博 士 論 文 の 要 約

博士の専攻分野の名称： 博 士（農学） 氏名 孫 菲

学 位 論 文 題 名

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)

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 of 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.

Diverse genotypic combinations at several flowering loci enable crops to maximize their productivity under various environmental conditions. The phytochrome A (PHYA)-E1 module is a main regulator in the photoperiodic flowering of soybean. 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 PHYA proteins encoded by *E3* and *E4*. 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. The LJ period appears to be controlled by

several recessive genes, including two major genes *j* and *e6*. 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.

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 RILs 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 the lowered floral induction activity and the suppression of flowering that was caused by red light-enriched LDs and higher temperatures.

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. QTL mapping with genome-wide markers has provided a way to explore the genetic system for quantitative traits. Until now, genome-wide and flowering gene-specific association analyses have detected associations of DNA polymorphisms in the orthologs of *Arabidopsis* flowering genes with flowering time. 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 Chapter 3, I carried out QTL analysis for the LF habit using F_6 progeny of a cross between AGS292 and AK16. 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. The flowering times of the F6 families correlated positively between the SD and R-LD conditions, and between the ND and FR-LD conditions. 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. 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

Genetic polymorphisms contribute to variations in phenotypes' response to environments. 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 have significantly improved throughput and dramatically reduced the cost as compared to capillary-based electrophoresis systems. 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. 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.

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. By using the genome-walking method, I determined 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).

The SNP at 3' UTR of *FT5a* gene between AGS292 and K3 detected in Chapter 4, 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. The expression of *MYC*-like genes is dependent on different temperature conditions. Moreover, the expression levels of *MYC* genes in 4th leaf stage at 25℃ were upregulated in AK16.

The E1 protein contains a putative binary nucleus localization signal and has a domain distantly related to the plant specific B3 domain. Several members of the B3 superfamily regulate flowering directly or indirectly. In soybean, the E1 and E1-like proteins regulate flowering time through down-regulation of *FT2a* and *FT5a*. However, how the 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*, which up-regulate the expression of CONSTANS through degradation of CYCLING DOF FACTOR and activate *FT* expression. It is also considered that GI directly binds to a cis-element of *FT* to activate the expression. Accordingly, it is also worth to explore whether the *E2* gene in soybean directly binds to cis-elements of *FT2a* and *FT5a* or not.

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 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 but the juvenile phases of 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 with 8-h to 16-h photoperiods at a constant 25℃ relative to the flowering time of AGS292 and the delay increased with R-enriched LDs of 20 h and 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. It did not respond to FR-LD conditions, like AGS292, and the flowering time was almost the same under ND and FR-LD conditions. However, AK16 retained sensitivity to photoperiods supplied by R-enriched lights. This was an unexpected result because the response of flowering to R-LD is controlled only by the *E3* gene. 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 and/or the blue light photoreceptor CRYPTOCHROME 2. The PHYA4 protein of Williams 82 is considered to be defective in function compared to the wild type. 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 amino acid. *PHYA1* showed no nonsynonymous substitution, but other four *PHY* genes showed several nonsynonymous substitutions between AGS292 and K3. 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. 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. It is intriguing to investigate whether these PHY genes are responsible for the residual

response of AK16 to R-light enriched LD condition.

In AK16, flowering was also inhibited by hot temperatures. Compared with the constant 25℃ condition, constant or daytime 32℃ conditions promoted flowering in AGS292, but adversely inhibited it in AK16. In contrast to the photoperiodic regulation, however, little is known about the genetic variability in the thermal regulation of flowering in soybean. Cover et al. (2001) found that a hot temperature (28℃) markedly delayed the flowering, with the effect being more pronounced in late maturing photoperiod-sensitive genotypes carrying two or more dominant alleles at the *E1*, *E3*, *E4* and *E7* loci. Whether the inhibition of flowering by hot temperatures observed in AK16 is a common characteristic of the LF habit of LJ cultivars 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.

Three QTLs (*qDTF-10*, *qDTF-16-1* and *qDTF-16-2*) identified in this study controlling the LF habit that originated from the Thai cultivar K3 in a photoperiod-insensitive genetic background. *qDTF-10*, *qDTF-16-1* and *qDTF-16-2* were co-localized with gene-specific DNA markers for *E2*, *FT5a* and *FT2a*, respectively. 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. 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. 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.

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. Intriguingly, however, the E1 family proteins could bind to the variant sequence of AGS292. It should be therefore confirmed whether the binding to the variant sequence in the 3' UTR of *FT5a* by E1 controlled the

flowering in AGS292. The *E1* expressions in both AGS292 and AK16 were very low in all conditions except for high temperature condition in AK16, due to the double recessive genotype *e3*/*e4*. However, the expression of *E1* was increased in AK16 at the 32°C condition. Therefore, *E1* may be involved in the flowering delay of AK16 in hot temperature conditions.

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.