Molecular Genetic Study on a Late-Flowering Habit from a Thai Soybean Cultivar in Photoperiod-Insensitive Genetic Background

（非感光性遺伝背景に導入されたタイダイズ品種の晩期開花性に関する分子遺伝学的研究）

Hokkaido University Graduate School of Agriculture

Division of Agrobiology Doctoral Course

Fei SUN

CONTENTS

Chapter I Introduction

As one of important agronomic traits affecting crop yield, the regulation of flowering time has been studied for over one hundred years (Fornara et al., 2010). For flowering plants, the process of floral initiation marks the transformation of plants from vegetative growth to reproductive growth (Piñeiro and Coupland, 1998; Brachi et al., 2010; Andrés and Coupland, 2012), and this transformation usually occurs when environmental and endogenous factors are most conducive to reproductive success and seed production (Ma, 1998; Pidkowich et al., 1999; Fornara et al., 2010; Srikanth and Schmid, 2011). This appropriate timing is the result of a well-designed regulatory network that coordinates external stimuli with endogenous cues to induce expression of genes that initiate flower transformation at the shoot apical meristem (SAM) (King and Zeevaart, 1973; Lang et al., 1977; Corbesier et al., 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007).

The study of the floral initiation process with *Arabidopsis thaliana* as a model system has indicated that 306 genes are involved in the regulation of the flowering process (Bouché et al., 2016). These genes belong to six major pathways including the vernalization pathway, the autonomous pathway, the gibberellin pathway, the photoperiod pathway, the ambient temperature pathway, and the age pathway (Fornara et al., 2010; Wils and Kaufmann, 2017; Chen et al., 2018). The photoperiod and vernalization pathways regulate flowering time by monitoring seasonal changes in day length and temperature, respectively. The ambient temperature pathway is responsible for regulating the response to daily growth temperatures. Signals from six pathways eventually converge on the regulation of several key genes called "floral integrator genes", including *LEAFY* (*LFY*), *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF CONSTANS1* (*SOC1*), which rapidly promote flowering process (Fornara et al., 2010). The upstream genes of *FT* and *SOC1* are *FLOWERING LOCUS C* (*FLC*) and *CONSTANS* (*CO*). The expression of both *FT* and *SOC1* are inhibited by high levels of *FLC* expression, which is released by vernalization treatment or autonomous development process (Fornara et al., 2010). *CO* regulated by the circadian rhythm promotes the expression of *FT* and *SOC1* to participate in the photoperiod pathway. The expression of *CO* is related to the circadian rhythm and the phytochromes also affect the stability of CO protein. Under LD conditions, the

amount of CO protein is increased, and the expression of *FT* is induced (Valverde et al., 2004; Calviño et al., 2005; Fornara et al., 2010).

According to the different responses to photoperiod, plants can be divided into different types. Long-day (LD) plants such as *Arabidopsis thaliana*, *Spinacia oleracea* and *Hordeum vulgare* form flower buds when the daylength becomes longer over a critical daylength, whereas short-day (SD) plants such as soybeans (*Glycine max* (L.) Merr.) and *Oryza sativa* flowers when the daylength becomes shorter than the critical daylengh. Day-neutral (DN) plants such as *Solanum lycopersicum* can initiate flowering independently of the daylength (Kobayashi and Weigel, 2007; Srikanth and Schmid, 2011).

Soybean is widely cultivated in the world as food crops and raw materials for vegetable oil processing and livestock industry (Graham and Vance, 2003). According to the data from FAOSTAT, soybeans are cultivated in more than 90 countries worldwide. In 2017, the top five countries in the world for the soybean production quantity are the United States of America (USA), Brazil, Argentina, China and India. Nevertheless, the soybean yield in China and India was only about half and one-third of the USA. In many countries in Africa, such as South Africa, Nigeria, and Zambia, the total soybean production quantity is also among the high level in the world, but in the same way, the unit yield is very low (Supplemental Table 1). The main reason for low seed yields of soybean in the tropics may be insufficient vegetative growth due to early flowering and maturation. Early flowering is usually caused by SD and high temperatures (Board and Hall, 1984). Consequently, the study on reproductive phenology, in particular, an extended vegetative period under short photoperiods and/or high temperatures is significant and valuable for the soybean production in those areas.

Soybean possesses various extents of sensitivity to photoperiod; each variety can be cultivated only in a limited latitudinal region (Scott, 1983; Watanabe et al., 2012). Time to flowering and maturity is an important trait to determine the adaptability and productivity of soybean in a region (Xia et al., 2012). Reduced or lack of sensitivity to long daylengths is a key character for soybean to adapt to higher latitudinal environments. However, the photoperiod-insensitivity often results in a reduction of

the duration of vegetative growth and final yield. Accumulation of genes controlling earlier-flowering often produced limited growth and seed production due to the shortened vegetative phases. In contrast, repression of flowering in SD conditions, inductive phase of flowering in soybean, is indispensable for soybean cultivars adapted to lower latitudes with high temperatures to maintain a sufficient duration of vegetative growth (Destro et al., 2001).

Owen (1927) detected a major pair of maturity genes and named them *E* and *e*. Since then, eleven major genes, *E1*-*E11*, and *J* for long juvenile period have been reported to associate with flowering in soybean; *E1*, *E2* (Bernard, 1971), *E3* (Buzzell, 1971), *E4* (Buzzell and Voldeng, 1980), *E5* (McBlain and Bernard, 1987), *E6* (Bonato and Vello, 1999), *E7* (Cober and Voldeng, 2001), *E8* (Cober et al., 2010), *E9* (Kong et al., 2014; Zhao et al., 2016), *E10* (Samanfar et al., 2017), *E11* (Wang et al., 2019) and *J* (Ray et al., 1995). Until now, the responsible genes for *E1*-*E4*, *E9*, *E10* and *J* have been identified by map-based cloning and/or candidate gene approaches, and their functions have been characterized (Liu et al., 2008; Watanabe et al., 2009, 2011; Xia et al., 2012; Zhao et al., 2016; Samanfar et al., 2017; Yue et al., 2017; Lu et al., 2017). However, the molecular bases for other genes remain unknown.

E1 is located on chromosome 6 (Chr06). The E1 protein contains a putative binary nucleus localization signal and has a domain distantly related to the plant specific B3 domain (Xia et al., 2012). *E1* plays a key role in inhibiting soybean flowering and delaying maturation by negatively regulating the expression of major soybean *FT* orthologs, *FT2a* and *FT5a*. Moreover, Xu et al. (2015) determined that in a cultivar lacking the *E1* gene, viral-induced *E1-like-a* (*E1La*) and *E1Lb* silencing also up-regulated the expression of *FT2a* and *FT5a* and led to early flowering. Zhu et al. (2019) found a loss-of-function allele of *E1Lb* that promoted the flowering in LD condition. Therefore, it is considered that *E1La* and *E1Lb* have the similar functions to *E1* in adjusting the flowering time (Xu et al., 2015; Zhu et al., 2019).

E2, located on Chr10, is an orthologue of *Arabidopsis GIGANTEA* (*GI*) (Watanabe et al., 2011), which encodes a nuclear-localized membrane protein. GI interacts with FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (FKF1) to up-regulate the expression of CONSTANS (CO) through degradation of CYCLING

DOF FACTOR (CDF) and activate FT expression (Huq et al., 2000). GI also directly binds to a cis-element of *FT* to activate the expression (Sawa et al., 2007).

E3 and *E4* are *PHYTOCHROME A* (*PHYA*) genes encoding PHYA3 and PHYA2 located on Chr19 and Chr20, respectively (Liu et al., 2008; Watanabe et al., 2009). PHYA-E1 pathway and GI-CO pathway are the two major partly interrelated pathways that control the flowering under LD conditions in soybean (Cao et al., 2017).

The genetic effect of *E5* is similar to that of *E2* in flowering time and maturation (McBlain and Bernard, 1987). However, Dissanayaka et al. (2016) pointed out that the QTL corresponding to *E5* was not found in Harosoy × PI 80837 population, from which *E5* was originally identified. They suggested that there may be no unique *E5* gene (Dissanayaka et al., 2016).

E7 is tightly linked to *E1* and *T* (pubescence color) (Cober and Voldeng, 2001). The alleles at *E7* locus have the weakest effect among all known *E* loci, delaying flowering by 4-5 days only. Molnar et al. (2003) found that Satt319 and Satt100 were the most reliable SSR markers located close to or within the *E1-T-E7* region of C2 linkage group (Chr06). Rozenzweig et al. (2008) reported that field trials of isolines differing in genotypes at *E7* revealed a yield gain equal to 21% in *E7* while ripening only 6 days later in *E7* relative to *e7*.

E8 was determined to be located between Sat_404 and Satt136 in C1 linkage group (Chr04) by linkage analyses (Cober et al., 2010). It is involved in the control of sensitivity to incandescent-induced LD conditions, particularly under the double-recessive *e3e3e4e4* genetic background; *E8E8* controls later maturity and *e8e8* dose early maturity.

E9 was fine-mapped to a genomic region of 245-kb on Chr16 and identified as a gene for early flowering introduced from a wild soybean accession (Liu et al., 2007, Kong et al., 2014). Zhao et al. (2016) revealed that *E9* is *FT2a* and its recessive allele delays flowering through lower transcript abundance due to an insertion of Ty1/copia-like retrotransposon, *SORE-1*, in the first intron by using a cross between two early-maturing soybean cultivars, Harosoy and Toyomusume.

E10 is located at the end of Chr08; the *e10e10* genotype results in 5-10 days earlier maturity than *E10E10* (Samanfar et al., 2017). The protein–protein interaction analysis predicted that *FT4* is most likely a candidate gene for *E10* and plays function downstream of *E1* and *E2*.

E11 was identified in the progeny of the cross between soybean cultivars Minsoy and Archer (Wang et al., 2019). It is located on Chr07, and fine-mapped to a genome region of 138 kb, including 11 genes based on the Williams 82 reference genome (http://www.phytozome.net). Among them, *Glyma.07g048500*, *Glyma.07g049000*, and *Glyma.07g049200* are most likely candidate genes. *E11* significantly promoted the flowering and maturity times than *e11* under LD conditions (Wang et al., 2019).

J controls long-juvenile (LJ) trait in soybean (Ray et al., 1995). It is an ortholog of *Arabidopsis EARLY FLOWERING 3* (*ELF3*) located on Chr04 (Yue et al., 2017; Lu et al., 2017). *J* is genetically dependent on *E1*, and physically binds to the *E1* promoter to downregulate its transcription; it attenuates the inhibition of two important *FT* genes by *E1*, and promotes flowering in SD (Lu et al., 2017). *E6* also regulates LJ traits (Bonato and Vello, 1999). It is mapped on Chr04 near a single-nucleotide polymorphism marker tagging the *J* gene, suggesting that *E6* and *J* might be tightly linked or allelic to each other (Lu et al., 2017).

In addition to these major genes, QTL and association mapping analyses have revealed novel genetic factors involved in the variation of flowering times. Molecular and genetic mechanisms for these major genes and QTLs have been gradually disclosed. More than a hundred flowering time QTLs, distributed on 17 chromosomes, have been reported and registered in SoyBase (https://soybase.org/sbt/). Most of the loci associated with flowering time were detected on Chr11, Chr16 and Chr20, and are sensitive to photothermal conditions (Keim et al., 1990; Mansur et al., 1993; Lee et al., 1996; Orf et al., 1999; Tasma et al., 2001; Yamanaka et al., 2001; Chapman et al., 2003; Wang et al., 2004; Watanabe et al., 2004; Zhang et al., 2004; Funatsuki et al., 2005; Pooprompan et al., 2006; Githiri et al., 2007; Liu et al., 2007; Khan et al., 2008; Komatsu et al., 2007; Liu and Abe 2009; Cheng et al., 2011; Liu et al., 2011; Li et al., 2017; Liu et al., 2018). Cheng et al. (2011) reported that two major-effect QTLs, *qRP-c-1* and *qRP-l-1*, which mapped on C1 and L linkage groups, respectively, were

detected under LD condition, and might associate with *E8* or *CRY1a* and the maturity gene *E3* or *PHYA3*, respectively. Liu et al. (2011) identified the QTLs for the variation of flowering time segregated in different latitudinal environments in a cross between early and late-flowering Korean cultivars, of which two QTLs located in Chr04 and Chr16 were specific to the SD condition of lower latitudes. Lu et al. (2015) reported that two maturity loci, *E2* and *E3*, and two novel QTLs located in Chr16 were involved in the genetic variation of flowering time segregated under SD conditions in a cross between photoperiod-insensitive cultivar AGS292 and a Thai LJ cultivar K3. Furthermore, the QTL mapping revealed two QTLs, one corresponding to the *E6*/*J* or linked novel gene and the other located in Chr02 (D1b linkage group), in the flowering time under SD segregated in a cross between LJ cultivars, Paranagoiana and PI159925 (Li et al., 2017).

In this study, I characterized photo-thermal responses of a late-flowering (LF) habit introduced from a Thai cultivar K3 into the photoperiod insensitive genetic background, determined the genetic basis using QTL mapping, and discussed possible molecular mechanisms underlying the LF habit, based on the results obtained from sequencing and expression analyses. In Chapter 2, I evaluated the flowering behavior and expression profiles for flowering genes of LF line in various photo-thermal conditions. In Chapter 3, I determined the genetic basis for the LF habit using QTL analysis. In Chapter 4, I compared the sequences of candidate genes for the QTLs by re-sequencing and genome walking analyses. In Chapter 5, I analyzed the binding of a soybean orthologue of bHLH-transcriptional factor MYC2 to a cis-element detected in the *FT5a* genome sequence of K3. I also analyzed the binding ability of the E1 family proteins to the DNA sequences by yeast-one-hybrid assays. Based on the results obtained, I discussed the probable molecular mechanism of the LF habit from K3 and the roles of *E1* family genes as floral repressors of flowering in soybean.

Chapter II Photo-Thermal Responses and Expression Profiles of Flowering Genes in a Photoperiod Insensitive Line with Late-Flowering Habit

2.1 Background and Purpose

The photo-thermal regulation of flowering and maturation determines the adaptability and productivity of crops. Diverse genotypic combinations at several flowering loci enable crops to maximize their productivity under various environmental conditions.

The PHYA-E1 module is a main regulator in the photoperiodic flowering of soybean, as demonstrated by night-break responses (Xu et al., 2015; Cao et al., 2017). The *E1* gene and its homologs, *E1La* and *E1Lb*, are legume-specific putative transcription factors that repress the transcription of soybean *FT* orthologs, *FT2a* and *FT5a*, during LDs under the control of phytochrome A proteins encoded by *E3* and *E4* (Xia et al., 2012; Xu et al., 2015). In particular, a reduced or lack of sensitivity to LD lengths is conferred by loss-of-function alleles at the *E1*, *E1Lb*, *E3*, and *E4* loci, as well as a transcriptionally upregulated *FT5a* allele (Xu et al., 2013; Takeshima et al., 2016; Zhu et al., 2019).

The delayed flowering under SD conditions has been referred to as LJ in soybean. Soybean cultivars with LJ periods can produce greater seed yields, compared with those having normal juvenile periods, by retaining sufficient vegetative growth levels under SD conditions (Destro et al., 2001). The LJ period appears to be controlled by several recessive genes (reviewed by Destro et al., 2001), including two major genes *j* (Ray et al., 1995) and *e6* (Bonato and Vello, 1999).

Lu et al. (2015) found a large variation of flowering time in recombinant inbred lines (RILs) developed from a cross between a photoperiod-insensitive cultivar AGS292 and a Thai late-flowering cultivar K3. They found QTLs corresponding to three maturity genes, *E2*, *E3* and *E4*, and novel QTLs in Chr16 in a wide range of latitudes. Approximately one-fourth of the RILs possessed a double recessive genotype (*e3*/*e4*) at *E3* and *E4* loci conferring the insensitivity of flowering to far-red (FR) light enriched LD (Lu et al., 2015).

In this chapter, I characterized the late-flowering (LF) habit of K3 introduced into a photoperiod-insensitive genetic background under different photo-thermal conditions. Since K3 does not possess the *j* allele, it is possible to identify novel QTLs for the LF habit. I selected the latest-flowering line (AK16) of photoperiod-insensitive RILs carrying the double recessive genotype at *E3* and *E4* loci (*e3*/*e4*) and evaluated flowering time and expression profiles for two floral integrators, *FT2a* and *FT5a*, and flowering genes negatively controlling flowering time as possible candidates for LF habits.

2.2 Materials and Methods

2.2.1 Plant materials

AGS292 and AK16 were used in this study. AK16 was the latest flowering line of 16 photoperiod-insensitive lines having the *e3*/*e4* genotype in the recombinant inbred line (RIL) population developed from a cross between AGS292 and the Thai late-flowering cultivar K3 (Pooprompan et al., 2006; Lu et al., 2015). The two lines possessed the same maturity genotype at *E1, E3, E4* and *E9* (*E1*/*e3*/*e4*/*E9*) but differed in the *E2* genotype; AGS292 and AK16 possessed the *e2* and *E2* alleles, respectively. AK16 possessed alleles from K3 at 120 (44%) of 276 simple sequence repeat (SSR) or insertions and deletions (indels) markers. The genomic region from K3 in AK16 covered the QTLs for flowering time under SD conditions in linkage groups (LG) J (Chr16), harboring *FT2a*, and *FT5a* (Lu et al., 2015).

2.2.2 Growing condition

AGS292 and AK16 were cultivated in different photoperiod and thermal conditions. Four day-length conditions of 8, 12, 16 and 20 h were set in a greenhouse in winter seasons of 2016 to 2018. High intensity discharge lamps (HONDA-T, Panasonic Co, Osaka, Japan) were used as light sources during the daytime. The average photosynthetically active photon flux density was 120 μmol m^{-2} s^{-1}, and the red-to-far red (R:FR) ratio was 4.5 at 1 m below the light source. Air temperatures in

the greenhouse were adjusted to 25℃, with fluctuations from a minimum of 22℃ to a maximum of 28℃.

Two lines were further cultivated under natural day (ND) conditions and incandescent-induced LD (ILD) conditions in an experimental farm of Hokkaido University, Sapporo (43°07′N, 141°35′E) in 2017. The ILD condition was generated by extending ND to 20 h using supplemental lighting of incandescent lamps with an R:FR ratio of 0.7 from 2:00 to 7:00 and 18:00 to 22:00 every day from sowing to August 10th.

Three consistent temperature conditions of 18℃, 25℃ and 32℃ and a varying temperature condition of 32℃ in the daytime and 25℃ in the nighttime were set in the growth chambers. Lighting was supplied for 16 h using a combination of fluorescent and incandescent lamps with an average photosynthetically active photon flux density of 150 μmol m^{-2} s^{-1} and an R:FR ratio of 7.0.

2.2.3 Cultivation methods

In the controlled experiments in the greenhouse and growth chambers, seeds were directly sown into plastic pots (15 cm in diameter and depth), and then, thinned to four plants per pot. In the field experiments, seeds were sown in paper pots (Paperpots No.2, Nippon Beet Sugar Manufacturing Co., Tokyo, Japan) on 28 May 2017 and put outdoors under ND or ILD conditions until transplanted into the field. The daily mean outdoor air temperature from the sowing date to the end of July was 18.9℃, with a minimum of 9.6℃ and a maximum of 27.3℃. Flowering times were recorded individually and expressed as days after sowing (DAS).

2.2.4 Expression analyses of flowering genes

Expression analyses were performed using fully expanded new leaves of AGS292 and AK16 grown in the greenhouse and in the growth chambers. Leaves were sampled at Zeitgeber times of 3, 12 and 21 h at two different growth stages, the 2nd and 4th trifoliate-leaf stages. All the samples were immediately frozen in liquid nitrogen and stored at −80°C. Total RNA was isolated from each sample using TRIzol

Reagent (Invitrogen, Waltham, MA, USA). cDNA was synthesized from total RNA (1 µg) using an oligo (dT) 20 primer with M-MLV Reverse Transcriptase (Invitrogen) in a 20-µL volume. The transcript abundance levels of *FT2a* and *FT5a* were determined using quantitative real-time PCR (qRT-PCR). Each qRT-PCR mixture (20 µL) contained 0.05 µL of the cDNA synthesis reaction mixture, 5 µL of 1.2 µM primer premix and 10 µL of SYBR Premix ExTaq Perfect Real Time (TaKaRa, Kyoto, Japan). A CFX96 Real-Time System (Bio-Rad, Hercules, CA, USA) was used to quantify expression levels. The PCR cycling conditions were 95°C for 3 min followed by 40 cycles of 95°C for 10 s, 58°C for 20 s, 72°C for 20 s and 78°C for 2 s. The mRNA for *Actin-2/7* was used for normalization. A reaction mixture without reverse transcriptase was also used as a control to confirm the absence of genomic DNA contamination. The amplification of a single DNA fragment was confirmed by a melting curve analysis and the gel electrophoresis of the PCR products. Averages and standard errors of relative expression levels were calculated for three independently synthesized cDNAs. Primers used in the expression analyses are listed in Supplemental Table 2.1.

2.3 Results and Discussion

2.3.1 AK16 responds differentially to R-enriched and FR-enriched LD conditions

AGS292 and AK16 are both the *e3*/*e4* genotype that conditions photoperiod insensitivity in soybean. Under outdoor ND conditions (maximum day length of 15.5 h) in Sapporo, AK16 flowered at 61.8 DAS, which was, on average, 22.3 d later than AGS292, which flowered at 39.5 DAS. However, under ILD conditions, they flowered at almost the same times as under ND conditions (at 62.3 DAS for AK16 and at 43.7 DAS for AGS292). Thus, the flowering times of both lines were not largely influenced by FR-enriched ILD (FR-LD) conditions, which was consistent with the photoperiod insensitive *e3*/*e4* genotype.

Intriguingly, the two lines responded differently to the R-enriched LD (R-LD) conditions in the greenhouse (Figure 2.1A). AGS292 flowered at almost the same time (28.9 to 31.8 DAS) under the four different day-length conditions, which ranged from 8 h to 20 h. The flowering of AK16 was delayed by 3.3 d and 5.6 d under 8-h

and 16-h day lengths, respectively, with an average of 4.9 d, compared with the flowering time of AGS292. The flowering delay in AK16 increased under R-LD conditions with a 20-h day length; AK16 flowered on average 11.7 d later than AGS292. Thus, AK16 retained the flowering sensitivity under the LD condition produced using high-intensity discharge lamps with greater R:FR ratios, although it did not respond to FR-LD under outdoor conditions.

2.3.2 Responses of the *e3*/*e4* LF line to different thermal conditions

The flowering times of AGS292 and AK16 were evaluated under three thermal conditions, constant 18℃, 25℃ and 32℃, at a day length of 16 h using growth chambers (Figure 2.1B). Flowering time at 25℃ was on average 24.6 and 32.3 DAS in AGS292 and AK16, respectively, with the difference (7.7 d) being significant at the 5% level. The flowering times were delayed at 18℃ by almost the same numbers of days in AGS292 (17.7 d) and AK16 (17.2 d) compared with at 25℃, indicating that both lines responded similarly to the lower temperature. In contrast, the responses to 32℃ differed between the two lines; AGS292 flowered at 22.4 DAS, which was slightly earlier than at 25℃ (24.6 DAS), while AK16 flowered at 42.5 DAS, which was 20.1 d later than AGS292. The flowering delay was also found in AK16 grown under the varying conditions of 32℃ during the daytime and 25℃ during the nighttime. AGS292 flowered at almost the same time (22.8 DAS) as with the constant 32℃ (22.4 DAS), while AK16 flowered at 38.4 DAS, which was, on average, 15.6 d later than AGS292.

2.3.3 Transcript profiles for *FT2a* and *FT5a*

Transcript abundances of *FT2a* and *FT5a* are closely related to the earliness of flowering under various environmental conditions (Watanabe et al., 2009; Takeshima et al., 2016; Zhao et al., 2016; Zhu et al., 2019). Expression profiles of *FT2a* and *FT5a* under the SD and R-LD conditions were examined at three time points, 3-, 12- and 21-h Zeitgeber times, at the 2nd and 4th trifoliate-leaf stages (Figure 2.2A–B). In both stages, the transcript abundances of *FT2a* and *FT5a* were almost the same between AGS292 and AK16 under SD conditions, although the expression level of *FT2a* was slightly greater in the former than in the latter. In contrast, the *FT2a* and

FT5a expression levels were strongly downregulated in AK16, compared with AGS292, under R-LD conditions (Figure 2.2).

Transcript abundances also varied with thermal conditions (Figure 2.3A–B). At 25°C, the expression levels of *FT5a* were almost the same between AGS292 and AK16, but *FT2a* was slightly upregulated in AK16 compared with AGS292. At 18°C, the expression levels of *FT2a* and *FT5a* were similar to those at 25°C in AGS292, but they were downregulated in AK16; the downregulation was greater at the 4th trifoliate-leaf stage. Intriguingly, the expression levels of *FT2a* and *FT5a* were lower at 32°C, compared with at 25°C, and this hot temperature-related suppression was greater in AK16.

2.3.4 Transcript profiles of repressor genes

AK16 exhibited distinct repressions of *FT2a* and *FT5a* expressions under R-LD and high temperature (32°C) conditions. Then I evaluated the expression profiles for possible repressors for *FT* genes. The repressor genes analyzed were *E2*, *MYC-Like* family, *CONSTANS-LIKE2b* (*COL2b*), *TARGET OF EAT4a* (*TOE4a*), and *REDUCED VERNALISATION RESPONSE 1* (*VRN1*). The *MYC-Like* family genes are the homologous genes of the *Arabidopsis MYC* gene (Wang et al., 2014), which is involved in temperature responses of flowering (Horvath, 2009; Kazan and Manners, 2013; Wang et al., 2017). *COL2b* showed a high sequence similarity to *Arabidopsis CO*, which plays a central role in photoperiodic flowering control of plants (Wu et al., 2014). *TOE4a* is a flowering time-related *APETALA2-like* gene mediated by photoperiod via maturity genes *E3* and *E4* and represses the flowering-related genes *FT2a* and *FT5a* under both LD and SD conditions (Zhao et al., 2015). *VRN1*, a B3 domain-containing transcription factor first be reported in wheat, responds vernalization and regulates flowering time by stably repressing the floral repressor *FLC* (Dubcovsky et al., 1998; Levy et al., 2002; Yan et al., 2003; Jung et al., 2012).

The transcript abundance of *E2* at the 4th trifoliate-leaf stage had a peak at 12 ZT and was repressed at 3 and 21 ZT in all conditions (Figure 2.4). AK16, which had the functional *E2* allele, exhibited higher expression levels than AGS292 having the dysfunctional *e2* allele in all of the conditions except for the R-LD at the second

trifoliate-leaf stage. The *E2* expression was upregulated at both growing stages in R-LD and the fourth trifoliate-leaf stage in 32℃ conditions, compared with SD and the other two thermal conditions (18℃ and 25℃), respectively. However, there was no marked difference in the responses to different photoperiod and thermal conditions between AGS292 and AK16.

Expression profiles of *MYCLa, MYCLb, and MYCLd* at the 4th trifoliate-leaf stage of AGS292 and AK16 under different thermal conditions (Figure 2.5) shows no significant difference in the expression patterns of these three genes, but *MYCLb* has the highest expression level. At 18℃, the expression levels of all three genes were almost the same in AGS292 and AK16. At 25℃, the expression levels in AK16 were slightly upregulated compared within AGS292, and at 32℃, the results were exactly the opposite.

The expression levels of *COL2b* at the 4th trifoliate-leaf stage in AGS292 decreased as temperature increased, but in AK16, the expression levels were slightly upregulated at 25℃ compared with 18℃ but obviously downregulated at 32℃. There was no significant difference in the expression levels of *TOE4a* between AGS292 and AK16; the expression levels increased as temperature increased (Figure 2.6). *VRN1* was hardly expressed under almost all conditions in both AGS292 and AK16.

The expression patterns of *MYCLb* and *COL2b* at the 2nd trifoliate-leaf stage were almost the same as those at the 4th trifoliate-leaf stage in AGS292 and AK16, although the expression levels were significantly lower at the 2nd trifoliate-leaf stage than the 4th trifoliate-leaf stage (Figure 2.7). In particular, the expression levels of *COL2b* at the 2nd trifoliate-leaf stage were extremely low. The transcript profiles of *MYCLb* and *COL2b* under the varying-temperature and low-temperature conditions at the 2nd trifoliate-leaf stage were almost the same as those under 32℃ and 18℃ conditions.

The expression levels of *E1* at the 4th trifoliate-leaf stages in both AGS292 and AK16 were very low under 16℃ and 25℃ conditions (Figure 2.8). It can be considered that *E1* is hardly expressed due to the double recessive genotype *e3*/*e4* under these temperature conditions. However, the expression of *E1* in AK16 was

increased at the 32°C condition, suggesting that the high expression of *E1* was involved in the delayed flowering at high temperatures (Figure 2.8).

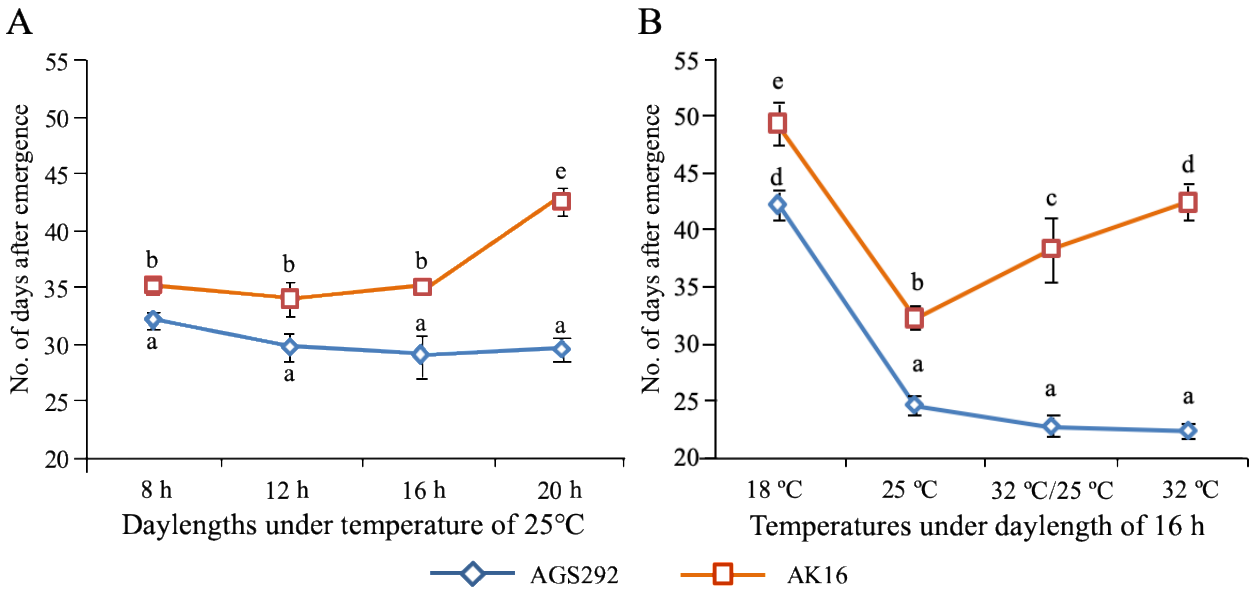

Figure 2.1. Flowering times in different photo-thermal conditions of photoperiod insensitive lines, AGS292 and AK16. Different alphabets indicate significant differences in flowering times between lines/conditions at 5 % level by Tukey-Kramer method.

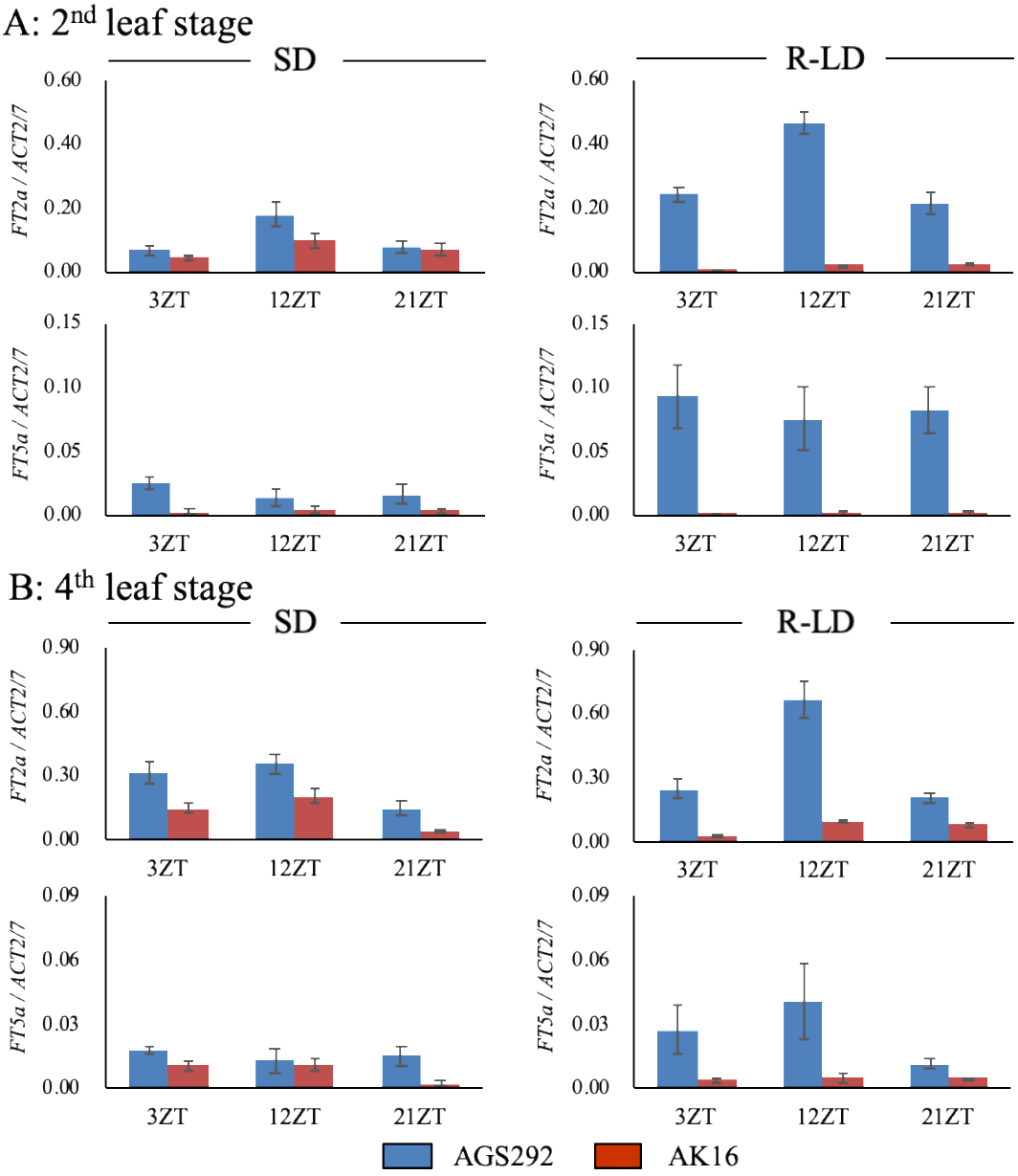

Figure 2.2. Expression profiles of *FT2a* and *FT5a* in AGS292 and AK16 under SD and R-LD conditions.

(A) 2nd leaf stage. (B) 4th leaf stage.

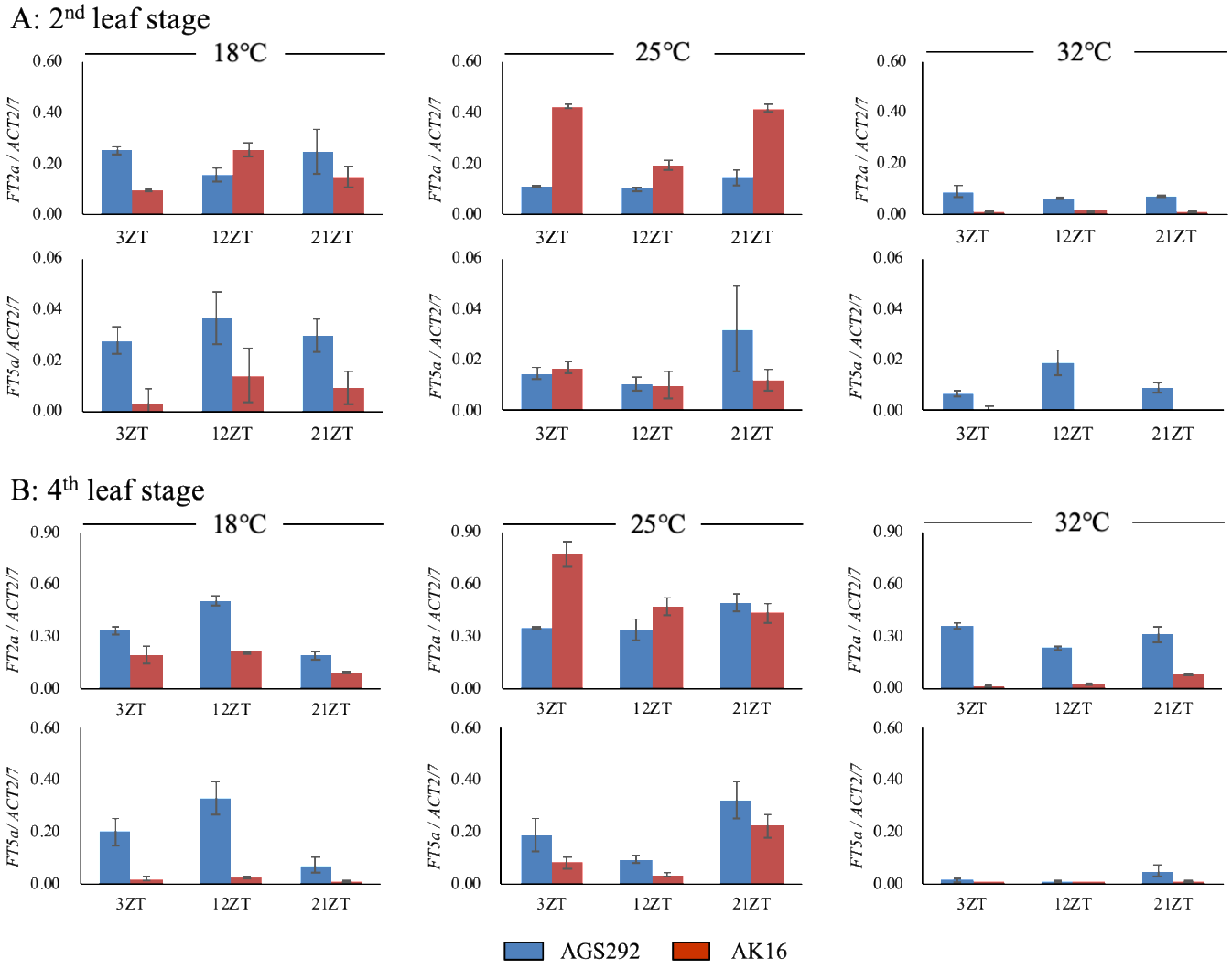

Figure 2.3. Expression profiles of *FT2a* and *FT5a* in AGS292 and AK16 under different thermal conditions.

(A) 2nd leaf stage. (B) 4th leaf stage.

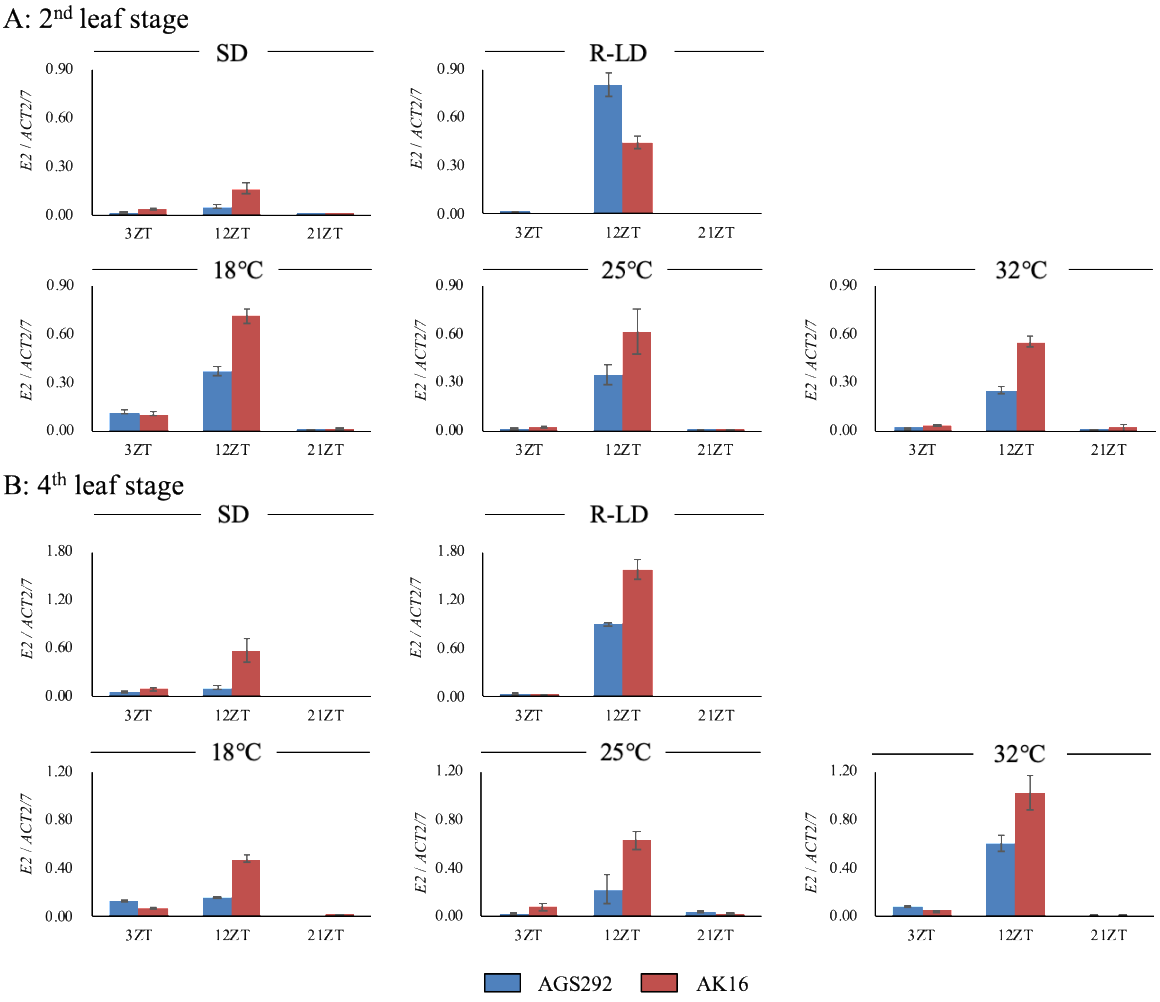

Figure 2.4. Expression profiles of *E2* in soybean lines AGS292 and AK16 under different photo-thermal conditions. (A) 2nd leaf stage. (B) 4th leaf stage.

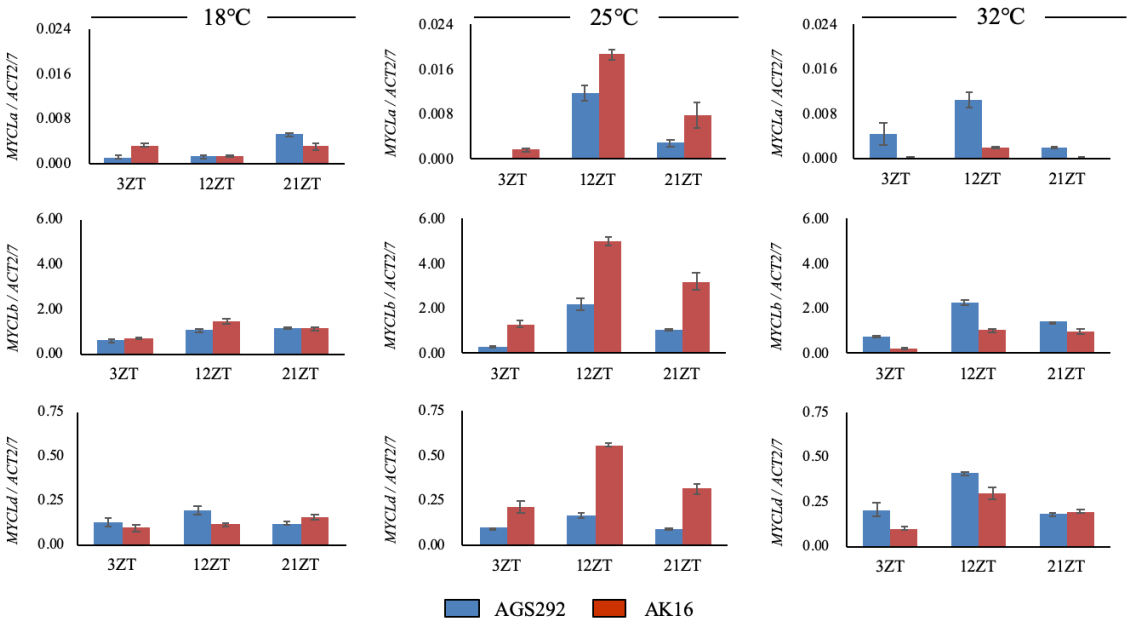

Figure 2.5. Expression profiles of *MYCLa, MYCLb and MYCLd* at the 4th leaf stage of AGS292 and AK16 under different thermal conditions.

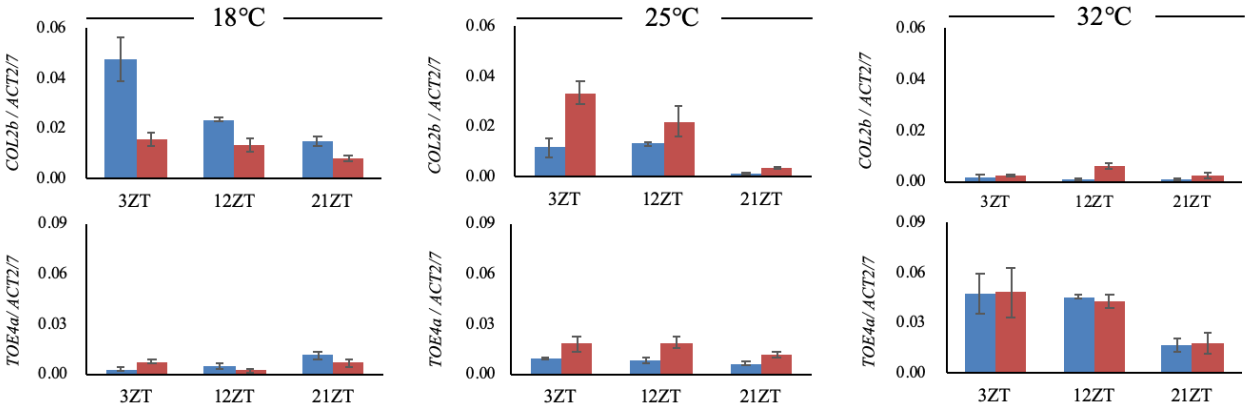

Figure 2.6. Expression profiles of *COL2b* and *TOE4a* at the 4th leaf stage of AGS292 and AK16 under different thermal conditions.

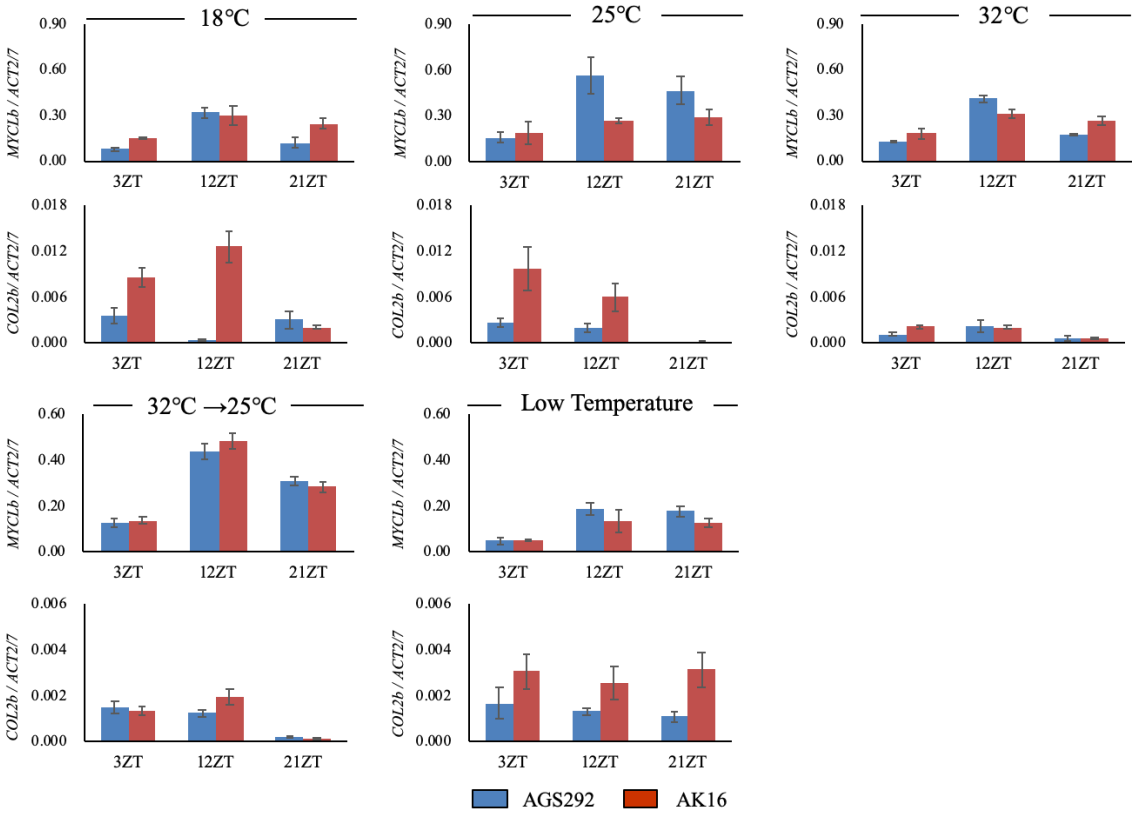

Figure 2.7. Expression profiles of *MYCLb* and *COL2b* at the 2nd leaf stage of AGS292 and AK16 under different thermal conditions.

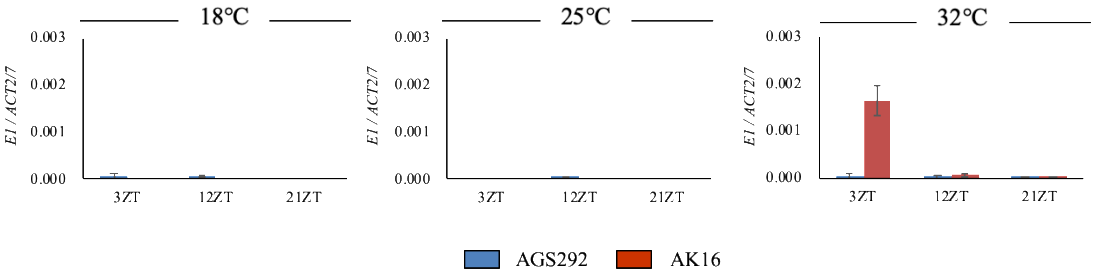

Figure 2.8. Expression profiles of *E1* at the 4th leaf stage of AGS292 and AK16 under different thermal conditions.

Chapter III Quantitative Trait Loci for Late Flowering

3.1 Background and Purpose

The wide adaptability of soybean is created by natural variations at a number of major genes and quantitative trait loci (QTLs) that control the timing of flowering and maturation (Watanabe et al., 2012; Cao et al., 2017). Since the early 1990s, so many linkage maps using molecular markers have been constructed and used in identification of QTLs for various agronomic traits in soybean. Until now, 104 QTLs detected in bi-parental crossings and 7 QTLs detected in genome-wide association studies (GWAS) have been reported for the R1 reproductive stage (days to flowering of the first flower) and registered in SoyBase (https://soybase.org/sbt/); there are distributed on 17 chromosomes except for Chr03, Chr05 and Chr17.

QTL mapping with genome-wide markers has provided a way to explore the genetic system for quantitative traits. Zhang et al. (2015) detected 27 QTLs for flowering time with GWAS. They found 18 and 10 candidate genes homologous to *Arabidopsis* flowering genes at or near the peak single nucleotide polymorphisms (SNPs), respectively. Since then, genome-wide and flowering gene-specific association analyses have detected associations of DNA polymorphisms in the orthologs of *Arabidopsis* flowering genes with flowering time (Contreras-Soto et al., 2017; Mao et al.,2017; Copley et al., 2018; Zhao et al., 2018; Jiang et al., 2019; Ogiso-Tanaka et al., 2019).

Although a number of QTLs associated with flowering time on soybean have been identified, our understanding of the molecular genetic mechanisms of delayed flowering time under SD conditions is still limited. In addition to the *j* and *e6* alleles, QTL mapping studies have revealed the involvement of novel *LJ* genes (Liu et al., 2011; Lu et al., 2015; Li et al., 2017). Two QTLs located in Chr04 and Chr16, responsible for the control of flowering under the SD conditions at lower latitudes, were identified in a cross between early and late flowering Korean soybean cultivars (Liu et al., 2011). In a cross between photoperiod-insensitive cultivar AGS292 and a Thai LJ cultivar K3, two novel QTLs controlling flowering time under SD conditions were detected in Chr16 (Lu et al., 2015). Moreover, Li et al. (2017) reported two

QTLs that regulated flowering time under SD conditions in a cross between the LJ cultivars Paranagoiana and PI159925.

In this chapter, I carried out QTL analysis for the LF habit of K3 in the progeny of a cross between AGS292 and AK16.

3.2 Materials and Methods

3.2.1 Plant materials and cultivation methods

75 F_6 families were developed using the single seed descent method from F_2 plants of a cross between AGS292 and AK16. These families, together with parental lines, were cultivated in four different photo-thermal conditions, as described in the previous chapter; SD and R-enriched LD conditions in the greenhouse and ND and FR-enriched LD conditions in the experimental field.

Two day-length conditions of 12 h (SD) and 20 h (R-LD) were set in a greenhouse in winter seasons of 2017 to 2018. High intensity discharge lamps (HONDA-T, Panasonic Co, Osaka, Japan) were used as light sources during the daytime. The average photosynthetically active photon flux density was 120 μmol m^{-2} s^{-1}, and the red-to-far red (R:FR) ratio was 4.5 at 1 m below the light source. Air temperatures in the greenhouse were adjusted to 25℃, with fluctuations from a minimum of 20℃ to a maximum of 28℃.

Two lines were also cultivated under natural day (ND) conditions and FR-light enriched incandescent-induced LD (FR-LD) conditions in an experimental farm of Hokkaido University, Sapporo (43°07′N, 141°35′E) in 2017. The FR-LD condition was generated by extending ND to 20 h using supplemental lighting of incandescent lamps with an R:FR ratio of 0.7 from 2:00 to 7:00 and 18:00 to 22:00 every day from sowing to August 10th.

3.2.2 DNA extraction and marker analysis

Total DNA was extracted by bulk from young leaves of four plants per family using the modified CTAB method (Doyle and Doyle, 1990). These RAD-Seq

procedures were carried out by Clockmics, Inc. (Izumi, Osaka, Japan). Cleaved amplified polymorphic sequence markers were developed to detect SNPs in the fourth exon of *FT2a* and the 3′ untranslated region (UTR) of *FT5a*. The products amplified by PCR using region-specific primers were digested with the restriction enzyme HinfI (*FT2a*) or PvuII (*FT5a*) and were separated by electrophoresis in a 2.5% agarose gel, stained with ethidium bromide and visualized under UV light. The genotypes at the *E2* locus were determined using the functional DNA marker developed by Tsubokura et al. (2014). The primers used are listed in Supplemental Table 3.1.

3.2.3 Restriction site-associated DNA sequence analysis

Total DNA was digested using the restriction enzymes BglII and EcoRI to create a DNA library for double-digest restriction site-associated DNA sequencing (ddRAD-Seq) (Bard et al., 2008; Peterson et al., 2012). Sequencing was performed with 51-bp single-end reads in one lane of a HiSeq2000 Sequencer (Illumina, San Diego, CA, USA) by Macrogen (Seoul, South Korea). The resulting reads were trimmed with Trimmomatic ver 0.3 (Bolger et al., 2014) using the following parameters: LEADING:19, TRAILING:19, SLIDINGWINDOW:30:20, AVGQUAL:20, and MINLEN:51. These ddRAD-Seq analyses were carried out by Clockmics, Inc. (Izumi, Osaka, Japan). The trimmed reads were mapped to the soybean reference genome Williams 82.v2 using Bowtie2 (Langmead et al., 2012) with the default parameter settings. SNP calling was performed using the Genome Analysis Toolkit (GATK) Unified Genotyper (McKenna et al., 2010). The imputation of missing genotypes in RILs based on the parental SNP data was performed using Beagle 4.0 (Browning et al., 2007). Filtering for monomorphic SNPs and SNPs having many missing calls was performed using TASSEL.5.2.31 (Bradbury et al., 2007) with the following parameters: a minimum call rate per SNP of 90% and a minimum allele frequency of 0.05 (to remove monomorphic SNPs). Using a custom script in R (https://www.R-project.org/), the nucleotide information was converted to the AB genotype with parents A and B being AGS292 and AK16, respectively. All the heterozygotes were converted as missing genotypes. Further filtering for duplicated markers or markers having switch alleles was performed in R/QTL (Broman et al., 2003), and resulted in a final set of 829 SNPs.

3.2.4 QTL mapping analysis

QTL IciMapping ver 4.1 (Wang et al., 2016) was used to construct a linkage map with three gene-specific and 829 SNP markers. The input algorithm, which re-estimated recombination frequencies and genetic distances without changing the marker order in the input file, was used to determine the order of the markers on the genetic map. The sum of the adjacent recombinant frequencies with a window size of 5 was used as a rippling criterion for fine tuning the markers. Recombination frequencies between linked loci were transformed into centimorgan (cM) distances using Kosambi's mapping function. The inclusive composite interval mapping of additive QTLs implemented in QTL IciMapping version 4.1 (Wang et al., 2016) was used to detect the QTLs at a 5% level after 1,000 genome-wide permutation tests.

3.3 Results and Discussion

3.3.1 Segregation of flowering time in the progeny

The flowering time in F_6 families segregated continuously within the ranges of the parental lines under SD and R-LD conditions in the greenhouse and under ND and FR-LD conditions in the experimental farm (Figure 3.1A–D). The flowering times of the F_6 families correlated positively between the SD and R-LD conditions ($r = 0.537$, $p < 0.01$), and between the ND and FR-LD conditions ($r = 0.818$, $p < 0.01$) (Figure 3.2A–B).

3.3.2 QTL mapping for the late flowering of AK16

The linkage map covering 868.4 cM was constructed with 829 SNPs obtained from the ddRADseq analysis and three gene-specific markers for *E2*, *FT2a* and *FT5a*.

The limited genome coverage resulted from AK16 being a progeny of a cross between AGS292 and K3. The composite interval mapping implemented in IciMapping detected three significant QTLs for days to flowering (DTF), *qDTF-10*, *qDTF-16-1* and *qDTF-16-2*, in the four photo-thermal conditions (Table 3.1, Figure

3.1). Of these, *qDTF-16-2* was detected in all four conditions, and it solely accounted for 23.9% to 56.1% of the total variation observed. *qDTF-16-1* also exhibited significant allelic effects in all the conditions, except for R-LD. The allelic effects were smaller than those of *qDTF-16-2* in the SD, R-LD and ND conditions, but greater in the FR-LD conditions (Table 3.1). *qDTF-10* was detected only in the outdoor ND and FR-LD conditions, although the effects were not significant in the latter. Collectively, the detected QTLs accounted for 31.2% in R-LD conditions (40.2% if non-significant *qDTF-16-1* was included) to 72.9% in ND conditions of the total variation observed. The *FT2a* and *E2* gene-specific markers had the greatest logarithm of odds (LOD) scores for *qDTF-16-2* and *qDTF-10*, respectively. In *qDTF-16-1*, the greatest LOD scores occurred in the interval of the *FT5a* gene-specific marker and SNP_4463558 in SD, R-LD and ND conditions, but occurred in the interval of the next two SNPs, SNP_4900158 and SNP_5373087, in FR-LD conditions. The intervals between left and right markers contained 30, 55 and 167 annotated genes for *qDTF-16-1*, *qDTF-16-2* and *qDTF-10*, respectively, in the reference genome Williams 82.v2 (Table 3.2).

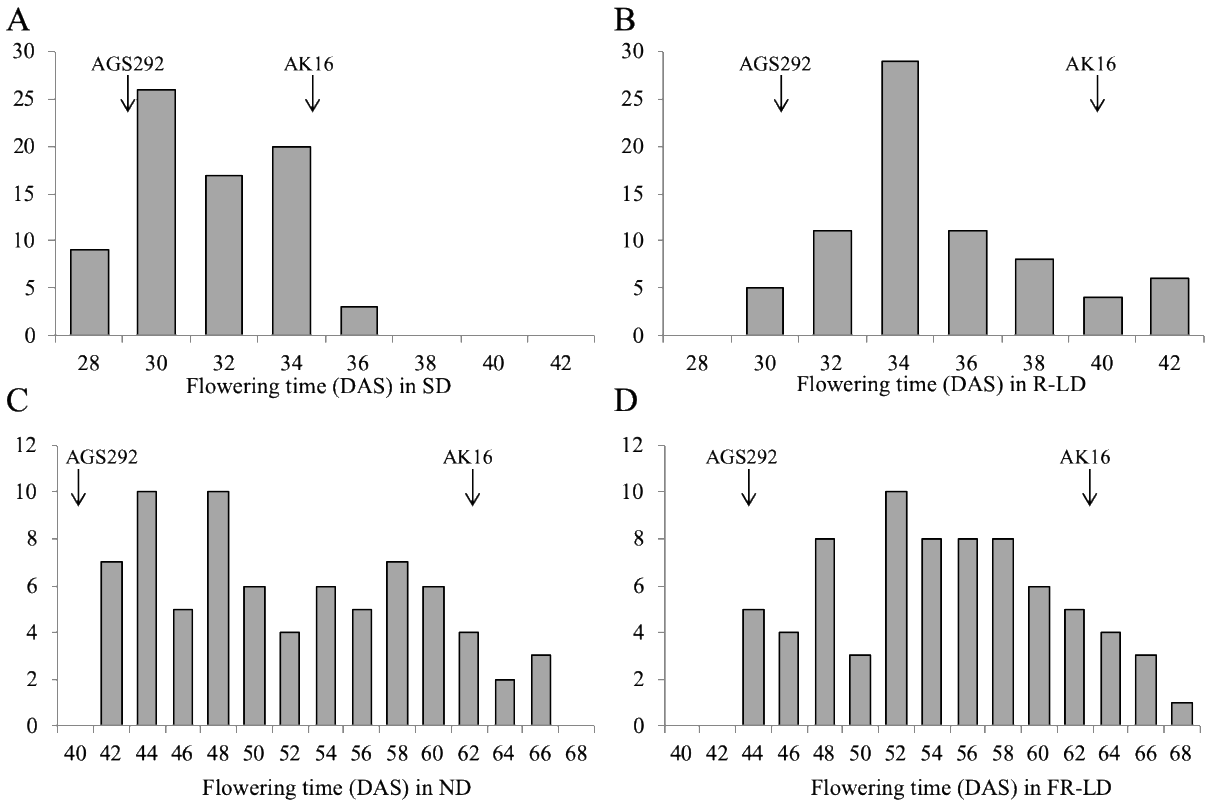

Figure 3.1. Frequency distribution of flowering time in the F_6 progeny of a cross between AGS292 and AK16 in four photo-thermal conditions. (A) SD condition (daylength of 12 h), (B) R-LD condition (daylength of 20 h) generated with red-light enriched high intensity discharge lamps, (C) ND condition in Sapporo (daylength of max. 15.5 h), (D) FR-LD (daylength of 20 h) generated with far red-light enriched incandescent lamps. The SD and R-LD conditions were set in the greenhouse where temperatures were adjusted to 25 ℃. The ND and FR-LD conditions were set outdoor. The outdoor average temperature from the sowing date to the end of July was 18.9 ℃.

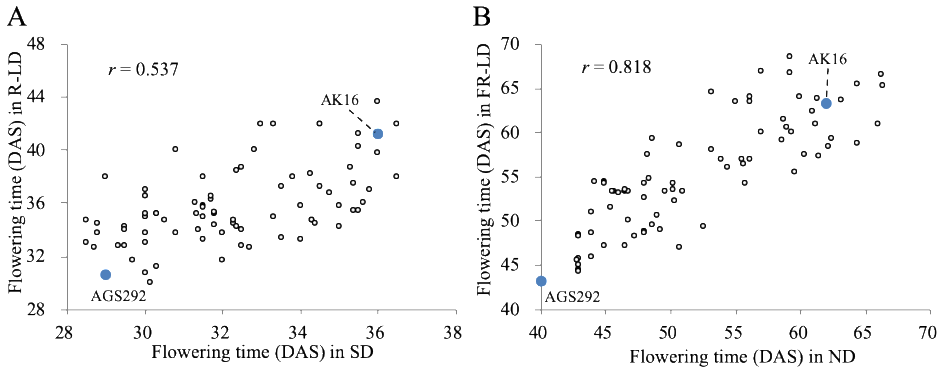

Figure 3.2. Scatter diagrams of flowering time in the F_6 progeny of the cross between AGS292 and AK16. (A) between SD and R-LD conditions and (B) between ND and FR-LD conditions

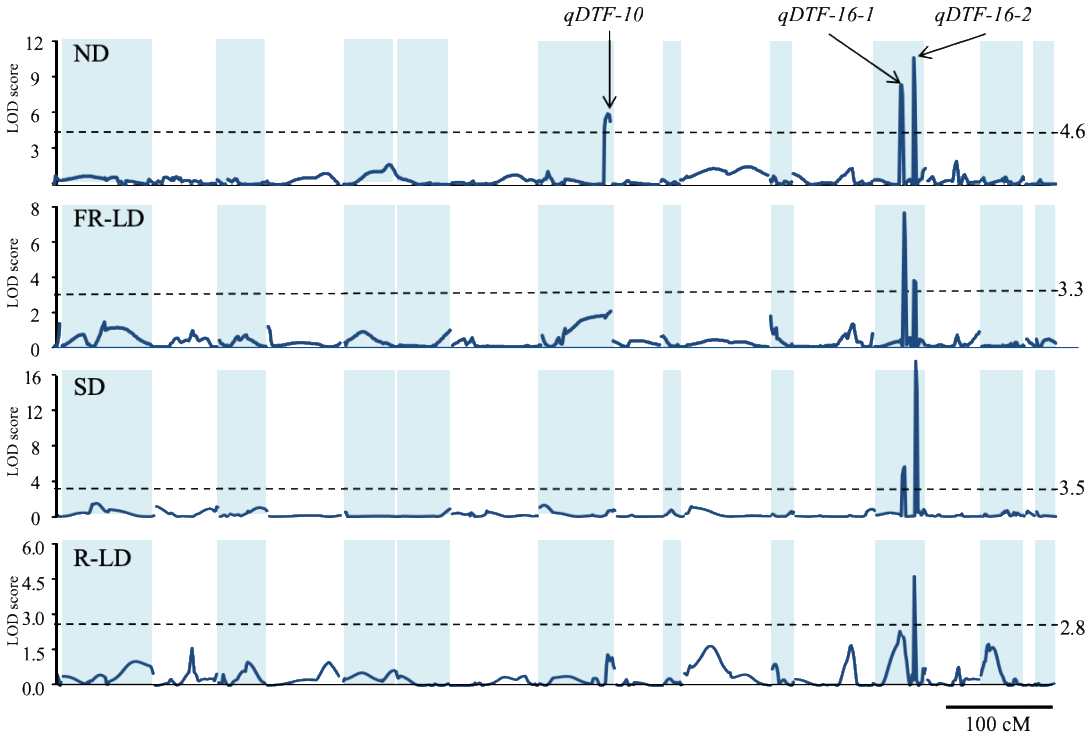

Figure 3.3. LOD scores for QTLs controlling the flowering time in four photo-thermal conditions in the F_6 progeny of the cross between AGS292 and AK16.

Table 3.1. Quantitative trait loci for flowering time in a cross between AGS292 and AK16

Environment	QTL	Chr	Left marker	Right marker	LOD	PVE	Additive effect
SD (12h)	*qDTF-16-1*	16	FT5a (4136378)	SNP_4463558	5.1	12.3	0.7
	qDTF-16-2	16	FT2a (31114633)	SNP_31642505	16	56.1	1.5
R-LD (20h)	*qDTF-16-1*	16	FT5a (4136378)	SNP_4463558	2.2 [ns]	9	0.9
	qDTF-16-2	16	FT2a (31114633)	SNP_31642505	4.6	31.2	1.5
ND	*qDTF-10*	10	E2 (45310798)	SNP_46678320	5.6	16	2.6
	qDTF-16-1	16	FT5a (4136378)	SNP_4463558	8	23.9	3.3
	qDTF-16-2	16	FT2a (31114633)	SNP_31642505	10.1	33	3.6
FR-LD (20h)	*qDTF-10*	10	E2 (45310798)	SNP_46678320	2.0 [ns]	5.7	1.6
	qDTF-16-1	16	SNP_4900158	SNP_5373087	7.6	36.5	3.6
	qDTF-16-2	16	FT2a (31114633)	SNP_31642505	3.7	16.4	2.3

Additive effect for the allele from AK16 (days)

Genomic positions for gene-specific markers are presented within parentheses

PVE; Percent of variation explained

ns; not significant at 5 % level based on the permutation test.

Table 3.2. Annotated genes in the intervals of left and right markers for QTLs for days to flowering

QTL		Location	Gene Symbol	Description
qDTF-16-1	FT5a maker	4136378		
	Glyma.16G044100	4135885-4137742	FT5a	FLOWERING LOCUS T-like protein FT5a
	Glyma.16G044200	4162525-4164824	FT3a	FLOWERING LOCUS T (FT) GmFT3a/GmFTL1
	Glyma.16G044300	4186007-4189436	LOC100805510	PTHR31642:SF40- DUAL TRANSCRIPTION UNIT AND ALTERNATIVE SPLICING PROTEIN GLAUCE
	Glyma.16G044400	4200389-4219290	LOC100808691	K15436- transportin-3 (TRPO3, MTR10),transportin MOS14
	Glyma.16G044500	4224380-4227731	LOC106796341	K03265- peptide chain release factor subunit 1 (ETF1, ERF1), eukaryotic peptide chain release factor subunit 1-3
	Glyma.16G044600	4230476-4238050	LOC100789257	K01895- acetyl-CoA synthetase (ACSS, acs)
	Glyma.16G044700	4242752-4244978	LOC100789790	pentatricopeptide repeat-containing protein At3g26630, chloroplastic
	Glyma.16G044800	4251406-4253186	LWD1	WD repeat-containing protein LWD1(ATAN11/AtLWD1, clock protein, regulate *Arabidopsis* photoperiodic flowering)
	Glyma.16G044900	4259754-4262894	LOC100815649	glyceraldehyde-3-phosphate dehydrogenase A, chloroplastic
	Glyma.16G045000	4272035-4274373	XTH1	endo-xyloglucan transferase
	Glyma.16G045100	4276764-4278580	LOC100779122	xyloglucan endotransglucosylase/hydrolase protein B-like
	Glyma.16G045200	4295497-4296132	LOC100781263	PTHR33450:SF5 - EMB
	Glyma.16G045300	4307544-4308679	LOC100790324	PF05553 - Cotton fibre expressed protein (DUF761)
	Glyma.16G045400	4311045-4312509	-	-
	Glyma.16G045500	4313909-4315249	-	-
	Glyma.16G045600	4317171-4317535	-	-
	Glyma.16G045700	4333771-4337989	LOC100790844	DNA cross-link repair protein SNM1
	Glyma.16G045800	4340643-4342789	LOC100791374	PTHR16007//PTHR16007:SF13- EPIDIDYMAL MEMBRANE PROTEIN E9-RELATED // SUBFAMILY NOT NAMED
	Glyma.16G045900	4343981-4346646	LOC100787141	syntaxin-22
	Glyma.16G046000	4348258-4360521	LOC100791900	ATP-dependent DNA helicase At3g02060, chloroplastic
	Glyma.16G046100	4361546-4361758	-	-
	Glyma.16G046200	4363668-4368542	LOC100790329	potassium transporter 4
	Glyma.16G046300	4383851-4384558	LOC100792421	ethylene-responsive transcription factor LEP

	Glyma.16G046400	4395909-4400837	LOC100794003	nudix hydrolase 12, mitochondrial-like
	Glyma.16G046500	4427265-4428578	LOC100306318	K02978 - small subunit ribosomal protein S27e (RP-S27e, RPS27)
	Glyma.16G046600	4431516-4439181	LOC100792955	calcium-binding mitochondrial carrier protein SCaMC-1
	Glyma.16G046700	4439905-4444017	LOC100793469	B3 domain-containing transcription factor VRN1
	Glyma.16G046800	4445226-4446974	LOC100804614	PF00098 - Zinc knuckle (zf-CCHC)
	Glyma.16G046900	4458888-4463093	LOC100807276	K11755- phosphoribosyl-ATP pyrophosphohydrolase / phosphoribosyl-AMP cyclohydrolase (hisIE)
	Glyma.16G047000	4458982-4459633	-	-
	SNP_4463558	4463558		
qDTF-16-2	FT2a maker	31114633		
	Glyma.16G150700	31109999-31114963	FT2a	FLOWERING LOCUS T-like protein FT2a
	Glyma.16G150800	31116770-31122254	-	PTHR12542:SF27 - EXOCYST SUBUNIT EXO70 FAMILY PROTEIN D3-RELATED
	Glyma.16G150900	31123710-31137197	LOC100787142	putative deoxyribonuclease TATDN3
	Glyma.16G151000	31148829-31151842	FT2b	FLOWERING LOCUS T-like protein FT2b
	Glyma.16G151100	31157887-31164776	LOC100815482	uncharacterized metal-dependent hydrolase YabD
	Glyma.16G151200	31173115-31175417	LOC100816016	syntaxin-related protein KNOLLE
	Glyma.16G151300	31181365-31186281	SIP1-2	aquaporin SIP1-2
	Glyma.16G151400	31186966-31189736	-	PTHR11782//PTHR11782:SF50 - ADENOSINE/GUANOSINE DIPHOSPHATASE // SUBFAMILY NOT NAMED
	Glyma.16G151500	31198908-31201131	LOC100817074	PF02365 - No apical meristem (NAM) protein (NAM),NAC domain-containing protein
	Glyma.16G151600	31208107-31208520	-	PF04398 - Protein of unknown function, DUF538 (DUF538)
	Glyma.16G151700	31214533-31219627	LOC100818134	exosome complex component RRP42
	Glyma.16G151800	31225865-31230698	LOC100818671	ynein light chain LC6, flagellar outer arm
	Glyma.16G151900	31232074-31238858	LOC100789267	glucan endo-1,3-beta-glucosidase 3
	Glyma.16G152000	31235374-31237966	-	-
	Glyma.16G152100	31264479-31266877	NAC23	NAC domain protein
	Glyma.16G152200	31289941-31293189	PHR28	MYB-CC domain-containing transcription factor PHR28（homolog of *AtPHR1*, PHOSPHATE STARVATION RESPONSE ）
	Glyma.16G152300	31298383-31301333	LOC100790331	probable pyruvate kinase, cytosolic isozyme
	Glyma.16G152400	31301770-31303592	LOC100790855	E3 ubiquitin-protein ligase SIRP1
	Glyma.16G152500	31305625-31319494	LOC100819202	general negative regulator of transcription

			subunit 3
Glyma.16G152600	31324673-31328127	LOC102666963	glucan endo-1,3-beta-glucosidase 8
Glyma.16G152700	31330128-31332137	LOC100776805	GATA transcription factor 9
Glyma.16G152800	31347028-31354847	-	PPR repeat (PPR) // PPR repeat family (PPR_2) // DYW family of nucleic acid deaminases (DYW_deaminase)
Glyma.16G152900	31358542-31362310	LOC100820627	protein SINE1
Glyma.16G153000	31363737-31366379	LOC100775728	PTHR11122:SF15 - NDH-DEPENDENT CYCLIC ELECTRON FLOW 5
Glyma.16G153100	31368331-31370455	LOC100791385	cell division cycle 20.1, cofactor of APC complex
Glyma.16G153200	31375013-31381634	LOC100819739	T-complex protein 1 subunit theta
Glyma.16G153300	31387861-31392064	LOC100777342	ADP-glucose phosphorylase
Glyma.16G153400	31393061-31398041	LOC100777869	glutamate dehydrogenase 1
Glyma.16G153500	31408142-31409499	LOC100526886	PTHR34132:SF1 - EMB
Glyma.16G153600	31420465-31421069	LOC100778937	arabinogalactan protein 14
Glyma.16G153700	31431020-31433072	LOC100779463	PTHR33646:SF2 - F20H23.8 PROTEIN-RELATED
Glyma.16G153800	31441158-31444423	LOC100499744	PTHR11960:SF18 - 4EHP, ISOFORM B
Glyma.16G153900	31445530-31449234	LOC100780008	superoxide dismutase [Cu-Zn] 2-like
Glyma.16G154000	31450077-31455500	LOC100780531	tubulin alpha-3 chain
Glyma.16G154100	31458458-31459539	LOC100792432	ethylene-responsive transcription factor ERN1
Glyma.16G154200	31463230-31469750	LOC100306329	SNARE superfamily protein
Glyma.16G154300	31472047-31475600	LOC100781076	shikimate kinase, chloroplastic-like
Glyma.16G154400	31477476-31484052	LOC100781621	septin and tuftelin-interacting protein 1 homolog 1
Glyma.16G154500	31482579-31483619	-	-
Glyma.16G154600	31484987-31489298	LOC100792965	alpha-(1,4)-fucosyltransferase
Glyma.16G154700	31491889-31496125	LOC100793479	tyrosine-specific transport protein
Glyma.16G154800	31501789-31505965	LOC102667386	WPP domain-associated protein
Glyma.16G154900	31508874-31509401	-	-
Glyma.16G155000	31513389-31517035	PIP2-10	aquaporin PIP2-10
Glyma.16G155100	31522889-31524889	PIP2-11	aquaporin PIP2-11
Glyma.16G155200	31533926-31545061	LOC100794533	protein ENHANCED DISEASE RESISTANCE 2
Glyma.16G155300	31555462-31561318	LOC100783228	GATA transcription factor 8
Glyma.16G155400	31566993-31569744	-	PTHR23091:SF208 - ACYL-COA N-ACYLTRANSFERASES SUPERFAMILY PROTEIN
Glyma.16G155500	31572611-31575650	LOC100306043	Domain of unknown function (DUF588) (DUF588)

	Glyma.16G155600	31579375-31590800	LOC100784829	mRNA decapping complex subunit 2-like
	Glyma.16G155700	31591570-31606167	LOC100783766	glutathione reductase, cytosolic
	Glyma.16G155800	31607630-31614865	LOC100784293	histidinol-phosphate aminotransferase, chloroplastic
	Glyma.16G155900	31610414-31611502	-	-
	Glyma.16G156000	31620672-31621151	LOC100795591	-
	Glyma.16G156100	31629372-31635425	LOC100796118	leucine-rich repeat receptor-like tyrosine-protein kinase PXC3
	SNP_31642505	31642505		
qDTF-10	E2 maker	45310798		
	Glyma.10G221500	45294735-45316121	E2	protein GIGANTEA
	Glyma.10G221600	45320219-45322143	LOC100306646	small subunit ribosomal protein S18e (RP-S18e, RPS18)
	Glyma.10G221700	45323071-45327485	LOC100776319	UDP-rhamnose/UDP-galactose transporter 6
	Glyma.10G221800	45329260-45331696	LOC100801119	-
	Glyma.10G221900	45339110-45345346	-	-
	Glyma.10G222000	45346296-45349474	-	-
	Glyma.10G222100	45346323-45349130	LOC100802195	-
	Glyma.10G222200	45351101-45356279	LOC100776854	probable serine/threonine-protein kinase At1g54610
	Glyma.10G222300	45358021-45358882	-	Domain of unknown function (DUF3511) (DUF3511)
	Glyma.10G222400	45366010-45367642	LOC100777390	peroxidase 12
	Glyma.10G222500	45373587-45376359	LOC100804126	peroxidase 12
	Glyma.10G222600	45378675-45380536	LOC100802728	phytosulfokine receptor 2
	Glyma.10G222700	45381000-45381544	-	Reverse transcriptase-like (RVT_3)
	Glyma.10G222800	45381577-45387369	LOC100804657	histone-lysine N-methyltransferase ASHH1（ABSENT, SMALL, OR HOMEOTIC DISCS 1 HOMOLOG 1)
	Glyma.10G222900	45393914-45399095	LOC100805189	putative DHHC palmitoyltransferase family protein
	Glyma.10G223000	45400648-45401217	LOC102664122	-
	Glyma.10G223100	45404603-45405334	LOC100305627	PMEI_like superfamily protein
	Glyma.10G223200	45415775-45417826	LOC100777915	ethylene-responsive transcription factor ERF110
	Glyma.10G223300	45430584-45438893	LOC100805732	polypyrimidine tract-binding protein homolog 3
	Glyma.10G223400	45440350-45443915	LOC100806256	60S ribosomal protein L3
	Glyma.10G223500	45444200-45451455	LOC100806786	cellulose synthase A catalytic subunit 2 [UDP-forming]
	Glyma.10G223600	45458315-45458750	-	Prolamin-like (Prolamin_like)
	Glyma.10G223700	45459518-45462243	LOC100808387	-
	Glyma.10G223800	45464900-45468155	LOC100808917	G-box-binding factor 4 : GBF4

Glyma.10G223900	45473669-45476195	LOC102663162	UPF0496 protein At1g20180
Glyma.10G224000	45479863-45493565	-	UDP-glucoronosyl and UDP-glucosyl transferase (UDPGT)
Glyma.10G224100	45494198-45496872	LOC100306580	thioredoxin Y1-like protein
Glyma.10G224200	45498206-45500186	LOC100499922	small subunit ribosomal protein S18e (RP-S18e, RPS18),40S ribosomal protein S18
Glyma.10G224300	45503642-45505214	LOC100809999	abscisic stress-ripening protein 5
Glyma.10G224400	45509340-45510452	LOC100779509	SF169 - ALPHA-CRYSTALLIN DOMAIN OF HEAT SHOCK PROTEIN-CONTAINING PROTEIN
Glyma.10G224500	45514702-45516815	LOC100780049	Protein of unknown function (DUF793) (DUF793)
Glyma.10G224600	45526661-45531424	-	-
Glyma.10G224700	45534516-45534867	-	-
Glyma.10G224800	45537703-45537876	-	-
Glyma.10G224900	45540859-45544972	LOC100810531	tubby-like F-box protein 8-like
Glyma.10G225000	45549152-45550892	-	-
Glyma.10G225100	45561745-45564354	-	SF9 - TRIHELIX TRANSCRIPTION FACTOR GT-2
Glyma.10G225200	45579021-45582548	LOC100811604	trihelix transcription factor GT-2
Glyma.10G225300	45598119-45603561	LOC100812134	SF11 - F16F4.11 PROTEIN
Glyma.10G225400	45605390-45609815	LOC100812684	fructose-1,6-bisphosphatase, cytosolic-like
Glyma.10G225500	45612670-45623310	LOC100813222	ubiquitin carboxyl-terminal hydrolase MINDY-3
Glyma.10G225600	45616985-45617457	-	-
Glyma.10G225700	45624809-45626702	LOC100814115	protein WHAT'S THIS FACTOR 1 homolog
Glyma.10G225800	45629065-45639027	LOC100814640	eukaryotic translation initiation factor 5B
Glyma.10G225900	45640990-45641657	-	-
Glyma.10G226000	45653088-45657459	LOC102664725	transcription factor VIP1
Glyma.10G226100	45658317-45660220	-	Histidine decarboxylase / L-histidine carboxy-lyase
Glyma.10G226200	45666030-45668827	LOC100781665	histidine decarboxylase (hdc, HDC), serine decarboxylase 1
Glyma.10G226300	45674356-45684646	LOC100782737	serine/threonine-protein kinase STY46
Glyma.10G226400	45682181-45682976	-	SF19 - ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT 1
Glyma.10G226500	45706001-45707052	-	-
Glyma.10G226600	45710673-45716672	LOC100815695	RNA-binding protein 2
Glyma.10G226700	45718946-45725503	LOC100816230	RNA-binding protein 2
Glyma.10G226800	45729028-45730338	-	-
Glyma.10G226900	45733893-45742369	LOC100783270	PH, RCC1 and FYVE domains-containing protein 1

Glyma.10G227000	45753863-45755920	LOC100783805	L-type lectin-domain containing receptor kinase S.6
Glyma.10G227100	45757751-45761394	LOC102666082	-
Glyma.10G227200	45763053-45767249	LOC100817132	-
Glyma.10G227300	45777447-45785455	LOC100784336	ABC transporter C family member 10
Glyma.10G227400	45788249-45796345	LOC100817665	ABC transporter C family member 10
Glyma.10G227500	45799329-45803586	-	Non-specific serine/threonine protein kinase / Threonine-specific protein kinase
Glyma.10G227600	45804245-45808712	LOC100784867	histidine decarboxylase (hdc, HDC)
Glyma.10G227700	45817689-45819155	LOC100818728	hevamine-A
Glyma.10G227800	45821419-45826466	LOC100818193	plastidial pyruvate kinase 2
Glyma.10G227900	45828656-45838014	LOC100785399	timeless (TIMELESS)
Glyma.10G228000	45840732-45843813	LOC100785934	receptor-like protein EIX2
Glyma.10G228100	45854630-45857975	LOC102667809	receptor-like protein EIX2
Glyma.10G228200	45862564-45865194	LOC100786447	receptor-like protein EIX2
Glyma.10G228300	45873599-45874166	LOC100819799	-
Glyma.10G228400	45874708-45875674	-	-
Glyma.10G228500	45876950-45878506	-	Protein kinase domain (Pkinase) // RING-type zinc-finger (zf-RING_UBOX)
Glyma.10G228600	45880533-45883776	LOC100788059	receptor-like protein EIX2
Glyma.10G228700	45886382-45892416	LOC100788591	WD repeat-containing protein 91 homolog（WDR91）
Glyma.10G228800	45893094-45895914	LOC102665288	receptor-like protein EIX2
Glyma.10G228900	45897641-45904061	LOC100775424	G-type lectin S-receptor-like serine/threonine-protein kinase At5g24080
Glyma.10G229000	45906187-45911386	LOC100775964	CAAX AMINO TERMINAL PROTEASE FAMILY PROTEIN
Glyma.10G229100	45911687-45912744	-	-
Glyma.10G229200	45914342-45915472	LOC100776507	probable membrane-associated kinase regulator 6-like
Glyma.10G229300	45923278-45927038	LOC100789128	S-type anion channel SLAH2
Glyma.10G229400	45951812-45959275	-	SF98 - AAA ATPASE // SUBFAMILY NOT NAMED
Glyma.10G229500	45961076-45964619	LOC100789652	F-box domain-containing protein
Glyma.10G229600	45968793-45973482	LOC100790190	reticulon-like protein B1
Glyma.10G229700	45984999-45986740	LOC102669208	probable membrane-associated kinase regulator 5
Glyma.10G229800	45996215-45998100	-	SF290 - ABC TRANSPORTER G FAMILY MEMBER 4-RELATED
Glyma.10G229900	46001071-46006711	LOC100777032	SPX domain-containing membrane protein At4g22990
Glyma.10G230000	46013366-46017442	LOC100777554	protein SGT1 homolog B-like

Glyma.10G230100	46018669-46021536	LOC100791770	pentatricopeptide repeat-containing protein At5g52850, chloroplastic
Glyma.10G230200	46024155-46026091	WRKY2	WRKY transcription factor 2
Glyma.10G230300	46036884-46037891	-	1.5.1.25 - Thiomorpholine-carboxylate dehydrogenase / Ketimine reductase
Glyma.10G230400	46039692-46040845	LOC100792820	Domain of unknown function (DUF4228)
Glyma.10G230500	46044438-46045534	LOC100793345	heavy metal-associated isoprenylated plant protein 39
Glyma.10G230600	46047209-46048615	LOC100306170	heavy-metal-associated domain-containing protein
Glyma.10G230700	46060022-46065739	LOC100794389	protein REVEILLE 6
Glyma.10G230800	46065779-46068770	LOC102659389	-
Glyma.10G230900	46072235-46075548	LOC100794916	probable polygalacturonase
Glyma.10G231000	46077416-46080227	LOC100779166	probable polygalacturonase
Glyma.10G231100	46085150-46088062	LOC100795453	probable polygalacturonase
Glyma.10G231200	46093079-46095496	BCH3	beta-carotene hydroxylase BCH3
Glyma.10G231300	46101504-46102415	LOC100306497	-
Glyma.10G231400	46106023-46108113	LOC100796498	deSI-like protein At4g17486
Glyma.10G231500	46111104-46113584	LOC100797031	probable receptor-like protein kinase At5g24010
Glyma.10G231600	46121774-46127235	LOC112998154	ferric reduction oxidase 4-like
Glyma.10G231700	46133598-46138435	LOC100779692	ferric reduction oxidase 2
Glyma.10G231800	46138896-46139641	-	SF109 - MATE EFFLUX FAMILY PROTEIN
Glyma.10G231900	46142748-46144819	LOC100780230	-
Glyma.10G232000	46154299-46156888	LOC100797571	scarecrow-like protein 8
Glyma.10G232100	46167448-46168037	-	-
Glyma.10G232200	46170280-46174862	LOC100780766	protein DETOXIFICATION 35
Glyma.10G232300	46180120-46186843	LOC100798100	methyltransferase-like protein 1
Glyma.10G232400	46187567-46191636	LOC100781301	SF3 - SEC-INDEPENDENT PROTEIN TRANSLOCASE PROTEIN TATB, CHLOROPLASTIC
Glyma.10G232500	46200481-46204422	LOC100798631	-
Glyma.10G232600	46208026-46215772	LOC100781849	silicon efflux transporter LSI2
Glyma.10G232700	46217284-46220027	LOC100782389	U3 small nucleolar ribonucleoprotein protein IMP3
Glyma.10G232800	46221921-46229878	LOC100782929	mitogen-activated protein kinase kinase kinase YODA
Glyma.10G232900	46231714-46235089	-	SF15 - NEDD8-SPECIFIC PROTEASE 1
Glyma.10G233000	46233655-46234189	-	SF15 - NEDD8-SPECIFIC PROTEASE 1
Glyma.10G233100	46237132-46242683	LOC100783465	WW domain-containing oxidoreductase-like
Glyma.10G233200	46244187-46249491	LOC100784001	short-chain dehydrogenase TIC 32, chloroplastic
Glyma.10G233300	46253499-46255020	LOC100784533	-

Glyma.10G233400	46256167-46262449	LOC100799153	SF1 - C2H2-LIKE ZINC FINGER PROTEIN
Glyma.10G233500	46266298-46273740	LOC100799675	U-box domain-containing protein 32
Glyma.10G233600	46281387-46285171	LOC100800208	uncharacterized protein At3g49055-like
Glyma.10G233700	46285252-46289700	LOC100800742	SF152 - CGI-141-RELATED/LIPASE CONTAINING PROTEIN // SUBFAMILY NOT NAMED
Glyma.10G233800	46291715-46292565	LOC102662566	SF332 - HISTONE-LYSINE N-METHYLTRANSFERASE SUVR3
Glyma.10G233900	46295074-46296709	LOC100785591	60S ribosomal protein L36-3
Glyma.10G234000	46296828-46302529	LOC100801281	-
Glyma.10G234100	46304593-46310226	LOC100786118	protein NLP6
Glyma.10G234200	46331672-46335217	LOC100786631	signal peptide peptidase-like 3-like
Glyma.10G234300	46335347-46337197	-	-
Glyma.10G234400	46346980-46355826	LOC100802358	pectin acetylesterase 9
Glyma.10G234500	46357820-46377913	LOC100787184	tripeptidyl-peptidase 2
Glyma.10G234600	46383827-46387200	LOC100802882	cytokinin hydroxylase
Glyma.10G234700	46404476-46406539	LOC100803409	cytokinin hydroxylase
Glyma.10G234800	46416494-46424985	LOC100803947	DNA-directed RNA polymerase III subunit RPC3
Glyma.10G234900	46430298-46432856	-	Reverse transcriptase-like (RVT_3)
Glyma.10G235000	46440992-46442617	LOC100787709	SF2 - PAR1 PROTEIN
Glyma.10G235100	46452974-46455556	LOC100788253	tubulin beta-1 chain
Glyma.10G235200	46460632-46465595	LOC102664068	proteoglycan 4
Glyma.10G235300	46468498-46473391	-	proteoglySF64 - CCHC-TYPE ZINC KNUCKLE PROTEINcan 4
Glyma.10G235400	46472588-46473158	-	-
Glyma.10G235500	46474976-46477770	-	SF30 - ACTIN-DEPOLYMERIZING FACTOR 10-RELATED
Glyma.10G235600	46486132-46486957	-	Legume lectin domain (Lectin_legB)
Glyma.10G235700	46491380-46492436	LOC100789139	heavy metal-associated isoprenylated plant protein 39
Glyma.10G235800	46493378-46494104	-	-
Glyma.10G235900	46494305-46498333	LOC100805547	pentatricopeptide repeat-containing protein At3g60050
Glyma.10G236000	46499079-46501955	LOC100806075	low-temperature-induced 65 kDa protein
Glyma.10G236100	46503152-46510614	LOC100806603	MAR-binding filament-like protein 1-1
Glyma.10G236200	46526782-46530177	LOC100527259	SF17 - CYTOCHROME B561-RELATED // SUBFAMILY NOT NAMED
Glyma.10G236300	46538040-46540831	LOC100305501	remorin family protein
Glyma.10G236400	46541417-46544051	LOC100807142	transcription factor MYB41
Glyma.10G236500	46546404-46549723	LOC100807669	SF97 - SNARE-LIKE SUPERFAMILY

			PROTEIN
Glyma.10G236600	46549220-46552230	LOC100790724	transcription factor LAF1-like
Glyma.10G236700	46557062-46557706	-	PRE-MRNA CLEAVAGE FACTOR IM, 25KD SUBUNIT
Glyma.10G236800	46559759-46563487	-	SF34 - DNA MISMATCH REPAIR PROTEIN MSH3
Glyma.10G236900	46571852-46575791	LOC100791249	F-box/LRR-repeat protein At3g48880
Glyma.10G237000	46580785-46583922	LOC100791779	membrane steroid-binding protein 1
Glyma.10G237100	46586624-46587350	-	-
Glyma.10G237200	46592131-46592616	-	-
Glyma.10G237300	46593154-46594226	LOC102666403	-
Glyma.10G237400	46598519-46600609	LOC100526865	putative protein phosphatase inhibitor 2
Glyma.10G237500	46600721-46601950	LOC100792301	3-hexulose-6-phosphate isomerase
Glyma.10G237600	46607534-46609838	LOC100792832	-
Glyma.10G237700	46614756-46616145	-	SF6 - PROTEIN Y87G2A.13
Glyma.10G237800	46621869-46623290	HSF-19	heat stress transcription factor 19
Glyma.10G237900	46627410-46631638	LOC100808726	leucine-rich repeat receptor protein kinase EMS1
Glyma.10G238000	46649889-46654561	LOC100793886	E3 ubiquitin-protein ligase XBAT33
Glyma.10G238100	46667755-46674270	LOC100794400	homeobox-leucine zipper protein ANTHOCYANINLESS 2
SNP_46678320	46678320		

-; uncharacterized

Chapter IV Sequence Polymorphism for Candidate Genes of QTLs

4.1 Background and Purpose

Genetic polymorphisms contribute to variations in phenotypes' response to the environment. Most quantitative traits are determined by a combination of multiple genes and their interactions with environmental factors and have complex genetic architectures. The molecular dissection of genetic variation between cultivars using whole-genome or target region resequencing may provide one of helpful clues for better understanding of the molecular bases for complex quantitative traits.

DNA sequencing technologies, such as Illumina Genome Analyzer (GA), have significantly improved throughput and dramatically reduced the cost as compared to capillary-based electrophoresis systems (Shendure et al., 2004). In a single experiment using one Illumina GA, the sequence of approximately 100 million reads of up to 50 bases in length can be determined. This ultrahigh throughput makes next-generation sequencing technologies particularly suitable for carrying out genetic variation studies by using large-scale resequencing of sizeable cohorts of individuals with a known reference (Bentley, 2006). These features and advantages have set the stage for large-scale genome-wide SNP surveys. DNA polymorphisms associated with variations in flowering time have been detected in some orthologs of *Arabidopsis* flowering genes by using genome-wide and flowering gene-specific association analyses in soybean (Zhang et al., 2015; Contreras-Soto et al., 2017; Mao et al., 2017; Copley et al., 2018; Zhao et al., 2018; Jiang et al., 2019; Ogiso-Tanaka et al., 2019).

In the previous chapter, the three major QTLs controlling the LF habit of K3 were detected. The intervals sandwiched by left and right markers contained a total of 167, 30 and 55 annotated genes for *qDTF-10, qDTF-16-1* and *qDTF-16-2*, respectively. In this chapter, I analyzed the sequence polymorphisms for flowering-related genes among the annotated genes (marked by light green in Table 3.2) in the QTL regions. I also analyzed the sequences of the genes for photoreceptors as candidates for different responses to the Red-light enriched LD condition.

4.2 Materials and Methods

4.2.1 Sequence analysis based on whole-genome resequencing data

The sequences from 6.0-kb upstream from the start codon to 1.0-kb downstream to the stop codon for the candidate genes were compared between AGS292 and K3. Raw reads of AGS292 and K3 from next-generation sequencing on Illumina HiSeq XTen were aligned to the soybean reference genome Williams 82.v2 (Schmutz et al., 2010). The alignment was performed using Bowtie2-2.2.9 (Langmead et al., 2009). The resulting alignment was further processed to remove duplicate reads and to fix mate information using Picard tools (http://broadinstitute.github.io/picard). GATK ver 3.8 (McKenna et al., 2010) was used to realign small indels. Subsequently, variants (SNP and indels) were called using the GATK Unified Genotyper function that filtered out reads having mapped base quality Phred scores of less than 20. Using the reference genome Williams 82.v2 and a SNP dataset for each variety, sequences were reconstructed using the FastaAlternateReferenceMaker function available in GATK.

A plant *cis*-acting regulatory DNA elements (Higo et al., 1999) analysis was carried out to detect known *cis*-elements in the *FT2a* and *FT5a* sequences.

4.2.2 Sequence analysis using a genome-walking method

The upstream and downstream sequences of *FT5a* were sequenced for AGS292 and K3 using the Universal GenomeWalker™ 2.0 kit (Takara Clontech, Otsu, Japan). The *FT5a* downstream sequences for Williams 82 were also be analyzed latter. The method for genome-walking (GW) followed with Shapter and Waters (2014) and GenomeWalker™ 2.0 User Manual (Takara Clontech, Otsu, Japan). The amplified fragments from GW 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 PRISM3100 Avant Genetic Analyzer (Applied Biosystems Japan, Tokyo, Japan). Primers sequences used in genome walking method are listed in Supplemental Table 4.1.

4.3 Results and Discussion

4.3.1 Sequence polymorphisms in candidates of *qDTF-16-1*

The genomic region of *qDTF-16-1* sandwiched by the left and right markers contained four flowering–related genes; two *FT* orthologs, *FT3a* and *FT5a*, *LIGHT-REGULATED WD1* (*LWD1*) and *VRN1*.

LWD1 protein (together with LWD2 protein) functions in close proximity to or within the circadian clock for photoperiodic flowering control (Wu et al., 2008).VRN1 protein is a B3 domain-containing transcription factor and binds DNA in vitro in a non–sequence-specific manner and functions in stable repression of the major target of the vernalization pathway, the floral repressor *FLC* (Levy et al., 2002).

The resequencing analysis detected no sequence variation for *FT3*a (Glyma.16G044200) and *LWD1* (Glyma.16G044800), but a number of sequence variations for *VRN1* (Glyma.16G046700) between AGS292 and K3 (Table 4.2).

The *FT5a* sequence was identical between K3 and Williams 82, and it differed by two SNPs, one in the promoter and one in the 3′ UTR from that of AGS292. The SNP in 3′ UTR generated a MYCCONSENSUSAT cis-element (CAGCTG) in K3 and Williams 82, but not in AGS292 (Figure 4.1). The genomic region of *FT5a* (6952 bp) in AGS292 and K3 from 4.6 kb upstream from the start codon to the 1.4 kb downstream to the stop codon were sequenced. The genomic sequence region (1858 bp) were sequenced by Dr. Ryoma Takeshima by using sequence analysis, and other regions were sequenced by using GW analysis in this study. I could determine the accurate sequence of 975 bp~1,101 bp downstream to the stop codon, including a sequence of 126 bp, which is presented as unconfirmed sequences in the latest Williams 82 reference genome sequence (http://www.phytozome.net) (Figure 4.2). As detected in the whole-genome resequencing data, there were two SNPs, one in the promoter region and the other in the 3′ UTR (Figure 4.1).

There were 44 polymorphisms for *VRN1* between AGS292 and K3. The sequences of AGS292 at polymorphic sites were mostly identical to that of Williams 82, except for a deletion of 2 bp in the 5′ UTR. The *VRN1* coding sequence of K3 had

an insertion of 3-bp and one nonsynonymous substitution in the exon 1, resulting in an insertion of histidine residue between 10th amino acid (aa) and 11th aa and a substitution of the alanine in AGS292 and Williams 82 to the threonine at 31th aa residue. In addition, a SNP in the exon 4 resulted in a nonsynonymous substitution; K3 possessed the asparagine in place of the lysine in AGS292 and Williams 82 at 234th aa residue (Table 4.2). The aa sequences in K3 at these three polymorphic sites were identical to those observed in *G. soja*, *Arabidopsis*, *Vigna unguiculata* and *Phaseolus vulgaris*, suggesting that the VRN1 protein from K3 is most likely a functional wild-type protein, and those from AGS292 and Williams 82 are a variant protein.

4.3.2 Sequence polymorphisms in candidates of *qDTF-16-2*

The genomic region of *qDTF-16-2* sandwiched by the left and right markers contained two *FT* orthologs, *FT2a* and *FT2b*.

The resequencing analysis detected a lot of sequence variations, including indels of 10 bp or more, between AGS292 and K3 for *FT2a*. The *FT2a* coding sequence of AGS292 was identical to that of Williams 82 (soybean reference genome sequence) but differed by a nonsynonymous substitution in the exon 4 from the coding sequence of K3. The glycine in AGS292 and Williams 82 at the 169th aa residue was converted to aspartic acid in K3 (Figure 4.1). AGS292 and K3 also showed a number of DNA polymorphisms, including four SNPs and four indels in the 2-kb promoter region, one indel in the 5′ UTR, 24 SNPs and six indels in the introns, and one SNP in the 3′ UTR (Figure 4.1; Table 4.1).

In *FT2b*, there were 58 polymorphisms, most of which were concentrated in the promoter, intron and downstream regions. The *FT2b* coding sequence of AGS292 had three synonymous substitutions, one in the exon 1 and two in the exon 4, compared with the Williams 82 sequence. In the non-coding region, there were four indels and four deletions, including a deletion of 23 bp in the downstream region of *FT2b* in AGS292. The coding sequences of K3 for *FT2b* were identical to that of Williams 82; there were three insertions and three deletions in the non-coding region, including a deletion of 63 bp in the downstream region (Table 4.3).

4.3.3 Sequence polymorphisms in candidates of *qDTF-10*

The genomic region of *qDTF-10* sandwiched by the left and right markers contained four *E2* and three *Arabidopsis* flowering genes, *ABSENT SMALL OR HOMEOTIC DISCS1 HOMOLOG1* (*ASHH1*), *G-BOX-BINDING FACTOR 4* (*GBF4*) and *WD REPEAT-CONTAINING PROTEIN 91* homolog (*WDR91*). *ASHH1* influence the rosette size and delay flowering by regulating the expressions of *FLC*, *MADS AFFECTING FLOWERING 5* (*MAF5*), *FT* and *SOC1* (del Pilar Valencia-Morales et al., 2012; Liu et al., 2016). *GBF4* encodes a transcription factor belong to bZIP classes, and functions in cryptochrome-mediated blue light signaling (Maurya et al., 2015). A homolog of *WDR91* in *Arabidopsis* is *WDR5a*, which is reported as a conserved core component of COMPASS-like H3K4 methyltransferase complex and repressed the floral transition via promoting the expression of *FLC* (Jiang et al., 2009, 2011).

There was no non-synonymous mutation in *GBF4* and *WDR91*. For *GBF4*, the sequence of K3 was identical to that of Williams 82; AGS292 had two SNPs in the exon 4, none of which caused nonsynonymous substitution. For *WDR91*, there were nine SNPs in the introns, 3′ UTR and downstream region. The sequences of AGS292 at polymorphic sites were mostly identical to those of Williams 82, except for a deletion of 1 bp in the intron 12 and an SNP in the 3′ UTR.

In *ASHH1*, there were six polymorphisms. The sequences of AGS292 at polymorphic sites were identical to those of Williams 82. One nonsynonymous substitution due to two polymorphisms in the exon 6 was detected; K3 possessed the alanine in place of the valine in AGS292 and Williams 82 at 144^{th} aa residue (Table 4.4). *G. soja*, *V. unguiculata* and *P. vulgaris* also possessed the alanine at the polymorphic aa residue as K3, suggesting that the ASHH1 protein from K3 is most likely a functional wild-type protein, and those from AGS292 and Williams 82 are a variant protein.

4.3.4 Sequence polymorphisms in genes encoding phytochromes

In Chapter 2, AGS292 and AK16 exhibited different responses to R-light enriched LD condition; the flowering time of AK16 was delayed in 20-h daylength condition, although AK16 flowered at almost the same time in ND and FR-light enriched LD conditions. This result was unexpected, because AK16 possessed dysfunctional alleles at *E3* (*PHYA3*) and *E4* (*PHYA2*) loci, as AGS292. Soybean possesses eight PHY genes, *PHYA1*, *PHYA2*, *PHYA3*, *PHYA4*, *PHYB1*, *PHYB2*, *PHYE1* and *PHYE2*, of which *PHYA3* and *PHYA2* corresponded to *E3* and *E4*, respectively (Watanabe et al., 2009; Liu et al., 2008; Wu et al., 2013). PHYA and PHYB represent the principal receptors involved in FR and R light-mediated development, respectively. Of two *PHYB* genes (*PHYB1* and *PHYB2*), heterologous expression analysis in *Arabidopsis* demonstrated that PHYB1 possessed the same functions as *Arabidopsis PHYB*; strong suppression of shade avoidance, early flowering under SD conditions, the shortening of the hypocotyls and the lengthening of the roots (Wu et al., 2011).

I compared aa sequences for six phytochrome proteins (*PHYA1*, *PHYA4*, *PHYB1*, *PHYB2*, *PHYE1* and *PHYE2*) between AGS292 and K3 (Table 4.5). RADseq data indicated that AK16 possessed the AGS292 allele for PHYA4 and the K3 allele for the remaining five PHY genes, so that PHYA4 was removed from the analysis. The aa sequences were estimated based on the resequencing data.

PHYA1 showed no nonsynonymous substitution. *PHYB1* and *PHYB2* possessed two and one nonsynonymous substitutions, respectively; the aa sequences in AGS292 were identical to those in Williams 82, *G. soja*, *Arabidopsis*, *V. unguiculata* and *P. vulgaris*, suggesting that the PHYB1 and PHYB2 proteins from AGS292 are most likely functional wild-type proteins, and those from K3 are variant proteins (Table 4.5).

PHYE1 exhibited six SNPs in the coding region, of which five resulted in non-synonymous substitutions (Table 4.5). In addition, a 3-bp deletion in the exon 3 converted two aa sequences (glutamine and isoleucine) in Williams 82 and K3 to a single aa (histidine) in AGS292. K3 possessed the identical coding sequence to Williams 82. Among six polymorphic sites, the aa sequences were identical at three sites (396, 639, and 904-905 aa residue) between K3 and Williams 82, and *G. soja*, *Arabidopsis*, *V.unguiculata* and *P. vulgaris*, and at the two sites (45 and 63 aa residue) between AGS292 and the latter four species, and at one site (1064 aa residue) between AGS292 and *G. soja*. *PHYE2* exhibited a non-synonymous substitution in the exon 2; AGS292 and Williams 82 possessed the histidine residue, whereas K3 did the tyrosine residue; the former was common to the aa residue observed in *G. soja*, *Arabidopsis*, *V. unguiculata* and *P. vulgaris* (Table 4.5).

Table 4.1. DNA polymorphisms in the genomic region of *FT2a* between AGS292 and K3

	Genomic position	Nucleotide[1)]			Amino acid position	Amino acid[2)]			
		Williams82	AGS292	K3		Williams82	AGS292	K3	*G. soja*[3)]
Promoter	-5885	A	A	G					
	-5831	C	C	T					
	-4827	T	T	A					
	-4359	T	C	T					
	-4118 ^ -4117	T/T	T/T	T/ATTAAT/T					
	-4074	T	C	T					
	-3797	A	G	A					
	-3729	T	T	C					
	-3684	C	T	C					
	-3591	G	G	T					
	-3554	C	A	C					
	-3527	A	T	A					
	-3489 ^ -3488	T/A	T/T/A	T/A					
	-3400	A	A	G					
	-3328	T	T	A					
	-3318~ -3312	TTTTTAT	AT	-					
	-2980	G	C	G					
	-2854	T	A	T					
	-2607	A	G	A					
	-2424 ~ -2409	16bp[4)]	16bp[4)]	-					
	-1949 ~ -1945	TTTTT	TTTT	-					

	-1905	C	A	C					
	-1770	T	T	A					
	-693	T	T	-					
	-448	G	G	A					
	-444 ^ -443	A/G	A/G	A/20bp[5)]/G					
	-340	G	G	-					
	-257	T	C	T					
5'UTR	-45 ~ -36	AAAGCATAAG	-	AAAGCATAAG					
	294	T	G	T					
	315	T	C	T					
	398	A	C	A					
	403	A	T	A					
IN1	475	T	T	C					
	710	T	T	C					
	942^943	A/G	A/G	A/TATA/G					
	979	T	T	A					
	988	T	C	T					
	1295	A	A	G					
	1461	T	T	C					
	1479~1482	ATGC	ATGC	-					
	1561	C	C	G					
	1734	G	G	A					
	1787^1788	A/G	A/G	A/A/G					
	2008	T	T	A					
	2812~2813	TT	TT	-					
	3188	T	T	A					
	3196	G	G	C					
IN3	3610	C	C	T					
	3721^3722	T/A	T/A	T/T/A					
	3767	C	C	A					
	3783~3796	AAAGAAAAA	-	AAAGAAAAA					
	4016	T	A	T					
	4203	T	T	C					
	4210	G	G	T					
	4215	T	T	G					
	4241	C	T	C					
	4269	T	T	C					
	4354	G	G	A					
EX4	4579	G	G	A	169	G	G	D	G
3'UTR	4876	G	G	T					

Downstream	4604~4613	GGAGTATT	GGAGTATT	-
	4758^4759	A/A	A/A	A/TATA/A
	5456	A	A	C
	5483	T	T	C

1) Nucleotides marked by red are different from nucleotides of Williams 82.

2) Amino acids marked by red are different from aminoacids of *Arabidopsis*, *Vigna unguiculata* and *Phaseolus vulgaris*.

3) https://www.ncbi.nlm.nih.gov/protein/XP_028206922.1

4) 16bp: ATTATCTTTTACACAA

5) 20bp: TAAGGAGGAGCAATGAAAGA

Table 4.2. DNA polymorphisms in the genomic region of *VRN1* between AGS292 and K3

	Genomic position	Nucleotide[1)]			Amino acid position	Amino acid[2)]			
		Williams82	AGS292	K3		Williams82	AGS292	K3	*G. soja*[3)]
Promoter	-2999	T	T	A					
	-2979	A	A	G					
	-2907	C	C	T					
	-2803 ^ -2802	A/C	A/C	A/AAAA/C					
	-2615	C	C	T					
	-2572	A	A	G					
	-2035	T	T	A					
	-1922	T	T	C					
	-1833	G	G	T					
	-1546	G	G	A					
	-1309	G	G	C					
	-1293	G	G	A					
	-1231	C	C	G					
	-1140	T	T	G					
	-1113	A	A	-					
	-947	T	T	C					
	-908	A	A	C					
	-863	T	T	-					
	-364	C	C	T					
	-360	T	T	A					
5'UTR	-265 ~ -264	GA	-	GA					

	-249	A	A	G					
	-93	T	T	A					
	-24	C	C	G					
EX1	30^31	C/A	C/A	C/ATC/A	11^12	-	-	H	H
	67	G	G	A	31	A	A	T	T
IN1	113^114	C/G	C/G	C/AC/G					
	410	T	T	C					
EX4	1797	G	G	C	234	K	K	N	N
IN4	2227	A	A	C					
	2496	A	A	G					
	2674	G	G	A					
	2739	G	G	A					
	2907	C	C	G					
Downstream	3907	C	C	T					
	4038	T	T	C					
	4085	C	C	T					
	4157	G	G	A					
	4225	T	T	A					
	4259	C	C	G					
	4328	T	T	C					
	4411	A	A	G					
	4478	T	T	-					
	4486	T	T	C					

1) Nucleotides marked by red are different from nucleotides of Williams 82.

2) Amino acids marked by red are different from aminoacids of *Glycine soja*.

3) Xie et al. (2019)

Table 4.3. Sequence polymorphisms in the genomic region of *FT2b* between AGS292 and K3

	Genomic position	Nucleotide[1)]			Amino acid position	Amino acid			
		Williams82	AGS292	K3		Williams82	AGS292	K3	*G. soja*[2)]
Promoter	-2176	C	C	T					
	-2071	T	T	-					
	-2025	T	C	T					
	-1370	A	A	T					
	-1246	T/A	T/T/A	T/A					
	-1231	T	C	T					
	-1101	C	C	A					
	-793	G	C	G					
	-765	T	T	C					
	-596	A	G	A					
	-404	T	T	C					
	-333	T	T	C					
	-78	G	G	A					
EX1	174	T	C	T	58	G	G	G	G
IN3	651^652	A/C	A/C	A/A/C					
	663^664	T/C	T/C	T/T/C					
	848	A	G	A					
	863^864	A/A	A/TA/A	A/A					
	970~972	GTA	GTA	-					
	1126	A	G	A					
	1158	A	C	A					
	1210	T	G	T					
	1321	G	G	C					
	1362~1363	AA	AA	TATATA					
	1602	C	C	T					
	1787~1789	ACA	-	ACA					
	2042	T	A	T					
	2111^2112	G/A	G/TTG/A	G/A					
	2179	T	C	T					
	2285	A	T	A					
	2507^2508	C/T	C/T/T	C/T					
	2544	T	C	T					
	2551	A	-	A					
	2553	A	G	A					
	2577	C	T	C					
	2598	C	C	T					
	2658	T	C	T					
	2782	G	G	A					

EX4	2864	A	G	A	127	Q	Q	Q	Q
	2996	T	C	T	171	G	G	G	G
	3022	A	A	G					
	3023	C	C	T					
	3080	A	A	T					
	3081	T	A	T					
	3183	A	A	G					
	3203	C	C	A					
	3218	T	C	T					
	3220	G	A	G					
	3226	G	A	G					
Downstream	3238~3299	54bp[3)]	54bp[3)]	-					
	3368	A	T	A					
	3407	A	G	A					
	3409	C	T	C					
	3431	A	-	A					
	3532	T	T	C					
	3583	A	G	A					
	3866~3888	23bp[4)]	C	23bp[4)]					
	4014	A	T	A					

1) Nucleotides marked by red are different from nucleotides of Williams 82.

2) https://www.ncbi.nlm.nih.gov/protein/NP_001276285.1

3) 54bp: CTGATGAAAGATTTAAAAACTCAAATTACATGTGATAGTCTCTTAGGCTCTTGAATGACGT

4) 23bp: AACTATCAAGAACAATACATATA

Table 4.4. Sequence polymorphisms in the genomic region of candidates of *qDTF-16-2* between AGS292 and K3

Gene (Glyma number)		Genomic position	Nucleotide[1)]			Amino acid position	Amino acid[2)]			
			Williams82	AGS292	K3		Williams82	AGS292	K3	*G. soja*[3,4)]
ASHH1 (Glyma.10G222800)	IN4	1163	T	T	G					
		1172	T	T	A					
		1180^1181	A/G	A/G	A/A/G					
	EX6	1497	T	T	C	144	V	V	A	A
		1498	G	G	A					
	Downstream	5206^5207	A/C	A/C	A/AA/C					
GBF4 (Glyma.10G223800)	EX2	674	T	A	T	188	A	A	A	A
		695	T	C	T	195	S	S	S	S
WDR91 (Glyma.10G228700)	IN12	4349	C	T	C					
	IN13	4714	T	A	T					
		4716	A	G	A					
	IN14	5239	A	A	-					
	3'UTR	5508	A	G	A					
		5514	T	A	T					
		5658	A	A	T					
	Downstream	5918	T	A	T					
		5936	T	A	T					

1) Nucleotides marked by red are different from nucleotides of Williams 82.

2) Amino acids marked by red are different from aminoacids of *Vigna unguiculata* and *Phaseolus vulgaris*.

3) https://www.ncbi.nlm.nih.gov/protein/XP_028222268.1

4) Xie et al. (2019)

Table 4.5. Sequence polymorphisms in the coding region of phytochrome genes between AGS292 and K3

Gene (Glyma number)		Genomic position	Nucleotide[1)]			Amino acid position	Amino acid[2,3)]			
			Williams82	AGS292	K3		Williams82	AGS292	K3	*G. soja*[4)]
PHYA1 (Glyma.10G141400)	EX4	3339	T	T	C	1113	S	S	S	S
		3348	T	T	C	1116	L	L	L	L
PHYB1 (Glyma.09G035500)	EX1	666	G	G	T	222	A	A	A	A
	EX4	3241	G	G	A	1081	D	D	N	D
		3259	C	C	G	1087	R	R	G	R
PHYB2 (Glyma.15G140000)	EX1	1180	G	G	A	394	V	V	I	V
PHYE1 (Glyma.15G196500)	EX1	117	T	C	T	39	I	I	I	I
		134	T	A	T	45	L	Q	L	Q
		187	T	C	T	63	S	P	S	P
	EX2	1186	G	A	G	396	G	S	G	G
		1907	C	T	C	636	A	V	A	A
	EX3	2712~2714	AAT	-	AAT	904~905	QI	H	QI	QI
	EX5	3191	G	A	G	1064	R	H	R	H
PHYE2 (Glyma.09G088500)	EX1	954	A	T	A	318	L	L	L	L
		963	T	A	T	321	P	P	P	P
	EX2	2122	C	C	T	708	H	H	Y	H

1) Nucleotides marked by red are different from nucleotides of Williams 82.

2) Amino acids marked by red are different from aminoacids of *Arabidopsis*, *Vigna unguiculata* and *Phaseolus vulgaris*.

3) Amino acids marked by red are different from aminoacids of *Vigna unguiculata* and *Phaseolus vulgaris*.

4) Qi et al. (2014)

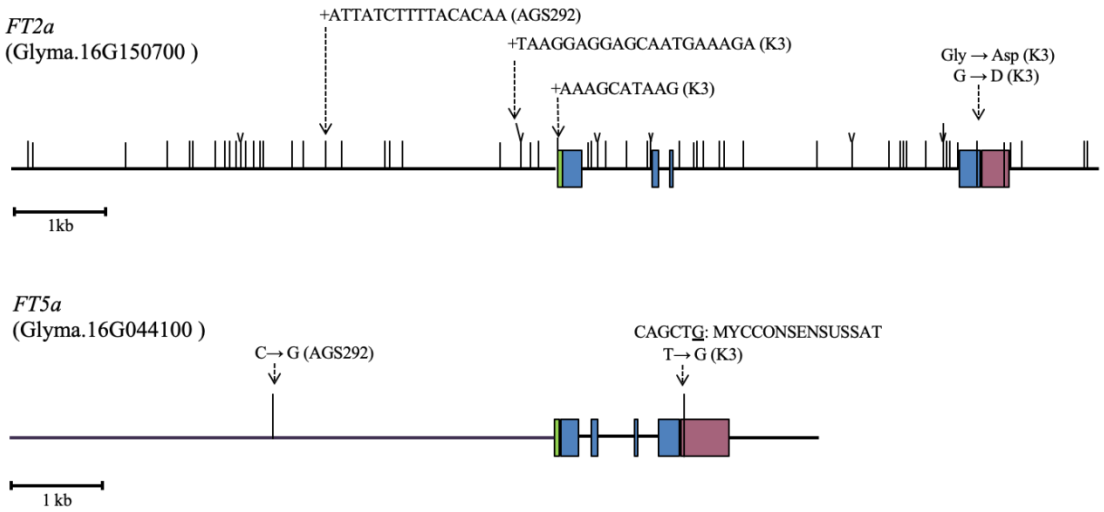

Figure 4.1. DNA polymorphisms for *FT2a* and *FT5a* between AGS292 and K3. AK16 possessed alleles from K3 at both *FT2a* and *FT5a* loci. Lines indicate SNPs and indels of 1 to 9 bp. Green, blue and purple boxes indicate 5′ UTR, exon and 3′ UTR, respectively.

```
                                                                  976
A:   AAGTTATCTT ATATATTATA ACTCGTCGCC ACCCTTTAGT ACCGCNNNNN NNNNNNNNNN
B:   AAGTTATCTT ATATATTATA ACTCGTCGCC ACCCTTTAGT ACCGCNNNNN NNNNNNNNNN
C:   AAGTTATCTT ATATATTATA ACTCGTCGCC ACCCTTTAGT ACCGCNNNNN NNNNNNNNNN
D:   AAGTTATCTT ATATATTAT- ---------- ---------- ---------- ----------

A:   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
B:   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
C:   NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
D:   ---------- ---------- ---CTTACAA AAATACAAGT TTGAATCTCA CGAAACAATA

A:   NNNNNNNNNN NNNNNNNNNN NNNNNAAAGA AAATTTGAAC TTTAAAATAT GTCTTAACTT
B:   NNNNNNNNNN NNNNNNNNNN NNNNNAAAGA AAATTTGAAC TTTAAAATAT GTCTTAACTT
C:   NNNNNNNNNN NNNNNNNNNN NNNNNAAAGA AAATTTGAAC TTTAAAATAT GTCTTAACTT
D:   TGTGAAGACC TATAAATATT TCTTAAAAGA AAATTTGAAC TTTAAAATAT GTCTTAACTT
                                                  1101
```

Figure 4.2. The sequences alignment data of 957 bp~1,136 bp downstream to the stop codon of *FT5a*. The red background area indicates the region of the difference between the four sequences. "N" indicates an unconfirmed base sequence. The red letters with underline indicate the sequences identified by genome walking in this study. (A) The sequences in Williams 82 from Glycine Max Wm82.a2.v1 reference genome. (B) The sequences in AGS292 estimated based on resequencing data. (C) The sequences in K3 estimated based on resequencing data. (D) The sequences of Williams 82 (same as AGS292 and K3) identified by genome walking.

Chapter V Binding Affinity of E1 Family and E2 Proteins to DNA Fragments

5.1 Background and Purpose

In the previous chapter, I detected an SNP at 3' UTR of *FT5a* gene between AGS292 and K3. This SNP created a cis-element of MYCCONSENSUSAT (CANNTG) in K3. The MYC recognition sequence was the binding site of *MYC2*, which encodes a bHLH transcriptional factor that negatively regulates blue light-mediated photomorphogenic development and the expression of blue and far-red light-regulated genes (Kazan and Manners, 2013). MYC2, together with MYC3 and MYC4, redundantly regulates flowering by modulating the transcription of *FT* under both LD and SD conditions (Wang et al., 2017). The expression of *MYC*-like genes is dependent on different temperature conditions (Figure 2.5). Moreover, the expression levels of *MYC* genes at 4th leaf stage at 25℃ were upregulated in AK16 (Figure 2.5).

The E1 protein contains a putative binary nucleus localization signal and has a domain distantly related to the plant specific B3 domain (Xia et al., 2012). B3 domain is a highly conserved domain that is widely found in the genomes of higher plants and has sequence-specific DNA binding activity that can play a regulatory role in plant growth and development by binding to specific DNA. Several members of the B3 superfamily regulate flowering directly or indirectly. For example, ABI3/VP1 (RAV1) in *Arabidopsis* affects flowering time (Hu et al., 2004). In soybean, E1 and E1-like proteins regulate flowering time through down-regulation of *FT2a* and *FT5a* (Xu et al., 2015). However, how E1 and E1-like proteins regulate the transcription of *FT2a* and *FT5a* has not yet been elucidated.

The *E2* gene is an ortholog of *Arabidopsis GIGANTEA* (*GI*) (Watanabe et al., 2011); the GI protein interacts with FLAVIN-BINDING, KELCH REPEAT, F-BOX 1 (FKF1) to up-regulate the expression of CONSTANS (CO) through degradation of CYCLING DOF FACTOR (CDF) and activate *FT* expression (Huq et al., 2000). It is also considered that GI directly binds to a cis-element of *FT* to activate the expression (Sawa et al., 2007). Accordingly, it is worth to explore whether E2 directly binds to cis-elements of *FT2a* and *FT5a* or not.

In this chapter, I conducted the yeast-one-hybrid assay to confirm whether MYC, E1, E1-like, and E2 proteins can bind to the MYCCONSENSUSAT cis-element in 3' UTR of *FT5a* gene in AGS292. In addition, I also analyzed the binding affinity of E1 family proteins to the various DNA sequences similar to the MYCCONSENSUSAT cis-element.

5.2 Materials and Methods

5.2.1 Construction of recombinant bait plasmids

Two DNA segments, in which a unit of 6 nucleotides including the SNP in the 3' UTR of FT5a, MYCCONSENSUSAT (CAGCTG) of K3 and its variant sequence (CAGCTT) of AGS292, was tandemly repeated six times, were ligated to restriction sites of Sac I or Xho I in each side of the segment, and their complimentary sequences were constructed. The segment and its complementary sequence were then gradually annealed according to the method of Matchmaker® Gold Yeast One-Hybrid Library Screening System User Manual (Takara, Clontech, Otsu, Japan). The double-strand DNA fragments for the cis-element RVA1-B (CACCTG) and six sequences of 6 bp (Supplemental Table 5.1) were also developed as E1-binding sites using the same procedure.

The pAbAi vector carrying the URA3 and AUR1-C genes obtained from Matchmaker® Gold Yeast One-Hybrid Library Screening System (Takara, Clontech, Otsu, Japan) were digested with Sac I and Xho I. The digested product of the pAbAi vector was purified using the Millipore® Amicon® Ultra-0.5 Centrifugal Filter Devices and then ligated with the synthesized double-stranded DNA fragments using T4 ligase. The ligated products were transformed into *E. coli* JM109 competent cells and the positive recombinant clones were screened using Luria-Bertani solid media containing ampicillin. Colony PCR electrophoresis using primers presented in Supplemental Table 5.1 was used to confirm the insertion of the target fragments. Positive clones were selected for plasmid extraction and then the sequence was confirmed by DNA sequencing.

5.2.2 Transformation of linearized bait plasmids into yeast cells

Yeast competent cells of Y1HGold were prepared using the lithium acetate (LiAc) method. The recombinant bait plasmid and the p53-AbAi control vector plasmid linearized by BtsB I were transformed and integrated into the genome of yeast, and then the transformed competent cells were transferred on to solid agar synthetic defined (SD) medium-Ura and incubated for 3 to 4 days at 30 °C. Single clones were then identified by colony PCR using Matchmaker® Insert Check PCR Mix 1 (Takara, Clontech, Otsu, Japan). The expected size of PCR products was 1.4 kb.

Three solid media were used to detect *AbA*^{r} expression for the bait strains (pAbAi-AGS, pAbAi-K3, pAbAi-RAV1B, pAbAi-V1, pAbAi-V2, pAbAi-V3, pAbAi-V4, pAbAi-V5, pAbAi-V6); SD/-Ura without aureobasidin A (AbA), SD/-Ura with 100 ng/mL AbA, and SD/-Ura with 200 ng/mL AbA. After incubation at 30 °C for 2 to 3 days, the minimum AbA concentration for each bait strain that completely inhibited colony growth was determined and used for further experiment.

5.2.3 Construction of recombinant prey plasmids

The yeast cloning vectors pGADT7 containing the fused GAL4 DNA activation domain (AD) used in the yeast one-hybrid assays were obtained from Matchmaker® Gold Yeast One-Hybrid Library Screening System (Takara, Clontech, Otsu, Japan). The full-length coding sequences of *MYC2b* were inserted into pGADT7 vectors and named pGADT7-MYC2b. The toxicity test was carried out in the yeast strain Y1H on SD/-Leu medium. The primer sequences used to clone *MYC2b* are listed in Supplemental Table 5.2.

pGADT7-E1, pGADT7-E1La, pGADT7-E1Lb and pGADT7-E2 (pGADT7 vectors containing the fused GAL4 DNA activation domain (AD) including the full-length cDNA of *E1*, *E1La*, *E1Lb* and *E2* were also used in the yeast one-hybrid assays. These AD vectors were developed by Dr. Meilan Xu and tested for the toxicity in the yeast strain Y187 on SD/-Leu medium by Dr. Jianghui Zhu.

5.2.4 DNA binding analysis for E1, E1La, E1Lb and E2

Yeast competent cells of the bait strains pAbAi-AGS292 and pAbAi-AK16 were prepared using the lithium acetate (LiAc) method. The cells were then transformed with plasmids of pGADT7-E1, pGADT7-E1La, pGADT7-E1Lb and pGADT7-E2.

The transformed competent cells were transferred on to solid agar SD/-Leu without AbA and SD/-Leu with 100 ng/mL AbA (or 200 ng/mL AbA), i.e, SD/-Leu and SD/-Leu AbA100 (or AbA200) for 3 to 4 days at 30 °C for incubation. Whether or not the four proteins are combined with the bait sequences can be determined depending on the growth situation of colony.

The monoclonal colonies that were initially confirmed to be bound, that is, the yeast monoclonals were grown in the plate containing AbA, were cultured in YPDA liquid medium for 1 hour at 30 °C and resuspended in physiological saline. The yeast cell suspension solution was diluted to 1/10 and 1/100. Then 1 μL of the diluted and undiluted yeast cell suspension solutions were applied to the solid medium plates of SD/-Leu AbA100 and SD/-Leu AbA200 to further confirm the binding of the protein to the DNA sequence.

5.3 Results and Discussion

5.3.1 MYC2b protein has toxicity to the yeast strain Y1HGold study

In the toxicity test of MYC2b AD vector, I found that even in the medium without antibiotic AbA screening pressure, yeast Y1HGold transformed with the *MYC2b* recombinant prey plasmid pGADT7-MYC2b could not grow. In order to rule out the possibility of experimental errors, I repeated the experiment twice with the control group, and the same results were obtained in the two experiments: the control group could grow normally, while the MYC2b group could not. The results indicate that MYC2b protein has toxicity to the yeast strain Y1HGold used in this study, and the yeast-one-hybrid assay could not be used for the assay of the binding affinity of MYC2b protein to the MYCCONSENSUSSAT cis-element. Therefore, I analyzed the binding affinity of E1, E1La, E1Lb and E2 to this cis element.

5.3.2 DNA binding analysis of E1, E1La, E1Lb and E2 proteins with yeast bait strain pAbAi-MYCCONSENSUSSAT (CAGCTG)

As shown in Figure 5.1, the yeast bait strains pAbAi-MYCCONSENSUSSAT (CAGCTG) transformed by pGADT7-E1, pGADT7-E1La, pGADT7-E1Lb and pGADT7-E2 could grow on SD/-Leu but not on SD/-Leu AbA100 mediums. Therefore, all the four proteins could not bind to the element CAGCTG/CAGCTG.

5.3.3 DNA binding analysis of E1, E1La, E1Lb and E2 proteins to the variant CAGCTT element

As shown in Figure 5.2A to C, the yeast bait strains pAbAi-CAGCTT transformed by pGADT7-E1, pGADT7-E1La and pGADT7-E1Lb grew well on dropout medium SD/-Leu AbA100. Therefore, they could grow well on the SD/-Leu AbA100 medium, even when the yeast cell suspension solution was diluted to 1/10 and 1/100. On the SD/-Leu AbA200 medium, 1/100-diluted solution of pAbAi-CAGCTT/pGADT7-E1 and pAbAi-CAGCTT/ pGADT7-E1La produced no colony, although all of the combinations of pAbAi-CAGCTT with three E1 family proteins could grow well when non-diluted or 1/10-diluted solutions were used. In contrast, the combination of pAbAi-CAGCTT/pGADT7-E1Lb grew well even when the suspension solution was diluted to 1/100 (Figure 5.1E). These results suggest that the E1, E1La and E1Lb proteins all could bind to the CAGCTT element.

On the other hand, the yeast pAbAi-CAGCTT/pGADT7-E2 could grow on SD/-Leu but not on SD/-Leu AbA100 (Figure 5.1D). The E2 protein could not thus bind to the CAGCTT element.

5.3.4 DNA binding analysis of E1, E1La and E1Lb proteins with yeast bait strains carrying various elements V1-V6

Since the E1 family protein could bind to the CAGCTT element, variant elements were analyzed for the binding affinity with the E1 family proteins. The variant elements were different in single nucleotides from the CAGCTT element (Table 5.1).

As shown in Figure 5.3, the yeast bait strains pAbAi-V1, pAbAi-V2, pAbAi-V4 and pAbAi-V6 transformed by pGADT7-E1, pGADT7-E1La and pGADT7-E1Lb grew well on dropout medium SD/-Leu AbA100 or SD/-Leu AbA200. Then, the binding affinities were further tested using 1/10- and 1/100-diluted suspension solutions. Some of them could not grow so well when the 1/100-diluted solution was tested on SD/-Leu AbA200, but all of them could grow well when the 1/10-deluted solution was used. These results suggest that E1, E1La and E1Lb proteins could also bind to these variant CAGCTT elements V1, V2, V4 and V6. In contrast, the yeast pAbAi-V3 and pAbAi-V5 transformed by pGADT7-E1, pGADT7-E1La and pGADT7-E1Lb could grow on SD/-Leu but not SD/-Leu AbA100 (Figure 5.3C and E), suggesting that the E1, E1La and E1Lb protein cannot bind to these two variants.

5.3.5 DNA binding analysis of E1, E1La and E1Lb proteins with yeast bait strain RAV1-B

RAV1, a member of the B3 superfamily in *Arabidopsis*, which is confirmed that affects flowering time (Hu et al., 2004). E1 and E1-likes also carried the minimal B3 domain structure required for DNA contact (Yamasaki et al., 2004; Swaminathan et al., 2008; Xia et al., 2012). Thence I analyzed the binding affinity of E1 family proteins to the RAV1-B element (CACCTG) to verify whether the E1 family proteins can bind to the binding site of B3 domain.

The yeast bait strains pAbAi-RAV1-B transformed by pGADT7-E1, pGADT7-E1La and pGADT7-E1Lb, respectively could grow on SD/-Leu but not on SD/-Leu AbA100 (Figure 5.5), suggesting that all the three proteins cannot bind to the RAV1-B element.

5.3.6 Distribution of E1-element and variants in the *FT2a* and *FT5a* genomic regions

The CAGCTT element and its variant sequences were renamed as shown in Table 5.1, depending on binding ability with the E1 protein. The *FT2a* and *FT5a* sequences from 6.0 kb upstream from the start codon to 1.0 kb downstream to the stop codon were constructed for 22 accessions, based on the whole-genome resequencing data. As shown in Figure 5.6, E1-elements that could be bound by the E1 family proteins

(E1-element C1~C5) were distributed widely in the coding, promoter and downstream regions of *FT2a* and *FT5a*. In the *FT2a* promoter there were two polymorphisms for E1-element C3(-) among the 22 accessions. In three wild soybean varieties, C3(-)-4 in *FT2a* was changed to the sequence AAAATG, which was not assayed for the binding with the E1 protein. E1-element C3(-)-5 in *FT2a* in AGS292 and some cultivars was changed to the E1-element V1(-), which could not be bound by the E1 proteins. For *FT5a*, one polymorphism of the E1-element C1 was observed in 3′ UTR among the 22 accessions; 19 accessions possessed the E1-element C1, as in AGS292, but the other three accessions did the MYC-CONSENSUSAT element, as Williams 82 and K3.

Table 5.1. E1-elements tested in the genomic region of *FT2a* and *FT5a*

Name used in yeast one-hybrid assays			+	-	Binding ability	FT5a		FT2a	
						(+)	(-)	(+)	(-)
E1 element		C1	CAGCTT	AAGCTG	+	1	0	4	1
E1AGS element V2		C2	AAGCTT	AAGCTT	+	4		4	
E1AGS element V4		C3	CATCTT	AAGATG	+	3	1	2	9
E1AGS element V6	E1-element	C4	CACCTT	AAGGTG	+	0	2	2	2
E1AGS element V1		C5	CAGATT	AATCTG	+	7	4	3	2
E1AGS element V3		V1	CAACTT	AAGTTG	-	9	5	7	4
E1AGS element V5		V2	CAGTTT	AAACTG	-	0	2	2	7
E1AK16 element	MYC-CONSENSUSAT		CAGCTG	CAGCTG	-	1		0	
RAV1-B	RAV1-B		CACCTG	CAGGTG	-	1	0	0	0

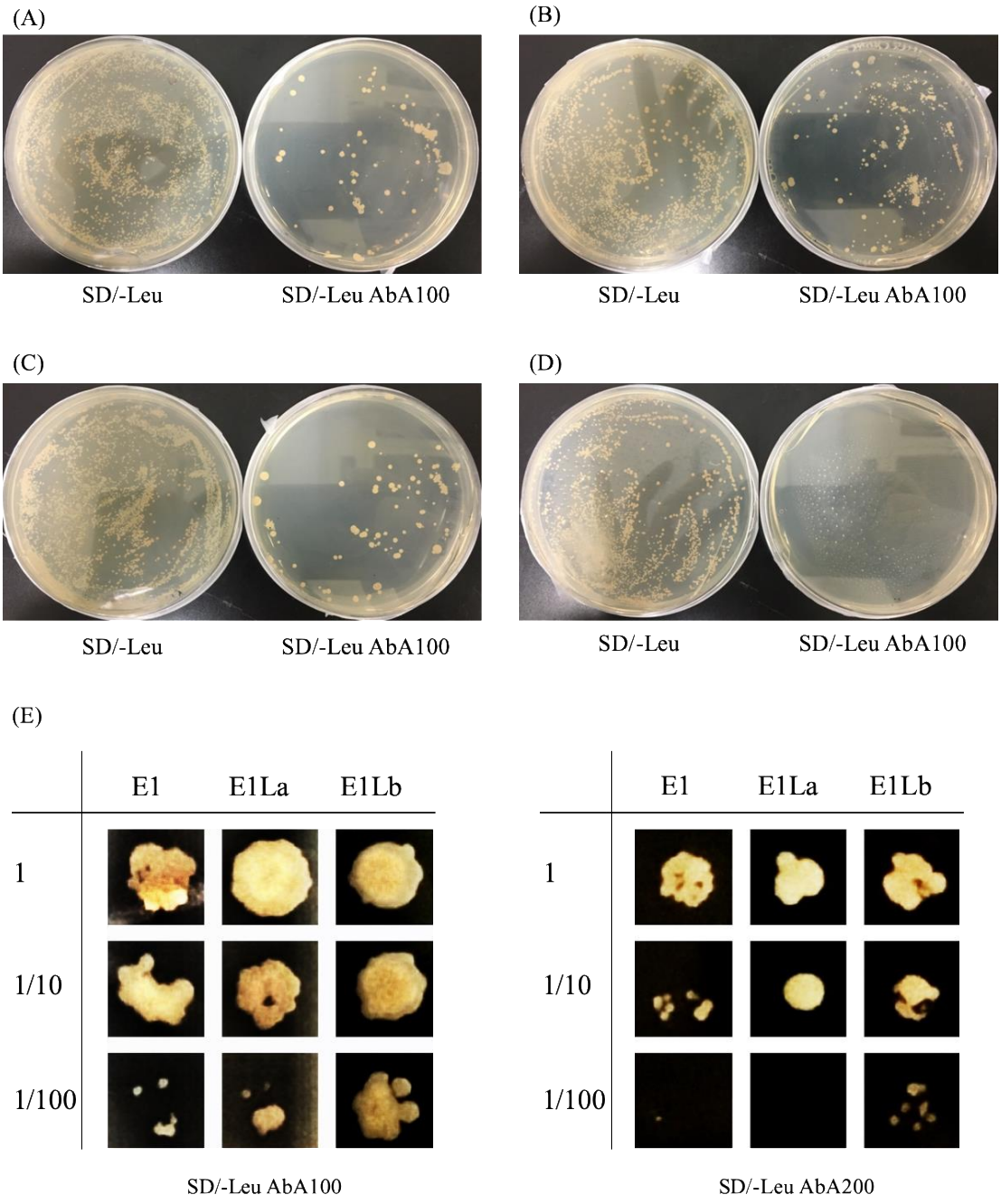

Figure 5.1. DNA binding analysis of E1, E1La, E1Lb, E2 proteins with yeast bait strain pAbAi-AGS292 on SD/-Leu and SD/-Leu AbA100 mediums. (A) E1; (B) E1La; (C) E1Lb; (D) E2; (E) 1 μL of 1, 1/10, 1/100 transformant yeast cells on SD/-Leu AbA100 and SD/-Leu AbA200 mediums. Growth of the transformants on dropout medium SD/-Leu and SD/-Leu AbA100 indicates that except E2, E1, E1La, E1Lb proteins all could bind to the E1AGS element.

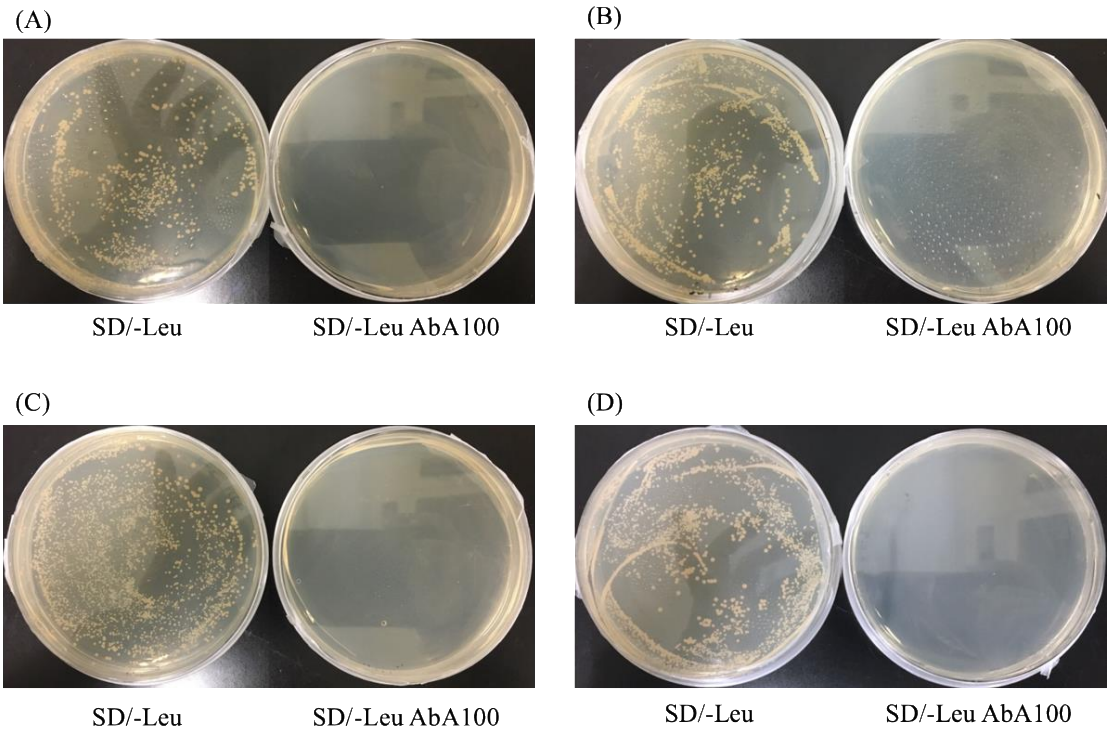

Figure 5.2. DNA binding analysis of E1, E1La, E1Lb, E2 proteins with yeast bait strain pAbAi-AK16 on SD/-Leu and SD/-Leu AbA100 mediums. (A) E1; (B) E1La; (C) E1Lb; (D) E2. The transformants did not grow on dropout medium SD/-Leu and SD/-Leu AbA100, suggesting that E1, E1La, E1Lb and E2 proteins all could not bind to the E1AK16 element.

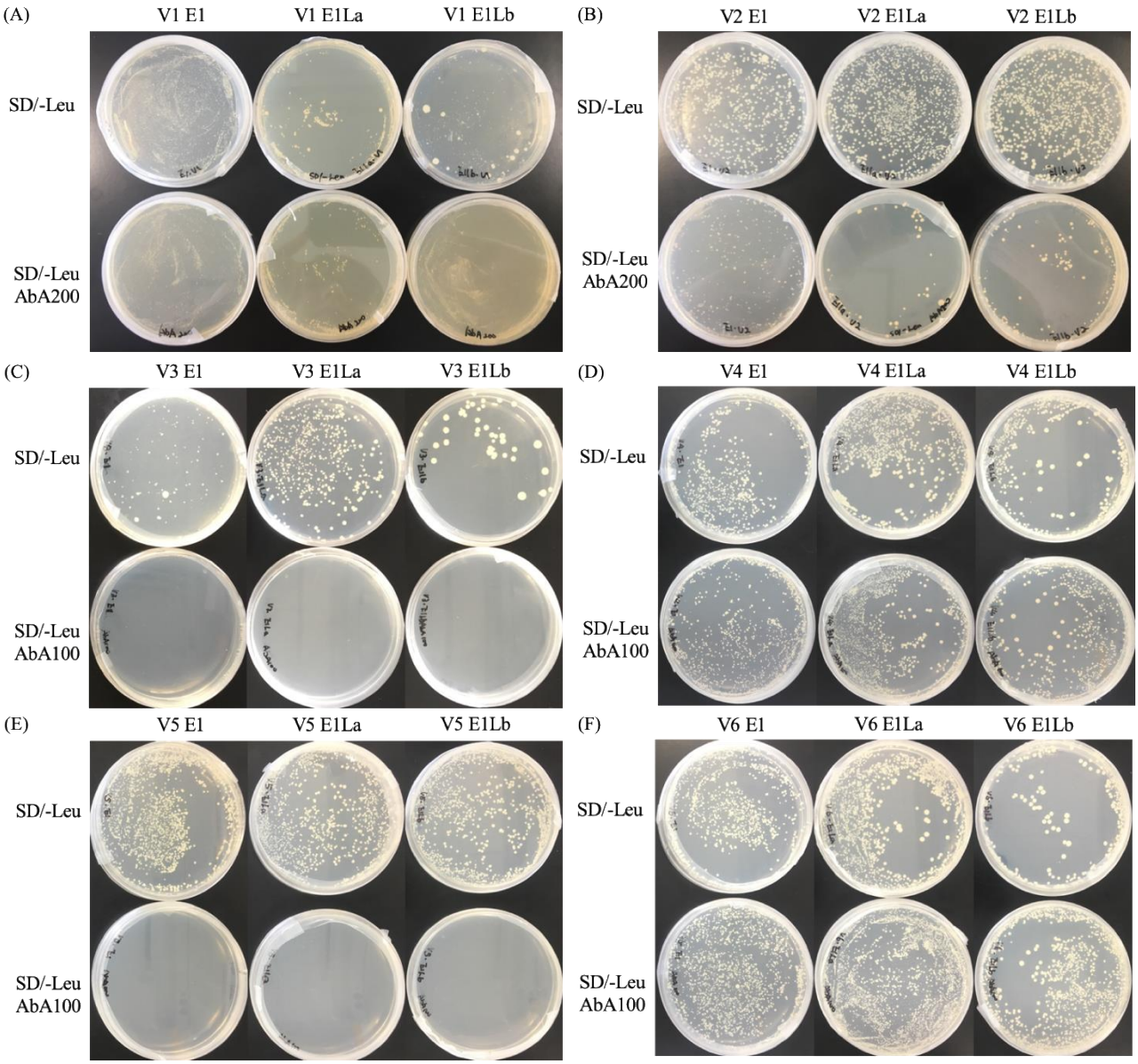

Figure 5.3. DNA binding analysis of E1, E1La, E1Lb, E2 proteins with yeast bait strains pAbAi-V1 to V6 on SD/-Leu and SD/-Leu AbA100 or SD/-Leu AbA200 mediums. (A) pAbAi-V1; (B) pAbAi-V2; (C) pAbAi-V3; (D) pAbAi-V4; (E) pAbAi-V5; (E) pAbAi-V6. Growth of the transformants on dropout medium SD/-Leu and SD/-Leu AbA100 indicates that E1, E1La and E1Lb proteins could bind to the variants of E1AGS element V1, V2, V4 and V6.

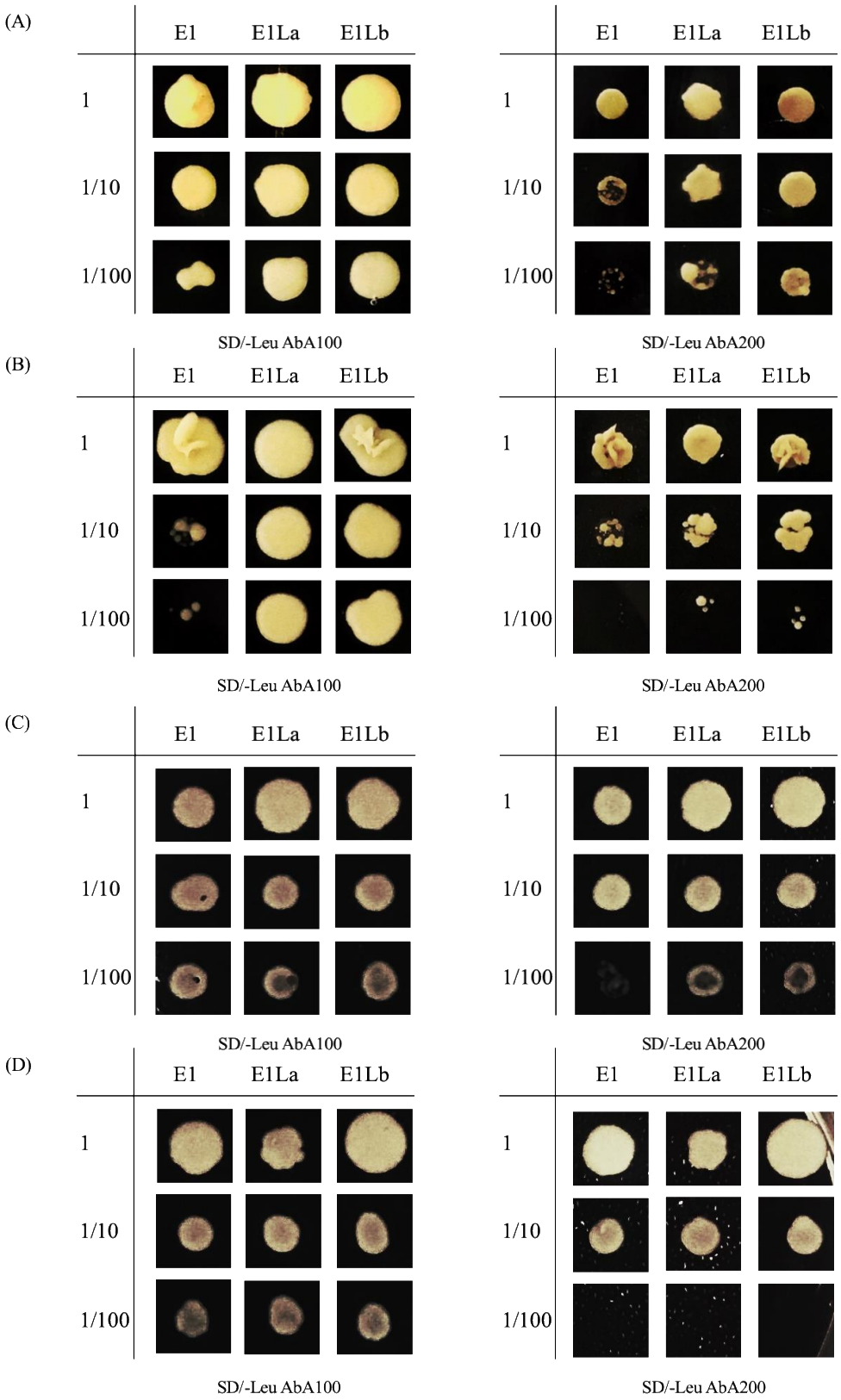

Figure 5.4. 1 µL of 1, 1/10, 1/100 transformant yeast cells on SD/-Leu AbA100 and SD/-Leu AbA200 mediums. (A) pAbAi-V1; (B) pAbAi-V2; (C) pAbAi-V4; (D) pAbAi-V6.

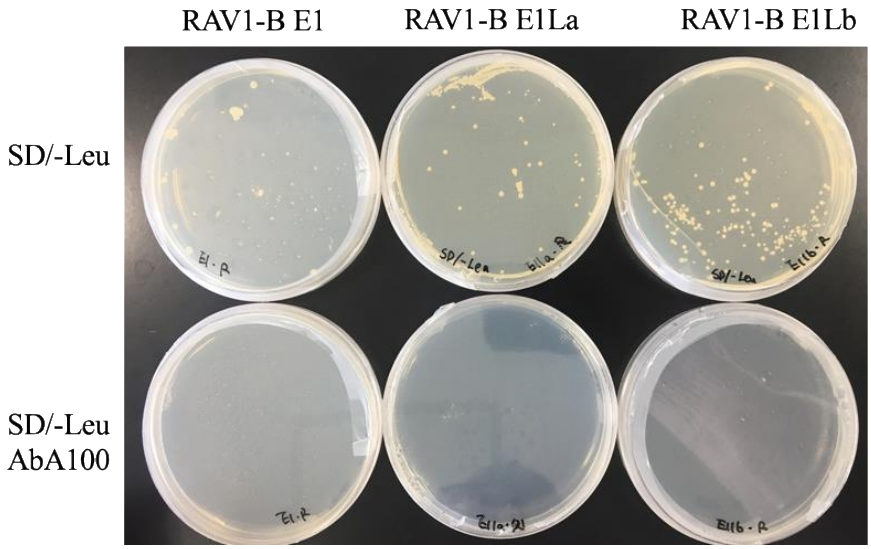

Figure 5.5. DNA binding analysis of E1, E1La, E1Lb, E2 proteins with yeast bait strain pAbAi-RAV1-B on SD/-Leu and SD/-Leu AbA100 mediums. Growth of the transformants on dropout medium SD/-Leu and SD/-Leu AbA100 indicates that E1, E1La and E1Lb proteins all could not bind to the RAV1-B element.

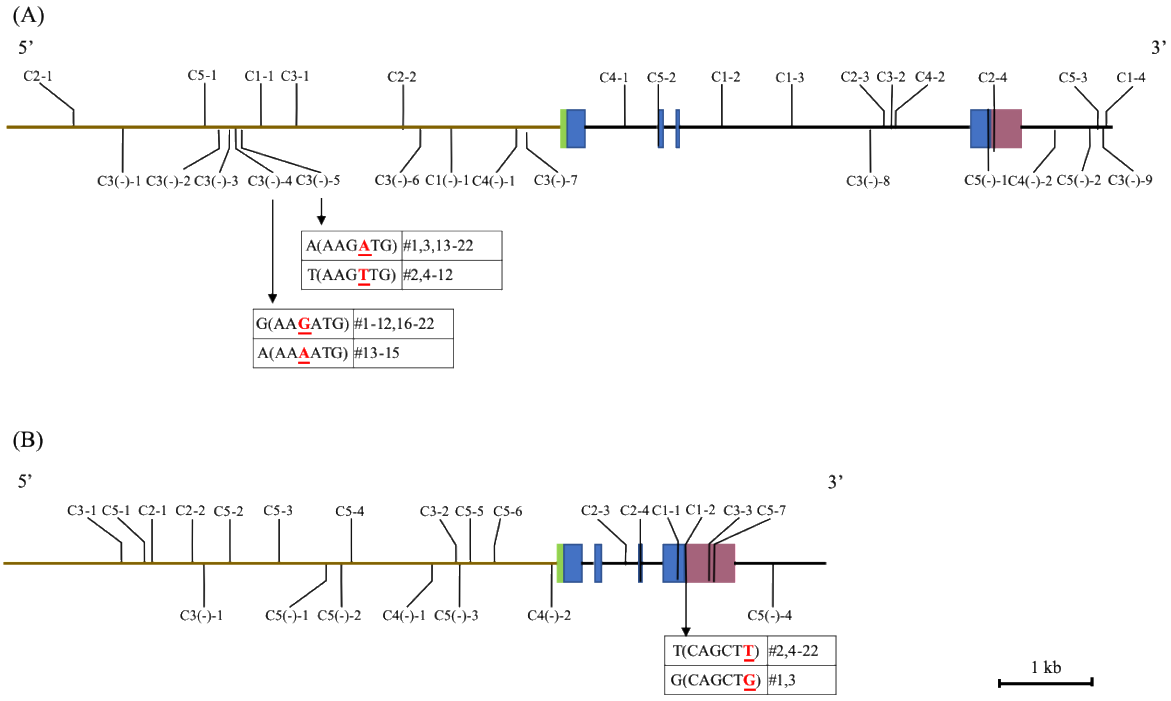

Figure 5.6. Presence of E1-element and variants in *FT2a* and *FT5a* from 6.0kb upstream from the start codon to the 1.0kb downstream to the end codon. (A) *FT2a*; (B) *FT5a*. 1,Williams82, 2,AGS292, 3,K3, 4,Otofuke, 5,Toyomusume, 6,TK780, 7,Tachinagaha, 8,Tokachinagaha, 9,Harosoy, 10,T106, 11,KAS, 12,R23, 13,B4009, 14,B5023, 15,B0090, 16,B0092, 17,B9040, 18,B8045, 19,Hidaka, 20,NL2, 21,Tanishi, 22,Kariyutaka.

Chapter VI General Discussion

6.1 Characteristics of the late-flowering habit from K3 in a photoperiod-insensitive genetic background

The long juvenile (LJ) habit in soybean has been characterized using reciprocal transfer experiments from inductive SD to non-inductive LD conditions or vice versa. The juvenile phase varies from ~10 to 30 d among soybean cultivars (Destro et al., 2001; Willkerson et al., 1989; Ellis et al., 1992; Collinson et al., 1993; Upadhyay et al., 1994). However, Cober (2011) demonstrated that the LJ cultivars Paranagoiana (*e6*/*e6*) and PI159925 (*j*/*j*) retained the sensitivity to shorter photoperiods of less than 12 h, while the conventional juvenile genotype lacking the juvenile phase did not. He assumed that the juvenile phases of these LJ cultivars lasted for a maximum of 5 d, and a flowering delay of more than 5 d was likely caused by photoperiod responses. AK16 used in this study did not possess the *j* allele but exhibited a flowering delay of 5 to 7 d with 8-h to 16-h photoperiods at a constant 25℃ relative to the flowering time of AGS292. The delay increased to 11.7 d with R-enriched LDs of 20 h and to 20.1 d with a 16-h photoperiod at a constant 32℃. Thus, the LF habit introduced from K3 into the photoperiod-insensitive genetic background may involve a basic difference in the activity of floral induction itself and the suppression of flowering by higher temperatures and longer day lengths generated by R-enriched light sources.

AK16 has the double-recessive genotype (*e3*/*e4*), which enables flowering under FR-enriched LD conditions (Saindon et al., 1989; Cober et al., 1996). It did not respond to FR-LD conditions, like AGS292, and the flowering time was almost the same under ND and FR-LD conditions (Figure 3.1). However, AK16 retained sensitivity to photoperiods supplied by R-enriched lights (Figure 2.1A). This was an unexpected result because the response of flowering to R-LD is controlled only by the *E3* gene (Buzzell et al., 1971; Kilen et al., 1971). *E3* and *E4* participate in different functions of PHYA in *Arabidopsis* (Liu et al., 2008; Watanabe et al., 2009). E4 and its homoeolog PHYA1 are redundantly responsible for the de-etiolation response of hypocotyls to FR light, and in an *e3*/*e3* genotype, *e4*/*e4* controls flowering under FR-LD conditions (Buzzell et al., 1971; Cober et al., 1996; Liu et al., 2008). In contrast, *E3* is responsible for PHYA's photoreceptor function for R-light with a high

photon irradiance level (Watanabe et al., 2009). The QTL mapping in the original RIL population from a cross between AGS292 and K3 segregating at the *E3* locus demonstrated that *E3* was involved in the control of flowering under a wide range of photoperiods, from the SD conditions of low latitudes to the LD conditions of high latitudes (Lu et al., 2015). Because AK16 lacks the functional E3 protein, the residual response to the R-LD conditions might be attributed to the other PHYA proteins, such as PHYA1 and PHYA4 (Liu et al., 2008; Watanabe et al., 2009), other phytochromes, such as PHYB and PHYE (Wu et al., 2013), and/or the blue light photoreceptor CRYPTOCHROME 2 (Zhang et al., 2008; Matsumura et al., 2009). The PHYA4 protein of Williams 82 is considered to be defective in function compared to the wild type, because of the lack of a long fragment (0.8 kb) covering the third exon and part of the following intron sequence, due to altered splicing caused by a single nucleotide substitution at the splicing site (Wu et al., 2013; Li et al., 2014). In this study, the *PHYA4* sequence was identical between AGS292 and AK16, although K3 possessed a dysfunctional *PHYA4* gene due to the frameshift that generated premature termination at the 302th aa (Table 4.5). *PHYA1* showed no nonsynonymous substitution, but other four *PHY* genes showed several nonsynonymous substitutions between AGS292 and K3 (Table 4.5). The sequences of *PHYB1* and *PHYB2* possessed two and one nonsynonymous substitutions, respectively; it was thought that PHYB1 and PHYB2 from AGS292 were most likely functional wild-type proteins, and those from K3 were variants (Table 4.5). Five of the six SNPs in the coding region of *PHYE1* resulted in non-synonymous substitutions, and *PHYE2* exhibited a non-synonymous substitution in the exon 2 (Table 4.5). It is intriguing to investigate whether these PHY genes are responsible for the residual response of AK165 to R-light enriched LD condition.

In AK16, flowering was also inhibited by hot temperatures (Figure 2.1B). Compared with the constant 25℃ condition, constant or daytime 32℃ conditions promoted flowering in AGS292, but adversely inhibited it in AK16. As in *Arabidopsis* (Balasubramanian et al., 2006), increasing temperature accelerates flowering in soybean (Destro et al., 2001). In contrast to the photoperiodic regulation, however, little is known about the genetic variability in the thermal regulation of flowering in soybean. Cober et al. (2001) found that different temperatures affected flowering under LD conditions of 16 h or longer. Compared with low temperature

(18℃), a hot temperature (28℃) markedly delayed the flowering, and the effect was more pronounced in late maturing photoperiod-sensitive genotypes carrying two or more dominant alleles at the *E1*, *E3*, *E4* and *E7* loci relative to the photoperiod-insensitive lines. Whether the inhibition of flowering by hot temperatures observed in AK16 is a common characteristic of the LF habit of LJ cultivars at low latitudes should be determined in a further study. AK16 could be used as a plant resource in the molecular and genetic dissection of the thermoregulation of flowering in soybean.

6.2 QTLs for the late-flowering habit in the photoperiod-insensitive genetic background

In this study, I identified three QTLs (*qDTF-10*, *qDTF-16-1* and *qDTF-16-2*) controlling the late0flowering (LF) habit that originated from the Thai cultivar K3 in a photoperiod-insensitive genetic background (Table 3.1 and Figure 3.3). *qDTF-10*, *qDTF-16-1* and *qDTF-16-2* were co-localized with gene-specific DNA markers for *E2*, *FT5a* and *FT2a*, respectively (Table 3.1). The intervals between left and right markers of respective QTLs contained 30 in *qDFT-16-1* to 167 in *qDFT-10* annotated genes including several orthologues of *Arabidopsis* flowering genes (Table 3.2). However, *FT5a*, *FT2a* and *E2* are considered as the most probable candidates responsible for the *qDTF-16-1*, *qDTF-16-2* and *qDTF-10*, respectively, because the previous studies have revealed the allelic effects on flowering of *FT5a*, *FT2a* and *E2* in diverse genetic backgrounds (Watanabe et al., 2011; Zhao et al., 2016; Takeshima et al., 2016).

Of the three QTLs, the *qDTF-10* (*E2*) was detected only under outdoor ND and FR-LD conditions, whereas the other two were detected in all the environments, although the effect at *qDTF-16-1* was not significant at the 5% level under R-LD conditions as assessed by genome-wide permutation tests. In the original RIL population of the cross between AGS292 and K3 segregating for *E3* and *E4*, four QTLs corresponding to *FT2a*, *FT5a*, *E2*, and *E3* were detected in the four SD environments tested (Lu et al., 2015). The effects of these QTLs, except for the QTL corresponding to *E3*, which was not targeted in this study, could be confirmed in the photoperiod-insensitive genetic background as well.

FT5a is an important *FT* ortholog in soybean and could be considered a candidate for *qDTF-16-1*. However, the DNA markers with the largest LOD scores varied with the environments tested. The intervals with the largest LOD scores included the gene-specific marker for *FT5a* under all the conditions, except R-LD conditions, where the greatest LOD score was located in the interval of the next two SNPs, SNP 4900158 and SNP, approximately 400 kb from the former (Table 3.1). The QTL analysis of the original RIL population also detected two adjacent QTLs in the genomic region of *qDTF-16-1* (Lu et al., 2015). It should be determined in a further study whether a common QTL or each of the two closely linked QTLs is involved in flowering times under different photo-thermal conditions.

The *FT5a* sequences were almost the same between AGS292 and K3, except for a SNP in the 3′ UTR of the latter, which generated a cis-element, MYCCON-SENSUSAT, a binding site for the basic-helix-loop-helix transcriptional factor MYC2 (Figure 4.1). The MYC2 protein in *Arabidopsis* is a master regulator in various jasmonate-regulated physiological and developmental pathways, such as lateral and adventitious root formation, lowering time and shade avoidance syndrome (Ho et al., 2014). MYC2, together with MYC3 and MYC4, redundantly regulates flowering by modulating the transcription of *FT* under both LD and SD conditions (Wang et al., 2015). Therefore, the MYCCONSENSUSAT element might function as a binding site of soybean MYC2 orthologs to control the *FT5a* expression in response to various environmental stresses. The *FT5a* expression level in AK16 was subject to repression by severe environmental stresses, such as lower or higher temperatures (Figures 2.4 and 2.5). However, as the MYC2 protein was toxic to the yeast strain Y1HGold used in the Y1H assay, it could not be verified whether the MYC2 protein binds to the MYCCONSENSUSAT element site in this study.

Another candidate gene of *qDTF-16-1* is *VRN1* (Table 3.2). The results of resequencing analysis showed a number of sequence variations for *VRN1* between AGS292 and K3. A comparison with the sequences of other species suggests that the VRN1 protein from K3 is most likely a functional wild-type protein and those from AGS292 and Williams 82 are a variant (Tables 4.2). However, *VRN1* was hardly expressed under almost all conditions in both AGS292 and AK16. Therefore, *FT5a* was considered as the most probable candidate of *q-DTF-16-1*.

FT2a is the most probable candidate responsible for the *q-DTF-16-2*, because the other candidate, *FT2b*, showed no amino acid sequence variation between AGS292 and K3 (Table 4.3). *FT2a* is an important floral inducer in soybean (Kong et al., 2010; Sun et al., 2011; Cai et al., 2018). When overexpressed using the Cauliflower mosaic virus 35S promoter or knocked out using CRISPR-Cas9, *FT2a* promotes and inhibits flowering, respectively (Sun et al., 2011; Nan et al., 2014; Cai et al., 2018). Furthermore, the DNA polymorphisms in the *FT2a* genomic region were associated with variation in flowering time in soybean germplasm having diverse flowering habits (Jiang et al., 2019; Ogiso-Tanaka et al., 2019). In particular, the non-synonymous nucleotide substitution at exon 4 from guanine to adenine, which converts the glycine to the aspartic acid at the 169th aa residue, was highly associated with the late-flowering habit in the world-mini soybean core collection (Ogiso-Tanaka et al., 2019) and 127 varieties evenly covering diverse maturity groups (Jiang et al., 2019). The resequencing analyses detected diverse DNA polymorphisms between AGS292 and K3, including the indels of 20 bp and 10 bp in the promoter and 5′ UTR, respectively, as well as the non-synonymous SNP at exon 4, in which K3 possessed adenine (Figure 4.1), as in the late-flowering cultivars tested by Jiang et al. (2019) and Ogiso-Tanaka et al. (2019). The aa substitutions in the external loop of the FT protein convert its function from a floral activator to a repressor (Hanzawa et al., 2005). The glycine at the 169th aa residue of FT2a in AGS292 and Williams 82 is highly conserved in FT-like phosphatidylethanolamine-binding proteins of diverse plant species (Wang et al., 2015). Thus, the conversion from glycine to aspartic acid at the 169th aa residue might result in a loss of FT2a's function as a floral inducer. Thus, the variant FT2a protein of K3 is most likely involved in the delay of flowering under SD conditions as shown in the original RIL population (Lu et al., 2015) and also under various photoperiod conditions in a photoperiod-insensitive genetic background in this study. The effect of the aa substitution from glycine to the aspartic acid on floral induction should be validated using ectopic expression analyses in a further study. In addition to this missense variant, another missense and one flame-shift variants were also detected in the accessions introduced from lower latitudes, such as in Taiwan, and South-eastern and South Asian countries (Ogiso-Tanaka et al., 2019). The dysfunctional variants of FT2a might, therefore, have been selected for as LJ genes during the adaptation of soybeans to lower latitudinal environments.

6.3 Molecular mechanisms underlying the long juvenile trait in soybean

J encodes the soybean ortholog of *Arabidopsis* EARLY FLOWERING 3, a component of the evening complex (Yue et al., 2017; Lu et al., 2017). It physically associates with the *E1* promoter to downregulate its transcription, relieving the repression of *FT2a* and *FT5a* for floral induction under SD conditions (Lu et al., 2017). The recessive allele *j* is a loss-of-function allele that cannot repress *E1* expression, resulting, in turn, in the downregulation of *FT* expression and the delay of flowering under SD conditions (Lu et al., 2017). In addition to the *j* gene, the missense variant of *FT2a* detected in K3 may also be involved in LJ. The LJ in soybean might therefore reflect a slower floral evocation process owing to the depressed function of *FT2a* resulting from a missense or nonsense mutations (Ogiso-Tanaka et al., 2019) or from the repressed transcription of *FT2a* and *FT5a* associated with the upregulation of *E1*, which is released from repression by *J* under SD conditions (Lu et al., 2017). Thus, the LJ trait could result partly from a combination of various dysfunctional alleles at *E9* (*FT2a*) and the loci involved in the PHYA-E1 module, which is an important pathway for photoperiodic flowering in soybean (Cao et al., 2017).

LJ can be considered as a basic difference in the activity of floral induction itself, and the suppression of flowering by higher temperature or longer day lengths generated by R-enriched light sources. Further molecular dissections will facilitate our understanding of the molecular mechanisms underlying the LF habit in soybean and aid in the marker-assisted breeding of soybean cultivars in a wide range of latitudes.

6.4 Binding affinity of E1 family and E2 proteins to DNA fragments

The two major partly-interrelated pathways that control the flowering under LD conditions in soybean, PHYA-E1 pathway and GI-CO pathway, are dependent on the two main soybean maturity genes *E1* and *E2*, respectively (Cao et al., 2017). The proteins of E1 family (E1, E1La and E1Lb) are containing putative binary nucleus localization signal and have domains distantly related to the plant-specific B3 domain,

which regulate flowering directly or indirectly (Xia et al., 2012; Xu et al., 2015). The *E2* gene is an ortholog of *Arabidopsis GI,* which encodes a nuclear-localized membrane protein (Watanabe et al., 2011). It is considered to directly bind to a cis-element of *FT* to activate the expression (Sawa and Kay, 2011).

As mentioned, there was an SNP in the 3′ UTR of the *FT5a* sequences of K3. The SNP created a MYCCONSENSUSAT cis-element (CAGCTG), which is a binding site for the basic-helix-loop-helix transcriptional factor MYC2 (Figure 4.1). To determine the roles of cis-elements in *FT5a* expression, the binding affinity of MYC2, E1 family and E2 proteins was assayed by the yeast-one-hybrid. MYC2 was predicted to bind to this cis-element, but unfortunately, experiments had shown that MYC2 is toxic to yeast, and the binding affinity could not be tested using this method. E2 could not bind to MYCCONSENSUSAT and its variant (CAGCTT) from AGS292. Similarly, E1, E1La and E1Lb proteins could not bind to MYCCONSENSUSAT, but could bind to the variant sequence of AGS292 (Figure 5.1, Figure 5.2). It should be confirmed whether the binding to the variant sequence in the 3′ UTR of *FT5a* by E1 controlled the flowering in AGS292.

Comparison of flowering time showed that the flowering time in AGS292 was earlier than that in AK16 under all conditions (Figure 2.1, Figure3.1~3.2). The expressions of *E1* in both AGS292 and AK16 were very low in all conditions (data not shown) except for high temperature condition in AK16 (Figure 2.8), due to the double recessive genotype *e3*/*e4*. Therefore, *E1* may not be involved in the flowering delay of AK16. However, considering that the expression of *E1* at the 4th trifoliate-leaf stages of AK16 was increased at the 32°C condition (Figure 2.8), it can be assumed that the high expression of *E1* is involved in delayed flowering at high temperatures.

References

Andrés F, Coupland G. The genetic basis of flowering responses to seasonal cues. Nature Reviews Genetics, 2012; 13(9): 627-639.

Balasubramanian S, Sureshkumar S, Lempe J, Weigel D. Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. PLoS Genetics, 2006; 2(7): e106.

Bentley D. Whole-genome re-sequencing. Current Opinion in Genetics & Development, 2006; 16(6): 545-552.

Bernard R. Two major genes for time of flowering and maturity in soybeans 1. Crop Science, 1971; 11(2): 242-244.

Board J, Hall W. Premature flowering in soybean yield reductions at nonoptimal planting dates as influenced by temperature and photoperiod1. Agronomy Journal, 1984, 76(4); 700-704.

Bolger A, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2014; 30(15): 2114-2120.

Bonato E, Vello N. *E6*, a dominant gene conditioning early flowering and maturity in soybeans. Genetics and Molecular Biology, 1999; 22(2): 229-232.

Bouché F, Lobet G, Tocquin P, Périlleux C. FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. Nucleic Acids Research, 2016; 44: Database issue D1167–D1171.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genetics, 2010; 6(5): e1000940.

Bradbury P, Zhang Z, Kroon D, Casstevens T, Ramdoss Y, Buckler E. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 2007; 23(19): 2633-2635.

Broman K, Wu H, Sen Ś, Churchill G. R/qtl: QTL mapping in experimental crosses. Bioinformatics, 2003, 19(7): 889-890.

Browning S, Browning B. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics, 2007; 81(5): 1084-1097.

Buzzell R. Inheritance of a soybean flowering response to fluorescent-daylength conditions. Canadian Journal of Genetics and Cytology, 1971; 13(4): 703-707.

Buzzell R, Voldeng H. Inheritance of insensitivity to long daylength. Soybean Genetics Newsletter, 1980; 7(1): 13.

Cai Y, Chen L, Liu X, Guo C, Sun S, Wu C, Jiang B, Han T, Hou W. CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soya bean. Plant Biotechnology Journal, 2018; 16(1): 176-185.

Calviño M, Kamada H, Mizoguchi T. Is the role of the short-day solely to switch off the CONSTANS in *Arabidopsis*? Plant Biotechnology, 2005; 22(3): 179-183.

Cao D, Takeshima R, Zhao C, Liu B, Abe J, Kong F. Molecular mechanisms of flowering under long days and stem growth habit in soybean. Journal of Experimental Botany, 2017; 68(8): 1873-1884.

Chapman A, Pantalone V, Ustun A, Allen F, Landau-Ellis D, Trigiano R, Gresshoff P. Quantitative trait loci for agronomic and seed quality traits in an F_2 and $F_{4:6}$ soybean population. Euphytica, 2003; 129(3): 387-393.

Chen D, Yan W, Fu L Y, Kaufmann K. Architecture of gene regulatory networks controlling flower development in *Arabidopsis* thaliana. Nature Communications, 2018; 9(1): 4534.

Cheng L, Wang Y, Zhang C, Wu C, Xu J, Zhu H, Zhang L. Genetic analysis and QTL detection of reproductive period and post-flowering photoperiod responses in soybean. Theoretical and Applied Genetics, 2011; 123(3): 421-429.

Cober E. Long juvenile soybean flowering responses under very short photoperiods. Crop Science, 2011; 51(1): 140-145.

Cober E, Molnar S, Charette M, Voldeng H. A new locus for early maturity in soybean. Crop Science, 2010; 50(2): 524-527.

Cober E, Tanner J, Voldeng H. Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. Crop Science, 1996; 36(3): 601-605.

Cober E, Voldeng H. A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. Crop Science, 2001; 41(3): 698-701.

Collinson S, Summerfield R, Ellis R, Roberts E. Durations of the photoperiod-sensitive and photoperiod-insensitive phases of development to flowering in four cultivars of soyabean [*Glycine max* (L.) Merrill]. Annals of Botany, 1993; 71(5): 389-394.

Contreras-Soto R, Mora F, Lazzari F, de Oliveira M, Scapim C, Schuster I. Genome-wide association mapping for flowering and maturity in tropical soybean: implications for breeding strategies. Breeding Science, 2017; 67(4): 435-449.

Copley T, Duceppe M, O'Donoughue L. Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. BMC Genomics, 2018; 19(1): 167.

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Swarle I, Giakountis A, Farrona S, Gissot L, Turnbull C, Coupland G. FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science, 2007; 316(5827): 1030-1033.

del Pilar Valencia-Morales M, Camas-Reyes J, Cabrera-Ponce J, Alvarez-Venegas R. The *Arabidopsis thaliana* SET-domain-containing protein ASHH1/SDG26 interacts

with itself and with distinct histone lysine methyltransferases. Journal of Plant Research, 2012; 125(5): 679-692.

Destro D, Carpentieri-Pípolo V, Kiihl RA de S, Almeida LA de. Photoperiodism and genetic control of the long juvenile period in soybean: A review. Embrapa Soja-Artigo em periódico indexado (ALICE), 2001; 1(1):72-92

Dissanayaka A, Rodriguez T, Di S, Yang F, Githiri S, Rodas F, Takahashi R. Quantitative trait locus mapping of soybean maturity gene *E5*. Breeding Science, 2016; 15160.

Doyle J, Doyle J. Isolation of plant DNA from fresh tissue. Focus, 1990; 12(13): 39-40.

Dubcovsky J, Lijavetzky D, Appendino L, Tranquilli G. Comparative RFLP mapping of Triticum monococcum genes controlling vernalization requirement. Theoretical and Applied Genetics, 1998; 97(5-6): 968-975.

Ellis R, Collinson S, Hudson D, Patefield W. The analysis of reciprocal transfer experiments to estimate the durations of the photoperiod-sensitive and photoperiod-insensitive phases of plant development: an example in soya bean. Annals of Botany, 1992; 70(1): 87-92.

Fornara F, de Montaigu A, Coupland G. SnapShot: control of flowering in *Arabidopsis*. Cell, 2010; 141(3): 550-550. e2.

Funatsuki H, Kawaguchi K, Matsuba S, Sato Y, Ishimoto M. Mapping of QTL associated with chilling tolerance during reproductive growth in soybean. Theoretical and Applied Genetics, 2005; 111(5): 851-861.

Githiri S, Yang D, Khan N, Xu D, Komatsuda T, Takahashi R. QTL analysis of low temperature–induced browning in soybean seed coats. Journal of Heredity, 2007; 98(4): 360-366.

Graham P, Vance C. Legumes: importance and constraints to greater use. Plant Physiology, 2003; 131(3): 872-877.

Hanzawa Y, Money T, Bradley D. A single amino acid converts a repressor to an activator of flowering. Proceedings of the National Academy of Sciences USA, 2005; 102(21): 7748-7753.

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

Ho WWH, Weigel D. Structural features determining flower-promoting activity of *Arabidopsis FLOWERING LOCUS T*. The Plant Cell, 2014; 26(2): 552-564.

Horvath D. Common mechanisms regulate flowering and dormancy. Plant Science, 2009; 177(6), 523-531.

Hu Y, Wang Y, Liu X, Li J. *Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. Cell Research, 2004; 14(1): 8.

Huq E, Tepperman J, Quail P. GIGANTEA is a nuclear protein involved in phytochrome signaling in *Arabidopsis*. Proceedings of the National Academy of Sciences USA, 2000; 97(17): 9789-9794.

Jaeger K, Wigge P. FT protein acts as a long-range signal in *Arabidopsis*. Current Biology, 2007; 17(12): 1050-1054.

Jiang B, Zhang S, Song W, Khan M, Sun S, Zhang C, Wu T, Wu C, Han T. Natural variations of *FT* family genes in soybean varieties covering a wide range of maturity groups. BMC Genomics, 2019; 20(1): 230.

Jiang D, Gu X, He Y. Establishment of the winter-annual growth habit via FRIGIDA-mediated histone methylation at FLOWERING LOCUS C in *Arabidopsis*. The Plant Cell, 2009; 21(6): 1733-1746.

Jiang D, Kong N, Gu X, Li Z, He Y. *Arabidopsis* COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. PLoS Genetics, 2011; 7(3): e1001330.

Jung C, Wong C, Singh M, Bhalla P. Comparative genomic analysis of soybean flowering genes. PloS One, 2012; 7(6): e38250.

Kazan K, Manners JM. MYC2: the master in action. Molecular Plant, 2013; 6(3): 686-703.

Keim P, Diers B, Olson T, Shoemaker R. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics, 1990; 126(3): 735-742.

Khan N, Githiri S, Benitez E, Abe J, Kawasaki S, Hayashi T, Takahashi R. QTL analysis of cleistogamy in soybean. Theoretical and Applied Genetics, 2008; 117(4): 479-487.

Kilen T, Hartwig E. Inheritance of a light-quality sensitive character in soybeans 1. Crop Science, 1971; 11(4): 559-561.

King R, Zeevaart J. Floral stimulus movement in Perilla and flower inhibition caused by noninduced leaves. Plant Physiology, 1973; 51(4): 727-738.

Kobayashi Y, Weigel D. Move on up, it's time for change—mobile signals controlling photoperiod-dependent flowering. Genes & development, 2007; 21(19): 2371-2384.

Komatsu K, Okuda S, Takahashi M, Matsunaga R, Nakazawa Y. Quantitative trait loci mapping of pubescence density and flowering time of insect-resistant soybean (*Glycine max* L. Merr.). Genetics and Molecular Biology, 2007; 30(3): 635-639.

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 Physiology, 2010; 154(3): 1220-1231.

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 Science, 2014; 54(6): 2529-2535.

Lang A, Chailakhyan M, Frolova I. Promotion and inhibition of flower formation in a dayneutral plant in grafts with a short-day plant and a long-day plant. Proceedings of the National Academy of Sciences USA, 1977; 74(6): 2412-2416.

Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods, 2012; 9(4): 357.

Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology, 2009; 10(3): R25.

Lee J, Jung J, Park C. INDUCER OF CBF EXPRESSION 1 integrates cold signals into *FLOWERING LOCUS C*-mediated flowering pathways in *Arabidopsis*. The Plant Journal, 2015; 84(1): 29-40.

Lee S, Bailey M, Mian M, Shipe E, Ashley D, Parrott W, Hussey R, Boerma H. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theoretical and Applied Genetics, 1996; 92(5): 516-523.

Levy Y, Mesnage S, Mylne J, Gendall A, Dean C. Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. Science, 2002; 297(5579): 243-246.

Li X, Fang C, Xu M, Zhang F, Lu S, Nan H, Su T, Li S, Zhao X, Kong L, Yuan X, Liu B, Abe J, Cober ER, Kong F. Quantitative trait locus mapping of soybean maturity gene *E6*. Crop Science, 2017; 57(5): 2547-2554.

Li Y, Zhou G, Ma J, Jiang W, Jin L, Zhang Z, Guo Y, Zhang, Sui Y, Zheng L, Zhang S, Zuo Q, Shi X, Li Y, Zhang W, Hu Y, Kong G, Hong H, Tan B, Song J, Liu Z, Wang Y, Ruan H, Yeung C, Liu J, Wang H, Zhang L, Guan R, Wang K, Li W, Chen S, Chang R, Jiang Z, Scott AJ, Li R, Qiu L. *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nature Biotechnology, 2014; 32(10): 1045.

Liu B, Abe J. QTL mapping for photoperiod insensitivity of a Japanese soybean landrace Sakamotowase. Journal of Heredity, 2009, 101(2): 251-256.

Liu B, Berr A, Chang C, Liu C, Shen W, Ruan Y. Interplay of the histone methyltransferases SDG8 and SDG26 in the regulation of transcription and plant flowering and development. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2016; 1859(4): 581-590.

Liu B, Fujita T, Yan Z, Sakamoto S, Xu D, Abe J. QTL mapping of domestication-related traits in soybean (*Glycine max*). Annals of Botany, 2007; 100(5): 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, 2008; 180(2): 995-1007.

Liu D, Yan Y, Fujita Y, Xu D. A major QTL (*qFT12. 1*) allele from wild soybean delays flowering time. Molecular Breeding, 2018; 38(4): 45.

Liu W, Kim MY, Kang YJ, Van K, Lee YH, Srinives P, Yuan DL, Lee SH. QTL identification of flowering time at three different latitudes reveals homeologous genomic regions that control flowering in soybean. Theor Appl Genet, 2011; 123, 545–553.

Lu S, Li Y, Wang J, Srinives P, Nan H, Cao D, Wang Y, Li J, Li X, Fang C, Shi X, Yuan X, Watanabe S, Feng X, Liu B, Abe J, Kong F. QTL mapping for flowering time in different latitude in soybean. Euphytica, 2015; 206(3): 725-736.

Lu S, Zhao X, Hu Y, S Liu S, Nan H, Li X, Fang C, Cao D, Shi X, Kong L, Su T, Zhang F, Li S, Wang Z, Yuan X, Cober E, Weller J, Liu B, Hou X, Tian Z, Kong F. Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. Nature Genetics, 2017; 49(5): 773.

Ma H. To be, or not to be, a flower—control of floral meristem identity. Trends in Genetics, 1998; 14(1) 26-32.

Mansur L, Lark K, Kross H, Oliveira A. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). Theoretical and Applied Genetics, 1993; 86(8): 907-913.

Mao T, Li J, Wen Z, Wu C, Sun S, Jiang B, Hou W, Li W, Song Q, Wang D, Han T. Association mapping of loci controlling genetic and environmental interaction of soybean flowering time under various photo-thermal conditions. BMC Genomics, 2017; 18(1): 415.

Mathieu J, Warthmann N, Küttner F, Schmid M. Export of FT protein from phloem companion cells is sufficient for floral induction in *Arabidopsis*. Current Biology, 2007; 17, 1055–1060.

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(3): 271-281.

Maurya J, Sethi V, Gangappa S, Gupta N, Chattopadhyay S. Interaction of MYC 2 and GBF 1 results in functional antagonism in blue light-mediated *Arabidopsis* seedling development. The Plant Journal, 2015; 83(3): 439-450.

McBlain B, Bernard R. A new gene affecting the time of flowering and maturity in soybeans. Journal of Heredity, 1987; 78(3), 160-162.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gariel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a

MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 2010; 20(9): 1297-1303.

Molnar S, Rai S, Charette M, Cober E. Simple sequence repeat (SSR) markers linked to *E1*, *E3*, *E4*, and *E7* maturity genes in soybean. Genome, 2003; 46(6), 1024-1036.

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; 9.5: e97669.

Ogiso-Tanaka E, Shimizu T, Hajika M, Kaga A, Ishimoto M. Highly multiplexed AmpliSeq technology identifies novel variation of flowering time-related genes in soybean (*Glycine max*). DNA Research, 2019; 26(3): 243-260.

Orf J, Chase K, Adler F, Mansur L, Lark K. Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Science,1999; 39(6), 1652-1657.

Owen F. Inheritance studies in soybeans. II. Glabrousness, color of pubescence, time of maturity, and linkage relations. Genetics, 1927; 12(6), 519.

Peterson B, Weber J, Kay E, Fisher H, Hoekstra H. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS One, 2012; 7(5): e37135.

Pidkowich M, Klenz J, Haughn G. The making of a flower: control of floral meristem identity in *Arabidopsis*. Trends in Plant Science, 1999; 4(2): 64-70.

Pooprompan P, Wasee S. Toojinda T, Abe J, Chanprame S, Srinives P. Molecular marker analysis of days to flowering in vegetable soybean (*Glycine max* (L.) Merrill). Kasetsart Journal, 2006; 40(3): 573-581.

Qi X, Li M W, Xie M, Liu X, Ni M, Shao G, Song C, Yim A, Tao Y, Wong F, Isobe, S, Wong F, Wong K, Xu C, Li C, Wang Y, Guan R, Sun F, Fan G, Xiao Z, Zhou F,

Phang T, Liu X, Tong S, Chan T, Yiu S, Tabata S, Wang J, Xu X, Lam H. Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nature Communications, 2014; 5(7): 4340.

Ray J, Hinson K, Mankono J, Malo, M. Genetic control of a long-juvenile trait in soybean. Crop Science, 1995; 35(4): 1001-1006.

Rozenzweig V, Aksyonova E, Milash S , Goloenko D, Davydenko O. Prospects of exploiting of photoperiod sensitivity gene *E7* in early soybean breeding and revealing of its sources with SSR-markers. Soybean Genetics Newsletter, 2008; 35.

Saindon G, Voldeng H, Beversdorf W, Buzzell R. Genetic control of long daylength response in soybean. Crop Science, 1989; 29(6): 1436-1439.

Samanfar B, Molnar S, Charette M, Schoenrock A, Dehne F, Golshani A, Belzile F, Cober E. Mapping and identification of a potential candidate gene for a novel maturity locus, *E10*, in soybean. Theoretical and Applied Genetics, 2017; 130(2): 377-390.

Sawa M, Kay S. GIGANTEA directly activates *Flowering Locus T* in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences, 2011; 108(28): 11698-11703.

Sawa M, Nusinow D, Kay S, Imaizumi T. *FKF1* and *GIGANTEA* complex formation is required for day-length measurement in *Arabidopsis*. Science,2007; 318(5848), 261-264.

Schmutz J, Cannon S, Schlueter J, Ma J, Mitros T, Nelson W, Hyten D, Song Q, Thelen J, Cheng J, Xu D, Hellsten U, May G, Yu Y, Sakurai T, Umezawa T, Bhattacharyya M, 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 X, Shinozaki K, Nguyen H, Wing R, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker R, Jackson S. Genome sequence of the palaeopolyploid soybean. Nature, 2010; 463(7278): 178.

Scott W. Modern soybean production. S&A Publications. Inc., Champaign, Illinois, 1983.

Shapter F, Waters D. Genome walking. *Cereal Genomics*. Humana Press, Totowa, NJ, 2014; 133-146.

Shendure J, Mitra R, Varma C, Church G. Advanced sequencing technologies: methods and goals. Nature Reviews Genetics, 2004; 5(5): 335.

Srikanth A, Schmid M. Regulation of flowering time: all roads lead to Rome. Cellular and Molecular Life Sciences, 2011; 68(12), 2013-2037.

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; 6.12: e29238.

Swaminathan K, Peterson K, Jack T. The plant B3 superfamily. Trends in Plant Science, 2008; 13(12): 647-655.

Takeshima R, Hayashi T, Zhu J, Zhao C, Xu M, Yamaguchi N, Sayama T, Ishimoto M, Kong L, Shi X, Liu B, Tian Z, Yamada T, Kong F, Abe J. A soybean quantitative trait locus that promotes flowering under long days is identified as *FT5a*, a *FLOWERING LOCUS T* ortholog. Journal of Experimental Botany, 2016; 67(17): 5247-5258.

Tasma I, Lorenzen L, Green D, Shoemaker R. Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. Molecular Breeding, 2001; 8(1), 25-35.

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*, *E*3 and *E4* in soybean. Annals of Botany, 2014; 113:429–41.

Upadhyay A, Ellis R, Summerfield R, Roberts E, Qi A. Characterization of photothermal flowering responses in maturity isolines of soyabean [*Glycine max* (L.) Merrill] cv. Clark. Annals of Botany, 1994; 74(1): 87-96.

Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science, 2004; 303(5660), 1003-1006.

Wang D, Graef G, Procopiuk A, Diers B. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theoretical and Applied Genetics, 2004; 108(3), 458-467.

Wang F, Nan H, Chen L, Fang C, Zhang H, Su T, Li S, Cheng Q, Dong L, Kong F. A new dominant locus, *E11*, controls early flowering time and maturity in soybean. Molecular Breeding, 2019; 39(5), 70.

Wang H, Ding C, Du H, Liu H, Wang Y, Yu D. Identification of soybean *MYC2-like* transcription factors and overexpression of *GmMYC1* could stimulate defense mechanism against common cutworm in transgenic tobacco. Biotechnology Letters, 2014; 36(9), 1881-1892.

Wang H, Li Y, Pan J, Lou D, Hu Y, Yu D. The bHLH transcription factors *MYC2*, *MYC3*, and *MYC4* are required for jasmonate-mediated inhibition of flowering in *Arabidopsis*. Molecular Plant, 2017; 10(11): 1461-1464.

Wang J, Li H, Zhang L, Bilyeu K, Song J, Lee J. User's manual of QTL IciMapping ver. 4.1. Quantitative Genetics group, Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS): Beijing/Genetic Resources Program, International Maize and Wheat Improvement Center (CIMMYT): Mexico City, 2016.

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*. The Plant Cell, 2015; 27(2): 323-336.

Watanabe S, Harada K, Abe J. Genetic and molecular bases of photoperiod responses of flowering in soybean. Breeding Science, 2012; 61(5): 531-543.

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(4): 1251-1262.

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. Breeding Science, 2004; 54(4), 399-407.

Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K, 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(2): 395-407.

Wils C, Kaufmann K. Gene-regulatory networks controlling inflorescence and flower development in *Arabidopsis thaliana*. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2017; 1860(1), 95-105.

Wu F, Fan C, Zhang X, Fu Y. The phytochrome gene family in soybean and a dominant negative effect of a soybean *PHYA* transgene on endogenous Arabidopsis PHYA. Plant Cell Reports, 2013; 32(12): 1879-1890.

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(1), e85754.

Wu F, Zhang X, Li D, Fu Y. Ectopic expression reveals a conserved PHYB homolog in soybean. PloS One, 2011; 6(11), e27737.

Wu J, Wang Y, Wu S. Two new clock proteins, *LWD1* and *LWD2*, regulate *Arabidopsis* photoperiodic flowering. Plant Physiology, 2008; 148(2), 948-959.

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. Proceedings of the National Academy of Sciences, 2012; 109(32): E2155-E2164.

Xie M, Chung C, Li M, Wong F, Wang X, Liu A, Wang Z, Leung A, Wong T, Tong S, Xiao Z, Fan K, Ng M, Qi X, Yang L, Deng T, He L, Chen L, Fu A, Ding Q, He J, Chung G, Isobe S, Tanabata T, Valliyodan B, Nguyen H, Cannon S, Foyer C, Chan T, Lam H. A reference-grade wild soybean genome. Nature Communications, 2019; 10(1): 1216.

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 Biology, 2013; 13(1): 91.

Xu M, Yamagishi N, Zhao C, Takeshima R, Kasai M, Watanabe S, Kanazawa A, Yoshikawa N, Liu B, Yamada T, Abe J. The soybean-specific maturity gene *E1* family of floral repressors controls night-break responses through down-regulation of *FLOWERING LOCUS T* orthologs. Plant Physiology, 2015; 168(4): 1735-1746.

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 Research, 2001, 8(2): 61-72.

Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Osanai T, Tanaka A, Seki M, Shinozaki K, Yokoyama S. Solution structure of the B3 DNA binding domain of the *Arabidopsis* cold-responsive transcription factor RAV1. The Plant Cell, 2004; 16(12): 3448-3459.

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. Proceedings of the National Academy of Sciences, 2003; 100(10), 6263-6268.

Yue Y, Liu N, Jiang B, Li M, Wang H, Jiang Z, Pan H, Xia Q, Ma Q, Han T, Nian H. A single nucleotide deletion in *J* encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. Molecular Plant, 2017; 10(4): 656-658.

Zhang J, Song Q, Cregan P, Nelson R, Wang X, Wu J, Jiang G. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. BMC Genomics, 2015; 16(1): 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. Proceedings of the National Academy of Sciences, 2008; 105(52): 21028-21033.

Zhang W, Wang Y, Luo G, Zhang J, He C, Wu X, Gai J, Chen S. QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theoretical and Applied Genetics, 2004; 108(6): 1131-1139.

Zhao C, Takeshima R, Zhu J, Xu M, Sato M, Watanabe S, Kanazawa A, Liu B, Kong F, Yamada T, Abe J. A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of *FT2a*, a *FLOWERING LOCUS T* ortholog. BMC Plant Biology, 2016; 16(1): 20.

Zhao L, Li M, Xu C, Yang X, Li D, Zhao X, Wang K, Li Y, Zhang X, Liu L, Ding F, Du H, Wang C, Sun J, Li W. Natural variation in *GmGBP1* promoter affects photoperiod control of flowering time and maturity in soybean. The Plant Journal, 2018; 96(1): 147-162.

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 Molecular Biology, 2015; 88(4-5), 343-355.

Zhu J, Takeshima R, Harigai K, Xu M, Kong F, Liu B, Kanazawa A, Yamada T, Abe J. Loss of function of the *E1-Like-b* gene associates with early flowering under long-day conditions in soybean. Frontiers in Plant Science, 2019; 9: 1867.

Summary

The timing of flowering and maturation determines crop adaptability and productivity. Soybean (*Glycine max*) is cultivated across a wide range of latitudes in the world, as food crops and raw materials for vegetable oil processing and livestock industry. The molecular-genetic mechanisms for flowering in soybean have been determined for photoperiodic responses to long days (LDs), but remain only partially determined for the delay of flowering under short-day (SD) conditions, known as long juvenility (Lj), an adaptive trait of cultivars grown in lower latitudes. Consequently, the study on reproductive phenology, in particular, an extended vegetative period under SDs and/or high temperatures, is significant and valuable for the soybean production in lower latitude areas. In this thesis, I studied the late-flowering (LF) habit introduced from a Thai cultivar K3 into a photoperiod-insensitive genetic background under different photo-thermal conditions and analyzed the molecular and genetic basis using quantitative trait locus (QTL) mapping, expression and resequencing analyses. The results obtained are summarized as follows.

In Chapter 2, I first compared flowering time under different photo-thermal conditions between photoperiod insensitive lines, AGS292 and AK16. AK16 was the latest-flowering line selected from the photoperiod-insensitive recombinant inbred lines derived from the cross between AGS292 and K3. The flowering time of AK16 was delayed by 5.6 days in 25℃/16-h condition to 11.7 days in 25℃/20-h Red-light enriched LD condition and 15.6 days in 32℃/16-h condition, compared with the flowering time of AGS292. The flowering of AK16 was thus inhibited by Red-light enriched LD and high ambient temperature. In addition, I performed the expression analysis for *FLOWERING LOCUS T* orthologs, *FT2a* and *FT5a*, and repressor genes for flowering. *FT2a* and *FT5a* were strongly down-regulated in R-LD and high temperature (32℃) conditions. The results obtained suggest that the LF habit of K3

consisted of a basic difference in the floral induction activity and the suppression of flowering, which was caused by red light-enriched LDs and higher temperatures.

In Chapter 3, I carried out QTL analysis for the LF habit using F_6 progeny of a cross between AGS292 and AK16. The linkage map was constructed with gene-specific markers for flowering genes *E2*, *FT2a* and *FT5a* and 829 single nucleotide polymorphisms obtained from restriction-site associated DNA sequencing. I identified three QTLs (*qDTF-10*, *qDTF-16-1* and *qDTF-16-2*) controlling the LF habit near the gene-specific markers, *E2*, *FT5a* and *FT2a*, respectively. Among them, *qDTF-16-2* had the largest and stable effect under all the conditions tested. The effect of *qDFT-10* was dependent on the environments and detected only in the field conditions.

In Chapter 4, I analyze the sequence polymorphisms of the candidate genes for the QTLs and the photoperiod response of AK16 to the red-light enriched LD condition, using whole-genome resequencing data on AGS292 and K3. I detected a nonsynonymous substitution in exon 4 of *FT2a* from K3, which converted the glycine conserved in FT-like proteins to the aspartic acid, suggesting a functional depression in the FT2a protein from K3. I also detected an SNP at 3' UTR of *FT5a* gene between AGS292 and K3, which created a cis-element of MYCCONSENSUSAT (CAGCTG) in K3.

In Chapter 5, I conducted the yeast-one-hybrid assay to confirm whether the cis-element of MYCCONSENSUSAT in K3 has a function in the control of flowering. MYC2 was predicted to bind to this cis-element, but experiments had shown that MYC2 was toxic to yeast and could not be tested for binding affinity using this method. E2 could not bind to MYCCONSENSUSAT and its variant (CAGCTT) from AGS292. Similarly, E1, E1La and E1Lb proteins could not bind to MYCCONSENSUSAT but they could bind to the variant sequence. The binding

affinity of E1 family proteins was then tested in various DNA sequences, which were different in a single nucleotide from the 6-bp AGS292 sequence (CAGCTT). All the three proteins could bind to four sequences (AAGCTT, CATCTT, CACCTT and CAGATT), but could not bind to the RAV1-B element, a binding site of B3 domain of RVA1, which is a member of the B3 superfamily in *Arabidopsis*. The sequences bound by E1 family proteins were distributed widely in the promoter, coding and downstream regions of *FT2a* and *FT5a*.

Based on the results obtained, I discussed on the characteristics and QTLs for LF habit from K3 in a photoperiod-insensitive genetic background, and the molecular mechanisms underlying the LJ trait in soybean. The functional depression in the FT2a protein caused by a nonsynonymous substitution in exon 4 from K3, and the down-regulation of two *FT* genes by red light-enriched LD conditions and high temperatures may be major factors conferring the LF habit of K3. In addition to the *j* gene, which has been reported to be related to LJ, the missense variant of *FT2a* detected in K3 may also be involved in LJ. The LJ in soybean might therefore reflect a slower floral evocation process owing to the depressed function of *FT2a* resulting from missense or nonsense mutations or from the repressed transcription of *FT2a* and *FT5a* associated with the upregulation of *E1*, which is released from repression by *J* under SD conditions.

Appendix

Supplemental Table 1. Production and Yield information of major soybean growing areas in 2017

Area Code	Area	Production(tonnes)	Yield(hg/ha)
3	Albania	500	22026
7	Angola	16081	5617
9	Argentina	54971626	31711
10	Australia	64546	21490
11	Austria	193416	30002
16	Bangladesh	96921	15416
18	Bhutan	204	5799
19	Bolivia (Plurinational State of)	3018872	23887
21	Brazil	114599168	33769
23	Belize	13037	35388
27	Bulgaria	20000	17346
28	Myanmar	209470	14990
29	Burundi	3185	8437
32	Cameroon	27022	13973
33	Canada	7716600	29310
38	Sri Lanka	14363	17272
41	China, mainland	13149485	17910
44	Colombia	75685	27553
52	Azerbaijan	45	4446
53	Benin	160000	10323
58	Ecuador	35006	12493
59	Egypt	45000	30000
60	El Salvador	5710	18839
68	France	412000	29220
73	Georgia	1812	33632
74	Gabon	4037	8774
79	Germany	61000	32105
80	Bosnia and Herzegovina	11740	15637
84	Greece	3300	22603
89	Guatemala	38000	25333
95	Honduras	1270	21332
97	Hungary	162075	209752
98	Croatia	207765	24405
100	India	10981000	10359
101	Indonesia	542000	15182
102	Iran (Islamic Republic of)	200000	24096
103	Iraq	26	7523
106	Italy	1019781	31629

107	Côte d'Ivoire	503	14393
108	Kazakhstan	252319	20106
110	Japan	253000	16844
113	Kyrgyzstan	3512	20526
114	Kenya	2012	8689
115	Cambodia	168000	16154
116	Democratic People's Republic of Korea	223325	14908
117	Republic of Korea	68992	16890
120	Lao People's Democratic Republic	7960	18685
123	Liberia	3356	4307
129	Madagascar	43	5711
130	Malawi	208556	11051
133	Mali	7875	4632
138	Mexico	432927	16486
143	Morocco	679	9658
146	Republic of Moldova	47000	13824
149	Nepal	29061	12333
154	North Macedonia	34	12593
157	Nicaragua	5693	17239
159	Nigeria	730000	9733
165	Pakistan	159	5162
166	Panama	110	3871
167	Czechia	37012	24121
169	Paraguay	10478000	31000
170	Peru	1581	21279
171	Philippines	666	12970
173	Poland	20297	21748
176	Timor-Leste	1033	12995
181	Zimbabwe	60000	10526
183	Romania	416370	25213
184	Rwanda	23934	4485
185	Russian Federation	3621344	14073
198	Slovenia	7713	26523
199	Slovakia	102441	23335
202	South Africa	1316000	22929
203	Spain	4599	27181
207	Suriname	6	10522
208	Tajikistan	23	3933
211	Switzerland	5642	33286
212	Syrian Arab Republic	300	23077
214	China, Taiwan Province of	3203	16087
215	United Republic of Tanzania	5741	10295

216	Thailand	54000	17419
217	Togo	2154	4429
223	Turkey	140000	44206
226	Uganda	30689	6864
230	Ukraine	3899370	19675
231	United States of America	119518490	32990
233	Burkina Faso	18500	9250
234	Uruguay	1316000	11899
236	Venezuela (Bolivarian Republic of)	8000	10000
237	Viet Nam	101856	14980
238	Ethiopia	84033	21535
250	Democratic Republic of the Congo	25000	5102
251	Zambia	351416	15594
272	Serbia	461272	22868
351	China	13152688	17910

Supplemental Table2.1. Primer sequences for expression analyses

Primer name		Primer sequence (5´ - 3´)	
ACT2/7	F	AATTCACGAGACCACCTACAAC	TC204150
	R	TGAGCCACCACTAAGAACAATG	(Glyma.06G150100/Glyma.04G215900)
FT2a	F	GGATTGCCAGTTGCTGCTGT	Glyma.16G150700
	R	GAGTGTGGGAGATTGCCAAT	
FT5a	F	GATTCTTGGATGCCTTACTCC	Glyma.16G044100
	R	GGCATGCTCTAGCATTGCAA	
E2	F	GTTGCAACCCCACTACAGCCT	Glyma.10G221500
	R	GAGGCAGAGCCAAAGCCTATG	
TOE4a	F	ATGAGAGCGAGATGAAACCT	Glyma.15G044400
	R	GGGAAAGAAAGTGGAATATAC	
MYCLa	F	CGCCAGTGAAGACAGTTTTCCAGA	Glyma.01G096600
	R	CGTACACGCCGATTAAATCGCTGA	
MYCLb	F	GGTAGCAGTAGCTGTAGCTTTCT	Glyma.08G271900
	R	CCCCTGGTTGATCATTCAGCTT	
MYCLd	F	GGGACGCGATGATAACGATCAC	Glyma.09G204500
	R	GCTCGAAGCTGCTCCTGCGTGTAA	
COL2b	F	GAGTTGTACCCGAATCAACAGTAAGTGG	Glyma.18G278100
	R	GAGGCATATCTTATTTTCTTCTCGAA	
VRN1	F	CGTCCCCCAATCATGAAGGA	Glyma.16G046700
	R	GTATGTGGCTGAGCCTCCTCT	

Supplemental Table 3.1 Primer sequences for gene-specific DNA markers

Primer name		Primer sequence (5′ - 3′)	
FT2a	F	GTCTATTTCAACATTCAGAGGAAAT	Glyma.16G150700
	R	GAGTGTGGGAGATTGCCAAT	
FT5a	F	GATTCTTGGATGCCTTACTCC	Glyma.16G044100
	R	GGCATGCTCTAGCATTGCAA	
E2[1)]	F	GAAGCCCATCAGAGGCATGTCTTATT	Glyma.10G221500
	R	GAGGCAGAGCAAAGCCTAT	

1) Tsubokura et al. (2014)

Supplemental Table 4.1 Primer sequences for genome walking sequencing analysis

Primer name		Primer sequence (5′ - 3′)
FT5a-2842up	R	GGAAACTTAGGGGGTGTTTACTTGCACTT
FT5a-3070up	R	GAAACCAAAATTACAAGTAAATGCCCTCTT
FT5a-3434up	R	GCG TGC GCA TGT GCG CAT GTT
FT5a-4064up	R	GAGAAGTAGCTATGACCAATCT
FT5a-EX4	F	GCCAAAGAGAGCGTGGTTGCGGTGGAAG
FT5a-296down	F	GCTAGAGCATGCCATTCAGGAACTTCTTCA
FT5a-885down	F	GCATATGGATTTGATTTCCA
FT5a-CDS[1)]	F	TAGAACCCCAAACACAAACA
	R	CAGCTGGAGTAAGGCATCCAA
FT5a-Pro1[1)]	F	GCAAGTTATAAGAGGGCATT
	R	TTGGGCACTGATTGATTGTG
FT5a-Pro2[1)]	F	CACAATCAATCAGTGCCCAA
	R	ATCGTGATCGGACGACAGAC
FT5a-Pro3[1)]	F	GTCTGTCGTCCGATCACGAT
	R	TGTTTGTGTTTGGGGTTCTAC
FT5a-3′UTR[1)]	F	ATTAATATTGATGCACAGGG
	R	TCAAACGTACATACGTTTCTCTG

1) Takeshima et al. (2016)

Supplemental Table 5.1 Sequences of primers used in construction of recombinant bait plasmids

Construction of recombinant bait plasmids		Primer sequence (5´ - 3´)
E1	F	CCAGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAGCTTC
	R	TCGAGAAGCTGAAGCTGAAGCTGAAGCTGAAGCTGAAGCTGGAGCT
K3	F	CCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGC
	R	TCGAGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGGAGCT
RAV1-B	F	CCACCTGCACCTGCACCTGCACCTGCACCTGCACCTGC
	R	TCGAGCAGGTGCAGGTGCAGGTGCAGGTGCAGGTGCAGGTGGAGCT
V1	F	CCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC
	R	TCGAGAATCTGAATCTGAATCTGAATCTGAATCTGAATCTG GAGCT
V2	F	CAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTC
	R	TCGAGAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTGAGCT
V3	F	CCAACTTCAACTTCAACTTCAACTTCAACTTCAACTTC
	R	TCGAGAAGTTGAAGTTGAAGTTGAAGTTGAAGTTGAAGTTGGAGCT
V4	F	CCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTC
	R	TCGAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGGAGCT
V5	F	CCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTC
	R	TCGAGAAACTGAAACTGAAACTGAAACTGAAACTGAAACTGGAGCT
V6	F	CCACCTTCACCTTCACCTTCACCTTCACCTTCACCTTC
	R	TCGAGAAGGTGAAGGTGAAGGTGAAGGTGAAGGTGAAGGTGGAGCT
Colony PCR Primer		Primer sequence (5´ - 3´)
E1	F	GTTCCTTATATGTAGCTTTCGACAT
	R	AGCTGAAGCTGAAGCTGGAGCT
K3	F	GTTCCTTATATGTAGCTTTCGACAT
	R	CAGCTGCAGCTGCAGCTGGAGCT
RAV1-B	F	GTTCCTTATATGTAGCTTTCGACAT
	R	TCGAGCAGGTGCAGGTGCAGGTGCA
V1	F	GTTCCTTATATGTAGCTTTCGACAT
	R	TCGAGAATCTGAATCTGAATCTGAA
V2	F	GTTCCTTATATGTAGCTTTCGACAT
	R	TCGAGAAGCTTAAGCTTAAGCTTAA
V3	F	GTTCCTTATATGTAGCTTTCGACAT
	R	TCGAGAAGTTGAAGTTGAAGT
V4	F	GTTCCTTATATGTAGCTTTCGACAT
	R	TCGAGAAGATGAAGATGAAGA
V5	F	GTTCCTTATATGTAGCTTTCGACAT
	R	TCGAGAAACTGAAACTGAAAC
V6	F	GTTCCTTATATGTAGCTTTCGACAT
	R	TCGAGAAGGTGAAGGTGAAGG

Supplemental Table 5.2 Sequences of primers used in construction recombinant prey plasmids

Full-length cloning primer		Primer sequence (5´ - 3´)
MYC2b step1	F	CCCCACTCAATCGCTCATAACCTT
	R	CTGCAGGGATCCCTATCGTAGTTCATC
MYC2b step2	F	GATACGCATATGATGACCGAGTACCGG
	R	CTGCAGGGATCCCTATCGTAGTTCATC

Acknowledgements

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

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

I also appreciate Professor Baohui Liu and Professor Fanjiang Kong (School of Life Science, Guangzhou University; The Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences) for their constructive suggestions and cooperation to my study.

I am grateful to all members of the Plant Genetics and Evolution laboratory. Special thanks to Dr. Meilan Xu and Dr. Dwiyanti for instructing me the yeast one-hybrid assay and sequence analysis. I am also greatly indebted to Dr. Zhu (Crop Biotechnology Research Unit, Division of Applied Genetics, Institute of Agrobiological Science, NARO), Dr. Li (Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences) and Miss. Zhang for their advice and experimental help.

Finally, I would like to express my heartfelt gratitude China Scholarship Council (CSC) and the Japanese Government Scholarship (MEXT) for their encouragement during my study life in Japan. And I want to thank Hokkaido University leaving me a wonderful and memorable study abroad experience.