Mol Genet Genomics (2016) 291:1561–1571
DOI 10.1007/s00438-016-1183-2

CrossMark

ORIGINAL ARTICLE

Identification of candidate genes for an early-maturing soybean mutant by genome resequencing analysis

Kyung Jun Lee[1,2] · Dong Sub Kim[1] · Jin-Baek Kim[1] · Sung-Hwan Jo[4] · Si-Yong Kang[1] · Hong-Il Choi[1] · Bo-Keun Ha[3]

Received: 19 November 2015 / Accepted: 15 February 2016 / Published online: 31 March 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Flowering is indicative of the transition from vegetative to reproductive phase, a critical event in the life cycle of plants. In this study, we performed whole genome resequencing by Illumina HiSeq to identify changes in flowering genes using an early-flowering phenotype of soybean mutant line Josaengserori (JS) derived from Korean landrace, Seoritae (SR), and we obtained mapped reads of 131,769,690 and 167,669,640 bp in JS and SR, respectively. From the whole genome sequencing results between JS and SR, we identified 332,821 polymorphic SNPs and 65,178 indels, respectively. Among these, 30 flowering genes were in SNPs and 25 were in indels. Among 30 flowering genes detected in SNPs, Glyma02g33040, Glyma06g22650, Glyma10g36600, Glyma13g01290, Glyma14g10530, Glyma16g01980, Glyma17g11040, Glyma18g53690, and Glyma20g29300 were non-synonymous substitutions between JS and SR. Changes in Glyma10g36600 (*GI*), Glya02g33040 (*AGL18*), Glyma17g11040 (*TOC1*), and Glyma14g10530 (*ELF3*) in JS affected the expression of *GmFT2a* and resulted in early flowering. These results provide insight into the regulatory pathways of flowering in soybean mutants and help to improve our knowledge of soybean mutation breeding.

Keywords Early flowering · Gamma-ray · Mutation · Resequencing · Soybean

Introduction

Mutation breeding is an important technique for creating genetic variation that contributes new sources of useful agronomic traits for improvement of crop varieties.

Communicated by R. K. Varshney.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-016-1183-2) contains supplementary material, which is available to authorized users.

✉ Jin-Baek Kim
jbkim74@kaeri.re.kr

✉ Bo-Keun Ha
bkha@jnu.ac.kr

Kyung Jun Lee
lkj5214@korea.kr

Dong Sub Kim
bioplant@kaeri.re.kr

Sung-Hwan Jo
shjo@seeders.co.kr

Si-Yong Kang
sykang@kaeri.re.kr

Hong-Il Choi
hichoi@kaeri.re.kr

1 Radiation Breeding Research Team, Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute, Jeongeup 580-185, Republic of Korea

2 National Agrobiodiversity Center, NAAS, RDA, Jeonju 560-500, Republic of Korea

3 Division of Plant Biotechnology, College of Agriculture and Life Sciences, Chonnam National University, Gwangju 500-757, Republic of Korea

4 Seeders Inc., Daejeon 305-509, Republic of Korea

During the past seventy years, more than 3000 mutant varieties have been released, most of which were induced by gamma-ray irradiation (Ahloowalia et al. 2004). Unlike chemical mutagens that induce point mutations, such as ethyl methane sulfonate (EMS), exposure to gamma-rays induces various types of DNA damage, such as single- or double-strand breaks and base pair substitutions (Wallace 2002; Wu et al. 2006). These broad-spectrum mutants generate many kinds of phenotypes. Gamma-irradiation treatment of soybeans can induce earlier maturity, seed coat color changes, triple null lipoxygenase, alteration of seed storage proteins, and high yield (Lee et al. 2011, 2014; Ha et al. 2014).

The causal mutations responsible for target agronomic traits induced by gamma-irradiation should be searched for direct use or for further cross breeding with marker-assisted selection or marker-assisted backcrossing (Ahloowalia et al. 2004). Map-based cloning has been the most commonly used strategy for identifying mutated genes. However, the multistep process of population development and genetic map construction is laborious and time consuming (Lukowitz et al. 2000; Salvi and Tuberosa 2005). Recently, next-generation sequencing (NGS) technology has allowed the de novo sequencing of crop species and the direct detection of genome-wide DNA polymorphisms within species by resequencing (Varshney et al. 2009). These NGS technologies can also be applied to identify the points of difference between mutant and wild-type genomes, which allows the rapid identification of causal mutations in combination with comparative functional candidate gene approaches (Zhu and Zhao 2007; Schneeberger 2014; Varshney et al. 2014). In rice, Hwang et al. (2015) detected genome-wide DNA polymorphism between an early maturing mutant and wild-type rice by whole genome resequencing. Then, a few hundred mutations that caused structural alterations of genes were further prioritized on the basis of their molecular functions related to flower development, and one putative causal mutation in a gene encoding a leucine-rich repeat receptor-like kinase was identified.

Flowering time is an important agronomic trait influencing crop yield. There are different mechanisms for controlling the timing of flowering in response to photoperiod length, temperature, and other environmental signals (Kim et al. 2009; Xia et al. 2012b). Soybean is classified as a short-day plant because it flowers in response to a short photoperiod (Kim et al. 2012). At present, nine major loci have been reported to control time to flowering and maturity in soybean: *E1* to *E8* and *J* (Xia et al. 2012b). Of these, the *E1*, *E2*, *E3*, and *E4* loci were related to photoperiod sensitivity under various light conditions. The gene for *E1* contains a putative nuclear localization signal and B3 domain and positively regulates flowering repressor *GmFT4*, a homolog of *Flowering Locus T*, and negatively regulates flowering promoters *GmFT2a*/*GmFT5a* (Xia et al. 2012a; Zhai et al. 2014). The gene responsible for *E2* was homologous to the circadian clock-controlled *GIGANTEA* (*GI*) gene in *Arabidopsis* (Watanabe et al. 2011). The *E3* and *E4* genes encode *PHYTOCHROME A* (*PHY A*) homologs *GmPHYA3* and *GmPHYA2*, respectively (Liu et al. 2008; Watanabe et al. 2009). In addition, the recent release of the soybean genome sequence has enabled the identification of 118 genes homologous to *Arabidopsis* flowering time-related genes throughout the soybean genome by comparative genomic analysis (Kim et al. 2012).

In a previous study, we selected an early-flowering soybean mutant line (Song et al. 2010). Here, to identify a casual mutation for the early flowering, we screened genome-wide DNA polymorphisms between the early-flowering mutant and wild-type soybean by whole genome resequencing and evaluated the different expression patterns of mutated genes involved in flower development.

Materials and methods

Plant materials

The early-flowering mutant cultivar *Glycine max* 'Josaengseori' (JS) was developed from the landrace Seoritae (SR) by gamma-ray treatment (250 Gy) at the Korea Atomic Energy Research Institute (KAERI) (Song et al. 2010). SR had a black seed coat and green cotyledon, suitable for mixed cooking with rice. However, it had late flowering (66 days to 50 % flowering) and maturity (164 days after sowing). In mutant cultivar Josaengseori, time to 50 % flowering was 10 days earlier, and time to pod maturity was 34 days earlier, than in wild-type SR (Song et al. 2010). Seeds of JS and SR were planted in 15-cm-diameter pots in a greenhouse, and leaves of a single plant per cultivar were used to isolate total DNA for whole genome sequencing. The leaves of JS and SR were harvested at the time to 50 % flowering (R1–R2) to analyze the expression levels of flowering genes.

DNA library construction and massively parallel sequencing

Genomic DNA was extracted from the fresh leaves according to the procedure described by Kim et al. (2010). The purified whole genomic DNA was randomly sheared by Covaris S2 (Covaris, Woburn, MA, USA) to yield DNA fragments in the target range of 400–500 bp, and average molecular sizes of the fragments were accessed using an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). Following fragmentation, the resulting overhangs were converted to blunt ends using TruSeq DNA

Sample Preparation Kits v2 (Illumina, CA, USA), followed by a clean-up step using AMPure XP beads (Beckman Coulter Genomics, Danvers, MA, USA). To increase the success of ligation between the fragmented DNA and index adapters and to reduce self-ligation of the blunt fragments, the 3′ ends were adenylated. Immediately following adenylation, the index adapters were ligated to the freshly adenylated, fragmented genomic DNA and were purified using AMPure XP beads. Sample ligation products were size selected on a 2 % agarose gel followed by gel extraction and column purification of the DNA. Successfully ligated DNA fragments that contained adapter sequences were enhanced via PCR using adapter-specific primers. DNA was re-isolated using AMPure XP beads, and average molecular sizes of the libraries were assessed with the Agilent Bioanalyzer 2100 to check for a sharp peak in the expected 500- to 600-bp range. Each library was loaded on the Hiseq 2000 platform (Illumina, CA, USA), and we performed high-throughput sequencing to ensure that each sample met the desired average sequencing depth. Image analysis and base calling were performed using the Illumina pipeline with default settings.

Preprocessing

After sequencing, the samples were demultiplexed and the indexed adapter sequences were trimmed using the Solexa QA package v.1.13 (Cox et al. 2010). It is common for the quality of bases from either end of Illumina reads to drop in quality; we, therefore, trimmed either end of the reads when the Phred quality score dropped below $Q = 20$ (or 0.05 probability of error). In addition, we trimmed all 5′ and 3′ stretches of ambiguous "N" nucleotides. Trimming resulted in reads with a mean length of 73.6 bp across all samples, and a minimum length of 25 bp was applied during sequence trimming.

Alignment and analysis of variants including SNPs

The Burrows–Wheeler Aligner (BWA) program (Li and Durbin 2009) was used to align the reads to the reference genome. The BWA default values for mapping were used, except for seed length = 32, maximum differences in the seed = 1, number of threads = 10, mismatch penalty = 6, gap open penalty = 15, gap extension penalty = 8. Mapped reads were extracted from the resulting BAM file for further analyses using SAMtools (Li et al. 2009). High mapping quality ensures reliable (unique) mapping of the reads, which important for variant calling.

Only reliable BWA-mapped reads were considered for single nucleotide polymorphism (SNP) calling. The positions of SNPs in the aligned reads compared to the reference were called using SAMtools (Li et al. 2009). Using the varFilter command, SNPs were called only for variable positions with a minimal mapping quality of 30. The minimum SNP quality was set 100. The minimum and maximum read depths were set 5 and 1000, respectively. A custom perl script was used to select significant sites in the called SNP positions.

RNA extraction and expression analysis of flowering genes

Total RNA was isolated from sampled fresh leaves with TRIzol Reagent (Invitrogen, CA, USA) according to the protocol provided. The RNA concentration was determined using the Nanodrop system (Nanodrop, DE). The DNA solution was then diluted to the working concentration with distilled water and stored at −20 °C until use.

The relative expression levels between JS and SR were obtained by quantitative RT-PCR using SYBR Green II Master Mix Kit (Takara, Japan). The specific primers were designed using PRIMER3 (http://frodo.wi.mit.edu/primer3/) based on the nine non-synonymous substitutions of flowering genes (Supplemental Table 1). An F-box family protein (*CD397253*) and an actin gene (*ACT2/7, TC204150*) were selected as housekeeping genes for comparison of gene expression levels (Gutierrez-Gonzalez et al. 2010). Total RNA (100 ng) was treated with DNase (Promega), and reverse transcription was performed according to the PrimeScript RT Reagent Kit instructions. For the qPCR, 100 ng of total RNA was used for reaction in the Eco™ Real-Time PCR System (Illumina), and the expression levels were analyzed with Eco Software v3.0.16.0 and normalized with the results of α-tubulin. Three replicate reactions were performed for each set of conditions, and the data were presented as mean ± SD ($n = 3$).

Results

Whole genome resequencing

Genome-wide variation between SR and JS was screened by whole genome resequencing with an Illumina Genome Analyzer. Illumina fragment libraries were made from genomic DNA isolated from a single homozygous plant. A total of 167 million reads produced approximately 15 Gb of sequence for SR, giving 15.5× genome coverage; 131 million reads produced 11 Gb of sequence for JS with 11.9× coverage (Table 1). In the wild-type JS genome, mapping of 158.4 million (94.5 %) reads to the reference resulted in the production of a 931.9-Mb consensus sequence, with 95.7 % genome coverage for the reference. In the mutant SR genome, mapping of 124.8 million (94.8 %) reads to the reference resulted in the production of a 926-Mb consensus sequence, with 95.2 % genome coverage for the reference.

Table 1 Results of genome resequencing between wild-type landrace Seoritae (SR) and early-flowering mutant Josaengseori (JS) by comparison with the reference genome (Gmax v1.1, 973 Mb)

	SR	JS
Total reads	167,669,640	131,769,690
Total size (bp)	15,069,751,712	11,616,072,608
Sequencing depth	15.5×	11.9×
Mapped reads	158,484,427 (94.5 %)	124,877,076 (94.8 %)
Properly paired reads	143,065,978 (85.3 %)	114,209,316 (86.7 %)
Consensus sequence (bp)	931,903,776	926,093,898
Genome coverage	95.7 %	95.2 %

Identification of SNPs on individual chromosomes between SR and JS

The frequencies of SNPs and insertions–deletions (indels) on individual chromosomes of SR and JS were surveyed by comparing with the reference sequence (Table 2). A total of 332,821 SNPs were identified throughout all soybean chromosomes. Chromosome 18 showed the highest number of SNPs (37,463), and very few SNPs (2591) were detected on chromosome 20. The frequency of SNPs in each chromosome varied from 1/1482 bp (chr. 16) to 1/18,051 bp (chr. 20), with an average value of 1/2925 bp; nucleotide diversity π (average number of SNPs per nucleotide) ranged from 5.54×10^{-5} (chr. 20) to 6.75×10^{-4} (chr. 16), with an average value of $\pi = 3.42 \times 10^{-4}$.

Among 332,821 polymorphic SNPs, 68,662 (20.6 %) were detected in genic regions, and 264,159 (79.4 %) were in intergenic regions. Among the polymorphic SNPs in genic regions, 21,084 were in coding sequence (CDS) regions, 13,755 were in untranslated regions (UTRs), and 69,140 were in introns. Of the 21,084 SNPs detected in CDS regions, 46.6 % were synonymous and 53.4 % were non-synonymous. The polymorphic SNPs between JS and SR included 12,519 genes.

The detected SNPs were classified into two groups: transitions (A/G and C/T) and transversions (A/C, A/T, C/G, and G/T), on the basis of nucleotide substitutions (Table 3). Among the 332,821 SNPs, 224,570 (67.5 %) were transitions and 108,251 (32.5 %) were transversions. Among transversions, the frequency of C/G was lowest (6.9 %).

Identification of indels on individual chromosomes between SR and JS

A total of 65,168 indels were identified in all soybean chromosomes, ranging from 875 (chr. 20) to 6793 (chr. 18) with an average value of 14,934 (Table 2). Chromosome 18 had the greatest number of SNPs (37,463) and also had a large number of indels. Among 65,178 polymorphic indels, 13,441 (20.6 %) were detected in genic regions, and 51,737 (79.4 %) were in intergenic regions. Among the polymorphic indels in genic regions, 741 were in CDS regions, 3884 were in UTRs, and 15,675 were in introns. The polymorphic indels between SR and JS included 7211 genes.

The lengths of indels and their frequency in SR and JS were calculated (Fig. 1). Among 655,178 indels, 32,571 insertions (1–5 bp), and 32,607 deletions (1–5 bp) were observed. The frequency of different types of indels varied and was negatively correlated with the number of nucleotides. Mononucleotide indels (49,614, 76.1 %) were the most frequent indels in JS and SR, following by di- (9821, 15.1 %) and trinucleotide indels (3390, 5.2 %).

Functional analysis of genes carrying non-synonymous SNPs and indels

To identify significantly affected biological processes, gene ontology (GO) analysis was performed using SNPs or indels between JS and SR. GO terms with a false discovery rate (FDR) of less than 0.01 were collected. Biological processes in which SNPs and indels were involved included reproduction (GO:0000003, GO:0022414), cellular processes (GO:0009987), organization of cellular components (GO:0016043), multicellular organismal processes (GO:0032501), development (GO:0032502), responses to stimulus (GO:0050896), localization (GO:0051177), establishment of localization (GO:0051234), and biological regulation (GO:0065007). Among these, cellular processes formed the largest categories of both SNPs (42 %) and indels (44.1 %) (Fig. 2).

Screening of genes related to flowering time in JS

Agronomic traits of mutant cultivar JS that showed major improvement over the wild-type SR were earlier time to 50 % flowering (10 days) and to pod maturity (34 days) (Song et al. 2010). To identify a casual mutation for early flowering, we screened the variation in functional DNA of 118 genes involved in the flowering pathway throughout the soybean genome (Kim et al. 2012). Of these, 30 flowering genes were in SNPs and 25 were in indels (Supplemental Tables 2 and 3). Especially, nine flowering time-related genes carried non-synonymous SNPs in coding regions (Table 4). Based on the reference genome sequence, five genes (Glyma02g33040, Glyma10g36600, Glyma13g01290, Glyma14g10530, and Glyma17g11040) showed non-synonymous SNPs in JS; Glyma06g22650, Glyma16g01980, Glyma18g53690, and Glyma20g29300

Table 2 Identification of polymorphisms on individual chromosomes between wild-type soybean landrace Seoritae (SR) and early-flowering mutant Josaengseori (JS) by comparison with the reference genome sequence

Chr. no.	No. of polymorphic SNPs	No. of intergenic regions	Genic region					No. of genes	No. of polymorphic indels	No. of intergenic regions	Genic region				No. of genes
			Total	UTR	Coding		No. of introns				Total	UTR	CDS	No. of introns	
					Synonymous	Non-synonymous									
1	27,950	24,671	3279	709	469	671	2903	651	4639	3976	663	255	44	717	366
2	16,444	13,998	2446	456	304	430	2396	564	2917	2447	470	130	21	487	273
3	21,293	17,252	4041	737	658	647	4084	745	3976	3195	781	216	43	936	419
4	15,408	12,285	3123	609	447	498	3439	524	3000	2410	590	164	32	699	326
5	2707	1442	1265	306	185	228	1296	243	937	670	267	79	18	328	165
6	19,326	14,485	4841	1,025	719	796	4402	924	4086	3156	930	296	46	1036	497
7	22,667	17,279	5388	1117	790	917	4747	992	4554	3525	1029	330	53	1100	549
8	6217	3853	2364	551	342	396	2542	440	1616	1148	468	158	18	635	244
9	19,823	15,289	4534	904	698	750	4435	875	3999	3143	856	227	58	1011	464
10	22,383	18,531	3852	844	546	672	4295	835	3971	3241	730	217	34	938	441
11	7407	4725	2682	614	350	374	2711	522	1903	1340	563	152	34	682	301
12	15,245	12,294	2951	506	396	528	2771	544	2716	2161	555	142	27	628	301
13	13,343	9719	3624	737	556	609	3816	705	3040	2276	764	216	40	920	412
14	26,085	22,605	3480	598	372	498	3856	510	4171	3503	668	164	34	792	330
15	5898	4034	1864	331	259	218	2270	290	1444	1043	401	110	21	489	209
16	25,238	20,204	5034	1010	755	813	4942	843	4898	3892	1006	267	68	1194	508
17	13,122	10,847	2275	464	315	381	2184	481	3026	2485	541	152	23	661	318
18	37,463	29,155	8308	1605	1200	1317	8777	1174	6793	5371	1422	384	89	1673	679
19	10,488	8381	2107	349	280	339	2140	403	2292	1808	484	164	23	486	252
20	2591	1576	1015	236	151	142	972	207	875	653	222	50	13	230	139
Scaffolds	1723	1534	189	47	30	38	162	47	325	294	31	11	2	33	18
Total	332,821	264,159	68,662	13,755	9822	11,262	69,140	12,519	65,178	51,737	13,441	3884	741	15,675	7211

Table 3 Identification of substitutions in SNPs detected in early-flowering mutant Josaengseori (JS) by comparison with wild-type landrace Seoritae (SR) genome sequence

	SNPs	Ratio (%)
Transitions		
A/G	111,850	33.6
C/T	112,720	33.9
Transversions		
A/C	28,591	8.6
A/T	27,824	8.3
C/G	22,922	6.9
G/T	28,914	8.7
Total	332,821	100

were in SR. Glyma02g33040 showed a change from nucleotide A in SR to C in JS, which resulted in changes in amino acids, from glutamic acid to aspartic acid. The G–A substitution was shown in Glyma06g22650 and caused a change from valine in SR to methionine in JS. Glyma10g36600 and Glyma13g01290 showed A–T substitutions, which changed from lysine to a stop codon and from threonine to serine, respectively. Glyma14g10530 had a C–T substitution, which caused a change from serine to leucine. Glyma16g01980 showed A–T substitutions that corresponded to cysteine being replaced by serine; Glyma17g11040 had T–C substitutions, resulting in changing from isoleucine to threonine. Glyma18g53690 showed a nucleotide change from C to A, which resulted in a change from alanine to aspartic acid. Glyma20g29300 showed two SNPs, T–C and A–G substitutions, which resulted in changes from glutamine to arginine and from valine to alanine, respectively.

Differential expressions of nine flowering genes detected in SNPs between SR and JR

We performed real-time quantitative PCR to determine whether non-synonymous SNPs might affect the expression of the nine flowering time-related genes (Fig. 3). Among these genes, expression of *ELF3* was increased in JS relative to SR, and six genes (*AGL18*, *GI*, *LHY*, *TOC1*, *TSF*, and *AGL42*) had decreased levels of expression. The expression level of *AGL8* was similar in JS and SR.

Discussion

Next-generation sequencing generates massive amounts of nucleotide reads at a low cost per Mb, and mapping of mutant reads to the wild-type genome sequence provides high sequencing depth for the prediction of mutation sites (Hwang et al. 2015). Ionizing radiation (e.g., gamma-ray) produces both DNA strand breaks and base substitutions (Morita et al. 2009). The induction of nucleotide substitutions and small deletions (2–16 bp) by gamma-ray treatment has been demonstrated in different plant species, but the frequency of these changes at the DNA level has been evaluated in only a few studies performed on rice (Sato et al. 2006; Morita et al. 2009) and *Arabidopsis* (Yoshihara et al. 2010).

In the present study, we acquired numerous nucleotide changes in a soybean mutant using NGS. Gamma-ray mutagenesis produced 332,821 polymorphic SNPs and 65,178 indels in early-flowering mutant JS compared to wild-type cultivar SR. In the detected SNPs, A/G and C/T transitions were predominant and accounted for 33.6 and 33.9 %, respectively. Similarly, in whole genome analysis

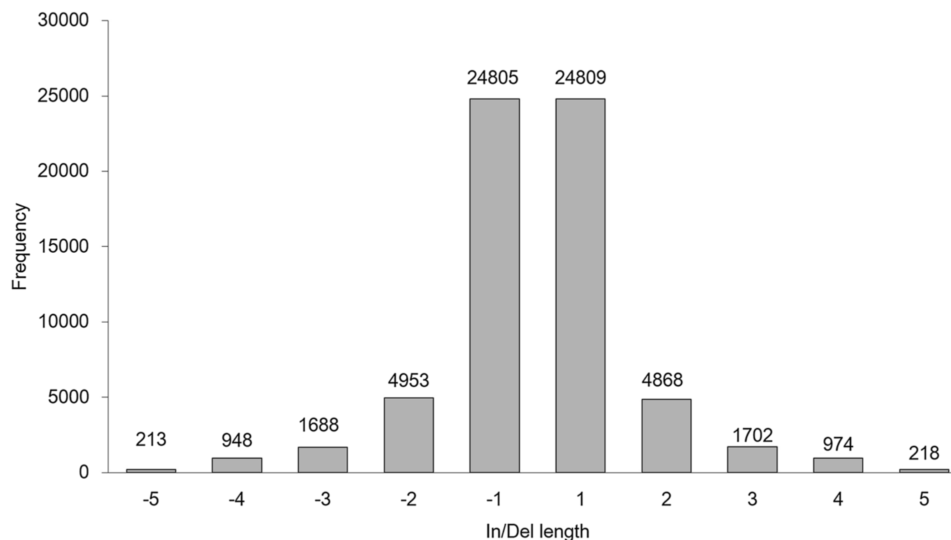

Fig. 1 Frequency and size of deletions and insertions in early-flowering soybean mutant Josaengseori (JS) compared with the wild-type landrace Seoritae (SR) genome sequence

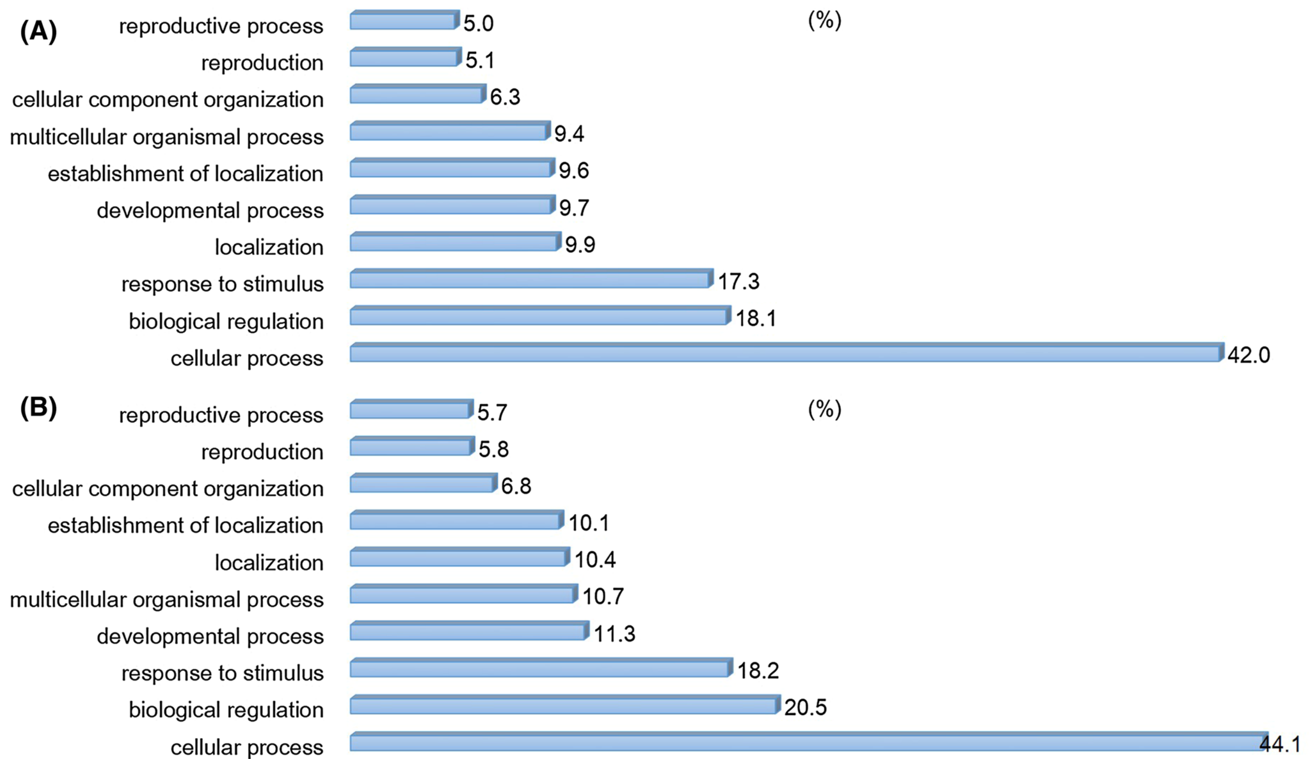

Fig. 2 GO term representation (%) of SNPs or indels between JS and SR

Table 4 Identification of functional mutations in soybean genes that are homologous with flowering time-related genes of *Arabidopsis*

#	Soybean gene model ID	TAIR ID	Symbol	Description	SNP position	Ref.	SR/JS	Syn/non-syn	Non-syn type
1	Glyma02g33040	AT3G57390	*AGL18*	AGAMOUS-like 18	36601115	A	A/C	Non-syn	E->D
2	Glyma06g22650	AT5G60910	*AGL8, FUL*	AGAMOUS-like 8	19383573	A	G/A	Non-syn	V->M
3	Glyma10g36600	AT1G22770	*GmGla, FB, GI*	gigantea protein (GI)	44732850	A	A/T	Non-syn	K->[a]
4	Glyma13g01290	AT5G57660	*COL5*	CONSTANS-like 5	960843	A	A/T	Non-syn	T->S
5	Glyma14g10530	AT2G25930	*ELF3, PYK20*	hydroxyproline-rich glycoprotein family protein	8738120	C	C/T	Non-syn	S->L
6	Glyma16g01980	AT1G01060	*LHY, LHY1*	Homeodomain-like superfamily protein	1510620	T	A/T	Non-syn	C->S
7	Glyma17g11040	AT5G61380	*APRR1, AtTOC1, PRR1, TOC1*	CCT motif-containing response regulator protein	8291448	T	T/C	Non-syn	I->T
8	Glyma18g53690	AT4G20370	*GmFT1b, TSF*	Phosphatidylethanolamine-binding protein (PEBP) family protein	61963221	A	C/A	Non-syn	A->D
9	Glyma20g29300	AT5G62165	*AGL42*	AGAMOUS-like 42	38198860	C	T/C	Non-syn	Q->R

Ref Gmax v1.1, 973 Mb, *SR* wild-type landrace Seoritae, *JS* early-flowering mutant Josaengseori, *Syn* synonymous mutant, *Non-syn* non-synonymous mutant

[a] Stop codon

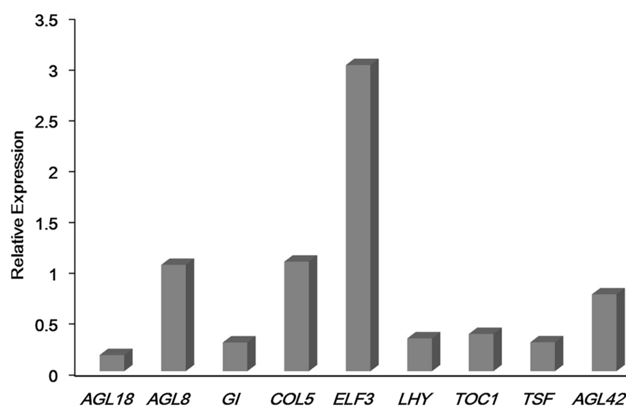

Fig. 3 Relative expression levels of flowering genes between wild-type landrace Seoritae (SR) and early-flowering mutant Josaengseori (JS)

of rice mutants generated by gamma-ray irradiation, the frequency of base transitions (70 %) was higher than that of base transversions (Hwang et al. 2014). In *Arabidopsis* dry seeds, gamma-ray exposure also generated a high frequency of G/C to A/T transitions (Yoshihara et al. 2010). Ionizing radiation induces the formation of free radicals (OH• and H•), which leads to oxidative DNA damage (Breen and Murphy 1995; Wallace 1998). The predominant form of oxidative damage to DNA in animal and bacterial cells is the formation of 8-OH-dG (Fuciarelli et al. 1990; Hirano et al. 2001; Wang et al. 1998). This molecule (8-OH-dG) can mispair with adenine, which induces major base substitutions involving G/C to T/A transversions (Shibutani et al. 1991). In our study, however, the incidence of transitions was higher than that of transversions (8.6 % A/C, 8.3 % A/T, 6.9 % C/G, and 8.2 % G/T). Yoshihara et al. (2010) suggested that mutations induced by the oxidation of guanine might be low in irradiated plant seeds and that the spectra of these mutation types varied according to the irradiation conditions and cell types. Conditions such as low water content and cell proliferation activity in the dry seeds might have affected the mutation spectrum in our soybean mutant as well as in rice and *Arabidopsis*.

The mutation rate in early-flowering mutant JS generated by 250-Gy gamma-irradiation was calculated on a genome basis using whole genome resequencing. The average frequency of SNPs in the soybean mutant genome was 1/2925 bp, and nucleotide diversity π was 3.42×10^{-4}. This was consistent with the rice mutant genome of 1/2736 bp found with 200–300 Gy of gamma-irradiation (Hwang et al. 2014). These results suggested that the degree of genetic alteration created by gamma-irradiation was similar among different species.

The genetic alterations (SNPs and indels) were predominantly (79 %) located in intergenic regions. Among the polymorphic SNPs in CDS regions, the frequency of non-synonymous substitutions was much higher than that of synonymous substitutions. Similar mutagenic effects, such as high frequency of non-synonymous changes, were observed in rice mutants (Hwang et al. 2014). Non-synonymous changes, which alter amino acid sequences, may contribute to phenotypic differences. Therefore, the high frequency of non-synonymous changes could help to explain why gamma-irradiation produces high mutant rates and spectrums in agronomic traits.

The early-flowering mutant JS carried SNPs and indels in approximately 19,700 genes from wild-type cultivar SR. To select potential candidate mutations, these variations can be prioritized according to the putative functions of the genes that are thought to be connected to the mutant phenotypes. The availability of the reference genome and comparative genomic analysis enabled the identification of 118 key gene sets involved in the flowering pathway (Kim et al. 2012). Among these, only mutated genes containing non-synonymous SNPs in coding regions that produce amino acid changes were further examined as potential candidate mutations for earl flowering mutant JS. Using this approach, we identified nine flowering genes containing non-synonymous SNPs in the CDS regions (Table 4). Among these, Glyma10g36600 (*GmGIa*, *E2*) of JS contained a premature stop codon (AAA → TAA) at the 10th exon, which was the same mutation described in a previous report (Watanabe et al. 2011). The *GI* gene plays an important role in flowering by controlling the mRNA expression levels of *CO* and *FT* under inductive conditions in a wide range of plant species, including monocots and dicots such as rice and *Arabidopsis* (Koornneef et al. 1998; Fowler et al. 1999; Hayama et al. 2003; Mizoguchi et al. 2005). Watanabe et al. (2011) suggested that early flowering in soybean caused by the loss of function of *GmGIa* was related to the expression level of *GmFT2a*. *GI* regulates *FT* expression via microRNAs that cooperate with other transcriptional factors (Jung et al. 2007). Overexpression of *OsGI* in transgenic rice suppressed the expression of *FT* orthologs and resulted in a late-flowering phenotype (Hayama et al. 2003). In our study, the expression level of *GmFT2a* was higher in JS than in SR (3.197, supplemental Fig. 1). It seems that the premature stop codon of Glyma10g36600 in JS affected the expression of *GmFT2a* and resulted in an early-flowering phenotype in JS.

TOC1 (TIMING OF CAB EXPRESSION 1) and *CCA1* (CIRCADIAN CLOCK ASSOCIATED 1)/*LHY* (LATE ELONGATED HYPOCOTYL) together make up the proposed central circadian loop (Ding et al. 2007). This positive–negative feedback loop between evening and morning factors led to the first genetic model of the plant clock (Alabadi et al. 2001). Both *toc1* and *cca1/lhy* have defects in flowering time and photomorphogenesis, which correlate with the respective mutant circadian phenotypes (Somers

et al. 1998; Strayer et al. 2000). In *toc1*, mutant plants have an early-flowering phenotype when grown under a short-day photoperiod. This phenotype is the result of clock-based misinterpretation of photoperiodic information rather than of the direct effect of *toc1* on floral-induction pathways (Somers et al. 1998; Strayer et al. 2000). *cca1* and *lhy* also exhibit an early-flowering phenotype under short-day conditions, and this was especially marked in the ccal/lhy double mutant, which is nearly insensitive to photoperiodic sensing (Mizoguchi et al. 2002). In our results, *TOC1* (Glyma17g11040) and *LHY* (Glyma16g01980) showed non-synonymous substitutions in JS and SR, respectively. They also showed lower expression levels in JS than in SR. Ding et al. (2007) reported that in *cca1/lhy/toc1*, *cca1/toc1*, and *lhy/toc1*, the phase of *GI* expression was shifted earlier, resulting in a correlative increase in *FT* expression level. We suggest that the early-flowering phenotype in JS is a result of the low expression of *TOC1* and *LHY*, which, in turn, leads to a phase shift of *GI* and an increase in *FT*.

Here, *AGL18* (Glyma02g33040) of JS was detected non-synonymous substitutions, but *AGL15* did not change. In the qPCR, *AGL18* of JS showed a lower expression level (0.159) than SR (Fig. 3). The MADS-domain factors *AGL15* and *AGL18* contribute to regulation of the floral transition (Fernandez et al. 2014). While single mutants have no phenotype, *agl15/agl18* double mutants flower earlier than the wild type (Adamczyk et al. 2007). Therefore, *AGL15* and *AGl18* appear to act in a redundant fashion as floral repressors in seedlings. The earlier flowering in *agl15/agl18* mutants under short-day conditions is associated with upregulation of *FT*, and both *AGL15* and *AGL18* are expressed in the vascular system and shoot apex of young seedlings (Adamczyk et al. 2007), which suggests that *AGL15* and *AGL18* act directly on *FT* in leaves and on other targets in the meristem (Fernandez et al. 2014).

ELF3 is a clock-associated gene that plays a pivotal role in the circadian gating pathway (Hicks et al. 1996; McWatters et al. 2000). *ELF3* has been shown to interact directly with both *COP1* and *GI* in vivo. In the present study, *ELF3* of JS showed higher expression (3.004) than SR, while *GI* had low expression (0.282). Yu et al. (2008) reported that *ELF3*-mediated interaction of *COP1* with *GI* may result in degradation of not only the protein target GI, but also of the substrate adaptor *ELF3*. They suggested that *ELF3* is degraded upon interaction with *COP1*, creating a negative feedback mechanism that limits the extent of *ELF3* activity. In our study, the expression level of *COP1* (Glyma14g05430) was lower in JS than in SR (0.535, Supplemental Fig. 1). The lower *COP1* levels appear to have caused the higher level of *ELF3* in JS. In addition, *ELF3* overexpression promoted constitutive degradation of *GI* (Yu et al. 2008). In our study, higher expression of *ELF3* in JS seemed to affect the lower expression of *GI*.

In this study, we used NGS to analyze an early-flowering soybean mutant generated by gamma-ray irradiation. The mutant, JS, showed numerous SNPs and indels compared to the original landrace, SR. JS contained changes in flowering genes related to the photoreceptor-mediated signaling pathways. We suggest that the early-flowering phenotype in JS was caused by changes in flowering genes generated by gamma-ray mutagenesis. Our results provide critical insights into the regulatory pathways associated with soybean flowering and help to improve our knowledge about mutation breeding.

Acknowledgments This work was supported by the Korea Atomic Energy Research Institute (KAERI) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2012M2A2A6010566) and Chonnam National University, 2014, Republic of Korea.

Compliance with ethical standards

This article does not contain any studies with human participants or animal performed by any of the authors.

Conflict of interest The authors declare that there are no competing interests.

References

Adamczyk BJ, Lehti-Shiu MD, Fernandez DE (2007) The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in *Arabidopsis*. Plant J 50:1007–1019

Ahloowalia BS, Maluszynski M, Nichterlein K (2004) Global impact of mutation-derived varieties. Euphytica 135:187–204

Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Mas P, Kay SA (2001) Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. Science 293:880–883

Breen AP, Murphy JA (1995) Reactions of Oxyl Radicals with DNA. Free Radic Biol Med 18:1033–1077

Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinform 11:485–490

Ding ZJ, Doyle MR, Amasino RM, Davis SJ (2007) A complex genetic interaction between *Arabidopsis thaliana* TOC1 and CCA1/LHY in driving the circadian clock and in output regulation. Genetics 176:1501–1510

Fernandez DE, Wang CT, Zheng YM, Adamczyk BJ, Singhal R, Hall PK, Perry SE (2014) The MADS-domain factors AGAMOUS-LIKE15 and AGAMOUS-LIKE18, along with SHORT VEGETATIVE PHASE and AGAMOUS-LIKE24, are necessary to block floral gene expression during the vegetative phase. Plant Physiol 165:1591–1603

Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Coupland G, Putterill J (1999) GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. EMBO J 118:4679–4688

Fuciarelli AF, Wegher BJ, Blakely WF, Dizdaroglu M (1990) Yields of radiation-induced base products in DNA—effects of DNA conformation and gassing conditions. Int J Radiat Biol 58:397–415

Gutierrez-Gonzalez JJ, Guttikonda SK, Tran LSP, Aldrich DL, Zhong R, Yu O, Nguyen HT, Sleper DA (2010) Differential expression of isoflavone biosynthetic genes in soybean during water deficits. Plant Cell Physiol 51:936–948

Ha BK, Lee KJ, Velusamy V, Kim JB, Kim SH, Ahn JW, Kang SY, Kim DS (2014) Improvement of soybean through radiation-induced mutation breeding techniques in Korea. Plant Genetic Plant Genet Resour C 12:S54–S57

Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K (2003) Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 422:719–722

Hicks KA, Millar AJ, Carre IA, Somers DE, Straume M, MeeksWagner DR, Kay SA (1996) Conditional circadian dysfunction of the *Arabidopsis* early-flowering 3 mutant. Science 274:790–792

Hirano T, Hirano H, Yamaguchi R, Asami S, Tsurudome Y, Kasai H (2001) Sequence specificity of the 8-hydroxyguanine repair activity in rat organs. J Radiat Res 42:247–254

Hwang SG, Hwang JG, Kim DS, Jang CS (2014) Genome-wide DNA polymorphism and transcriptome analysis of an early-maturing rice mutant. Genetica 142:73–85

Hwang WJ, Kim MY, Kang YJ, Shim S, Stacey MG, Stacey G, Lee SH (2015) Genome-wide analysis of mutations in a dwarf soybean mutant induced by fast neutron bombardment. Euphytica 203:399–408

Jung JH, Seo YH, Seo PJ, Reyes JL, Yun J, Chua NH, Park CM (2007) The GIGANTEA-regulated MicroRNA172 mediates photoperiodic flowering independent of CONSTANS in *Arabidopsis*. Plant Cell 19:2736–2748

Kim DH, Doyle MR, Sung S, Amasino RM (2009) Vernalization: winter and the timing of flowering in plants. Annu Rev Cell Dev Biol 25:277–299

Kim DS, Lee KJ, Kim JB, Kim SH, Song JY, Seo YW, Lee BM, Kang SY (2010) Identification of Kunitz trypsin inhibitor mutations using SNAP markers in soybean mutant lines. Theor Appl Genet 121:751–760

Kim MY, Shin JH, Kang YJ, Shim SR, Lee SH (2012) Divergence of flowering genes in soybean. J Biosci 37:857–870

Koornneef M, Alonso-Blanco C, Blankestijn-de Vries H, Hanhart CJ, Peeters AJM (1998) Genetic interactions among late-flowering mutants of *Arabidopsis*. Genetics 148:885–892

Lee KJ, Kim JB, Kim SH, Ha BK, Lee BM, Kang SY, Kim DS (2011) Alteration of seed storage protein composition in soybean [*Glycine max* (L.) Merrill] mutant lines induced by gamma-irradiation mutagenesis. J Agric Food Chem 59:12405–12410

Lee KJ, Hwang JE, Velusamy V, Ha BK, Kim JB, Kim SH, Ahn JW, Kang SY, Kim DS (2014) Selection and molecular characterization of a lipoxygenase-free soybean mutant line induced by gamma irradiation. Theor Appl Genet 127:2405–2413

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, Abe J (2008) Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A Gene. Genetics 180:995–1007

Lukowitz W, Gillmor CS, Scheible WR (2000) Positional cloning in arabidopsis. Why it feels good to have a genome initiative working for you. Plant Physiol 123:795–805

McWatters HG, Bastow RM, Hall A, Millar AJ (2000) The ELF3 zeitnehmer regulates light signalling to the circadian clock. Nature 408:716–720

Mizoguchi T, Wheatley K, Hanzawa Y, Wright L, Mizoguchi M, Song HR, Carre IA, Coupland G (2002) LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in *Arabidopsis*. Dev Cell 2:629–641

Mizoguchi T, Wright L, Fujiwara S, Cremer F, Lee K, Onouchi H, Mouradov A, Fowler S, Kamada H, Putterill J, Coupland G (2005) Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in *Arabidopsis*. Plant Cell 17:2255–2270

Morita R, Kusaba M, Iida S, Yamaguchi H, Nishio T, Nishimura M (2009) Molecular characterization of mutations induced by gamma irradiation in rice. Genes Genet Syst 84:361–370

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Sato Y, Shirasawa K, Takahashi Y, Nishimura M, Nishio T (2006) Mutant selection from progeny of gamma-ray-irradiated rice by DNA heteroduplex cleavage using Brassica petiole extract. Breed Sci 56:179–183

Schneeberger K (2014) Using next-generation sequencing to isolate mutant genes from forward genetic screens. Nat Rev Genet 15:662–676

Shibutani S, Takeshita M, Grollman AP (1991) Insertion of specific bases during DNA-synthesis past the oxidation-damaged base 8-Oxodg. Nature 349:431–434

Somers DE, Webb AAR, Pearson M, Kay SA (1998) The short-period mutant, toc1-1, alters circadian clock regulation of multiple outputs throughout development in *Arabidopsis thaliana*. Development 125:485–494

Song HS, Kim J-B, Lee KJ, Kim DS, Kim SH, Lee SJ, Kang S-Y (2010) A new improved soybean variety, "Josaengseori" by mutation breeding. Korean J Breed Sci 42:222–225

Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Mas P, Panda S, Kreps JA, Kay SA (2000) Cloning of the *Arabidopsis* clock cone TOC1, an autoregulatory response regulator homolog. Science 289:768–771

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. Plos Biol 12:e1001883

Wallace SS (1998) Enzymatic processing of radiation-induced free radical damage in DNA. Radiat Res 150:S60–S79

Wallace SS (2002) Biological consequences of free radical-damaged DNA bases. Free Radic Biol Med 33:1–14

Wang D, Kreutzer DA, Essigmann JM (1998) Mutagenicity and repair of oxidative DNA damage: insights from studies using defined lesions. Mutat Res 400:99–115

Watanabe S, Hideshima R, Xia ZJ, Tsubokura Y, Sato S, Nakamoto Y, Yamanaka N, Takahashi R, Ishimoto M, Anai T, Tabata S, Harada K (2009) Map-based cloning of the gene associated with the soybean maturity locus E3. Genetics 182:1251–1262

Watanabe S, Xia ZJ, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata S, Kitamura K, Harada K (2011) A Map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. Genetics 188:U260–U395

Wu JY, Morimyo M, Hongo E, Higashi T, Okamoto M, Kawano A, Ohmachi Y (2006) Radiation-induced germline mutations detected by a direct comparison of parents and first-generation offspring DNA sequences containing SNPs. Mutat Res 596:1–11

Xia ZJ, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lu SX, Wu HY, Tabata S, Harada K (2012a) Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proc Natl Acad Sci USA 109:E2155–E2164

Xia ZJ, Zhai H, Liu BH, Kong FJ, Yuan XH, Wu HY, Cober ER, Harada K (2012b) Molecular identification of genes controlling flowering time, maturity, and photoperiod response in soybean. Plant Syst Evol 298:1217–1227

Yoshihara R, Hase Y, Sato R, Takimoto K, Narumi I (2010) Mutational effects of different LET radiations in rpsL transgenic *Arabidopsis*. Int J Radiat Biol 86:125–131

Yu JW, Rubio V, Lee NY, Bai SL, Lee SY, Kim SS, Liu LJ, Zhang YY, Irigoyen ML, Sullivan JA, Zhang Y, Lee I, Xie Q, Paek NC, Deng XW (2008) COP1 and ELF3 control circadian function and photoperiodic flowering by regulating GI stability. Mol Cell 32:617–630

Zhai H, Lu SX, Liang S, Wu HY, Zhang XZ, Liu BH, Kong FJ, Yuan XH, Li J, Xia ZJ (2014) GmFT4, a Homolog of FLOWERING LOCUS T, Is Positively Regulated by E1 and Functions as a Flowering Repressor in Soybean. Plos One 9:e89030

Zhu MJ, Zhao SH (2007) Candidate gene identification approach: progress and challenges. Int J Biol Sci 3:420–427