RESEARCH ARTICLE

Genetic relationship, population structure analysis and allelic characterization of flowering and maturity genes *E1*, *E2*, *E3* and *E4* among 90 Indian soybean landraces

Giriraj Kumawat¹ · Arti Yadav² · Gyanesh K. Satpute¹ · C. Gireesh¹ · Rakesh Patel¹ · M. Shivakumar¹ · Sanjay Gupta¹ · Suresh Chand² · Virender Singh Bhatia¹

Received: 16 August 2018 / Revised: 3 October 2018 / Accepted: 8 October 2018 / Published online: 1 January 2019
© Prof. H.S. Srivastava Foundation for Science and Society 2019

Abstract A set of 90 Indian soybean landraces were analysed for polymorphism at 43 SSRs and five allele specific markers of four major genes involved in regulating flowering and photoperiod response. A total of 42 polymorphic SSRs had amplified 126 alleles which served as raw data for estimation of genetic relationship and population structure among 90 accessions. Rare alleles of four and three SSRs were detected in accessions IC18768 and IC15089, respectively. Gene diversity in the population ranges from 0.065 to 0.717 with a mean value of 0.411. The polymorphism information content of 42 SSRs varied from 0.063 to 0.668. Hierarchical clustering based on neighbour-joining method identified three major clusters among 90 soybean accessions. Model based population structure analysis divided the 90 soybean accessions into four populations ($K = 4$). Mean value of F_{st} for different populations ranged between 0.4143 and 0.7239. Genotyping of 90 accessions with allele specific markers had identified accession IC15089 as triple recessive mutant of flowering genes *E1*, *E2* and photoperiod sensitivity gene *E3*. The triple mutant IC15089 (*e1*, *e3*, *e3*) had been characterized phenotypically and identified as early maturing (88 days) and photoperiod insensitive genotype under extended photoperiod. The present study

characterized genetic relationship among 90 Indian soybean landraces and had identified a few diverse and unique genotypes for utilization in soybean breeding programmes targeting development of short duration and photoperiod insensitive varieties through marker assisted selection.

Keywords Genetic relationship · Population structure · Landraces · Flowering · Photoperiod · Soybean

Introduction

Soybean [*Glycine max* (L.) Merrill] is major oilseed crop of India grown over diverse climatic conditions with a total production of 14.0 MT (FAOSTAT 2018). Breeding programmes which aims to increase the soybean productivity and sustainable production requires utilization of diverse genotypes harbouring desirable alleles of agronomic traits. India possesses 3637 germplasm accessions of soybean comprising exotic collection (EC), Indigenous collections (IC), landraces and local cultivars (<http://www.nbpgr.ernet.in:8080/PGRPortal>). A total of 3443 accessions from this collection were characterized for eight agro-morphological characters and an integrated core collection of 713 accessions was developed by Gireesh et al. (2017). Although majority of the Indian soybean germplasm collection have been characterized for agro-morphological characters, molecular characterization of this collection is lacking. Molecular characterization of germplasm collection can facilitate identification of diverse genotypes as well as sources for unique and rare alleles which are very useful for molecular mapping, genotype identification and gene introgression to cultivated soybean. Even though, several reports are available on molecular characterization of cultivated varieties of soybean in India (Satyavathi et al. 2006; Kondetti

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12298-018-0615-3>) contains supplementary material, which is available to authorized users.

✉ Giriraj Kumawat
Giriraj.Kumawat@icar.gov.in

¹ ICAR-Indian Institute of Soybean Research, Indore, Madhya Pradesh 452001, India

² School of Life Sciences, Devi Ahilya Vishwavidyalaya, Indore, Madhya Pradesh 452001, India

et al. 2012; Ghosh et al. 2014; Bisen et al. 2015; Chauhan et al. 2015, 2017; Gupta and Manjaya 2017), only few studies are available on molecular characterization of soybean germplasm accessions (Singh et al. 2008a, b, 2010; Kumawat et al. 2015; Hipparagi et al. 2017).

Singh et al. (2008b, 2010) characterized a set of 44 soybean accessions differing in photoperiod sensitivity by using RAPD, SSR and AFLP markers. Kumawat et al. (2015) characterized a set of 82 soybean accessions comprising EC, IC and local cultivars by using simple sequence repeat (SSR) markers. These two studies could identify few diverse genotypes from random sets of accessions selected from entire germplasm population. But the genetic relationship within Indigenous collection subpopulation which comprises of Indigenous landraces has not been well characterized. Recently, Hipparagi et al. (2017) used a set of 75 soybean genotypes collected from Uttarakhand state of India to estimate the genetic variability and population structure using SSR markers. However, this study was also limited to only bhat soybean landraces, a local collection of soybean grown in Uttarakhand state of India. In the present study, a larger set of 90 IC accessions originally collected from different states of India has been used for identifying genetic relationship, diversity and population structure among these IC accessions using SSR markers. Further, we have also characterized these IC accessions for allelic status of four major genes *E1* (B3 DNA binding protein), *E2* (*GIGANTEA*), *E3* (*Phytochrome A3*) and *E4* (*Phytochrome A2*), which are involved in regulating flowering, maturity and photoperiod response in soybean.

Soybean crop has adaptation to a wide range of latitudes owing to natural variation in the major genes and quantitative trait loci (QTLs) regulating flowering and maturity duration. Ten flowering and maturity loci, designated as *E* loci, have been characterized in soybean (Tsubokura et al. 2014; Gupta et al. 2017). Of these 10 major genes reported, dominant alleles at *E1*, *E2*, *E3*, *E4*, *E5*, *E7* and *E8* loci are known to delay flowering time, whereas, dominant alleles at *E6*, *E9* and *J* loci promote flowering to different extents (Watanabe et al. 2012; Kong et al. 2014). The *E1*, *E3*, *E4* and *E7* loci were reported as quantitative photoperiodic loci (Cober and Voldeng 2001; Watanabe et al. 2012). *E3* locus controls flowering under long day (LD) conditions with a high red (R) to far red (FR) quantum ratio, created by R-enriched fluorescent lamps, where *e3e3* recessive homozygous genotypes can initiate flowering under such conditions (Buzzell 1971). *E4* locus is involved in flowering under LD conditions with a low R: FR quantum ratio generated by FR-enriched incandescent lamps, where recessive *e4e4* genotype is necessary to initiate flowering under such conditions (Buzzell and Voldeng 1980; Cober et al. 1996). The underlying genes of *E1*, *E2*, *E3*, *E4*, and *E9* loci were identified as B3 DNA binding protein (*E1*), *GIGANTEA* (*E2*),

Phytochrome A3 (*E3*), *Phytochrome A2* (*E4*) and *GmFT2a* (*E9*) (Liu et al. 2008; Watanabe et al. 2009, 2011; Xia et al. 2012; Zhao et al. 2016). Based on nucleotide sequence variation, mutant alleles of these genes have been characterized and allele specific markers were developed (Watanabe et al. 2009, 2011; Xia et al. 2012; Xu et al. 2013; Zhao et al. 2016). These allele specific markers of *E1*, *E2*, *E3* and *E4* genes have been used in present study to identify recessive mutants of respective genes in 90 IC accessions.

Materials and methods

Plant material and genotyping of SSRs

A set of 90 IC accessions originally collected from different states of India as landraces were obtained from ICAR-National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India and grown at ICAR-Indian Institute of Soybean Research, Indore, India. A widely adapted cultivar JS335 was also included in this set. The details of each genotype along with genebank ID i.e. Indigenous Collection (IC) number, biological status, Indian State of origin/collection, days to flowering and days to maturity is given in Table 1. Genomic DNA from leaf samples was isolated using CTAB method (Doyle and Doyle 1990). A total of 43 SSR markers randomly selected from 20 chromosomes were used for genotyping (Additional file 1: Table S1). Genotyping of SSRs was performed as described in Kumawat et al. (2015).

Genotyping with allele specific markers

Allele specific derived cleaved amplified polymorphic sequence (dCAPS) markers for flowering and maturity genes *E1* and *E2* were genotyped as reported in Xia et al. (2012) and Watanabe et al. (2011), respectively. Briefly, *E1* gene fragment containing mutant allele site was amplified by polymerase chain reaction (PCR) and treated with restriction enzyme *TaqI* for detection of *e1-as* allele. Similarly, *E2* gene fragment containing mutant allele site was amplified by PCR and treated with restriction enzyme *DraI* for detection of *e2-ns* allele. The restriction enzyme treated samples were visualized on 3% metaPhor agarose for detection of mutant alleles. Genotyping for detection of *e1-nl* was performed with the PCR primers reported by Xu et al. (2013). Allele specific fragment length polymorphism (FLP) markers of two photoperiod sensitivity genes *E3* and *E4* were genotyped as reported by Liu et al. (2008) and Watanabe et al. (2009), respectively. PCR primer sequences of the allele specific markers used for genotyping of *E1*, *E2*, *E3* and *E4* genes are given in Table S2 of Additional file 1.

Table 1 List of soybean accessions used in the study with their IC numbers, biological status, collection state along with days to flowering and maturity at Indore (22°44'N, 75°50'E)

S. no.	IC number	Biological status	Collection state	Days to flowering	Days to maturity
1	IC191	Landrace	Bihar	49	105
2	IC192	Landrace	Bihar	62	121
3	IC195	Landrace	Bihar	58	121
4	IC196	Landrace	Bihar	–	–
5	IC202	Landrace	Bihar	52	104
6	IC206	Landrace	Bihar	52	104
7	IC207	Landrace	Bihar	52	113
8	IC208	Landrace	Bihar	52	113
9	IC209	Landrace	Bihar	68	121
10	IC210	Landrace	Bihar	68	113
11	IC212	Landrace	Bihar	68	104
12	IC217	Landrace	Bihar	52	104
13	IC222	Landrace	Bihar	68	113
14	IC224	Landrace	Bihar	44	104
15	IC574	Landrace	Punjab	52	107
16	IC2061	Landrace	West Bengal	34	108
17	IC7217	Landrace	Unknown	44	113
18	IC7469	Landrace	Uttar Pradesh	42	104
19	IC7842	Landrace	Odisha	36	92
20	IC7844	Landrace	Odisha	36	92
21	IC7846	Landrace	Unknown	36	89
22	IC7848	Landrace	Unknown	36	89
23	IC9444	Landrace	Sikkim	47	101
24	IC9445	Landrace	Sikkim	44	113
25	IC9451	Landrace	Sikkim	44	113
26	IC9459	Landrace	Sikkim	47	108
27	IC9461	Landrace	Sikkim	44	108
28	IC9468	Landrace	Sikkim	52	108
29	IC9473	Landrace	Sikkim	47	104
30	IC10038	Landrace	Himachal Pradesh	34	96
31	IC10678	Landrace	Arunachal Pradesh	36	101
32	IC13005	Landrace	West Bengal	52	104
33	IC13007	Landrace	West Bengal	47	108
34	IC13009	Landrace	West Bengal	47	108
35	IC13047	Landrace	West Bengal	–	–
36	IC13049	Landrace	West Bengal	47	108
37	IC13050	Landrace	West Bengal	–	–
38	IC13052	Landrace	West Bengal	44	108
39	IC13053	Landrace	West Bengal	49	104
40	IC13055	Landrace	West Bengal	47	104
41	IC13056	Landrace	West Bengal	49	107
42	IC14683	Landrace	Assam	44	121
43	IC14684	Landrace	West Bengal	71	121
44	IC15089	Landrace	Himachal Pradesh	26	88
45	IC15092	Landrace	Himachal Pradesh	–	–
46	IC15539	Landrace	Meghalaya	44	113
47	IC15541	Landrace	Meghalaya	38	108
48	IC15544	Landrace	Meghalaya	–	–
49	IC15545	Landrace	Meghalaya	34	96
50	IC15546	Landrace	Meghalaya	42	113

Table 1 continued

S. no.	IC number	Biological status	Collection state	Days to flowering	Days to maturity
51	IC15552	Landrace	Meghalaya	47	108
52	IC15553	Landrace	Meghalaya	42	108
53	IC15759	Landrace	Jharkhand	49	108
54	IC15767	Landrace	Jharkhand	40	96
55	IC15959	Landrace	Madhya Pradesh	52	113
56	IC16171	Landrace	Maharashtra	58	113
57	IC16438	Landrace	Maharashtra	68	113
58	IC16825	Landrace	Manipur	47	108
59	IC16832	Landrace	Manipur	58	108
60	IC16833	Landrace	Manipur	52	105
61	IC16971	Landrace	Manipur	58	113
62	IC16974	Landrace	Manipur	47	105
63	IC16975	Landrace	Manipur	49	101
64	IC18384	Landrace	Himachal Pradesh	44	105
65	IC18733	Landrace	Himachal Pradesh	38	96
66	IC18734	Landrace	Himachal Pradesh	36	96
67	IC18736	Landrace	Himachal Pradesh	44	105
68	IC18739	Landrace	Himachal Pradesh	42	93
69	IC18743	Landrace	Himachal Pradesh	42	93
70	IC18746	Landrace	Himachal Pradesh	34	79
71	IC18748	Landrace	Himachal Pradesh	38	96
72	IC18749	Landrace	Himachal Pradesh	36	96
73	IC18750	Landrace	Himachal Pradesh	44	96
74	IC18753	Landrace	Himachal Pradesh	38	96
75	IC18754	Landrace	Himachal Pradesh	38	96
76	IC18756	Landrace	Himachal Pradesh	38	101
77	IC18757	Landrace	Himachal Pradesh	42	84
78	IC18759	Landrace	Himachal Pradesh	38	96
79	IC18761	Landrace	Himachal Pradesh	38	96
80	IC18765	Landrace	Himachal Pradesh	38	96
81	IC18767	Landrace	Himachal Pradesh	44	101
82	IC18768	Landrace	Himachal Pradesh	34	89
83	IC19486	Landrace	Himachal Pradesh	38	105
84	IC19917	Landrace	Bihar	71	108
85	IC21748	Landrace	Madhya Pradesh	36	108
86	IC24057	Landrace	Uttar Pradesh	49	105
87	IC24064	Landrace	Uttar Pradesh	49	107
88	IC24068	Landrace	Uttar Pradesh	34	101
89	IC24069	Landrace	Uttar Pradesh	47	105
90	JS335	Cultivar	Madhya Pradesh	45	102

SSR allele scoring and statistical analysis

The fragment size of amplified DNA products was determined by comparing with 50 bp DNA ladder (Thermo scientific, USA). The SSR bands scored in soybean accessions were used for genetic diversity analysis. Major

allele frequency, gene diversity, polymorphic information content (PIC) and heterozygosity for each SSR locus were calculated using Power Marker 3.25 and a phylogenetic tree was constructed using neighbour-joining method (Liu and Muse 2005). Population structure was inferred by using the model-based program, STRUCTURE 2.3.4

(Pritchard et al. 2000). For each K , three replications were run where each run was implemented over a burn-in period of 50,000 steps and 100,000 Monte Carlo Markov Chain replicates. The membership of each genotype was run for $K = 1$ –20 by taking admixture model and correlated allele frequency into account. The final population was determined by applying the second-order statistics (ΔK) developed by Evanno et al. (2005) using ‘Structure harvester’ software (Earl and vonHoldt 2012). Venn diagram analysis was performed to identify common accessions between clusters and populations using online software Venny 2.1 (Oliveros 2007).

Phenotypic analysis of flowering and photoperiod sensitivity

Based on allele specific marker genotyping results, one genotype, IC15089 was identified as triple recessive mutant of flowering and photoperiod sensitivity genes *E1*, *E2* and *E3*. The triple mutant IC15089 (*e1e1e2e2e3e3*) was hybridized with previously identified single mutant JS95-60 (*e1e1*) and double mutant EC538828 (*e2e2e3e3*) for phenotypic analysis in F_1 s (Gupta et al. 2017). The identified triple recessive mutant of *e1*, *e2*, *e3* (IC15089) along with four F_1 s of IC15089 $\times$ EC538828 (*e1e1e2e2e3e3*), four F_1 s of IC15089 $\times$ JS95-60 (*e1e1E2e2E3e3*) and two F_1 s of EC538828 $\times$ JS95-60 (*E1e1E2e2E3e3*), were phenotyped under fluorescent light in a glasshouse where photoperiod was extended for 24 h, starting from 20 days after sowing till 60 days after sowing. Single recessive mutants for *e1* (JS95-60), double recessive mutants for *e2*, *e3* (EC538828) and wild type genotypes for *E1*, *E2*, *E3* (JS335) were used as control. Ninety soybean accessions were planted in augmented design on 09th July 2014 at Indore, India (22°44'N, 75°50'E). Each plot consisted of single 3 m row having 50–60 plants. Phenotypic observations for flowering and maturity duration were recorded as reported by Gireesh et al. (2017) and available on National Bureau of Plant Genetic Resources portal (<http://www.nbpgri.ernet.in:8080/PGRPortal>).

Results

Genetic diversity of SSRs

A total of 43 SSR markers were used for genotyping of 90 soybean accessions. Out of 43 SSR markers, 42 SSRs were found polymorphic. These 42 polymorphic SSRs amplified a total of 126 alleles. The number of alleles amplified per SSR locus ranged from 2 to 4 with an average of 3 alleles per SSR locus (Table 2). The major allele frequency was lowest for sat_252 (0.36) and highest for satt164 (0.966).

Gene diversity in the population varied from 0.717 (sat_328) to 0.065 (satt164) with a mean value of 0.411. The polymorphism information content (PIC) varied from 0.063 (satt164) to 0.668 (satt328) with a mean value of 0.356. Maximum heterozygosity observed was 0.157 for sat_390 whereas 35 SSRs showed no heterozygosity. Four SSRs had displayed unique alleles among five genotypes (Table 3). In addition, 16 rare alleles (allele having frequency of less than 5%) were detected for 13 SSR loci. Rare alleles of four and three SSRs were detected in accessions IC18768 and IC15089, respectively.

Hierarchical cluster analysis

Hierarchical clustering based on neighbour-joining method identified three major clusters among 90 soybean accessions (Fig. 1). These IC accessions were originally collected from 13 states of India which can be divided into four geographical regions namely Northern region [23° to 33°N, 74° to 84°E (Himachal Pradesh, Uttar Pradesh and Punjab)], Eastern region [17°–27°N, 81°–90°E (Bihar, West Bengal and Odisha)], North-Eastern region [20°–29°N, 85°–97°E (Manipur, Meghalaya, Assam, Arunachal Pradesh and Sikkim)] and Central region [16°–27°N, 72°–83°E (Madhya Pradesh and Maharashtra)]. The cluster analysis distinguished majority of collection according to geographical origin. The first major cluster contains 4 accessions (IC195, IC13009, IC13053, IC15089), majority of which were collected from Eastern part (3 out of 4). Accession IC195 was collected from Bihar, IC13009 and IC13053 from West Bengal and IC15089 from Himachal Pradesh. Majority of accessions (24 out of 29) collected from Northern part were grouped in cluster 2 (Fig. 1). Whereas majority of accessions (17 out of 22) collected from North-Eastern part were grouped in cluster 3. Although majority of accessions (16 out of 31) collected from Eastern part were grouped in cluster 3 but had also good representation in cluster 1 and 2. This indicates the genetic base of germplasm collected from Eastern part is originally derived from Northern and North-Eastern gene pool.

Population structure analysis

Allelic information of 42 SSRs among 90 IC accessions was used to infer the population structure using Structure software. The Structure results were used in Structure harvester software to infer the best K . A sharp peak in ΔK at $K = 4$ was observed on applying the second-order statistics (ΔK), which suggested the presence of four major populations (Fig. 2a). Thus, population structure analysis divided the 90 soybean accessions into four populations consisting of 10–37 accessions (Fig. 2b). A total of seventeen pure and 20 admixed accessions were present in population (pop) 1; nine pure and ten admixed accessions

Table 2 Genetic diversity parameters estimated for 42 SSR markers in 90 soybean accessions

S. no.	SSR name	Major Allele frequency	Allele number	Gene diversity	Heterozygosity	PIC
1	Satt155	0.8667	4	0.2427	0.0000	0.2334
2	Satt200	0.7614	2	0.3634	0.0000	0.2974
3	Satt717	0.8929	2	0.1913	0.0000	0.1730
4	Sat_406	0.6292	4	0.5300	0.0000	0.4703
5	Satt409	0.7111	3	0.4279	0.0000	0.3606
6	BE806308	0.8636	2	0.2355	0.0909	0.2078
7	Satt484	0.8444	2	0.2627	0.0000	0.2282
8	Satt126	0.5556	2	0.4938	0.0000	0.3719
9	Satt687	0.8000	3	0.3360	0.0000	0.3053
10	Satt164	0.9663	2	0.0651	0.0000	0.0630
11	Satt396	0.7667	3	0.3627	0.0000	0.3045
12	Sat_251	0.7108	2	0.4111	0.0000	0.3266
13	Sat_252	0.3605	3	0.6655	0.0000	0.5914
14	Satt557	0.5111	4	0.6368	0.0000	0.5790
15	Satt129	0.4651	4	0.6520	0.0000	0.5875
16	Satt179	0.5172	4	0.5578	0.0000	0.4632
17	Satt184	0.4770	4	0.5540	0.0805	0.4539
18	Sat_201	0.5783	3	0.4976	0.0000	0.3852
19	Satt580	0.9419	3	0.1111	0.0000	0.1080
20	Sat_227	0.6180	3	0.5441	0.0000	0.4850
21	Satt459	0.7816	2	0.3414	0.0000	0.2831
22	Satt310	0.7444	2	0.3805	0.0000	0.3081
23	Satt328	0.4045	4	0.7171	0.1348	0.6687
24	Satt411	0.5056	4	0.5214	0.0674	0.4073
25	Sat_390	0.4663	3	0.6318	0.1573	0.5573
26	Satt163	0.8556	3	0.2546	0.0000	0.2348
27	Satt517	0.7931	4	0.3517	0.0000	0.3273
28	Sat_218	0.9438	4	0.1078	0.0000	0.1053
29	Satt666	0.5632	3	0.5493	0.0000	0.4630
30	Satt127	0.9659	3	0.0664	0.0000	0.0652
31	Satt270	0.4659	3	0.5860	0.0000	0.4966
32	Satt414	0.6333	4	0.5267	0.0000	0.4672
33	Sat_393	0.5111	3	0.5894	0.0000	0.5073
34	Sat_167	0.7102	3	0.4528	0.0114	0.4088
35	Satt055	0.6420	4	0.5381	0.0568	0.4960
36	Satt076	0.9318	2	0.1271	0.0000	0.1190
37	Satt286	0.5056	2	0.4999	0.0000	0.3750
38	Sat_316	0.8556	3	0.2526	0.0000	0.2300
39	Sct_195	0.9101	2	0.1636	0.0000	0.1502
40	Satt022	0.6292	3	0.4893	0.0000	0.3982
41	Sat_190	0.6333	3	0.5242	0.0000	0.4625
42	Sat_196	0.6854	3	0.4767	0.0000	0.4254
Mean		0.6899	3.0	0.4117	0.0143	0.3560

in pop2; five pure and five admixed accessions in pop3; eleven pure and thirteen admixed accessions were present in pop 4, where membership fraction value of more than 0.90 was considered as pure and less than 0.90 as admixture. Overall proportion of membership of the sample in

the four populations ranged from 0.148 for population 3 to 0.366 for pop1, whereas average distances between individuals in same cluster ranged from 0.151 for pop2 to 0.334 for pop3 (Table 4). Mean value of Wrights fixation index (F_{st}) ranged between 0.7239 (highest) for pop2 and

S. no.	SSR name	Unique allele size (bp)	Accessions showing unique allele
1	Satt396	120	IC21748
2	Satt411	120	IC24057
3	Sat_218	190	IC18743
		270	IC15541
4	Satt127	180	IC19917

Legend:

- Green circle: Northern region
- Dark blue circle: Eastern region
- Red circle: North Eastern region
- Yellow circle: Central region
- White circle: Unknown

pop1 and 90 accessions of three clusters showed that 33 accessions (89.18%) of pop1 were common with cluster III (Fig. 3a). Similarly, Venn diagram showed that 17 of the 19 accessions (89.47%) were common between pop2 and cluster II (Fig. 3b). All of the 10 accessions of pop3 were common with cluster II (Fig. 3c). The Venn diagram of 24 accessions of pop4 and three clusters showed that 18 accessions (75%) of pop4 were common with cluster II (Fig. 3d). Though co-linearity of grouping soybean accessions based on hierarchical cluster and model based population approach is not very high, possibly due to higher admixture, but has shown fairly good level of similarity ($> 75\%$).

Allelic characterization for flowering and photoperiod sensitivity genes

Allele specific dCAPS marker genotyping for *E1* gene had identified two accessions with recessive *e1-as* allele (IC15089, IC10038), whereas dCAPS analysis of *E2* gene identified one accession having recessive *e2-ns* (IC15089) among 90 accessions (Fig. 4; Additional file 2: Figure S1

Fig. 2 **a** Determination of final population using $\ln P(D)$ derived ΔK for K sizes of 1–20; **b** Population structure of 90 soybean accessions based on SSR genotyping; each accession is represented by a single vertical line and numbers represent soybean accessions as listed in Table 1

Table 4 Details of membership parameters observed among four populations

Population	No. of accessions	Overall proportion of membership of the sample	Average distances between individuals in same cluster	Mean value of F_{st}
Pop1	37	0.366	0.3053	0.4143
Pop2	19	0.228	0.1515	0.7239
Pop3	10	0.148	0.3347	0.4407
Pop4	24	0.257	0.2864	0.4217

Table 5 Allele-frequency divergence among populations computed using estimates of P (Model based approach)

	POP1	POP2	POP3	POP4
POP1	–	0.2100	0.2336	0.1466
POP2	0.2100	–	0.2968	0.1466
POP3	0.2336	0.2968	–	0.2283
POP4	0.1466	0.1466	0.2283	–

and S2). Marker genotyping for *e1-nl* allele had not detected any genotype with *e1-nl* allele (Additional file 2: Figure S3). Allele specific FLP genotyping of 90 accessions for *E3* gene had identified recessive allele *e3-tr* in accession IC15089 (Fig. 4; Additional file 2: Figure S4). These results identified that accession IC15089 is a triple recessive for flowering genes *E1*, *E2* and photoperiod sensitivity gene *E3*. Genotyping with allele specific FLP marker for *E4* had not detected recessive allele in any of

the 90 accessions (Additional file 2: Figure S5). Phenotypic observations of triple recessive genotype IC15089 for flowering and maturity time were recorded as 26 days and 88 days, respectively. The phenotype of IC15089 for photoperiod insensitivity was confirmed by phenotyping in extended photoperiod with fluorescent light. Since all the three genes have dominant effect, their phenotypic effect can be studied in F_1 s. F_1 s were produced in three different combinations of three genes having heterozygous alleles in F_1 and homozygous recessive genotypes were used as control for respective gene combination (Table 6, Additional file 2: Figure S6). It was observed that when all three genes were in heterozygous conditions (F_1 of JS95-60 $\times$ EC538828), flowering was not initiated until extended photoperiod was withdrawn; similar is the case when *E1* gene was in homozygous recessive condition (F_1 of IC15089 $\times$ JS95-60). However, when *E2* and *E3* gene was

Fig. 3 **a** Venn diagram depicting co-linearity between pop1 and Cluster I, II and III; **b** Between pop2 and cluster I, II and III; **c** Between pop3 and cluster I, II and III; **d** Between pop4 and cluster I, II and III

Fig. 4 Detection of mutant alleles for *E1*, *E2* and *E3* genes by dCAPS and FLP markers; *M* = 50 bp DNA ladder; 1, 3 and 2, 4 = *E1* gene in IC10038 and IC15089 before and after restriction digestion with *TaqI*, respectively; 5 and 6 = *E2* gene in IC15089 before and after restriction digestion with *DraI*; 7 = *e3-tr* allele in IC15089

in recessive condition (F_1 of IC15089 $\times$ EC538828), there was no significant delay in initiation of flowering (Table 6).

Discussion

Identification of diverse genetic sources is desirable for broadening genetic base and incorporation of new alleles for creating better recombinants in crops. SSR markers

have been used extensively for evaluation of genetic diversity and population structure analysis in crop plant due to their higher polymorphism rate, multi-allelic nature, relative abundance and high reproducibility. Also, the fact that molecular marker based diversity and population structure analysis is not influenced by environmental factors, gives another reason for undertaking molecular marker based studies (Vinu et al. 2013). SSR markers have been used in several studies for characterization of genetic diversity in soybean germplasm and cultivars (Abe et al. 2003; Wang et al. 2006, 2010; Fu et al. 2007; Wang and Takahata 2007; Li et al. 2008; Singh et al. 2010; Ghosh et al. 2014; Kumawat et al. 2015; Bisen et al. 2015; Chauhan et al. 2015, 2017; Gupta and Manjaya 2017; Hipparagi et al. 2017).

In this study, 42 polymorphic SSRs used for characterization of 90 Indian soybean landraces had amplified an average of 3 alleles per SSR locus and observed mean gene diversity value of 0.41. Hipparagi et al. (2017) reported an average of 2.61 alleles per locus with observed mean gene diversity value of 0.43 among 75 bhat soybean landraces of India. The allele number per locus amplified in this study was comparable with 2.97 reported by Kumawat et al. (2015) but lower than earlier reports by Singh et al. (2010)

Table 6 Phenotypic evaluation of recessive mutants and F_1 s for days to flowering under extended photoperiod created by illuminating 24 h continuous light with Red enriched fluorescent lamps

Genotype name	Allelic status of <i>E1</i> , <i>E2</i> and <i>E3</i>	Days to flower
IC15089	<i>e1e1e2e2e3e3</i>	26
EC538828	<i>E1E1e2e2e3e3</i>	28
JS95-60	<i>e1e1E2E2E3E3</i>	67
JS335	<i>E1E1E2E2E3E3</i>	70
F_1 (IC15089 $\times$ JS95-60)	<i>e1e1E2e2E3e3</i>	68
F_1 (JS95-60 $\times$ EC538828)	<i>E1e1E2e2E3e3</i>	70
F_1 (IC15089 $\times$ EC538828)	<i>E1e1e2e2e3e3</i>	28

among 44 germplasm accessions, Chauhan et al. (2017) and Gupta and Manjaya (2017) among 78 and 90 soybean cultivars of India, respectively. Multiple alleles at SSR loci are generated by variation in repeat numbers created by a combined mechanism of length dependent DNA replication slippage and indel slippage with base substitutions (Dieringer and Schlötterer 2003). In our study, unique alleles were identified for four SSRs which are lower than five unique alleles detected by Kumawat et al. (2015) and 19 unique alleles detected by Gupta and Manjaya (2017). However, there were 16 rare alleles detected for 13 SSRs which is higher than five rare alleles reported by Kumawat et al. (2015), but lower than 54 rare alleles reported by Gupta and Manjaya (2017). In this study, two accessions, IC18768 and IC15089 were identified as good source of rare alleles and thus promising for soybean breeding. These results showed that Indian soybean landraces are an important source of rare alleles for increasing genetic diversity in soybean breeding programmes.

The mean PIC value observed in our study was 0.356, which is comparable with 0.36 reported by Hipparagi et al. (2017) but less than earlier reports (Kumawat et al. 2015; Chauhan et al. 2017; Gupta and Manjaya 2017). In our study, a very low level of heterozygosity (0.014) was observed which is comparable with 0.014 reported by Li et al. (2008) but less than other reports (Wang et al. 2015; Gupta and Manjaya 2017; Hipparagi et al. 2017). The gene diversity observed in this germplasm set was 0.41, which is comparable with 0.43 reported by Hipparagi et al. (2017) but lower than reported by Wang and Takahata (2007) and Hudcovicova and Kraic (2003) on two different sets of soybean genotypes. Thus the PIC, heterozygosity and gene diversity values are highly dependent on type and size of germplasm population being evaluated.

Hierarchical clustering had divided 90 Indian soybean landraces into three clusters and a good correspondence was observed with respect to genotypes geographical origin. Genotypes collected from Northern region were grouped mainly in cluster 2, whereas genotypes collected from North-Eastern region were grouped in cluster 3 and genotypes collected from Eastern region were distributed mainly in cluster 1 and 3. Earlier studies on SSR based

grouping of Indian soybean germplasm also reported three major clusters among analysed population (Kumawat et al. 2015; Hipparagi et al. 2017), except Singh et al. (2010) which reported four clusters. However, SSR based grouping of Indian soybean cultivars reported two to five clusters (Ghosh et al. 2014; Bisen et al. 2015; Chauhan et al. 2015, 2017; Gupta and Manjaya 2017). Model based population structure analysis grouped the 90 Indian soybean landraces into four different populations. While individuals of Northern region and North-Eastern region were grouped mainly in pop2 and pop1, respectively, accessions collected from Eastern region were distributed among pop1 and pop4. Gyu et al. (2008) identified three populations among 260 Korean land races. Hipparagi et al. (2017) reported six populations in 75 Indian landraces. Chauhan et al. (2017) reported 5 populations in 78 Indian soybean cultivars, whereas Gupta and Manjaya (2017) reported two populations in 90 Indian soybean cultivars. The Venn diagram analysis shows good level of similarity between hierarchical clustering and model based population structure.

Early maturity and photoperiod insensitivity are desirable traits in soybean breeding to breed varieties suitable for different latitudes, cropping patterns and to mitigate the climate change induced variable rainfall periods. Identification of desirable and rare alleles of genes for flowering and photoperiod sensitivity will help breeders to develop short duration and photoperiod insensitive varieties using marker assisted selection. Using allele specific markers of four known flowering and photoperiod sensitivity genes, we have identified genotypes with mutant alleles of two flowering and maturity genes (*E1*, *E2*) and one photoperiod responsive flowering gene (*E3*), by a reverse genetics approach. Although several studies are available on allelic characterization of flowering and maturity genes in germplasm collection of Japan, USA and China but for Indian soybean germplasm collection, only one study provided information on allelic status of these genes among 11 soybean genotypes (Tsubokura et al. 2013; Xu et al. 2013; Tsubokura et al. 2014; Jiang et al. 2014; Langewisch et al. 2014, 2017; Zhai et al. 2014; Zhou et al. 2015; Gupta et al. 2017). Gupta et al. (2017) reported

allelic characterization of six incandescent light insensitive accessions identified by Singh et al. (2008a) and reported one triple mutant (*e2*, *e3*, *e4*) and one tetra mutant (*e1*, *e2*, *e3*, *e4*) accession of flowering and photoperiod sensitivity genes. Both these genotypes were exotic collection and were originated from Hungary (Gupta et al. 2017). The present study had identified a new genetic source of mutant alleles of flowering and photoperiod sensitivity genes in a new combination (*e1*, *e2*, *e3*). The triple mutant genotype IC15089, will be very useful for precision breeding of short duration and photoperiod insensitivity traits using allele specific markers. As the new genotype is identified from Indian landrace population, it will be more useful for soybean breeding in India because of their better regional adaptation.

Conclusion

This study depicted genetic relationship and population structure of Indian soybean landraces. SSR based molecular characterization had identified some genetically diverse Indian soybean landraces having unique and rare alleles which can be used in breeding programmes for broadening genetic base and development of new soybean cultivars. Allelic characterization of four genes regulating flowering, maturity and photoperiod sensitivity among Indian soybean landraces had identified genotypes with recessive alleles of these genes. These genotypes can be used as donors in molecular breeding programmes for development of short duration and photoperiod insensitive cultivars having better adaptability to Indian environment.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abe J, Xu DH, Suzuki Y, Kanazawa A, Shimamoto Y (2003) Soybean germplasm pools in Asia revealed by nuclear SSRs. *Theor Appl Genet* 106:445–453
- Bisen A, Khare D, Nair P, Tripathi N (2015) SSR analysis of 38 genotypes of soybean [*Glycine max* (L.) Merr.] genetic diversity in India. *Physiol Mol Biol Plants* 21:109–115
- Buzzell RI (1971) Inheritance of a soybean flowering response to fluorescent day length conditions. *Can J Genet Cytol* 13:703–707
- Buzzell RI, Voldeng HD (1980) Inheritance of insensitivity to long day length. *Soybean Genet Newsl* 7:26–29
- Chauhan DK, Bhat JA, Thakur AK, Kumari S, Hussain Z, Satyawathi CT (2015) Molecular characterization and genetic diversity assessment in soybean [*Glycine max* (L.) Merr.] varieties using SSR markers. *Ind J Biotechnol* 14:504–510
- Chauhan DK, Bhat JA, Thakur AK et al (2017) Understanding genetic relationship and population structure of Indian soybean varieties using microsatellite markers. *Proc Natl Acad Sci India, Sect B Biol Sci*. <https://doi.org/10.1007/s40011-017-0847-y>
- Cober ER, Voldeng HD (2001) A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. *Crop Sci* 41:698–701
- Cober ER, Tanner JW, Voldeng HD (1996) Soybean photoperiod-sensitivity loci respond differentially to light quality. *Crop Sci* 36:606–610
- Dieringer D, Schlötterer C (2003) Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res* 13:2242–2251
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- FAOSTAT (2018) <http://www.fao.org/faostat/en/#data/QC>. Accessed 12 Apr 2018
- Fu YB, Peterson GW, Morrison MJ (2007) Genetic diversity of Canadian soybean cultivars and exotic germplasm revealed by simple sequence repeat markers. *Crop Sci* 47:1947–1954
- Ghosh J, Ghosh PD, Choudhury PR (2014) An assessment of genetic relatedness between soybean [*Glycine max* (L.) Merrill] cultivars using SSR markers. *Am J Plant Sci* 5:3089–3096
- Gireesh C, Husain SM, Shivakumar M, Satpute G, Kumawat G, Arya M, Agarwal DK, Bhatia VS (2017) Integrating principal component score strategy with power core method for the development of core collection in Indian soybean germplasm. *Plant Genetic Resour* 15:230–238. <https://doi.org/10.1017/S1479262115000556>
- Gupta S, Manjaya JJ (2017) Genetic diversity and population structure of Indian soybean [*Glycine max* (L.) Merr.] revealed by simple sequence repeat markers. *Crop Sci Biotechnol* 20:221. <https://doi.org/10.1007/s12892-017-0023-0>
- Gupta S, Bhatia VS, Kumawat G, Thakur D, Singh G, Tripathi R et al (2017) Genetic analyses for deciphering the status and role of photoperiodic and maturity genes in major Indian soybean cultivars. *J Genet* 96:147–154
- Gyu TC, Jeongran L, Jung KM, MunSup Y, Hyung JB, Jung HK, TaeSan K, NamChon P, Paek P (2008) Genetic diversity and population structure of Korean soybean landrace [*Glycine max* (L.) Merr.]. *J Crop Sci Biotech* 11:83–90
- Hipparagi Y, Singh R, Choudhury DR, Gupta V (2017) Genetic diversity and population structure analysis of Kala bhat [*Glycine max* (L.) Merrill] genotypes using SSR markers. *Hereditas* 154:9. <https://doi.org/10.1186/s41065-017-0030-8>
- Hudcovicova M, Kraic J (2003) Utilisation of SSRs for characterization of the soybean [*Glycine max* (L.) Merr.] genetic resources. *Czech J Genet Breed* 39:120–126
- Jiang B, Nan H, Gao Y, Tang L, Yue Y et al (2014) Allelic combinations of soybean maturity loci *E1*, *E2*, *E3* and *E4* result in diversity of maturity and adaptation to different latitudes. *PLoS ONE* 9(8):e106042. <https://doi.org/10.1371/journal.pone.0106042>
- Kondetti P, Jawali N, Apte SK, Shitole MG (2012) Comparative study of genetic diversity in Indian soybean [*Glycine max* L. Merr.] by AP-PCR and AFLP. *Ann Biol Res* 3:3825–3837
- Kong F, Nan H, Cao D, Li Y, Wu F, Wang J et al (2014) A new dominant gene *E9* conditions early flowering and maturity in soybean. *Crop Sci* 54:2529–2535

- Kumawat G, Singh G, Gireesh C, Shivakumar M, Arya M, Agarwal DK, Husain SM (2015) Molecular characterization and genetic diversity analysis of soybean [*Glycine max* (L.) Merr.] germplasm accessions in India. *Physiol Mol Biol Plants* 21:101–107
- Langewisch T, Zhang H, Vincent R, Joshi T, Xu D et al (2014) Major soybean maturity gene haplotypes revealed by SNPviz analysis of 72 sequenced soybean genomes. *PLoS ONE* 9(4):e94150. <https://doi.org/10.1371/journal.pone.0094150>
- Langewisch T, Lenis J, Jiang GL, Wang DC, Pantalone V, Bilyeu K (2017) The development and use of a molecular model for soybean maturity groups. *BMC Plant Biol* 17:91. <https://doi.org/10.1186/s12870-017-1040-4>
- Li Y, Guan R, Liu Z, Ma Y, Wang L, Li L, Lin F, Luan W, Chen P, Yan Z, Guan Y, Zhu L, Ning X, Smulders MJM, Li W, Piao R, Cui Y, Yu Z, Guan M, Chang R, Hou A, Shi A, Zhang B, Zhu S, Qiu L (2008) Genetic structure and diversity of cultivated soybean [*Glycine max* (L.) Merr.] landraces in China. *Theor Appl Genet* 117:857–871
- Liu K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. *Bioinformatics* 21:2128–2129
- 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. *Genetics* 180:996–1007
- Oliveros JC (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>. Accessed 25 Sept 2018
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Satyavathi CT, Bhat KV, Bharadwaj C, Tiwari SP, Chaudhary VK (2006) AFLP analysis of genetic diversity in Indian soybean [*Glycine max* (L.) merr.] varieties. *Genet Resour Crop Evol* 53:1069. <https://doi.org/10.1007/s10722-005-0779-x>
- Singh RK, Bhat KV, Bhatia VS, Mohapatra T, Singh NK (2008a) Association mapping for photoperiod insensitivity trait in soybean. *Natl Acad Sci Lett* 31:281–283
- Singh RK, Bhatia VS, Yadav S et al (2008b) Identification of genetically diverse genotypes for photoperiod insensitivity in soybean using RAPD markers. *Physiol Mol Biol Plants* 4:369–375. <https://doi.org/10.1007/s12298-008-0036-9>
- Singh RK, Bhatia VS, Bhat KV, Mohapatra T, Singh NK, Bansal KC, Koundal KR (2010) SSR and AFLP based genetic diversity of soybean germplasm differing in photoperiod sensitivity. *Genet and Mol Bio* 33:319–324
- Tsubokura Y, Matsumura H, Xu M, Liu B, Nakashima H et al (2013) Genetic variation in soybean at the maturity locus *E4* is involved in adaptation to long days at high latitudes. *Agronomy* 3:117–134
- Tsubokura Y, Watanabe S, Xia Z, Kanamori H, Yamagata H, Kaga A et al (2014) Natural variation in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. *Ann Bot* 113:429–441
- Vinu V, Singh N, Vasudev S, Yadava DK, Kumar S, Naresh S, Bhat SR, Prabhu KV (2013) Assessment of genetic diversity in *Brassica juncea* (Brassicaceae) genotypes using phenotypic differences and SSR markers. *Rev Biol Trop* 61:1919–1934
- Wang KJ, Takahata Y (2007) A preliminary comparative evaluation of genetic diversity between Chinese and Japanese wild soybean (*Glycine soja*) germplasm pools using SSR markers. *Genet Resour Crop Evol* 54:157–165
- Wang L, Guan R, Zhangxiong L, Chang R, Qiu L (2006) Genetic diversity of Chinese cultivated soybean revealed by SSR markers. *Crop Sci* 46:1032–1038
- Wang M, Li R, Yang W, Du W (2010) Assessing the genetic diversity of cultivars and wild soybeans using SSR markers. *Afr J Biotechnol* 9:4857–4866
- Wang YH, Zhang XJ, Fan SJ (2015) Genetic diversity of wild soybean populations in Dongying, China, by simple sequence repeat analysis. *Genet Mol Res* 14:11613–11623
- Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y et al (2009) Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics* 182:1251–1262
- Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, 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* 188:395–407
- Watanabe S, Harada K, Abe J (2012) Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breed Sci* 61:531–543
- Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H et al (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc Natl Acad Sci* 109:E2155–E2164
- Xu M, Xu Z, Liu B, Kong F, Tsubokura Y, Watanabe S et al (2013) Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. *BMC Plant Biol* 13:91
- Zhai H, Lü S, Wang Y, Chen X, Ren H et al (2014) Allelic variations at four major maturity genes and transcriptional abundance of the *E1* gene are associated with flowering time and maturity of soybean cultivars. *PLoS ONE* 9(5):e97636. <https://doi.org/10.1371/journal.pone.0097636>
- Zhao C, Takeshima R, Zhu J, Xu M et al (2016) A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of *FT2a*, a flowering locus *T* ortholog. *BMC Plant Biol* 16:20. <https://doi.org/10.1186/s12870-016-0704-9>
- Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W et al (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* 33:408–414