Applied Biological Research 25(2): 133-142; (2023)
DOI: 10.5958/0974-4517.2023.00016.2

MOLECULAR CHARACTERIZATION OF *E1* and *E3* MATURITY LOCI OF SOYBEAN [G*lycine max* (L.) Merrill] GENOTYPES WITH CONTRASTING MATURITY

Manjushri Dongare[1], V.P. Chimote[1], M.P. Deshmukh[2] and A.A. Kale[1]

[1]*State Level Biotechnology Centre, Mahatma Phule Krishi Vidyapeeth (M.P.K.V.), Rahuri, Ahmednagar – 413 732, Maharashtra (India)*
[2]*Agricultural Research Station, M.P.K.V., Kasbe Digraj, Sangli – 416 305, Maharashtra (India)*
*****e-mail: *vpchimote@gmail.com*

(Received 6 October, 2022; accepted 30 January, 2023)

ABSTRACT

Soybean has ten maturity loci governing its flowering time, maturity and photoperiodism. The *E1* and *E3* maturity loci delay maturity and induce photoperiod sensitivity, while their mutants are *vice versa*. In a field trial of 17 genotypes during *kharif* 2016, two early (JS 95-60 and JS 20-34) and two late maturing genotypes (MAUS 61 and MACS 1188) were identified for further molecular analysis. Full-length 525 bp *E1* gene and partial 1688 bp *E3* gene coding sequence were amplified, and sequence analysed along with reference sequence of early maturing Harosoy. In the partial *E3* gene coding sequence studied, 26 substitution SNPs were observed, none of which was maturity specific. In *E1* coding sequences, 87 substitution SNPs were observed, of which 17 SNPs were earliness specific conserved among JS 20-34, JS 95-60 and early Harosoy. Similarly, six amino acid frame polymorphisms were found to be earliness specific shared amongst them. Typical B3-like conserved domains with 11 trimeric DNA binding sites were found in the *E1* gene within 172-504 and 178-504 bases interval intervals in early and late varieties, respectively. Thus, the molecular characterization of *E1* maturity locus in local genotypes of soybean can contribute for more efficient plant breeding.

Keywords: B3 conserved domain, maturity locus *E1*, phytochrome, soybean, SNPs

INTRODUCTION

Soybean [*Glycine max* (L.) Merrill] is a most important oilseed crop. In India, Maharashtra (38% area) and Madhya Pradesh (46%) states share the largest area under soybean crop, followed by Rajasthan (8.8%). Soybean crop is grown worldwide over an extended range of latitudes. However, individual cultivars are confined to specific geographical regions under certain latitudes defined by their maturity groups and photoperiod requirements for flowering (Watanabe *et al*., 2012). Maturity control is an important factor in optimizing geographical compatibility and agronomic efficiency (yield and early maturity). From agricultural perspective, the regulation of both flowering time and photoperiod is critical for getting higher yield in soybean in a shorter duration. Soybean being a short-day plant flowers in response to a short photoperiod. Photoperiod sensitiveness is preferred under subtropical conditions with shorter day-length of Central India which covers approximately 95% of soybean area in India. As against this photoperiod insensitivity induces early flowering is preferred under temperate summer crop with longer day conditions (Lin *et al*., 2021). Flowering in soybean is regulated by a complex but coordinated mechanism. Under long day high latitude temperate conditions, the

photoperiod insensitivity in soybean helps the crop to mature prior to the incoming winter due to the mutant alleles at maturity loci (Tsubokura *et al.*, 2014). However, under rainfed, semi-arid, short day, and sub-tropical Indian conditions these mutant alleles confer advantage of earliness (Tripathi *et al.*, 2021). Of the 101 cultivated Indian soybean varieties, Tripathi *et al*. (2021) found that 15 early maturing and photo-insensitive varieties had recessive alleles at these loci. When soybean varieties, adapted from temperate long day world, are introduced in relatively shorter day length subtropical world they mature early by entering into the reproductive phase before attaining the sufficient vegetative growth (Islam *et al*., 2018; Gupta *et al.*, 2022). Photoperiod conditions restricted the soybean cultivation to the areas of outside 22º latitude until 1970s when Brazil introduced the long juvenile soybean technology (Carpentieri-Pı´polo *et al*., 2000, 2002).

To date about twelve major maturity loci (*E1* to *E11* and *J*) have been identified that control over-flowering and maturity in soybean (Hou *et al*., 2022). Dominant alleles at 7 of E-loci (*E1*, *E2*, *E3*, *E4*, *E7*, *E8* and *E10*) confer late maturity, while the alleles at *E6*, *E9*, *E11* and *J* genes induce early flowering (Hou *et al*., 2022). Four out of them *viz*., *E1*, *E3*, *E4*, and *E7* also govern photoperiod sensitivity trait (Watanabe *et al.,* 2012). Amongst flowering locus T (FT), FT2a and FT5a are the inducers of floral transition under short days in soybean (Kong *et al*., 2010; Thakare *et al.*, 2011). *E1* is a major gene associated with flowering time and maturity. *E1* is a small intron-free gene which encodes flowering suppressor protein with a putative bipartite nuclear localizing signal and a plant-specific DNA binding B3-domain (Xia *et al.,* 2012). Mutations in *E1* reportedly lead to increased expression of both *FT* orthologs (Thakare *et al.*, 2011), but mutations in *E2* gene resulted in an increase in GmFT2a but not *GmFT5a* (Kong *et al.*, 2010), suggesting that FT orthologs fulfil different roles in floral initiation. *E3* gene causes delay in flowering under long day with a variety of light conditions (Hou *et al*., 2022). Both *E3* and *E4* genes encode twin photoreceptor phytochrome*j*, PhyA2 and PhyA3, respectively (Watanabe *et al*., 2009; Lin *et al*., 2022;). These phytochrome receptors are also associated with photosensitivity; however, they show different floral responses to long day conditions. Both PhyA2 and PhyA3 regulate expression of *E1* both at transcriptional and translational levels (Lin *et al*., 2022).

Breeding for early maturity and adaptation to challenging climate conditions are the prime objective of soybean breeding. Though during breeding for early maturity several studies have been conducted in past to identify soybean maturity loci, still the identification and defining the role of these genes is critically desired to enhance our ability to understand and improve soybean for maturity traits. Molecular characterization of genes involved in maturation can provide significant insights into the complexity of soybean. Thus, understanding the molecular mechanism of flowering and identification of relevant genes should be more efficient for molecular breeding (Roux *et al.,* 2006; Lin *et al*., 2021). The *E1* and *E3* genes are the major contributors to both photoperiod sensitivity and plant maturity. However, information on the allelic status of these genes from different genotypes is limited. Therefore, the objective of this study was to identify the alleles of *E1* and *E3* genes through specific SNPs haplotypes in selected soybean genotypes having contrasting maturity.

MATERIALS AND METHODS

Plant material and field trials

The seeds of seventeen genetically diverse genotypes were obtained from the Soybean Breeder, Agricultural Research Station, Kasbe Digraj, district Sangli, Maharashtra (India). The experiment was laid out in a randomized block design with three replications at the Post Graduate Institute Farm, Department of Botany, M.P.K.V., Rahuri, Maharashtra (India) during *Kharif*, 2016. In each replication, the genotypes were grown in 3 m long rows with spacing of 45 cm × 10 cm for the row to row and plant to plant, respectively.

Phenotypic analysis

Five plants of each genotype were randomly selected and data recorded for three characters *viz*., days to 50% flowering, days to pod formation and days to physiological maturity from the date of sowing for the screening of 17 varieties to identify the most-early and late maturing varieties.

DNA isolation

Genomic DNA was extracted from young leaves, based on the standard cetyl trimethyl ammonium bromide method with some modifications described by Doyle and Doyle (1987). The quality of the genomic DNA was checked by using 0.8% (w/v) agarose gel electrophoresis, while quantification was done by spectrophotometry.

Gene-specific marker and PCR amplification

The *E1* gene-specific primers were designed using the primer BLAST online primer design tool to cover the entire coding sequence of *E1* gene. However, for larger *E3* gene (phytochrome A3; 8651 bp), four *E3* gene-specific primer pairs *viz*., PhyA3I, PhyA3II, PhyA3III and PhyA3IV, respectively, were designed for PCR amplification of continuous region from the start of coding region to the end of 3rd exon. Initial 1252 bp UTR and the terminal 3531 bp were excluded from this study. Genomic DNA (30-50 ng) samples were used as templates. The PCR reaction was performed using a PCR thermal cycler with the thermal regime of an initial 5 min denaturation at 95°C, followed by 35 cycles of 45 sec denaturation at 94°C, 45 sec annealing (at 57°C for *E1* primers; at 52 to 60°C for *E3* region specific primers) and 1 min extension at 72°C.

Sequence analysis

The PCR amplified products were resolved on 2.5 and 1.0% agarose gel, for *E1* and *E3* genes, respectively. The amplicon was initially eluted by phenol extraction and purified by HiPurA™ PCR product purification kit (Xu *et al.,* 2013). The gel-eluted PCR products were used as templates for forward and reverse custom sequencing using gene-specific primers.

Bioinformatics analysis

The sequences of amplicons obtained from both the ends with *E1/PhyA3II /PhyA3IV* gene-specific primers were analysed using sequence analysis tool ChromasLite 2.01 software. The resultant orientation sequences were merged and analysed using bioinformatics tools (viz. BLASTN, Clustal Omega, ExPasy translation and NCBI conserved domain search). In BLASTN, the nucleotide query was used to search homologous nucleotide sequence database (Johnson *et al.*, 2008). Clustal Omega was used as multiple sequence alignment tool, available on EMBL database (*www.ebi.ac.uk/Tools/ msa/clustalo/*; Sievers and Higgins, 2014). ExPasy translation tool (*www.expasy.org/tran;* Artimo *et al.*, 2012) was used to predict amino acid sequences based on DNA sequence input query, from which the closest frame matching with peptide sequence was selected for multiple sequence alignment. NCBI conserved domain search (*https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi*; Marchler-Bauer *et al.*, 2015) was used to find out the conserved domains using the nucleotide query.

RESULTS AND DISCUSSION

Identification of most early and most late maturing varieties

Based on preliminary screening of soybean genotypes during *kharif* 2016, the most early- and late-maturing genotypes were identified (Table 1). Genotypes JS9560 and JS2034 were identified as earliest maturing (69 days) genotypes; while on contrary, genotypes MAUS61 and MACS1188 were identified as late maturing (121 and 129 days, respectively) genotypes with average maturity duration of 102 days. The average duration needed for flowering was 44 days with both JS9560 and JS2034 being first to flower in 32 days. These varieties were used for molecular analysis to identify the loci responsible for respective trait.

PCR amplification and sequencing

The PCR amplification of *E1* gene revealed a complete coding region of 525 bp in all the 4 soybean genotypes (Table 2, Fig. 1). These PCR products were gel eluted and sequenced. The *E3* gene being larger (8651 bp) was amplified with 4 sets of primer pairs (3.4 kb terminal region including 4th exon). As expected, the *E3* specific *PhyA3*I, *PhyA3*II and *PhyA3*IV regions amplified 1008 bp (1st exon-1st half 1247-2254), 1091 bp (1st exon-2nd half 2233-3323) and 720 bp (3rd exon with a part of previous intron; 4416-5121) fragments, respectively, in the studied genotypes (Fig. 2a-d). The *PhyA3*III region (816 bp 2nd exon and 294 bp preceding intron 3307-4435) amplified 1129 bp fragment in JS2034 and MAUS61 only, and hence was not sequenced (Fig. 2c). The *GmPhyA3*-I region gave poor quality sequence reads with no match to reference accessions, so was not included in subsequent study. Xia *et al*. (2012) reported *E1* as a small intronless gene which regulates photoperiodic flowering.

Table 1: Maturity associated traits of soybean genotypes

Genotypes	Days to 50% flowering	Days to pod formation	Days to maturity
MAUS61	48	59	121
JS 335	36	43	99
KDS 904	47	56	97
KDS 730	45*	51	89
KDS 726	50	58	109
KDS 921	46	52	110
MAUS 162	48	53	115
MACS1188	49*	59	129
JS 9305	47	50	101
JS 9560	32	38	69
KDS 753	48	53	110
EC241780	43	49	109
KDS 1045	49	54	103
KDS 344	50	55	100
JS 2069	38*	43	99
JS 2034	32*	38	69
JS 2029	39*	43	99

Nucleotide sequence homology search of E1 gene

E1 and *E2* (GIGANTEA) are the primary negative regulator of flowering or floral repressors in soybean (No *et al*., 2021). Complete 525 bp *E1* gene coding region sequences of JS9560, JS2034, MAUS61 and MACS1188 varieties showed 78, 72, 72 and 72 hits, respectively, when searched using BLASTN tool (highly similar). The *E1* gene of the genotypes studied showed 91-94% identity (481-499 out of 525 bp) with *E1* gene from GenBank Accession AB795391.1 (genotype Harosoy). Two *E1* gene paralogues have been reported to show 94% nucleotide sequence identity (Xia *et al.*, 2012).

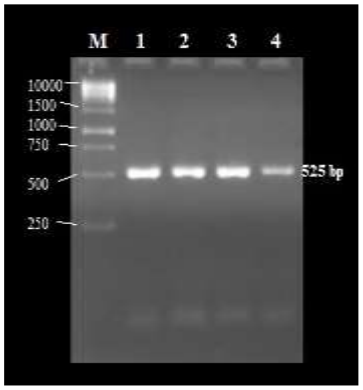

Fig. 1: Amplification of *E1* gene of 1 (JS9560), 2(JS2034), 3(MAUS61) and 4 (MACS1188) by using *E1* gene specific primer pair

Nucleotide sequence homology of E3 PhyA3II and PhyA3IV gene region

The phytochromes (PhyA2/PhyA3) act as photoreceptors of light signals (Tang *et al*., 2022). Phytochrome A and E1 protein together interact to regulate flowering mechanism in soybean (Gupta *et al.*, 2023). Watanabe *et al*. (2009) used map-based cloning to report that *E3 GmPhyA3* gene of 8651 bp has five exons with 3390 bp complete coding sequence. Kim *et al*. (2012) reported that BLAST analysis can be used for the identification of soybean homologs of flowering-related genes of *Arabidopsis thaliana*. Out of the four *E3* gene region specific primer pairs, *PhyA3 III* region was not sequenced as it failed to amplify in two of the four genotypes studied; while *PhyA3-I* region yielded poor quality sequence reads. Out of 1091 bp *E3 PhyA3*II gene region, clear reads for initial and last ~50 bp were missing in all the studied genotypes. The *E3-PhyA3-II* gene region of all genotypes when compared with the same region of genotype Harosoy [GenBank Accession AB468154.1], Watanabe

Table 2: Details of *E1* and *E3* gene specific primers designed and used in the study

Primer name	Sequences	Gene position* (from to 5' to 3')	Annealing temperature (°C)	Expected PCR product size (bp)
E1 F	ATGAGCAACCCTTCAGAT	1 to 18	57	525
E1 R	TTAATTCTCTGGCATAGCTTGTT	525 to 502		
PhyA3 F1	TGACAATGTCCTCTTCAAGG	1247 to 1266	52	1008
PhyA3 R1	CCATGTATTGCAAGTGGC	2254 to 2237		
PhyA3 F2	AGTTGCCACTTGCAATACATGG	2233 to 2254	59	1091
PhyA3 R2	TGCAATGCCATGTCCAACATC	3323 to 3303		
PhyA3 F3	TTGGACATGGCATTGCAG	3307 to 3324	60	1129
PhyA3 R3	CAATGATGCTGTCAAGATCC	4435 to 4416		
PhyA3 F4	GGATCTTGACAGCATCATTG	4416 to 4435	52	706
PhyA3 R4	ACCTGAACTCCAAATTAGCAA	5121 to 5101		

*N is nucleotide position as compared to the reference GenBank Accession, AB468154.1

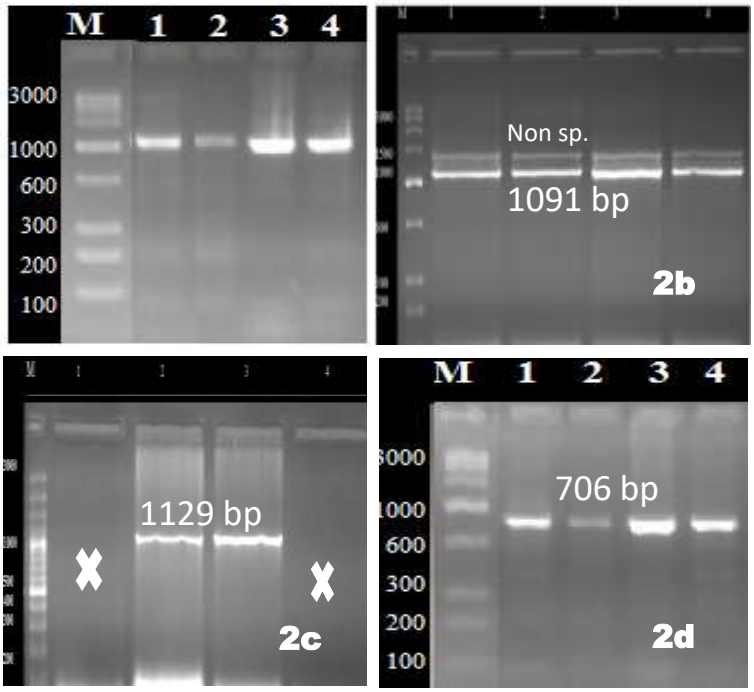

Fig. 2: Amplification of *E3 PhyA3* gene by using four sets of *E3* gene specific PhyA3 I (2a); PhyA3 IV primer pairs; *E3* gene specific PhyA3 I (Fig. 2a), PhyA3 II (Fig. 2b), PhyA3 II ((Fig. 2c), PhyA3 IV (Fig. 2d) primer pairs. Lane M: 250bp DNA ladder, Lane 1 (JS9560), Lane 2(JS2034), Lane 3 (MAUS61) and Lane 4 (MACS1188).

et al. (2009) showed 99.2-99.7% identity (999 of 1002 bp with JS9560 with no gaps; 992 of 995 bp with JS2034 with 1 gap; 988 of 991 bp with MAUS61 with 0 gap; 990 of 998 bp with MACS1188 with 0 gap). *E3-PhyA3-II* gene region sequences of the studied genotypes showed 101-103 hits [all Fabaceae with 49 of them all glycine) having 99% coverage and similarity] when searched using BLASTN tool (highly similar).

Only 708 bp out of 720 bp in *E3 PhyA3*-IV gene region could be assembled in all test genotypes. The *E3-PhyA3-IV* gene region of test genotypes when compared with *E3PhyA3* region from genotype Harosoy (GenBank Accession AB468154.1) showed 97.74-99.15% identity (699, 702, 692 and 701 out of 708 bp with genotypes JS9560; JS2034; MAUS61 and MACS1188, respectively, with each showing 1 gap. *E3-PhyA3-IV* gene region sequences of genotypes JS9560, JS2034, MAUS61 and MACS1188 showed 72-81 hits [all Fabaceae from *Glycine*, *Abrus*, *Cajanus*, *Vigna* and, *Cicer*) with 39 of them (all *Glycine*) having 100% coverage and > 98.59% similarity] when searched using BLASTN tool (highly similar). In case of 39 queries with 100% coverage, the sequence homology observed was 98.59-98.87% similar in JS9560 (of 72 hits), 99.01-99.29% in JS2034 (of 79 hits); 97.74-97.88% in MAUS61 (of 73 hits) and 98.87-99.15% identical in MACS1188 (of 81 hits).

Multiple sequence alignment of E1 gene

E1 gene of all the four test varieties when aligned alongside reference Harosoy (GenBank Accession AB795391.1) had no gaps or insertion/deletion (Fig. 3). Out of 525 bases, 422 bases completely

matched in all the 5 sequences. A total of 103 SNPs were observed in *E1* gene sequences at 52 positions of which 71 SNPs were genotype-specific *i.e.* 43 JS2034 specific, 17 Harosoy specific, 9 MAUS61 specific, and 1 each JS9560 and MACS1188 specific. Amongst the JS2034 specific SNPs, 41 were observed in the initial 101 bases with poor sequence reads and the rest 32 *E1* SNPs were shared by two or more genotypes, of which 19 were specific to all early- or late-genotypes at the same base positions; 6 were shared by JS2034 and Harosoy; 5 were shared only by JS2034 and JS9560; two SNPs were shared by genotypes, irrespective of their maturity type. E1 is a legume-specific transcriptional repressors of flowering in soybean under long day conditions (No *et al*., 2021), and mutation in *E1* gene induces early flowering. Soybean varieties adjust to temperate long-day conditions through the defective/mutant copies of either flowering suppressor encoding *E1* gene, or two phytochromes encoding *E3 /E4* genes (Lin *et al*., 2022). PhyA-E1 photoperiodic pathway mediates delayed flowering at 35°C under short day conditions, however, induction of early flowering at 30°C is independent of E1 suppressor (Tang *et al*., 2022).

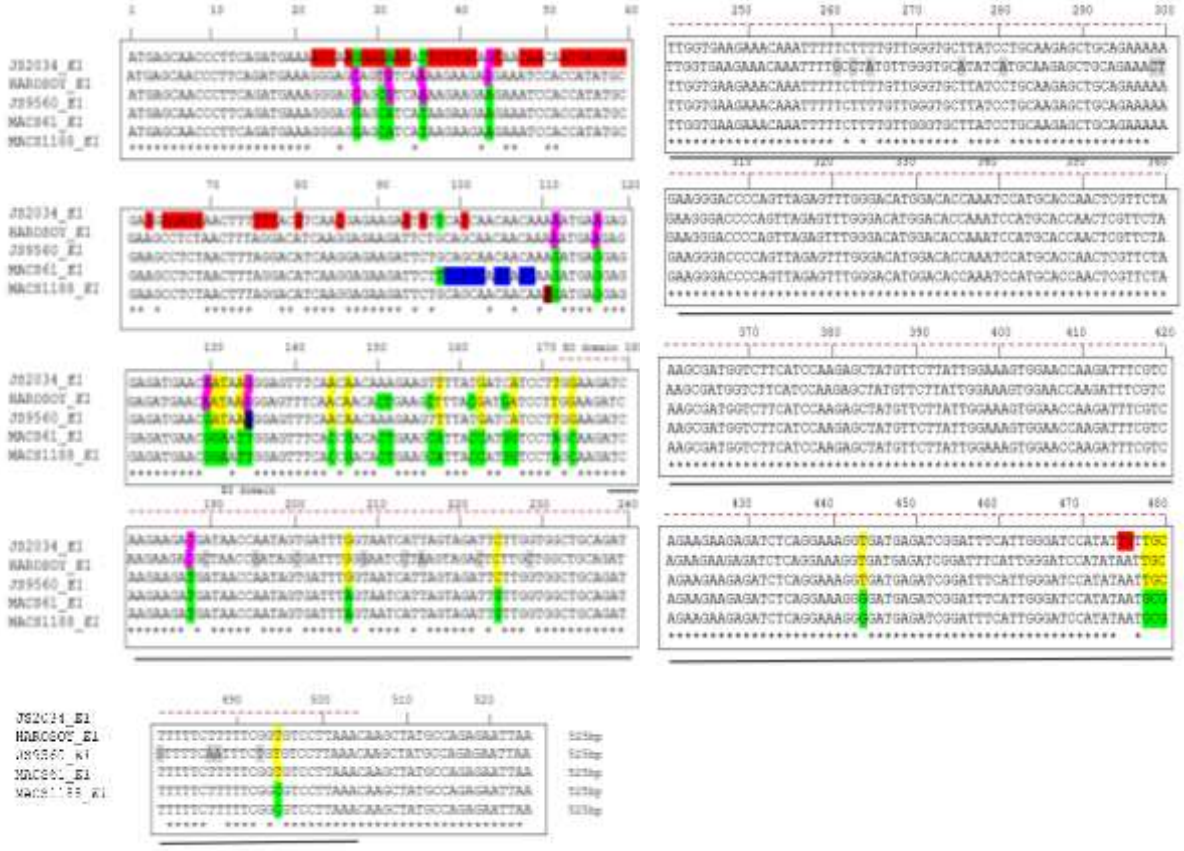

Fig. 3: Multiple sequence alignment of *E1* gene nucleotide sequence of JS2034, JS9560, MAUS61 and MACS1188 with Harosoy by Clustal Omega (*https://www.ebi.ac.uk/Tools/msa/clustalo/*). Nucleotide bases are identical or similar in all sequences as indicated by *asterisk* respectively. Only JS2034 specific SNPs coloured in red, JS 9560 specific SNPs coloured in violet with white letter, blue colour with white letter indicates MAUS 61 specific SNPs, MACS1188 specific SNPs coloured in brown with white letter, Harosoy specific SNPs coloured in grey. SNPs were shared by two or more sequences like only earliness specific (JS2034 & JS9560) indicated by red letter, pink colour indicates SNPs shared by JS 2034 + Harosoy/JS9560 + Harosoy specific, green colour indicates SNPs specific to lateness specific (MAUS61 & MACS1188) or shared by one of two early genotypes or also present in Harosoy (early maturing) and yellow colour indicate SNPs shared by earliness genotypes (JS2034 & JS9560) with Harosoy respectively. Red coloured dotted line above the box indicates B3 like conserved domain in *E1* gene of early maturing genotypes (JS2034 and JS9560). Black coloured straight line below the box indicates B3 like conserved domain in *E1* gene of MAUS61 and MACS 1188.

Twenty-two SNPs were earliness specific shared by genotypes JS 20-34 and JS 95-60, out of which 17 were also shared with early Harosoy. Ten such maturity specific SNPs were distributed in between 121-180 base position, 2 SNPs were distributed in between 181-240 base position, 5 such SNPs were present in between 440-500 base position. Late maturing genotype-specific 17 SNPs were shared by both MAUS 61 and MACS 1188. Additional five SNPs were shared by them with early Harosoy. Thus, in overall genotypes MAUS 61 and MACS 1188 differed from both early maturing JS 95-60 + JS 20-34 at 23 SNP positions (Fig. 3).

Earlier six allelic variations (*e1-as, e1-fs, e1-nl, e1-re, e1-p* and *e1-b3a*) have been identified at *E1* locus (Xai *et al.,* 2012; Zhai *et al.,* 2015; Tsubokura *et al.,* 2014). In *e1-as,* a SNP at nucleotide 44 lead to a missense mutation, while in *e1-fs* another base deletion leads to a premature stop codon (Abe *et al.,* 2003). A null allele (*e1-nl*) with entire *E1* gene deleted has been reported (Xia *et al.,* 2012). They also detected 15 SNPs in *e1-as* allele. Tsubokura *et al.* (2014) reported two new alleles at *E1* loci., with *e1-re* having a long interspersed nuclear element (LINE) insertion upstream from start codon of the *E1* gene (Xu *et al.,* 2013) and *e1-P* having mutation only at 5'UTR region.

Soybean varieties adjust to temperate long-day conditions through the defective/mutant copies of either the flowering suppressor encoding *E1* gene, or two phytochromes encoding *E3 /E4* genes (Lin *et al*., 2022). PhyA-E1 photoperiodic pathway mediates delayed flowering at 35°C- under short day conditions, however, induction of early flowering at 30°C is independent of E1 suppressor (Tang *et al*., 2022).

Multiple sequence alignment of E3 PhyA3 gene regions

Out of the 1091 bp *E3 PhyA3*II gene region studied, only 991 bp clearly resolved the sequences of *E3 PhyA3*II region and were further investigated for variations. Initial 22 bases after *E3 PhyA3*II F primer sequence and terminal 34 base sequences before *E3 PhyA3*II R primer sequence were missing or not shared, so were not included in further studies. Out of the 991 base-stretch that showed continuous alignment, *E3 PhyA3*II region showed 99.19% homology in all the 5 sequences studied thereby indicating it to be highly conserved in soybean. A total of 8 SNP were observed in *E3 PhyA3*II region, of which 6 SNPs were genotype-specific substitutions (with 5 being MACS1188 specific). Another 210th SNP was earliness-specific shared by JS2034 and JS9560 (A), while early reference Harosoy shared G with both late maturing varieties. Another 390th base SNP was shared by JS9560 and MACS 61 having contrasting maturity.

In *E3 PhyA3*IV region (720 bp), 708 bp read length could be assembled in all the 4 genotypes. Out of these 691 bases matched (97.6%) in all the 5 sequences studied. This gene is highly conserved in soybean. A total 17 substitution SNPs were observed of which 15 were genotype-specific; while the rest were shared by 2 of the 3 early genotypes *viz*., 643rd and 653rd positions (shared by early reference genotype Harosoy with genotypes JS2034 and JS9560, respectively). Watanabe *et al*. (2009) and Xu *et al*. (2013) reported five different *E3* gene alleles in soybean *viz., E3Ha, e3-tr, E3Mi, e3-fs* and *e3-ns*. Gupta *et al*. (2017) studied genotypes of *E1* to *E4* loci in photoperiod insensitive genotypes to observe the recessive *e3-tr e3-tr* in 5 photoperiod insensitive genotypes (MACS 330, EC325097, EC34101, EC325118 and EC 390977*)* and a sensitive EC538828 accessions. They reported that widely adopted photoperiod sensitive JS 335 variety had *e3-fs e3-fs* alleles. Langewisch *et al*. (2017) investigated different Asian and American soybean genotypes for sequence variation of maturity loci and concluded that non-functional (di- and tri-genic) *e1/e2/e3/e4* allelic combinations contributed towards earliness with non-functional *e4* allele being rare and limited to north Japanese genotypes. The prevalence of *E3* in landraces was maintained in Chinese cultivars (83%) as well as in North American ancestors and, subsequently, in North American cultivars (73%).

Amino acids position polymorphisms (AAPP) in E1 protein

Out of the total 174 amino acid (aa) sequence, 126 aa were conserved in all the 5 protein sequences, with 49 aa positions being polymorphic. Earliness-specific 10 AAPPs were found to be shared by genotype JS2034 and JS9560 protein, and 6 of them also shared with early reference Harosoy. Five

maturity-specific AAPPs were distributed in between 37-60 aa position and one present at 160th aa position. Five polymorphisms were specific to all the 3 early varieties (including Harosoy) and both late varieties simultaneously. Two polymorphisms (51st and 75th position) were specific only to early Indian genotypes JS2034 and JS9560 (with Harosoy matching with late genotypes MAUS61 and MACS1188). Late maturing genotype-specific 13 AAPPs were shared by genotypes MAUS61 and MACS1180; with additional two also shared with early maturing Harosoy. Thus, in overall MAUS61 and MACS1180 differed from early maturing genotypes JS9560 + JS2034 at 15 AAPPs (Fig. 4). Xia *et al.* (2012) found that the conventional recessive allele (*e1-as*) differed from dominant allele *E1* by a single aa *i.e.* threonine to arginine substitution lead to the altered subcellular localization of protein. They also reported EMS mutants with missense mutations having early flowering phenotype *i.e.* serine to phenylalanine (*e1-m1*), arginine to lysine/ (*e1-m2*) and threonine to isolucine (*e1-m3*).

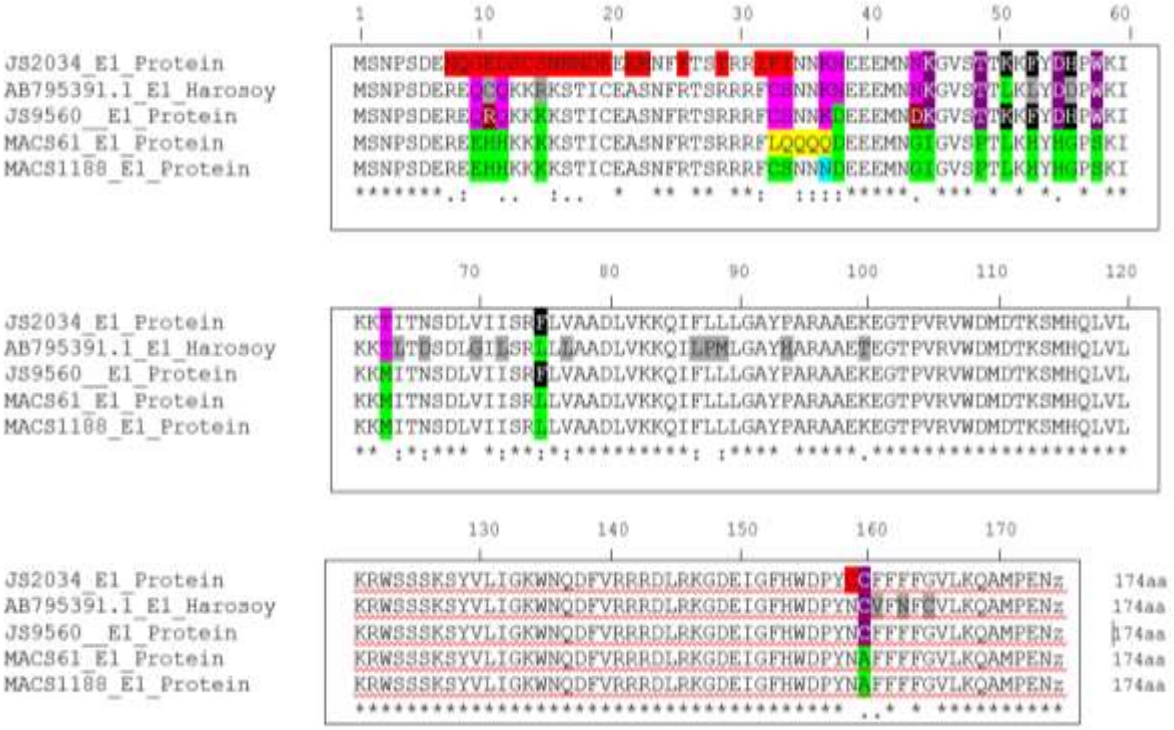

Fig. 4: Multiple sequence alignment of amino acid sequence of E1- JS2034, JS9560, MAUS61 and MACS1188 with E1 protein of Harosoy by using Clustal Omega (*https://www.ebi.ac.uk/Tools/msa/clustalo/*). Amino acid residues identical or similar in all sequences are indicated by an *asterisk* or *dot* respectively. Red colour indicates JS2034 specific polymorphism, gray colour shows Harosoy specific, MAUS61 specific polymorphism indicated in yellow colour, blue colour indicate MACS1188 specific only and brown colour with white letter gives JS9560 specific polymorphism. Polymorphism shared by two or more E1 amino acid residues *viz.*, earliness specific (including Harosoy) indicated in purple colour with white letter, early varieties specific (JS2034 & JS9560) indicated in black colour with white letter, pink colour indicate polymorphism shared by Harosoy+JS2034 or JS9560 or both early and green colour indicates lateness specific (MAUS61 & MACS1188) or lateness + Harosoy or JS9560.

Amino acids frame polymorphisms (AAPPs) in phytochrome A3 protein regions

Watanabe *et al.* (2009) suggested that phytochrome A contributes to the complex system of soybean flowering responses and geographic adaptation. As per reference Harosoy protein Accession BAH57344.1, the phytochrome A3 is 1130 aa long protein. High homology (98.18%) was observed in phytochrome A3-II peptide region (in between 342nd to 672nd aa) based on translated sequence with 323 of 329 aa being conserved. Six genotype-specific AAPPs were observed; of which 5 were MACS1188 specific and 1 MAUS61 specific. In PhyA3 IV peptide region (965th to 1060th aa), 88 out

of 96 aa were conserved in all the sequences studied with 6 AAPPs being genotype-specific; while 1 AAPP was shared by Harosoy with JS2034 (both early) and MAUS61 (late) each.

NCBI conserved domain analysis of E1 protein

In translated nucleotide sequence, a total of 11 conserved DNA binding sites of 3 bases each were present in B3 domain region. The domain was located in between 172 to 504 base interval in JS9560 and JS2034; while in MAUS61 and MACS1188 they were located in between 178-504 base intervals. B3 domain is a plant-specific DNA binding domain involved in a wide variety of plant development processes (Swaminathan *et al.,* 2008). Xia *et al*. (2012) demonstrated that B3 domain mutation (in *e1-as*) results in altered subcellular localization. Zhai *et al.* (2015) reported that in *e1-b3a* allele, 5 bp mutation (3 SNPs and 2 deletions) occurred in middle of B3 domain.

Conclusions: In present study, *E1* gene sequence analysis revealed 103 substitution SNPs and 22 of them were earliness-specific. Both late maturing varieties differed from early maturing varieties at 23 SNP positions. Moreover, sequence based markers can be designed from the identified *E1* gene polymorphic sequences and exploited for the selection of early maturing soybean genotypes.

Acknowledgements: The authors are thankful to the university authorities of Mahatma Phule Krishi Vidyapeeth, Rahuri, Maharashtra (India) for providing necessary support to conduct the present work. They are also grateful to the Soybean Breeder, V.N.M.KV., Parbhani and A.R.I., Pune (India) for providing a few soybean genotypes for this study.

REFERENCES

Abe, J., Xu, D.H., Miyano, A., Komatsu, K., Kanazawa, A. and Shimamoto, Y. 2003. Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. *Crop Science*, **43**: 1300-1304.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E. and Grosdidier, A. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, **40**(W1): W597-W603.

Doyle, J.J. and Doyle, J.V. 1987. A rapid DNA isolation procedure for small quantity of fresh leaf tissue. *Phytochemical Bulletin*, **19**: 11-15.

Gupta, S., Bhatia, V.S., Kumawat, G., Thakur, D., Singh, G., Tripathi, R., Satpute, G., Devadas, R., Husain, S.M. and Chand, S. 2017. Genetic analyses for deciphering the status and role of photoperiodic and maturity genes in major Indian soybean cultivars. *Journal of Genetics*, **96**(1): 147-154.

Hou, Z., Liu, B. and Kong, F. 2022. Regulation of flowering and maturation in soybean. *Advances in Botanical Research*, **102**: 43-75.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L. 2008. NCBI BLAST: A better web interface. *Nucleic Acids Research*, **36**(suppl.2): W5-W9. [*https://doi.org/10.1093/nar/gkn201*].

Kim, M., Shin, H., Kang, Y., Shim, S. and Lee, S. 2012. Divergence of flowering genes in soybean. *Journal of Bioscience*, **37**(5): 857-870.

Kong, F.J., Liu, B.H., Xia, Z.J., Sato, S. and Kim, B.M. 2010. Two co-ordinately regulated homologs of flowering locus T are involved in control of photoperiodic flowering in soybean. *Plant Physiology*, **154**: 1220-1231.

Langewisch, T., Lenis, J., Jiang, G.L., Wang, D., Pantalone, V. and Bilyeu, K. 2017. The development and use of a molecular model for soybean maturity groups. *BMC Plant Biology*, **17**(1): 91. [*https://doi.org/10.1186/s12870-017-1040-4*].

Lin, X., Dong, L., Tang, Y., Li, H., Cheng, Q., Li, H., Zhang, T., Ma, L., Xiang, H., Chen, L. and Nan, H. 2022. Novel and multifaceted regulations of photoperiodic flowering by phytochrome

A in soybean. *Proceedings of the National Academy of Sciences*, **119**(41): p.e2208708119. [*doi:10.1073/pnas.2208708119. Epub 2022 Oct 3. PMID: 36191205; PMCID: PMC9565047*].

Lin, X., Liu, B., Weller, J.L., Abe, J. and Kong, F. 2021. Molecular mechanisms for the photoperiodic regulation of flowering in soybean. *Journal of Integrative Plant Biology*, **63**(6): 981-994.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I. and Lanczycki, C.J. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research*, **43**(D1): D222-D226.

No, D.H., Baek, D., Lee, S.H., Cheong, M.S., Chun, H.J., Park, M.S., Cho, H.M., Jin, B.J., Lim, L.H., Lee, Y.B. and Shim, S.I. 2021. High-temperature conditions promote soybean flowering through the transcriptional reprograming of flowering genes in photoperiod pathway. *International Journal of Molecular Sciences*, **2B2**(3): 1314 [*https://doi.org/10.3390/ijms22031314*].

Roux, E., Touzet, P., Cuguen, J. and Corre, V.L. 2006. How to be early flowering: An evolutionary perspective. *Trends in Plant Science*, **11**: 375-381.

Sievers, F. and Higgins, D.G. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. pp. 105-116. **In**: *Multiple Sequence Alignment Methods*. Springer, [*doi:10.1002/0471250953.bi0313s48*].

Swaminathan, K., Peterson, K. and Jack, T. 2008. The plant B3 superfamily. *Trends in Plant Science*, **13**: 647-655.

Tang, Y., Lu, S., Fang, C., Liu, H., Dong, L., Li, H., Su, T., Li, S., Wang, L., Cheng, Q. and Liu, B. 2022. Diverse flowering responses subjecting to ambient high temperature in soybean under short-day conditions. *Plant Biotechnology Journal*, **21**(4): 782-791.

Thakare, D., Kumudini, S. and Dinkins, R.D. 2011. The alleles at the E1 locus impact the expression pattern of two soybean FT like genes shown to induce flowering in *Arabidopsis*. *Planta*, **234**: 933-943.

Tsubokura, Y., Watanabe, S., Xia, Z., Kanamori, H., Yamagata, H., Kaga, A., Katayose, Y., Abe, J., Ishimote, M. and Harada, K. 2014. Natural variation in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. *Annals of Botany*, **113**: 429-411.

Watanabe, S., Harada, K. and Abe, J. 2012. Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breeding Science*, **61**: 531-543.

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

Xia, Z., Watanabe, S., Yamadac, T., Tsubokurab, Y., Nakashimad, H., Zhaia, H., Anaid, T., Sato, S., Yamazakif, T., Lua, S., Wua, H., Tabatae, S. and Haradab, K. 2012. Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proceeding of the National Academy of Science (USA)*. **109**: 2155-2164.

Xu, M., Xu, Z., Liu, B., Kong, F., Tasubokaru, Y., Watanabe, S., Xiu Z., Harada, K., Kanazawa, A., Yamada, T. and Abe, J. 2013. Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA regulated post flowering responses in soybean. *BMC Plant Biology*, **13**: 1-14.

Zhai, H., Lu, S., Wu, H., Zhang, Y., Zhang, X., Wang, Y., Yang, G., Cui, T. and Xia, Z. 2015. Diurnal expression pattern, allelic variation, and association analysis reveal functional feature of *E1* gene in control of photoperiodic flowering in soybean. *PLoS One*, **10**: 1-16.