Abstract

Division: Agrobiology Name: Jianghui Zhu

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

Soybean (*Glycine max* (L.) Merr.) is a typical short-day plant. Eleven major genes for flowering have been reported so far. Among them, four maturity genes, *E1* to *E4*, are the main contributors to soybean adaptation to a wide range of latitudes

The floral repressor *E1* is a possible transcription factor that represses the expression of major soybean *FT* orthologues *FT2a* and *FT5a*. *E1* expression is up-regulated under long day (LD) conditions under the control of the phytochrome A (phyA) proteins E3 and E4. *E1* has two homologues, *E1-Like-a* (*E1La*) and *E1Lb*. Down-regulation of the *E1L* genes by virus-induced gene silencing (VIGS) in a cultivar deficient in the *E1* gene leads to early flowering and abolishes the night-break response, suggesting that the two *E1L* genes are also involved in the photoperiod responses of soybean.

Photoperiod-insensitivity in soybean is conditioned by combinations of various alleles at *E1*, *E3*, and *E4*. On the basis of the functions of alleles at the three loci, Xu et al. (2013) classified ILD-insensitive cultivars into three genotypic groups: (group 1) the dysfunction of both *E3* and *E4*; (group 2) the dysfunction of *E1* in combination with that of either *E3* or *E4*; and (group 3) a combination of *e1-as* (hypomorphic allele), *e3*, and *E4*. Because *E4* inhibits flowering under ILD conditions, the group 3 cultivars are predicted to have novel genes that abolish or reduce ILD-sensitivity. One such gene is an early-flowering allele (*FT5a-ef*) at *qDTF-J*, a QTL for days to flowering in linkage group J, which encodes FT5a; early flowering is caused by its increased transcriptional activity or mRNA stability associated with an insertion in the promoter and/or deletions in the 3′-untranslated region (UTR).

In the thesis, I studied a molecular-genetic mechanism of photoperiod-insensitivity in group 3 soybean cultivars introduced from Far-Eastern Russia. Firstly, I identified a novel loss-of-function allele at the *E1Lb* locus as a most likely causal factor for the photoperiod-insensitivity, by the genetic analysis including the association test, fine-mapping and sequence analysis. Secondly, I characterized the function and roles of the loss-of-function allele on flowering under LD conditions, by using near-isogenic lines. I also determined the interaction between the E1 family proteins, E1, E1La and E1Lb, by yeast two-hybrid assay. At last, I surveyed the molecular diversity for the *E1La* and *E1Lb* genes.

1. Identification of a novel gene for photoperiod-insensitivity

1.1 Materials

The Far Eastern Russian soybean cultivars Zeika (ZE), Yubileinaya (YU), and Sonata were crossed with the Canadian cultivar Harosoy (L58-266; HA); ZE and YU were also crossed with a Harosoy near-isogenic line for *e3* (PI547716; H-*e3*). The three Russian cultivars have the same genotype as H-*e3* at five maturity loci, *E1*, *E2*, *E3*, *E4*, and *E9* (*e1-as*/*e2*/*e3*/*E4*/*E9*), but unlike H-*e3* they flower without any marked delay under ILD conditions in comparison with natural daylength (ND) conditions (maximum daylength, 15.2 h) in Sapporo, Japan (43°07′N, 141°35′E).

1.2 Results

Flowering time under the ILD condition in F_2 populations of the H-*e3* × ZE and H-*e3* × YU crosses varied continuously from that of ILD-insensitive parents (45 DAS for ZE and 46 DAS for YU) to the end of light supplementation; 48 H-*e3* × ZE F_2 plants were randomly selected and the progeny for flowering time segregation under the ILD condition were tested. Based on the segregation pattern, the 48 F_2 plants could be classified into three groups: (1) plants fixed for ILD insensitivity (all F_3 plants tested flowered as ZE did; *e*/*e*); (2) those segregating for flowering time (*E*/*e*) and (3) those fixed for ILD sensitivity (all F_3 plants tested showed delayed or no flowering; *E*/*E*). The number of plants was 8 in *e*/*e*, 23 in *E*/*e*, and 17 in *E*/*E*, in consistence with a monogenic 1:2:1 ratio (χ^2 = 3.81, *df* = 2, *p* = 0.18), suggesting the involvement of a single recessive gene for ILD insensitivity.

Based on the results of the progeny test, 16 F_2 plants from the H-*e3* × ZE cross were selected, 8 homozygous for ILD insensitivity (*e*/*e*), and 8 homozygous for ILD sensitivity (*E*/*E*). Among the SSR markers tested, Satt190 and Sat_085 in linkage group C1 (chromosome 4; Chr04) showed genotypic variation in complete accordance with the ILD sensitivity. Then the genotypes of the two markers in the whole F_2 plants of H-*e3* × ZE and H-*e3* × YU populations were determined. The two markers were tightly linked to each other with a recombination value of 2.1, and were closely associated with ILD sensitivity. Satt190 and Sat_085 are located 17.3 Mb from each other in the pericentromeric region of chromosome 4.

To delimit the genomic region of the gene for ILD-insensitivity more precisely, plants with recombination between the two markers (7 from 306 F_2 plants from the H-*e3* × ZE and H-*e3* × YU crosses and 3 from 492 F_3 plants from the H-*e3* × ZE cross) were selected and constructed their graphical genotypes with 11 SSR markers. A comparison of the graphical genotypes with the genotype of ILD-insensitivity estimated by the progeny test revealed that the gene for ILD-insensitivity was located between SSR markers BARC-18g-0889 and BARC-18g-0895. The physical distance between the two markers was 842 kb, and the delimited region contained only 6 annotated genes. RNA-sequencing Atlas in Phytozome v12.1/*Glycine max Wm82.a2. v1* indicates that Glyma.04G143000, Glyma.04G143100 and Glyma.04G143200 are expressed

only in flower or root tissues, whereas Glyma.04G143300, Glyma.04G143400 and Glyma.04G143500 are expressed in leaves. I focused on the three genes expressed in leaves as a possible candidate of the gene for ILD-insensitivity that upregulates the two *FT* genes.

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

2. Characterization of the *e1lb* allele in the control of flowering

2.1 Materials

Four sets of NILs, each including one NIL for ILD insensitivity and another for sensitivity, were developed from heterozygous inbred F_5 plants derived from different F_2 plants from the H-*e3* × ZE cross and those from the HA × ZE cross. The former two sets of NILs had the recessive *e3* allele, whereas the latter two had the dominant *E3* allele. These lines, together with parents and an ILD-insensitive NIL of HA for *e3* and *e4* (PI546043; H-*e3e4*), were cultivated in a growth chamber (25°C, 20-h daylength) with an average photon flux of 120 μmol m^{-2} s^{-1} and an R:FR ratio of 2.2 at 1 m below light sources, or in the field under the ILD condition. For comparison, NILs for alleles, *e1-nl* and *e1-as*, at *E1* (NIL-*E1*; *e2*/*E3*/*E4*/*E9*), which were developed from a heterozygous inbred F_5 plant derived from a cross between the Japanese cultivar Toyomusume (*e1-nl*/*e2*/*E3*/*E4*/*e9*) and HA, were included in the evaluation of flowering under the ILD condition.

2.2 Results

2.2.1 Comparison of flowering time and gene expression among NILs

The allelic effects of *E1Lb* and *e1lb* on flowering under the R-enriched LD condition (daylength, 20 h) were evaluated in four sets of NILs. In the two sets of the *e3*/*E4* NILs, each NIL for *e1lb* flowered at the same or almost the same time as ZE; this was on average 6.7 to 7.6 days earlier than the respective NILs for *E1Lb*, which flowered at almost the same time as H-*e3*. Flowering times of the *E3*/*E4* NILs were around 20 days or more later than those of the *e3*/*E4* NILs. *e1lb* also promoted flowering in the *E3*/*E4* background: each NIL for *e1lb* flowered around 10 days

earlier than the respective NIL for *E1Lb* and HA. This flowering-promoting effect of *e1lb* versus *E1Lb* under the R-enriched LD condition was smaller than that of *e4* vs. *E4* and that of *e3* vs. *E3*, because H-*e3e4* and H-*e3* flowered, on average, 13 and 25 days earlier than H-*e3* and HA (*E3E4*), respectively.

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

The expression levels of *E1*, two *E1L* genes, and two *FT* orthologues were tested in the *e3*/*E4* NILs grown under the R-enriched LD condition. The expression levels of *E1* and *E1La* were similar between the NILs for *E1lb* and *e1lb*. While, the expression of *E1Lb* was significantly down-regulated in the NILs for *e1lb*. The extremely low expressions of *E1Lb* gene in NILs for *e1lb* were considered to be due to nonsense-mediated mRNA decay or different promoter activity. In contrast, the expressions of both *FT2a* and *FT5a* were up-regulated in the NILs for *e1lb* relative to those for *E1Lb* in both NIL sets. The similar effect of *e1lb* vs. *E1Lb* on the expression of *FT2a* and *FT5a* was also observed in both sets of *E3*/*E4* NILs. As observed in the *e3*/*E4* NILs for *e1lb*, the expression levels of *FT2a* and *FT5a* were significantly upregulated in the *E3*/*E4* NIL for *e1lb*.

2.2.2 Interaction between the E1 family proteins

A total of six combinations between the E1 family proteins were tested. All of the yeasts except positive control died in the selection medium, indicating that the three E1 family proteins did not interact physically with one another and themselves. These results suggest that the *E1* family genes control the *FT2a* and *FT5a* expressions independently of one another.

3. Molecular diversity of *E1Lb* and its homologue *E1La*

3.1 Materials

A total of 59 ILD-insensitive accessions were surveyed for the *E1Lb* genotype using the allele-specific DNA marker. They included 9 accessions from northern Japan, 26 from North-Eastern China, 13 from Far Eastern Russia, 8 from Ukraine, and 3 from Poland. Fourteen ILD-insensitive accessions, together with H-*e3* and Zeika, and eight wild soybean (*Glycine*

soja) accessions introduced from the Far Eastern Russia were surveyed for sequence variation of *E1La* and *E1Lb*.

3.2 Results

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

To determine whether or not the deletion in the *E1Lb* gene is region-specific, I surveyed the polymorphism in the ILD-insensitive soybean accessions by using dCAPS marker. In addition to the three Russian cultivars used in the genetic analysis, another two Russian cultivars, Salyut 216 and DYA-1, were found to have the *e1lb* allele, whereas all the other accessions had the functional *E1Lb* allele. I also examined the *E1Lb* genotype for 8 wild accessions introduced from the Far Eastern Russia, because the Russian cultivar Zeika was developed from the crossing between cultivated and wild soybeans. I expected that the *e1lb* allele had been introduced from the wild soybean in Russia. However, there was no wild accession with the *e1lb* allele.

3.2.2 Sequence polymorphisms in *E1La*

Eight wild soybean accessions and fourteen ILD-insensitive accessions were surveyed for sequence diversity of *E1Lb*. However, no novel polymorphism was detected.

3.2.3 Sequence polymorphisms in *E1La*

Fourteen ILD-insensitive accessions, together with H-*e3* and Zeika, and 8 wild soybean accessions were surveyed for sequence diversity of *E1La*. Besides the reference sequence of Williams 82 (designated haplotype 1; Hap 1), one haplotype (haplotype 2; Hap 2) in wild soybean accessions, two haplotypes (haplotype 3 and 4; Hap 3 and Hap 4) in cultivated accessions were detected. Compared with Hap 1, Hap 2 contained three SNPs, synonymous and non-synonymous nucleotide substitutions in exon and a SNP in 3'-UTR. Hap 3 contained the same SNPs as Hap 2 at the synonymous nucleotide substitution and the nucleotide substitution in 3 '-UTR. Hap 4 contained the SNP in 3'-UTR as Haplotypes 2 and 3 did.

To confirm whether the non-synonymous substitution in Hap 2 affects the protein structure of E1La, the putative 3D protein structures were constructed with the I-TASSER program. The non-synonymous substitution located in the B3-DNA binding domain of E1La protein. The structure analysis for the wild type and variant type E1La proteins indicated that the nonsynonymous substitution changed the pattern of hydrogen bonds between neighboring amino acids. The change of hydrogen bonding pattern was predicted to result in the structure change in the variant protein. These results suggested that the amino acid substitution detected in E1La Hap2 produced an adverse effect on the DNA-binding ability of E1La to the *cis*-elements in *FT2a* and *FT5a*. A further study is needed to determine the effect of this non-synonymous substitution on flowering in segregating populations.

In this study, I detected a novel loss-of-function allele that resulted from a frameshift mutation at the *E1Lb* locus in the Far-Eastern Russian photoperiod-insensitive cultivars. Genotyping with an allele-specific DNA marker revealed that *e1lb* is a rare and region-specific allele even in early maturing photoperiod-insensitive cultivars. Only Russian cultivars possessed *e1lb.* The *e1lb* allele most likely has neither largely contributed to the diversity of flowering behaviors nor been used widely in current soybean breeding. The *e1lb* allele may thus be useful as a new resource to broaden the genetic variability of soybean cultivars for flowering under LD conditions at high latitudes.

I also discovered a nonsynonymous substitution in the B3-like DNA-binding domain of E1La in wild soybean accessions which changes hydrogen bonding patterns between neighboring amino acids, and thus may influence the function of DNA-binding of E1La.

The loss-of-function allele at *E1Lb* and the putative missense variant of *E1La* detected in this study may be useful to better comprehend the network among *E1* family genes in the control of soybean flowering. The missense *E1La* variant can also be used to breed the cultivars toward unexplored regions of higher latitudes for soybean cultivation.