

Genetic Diversity and Its Use in Soybean

Abe Jun

Hokkaido University

Introduction

Cultivated and wild soybeans form a primary gene pool (Hymowitz, 2004). Both usually produce fertile hybrids, and there is no marked obstacle to gene exchange, although the sterility due to abnormal chromosome pairing has often been observed. Molecular assays have revealed that genetic variability of the cultivated soybean is low compared to that of the wild soybean (Xu *et al.* 2002; Hyten *et al.* 2006). As suggested by the results of chloroplast and mitochondrial genomes analyses, however, the introgressions from the wild to the cultivated germplasm most likely has occurred repeatedly in various regions of East Asia in the past (Xu *et al.* 2002). This finding may predict that the cultivated soybean possesses a relatively rich genetic variability in contrast to a general trend that the diversity in cultivated crops that is limited by the bottle-neck that occurred in the domestication process (Tanksley and McCouch, 1996). The most easily-accessible and useful genetic resources for soybean breeding may be landraces/local varieties that have been established in various regions of Asia. Here, I present an example that shows diverse genetic variability embedded in a single locus controlling important adaptative traits in soybean.

Genetic base of flowering under long days in soybean

Soybean is cultivated in a wide range of latitudes from the equator to 50° north. 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 plant, and soybean cultivars adapted to high latitude environments possess insensitivity to photoperiod. 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; 2001b). 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 when the day length is extended to 20 h using fluorescent lamps with a high R:FR ratio (Buzzell 1971). The *E4* locus was identified by extending the natural daylength to 20 h with incandescent lamps with a low R:FR ratio (Buzzell and Voldeng 1980). A recessive allele at the *E4* locus can not simply confer 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* background (Cober *et al.* 1996; Cober and Voldeng 2001b). Using near isogenic lines (NILs) for those maturity genes, Cober *et al.* (1996) found different responses to each photoperiod-sensitivity gene to 20-h LD conditions with different R:FR ratios, suggesting that some of these genes may belong to the phytochrome family.

Classification of ILD-insensitive cultivars based on isozyme and SSR markers

We evaluated the responses of soybean accessions introduced from various countries of East Asia to incandescent-induced long daylength (ILD) where the natural day length was extended to 20 h with incandescent lamps. We defined the ability of plants to initiate flowering under ILD without any delay relative to natural daylength condition as ILD-insensitive. A total of 29 accessions were determined as ILD-insensitive. They were further classified into several groups, based on the result of UPGMA for genotypes at 7 isozyme and 9 SSR markers. There are three clearly-separated groups (Fig.1). The first group (designated group I) mostly consists of the landraces that are adapted to the cool summer of northern and northeastern Hokkaido of Japan and Sakhalin, where the frost-free season is limited to less than 130 d. The second group (designated group II) consists of the landraces that have been cultivated as a short-season crop in wide areas of Japan and

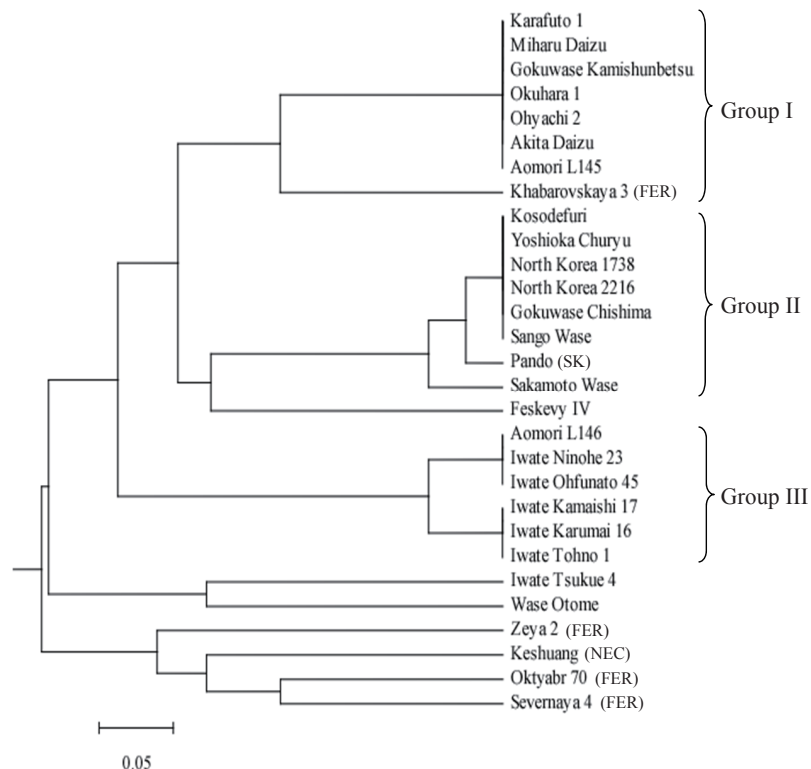


Fig. 1. Classification of ILD insensitive soybean accessions in East Asia, based on genotypes at seven isozyme and nine SSR markers. NEC; Northeast China, FER; Far-east Russia, SK; South Korea (Abe *et al.* unpublished data)

the Korean peninsula. The third group (designated group III) consists of the landraces collected in Iwate Prefecture of northern Japan. All of the accessions within each group had almost identical genotype. Another two accessions collected in Iwate Prefecture formed a different clade, which are connected to the above three groups. The accessions from North-East China and Far-East Russia are loosely related to each other, and form the fifth group.

The identity of the genetic mechanisms underlying the ILD insensitivity was tested between the cultivar groups I and II. The F_2 progeny of the cross between Miharudaizu (group I) and Sakamotowase (group II) exhibited transgressive segregation toward ILD-insensitive late flowering, suggesting that both possessed different genetic systems (Abe *et al.*, 1998). From the genetic analyses for NILs for the ILD insensitivity, Abe *et al.* (2003) determined the genotypes at the three loci, *E1*, *E3* and *E4*, of Miharudaizu and Sakamotowase as *E1E1e3e3e4e4* and *e1e1e3E4E4*, respectively. The genotype of Miharudaizu was therefore the same as the double recessive genotype for *E3* and *E4*, which was determined in previous studies to condition ILD-insensitivity (Buzzell, 1971; Buzzell and Voldeng, 1980). However, the genotype of Sakamotowase was the same as a Harosoy NIL for the *e3* allele, which develops

no flower buds under the ILD treatment. Therefore, a novel gene(s) may be needed for Sakamotowase to initiate flowering under ILD. The test crosses with the Harosoy isolate were then carried out to identify the novel gene for the photoperiod-insensitivity of Sakamotowase. By marker-assisted analyses, we detected a major QTL for insensitivity near a SSR marker in linkage group C2 and a minor QTL in linkage group L. It was estimated from the position of tagging marker that the novel gene may be an allele at the *E1* locus or a gene tightly linked to the *E1* locus (Liu and Abe, 2009).

Identification of the maturity gene *E4* as a paralog of *phytochrome A*

As suggested by Cober *et al.* (1996), the *E4* gene appears to encode phytochrome A protein. We isolated two *phyA* genes designated as *GmPHYA1* and *GmPHYA2* from the soybean genome to test the association between *phyA* genes and soybean maturity genes. Analysis of the *GmPHYA2* gene from the photoperiod insensitive lines carrying the recessive allele *e4* revealed an insertion of a retrotransposon in exon 1 of the gene, which resulted in dysfunction of the gene (Fig.2A and 2B; Liu *et al.* 2008). Genetic mapping allocated *GmPHYA1* and *GmPHYA2* into homoeologous regions of linkage groups O and I,

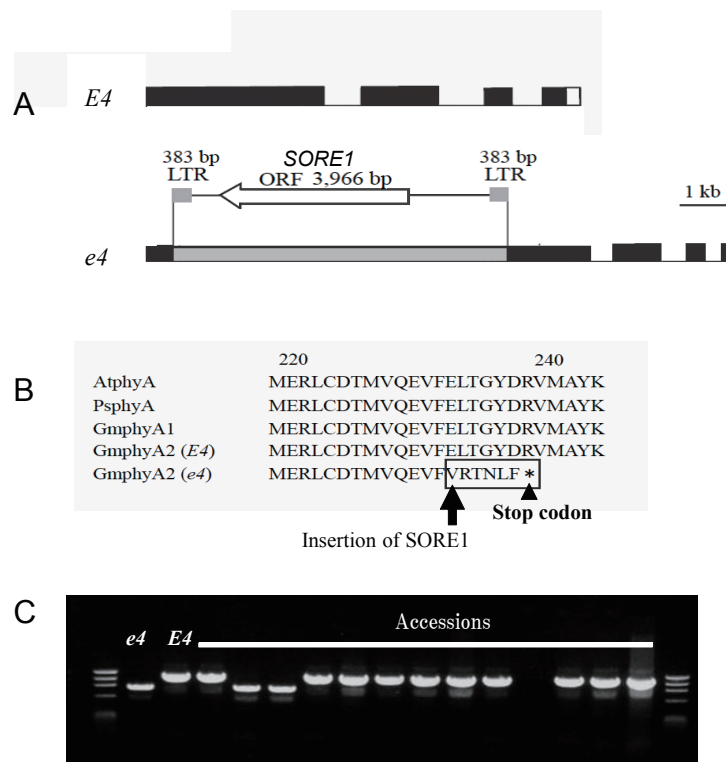


Fig. 2. The dysfunctional allele at the *E4* locus and PCR amplification using allele-specific primers. A) the *e4* allele possesses a TY1/copia-like retrotransposon, SORE1, in the first exon of GmPHYA2. B) Insertion of SORE1 that causes a premature stop codon at the *e4* allele. C) Identification of accessions carrying the *e4* allele by PCR. (Liu *et al.* 2008)

respectively, suggesting that *GmPHYA1* and *GmPHYA2* are homoeologs which resulted from ancient chromosomal duplications and rearrangements in soybean. Of these, *GmPHYA2* was mapped at the position corresponding to the *E4* locus previously mapped (Abe *et al.* 2003), and cosegregated with the ILD insensitivity. Taken together, the *E4* locus encodes the *GmPHYA2* protein, and the dysfunctional allele due to the insertion of retrotransposon conditions the ILD-insensitivity.

The phytochrome A protein 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 *et al.* 2001; 2005). *phyA* and *phyB* are involved in de-etiolation of seedlings under FR-light and R-light conditions, respectively. The *phyA* mutants of Arabidopsis, pea and rice show a complete loss of the de-etiolation response under the continuous FR-light condition. 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* al-

lele 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 (Fig. 3A; Liu *et al.* 2008). The similar elongated internodes are also observed in higher nodes of plants (Fig. 3B). However, the *phyA* function of the *e4* allele was lost partially, not completely, possibly due to the presence of another copy, *GmPHYA1*. The presence of duplicated copies of *phyA* genes may account for the generation of photoperiod insensitivity, while protecting against the deleterious effects of mutation (Liu *et al.* 2008). 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; Neff and Chory 1998; Takano *et al.* 2001; Weller *et al.* 2001; Takano *et al.* 2005).

Distribution of soybean cultivars carrying the SORE-1-inserted GmPHYA2 allele

Inactivation of *GmPHYA2*, which constitutes the *e4* allele that confers insensitivity to a long-day condition, is caused by the insertion of *SORE-1* in exon 1 of the gene (Liu *et al.* 2008; Kanazawa *et al.* 2009). Based on these findings, we hypothesized that

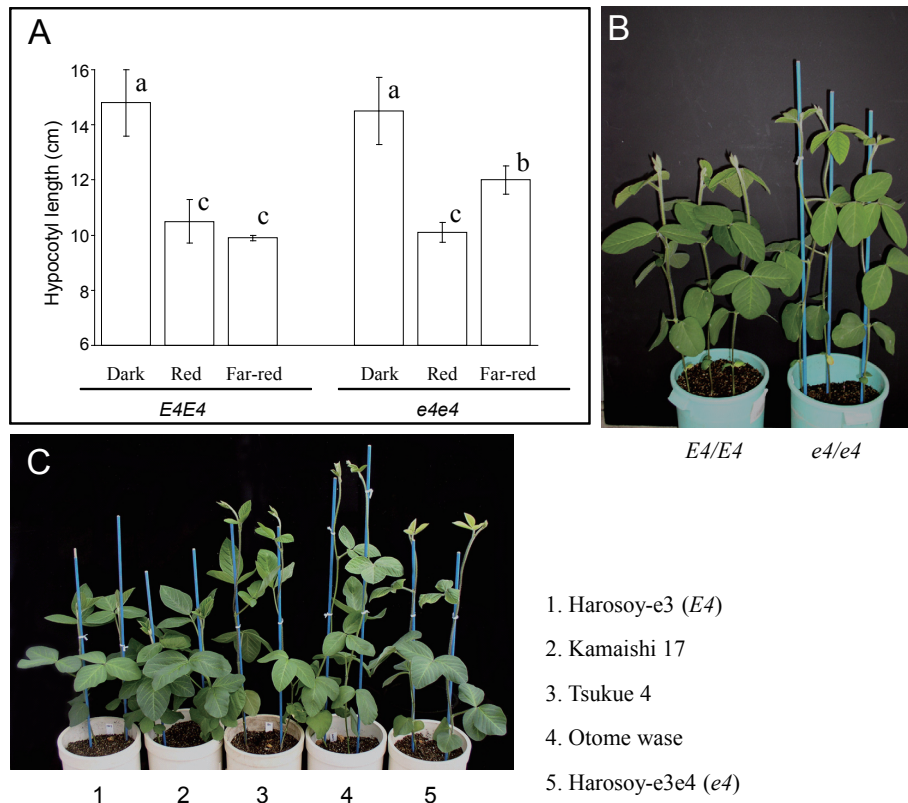


Fig.3. *E4* and *e4* alleles exhibit different photomorphogenesis under long day.

- A) De-etiolation response under dark, red light and far-red light, Different alphabets indicate statistically significant differences at a level of 5 %
- B) Plant morphology of Harosoy isolines for *E4* and *e4*
- C) Plant morphology of various ILD-insensitive soybean accessions carrying different dysfunctional alleles.(Liu *et al.* 2008 Abe *et al.* unpublished data)

the insertion of *SORE-1* in exon 1 of *GmPHYA2* is one of the major genetic changes that allowed distribution of soybean in high latitude regions. To test this hypothesis, we analyzed the presence or absence of *SORE-1* at this locus in various cultivated and wild soybean accessions.

PCR that amplifies a region encompassing a portion of *SORE-1* and its flanking genomic region (Fig.2C) was performed using DNA isolated from 332 cultivated soybean accessions introduced from various East Asian countries that cover a wide range of latitude and include regions where cultivated soybean originated (Kanazawa *et al.* 2009). We also analyzed 85 wild soybean (*ssp. soja*) accessions that were collected from natural populations in various regions of Japan. While no plants that harbor *SORE-1* at the locus were found in wild soybean lines examined, the *SORE-1* insertion at the locus was detected in 10 accessions of cultivated soybean, nine of which were distributed in Hokkaido, Japan (Fig. 4). Most of the nine accessions belong to the group I (Fig.1). A remaining accession (Col/Aomori/1981/L145) is a collection from Aomori Prefecture, the nearest prefecture to Hokkaido Island. All these ac-

cessions had an ILD-insensitive and early-maturing habit. In addition, the historical record indicates that a local variety named ‘Ohyachi’, an introduction by an immigrant from the North East region of Japan, enabled the expansion of soybean cultivation into inland, northern and eastern areas of Hokkaido with harsher environments in the late 19 century, where different named landraces, as those of group I, had been established. These results are consistent with the notion that disruption of *GmPHYA2* by the insertion of *SORE-1* contributed to the expansion of cultivated region of soybean toward higher latitude regions (Kanazawa *et al.* 2009).

Independent mutations that result in dysfunctional alleles at the *E4* loci

It remains unsolved what mechanisms are involved in the genetic controls for ILD insensitivity for other accessions. We determined the sequences of two *phyA* genes for representative accessions of groups. No non-synonymous mutation was detected for *GmPHYA1* among the accessions sequenced. On the other hand, sequencing the *GmPHYA2* gene for these accessions revealed four additional recessive

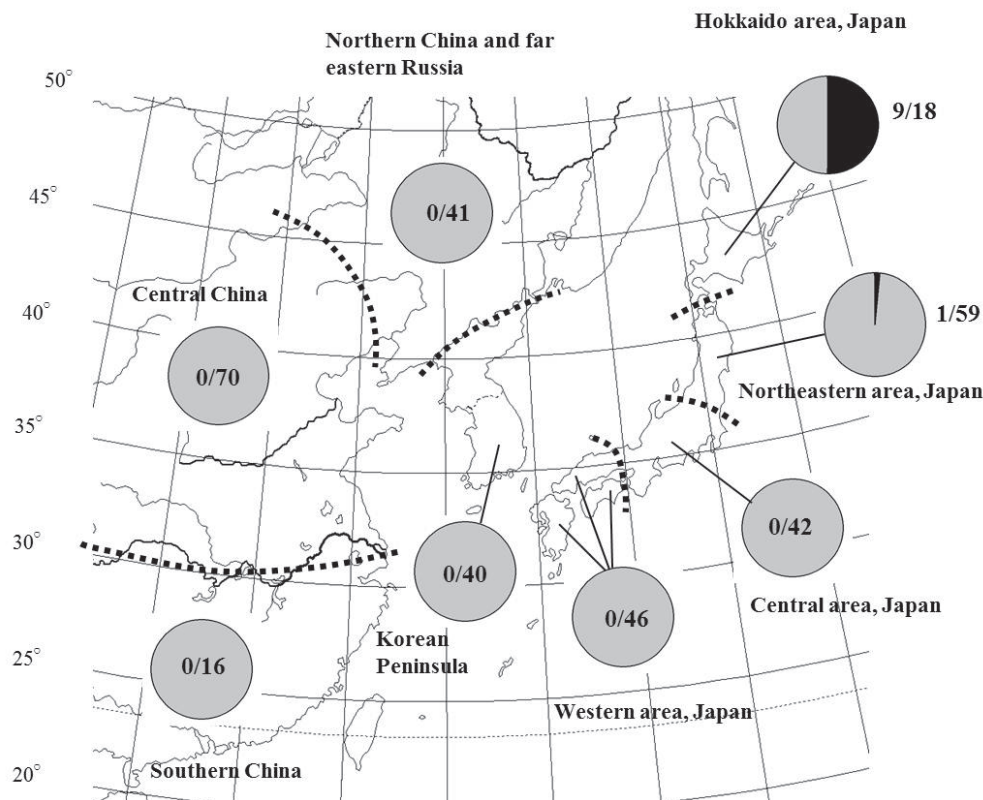


Fig. 4. Geographical distribution of the *e4* allele in cultivated soybeans in East Asia. Figures indicate the numbers of accessions carrying the *e4* allele per those of total accessions tested in each region (Kanazawa *et al.* 2009).

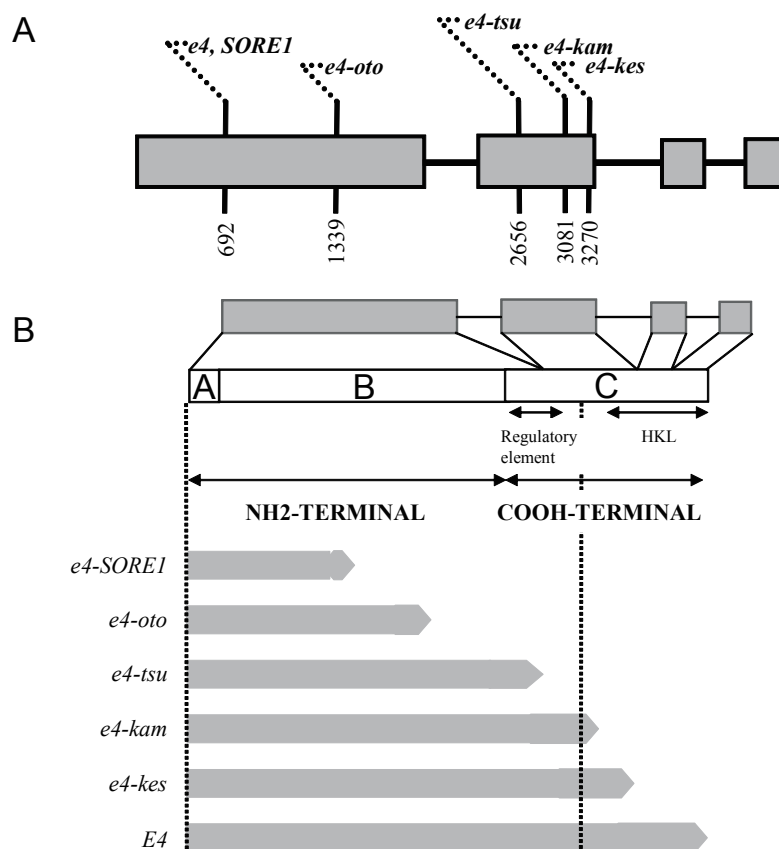


Fig. 5. Independent mutations at the *E4* locus that produced premature stop codons and truncated proteins in the ILD-insensitive accessions.

A) Positions of single-base or two-base deletions in the exons 1 and 2 of *GmPHYA2* gene.

B) Domain structure of phytochrome A gene and the deduced amino acid regions translated from the mutant alleles (Abe *et al.* unpublished data).

alleles, *e4-kes* (in Keshung from northeast China), *e4-kam*, *e4-tsuk* and *e4-oto* (all from Iwate Prefecture), all of which exhibited single-base or two-base deletions in the first and second exons, resulting in premature truncated proteins (Fig. 5A). PCR analyses using allele-specific primers further revealed that all of the ILD-insensitive accessions in group III carried the *e4-kam* allele, and three accessions from Far-east Russia carried the *e4-kes* allele as did the Chinese accession Keshung. Therefore, all of the ILD accessions except for the accessions of group II possessed dysfunctional alleles at the *E4* locus. These accessions possessed the recessive alleles at both the *E3* and *E4* loci to condition the ILD insensitivity. A swapping experiment for the domains of *phyA* and *phyB* indicated that the NH₂-terminal domain and core region for regulatory activity possessed important functions for de-etiolation responses for both *Arabidopsis* phytochrome A and B genes (Quail, 1997). The deduced amino acids differ in length among the alleles detected at the *E4* locus (Fig. 5B). In particular, the *e4-kes* and *e4-kam* alleles are expected to produce amino acids that cover the regions important for the de-etiolation function. In fact, the accession Kamaishi-17 having the *e4-kam* allele showed almost the same internode growths as plants having the *E4* allele under LD (Fig. 3C). Different *GmPHYA2* alleles may regulate responses of soybeans to photoenvironments, such as photoperiod-sensitivity and photomorphogenesis, differently. The alleles from Kamaishi-17 and Keshung at the *E4* locus might therefore make it possible for the soybean cultivars to flower under long day without defective morphogenesis.

Concluding remarks

Soybean is considered to be a paleopolyploid species with a complex genome (reviewed by Shoemaker *et al.* 2006). Most of genes possess their homoeologous partners, like the two phytochrome A genes, *GmPHYA1* and *GmPHYA2*. The genetic redundancy may have generated genetic materials through the accumulation of mutations, which have been subjected to selections by farmers. The duplicated genes may have a potential to acquire a novel function via sub-functionalization and/or neo-functionalization of genes and to adjust the phenotype to the appropriate levels by the change of gene dosages as caused by the null allele at either of the two loci. Molecular dissection, together with phenotypic evaluation of diverse genetic resources for

targeted traits and their detailed characterization, may enable us to construct a new breeding strategy for improvements of yield, quality and adaptation in soybean.

References

- Abe J, Komatsu K, Shimamoto Y, 1998. A new gene for insensitivity of flowering to incandescent long daylength. *Soybean Genet Newsl.* 25: 92.
- Abe J, Xu D, Miyano A, Komatsu K, Kanazawa A, Shimamoto Y, 2003. Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. *Crop Sci.* 43:1300-1304.
- Buzzell RI, 1971. Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Can J Cytol.* 13: 703-707.
- Buzzell RI, Voldeng HD, 1980. Inheritance of insensitivity to long daylength. *Soybean Genet Newsl.* 7: 26-29.
- Casal JJ, Sanchez RA, Yanovsky MJ, 1997. The function of phytochrome A. *Plant Cell Environ.* 20: 813-819.
- Cober ER, Tanner JM, Voldeng HD, 1996a. Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. *Crop Sci.* 36: 601-605.
- Cober ER, Tanner JM, Voldeng HD, 1996b. Soybean photoperiod-sensitivity loci respond differentially to light quality. *Crop Sci.* 36: 606-610.
- Cober ER, Voldeng HD, 2001a. A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. *Crop Sci.* 41: 698-701.
- Cober ER, Voldeng HD, 2001b. Low R:FR light quality delays flowering of *E7E7* soybean lines. *Crop Sci.* 41: 1823-1826.
- Hymowitz, T, 2004. Speciation and cytogenetics, pp.97-136 in *Soybeans: Improvement, Production, and Uses*, Ed 3, edited by Boerma, HR., and Specht, JE. Agronomy Monograph No.16. American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc. Publishers, Madison, USA.
- Hyten, DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB, 2006. Impacts of genetic bottleneck on soybean genome diversity. *PNAS*, 103: 16666 - 16687.
- Kanazawa A, Liu B, Kong F, Arase S, Abe J, 2009. Adaptive evolution involving gene duplication and insertion of a novel Ty1/copea-like retro transposon in soybean. *Journal of Molecular*

- Evolution, 69: 164-175.
- Liu B, Abe J, 2010. QTL Mapping for Photoperiod-Insensitivity of a Japanese Soybean Landrace Sakamotowase. *Journal of Heredity*,
- Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, Abe J, 2008. Genetic redundancy in soybean photoresponses associated with duplication of phytochrome A gene. *Genet.* 180: 996-1007.
- Neff MM, Chory J, 1998 Genetic interactions between phytochrome A, phytochrome B, and cryptochrome 1 during *Arabidopsis* development. *Plant Physiol.* 118: 27-36.
- Quail PH, 1997. An emerging molecular map of the phytochromes. *Plan Cell Environ* 20: 657-665.
- Saindon G, Voldeng HD, Beversdorf WD, Buzzell RI, 1989. Genetic control of long daylength response in soybean. *Crop Sci.* 29: 1436-1439.
- Shoemaker RC, Schlueter JA, Doyle JF, 2006. Paleopolyploidy and gene duplication in soybean and other legumes. *Curr. Opin. Plant Biol.* 9: 104-109.
- Takano, M, Inagaki, N, Xie, X, Yuzurihara, N, Hirata, F *et al.*, 2001. Isolation and characterization of rice phytochrome A mutants. *Plant Cell* 13: 521-534.
- Takano, M, Inagaki, N, Xie, X, Yuzurihara, N, Hirata, F *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.
- Tanksley SD, McCouch SR, 1997. Seed banks and molecular maps: unlocking genetic potential from the wild, *Science*, 277: 1063 - 1066.
- Weller, JL, Murfet, IC, Reid JB, 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, JL, Beauchamp N, Kerckhoffs, J, Platten, JD, Reid JB, 2001. Interaction of phytochromes A and B in the control of de-etiolation and flowering in pea. *Plant J.* 26: 283-294.
- Xu DH, Abe J, Gai JY, Shimamoto Y, 2002. Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean, *Theor. Appl. Genet.*, 105: 645-653.