



**Figure 4.5 *FT2a* transcript abundance in photoperiod-insensitive *e3 e4* cultivars under SD conditions.** (A) DNA polymorphisms in the 10-bp indel, SNP #17, and *SORE-1* insertion. (B) *FT2a* expression at the first (12 days after emergence: DAE), second (20 DAE), and third (24 DAE) leaf stages. Relative mRNA levels are expressed as the ratios to *β-tubulin* transcript levels. KA, Karafuto 1; GK, Gokuwase-Kamishunbetsu; NA, Napoli; H13, Heihe 13; KI, Kitamusume; TO, Toyomusume; HY, Hayahikari

Next, I surveyed the allelic variation at the *E9* locus in 31 cultivars and landraces/pure-line selections introduced from Hokkaido, Sakhalin of Far-east Russia,

Northern Tohoku, North-east China, Ukraine and Poland with the DNA marker for the detection of *SORE-1* insertion. I found that three cultivars of Hokkaido, Toyoharuka, Yukihomare and Toyohomare and five accessions, two from Sakhalin, Karafuto 1 and Karafuto (Toyohara) (or Toshidai-7910), two from the Northern Hokkaido, Horokanai Zairai and Denko OKuhara, and one from the Northern Tohoku, Aomori Zairai L145, processed the *e9* allele (Table 4.2).

**Table 4.2 Allelic variation at the *E9* (*FT2a*) locus in early-maturing soybean cultivars and land races/pure-line selections (*e9* is a *SORE-1*-inserted *FT2a* allele).**

| | Region | Cultivar/pure line selection/land races | *E9* |
|---|---|---|---|
| Toyomusume | Hokkaido | Cultivar | *e9* |
| Harosoy | Canada | Cultivar | *E9* |
| Hayahikari | Hokkaido | Cultivar | *E9* |
| Bekkai Zairai | Hokkaido | Land race | *E9* |
| Denko Okuhara | Hokkaido | Pure-line selection | *e9* |
| Gokuwase Kamoishunbetu zairai | Hokkaido | Land race | *E9* |
| Horokanai Zairai | Hokkaido | Land race | *e9* |
| Karafuto (Toyohara); Toshidai -7910 | Sakhalin | Land race | *e9* |
| Karafuto 1 | Sakhalin | Land race | *e9* |
| Konsen Zairai 1 | Hokkaido | Land race | *E9* |
| Konsen Zairai 3 | Hokkaido | Land race | *E9* |
| Ohyachi 1 | | Pure-line selection | *E9* |
| Ohyachi 2 | Hokkaido | Pure-line selection | *E9* |
| Okuhara E | Hokkaido | Pure-line selection | *E9* |
| Okuhara 1 | Hokkaido | Pure-line selection | *E9* |
| Okuihara daizu 1 | Hokkaido | Pure-line selection | *E9* |
| Sakamooto wase | Hokkaido | Land race | *E9* |
| Kitamusume | Hokkaido | Cultivar | *E9* |
| Tokachi nagaha | Hokkaido | Cultivar | *E9* |
| Toyoharuka | Hokkaido | Cultivar | *e9* |
| Toyohomare | Hokkaido | Cultivar | *e9* |
| Yukihomare | Hokkaido | Cultivar | *e9* |
| Aomiri L145 | Northern Tohoku | Land race | *e9* |
| Kamaishi 17 | Northern Tohoku | Land race | *E9* |
| Ohfunato 45 | Northern Tohoku | Land race | *E9* |
| Wase Otome | Northern Tohoku | Land race | *E9* |
| Heihe 1 | North-east China | Cultivar | *E9* |
| Heihe 13 | North-east China | Cultivar | *E9* |
| Heihe 33 | North-east China | Cultivar | *E9* |
| Heihe 34 | North-east China | Cultivar | *E9* |
| Heijian 1 | North-east China | Cultivar | *E9* |
| Heinong 37 | North-east China | Cultivar | *E9* |
| Jagataqi 13 | North-east China | Cultivar | *E9* |
| Jagataqi 3 | North-east China | Cultivar | *E9* |
| Darta | Poland | Cultivar | *E9* |
| Nawiko (=S23) | Poland | Cultivar | *E9* |
| Kiev 242 WH | Ukraine | Cultivar | *E9* |

# Chapter Ⅴ Molecular mechanism of delayed flowering mediated by the *e9* allele

## 5.1 Background and purpose

Transposable elements (TEs) are a primary DNA source that causes insertion-mediated dysfunction of a gene. Of these TEs, retrotransposons are the most abundant class and can generate stable mutations when they are inserted within or near genes because they transpose via replication and the sequence at the insertion site is retained (Kumar and Bennetzen 1999). Retrotransposons consist of the LTR and the non-LTR retrotransposons (Wicker *et al*. 2007). The LTR retrotransposons have direct LTRs that can range from a few 100 bp to over 5 kb in size. The LTRs do not encode any known proteins, but they do contain the promoters and terminators associated with the transcription of LTR retrotransposons. LTR retrotransposons are further sub-classified into the *Ty1*/*copia* and *Ty3*/*gypsy* groups that differ from each other in both their degree of sequence similarity and the order of encoded gene products (Doolittle *et al*. 1989; Xiong *et al*. 1990).

A *Ty1*/*copia*-like retrotransposon, *SORE-1*, was firstly identified in the exon 1 of *E4* in a photoperiod insensitivity line of soybean (Liu *et al*. 2008a). The *SORE-1* element comprised two 383-bp LTRs and a 5,472-bp internal coding region, and was flanked by a 5-bp target-site duplication sequence (5′-AAAAC-3′) (Kanazawa *et al*. 2009). The *SORE-1* is transcriptionally active and is partially silenced, and then causes the dysfunction of *E4* to allow soybean cultivation at high latitudes (Kanazawa *et al*. 2009). TEs inserted in intron of gene are known to affect RNA processing or render its host gene susceptible to siRNA- mediated transcriptional gene silencing and this mechanism has been adopted to evolve new developmental traits (Liu *et al*. 2004).

In this chapter, I studied RNA processing of the *FT2a-TO* haplotype in transcription and DNA methylation of *SORE-1* inserted and its flanking genomic

region to determine what mechanisms were involved in the lowered expression of *FT2a* in *TO*, and, if the *SORE-1* is a most likely candidate factor, how *SORE-1* attenuated the *FT2a* expression.

## 5.2 Materials and methods

### 5.2.1 Plant materials

Two sets of $F_6$ NILs for the *E9* gene were used for expression analyses: TO × HY-NILs #34 and TO × HY-NILs #115. In addition, TO × HA-NILs #5 and their heterozygous siblings were used.

### 5.2.2 Analysis of RNA processing

Plants of TO × HY-NILs #34 and TO × HY-NILs #115 were grown in growth cabinets under LD (18 h) conditions at 24 ℃. Fully developed trifoliate leaves of 20-day-old plants were sampled as a bulk at Zeitgeber time 3 to compare the expression levels in both sets of NILs. Leaves were immediately frozen in liquid N2, and stored at −80°C. RNA extraction, cDNA synthesis and qRT-PCR were performed as previously described in chapter IV. Three pairs of primers were used to detect expression in different cDNA areas of *FT2a*: *FT2a-a* from 5′ UTR to exon1; *FT2a-b* from exon1 to exon2; *FT2a-c* from exon4 to 3′ UTR.

Total genomic DNA was also extracted from trifoliate leaves as previously described in chapter three. Genomic DNA and cDNA from TO × HY-NILs #115 were used in semi-quantitative RT-PCR to determine if parts or full *SORE-1* is inserted into cDNA. Each of semi-quantitative RT-PCR contained 0.5μg of cDNA (0.1μg for total genomic DNA used as a control) as template, 1 μl of each primer (10 μM) and dNTP (2.5 mM), 0.5 μl of ExTaq polymerase, and 2.5 μl of 10× ExTaq buffer in a total volume of 25 μl; amplification conditions were 33 cycles at 94°C for 30 s, 60°C or 64 ℃ for 30 s (depending on the primers used), and 72 ℃ for 30 s–8 mins (depending on the sizes of amplified fragments).

To identify whether allele-specific transcriptional repression exist in *FT2a-TO*, cDNA was synthesized from trifoliate leaves of 20-day-old plants of NILs for the *FT2a-TO*, *FT2a-HY* and *FT2a-TO*/*HY* haplotypes grown under LD. DNA marker was developed based on SNP #28 after the stop codon: PCR was performed with the primers 5′-ATTTCGTCAACTGGGTAGGGAG-3′ and 5′-GGGAGATTGCCAATTAATATTTCTGAAA-3′, and the amplified products were digested with *Dde* I.

### 5.2.3 Methylation analysis

Genomic DNA was extracted from trifoliate leaves of 20-day-old plants of NILs for the *FT2a-TO* and *FT2a-HY* haplotypes, grown under SD and LD conditions. DNA samples were digested with McrBC. Digested and undigested samples were used for semi-quantitative PCR amplification of different regions of *FT2a* genomic and *SORE-1* regions. Primer sequences used in methylation analyses are listed in Additional table 4.

## 5.3 Results and Discussion

### 5.3.1 RNA processing of the *E9* gene

TEs in introns often affect chromatin structure and modify RNA processing of the host gene and, therefore, influence its expression patterns (Varagona *et al*. 1992; Liu *et al*. 2004; Iwata *et al*. 2012). Using qRT-PCR on cDNA synthesized with random primers, which targeted different regions, *FT2a* expression was analyzed in two sets of NILs for *FT2a-TO* and *FT2a-HY* grown in SD. In all three targeted regions (a–c in Figure 5.1A), the *FT2a* transcript abundance was considerably lower (1/5 to <1/10) in NILs for *FT2a-TO* than in NILs for *FT2a-HY* (Figure 5.1B).

To analyze *FT2a* RNA processing in *FT2a-TO*, semi-quantitative RT-PCR was performed on cDNAs synthesized with random primers. No amplicon was detected in

regions a (from exon 1 to intron 1), b and c (from exon 1 to *SORE-1*), or d and e (from *SORE-1* to exon 2), although the expected amplicons were observed in PCR on genomic DNA of the NIL for *FT2a-TO* (Figure 5.2). For region f (from exon 1 to exon 2), a fragment (~150 bp) was amplified in both NILs, although signal intensity was much higher in the NIL for *FT2a-HY* than in the NIL for *FT2a-TO*; as expected, genomic PCR produced fragments of 7,293 bp in the NIL for *FT2a-TO* and 1,064 bp in the NIL for *FT2a-HY* (Figure 5.2B). These results suggest that intron 1 with the *SORE-1* insertion could be spliced out in the NIL for *FT2a-TO*.

Next, *FT2a* expression was examined in heterozygous siblings of NILs; this analysis was based on the fact that SNP #28 after the stop codon (Figure 5.2) created a *Dde*I restriction site in *FT2a-HA*, but not in *FT2a-TO* and *FT2a-HY*. By performing RT-PCR and digesting the product with *Dde*I, expression of *FT2a-TO* can be distinguished from that of *FT2a-HA* in heterozygous plants. In the NILs-#5 for *FT2a-TO* and *FT2a-HA*, and its siblings, the *FT2a* transcript level was high in homozygotes for *FT2a-HA*, slightly lower in heterozygotes, and very low in homozygotes for *FT2a-TO* (Figure 5.3). Digestion of PCR products revealed that in heterozygotes, the transcript level of *FT2a-HA* was much higher than that of *FT2a-TO*. This difference suggests that the lower expression of *FT2a-TO* was caused by allele-specific transcriptional repression rather than sequence-specific RNA degradation of RNA silencing that decreases the levels of transcripts from both alleles.

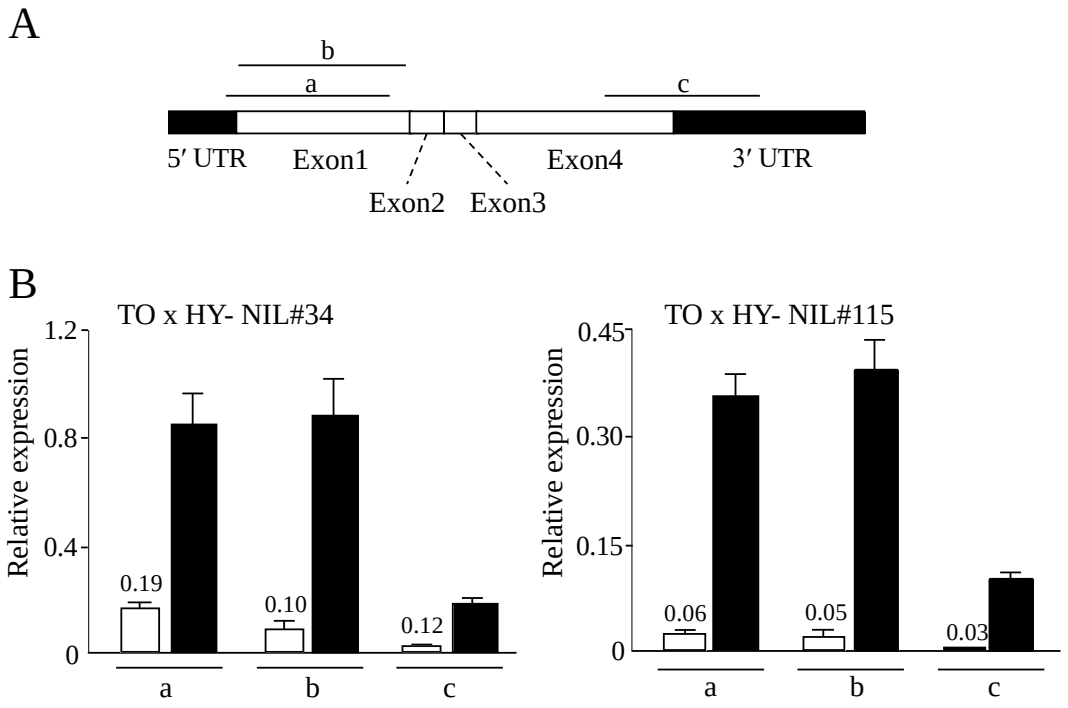



**Figure 5.1 *FT2a* transcript abundance in two sets of NILs for *FT2a-TO* and *FT2a-HY* haplotype.** (A) Three regions (a–c) in the *FT2a* coding region used to assess transcript abundance. The 5′ UTR and 3′ UTR are a part of exon 1 and exon 4, respectively. (B) *FT2a* expression analyzed in 20-DAS plants under SD conditions. Relative mRNA levels are expressed as the ratios to *β-tubulin* transcript levels. cDNA was synthesized with random primers. Numbers above the white bars are the ratios of the expression levels in NIL for *FT2a-TO* (white bars) to those in *FT2a-HY* (black bars).

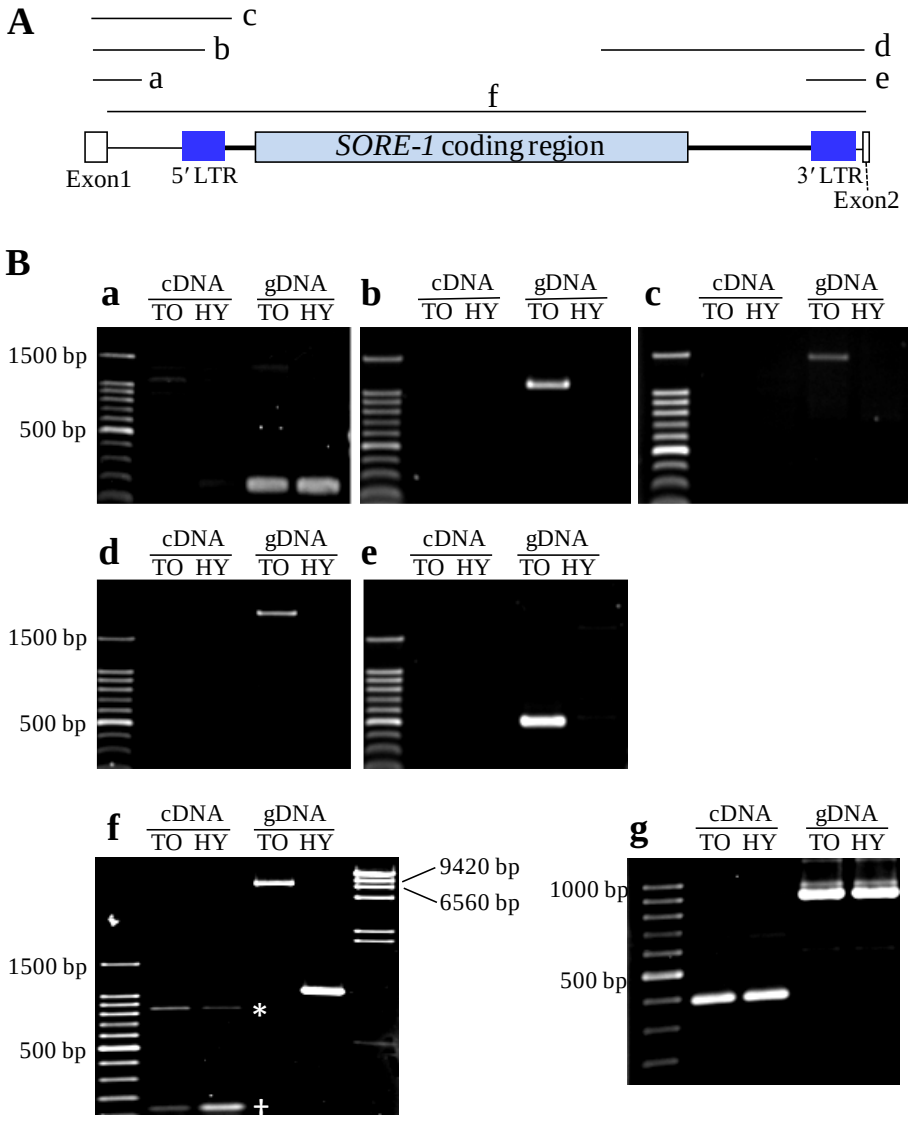



**Figure 5.2 *FT2a* RNA processing in the first intron with *SORE-1* insertion.** (A) a–f, Regions examined. (B) Semi-quantitative PCR analysis of *FT2a* expression in NIL #115 for *FT2a-TO* (TO) and *FT2a-HY* (HY) in 20-DAS plants under SD conditions. cDNA was synthesized with random primers. g, amplification of the *β-tubulin* transcript. *, nonspecific amplification. †, the targeted amplicons from cDNA (125 bp).

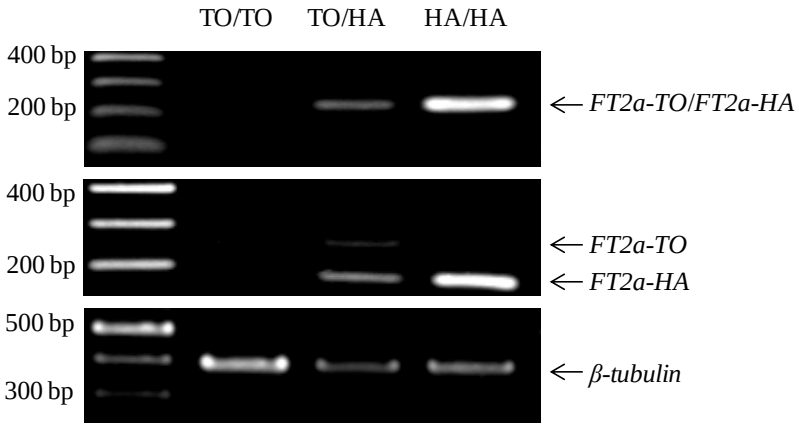



**Figure 5.3 Transcript abundance of different alleles.** Transcript abundance of different *FT2a* alleles was assayed by allele-specific restriction digestion.

### 5.3.2 DNA methylation at the *E9* locus

The methylation level of *FT2a* in TO and HY was evaluated. Methylation-dependent McrBC restriction digestions and mock digestions of genomic DNA were used to analyze cytosine methylation in NILs for *FT2a* in TO and HY. There was no difference in PCR amplification of genomic regions a–f and h–k in the McrBC-digested and mock-digested samples in both NILs under SD (Figure 5.4A) and LD conditions (Figure 5.4B). In contrast, no amplicons were detected for regions S1–S3 (which include the LTRs of *SORE-1* and *FT2a* regions flanking the LTRs) after McrBC digestion in the NIL for *FT2a* in TO, although fragments of expected sizes were amplified from mock-digested DNA. PCR on both McrBC-digested and mock-digested DNAs produced the expected amplicons in region S4 (which did not include the LTR sequence) of the NIL for *FT2a-TO* haplotype and in genomic region g (which did not contain *SORE-1*) of the NIL for *FT2a-HY* haplotype. Taken together, these data indicate that *SORE-1* was highly methylated, but methylation appeared not to extend to the *FT2a* genomic region flanking *SORE-1*. The same result was obtained for plants grown in LD (data not shown), which indicates that lower mRNA level of *FT2a* in TO is associated with *SORE-1* methylation in both SD and LD conditions.

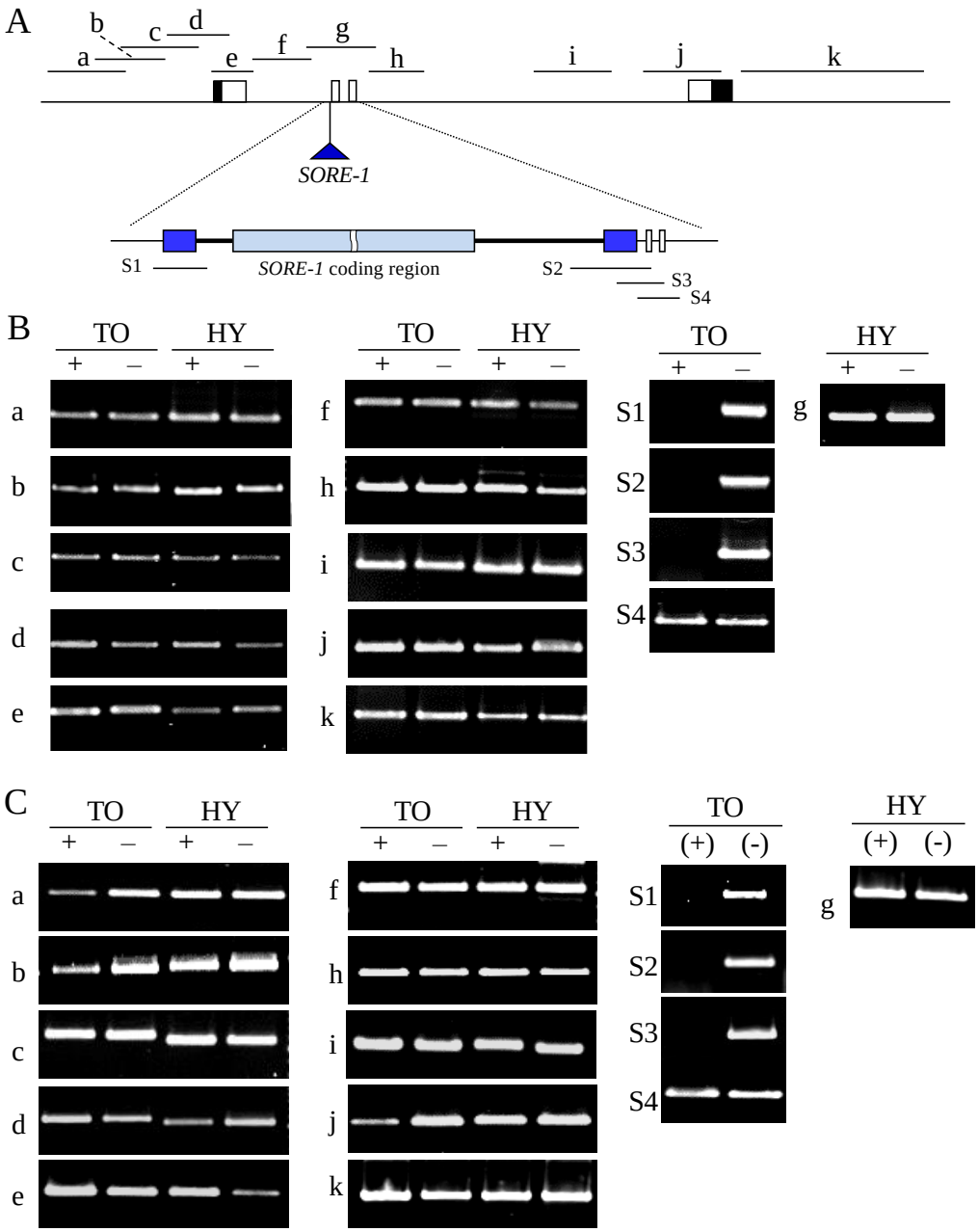



**Figure 5.4 DNA methylation in the *FT2a* genomic region.** (A) Diagram of the *FT2a* genomic region showing the position of *SORE-1* insertion. Amplicons were analyzed by semi-quantitative PCR after McrBC or mock digestion; the amplified regions are designated as a to k and S1 to S4. Exons, white; UTRs, black; LTRs of *SORE-1*, gray. (B) and (C) Genomic DNA from leaves of 20-day-old plants of NILs for *FT2a-TO* (TO) and *FT2a-HY* (HY) grown under SD conditions (B) and LD conditions (C) was digested with McrBC (+) or mock-digested (–) and amplified by PCR. Amplicons were visualized in agarose gels.

# Chapter Ⅵ Discussion

## 6.1 Flowering gene *E9* is *FT2a*

Recent studies concerning the genetics of flowering time in soybean have detected a QTL on LG J (Gm16) associated with flowering and maturity as a new major gene *E9* (Liu *et al.* 2007; Kong *et al.* 2014). This new locus *E9* was fine mapped to a 245-kb region on Gm16, where 24 predicted genes located according to the reference Williams 82 sequence, among which two previously characterized flowering homologous genes, *FT2a* or *FT2b* are considered as the candidate genes of *E9* locus (Kong *et al*. 2014). However, the responsible gene for *E9* remains to be determined.

In this study, flowering time in the $F_2$ and $F_3$ progeny of a cross between Toyomusume and Harosoy co-segregated with the alleles at the *E1* and *E9* loci. Fine mapping delimited *E9* to a 40.1-kb region that contained three genes, including *FT2a*, a soybean ortholog of *FT* (Figure 3.1). Sequencing and expression analysis suggested that *FT2a* is the most likely candidate for *E9*, and delayed flowering due to *e9* is most likely caused by the reduced *FT2a* transcript abundance. Despite sequence identity in the coding regions, several SNPs and indels of 4−43 bp in the promoter and 5′ UTR were detected among cultivars and accessions tested, which is consistent with a previous report (Zhang *et al*. 2015). However, expression analysis of NILs and photoperiod-insensitive accessions carrying different *FT2a* haplotypes revealed that the polymorphisms in the promoter and 5′ UTR were not responsible for different *FT2a* expression levels (Figures 3.4 and 3.5). Toyomusume also differed from Harosoy and Hayahikari by a SNP and a *SORE-1* insertion in the first intron, of which the latter was solely associated with the *FT2a* expression levels (Figure 3.5). Thus, this study reveals that the insertion of *SORE-1* attenuated *FT2a* expression and delayed flowering, implicated the role of retrotransposon in the regulation of the floral transition in soybean.

The soybean genome possesses a total of ten *FT* orthologs, among which six retain the *FT* function and can promote flowering of *Arabidopsis ft* mutants (Thakare *et al*. 2011; Wang *et al*. 2015) or *Col-0* (Kong *et al*. 2010; Fan *et al*. 2014). All of the six homologs could therefore function as potential floral inducers in soybean, although only two of them, *FT2a* and *FT5a*, have been extensively characterized in studies of molecular mechanisms of flowering (Kong *et al*. 2010; Sun *et al*. 2011; Xia *et al*. 2012; Xu *et al*. 2015; Watanabe *et al*. 2011; Jiang *et al*. 2013; Nan *et al*. 2014; Zhao *et al*. 2015; Thakare *et al*. 2011). This study demonstrates that different levels of *FT2a* expression directly regulate natural variation in flowering time in soybean.

### 6.2 Factors responsible for attenuation of *FT2a* expression

Plant TEs inserted in introns may affect RNA processing (Varagona *et al*. 1992; Iwata *et al*. 2012) or render their host genes susceptible to short interfering RNA (siRNA)-mediated silencing (Liu *et al*. 2004). The results of this study show that the first intron (including *SORE-1*) is spliced out, because no primary RNA transcripts that would cover *FT2a* exons and *SORE-1* were detected while the spliced products were detected (Figure 4.2). Thus, *SORE-1* insertion did not markedly interfere with *FT2a* RNA processing.

In this study, I found that the reduction in *e9* transcript abundance was caused by allele-specific transcriptional repression due to the insertion of *SORE-1*, the LTRs and adjacent sequences of which were highly methylated (Figure 4.4). Therefore, epigenetic mechanisms likely account for the reduction in *FT2a-TO* transcript levels. RNA-directed DNA methylation or the resulting chromatin modifications regulate gene expression by interfering with transcription factor binding, leading to different expression profiles for different transcription factors (Liu *et al*. 2004; Shibuya *et al*. 2009; Deng and Chua 2015). PLACE analysis detected two *cis*-elements, RBCSCONSENSUS and ARR1AT, in the region flanking the *SORE-1* integration site in the first intron. However, the functions of the two elements in *FT2a* expression are

unclear. A further test is thus needed to determine the functions of the two *cis*-elements or nearby unknown elements in the regulation of *FT2a* expression and whether *SORE-1* insertion interrupts binding of a transcriptional factor(s) to these *cis*-elements.

Methylation-mediated gene repression by intronic TEs is well characterized in *Arabidopsis FLOWERING LOCUS C* (*FLC*), which encodes a transcription factor containing a MADS domain that inhibits *FT* expression (Michaels and Amasino 1999; Michaels *et al*. 2003). In Col-0, the functional *FLC* allele is highly expressed in the presence of FRIGIDA and causes extremely late flowering (Michaels *et al*. 2003). In contrast, in ecotype *Landsberg erecta* (*Ler*), the *FLC* allele has a 1,224-bp non-autonomous Mutator-like TE in intron 1 and is expressed at low levels due to its transcriptional silencing through histone H3-K9 methylation, which is triggered by siRNA generated from homologous TEs (Liu *et al*. 2004). *FLC-Ler*, however, can still be regulated by genes in the autonomous flowering pathway and by genes involved in vernalization, because the TE insertion does not affect the transcription factor–binding sites in intron 1 (Liu *et al*. 2004). Similarly to the *FLC-Ler* allele, the expression of *FT2a-TO* is repressed due to epigenetic modification caused by the insertion of *SORE-1* in intron 1. However, *FT2a-TO* expression was still higher in SD than in LD (Figures 3.3). Virus-induced silencing of *E1-like* genes (repressors of *FT2a* and *FT5a*) lowers photoperiod sensitivity of Toyomusume by up-regulating the expression of both *FT2a* and *FT5a* (Xu *et al*. 2015). The regulation of *FT2a* expression by *E1-like* and other genes involved in photoperiod responses is thus retained in *FT2a-TO* plants. The *FT2a-TO* haplotype may thus be involved in flowering as a leaky allele, not a dysfunctional allele.

### 6.3 Origin and adaptive role of the *e9* allele

*SORE-1* was first detected in a recessive allele at the *E4* locus encoding phytochrome A; its insertion in the first exon caused a premature stop codon and

resulted in a dysfunctional truncated protein (Liu *et al*. 2008a). DNA marker analysis revealed that the *e4* allele with the *SORE-1* insertion is present mainly in landraces from northern Japan (Kanazawa *et al*. 2009), although it has been used in breeding of photoperiod-insensitive cultivars in high-latitude regions of other countries (Xu *et al*. 2015). This insertion in the *E4* gene may thus have played an adaptive role in expanding the areas of soybean cultivation to higher latitudes. The survey for the insertion of *SORE-1* in the *FT2a* allele suggests that the *SORE-1*-inserted *FT2a* is a region-specific allele, which was detected in only a few local varieties established in Sakhalin, northern Hokkaido and northern Honshu among photoperiod-insensitive landraces and cultivars having the *e4* allele with the *SORE-1* insertion. Therefore, the insertion of *SORE-1* in the first intron of *FT2a-TO* may be of recent origin.

The landrace Karafuto (Toyohara) (Toshidai-7910), introduced from Sakhalin with Karafuto 1, most likely was the source of the *FT2a* allele with the *SORE-1* insertion in Toyomusume, because this landrace was used as a parent in the breeding of Toyosuzu, a parent in the breeding of Toyomusume (Tanaka *et al*. 2003). Similar to Toyomusume, both Toshidai-7910 and Karafuto 1 have a null allele at the *E1* locus, but, unlike Toyomusume, they have recessive alleles at *E3* and *E4* (Tsubokura *et al*. 2014; this study). This is a maturity genotype that permits extremely early flowering and maturation and enables seed production in cold climates with a limited frost-free season. Because *FT2a* and *FT5a* control flowering redundantly (Kong *et al*. 2010; Nan *et al*. 2014), the *e9* (*FT2a-TO*) allele could have been selected in the presence of functional *FT5a* because it maintains vegetative growth. It is thus another example of the adaptive role of *SORE-1* insertion as indicated by Kanazawa *et al*. (2009). The *e9* allele was further detected in three cultivars of Hokkaido, Toyoharuka, Toyohomare, and Yukihomare, all of which possessed the *e1-nl* allele like Toyomusume. The characteristic combination of *e9* and *e1-nl* may therefore indicate that the *e9* allele possesses the adaptive significance even in modern soybean breeding in Hokkaido, to which the early-maturing habit conditioned by *e1-nl* has been preferable.

In addition to the adaptive role in the limited frost-free growing season of

Hokkaido, the *e9* allele may also be useful for developing cultivars adapted to a shorter photoperiod in low-latitude environments where flowering is strongly promoted. In such environments, a longer vegetative phase, a so-called long-juvenile trait, is desirable. A leaky allele similar to *e9* may be useful for reducing the transcript levels of *FT2a* under SD conditions, in addition to long-juvenile genes reported so far, such as *E6* (Bonato and Vello *et al*. 1999) and *j* (Ray *et al*. 1995). A further study is therefore needed to evaluate the adaptive significance of *e9* under SD conditions.

## Summary

Knowledge of molecular mechanisms of flowering and maturity is important for understanding the phenology of seed crops and for maximizing yield in a given environment. Soybean is cultivated in wide regions of the world as an important crop to supply vegetable proteins and oil. This wide adaptability has been created by natural variations in major genes and quantitative trait loci to control flowering. However, the molecular basis of flowering and maturity still remains poorly understood. In this thesis, I studied molecular-genetic bases of flowering in soybean, in particular focusing on a maturity gene *E9*, which has been identified in the progeny of a cross between cultivated and wild soybeans.

I first analyzed genetic bases on flowering times segregated in the progeny of a cross between two early-maturing soybean cultivars, Toyomusume and Harosoy. The two cultivars flowered almost at the same time, but flowering times in the $F_2$ and $F_3$ populations varied widely. They differed in the genotypes at a maturity *E1* locus, a repressor for soybean orthologs of *FLOWERING LOCUS T* (*FT*), which produces the most marked effect on flowering. Expectedly, the progeny exhibited segregation in flowering times closely associated with the *E1* genotypes. Further, I found by marker-assisted analysis that the progeny segregated for the *E9* locus, at which Toyomusume had the recessive *e9* allele for late flowering.

Then, I carried out fine-mapping of *E9* and delimited it to a 40.1-kb genomic region, which contained *FT2a*, an ortholog of *FT*. As a result of expression analyses in the progeny of 10 heterozygous $F_2$ plants, plants homozygous for *e9* exhibited reduced *FT2a* expressions, compared with those for *E9* in each family. Sequence analyses further revealed that the two cultivars had the identical sequence in the coding region, but differed in a total of 30 DNA polymorphisms in the genomic sequence. From a comparison among different cultivars, I found the three DNA

polymorphisms that differed consistently between the *E9* and *e9* alleles. Expression analysis for near-isogenic lines and photoperiod-insensitive cultivars for different *FT2a* haplotypes demonstrated that low *FT2a* expression of *e9* was caused by the insertion of a *Ty1/copia*-like retrotransposon designated *SORE-1*.

Transposable elements in introns are well known to affect chromatin structure and modify RNA processing of the host gene and thereby influence its expression pattern. From the expression analysis for plants heterozygous for the *E9* locus, I found that the lower expression of *e9* was caused by allele-specific transcriptional repression. I further confirmed that *SORE-1* did not influence the RNA processing. Rather, lower mRNA level of *e9* was associated with *SORE-1* methylation, because the inserted *SORE-1* was highly methylated.

Based on the results obtained, I discussed on roles and functions of the *e9* allele in soybean flowering and adaptation. This allele is a leaky allele, because its regulation by other genes involved in photoperiod response was retained. The *FT2a* transcript abundance is directly associated with the variation in flowering time in soybean, and lower transcript abundance due to the insertion of *SORE-1* produce delayed flowering. It is considered that the *e9* allele had been selected for by its action to maintain vegetative growth which enables higher yield in early-flowering genetic backgrounds. It may also be useful as a long-juvenile allele in cultivar development in low-latitude regions, where flowering is strongly promoted.

## References


Abe J, Xu D, Miyano A, Komatsu K, Kanazawa A, Shimamoto Y. Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. Crop Sci. 2003;43:1300–4.

Adrian J, Farrona S, Reimer JJ, Albani MC, Coupland G, Turck F. *Cis*-regulatory elements and chromatin state coordinately control temporal and spatial expression of *FLOWERING LOCUS T* in *Arabidopsis*. Plant Cell 2010;22:1425–40.

Andrés F and Coupland G. The genetic basis of flowering responses to seasonal cues. Nat. Rev. Genet. 2012;13:627–39.

Bernard RL. Two genes for time of flowering in soybeans. Crop Sci. 1971;11:242–4.

Bonato ER and Vello NA. *E6*, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. 1999;22:229–32.

Bünning E. Circadian rhythms and the time measurement in photoperiodism. Cold Spring Harbor Symp. Quant. Biol. 1960;25:249–56.

Bünning E. Die endogene Tagesrhythmik als grundlage der photoperiodischen reaktion. Ber. Dtsch. Bot. Ges. 1936;54:590–607.

Butler WL, Norris KH, Siegelman HW, Hendricks SB. Detection, assay, and preliminary purification of the pigment controlling photoresponsive development of plants. Proc. Natl. Acad. Sci. USA 1959;45:1703–8.

Buzzell RI and Voldeng HD. Inheritance of insensitivity to long day length. Soybean Genet. Newsl. 1980;7:26–9.

Buzzell RI. Inheritance of a soybean flowering response to fluorescent-daylength conditions. Can J. Genet. Cytol. 1971;13:703–7.

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

Cao S, Kumimoto RW, Gnesutta N, Calogero AM, Mantovani R, Holt BF 3rd. A distal *CCAAT*/NUCLEAR FACTOR Y complex promotes chromatin looping at the *FLOWERING LOCUS T* promoter and regulates the timing of flowering in *Arabidopsis*. Plant Cell 2014;26:1009-17.

Cashmore AR. Cryptochromes: enabling plants and animals to determine circadian time. Cell 2003;114:537–43.

Castillejo C, Pelaz S. The balance between CONSTANS and TEMPRANILLO activities determines *FT* expression to trigger flowering. Curr. Biol. 2008;18: 1338–43.

Chaves I, Pokorny R, Byrdin M, Hoang N, Ritz T, Brettel K, Essen LO, van der Horst GT, Batschauer A, Ahmad M. The cryptochromes: Blue light photoreceptors in plants and animals. Annu. Rev. Plant Biol. 2011;62:335–64.

Cober ER and Voldeng HD. A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. Crop Sci. 2001;41:698–701.

Cober ER, Molnar SJ, Charette M, Voldeng HD. A new locus for early maturity in soybean. Crop Sci. 2010;50:524–7.

Cober ER, Tanner JW, Voldeng HD. Soybean photoperiod-sensitivity loci respond differentially to light quality. Crop Sci. 1996;36:606–10.

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, vanToai TT, Lohnes DG, Chung J, Specht JE. An integrated genetic linkage map of the soybean genome. Crop Sci. 1999;39:1464–90.

Deng S and Chua NH. Inverted-repeat RNAs targeting *FT* intronic regions promote *FT* expression in *Arabidopsis*. Plant Cell Physiol. 2015;56:1667-78.

Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, Yano M, Yoshimura A. *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. 2004;18:926-36.

Doolittle WF, Feng DF, Johnson MS, McClure MA. Q Rev Biol. 1989;64:1–29.

Dwiyanti MS, Yamada T, Sato M, Abe J, Kitamura K. Genetic variation of γ-tocopherol methyltransferase gene contributes to elevated α-tocopherol content in soybean seeds. BMC Plant Biol. 2011;11:152.

El-Din El-Assal S, Alonso-Blanco C, Peeters AJ, Raz V, Koornneef M. A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. Nat. Genet. 2001;298:435–40.

Fan C, Hu R, Zhang X, Wang X, Zhang W, Zhang Q, Ma J, Fu YF. Conserved CO-FT regulons contribute to the photoperiod flowering control in soybean. BMC Plant Biol. 2014;7:14:9.

Fehr WR, Caviness CE, Burmood DT, Pennington JS. Stage of development

descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci. 1971;11:929–31.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations, Genetics 1999;151:1531-45.

Fornara F, Panigrahi KC, Gissot L, Sauerbrunn N, Rühl M, Jarillo JA, Coupland G. *Arabidopsis* DOF transcription factors act redundantly to reduce *CONSTANS* expression and are essential for a photoperiodic flowering response. Dev. Cell 2009;17:75–86.

Foucher F, Morrin J, Courtiade J, Cadioux S, Ellis N, Banfield MJ and Pameau C. *DETERMINATE* and *LATER FLOWERING* are two *TERMINAL FLOWER1*/*CENTRORADIALIS* homologs that control two distinct phases of flowering initiation and development in pea. Plant Cell 2003;15:2742–54.

Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Morris B, Coupland G, Putterill J. *GIGANTEA*: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. EMBO J. 1999;18:4679–88.

Franklin KA and Quail PH. Phytochrome functions in *Arabidopsis* development. J. Exp. Bot. 2010;61:11–24.

Garner WW and Allard HA. Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. J. Agric. Res. 1920;18:553–606.

Guo G, Xu K, Zhang X, Zhu J, Lu M, Chen F, Liu L, Xi ZY, Bachmair A, Chen Q, Fu YF. Extensive analysis of *GmFTL* and *GmCOL* expression in northern soybean

cultivars in field conditions. PLoS One 2015;10:e0136601.

Guo H, Yang H, Mockler TC, Lin C. Regulation of flowering time by *Arabidopsis* photoreceptors. Science 1998;279:1360-3.

Hamner KC and Bonner J. Photoperiodism in relation to hormones as factors in floral initiation and development. Bot. Gaz. 1938;100:388-31.

Hamner KC. Photoperiodism and circadian rhythms. Cold Spring Harb. Symp. Quant. Biol. 1960;25:269–77.

Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 2003;422:719-22.

Hecht V, Knowles CL, Vander Schoor JK, Liew LC, Jones SE, Lambert MJ and Weller JL. Pea *LATE BLOOMER1* is a *GIGANTEA* ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. Plant Physiol. 2007;144:648–61.

Hecht V, Laurie RE, Vander Schoor JK, Ridge S, Knowles CL, Liew LC, Sussmilch FC, Murfet IC, Macknight RC, Weller JL. The pea *GIGAS* gene is a *FLOWERING LOCUS T* homolog necessary for graft-transmissible specification of flowering but not for responsiveness to photoperiod. Plant Cell 2011;23:147-61.

Helliwell CA, Wood CC, Robertson M, James Peacock W, Dennis ES. The *Arabidopsis* FLC protein interacts directly *in vivo* with *SOC1* and *FT* chromatin and is part of a high-molecular-weight protein complex. Plant J. 2006;46:183–92.

Higo K, Ugawa Y, Iwamoto M, Korenaga T. Plant *cis*-acting regulatory DNA elements (PLACE) database. Nucl. Acids Res. 1999;27:297-300.

Hisada A, Hanzawa H, Weller JL, Nagatani A, Reid JB, Furuya M. Light-induced nuclear translocation of endogenous pea phytochrome A visualized by immunocytochemical procedures. Plant Cell 2000;12:1063-78.

Hopkins WG. Introduction to Plant Physiology, Ed2. John Wiley & Sons, Inc., New York, 1999;261–3.

Hu Q, Jin Y, Shi H, Yang W. *GmFLD*, a soybean homolog of the autonomous pathway gene *FLOWERING LOCUS D*, promotes flowering in *Arabidopsis thaliana*. BMC Plant Biol. 2014;14:263.

Huang F, Xu G, Chi Y, Liu H, Xue Q, Zhao T, Gai J, Yu D. A soybean MADS-box protein modulates floral organ numbers, petal identity and sterility. BMC Plant Biol. 2014;14:89.

Hughes RM, Vrana JD, Song J, Tucker CL. Light-dependent, dark-promoted interaction between *Arabidopsis* cryptochrome 1 and phytochrome B proteins. J. Biol. Chem. 2012;287:22165-72.

Imaizumi T, Tran HG, Swartz TE, Briggs WR, Kay SA. *FKF1* is essential for photoperiodic-specific light signalling in *Arabidopsis*. Nature 2003;426:302-6.

Ishikawa R, Aoki M, Kurotani KI, Yokoi S, Shinomura T, Takano M, Shimamoto K. Phytochrome B regulates *Heading date 1* (*Hd1*)-mediated expression of rice florigen *Hd3a* and critical day length in rice. Mol. Genet. Genomics 2011;285:461–70.

Itoh H, Izawa T. The coincidence of critical day length recognition for florigen gene expression and floral transition under long-day conditions in rice. Mol. Plant. 2013;6:635–49.

Itoh H, Nonoue Y, Yano M, Izawa T. A pair of floral regulators sets critical day length for *Hd3a* florigen expression in rice. Nat. Genet. 2010;42:635–8.

Iwata H, Gaston A, Remay A, Thouroude T, Jeauffre J, Kawamura K, Oyant LH, Araki T, Denoyes B, Foucher F. The *TFL1* homologue *KSN* is a regulator of continuous flowering in rose and strawberry. Plant J. 2012;69:116-25.

Izawa T, Oikawa T, Sugiyama N, Tanisaka T, Yano M, Shimamoto K. Phytochrome mediates the external light signal to repress *FT* orthologs in photoperiodic flowering of rice. Genes Dev. 2002;16:2006–20.

Jackson SD. Plant responses to photoperiod. New Phytol. 2009;181:517-31.

Jang S, Marchal V, Panigrahi KC, Wenkel S, Soppe W, Deng XW, Valverde F, Coupland G. *Arabidopsis* COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. EMBO J. 2008;27:1277-88.

Jia Z, Jiang B, Gao X, Yue Y, Fei Z, Sun H, Wu C, Sun S, Hou W, Han T. *GmFULa*, a *FRUITFULL* homolog, functions in the flowering and maturation of soybean. Plant Cell Rep. 2015;34:121-32.

Jiang B, Yue Y, Gao Y, Ma L, Sun S, Wu C, Hou W, Lam HM, Han T. *GmFT2a* polymorphism and maturity diversity in soybeans. PLoS One 2013;8:e77474.

Jung JH, Seo YH, Seo PJ, Reyes JL, Yun J, Chua NH, Park CM. The *GIGANTEA*-regulated microRNA172 mediates photoperiodic flowering independent of *CONSTANS* in *Arabidopsis*. Plant Cell 2007;198:2736–48.

Kanazawa A, Liu B, Kong F, Arase S, Abe J. Adaptive evolution involving gene

duplication and insertion of a novel *Ty1*/*copia*-like retrotransposon in soybean. J. Mol. Evol. 2009;69:164-75.

Kendrick RE, Kronenberg G. Photomorphogenesis in plants. Martinus Nijhoff, Dordrecht, Netherlands. 1994.

Kevei E, Schafer E, Nagy F. Light-regulated nucleo-cytoplasmic partitioning of phytochromes. J. Exp. Bot. 2007;58:3113-24.

Kircher S, Kozma-Bognar L, Kim L, Adam E, Harter K, Schafer E, Nagy F. Light quality-dependent nuclear import of the plant photoreceptors phytochrome A and B. Plant Cell 1999;11:1445-56.

Knott JE. Effect of a localized photoperiod on spinach. Proc. Am. Soc. Hortic. Sci. 1934;31:152–4.

Kobayashi Y and Weigel D. Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. Genes Dev. 2007;21:2371-84.

Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M. *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol. 2002;43: 1096–105.

Komiya R, Ikegami A, Tamaki S, Yokoi S, Shimamoto K. *Hd3a* and *RFT1* are essential for flowering in rice. Development 2008;135:767-74.

Komiya R, Yokoi S, Shimamoto K. A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. Development. 2009;136:3443-50.

Kong F, Liu B, Xia Z, Sato S, Kim BM, Watanabe S, Yamada T, Tabata S, Kanazawa A, Harada K, Abe J. Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. Plant Physiol. 2010;154:1220–31.

Kong F, Nan H, Cao D, Li Y, Wu F, Wang J, Lu S, Yuan X, Cober ER, Abe J, Liu B. A new dominant gene *E9* conditions early flowering and maturity in soybean. Crop Sci. 2014;154:1220-31.

Kumar A and Bennetzen JL. Plant retrotransposons. Annu. Rev. Genet. 1999;33:479-532.

Kumar SV, Lucyshyn D, Jaeger KE, Alós E, Alvey E, Harberd NP, Wigge PA. Transcription factor *PIF4* controls the thermosensory activation of flowering. Nature 2012;484:242-5.

Kumar V and Rani S. Light sensitivity of the photoperiodic response system in higher vertebrates: wavelength and intensity effects. Indian J. Exp. Biol. 1999;37:1053-64.

Kumimoto RW, Adam L, Hymus GJ, Repetti PP, Reuber TL, Marion CM, Hempel FD, Ratcliffe OJ. The nuclear factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in *Arabidopsis*. Planta 2008;228:709–23.

Kumimoto RW, Zhang Y, Siefers N, Holt BF. NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*. Plant J. 2010;63:379–91.

Langewisch T, Zhang H, Vincent R, Joshi T, Xu D, Bilyeu K. Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes.

PLoS One 2014;9:e94150.

Lariguet P, Dunand C. Plant photoreceptors: phylogenetic overview. J. Mol. Evol. 2005;61:559-69.

Laubinger S, Marchal V, Gentilhomme J, Wenkel S, Adrian J, Jang S, Kulajta C, Braun H, Coupland G, Hoecker U. *Arabidopsis* SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. Development 2006;133:3213–22.

Lazaro A, Valverde F, Pineiro M, Jarillo JA. The *Arabidopsis* E3 ubiquitin ligase HOS1 negatively regulates CONSTANS abundance in the photoperiodic control of flowering. Plant Cell 2012;24:982–99.

Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH. Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. Genes Dev. 2007;21:397–402.

Li F, Zhang X, Hu R, Wu F, Ma J, Meng Y, Fu Y. Identification and molecular characterization of *FKF1* and *GI* homologous genes in soybean. PLoS One 2013;8:e79036.

Li J, Li G, Wang H, Wang Deng X. Phytochrome signaling mechanisms. *Arabidopsis* Book. 2011;9:e0148.

Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang SS, Zuo Q, Shi XH, Li YF, Zhang WK, Hu Y, Kong G, Hong HL, Tan B, Song J, Liu ZX, Wang Y, Ruan H, Yeung CK, Liu J, Wang H, Zhang LJ, Guan RX, Wang KJ, Li WB, Chen SY, Chang RZ, Jiang Z, Jackson SA, Li R, Qiu LJ. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol. 2014;32:1045-52.

Lian HL, He SB, Zhang YC, Zhu DM, Zhang JY, Jia KP, Sun SX, Li L, Yang HQ. Blue-light-dependent interaction of cryptochrome 1 with SPA1 defines a dynamic signaling mechanism. Genes Dev. 2011;25:1023-8.

Lin C. Photoreceptors and regulation of flowering time. Plant Physiol. 2000;123: 39–50.

Liscum E, Hangarter RP. *Arabidopsis* mutants lacking blue light-dependent inhibition of hypocotyl elongation. Plant Cell 1991;3:685-94.

Liu B, Fujita T, Yan ZH, Sakamoto S, Xu D, Abe J. QTL mapping of domestication-related traits in soybean (*Glycine max*). Ann. Bot. (Lond.) 2007;100:1027–1038.

Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, Abe J. Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. Genetics 2008a;180:995–1007.

Liu B, Zuo Z, Liu H, Liu X, Lin C. *Arabidopsis* cryptochrome 1 interacts with SPA1 to suppress COP1 activity in response to blue light. Genes Dev. 2011;25:1029-34.

Liu H, Yu X, Li K, Klejnot J, Yang H, Lisiero D, Lin C. Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in *Arabidopsis*. Science 2008b;3228:1535–9.

Liu J, He Y, Amasino R, Chen X. siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. Genes Dev. 2004;18:2873-8.

Liu LJ, Zhang YC, Li QH, Sang Y, Mao J, Lian HL, Wang L, Yang HQ.

COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in *Arabidopsis*. Plant Cell 2008c;20:292–306.

Lu Q, Zhao L, Li D, Hao D, Zhan Y, Li W. A *GmRAV* ortholog is involved in photoperiod and sucrose control of flowering time in soybean. PLoS One 2014;9:e89145.

Macinnes CB, Albert LS. Effect of light intensity and plant size on rate of development of early boron deficiency symptoms in tomato root tips. Plant Physiol. 1969;44:965-7.

Mathews S. Evolutionary studies illuminate the structural-functional model of plant phytochromes. Plant Cell. 2010;22:4-16.

Mathieu J, Yant LJ, Murdter F, Kuttner F, Schmid M. Repression of Flowering by the miR172 target *SMZ*. PLoS Biol 2009;7:e1000148.

Matsumura H, Kitajima H, Akada S, Abe J, Minaka N, Takahashi R. Molecular cloning and linkage mapping of cryptochrome multigene family in soybean. Plant Genome 2009;2:1–11.

McBlain BA and Bernard RL. A new gene affecting the time of flowering and maturity in soybean. J. Hered. 1987;78:160–2.

Meng Q, Zhang C, Huang F, Gai J, Yu D. Molecular cloning and characterization of a *LEAFY-like* gene highly expressed in developing soybean seeds. Seed Sci. Res. 2007;17:297–302.

Meng Y, Li H, Wang Q, Liu B, Lin C. Blue light-dependent interaction between cryptochrome2 and CIB1 regulates transcription and leaf senescence in soybean. Plant

Cell 2013;25:4405-20.

Michaels SD and Amasino RM. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell 1999;11:949-56.

Michaels SD, He Y, Scortecci KC, Amasino RM. Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 2003;100:10102-7.

Miyoshi K, Ito Y, Serizawa A, Kurata N. *OsHAP3* genes regulate chloroplast biogenesis in rice. Plant J. 2003;36:532–40.

Monte E, Alonso JM, Ecker JR, Zhang Y, Li X, Young J, Austin-Phillips S, Quail PH. Isolation and characterization of phyC mutants in *Arabidopsis* reveals complex crosstalk between phytochrome signaling pathways. Plant Cell 2003;15:1962-80.

Na X, Jian B, Yao W, Wu C, Hou W, Jiang B, Bi Y, Han T. Cloning and functional analysis of the flowering gene *GmSOC1-like*, a putative *SUPPRESSOR OF OVEREXPRESSION CO1*/*AGAMOUS-LIKE 20* (*SOC1*/*AGL20*) ortholog in soybean. Plant Cell Rep. 2013;32:1219–29.

Nagatani A. Light-regulated nuclear localization of phytochromes. Current Opinion in Plant Biol. 2004;7:708-11.

Nan H, Cao D, Zhang D, Li Y, Lu S, Tang L, Yuan X, Liu B, Kong F. *GmFT2a* and *GmFT5a* redundantly and differentially regulate flowering through interaction with and upregulation of the bZIP transcription factor *GmFDL19* in soybean. PLoS One 2014;9:e97669.

Napoli C, Lemieux C, Jorgensen R. Introduction of a chimeric chalcone synthase

gene into *Petunia* results in reversible co-suppression of homologous genes in trans. Plant Cell 1990;2:279-89.

Ogiso E, Takahashi Y, Sasaki T, Yano M, Izawa T. The role of casein kinase II in flowering time regulation has diversified during evolution. Plant Physiol. 2010;152:808–20.

Okamoto K, Yanagi T, Takita S, Tanaka M, Higuchi T, Ushida Y, Watanabe H. Development of plant growth apparatus using blue and red LED as artificial light source. Acta Hortic. 1996;440:111-6.

Owen FV. Inheritance Studies in Soybeans. II. Glabrousness, Color of Pubescence, Time of Maturity, and Linkage Relations. Genetics 1927;12:519-29.

Perilleux C, Bernier G, Kinet J-M. Circadian rhythms and the induction of flowering in the long-day grass *Lolium temulentum* L. Plant, Cell Environ. 1994; 17:755-61.

Pin PA and Nilsson O. The multifaceted roles of FLOWERING LOCUS T in plant development. Plant Cell Environ. 2012;35:1742–55.

Putterill J, Robson F, Lee K, Simon R, Coupland G. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 1995;80:847–57.

Quach TN, Nguyen HT, Valliyodan B, Joshi T, Xu D, Nguyen HT. Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response. Mol. Genet. Genomics. 2015;290:1095-115.

Ray JD, Hinson K, Mankono EB, Malo FM. Genetic control of a long-juvenile trait in soybean. Crop Sci. 1995;35:1001–6.

Roberts EH and Summerfield RJ. Measurement and predic tion of flowering in annual crops. In J.G. Atherton (ed.) Manipulation of flowering. Butterworths, London. 1987;17–50.

Roden LC, Song HR, Jackson S, Morris K, Carre IA. Floral responses to photoperiod are correlated with the timing of rhythmic expression relative to dawn and dusk in *Arabidopsis*. Proc. Natl. Acad. Sci. U S A. 2002;99:13313-8.

Saindon G, Voldeng HD, Beversdorf WD, Buzzell RI. Genetic control of long daylength response in soybean. Crop Sci. 1989;29:1436–9.

Salisbury FB. Photoperiodism. Hortic Rev (Am Soc Hortic Sci). 1985;4:66-105.

Sancar A. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. Chem. Rev. 2003;103:2203–37.

Sawa M, Kay SA. GIGANTEA directly activates *Flowering Locus T* in *Arabidopsis thaliana*. Proc. Natl Acad. Sci. USA 2011;108:11698–703.

Sawa M, Nusinow DA, Kay SA, Imaizumi T. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. Science 2007;318:261-5.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA.

Genome sequence of the palaeopolyploid soybean. Nature 2010;463:178-83.

Schuerger AC, Brown CS, Stryjewski EC. Anatomical features of pepper plants (Capsicum annuum L.) grown under red light-emitting diodes supplemented with blue or far-red light. Ann. Bot. 1997;79:273-82.

Searle I, He Y, Turck F, Vincent C, Fornara F, Kröber S, Amasino RA, Coupland G. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. Genes Dev. 2006;20:898–912.

Sharrock RA. The phytochrome red/far-red photoreceptor superfamily. Genome Biol. 2008;9:230.

Shibuya K, Fukushima S, Takatsuji H. RNA-directed DNA methylation induces transcriptional activation in plants. Proc. Natl. Acad. Sci. USA 2009;106:1660–5.

Shoemaker RC, Schlueter J, Doyle JJ. Paleopolyploidy and gene duplication in soybean and other legumes. Curr. Opin. Plant Biol. 2006;9:104–9.

Shrestha R, Gómez-Ariza J, Brambilla V, Fornara F. Molecular control of seasonal flowering in rice, *arabidopsis* and temperate cereals. Ann. Bot. 2014;114:1445–58.

Somers DE, Devlin PF, Kay SA. Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock. Science 1998;282:1488–90.

Somers DE. The physiology and molecular bases of the plant circadian clock. Plant Physiol. 1999;121:9-21.

Song Qj, Jia GF, Zhu YL, Grant D, Nelson RT, Hwang EY, DY, Hyten DY, Hyten DY,

Cregan PB. Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci. 2010;50:1950–60.

Song YH, Estrada DA, Johnson RS, Kim SK, Lee SY, MacCoss MJ, Imaizumi T. Distinct roles of FKF1, Gigantea, and Zeitlupe proteins in the regulation of Constans stability in *Arabidopsis* photoperiodic flowering. Proc. Natl. Acad. Sci. U S A. 2014;111:17672-7.

Song YH, Smith RW, To BJ, Millar AJ, Imaizumi T. FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. Science 2012;336:1045–9.

Stasinopoulos TC, Hangarter RP. Preventing photochemistry in culture media by long-pass light filters alters growth of cultured tissues. Plant Physiol. 1990;93:1365-9.

Sun H, Jia Z, Cao D, Jiang B, Wu C, Hou W, Liu Y, Fei Z, Zhao D, Han T. *GmFT2a*, a soybean homolog of *FLOWERING LOCUS T*, is involved in flowering transition and maintenance. PLoS One 2011;6:e29238.

Takano M, Inagaki N, Xie X, Yuzurihara N, Hihara F, Ishizuka T, Yano M, Nishimura M, Miyao A, Hirochika H, Shinomura T. Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. Plant Cell 2005;17:3311–25.

Tanaka Y, Tomita K, Yumoto S, Kurosaki H, Yamazaki H, Suzuki C, Matsukawa I, Tsuchiya T, Shirai K. A new soybean variety Yukihomare. Bull. Hokkaido Prefect.Agric. Exp. Stn. 2003;84:13-24.

Thakare D, Kumudini S, Dinkins RD. The alleles at the *E1* locus impact the expression pattern of two soybean *FT-like* genes shown to induce flowering in

*Arabidopsis*. Planta 2011;234:933-43.

Tibbitts TW, Morgan DC, Warrington IJ. Growth of lettuce, spinach, mustard and wheat plants under four combinations of high-pressure sodium, metal halide, and tungsten halogen lamps at equal PPFD. J. Amer. Soc. Hort. Sci. 1983;108:622-30.

Tiwari SB, Shen Y, Chang HC, Hou Y, Harris A, Ma SF, McPartland M, Hymus GJ, Adam L, Marion C, Belachew A, Repetti PP, Reuber TL, Ratcliffe OJ. The flowering time regulator CONSTANS is recruited to the *FLOWERING LOCUS T* promoter via a unique *cis*-element. New Phytol. 2010;187:57–66.

Tsubokura Y, Matsumura H, Xu M, Liu B, Nakashima H, Anai T, Kong F, Yuan X, Kanamori H, Katayose Y, Takahash R, Harada K, Abe J. Genetic variation in soybean at the maturity locus E4 is involved in adaptation to long days at high latitudes. Agronomy 2013;3:117-34.

Tsubokura Y, Watanabe S, Xia Z, Kanamori H, Yamagata H, Kaga A, Katayose Y, Abe J, Ishimoto M, Harada K. Natural variation in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. Ann. Bot. 2014;113:429–41.

Turck F, Roudier F, Farrona S, Martin-Magniette ML, Guillaume E, Buisine N, Gagnot S, Martienssen RA, Coupland G, Colot V. *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genet 2007;3:e86.

Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science 2004;303:1003–06.

Varagona MJ, Purugganan M, Wessler SR. Alternative splicing induced by insertion

of retrotransposons into the maize waxy gene. Plant Cell 1992;4:811-20.

Wang M, Xu X, Zhang X, Sun S, Wu C, Hou W, Wang Q, Han T. Functional analysis of *GmCPDs* and investigation of their roles in flowering. PLoS One 2015;10:e0118476.

Wang Z, Zhou Z, Liu Y, Liu T, Li Q, Ji Y, Li C, Fang C, Wang M, Wu M, Shen Y, Tang T, Ma J, Tian Z. Functional evolution of phosphatidylethanolamine binding proteins in soybean and *Arabidopsis*. Plant Cell 2015;27:323-36.

Watanabe S, Harada K, Abe J. Genetic and molecular bases of photoperiod responses of flowering in soybean. Breed. Sci. 2012;61:531-43.

Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y, Yamanaka N, Takahashi R, Ishimoto M, Anai T, Tabata S, Harada K. Map-based cloning of the gene associated with the soybean maturity locus *E3*. Genetics 2009;182:1251-62.

Watanabe S, Tajuddin T, Yamanaka N, Hayashi M, Harada K. Analysis of QTLs for reproductive development and seed quality traits in soybean using recombinant inbred lines. Breed. Sci. 2004;54:399–407.

Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Harada K. A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. Genetics 2011;188:395-407.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 2007;8:973–82.

Wilczek AM, Burghardt LT, Cobb AR, Cooper MD, Welch SM, Schmitt J. Genetic

and physiological bases for phenological responses to current and predicted climates. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2010;365:3129-47.

Wu F, Price BW, Haider W, Seufferheld G, Nelson R, Hanzawa Y. Functional and evolutionary characterization of the *CONSTANS* gene family in short-day photoperiodic flowering in soybean. PLoS One 2014;9:e85754.

Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lü S, Wu H, Tabata S, Harada K. Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. Proc. Natl. Acad. Sci. USA 2012;109:E2155-64.

Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 1990;9:3353–62.

Xu M, Xu Z, Liu B, Kong F, Tsubokura Y, Watanabe S, Xia Z, Harada K, Kanazawa A, Yamada T, Abe J. Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. BMC Plant Biol. 2013;13:91.

Xu M, Yamagishi N, Zhao C, Takeshima R, Kasai M, Watanabe S, Kanazawa A, Yoshikawa N, Liu B, Yamada T, Abe J. Soybean-specific maturity gene *E1* family of floral repressors controls night-break responses through down-regulation of *FLOWERING LOCUS T* orthologs. Plant Physiol. 2015;168:1735-46.

Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, Zhang Q. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nat. Genet. 2008;40:761–7.

Xue ZG, Zhang XM, Lei CF, Chen XJ, Fu YF. Molecular cloning and functional

analysis of one *ZEITLUPE* homolog *GmZTL3* in soybean. Mol. Biol. Rep. 2012;39:1411-8.

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T and Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. Proc. Natl. Acad. Sci. USA 2003;100:6263–8.

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S and Dubcovsky J. The wheat and barley vernalization gene *VRN3* is an ortholog of *FT*. Proc. Natl. Acad. Sci. USA 2006;103:19581–6.

Yamanaka N, Ninomiya S, Hoshi M, Tsubokura Y, Yano M, Nagamura Y, Sasaki T, Harada K. An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. DNA Res. 2001;8:61–72.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. Plant Cell 2000;12:2473–84.

Zhai H, L ü S, Wu H, Zhang Y, Zhang X, Yang J, Wang Y, Yang G, Qiu H, Cui T, Xia Z. Diurnal expression pattern, allelic variation, and association analysis reveal functional features of the *E1* gene in control of photoperiodic flowering in soybean. PLoS One 2015;10:e0135909.

Zhang J, Song Q, Cregan PB, Nelson RL, Wang X, Wu J, Jiang GL. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. BMC Genomics 2015;16:217.

Zhang Q, Li H, Li R, Hu R, Fan C, Chen F, Wang Z, Liu X, Fu Y, Lin C. Association

of the circadian rhythmic expression of *GmCRY1a* with a latitudinal cline in photoperiodic flowering of soybean. Proc. Natl. Acad. Sci. USA 2008;1058:21028–33.

Zhao X, Cao D, Huang Z, Wang J, Lu S, Xu Y, Liu B, Kong F, Yuan X. Dual functions of *GmTOE4a* in the regulation of photoperiod-mediated flowering and plant morphology in soybean. Plant Mol. Biol. 2015;88:343-55.

Zhong X, Dai X, Xu J, Wu H, Liu B, Li H. Cloning and expression analysis of *GmGAL1*, *SOC1* homolog gene in soybean. Mol. Biol. Rep. 2012;39:6967-74.

## Lists for figures

Figure 3.1 Flowering time in the progeny of the cross between Toyomusume and Harosoy. ..... 22

Figure 4.1 Fine mapping of the *E9* locus and annotated genes in the delimited genomic region. ..... 29
Figure 4.2 DNA polymorphisms detected in the *FT2a* genomic region for four soybean cultivars or breeding line and a wild soybean accession. ..... 32
Figure 4.3 *FT2a* expression in the progeny of $F_2$ plants from a cross between Toyomusume and Harosoy. ..... 33
Figure 4.4 DNA polymorphisms that discriminate between the *FT2a* alleles and *FT2a* transcript abundance in their NILs. ..... 35
Figure 4.5 *FT2a* transcript abundance in photoperiod-insensitive *e3 e4* cultivars under SD conditions. ..... 36

Figure 5.1 *FT2a* transcript abundance in two sets of NILs for *FT2a-TO* and *FT2a-HY* haplotype. ..... 42
Figure 5.2 *FT2a* RNA processing in the first intron with *SORE-1* insertion. ..... 43
Figure 5.3 Transcript abundance of different alleles. ..... 44
Figure 5.4 DNA methylation in the *FT2a* genomic region. ..... 45

## List for tables

Table 2.1 Correspondences between different naming systems for *GmFTs*, *GmCOs*, and *GmGIs*. .....18

Table 3.1 Association tests of SSR marker genotypes with flowering time. .....23

Table 4.1 Association tests of ID1, a tagging marker of *E9*, with flowering. .....30

Table 4.2 Allelic variation at the *E9* (*FT2a*) locus in early-maturing soybean cultivars and land races/pure-line selections (*e9* is a *SORE-1*-inserted *FT2a* allele). 37

## List for abbreviations

| | |
|---|---|
| AGL24 | AGAMOUS-LIKE24 |
| AP1 | APETALA1 |
| ATC | ARABIDOPSIS THALIANA CENTRORADIALIS |
| BFT | BROTHER OF FT AND TFL1 |
| CAL | CAULIFLOWER |
| CO | CONSTANT |
| COP1 | CONSTITUTIVE PHOTOMORPHOGENIC 1 |
| CIB1 | cryptochrome-interacting basic-helix-loop-helix 1 |
| CRYs | cryptochromes |
| CDF | CYCLING DOF FACTOR |
| DELLAs | DELLA growth inhibitors |
| DCL1 | Dicer Like Enzyme1 |
| DsRNA | double-strand RNA |
| SLY1 | F-box proteins SLEEPY1 |
| FKF1 | FLAVIN KELCHF BOX 1 |
| FLC | FLOWERING LOCUS C |
| FLC | FLOWERING LOCUS C |
| FT | FLOWERING LOCUS T |
| FRI | FRIGIDA |
| FUL | FRUITFULL |
| GA | gibberellin |
| GID1 | GIBBERELLIN INSENSITIVE DWARF1 |
| GI | GIGANTEA |
| Hd1 | Heading date 1 |
| Hd3a | Heading date 3a |
| LHP1 | HETEROCHROMATIN PROTEIN1 |
| HOS1 | HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 1 |
| LFY | LEAFY |
| LD | long day |

| | |
|---|---|
| lsiRNAs | long siRNAs |
| MiRNAs | microRNAs |
| MFT | MOTHER OF FT AND TFL1 |
| natsiRNAs | natural antisense transcript-derived siRNAs |
| NF-Y | NUCLEAR FACTOR Y |
| PEBP | phosphatidylethanolamine binding protein |
| PIF4 | PHYTOCHROME-INTERACTING FACTOR 4 |
| PHYs | phytochromes |
| casiRNAs | plant-specific cis-acting siRNAs |
| PRC | polycomb repressive complex |
| PTGS | post-transcriptional gene regulation |
| QTLs | quantitative trait loci |
| R/FR | red /far-red light |
| hc-siRNAs or ra-siRNAs | repeat-associated siRNAs |
| R stages | reproductive stages |
| RNAi | RNA interference |
| RdDM | RNA-directed DNA methylation |
| SAM | shoot apical meristem |
| SD | short day |
| SVP | SHORT VEGETATIVE PHASE |
| SVP | SHORT VEGETATIVE PHASE |
| HpRNAs | single-stranded hairpin RNAs |
| SiRNAs | small interfering RNAs |
| SNZ | SNEEZY |
| SOC1 | SUPPRESSOR OF CONSTANS1 |
| SPA1 | SUPPRESSOR OF HYTOCHROME A |
| TEM | TEMPRANILLO |
| TFL1 | TERMINAL FLOWER 1 |
| tasiRNAs | trans-acting siRNAs |
| TGS | transcriptional gene silencing |
| TE | transposable elements |

| | |
|---|---|
| V stages | vegetative stages |
| VRN1 | VERNALIZATION1 |
| VRN2 | VERNALIZATION2 |
| VRN3 | VERNALIZATION3 |

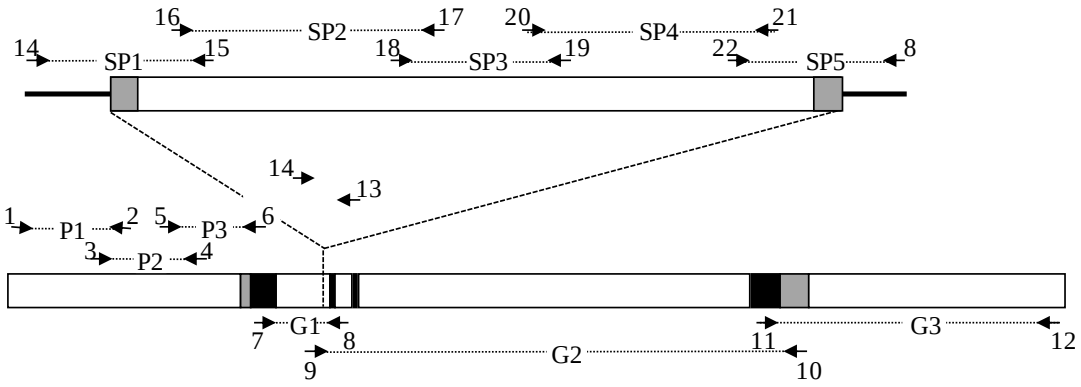
**Additional table 1. Primers, PCR conditions and expected amplified fragment sizes for allele-specific DNA markers.**

| Locus / Alleles | Primer Sequence (5' − 3') | | Annealing tempreture (°) | Size of fragment (bp) | |
|---|---|---|---|---|---|
| *E1* Xu *et al* . (2013) | F | CACTCAAATTAAGCCCTTTCA | 60 | *e1-as* | 222 |
| *e1-as* /*e1-nl* | R | TTCATCTCCTCTTCATTTTTGTTC | | *e1-nl* | no products |
| *E3* Tsubokura *et al* . (2013) | E3_08557FW | TGGAGGGTATTGGATGATGC | 58 | *E3* | 558 |
| *E3* /*e3-tr* | E3Ha_1000RV | CGGTCAAGAGCCAACATGAG | | *e3-tr* | 275 |
| | e3tr_0716RV | GTCCTATACAATTCTTTACGACC | | | |
| *E4* Xu *et al* . (2013) | F | AGACGTAGTGCTAGGGCTAT | 55 | *E4* | 1,229 |
| *E4* /*e4-SORE-1* | R2 | GCATCTCGCATCACCAGATCA | | *e4-SORE-1* | 837 |
| | R3 | GCTCATCCCTTCGAATTCAG | | | |

**Additional table 2. Sequences of primers used in expression and RNA processing analyses of *FT2a*. Targeted *FT2a* regions are shown in Figure 5.1 (a to c) and Figure 5.2 (a to f).**

| Areas targeted | Primer Sequence (5' − 3') | | Annealing tempreture (℃) | Extension times |
|---|---|---|---|---|
| FT2a-a | F | TAAACTAGTGTGCACACTATCCC | 60℃ | 25s |
| | R | TATAGAAGTTCCTGAGGTCATCACCA | | |
| FT2a-b | F | GGGGAGTAATTGGGGATGTATTGG | 60℃ | 25s |
| | R | AAACTAGCCCCTGTTGTTGC | | |
| FT2a-c | F | GGATTGCCAGTTGCTGCTGT | 60℃ | 25s |
| | R | GAGTGTGGGAGATTGCCAAT | | |
| | | | | |
| tubulin | F | GAGAAGAGTATCCGGATAGG | 60℃ | 25s |
| | R | GAGCTTGAGTGTTCGGAAAC | | |
| | | | | |
| a | F | GGGGATAATTGGGGATGTATTGG | 60℃ | 20s |
| | R | CAAAAAGAGTACTTGGACAA | | |
| b | F | GGGGAGTAATTGGGGATGTATTGG | 60℃ | 1min20s |
| | R | CACCATCTAGAGAGTGGAAGAGAGAG | | |
| c | F | GGGGAGTAATTGGGGATGTATTGG | 60℃ | 1min36s |
| | R | ACCAAGAATAAACATTGGGT | | |
| d | F | AACTTCATGATCAAGGTGCT | 60℃ | 2min |
| | R | AAACTAGCCCCTGTTGTTGC | | |
| e | F | GCTCTCTCTCTTCCACTCTCTAGATGG | 60℃ | 40s |
| | R | AAACTAGCCCCTGTTGTTGC | | |
| f | F | GGGTAAATATCGGTGGTGATGAC | 64℃ | 8min |
| | R | AAACTAGCCCCTGTTGTTGC | | |

**Additional table 3. Genomic positions and sequences (5′ – 3′) of primers used in sequencing of the *FT2a* genomic region and Ty1/copia-like retrotransposon, SORE-1. Primer 13 was used for nested PCRs for genome libraries of Toyomusume. The PCR products amplified with primers 14 and 8 were used as temperate in PCR amplifications for the SORE-1 and its flanking region (SP1 to SP5).**




| Areas targeted | | Primer sequence (5' – 3') | Annealing tempreture (℃) | Extension times |
|---|---|---|---|---|
| P1 | 1 | AACAAAGAGAGAGAGAGGAAC | 57℃ | 1min |
| | 2 | GCTTTTCCCTCTCAATTAAGAATC | | |
| P2 | 3 | CCTACAAACAAAATTGATGATGTTTCTAG | 57℃ | 1min |
| | 4 | TTCAAACATTTTCACATCCCTTCC | | |
| P3 | 5 | AGTCATACATTACGGAATCTAGTT | 57℃ | 45s |
| | 6 | CTTCCACTAGGCATGGGATA | | |
| G1 | 7 | GGGGAGTAATTGGGGATGTATTGG | 56℃ | 1min20s |
| | 8 | CTAGGTGCATCGGGATCAAC | | |
| G2 | 9 | TATATTCTTAGTGCAATCGAGGATCATTAG | 62℃ | 4min |
| | 10 | GGGAGATTGCCAATTAATATTTCTGAAA | | |
| G3 | 11 | ATTTCGTCAACTGGGTAGGGAG | 55℃ | 3min |
| | 12 | GCAGATTCTCATAACGTCGGAG | | |
| Genomic walker primer | 13 | AGTCACCAACCTGACACATTGAGAGAAGAT | 63℃ | 4min |
| Full SORE-1 | 14 | ACCCTCTCAAGTGGACATGT | 62℃ | 8min30s |
| | 8 | CTAGGTGCATCGGGATCAAC | | |
| SP1 | 14 | ACCCTCTCAAGTGGACATGT | 60℃ | 1min30s |
| | 15 | CGATAACTCAGCTGCTGTCT | | |
| SP2 | 16 | CCTTAAATTTAAACGGTGGC | 55℃ | 2min |
| | 17 | ACATGATTGGCATTCCACAA | | |
| SP3 | 18 | CAGGATCATTTTCAGCTCCA | 55℃ | 1min30s |
| | 19 | TTCCTCTATGTTATGGCACC | | |
| SP4 | 20 | CATCCAATGCTTCGTATTTG | 55℃ | 1min30s |
| | 21 | GACATAGATTATGCTATAAGG | | |
| SP5 | 22 | AACTTCATGATCAAGGTGCT | 60℃ | 2min |
| | 8 | CTAGGTGCATCGGGATCAAC | | |

**Additional table 4. Sequences of primers used in methylation analysis of *FT2a*. Targeted regions are presented in Figure 5.4.**

| Areas targeted | Primer Sequence (5' − 3') | | Annealing tempreture (℃) | Extension times |
|---|---|---|---|---|
| a | F | TGAAGTCTCTGAACATGCACGC | 58℃ | 1min |
| | R | CACCTTTATATATTCCTCATATCTGTC | | |
| b | F | AAAGGTGAAATATATATTGTTGG | 64℃ | 40s |
| | R | CTTCCACTAGGCATGGGATA | | |
| c | F | TATAGAAATTTCAGAAATTTCTAGCAGCG | 60℃ | 40s |
| | R | TTCAAACATTTTCACATCCCTTCC | | |
| d | F | GATTCTTAATTGAGAGGGAAAAGC | 60℃ | 40s |
| | R | AAAGGAGCAGCAAAACGCTA | | |
| e | F | GGAATCGAGGCTATTGACTA | 58℃ | 1min |
| | R | CAAAAAGAGTACTTGGACAA | | |
| f | F | ACCAAGCTAGAATAATTTTTGTGAG | 60℃ | 40s |
| | R | TGAGTGGTGGGGTTTTCTTT | | |
| g | F | ACCCTCTCAAGTGGACATGT | 64℃ | 40s |
| | R | AAACTAGCCCCTGTTGTTGC | | |
| h | F | ACACGCCATGAATGCAAACA | 64℃ | 40s |
| | R | GCAAGCAAACATTGTGCGTT | | |
| i | F | CCGAATTCTTAAAGTGCATGCA | 64℃ | 40s |
| | R | GCCAGAAACACTCATACATACT | | |
| j | F | GGACAGCAGTAGTAGGACCA | 64℃ | 1min35s |
| | R | GGGAGATTGCCAATTAATATTTCTGAAA | | |
| k | F | TGATAGTCTATGGGATTGTGC | 64℃ | 40s |
| | R | CCAAGCTTCCAACCGTGAAA | | |
| S1 | F | AATATAGTAGTAGTTTCTAGCTT | 60℃ | 40s |
| | R | ACCAAGAATAAACATTGGGT | | |
| S2 | F | TATATCATGATGCCTGTGGG | 64℃ | 1min35s |
| | R | CTAGGTGCATCGGGATCAAC | | |
| S3 | F | GCTCTCTCTCTTCCACTCTCTAGATGG | 64℃ | 40s |
| | R | AAACTAGCCCCTGTTGTTGC | | |
| S4 | F | CCAACTTGATTTCATTTCAT | 64℃ | 1min35s |
| | R | GCAAGCAAACATTGTGCGTT | | |

## Acknowledgements

Firstly, I would like to thank my supervisor Professor Jun Abe. Thanks for giving me the opportunity to participate in this great research area, and also thanks for his guidance, great support and kind advice throughout this research. I learnt from him a great deal, not only his exceptional scientific knowledge but also his extraordinary human qualities of kindness. Additionally, I also would like to thank my Associate Professor Akira Kanazawa and Lecturer Tetsuya Yamada for their kind, availability and constructive suggestions, which were determinant for the accomplishment of my doctoral study.

I am very much grateful to Dr. Meilan Xu for the great support not only on research but also my life in Hokkaido. Thanks to all the collaboration partners in Division of Agrobiology and Bioresources, especially Mr. Ryoma Takeshima for RNA extraction and Mrs. Masako Sato for fine-mapping.

I gratefully acknowledge the financial support provided by China Scholarship Council to undertake my doctoral study. I also heartily thank Hokkaido University for waiving all of my tuition fees and giving me this opportunity to study in Japan.

2016/01/08

Zhao Chen