ScienceAsia 28 (2002) : **227-239**

Genetic Relationship among Exotic Soybean Introductions in Thailand: Consequence for Varietal Registration

Abul Kashem Chowdhury[a], Peerasak Srinives[b,]*, Panie Tongpamnak[c], Panapa Saksoong[d] and Presert Chatwachirawong[b]

[a] Department of Genetics and Plant Breeding, Faculty of Agriculture, Patuakhali Science and Technology University, Patuakhali-8602, Bangladesh.

[b] Department of Agronomy, Faculty of Agriculture, Kasetsart University, Kamphaeng Saen, Nakhon Pathom 73140, Thailand.

[c] Central Laboratory and Greenhouse Complex, Kasetsart University Research and Development Institute, Kamphaeng Saen, Nakhon Pathom 73140, Thailand.

[d] Department of Genetics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand.

* Corresponding author, E-mail: agrpss@ku.ac.th

Received 29 Oct 2001
Accepted 11 Dec 2001

ABSTRACT This research was designed to identify forty-eight exotic germplasm lines that were genetically and geographically distinct from the existing Thai soybean lines. Using 11 morphological descriptors, all genotypes were classified according to 37 morphological markers which allowed fully discrimination of the cultivars. Similarity indices between cultivars were calculated from 37 binary character states using Dice coefficient, which varied from 0.0 to 0.92 with an average of 0.449. The UPGMA cluster analysis revealed two groups, one formed by 32 cultivars and the other by the remaining 16 cultivars. DNA samples from forty-eight exotic soybean cultivars was examined to determine the efficiency of randomly amplified polymorphic DNA (RAPD) markers in identifying cultivars and determining level of genetic similarity. Out of 80 random primers, 37 generated highly reproducible polymorphic RAPD fragments. With these primers, 274 clear-cut RAPD markers were produced and only 85 (31%) were polymorphic, which indicated that high level of genetic similarities existed in these exotic cultivars. One to six alleles per primer were detected with a polymorphic information content varying from 0.04 to 0.50. The use of only 14 RAPD markers amplified from five primers was sufficient to identify uniquely all the cultivars, indicating that RAPD markers are efficient for use in genetic fingerprinting in soybean. Genetic similarities of 85 RAPD profiles were estimated via the DICE coefficient and then the data were processed using UPGMA clustering method. Each genotype was clearly identified and separated from the others. RAPD based dendrogram revealed that the 48 cultivars could be classified into four groups at 0.57 similarity scale, between which the similarity coefficient was as low as 0.51, even though the cultivars are morphologically or geographically very close. The Principal Component Analysis (PCA) yielded rather similar results to the UPGMA dendrogram. RAPD genetic similarity coefficients were correlated with morphological similarity coefficients (r = 0.241). Comparing agronomic performance and RAPD analysis via dendrogram, a total of 11 cultivars were ear-marked for crossing program. These genotypes can be useful to soybean breeders in Thailand who want to utilize genetically diverse introductions in soybean improvement.

KEYWORDS: exotic soybean germplasm, genetic similarity, morphological markers, RAPD markers, cluster analysis.

INTRODUCTION

Soybean (*Glycine max* (L) Merr) originated in China but has been grown in Korea and Japan for more than 2,000 years. These three countries are thus considered as major sources of soybean germplasm. With the advent of genetics and plant breeding, selection has been intensified for high yield potential with broader adaptation.[1] In general, to improve the genetic gain, breeders need the existing genetic diversity. Unfortunately in Thailand, there is no indigenous genetic pool or land races of soybean and their possibilities of being used as source material in breeding programs. Expanding the genetic base of soybean may introduce unique favorable alleles for polygenic traits.[2] This can be done by incorporating Plant Introductions (PIs) with agronomic merits into breeding programs.[2-4]

Previous research has indicated that the best outcome of introgressing PI germplasm into the current soybean genetic base would be to increase genetic diversity without reducing yield.[4] Furthermore, to utilize introduced germplasm to increase productivity and provide new sources of genetic variation for future gain, selection criteria for parental stock need to consider genetic relationship as well as agronomic value. Agronomic performance of exotic germplasm in the target environment may be taken into account in parental selection, but it could not predict the probability of obtaining new allelic diversity.

Today's soybean breeders observe a limited variability of major characters such as agro-morphological traits among the modern cultivars,[5] due to the lack of genetic variability. Genetic markers are being increasingly utilized in cultivar development, quality control of seed production, measurement of genetic diversity for conservation management, and varietal identification. Methods of varietal identification or similarity estimation are most frequently based on assessment of a range of morphological characteristics. In soybean improvement, many morphological markers express undesirable effects on plant phenotype and their use in crop improvement is limited. Recently, DNA markers were introduced for a more precise characterization and thus offer the potential for unique identification of self-pollinated crop varieties like soybean.

Better knowledge on genetic similarity of breeding materials helps maintaining genetic diversity and sustaining long-term selection gain. Decisions on registration and/or protection of a new candidate variety have crucial economic consequence for breeders and farmers. Moreover, variety identification assures farmers and processors of correct genotype and special provenance of the varieties offered for sale. Genetic identity of cultivated soybeans has been assayed on the basis of morphological traits,[5, 6] allozymes,[7] restriction fragment length polymorphisms (RFLP),[8, 9] random amplified polymorphic DNA (RAPD),[2, 10-12] amplified fragment length polymorphism (AFLP),[9] and simple sequence repeat (SSR).[9, 13]

Among the available DNA molecular techniques, RAPD has many advantages over others. As for example, RAPD technique is simple, quick, inexpensive, requires only small amount of DNA, largely automatable, and requires neither known DNA sequence information nor radio-isotope labeling for sample detection.[14] This technique has been successfully applied in registration activities.[15-16] In spite of these benefits consistency of results within and between laboratories is the main obstacle. This problem, however, has become less problematic as the mechanism of PCR generating RAPD fragments is more understood.[17]

Cultivar identification based on plant phenotype-derived markers is very limited, time consuming, and only visible at a distinct stage of plant development. But some of them are highly heritable and stable which could be utilized in varietal registration[18], especially in combination with extensive morphological analysis. Therefore the objectives of this study were to (i) identify some elite soybean Plant Introductions (PIs) and (ii) analyze the structure of the genetic similarity revealed by morphological and RAPD markers.

Materials and Methods

Plant materials

Based on previous yield testing, forty-eight field soybean varieties (Table 1) were selected for evaluation in this study. All varieties were sown in the nursery of Asian Regional Center- Asian Vegetable Research and Development Center (ARC-AVRDC), Kasetsart University, Kamphaeng Saen Campus. Variety-wise, six agronomic and 11 morphological characters were recorded in a field trial using standard descriptors.[19]

Bulk DNA isolation

Ten days after emergence, the first trifoliate leaves from eight young seedlings from each of the 48 cultivars were sampled and subjected to DNA extraction. Total genomic DNA was isolated from bulked leaf tissue using the protocol of Doyle and Doyle.[20] The presence of DNA was monitored by subjecting samples to 1% agarose gel electrophoresis in 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA) with 2 μL ethidium bromide (0.1 g/mL) and by visual assessment of band intensities compared with λ phage DNA standards. The exact DNA concentration and purity was determined by Spectrophotometry and the concentration was adjusted to 5 ng/μL.

Individual plant DNA isolation

Plant-wise genomic DNA was isolated from 8 individual plants each of 10 selected varieties. Two varieties from China (Wea, and PI 68481), two from Japan (Wakajima and Tastee 824), two from the Philippines (Sl-6 and Multivar), two from Taiwan (HS1 and KS-519), and two from USA (Acadian and Palmetto) were extracted for leaf DNA using the afore-mentioned protocol.

Table 1. Forty-eight soybean genotypes used in the study with their major agronomic traits.

Cultivar no.	Cultivar name	AVRDC Entry no.	Seed origin*	Days to- Flowering	Maturity	Plant height (cm)	Seeds/ pod (no)	100-seed weight (g)	Yield/ plant (g)
1	Acadian	G0001	USA	33	81	55.6	2.3	11.0	10.2
2	Palmetto	G0002	USA	33	81	58.0	2.4	11.0	18.6
3	Wakajima	G0004	Japan	27	82	40.0	2.0	19.5	21.5
4	Native Variety.	G0005	Taiwan	30	80	27.0	2.4	16.0	10.0
5	HS1	G0009	Taiwan	29	86	46.2	2.5	19.8	11.0
6	KS-519	G0014	Taiwan	32	86	33.0	2.5	22.0	15.3
7	66-G-3	G0020	Taiwan	33	82	38.0	2.8	17.0	17.2
8	SP Soybean	G0025	Hong Kong	30	84	27.6	2.1	16.0	18.5
9	TE 32	G0035	Philippines	30	82	29.0	2.5	10.5	23.6
10	Shih Shih	G0038	Taiwan	29	82	28.8	2.0	18.0	16.7
11	Huang-Pau-Tsu	G0040	Taiwan	29	82	32.6	2.3	21.0	20.4
12	PI 153212	G0043	USA	28	81	28.4	2.2	19.0	19.7
13	Wilken	G0048	USA	29	85	19.4	2.0	17.0	8.0
14	Kaohshiung #3	G0055	Taiwan	30	82	20.0	2.0	20.0	18.5
15	KS 419	G0062	Taiwan	36	85	41.4	1.9	19.0	16.3
16	Chung-Hsing #3	G0068	Taiwan	36	85	55.4	2.0	12.0	23.0
17	Wayne	G0072	USA	35	85	69.2	1.9	13.0	15.9
18	SL-6	G0075	Philippines	27	82	42.6	2.5	17.0	14.3
19	Hill	G0081	USA	30	82	28.0	2.1	21.0	15.3
20	I-113	G0095	Philippines	32	85	49.0	1.9	15.0	9.2
21	Kanrich	G0124	USA	32	84	35.0	2.0	13.0	9.3
22	WI-4243	G0129	Philippines	29	84	28.4	2.2	13.0	16.3
23	Multivar	G0132	Philippines	32	89	58.4	2.0	15.0	4.6
24	PI 189860	G0188	France	30	81	29.4	2.0	11.0	17.3
25	Norchief	G0244	USA	29	81	34.0	2.0	19.0	13.6
26	Polland yellow	G0245	UK	28	80	35.0	2.4	14.0	8.2
27	Chippewa	G0377	USA	30	81	39.0	2.4	15.0	20.7
28	PI 153253	G0515	Belgium	31	81	41.0	2.4	15.0	22.2
29	PI 184045	G0543	Yugoslavia	29	81	35.0	2.2	18.0	15.6
30	Funman	G0605	USA	30	82	29.0	1.9	13.0	7.8
31	Harosoy	G0606	USA	29	82	34.0	2.4	18.0	15.3
32	Lindarin	G0610	USA	29	84	35.0	2.0	18.0	8.2
33	Manchu 3 Wis	G0612	USA	31	90	32.0	2.0	20.0	6.3
34	Sousei 823	G0616	Japan	28	81	21.0	2.3	18.0	6.5
35	Tastee 824	G0617	Japan	29	84	42.0	2.0	17.0	10.2
36	Wea	G0619	China	32	81	18.8	2.4	16.0	8.1
37	Yellow Marvel 893	G0620	UK	29	81	16.0	2.0	17.0	6.5
38	PI 68481	G0658	China	30	81	31.4	2.5	18.0	16.3
39	PI 68482	G0659	China	31	81	44.6	2.0	13.0	8.1
40	PI 68543 501	G0668	China	30	87	37.8	2.4	16.0	18.0
41	PI 68683 527	G0694	China	30	82	50.2	2.5	10.0	12.6
42	Bavender special	G0998	USA	30	85	55.0	2.2	13.0	18.6
43	Illington	G1001	Japan	28	82	22.4	2.4	21.0	11.2
44	Litteuiorder 449	G1003	UK	28	82	40.0	2.0	20.0	17.8
45	Manchu 390	G1005	USA	28	82	48.0	2.4	16.0	20.0
46	Manchuria 391	G1006	USA	30	84	36.0	2.2	16.0	8.3
47	Shelby	G1012	USA	31	88	46.4	2.0	18.0	10.0
48	PI 89146-4	G1204	Korea	30	85	35.0	2.0	16.0	12.4
Average				30.2	83.0	36.7	2.2	16.3	13.9
±SD				±2.03	±2.34	±11.2	±0.22	±3.1	±5.98

* Based on the seed receipt record of AVRDC, Taiwan.

Polymerase chain reaction (PCR)

Amplification reactions were performed following the protocol reported by Williams *et al*[21] with minor modifications. PCR reaction mixtures were in volumes of 10 µL containing 2 µL extracted genomic DNA, 1 µL 10x PCR buffer (100 mM Tris-Cl buffer, 500 mM KCl, 20 mM $MgCl_2$, and 0.01%Gelatin), 1 µL 1 mM dNTPs (Promega, USA), 1 µL 2 mM primer, 0.2 µL (1 unit/µL) Taq DNA polymerase (Promega) and 4.8 µL sterile water. Each tube was added with 30 µL of sterile mineral oil to seal the reaction mixture and to prevent evaporation. The PCR was carried out in a DNA Thermal Cycler (Bio Oven III) programmed to run the following temperature profile: 1 cycle of 1 min at 94 °C; 44 cycles of 1 min at 91 °C for denaturation; 1 min at 36 °C for annealing; 2 min at 72 °C for extension; 1 cycle of 7 min at 72 °C as the final extension.

Electrophoresis

The amplification products were size-separated by electrophoresis in 1.6% agarose gels containing 0.6 mL 50x TAE buffer (Tris-Acetate-EDTA, 2M Tris aminomethane, 5.71% acetic acids, 50 mM EDTA) and 1 µL ethidium bromide (10 mg/mL). The PCR product was mixed with 3 µL of BFF (1.2 mg/mL bromophenol; 125 mg/mL Ficoll) and the whole mixture was applied in each well of the gel. DNA molecular weight markers (λ DNA digested with *Hin*d III and *Eco*RI) were added and the gels were electrophoresised in 1x TAE buffer with 2 µL ethidium bromide at the electric potential of 100 volts until the front marker of BFF had reached 1 cm from the end of the gel. Gels stained with ethidium bromide solution (0.08 mg/ml) were exposed to UV light and the images were photographed using polaroid 667 film. Observations were made from photographs. Polymorphisms at all loci were confirmed by three repeating tests at different times.

Data analysis

Statistical analyses of the morphological data were performed using a binary data matrix. The matrix described the 11 qualitative morphological characteristics and can be used to determine the aggregate morphological similarities among genotypes. The term "RAPD band" was used here to describe a set of unit character amplified by the same RAPD primer. Each variable RAPD band was considered as a locus so that every locus had two alleles and scored as present (1) or absent (0). For data analysis, only polymorphic, reproducible, and clear-cut bands were kept. The polymorphism information content (PIC) of each RAPD marker was determined as described by Weir.[22] NTSYS-pc, version 2.01 d was used to calculate the genetic similarity matrices based on Nei and Li's Dice coefficients,[23] Jaccar d's coefficients (J)[24] and Sneath and Sokal's simple matching (SM) coefficients.[25] Dendrograms were constructed by the unweighted pair-group method using arithmetic averages (UPGMA) algorithm as described by Sneath and Sokal.[26] Principal Component Analysis (PCA) of Dice similarity values was calculated by MATLAB for windows program.

Results

Agro-morphological classification

The 48 genotypes displayed polymorphism for both quantitative (Table 1) and qualitative (Table 2) agro-morphological characteristics. As expected, significant variability was observed among the genotypes. Using eleven morphological traits, 39 cultivars and 9 PIs produced 37 morphological markers (binary character states). Two cultivars revealed specific morphological markers, namely PI 189860 (24) characterized by reddish brown seed coat and Litteuiorder 449 (44) by brown seed coat color. SP Soybean (8) and WI-4243 (22) characterized by black seed coat. Cultivars possessing green hypocotyl showed white flowers while those having purple hypocotyl had purple flowers since both traits are governed by the same locus of gene. The remaining morphological traits were scored in differential combinations of cultivars and PIs.

Optimization of RAPD protocol

Eighty 10-mer primers from Operon Kits L, N, O, and P were initially screened against 16 varieties (see variety no 1-16 in Table 1). Since the RAPD-PCR protocol is sensitive to reaction conditions, the effects of magnesium and template DNA concentrations, pH values and duration of time during the denaturation step of amplification were examined. Under the optimal conditions cited in the Materials and Methods, 73 out of 80 primers (91.3%) generated RAPD bands, typically with 1 to 14 major bands and a number of minor bands of less intensity, while 7 did not produce any amplified products. Of the set of 73 primers, 36 revealed monomorphic RAPD bands across all screened varieties. Consequently, the remaining 37 primers (Table 3) were chosen for further analyses based on the existence of polymorphic bands.

Table 2. Morphological characterization of 48 soybean genotypes using 11 qualitative characters.

Cultivar no.	Cultivar name	Morphological description -										
		A	B	C	D	E	F	G	H	I	J	K
1	Acadian	2	7	7	2	2	3	7	2	5	3	2
2	Palmetto	2	7	5	2	2	7	5	7	7	6	2
3	Wakajima	2	7	7	2	2	5	7	7	5	3	2
4	Native Variety.	2	7	3	3	1	3	0	7	7	8	2
5	HS1	1	3	9	1	2	7	0	7	5	6	2
6	KS-519	2	7	7	3	1	3	0	7	3	6	2
7	66-G-3	1	3	7	2	1	7	0	7	7	6	2
8	SP Soybean	1	3	7	3	1	3	5	1	5	7	8
9	TE 32	2	7	9	1	1	7	0	7	7	6	2
10	Shih Shih	2	7	7	1	1	5	0	7	5	3	2
11	Huang-Pau-Tsu	2	7	7	1	1	3	3	2	5	1	2
12	PI 153212	1	3	7	3	1	3	3	7	3	6	5
13	Wilken	1	3	7	1	1	7	0	5	3	1	2
14	Kaohshiung #3	1	3	7	2	1	7	0	5	5	8	2
15	KS 419	1	3	7	2	1	7	0	7	5	8	2
16	Chung-Hsing #3	2	7	7	2	1	5	7	5	5	7	2
17	Wayne	1	3	9	3	1	7	5	7	7	8	2
18	SL-6	2	7	9	2	1	3	0	2	5	6	2
19	Hill	1	3	9	2	1	7	0	2	5	6	2
20	I-113	2	7	7	2	1	5	0	5	5	6	2
21	Kanrich	1	3	7	2	1	3	97	3	1	2	
22	WI-4243	2	7	7	2	1	3	5	2	3	7	8
23	Multivar	1	3	7	1	1	3	0	2	3	1	2
24	PI 189860	1	3	7	2	1	7	0	7	5	3	5
25	Norchief	2	7	7	1	1	3	0	5	3	3	2
26	Poland-yellow	2	7	7	2	1	7	0	2	3	6	2
27	Chippewa	2	7	9	2	1	5	0	5	5	7	2
28	PI 153253	2	7	7	1	1	7	0	2	5	3	2
29	PI 184045	2	7	7	1	1	5	0	7	3	3	2
30	Funman	2	7	9	1	1	3	5	2	3	6	2
31	Harosoy	2	7	9	1	1	7	0	2	7	1	2
32	Lindarin	2	7	5	1	1	7	0	1	3	3	2
33	Manchu 3 Wis	2	7	7	2	1	5	9	1	5	6	2
34	Sousei 823	1	3	7	1	2	3	0	7	7	3	2
35	Tastee	2	7	7	1	1	5	3	7	5	8	2
36	Wea	1	3	9	1	1	7	0	2	3	8	2
37	Yellow Marvel	1	3	7	1	1	5	0	5	3	3	2
38	PI 68481	1	3	7	1	2	7	7	1	5	8	2
39	PI 68482	2	7	7	2	1	5	7	5	5	3	5
40	PI 68543 501	2	7	7	1	1	7	7	1	7	8	2
41	PI 68543 527	2	7	7	3	1	7	7	2	3	6	2
42	Bavender sp 38	2	7	3	1	1	7	3	7	7	8	2
43	Illington 318	1	3	9	1	1	5	0	7	7	1	2
44	Litteuiorder 449	2	7	9	1	1	7	7	1	3	3	3
45	Manchu 390	2	7	7	2	1	7	5	2	7	7	2
46	Manchuria 391	2	7	9	1	1	7	7	2	5	3	2
47	Shelby 398	2	7	7	1	1	5	5	5	5	6	2
48	PI 89146-4	2	7	7	2	1	7	0	5	3	8	2

A = Hypocotyl color: 1 = Green; 2 = Purple.
B = Corolla color: 3 = White; 7 = purple.
C = Pubescence density: 3 = Sparse; 5 = Semi-sparse; 7 = Normal; 9 = Dense.
D = Pubescence color: 1 = Grey; 2 = Light brown, 3 = Brown/Tawny.
E = Pubescence type: 1 = Erect; 2 = Semi-appressed.
F = Stem determination: 3 = Determinate; 5 = Semi-determinate; 7 = Indeterminate.
G = Lodging: 0 = None; 3 = Slight; 5 = Moderate; 7 = Severe; 9 = Very severe.
H = Shattering: 1 = No shattering; 2 = Slight shattering; 5 = Medium shattering; 7 = Shattering.
I = Seed quality: 3 = Poor; 5 = Medium; 7 = Good.
J = Hilum color: 1 = Yellow; 3 = Brown; 6 = Imperfect black; 7 = Black; 8 = Dark brown.
K = Seed coat color: 2 = Yellow; 3 = Brown; 5 = Reddish brown; 8 = Black.

Table 3. List of selected Operon primers, their sequences, number of bands, positions of polymorphic fragments, polymorphism (%) and polymorphism information content (PIC) of the RAPD analysis results in 48 field soybean cultivars.

Sl no.	Operon code	Sequences (5' to 3')	Total bands**	Polymorphic fragment	Polymorphism (%)	PIC
1	OPL-01	GGCATGACCT	9	OL-011300	11.11	0.486
2	OPL-03	CCAGCAGCTT	6	OL-031800	33.33	0.469
				OL-031650		0.486
3	OPL-04	GACTGCACAC	4	OL-04500	25.0	0.444
4	OPL-12	GGGCGGTACT	9	OL-121600	22.22	0.457
				OL-121100		0.444
5	OPL-13	ACCGCCTGCT	10	OL-131150	20.0	0.117
				OL-13947		0.249
6	OPL-14	GTGACAGGCT	8	OL-141000	25.0	0.420
				OL-14525		0.492
7	OPL-17	AGCCTGAGCC	8	OL-171584	25.0	0.187
				OL-17500		0.395
8	OPL-19	GAGTGGTGAC	6	OL-191800	33.33	0.041
				OL-191375		0.492
9	OPN-03	GGTACTCCCC	4	ON-031350	50.00	0.305
				ON-031150		0.457
10	OPN-04	GACCGACCCA	6	ON-04989	33.33	0.249
				ON-04750		0.457
11	OPN-08	ACCTCAGCTC	6	ON-081584	33.33	0.330
				ON-08600		0.499
12	OPN-09	TGCCGGCTTG	10	ON-091375	30.00	0.413
				ON-091200		0.278
				ON-09584		0.497
13	OPN-11	TCGCCGCAAA#	5	ON-111584	100	0.478
				ON-111375		0.305
				ON-111050		0.278
				ON-11831		0.330
				ON-11500		0.444
14	OPN-14	TCGTGCGGGT	7	ON-14989	14.28	0.187
15	OPN-16	AAGCGACCTG	14	ON-161100	21.43	0.330
				ON-16750		0.413
				ON-16500		0.444
16	OPN-18	GGTGAGGTCA	8	ON-181375	25.00	0.413
				ON-181050		0.430
17	OPN-20	GGTGCTCCGT	7	ON-202000	42.85	0.353
				ON-201350		0.499
				ON-20987		0.444
18	OPO-01	GGCACGTAAG#	4	OO-011375	100.00	0.305
				OO-011050		0.457
				OO-01947		0.375
				OO-01831		0.413
19	OPO-02	ACGTAGCGTG	7	OO-021650	14.28	0.499
20	OPO-05	CCCAGTCACT	8	OO-051375	12.50	0.330
21	OPO-09	TCCCACGCAA	5	OO-09947	40.00	0.353
				OO-09831		0.478
22	OPO-11	GACAGGAGGT	7	OO-111350	14.28	0.219
23	OPO-13	GTCAGAGTCC	6	OO-131375	16.66	0.444
24	OPO-15	TGGCGTCCTT#	7	OO-151584	57.44	0.492
				OO-151100		0.469
				OO-15947		0.499
				OO-15831		0.153
25	OPO-16	TCGGCGGTTC	6	OO-161584	33.33	0.497
				OO-161375		0.497
26	OPO-18	CTCGCTATCC	4	OO-181450	25.00	0.499
27	OPO-19	GGTGCACGTT#	11	OO-191800	54.54	0.486
				OO-191500		0.305
				OO-191200		0.444
				OO-191150		0.430
				OO-19700		0.430
				OO-19500		0.497

Table 3. List of selected Operon primers, their sequences, number of bands, positions of polymorphic fragments, polymorphism (%) and polymorphism information content (PIC) of the RAPD analysis results in 48 field soybean cultivars. (cont)

Sl no.	Operon code	Sequences (5' to 3')	Total bands※※	Polymorphic fragment	Polymorphism (%)	PIC
28	OPO-20	ACACACGCTG	6	OO-201584 OO-20947	33.33	0.249 0.249
29	OPP-01	GTAGCACTCC	7	OP-011375 OP-01870	28.57	0.353 0.187
30	OPP-06	GTGGGCTGAC	7	OP-061584 OP-061050	28.57	0.305 0.469
31	OPP-07	GTCCATGCCA	9	OP-07564	11.11	0.277
32	OPP-08	ACATCGCCCA	11	OP-081300 OP-081100 OP-08947 OP-08564	36.36	0.277 0.469 0.499 0.375
33	OPP-09	GTGGTCCGCA#	10	OP-091584 OP-091400 OP-091200 OP-091050	40.00	0.457 0.497 0.117 0.080
34	OPP-11	AACGCGTCGG	7	OP-112000 OP-111800 OP-111584 OP-111100	57.14	0.330 0.444 0.430 0.249
35	OPP-14	CCAGCCGAAC	7	OP-141375 OP-14947	28.57	0.499 0.117
36	OPP-17	TGACCCGCCT	8	OP-171375 OP-171050 OP-17564	37.50	0.330 0.430 0.278
37	OPP-18	GGCTTGGCCT	10	OP-18831	10.0	0.500
Total			**274**	**85**	**---**	**---**
Average ± SD			**7.4 ± 2.19**	**2.3 ± 1.23**	**31.02**	**0.377 ± 0.12**

※※ Total number of scorable bands detected
\# Primer used for intra-varietal variation analyses

RAPD polymorphism and power of discrimination

Thirty-seven primers were scored for their consistent production of strong amplification and reproducible band criteria in three replicated PCRs across 48 soybean lines (Table 3). A total of 274 bands were generated using the 37 selected primers. The number of bands produced by each primer varied from 4 (OPL-04, OPN-03, OPO-01, and OPO-18) to 14 (OPN-16) with the mean ± SD of 7.4 ± 2.19 bands per primer. Sizes of the amplified fragments ranged from 300 bp to more than 2 kbp. Out of 274 bands observed, 137 (50 %) were monomorphic for all varieties examined in this study. Of the remaining 137 variable bands, 85 (31.02 %) were reproducible polymorphic and thus regarded as informative RAPD markers for the genetic structure study. These informative markers were able to differentiate all varieties. Each variety could be distinguished by at least four RAPD markers. The remaining 52 (18.98%) bands were unstable, ie non-reproducible when the amplifications were repeated at different times, and thus were excluded from further study. Among 85 clear-cut reproducible bands, only 14 RAPD markers amplified by primers OPN-03, OPN-04, OPN-08, OPN-11, and OPN-16 successfully distinguished all the genotypes. OL-191800 fragment was PI 189860 line specific. In this experiment, different cultivars revealed different banding patterns which were generated by different primers. An example of the banding pattern and polymorphism detected with primer OPN-16 was shown in Fig. 1. Percentages of polymorphic bands for each primers and polymorphic information content (PIC) for each marker were shown in Table 3. Primer OPN-11 and OPO-1 exhibited the greatest level of polymorphism (100%), whereas primer OPP-18 exhibited the lowest level (10%). Analysis of 85 RAPD loci among all cultivars of soybean showed that the PIC among all polymorphic loci ranged from 0.041 to 0.50 with a mean value ± SD of 0.377 ± 0.117. Sixty-five RAPD fragments from 32 primers with PIC > 0.30 were observed and considered to be informative.

Genetic similarity

Eleven morphological traits described by 37 binary character states and 85 polymorphic RAPD markers of the 48 soybean varieties amplified by 37 primers were separately used to calculate genetic similarities for all possible 1,128 pairwise comparisons (data not shown). Using these traits, the genetic similarity between any two cultivars was between 0.0 [HS1 (5) vs WI-4243 (22)] and 0.917 [Wilken (13) vs Multivar (23)] with the mean ± SD value of 0.449 ± 0.114. On the other hand, the RAPD based lowest genetic similarity of 0.345 was found between the pair HS1 (5) vs Manchu 3 Wis (33) with 44 RAPD marker difference. The highest similarity coefficient of 0.941 was found between Sousei (34) vs Yellow Marvel (37) with four marker differences. Using 85 RAPD markers, the average genetic similarity coefficient recorded at about 0.577 ± 0.082, which was higher than that of the morphological markers. Substantial genetic variation existed among the soybean cultivars with an average of 0.513 using both types of marker. Correlation coefficient between the Dice coefficient determined by morphologic means and RAPD was 0.241** (P<0.01).

Ranging of similarity values among cultivars based on the Dice, Jaccard and SM coefficients were shown in Table 4. Similarity matrices based on the three similarity coefficients were highly correlated (r = 0.986 for Dice and Jaccard, r = 0.941 for Dice and SM, and r = 0.953 for Jaccard and SM; all are significant at P<0.001). However, the overall pairwise similarity values in Jaccard coefficient were lower than that of the others. Regardless of the estimators considered, Sousei (34) and Yellow marvel

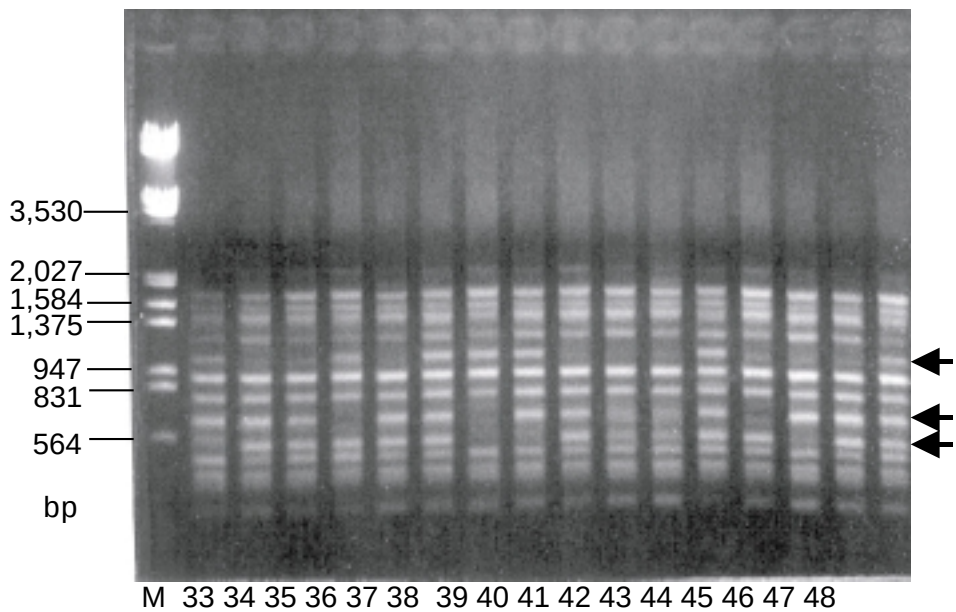

Fig 1. RAPD profiles of a subset of the soybean genotypes amplified by primer OPN-16. M is the DNA marker (λ DNA digested with *Hin*dIII and *Eco*RI). Lanes 33 to 48 correspond to series number for cultivars listed in Table 1. The polymorphic bands are marked by arrow.

Table 4. Range, mean ± SD, and marker difference for the three genetic similarity estimators calculated from 37 morphological and 85 RAPD markers of 48 field soybean cultivars.

Estimator#	Morphological markers			RAPD markers		
	Range		Mean ± SD	Range		Mean ± SD
	Min	Max		Min	Max	
J	0.0	0.846	0.279 ± 0.128	0.210	0.889	0.431 ± 0.109
NL	0.0	0.917	0.449 ± 0.114	0.345	0.941	0.577 ± 0.082
SM	0.405	0.946	0.655 ± 0.091	0.424	0.953	0.624 ± 0.072

J = Jaccard's coefficient, NL= Nei and Li's Dice coefficient and SM = Sneath and Sokal's Simple matching coefficient.

(37) were the closest cultivars with coefficients of 0.889, 0.941, and 0.953 for J, Dice, and SM, respectively (Table not shown). Acadian (1) vs Shelby (47), HS1 (5) vs Manchu 3 Wis (33), Manchu 390 (45) vs Chung-Hsing (16), Illington 318 vs Chung-hsing and Chung-Hsing (16) vs Shelby (47) gave low similarity values (<0.350), indicating that they were genetically distant cultivars.

Dendrogram analyses

Dendrograms were constructed on the basis of Dice similarity matrices using UPGMA method to show the genetic structure based on morphology and RAPD markers. Forty-eight cultivars were formed into two main groups in the dendrogram constructed through morphological markers, containing 32 and 16 lines, respectively (Fig. 2). Both groups were further classified into large number of sub-clusters at different similarity coefficient levels having a clear-cut discrimination of all 48 cultivars. On the other hand, the dendrogram constructed by 85 RAPD markers revealed that these soybean cultivars fell into two main groups (Fig. 3). One was formed by Acadian, Palmetto, and Chung-hsing #3. The other group can be further separated into three sub-groups at the 0.57 level of similarity, containing 18, 14 and 13 accessions, respectively. Each of the sub-groups could be further divided into two well defined clusters. However, the relative positions of the cultivars in morphological and RAPD markers based dendrograms were rather different (Fig. 2 and 3).

Principal Component Analyses (PCA)

Figure 4 shows association among the 48 cultivars revealed by PCA. The first (PCA1) and second (PC2) principal component explained 41.43

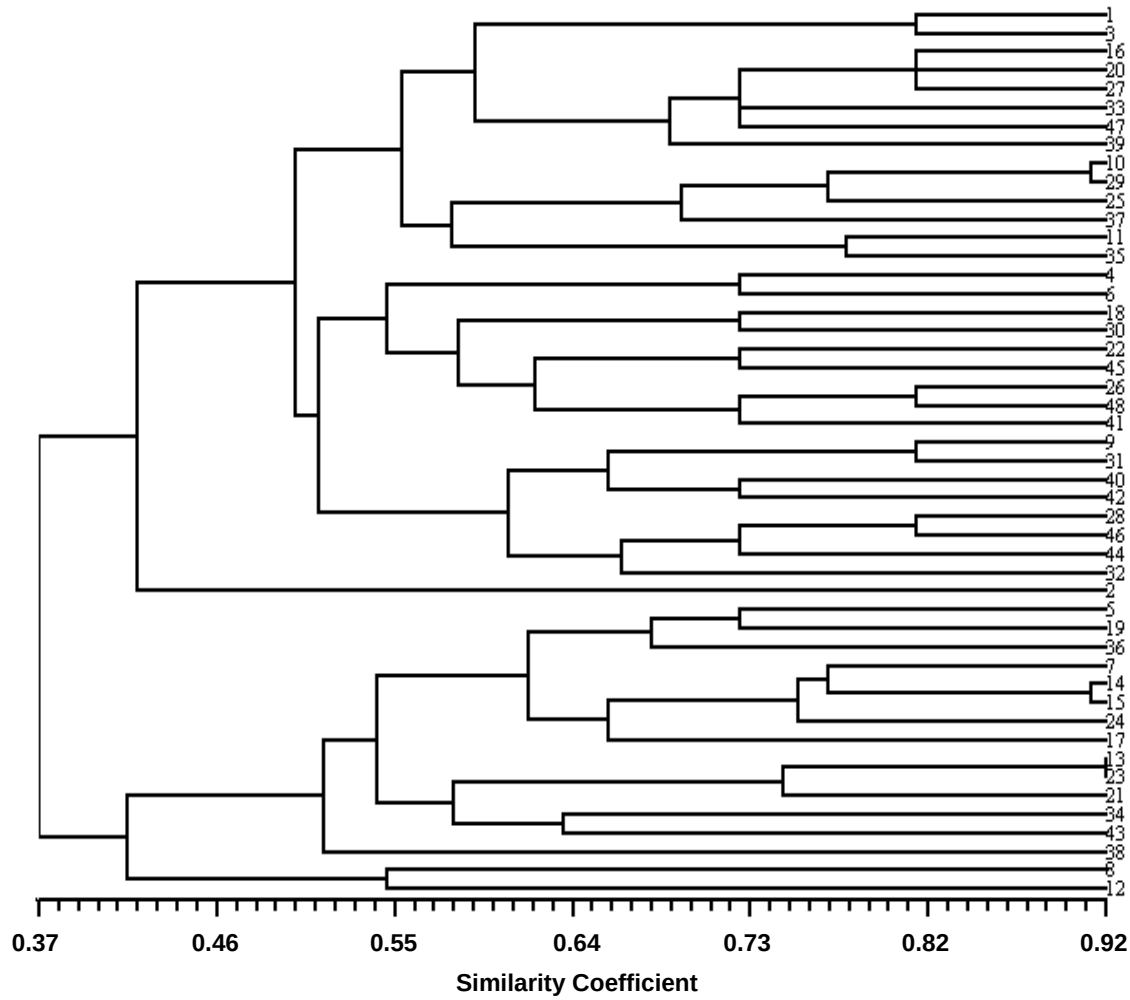

Fig 2. Dendrogram illustrating genetic relationship among 48 exotic soybean cultivars generated by UPGMA cluster calculated from 11 morphological traits described by 37 binary character states listed in Table 2. Scale at the bottom is Dice coefficient of similarity. The genotypes are numbered and defined as in Table 1.

and 4.52 % of the total variation in RAPD data, respectively. The PCA analysis yielded rather similar results to the UPGMA dendrograms, ie there were two main groups. The first major split showed three distinct subgroups, viz. Acadian (V1), Palmetto (V2) and Chung - Hsing #3 (V16). The second group showed that several sectional groups of the exotic cultivars were dispersed, they were clearly distinct from each other and could be sub-divided at different variation levels.

Analysis of genetic variation within cultivars

Eight genotypes were assayed for genetic uniformity, using five primers namely OPN-11, OPO-01, OPO-15, OPO-19 and OPP-09. No polymorphism was detected in the 8 tested cultivars, since soybeans are self-pollinated crop and thus the cultivars are maintained in pure-line state of homozygous genotype.

DISCUSSION

Genetic discrimination among 48 selected accessions of exotic soybean was assessed with 11 morphologic traits and 85 RAPD markers to test their possible duplication and to estimate their genetic similarity. Sufficient discriminatory power of 11 morphological characters revealed in this study, had previously been noted in soybean by Gizlice *et al*[6] using 10 morphological traits under controlled conditions. This finding concurred with the previous studies in which low levels of polymorphism were detected among USDA Soybean Germplasm Collection[2, 10-11] and among ARC-AVRDC Vegetable Soybean Germplasm Collection on the basis of RAPD analysis.[12] The range of polymorphisms or diversity detected by RAPD markers within the annual *Glycine* species is unknown. But most reports on soybean genetic variation concluded that the diversity is low, even compared with other self-pollinated legume species.[10]

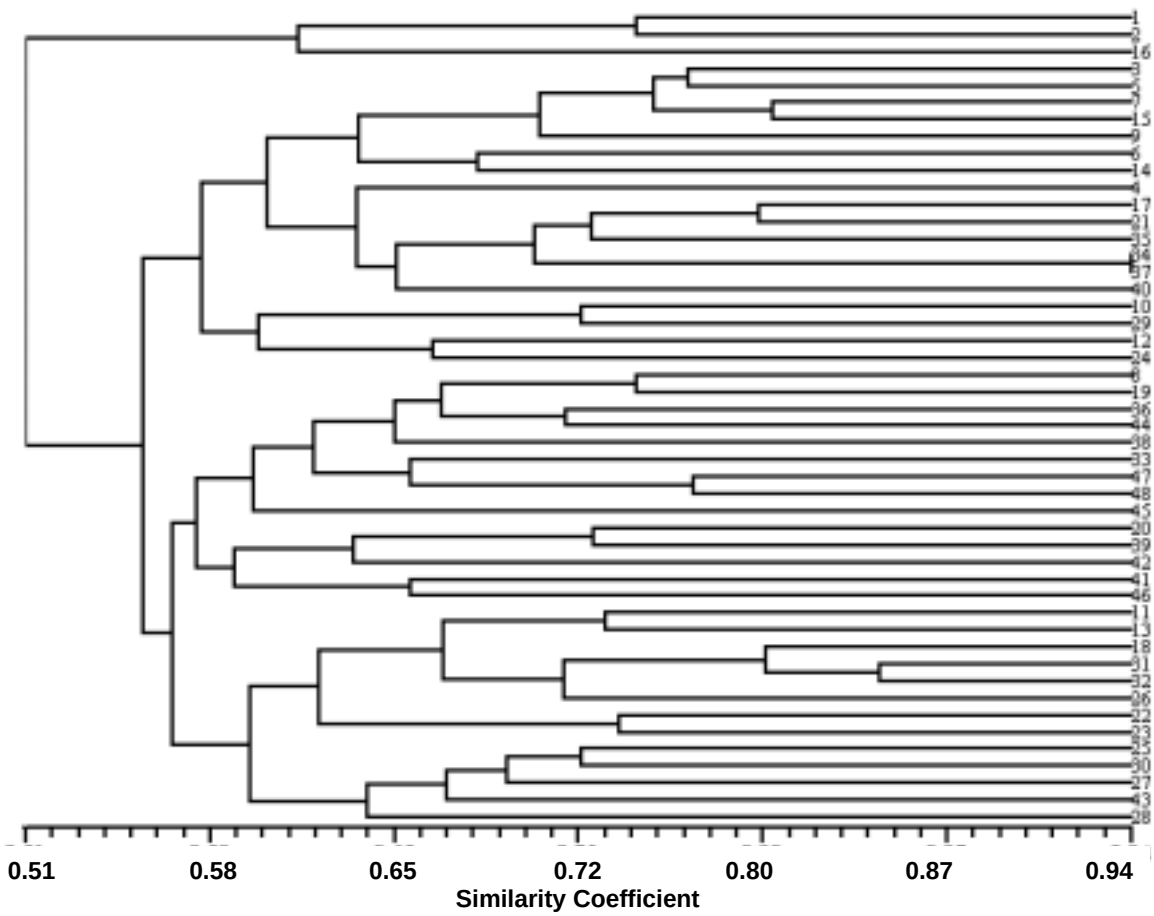

Fig 3. Dendrogram illustrating genetic relationship among 48 exotic soybean cultivars generated by UPGMA cluster calculated from 85 polymorphic RAPD markers amplified by 37 Operon primers as listed in Table 3. Scale at the bottom is genetic relatedness derived from Dice coefficient of similarity. The genotypes are numbered and defined as in Table 1.

The smaller number of pairwise differences (high genetic similarity values) among some cultivars is likely due to their genetical relatedness. On the other hand, large number of pairwise differences (low genetic similarity values) should be observed among those cultivars developed from genetically distant parental lines. The average pairwise difference of 24.6 indicated that RAPD analysis gave a high degree of identity among the cultivars examined in this study. In some cases, very few genetic differences were identified with minimum RAPD marker differences. For example, Yang and Quiros[27] reported that one celery cultivar was distinguished from the others by only a single band difference, despite over 300 bands being scored. Such types of low marker difference may create unreliable information for cultivar registration or identification. Since RAPD analyses can amplify low incidence of non-heritable bands, which are probably PCR artifacts. Heun and Helentjaris[28] reported this problem for a small percentage of RAPD bands in maize and noted that similar patterns have been found in other plants. While the great majority of RAPD bands are known to be inherited as Mendelian markers, care is needed when drawing conclusion based on a small number of band differences. If the genetic basis for cultivar difference is dependent on point mutation(s) or the cultivar are heterozygous, the identification of genetic difference by RAPD profiling technique is very difficult.[12, 29] However, the results of this study demonstrated that all cultivars were identified by at least four RAPD markers using 37 primers which generated 222 reproducible bands. Stepwise removal of all data from individual primers resulted in deletion of data from 37 primers affecting the outcome of the cluster analysis and reduced the

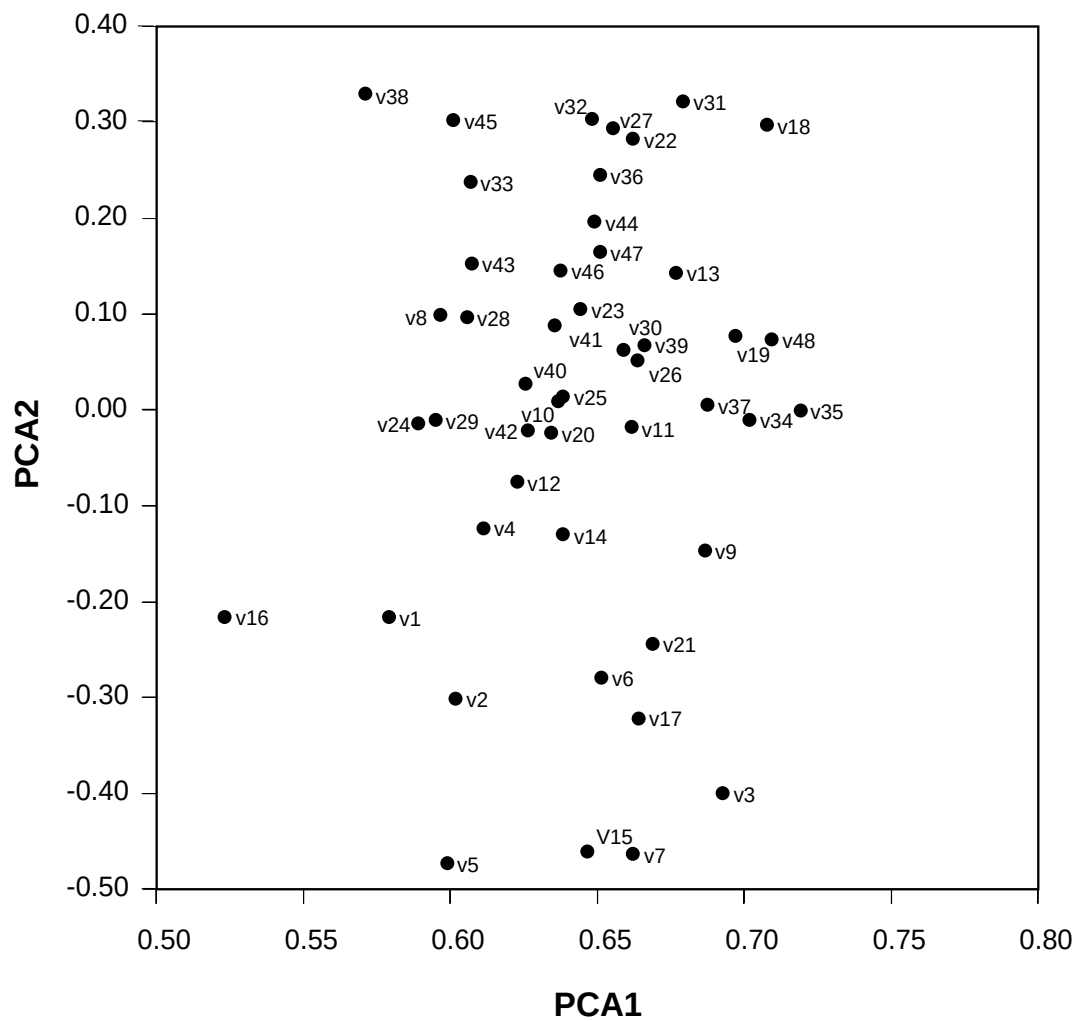

Fig 4. Two-dimension principle component analysis (PCA) using similarity values of RAPD binary data of the 48 exotic soybean cultivars. Numbers of the cultivars are listed in Table 1.

number of primers containing critical information to 5. This set of 5 primers containing 14 polymorphic fragments could be successfully used to discriminate all of the cultivars with at least one RAPD marker difference. This compares to the 14 morphological markers which could not discriminate between the cultivars. Therefore, RAPD markers can identify cultivars despite they are morphologically very similar.

The correlation coefficient between genetic similarities calculated from RAPD and morphological data was significant. This result suggested that morphological similarity might be used to predict RAPD similarity values, but additional analysis would be required. The highly significant correlation between Dice and J, J and SM, and Dice and SM showed that an allelic relationship between the absence and the presence of a given band can be assumed. This result was reported in other studies where genotypes came from the same species.[30] In the context of essential derivation which is always the issue in plant breeder's right, the choice of a genetic similarity is crucial for estimating the level of relatedness between cultivars. Genetic similarities between cultivars estimated with Dice, J, and SM coefficients of similarity were very well correlated and led to a very similar assessment of relationships between cultivars. However, Dice method has been widely used in genetic similarity assessment.[29]

The cluster analysis based on 85 polymorphic bands separated this 48 soybean cultivars into two main groups, each included cultivars from all other origins. To investigate sensitivity of the dendrogram against changes in the computational methods, more dendrograms were constructed using different formulae for the relative genetic similarities and two different clustering methods (single and complete linkage). All dendrograms differed only in the arrangement of few genotypes within the 2nd main group. However, several factors may affect the genetic relationship among cultivars, such as number of markers used, distribution of markers in the genome of working samples, and the nature of evolutionary mechanisms underlying the variation measured.[9] Using more markers will affect the variance of the estimated similarity. If linkage disequilibrium is present, equally spaced markers will afford a better estimate than randomly distributed markers. This appeared to be the case for the 48 lines of field soybean evaluated in this study as all cultivars were successfully separated from one another with high similarity value of 0.94 which is still far from 1.0. From this study, using RAPD based dendrogram cultivars Chung-Hsing # 3 (16), TE 32 (9), KS-519 (6), Kaoshiung #3 (14), Native variety (4), PI 68543-501 (40), PI 153212 (12), PI 189860 (24), PI 68481 (38), Manchu 3 Wis (33), Manchu 390 (45), Bevender Special 383 (42), PI 68683 527 (41), Manchuria 391 (46), Polland Yellow (26), Illington (43), and PI 143253 (28) would be ear-marked for use in future breeding programs. Of these candidate lines, cultivars KS-519, TE 32, PI 153212, Kaoshiung #3, Chung-Hsing #3, PI 189860, PI 153253, PI 68481, PI 68543 501, Bavender special 382, Manchu 390 yielded more than the average yield per plant (13.6 g). Thus genetic relationships complemented with phenotypic data can reveal sources of desirable characteristics in closely related individuals. RAPD could be used routinely by plant breeders to identify genetic variation, locate regions of the genome linked to agronomically important genes and facilitate introgression of desirable genes into commercial accessions.

In order to assess whether the grouping of cultivars based on RAPDs could be further resolved, the Principal Component Analysis (PCA) was used to examine common band data available among genotypes. It was found that PCA provided a clear separation of the cultivars in different groupings containing distinct geographical origin. This supported the conclusion drawn from the previous form of UPGMA dendrogram analysis. The application of PCA for evaluation of the relationships between accessions nevertheless depended on the level of resolution desired. While UPGMA clustering provided the best indication of relationship among closely related accessions, ordination appeared to provide a representation of the relationships among major groups. However, the dispersion of the cultivars in PCA plot indicated that the cultivars were diversified within species and also grown in vast geographical area. This result is very much consistent with that of microsatellite markers analysis in cowpea.[31]

In conclusion, (1) RAPD marker can detect varietal difference in soybean though they are morphologically more or less alike, (2) genetically, the soybean varieties used in this study are diversified, (3) RAPD approach is particularly beneficial in establishing criteria of distinctness and uniformity for regulation of varieties under Plant Variety Protection (PVR) legislation which is under development in Thailand, and (4) RAPD technique is cheap and suitable for developing countries to help identifying parents for soybean breeding programs.

ACKNOWLEDGEMENT

This work was supported by Agricultural Research Management Project (ARMP) of Bangladesh, KURDI, ARC-AVRDC and The Thailand Research Fund for Ph.D. research of the senior author at Kasetsart University.

REFERENCES

1. Simmonds NW (1979) *Principles of Crop Improvement*. Longman, London.
2. Thompson JA and Nelson RL (1998) Core set of primers to evaluate genetic diversity in soybean. *Crop Sci* 38, 1356-62.
3. Thorne JC and Fehr WR (1970) Incorporation of high protein, exotic germplasm into soybean population by 2- and 3-way crosses. *Crop Sci* 10, 652-5.
4. Vello NA, Fehr WR and Bahrenfus JB (1984) Genetic variability and agronomic performance of soybean populations developed from plant introductions. *Crop Sci* 24, 511-4.
5. Chowdhury AK, Srinives P, Tongpamnak P and Saksoong P (2001) Genetic diversity based on morphology and RAPD analysis in vegetable soybean. *Korean J Crop Sci* 46, 112-20.
6. Gizlice Z, Carter TE, Jr and Burton JW (1993) Genetic diversity in North American soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. *Crop Sci* 33, 614-20.
7. Griffin JD and Palmer RG (1995) Variation of thirteen isozyme loci in the USDA soybean germplasm collections. *Crop Sci* 35, 897-904.
8. Keim P, Beavis W, Schupp J and Freestone R (1992) Evaluation of soybean RFLP marker diversity in adapted germplasm. *Theor Appl Genet* 85, 205-12.
9. Powell W, Morgantee M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2, 225-38.
10. Thomson JA, Nelson RL and Vodkin LO (1998) Identification of diverse soybean germplasm using RAPD markers. *Crop Sci* 38, 1348-55.
11. Brown-Guedira GL, Thomson JA, Nelson RL and Warburton ML (2000) Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. *Crop Sci* 40, 815-23.
12. Chowdhury AK, Srinives P, Tongpamnak P and Saksoong P (2000) Identification of cultivars of vegetable soybeans [*Glycine max* (L) Merr] by RAPD markers *SABRAO J Breed Gen* 32, 63-72.
13. Diwan N and Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor Appl Genet* 95, 723-33.
14. Hedrick P (1992) Shooting the RAPDs. *Nature* 355, 679-80.
15. Mailer RJ, Scarth R and Fristensky B (1994) Discrimination among cultivars of rape seed (*Brassica napus* L) using DNA polymorphism amplified from arbitrary primers. *Theor Appl Genet* 87, 853-8.
16. Ford R and Taylor PWJ (1997) The application of RAPD markers for potato cultivar identification. *Aust J Agric Res* 48, 1213-7.
17. Weising K, Nybom H, Wolf K and Mayer W (1995) *DNA Fingerprinting in Plants and Fungi*. CRC Press, Inc London.
18. Tatineni V, Cantell RG and Davis DD (1996) Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. *Crop Sci* 36, 186-92.
19. International Board for Plant Genetic Resources (IBPGR) (1984) *Descriptors for Soybean.* IBPGR Secretariat, Rome, Italy.
20. Doyle LJ and Doyle JJ (1990) Isolation of plant DNA from fresh tissue. *Focus* 12, 13 -4.
21. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are used as genetic markers. *Nucleic Acids Res* 18, 6531-5.
22. Weir BS (1996) Genetic data analysis II, 2nd ed Sinauer Associates, Inc, Sunderland, MA.
23. Nei M and Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci (USA*) 76, 5269-73.
24. Jaccard P (1908) Nouvelles recherches sur ladistribution florale. *Soc Vaud Sci Nat Bull* 44, 223-70.
25. Puterka GJ, Black WC, Steiner MW and Burton RL (1993) Genetic variation and phylogenetic relationships among worldwide collections of the Russian wheat aphid, *Diuraphis noxia* (Mordvilko), inferred from allozyme and RAPD-PCR markers. *Heredity* 70, 604-18.
26. Sneath PHA and Sokal RR (1973) *Numerical Taxonomy*. WH Freeman. San Francisco.
27. Yang X and Quiros C (1993) Identification and classification of celery cultivars with RAPD markers. *Theor Appl Genet* 86, 205-12.
28. Heun M and Helentjaris T (1993) Inheritance of RAPDs in F1 hybrids of corn. *Theor Appl Genet* 85, 961-8.
29. Morell MK, Peakall R, Appels R, Preston LR and Lioyd HL (1995) DNA profiling techniques for plant variety identification. *Aust J Exp Agric* 35, 807-19.
30. Lombard V, Baril CP, Dubreuil P, Blouet F and Zhang D (2000) Genetic relationships and fingerprinting of rape seed cultivars by AFLP: consequences for varietal registration. *Crop Sci* 40, 1417-25.
31. Li C, Fatokun CA, Ubi B, Singh BB and Scoles GJ (2001) Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. *Crop Sci* 4, 189-97.