This is to certify that the

dissertation entitled
GENETIC DIVERSITY FOR RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP) MARKERS WITHIN SOYBEAN (GLYCINE MAX L. MERR.) GERM PLASM AND ITS USE AS A SELECTION CRITERION FOR PARENTS IN A BREEDING PROGRAM.

presented by

Theodore J. Kisha

has been accepted towards fulfillment
of the requirements for

Doctoral degree in Plant Breeding & Genetics - Crop & Soil Sciences

Major professor

Date August 21, 1996

GENETIC DIVERSITY FOR RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP) MARKERS WITHIN SOYBEAN (*GLYCINE MAX* L. MERR.) GERM PLASM AND ITS USE AS A SELECTION CRITERION FOR PARENTS IN A BREEDING PROGRAM.

By

Theodore James Kisha

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Crop and Soil Sciences

1996

ABSTRACT

GENETIC DIVERSITY FOR RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP) MARKERS WITHIN SOYBEAN (*GLYCINE MAX* L. MERR.) GERM PLASM AND ITS USE AS A SELECTION CRITERION FOR PARENTS IN A BREEDING PROGRAM.

By

Theodore James Kisha

Genetic diversity is limited in soybean in the US because only a few early plant introductions formed the original breeding pool. This study examined RFLP markers among samples of ancestral plant introductions, more recent plant introductions, and cultivars and elite lines from the northern US. Markers uniquely identified all lines examined. Cluster analysis grouped ancestors according to area of origin, while other lines formed groups in agreement with their pedigrees. Genetic distances among lines determined with RFLP, Random amplified polymorphic DNA (RAPD), and coefficient of parentage data were compared. Correlations between genetic distance and genetic variance of several agronomic traits were examined in two population sets over two years. Distance measures were generally positively correlated with genetic variances. There was a negative correlation with yield variance in one population set in one year. A multiple regression model using mid-parent yield and marker genetic distance predicted the highest yielding progeny. The relationship to mid-parent yield was always positive, but highest yielding progeny were negatively associated with genetic distance for one population set. The data herein suggest that using RFLP distance estimates for parent selection can increase the probability of producing transgressive segregates for yield.

This work is dedicated to the memory of my father

George Kisha

ACKNOWLEDGEMENTS

I would like to express my appreciation to Dr. Brian Diers, whose guidance and friendship made my education a truly enjoyable experience. I would also like to acknowledge the assistance of my graduate committee members; Dr. Jim Kelly, Dr. Jim Hancock, and Dr. Mike Thomashow. Their guidance and discussion has proven invaluable to my education. I would also like to express my appreciation for the support given by friends and colleagues, especially Dr. Bob Olien, during some difficult moments. Finally, and above all, I would like to thank my wife, Linda, whose strength and love during these trying years has been phenomenal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

The genetic distance between individuals is a quantitative estimate of the difference of their genetic makeup. Genetic distance can be measured in terms of probability, using coefficients of parentage (CP) where pedigrees are known (Falconer, 1989), indirectly by measuring differences in expressed genetic traits, more directly by measuring differences in gene products such as isozymes, or directly by analysis of DNA. Indirect measurements may be qualitative, such as flower color, hairy versus glabrous stems, hilum color; or quantitative, such as differences in plant height, leaf size, and days to maturity. Since any distance measurement must be related to differences in genes, no characters should be used which are not a reflection of differences in genes. Sneath and Sokal (1973) list as inadmissable characters that are environmentally determined and characters that are to any degree correlated. The former are not related to the genetic makeup and the latter bias the distance by summing multiple measurements on the same character.

Genetic distance based on quantitative characters can be expressed geometrically in *n* dimensions of Euclidean hyperspace, where *n* is the number of characters measured. The Euclidean distance between individuals (Sneath and Sokal, 1973) is given as:

$$d_{jk} = [\Delta_{jk}/n]$$

where:

$$\Delta_{jk} = [\Sigma_{i=1..n}(X_{ij}-X_{ik})^2]^{1/2}$$

The *n* characters are assumed independent and normally distributed and are standardized by giving them a mean of zero and a variance of unity. Equal weighting of characters may introduce an indeterminable amount of error when characters are a result of different numbers of segregating genes. Error will also occur if different combinations of genes result in the same phenotypic effect. When correlations exist among a set of *n* characters, distance can be expressed as a function of a subset of $m < n$ principal components (Sneath and Sokal, 1973). Euclidean distances can then be calculated on the basis of *m* orthogonal axes in hyperspace.

Relationships among individuals can also be based on the correlation of standardized quantitative characters between two individuals (Sneath and Sokal, 1973). The distance is given by the compliment of the correlation coefficient (1 - r).

Genetic distance based on qualitative characters begins with expressing the data in the form of an association coefficient (Sneath and Sokal, 1973), which is a measure of character matches relative to the number of possible matches. These pair-wise comparisons take the form of a 2 X 2 matrix for each line in the overall *n X m* data matrix in which *n* individuals are compared over *m* possible character states:

		Individual *j*	
		1	0
Individual *k*	1	a	b
	0	c	d

The row or column corresponding to the number 1 indicates a character is present, while the row or column corresponding to the number 0 indicates a character is absent. In the case of a two-state character, a match may be defined by either *a* or *d* and a mismatch by either *b* or *c*, but for multi-state characters, *d* provides no useful information, since it gives no indication whether the individuals are similar or different for the other character states. In this case, an association coefficient which ignores *d* would be appropriate. The coefficient of Jaccard (Sneath and Sokal, 1973; Rohlf, 1992), for example, does not consider matches based on mutual lack of a trait (*d*). Similarity is based on $a/(a+b+c)$, and distance is determined by the compliment of similarity. Sneath and Sokal (1973), as well as Rohlf (1992), provide lists of a number of association coefficients which differ in the way the results *a*,*b*,*c*, *and* *d* are handled.

Plant breeders have used some of the distance measures defined above in an attempt to predict the outcome of matings. Generally, the goal is to predict which crosses will have the greatest genetic variance of progeny, or the highest performing transgressive segregants or

hybrids. Cowen and Frey (1987a) examined the relationship of genealogical distance between parents with progeny performance in oat (*Avena sativa* L.) using a diallel mating design without recipricals. They evaluated progeny populations for generalized genetic variance and transgressive segregation for bundle weight, grain yield, straw yield, harvest index, height, and heading date. The generalized genetic variance (Goodman, 1968) was calculated from the genetic variance-covariance matrices of mean squares and cross products for genotype and genotype X location interaction for bundle weight, grain yield, and harvest index. Significant positive correlations were found for genealogical distance with generalized genetic variances and with transgressive segregates for height.

The same populations were later used to examine the relationships between several other distance measurements and progeny performance (Cowen and Frey, 1987b). A Euclidean distance was calculated using the first five principal components based on the correlation matrix of 12 quantitative agronomic traits measured for the nine parents. This distance proved to be negatively correlated with both transgressive segregation and generalized genetic variances. The second distance Cowen and Frey used was calculated from the 9 X 9 matrix of parental and population mating means for grain yield. This distance is based on the assumption that heterotic effects are proportional to diversity (Hanson and Casas, 1968). These distances were positively correlated with transgressive segregation in one year and with generalized genetic variance in both years. The third distance measure used by Cowen and Frey was calculated using the correlation of general combining ability (GCA) effects (Cervantes et al., 1978) over all the traits measured. The

distance was taken as 1-r. This distance measure was positively correlated with mid-parent heterosis in one year.

Souza and Sorrells (1991a) used the first six principal components from the correlation of 13 quantitative traits and the covariance of 15 discrete qualitative traits (1991b) to estimate genetic distance among oat genotypes. They found that classification using quantitative traits was according to area of adaptation. This method of classification did not agree well with that taken by coefficient of parentage, while classification using qualitative traits clustered lines according to common ancestors in their pedigrees. Genetic distances between parents based on either quantitative or qualitative traits were poor predictors of progeny genetic variance (Souza and Sorrells, 1991c). Only distance based on coefficient of parentage was significantly related to genetic variance and, for all agronomic traits measured except biomass, this relationship was negative.

A common factor in the estimations of genetic distance in the above examples is the complexity of their calculation and the time and effort required to collect the necessary data. A method of estimating genetic distance which precludes the *a priori* knowledge of the genetic effects of the parents is the ideal goal. Molecular markers can provide these estimates.

Isozymes (Lewontin and Hubby, 1966) are molecular markers whose variants can be used in a qualitative estimation of genetic distance. Non-denatured proteins are separated by electrophoresis and visualized using specific staining techniques. Isozyme variants are theoretically catagorized on the basis of charge:mass ratio, so their detection relies heavily on amino acid substitutions which result in net loss or gain in

charge. Substitutions not resulting in charge differences or large changes in mass should be more difficult to detect. Ramshaw et al. (1979) found cryptic differences within electrophoretic variants of isozymes of hemoglobin. Twenty known variants were separated into only eight electromorphs under "standard" conditions of pH 8.9 and 4.5% acrylimide. Further manipulations of pH, acrylimide concentration, and increased running time eventually were able to discriminate 17 classes for an efficiency of 85%. Chemically similar substitutions in different parts of the protein were discriminated 77% of the time under standard conditions. Further manipulation increased this efficiency to 90%. Four out of five chemically different but charge-equivalent substitutions at the same location on the protein were distinguishable under standard conditions, but one was not distinguishable under any of the conditions used. These results show that while several stringent analyses may separate isozymes with an acceptable degree of reliability, results using only a single protocol may lead to errors due to isozymes scored as identical but which are merely alike in state.

Cox et al. (1985) found significant correlations between genealogical distance and isozyme distance using 11 enzymes in groups of soybean (*Glycine max* L. Merr.). The correlation was higher for groups with lower mean genealogical distance. Lamkey et al. (1987) estimated genetic distance among 35 maize lines using isozyme differences at 9 loci. They found that isozyme genetic distance between parents was unable to predict hybrid performance. Damerval et al. (1987) tested the hypothesis that quantitative differences in gene products could be more important sources of genetic variability in maize than qualitative differences based on the presence or absence of a particular gene

product variant. They found that quantitative differences in enzymes in maize were more related to Mahalanobis distances (Mahalanobis, 1936) than were qualitative differences. The Mahalanobis distances were calculated on the basis of general combining ability for 14 heritable quantitative characters. This suggested that regulatory processes may play an important role in genetic diversity. If this is the case, direct qualitative analysis of differences in DNA sequences would provide more useful information than qualitative analysis of gene products, because differences in regulatory regions of DNA would be randomly sampled along with differences in coding regions. Direct analysis of DNA increases the extent of genome sampling by including introns and flanking sequences which may include promoters or enhancers. Additionally, direct DNA analysis, compared with isozyme analysis, does not rely on changes solely within coding regions which result in amino acid substitutions.

Differences between individuals at the DNA level can be estimated using restriction fragment length polymorphism (RFLP) (Southern, 1975). Genomic DNA is digested with a restriction enzyme, separated by size on an agarose gel, denatured, and transferred to a nylon membrane. The DNA on the membrane can then be probed with a radioactively labelled (Feinberg and Vogelstein, 1984) DNA clone, and the fragment to which the probe hybridizes visualized on x-ray film. Qualitative differences in RFLP banding patterns coded as one of a number of available association coefficients (Sneath and Sokal, 1973; Rohlf, 1992), as discussed earlier, are used to calculate genetic distance.

Genetic distances estimated using RFLP markers may be subject to error. Size differences of the genomic DNA to which the clone hybridizes may be due to point mutations which either eliminate or create new

restriction sites, or DNA rearrangements (Borst and Greaves, 1987). These rearrangements may be inversions, deletions or insertions. Polymorphism that arises from DNA rearrangement is a macromolecular difference which may be superimposed over micromolecular differences. Roth et al. (1989) propose that genetic variation may be generated within inbreeding plants by rearrangements due to specific recombinational processes in response to stress. They found that tissue culture of soybean root resulted in changes in RFLP markers arising from DNA rearrangement. Genetic alterations in plants regenerated from tissue culture is well documented (Mein, 1983; Evans et al., 1984). The surprising aspect of the results of this work was that the rearrangements resulted in previously characterized RFLP fragments. The majority of RFLP alleles characterized in soybean are dimorphic (Keim et al., 1989; Keim et al., 1992;) and are due to rearrangements of DNA (Apuya et al., 1988). Instead of generating unique alleles, the rearrangements which occurred during tissue culture resulted in conversion from one allele to the other previously characterized allele. Such rearrangements arising in whole plants would result in errors in genetic distance estimates if alleles alike in state are assumed to be identical by descent.

In general, RFLP's have proven superior to isozymes for the estimation of genetic diversity. McGrath and Quiros detected nearly three times the number of alleles at RFLP loci than at isozyme loci in *Brassica campestris* L. (syn. *B. rapa* Metz.). Messmer et al. (1991) detected polymorphism at 94% of RFLP loci examined compared with 68% of isozyme loci. The maximum number of isozyme alleles at a given locus was three compared with a maximum of eight alleles at a given RFLP locus.

The level of RFLP diversity was also twice that for isozyme diversity of common bean (*Phaseolus valgaris* L.) (Velasquez and Gepts, 1994).

Genetic distance estimated using RFLP data has been tested extensively as a predictor of progeny performance in maize. Smith et al. (1990) showed a close relationship between hybrid performance and RFLP distance of parents in maize using parents representing a wide range of related and unrelated elite corn belt germ plasm. Lee at al. (1989) found significant correlations of RFLP distance with both hybrid grain yield (r = .46) and specific combining ability (SCA) (r = .74) in maize (Zea mays L.). Godshalk et al. (1990), however, found no such relationship. Whereas Lee's group tested crosses both within and among heterotic groups, Godshalk's group selected crosses which minimized matings within heterotic groups. Melchinger et al. (1990) found only moderate relationships between RFLP distance and hybrid grain yield (r = .32) and SCA (r = .39). They concluded that RFLP's have only limited use in predicting progeny performance in maize, especially among unrelated lines.

Genealogical distance was significantly correlated with RFLP distance in oat (*Avena sativa* L.), but not with a distance calculated using the first five principal components of the parental correlation matrix for 12 agronomic traits (Moser and Lee, 1994). There were no correlations of RFLP distance between parents with progeny genetic variance for grain yield, biological yield, harvest index, height, or heading date. There was a small but significant (r = .32) correlation of RFLP distance with straw yield genetic variance in one year. Parental distance based on RFLP markers was unable to predict either heterosis or population genetic variance for grain yield in oats.

Another type of molecular marker for estimating differences at the DNA level is random amplified polymorphic DNA (RAPD) (Williams et al., 1990; Welsh and McClelland, 1990; Rafalski et al., 1991). These markers are DNA fragments arising from a mixture of short oligodeoxynucleotide primers of a single randomly chosen sequence mixed with genomic DNA and subjected to the polymerase chain reaction (Mullis and Faloona, 1987). The RAPD estimation of genetic distance is simpler than that using RFLP markers because it requires no development of specific clones to be used as probes.

Although RAPD markers are easy to generate, genetic distances estimated from RAPD markers may be subject to error. Primer binding sites on the genomic DNA template at a distance that can be overlapped during the extension phase of the PCR reaction should result in amplification of the intervening DNA sequence; however, Williams et al. (1990) have shown that the final amplification products may be a result of competition among binding sites rather than the actual number of available sites. Thus, template and primer DNA concentrations must be identical for each reaction mixture for reliable comparison of the resulting markers.

Smith et al. (1994), in a phylogenetic analysis of bacterial strains, found that presence or absence of a RAPD phenotype arose from either the absence of the primer binding site or competition from a preferred alternative RAPD product. They also detected co-migrating RAPD products from unrelated loci, as well as multiple, related products within a given reaction mixture. F_1 hybrids from crosses between maize inbreds did not always reveal simple inheritance of a dominant RAPD marker (Heun and Helentjaris, 1993). This indicates that amplification

of a given RAPD product could be dependent upon the genetic background, rather than the presence or absence of the DNA segment corresponding to the actual RAPD product. The problems encountered above should not preclude the use of RAPD markers to measure intraspecific genetic distance among inbred lines however, providing reaction conditions are carefully controlled (Ellsworth et al., 1993).

Genetic relationships using RAPD markers have been estimated in rice (*Oryza sativa* L.) (Yu and Nguyen, 1994), *Brassica* species (Mailer et al., 1994; dos Santos et al., 1994; Jain et al., 1994; Thormann et al., 1994; Hallden et al., 1994), tomato (*Lycopersecon esculaentum* Mill.) (Williams and St. Clair, 1993), wild oat (*Avena sterilis* L.) (Heun et al., 1994), and barley (*Hordeum vulgare* L.) (Tinker et al., 1993). Heun et al. (1994) compared RAPD markers to isozymes for determining relationships among wild oat accessions. Both isozyme and RAPD markers were able to distinguish all 24 of the wild oat accessions studied. Cluster analyses produced similar groupings among the accessions, but overall correlation of distance estimates was only moderate ($r = .36$). Principal component analysis resulted in more definitive groupings for the RAPD markers. A comparison of RAPD and RFLP markers in *Brassica oleracea* (L.) genotypes (dos Santos et al., 1994) gave equal coefficients of variance (CV) of the genetic distance estimates for equal sample size for both marker types. Both marker types identified distinct groupings for the sub-species cabbage, broccoli, and cauliflower. The observed differences in genetic distance estimates were concluded to be the result of sampling error rather than inherent DNA-based differences in how RAPDs and RFLPs reveal polymorphism. Thormann et al. (1994) estimated genetic relationships within and among

cruciferous species using RAPDs and RFLPs based on either genomic DNA (gDNA) or cDNA clones. The number of markers required for a CV of 10% was approximately 300 for each marker type. The correlations between distances among the three marker types were all high ($r > .90$). Dendrograms were compared using matrices based on cophenetic values and the Mantel test for matrix correspondence (Mantel, 1967). The correlation between the gDNA dendrogram and the cDNA dendrogram was higher than either correlation between RFLP dendrograms with the RAPD dendrogram. Although all three correlations were high (r = approximately .90) for intraspecific comparisons, the correlations between RFLP-based and RAPD-based dendrograms was low ($r < .37$) for interspecific comparisons. Hybridization tests using the RAPD fragments as probes demonstrated that some of the fragments scored as identical were not actually homologous at the interspecific level.

Jain et al. (1994) examined the use of RAPD genetic distance estimates to predict heterosis among crosses of Indian mustard (*Brassica juncea* L. Czern and Coss). They tested 12 Indian and 11 exotic *B. juncea* genotypes. Although they found no direct relationship between RAPD genetic distance and hybrid performance, RAPD analysis was able to classify the genotypes into two distinct groups comprised almost exclusively of the Indian and exotic genotypes, respectively. Crosses between groups exhibited more overall heterosis than crosses within groups.

Soybean is a self pollinated crop with limited genetic diversity in the elite germ plasm used by applied breeders in North America (Delanney et al., 1983). This limited genetic diversity makes research to exploit the existing diversity very important for continued

improvement of the crop. Delanney et al. (1983) calculated that ten ancestors contributed more than 80% of the gene pool for the northern soybean germ plasm. Continued improvement of soybean yield could be facilitated by identification of diverse parents within adapted germplasm for making cross pollinations, or the identification of unique diversity from among more recent plant introductions. Molecular markers could provide the necessary tools to make this identification.

The lack of diversity in soybean assumed by genealogical analysis is reflected in the low number of RFLP alleles found. Most RFLP loci have only two alleles and, in some cases, the second allele is rare (Keim et al., 1989; Keim et al., 1992). Despite this, enough RFLP diversity has been found to uniquely identify and establish relationships among large numbers of soybean lines (Skorupska et al., 1993). The large degree of relatedness among elite soybean lines may actually increase the effectiveness of molecular distance estimates among parents in predicting progeny performance. Some studies (Smith et al., 1990; Lee et al., 1989) have indicated that there is a high correlation of molecular genetic distance with progeny performance among closely related parents. The work presented here was undertaken to examine 1) the relationship between molecular markers and coefficients of parentage and 2) the relationship between parent genetic distance and progeny performance in soybean. Because of the close relationships among soybean lines in the Northern U.S., parent genetic distance may predict progeny genetic variance. Additionally, since pedigree information is not available for the early ancestral lines from which North American lines were developed, molecular marker distance may be more accurate than genealogical distance for this purpose.

SECTION ONE

RESTRICTION FRAGMENT LENGTH POLYMORPHISM RELATIONSHIPS AMONG SOYBEAN LINES IN THE NORTHERN UNITED STATES

INTRODUCTION

The continued improvement of soybean (*Glycine max* L. Merr.) yield in the northern United States may be limited by lack of genetic diversity. Only a few of the plant introductions brought from eastern Asia in the early twentieth century were suitable for seed production in the U.S., and these formed the original gene pool from which present soybean cultivars have been derived (Committee on Genetic Vulnerability of Major Crops, 1972). Delanney et al. (1983) calculated that ten ancestors contributed more than 80% of the gene pool for northern soybean germplasm. The genetic base does not appear to have changed in recent years (Gizlice et al., 1994), even with the inclusion of proprietary cultivars (Sneller, 1994). St. Martin (1982) compared 50 years of soybean breeding in the U.S. to a program of recurrent selection. He estimated the effective number of lines recombined each cycle to be between 11 and 15. This suggests that there has been a loss of genetic variability in soybean through selection in breeding programs and random drift. Gizlice et al. (1994) estimated that the genetic diversity in public cultivars was down 21% from that of the original ancestral plant introductions.

Relatedness of soybean genotypes can be estimated using pedigrees to calculate coefficient of parentage, or by analyzing each genotype for morphological or molecular markers. Cox et al. (1985) compared genetic distance estimates among soybean lines calculated using coefficient of parentage, morphological characters, and isozyme markers. Rank correlation coefficients of estimated genetic distances among all types of measurements, including a combination of both isozyme and morphological traits were statistically significant, but ranged from 0.15 to 0.60. This wide range may have been a result of the few isozymes or morphological traits used to estimate distance.

Keim et al. (1989) compared 58 soybean accessions using 17 restriction fragment length polymorphism (RFLP) loci. These included 48 accessions from the species *G. max*, 8 from *G. soja* Sieb. and Zucc., and 2 from *"Glycine gracilis"* Skvortz. The *G. max* accessions included 18 cultivars, 10 plant introductions, and 20 ancestral lines. Polymorphic loci generally had only two alleles, and for one-third of these loci, the second allele was rare, occurring in only one or two of the accessions characterized. On the average, any two cultivars differed at only 16% of the loci. Seven of the cultivars were identical at all 17 RFLP loci. The average within group diversity was greatest among the *G. max* plant introductions.

Keim et al. (1992) screened 16 ancestral and 22 adapted lines of *G. max* at 128 RFLP marker loci. Seventy percent of the clones were polymorphic, and their average polymorphism information content (PIC) was 0.30. Only one in five markers was informative between any two soybean genotypes. The polymorphism frequency among adapted lines was

lower using clones selected by screening interspecific germ plasm than when using clones selected using intraspecific germ plasm.

Skorupska et al. (1993) characterized 108 genotypes of *G. max.* using 83 molecular probes. These included ancestral genotypes, breeding lines, and elite cultivars encompassing maturity groups V-IX. The majority of the probes were uninformative, and only 35% detected polymorphism between any two lines with a frequency greater than 0.30. The greatest genetic distances were among the ancestral genotypes, while recently developed lines had a relatively narrower range of diversity. Genotypes within maturity groups were associated by principal component analysis, suggesting that molecular diversity was diminished through selection within geographical regions.

The studies outlined above included probes which had not previously been screened for levels of polymorphism revealed in adapted germ plasm. While the average marker diversity was low, some probes revealed no polymorphism, while others revealed above average marker diversity. In this study, only clones which had previously been determined to reveal high levels of polymorphism within elite soybean germ plasm were used as probes. The RFLP markers from these probes were used to 1)determine the relationships among ancestral plant introductions 2)estimate genetic distances among Northern soybean genotypes, 3)assess whether genetic relationships based on RFLP data are related to those based on known pedigree relationships, 4)determine whether RFLP allelic diversity has been lost in modern, elite lines from the Northern U.S. compared with the ancestral plant introductions, 5)examine more recent plant introductions as a source of exploitable genetic diversity, 6)estimate the effect of selection on the

contribution of alleles from parents compared to that expected from the coefficient of parentage.

MATERIALS AND METHODS

One hundred and three soybean cultivars and lines (Table 1.1) were evaluated using 57 RFLP markers. Seventy cultivars or elite lines from the northern U.S. (referred to hereafter as northern elites) were evaluated because they were important regional cultivars, or because they were parents in the Michigan State University breeding program. The 20 ancestral plant introductions (referred to hereafter as ancestors) were evaluated because they contributed approximately 80% of both the Northern and Southern soybean germ plasm parentage (Delanney et al., 1983; Gizlice et al., 1994; Sneller, 1994). A sample of 13 plant introductions (PI's) were selected because they performed well as parents when crossed with adapted genotypes from the northern U.S. (Nelson, 1994). The 70 cultivars and lines included 'Williams', 'Essex', and 'Ransom' , 10 cultivars selected from the cross Williams by Essex, and 5 cultivars selected from the cross Williams by Ransom. The progeny of these crosses were not included in the estimates of genetic distance mean and variance for the northern elites because these closely related lines would have biased the results. Some of the lines were not analyzed at all 57 marker loci.

Soybean DNA was extracted from greenhouse grown plants according to Keim and Shoemaker (1988) with modifications. Ten seed were sown for each genotype, but, in some cases, tissue was collected from as few as

Table 1.1 Soybean cultivars and lines analyzed.

Cultivars and Elite Lines

Asgrow

A2234(II)	A3127[†](III)
A2396(II)	A3860[†](III)
A2543(II)	A3966[†](III)
A2943(II)	A4268[†](IV)
	A5308[†](V)

Agripro

AP 1989(I)

Iowa State Univ.

A81-356022[§](III)	AC89-241029(II)
A84-185032(II)	AC90-115043[§](I)
A85-293033(II)	IA 2007(II)
A86-103027(II)	IA 2008(II)
A88-221013(II)	
AC89-145013(I)	

Michigan State Univ.

E90006(II)	E90012(II)
E90009(II)	E90013(III)
E90010(II)	E87223(II)

Northrup King

NKB-335[†](III)	NKS 20-26(II)
NKC-393[†](III)	NKS 23-12(II)
NKS 13-46(I)	NKS 25-96[§](II)
NKS 19-90(I)	NKS 42-40[†](IV)
NKS 20-20[§](II)	NKS 48-84(IV)

Pioneer HiBred

P9273(II)	P9441[†](IV)
P9341[§](III)	P9471[†](IV)

Univ. of Minn.

M82-946(I)

Ohio St. Univ.

HC84-2001(II)

Univ. of Ill.

LN86-983(II)

Purdue Univ.

C1786(II)	C1817(II)
C1797(II)	

Public Cultivars

Archer(I)	Hack(II)
Beeson 80(II)	Haroson(I)
Bert(I)	Hobbit[‡](III)
Brock(I)	Hoyt(II)
Burlison(II)	Kenwood(II)
Century 84(II)	Pella 86(III)
Conrad(II)	Pixie[‡](IV)
Dimon(II)	Ransom(VII)
Elf[‡](III)	RCAT Angora(II)
Elgin 87(II)	Sibley(I)
Essex(V)	Sprite[‡](III)
Gnome[‡](I)	Williams(III)

Table 1.1 (Cont'd)

Plant Introductions			
Ancestral Introductions		Other Plant Introductions	
AK(Harrow(III)¶	Mejiro(IV)	PI 68508(II)	PI 427099(I)
Biloxi-3(VII)¶	Mukden(II)¶	PI 297515(II)	PI 445830(I)
CNS(VII)¶	Palmetto(VII)	PI 297544(II)	PI 391594(II)
Dunfield(III)¶	Patoka(IV)¶	PI 361064(II)	PI 68522(II)
Flambeau(00)¶	Richland(II)¶	PI 54610(III)	PI 384474(II)
Lincoln(III)¶	Roanoke(VII)¶	PI 407710(I)	PI 90566-1(III)
Manchu(III)	S-100(V)¶	PI 68658(II)	PI 290126-b(II)
Mandarin(I)	Seneca(II)¶		
Mandarin (Ottawa)(0)	Tokyo(VII)		
Manitoba Brown(00)			

† Progeny of Williams by Essex, ‡ Progeny of Williams by Ransom, § Analyzed using only 38 marker loci, ¶ Ancestral lines which contributed parentage to the cultivars and elite lines examined in this study, # Ancestral lines which did not contribute to northern soybean germ plasm. Maturity groups are given in parenthesis.

four plants because of poor seed germination. Freeze-dried leaf tissue was pulverized using a paint shaker modified to hold 50ml disposable polypropylene centrifuge tubes. The dry tissue was placed in the tube along with 5ml of glass beads and shaken for two minutes. Pulverized tissue was incubated for one hour at 65°C with CTAB extraction buffer (2% CTAB, 1.4M NaCl, 0.2M EDTA, 0.1M Tris-HCl pH 8.0, 1% 2-mercapto-ethanol). The aqueous phase was then extracted twice with chloroform:isoamyl alcohol (24:1) and the nucleic acid precipitated with ice-cold isopropanol. DNA that proved difficult to cut with restriction enzyme was dissolved in a high salt solution and precipitated again to remove bound carbohydrate (Fang et al., 1991). Restriction enzyme digestions, electrophoresis, Southern blotting and hybridizations were

done according to Maniatis et al. (1982) with adaptation described by Diers and Osborn (1994).

The soybean genotypes were evaluated by RFLP analysis using 50 clones as hybridization probes. The clones (Table A.1) were obtained from Iowa State University and the University of Utah (Keim and Shoemaker, 1988). The clones were selected because they were previously shown to reveal a high frequency of polymorphism in elite germplasm (Webb, 1992, Skorupska et al., 1993). Each polymorphic RFLP fragment was scored as present or absent and genetic distance (RD) among the genotypes was calculated using a the compliment of the simple matching coefficient (1-(n'/n), where n' is the number of alleles two lines have in common and n is the total number of alleles scored in each comparison). Cluster analysis was performed on the similarity matrix using the unweighted pair-group method, arithmetic average (UPGMA).

Principal component analysis was done by first calculating a correlation matrix of alleles from the RFLP data. Genotypes were then plotted using eigenvectors calculated from the correlation matrix. Genetic similarity calculations, cluster analyses, and principal component analyses were done using NTSYS-pc software (Rohlf, 1992).

Polymorphism information content (PIC) at each locus was computed using the formula $1-\Sigma p_{ij}^2$, where p_{ij} is the frequency of the *j*th RFLP allele at the *i*th locus (Anderson et al., 1993). PIC is a measure of the genetic diversity. PIC increases with both the number of alleles at a locus and the equality of frequency of those alleles.

Genealogical distance (GD) was calculated as the compliment of the coefficient of parentage (CP). GD values used in clustering and correlation analyses were calculated with the assumed relations among

ancestors as described by Carter et al. (1993). Other ancestors were assumed to be unrelated, each parent was assumed to contribute equally to all progeny, and all lines were assumed to be completely inbred. The CP between any line and a line derived from a random mating population was calculated as:

$$r_{x,RM}=1/n\ \Sigma_{(i=1,n)}r_{x,Zi}$$

where $r_{x,RM}$ is the CP between line x and a line from a particular random mating population, n is the number of parents used to form the population, and $r_{x,zi}$ is the CP between x and the i^{th} parent of the population. All CP values were calculated with SAS programs (Sneller, 1994b).

RESULTS AND DISCUSSION

Fifty clones were hybridized onto the soybean DNA (Table 1.2). Seven of the clones revealed two independent polymorphic loci, whereas the remainder revealed only one polymorphic locus. Thus, a total of 57 marker loci were scored. Fifty-three marker loci had only two alleles, two loci had three alleles, and two loci had four alleles. Where three or four alleles were present, the least common allele(s) was observed only in the ancestral lines and/or the plant introductions. The allelism of fragments was readily identified because of the predominance of only two alleles at any locus and the inbred nature of the genotypes. Previous studies with soybean have shown a similar number of alleles for polymorphic markers (Keim et al., 1989; Keim et al., 1992; Skorupska et al., 1993)

The mean and range of PIC for loci in this study were 0.39 and 0.10-0.61 for the ancestors, 0.29 and 0.00-0.57 for the PI's, 0.37 and 0.0-0.50 for the northern elites, and 0.39 and 0.04-0.54 overall (Table A.1). This is an increase over average PIC values previously reported for soybean of 0.28 (Keim et al., 1989), 0.30 (Keim et al., 1992), and 0.24 (Skorupska et al., 1993). The greater PIC values in our study were probably the result of prior screening for high values within elite germ plasm.

According to the Committee for the World Atlas of Agriculture (1973), the soybean production region of China is found within three

agricultural areas defined by climate (Figure 1.1). These are the Northeast Cold Temperate Area (NECTA), the North Temperate Area (NTA), and the Central Subtropical Area (CSA).

Cluster analysis (Figure 1.2) grouped the ancestors according to place of origin as listed by Bernard et al. (1987a). 'Palmetto', 'CNS', and 'Biloxi-3', are ancestors from the CSA near the Yangtze delta (below 32N latitude) and clustered apart from all the other ancestors examined. These three ancestors and 'Mejiro' (PI 80837, from the Rikuu AES, Japan) have the 'Arksoy' cytoplasm (Grabau et al., 1992; Hanlon and Grabau, 1995). The remaining ancestors have 'Bedford' cytoplasm, except for Lincoln, whose cytoplasm is unique among the ancestors in this study. Most ancestors from the NECTA of China, which includes the Heilungjiang and Jirin provinces between 42N and 49N latitude, were clustered together (PI54610, 'Dunfield', 'Manchu', 'Patoka', and 'Richland'). This cluster also includes 'Flambeau', an introduction from Russia whose origin is likely from near this region, 'A.K.(Harrow)' and 'S-100', which are selections from 'A.K.', which probably originated from within the NECTA, and 'Lincoln', whose parents are unknown. Although Mandarin was introduced from Sui Hua, a town in the Heilungjiang province (NECTA) near 47N latitude, it and the selection 'Mandarin(Ottawa)' are clearly separated from other ancestors from the NECTA. These two ancestors are more closely associated with those originating from latitudes between 32N and 42N, which form separate clusters. 'Tokyo' (Yokohama, Japan, 36N latitude) and 'Roanoke' (a rogue from 'Nanking') are loosely associated with ancestors of the NECTA. 'Mukden' (from the NTA)in the Liaoning Province, 42N latitude), 'Seneca' (origin unknown), Mejiro (37N

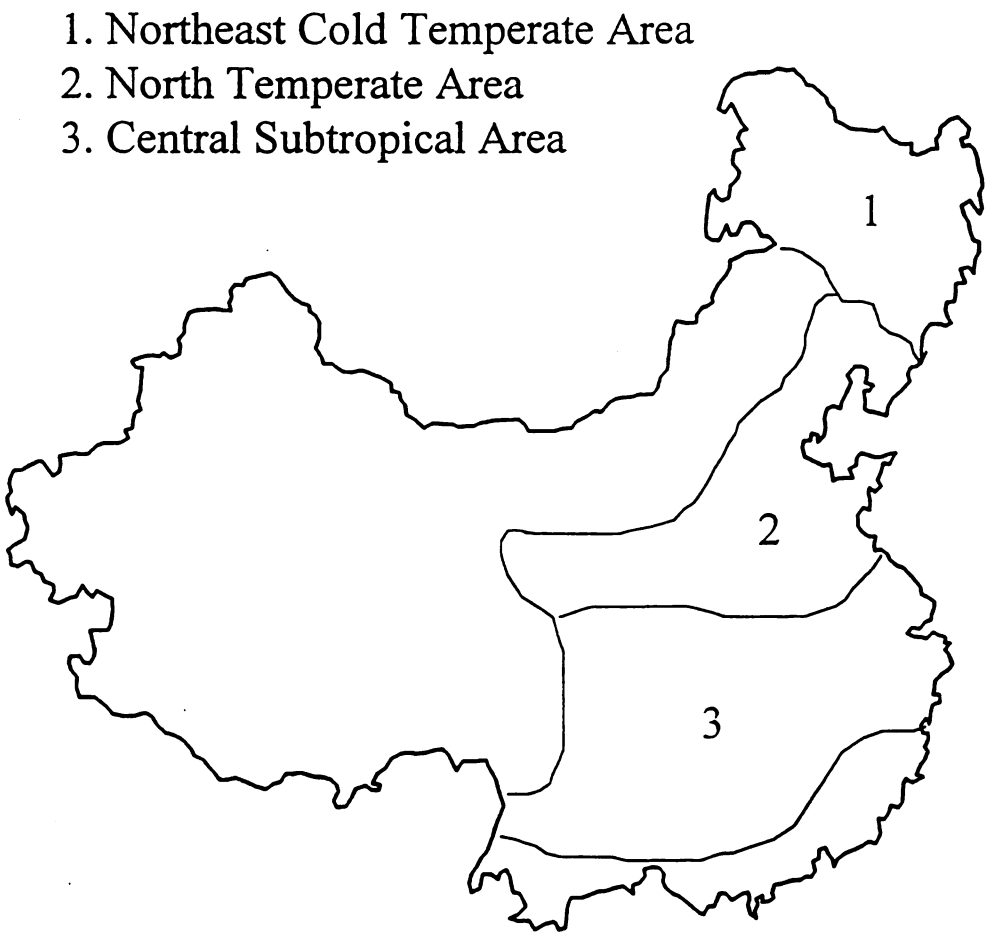

Figure 1.1 Agricultural areas associated with soybean production in China (Committee for the World Atlas of Agriculture, 1973).

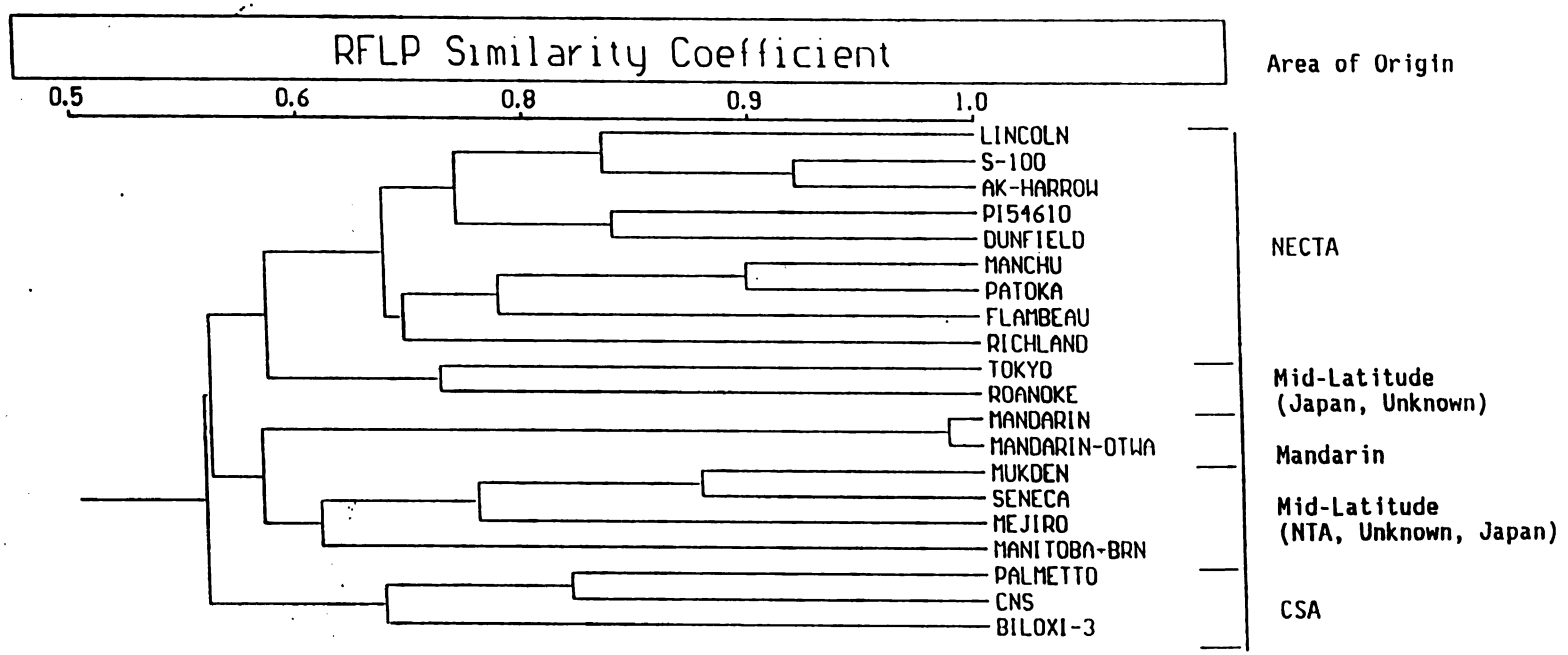

Figure 1.2 Phenogram showing the relationships of 20 ancestral plant introductions, based on RFLP analysis.

latitude) and 'Manitoba Brown' (origin unknown) are clustered and loosely associated with Mandarin and Mandarin(Ottawa).

Lincoln was believed to be a selection from the cross Mandarin by Manchu (Bernard, 1987), but molecular evidence disputes this. Lincoln cytoplasm differs from that found in either Mandarin or Manchu (Grabau et al., 1989). Our study provides further evidence that Lincoln is not a progeny of Mandarin by Manchu. We found that Lincoln has alleles for 17 markers not found in either Mandarin or Manchu. Given a mean value of 61% shared alleles within the ancestral introductions, the average value of common alleles between parents and progeny from crosses among these lines would be 81%. Lincoln and Mandarin shared only 36% of their RFLP alleles and were widely separated in a three dimensional principal component analysis (Data not shown). Manchu shared 65% of its alleles with Lincoln, however, Manchu was a heterogeneous introduction which gave rise to numerous pure line selections.

The close relationship between AK(Harrow) and Lincoln provides a possible clue to the origin of Lincoln. The cultivars 'Illini' and AK(Harrow) are selections from A.K., are phenotypically indistinguishable (Carter et al., 1993), and were found by Keim et al. (1992) to differ at only 1 in 129 RFLP loci. In this study, Lincoln and AK(Harrow) shared common alleles at 83% of the RFLP loci examined. Because both Illini and Lincoln were released by C. M. Woodworth at the Illinois Agricultural Experiment Station (Bernard et al., 1987), it raises the possibility that Illini or another selection from A.K. could be a parent of Lincoln.

The marker information provided insight into other relationships among ancestors. Bernard et al. (1987) listed CNS as probably equivalent

to Nanking, and Roanoke as a rogue from Nanking. CNS and Roanoke shared only 40% of the RFLP alleles examined. If CNS is equivalent to Nanking, the markers indicate that Roanoke is probably unrelated to Nanking.

The average RD among the ancestral lines Lincoln, AK(Harrow), PI 54610, S-100, and Dunfield was 0.24. These five ancestors contributed 38.5% and 35.4% of the elite parentage of soybean lines in the northern and southern US, respectively (Gizlice et al., 1994). Ancestral lines are typically assumed to be unrelated when GD is calculated among lines. When GD estimates among the northern elite lines in this study were adjusted by replacing a GD of one with the calculated RD among the ancestors, the average GD was reduced from 0.82 to 0.43.

The RFLP markers distinguished all lines evaluated, and clustering was generally in agreement with known genealogical relationships (Figure 1.3). The selection Mandarin(Ottawa) is closely paired with its ancestral line Mandarin. 'A2234', 'A2543', 'Century 84', and 'Burlison' are clustered together and each share the cultivar 'Century' as a parent. 'NKS20-26' and 'NKS19-90' are linked through their common parent 'Pride B152', which is a progeny of a cross with 'NKS13-46', also in the cluster. 'C1797' and 'C1786' are half-sibs and are paired together. 'E90012' and 'E90013', full-sibs, are paired and are clustered with their half-sib 'E90010'.

RFLP distance (RD) analysis resulted in association of genotypes into clusters previously defined by their ancestors. Biloxi-3 (maturity group (MG) VIII), Palmetto (MG VII), CNS (MG VII), and Manitoba Brown (MG 00) formed a cluster separate from all of the other lines examined (Group 5, Figure 1.3). Biloxi-3, Palmetto, and CNS were the only ancestral introductions evaluated from southern China and were among the

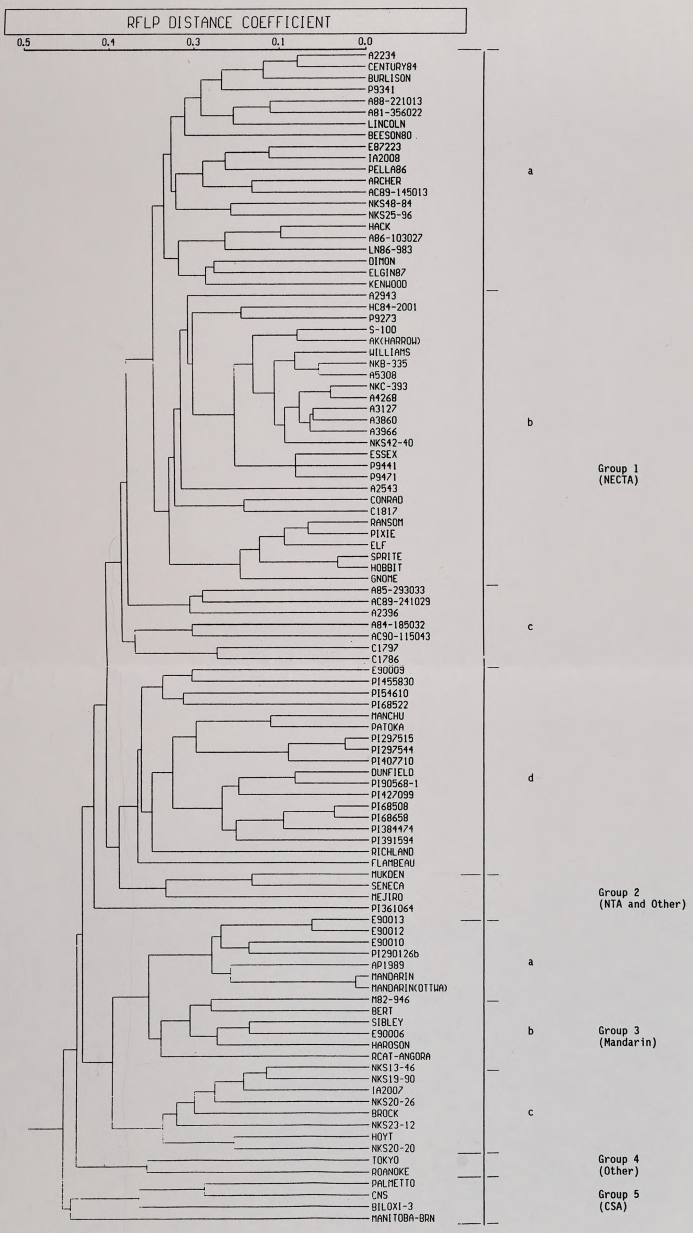

Figure 1.3 Phenogram showing relationships of a sample of soybean lines, based on RFLP analysis.

latest maturing lines in this study. Manitoba Brown may have clustered with the southern ancestors only by virtue of its differences with all the other lines.

Another prominent cluster (Group 3) is defined by Mandarin and Mandarin(Ottawa). This cluster is divided into three smaller clusters. Cluster 3a included Mandarin and Mandarin(Ottawa) along with three Michigan lines (E90010, E90012, E90013) and 'AP 1989', whose pedigrees traced back to Mandarin(Ottawa). It also included PI290126b, introduced in 1963 from Hungary. Cluster 3b includes lines which trace back to Mandarin(Ottawa) through cultivars and breeding lines from Minnesota. Cluster 3c included five lines from Northrup King, 'IA 2007', and 'Brock', which all have PI257435, an introduction from West Germany, in their pedigree. Cluster 3c also includes 'Hoyt', which traces back to Mandarin(Ottawa).

The majority of the northern elites clustered with the group of ancestors originating within the NECTA (Group 1). This group can be broken down further into associations with Lincoln (Group 1a), S-100, and AK(Harrow) (Group 1b), a cluster of six lines from Iowa State and Purdue Universities along with an Asgrow cultivar (Group 1c), and the remainder of the ancestors from this geographical area (Group 1d). Group 1b is comprised almost entirely of the parents and progeny of the crosses Williams by Essex and Williams by Ransom. Other lines within this group have Williams in their pedigree. The close relationship among these lines set them apart from other lines within Group 1 and separates Group 1c from Group 1a. The separation of group 1b from 1a is not consistent with the close RD between Lincoln and AK(Harrow) of 0.17. Cluster analysis without the progeny of Williams, Essex, and Ransom (not

shown) merges the members of Group 1c with Group 1a and places Lincoln close to AK(Harrow), S-100, and Williams. Cluster 1d includes most of the PI's which are grouped with the ancestors Manchu, Patoka, Dunfield and Richland, and, a Michigan breeding line, E90009. Ancestors whose origins are intermediate to the extreme northern and southern ancestors are associated outside of the groupings defined above.

Both RFLP and pedigree analysis grouped the progeny of the crosses Williams by Essex and Williams by Ransom with one or the other of the parents (Figures 1.3 and 1.4). However, pedigree relationships failed to account for the close RD relationship (RD = 0.30) between Williams and Essex. The CP between Williams and Essex is near zero, but Lincoln and S-100, ancestors of Williams and Essex, respectively, are closely related by RFLP analysis. RFLP data (Figure 1.3) associated Williams and Essex, while genealogical data (Figure 1.4) places Williams apart from Essex. GD analysis also failed to account for unequal allele contributions from the parents. GD analysis grouped most of the progeny of Williams by Essex with Essex, while RD analysis showed that they were more related to Williams. GD analysis was unable to associate lines of unknown pedigree, such as the ancestors, and account for allele contributions deviating from probability estimates. These are the likely reasons for disagreement between cluster relationships based on the two types of distance information.

The correlation between genealogical distances (GD) and RFLP distances (RD) among elite lines is highly significant ($P<0.001$, $R = 0.68$) (Figure 1.5). The low correlation coefficient may be a result of the downward bias of GD associated with alleles alike in state as well as the upward bias of GD from pre-existing relationships among the

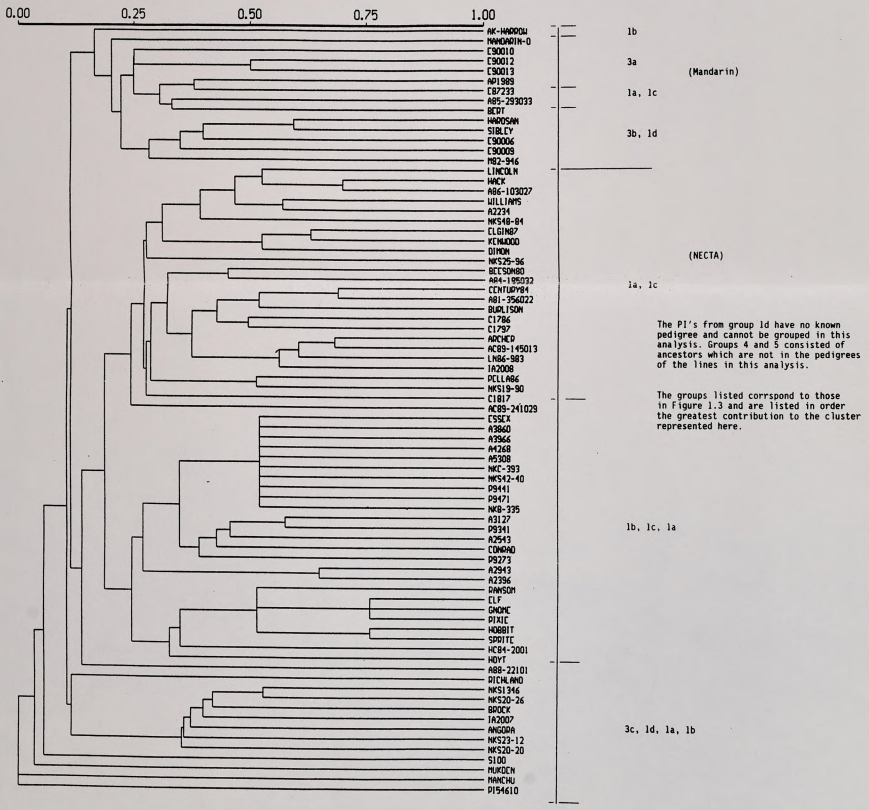

Figure 1.4 Phenogram showing relationships of a sample of soybean lines, based on coefficient of parentage analysis.

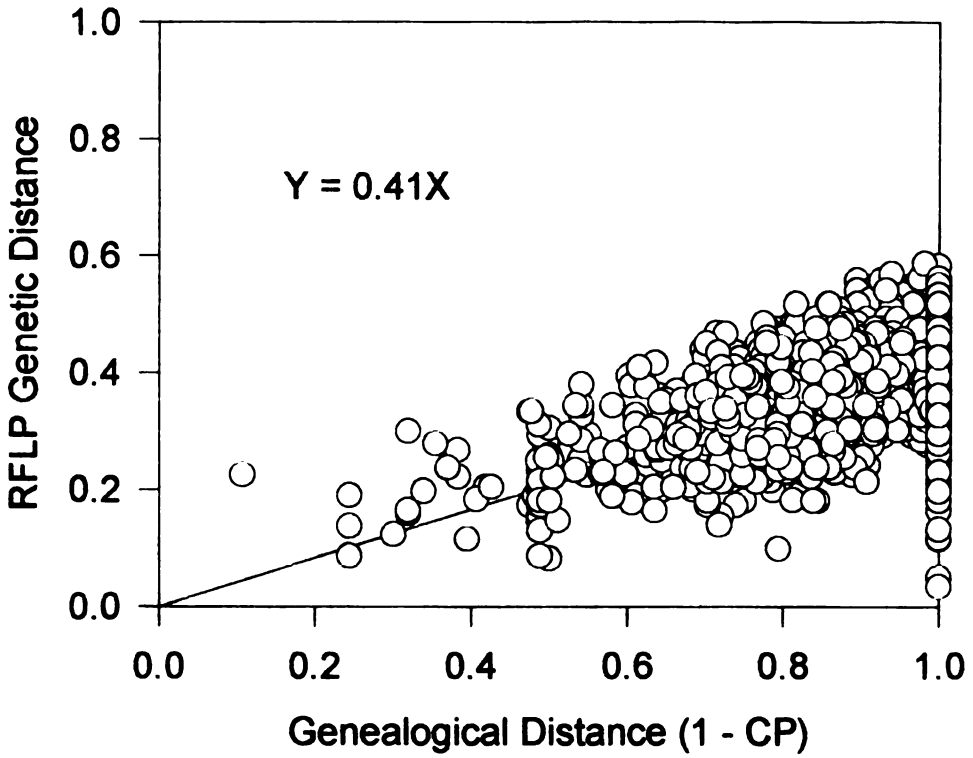

Figure 1.5 Scatterplot of the correlation between genetic distances based on RFLP and coefficient of parentage analyses.

ancestors. The ability of RD analysis to cluster the ancestors by area of origin is evidence that associations among the ancestors are real and likely to account for alleles identical by descent among ancestors previously believed unrelated. The effects of these biases should be greatest as GD nears 1.0.

The mean genetic distance was 0.39 among the ancestors and 0.36 among the cultivars and elite lines. This represents a statistically significant decrease in diversity of 8% in the cultivars and elite lines compared to the ancestors. There were five alleles (A4,pA59; A3/A4,pB142; A3,pK258; A3,pR92; Table A.1) present in the ancestors that were not present in the cultivars or elite lines examined. However,

these ancestors were not in the pedigrees of this sample of northern elites. The 20 ancestral lines we examined contributed 81% of the parentage within northern elite germ plasm according to Gizlice et al. (1994). However, pedigree analysis of the northern elites examined in this study revealed that the 20 ancestors comprised only 74.4% of their parentage, and that several ancestors not in the pedigrees of these northern elites were included (Table 1.1).

There were no unique RFLP alleles present in the PI's. Since these lines were selected on the basis of their performance as parents when crossed with adapted lines, they should not be considered a random sample of diversity within the available gene pool of PI's. The smaller distances (average = 0.30) within this group of PI's may be the result of their selection for good performance as parents when crossed with elite lines from the northern US. Although these lines were acquired from China (PI68508, PI 68522, PI 68658, PI 90566-1, PI391594, PI407710, PI427099), Hungary (PI297515), Russia (PI297544, PI384474), Yugoslavia (PI361064), and Romania (PI445830), the diversity implied by range of source countries may be deceiving. The Chinese lines are all from the northeast provinces of Heilungjiang and Jirin (Bernard et al., 1989a,b), the Russian lines come from the far east region bordering Northeast China, and the lines from eastern Europe were developed from imported germplasm which likely has origins in northeast China (Nelson, 1995).

Ten cultivars from the cross Williams by Essex and five cultivars from the cross Williams by Ransom were evaluated in this study because these crosses were so productive in generating new cultivars. Both crosses were between a Northern and a Southern cultivar. The coefficients of parentage between the parents were near zero for each

cross (Carter et al., 1993), however, the RD's did not reflect this for either cross. Williams differed from Essex at 17 of 55 loci examined (RD = 0.31) and Williams differed from Ransom at 21 of 55 loci (RD = 0.38). The cultivars 'NKB-335', 'P9471', 'A3127', and Pixie differed significantly from an equal contribution of alleles from each parent (Table 1.3). When cultivars are grouped by company or university of origin, all groups, except for cultivars developed by Asgrow are significantly different than that expected had there been an equal contribution of alleles from each parent. The Asgrow lines span maturity groups III, IV, and V; and may represent a broader range of adaptation than do the Williams by Essex lines in the other breeding programs.

Cluster analysis of parents and progeny of the cross Williams by Essex (Figure 1.6) shows an association of the majority of progeny with Williams. Grouping of lines implies common alleles are shared among them. The five maturity group (MG) III lines out of the cross Williams by Essex shared a common allele at 5 out of 17 loci (pA89, pK14, pK385, pA203, and pR201). Four out of five of these were Williams (MGIII) alleles. At the one locus where they shared an Essex allele (pA203), that allele was found in all the Williams by Essex progeny examined. No common alleles were shared by all group IV lines except the Essex allele of pA203. In 21 out of 51 possible cases (3 breeding programs by 17 clones), Williams by Essex lines within a breeding program shared the same allele. Within breeding programs. common alleles were shared by all lines at 5 (Northrup King), 11 (Pioneer HI-BRED), and 5 (Asgrow) loci out of 17. Although too few lines and alleles were examined for precise frequency estimates, generally, contribution of alleles from the parents were either bimodal or skewed toward a greater contribution from a

Table 1.2. Contribution of alleles from parent cultivars to selected progeny of the crosses Williams by Essex and Williams by Ransom.

Williams by Essex

Progeny Name	Williams alleles		Essex alleles		
	Number	Percent	Number	Percent	Prob.†
NKC-393	11	65	6	35	0.09
NKB-335	12	71	5	29	0.05*
NKS42-40	10	59	7	41	0.15
Total Northrup King	33	65	18	35	0.01**
P9441	6	35	11	65	0.09
P9471	5	29	12	71	0.05*
Total Pioneer	11	32	23	68	0.02*
A3127	5	29	12	71	0.05**
A3860	8	47	9	53	0.19
A3966	10	59	7	41	0.15
A4268	10	59	7	41	0.15
A5308	11	65	6	35	0.09
Total Asgrow	44	52	41	48	0.08

Williams by Ransom

Progeny	Williams alleles		Ransom alleles		
	Number	Percent	Number	Percent	Prob.
Gnome	9	43	12	57	0.14
Elf	9	43	12	57	0.14
Pixie	5	24	16	76	0.01**
Sprite	8	38	13	62	0.10
Hobbit	8	38	13	62	0.10
Total Ohio St. Univ.	39	37	66	63	<0.01**

† The probability value is for the allele distributions given and is calculated from the binomial frequency distribution assuming the null hypothesis of allele frequencies of 0.5.

single parent within both breeding programs and maturity groups (data not shown). This suggests that there likely had been selection during breeding for traits associated with specific alleles.

Progeny of the cross Williams by Ransom were selected for high yield, lodging resistance, and determinant growth habit, with specific adaptation to highly productive environments (Cooper, 1995). Lines were selected using a modified, early generation testing procedure (Cooper, 1990) which resulted in selection from within inbred lines. 'Elf', 'Gnome' and Pixie were selected from a common same F_2 line, as were 'Sprite' and 'Hobbit' (Carter et al., 1993). Elf and Gnome were selections from the same F_3 line (Cooper and Martin, 1981). Cluster analysis of parents and progeny of the Williams by Ransom cross (Figure 1.7) shows greater association of progeny with Ransom than Williams.

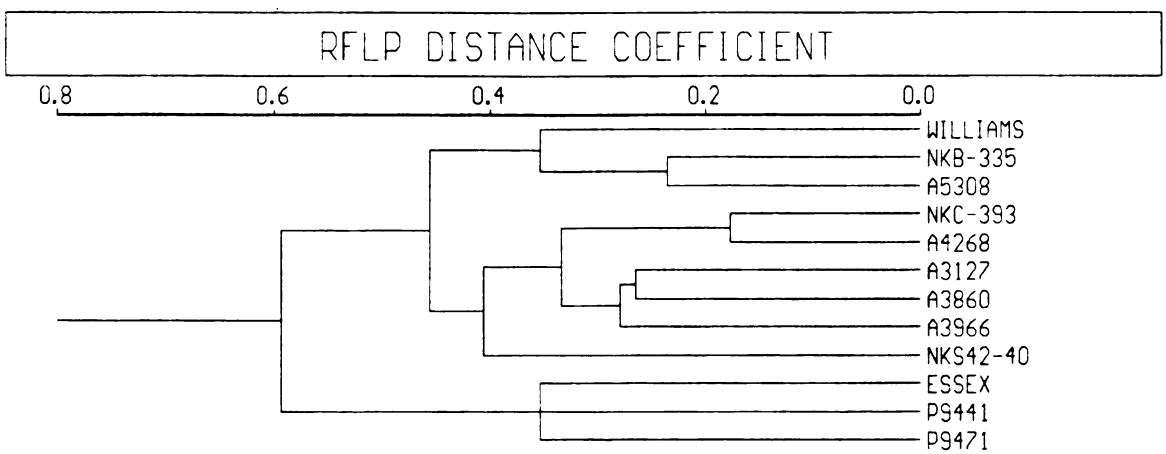

Figure 1.6 Phenogram showing the relationships of the parents and selected progeny of the cross Williams by Essex.

This was not expected because the progeny (MGI-IV) were selected for adaptation to an early maturing environment more amenable to Williams (MGIII) than to Ransom (MGVII). The lines were similar in yield to Williams when grown in this environment. This shows that favorable gene combinations can be introgressed from gene pools outside areas of adaptation, especially in conjunction with specific traits, such as determinate growth habit.

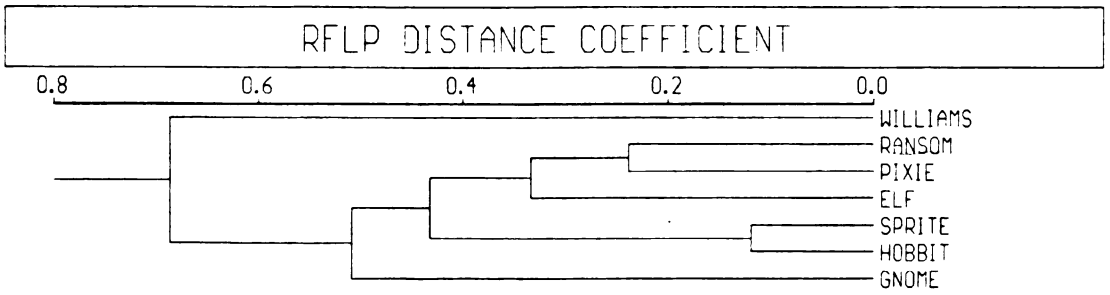

Figure 1.7 Phenogram showing the relationships of the parents and selected progeny of the cross Williams by Ransom.

CONCLUSIONS

RFLP fingerprinting can be a valuable tool for cultivar identification. The 50 clones revealed 57 independently segregating loci that completely distinguished the 95 lines evaluated. The lower diversity exhibited by RFLP alleles in soybean necessitates sampling more loci for identification by genetic fingerprinting or estimation of genetic distance than would be required for crops such as maize, which exhibits PIC values closer to 0.80 (Smith et al., 1990). Uniquely identifying closely related lines using molecular markers further increases the number of loci required. The number of markers required to differentiate two lines for at least two loci is increased by a factor of 1/(1-F) for any comparison, where F is the coefficient of parentage between the lines in question. The probability of detecting differences between lines with a given marker set is a function of allelic diversity (PIC) revealed by the markers used. For example, finding a marker difference 99.99% of the time in two unrelated (coefficient of parentage = 0) lines would require analysis at 14 loci with a PIC of 0.5 ($p = [0.5]^{14} < .0001$). However, if the two lines were related (e.g., F = 0.8), the number of loci required to distinguish the two lines with the same probability would be 14/(1-F) = 70. Thus, genetic fingerprinting must be

defined on the basis of both the marker set and the probability of relatedness among the individuals being examined.

Although selection practiced in breeding programs and random fixation may have reduced variability among elite lines, we found few RFLP alleles have been lost in elite germplasm. At the same time, the PI's evaluated contained no alleles not already present in the elite gene pool or available in the original ancestral lines. It is possible that these results could be an artifact of the clones selected for use in this study. Because RFLP markers seldom reveal more than two alleles, even across diverse soybean types, and we used clones that were previously shown to be polymorphic in elite germplasm, there would be only a limited chance of finding new alleles in ancestral lines or PI's. Perhaps clones that were monomorphic in elite material would reveal new alleles in the ancestral lines or PI's.

The relationships given by cluster analysis using RFLP data are generally in agreement with known genetic relationships estimated by pedigree. Cultivars and elite lines were associated with ancestor(s), which in turn were clustered according to geographical area of origin. However, RD's should be more accurate than GD's since they account for unknown relationships among primary breeding parents. A wide range of genetic distances was obtained using RFLP's. If RFLP genetic distances are accurate measures of true genetic differences, significant progress in performance from recombinant inbreds should still be possible from crosses using parents selected from within elite germplasm. The usefulness of either RFLP genetic distance estimates or coefficients of parentage for selecting parents in breeding programs remains to be

determined by the relative successes of the two methods in the production of transgressive segregants for release as new cultivars.

Selection within breeding programs for adaptation to particular growing areas or specific traits can result in a significant deviation from genetic relationships estimated by the coefficient of parentage. RFLP molecular distance coefficients are probably more accurate measures of genetic relationship than coefficients of parentage, even when pedigree information is available and accurate.

The association of progeny with a parent outside their maturity group in selection for specific traits such as semi-dwarf, determinate genotypes shows that unadapted germplasm can be a source for new, favorable gene combinations. The clustering of ancestors or cultivars and elite lines associated with a defined geographical area may be a result of random fixation during long term recurrent selection within adapted gene pools. It is not necessarily indicative of selection of alleles required for that environment.

The northern elites used in our study were a sample of lines from breeding programs in the northern US, and many of these were selected for use as parents in our breeding program. If the limited parentage of the lines we used is representative of the kind of selection occurring in breeding programs in general, it indicates a trend toward reduction in diversity within soybean germ plasm in the northern US.

SECTION TWO

THE RELATIONSHIP BETWEEN GENETIC DISTANCE AND GENETIC VARIANCE

INTRODUCTION

Plant breeders that develop inbred cultivars take advantage of genetic diversity between parents to produce agronomically superior cultivars through new combinations of available genes. According to quantitative theory, population genetic variance of a metric trait, such as yield, is the result of simultaneous segregation of many genes affecting that trait. Assuming no epistasis, a random population of inbred lines resulting from a cross between two highly inbred lines has genetic variance equal to $\Sigma(a_i^2)$, where a_i is the genotypic value of the homozygote at locus *i*, a quantitative trait locus affecting yield (Falconer, 1989). Assuming that genes affecting yield are many and randomly distributed throughout the genome, crossing high-performing lines from distinct genetic backgrounds should provide the greatest chance of pyramiding genes in combinations which result in progeny that out perform either parent. Accurate genetic distance measurements would then aid breeders in selecting diverse parental combinations. However, genetic distance calculated from random sampling of the genome will fail to account for either epistasis or loci with relatively large effects on yield.

Souza and Sorrells (1991), found that variance for biomass among F_3-derived families in oat (*Avena sativa* L.) initially increased with genetic distance between the parents of the crosses based on coefficient of parentage, but decreased as distances increased beyond a certain

point. Genetic variance for grain yield, test weight, heading date, maturity date, and grain filling period all decreased with increasing genetic distance. They suggested this negative relationship might have been a result of including parents unadapted for the region in which the experiments took place.

Restriction Fragment Length Polymorphism (RFLP) (Southern, 1975) and Random Amplified Polymorphic DNA (RAPD) (Williams et al., 1990) molecular markers can provide estimates of genetic distance if they relate to average differences in coding or regulating sequences. Distance estimates between parents using RFLP's have successfully predicted progeny performance in some cases. Lee et al. (1989) showed that RFLP genetic distance between parents was correlated with grain yield (R = 0.46) and specific combining ability (R = 0.74) in resulting maize hybrids. Smith et al. (1990) found r^2 values from regressing hybrid grain yield and grain yield heterosis on RFLP genetic distances between maize parent lines to be 0.87 and 0.77, respectively.

Other studies have shown little association between RFLP genetic distance and progeny performance. Melchinger et al. (1990) found that the correlation of parent distances with F_1 performance in maize was positive and significant, but too small to be of predictive value, especially between crosses of unrelated lines. Godshalk et al. (1990) investigated the relationship between hybrid performance and RFLP-derived genetic distances using inbred maize lines crossed with four testers. They found that while RFLP markers could be used to assign maize inbreds to heterotic groups, there was no relationship between RFLP genetic distance and hybrid performance. Moser and Lee (1994) found no significant relationship between marker genetic distances of parents

and hybrid grain yield in oats. The only significant relationship between RFLP distance among parents and progeny genetic variance was for straw yield in one of two years. Martin et al. (1995) examined the relationship between molecular marker diversity and hybrid yield in wheat using sequence tagged sites (Olson et al., 1989). Genealogical distance was significantly correlated with marker genetic distance (r = 0.68), but they found no significant association for either of the genetic distance estimates with F_1 grain yield, SCA effects, or heterosis.

Thormann et al. (1994) showed RFLP and RAPD markers to be very similar for estimating genetic distances within cruciferous species (r = .96), although the number of markers required for a coefficient of variation (CV) of the distance estimate of 10% was 327 for RAPD markers and 294 and 288 for RFLP markers selected from a genomic and a cDNA library, respectively. Comparison of genetic relationships among *Brassica oleracea* L. genotypes by dos Santos et al. (1994) also showed that RFLP and RAPD markers provide equal resolution. Bootstrap estimates of the CV of either marker type showed no significant differences for either the slope or intercept of the plot of CV vs number of markers. Jain et al. (1994) showed no direct correlation of RAPD genetic distances of parents with heterosis in *Brassica juncea* L. (Czern and Coss), but cluster analysis was useful in identifying heterotic groups.

There is no published information to date on the relationship of the genetic distance between parents of crosses and the genetic variation in the progeny for soybean (*Glycine max* L. Merr.). The objective of this research was to study this relationship. The extent of relatedness among elite soybean lines may increase the effectiveness of

genetic distance estimates among parents in predicting progeny performance.

MATERIALS AND METHODS

Distance Analysis

Genetic distances were estimated for forty-six soybean cultivars and lines (Tables 2.1) using RFLP, RAPD, a combination of RFLP and RAPD markers, and pedigree analyses. The cultivars and lines evaluated had previously been used as parents in the Michigan State University soybean breeding program. Fifty-seven polymorphic RFLP loci were obtained by hybridizing each of 50 clone/restriction enzyme combinations (Table 1.2) to total genomic DNA digested with one of five restriction enzymes. Detailed protocols are given in section one, materials and methods.

RAPD analyses were performed using 43 decamer primers (Table 2.2) obtained from Operon Technologies Inc., Alameda, CA (kits AA-AZ). Primers were screened prior to use for ability to reveal polymorphism among a sample of eight soybean lines from various breeding programs from the northern US. Reactions were performed in 25 μl volumes containing 50 mM Tris, pH 8.5, 3 mM $MgCl_2$, 200 μl each dNTP, 2 units Stoffel fragment (Perkin-Elmer, Norwalk CT) and 25ng each of primer and template DNA. The reactions were loaded into 200 μl thin-walled reaction tubes and placed in a Gene-Amp 9600® thermo-cycler (Perkin-Elmer Cetus Corp., Norwalk, CT). DNA was amplified using a cycling profile of 4 min at 94° C followed by 3 cycles of 15 s/94° C, 15 s/35° C, 45 s ramp to 72° C, 75 s/72° C; 34 cycles of 15 s/94° C, 15 s/40° C, 45 s ramp to 72° C, 75

Table 2.1 Cultivars and lines used as parents.

Population set A		Population set B	
A2234	E90013	A2234†	HAROSON
A2943	E87223	A2543	IA 2007†
A84-185032	ELGIN 87	A2936	IA 2008
A85-293033	HC84-2001	A88-221013	LN86-983
A86-103027	HACK	AC89-145013	NKS19-90†
AP 1989	HOYT	AC89-221013	NKS20-26
ARCHER	IA 2007	BERT	P9273
BEESON 80	KENWOOD	BROCK	RCAT-ANGORA
BURLISON	M82-946	C1786	
CENTURY 84	NKS19-90	C1797	
CONRAD	NKS23-12	C1817	
E90009	PELLA 86	E90006	
E90012	SIBLEY	E90010	

† Parent was used in both populations

Table 2.2 Primers† used in RAPD analysis.

Primer Number	Sequence 5′ 3′	Primer Number	Sequence 5′ 3′	Primer Number	Sequence 5′ 3′
AA 01	AGACGGCTCC	AD 05	ACCGCATGGG	AI 11	ACGGCGATGA
AA 02	GAGACCAGAC	AD 08	GGCAGGCAAG	AI 12	GACGCGAACC
AA 15	ACGGAAGCCC	AD 11	CAATCGGGTC	AI 15	GACACAGCCC
AA 17	GAGCCCGACT	AE 03	CATAGAGCGG	AI 16	AAGGCACGAG
AA 18	TGGTCCAGCC	AE 05	CCTGTCAGTG	AI 19	GGCAAAGCTG
AB 01	CCGTCGGTAG	AE 09	TGCCACGAGG	AJ 02	TCGCACAGTC
AB 04	GGCACGCGTT	AE 19	GACAGTCCCT	AJ 06	GTCGGAGTGG
AB 09	GGGCGACTAC	AG 04	GGAGCGTACT	AJ 09	ACGGCACGCA
AB 20	CTTCTCGGAC	AG 08	AAGAGCCCTC	AJ 11	GAACGCTGCC
AC 02	GTCGTCGTCT	AH 06	GTAAGCCCCT	AJ 12	CAGTTCCCGT
AC 05	GTTAGTGCGG	AH 08	TTCCCGTGCC	AJ 15	GAATCCGGCA
AC 06	CCAGAACGGA	AH 09	AGAACCGAGG		
AC 08	TTTGGGTGCC	AH 14	TGTGGCCGAA		
AC 12	GGCGAGTGTG	AH 17	CAGTGGGGAG		
AC 19	AGTCCGCCTG	AH 18	GGGCTAGTCA		
AD 01	CAAAGGGCGG	AI 09	TCGCTGGTGT		

† Primers were obtained from Operon Technologies Inc., Alameda, CA. Operon primer numbers are given, followed by their nucleotide sequence from the 5′ to 3′ direction.

s/72° C; and a final extension period of 7 min at 72° C. Reactions were kept at 4° C overnight, and 20 μl of the completed amplification reaction mixture were run in 1.4% agarose gels.

Each polymorphic RFLP or RAPD fragment was scored as present or absent and genetic distance among the genotypes was calculated using the compliment of the simple matching coefficient (1 - n'/n, where n' is the number of alleles two lines have in common and n is the total number of alleles scored in each comparison). Combined distances were calculated from a matrix of all RFLP and RAPD marker data. Because the RAPD markers were mostly dominant and the RFLP markers were mostly codominant, the RFLP markers were scored as either present or absent for one allele per locus to give equal weight to each marker type. Where both RFLP alleles were present in a heterogeneous mixture, the marker was scored as present. This occurred in 31 out of a total of 2668 cases, and resulted in a small amount of error compared with RFLP analysis where both alleles were scored. This same error is inherent in RAPD markers that are dominant. Genealogical distance (GD) was calculated as the compliment of the coefficient of parentage as previously described in section one.

Field Evaluation

Two sets of single seed descent populations were evaluated in field tests. The populations were all derived from two-parent crosses. Set A included 22 populations of $F_{3:4}$ lines evaluated in 1993 and a subgroup of fourteen populations evaluated as $F_{3:5}$ lines in 1994. Subgroup populations were selected to provide a wide range in genetic

distance and germ plasm diversity. For each population, 28 lines and the two parents were tested in each year. The tests were blocked by population and lines were randomized within each population. In 1993, the populations were sown on May 20 at the Michigan State University farm near Mason, MI. Thirty seeds of each line were sown in plots 91 cm long with a 76 cm row spacing and a 91 cm alley between ranges. Rows of plots were bordered on each side with a continuous row of 'Dimon'. The test was replicated 3 times using a randomized complete block design. Plots were harvested for yield measurement over a period of several weeks beginning in the middle of October. In 1994, populations were evaluated at 2 locations; at the Michigan State University farm in East Lansing, MI and near Britton, MI with 2 replications at each location. The planting dates were May 13 for Britton and May 16 for East Lansing. Plots consisted of two 2.74 m rows with 91 cm between ranges and row spacing of 76 cm. Both rows were harvested to estimate yield. Harvest dates were October 13 for Britton and October 18 and October 22 for East Lansing.

Set B included 25 populations of $F_{4:5}$ lines evaluated in 1994 and a subgroup of ten populations evaluated as $F_{4:6}$ lines in 1995. For each population, 48 lines and the two parents were tested in each year. In 1994, the populations were sown on May 24 at the Michigan State University farm in East Lansing, MI, using the same experimental design as the 1993 test for set A. Plots were harvested over a period of several weeks, beginning in the middle of October. In 1995, populations were evaluated at the Michigan State University farm near Mason, MI and near Britton, MI with 2 replications at each location. The experimental design and plot layout was the same as that for the 1994 test for set A.

Planting dates were May 22 for near Britton and June 1 at Michigan State University, and harvest dates were November 8 and Oct 13, respectively.

In all plots, fertilizer rates per ha were 6.7 kg N, 26.9 kg P, and 26.9 kg K. All plots except Mason, MI received .56 kg/ha Lexone® (4-Amino-6-(1,1-dimethylethyl)-3-(methylthio)-1,2,4-triazin-5(4H)-one) and 4.7 l/ha Lasso® (2-chloro-2′,6′-diethyl-N-(methoxymethyl) acetanilide) incorporated into the soil prior to planting. Basogran® (3-(1-methylethyl)1H-2,1,3-benzothiadiazin-4(3H)-one 2,2-dioxide) (1.2 l/ha), Concentrated Crop Oil (1.2 l/ha), and Assure® (2-[4-[(6-chloro-2-quinoxalinyl)oxy]phenoxyl] propionic acid, ethyl ester) (0.37 l/ha) were applied post-emergence. The plots at Mason, MI were treated as above, except 2.3 l/ha Dual® (2-chloro-N-(2-ethyl-6-methylphenyl)-N-(2-methoxy-1-methylethyl acetamide) was applied in place of Lasso.

Maturity date was recorded as the day on which 95% of the pods had reached mature pod color (R8) (Fehr and Caviness, 1977). Plant height was recorded as inches from the ground to the average terminal node of a group of plant randomly chosen toward the center of the plot. Lodging index was a subjective score of 1 through 5, where 1 indicated that plants were almost completely vertical, and 5 indicated that the main stem was lying flat on the ground. Both plant height and lodging were measured at full maturity, just prior to harvest.

Genetic variances of populations were estimated from algebraic combinations of mean squares (MS) (Johnson et al., 1955). The algebraic estimate of genetic variance for the single row plots at one location was (MS(genotype)-(MS(error))/3). The estimate taken over two locations for the 2-row plots was MS(genotype by location)-MS(genotype))/4 where the genotype by environment interaction was significant. Otherwise,

(MS(genotype)-(MS(error))/4) was used. Negative estimates of genetic variance are listed, but were assumed to be zero for correlation and regression analyses. Non-significant, but positive genetic variances were analyzed at their calculated value. The standard error of the genetic variance was estimated using the formula: $[(2(MS1)^2/df+2 + 2(MS2)^2/df+2)/(rl)^2]^{1/2}$, where MS1 and MS2 are the mean squares used in the algebraic determination of the genetic variance and denominator *df* is the degree of freedom for that mean square. The terms in the overall denominator are replications (*r*) and locations (*l*).

RESULTS

Fifty clones revealed fifty-seven independent polymorphic RFLP loci (Table 1.2). The clones mapped to 14 of 23 linkage groups (Lorenzen et al., 1995). Nine linkage groups were not covered (I,N,O,Q,S,U,W,X), although 6 of the 9 (Q,S,U,V,W,X) were small and had only 2-4 markers per group. Forty-three decamer primers revealed 78 polymorphic RAPD loci. The location of markers is not known because they have not been incorporated into a map of the soybean genome. Genetic distances between the parents and genetic variances for several agronomic traits are given in Table 2.3 through Table 2.6.

For the parents of population set A, RFLP genetic distances (RFD) averaged 0.38 with a range of 0.24 to 0.54, RAPD distances (RPD) averaged 0.31 with a range of 0.23 to 0.39, while distance estimates from a combination of both RFLP and RAPD marker data (CMD) averaged 0.36 with a range of 0.27 to 0.43 (Table 2.3). Genealogical distances (GD) between parents of populations averaged 0.84 and ranged from 0.73 to 0.95. Genealogical distance was not correlated with either RFD or RPD, although it was significantly correlated with CMD (Table 2.7). RPD was significantly correlated with RFD, and CMD was significantly correlated with both RFD and RPD.

The parents of population set B were more closely related than the parents of population set A. The RFD between parents of population set B averaged 0.36 with a range of 0.20 to 0.48, RPD averaged 0.26

Table 2.3 Parents, genetic distance estimates, and genetic variances for several agronomic traits for the 1993 single-row plots.

	Parents		Distance			Yield		Maturity		Height		Lodge	
Pop.	Female	Male	GD†	RFD¶	RPD§	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.
1	A84-185032	ELGIN 87	0.79	0.33	0.27	103	420	12.9**	5.2	2.1*	1.1	0.43**	0.16
2	E90009	ARCHER	0.93	0.29	0.35	1287*	691	69.4**	19.6	2.7*	1.7	0.16	0.12
3	A85-293033	PELLA 86	0.80	0.24	0.25	2724**	997	78.3**	22.3	16.6**	5.0	0.71**	0.20
4	NKS 23-12	PELLA 86	0.89	0.44	0.33	1298**	573	15.0**	4.8	4.1**	1.9	0.04*	0.02
5	HC84-2001	A96-103027	0.76	0.34	0.23	974*	677	7.2*	4.6	9.8**	4.0	0.13	0.16
6	HC84-2001	NKS 19-90	0.89	0.38	0.27	888*	556	77.4**	22.2	23.2**	6.7	0.24**	0.11
7	E90012	KENWOOD	0.89	0.46	0.34	1116*	694	91.1**	25.6	2.0	2.0	0.54**	0.20
8	BURLSISON	M82-946	0.80	0.43	0.31	866**	420	16.1**	6.2	2.7**	1.4	0.05**	0.02
9	A84-185032	A2234	0.76	0.35	0.28	915*	555	64.7**	18.1	7.1**	2.3	0.07**	0.04
10	BURLISON	A84-185032	0.77	0.34	0.37	907	688	19.8**	7.3	-0.2	1.0	0.33**	0.16
11	M82-946	CENTURY 84	0.73	0.41	0.31	-227	318	27.1**	8.1	1.6	1.2	0.03	0.02
12	HOYT	CONRAD	0.83	0.29	0.27	1991**	781	53.7**	15.0	34.0**	9.7	0.41**	0.14
13	HACK	AP 1989	0.81	0.39	0.24	1940**	636	47.2**	14.9	6.1**	2.2	0.02	0.31
14	E90012	ARCHER	0.92	0.48	0.39	1622**	674	56.3**	15.9	4.9**	2.0	0.10	0.07
15	E90009	NKS 23-12	0.95	0.43	0.29	620*	369	51.6**	14.8	2.8*	1.3	0.15**	0.06
16	SIBLEY	PELLA 86	0.87	0.51	0.37	382	325	18.0**	7.3	4.5**	1.6	0.61**	0.20
17	E87223	IA 2007	0.81	0.25	0.26	1165**	582	46.4**	13.8	1.8*	1.0	0.07**	0.04
18	A84-185032	NKS 19-90	0.84	0.35	0.33	1317**	601	18.4**	6.3	4.3**	1.6	0.04	0.06
19	E90013	CENTURY 84	0.89	0.54	0.33	1075**	562	33.6**	10.5	15.0**	4.7	0.43**	0.18
20	E90013	BEESON 80	0.93	0.42	0.30	1022**	519	43.2**	13.4	8.8**	3.2	0.84**	0.27
21	BURLISON	E87223	0.83	0.30	0.28	2367**	935	64.4**	18.7	7.6**	2.5	1.18**	0.38
22	SIBLEY	A2943	0.85	0.40	0.34	2155**	789	68.2**	19.2	20.7**	6.0	0.93**	0.07

*, ** Significant at the 0.05 and 0.01 levels, respectively. †, GD = genealogical distance; ¶, RFD = RFLP distance; §, RPD = RAPD distance.

Table 2.4 Parents, genetic distance estimates, and genetic variances for several agronomic traits for the 1994 two-row plots.

	Parents		Distance			Yield		Maturity		Height		Lodge	
Pop.	Female	Male	GD†	RFD¶	RPD§	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.
2	E90009	ARCHER	0.93	0.29	0.35	-2.5	5.1	39.7**	12.7	15.9**	6.0	0.37**	0.16
4	NKS 23-12	PELLA 86	0.89	0.44	0.33	6.9**	3.1	8.8**	2.9	9.6**	3.0	0.16**	0.06
7	E90012	KENWOOD	0.89	0.46	0.34	-3.1	4.5	41.6**	11.4	10.2**	3.7	0.15**	0.07
8	BURLISON	M82-946	0.80	0.43	0.31	0.9	3.6	8.1**	2.5	2.3**	1.0	0.14**	0.06
11	M82-946	CENTURY 84	0.73	0.41	0.31	5.8**	2.8	16.1**	4.7	7.2**	2.3	-0.02	0.04
12	HOYT	CONRAD	0.83	0.29	0.27	17.4**	6.6	18.7**	5.4	63.9**	17.9	0.36**	0.11
13	HACK	AP 1989	0.81	0.39	0.24	3.6	2.8	29.8**	8.7	9.7**	2.4	0.03	0.04
14	E90012	ARCHER	0.92	0.48	0.39	1.5	2.8	27.9**	8.2	16.1**	5.7	0.21	0.08
15	E90009	NKS 23-12	0.95	0.43	0.29	4.4*	2.7	18.6**	5.5	8.9**	3.2	0.23**	0.09
16	SIBLEY	PELLA 86	0.87	0.51	0.37	-1.1	4.0	5.0**	2.5	7.9**	3.6	0.23*	0.13
19	E90013	CENTURY 84	0.89	0.54	0.33	-2.4	3.3	24.2**	7.4	48.1**	13.4	0.42**	0.17
20	E90013	BEESON 80	0.93	0.42	0.30	-4.3	3.6	24.6**	8.2	18.6**	6.3	0.34**	0.12
21	BURLISON	E87223	0.83	0.30	0.28	14.4**	5.9	39.8**	11.0	13.5**	4.3	0.44**	0.14
22	SIBLEY	A2943	0.85	0.40	0.34	-9.8	6.4	32.3**	10.1	28.3**	8.7	0.19	0.11

*, ** Significant at the 0.05 and 0.01 levels, respectively. †, GD = genealogical distance; ¶, RFD = RFLP distance; §, RPD = RAPD distance.

Table 2.5 Parents, genetic distance estimates, and genetic variances for several agronomic traits for the 1994 single-row plots.

	Parents		Distance			Yield		Maturity		Height		Lodge	
Pop.	Female	Male	GD†	RFD¶	RPD§	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.
1	E90010	C1817	0.94	0.47	0.30	1813**	521	68.6**	15.0	28.5**	6.8	0.57**	0.14
2	E90010	A2543	0.94	0.47	0.32	1481**	393	19.2**	4.2	9.2**	2.2	0.09**	0.03
3	BROCK	P9273	0.88	0.44	0.33	1998**	679	48.1**	10.3	10.4**	2.5	0.13**	0.05
4	C1797	IA 2007	0.84	0.48	0.28	1628**	590	19.8**	4.4	6.9**	2.1	0.20**	0.07
5	IA 2008	P9273	0.78	0.38	0.26	2792**	703	52.8**	11.0	23.6**	5.3	0.30**	0.09
6	NKS 20-26	A2396	0.87	0.42	0.32	3078**	1156	57.7**	13.7	15.7**	4.3	0.46**	0.12
7	RCAT-ANGORA	P9273	0.85	0.40	0.28	2814**	759	71.7**	15.0	24.0**	5.3	0.41**	0.11
8	P9273	NKS 19-90	0.86	0.44	0.36	3863**	922	58.9**	12.4	20.3**	4.6	0.44**	0.11
9	NKS 19-90	A2234	0.78	0.39	0.26	1765**	502	3.1**	0.8	1.7**	0.6	0.01	0.01
10	A2396	E90010	0.91	0.48	0.25	3480**	899	58.4**	12.3	21.2**	4.8	0.30**	0.08
11	A2396	E90010	0.91	0.48	0.25	3558**	1040	37.9**	8.6	8.8**	2.5	0.10**	0.05
12	AC89-145013	P9273	0.80	0.38	0.20	3580**	872	67.6**	14.0	22.2**	4.9	0.57**	0.14
13	A2396	A88-221013	0.89	0.35	0.26	876**	408	45.7**	10.1	16.6**	4.4	0.70**	0.18
14	LN86-983	NKS 19-90	0.77	0.35	0.19	1536**	546	11.4**	2.9	8.2**	2.7	0.03	0.05
15	E90010	AC89-241029	0.90	0.34	0.24	1070**	443	81.3**	17.6	37.9**	8.5	0.46**	0.12
16	BROCK	C1786	0.86	0.34	0.33	820**	372	47.3*	11.1	17.9*	4.9	0.12*	0.06
17	A2234	P9273	0.71	0.33	0.30	7181**	1832	76.7**	17.5	26.4**	6.9	0.56**	0.17
18	C1817	P9273	0.72	0.27	0.13	787**	334	10.9**	2.5	5.3**	1.5	0.03	0.04
19	HAROSON	AC89-241029	0.82	0.33	0.27	647*	375	21.9**	4.7	10.9**	2.9	0.48**	0.13
20	IA 2008	BROCK	0.80	0.23	0.31	461	335	7.2**	1.8	9.0**	3.1	0.18**	0.07
21	BERT	E90010	0.76	0.26	0.22	1430**	452	21.3**	4.8	14.5**	3.8	0.29**	0.09
22	LN86-983	C1817	0.71	0.22	0.27	904**	415	22.9**	5.5	23.9**	6.0	0.52**	0.14
23	E90006	E90010	0.62	0.25	0.17	1544**	483	8.2**	2.2	11.2**	2.9	0.11**	0.04
24	NKS 19-90	BROCK	0.58	0.26	0.24	1216**	499	13.4**	3.0	4.4**	1.5	0.36**	0.10
25	BROCK	IA 2007	0.63	0.20	0.24	908**	356	9.3**	2.1	4.5**	1.3	0.23**	0.07

*, ** Significant at the 0.05 and 0.01 levels, respectively. †, GD = genealogical distance; ¶, RFD = RFLP distance; §, RPD = RAPD distance.

Table 2.6 Parents, genetic distance estimates, and genetic variances for several agronomic traits for the 1995 two-row plots.

	Parents		Distance			Yield		Maturity		Height		Lodge	
Pop.	Female	Male	GD†	RFD¶	RPD§	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.	σ^2	S.E.
1	E90010	C1817	0.94	0.47	0.30	17.5**	5.5	28.5**	6.1	17.6**	4.2	.44**	.12
4	C1797	IA 2007	0.84	0.48	0.28	11.0**	3.9	10.3**	2.2	4.8**	1.3	.28**	.07
6	NKS 20-26	A2396	0.87	0.42	0.32	14.9	9.9	30.8**	7.9	24.6**	6.5	.65**	.19
8	P9273	NKS 19-90	0.86	0.44	0.36	28.9**	6.7	37.5**	8.0	27.2**	5.8	.18**	.05
17	A2234	P9273	0.71	0.33	0.30	48.4**	13.5	57.7**	12.1	59.2**	12.6	.57**	.14
18	C1817	P9273	0.72	0.27	0.13	5.5**	1.8	5.0**	1.1	4.6**	1.2	.18**	.05
20	IA 2008	BROCK	0.80	0.23	0.31	9.4**	3.7	7.0**	1.6	8.8**	2.2	.10**	.03
22	LN86-983	C1817	0.71	0.22	0.27	6.2**	2.3	14.6**	3.6	14.5**	3.5	.81**	.18
23	E90006	E90010	0.62	0.25	0.17	8.9**	3.4	4.7**	1.1	8.4**	2.8	.01	.02
25	BROCK	IA 2007	0.63	0.20	0.24	3.5*	2.1	4.6**	1.1	4.2**	1.2	.01**	.01

*, ** Significant at the 0.05 and 0.01 levels, respectively. †, GD = genealogical distance; ¶, RFD = RFLP distance; §, RPD = RAPD distance.

Table 2.7 Correlations and P-values among genetic distance measures' for the parents of population sets.

	Population set A			
	RFD	RPD	CMD	GD
RFD	-	.55** <.009	.93** <.001	.42 <.053
RPD		-	.78** <.001	.41 <.06
CMD			-	.44* <.04

	Population set B			
	RFD	RPD	CMD	GD
RFD	-	.42* <.04	.88** <.001	.79** <.001
RPD		-	.70** <.001	.50* <.02
CMD			-	.75** <.001

*, ** Significant at the 0.05 and 0.01 levels, respectively. † RFP = RFLP Distance, RPD = RAPD Distance, CMD = Combined RFLP and RAPD Distance, GD = Genealogical Distance (1-Coefficient of Parentage.

with a range from 0.13 to 0.36, and CMD averaged 0.28 with a range of 0.18 to 0.39 (Table 2.4). GD between parents of population set B averaged 0.81 with a range of 0.58 to 0.94. In contrast to parents in set 1, all distances calculated between parents in set 2 were significantly correlated with one another (Table 2.7).

Two populations in set B were not included in the analysis. In 1995, population 6 was not included in the 1995 analysis because, at one location, 14 of the 48 progeny lines along with the parent 'NKS20-26' were devastated by a disease which was not diagnosed. The algebraic estimate of the yield genetic variance using the remaining progeny in population 6 fit well within the linear regression model of yield genetic variance versus RFLP distance (data not shown), but the variance was non-significant according to the F-test. This could have been a result of the loss of degrees of freedom from the reduced number of progeny included in the analysis. Also, analysis of variance showed significant genotype by environment interaction among the remaining progeny. Therefore, population 6 was not included in the analysis in 1995.

The yield genetic variance of population 17 was almost twice that of any other population in set 2 in both 1994 and 1995 (Table 2.4 and Table 2.6), although the genetic distance between the parents was moderate. Because of its disproportionately large yield genetic variance, population 17 was tested as an outlier according to the procedures given by Snedecor and Cochran (1967). Using the standard error of the individual estimate:

$$S_{Y\cdot\hat{Y}} = S_{y.x}[1 + 1/(n-1) + (X_0-\bar{X})^2/\Sigma(X_i-\bar{X})^2]^{1/2}$$

and:

$$t = (Y-\hat{Y})/S_{Y\cdot\hat{Y}}$$

where $S_{y.x}$ is the standard deviation from regression, *n* is the number of the data points including the outlier, $\bar{X}$ is the mean genetic distance, and X_0 is the distance associated with the outlier. The P-value associated with *t* is set to *n*P. In all cases where the regression of yield genetic variance on genetic distance was significant (population 17 omitted), population 17 was a significant outlier (*n*P < 0.05). Correlations were calculated with and without population seventeen.

In all the experiments, most populations exhibited significant genetic variance for all traits measured (Table 2.3 through Table 2.6). The exception was yield genetic variance in population set A in 1995. Only 5 out of 14 populations had significant yield genetic variance in the 1995 two-row plots.

There were no significant correlations between any of the distances and genetic variance estimates from the populations for set 1 in the 1993 1-row plots (Table 2.8). In the 1994 evaluation of the set A populations in two-row plots, RFD (Figure 2.1), RPD (Figure 2.2) and CMD (Figure 2.3) were both negatively correlated with yield.

Yield genetic variance of populations in set B was significantly related to RFD in the 1994 single-row plots, with an *r* of 0.41 (Table 2.9). Maturity genetic variance was also significantly correlated with both RFD and GD for these populations in 1994. In the 1995 two-row plots, there were no significant correlations between genetic distance and genetic variance for any trait.

When population 17 was excluded from the analysis of set B populations, the correlations between genetic distance and genetic variance for yield and maturity generally increased (Table 2.10). The

Table 2.8 Correlation coefficients and P-values of genetic distance estimates† between parents with genetic variances of several agronomic traits for population set one.

Genetic Distance	1993 Genetic Variance				1994 Genetic variance			
	Yield	Height	Maturity	Lodging	Yield	Height	Maturity	Lodging
RFD	-0.38	-0.16	-0.20	-0.06	-0.62*	-0.18	-0.40	-0.25
	< .09	< .47	< .37	< .80	< .02	< .55	< .17	< .39
RPD	-0.17	-0.30	-0.06	0.04	-0.56*	-0.15	-0.04	0.06
	< .45	< .17	< .79	< .86	< .04	< .62	< .88	< .86
CMD	-0.40	-0.27	-0.16	-0.07	-0.74**	-0.20	-0.27	-0.20
	< .07	< .23	< .49	< .78	< .003	< .51	< .36	< .50
GD	0.09	0.07	0.34	0.16	-0.35	0.05	0.22	0.48
	< .69	< .78	< .13	< .48	< .22	< .88	< .46	< .08

*,** Significant at the 0.05 and 0.01 levels, respectively. † RFP = RFLP Distance, RPD = RAPD Distance, CMD = Combined RFLP and RAPD Distance, GD = Genealogical Distance (1-Coefficient of Parentage.

Table 2.9 Correlation coefficients and P-values of genetic distance estimates† between parents with genetic variances of several agronomic traits for population set two.

	1994				1995			
	Genetic Variance				Genetic variance			
Genetic Distance	Yield	Height	Maturity	Lodging	Yield	Height	Maturity	Lodging
RFD	0.41*	0.16	0.47*	0.01	0.39	0.25	0.48	0.24
	< .05	< .44	< .02	< .99	< .27	< .49	< .16	< .51
RPD	0.22	0.21	0.36	0.23	0.50	0.48	0.61	0.37
	< .29	< .32	< .08	< .28	< .15	< .17	< .06	< .29
CMD	0.32	0.13	0.38	0.04	0.44	0.34	0.56	0.36
	< .13	< .55	< .07	< .86	< .21	< .34	< .10	< .31
GD	0.10	0.36	0.51**	0.13	0.21	0.12	0.36	0.31
	< .65	< .08	< .009	< .54	< .57	< .74	< .31	< .38

*,** Significant at the 0.05 and 0.01 levels, respectively. † RFP = RFLP Distance, RPD = RAPD Distance, CMD = Combined RFLP and RAPD Distance, GD = Genealogical Distance (1-Coefficient of Parentage.

Table 2.10 Correlation coefficients and P-values of genetic distance estimates† between parents with genetic variances of several agronomic traits for population set two. Population 17 is omitted.

	1994				1995			
	Genetic Variance				Genetic variance			
Genetic Distance	Yield	Height	Maturity	Lodging	Yield	Height	Maturity	Lodging
RFD	0.65**	0.19	0.52**	0.02	0.72*	0.46	0.68	0.14
	< .0001	< .38	< .01	< .93	< .05	< .27	< .07	< .74
RPD	0.17	0.18	0.33	0.20	0.68	0.68	0.72*	0.25
	< .43	< .40	< .12	< .35	< .07	< .07	< .05	< .55
CMD	0.47*	0.13	0.40	0.05	0.75*	0.58	0.77*	0.27
	< .03	< .54	< .06	< .85	< .04	< .14	< .03	< .52
GD	0.35	0.43*	0.62**	0.20	0.70	0.55	0.73*	0.30
	< .10	< .04	< .002	< .37	< .06	< .16	< .04	< .47

*,** Significant at the 0.05 and 0.01 levels, respectively. † RFP = RFLP Distance, RPD = RAPD Distance, CMD = Combined RFLP and RAPD Distance, GD = Genealogical Distance (1-Coefficient of Parentage.

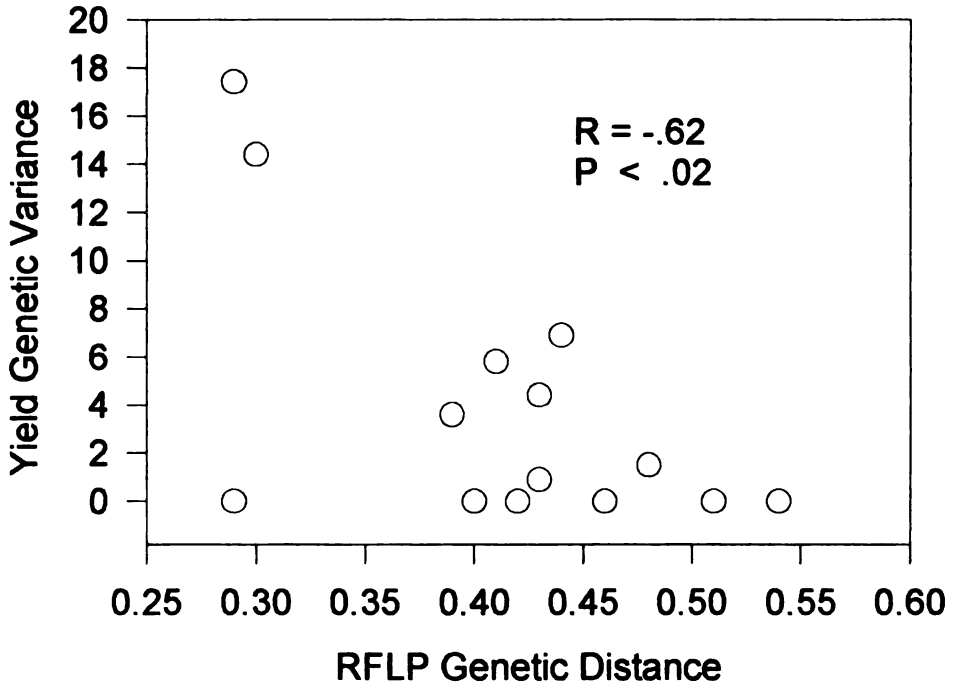

Figure 2.1 Scatterplot of yield genetic variance versus RFLP genetic distance for population set A in the 1994 two-row plots.

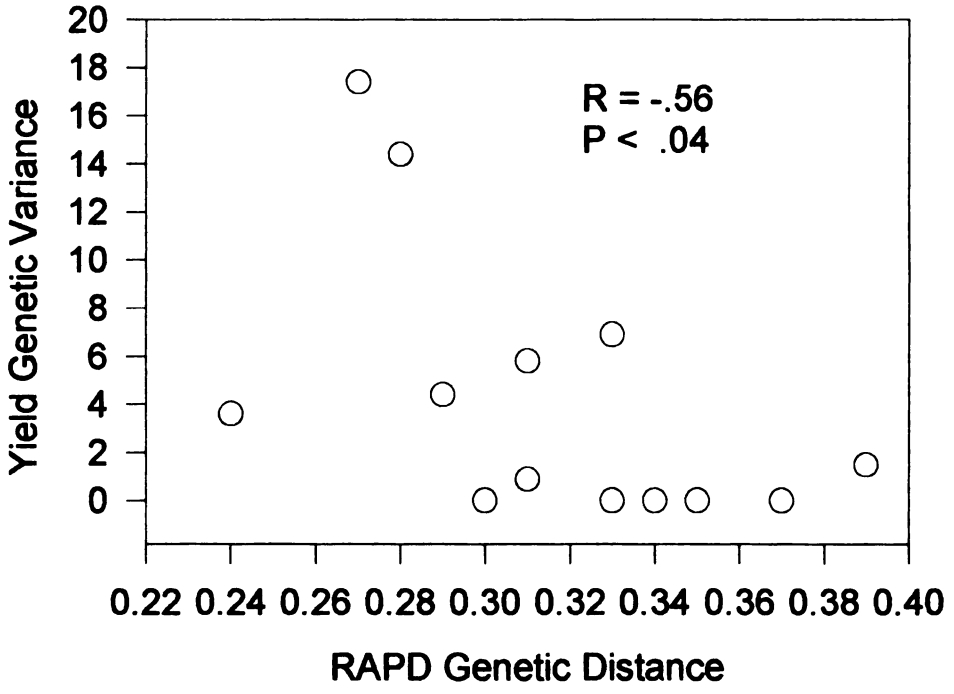

Figure 2.2 Scatterplot of yield genetic variance versus RAPD genetic distance for population set A in the 1994 two-row plots.

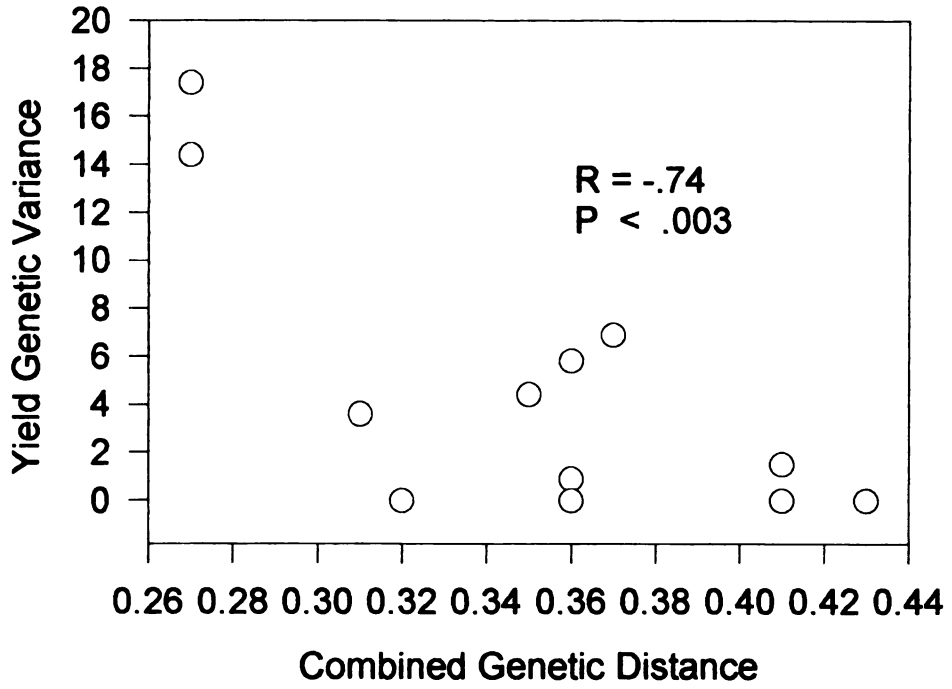

Figure 2.3 Scatterplot of yield genetic variance versus genetic distance from the combined analysis of RFLP and RAPD data for population set A in the 1994 two-row plots.

correlations of yield genetic variance with RFD (Figure 2.4) and CMD (Figure 2.5) were now significant in 1994 and 1995. Maturity genetic variance remained significantly correlated with RFD (Figure 2.6) and GD (Figure 2.7) for the 1994 tests, and was significantly correlated with RPD (Figure 2.8), CMD (Figure 2.9), and GD (Figure 2.10) in the 1995 tests.

While the genetic variance of a population may be dependent on the allelic difference between the two parents, the population mean is usually a function of the parent means. Regression of mean yield of each population with its mid-parent yield was positive and significant for parents of both population sets in 1994 and 1995 two-row plots (Figure 2.11). This relationship was not evaluated for the 1993 and 1994 1-row

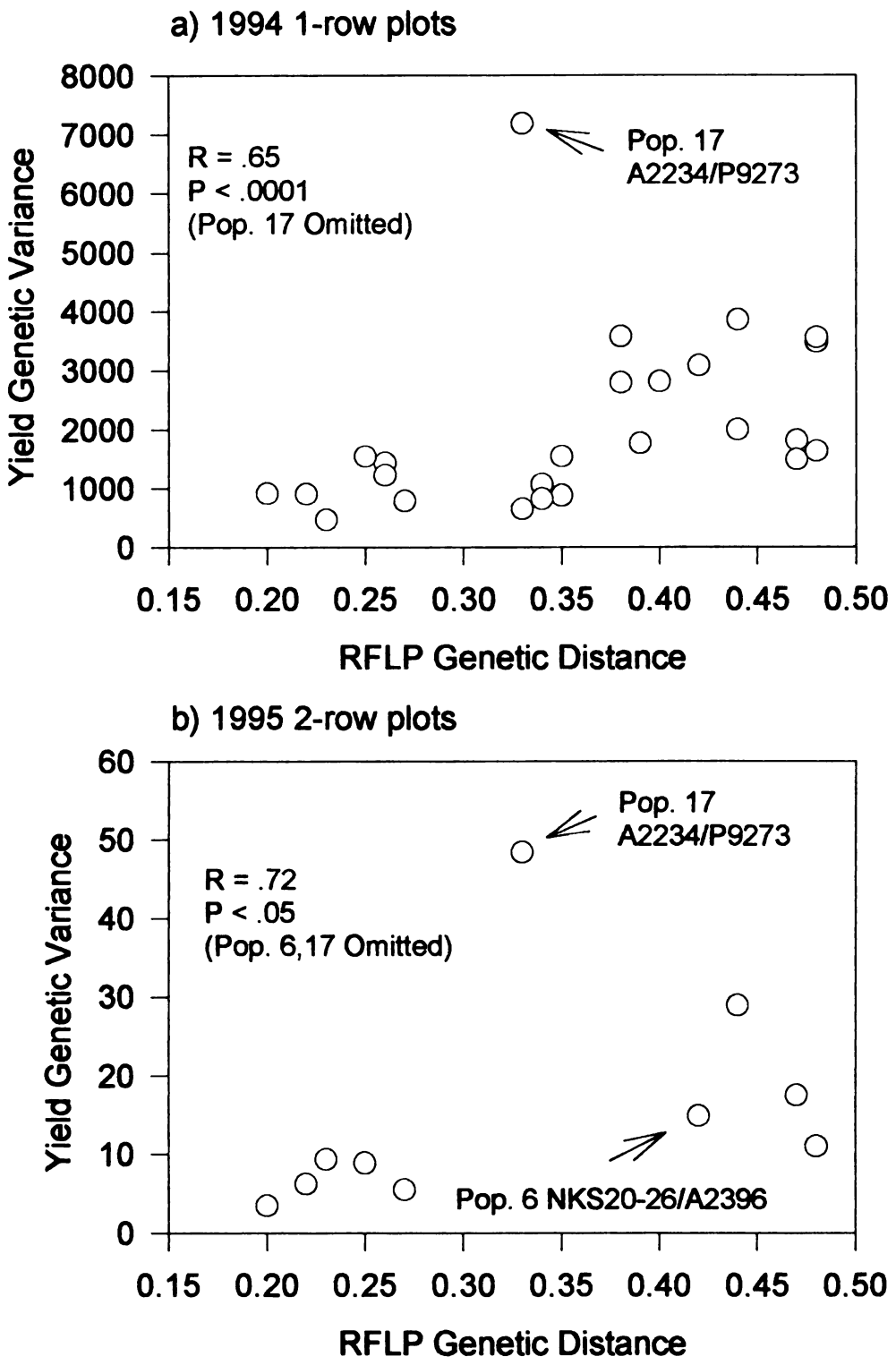

Figure 2.4 Scatterplot of yield genetic variance versus RFLP genetic distance for population set B. a) 1994 single row plots b) 1995 two-row plots.

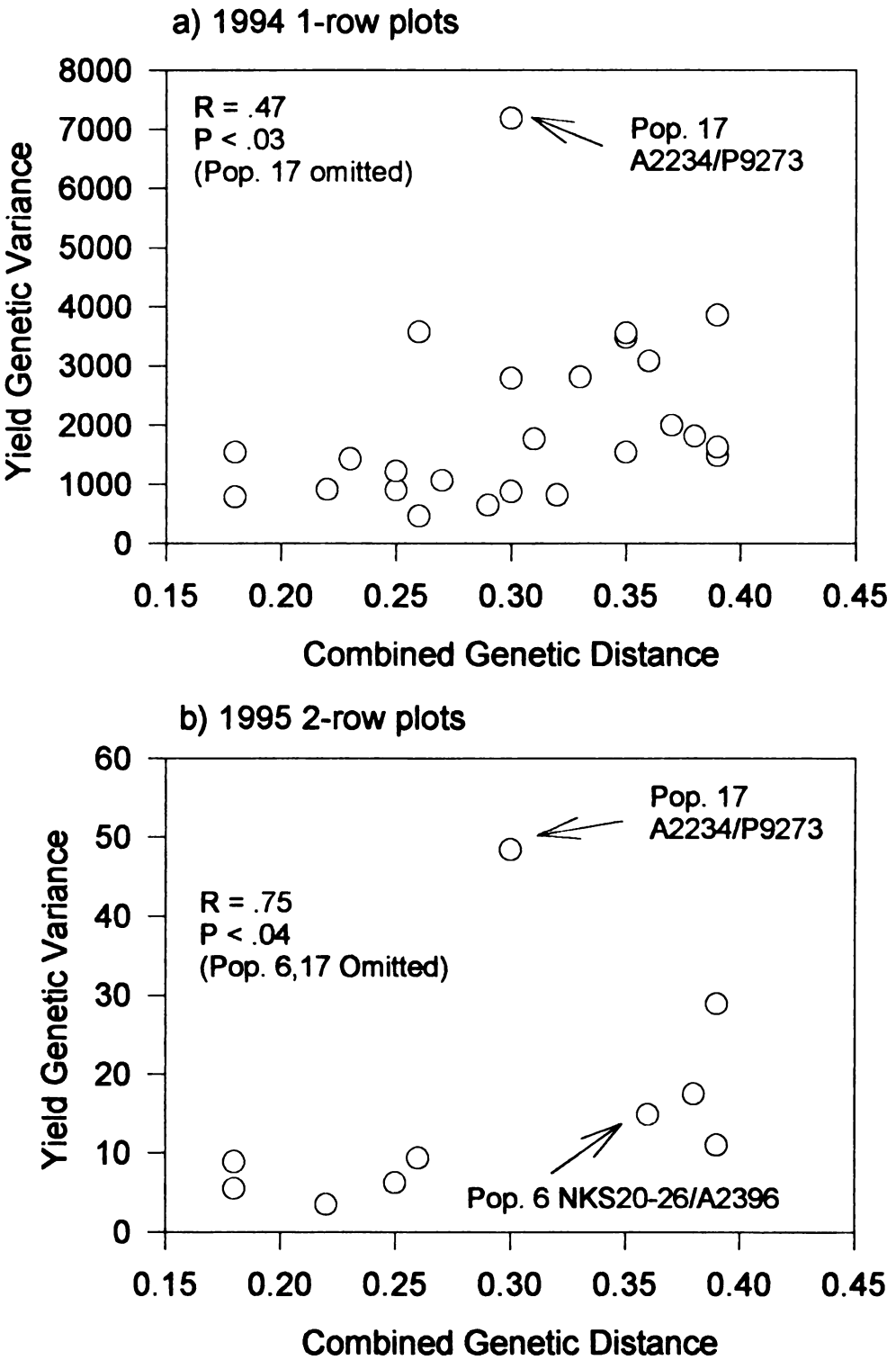

Figure 2.5 Scatterplot of yield genetic variance versus genetic distance from the combined analysis of RFLP and RAPD data for population set B. a) 1994 single-row plots b) 1995 two-row plots

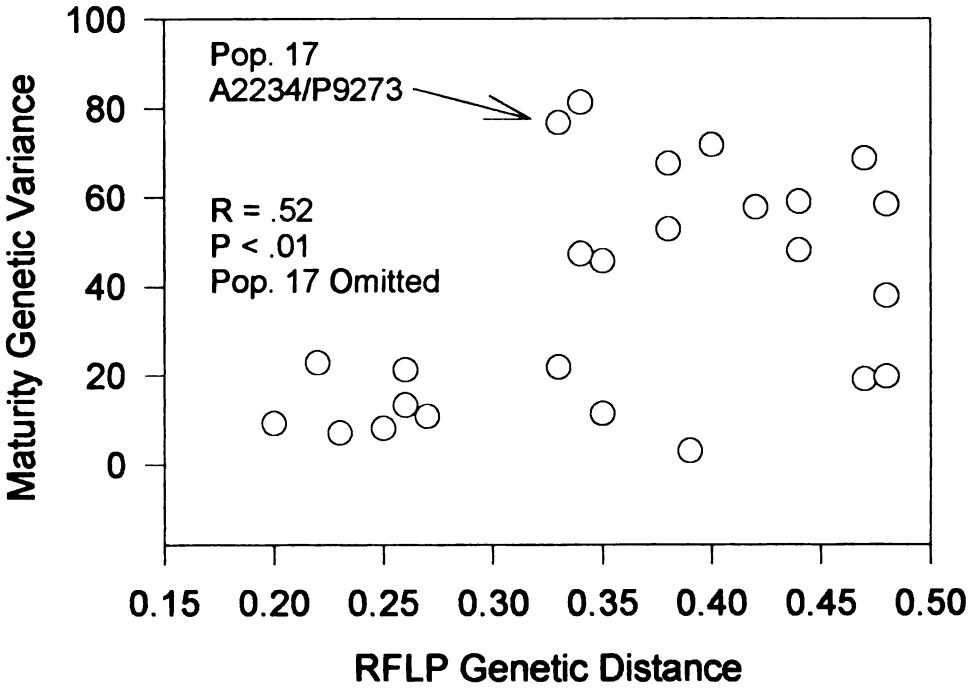

Figure 2.6 Scatterplot of maturity genetic variance versus RFLP genetic distance for population set B in the 1994 single-row plots.

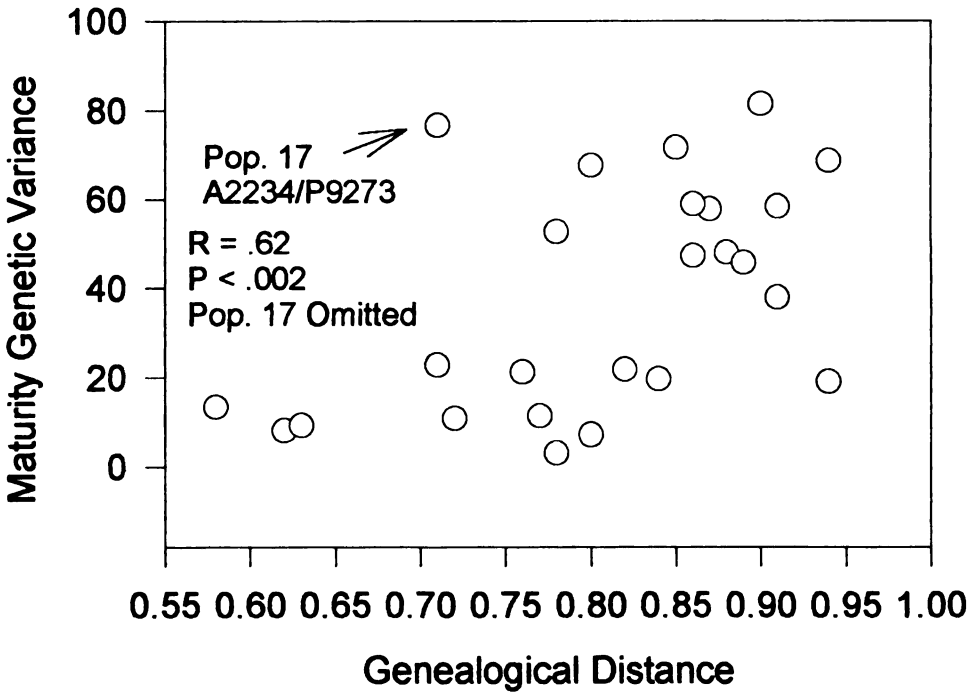

Figure 2.7 Scatterplot of maturity genetic variance versus genealogical distance for population set B in the 1994 single-row plots.

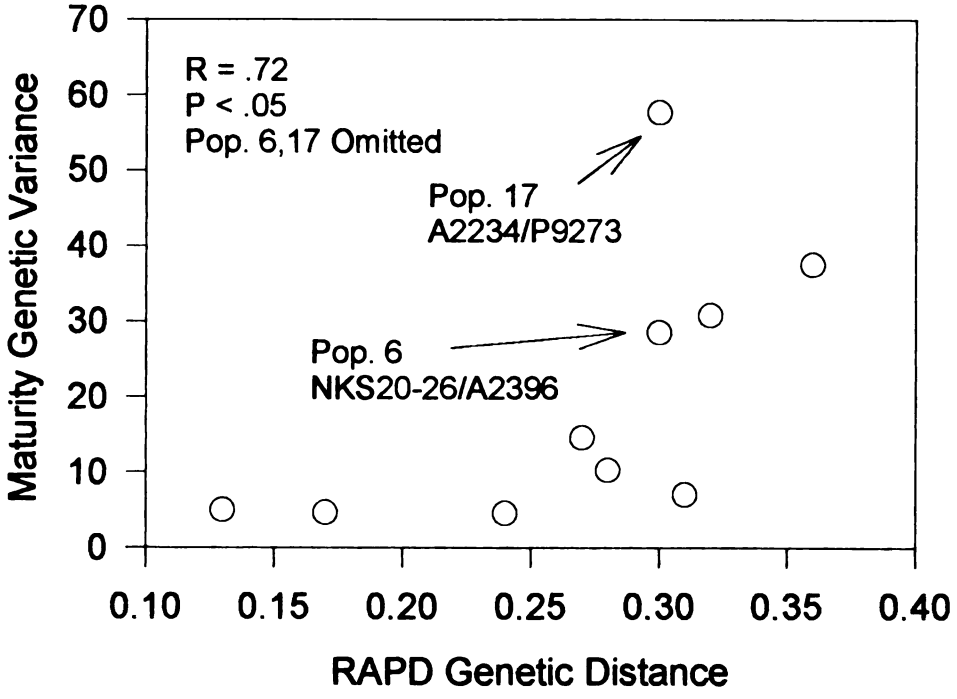

Figure 2.8 Scatterplot of maturity genetic variance versus RAPD genetic distance for population set B in the 1995 two-row plots.

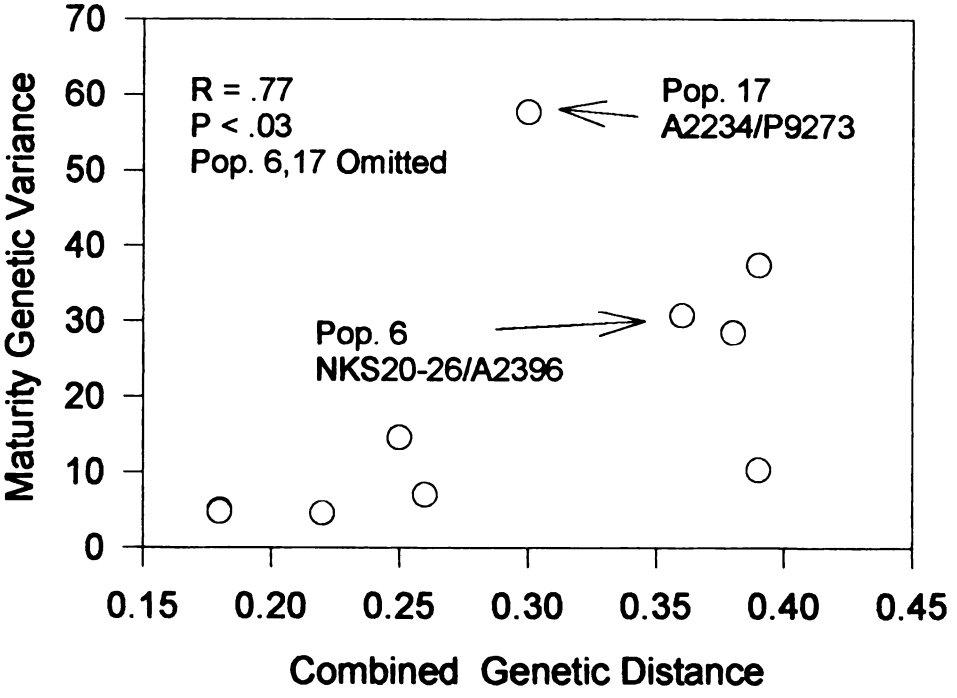

Figure 2.9 Scatterplot of maturity genetic variance versus genetic distance from the combined analysis of RFLP and RAPD data for population set B in the 1995 two-row plots.

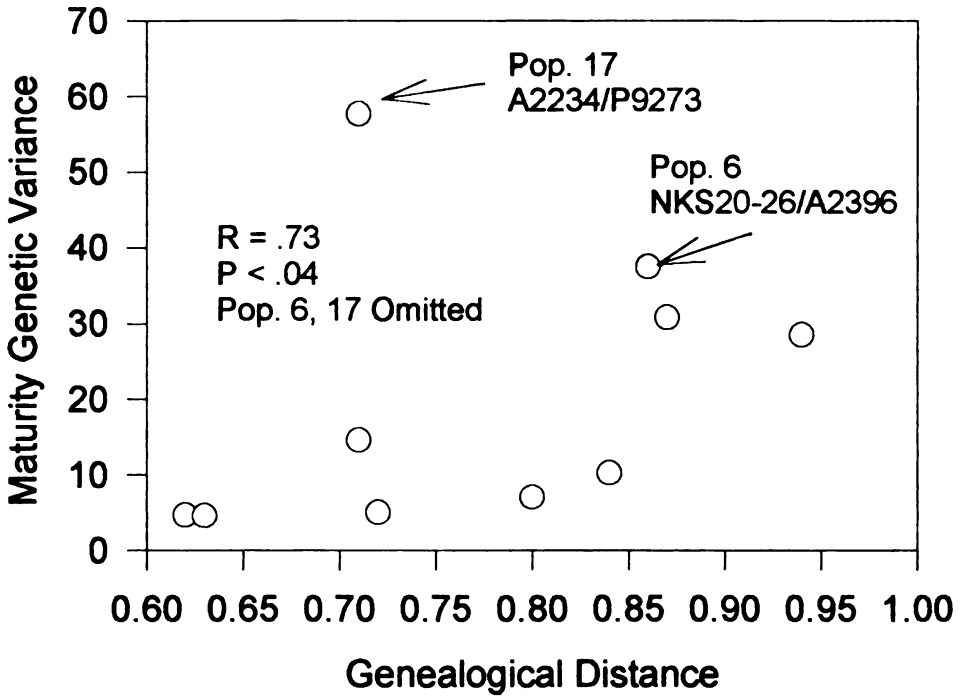

Figure 2.10 Scatterplot of maturity genetic variance versus genealogical distance for population set B in the 1995 two-row plots.

plots because of poor germination of the parent seed. A multiple regression model including both mid-parent yield and genetic distance was tested as a predictor of the mean of the top five yielding progeny (MY5) within each population using the 1994 and 1995 2-row plots. The model was a significant predictor of MY5 for the populations in both years. The relationship between RFD and yield potential was negative for parent set A after the effects of mid-parent yield were removed (Figure 2.12). The combination of RFD and mid-parent yield (Figure 2.13) and the combination of CMD and mid-parent yield (Figure 2.14) both provided a predictive model in which yield potential was directly proportional to both mid-parent yield and genetic distance for population set B.

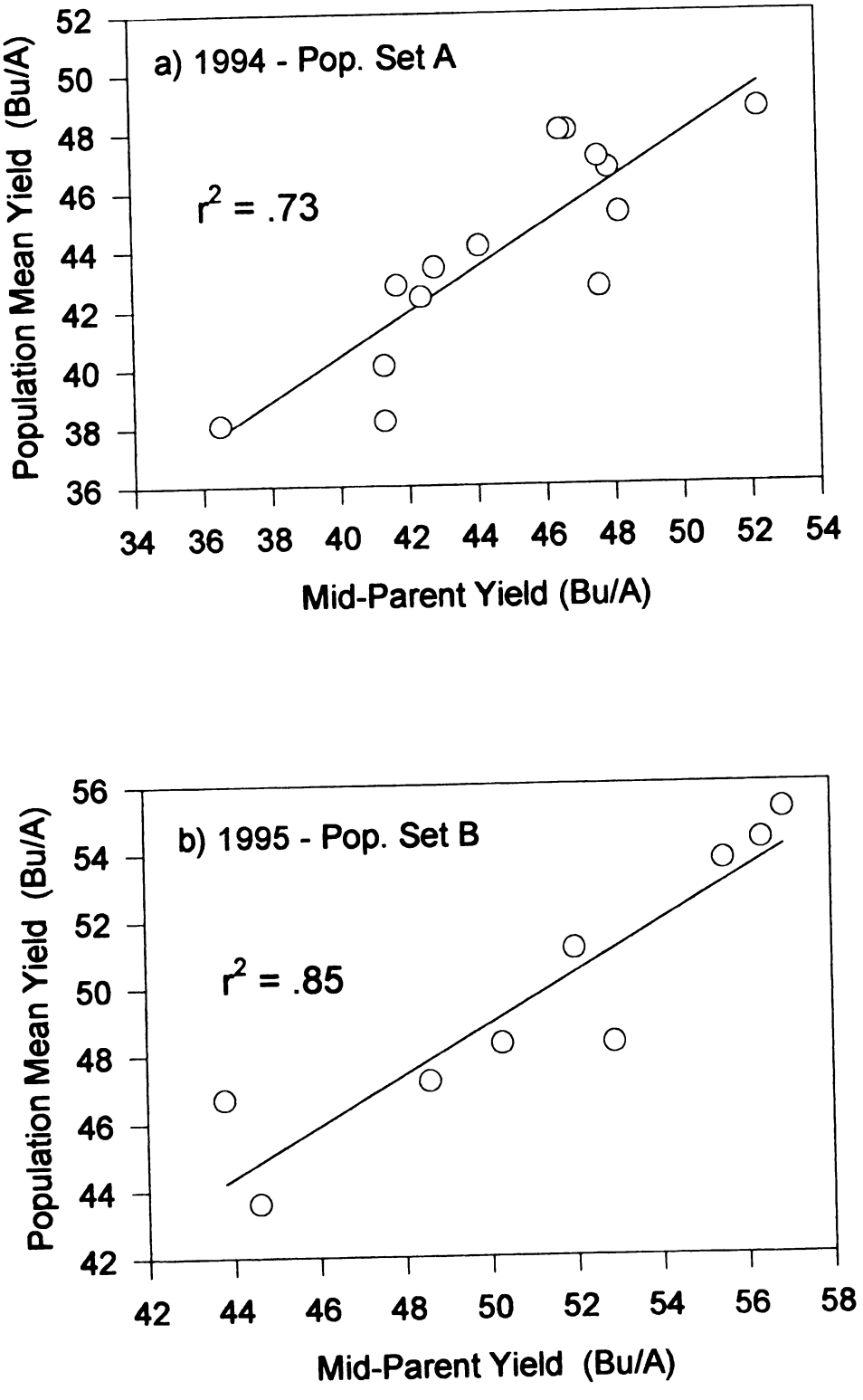

Figure 2.11 Regression of population mean yield on mid-parent yield for the two-row plots. a) Population set A b) populaton set B

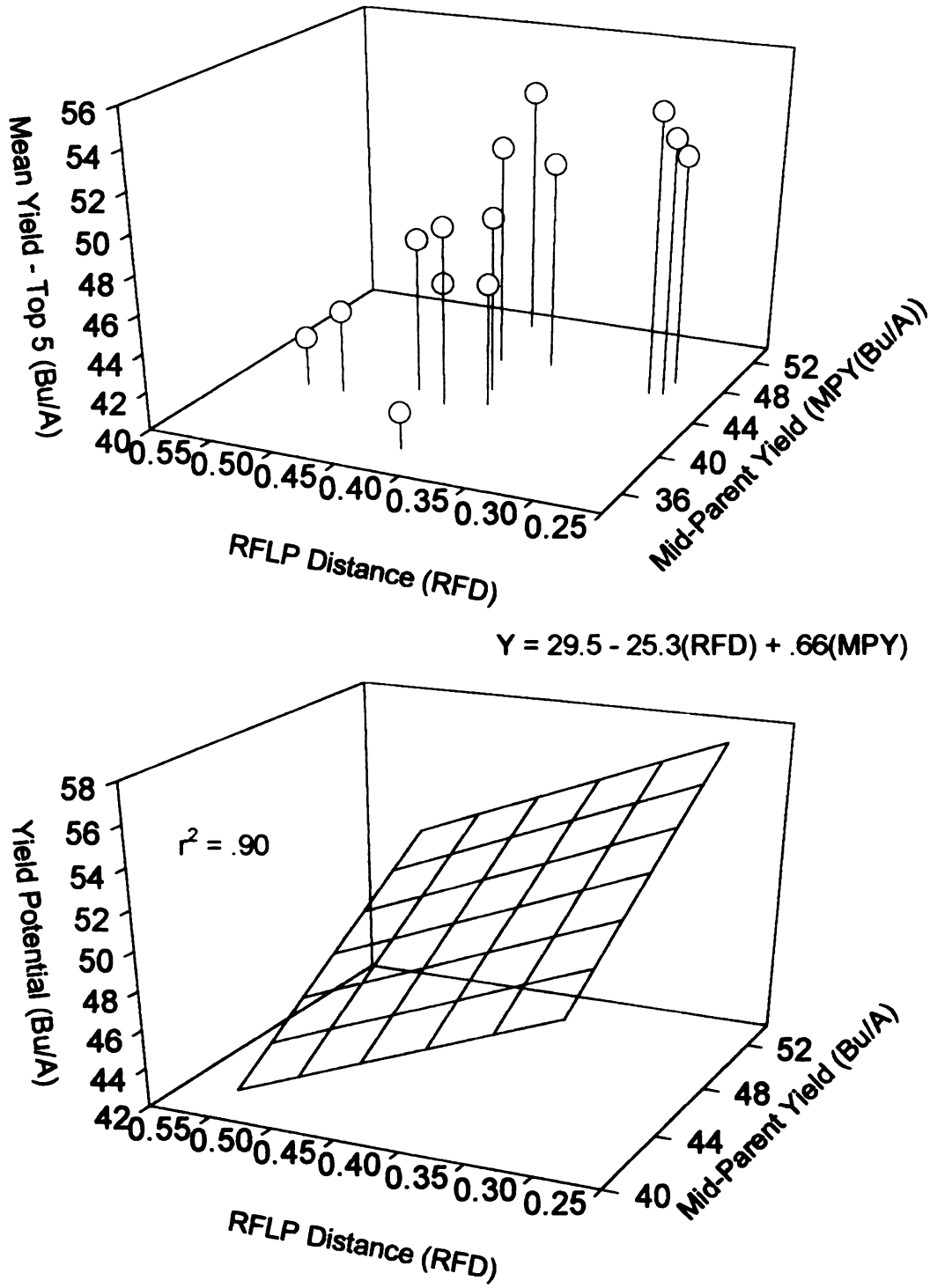

Figure 2.12 Multiple regression model for the prediction of the top five yielding progeny of the 1994 two-row plots as a function of mid-parent yield and RFLP genetic distance.

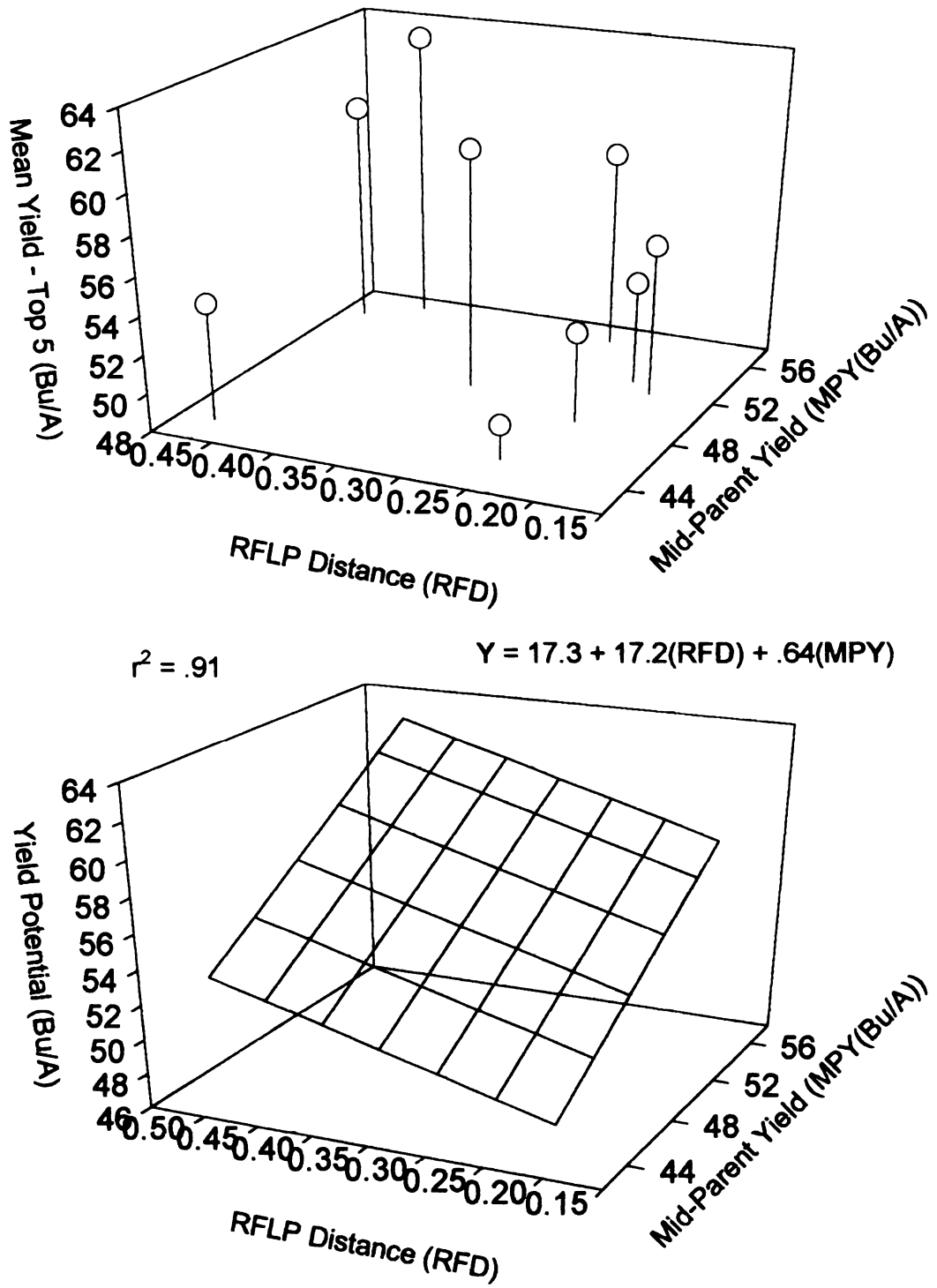

Figure 2.13 Multiple regression model for the prediction of the top five yielding progeny of the 1995 two-row plots as a function of mid-parent yield and RFLP genetic distance.

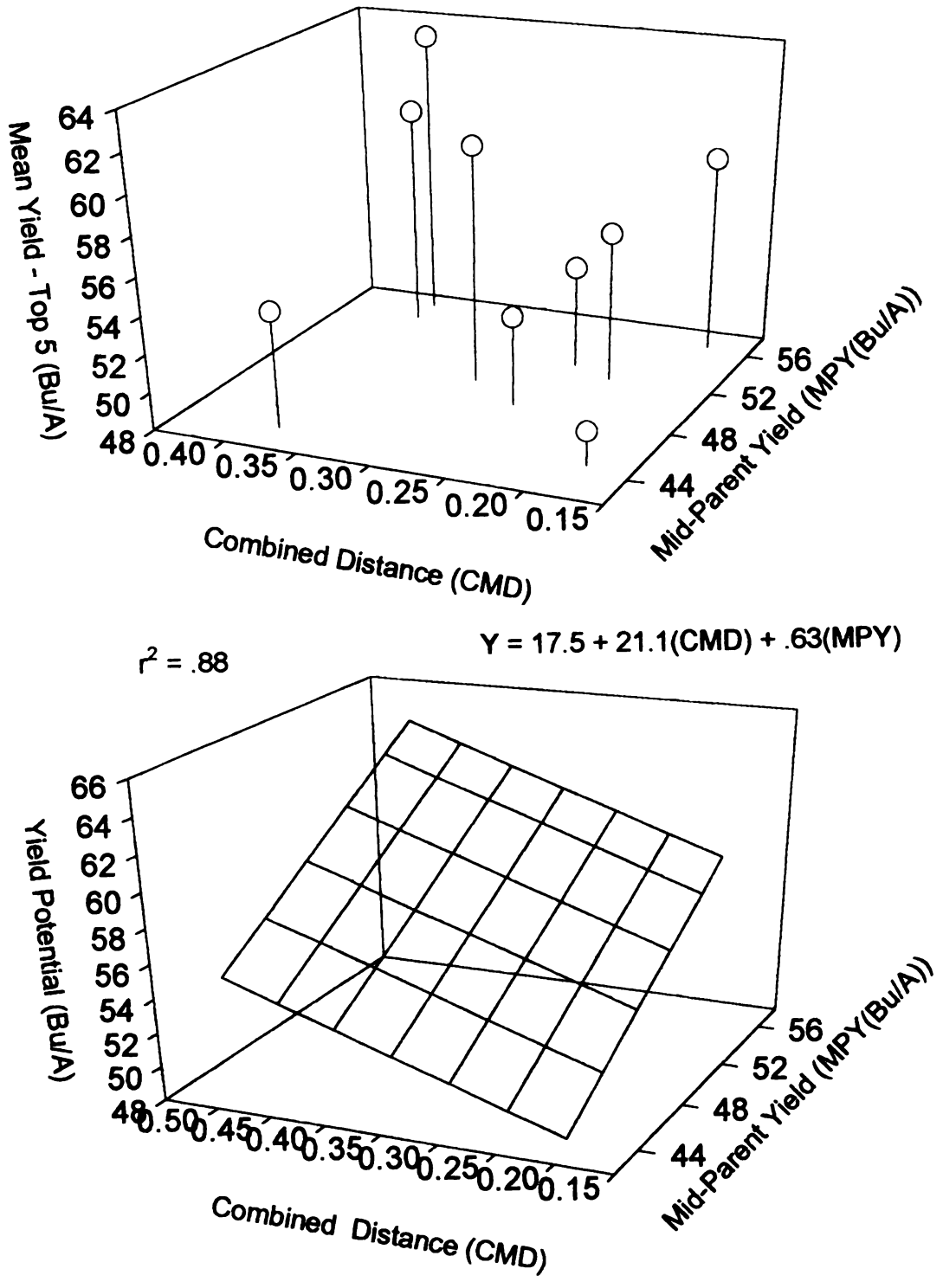

Figure 2.14 Multiple regression model for the prediction of the top five yielding progeny of the 1995 two-row plots as a function of mid-parent yield and genetic distance from the combined anaylsis of RFLP and RAPD data.

DISCUSSION

Genetic distance estimates were significant predictors of genetic variation for set B populations but not for set A populations. Poor estimates of genetic variances for set A populations, especially in 1994, and the use of the parents with 50% plant introduction (PI) in their pedigree for set A are possible explanations for this inconsistency.

The coefficients of variation (CV) from the statistical analyses of population set A and population set B were not greatly different. The CV's averaged 15.6% for population set A in 1993 and 14.4% for population set B in 1994 in the analyses using single row plots. The CV's averaged 11.5% for population set A in 1994 and 10.1% for population set B in 1995 in the analyses using two-row plots.

The standard errors of the yield genetic variance (SEYV) did not differ greatly from one population set to the other. In fact, SEYV were somewhat higher for population set B than for population set A. For the single row plots (yield given in grams per plot), the 1993 SEYV for population set A was 607, while the 1994 SEYV for population set B was 635. However, the corresponding mean yield genetic variance was 1204 for population set A and 2049 for population set B. Only 18 of 22 populations from set A showed significant genetic variance for yield in 1993 (Table 2.3), while all but one of the 25 populations in set 2 showed significant yield genetic variance in 1994 (Table 2.4). The

situation was similar for the two-row plots (yield given in Bu/A). The 1994 SEYV for population set A averaged 4.1, while the 1995 SEYV for population set B averaged 5.3. The corresponding mean yield genetic variance for population set A was only 3.5, while that for population set B was 15.4 (11.4 without population 17). Yield genetic variance was significant for only 5 out of 14 populations in set A in 1994 (Table 2.5); while, for set B, all but population 6, which was severely affected by disease, exhibited significant yield genetic variance in 1995 (Table 2.6). There was no correlation between yield genetic variance from single-row to two-row plots for population set A, but the correlation was high for population set B (Figure 2.15).

Six out of 14 populations in set A exhibited significant genotype by environment variance in 1994 (GEV), while the only two populations from set B that had significant GEV in 1995 were population 17, whose yield genetic variance was almost twice that of the other populations in set B, and population 6, which was severely affected by disease at one of the two locations. The genotype by environment interactions may have reduced the accuracy of the genetic variance estimates.

The inclusion of unadapted germplasm in the pedigrees of the population sets tested may have reduced performance among the progeny. Six of the 14 populations tested in two-row plots from parent set A contained 25% plant introduction germ plasm in their pedigree, while only 2 out of 10 row plot populations from parent set 2 contained 25% plant introduction germ plasm. Schoener and Fehr (1979) showed that as little as 25% plant introduction germ plasm within a population can significantly reduce population performance. Souza and Sorrells (1991)

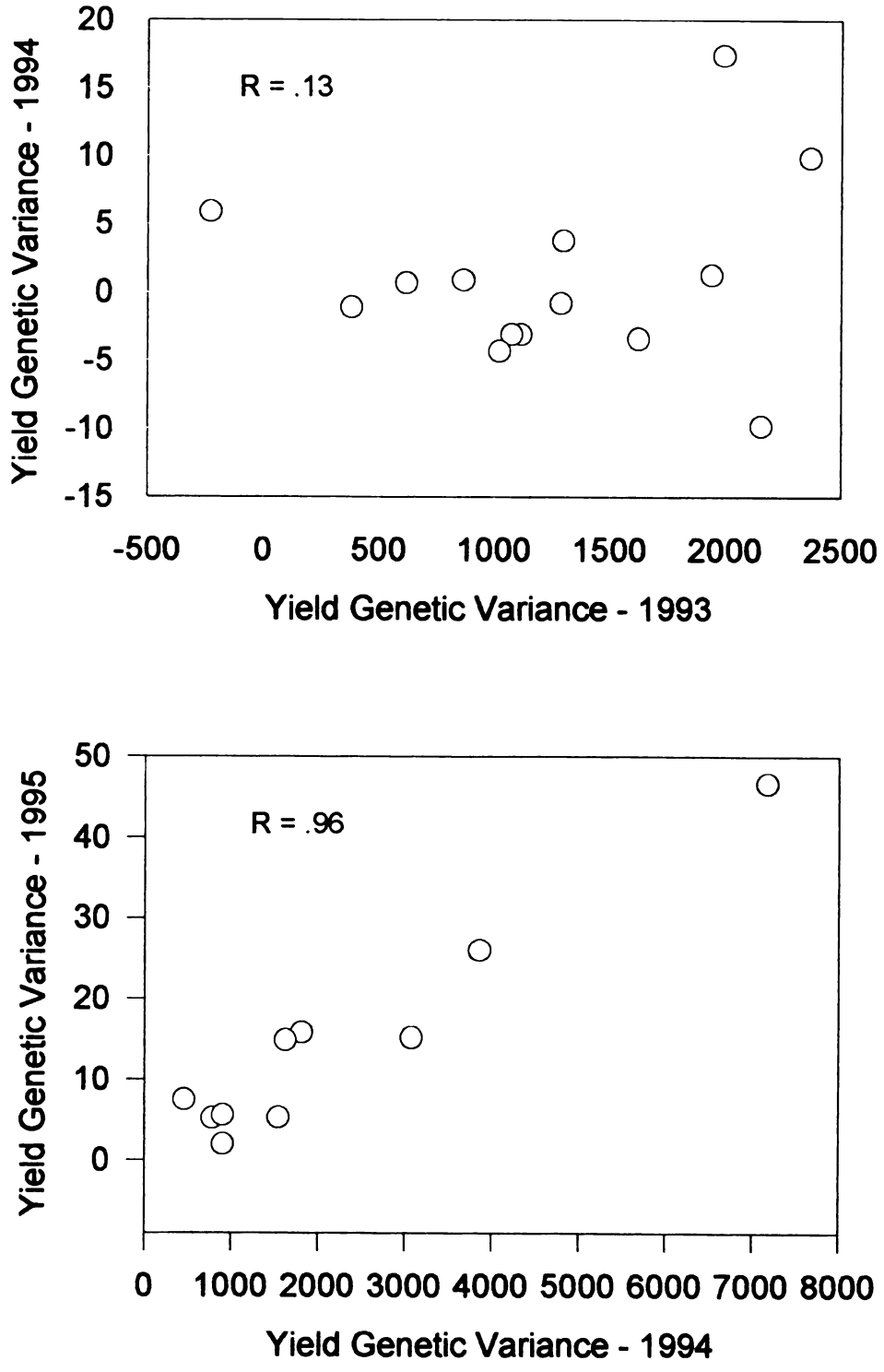

Figure 2.15 Correlation between yield genetic variance from the single-row plots to the two-row plots. a) Population set A. b) Population set B.

showed declining variability in progeny with an increase in parent genetic distance for grain yield, test weight, heading date, and maturity date in oat. Populations in that study were a result of an adapted parent crossed with an unadapted parent. The greater the genetic distance between the two parents, the less well adapted one of the parents was. Variance for biomass was positively correlated with parent genetic distance, but the effects of distance lessened as large distances were approached. They suggested that plant biomass may not have been as environmentally sensitive as the other traits. While the use of PI germ plasm in the pedigree of parents of population set A may have been a contributing factor to high GEV and less precise estimates of yield genetic variance, it should be noted that set B populations containing 25-50% PI germ plasm did quite well in 1994 and 1995. In 1995, both set B populations containing PI germ plasm had significant yield genetic variance, while neither had significant GEV. One of these populations (23) was a cross between two parents with 50% PI germ plasm.

The mean yields of populations in set A averaged 347g in the 1993 single row plots and 44.0 Bu/A in the 1994 two-row plots, while those of population set B averaged 356g in the 1994 single row plots and 49.6 Bu/A in the 1995 two-row plots. Because these means are from separate years, the two population sets cannot be directly compared. However, the lower yield potential exhibited by population set A in 1994, whether from environmental or genetic causes, may have been a contributing factor to the inability to accurately determine yield genetic variance in that year.

The range of genealogical distances of the parents in set B was 0.94 to 0.62., while the range for parent set A was 0.95 to 0.73.

Genealogically distant parents share a greater proportion of alleles alike in state than parents with a close genealogical relationship, whose alleles are primarily identical by descent. The effects of identity by descent versus alikeness in state on genetic distance are unknown. Melchinger et al. (1990) found that there was no significant relationship between parent genetic distance and hybrid performance when only crosses between unrelated lines were considered. Marker distance among unrelated lines is based entirely on alleles alike in state.

Field conditions for the 1994 two-row plots were less than ideal. One location suffered early drought, causing uneven germination, followed by heavy rains and hail. The other location had standing water for a period of several days, and later exhibited substantial levels of brown stem rot (*Phialophora gregata*). These conditions likely contributed to genotype by environment interactions within the populations, further reducing the amount of genetic variance calculated among lines within the populations.

The population 'A2234' by 'P9273' (No. 17, Set B) exhibited yield genetic variance disproportionate to the marker distance between the two parents. This may have been because more of the markers were linked to alleles of quantitative trait loci for yield that were different between the two parents. It may also have been the result of greater epistatic variation in this population than the others.

Other factors which may have contributed to lower precision of genetic variance estimates in population set A compared with population set B were the degree of inbreeding and the difference in progeny number used in the two population sets. The populations developed from parent set A were F_3-derived, while populations from parent set B were F_4-

derived. This would have resulted in a 14% smaller ratio of among-line:within-line variance in parent set A populations compared to parent set B populations (Falconer, 1989). Set B populations each contained 48 progeny, while set A populations contained only 28 progeny. The lower number of entries in set A populations should have resulted in greater error in the variance estimates.

CONCLUSIONS

Our data suggests that marker genetic distance estimates can assist soybean breeders in choosing parents which will increase the probability of transgressive segregation for yield in their progeny. Population set B exhibited significant yield genetic variance within almost every population, and this variance was significantly correlated between years. Genetic distance from RFLP markers was positively and significantly correlated with yield genetic variance in both years, while genealogical distance was not. Marker data alone, however, will not take the place of accuracy in field testing of both putative parents and progeny. A multiple regression model based on RFLP or CMD marker distance and performance data of parents was able to predict which populations had the highest yielding progeny across a wide range of marker distances and mid-parent yields for parents of population set 2. Strict adherence to this model, however, may exclude some parent combinations whose specific combining ability will exceed expectations based on the data. Population 17 of set B had the highest yield genetic variance of all the populations in that set, yet its parents' RFD was lower than the average RFD between parents of populations in set B.

GENERAL CONCLUSIONS

CONCLUSIONS

Despite the limited genetic base in soybean germ plasm in the northern United States, there was enough RFLP diversity to distinguish among cultivars and lines and establish genetic relationships. Ancestral soybean introductions were clustered according to area of origin, which indicates that shared alleles are likely identical by descent. Genetic distances calculated using sufficient RFLP markers are probably more accurate than those taken from pedigree relationships, since marker data can account for selection practiced in breeding programs. There was little diversity lost within modern germ plasm compared with the ancestors, and no new alleles were found unique to a set of selected newer plant introductions.

Genetic distance between parents was generally positively correlated with progeny genetic variance among lines with good yield potential. In a population set with lower yields, whether due to environmental conditions or limited genetic yield potential, correlations were low and sometimes negative. A multiple regression approach using RFLP genetic distance and mid-parent yield to predict the highest yielding progeny shows promise, but was not consistent between the two population sets examined. More data is required before a general conclusion can be drawn, but the data presented here suggests that genetic distance estimates based on markers, especially those obtained with RFLP markers, can assist soybean breeders in choosing parents with

the greatest probability of producing transgressive segregates for yield.

APPENDIX

Table A1. Allele frequencies and polymorphism information content (PIC[§]) per locus for clone/enzyme combinations among all lines and cultivars or within groups.

Clone	Allele[†]	N[¶]	Freq	PIC[§]	N[¶]	Freq	PIC[§]	N[¶]	Freq	PIC[§]	N[¶]	Freq	PIC[§]
		All Lines			Ancestral			Elite			PI's		
A059	A1	80	0.01	0.54	12	0.00	0.59	55	0.00	0.49	13	0.08	0.57
	A2	80	0.56		12	0.58		55	0.60		13	0.46	
	A3	80	0.41		12	0.34		55	0.40		13	0.46	
	A4	80	0.01		12	0.08		55	0.00		13	0.00	
A063	A1	76	0.60	0.49	12	0.50	0.50	52	0.74	0.38	12	0.08	0.15
	A2	76	0.40		12	0.50		52	0.26		12	0.92	
	B1	76	0.83	0.36	12	0.75	0.50	52	0.89	0.20	12	0.58	0.49
	B2	76	0.17		12	0.25		52	0.11		12	0.42	
A064	A1	80	0.52	0.50	12	0.50	0.48	55	0.42	0.49	13	1.00	0.00
	A2	80	0.48		12	0.50		55	0.58		13	0.00	
A073	A1	77	0.63	0.46	12	0.67	0.42	52	0.70	0.42	13	0.31	0.43
	A2	77	0.37		12	0.33		52	0.30		13	0.69	
A077	A1	80	0.27	0.41	12	0.25	0.46	55	0.25	0.38	13	0.38	0.47
	A2	80	0.73		12	0.75		55	0.75		13	0.62	
A085	A1	80	0.56	0.49	12	0.25	0.48	55	0.64	0.46	13	0.54	0.50
	A2	80	0.44		12	0.75		55	0.36		13	0.46	
A086	A1	80	0.19	0.31	12	0.08	0.32	55	0.24	0.36	13	0.00	0.00
	A2	80	0.81		12	0.92		55	0.76		13	1.00	

Table A1. (Cont'd)

A089	A1	79	0.32	0.46	11	0.50	0.48	55	0.25	0.38	13	0.46	0.50
	A2	79	0.68		11	0.50		55	0.75		13	0.54	
	B1	79	0.14	0.24	11	0.18	0.27	55	0.16	0.27	13	0.00	0.00
	B2	79	0.86		11	0.82		55	0.84		13	1.00	
A095	A1	80	0.64	0.44	12	0.71	0.28	55	0.58	0.49	13	0.85	0.26
	A2	80	0.36		12	0.29		55	0.42		13	0.15	
A096	A1	75	0.21	0.33	12	0.08	0.16	50	0.13	0.23	13	0.61	0.47
	A2	75	0.79		12	0.92		50	0.87		13	0.39	
A111	A1	73	0.58	0.50	12	0.58	0.46	50	0.64	0.46	12	0.25	0.38
	A2	73	0.42		12	0.42		50	0.36		12	0.75	
	B1	73	0.07	0.34	12	0.17	0.26	50	0.06	0.11	12	0.00	0.00
	B2	73	0.93		12	0.83		50	0.94		12	1.00	
A124	A1	79	0.28	0.40	12	0.17	0.32	54	0.35	0.46	13	0.08	0.15
	A2	79	0.72		12	0.83		54	0.65		13	0.92	
A131	A1	79	0.78	0.34	11	0.64	0.48	55	0.80	0.32	13	0.85	0.26
	A2	79	0.22		11	0.36		55	0.20		13	0.15	
A135	A1	74	0.05	0.33	12	0.08	0.18	50	0.06	0.11	12	0.00	0.00
	A2	74	0.95		12	0.92		50	0.94		12	1.00	
A176	A1	73	0.19	0.35	11	0.27	0.49	50	0.14	0.24	12	0.33	0.44
	A2	73	0.81		11	0.73		50	0.86		12	0.67	
A186	A1	80	0.83	0.27	12	0.58	0.42	55	0.88	0.21	13	0.85	0.26
	A2	80	0.17		12	0.42		55	0.12		13	0.15	

Table A1. (Cont'd)

A203	A1	77	0.47	0.50	10	0.65	0.48	54	0.48	0.50	13	0.31	0.43
	A2	77	0.53		10	0.35		54	0.52		13	0.69	
A233	A1	80	0.47	0.50	12	0.50	0.50	55	0.46	0.50	13	0.46	0.50
	A2	80	0.53		12	0.50		55	0.54		13	0.54	
A235	A1	80	0.68	0.45	12	0.58	0.50	55	0.69	0.43	13	0.69	0.43
	A2	80	0.32		12	0.42		55	0.31		13	0.31	
A264	A1	80	0.39	0.47	12	0.42	0.46	55	0.47	0.50	13	0.00	0.00
	A2	80	0.61		12	0.58		55	0.53		13	1.00	
A378	A1	80	0.55	0.50	12	0.58	0.50	55	0.56	0.49	13	0.46	0.50
	A2	80	0.45		12	0.42		55	0.44		13	0.54	
A398	A1	74	0.72	0.41	12	0.75	0.42	50	0.71	0.41	12	0.75	0.38
	A2	74	0.28		12	0.25		50	0.29		12	0.25	
A487	A1	74	0.14	0.23	12	0.33	0.38	50	0.06	0.11	12	0.25	0.38
	A2	74	0.86		12	0.67		50	0.94		12	0.75	
A493	A1	79	0.65	0.49	12	0.33	0.32	55	0.82	0.32	12	0.17	0.28
	A2	79	0.35		12	0.67		55	0.18		12	0.83	
A505	A1	74	0.49	0.50	12	0.50	0.50	50	0.42	0.49	12	0.75	0.38
	A2	74	0.51		12	0.50		50	0.58		12	0.25	
A517	A1	73	0.47	0.50	11	0.36	0.46	50	0.40	0.48	12	0.83	0.28
	A2	73	0.53		11	0.64		50	0.60		12	0.17	

Table A1. (Cont'd)

A586	A1	80	0.48	0.50	12	0.50	0.50	55	0.60	0.48	13	0.00	0.00
	A2	80	0.52		12	0.50		55	0.40		13	1.00	
	B1	80	0.40	0.48	12	0.38	0.47	55	0.41	0.48	13	0.39	0.47
	B2	80	0.60		12	0.62		55	0.59		13	0.61	
A635	A1	74	0.55	0.50	12	0.33	0.42	50	0.71	0.41	12	0.08	0.15
	A2	74	0.45		12	0.67		50	0.29		12	0.92	
A681	A1	80	0.44	0.49	12	0.50	0.50	55	0.43	0.49	13	0.46	0.50
	A2	80	0.56		12	0.50		55	0.57		13	0.54	
A708	A1	74	0.73	0.42	12	0.42	0.48	51	0.84	0.27	12	0.58	0.49
	A2	74	0.27		12	0.58		51	0.16		12	0.42	
A715	A1	75	0.61	0.48	12	0.83	0.38	50	0.55	0.50	13	0.62	0.47
	A2	75	0.39		12	0.17		50	0.45		13	0.38	
A847	A1	80	0.61	0.48	12	0.17	0.42	55	0.79	0.33	13	0.15	0.26
	A2	80	0.39		12	0.83		55	0.21		13	0.85	
A947	A1	72	0.57	0.49	12	0.33	0.46	47	0.60	0.48	13	0.69	0.43
	A2	72	0.43		12	0.67		47	0.40		13	0.31	
B030	A1	80	0.64	0.48	12	0.58	0.50	55	0.72	0.44	13	0.38	0.47
	A2	80	0.36		12	0.42		55	0.28		13	0.62	
B032	A1	80	0.66	0.43	12	0.83	0.18	55	0.54	0.50	13	1.00	0.00
	A2	80	0.34		12	0.17		55	0.46		13	0.00	
B046	A1	76	0.57	0.49	12	0.46	0.50	51	0.52	0.50	13	0.85	0.26
	A2	76	0.43		12	0.54		51	0.48		13	0.15	

Table A1. (Cont'd)

B142	A1	74	0.34	0.54	12	0.17	0.61	49	0.41	0.48	13	0.23	0.54
	A2	74	0.59		12	0.58		49	0.59		13	0.62	
	A3	74	0.07		12	0.25		49	0.00		13	0.15	
	B1	74	0.78	0.34	12	1.00	0.00	49	0.78	0.34	13	1.00	0.00
	B2	74	0.22		12	0.00		49	0.22		13	0.00	
K003	A1	80	0.58	0.48	12	0.50	0.50	55	0.56	0.49	13	0.69	0.43
	A2	80	0.42		12	0.50		55	0.44		13	0.31	
K007	A1	79	0.85	0.24	12	0.92	0.18	54	0.86	0.24	13	0.77	0.35
	A2	79	0.15		12	0.08		54	0.14		13	0.23	
	B1	79	0.48	0.49	12	0.25	0.26	54	0.63	0.47	13	0.08	0.15
	B2	79	0.52		12	0.75		54	0.37		13	0.92	
K014	A1	80	0.98	0.06	12	1.00	0.00	55	0.98	0.04	13	0.92	0.15
	A2	80	0.02		12	0.00		55	0.02		13	0.08	
	B1	80	0.42	0.49	12	0.25	0.46	55	0.40	0.48	13	0.69	0.43
	B2	80	0.58		12	0.75		55	0.60		13	0.31	
K069	A1	80	0.06	0.16	12	0.17	0.38	55	0.05	0.10	13	0.00	0.00
	A2	80	0.94		12	0.83		55	0.95		13	1.00	
K258	A1	80	0.24	0.09	12	0.34	0.56	55	0.22	0.34	13	0.15	0.26
	A2	80	0.75		12	0.58		55	0.78		13	0.85	
	A3	80	0.01		12	0.08		55	0.00		13	0.00	
K385	A1	80	0.34	0.47	12	0.50	0.48	55	0.28	0.40	13	0.46	0.50
	A2	80	0.66		12	0.50		55	0.72		13	0.54	
K400	A1	73	0.90	0.16	11	0.91	0.10	50	0.88	0.21	12	1.00	0.00
	A2	73	0.10		11	0.09		50	0.12		12	0.00	

Table A1. (Cont'd)

K411	A1	78	0.59	0.34	10	0.80	0.18	55	0.69	0.43	13	0.00	0.00
	A2	78	0.41		10	0.20		55	0.31		13	1.00	
K443	A1	80	0.15	0.28	12	0.00	0.00	55	0.13	0.23	13	0.39	0.47
	A2	80	0.85		12	1.00		55	0.87		13	0.61	
R013	A1	73	0.82	0.28	11	0.73	0.27	50	0.81	0.31	12	0.92	0.15
	A2	73	0.18		11	0.27		50	0.19		12	0.08	
R017	A1	74	0.73	0.39	11	0.91	0.16	50	0.76	0.36	13	0.46	0.50
	A2	74	0.27		11	0.09		50	0.24		13	0.54	
R092	A1	74	0.83	0.30	12	0.50	0.54	50	0.87	0.23	12	1.00	0.00
	A2	74	0.14		12	0.33		50	0.13		12	0.00	
	A3	74	0.03		12	0.17		50	0.00		12	0.00	
R201	A1	74	0.48	0.50	11	0.55	0.47	50	0.47	0.50	13	0.46	0.50
	A2	74	0.52		11	0.45		50	0.53		13	0.54	
Mean				0.39			0.39			0.37			0.29

† Molecular markers are designated a#, where a is the locus and # is the allele at that locus. The highest molecular weight marker is assigned the number one, and the markers allelic to it are assigned 2,3,etc. in the order of decreasing molecular weight.
¶ N is the number of lines in a group for which the clone was scored. The progeny of the crosses Williams by Essex and Williams by Ransom were not included in the computation of allele frequencies. Marker data for some clones is missing for some lines. § PIC = $1-\Sigma p_{ij}^2$, where p is the frequency of the jth allele at the ith locus.

LIST OF REFERENCES

LIST OF REFERENCES

Anderson, J. A., G. A. Churchill, J. E. Autrique, S. D. Tanksley, and M. E. Sorrells. 1993. Optimizing parental selection for genetic linkage maps. Genome. 36:181-186.

Bernard, R. L., G. A. Juvik, R. L. Nelson. 1987a. USDA soybean germplasm collection survey. Vol. I. International Soybean Program. Urbana, IL 80pp.

Bernard, R. L., G. A. Juvik, R. L. Nelson. 1987b. USDA soybean germplasm collection survey. Vol. II. International Soybean Program. Urbana, IL 203pp.

Borst, P. and D. R. Greaves. 1987. Programmed gene rearrangements altering gene expression. Science. 235:658-667.

Carter, Jr., T. E., Z. Gizlice, and J. W. Burton. 1993. Coefficient-of-parentage and genetic-similarity estimates for 258 North American soybean cultivars released by public agencies during 1945-1988. U. S. Department of Agriculture, Technical Bulletin No. 1814, 169 pp.

Cervantes, T., M. M. Goodman, E. Casas, and J. O. Rawlings. 1978. Use of genetic effects and genotype by environment interactions for the classification of Mexican races of maize. Genetics. 90:339-348.

Cowen, N. M. and K. J. Frey. 1987a. Relationship between genealogical distance and breeding behavior in oats (*Avena sativa* L.). Euphytica. 36:413-424.

Cowen, N. M. and K. J. Frey. 1987b. Relationships between three measures of genetic distance and breeding behavior in oats (*Avena sativa* L.). Genome. 29:97-106.

Committee on Genetic Vulnerability of Crops. 1972. Genetic vulnerability of major crops. Natl. Acad. Sci. Washington D.C.

Cooper, R. L. 1990. Modified early generation testing procedure for yield selection in soybean. Crop Sci. 30(2):417-419.

Cooper, R. L. and R. J. Martin. 1981. Registration of Gnome soybean. Crop Sci. 21:634.

Cox, T. S., Y. T. Kiang, M. B. Gorman, and D. M. Rodgers. 1985. Relationship between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci. 25:529-532.

Damerval, C., Y. Hébert, and D. de Vienne. 1987. Is the polymorphism of protein amounts related to phenotypic variability? A comparison of two-dimensional electrophoresis data with morphological traits in maize. Theor. Appl. Genet. 74:194-202.

Delannay X., D. M. Rodgers, and R. G. Palmer. 1983. Relative genetic contributions among ancestral lines to North American soybean cultivars. Crop Sci. 23:944-949.

Diers, B. W. and Osborn 1994. Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. Theoret. Appl. Genet. 88:662-668.

dos Santos, J. B., J. Nienhuis, P. Skroch, J. Tivang, and M. K. Slocum. 1994. Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. Theor. Appl. Genet. 87:909-915.

Ellsworth, D. L., K. D. Rittenhouse, and R. L. Honeycutt. 1993. Artifactual variation in randomly amplified polymorphic DNA banding patterns. BioTechniques. 14(2):214-217.

Falconer, D. S. 1989. Introduction to quantitative genetics. John Wiley and Sons, Inc. pp. 264-270.

Fang, G., S. Hammar, and R. Grumet. 1992. A quick and inexpensive method for removing polysaccharides from plant genomic DNA. Biofeedback. 13(1):52-55.

Frei, O. M., C. W. Stuber, and M. M. Goodman. 1986. Use of allozymes as genetic markers for predicting performance in maize single cross hybrids. Crop Sci. 26:37-42.

Gizlice, Z., T. E. Carter, Jr., and J. W. Burton. 1994. Genetic base for North American soybean cultivars released between 1947 and 1988. Crop Sci. 34(5):1143-1151.

Godshalk, E. B., M. Lee, and K. R. Lamkey. 1990. Relationship of restriction fragment length polymorphisms to single-cross hybrid performance in maize. Theor. Appl. Genet. 80:273-280.

Goodman, M. M. 1968. A measure of 'overall variability' in populations. Biometrics. 24:189-192.

Grabau, E. A., W. H. Davis, and B. G. Gengenbach. 1989. Restriction fragment length polymorphism in a subclass of the 'Mandarin' soybean cytoplasm. Crop Sci. 29:1554-1559.

Grabau, E. A., W. H. Davis, N. D. Phelps, and B. G. Gengenbach. 1992. Classification of soybean cultivars based on mitochondrial DNA restriction fragment length polymorphism. Crop Sci. 32:271-274.

Hallden, C., N. O. Nilsson, I. M. Rading, and T. Sall. 1994. Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. Theor. Appl. Genet. 88:123-128

Hanlon, R. and E. A. Grabau. 1995. Cytoplasmic diversity in old domestic varieties of soybean using two mitochondrial markers. Crop Sci. 35:1148-1151.

Hanson, W. D. and E. Casas. 1968. Spatial relationship among eight population of *Zea mays* L. utilizing information from a diallel mating design. Biometrics. 24:867-880.

Huen, M., J. P. Murphy, and T. D. Phillips. 1994. A comparison of RAPD and isozyme analyses for determining the genetic relationships among *Avena sterilis* L. accessions. Theor. Appl. Genet. 87:689-696.

Huen, M., and T. Helentjaris. 1993. Inheritance of RAPDs in F1 hybrids of corn. Theor. Appl. Genet. 85:961-968.

Jain, A., S. Bhatia, S. S. Banga, S. Prakash, and M. Lakshmikumaran. 1994. Potential use of random amplified polymorphic DNA (RAPD) technique to study the genetic diversity in Indian mustard (*Brassica juncea*) and its relationship to heterosis. 1994. Theor. Appl. Genet. 88:116-122.

Johnson, W. J., H. F. Robinson, and R. E. Comstock. 1955. Estimates of genetic and environmental variability in soybeans. Agron. J. 47:314-318.

Keim, P., and R. C. Shoemaker. 1988. Construction of a random recombinant DNA library that is primarily single copy sequence. Soybean Genet. Newsl. 15:147-148.

Keim, P., R. C. Shoemaker, and R. G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-792.

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.

Lamkey, K. R., A. R. Hallauer, and A. L. Kahler. 1987. Allelic differences at enzyme loci and hybrid performance in maize. Journal of Heredity. 78:231-234.

Lee, M., E. B. Godshalk, K. R. Lamkey, and W. W. Woodman. 1989. Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. Crop Sci. 29:1067-1071.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. Proc. Natl. Inst. Sci. India. 2:49-55.

Mailer, R. J., R. Scarth, and B. Fristensky. 1994. Discrimination among cultivars of rapeseed (*Brassica napus* L.) using DNA polymorphisms amplified from arbitrary primers. Theor. Appl. Genet. 87:697-704.

Maniatas, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York. 545 pp.

Martin, J. M., L. E. Talbert, S. P. Lanning, and N. K. Blake. 1995. Hybrid performance in wheat as related to parental diversity. Crop Sci. 35:104-108.

McGrath, J. M. and C. F. Quiros. 1992. Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. Theor. Appl. Genet. 83:783-790.

Melchinger, A. E., M. Lee, K. R. Lamkey, and W. L. Woodman. 1990. Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. Crop Sci. 30:1033-1040.

Melchinger, A. E., J. Boppenmaier, B. S. Dhillon, W. G. Pollmer, and R. G. Herrmann. 1992. Genetic diversity for RFLPs in European maize inbreds: relation to performance of hybrids within versus between heterotic groups for forage traits. Theor. Appl. Genet. 84:672-681.

Melchinger, A. E., M. Lee, K. R. Lamkey, A. R. Hallauer, and W. L. Woodman. 1990a. Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. Theor. Appl. Genet. 80:488-496

Melchinger, A. E., M. Lee, K. R. Lamkey, and W. L. Woodman. 1990b. Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize. Crop Sci. 30:1033-1040.

Melchinger, A. E., M. M. Messmer, M. Lee, W. L. Woodman, and K. R. Lamkey. 1991. Diversity and relationships among U. S. maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci. 31:669-678.

Messmer, M. M., A. E. Melchinger, M. Lee, W. L. Woodman, E. A. Lee, and K. R. Lamkey. 1991. Genetic diversity among progenitors and elite lines from the Iowa stiff stalk synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor. Appl. Genet. 83:97-107.

Moser, H. and M. Lee. 1994. RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats. Theor. Appl. Genet. 87:947-956.

Mullis, K. B., and F. Faloona. 1987. Specific synthesis of DNA *in vitro* via a polymerase catalyzed reaction. Meth. Enzymol. 155:335-350.

Olson, M., L. Hood, C. Cantor, and D. Botstein. 1989. A common language for physical mapping of the human genome. Science 254:1434-1435.

Rafalski, J. A., S. V. Tingey, and J. G. K. Williams. 1991. RAPD markers - a new technology for genetic mapping and plant breeding. AgBiotech News and Information. 3(4):645-648.

Ramshaw, J. A. M., J. A. Coyne, and R. C. Lewontin. 1979. The sensitivity of gel electrophoresis as a detector of genetic variation. Genetics. 93:1019-1037.

Rohlf, F. J. 1992. NTSYS-pc: Numerical taxonomy and multivariate analysis system (version 1.7). Exeter Software. Setauket, NY. pp 7.5-7.7.

Roth, E. J., B. L. Frazier, N. R. Apuya, and K. G. Lark. 1989. Genetic variation in an inbred plant: variation in tissue cultures of soybean [*Glycine max* (L.) Merrill]. Genetics. 121:359-368.

Skorupska, H. T., R. C. Shoemaker, A. Warner, E. R. Shipe, and W. C. Bridges. 1993. Restriction Fragment Length Polymorphism in soybean germplasm of the Southern USA. Crop Sci. 33:1169-1176.

Smith, J. J., J. S. Scott-Craig, J. R. Leadbetter, G. L. Bush, D. L. Roberts, and D. W. Fulbright. 1994. Characterization of random amplified DNA (RAPD) products from *Xanthomonas campestris* and some comments on the use of RAPD products in phylogenetic analysis. Mol. Phylogenet. Evol. 3(2):135-145.

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

Sneller, C. H. 1994b. SAS programs for calculating coefficient of parentage. Crop. Sci. 34:1679-1680.

St. Martin, S. K. 1982. Effective population size for the soybean improvement program in maturity groups 00 to IV. Crop Sci. 22:151-152.

Schoener, C. S. and W. R. Fehr. 1979. Utilization of plant introductions in soybean breeding populations. Crop Sci. 19:185-188.

Southern, E. M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

Souza, E. and M. E. Sorrells. 1991a. Relationships among 70 North American oat germplasms: I. Cluster analysis using quantitative characters. Crop Sci. 31:599-605.

Souza, E. and M. E. Sorrells. 1991b. Relationships among 70 North American oat germplasms: II. Cluster analysis using qualitative characters. Crop Sci. 31:605-612.

Souza, E. and M. E. Sorrels. 1991. Prediction of progeny variation in oat from parental genetic relationships. Theor. Appl. Genet. 82:233-241.

Smith, O. S., J. S. C. Smith, S. L. Bowen, R. A. Temborg, and S. J. Wall. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, heterosis, and RFLPs. Theor. Appl. Genet. 80:833-840.

Snedecor, G. W. and W. G. Cochran. 1967. Statistical Methods. The Iowa State University Press, Ames, Iowa. pp. 157-158.

Talbert, L. E., N. K. Blake, P. W. Chee, T. K. Blake, and G. M. Magyar. 1994. Evaluation of "sequence-tagged-site" PCR products as molecular markers in wheat. Theor. Appl. Genet. 87:789-794.

Thormann, C. E., M. E. Ferreira, L. E. A. Camargo, J. G. Tivang, and T. C. Osborn. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theor. Appl. Genet. 88:973-980.

Tinker, N. A., M. G. Fortin, and D. E. Mather. 1993. Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor. Appl. Genet. 85:976-984.

Velasquez, V. L. B., and P. Gepts. 1994. RFLP diversity of common bean (*Phaseolus vulgaris* L.) in its centres of origin. Genome. 37:256-263.

Welsh, J., and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 19(2):303-306.

Williams, J. G. K., A. R. Kubelik, K. J. Livak. J. A. Rifalski, and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acid Res. 18:6531-6535.

Williams, C. E. and D. A. St. Clair. 1993. Phenetic relationships and levels of variability detected by restriction fragment lenth polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. Genome. 36:619-630.

Yu, L. X. and H. T. Nguyen. 1994. Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa* L.). Theor. Appl. Genet. 87:668-672.

OTHER REFERENCES

Cooper, R. L., 1995. USDA/ARS, OSU/OARDC, Wooster, Ohio. Personal Communication.

Neslon, R. 1994. USDA/ARS, Urbana, Illinois. Personal Communication.

Webb, D. 1992. Pioneer HI-BRED International. Personal Communication.