Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407–4414.

Whitkus, R., M. de la Cruz and L. Mota-Bravo. 1998. Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. Theor. Appl. Genet. 96:621–627.

Xu, R.Q , N. Tomooka, D.A. Vaughan, and K. Doi. 2000. The *Vigna angularis* complex: genetic variation and relationships revealed by RAPD analysis, and their implication for in situ conservation and domestication. Genet. Res. Crop Evol. (in press).

Yamaguchi, H. 1990. A note on the distribution of semi-wild azuki beans in south-western Japan and their variation in seed color and size (in Japanese). Rep. Soc. Crop Breed. Kinki 35: 36–39.

Yamaguchi, H. 1992. Wild and weed azuki beans in Japan. Econ. Bot. 46: 384–394.

Yee, E., K.K. Kidwell, G.R. Sills, and T.A. Lumpkin. 1999. Diversity among selected *Vigna angularis* (Azuki) accessions on the basis of RAPD and AFLP markers. Crop Sci. 39:268–275.

Zar, J.H. 1984. Biostatistical analysis. Prentice-Hall International, NJ.

Evaluation of Genetic Diversity of Soybean Introductions and North American Ancestors Using RAPD and SSR Markers

G. L. Brown-Guedira,* J. A. Thompson, R. L. Nelson, and M. L. Warburton

ABSTRACT

The genetic base of soybean [*Glycine max* (L.) Merr.] cultivars developed for North America is very narrow. This may threaten the ability of breeders to sustain improvement and increase vulnerability of the crop to pests. The objective of this research was to assess the relationship of 18 major ancestors of North American soybean germplasm with 87 plant introductions (PIs) that are potential new sources of genetic variation for soybean breeding programs. Genetic distances (GD) among the 105 genotypes analyzed were calculated from 109 polymorphic DNA fragments amplified with random oligonucleotide primers and simple sequence repeat (SSR) primer pairs. Two hierarchical clustering algorithms were combined with data resampling and multidimensional scaling (MDS) to evaluate relationships among the genotypes. Genetic distances ranged from 0.08 to 0.76, with a mean of 0.52. Genotypes were placed in 11 clusters on the basis of a consensus of the different methods utilized. Co-occurrence values calculated from the resampling iterations showed that the stability of clusters varied. The most stable grouping was among ancestors that corresponded with known relationships based on pedigree and maturity. Several groups of PIs are distinct from the majority of the ancestors. These genotypes may be useful to breeders wanting to utilize genetically diverse introductions in soybean improvement.

Soybean breeding programs in the USA have successfully developed hundreds of improved cultivars through hybridization of elite cultivars and breeding lines that trace back to a small number of original plant introductions and selections. The narrowness of the North American soybean germplasm base has been well documented by pedigree analysis (Gizlice et al., 1994; Sneller, 1994). In an analysis of the pedigrees of 258 North American cultivars released between 1947 and 1988, Gizlice et al. (1994) determined that only 35 ancestors contributed more than 95% of all alleles. As few as five lines account for more than 55% of the genetic background of public cultivars in North America. An increase in the coefficient of parentage has been noted when ancestry of cultivars developed for the southern and northern growing regions of North America are examined separately, indicating an even greater restriction of the genetic base of cultivars within these regions (Gizlice et al., 1994; Sneller, 1994).

The limited germplasm base of North American soybean cultivars threatens the ability of breeders to sustain genetic improvement. It also increases vulnerability of the crop to changes in pathogen and pest populations. Introgression of new genetic diversity through hybridization with introduced germplasm is one way to increase genetic variation in breeding populations, the base upon which gain from selection depends. More than 15 000 introduced accessions of *G. max* are maintained in the USDA soybean germplasm collection, the vast majority of which do not appear in the pedigree of any released cultivar. These introductions potentially represent a rich source of allelic variation not present in current North American soybean cultivars. At present, the use of exotic germplasm in soybean cultivar development generally has been limited to a small number of introductions that have served as sources of genes for resistance to disease and insect pests and have contributed little to overall genetic diversity.

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 diversity as well as agronomic value. Agronomic performance of exotic germplasm in the target environment may be taken into account in parental selection; but, it is not known what effect this has on the probability of obtaining new allelic diversity. Geographic origin and plant morphology data are available for most of the introductions in the USDA collection and frequently serve as criteria for selection of genetically diverse parents. However, morphological

G.L. Brown-Guedira, USDA-ARS, Plant Science and Entomology Research Unit, Dep. of Agronomy, 2001 Throckmorton Plant Science Center, Kansas State Univ., Manhattan, KS 66506; J.A. Thompson, Pioneer Hi-Bred Intl., P.O. Box 328, Hamel, IL 62046; R.L. Nelson, USDA-ARS, Plant Physiology and Genetics Research Unit, Dep. of Crop Sciences, 1101W. Peabody Dr., Univ. of Illinois, Urbana, IL 61801; M.L. Warburton, CIMMYT Applied Biotechnology Center, Lisboa 27, Apdo Postal 6-641 06600 Mexico DF, Mexico. Joint contribution of the USDA-ARS, and Illinois Agric. Exp. Stn. Received 15 April 1999. *Corresponding author (gbg@ksu.edu).

Published in Crop Sci. 40:815–823 (2000).

Abbreviations: cM, centimorgans; GD, genetic distance; MG, maturity group; PI, Plant Introduction; PCR, polymerase chain reaction; RAPD, random amplified polymorphic DNA; SSR, simple sequence repeat; RFLP, restriction fragment length polymorphism; UPGMA, Unweighted Pair Group Method Using Arithmetic Averages.

differences usually are determined by a small number of genes and may not be representative of genetic divergence in the entire genome. Geographic origin also may not be an adequate indicator of genetic diversity. The original source of many soybean accessions introduced into the collection from secondary sources in Europe, Africa, and Asia are not known. In addition, a large number of the accessions in the USDA soybean collection are from the same regions of China and Korea as the introductions that make up the base of the North American germplasm. The genetic relationships of these materials with the major ancestors of modern U.S. cultivars are not known.

An understanding of the relationship of soybean introductions with the ancestors of North American cultivars based on selectively neutral DNA markers could be useful to breeding programs for selection of diverse parents. Using elite lines and culitvars, Kisha et al. (1997) demonstrated that greater diversity among parental lines as measured by restriction fragment length polymorphism (RFLP) makers produced greater variance for seed yield in the resulting populations. The availability of polymerase chain reaction (PCR) based molecular markers, such as randomly amplified polymorphic DNAs (RAPDs) and simple sequence repeats (SSRs), allows one to survey a large number of loci from many accessions. Thompson and Nelson (1998a) identified a set of 35 random primers that reliably produced RAPDs with gene heterozygosity scores of 0.30 or greater when a group of 35 genotypes were analyzed. Using 20 SSR loci, Diwan and Cregan (1997) observed a mean gene diversity of 0.80 in a survey of 35 major ancestors of North American soybean cultivars, much greater than that observed in similar studies using RFLP markers (Keim et al., 1992). Using the 20 SSR loci, they were able to distinguish several modern soybean cultivars considered identical on the basis of RFLPs, morphological, and pigmentation traits.

We have used 87 soybean introductions successfully in a our germplasm enhancement program to increase yield. The objective of this research was to evaluate the relationships of these introduced lines with the major ancestors of North American cultivars.

MATERIALS AND METHODS

Seventy soybean plant introductions that have been used as parents and appear in the pedigrees of high yielding breeding lines in the USDA-ARS soybean germplasm enhancement program at the University of Illinois, Urbana-Champaign, were selected for characterization in this study. Maturity group (MG) of the selected introductions ranged from 00 to IV. Whereas the majority of plant introductions were originally obtained from Asia (China, Korea, and Japan), lines were included in the study that came into the collection from sources in eastern Europe, France, and Morocco (Table 1). The materials include soybean landraces and newer lines from breeding programs in China and Japan. The plant introductions were selected as parents in the breeding program on the basis of agronomic performance, primarily seed yield and maturity, when grown at Urbana, IL. For comparison, the genotypes analyzed by Thompson et al. (1998) were also included. These included 18 soybean ancestors and first progeny, U.S. developed cultivars with uncertain pedigrees, and 17 additional plant introductions used as parents in our germplasm enhancement program. A total of 105 soybean genotypes were analyzed in the current research.

Genomic DNA was isolated from up to 10 greenhouse-grown seedlings of each plant introduction using either the rapid isolation protocol of Oard and Dronavalli (1992) or a modified CTAB (hexadecyltrimetyl ammonium bromide) extraction protocol. For RAPD analysis, DNA amplifications were carried out with 46 random decamer primers obtained from Operon (Operon Technologies, Alameda, CA) (Table 2). Included were 31 of 35 primers of a core set identified by Thompson and Nelson (1998b) that amplified highly polymorphic RAPD markers useful for diversity analysis in soybean. The PCR amplification reactions contained 1× PCR buffer, 2.0 m*M* $MgCl_2$, 200 μ*M* of each dNTP, 0.4 μ*M* 10-mer primer, 50 ng template DNA, and 1.25 units *Taq* DNA polymerase in 25 μL. Amplification protocol was as described by Kresovich et al. (1994). Controls run with each amplification included at least one genotype that had been previously amplified and/or one sample of reaction mix with no template DNA. The PCR products were separated on 1% (w/v) agarose gels in 1× TBE buffer and visualized by ethidium bromide staining.

DNA of all genotypes was amplified with SSR loci primer pairs, BARC-Satt 002, BARC- Satt 006, and BARC-Satt 080, hereafter referred to as Satt 002, Satt 006, and Satt 080, respectively (Akkaya et al., 1995; Research Genetics, Huntsville, AL). The PCR amplification reactions of SSR markers were similar to reactions for RAPD markers, except 0.15 μ*M* of 3′ and 5′ primers was used and reaction volumes were 10 μL. Cycling consisted of a 1 min denaturation at 94°C, 1 min annealing at 50°C and 2 min extension at 72°C for 35 cycles. The PCR products were denatured in formamide loading buffer for 5 min, then separated on a standard DNA sequencing gel containing 6% (w/v) polyacrylamide, 5.6 *M* urea, and 1× TBE for approximately 3 h at 30 W constant power. Bands were visualized with a Silver Sequence staining kit (Promega, Madison, WI).

Polymorphic DNA segments amplified with each random and microsatellite primer pair were assigned a letter and each band was scored as present (1) or absent (0). A matrix of all possible pair-wise GD was calculated by PROC IML in PC SAS (SAS Institute, 1989) on the basis of the following formula: $GD = 1 - a/(n - d)$, where a is the number of shared bands for each pair of genotypes (1,1), n is the number of possible matches (1,1 and 0,0) and mismatches (0,1 and 1,0) between genotypes, and d is the number of (0,0) matches between genotypes. The distance measure used is equivalent to the compliment of Jaccard's coefficient of similarity (Jaccard, 1908). The 0,0 matches were not treated as information for two reasons: (i) lack of a RAPD band in two genotypes may not be due to a common evolutionary event, and (ii) the presence of multiple alleles at microsatellite loci inflates genetic similarity if 0,0 matches are treated as information. Gene diversity scores were calculated for each marker as $1-\Sigma p_{ij}^2$ where p_{ij} is the frequency of the jth allele of Marker i (Weir, 1990). Each RAPD fragment was considered to have two forms (present or absent) whereas multiple alleles were detected at SSR loci.

Cluster analysis was performed on the 105 × 105 distance matrix using the AVERAGE and WARD options of PROC CLUSTER of PC SAS (SAS Institute, 1989). Mean distances within and between clusters were calculated from the genetic distance matrix. An acceptable cluster was defined as a group of two or more genotypes with a within-cluster genetic distance less than the overall mean genetic distance, and between clus-

Table 1. Soybean introductions and ancestors analysed, the assigned entry number, cluster assignment of each genotype using UPGMA and Ward's methods, average proportion of 100 resampling iterations utilizing UPGMA that a genotype was clustered with every other genotype in the cluster (mean co-occurence), the concensus cluster (C.C.) assignment, maturity group, country of origin, and mean genetic distance of genotype with all other genotypes.

Entry	Genotype	UPGMA	Ward's	Mean co-occurence	C.C.	MG	Origin	Mean GD
1	Korean†	A	A	1.00	A	II	North Korea	0.49
2	PI 189930	A	A	0.99	A	II	France	0.48
3	PI 227328	A	F	0.46	A	III	Japan	0.50
4	PI 227333	A	A	0.99	A	II	Japan	0.48
5	PI 253665D	A	A	0.56	A	III	China	0.48
6	PI 261474	A	A	0.99	A	II	China	0.48
7	PI 290126B	A	A	0.51	A	II	China	0.54
8	PI 297515	A	F	0.54	A	II	Hungary	0.49
9	PI 297544	A	A	0.99	A	II	Hungary	0.48
10	PI 347560	A	A	0.74	A	I	China	0.46
11	PI 370059	A	A	0.95	A	I	Soviet Union	0.45
12	PI 372415A	A	A	0.99	A	II	Korea	0.50
13	PI 393999	A	A	0.99	A	II	China	0.49
14	PI 404157	A	A	0.94	A	I	Soviet Union	0.49
15	PI 407710	A	A	0.88	A	I	China	0.47
16	PI 437909B	A	A	0.62	A	II	China	0.46
17	PI 506920	A	A	0.93	A	II	Japan	0.46
18	Mandarin (Ottawa)†	B	B	0.78	B	I	China	0.48
19	PI 84657	B	B	0.66	B	III	Korea	0.53
20	PI 189061-2	B	B	0.76	B	III	China	0.47
21	PI 189916	B	B	0.67	B	I	China	0.53
22	PI 290116A	B	B	0.80	B	O	Hungary	0.51
23	PI 317335	B	B	0.76	B	I	Japan	0.53
24	PI 437851A	B	B	0.54	B	I	China	0.47
25	PI 468377	B	B	0.70	B	OO	China	0.50
26	PI 507373	B	B	0.81	B	I	Japan	0.50
27	Dunfield†	C	C	0.58	C	III	China	0.48
28	Mukden†	C	C	0.52	C	II	China	0.50
29	PI 69507	C	C	0.61	C	I	China	0.52
30	PI 88310	C	C	0.68	C	III	China	0.49
31	PI 90566-1	C	C	0.59	C	III	China	0.48
32	PI 361075	C	C	0.64	C	I	China	0.50
33	PI 383276	C	C	0.68	C	I	China	0.46
34	PI 391583	C	C	0.55	C	II	China	0.52
35	PI 404161	C	C	0.51	C	IV	Soviet Union	0.53
36	PI 423950	C	C	0.42	C	II	Japan	0.51
37	PI 427099	C	C	0.62	C	I	China	0.47
38	PI 464878	C	C	0.65	C	II	China	0.49
39	PI 464920A	C	C	0.42	C	III	China	0.52
40	PI 476352C	C	C	0.44	C	II	Soviet Union	0.53
41	PI 491579	C	C	0.63	C	I	China	0.48
42	PI 503338	C	C	0.60	C	II	China	0.48
43	PI 506945	C	C	0.64	C	II	Japan	0.51
44	PI 507296	C	C	0.68	C	III	Japan	0.49
45	PI 507543	C	C	0.56	C	II	China	0.53
46	PI 68508	D	D	0.65	D	II	China	0.51
47	PI 68522	D	D	0.53	D	II	China	0.50
48	PI 68658	D	D	0.65	D	II	China	0.50
49	PI 361067	D	F	0.43	D	II	Yugoslavia	0.54
50	PI 378664A	D	F	0.53	D	I	Soviet Union	0.54
51	PI 384469A	D	D	0.51	D	I	Soviet Union	0.53
52	PI 384471	D	D	0.50	D	II	Soviet Union	0.57
53	PI 384474	D	D	0.51	D	II	Soviet Union	0.53
54	PI 391584	D	D	0.56	D	I	China	0.51
55	PI 391594	D	D	0.58	D	II	China	0.51
56	PI 436682	D	F	0.40	D	I	China	0.53
57	PI 436684	D	C	0.51	D	III	China	0.48
58	Pi 437697	D	F	0.46	D	II	China	0.54
59	PI 491548	D	D	0.66	D	II	China	0.53
60	Jackson†	F	F	0.89	E	VII	‡	0.55
61	Ogden†	F	F	0.85	E	VI	‡	0.56
62	Roanoke†	F	F	0.89	E	VII	China	0.51
63	Capital†	F	F	0.47	F	O	China	0.54
64	PI 291306A	F	F	0.69	F	II	China	0.56
65	PI 297500	F	F	0.70	F	I	Hungary	0.56

Continued.

ter distances greater than either within cluster distance of the two clusters involved.

To test the reliability of cluster assignments, resampling of the original data coupled with cluster analysis was performed using a SAS macro provided by D.Z. Skinner (1998, personal communication). The macro couples the IML and CLUSTER procedures of SAS to sample the data set, calculate a measure of genetic distance based on the sample, and perform cluster analysis. One hundred iterations of resampling and clustering were performed on the original data with the AVERAGE option of PROC CLUSTER. A 105 × 105 matrix was produced showing all possible pair-wise combinations of lines and

Table 1. Continued.

Entry	Genotype	UPGMA	Ward's	Mean co-occurence	C.C.	MG	Origin	Mean GD
66	PI 407654	F	F	0.65	F	I	China	0.58
67	PI 438205	F	F	0.42	F	I	China	0.53
68	PI 438206	F	F	0.41	F	I	China	0.55
69	PI 200485	H	F	0.68	F	III	Japan	0.53
70	PI 398763	H	F	0.56	F	III	South Korea	0.53
71	PI 399016	H	F	0.64	F	IV	South Korea	0.52
72	Arksoy†	G	G	0.77	G	VI	North Korea	0.56
73	Perry†	G	G	0.77	G	IV	‡	0.55
74	Ralsoy†	G	G	0.77	G	VI	North Korea	0.56
75	FC 04007B	H	G	0.63	H	III	Unknown	0.52
76	PI 467307	H	H	0.43	H	I	China	0.52
77	PI 507295	H	H	0.46	H	III	Japan	0.56
78	Haberlandt†	H	H	0.62	H	VI	North Korea	0.53
79	PI 91730-1	H	H	0.54	H	III	China	0.56
80	PI 189893	H	H	0.57	H	O	France	0.53
81	PI 404192C	H	H	0.48	H	I	China	0.56
82	PI 407720	H	H	0.43	H	II	China	0.53
83	PI 417510	H	H	0.59	H	I	Germany	0.53
84	Richland†	H	H	0.61	H	II	China	0.53
85	AK(Harrow)†	I	I	0.62	I	III	China	0.50
86	FC 31571	I	H	0.49	I	III	China	0.51
87	Illini†	I	I	0.62	I	III	China	0.50
88	Lincoln†	I	I	0.62	I	III	‡	0.49
89	PI 361064	I	I	0.52	I	II	Yugoslavia	0.53
90	PI 361066A	I	I	0.39	I	II	Yugoslavia	0.57
91	PI 404188A	I	I	0.60	I	II	China	0.53
92	PI 424159B	I	I	0.62	I	IV	South Korea	0.52
93	PI 445830	I	I	0.44	I	I	Romania	0.49
94	PI 445837	I	I	0.43	I	I	Romania	0.53
95	S-100†	I	I	0.45	I	V	China	0.54
96	PI 87588	J	J	0.59	J	IV	North Korea	0.58
97	PI 283331	J	J	0.57	J	III	Morocco	0.62
98	PI 464887	J	J	0.66	J	II	China	0.56
99	PI 475814	J	J	0.54	J	II	China	0.58
100	PI 427088B	K	J	0.64	K	I	China	0.58
101	PI 458511	K	J	0.64	K	II	China	0.60
102	CNS†	H	F	–	out§	VII	China	0.56
103	PI 437578	out	I	–	out§	III	China	0.56
104	PI 68600	out	B	–	out§	II	China	0.53
105	PI 91091	out	C	–	out§	II	China	0.56

† Major contributing ancestors and first progeny of modern North American soybean cultivars.
‡ First progeny are U.S.-developed cultivars with unknown parentage.
§ Outlier in the analysis.

Table 2. Random oligonucleotide primers used to amplify DNA from 105 soybean genotypes and the number of polymorphic fragments scored with each primer.

Primer	Sequence†	Fragments	Primer	Sequence	Fragments
OA20	GTTGCGATCC	2	OE01	CCCAAGGTCC	1
OF04	GGTGATCAGG	1	OF10	GGAAGCTTGG	1
OG06	GTGCCTAACC	6	OH02	TCGGACGTGA	1
OH12	ACGCGCATGT	2	OH13	GACGCCACAC	3
OH15	AATGGGGCAG	1	OK03	CCAGCTTAGG	4
OK14	CCCGCTACAC	3	OK16	GAGCGTCGAA	3
OL09	TGCGAGAGTC	2	OL13	ACCGCCTGCT	2
OL17	AGCCTGAGCC	1	OL18	ACCACCCACC	1
ON03	GGTACTCCCC	2	ON07	CAGCCCAGAG	5
ON08	ACCTCAGCTC	1	ON09	TGCCGGCTTG	2
ON11	TCGCCGCAAA	4	ON18	GGTGAGGTCA	1
ON19	GTCCGTACTG	1	OO01	GGCACGTAAG	3
OO02	ACGTAGCGTC	2	OO04	AAGTCCGCTC	2
OO05	CCCAGTCACT	2	OO10	TCAGAGCGCC	1
OO14	AGCATGGCTC	2	OO15	TGGCGTCCTT	1
OO16	TCGGCGGTTC	3	OO19	GGTGCACGTT	1
OO20	ACACACGCTG	1	OP07	GTCCATGCCA	1
OP09	GTGGTCCGCA	2	OR10	CCATTCCCCA	4
OR12	ACAGGTGCGT	4	OR13	GGACGACAAG	2
OS01	CTACTGCGCT	1	OS03	CAGAGGTCCC	2
OS05	TTTGGGGCCT	1	OS07	TCCGATGCTG	2
OS09	TCCTGGTCCC	1	OS11	AGTCGGGTGG	1
OS13	GTCGTTCCTG	2	OS14	AAAGGGGTCC	4

† Sequence of each decamer primer reads from 5′ to 3′.

the number of iterations for which any two lines were placed in the same cluster, which is referred to as the co-occurrence value (Vera Cruz et al., 1996). These data and the cluster assignments in the original analysis were used to determine the number of iterations that a line was grouped with all other members of a cluster. Cluster assignment by the AVERAGE and WARD options of PROC CLUSTER and average number of iterations a genotype was clustered with other members were used to place lines in a consensus cluster.

The distance matrix was subjected to MDS (Shepard, 1974) by the MDS procedure of PC SAS (SAS Institute, 1992) and criteria similar to that described by Gizlice et al. (1996) and Thompson et al. (1998). To maintain the scale of 0 to 1 used for the genetic distances, the ABSOLUTE option was used. The goodness of fit criteria (R^2) between the original data and the predicted values that were derived from the MDS coordinates was used to evaluate the effectiveness of 2 to 20 dimensions. The most effective analysis was defined as the fewest dimensions resulting in an $R^2 > 0.95$ with the original genetic distance matrix.

RESULTS AND DISCUSSION

Each of the 46 random oligonucleotide primers detected DNA polymorphism, resulting in 94 polymorphic fragments. Diversity scores of the RAPD fragments ranged from 0.02 to 0.50 with a mean of 0.35. The diversity scores of the three SSR primer pairs were greater than that observed for RAPDs because of the presence of multiple alleles. Satt 002, Satt 006, and Satt 080 detected four, five, and six alleles, respectively, and had diversity scores of 0.41, 0.62, and 0.73. The mean GD of all pairwise comparisons of genotypes based on Jaccard's coefficient was 0.52.

Relationship of Ancestors and Plant Introductions

The UGPMA and Ward's methods both assigned genotypes into nine clusters (Table 1) that were judged reasonable on the basis of the within and between cluster distances. The cluster assignments were similar for the two algorithms although some rearrangement occurred. Thompson et al. (1998) identified 10 major clusters among 32 genotypes using the RAPD marker data included in this study. The addition of microsatellite marker data and 70 additional plant introductions resulted in some changes in the diversity patterns observed in the earlier study. Most notably, clusters were larger, and less distinction was observed between groups.

Members of Cluster A were grouped together in the majority of resampling iterations (Table 1). This cluster consisted of the ancestral line 'Korean' and 16 plant introductions originating from France, Japan, northern China, Hungary, and the far eastern provinces of the former USSR (Table 1). Korean contributed less than 1% to the genetic base of northern cultivars and was not involved in the development of any cultivars in the southern growing region, so this cluster has little relationship to current U.S. varieties. Although members of the cluster were of very diverse origin, it was one of the most consistent clusters. Ten of the 17 members were placed in this cluster in more than 90% of the iterations. Cluster A also had the smallest within cluster mean GD of 0.32.

The ancestor 'Mandarin (Ottawa)', which contributed over 17% to the genetic base of northern U.S. cultivars, was clustered with eight introductions from northern China, Korea, Japan, and Hungary in 85% of clustering iterations to form Cluster B (Table 1). Thompson et al. (1998) put Mandarin (Ottawa), 'Richland' and 'Capital' in the same cluster but that cluster was the least stable of any of their ancestral clusters. The two hierarchical clustering procedures in this analysis split these three ancestors into three different clusters.

The ancestors 'Dunfield' and 'Mukden', each of which account for about 3.5% of the genetic base of North American cultivars, were included in Cluster C, the largest cluster identified. Thompson et al. (1998) found that Dunfield did not fit well into any single cluster. In our study, 16 of the nineteen members of Cluster C, including Mukden and Dunfield, were grouped together in a majority of resampling iterations (Table 1). This cluster contained seven cultivars released in China in Liaoning and Jilin provinces between 1963 and 1983. Mukden and Dunfield originated from Liaoning and Jilin, respectively. Two of the Chinese lines were received with the name Jilin No. 3, one from Romania in 1971 (PI 361075) and one from Jilin province in 1978 (PI 427099). The physical appearance of these lines had been compared in the field and although the major qualitative descriptors were identical, the plant types were judged to be different enough to warrant maintaining both lines in the collection. This decision was supported by our DNA data which showed that these two lines differed at 16% of the loci compared. Pedigree information relates some of the cultivars in this cluster. Jilin No. 6 (PI 383276) shares a common parent with Jilin No. 3 and was most closely related to the two accessions named Jilin No. 3, PI 361075 and PI 427099, with GD of 0.24 and 0.20, respectively.

Cluster D was mostly MG I and II accessions developed in the former Soviet Union or originating from China. It includes three introductions (PI 68508, PI 68522, and PI 68658) that were imported into the USA at about the same time and from approximately the same region of China as many of the major ancestors of current N.A. cultivars. Also in this group were four cultivars released in Jilin and Liaoning provinces between 1970 and 1978. Some of these Chinese cultivars are very closely related to the Chinese cultivars in Cluster C. Jilin 10 in Cluster C has the same parents as Jilin 11 (PI 391584) in Cluster D. Jilin No. 6 in Cluster C has one of the same parents as Jilin No. 8 (PI 391594) in Cluster D. Jilin No. 3 in Cluster C is a parent of Jilin No. 10 (PI 391584) in Cluster D. The genetic distance between Clusters C and D was 0.50, less than the distance between most other clusters.

The MG VI and VII lines 'Jackson', 'Roanoke', and 'Ogden', which together account for 24% of the genetic base of cultivars developed in the southern USA, were placed in the same cluster in more than 78% of iterations and also formed one of the major ancestral clusters in the analysis of Thompson et al. (1998). The UPGMA

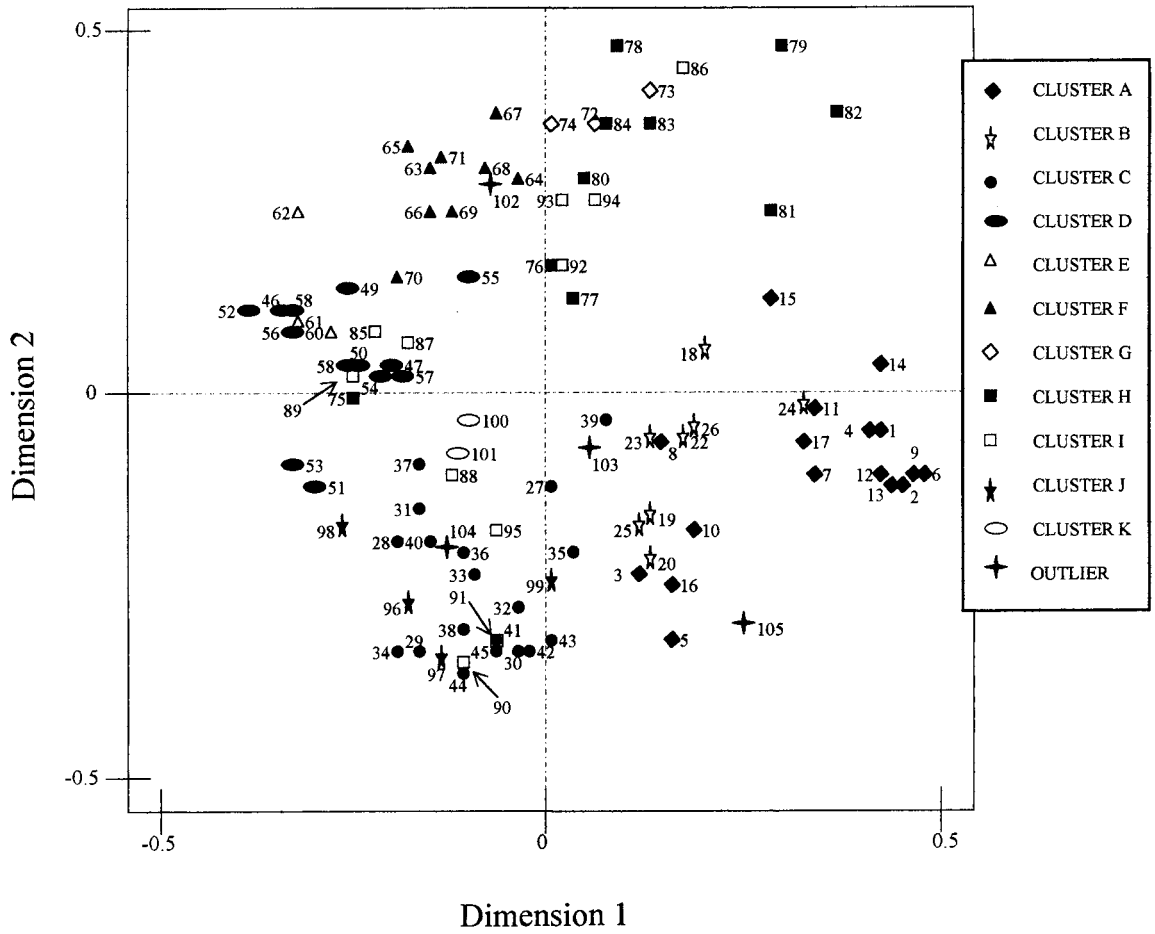

Fig. 1. Two-dimension multidimensional scaling scatter plot depicting patterns of diversity among the 18 major ancestors of North American cultivars and 88 soybean introductions. Diversity estimates are based on 109 polymorphic fragments amplified by RAPD and SSR primers. Co-ordinates are labeled with entry numbers and cluster assignments correspond to consensus clusters.

and Ward's methods clustered these genotypes in Cluster F with five and 14 introductions, respectively, and the ancestor Capital. The cluster classification of Capital was difficult to assess since it was not placed in any one cluster in greater than 50% of resampling iterations. On the basis of the mean genetic distance between Capital and PIs 291306A, 297500, and 407654 (entries 64, 65, and 66, respectively) of 0.48 and proximity of Capital to these introductions on the scatter plot of the first two dimensions of MDS (Fig. 1), Capital (entry 63) was placed in consensus Cluster F.

Whereas the two clustering algorithms placed the lines Jackson (entry 60), Ogden (entry 61), and Roanoke (entry 62) in Cluster F, the two dimensional MDS plot clearly separated these accessions from the other genotypes in the cluster (Fig. 1). The inclusion of these three genotypes in the cluster inflated the within cluster mean GD of Cluster F and resulted in low co-occurrence values for the cluster. Jackson, Ogden, and Roanoke were therefore placed in a separate cluster labeled E. The within cluster mean GD of Clusters E and the new Cluster F were 0.32 and 0.44, respectively.

The ancestors 'Arksoy' (entry 72) and 'Ralsoy' (entry 74) were always placed in the same cluster. This is consistent with the results of Thompson et al. (1998) but their analysis added 'Haberlandt' to this group. These two closely related MG VI lines were clustered with the MG IV line 'Perry' (entry 73), defined as a first progeny by Gizlice et al. (1994), in 77% of iterations to form Cluster G. Some overlap was observed in the MDS scatter plot between this group of genotypes and Cluster H (Fig. 2) which was defined by the ancestors Haberlandt (entry 78), and Richland (entry 84), and introductions PI 91730-1, PI 189893, PI 417510, and FC 04007B (entries 79, 80, 83, and 75, respectively) that were clustered together in approximately 75% of resampling iterations. Four other soybean introductions were included in Cluster H that were not stable members of the cluster (Table 1).

Cluster I was defined by the major ancestral lines 'A.K. (Harrow)', 'Illini', and 'Lincoln', which were clustered together in more than 87% of resampling iterations and were placed with all other members of the cluster in 62% of the iterations (Table 1). This relationship was observed by Thompson et al. (1998) who noted that the close relationship of Lincoln to A.K.(Harrow) and Illini may be cause for concern since the three genotypes together account for 34% of the genetic base of cultivars adapted to the northern growing region of North America. The ancestor S-100 was placed in Clus-

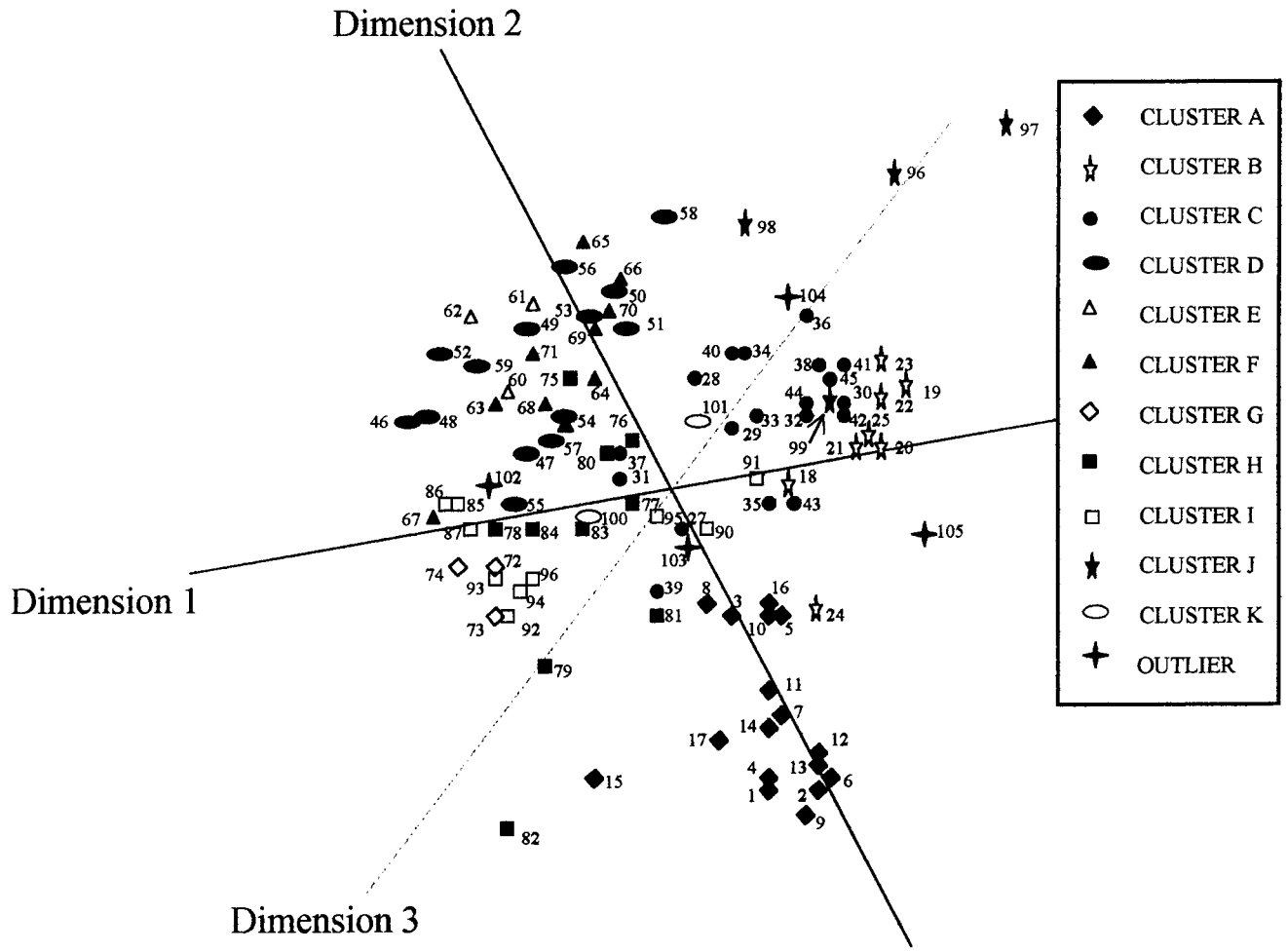

Fig. 2. Three-dimension multidimensional scaling scatter plot depicting patterns of diversity among the 18 major ancestors of North American cultivars and 88 soybean introductions. Diversity estimates are based on 109 polymorphic fragments amplified by RAPD and SSR primers. Co-ordinates are labeled with entry numbers and cluster assignments correspond to consensus clusters.

ter I in 45 iterations and in Cluster H in 37 iterations. S-100 was most closely related to Illini and A.K. (Harrow), with genetic distances of 0.33 and 0.34, respectively. S-100, which accounts for 21% of the genetic base of cultivars developed for the southern USA, is reported to be a Maturity Group V selection from the MG III line Illini. Thompson et al. (1998) suggested that S-100 may be a progeny of rather than a selection from Illini. Lorenzen and Shoemaker (1996) found differences between S-100 and Illini at 7 of 48 loci surveyed with RFLPs. The four ancestral lines in Cluster I have contributed nearly one-third of the genes to current North American cultivars.

Clusters J and K were comprised entirely of plant introductions. Three of the most diverse genotypes in the study, PI 283331 (from Morroco through Australia in 1962), PI 87588 (from Korea in 1930), and PI 475814 (from Xinjiang, China in 1982) were placed in Cluster J in a majority of iterations. There is little soybean production in Xinjiang and the history of soybean in this province is much shorter than is most places in China. The soybean cultivars grown in Xinjaing probably originated in Northeast China. The fourth member of this cluster (PI 464887) is from Jilin province. All of the accessions in Cluster J may have originally come from Northeast China but they are genetically distinct from the other accessions from the region that were included in this research. PI 458511, which is Kai yu No. 3 released in Liaoning province in 1976 (Cui et al., 1999), had the second largest mean genetic distance from all other genotypes. It was grouped with another diverse introduction, PI 427088B, to form Cluster K. PI 427088B is an unknown Chinese cultivar obtained from a soybean crushing plant in Jilin province in 1978 (Bernard et al., 1989). Clusters J and K had the greatest mean GD with all other clusters.

Kisha et al. (1998) analyzed 9 ancestors and 12 introductions that are also a part of this study. Because of the few lines in common between the two studies it is difficult to make meaningful comparisons. We analyzed 10 of the 15 members of the Cluster 4 defined by Kisha et al. (1998). Our results placed those 10 accessions in three different clusters. We also had five members of the Cluster 6 defined by Kisha et al. (1998) and we agreed with the grouping of A.K. (Harrow), Lincoln, PI 445830, and S-100 but our results put Richland in a different cluster. Cluster formation is a comparative process so changing some of the lines in the analysis

can change cluster members and Kisha et al. (1998) used only 65 RFLP alleles to classify 165 entries

Genetic Diversity in Exotic Soybean Germplasm

No discernable geographical patterns of variation were found in this research. However, lines included in the study were not selected to accurately represent the major regions of soybean production and many of the entries were products of modern breeding programs. Griffin and Palmer (1995) assumed that the long history of domestication and commerce of soybean in Asia has contributed to the dispersion of alleles throughout the region, lessening the influence of geography on patterns of variation among Asian soybean accessions. Nelson and Li (1998) did not find this to be true when comparing primitive soybean accessions. A unique allele was detected in the ancestral line CNS at the Satt 080 locus. Another unique allele was present at the Satt 006 locus in the genotype PI 87588. Satt 002 detected a rare allele in Arksoy, Ralsoy, and PI 399016. All of the lines for which unique alleles were detected originated from the Korean peninsula, except for CNS.

The level of genetic diversity observed in this study was higher than that reported in studies using RFLPs. This may be due to the selection of RAPD markers known to have high levels of gene diversity in soybeans and the inclusion of a small number of SSR loci that are more highly polymorphic than RFLPs. The analysis was able to identify related groups of genotypes and many of the soybean plant introductions included in the study were found to be genetically distinct from the founding stock of North American soybean breeding programs. For example, Clusters A and F were composed mainly of plant introductions, including only two ancestors, Korean and Capital, that have made only small contributions to the pedigrees of soybean cultivars in the northern USA. Clusters D, J, and K consisted exclusively of introduced germplasm. These results identify three new genetic groups (Clusters F, J, and K) in addition to those already found by Thompson et al. (1998) that are distinct from the ancestral base of U.S. soybean cultivars.

This study will aid breeders interested in selecting genetically diverse germplasm, both in relation to the ancestors of North American cultivars and to other plant introductions. All of the introductions analyzed have successfully contributed to a breeding program focused on increasing yield and genetic diversity. The fact that many of the genetically diverse plant introductions in this study were used to produce lines having from 25% to 100% plant introduction parentage and yields comparable to contemporary cultivars of similar adaptation (Thompson and Nelson, 1998b) illustrates the potential useful of these lines.

We have made the assumption that lines that have yields similar to commercial cultivars and that have a parent or parents that are genetically distinct from those same cultivars are good candidates for inclusion in a breeding program to expand genetic diversity and increase yield. We do not know if the plant introductions in this research have unique alleles for traits of economic importance. None of the RAPD markers used in the study have been mapped. Satt 002 is on linkage group D2 (Cregan et al., 1999) and is within 10 centimorgans (cM) of an unknown malate dehydrogenase locus (Palmer et al., 1992). Satt 006 is on Linkage Group L (Cregan et al., 1999) less than 3 cM from the *Dt1* locus that controls stem termination. All genetic groups had members with both determinate and indeterminate phenotypes except for the two member K cluster that was all indeterminate and the three member E cluster that was all determinate ancestral lines. Satt 080 is on Linkage Group N (Cregan et al., 1999) and close to the *Rps1* locus. It is unlikely that any of these known linkages influenced the cluster analyses of this research. At some point in the future, it is likely that we will be able be identify the exact loci and the unique alleles that can increase yield. Until we do and to help in that process, we think that identifying genetically distinct introductions that have the potential to produce high yielding lines is a productive strategy. Thompson and Nelson (1998b) demonstrated that these diverse soybean introductions can contribute genes that can increase the yield of U.S. cultivars. The literature shows that there is a relationship between marker diversity of the parents and genetic variance of the resulting progeny. Collecting data on genetic diversity in parents and progeny is time consuming and expensive. By identifying genetically diverse introductions that have the potential to produce high yielding progeny, we are making available to breeders and geneticists important germplasm resources that have a high potential to contribute not only to increasing yield but also to the process of understanding the genetic basis of yield improvement.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. D.Z. Skinner for use of the program for resampling, cluster analysis, and calculation of co-occurrence values. This research was supported by a grant from the Illinois Soybean Program Operating Board.

REFERENCES

Akkaya, M.S., R.C. Shoemaker, J.E. Specht, A.A. Bhagwat, and P.B. Cregan. 1995. Integration of simple sequence repeat (SSR) DNA markers into soybean linkage map. Crop Sci. 35:1439–1445.

Bernard, R.L., G.A. Juvik, and R.L. Nelson. 1989. USDA soybean germplasm collection inventory. Vol. 2. International agricultural publications. INTSOY Ser. no. 31. Univ. of Illinois at Urbana-Champaign.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464–1490.

Cui, Z., T.E. Carter, Jr., J. Gai, J. Qui, and R.L. Nelson. 1999. Origin, description, and pedigree of Chinese soybean cultivars released from 1923 to 1995. U.S. Dep. of Agriculture, Agricultural Research Service, Tech. Bull. No.1871. U.S. Gov. Print. Office, Washington, DC.

Diwan, N., and P.B. Cregan. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. 95:723–733.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1993. Genetic diversity in North American soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. Crop Sci. 33:614–620.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34:1143–1151.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1996. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. Crop Sci. 36:753–756.

Griffin, J.D., and R.G. Palmer. 1995. Variability of thirteen isozyme loci in the USDA soybean germplasm collections. Crop Sci. 35: 897–904.

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaudoise Sci. Nat. 44:223–270.

Johnson, R.A., and D.W. Wichern. 1992. Applied multivariate statistical analysis. Prentice-Hall, Englewood Cliffs, NJ.

Keim, P., W. Beavis, J. Schupp, and R. Freestone. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. Theor. Appl. Genet. 85:205–212.

Kisha, T.J., B.W. Diers, J.M. Hoyt, and C.H. Sneller. 1998. Genetic diversity among soybean plant introductions and North American germplasm. Crop Sci. 38:1669–1680.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317–1325.

Kresovich, S., W.F. Lamboy, R. Li, J. Ren, A.K. Szewc-McFadden, and S.M. Bliek. 1994. Application of molecular methods and statistical analysis for discrimination of accessions and clones of vetiver grass. Crop Sci. 34:805–809.

Lorenzen, L.L., S. Boutin, N. Young, J.E. Specht, and R.C. Shoemaker. 1995. Soybean pedigree analysis using map-based markers: I. Tracking RFLP markers in cultivars. Crop Sci. 35:1326–1336.

Lorenzen, L.L., and R.C. Shoemaker. 1996. Genetic relationship within old U.S. soybean cultivar groups. Crop Sci. 36:743–752.

Nelson, R.L., and Z. Li. 1998. RAPD marker diversity among soybean and wild soybean accessions from four chinese provinces. p. 164. *In* Agronomy abstracts. ASA, Madison, WI.

Oard, J.D., and S. Dronavilli. 1992. Rapid isolation of rice and maize DNA for analysis by random-primer PCR. Plant Mol. Bio. Reporter 10:236–241.

Palmer, R.G., Sung M. Lim, and Bradley R Hedges. 1992. Testing for linkage between the *Rxp* locus and nine isozyme loci in soybean. Crop Sci. 32:681–683.

SAS. 1989. SAS/STAT user's guide, Version 6, Fourth ed. SAS Institute, Inc., Cary, NC.

SAS. 1992. Technical report P-229. SAS Institute, Cary, NC.

Shepard, R.N. 1974. Representation of structure in similarity data: Problems and prospects. Psycometrika 39:373–421.

Sneller, C.H. 1994. Pedigree analysis of elite soybean lines. Crop Sci. 34:1515–1522.

Thompson, J.A., and R.L. Nelson. 1998a. Core set of primers to evaluate genetic diversity in soybean. Crop Sci. 38:1356–1362.

Thompson, J.A., and R.L. Nelson. 1998b. Utilization of diverse germplasm for soybean yield improvement. Crop Sci. 38:1362–1368.

Thompson, J.A., R.L. Nelson, and L.O. Vodkin. 1998. Identification of diverse soybean germplasm using RAPD markers. Crop Sci. 38:1348–1355.

Vera Cruz, C.M., E.Y. Ardales, D.Z. Skinner, J. Talag, R.J. Nelson, F.J. Louws, H. Leung, T.W. Mew, and J.E. Leach. 1996. Measurement of haplotypic variation in Xanthomonas oryzea pv. oryzae within a single field by rep-PCR and RFLP analyses. Phytopathology 86:1352–1359.

Weir, B. 1990. Genetic data analysis: Methods for discrete population genetic data. Sinauer Assoc., Sunderland, MA.