Australian Journal of Crop Science AJCS
AJCS 9(9):844-852 (2015) ISSN:1835-2707

Genetic diversity of soybean genotypes differing in isoflavones content as revealed by HPLC and SSR markers

Vineet Kumar*, Anita Rani, Reena Rawal, Vaishali Mourya, Swati Putrevu, Jaya Jhawar, Amit Dixit

ICAR-Directorate of Soybean Research, Khandwa Road, Indore (MP), India, 452 001

***Corresponding author: vineetksahni@yahoo.com**

Abstract

High-isoflavones soybean genotypes are desired in nutraceutical industry. Conversely, low-isoflavones soybean genotypes are preferred to regular soybean in designing soy-based infant formula and in developing soy food products with reduced astringent taste. Concentration of individual form of isoflavones viz. daidzein, glycitein and genistein was determined in the seeds of 46 Indian and exotic soybean genotypes using high performance liquid chromatography. The study exhibited a 9-fold (234.3-2092.5 µg/g of seed) genetic variation for total isoflavones content, with 19 genotypes falling in high isoflavones (>1200 µg/g), and 14 genotypes in low isoflavoes category (<600 µg/g). For developing genotypes with further high or low values of isoflavones, it is critical to hybridize genetically diverse parents with-in high or low-isoflavones genotypes as analysed by HPLC. Genetic diversity analysis carried out using 58 simple sequence (SSR) markers exhibited 144 alleles with polymorphic information content (PIC) varying from 0.00 to 0.773. The pair-wise genetic similarity value between soybean genotypes varied from 0.24 to 0.95. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) allocated the genotypes in 5 clusters with fairly good bootstrap support. Mantel's test for cophenetic correlation with r = 0.810 indicated a good fit of the soybean genotypes in a group in the cluster analysis. Genetically diverse parents identified in low- and high-isoflavones category can be crossed to obtain trangressive segregants.

Keywords: Genetic diversity; isoflavones; soybean; simple sequence repeat marker.
Abbreviations: HPLC_High Performance Liquid Chromatography; PIC_Polymorphism Information Content; SSR_Simple Sequence Repeats; UNJ_Unweighted Neighbour joining; UPGMA_Unweighted Pair-Group Method with Arithmetic Mean.

Introduction

Soybean has acquired the sobriquet of "functional food of the century" because of the wide array of health-promoting bio-molecules associated with this wonder bean. Much of the importance soybean has gained in nutraceutical industry is attributed to the isoflavones present in seeds, which possess estrogenic and antioxidative properties. The estrogenic activity of these biomolecules has been associated with mitigation of postmenopausal blues (Ye et al., 2012), the reduced incidence of hormone-dependent breast cancer (Messina and Wood, 2008) and osteoporosis (Civitelli, 1997). As antioxidants, soy isoflavones have been implicated in reducing the risk of onset of killer diseases like prostate and colon cancer (Raju et al., 2009), atheroscelerosis (Damasceno et al., 2007) and diabetes (Nordentoft et al., 2008). This has led to the boom in the manufacturing of dietary supplements enriched with soy isoflavones concentrate, for which high-isoflavones soybean genotypes are the most suitable raw material. Conversely, use of soybean seeds with high level of isoflavones as the raw material has been reported to impart astringent taste in soy products (Kudou et al., 1991), this deters the utilization of soybean in food uses in many countries including India. Therefore, low-isoflavones soybean genotypes are always preferred to regular soybean seeds for delivering the soy products with wider organoleptic acceptance. Further, in the backdrop of the concerns raised about the possible adverse effects of isoflavones on the infants fed on soy-based formulations (Mendez et al., 2002; Chen and Rogan, 2004), low-isoflavones soybean genotypes are much-sought-after commodity for tailoring the soy formula for infants. Therefore, high- and low-isoflavones specialty soybean genotypes are in demand to address the requirements of pharmaceutical and soy-food processing industry, respectively. This can be achieved by either identifying genotypes with extreme concentrations (high/low) of isoflavones or by developing special genotypes for this purpose through hybridization programme between genotypes of diverse genetic background. More recently, simple sequence repeats (SSRs), because of their codominance, polymorphic and reproducible properties have been employed to assess genetic diversity in different crops (Li et al., 2010; Guan et al., 2010; Tantasawat et al., 2011). They have been used for the selection of genetically diverse parents for initiating hybridization programme for improvement of particular traits across different crops including soybean (Varshney and Tuberosa 2010; Singh et al., 2010; Rakshit et al., 2011). Genetic variability for isoflavones content in soybean seed has been reported in the literature (Cvejic et al., 2009; Kumar et al., 2010; Tepancevic et al., 2010; Cvejic et al., 2011). QTLs (quantitative trait loci) regions, though a limited number, contributing to the isoflavones content have also been reported (Meksem et al., 2001; Zeng et al., 2009; Primomo et al., 2005; Zhang et al., 2014a), thereby indicating the possibility of employing marker assisted selection (MAS) approach for manipulating the concentration of this trait in soybean. In the present investigation, seeds of 46 genotypes were analysed for isoflavones content through high performance liquid

chromatography to assess the genetic variability, and genetic diversity analysis was carried out using SSR markers to identify diverse parents for initiating breeding programme for development of soybean genotypes with extreme levels of isoflavones (very high /very low) to address the needs of niche specific market.

Results and Discussion

Genetic variability for isoflavones content as assessed by HPLC

Fig 1 depicts the separation of daidzein, glycitein and genistein by high performance liquid chromatography at retention times 6.19, 6.59 and 9.74 min, respectively. Contents of daidzein, glycitein and genistein were in the ranges of 72.8-776.8, 72.1-737.9 and 54.4-1177.2 µg/g, respectively. In most of the genotypes analyzed, genistein content was the dominant among three forms of isoflavones. Total isoflavones content in the seeds of 46 soybean genotypes ranged from 234.4 for Fukuyutaka to 2092.5 µg/g for SL795, thereby exhibiting approximately 9-fold genotypic variation (Table 2). To allocate soybean genotypes into high, normal and low isoflavones category, we took into consideration the recommendations of Song et al. (2007) and Food and Drug Administration (FDA) of United States. Song et al.(2007) reported daily intake of 75,000 µg of soy isoflavones as upper safe limit, while FDA has recommended a daily intake of 25 g soybean protein, which is equivalent to 62.5 g of soybean seeds (assuming 40% protein in seeds) necessary to avail all the health benefits of soybean. Therefore, a genotype with more than 75,000 µg isoflavones in 62.5 g seeds *i.e.*1200 µg isoflavones/g of seeds was considered as high isoflavones-containing genotype. In the present study, soybean genotypes with total isoflavones content exceeding 1200 µg/g of soy flour were grouped in high isoflavones-containing category, while genotypes with 600-1200 µg/g and less than 600 µg/g of soy flour were grouped in medium and low isoflavones-containing category, respectively. Of the 46 selected genotypes, 14 genotypes were in low isoflavones category, 16 genotypes each in high and medium isoflavones category. Our results showed that JS79-264, UPSM534, PK564, Samrat, Shivalik, DS97-12, Pedegra, Kegone, Boiling type, Fukuyutaka, Table variety, PK416, PK1042, and NRC7 were low isoflavones-containing soybean genotypes. Of the 16 high isoflavones-containing soybean genotypes mentioned in Table 3, genotypes Hardee, SL795, and MACS22 exhibited total isoflavones content of approximately 2000 µg/g. Soybean genotypes with total isoflavones content as high as 6115.5 µg/g has been reported in Chinese germplasm (Yan-Wei et al., 2013).

Polymorphism as revealed by SSR primer pair screening

A total of 58 SSR primer pairs, distributed across 20 linkage groups of soybean (Table 4) were used to amplify specific loci from the genomic DNA of each of the 46 soybean genotypes. The amplified products obtained from each of these primers were resolved on 3% rezophore agarose gel and scrutinized for the polymorphism. Fifty one primers detected polymorphism while seven primers (Satt258, Satt143, Satt050, Satt558, Satt459, Satt575 and Satt314) yielded monomorphic bands. A total of 144 alleles were amplified with an average of 2.48 alleles per locus. The number of alleles per primer pair (locus) ranged from 1 (Satt258, Satt143, Satt050, Satt558, Satt459, Satt575, Satt314) to 5 (Sct_199, Satt009, Satt281). Three loci amplified 4 alleles, 20 loci showed 3 alleles and 25 loci exhibited 2 alleles. Representative banding pattern of PCR product profiles at Sct_199 with 5 distinct alleles is shown in Fig. 2. Of the total number of 144 alleles, 63 alleles showed a frequency of 0.25 or less, 24 alleles exhibited a frequency of 0.75 or higher and remaining 57 alleles exhibited a frequency between 0.25-0.75. The size of the allele fragments ranged from 80 to 380 bp. The PIC value, which is a measure of allelic diversity, for the 58 SSR markers ranged from 0.00 to 0.773 (Sct_199) with 0.372 as the PIC/locus. Nei's expected heterozygosity (H_E) of the 58 primer pairs employed ranged from 0.000 (Satt258, Satt050, Sattt558, Satt459, Satt575, Satt314 and Satt143) to 0.7723 (Sct_199); while effective number of alleles (N_E) ranged from 1.0 (Satt258, Satt050, Satt558, Satt459, Satt575, Satt314 and Satt143) to 4.3926 (Sct_199). Of the total 144 alleles identified in the present investigation, 11 alleles (7.63%) were unique which were amplified in single genotype. Satt197, Sat_424, Satt190, Satt548, AI856415, Satt181, Satt240, Satt552, Satt260 generated one unique allele of fragment size 140, 170, 245, 230, 200, 240, 260, 150, 250 bp, respectively. Satt281 on LGp C2 amplified two unique alleles of fragment size 190 bp (Shivalik) and 210 bp (Fukuyutaka). A high percentage of polymorphism (87.93%) detected in this study was consistent with the previous studies (Singh et al., 2010; Tantasawat et al., 2011). In general, the primer pair showing high number of alleles (4 or 5) also showed high PIC value. However, Sat_140 (on LGp C1) which showed 4 alleles was found to have low PIC value (0.385). Two SSR markers (Sct_199, Satt009) with 5 alleles showed PIC values greater than 0.7. Hence, these 2 SSR markers were the most informative for distinguishing the soybean genotypes. The SSR allelic diversity observed in the present study was low (2.48 alleles per locus) compared to the previous reports. Wang et al. (2006) reported 12.20 alleles per locus with average PIC value of 0.78 in assessment of genetic diversity of 129 soybean genotypes using 60 SSR markers. Similarly, Chotiyarnwong et al. (2007) reported 11.83 alleles per locus in an analysis of 149 Thai soybean genotypes using 18 SSR markers. Compared to these studies, low diversity in our study may be because of the fact that a fair number of the genotypes selected in the diversity analysis were Indian genotypes, which have common parentage or one of the parents from the late maturity group of United States soybean.

Genetic diversity and relationship among soybean genotypes

All 144 SSR alleles were used for the genetic diversity analysis. Jaccard's similarity coefficient was calculated to assess the genetic proximity among the genotypes and the similarity coefficient matrix was used for UPGMA (Unweighted Pair Group Method with Arithmatic Average) cluster analysis. The pair-wise genetic similarity value among soybean genotypes varied from 0.24 to 0.95 with Boiling type *vs.* Pedegra (0.95) and JS90-41 *vs.* Pedegra (0.24) as the most close and diverse pair, respectively. Cluster analysis based upon the coefficient of similarity classified 46 soybean genotypes into 5 groups (Fig. 3). Cluster I comprised of 4 genotypes viz. Boiling type, Pedegra, Fukuyutaka and Kegone. All these genotypes were exotic (Table 1). Of these 4 genotypes, Boiling type and Pedegra are difficult to be distinguished genetically as they are very closely related with 95% similarity and 100% bootstrap value (Fig. 4). The cluster II contained only 2 genotypes *viz.* NRC7 and JS93-05 with 54% genetic similarity. Cluster III contained only one

Table 1. Soybean genotypes (46) with their respective country of origin and pedigree, analyzed for isoflavones content.

Genotype	Pedigree	Origin	Genotype	Pedigree	Origin
EC537960	EC537960	India	PK1092	PK327 × PK416	India
JS79-264	JS79-264	India	JS97-52	PK327 × L129	India
UPSM534	Germplasm accession	China	JS79-81	Braggx Harsoy deciduous	India
EC391349	EC391349	India	JS90-41	PS73-7 × Hark	India
MACS22	MACS22	India	JS93-05	Selection from PS73-22	India
SL688	PK416 × SL317	India	SL295	PK416 × PK564	India
SL744	SL457 × SL459	India	SL525	PK416 × PK1023	India
SL900	PK1241 × JS335	India	Samrat	Farmers' selection	India
SL982	SL525 × DS98-41	India	Shivalik	Selection from PK 73-55	India
SL799	JS90-29 ×E4	India	MAUS32	Selection from JS80-21	India
SL871	DS97-12 × SL798	India	LSb1	Selection from MACS330	India
SL795	PK1162 × E4	India	NRC7	Selection from S 69-96	India
SL958	SL525 ×SL706	India	ADT1	Selection from "Hill" variety	Introduction from USA
SL794	PK1162 × SL459	India	Lee	S-100 × CNS	Introduction from USA
PS1347	PS1024 × PK472	India	Table variety	Germplasm accession	Taiwan
PS1241	PK1039 × PK327	India	DS97-12	Mutant of DS 74	India
PK1024	PK308 × PK317	India	Kegone	Japan	Japan
PK1029	PK262 × PK317	India	Hardee	D49-77×Improved pelican	Introduction from USA
PK416	UPSM534 × S38	India	JS95-60	Selection from PS7322	India
PK472	Hardee × Punjab1	India	Fukuyutaka	Germplasm accession	Japan
PK308	T31 × Hardee	India	JS335	JS78-77 × JS 71-05	India
PK564	UPSM534× Ankur × Bragg	India	Boiling type	Germplasm accession	Japan
PK1042	Bragg × PK416	India	Pedegra	Germplasm accession	Taiwan

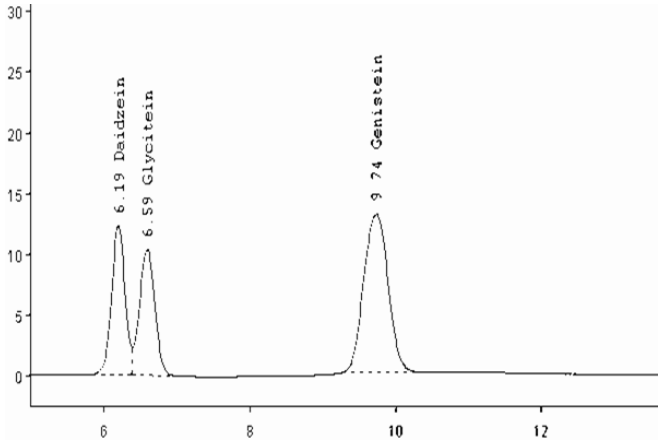

Fig 1. Three forms of isoflavones viz. daidzein, glycitein and genistein as resolved by high performance liquid chromatography.

genotype MACS22. Cluster IV was comprised of 14 genotypes of both Indian (JS335, PK472, JS95-60, LSb1, SL795, SL794, Shivalik, MAUS32 and JS79-264) and United States (Lee, Hardee and ADT1) origin. It is explainable as most of Indian soybean varieties released in 1970s and 1980s were from the maturity group VII of American soybean or developed using American genotype as one of the parents. Cluster V contained maximum number of genotypes (25). Cophenetic correlation of the clustering with similarity matrix was high (0.810), suggesting thereby a good fit of the two data sets. Of the total 14 genotypes in low isoflavones category, 4 genotypes were present in Cluster I, 1 genotype in Cluster II, 3 genotypes in IV and 6 genotypes in V. The most diverse combination among the low-isoflavones genotypes was Kegone *vs*. NRC7 with similarity value (0.301). Kegone is a Japanese genotype, while NRC7 is an Indian soybean genotype. Further, UPSM534 is a low-isoflavones containing soybean genotype and the pedigree data in Table 1 shows that this genotype is maternal parent of PK416 and PK564. Interestingly, both these genotypes also exhibited low levels of isoflavones content. Similarly, PK1042 which has genomic content of UPSM 534 through PK416 as one of its parent also showed low levels of isoflavones content. These results are supported by the observations of Cvejik et al. (2011) who reported significant correlation between total isoflavones content in parents and the F_1 progenies. With regards to 19 genotypes in high-isoflavones category, one genotype each was present in cluster II and cluster III, 6 genotypes in cluster IV and maximum of 11 genotypes in cluster V. The most diverse combination in high-isoflavone category was MACS22 *vs.* JS93-05 with 0.318 similarity coefficient followed by Hardee *vs.* SL295 with 0.345 similarity coefficient. Distribution of 46 genotypes revealed through principal co-ordinate analysis (Fig. 5) conformed to UPGMA cluster analysis. Similar to UPGMA dendrogram, 4 exotic genotypes viz. Pedegra,

Table 2. Isoflavones' content (µg/g) of forty six soybean genotypes.

S.N.	Genotype	Daidzein	Glycitein	Genistein	Total Isoflavones
1	EC537960	198.1±7.4^{g}	230.2±10.6^{i}	211.1±10.9^{f}	639.4±18.3^{g}
2	JS79-264	136.5 ±6.4^{c}	131.4±7.8^{d}	145.4±8.9^{cd}	412.3±15.6^{c}
3	UPSM534	183.6± 9.6^{f}	103.3±6.3^{c}	96.3±1.2^{b}	383.2±12.6^{b}
4	EC391349	182.2±10.6^{f}	180.5±8.5^{g}	421.9±10.7^{k}	784.6±14.9^{i}
5	MACS22	400.2±12.4^{o}	737.9±14.5^{u}	857.5±16.7^{v}	1995.6±38.9^{x}
6	SL688	301.1±8.7^{l}	258.4±12.1^{kl}	683.6±13.4^{t}	1243.1±24.8^{p}
7	SL744	280.9±13.4^{k}	283.3±8.8^{mn}	412.7±12.7^{k}	976.9±17.9^{l}
8	SL900	377.3± 8.5^{n}	158.6±6.4^{ef}	413.5±12.5^{k}	949.4±18.3^{k}
9	SL982	456.4± 12.6^{p}	445.5±8.5^{r}	1026.1±20.3^{w}	1927.0±38.6^{w}
10	SL799	211.7±10.8^{h}	262.5±10.5^{l}	268.6±13.6^{h}	741.8±14.6^{hi}
11	SL871	322.9±9.9^{m}	341.7±12.3^{o}	614.2±10.9^{q}	1278.5±20.8^{q}
12	SL795	609.4±12.4^{u}	305.9±9.8^{n}	1177.2±12.5^{z}	2092.5±30.3^{y}
13	SL958	303.0±13.5^{l}	192.4±9.6^{gh}	1096.8±16.5^{x}	1591.2±29.9^{u}
14	SL794	505.8±10.5^{r}	292.7±12.4^{n}	1140.3±22.8^{y}	1938.8±38.7^{w}
15	PS1347	393.4±9.5^{o}	362.2±12.5^{p}	486.9±12.3^{m}	1242.5±24.9^{p}
16	PS1241	589.8±15.4^{t}	386.8±11.4^{q}	454.8±12.6^{l}	1432.4±28.5^{s}
17	PS1024	274.6±12.1^{k}	249.2±12.6^{kl}	246.9±10.9^{g}	770.7±21.2^{i}
18	PS1029	234.5±10.5^{i}	264.9±10.3^{lm}	146.5±5.5^{cd}	642.9±12.8^{g}
19	PK416	172.4±6.4^{ef}	100.4±5.9^{c}	315.2±9.9^{i}	588.0±10.7^{f}
20	PK472	274.1±10.2^{k}	168.9±6.3^{fg}	542.9±10.1^{o}	985.9±18.9^{l}
21	PK308	440.2±12.6^{p}	288.2±12.5^{mn}	643.3±13.9^{r}	1371.7±26.8^{r}
22	PK564	154.7±7.5^{d}	145.7±5.8^{e}	192.8±5.6^{ef}	493.2±12.8^{d}
23	PK1042	159.9±6.6^{de}	217.3±10.8^{i}	161.6±6.1^{d}	538.8±10.9^{e}
24	PK1092	294.6±8.8^{l}	100.2±9.8^{c}	372.2±15.7^{j}	766.9±14.2^{i}
25	JS97-52	402.0±10.9^{0}	258.2±12.8^{kl}	611.3±12.8^{q}	1271.5±23.9^{q}
26	JS79-81	443.9±11.1^{p}	465.8±12.3^{s}	535.9±10.8^{no}	1445.6±28.1^{s}
27	JS90-41	405.9±10.8^{o}	229.9±12.6^{i}	637.9±12.2^{r}	1273.7±23.9^{q}
28	JS93-05	480.4±12.5^{q}	218.4±11.9^{i}	497.2±12.3^{m}	1195.0±22.8^{o}
29	SL295	622.7±12.7^{v}	272.9±13.3^{lm}	646.9±13.6^{r}	1542.5±29.5^{t}
30	SL525	258.9±12.1^{j}	367.1±13.7^{p}	427.5±12.3^{k}	1053.5±22.5^{m}
31	Samrat	72.8±1.5^{a}	166.9±8.8^{f}	182.4±7.7^{e}	422.1±12.9^{c}
32	Shivalik	133.1±7.5^{c}	159.9±6.7^{f}	281.8±13.5^{h}	574.8±15.6^{f}
33	MAUS32	241.1±10.6^{i}	145.9±5.5^{e}	423.7±12.8^{k}	810.7±14.1^{j}
34	LSb1	545.3±10.8^{s}	284.3±10.6^{mn}	783.9±14.3^{u}	1613.5±22.8^{v}
35	NRC7	194.8±5.8^{fg}	104.5±7.5^{c}	240.4±12.6^{g}	539.7±15.5^{e}
36	ADTI	406.3±12.5^{o}	147.9±7.8^{ef}	203.9±12.2^{f}	758.1±13.9^{hi}
37	Lee	376.7±12.5^{n}	203.6±12.2^{h}	569.5±15.4^{p}	1149.8±22.1^{n}
38	Table variety	140.6±5.3^{c}	96.8±1.9^{bc}	195.1±6.6^{ef}	432.5±9.9^{c}
39	DS97-12	168.3±6.4^{e}	86.4±1.8^{b}	179.7±5.8^{e}	434.4±8.9^{c}
40	Kegone	155.7±6.5^{d}	284.2±12.3^{mn}	130.5±4.9^{c}	570.4±15.1^{f}
41	Hardee	776.8±14.3^{w}	585.8±10.3^{t}	666.6±12.3^{s}	2029.2±31.1^{f}
42	JS95-60	172.3±5.8^{ef}	334.8±12.5^{o}	456.2±8.9^{l}	963.3±18.2^{kl}
43	Fukuyutaka	107.9±5.6^{b}	72.1±1.2^{a}	54.4±1.2^{a}	234.4±13.9^{a}
44	JS335	475.9±12.3^{q}	276.9±12.6^{m}	519.9±10.4^{n}	1269.7±24.3^{q}
45	Boiling type	132.4±5.5^{c}	247.5±10.3^{k}	167.6±7.3^{de}	547.5±10.9^{e}
46	Pedegra	132.1±7.7^{c}	96.9±1.8^{bc}	267.3±10.4^{h}	496.3±12.8^{d}
	LSd (p <0.05)	13.4	13.4	16.7	24.6

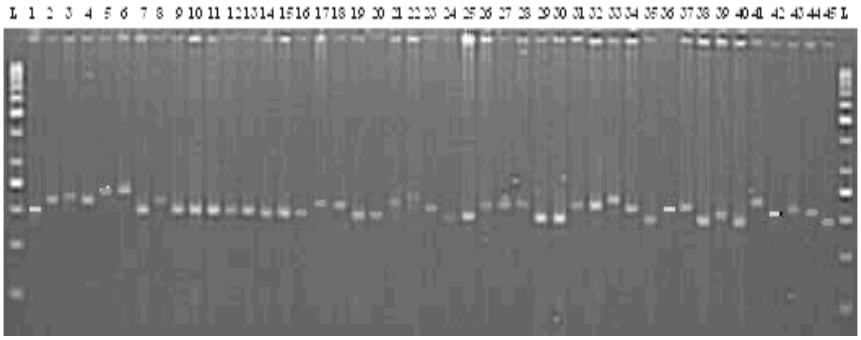

Fig 2. SSR profile of soybean genotypes showing allelic variation at loci Sct_199 (LGp G). L denotes 50 bp ladder.

Table 3. Distribution of 46 soybean genotypes according to total isoflavones'content in the seeds.

High Isoflavones (1200µg/g)	Medium Isoflavones (600-1200 µg/g)	Low Isoflavones (<600 µg/g)
MACS22 (1995.6±38.9)	EC537960 (639.4±18.3)	JS79-264 (412.3±15.6)
SL688 (1243.1±24.8)	EC391349 (784.6±14.9)	UPSM534 (383.2±12.6)
SL982 (1927.0±38.6)	SL744 (976.9±17.9)	PK564 (493.2±12.8)
SL871 (1278.7±20.8)	SL900 (949.4±18.3)	Samrat (422.1±12.9)
SL795 (2092.5±30.3)	SL799 (741.8±14.6)	Shivalik (574.8±15.6)
SL794 (1938.8±38.7)	PK1024 (770.7±21.2)	DS97-12 (434.4±8.9)
SL295 (1542.6±29.5)	PK1029 (642.9±12.8)	Pedegra (496.3±12.8)
SL958 (1591.2±29.9)	PK472 (985.9±18.9)	Kegone (570.4±15.1)
PS1347 (1242.5±24.9)	PK1092 (766.9±14.2)	Boiling type (547.5±10.9)
PS1241 (1432.3±28.5)	MAUS32 (810.7±14.1)	Fukuyutaka (234.4±13.9)
PK308 (1371.7±26.8)	ADT1 (758.1±13.9)	Table variety (432.5±9.9)
JS97-52 (1271.5±23.9)	JS95-60 (963.3±18.2)	PK416 (588.0±10.7)
JS79-81 (1445.6±28.1)	SL525 (1053.5±22.5)	PK1042 (538.8±10.9)
JS90-41 (1273.7±23.9)		NRC7 (539.7±15.5)
JS335 (1269.7±24.3)		
JS93-05 (1195.0±22.8)		
LSb1 (1613.5±22.8)		
Lee (1149.8±22.1)		
Hardee (2029.2±31.1)		

Values given in the parenthesis indicate the average total isoflavones content

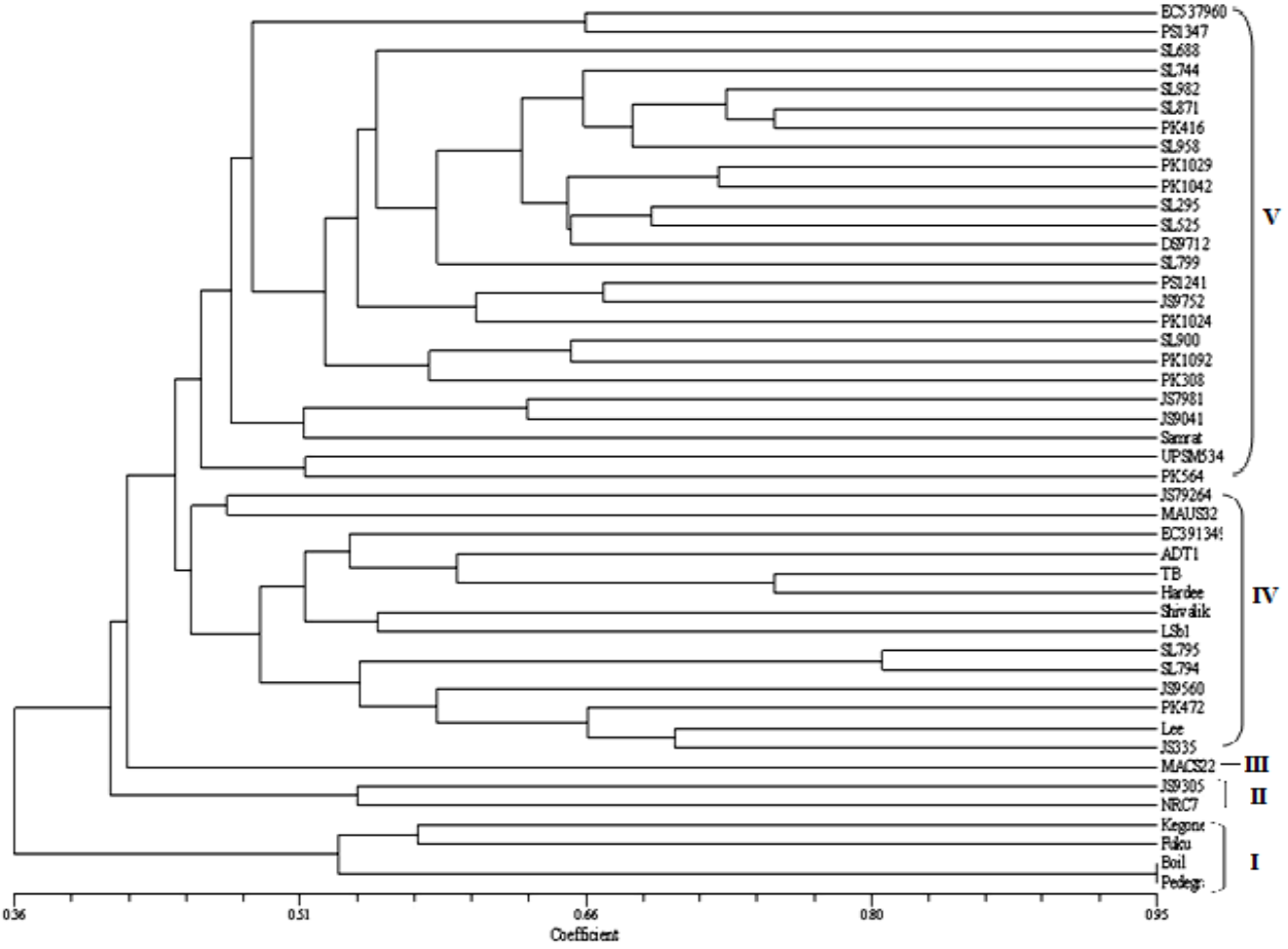

Fig 3. UPGMA dendrogram showing similarity coefficients and genetic relationships among 46 soybean genotypes based on SSR profile.

Table 4. Primer sequence, linkage group, number of alleles, effective number of alleles, polymorphism information content (PIC), Nei's expected heterozygosity of SSR primers employed in the genetic diversity analysis of 46 soybean genotypes.

SSR primer	LGp	No. of alleles	N_E	PIC	H_E	SSR primer	LGp	No. of alleles	N_E	PIC	H_E
Satt258	A1	1	1.000	0.00	0.0000	Satt256	D2	2	1.4706	0.293	0.3200
Satt200	A1	2	1.8772	0.468	0.4673	Satt598	E	3	2.2274	0.552	0.5510
Satt050	A1	1	1.000	0.000	0.0000	Sat_124	E	3	2.4884	0.599	0.5981
Satt228	A2	3	2.3519	0.513	0.5748	Satt575	E	1	1.0000	0.000	0.0000
Satt538	A2	3	1.8872	0.529	0.4701	Satt146	F	3	2.2851	0.563	0.5624
AW132402	A2	2	1.8772	0.467	0.4673	Satt114	F	4	3.3251	0.693	0.6993
Sat_270	B1	3	1.8438	0.458	0.4576	Satt586	F	3	1.7963	0.444	0.4433
Satt509	B1	2	1.3331	0.250	0.2499	Sct_199	G	5	4.3926	0.773	0.7723
Satt197	B1	2	1.0444	0.043	0.0425	Satt612	G	2	1.9756	0.494	0.4938
Satt577	B2	3	2.0522	0.575	0.5127	Satt352	G	3	2.1939	0.545	0.5442
Sat_424	B2	3	1.8093	0.448	0.4473	Satt181	H	3	2.0868	0.521	0.5208
Satt189	B2	2	1.9850	0.497	0.4962	Satt541	H	3	2.4506	0.592	0.5919
Sat_140	C1	4	1.6242	0.385	0.3843	Satt314	H	1	1.0000	0.000	0.0000
Satt399	C1	3	1.7176	0.450	0.4178	Satt587	I	2	1.9018	0.475	0.4742
Satt190	C1	2	1.0454	0.045	0.0435	Satt571	I	3	1.9527	0.488	0.4879
GMAC7L	C2	2	1.9912	0.493	0.4978	AW310961	J	2	1.6148	0.381	0.3807
Satt305	C2	3	1.9360	0.484	0.4835	Sat_366	J	2	1.1389	0.122	0.1219
Satt281	C2	5	2.6540	0.631	0.6232	Satt285	J	2	1.5857	0.361	0.3694
Satt457	C2	3	1.3848	0.284	0.2779	Satt240	K	4	2.7721	0.640	0.6393
Sat_246	C2	2	1.3478	0.193	0.2580	Satt552	K	3	1.3770	0.274	0.2738
Satt643	C2	2	1.5283	0.346	0.3457	Satt260	K	2	1.0444	0.043	0.0425
Satt658	C2	2	1.1389	0.159	0.1219	Satt229	L	2	1.9707	0.498	0.4926
Satt267	D1a	3	2.3151	0.568	0.5681	Satt143	L	1	1.0000	0.000	0.0000
Satt502	D1a	2	1.8872	0.471	0.4701	Satt523	L	2	1.5857	0.340	0.3694
Satt548	D1a	3	1.3582	0.736	0.2637	Satt551	M	2	1.9600	0.490	0.4898
Satt558	D1b	1	1.000	0.000	0.000	Satt009	N	5	3.7519	0.734	0.7335
AI856415	D1b	2	1.0444	0.044	0.0425	Sat_132	O	2	1.9988	0.500	0.4997
Sat459	D1b	1	1.000	0.000	0.000	Sat_318	O	2	1.5283	0.346	0.3457
Satt002	D2	2	1.8459	0.459	0.4583	Satt345	O	3	1.4491	0.310	0.3099

N_E: Effective number of alleles; H_E: Nei's expected heterozygosity.

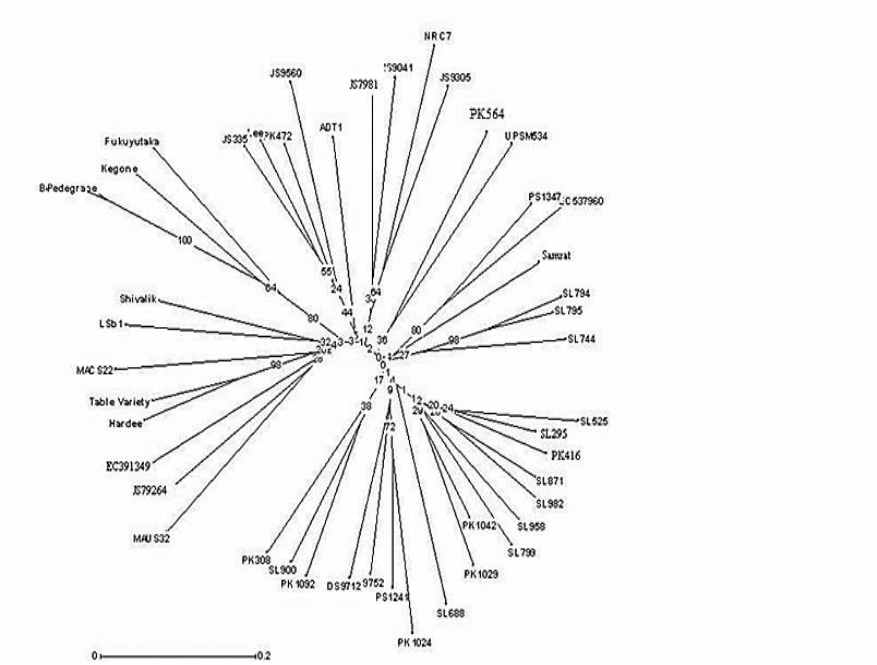

Fig 4. Unweighted Neighbor-Joining dendrogram prepared using DARwin showing clustering pattern of 46 soybean genotypes based on SSR data. Values at nodes are percentage over 10,000 bootstrap replicates.

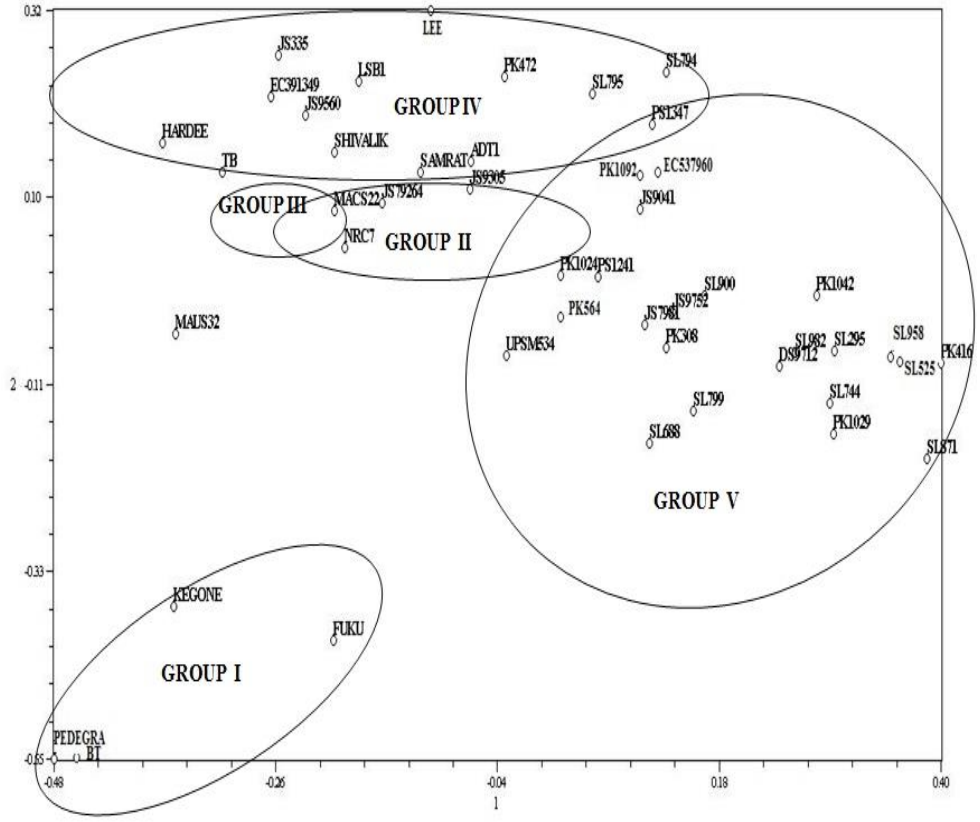

Fig 5. Two-dimensional plot generated through principal co-ordinate analysis of 46 soybean genotypes using 58 SSR markers as carried out using NTSYS 2.02 (Rohalf, 1998).

Kegone, Fukutyutaka, and Boiling type formed a distinct cluster (Group I) and distinct separation between two large groups viz. Group IV and Group V, was observed. However, PS1347 was intermixed between groups IV and V; while genotypes MACS22 and JS79264 of group III and IV, respectively, appeared in group II. Soybean genotypes with diverse genetic background within low- or high-isoflavones group are the most appropriate for selection as parents in embarking upon a plant breeding programme as the new allelic combinations can give rise to transgressive seggregants with further low or high levels of isoflavones compared to either of the parents. Kegone and NRC7 would be the ideal parents for development of genotypes with further low isoflavones content, while MACS22 and JS93-05 for breeding genotypes with isoflavones content higher than the maximum value observed in the present study. Further, influence of environment on the accumulation of isoflavones in soybean seeds renders phenotypic selection a complex process in a breeding programme focusing on this trait (Zhang et al., 2014b). QTLs have been reported for isoflavones content (Primomo et al., 2005; Gutierrez-Gonzalez et al., 2009; Zhang et al., 2014a); however, most of them contribute minor additive effects. For development of mapping population for the identification of novel QTLs for isoflavones content and validation of the reported QTLs in a new population, genetically divergent parents with extreme values for isoflavones content is the first pre-requisite. Our results showed that JS90-41 and Pedegra not only exhibit a wide variation for isoflavones content but they are genetically divergent with pair wise distance of 0.73. Therefore, these two genotypes would be the most appropriate parents to mapping population for identification of Marker Assisted Breeding.

Materials and Methods

Plant Material

The pedigree and the centre of origin of 46 soybean genotypes undertaken for the study are given in Table 1. Forty six genotypes were raised in the field and subjected to isoflavone analysis through high performance liquid chromatography. Daidzein, glycitein and genistein were resolved and the concentration of each of the individual isoflavone was determined. The total isoflavone concentration was computed by summing up all the individual isoflavones (Table 2).

Isoflavones extraction and estimation by HPLC

The isoflavones were extracted following Vyn et al. (2002) and concentration of daidzein, glycitein and genistein was determined through high performance liquid chromatography. HPLC conditions have been described in our earlier report (Kumar et al., 2010). Values presented in table 2 are mean of triplicate samples ± standard deviation.

DNA isolation

Genomic DNA was isolated from the finely ground young leaf tissues following cetyl trimethyl ammonium bromide procedure (Doyle and Doyle., 1990). Purification of DNA was done through phenol: chloroform: isoamylalcohal

method. Purified DNA was quantified through spectroscopic method.

Simple sequence repeats (SSR) analysis

A total of 58 SSR markers were randomly chosen for the analysis from twenty linkage groups of soybean genome (Table.4). Quantified DNA was subjected to PCR amplification in 10 µl reaction mixture containing 2 µl DNA (25 ng/µl), 1 µl PCR 10x buffer, 1.1 µl $MgCl_2$ (25 mM), 0.1 µl dNTPs (25 mM), 0.4 µl each forward and reverse SSR primers (30 ng/µl), 0.068 µl *Taq* DNA polymerase (3 units/µl) and 4.932 µl distilled water. DNA was denatured at 94^{o}C for 2 min followed by 30 cycles each consisting of denaturation at 94^{o}C for 1 min, primer annealing at 50^{o}C for 2 min, primer elongation at 72^{o}C for 3 min and final elongation at 72^{o}C for 10 min in the thermocycler (MJ Research, model PTC100). Amplified products so obtained were resolved on 3% rezophore agarose gel.

Data analysis

Computation was facilitated by the PC based programme NTSYS 2.02 (Rohalf, 1998). Presence and absence of an SSR allele was scored as 1 and 0, respectively. The data of all the SSR alleles was imported to NTedit and created into binary data matrix. Similarity coefficients between paired genotypes were computed using Jaccard's similarity formula through SIMQUAL module. Cluster analysis was carried out to construct dendrogram using Unweighted Pair Group Method with Arithmetic Average (UPGMA). To complement the information obtained from the cluster analysis, bootstrap values over 10000 permutations were determined through Unweighted Neighbor Joining (UNJ) using DARwin 5.0 software (Perrier and Jacquemoud-Collet, 2006). The allelic diversity at a locus was measured by polymorphic information content (PIC), which was determined as $PIC_i = 1 - \sum^{n}_{j=1} p^2_{ij}$ where *i* denotes the SSR marker while p_{ij} is frequency of j^{th} allele. Any allele appearing in only one genotype was treated as unique allele. Nei's expected heterozygosity and effective number of alleles were calculated in POPGENE (Yeh et al., 1997). Principal co-ordinate analysis was also carried out by NTSYS 2.02 (Rohalf, 1998).

Conclusion

The results obtained in the present study showed wide genetic variability for total isoflavones content in 46 soybean genotypes. The data obtained from the SSR markers assisted assessment of genetic diversity of the soybean genotypes can help plant breeders to select the parents for development of soybean genotypes with isoflavones content beyond the highest and the lowest extremes of the observed values and can also be very useful in generating mapping population for identification of new QTLs associated with the trait.

Acknowledgements

Authors acknowledge germplasm repository, Directorate of Soybean research for proving the seeds of exotic lines.

References

Chen AM, Rogan WJ (2004) Isoflavones in soybean infant formula: A review of evidence for endocrine and other activity in infants. Annual Rev Nutr. 24:33-54.

Chotiyarnwong O, Chatwachirawong P, Chanprame S, Srinives P (2007) Evaluation of genetic diversity in Thai indigenous and recommended soybean varieties by simple sequence repeat markers. Thai J Agri Sci. 40:119-126.

Civitelli R (1997) *In vitro* and *in vivo* effects of isoflavones on bone formation and bone biomechanics. Cacif Tissue Inter. 61:12s-14s.

Cvejic J, Malencic D, Tepavcevic V, Posa M, Milandinovic J (2009) Determination of phytoestrogen composition in soybean cultivars in Serbia. Nat Pro Commun. 4(8):1069-1074.

Cvejic J, Tepavcevic V, Bursac M, Milandinovic J, Malencic D (2011) Isoflavone composition in F1 soybean progenies. Food Res Inter. 44:2698-2702.

Damasceno NRT, Apolinario E, Flauzino FD, Fernandes I, Abdall DSP (2007) Soy isoflavones reduce electronegative low-density lipoprotein (LDL) and anti-LDL auto-antibodies in experimental atherosclerosis. Eu J Nutr. 46:125-132.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from the fresh tissue. Focus 12:14-15.

Guan R, Chang R, Li Y, Wang L, Liu Z, Qiu L (2010) Genetic diversity comparison between Chinese and Japanese soybeans (*Glycine max* (L.) Merr.) revealed by nuclear simple sequence repeats. Genet Resour Crop Ev. 57:229-242.

Gutierrez-Gonzalez JJ, Wu X, Zhang J, Lee JD, Ellersieck M, Shannon JG, Yu O, Nguyen HT, Sleper DA (2009) Genetic control of seed isoflavone content: importance of statistical model amd epistasis in complex traits. Theor Appl Genet. 119(6):1069-1083.

Kudou S, Fleury Y, Welti D, Magnolato D, Uchida T (1991) Malonyl isoflavones glucosides in soybean seeds. Agric Biol Chem. 55(9):2227-2233.

Kumar V, Rani A, Dixit A, Pratap D, Chauhan GS (2010) A comparative assessment of total phenolic content, ferric reducing-antioxidative power, free radical-scavenging activity, isoflavones and vitamin C content in soybean with varying seed coat colour. Food Res Int. 43:323-328.

Li YH, Li W, Zang C, Yang L, Chang RZ, Gaut BS, Qiu LJ (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine Soja*) for simple sequence repeat and single nucleotide polymorphism loci. New Phytol. 188:242-253.

Meksem K, Njiti VN, Banz WJ, Iqbal MJ, Kassem MyM, Hyten DL, Yuang J, Winters TA, Lightfoot DA (2001) Genomic regions that underlie soybean seed isoflavone content. J Biomed Biotechnol. 1:38-44.

Mendez MA, Anthony MS, Ara L (2002) Soy-based formulae and infant growth and development: A review. J Nutr. 132:2127-2130.

Messina MJ, Wood CE (2008) Soy Isoflavones, estrogen therapy and breast cancer risk: analysis and commentary. Nutr J. 7-17.

Nordentoft I, Jeppesen PB, Hong J, Abdula R, Hermansen K (2008) Increased insulin sensitivity and changes in the expression profile of key insulin regulatory genes and beta cell transcription factors in diabetic kkAy-mice after feeding with a soybean protein rich diet high in isoflavones content. J Agric Food Chem. 56:4377-4385.

Perrier X, Jacquemoud-Collet JP (2006) DARwin software, http://darwin.cirad.fr/darwin

Primomo VS, Poysa V, Ablett GR, Jackson CJ, Gijzen M Rajcan I (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci. 45:2454-2464.

Raju J, Bielecki A, Caldwell D, Lok E, Taylor M, Kapal K, Curran I, Cooke GM, Bird RP, Mehta R (2009) Soy Isoflavone modulate azoxymethane-induced rat colon carcinogenesis exposed pre-and post-natally and inhibit growth of DLD-1 human colon adenocarcinoma cells by increasing the expression of estrogen receptor-β^{1-3}. J Nutr. 139:474-481.

Rakshit S, Santosh HB, Sekhar JC, Nath R, Shekhar M, Chikkappa GK, Gadag RN, Dass S (2011) Analyses of genetic diversity among maize inbred lines differing for resistance to pink borer and post-flowering stalk rot. J Plant Biochem Biotech. 20(2):173-181.

Rohalf FJ (1998) NTSYS pc, Numerical taxonomy and multivariate analysis system, Version 2.02. Exter Software, New York.

Singh RK, Bhatia VS, Bhat KV, Mohapatra T, Singh NK, Bansal KC, Koundal KR (2010) Simple sequence repeat and amplified fragment length polymorphism based genetic diversity of soybean germplasm differing in photoperiod sensitivity. Genet Mol Biol. 33(2):319-324.

Song WO, Chun OK, Hivang I, Shin HS, Kim BG, Kim KS, Lee SY, Shin D, Lee SG (2007) Soy isoflavones as safe functional ingredients. J Med Food 10(4):571-580.

Tantasawat P, Trongchuen J, Prajongjai T, Jenweerawat S, Chaowiset W (2011) Simple sequence repeat analysis of soybean (*Glycine max* (L.) Merr.) genetic relationship and variety identification in Thailand. Aus J Crop Sci. 5(3):283-290.

Tepavcevic V, Atanackovic M, Milandinovic J, Malencic DJ, Popovic J, Cvejic J (2010) Isoflavone composition, total phenolic content and antioxidant activity in soybeans of different origin. J Med Food 13(3):1-8.

Varshney RK, Tuberosa R (2010) Genomics-assisted crop improvement. In: Roberto (ed.) Genomics applications in crops Vol 2. A Springer publication.

Vyn TJ, Yin X, Bruulsema TW, Jackson C-JC, Rajcan I, Bruder SM (2002) Potassium fertilization effects on isoflavone concentratons in soybean. J Agri Food Chem. 50:3501-3506.

Wang L, Guan R, Zhangxiong L, Chang R, Qiu L (2006) Genetic diversity of Chinese cultivated soybean revealed by simple sequence repeat markers. Crop Sci. 46:1032-1038.

Yan-Wei C, Xi-Huan L,Wen-Long L, Wen-Suo C, Cai-Ying Z (2013) Genetic variability analysis and elite germplasm selection of isoflavone content in soybean from HUangHuaiHai ecotype region. J Plant Genet Res. 14(6):1167-1172.

Ye YB, Wang ZL, Zhuo SY, Lu W, Liao HF, Verbruggen MA, Fang S, Mai HY, Chen YM, Su YX (2012) Soy germ isoflavones improve menopausal symptoms but have no effect on blood lipids in early postmenopausal Chinese women: a randomized placebo-controlled trial. Menopause 19(7):791-8.

Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Rao JX (1997) POPGENE, the user friendly shareware for population genetic analysis. Mol.Biol. Biotechnology. Ctr.,University of Alberta, Canada.

Zeng G, Li D, Han Y, Weili T, Wang J, Qiu L, Li W (2009) Identification of quantitative trait loci underlying isoflavone contents in soybean seeds among multiple environments. Theor App Genet. 118(8):1455-1463.

Zhang HJ, Li JW, Liu YJ,Jiang WZ, Du XL, Li L, Li XW, Su LT, Wang QY, Wang Y (2014a) Quantitative trait loci analysis of individual and total isoflavone contents in soybean seeds. J Genet. 93:331-338.

Zhang J, Yinan G, Han F, Li B, Yan S, Sun J (2014b) Isoflavones content of soybean cultivars from maturity group 0 to VI grown in northern and southern China. J Am Oil Chem Soc. 91(6):1019–1028.