DOI: 10.1534/genetics.108.092742

Genetic Redundancy in Soybean Photoresponses Associated With Duplication of the Phytochrome A Gene

Baohui Liu,* Akira Kanazawa,* Hisakazu Matsumura,† Ryoji Takahashi,‡ Kyuya Harada§ and Jun Abe*,[1]

**Research Faculty of Agriculture, Hokkaido University, Sapporo, Hokkaido 060-8589, Japan, †University of Tsukuba, Tsukuba, Ibaraki 305-8518, Japan, ‡National Institute of Crop Science, Tsukuba, Ibaraki 305-8518, Japan and §National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan*

Manuscript received June 16, 2008
Accepted for publication July 24, 2008

ABSTRACT

Gene and genome duplications underlie the origins of evolutionary novelty in plants. Soybean, *Glycine max*, is considered to be a paleopolyploid species with a complex genome. We found multiple homologs of the *phytochrome A* gene (*phyA*) in the soybean genome and determined the DNA sequences of two paralogs designated *GmphyA1* and *GmphyA2*. Analysis of the *GmphyA2* gene from the lines carrying a recessive allele at a photoperiod insensitivity locus, *E4*, revealed that a *Ty1/copia*-like retrotransposon was inserted in exon 1 of the gene, which resulted in dysfunction of the gene. Mapping studies suggested that GmphyA2 is encoded by *E4*. The *GmphyA1* gene was mapped to a region of linkage group O, which is homeologous to the region harboring *E4* in linkage group I. Plants homozygous for the *e4* allele were etiolated under continuous far red light, but the de-etiolation occurred partially, indicating that the mutation alone did not cause a complete loss of phyA function. The genetic redundancy suggests that the presence of duplicated copies of *phyA* genes accounts for the generation of photoperiod insensitivity, while protecting against the deleterious effects of mutation. Thus, this phenomenon provides a link between gene duplication and establishment of an adaptive response of plants to environments.

GENE duplication is a major evolutionary force for generating new genetic materials under the control of natural and artificial selection. Under the classic model of duplicate gene evolution, one of the duplicated genes is free to accumulate mutations, which results in either transcriptional and/or functional inactivation (pseudogenization) or gain of a new function (neofunctionalization) as long as another copy retains the requisite physiological functions. However, empirical data suggest that a much greater proportion of gene duplicates is preserved than predicted by the classical model (Force *et al.* 1999). Recent advances in genome study have led to the formulation of evolutionary models: a model proposed by Hughes (1994) suggests that the phenomenon of gene sharing, in which a single gene encodes a protein having two distinct functions, precedes the evolution of new proteins; the duplication–degeneration–complementation model suggests that duplicate genes acquire debilitating yet complementary mutations that alter one or more subfunctions of the single gene progenitor, an evolutionary consequence for duplicated loci referred to as subfunctionalization (Force *et al.* 1999; Lynch and Force 2000; reviewed by Moore and Purugganan 2005).

Polyploidization is a well-known mechanism of gene duplication in plants. Approximately 70–80% of angiosperm species have undergone polyploidization at some point in their evolutionary history (Moore and Purugganan 2005). Genome analyses have revealed that even plants with relatively small genomes have been impacted by polyploidy (Adams and Wendel 2005). In *Arabidopsis thaliana*, for example, up to 90% of loci are duplicated (Moore and Purugganan 2005), and at least four whole-genome duplication events appear to have occurred (Vision *et al.* 2000).

Soybean, *Glycine max* (L.) Merr., is considered to be a paleopolyploid species with a complex genome (reviewed by Shoemaker *et al.* 2006). It belongs to the Papilionoideae subfamily Leguminosae, which contains nearly all commercially important crop legumes. The Papilionoideae subfamily falls into two clades, Galegoid and Phaseoloid (Goldblatt 1981; Zhu *et al.* 2005). Soybean is a member of the Phaseoloid clade, together with the common bean (*Phaseolus vulgaris*) and mungbean (*Vigna radiata*), whereas the Galegoid clade includes pea (*Pisum sativum*), alfalfa (*Medicago sativa*), and two model legumes, *Lotus japonicus* and *Medicago truncatula*. Most papilionoids are considered to be cytological

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AB370252 and AB370253 for the genomic sequences of the 130S allele (*E4*) of *GmphyA1* and *GmphyA2*, respectively, and under accession no. AB370254 for the retrotransposon inserted in the genomic sequence of the 130I allele (*e4*) of *GmphyA2*.

[1]*Corresponding author:* Laboratory of Plant Genetics and Evolution, Research Faculty of Agriculture, Hokkaido University, Kitaku, Kita 9 Nishi 9, Sapporo, Hokkaido 060-8589, Japan.
E-mail: jabe@res.agr.hokudai.ac.jp

diploids, and except for soybean, possess a base chromosome number (x) of 6–8 in the Galegoid and 10–11 in the Phaseoloid clade, respectively (Goldblatt 1981). The doubling of the base chromosome number in soybean ($x = 20$) suggested the hypothesis that soybean is a diplodized tetraploid species (Lackey 1980; Hymowitz 2004). This hypothesis was confirmed by extensive studies utilizing DNA hybridization and genetic mapping with restriction fragment length polymorphism (RFLP) markers, which demonstrated that soybean possesses a high level of duplicate sequences, and furthermore, homeologous duplicated regions anchored by RFLP markers are scattered across different linkage groups in a complex manner (Zhu *et al.* 1994; Shoemaker *et al.* 1996; Lohnes *et al.* 1997; Lee *et al.* 1999). On the basis of the genetic distances estimated by synonymous substitution measurements for the pairs of duplicated transcripts from the EST collections of soybean and *M. truncatula*, Schlueter *et al.* (2004) estimated that soybean probably underwent two major genome duplication events: one which took place 15 million years ago (MYA) and another 44 MYA. Soybean may therefore be a suitable crop among the legumes for determining the evolutionary role of gene duplications brought about by paleopolyploidy.

Soybean is cultivated in a wide range of latitudes from equatorial to high-latitude regions of up to 50 or more degrees of latitude. This wide adaptability most likely has been created by genetic diversity at a large number of the major genes and quantitative trait loci (QTL) controlling flowering behavior. Soybean is basically a short-day (SD) plant, and soybean cultivars adapted to high-latitude environments possess insensitivity to photoperiods. Four major loci, *E1*, *E3*, *E4*, and *E7*, are known to be involved in the control of this insensitivity, particularly to long-day (LD) conditions (Buzzell 1971; Buzzell and Voldeng 1980; Cober *et al.* 1996; Cober and Voldeng 2001a,b). Soybean plants also respond differentially to light quality controlled artificially by fluorescent and incandescent lamps with different red-to-far-red quantum (R:FR) ratios (Cober *et al.* 1996). The *E3* locus was first identified with the use of fluorescent lamps to extend day length; the *e3e3* recessive homozygote can initiate flowering under LD conditions where the day length was extended to 20 hr using fluorescent lamps with a high R:FR ratio (Buzzell 1971). The *E4* locus was identified by extending the natural day length to 20 hr with incandescent lamps with a low R:FR ratio (Buzzell and Voldeng 1980). A recessive allele at the *E4* locus cannot singly confer the insensitivity to LD conditions induced by both fluorescent and incandescent lamps, but is necessary for plants homozygous for the *e3* allele to flower under the LD condition with a low R:FR ratio (Buzzell and Voldeng 1980; Saindon *et al.* 1989; Cober *et al.* 1996). The *E1* and *E7* loci are also involved in the control of the insensitivity to artificially induced LD conditions in the *e3* and *e4* backgrounds (Cober *et al.* 1996; Cober and Voldeng 2001b). Using near isogenic lines (NILs) for those maturity genes, Cober *et al.* (1996) found different responses of each photoperiod-sensitivity gene to 20-hr LD conditions with different R:FR ratios, suggesting that some of these genes may belong to the phytochrome family.

The role of phytochromes, *i.e.*, the R-light- and FR-light-absorbing photoreceptors, in flowering has been investigated in LD plants in Arabidopsis (Goto *et al.* 1991; Johnson *et al.* 1994; Devlin *et al.* 1996; Aukerman *et al.* 1997; Neff and Chory 1998; Devlin *et al.* 1999a,b; Mockler *et al.* 1999) and pea (Weller *et al.* 1997, 2001) and in SD plants in rice (Takano *et al.* 2001, 2005) and sorghum (Childs *et al.* 1997). Studies using loss-of-function mutants have revealed that phytochrome A (phyA) promotes flowering, whereas phytochrome B (phyB) plays an inhibitory role in floral initiation, and other phytochromes, such as phytochrome D (phyD) and phytochrome E (phyE), mediate responses similar to those involving phyB in a redundant manner (reviewed by Lin 2000). The phyA functions in rice, especially in combination with other phytochromes, are quite different from those in Arabidopsis. Takano *et al.* (2005) characterized the photomorphogenic roles of the rice phytochromes phyA, phyB, and phytochrome C (phyC), using various combinations of mutant genes. They found that the *phyA* monogenic mutant exhibited the same flowering time as the wild type under LD conditions, but, in the *phyB* and *phyC* mutant background, the flowering was greatly accelerated relative to *phyB* and *phyC* monogenic mutants. The accelerated flowering of the *phyA* mutant in the *phyB* and *phyC* mutant background is not observed in the LD plants of Arabidopsis (Reed *et al.* 1994; Devlin *et al.* 1996; Neff and Chory 1998) and pea (Weller *et al.* 1997). Nonadditive interactions observed between *phyA* and *phyB* or *phyC* in rice appear to rather resemble the different functions of the soybean maturity genes *E3* and *E4* in floral induction under LD conditions with different R:FR ratios (Buzzell and Voldeng 1980; Saindon *et al.* 1989; Cober *et al.* 1996). However, there is no report that addresses the possible relationships between the soybean maturity genes and phytochromes.

In this study, we isolated two *phyA* genes from the soybean genome to test the association between *phyA* genes and soybean maturity genes. We report here that the two *phyA* genes are mapped to homeologous regions of different linkage groups, and early maturing photoperiod-insensitive breeding lines with the *e4* gene possess a truncated *phyA* gene that resulted from an insertion of a *Ty1/copia*-like retrotransposon at one of the two homeologs. The data further suggest that the presence of duplicated copies of the *phyA* gene may have enabled the creation of photoperiod insensitivity through a loss-of-function mutation on one of the duplicated genes while protecting against any deleterious effect on photomorphogenesis.

MATERIALS AND METHODS

Plant materials: The cultivated soybean (*G. max*) line TK780 (TK) and a wild soybean (ssp. *soja*) line Hidaka 4 (H4) were used in this study. TK is an early maturing line with insensitivity to incandescent-lamp-induced LD (ILD; Cober *et al.* 1996), whereas H4 is a late-maturing ILD-sensitive wild soybean. Ninety-six recombinant inbred lines derived from a cross between TK and H4 were used for gene mapping. Three sets of NILs for the *E4* locus were also used in this study: 130S (ILD sensitive, *E1E1e3e3E4E4*) and 130I (ILD insensitive, *E1E1e3e3e4e4*), Harosoy (L58-266, *e1e1E3E3E4E4*) and its NILs for *e4* (OT94-41, *e1e1E3E3e4e4*), *e3* (L62-667, *e1e1e3e3E4E4*), and *e3e4* (OT89-5, *e1e1e3e3e4e4*). 130S and 130I were developed by means of a repetitive heterozygote selection method from an F_2 plant derived from a cross between the photoperiod-insensitive cultivars Miharudaizu and Sakamotowase, and their genotypes at three maturity loci, *E1*, *E3*, and *E4*, were determined (Abe *et al.* 2003). An $F_{8:9}$ family segregating for the *E4* locus, from which the NILs were derived, was also used for mapping. The genotype at the *E4* locus in the family was determined by the progeny test (Abe *et al.* 2003).

Isolation and sequence analysis of the *phyA* genes: Four steps (S1–S4) of PCR amplification and subsequent sequence analysis were performed to determine the nucleotide sequences of genomic DNA regions covering a coding region of the *phyA* genes. First, a genomic DNA corresponding to the 5′ portion of exon 1 in the upstream region of the putative soybean *phyA* gene (SOYPHYA: L34842) was amplified by PCR using the genomic DNA of TK and H4 as a template. The amplified fragments were cloned into a pGEM^R-T Easy vector (Promega). Sequence analysis of plasmid DNAs was performed using an ABI 377 sequencer (Applied Biosystems). Alignment of the DNA sequence was carried out using the CLUSTAL W Multiple Sequence Alignment Program version 1.8 (http://clustalw.genome.jp) (Thompson *et al.* 1994). Sequence analysis of the clones identified two *phyA* paralogs (*GmphyA1* and *GmphyA2*).

In step 2, a genome-walking technique using a BD Genome-Walker Universal kit (Clontech) was applied to obtain 3′ sequences of the paralogs. Briefly, genomic DNA from TK and H4 was digested in separate reactions with blunt-end endonucleases: *Dra*I, *Eco*RV, *Pvu*II, and *Stu*I. The ends of the DNA in each digested pool were ligated to an adaptor sequence. Long PCR was then performed using adaptor primers and paralog-specific primers according to the manufacturer's instructions. The products from the PCR reaction were cloned and sequenced.

In step 3, sequences from intron 1 to exon 4 were amplified using a forward primer common to both paralogs and a reverse primer designed to anneal the 3′ portion of exon 4 of SOYPHYA.

In the last step, 3′ rapid amplification of cDNA ends using cDNA from etiolated leaves was applied to obtain the sequences of the 3′ untranslated region (UTR) for both paralogs. Total RNA was isolated according to the method of Napoli *et al.* (1990), except that we removed genomic DNA from the RNA fraction using DNase I (Takara Bio). The cDNA was synthesized from total RNA essentially as described previously (Koseki *et al.* 2005). The cDNA synthesis reaction mixture was prepared by mixing 4 μl of 5× reaction buffer [250 mm Tris–HCl (pH 8.3), 375 mm KCl, 15 mm $MgCl_2$], 2 μl of 0.1 m DTT, 0.5 μl of RNaseOUT inhibitor (Invitrogen), 1 μl of 100 μm B26 primer, 4 μl of 2.5 mm dNTPs, the RNA solution, and water to a final volume of 19 μl. After the addition of 1 μl of reverse transcriptase (M-MLV, Invitrogen), the cDNA synthesis was performed at 42° for 1 hr. The reverse transcriptase was inactivated by heating the sample at 99° for 1 min. Amplification reactions were performed using the cDNA as a template with a forward primer designed to anneal exon 3 of both paralogs and B25 reverse primer. Amplified products were cloned and sequenced.

The transcripts covering the entire coding regions of the *phyA* paralogs were also amplified by the reverse-transcription-mediated PCR (RT–PCR) using a set of primers designed to anneal the 5′- and 3′-UTRs. Sequence analysis of the PCR products as well as subsequent DNA gel-blot analysis demonstrated that no amplification of an artifact (*i.e.*, recombinant PCR) occurred during the process of genome walking.

An additional two steps of PCR amplifications and sequence analyses (S5 and S6) were performed to determine the sequences of the retroelement and 3′ sequence of the *GmphyA2* gene in TK. The sequences for 130S and 130I were also determined by the same procedures as were applied to gene isolation in H4 and TK except for step 2, in which, in place of adaptor primer and enzyme-digested DNA libraries, paralog-specific primers and genomic DNA were used for the PCR reactions. The primers used for isolation of the *phyA* genes are listed in supplemental Table S1.

DNA gel-blot analysis: The genomic DNA of 130S and 130I was digested with three restriction enzymes, *Eco*RI, *Eco*RV, and *Dra*I. The digested DNA of 5 μg/lane was separated by electrophoresis and blotted onto nylon membranes (Hybond-N+, Amersham Pharmacia Biotech). Southern hybridization was performed at 55° for 16 hr with a probe labeled by AlkPhos Direct (GE Healthcare). A 1414-bp fragment covering the 5′ portion of exon 1 of *GmphyA2* was amplified by PCR and used as a probe. Hybridization signals were detected by chemiluminescence on X-ray film.

Genetic mapping of the *GmphyA1* and *GmphyA2* genes: Two *phyA* paralogs were mapped onto genetic maps that had been constructed using the recombinant inbred (RI) mapping population and the progeny of a cross between Miharudaizu and Sakamotowase. The former map consists of 20 linkage groups, covering 2383 cM in its entire length (Liu *et al.* 2007). The linkage map is available via Legume Base Glycine max/soja (http://www.shigen.nig.ac.jp/bean/glycinesoja.jp). The latter map comprises 5 markers, including the *E4* locus, and spans a region of ~30 cM of linkage group I (LG I) (Abe *et al.* 2003).

A two-base insertion/deletion (indel) in intron 3 was used for mapping the *GmphyA1* gene in the RI population. The regions harboring the indel were amplified by PCR using a forward primer (PhyA1-For) and a reverse primer (PhyA1-Rev), of which the forward primer was labeled by a fluorescent dye and detected by the ABI 377 sequencer with GeneScan software (Applied Biosystems). For mapping the *GmphyA2* gene, the presence or absence of the retroelement was analyzed by PCR using a common forward primer in exon 1 (PhyA2-For) and allele-specific reverse primers in the retroelement (PhyA2-Rev/e4) or in exon 1 (PhyA2-Rev/E4). PCRs performed in the presence of the three primers resulted in the amplification of a 837-bp fragment from TK and 130I and a 1229-bp fragment from H4 and 130S.The amplified products were separated by electrophoresis on a 0.8% agarose gel and visualized under UV light. The primers used for mapping of the *phyA* genes are listed in supplemental Table S1.

The data were incorporated into the two maps using the Map Manager program QTXb17 (http//:mapmgr.roswellpark.org/mapmgr.hyml). Marker order and distance were determined using the Kosambi function and a criterion of 0.001 probability.

Analysis of gene expression by RT–PCR: Total RNA was isolated from cotyledon, hypocotyl, and the etiolated and green primary leaves of 130S and 130I plants, as mentioned above. The cDNA was synthesized from total RNA as described above. Transcripts of the β-*tubulin* gene were amplified by PCR reaction with the primers β-tub-For and β-tub-Rev from synthesized cDNA as a control for the RT–PCR of *GmphyA1* and *GmphyA2* transcripts. The PCR cycle was 94° for 30 sec, 61°

for 30 sec, 72° for 30 sec, and 72° for 4 min. This cycle was repeated 28 times. RT–PCR was also performed using a common forward primer in exon 3 of both paralogs (PhyA1/2-RT-For) and a paralog-specific reverse primer in the 3′-UTR (PhyA1-RT-Rev for *GmphyA1* and PhyA2-RT-Rev for *GmphyA2*). PCR with these primers amplifies a 451-bp fragment and a 325-bp fragment for *GmphyA1* and *GmphyA2*, respectively, which are distinct from those amplified from genomic DNAs (a 789-bp fragment for *GmphyA1* and a 654-bp fragment for *GmphyA2*). The PCR cycle was 94° for 30 sec, 52° (for *GmphyA1*) or 58° (for *GmphyA2*) for 30 sec, 72° for 30 sec, and 72° for 4 min. The reaction products were separated by electrophoresis on a 1% agarose gel and visualized under UV light. Digestion of RCR products by the restriction enzymes *Hin*dIII and *Mbo*I was used to distinguish the transcripts from the *GmphyA1* and *GmphyA2* genes. Ten microliters of the PCR products was digested and separated by electrophoresis on a 0.8% agarose gel and visualized under UV light.

Quantitative RT–PCR was done essentially as described previously (Nagamatsu *et al.* 2007). The quantitative RT–PCR mixture was prepared by mixing a 1-μl aliquot of the reaction mixture of cDNA synthesis, 5 μl of 1.2 μM primer premix, 10 μl SYBR Premix ExTaq Perfect Real Time (TaKaRa Bio), and water to a final volume of 20 μl. The analysis was done using the DNA Engine Opticon 2 system (MJ Research). The PCR cycling conditions were 95° for 10 sec, 52° (for *GmphyA1* and its controls) or 58° (for *GmphyA2* and its controls) for 20 sec, 72° for 20 sec, and 78° for 2 sec. This cycle was repeated 40 times. Fluorescence quantification was carried out before and after the incubation at 78° to monitor the formation of primer–dimers. The mRNA level of the β-*tubulin* gene was used as a control for the analysis. A reaction mixture without reverse transcriptase was also used as a control to confirm that no amplification occurred from genomic DNA contaminants in the RNA sample. In all PCR experiments, amplification of a single DNA species was confirmed by both melting curve analysis of quantitative PCR and gel electrophoresis of PCR products. The primers used are a common forward primer in exon 4 of both paralogs (PhyA1/2-realtime-For) and paralog-specific reverse primers in the 3′-UTR of *GmphyA1* (PhyA1-RT-Rev) and *GmphyA2* (PhyA2-RT-Rev). Primers β-tub-realtime-For and β-tub-Rev were also used for amplifying the β-*tubulin* transcripts. The primers used for RT–PCR are listed in supplemental Table S1.

Measurement of hypocotyl elongation: Three sets of NILs for the *E4* locus were assayed for de-etiolation responses to different light conditions. Three-day-old germinated seeds with a primary root of ~1 cm in length were transplanted into plastic pots filled with vermiculite, and then plants were grown at 25° in darkness, continuous R-light or FR-light conditions. Monochromatic light was supplied by an R-light-emitting diode panel (model LED-R; EYELA) or an FR-light-emitting panel (model LED-FR; EYELA). The average photon flux was 2.5 μmol photons m^{-2} s^{-1} for R and 7.6 μmol photons m^{-2} s^{-1} for FR, as measured using a LI-COR quantum sensor (model LI-4800C, LI-COR). Hypocotyl length was measured 8 days after seeding. Three independent experiments with a replication of six plants were carried out. Two-way analysis of variance was performed to test the statistical significance for effects of light sources on the *E4* and *e4* genotypes.

RESULTS

Isolation and sequence analysis of the *phyA* genes: We analyzed the nucleotide sequences of genomic DNA containing the entire regions of the soybean *phyA* genes in the *max* line TK780 (TK; photoperiod insensitive) and *soja* line Hidaka 4 (H4; photoperiod sensitive). We first aimed to find a polymorphism in the *phyA* gene region between the TK and H4 lines, which would enable analysis of the genetic relationships between *phyA* gene(s) and previously reported major genes as well as QTL for flowering using an RI mapping population derived from a cross between these lines. The *phyA* genes were isolated by a method comprising multiple steps of PCR amplification. The first-step PCR (S1 in Figure 1) was targeted to a 1024-bp region from a 5′ upstream site to the 5′ portion of exon 1 on the basis of the nucleotide sequence of a putative soybean *phyA* gene (SOYPHYA: L34842). This yielded two fragments of different sizes (1021 and 810 bp) from both the TK and H4 genomic DNA. Sequence analyses of these fragments revealed that both contained a portion of the *phyA*-coding region and its 5′ upstream region (supplemental Figure S1). The sequence of the 1021-bp fragment from TK and H4 was almost identical to the upstream sequence of SOYPHYA, whereas the sequence of the 810-bp fragment from TK and H4 differed from the sequences of the 1021-bp fragment and SOYPHYA by many base substitutions and indels ranging from 1 to 166 bp, the

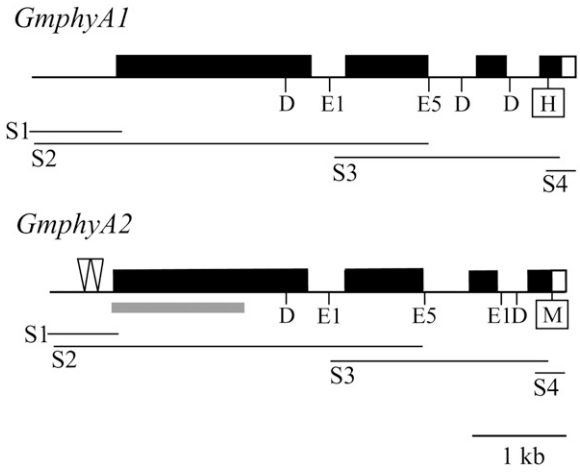

Figure 1.—Structure of the soybean *phytochrome A* paralogs, *GmphyA1* and *GmphyA2*. Exons are represented as boxes of which solid and open parts indicate the coding regions and the 3′-untranlated regions, respectively. Lines indicate genome-walking steps (S1–S4). Arrowheads indicate two major deletions (28 and 166 bp) in the *GmphyA2* sequence relative to the *GmphyA1* sequence (see supplemental Figure S1). The shaded box indicates a 1414-bp fragment used as a probe in DNA gel-blot analysis. D, *Dra*I; E1, *Eco*RI; E5, *Eco*RV; H, *Hin*dIII; M, *Mbo*I. For *Hin*dIII and *Mbo*I, only the restriction sites that exist in exon 4, which were used to distinguish the RT–PCR products of transcripts from these paralogs, are presented (see Figure 6B). These restriction sites are marked by a square. The structures of *GmphyA1* and *GmphyA2* were highly conserved among the four plant lines (TK780, Hidaka 4, 130S, and 130I) and between the two photoperiod-sensitive lines (Hidaka 4 and 130S), respectively. In contrast, the structure of *GmphyA2* in Hidaka 4 and 130S was different from that in TK780 or 130I by a 6238-bp insertion in exon 1 (see Figure 2A).

latter of which mainly accounted for the difference in fragment size (Figure 1, supplemental Figure S1). The 1021- and 810-bp fragments were considered to contain the portions of the paralogs of SOYPHYA, and the corresponding genes were designated as *GmphyA1* and *GmphyA2*, respectively.

To isolate the entire gene regions of *GmphyA1* and *GmphyA2*, a genome-walking technique comprising ligation of restriction-enzyme-digested DNA to adaptors followed by amplification with nested PCR using adaptor-specific primers and gene-specific primers (for details, see MATERIALS AND METHODS). The genome-walking approaches (S2 in Figure 1) successfully produced amplified fragments for both paralogs when applied to the *Dra*I-digested and *Eco*RV-digested DNA libraries of H4. Fragments amplified from the *Eco*RV-digested DNA libraries were subjected to further sequence analyses. Through additional steps of PCR amplifications and sequence analyses (S3 and S4 in Figure 1), we finally obtained contigs of 5827- and 5557-bp sequences containing the entire gene regions of *GmphyA1* and *GmphyA2*, respectively. The cDNAs of these genes were also isolated by RT–PCR using RNA extracted from etiolated tissues. Comparisons between genomic DNA and cDNA sequences revealed that the *GmphyA1* and *GmphyA2* genes are composed of four exons, as in the *phyA* genes of other plant species, and are predicted to encode a 1130-amino-acid protein and a 1123-amino-acid protein, respectively (Figure 1; supplemental Figure S2). The amino acid sequences deduced from the nucleotide sequences of the *GmphyA1* and *GmphyA2* genes displayed a high similarity with those of the *phyA* genes of other plant species (supplemental Figure S2). These sequences also contained conserved motifs essential for normal function of the protein (supplemental Figure S2).

The *GmphyA1* and *GmphyA2* genes were also isolated from TK and NILs for the *E4* locus, 130S and 130I. The genotypes at the *E4* locus in the 130S and 130I plants are known (ABE *et al.* 2003). The structure of the *GmphyA1* gene isolated from TK was almost identical to that from H4. On the other hand, sequence analysis of the portion of the *GmphyA2* gene in TK obtained by step 2 (S2 in Figure 2A) revealed that the portion consists of a chimeric sequence comprising exon 1 and an unknown sequence. As a result of two additional steps of PCR amplification and sequence analysis (S5 and S6 in Figure 2A), a sequence of 6238 bp was found to be inserted at nucleotide 692 from the start codon of the *GmphyA2* gene in TK. Sequence analyses of *GmphyA2* in 130S and 130I indicated that the 6238-bp segment was also inserted in exon 1 in 130I harboring the *e4* allele as in TK, but was not inserted in 130S harboring the *E4* allele, as it was not in H4.

The insertion of the segment in exon 1 of *GmphyA2* was confirmed in two sets of NILs for the *E4* locus: Harosoy and Harosoy-*e4* and Harosoy-*e3* and Harosoy-*e3e4*. PCRs performed with a common forward primer (PhyA2-For) and two allele-specific reverse primers that were designed

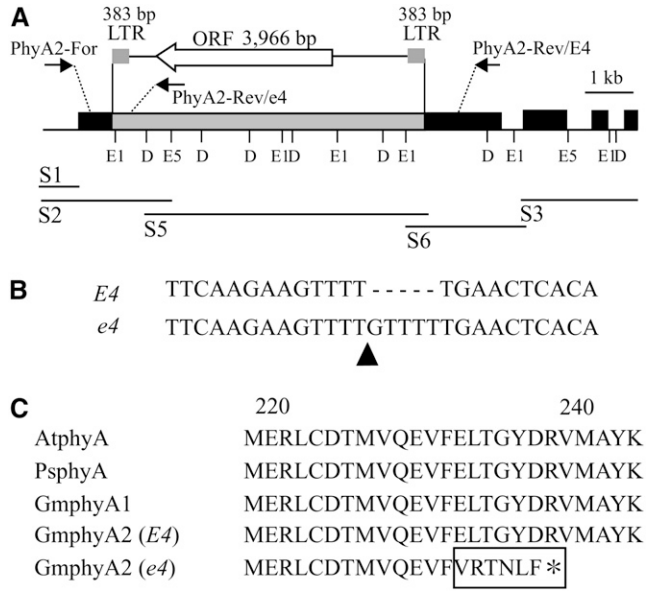

FIGURE 2.—Structure of *GmphyA2* in the photoperiod-insensitive line TK780 and the generation of a new stop codon due to the insertion of a *Ty1/copia*-like retrotransposon. (A) Structure of *GmphyA2* from TK780 and orientation of the retrotransposon. Solid and shaded boxes indicate exons and the *Ty1/copia*-like retrotransposon, respectively. Lines indicate genome-walking steps (S1–S3, S5, and S6). The positions of primers used to amplify the portion of exon 1 containing no insertion of the retrotransposon (the *E4* allele; primers PhyA2-For and PhyA2-Rev/E4) or a portion of exon 1 containing the retrotransposon (the *e4* allele; primers PhyA2-For and PhyA2-Rev/e4) are shown by arrows. The open arrow indicates the position and direction of the ORF in the retrotransposon. D, *Dra*I; E1, *Eco*RI; E5, *Eco*RV. The structure of the gene in TK780 was identical to that in 130I (*e4e4*). (B) Target-site duplication at the position of insertion in *GmphyA2* in the NILs for the *e4* allele. A dash indicates lack of nucleotide. The arrowhead indicates the position of the insertion of a retrotransposon. (C) Changes in amino acid sequence that were caused by the insertion of a retrotransposon in *GmphyA2* in the NILs for the *e4* allele. Amino acid sequences deduced from the DNA sequences of *AtphyA*, *PsphyA*, *GmphyA1*, and *GmphyA2* from *E4* NILs and of *GmphyA2* from *e4* NILs were aligned. Amino acid sequence generated by the retrotransposon is marked by a square. The asterisk indicates the stop codon. Positions of amino acids are indicated above the sequences.

to anneal the segment (PhyA2-Rev/e4) or exon 1 (PhyA2-Rev/E4) (Figure 2A) amplified the 837-bp fragment in homozygous lines for the *e4* allele and the 1229-bp fragment in those for the *E4* allele, respectively (Figure 3A). These results suggest that plants carrying the *e4* allele possess the 6238-bp insertion in exon 1 of *GmphyA2*.

Identification of a *Ty1/copia*-like retrotransposon in exon 1 of *GmphyA2*: The 6238-bp sequence inserted in exon 1 of the *GmphyA2* gene was subjected to a further characterization. The inserted sequence comprised two 383-bp long terminal repeats (LTRs) and an internal domain of 5472 bp (Figure 2A). The element was flanked by a 5-bp target-site duplication sequence (5′-GTTTT-3′) (Figure 2B). The nucleotide sequences of the two LTRs

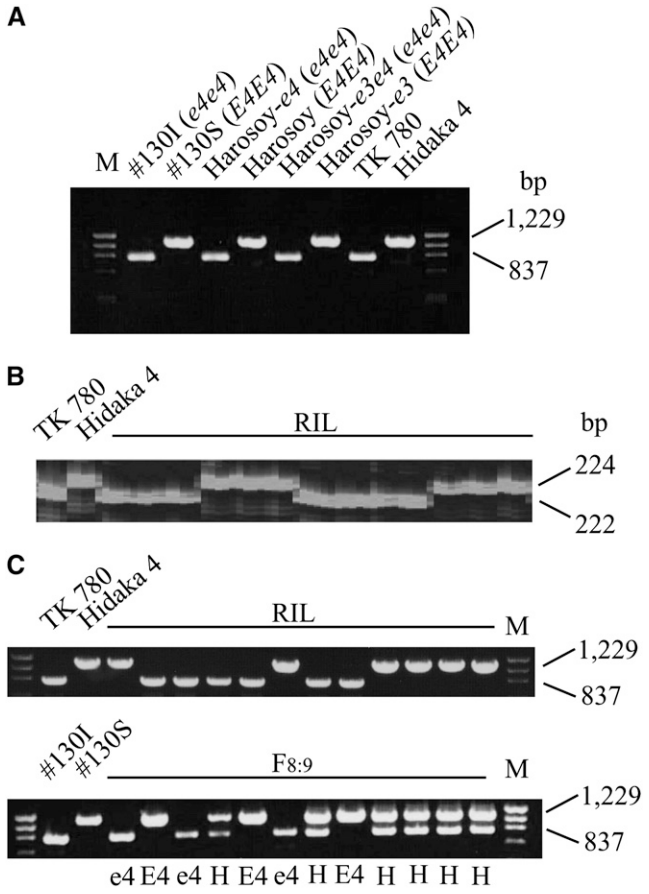

FIGURE 3.—PCR amplification of the *GmphyA2* fragments specific to homozygous plants for the *E4* or *e4* alleles and segregations of PCR-amplified fragments for *GmphyA1* and *GmphyA2* in mapping populations. (A) PCR amplification of the *GmphyA2* fragments specific to homozygous plants for the *E4* or *e4* alleles. A PCR with a common primer (PhyA2-For) and two reverse primers (PhyA2-Rev/e4 and PhyA2-Rev/E4) produces fragments of 837 and 1229 bp in the lines homozygous for the *e4* allele and the *E4* allele, respectively. Positions of the primers used are presented in Figure 2. (B) Polymorphism of PCR-amplified fragments for *GmphyA1* and segregation in the RI mapping population. (C) Segregations of PCR-amplified fragments for *GmphyA2* in the RI mapping population and a segregating $F_{8:9}$ sib family for the NILs, 130I and 130S. The letters below the electrogram indicate the genotypes at the *E4* locus, which were determined by the progeny test (ABE *et al.* 2003); *e4e4* (e4), *E4e4* (H), and *E4E4* (E4). M, φx174 DNA digested with *Hae*III loaded for size markers.

were 100% identical with each other and contained 2-bp inverted repeats (5′-TG . . . CA-3′) in their end portions. These features are canonical for retrotransposons (reviewed by KUMAR and BENNETZEN 1999).

In the internal sequence of the element, the sequences of the primer-binding sites and the polypurine tract were identified adjacent to the LTRs (not shown). A single large open reading frame (ORF) comprising 3966 bp was also identified (Figure 2A). Both the sequence similarity with other retroelements and the allocation of motifs in the ORF revealed the presence of Gag-protease-integrase-RT-RNaseH domains in this order, which indicated that the retroelement (database accession no.AB370254) belongs to the *Ty1/copia*-like retrotransposon (the details will be published elsewhere). These analyses also indicated that the element was inserted in exon 1 of *GmphyA2* in an orientation opposite to the transcription of the *GmphyA2* gene (Figure 2A). The insertion resulted in the production of a truncated GmphyA2 protein comprising 237 amino acids in the presence of a termination codon in the inserted sequence (Figure 2C).

Polymorphism and genetic mapping of *GmphyA1* and *GmphyA2*: Polymorphisms in the gene regions of *GmphyA1* and *GmphyA2* between the TK and H4 lines were detected in addition to the presence or absence of the insertion in exon 1 of the *GmphyA2* gene (supplemental Figure S3). The *GmphyA1* gene exhibited five SNPs, a 2-bp indel, and a difference in the repeat number of a mononucleotide SSR between TK and H4. All of these differences were localized in introns. In the *GmphyA2* gene, in addition to the large insertion in exon 1 in TK, a total of seven SNPs and a difference in the repeat number of the mononucleotide SSR were present between TK and H4. Of these, five SNPs were located in exons 1 and 2, but only a SNP at nucleotide 452 from the ATG start codon led to an amino acid substitution (leucine to serine) in H4 relative to TK.

We used the 2-bp indel in intron 3 for mapping the *GmphyA1* gene. The PCR amplification produced 222- and 224-bp fragments for TK and H4, respectively (Figure 3B). Mapping these fragments in the RI population allocated the *GmphyA1* gene to the vicinity of a SSR marker (Satt262) in LG O (Figure 4). The presence or absence of a *Ty1/copia*-like retrotransposon was used to map the *GmphyA2* gene (Figure 3C). The *GmphyA2* gene was assigned between an AFLP marker (ACC/CAA-216) and an SSR marker (Satt354) in LG I (Figure 4), in which the *E4* locus had been mapped previously (ABE *et al.* 2003; MOLNAR *et al.* 2003).

To confirm that the *GmphyA2* gene is mapped at the same position as the *E4* locus, cosegregation of the *GmphyA2* and *E4* genes was tested in an $F_{8:9}$ segregating family from which 130S and 130I were derived (ABE *et al.* 2003). The segregation data of the *E4* locus taken from ABE *et al.* (2003) was used for the analysis. Mapping of the fragments according to the presence or absence of the retrotransposon showed that the *GmphyA2* gene cosegregated with the *E4* locus (Figure 3C), and both were allocated between a SSR marker (Satt496) and an isozyme marker (*Enp*) in LG I (Figure 4). Thus, these results suggest that GmphyA2 is encoded by the *E4* locus.

Soybean possesses a complicated genome in which homeologous duplicated regions are scattered across different linkage groups (SHOEMAKER *et al.* 1996; LEE *et al.* 1999). Mapping studies with RFLP markers have revealed that LGs I and O share such homeologous duplicated regions, which are anchored by seven RFLP

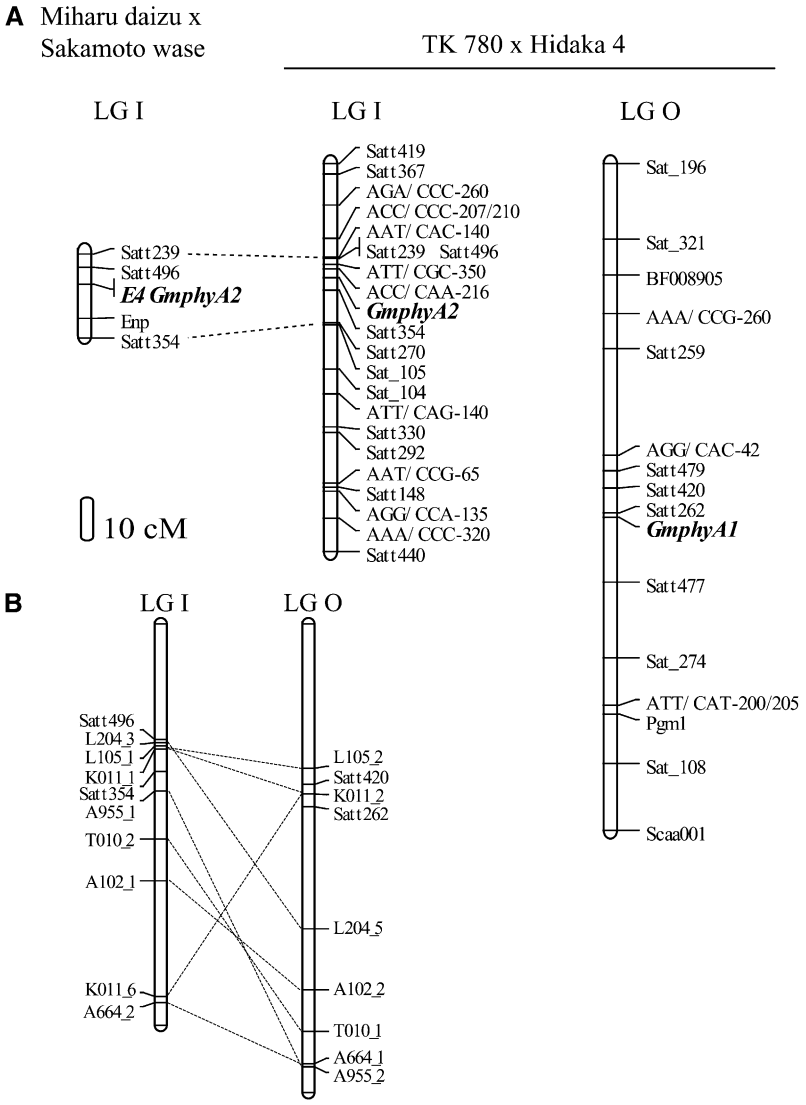

FIGURE 4.—Locations of *GmphyA1* and *GmphyA2* on the soybean linkage map. (A) Locations of *GmphyA1* (in linkage group O) and *GmphyA2* (in linkage group I) mapped by using two segregating populations. (B) Homeologous regions between linkage groups I and O, which are anchored by seven common RFLP probes, taken from SoyBase (http://www.SoyBase.org). *GmphyA1* is located in the vicinity of Satt420 and Satt262 in LG O, while *GmphyA2* is located in the vicinity of Satt496 and Satt354 in LG I (A), and all these markers are located in the homeologous regions.

markers through large regions of their linkage groups (Figure 4B; see SoyBase: http://www.SoyBase.org). Locations of markers closely linked with *GmphyA1* (Satt 420 and Satt262) in LG O and those linked with *GmphyA2* (Satt496 and Satt354) in LG I indicated that *GmphyA1* and *GmphyA2* are located in the homeologous chromosomal region, in which these genes were mapped near a common RFLP marker (K011) (Figure 4B).

Organization of the *phyA* genes in the soybean genome: Polymorphisms of the *phyA* genes between homozygotes for the *E4* or *e4* alleles were also analyzed by DNA gel-blot analysis. DNA digested with each of three restriction enzymes, *Dra*I, *Eco*RI, and *Eco*RV, was hybridized to a labeled DNA fragment of 1.4 kb covering a 5′ portion of exon 1 of *GmphyA2* (Figure 1). The *GmphyA2* gene in 130I (*e4*) possesses the retroelement, which is expected to divide the hybridized region in exon 1 into two fragments as a consequence of digestion at the restriction sites within the element (Figure 2A). As expected, one hybridization signal specific to 130S (*E4*) and two hybridization signals specific to 130I were detected (Figure 5).

DNA gel-blot analysis further provided an estimate of the copy number of the *phyA* gene in the soybean genome. In addition to the polymorphic hybridization signals, three or four nonpolymorphic hybridization signals, excluding weakly hybridized ones, were detected per lane (Figure 5). These results suggest that there are at least two more copies of *phyA* genes in the soybean genome in addition to the *GmphyA1* and *GmphyA2* genes.

Expression of the *GmphyA1* and *GmphyA2* genes: We analyzed the transcription profiles of the *GmphyA1* and *GmphyA2* genes in cotyledons, hypocotyls, and etiolated and green primary leaves of 130S (*E4E4*) and 130I (*e4e4*) by RT–PCR. The *GmphyA1* gene was expressed in all the tissues examined in the two lines (Figure 6A); the transcripts accumulated at slightly higher levels in etiolated leaves than in the other tissues. The gene expression profiles of *GmphyA2* were similar to those observed in *GmphyA1* in 130S, but no transcripts were detected in 130I.

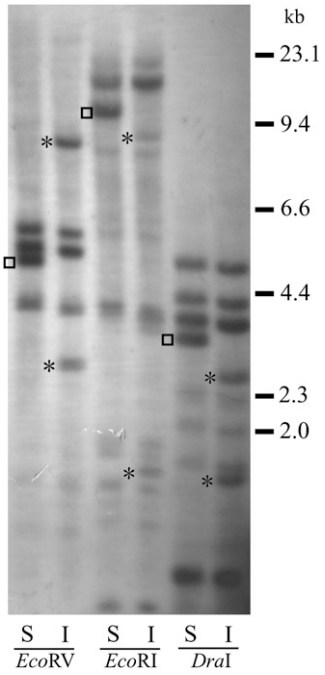

FIGURE 5.—Gel-blot analysis of the DNA of the NILs for the *E4* locus, 130S (*E4*) and 130I (*e4*). Genomic DNA of 130S (S) and 130I (I) was digested with *Eco*RI, *Eco*RV, and *Dra*I, separated by agarose gel electrophoresis, and blotted to a membrane. The blot was hybridized with a labeled 1.4-kb fragment corresponding to the 5′ portion of exon 1 of *GmphyA2* (see Figure 1). Squares and asterisks indicate hybridized fragments specific to the *E4* allele and the *e4* allele, respectively. Sizes of the marker fragments are represented in kilobases.

The absence of the *GmphyA2* transcripts at a detectable level in 130I may be due to destabilization of the mRNA caused by the insertion of a retroelement that led to premature termination of translation, which is known as nonsense-mediated mRNA decay (reviewed by MAQUAT 2004). Alternatively, it may be simply due to an inhibition of the transcription by the insertion of a retroelement. The fragments amplified by the RT–PCR were expected to contain restriction sites of *Hin*dIII for *GmphyA1* and of *Mbo*I for *GmphyA2* (Figure 1). Digestion of the *GmphyA1* and *GmphyA2* products by the respective restriction enzymes confirmed that no cross-amplification of the transcripts from these genes occurred by RT–PCR (Figure 6B).

Differences in the mRNA levels of *GmphyA1* and *GmphyA2* between different tissues or genotypes were confirmed by quantitative RT–PCR analysis (Figure 6C). This analysis revealed that the mRNA levels of both *GmphyA1* and *GmphyA2* were more than twofold higher in etiolated leaves than in green leaves (Figure 6C). These results are consistent with previous observations that the mRNA level of *phyA* is downregulated by light and is higher in etiolated tissues in other plants, *e.g.*, Arabidopsis (CLARK *et al.* 1994), tomato (HAUSER *et al.* 1998), morning glory (CARTER *et al.* 2000), and rice (BASU *et al.* 2000).

Effects of the *E4* and *e4* alleles on hypocotyl elongation under FR-rich light condition: The roles of phytochromes in photomorphogenesis have been well characterized by measuring the seedling growth of mutants deficient in phytochromes under different light conditions. Mutants deficient in *phyA* cannot de-etiolate under continuous FR light, whereas those deficient in *phyB* do not exhibit the de-etiolation response under continuous R light in Arabidopsis (NEFF and CHORY 1998), pea (WELLER *et al.* 1997; 2001), and rice (TAKANO *et al.* 2001, 2005). To know whether soybean could exhibit a similarity of response in photomorphogenesis depending on the genotypes at the *E4* locus, the growth response of soybean seedlings to different types of light was analyzed.

We examined the hypocotyl growth of 8-day-old seedlings grown under darkness, continuous R, or FR monochromic light in three sets of NILs for the *E4* locus (Figure 7). When grown in complete darkness, the growth of seedlings in the 130I (*e4*) plants did not differ significantly from the dark-grown 130S (*E4*) plants (Figure 7A). The growth of seedlings in the 130I (*e4*) plants grown in R light also did not differ significantly from that in the 130S (*E4*) plants when grown under R light (Figure 7A). Hypocotyl length was the longest in the seedlings grown in complete darkness and was reduced in those grown in continuous R light in both of the NILs for the *E4* and *e4* alleles (Figure 7A). The hypocotyls of 130I (*e4*) were significantly longer than those of 130S (*E4*) when seedlings were grown under continuous FR light, and in 130I (*e4*), hypocotyls were longer in the seedlings grown under continuous FR light than in the seedlings grown under continuous R light. However, the hypocotyls of the 130I (*e4*) plants grown under continuous FR light were still shorter than those of dark-grown seedlings (Figure 7A). A similar result was obtained in the other two sets of NILs for the *E4* locus in the Harosoy background, independently of the genotype at the *E3* locus (Figure 7, B and C). Two-way analysis of variance revealed that the effects of both light sources and their interactions with the *E4* genotypes were significant at a 5% level in all three sets. These results suggest that the function of phyA is lost, not completely but partially, in the NILs for *e4*.

DISCUSSION

Multiple homologs of the *phyA* gene are present in the soybean genome: We isolated two paralogs of the *phyA* gene, designated *GmphyA1* and *GmphyA2* in soybean. These genes showed a high similarity in amino acid sequences and similar transcription profiles in the tissues examined, despite their divergent sequences in the 5′ upstream region. Genetic mapping allocated *GmphyA1* and *GmphyA2* into the homeologous regions of LGs O and I, respectively. Therefore, *GmphyA1* and *GmphyA2* are not solely paralogs, but rather are homeo-

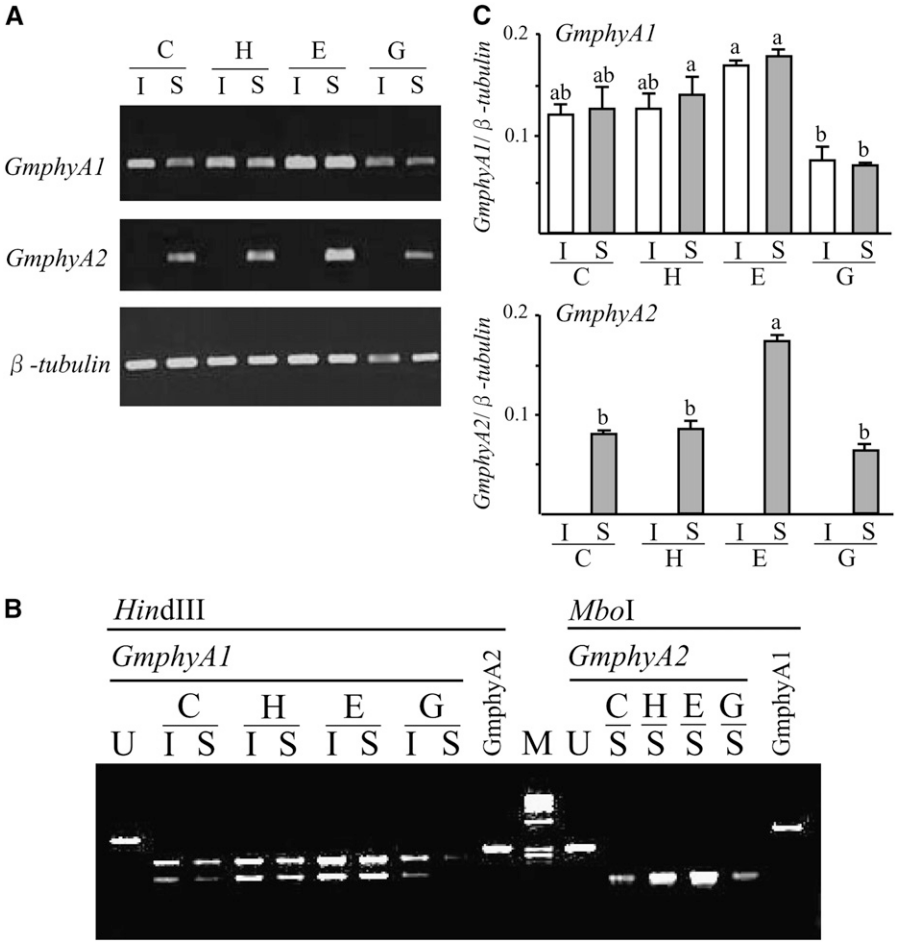

FIGURE 6.—Expression analysis of *GmphyA1* and *GmphyA2* in the NILs for the *E4* locus, 130S (*E4*) and 130I (*e4*), by RT–PCR. (A) Gene expressions in cotyledon (C), hypocotyl (H), etiolated primary leaves (E), and green primary leaves (G) of 130I (I) and 130S (S) plants. Transcripts of the β-*tubulin* gene were amplified as a control. (B) The presence or absence of *Hin*dIII or *Mbo*I restriction sites (see Figure 1) in the products of amplification from *GmphyA1* and *GmphyA2* transcripts by RT–PCR. The products of RT–PCR were separated by agarose gel electrophoresis after treatment with restriction enzymes. Complete digestion of the *GmphyA1* and *GmphyA2* products with *Hin*dIII and *Mbo*I, respectively, indicates that no cross-amplification of the transcripts from these genes occurred by RT–PCR. Primers were designed to anneal the 3′-UTR and exon 3, which amplify DNA fragments of different sizes from cDNA and genomic DNA, to confirm that no amplification occurred from genomic DNA contaminants in the RNA sample (for details, see text). U and M indicate undigested fragments and a size marker (φx174 DNA/*Hae*III), respectively. (C) Quantitative analysis of the mRNA levels of *GmphyA1* and *GmphyA2*. The mRNA levels of *GmphyA1* and *GmphyA2* relative to the β-*tubulin* mRNA level were assessed by quantitative RT–PCR. The data shown represent means and standard errors obtained from three replicates of the analysis. Means that are not indicated by the same letter are significantly different ($P < 0.01$). *GmphyA1* and *GmphyA2* were expressed at a higher level in etiolated leaves.

logs that resulted from ancient chromosomal duplications and rearrangements in soybean. The DNA gel-blot analysis further suggested that, in addition to the two *phyA* homeologs, there were at least two additional *phyA* homologs in the soybean genome (Figure 5). We actually found four *phyA* homologs, including the two *phyA* homeologs isolated in this study, in a recently released genome sequence database of soybean (http://www.phytozome.net/soybean). We have performed RT–PCR analyses of the two *phyA* genes other than *GmphyA1* and *GmphyA2* using specific primers designed on the basis of the sequence information of the database. No product of amplification was obtained for these genes (data not shown), suggesting that *GmphyA1* and *GmphyA2* are the only *phyA* genes expressed in the soybean lines used in this study. The presence of multiple copies of the *phyA* gene in the soybean genome is in a sharp contrast to the two legume model species, *L. japonicus* and *M. truncatula*, in which *phyA* is represented by a single gene in their EST databases (HECHT *et al.* 2005). The increased number of *phyA* homologs in the soybean genome has most likely derived from genome duplication due to paleopolyploidy, as suggested by SHOEMAKER *et al.* (2006).

***GmphyA2* is a candidate for the photoperiod sensitivity gene *E4*:** Of the two *phyA* homeologs identified, *GmphyA2* cosegregated with the photoperiod sensitivity gene *E4*. Sequence analyses of *GmphyA2* from the NILs for the *E4* locus revealed that the *e4* allele possessed a *Ty1/copia*-like retrotransposon in exon 1, an insertion that is absent in the *E4* allele. This insertion resulted in a stop codon shortly after the insertion and a truncated protein comprising 237 amino acids that lacks a chromophore-binding site and other cassettes necessary for phyA function. The results obtained in this study therefore strongly suggest that the *E4* locus encodes GmphyA2 and that the *Ty1/copia*-like retrotransposon disrupts the function of *GmphyA2* to condition the insensitivity to long days induced by incandescent lamps with low R:FR ratios.

phyA is involved in various developmental processes that are regulated by different R:FR ratios, such as germination, de-etiolation, early neighbor detection, shade avoidance, resetting of the circadian clock, and flowering (CASAL *et al.* 1997). Studies using mutants deficient in phytochromes in Arabidopsis, pea, and rice have revealed different roles of phyA and phyB in the de-etiolation responses to different light conditions (WELLER *et al.* 1997; NEFF and CHORY 1998; TAKANO

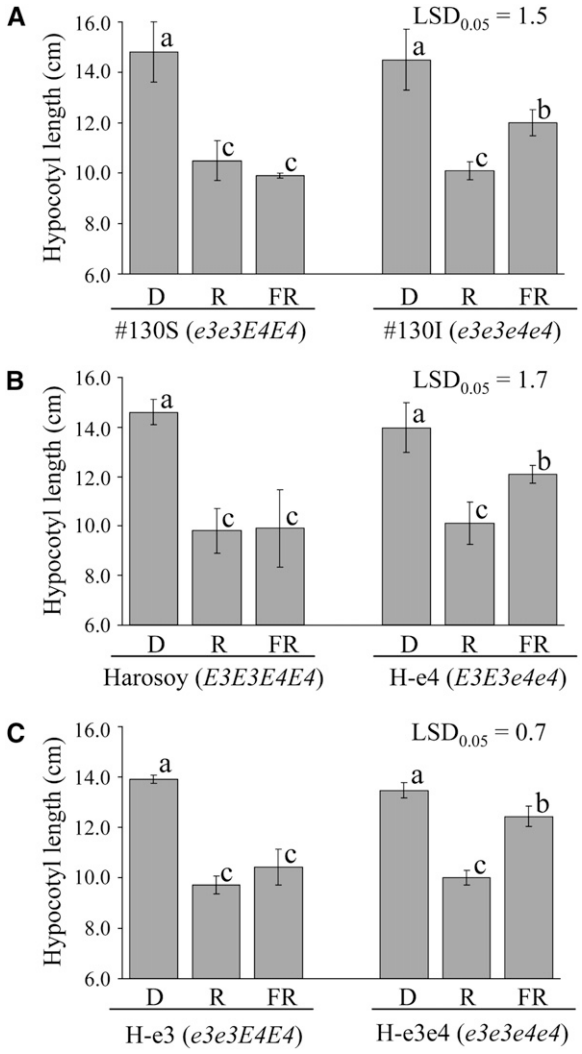

FIGURE 7.—Hypocotyl growths in three sets of NILs for the *E4* locus. Eight-day-old seedlings were grown in complete darkness (D), red light (R), and far red light (FR). Comparisons are made between 130S and 130I (A), between Harosoy and Harosoy-*e4* (H-e4) (B), and between Harosoy-*e3* (H-e3) and Harosoy-*e3e4* (H-e3e4) (C). The data represent the mean and standard errors obtained from three independent experiments with a replication of six plants. Means that are not indicated by the same letter are significantly different ($P < 0.05$) in A–C.

et al. 2001, 2005). phyA and phyB are involved in de-etiolation of seedlings under FR-light and R-light conditions, respectively. Like the *phyA* null mutants of these species, the *e4* allele impaired the de-etiolation response to the continuous FR-light condition. Plants homozygous for the *e4* allele produced significantly elongated hypocotyls under continuous FR light, when compared with those grown under continuous R light, whereas plants homozygous for the *E4* allele exhibited similar hypocotyl growth under both FR and R light (Figure 7). The impaired de-etiolation conferred by the *e4* allele under FR light is thus consistent with the notion that *E4* encodes GmphyA2.

Another important function of phyA is the photoperiodic control of flowering. In Arabidopsis, an LD plant, phyA promotes flowering, whereas phyB, phyD, and phyE redundantly play an inhibitory role in floral initiation (reviewed by LIN 2000). The promotive function of phyA in flowering is also observed in pea (LD plant) and rice (SD plant), when grown under inductive light conditions (WELLER *et al.* 1997; TAKANO *et al.* 2005). However, the phyA functions in rice are quite different from those in Arabidopsis, especially when combined with other phytochromes. The rice *phyA* monogenic mutants exhibited the same flowering time as the wild type under LD conditions, but when combined with the *phyB* or *phyC* null mutants, the *phyA* null mutants greatly accelerated flowering (TAKANO *et al.* 2005). In soybean, the *e4* allele, a loss-of-function allele in *GmphyA2*, confers no flowering delay under SD conditions (COBER *et al.* 1996; COBER and VOLDENG 2001b), unlike the *phyA* mutant in rice. However, the *E3* and *E4* loci exhibit interactions in responses to LD conditions with different R:FR ratios, which are similar to those observed between phyA and phyB or phyC in rice (TAKANO *et al.* 2005). Namely, the *E3* locus is responsible for the flowering under R-enriched LD conditions, whereas the *E4* locus is involved in floral induction under FR-enriched LD conditions. As a consequence, the *e3e3* recessive homozygote can initiate flowering under the R-enriched LD conditions, but the *e4* allele is necessary for plants homozygous for the *e3* allele to flower under the FR-enriched LD conditions, while the *e4* allele cannot singly confer insensitivity to either of the R- and FR-enriched LD conditions (BUZZELL and VOLDENG 1980; SAINDON *et al.* 1989; COBER *et al.* 1996). The different roles of the *E3* and *E4* loci in flowering under LD conditions with different R:FR ratios and the epistasis of the *e3* allele to the *E4* allele imply that the relationship between the *E3* and *E4* loci in soybean may be equivalent to the relationship between *phyA* and *phyB* or *phyC* in rice. As in other species, the *phyA* gene is thus involved in the control of the photoperiod responses in soybean. Together with the *e3* allele, the *e4* allele most likely conditions the photoperiod insensitivity in soybean.

The function of the *E3* locus has not been determined. The de-etiolation responses of the NILs for *e3* to R light were similar to those of Harosoy (*E3E3*) in both the *E4* and *e4* backgrounds (Figure 7, B and C), suggesting that the *E3* locus was not involved in the control of the de-etiolation responses under R light. This is inconsistent with the function of phyB in de-etiolation responses as would be expected if the *E3* gene encodes phyB. It is thus likely that the *E3* locus encodes one of the phytochrome species other than phyB.

Functional redundancy in phyA enables adaptation to long-day-length environments: The functions of *GmphyA1* in seedling de-etiolation and flowering remain undetermined. None of the classical major genes

controlling flowering and maturing times that have been mapped so far is located in LG O. Many QTL influencing the flowering and maturing times have been allocated to various LGs (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; Watanabe *et al.* 2004; Zhang *et al.* 2004; Funatsuki *et al.* 2005; Liu *et al.* 2007). However, only a maturity-controlling QTL has been located in LG O (Zhang *et al.* 2004). The tagging markers of the QTL (Satt592 and Satt581) are located 43 to 49 cM distant from Satt262, the closest marker to *GmphyA1*, in the soybean consensus map (Song *et al.* 2004). Therefore, no major genes and QTL for flowering and maturity thus far reported are present in the regions flanking *GmphyA1*.

De-etiolation response investigation revealed an impaired de-etiolation in plants homozygous for the *e4* allele under FR light. The FR-light-grown NILs for *e4* produced longer hypocotyls than the R-light-grown NILs, but their hypocotyl lengths were still shorter than the NILs grown in complete darkness (Figure 7). The phyA function of the *e4* allele therefore was lost partially, not completely. Such an intermediate phenotype suggests a genetic redundancy in the de-etiolation responses under FR light. This is in contrast to a complete loss of the de-etiolation response under the continuous FR-light condition that is observed in the *phyA* mutants of Arabidopsis, pea, and rice, in which the *phyA* gene is present as a single-copy gene (Weller *et al.* 1997, 2001; Neff and Chory 1998; Takano *et al.* 2001, 2005). It is likely that the genetic redundancy in soybean can be attributed to the presence of multiple copies of the *phyA* gene. Both the protein sequence and the gene expression patterns are highly conserved between the *GmphyA1* and *GmphyA2* genes, which suggests that at least GmphyA1 has a function similar to that of GmphyA2 and is involved in this phenomenon.

The genetic redundancy in the de-etiolation responses resembles flowering behaviors previously recognized under FR-enriched LD conditions. Cober and Voldeng (2001b) found that plants homozygous for the recessive alleles at all of the four photoperiod-sensitive loci, *E1*, *E3*, *E4*, and *E7*, still exhibit delayed flowering under LD conditions with a low R:FR ratio compared with those grown under SD conditions, suggesting the presence of an unknown locus influencing the response. Since the *E1* and *E7* loci are closely linked to each other in LG C2 (Cober and Voldeng 2001a), and the *E3* and *E4* loci are located in LGs L and I, respectively (Abe *et al.* 2003; Molnar *et al.* 2003), the redundancy in the flowering of the recessive genotype *e1e1e3e3e4e4e7e7* under the FR-enriched LD conditions may be attributable to the *phyA* homologs in the other LGs, such as *GmphyA1* in LG O, or to the *phyA* genes, whose loci have not yet been identified.

The functional redundancy may be important in the generation of novel phenotypes, particularly for genes that are involved in the regulation of various developmental processes, such as *phyA* genes. In Arabidopsis, phyA inhibits the growth extension under FR enrichment that is conditioned by other phytochromes to be shade-avoidance responses and consequently prevents excessive elongation (Smith and Whitelam 1990, 1997). Mutants deficient in phyA functions therefore display extreme hypocotyl elongation and die when grown in the field under dense vegetational shade, an FR-enriched environment (Yanovsky *et al.* 1995). However, the *e4* allele, a naturally occurring loss-of-function allele, may have been retained in the soybean gene pool, possibly because it effects only a subtle change under the functional redundancy afforded by other *phyA* homeologs, which buffers its deleterious effects on seedling growth. Only when combined with the *e3* allele does it condition photoperiod insensitivity, which enables soybean to adapt to long day length in high-latitude environments.

This study indicated that both *GmphyA1* and *GmphyA2* were expressed at a considerable level in the *E4* homozygous plants whereas only *GmphyA1* was expressed in the *e4* homozygous plants. It is tempting to speculate that multiple copies of the *phyA* gene, including *GmphyA1* and *GmphyA2*, have been retained in the soybean genome via subfunctionalization, although no obvious difference in transcription pattern was detected between these two genes. In this scenario, lack of certain function(s) specifically conferred by the protein encoded by *GmphyA2* in the *e4* homozygous plants may explain the phenotypic difference between the *E4* and *e4* homozygous plants. Alternatively, it is also possible that the complete phyA function in soybean is composed of the sum of the activities conferred by multiple *phyA* genes whose function has not diverged. A difference in the total expression level of the *phyA* genes between the *E4* and *e4* homozygous plants is consistent with this notion. Dissection of the roles of *GmphyA1*, *GmphyA2*, and other *phyA* paralogs in various photo-environment-dependent plant growth situations should reveal the relevance of these hypotheses as well as the adaptive strategy of soybean in diverse cultivation environments.

We thank Y. Kishima for valuable suggestions for our study and E. R. Cober and R. L. Nelson for supplying the seeds of the Harosoy isolines for maturity genes. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

LITERATURE CITED

Abe, J., D. H. Xu, A. Miyano, K. Komatsu, A. Kanazawa *et al.*, 2003 Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. Crop Sci. **43:** 1300–1304.

Adams, K. L., and J. F. Wendel, 2005 Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. **8:** 135–141.

Aukerman, M. J., M. Hirschfeld, L. Wester, M. Weaver, T. Clark *et al.*, 1997 A deletion in the PHYD gene of the *Arabidopsis* Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. Plant Cell **9:** 1317–1326.

Basu, D., K. Dehesh, H. Schneider-Poetsch, S. E. Harrington, S. R. McCouch *et al.*, 2000 Rice *PHYC* gene: structure, expression, map position and evolution. Plant Mol. Biol. **44:** 27–42.

Buzzell, R. I., 1971 Inheritance of a soybean flowering response to fluorescent-daylength conditions. Can. J. Genet. Cytol. **13:** 703–707.

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

Carter, C. E., A. Szmidt-Jaworska, M. Hughes, B. Thomas and S. Jackson, 2000 Phytochrome regulation of phytochrome A mRNA levels in the model short-day-plant *Pharbitis nil*. J. Exp. Bot. **51:** 703–711.

Casal, J. J., R. A. Sanchez and M. J. Yanovsky, 1997 The function of phytochrome A. Plant Cell Environ. **20:** 813–819.

Chapman, A., V. R. Pantalone, A. Ustun, F. L. Allen, D. Landau-Ellis *et al.*, 2003 Quantitative trait loci for agronomic and seed quality traits in an F_2 and $F_{4:6}$ soybean population. Euphytica **129:** 387–393.

Childs, K. L., F. R. Miller, M. M. Cordonnier-Pratt, L. H. Pratt, P. W. Morgan *et al.*, 1997 The sorghum photoperiod sensitivity gene, Ma_3, encodes a phytochrome B. Plant Physiol. **113:** 611–619.

Clark, T., S. Mathews and R. A. Sharrock, 1994 The phytochrome apoprotein family in *Arabidopsis* is encoded by five genes: the sequences and expression of *PHYD* and *PHYE*. Plant Mol. Biol. **25:** 413–427.

Cober, E. R., and H. D. Voldeng, 2001a A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. Crop Sci. **41:** 698–701.

Cober, E. R., and H. D. Voldeng, 2001b Low R:FR light quality delays flowering of *E7E7* soybean lines. Crop Sci. **41:** 1823–1826.

Cober, E. R., J. W. Tanner and H. D. Voldeng, 1996 Soybean photoperiod-sensitivity loci respond differentially to light quality. Crop Sci. **36:** 606–610.

Devlin, P. F., K. J. Halliday, N. P. Harberd and G. C. Whitelam, 1996 The rosette habit of *Arabidopsis thaliana* is dependent upon phytochrome action: novel phytochromes control internode elongation and flowering time. Plant J. **10:** 1127–1134.

Devlin, P. F., S. R. Patel and G. C. Whitelam, 1999a Phytochrome E influences internode elongation and flowering time in *Arabidopsis*. Plant Cell **10:** 1479–1488.

Devlin, P. F., P. R. Robson, S. R. Patel, L. Goosey, R. A. Sharrock *et al.*, 1999b Phytochrome D acts in the shade-avoidance syndrome in *Arabidopsis* by controlling elongation growth and flowering time. Plant Physiol. **119:** 909–915.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y.-I. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545.

Funatsuki, H., K. Kawaguchi, S. Matsuba, Y. Sato and M. Ishimoto, 2005 Mapping of QTL associated with chilling tolerance during reproductive growth in soybean. Theor. Appl. Genet. **111:** 851–861.

Goldblatt, P, 1981 Cytology and the phylogeny of Leguminosae, pp. 427–463 in *Advances in Legume Systematics*, Part 2, edited by R. M. Polhill and P. H. Raven. Royal Botanic Gardens, Kew, United Kingdom.

Goto, N., T. Kumagai and M. Koornneef, 1991 Flowering responses to light-breaks in photomorphogenic mutants of *Arabidopsis thaliana*, a long-day plant. Physiol. Plant. **83:** 209–215.

Hauser, B. A., M. Cordonnier-Pratt and L. H. Pratt, 1998 Temporal and photoregulated expression of five tomato phytochrome genes. Plant J. **14:** 431–439.

Hecht, V., F. Foucher, C. Ferrandiz, R. Macknight, C. Navarro *et al.*, 2005 Conservation of *Arabidopsis* flowering genes in model legumes. Plant Physiol. **137:** 1420–1434.

Hughes, A. L., 1994 The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. Lond. B Biol. Sci. **256:** 119–124.

Hymowitz, T., 2004 Speciation and cytogenetics, pp. 97–136 in *Soybeans: Improvement, Production, and Uses* (Agronomy monograph no. 16), Ed. 3, edited by H. R. Boerma and J. E. Specht. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI.

Johnson, E., J. M. Bradley, N. P. Harberd and G. C. Whitelam, 1994 Photoresponses of light-grown *phyA* mutants of *Arabidopsis*: phytochrome A is required for the perception of daylength extensions. Plant Physiol. **105:** 141–149.

Keim, P., B. W. Diers, T. C. Olson and R. C. Shoemaker, 1990 RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics **126:** 735–742.

Koseki, M., K. Goto, C. Masuta and A. Kanazawa, 2005 The star-type color pattern in *Petunia hybrida* 'Red Star' flowers is induced by the sequence-specific degradation of the chalcone synthase RNA. Plant Cell Physiol. **46:** 1879–1883.

Kumar, A., and J. L. Bennetzen, 1999 Plant retrotransposons. Annu. Rev. Genet. **33:** 479–532.

Lackey, J. A., 1980 Chromosome numbers in the Phaseoleae (Fabacceae: Faboideae) and their relation to taxonomy. Am. J. Bot. **67:** 595–602.

Lee, J. M., A. L. Bush, J. E. Specht and R. C. Shoemaker, 1999 Mapping of duplicate genes in soybean. Genome **42:** 829–836.

Lee, S. H., M. A. Bailey, M. A. R. Mian, E. R. Shipe, D. A. Ashley *et al.*, 1996 Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor. Appl. Genet. **92:** 516–523.

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

Liu, B., T. Fujita, Z. H. Yan, S. Sakamoto, D. Xu *et al.*, 2007 QTL mapping of domestication related traits in soybean (*Glycine max*). Ann. Bot. **100:** 1027–1038.

Lohnes, D. G., J. E. Specht and P. B. Cregan, 1997 Evidence for homoeologous linkage groups in the soybean. Crop Sci. **37:** 254–257.

Lynch, M., and A. Force, 2000 The probability of duplicate gene preservation by subfunctionalization. Genetics **154:** 459–473.

Mansur, L. M., K. G. Lark, H. Kross and A. Oliveira, 1993 Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). Theor. Appl. Genet. **86:** 907–913.

Maquat, L. E., 2004 Nonsense-mediated mRNA decay: splicing, translation and mRNA dynamics. Nat. Rev. Mol. Cell Biol. **5:** 89–99.

Mockler, T. C., H. Guo, H. Yang, H. Duong and C. Lin, 1999 Antagonistic actions of *Arabidopsis* cryptochromes and phytochrome B in the regulation of floral induction. Development **126:** 2073–2082.

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

Moore, C., and M. D. Purugganan, 2005 The evolutionary dynamics of plant duplicate genes. Curr. Opin. Plant Biol. **8:** 122–128.

Nagamatsu, A., C. Masuta, M. Senda, H. Matsuura, A. Kasai *et al.*, 2007 Functional analysis of soybean genes involved in flavonoid biosynthesis by virus-induced gene silencing. Plant Biotechnol. J. **5:** 778–790.

Napoli, C., C. Lemieux and R. Jorgensen, 1990 Introduction of a chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes in trans. Plant Cell **2:** 279–289.

Neff, M. M., and J. Chory, 1998 Genetic interactions between phytochrome A, phytochrome B, and cryptochrome 1 during *Arabidopsis* development. Plant Physiol. **118:** 27–36.

Orf, J. H., K. Chase, F. R. Adler, L. M. Mansur and K. G. Lark, 1999 Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Sci. **39:** 1652–1657.

Reed, J. W., A. Nagatani, T. D. Elich, M. Fagan and J. Chory, 1994 Phytochrome A and phytochrome B have overlapping but distinct functions in *Arabidopsis* development. Plant Physiol. **104:** 1139–1149.

Saindon, G., H. D. Voldeng, W. D. Beversdorf and R. I. Buzzell, 1989 Genetic control of long daylength response in soybean. Crop Sci. **29:** 1436–1439.

Schlueter, J. A., P. Dixon, C. Granger, D. Grant, L. Clark *et al.*, 2004 Mining EST databases to resolve evolutionary events in major crop species. Genome **47:** 868–876.

Shoemaker, R. C., K. Polzin, J. Labate, J. E. Specht, E. C. Brummer *et al.*, 1996 Genome duplication in soybean (*Glycine* subgenus *Soja*). Genetics **144:** 329–338.

Shoemaker, R. C., J. A. Schlueter and J. F. Doyle, 2006 Paleopolyploidy and gene duplication in soybean and other legumes. Curr. Opin. Plant Biol. **9:** 104–109.

Smith, H., and G. C. Whitelam, 1990 Phytochrome, a family of photoreceptors with multiple physiological roles. Plant Cell Environ. **13:** 695–707.

Smith, H., and G. C. Whitelam, 1997 The shade avoidance syndrome: multiple responses mediated by multiple phytochromes. Plant Cell Environ. **20:** 840–844.

Song, Q. J., L. F. Marek, R. C. Shoemaker, K. G. Lark, V. C. Concibido *et al.*, 2004 A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. **109:** 122–128.

Takano, M., H. Kanegae, T. Shinomura, A. Miyano, H. Hirochika *et al.*, 2001 Isolation and characterization of rice phytochrome A mutants. Plant Cell **13:** 521–534.

Takano, M., N. Inagaki, X. Xie, N. Yuzurihara, F. Hihara *et al.*, 2005 Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. Plant Cell **17:** 3311–3325.

Tasma, I. M., L. L. Lorenzen, D. E. Green and R. C. Shoemaker, 2001 Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. Mol. Breed. **8:** 25–35.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Vision, T. J., D. G. Brown and S. D. Tanksley, 2000 The origins of genomic duplications in *Arabidopsis*. Science **290:** 2114–2117.

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

Weller, J. L., I. C. Murfet and J. B. Reid, 1997 Pea mutants with reduced sensitivity to far-red light define an important role for phytochrome A in day-length detection. Plant Physiol. **114:** 1225–1236.

Weller, J. L., N. Beauchamp, L. H. J. Kerckhoffs, J. D. Platten and J. B. Reid, 2001 Interaction of phytochromes A and B in the control of de-etiolation and flowering in pea. Plant J. **26:** 283–294.

Yamanaka, N., S. Ninomiya, M. Hoshi, Y. Tsubokura, M. Yano *et al.*, 2001 An informative linkage map of soybean reveals QTLs for flowering time, leaf morphology and region of segregation distortion. DNA Res. **8:** 61–72.

Yanovsky, M., J. J. Casal and G. C. Whitelam, 1995 Phytochrome A, phytochrome B and HY4 are involved in hypocotyl growth responses to natural radiation in *Arabidopsis*: weak de-etiolation of the *phyA* mutant under dense canopies. Plant Cell Environ. **18:** 788–794.

Zhang, W. K., Y. J. Wang, G. Z. Luo, J. S. Zhang, C. Y. He *et al.*, 2004 QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor. Appl. Genet. **108:** 1131–1139.

Zhu, H., H. K. Choi, D. R. Cook and R. C. Shoemaker, 2005 Bridging model and crop legumes through comparative genomics. Plant Physiol. **137:** 1189–1196.

Zhu, T., J. M. Schupp, A. Oliphant and P. Keim, 1994 Hypomethylated sequences: characterization of the duplicated soybean genome. Mol. Gen. Genet. **244:** 638–645.

Communicating editor: A. H. Paterson