Theor Appl Genet (1997) 94:20–26 

T. Helms · J. Orf · G. Vallad · P. McClean

# Genetic variance, coefficient of parentage, and genetic distance of six soybean populations

Received: 24 April 1996 / Accepted: 17 May 1996

**Abstract** Plant breeders would like to predict which biparental populations will have the largest genetic variance. If the population genetic variance could be predicted using coefficient of parentage or genetic distance estimates based on molecular marker data, breeders could choose parents that produced segregating populations with a large genetic variance. Three biparental soybean {*Glycine max* (L.) Merr.} populations were developed by crossing parents that were closely related, based on pedigree relationships. Three additional biparental populations were developed by crossing parents that were assumed to be unrelated. The genetic variance of each population was estimated for yield, lodging, physiological maturity, and plant height. Coefficient of parentage was calculated for each pair of parents used to develop the segregating populations. Genetic distance was determined, based on the number of random amplified polymorphic markers (RAPD) that were polymorphic for each pair of parents. Genetic distance was not associated with the coefficient of parentage or the magnitude of the genetic variance. The genetic variance pooled across the three closely related populations was smaller than the genetic variance pooled across the three populations derived from crossing unrelated parents for all four traits that were evaluated.

**Key words** *Glycine max* · DNA · Pedigree analysis · Genetic variance

Communicated by E. J. Eisen

T. Helms (✉) · G. Vallad · P. McClean
Department of Plant Sciences, North Dakota State University, Fargo, ND 58105, USA

J. Orf
Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

## Introduction

Plant breeders that develop segregating populations from biparental crosses would like to predict the mean and genetic variance of experimental lines derived from each two-parent combination. Populations with a high mean and large genetic variance would be excepted to be the most useful for developing improved cultivars.

Prediction of the population mean from the mid-parent mean has been evaluated by several investigators. Busch et al. (1974) found that the mean yield of 25 populations of experimental lines of wheat (*Triticum aestivum* L.) was highly correlated ($r = 0.73$, $P = 0.01$) with mid-parent means. Souza and Sorrells (1991) found that means of three traits were not significantly different than the mid-parent means for 20 oat populations. In the oat populations studied by Cowan and Frey (1987), the mid-parent mean generally was higher than the progeny mean.

Predicting the genetic variance within a biparental population is more difficult and less reliable than predicting the population mean. The coefficient of parentage (Kempthorne 1969) has been suggested as one measure of genetic diversity. However, the coefficient of parentage requires simplifying assumptions and quantifies only the probability of two alleles being identical by descent. Breeders are perhaps more concerned with the probability of two allelles being alike in state, because the genetic variance among progeny would be determined by the number of alleles that are alike in state.

The association between the genetic variance and measures of genetic diversity among the parents has been evaluated. Souza and Soorrells (1991) determined that the coefficient of parentage was a better predictor of genetic variances than similarity indexes derived from discrete morphological or biochemical markers. Cowen and Fry (1987) found a significant correlation ($r = 0.41$) between the log of the generalized genetic variance and genealogical distance.

Cox et al. (1985a) suggested that the ideal breeding population should have a high mean for the traits of

interest and a high genetic variance. They further suggested that a similarity index which combines coefficient of parentage and data derived from biochemical and morphological markers could be a useful measure of genetic diversity. In their analysis of soybean cultivars released after 1970, they found a correlation of $r = 0.60$ ($P = 0.1$) between coefficient of parentage and a similarity index measured with morphological and biochemical markers.

In a separate study Cox et al. (1985b) compared the coefficient of parentage of selected hard red winter wheat (*Triticum aestivum* L) parents to a similarity index derived from polyacrylamide gel electrophoresis data of gliadin storage protein patterns. Because the ancestors of hard red winter wheat were heterogeneous landraces, the correlation between these two parameters was only 0.27 ($P = 0.01$).

If the mean and genetic variance of biparental populations could be predicted, the best two parent combinations could be selected for crossing. Either the coefficient of parentage or biochemical markers might be used to predict the genetic similarity between parents. Our objectives were to: (1) compare the coefficient of parentage to the genetic distance estimated from random amplified polymorphic DNA (RAPD) data; (2) compare the genetic distance estimated from RAPD markers to the genetic component of variance estimate; (3) compare the coefficient of parentage to the genetic component of variance estimate for populations derived from closely and distantly related parents.

## Materials and methods

### Genetic relationships

Six soybean populations were developed, three from crosses among parents that were closely related as determined by published pedigree relationships and three additional populations among unrelated parents. The coefficient of parentage among these unrelated parents was assumed to be zero based on pedigree relationships.

Cultivars 'Evans' (Lambert and Kennedy 1975) and 'Wilkin' (Lambert and Kennedy 1973) are derived from the cross of 'Merit' (Johnson 1960) and 'Harosoy' (Weiss and Stevenson 1955). The cultivar 'Ozzie' (Orf et al. 1985) is derived from a cross of 'Wilkin' × M63-217Y. M63-217Y is a yellow hilum selection from 'Hodgson' (Lambert and Kennedy 1975). 'Pioneer 9061' has the pedigree 'Wells' × ('Corsoy'/2 × 'Rampage') (Wilcox et al. 1973; Weber and Fehr 1970a; Weber and Fehr 1970b) and 'Pioneer 0877' has the pedigree 'Corsoy' × ('Clark' × 'Chippewa 64') (Johnson 1958; Bernard 1964). The experimental line ND867 is a North Dakota State University-derived line with the pedigree 'Wilkin' × L62-361. L62-361 is a Dt2 semi-determinate near-isogenic derivate of 'Harosoy'. 'Ozzie', 'Evans', Pioneer 9061, and 'Pioneer 0877' are Maturity Group 0 cultivars. 'Wilkin' and 'ND867' are Maturity Group 00 genotypes. These genotypes are considered to be adapted and high yielding.

PI 238924 is a Maturity Group 0 accession from Czechoslovakia with the name 'Kirches Stamm 2008' and was added to the USDA Soybean Germplasm Collection in 1957 (Bernard et al. 1989). PI417511 is a Maturity Group 000 accession from France with the name 'Grignon 48' and was added to the USDA Soybean Germplasm Collection in 1977 (Bernard et al. 1989). PI 261475 is a Maturity Group 0 accession from northeast China with the name 'Shika No. 1' and was added to the USDA Soybean Germplasm Collection in 1959 (Bernard et al. 1989). None of these plant introductions (PI) are in the pedigrees of the adapted cultivars or experimental lines used as parents in this experiment.

The following crosses were made in the summer of 1989 at Fargo, ND: 'Ozzie' × 'Wilkin', 'Wilkin' × 'Evans', 'Ozzie' × 'ND867', 'Evans' × PI 417511, 'Pioneer 9061' × PI 238924, and 'Pioneer 0877' × PI 261475. The resulting $F_1$ plants were grown in the winter of 1989–1990 in Chile. The $F_2$ populations were grown in the 1990 Fargo summer nursery. The $F_3$ popplulations were grown in the 1990–1991 Chile winter nursery, and $F_{3:4}$ lines were developed by single-plant threshing appproximately 100 individual plants from each population. No selection was applied to these populations, and the experimental lines derived from each population are random lines. The $F_{3:4}$ lines were evaluated in progeny rows in the summer of 1991. Ninety-nine lines were harvested from each population with the exception that 94 lines were derived from the population 'Pioneer 0877' × PI 261475. The experimental lines were evaluated for physiological maturity in 1991, based on data from a single progeny row.

### Experimental design and statistical analysis

Eleven sets were formed by grouping lines of similar maturity into a set. Each set consisted of 9 lines from each population, except for the latest maturing set. The set of latest maturing lines contained 9 lines from five populations and 4 lines from the 'Pioneer 0877' × PI 261475 population. Each set was arranged in a square lattice design and included 54 experimental lines, the nine parental genotypes and a Maturity Group 00 cultivar 'McCall'. 'McCall' was used to complete the required 64 entries for an 8 × 8 lattice experiment.

Each set was arranged in a separate 8 × 8 square lattice design to form 11 lattice sets per environment. Each lattice design was replicated twice at each location. The experiment was evaluated at Casselton, N.D. and Morris, Minn. in 1993 and 1994. Plots at Casselton, N.D. had a planted length of 6.4 m. The row spacing within and between plots was 0.76 m and the center 4.6 m of a two-row plot was harvested. Plots at Morris, Minn. had a planted length of 3 m with a between-row spacing of 0.25 m. The center 2.4 m of a four-row plot was harvested.

Grain yield, plant height, physiological maturity, and lodging were evaluated. Grain yield was converted to kilograms per hectare at 13% water content. Plant height, measured at maturity, was recorded as the distance (in centimeters) from the base of the plant to the terminal node. Dates when 95% of the pods were the mature plant color were recorded at all locations, except Casselton in 1993. Grain yield and plant height were evaluated in all four environments. Lodging was evaluated at Casselton and Morris in 1994. In 1993, there was not an appreciable amount of lodging at either location.

Within each environment and set the data were adjusted for lattice block effects using the Statistical Analysis Package (SAP) computer program (Hammond 1993). The adjusted means of all environments were included in a combined analysis across environments, and entries were pooled across sets. Environments, lines and sets were considered to be random effects. Populations were considered to be fixed effects. Mean squares were equated to expected mean squares, and the genetic component of variance was calculated for each population. Standard errors of estimates for the genetic variance components were calculated, based on the formula provided by Hallauer and Miranda (1981).

Further statistical analyses included pooling the sum of squares from the three closely related populations into group one. The sum of squares from the three populations derived from the crosses of the adapted × PI parents were pooled into group two. The sums of squares for the genotype × environment interaction were pooled across the three populations within each group. Lines derived from each population were nested within that population. Three populations were nested within each group. The population mean was susbtracted from the mean of each line within that population before squaring and summing these deviations. These sums of squares for each population within a group were then pooled and divided by the pooled degrees of freedom.

The mean square for populations within each group was then equated to the expected mean square, and the genetic component of variance was estimated. The standard error of the genetic variance for

each group was calculated, based on the formula provided by Hallauer and Miranda (1981). When the difference between group-one and group-two genetic variance components was greater than the sum of the standard errors, the genetic variance components were considered to be significantly different.

This experiment was designed to compare populations derived from closely related parents to populations derived from unrelated parents. Each group of three populations were considered to be two levels (groups) of a genetic similarity factor. The coefficients of parentage were derived using a computer program provided by Cox et al. (1986). Cox et al. (1985b) stated the assumptions used for the calculations of the coefficients of parentage.

### RAPD procedures

DNA was isolated from leaf material of each genotype by the technique of Doyle and Doyle (1990). The DNA was amplified using Operon Technology random decamer synthetic oligonucleotide primers using a modification of the procedure described by Martin et al. (1991). The reactions contained 20 ng template DNA, 400 n*M* primer, 20 m*M* Tris-HCl, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 200 n*M* each dNTP, and 1.0 unit *Taq* DNA polymerase in a reaction volume of 10 ul. The thermal cycle used was: 44 cycles at 94°C (1 min) for DNA denaturation, 35°C (1 min) to anneal primers, and 72°C (2 min) for primer extension, followed by a final cycle of 94°C (1 min), 35°C (1 min), and 72°C (7 min) for final extension. After the final cycle, samples were held at 4°C until analysis. Fragments generated by amplification were separated according to size on 2% agarose gels run in 1 × TAE (40 m*M* Tris-Acetate, pH 9.0, 1.0 m*M* EDTA), stained with ethidium bromide and visualized by illumination with ultraviolet light. A total of 80 primers were sampled, which resulted in 470 markers, of which 277 were polymorphic. One genotype, 'Pioneer 9061', was included twice in each amplification set as an internal check to evaluate the repeatability of the banding patterns. The duplicate lanes of 'Pioneer 9061' showed complete agreement for all marker bands.

### Genetic distance

A pairwise comparison of the RAPD pattern of the two parental genotypes for each population over all primers was made, and the number of products shared by the two genotypes and the number of unique products for each genotype were counted. From this data the proportion of products that was shared by the two genotypes ($S_{xy}$) was determined to estimate the genetic similarity. The following formula developed by Nei and Li (1979) was used:

$$S_{xy} = 2n_{xy}/(n_x + n_y)$$

where $n_x$ and $n_y$ are the total number of amplification products obtained with the 80 primers for genotypes X and Y, respectively. $n_{xy}$ is the number of amplification products shared between the two parents.

The genetic distance ($D_{xy}$) is the proportion of products that are not in common between two genotypes X and Y and was defined $1 - S_{xy}$. The standard error (SE) of $D_{xy}$ was calculated as: (Keim et al. 1992)

$$SE(D_{xy}) = [S_{xy}(D_{xy})/n]^{1/2}$$

where $n$ is the number of markers. Markers that were monomorphic across all nine genotypes were excluded from the calculation of genetic distance. There were 277 non-monomorphic markers. If the sum of the standard errors of two genetic distances was smaller than the difference between the $D_{xy}$ estimates between two combinations of two parents, then the difference was considered significant. A bootstrap technique was used to determine the number of polymorphic markers necessary to evaluate genetic distance. There was a negligible change in the genetic distance when calculations using 80% of the marker data were compared to 100% of the marker data.

## Results and discussion

Mean yield of the 'Evans*Wilkin' population was higher than that of any of the three adapted × PI populations (Table 1). However, the 'Pioneer 0877'*PI 261475 population was higher yielding than the 'Ozzie'*ND867 population. The results show that adapted × PI populations can be developed that yield as much as

**Table 1** Mean yield, plant height, physiological maturity, and lodging of six bi-parental derived soybean populations and parents

| Population or parent | Yield (kg/ha) | Plant height (cm) | Physiological maturity (days[a]) | Lodging score[b] |
|---|---|---|---|---|
| Ozzie | 2690 | 65 | 47 | 1.1 |
| Wilkin | 2610 | 62 | 46 | 1.2 |
| Evans | 3130 | 74 | 51 | 1.9 |
| ND867 | 2060 | 61 | 39 | 1.4 |
| PI 417511 | 1370 | 48 | 40 | 1.5 |
| Pioneer 0877 | 2790 | 69 | 54 | 2.5 |
| PI 261475 | 2670 | 65 | 51 | 3.1 |
| Pioneer 9061 | 3190 | 65 | 50 | 1.2 |
| PI 238924 | 2110 | 56 | 48 | 2.3 |
| Ozzie*Wilkin | 2710 | 65 | 48 | 1.3 |
| Evans*Wilkin | 2820 | 68 | 49 | 1.5 |
| Ozzie*ND867 | 2580 | 68 | 50 | 1.4 |
| PI 417511*Evans | 2460 | 65 | 52 | 2.1 |
| Pioneer 0877* PI 261475 | 2650 | 68 | 54 | 2.6 |
| Pioneer 9061* PI 238924 | 2590 | 60 | 50 | 1.7 |
| LSD (0.05)[c] | 353 | 6 | 3 | 0.7 |
| LSD (0.05)[d] | 36 | 1 | 0.3 | 0.1 |

[a] Days from August 1
[b] Lodging score, 1 (all plants upright) to 5 (all plants prostrate)
[c] Least significant difference between parents at the 0.05 probability level
[d] Least significant difference between population means at the 0.05 probability level

adapted × adapted population crosses. All six populations were adapted to this northern USA soybean growing region. The maturity among these populations only varied by a range of 6 days. Populations that are unadapted due to late maturity may not express their yield potential in northern environments.

There was no apparent relationship between the coefficient of parentage and the genetic distance, based on RAPD markers (Table 2). For example, 'Evans' and 'Wilkin' had a coefficient of parentage of 0.53, and the genetic distance between these two parents was 0.29. 'Evans' and PI 417511 have an assumed coefficient of parentage of zero, and the genetic distance between these two parents was 0.31. 'Ozzie' and 'Wilkin'; 'Ozzie' and ND867; and 'Pioneer 9061' and PI238924 all have a similar genetic distance. However, the coefficient of parentage of 'Pioneer 9061' and PI 238924 is very different from that of 'Ozzie' and 'Wilkin' or 'Ozzie' and ND867.

The results suggest that coefficient of parentage and genetic distance are different measures of genetic diversity between two genotypes. Coefficient of parentage is a measure of identity by descent. Genetic distance measured with RAPD data is a measure of alike in state for predominantly non-coding regions of the genome. Because the soybean genome consists of approximately 40% single-copy and 60% repetitive DNA (Goldberg 1978), the DNA markers should be distributed in the same ratio among these two DNA classes. Although the single-copy DNA and repetitive DNA classes contain expressed sequences, the majority of the genome complexity is not associated with expressed genes (Goldberg et al. 1978). Therefore, the majority of RAPD markers would be associated with non-expressed and neutral sequences that should not be affected by selection. Thus, the marker data as a whole should be reflective of the genetic distance among two parents.

**Table 2** Coefficient of parentage and genetic distance and standard error (SE) of genetic distance for six combinations of two parents. Genetic distance was determined by the proportion of markers that were not common to the two parents. Monomorphic bands for all nine genotypes were excluded from genetic distance calculations

| Parental combination | Coefficient of parentage | Genetic distance[a] | SE |
|---|---|---|---|
| Ozzie, Wilkin | 0.63 | 0.14 | 0.02 |
| Evans, Wilkin | 0.53 | 0.29 | 0.02 |
| Ozzie, ND867 | 0.49 | 0.17 | 0.02 |
| Evans, PI 417511 | 0 | 0.31 | 0.03 |
| Pioneer 0877, PI 261475 | 0 | 0.23 | 0.02 |
| Pioneer 9061, PI 238924 | 0 | 0.20 | 0.02 |

[a] Based on 277 polymorphic RAPD markers from 80 primers

It is important to our conclusions that we can detect significant differences between genetic variances. Genetic components of variance were estimated by equating mean squares (Table 3) to expected mean squares and solving. We defined a difference between two genetic variance components to be significant when the difference was greater than the sum of the standard errors.

Genetic distance was not an indication of genetic variance. The genetic distance of 'Ozzie' and 'Wilkin' was 0.14 (Table 2) and was similar to the genetic distance of 'Ozzie' and ND867, which was 0.17. However, the genetic variance estimate for yield of the 'Ozzie' × 'Wilkin' population was less than one-half the magnitude of the genetic variance estimate for the 'Ozzie' × ND867 population (Table 4). Additionally, 'Evans' and 'Wilkin' had a similar genetic distance as 'Evans' and PI 417511 (Table 2). However, the genetic variance estimate for yield of the 'Evans' × 'Wilkin' population is about one-sixth as large as the genetic variance estimate of the 'Evans' × PI 417511 population (Table 4).

The results show that genetic distance measured with RAPD markers was not a good indication of genetic

**Table 3** Mean squares for six populations (POP) and two groups for both genotype and genotype × environment (G × E) for four traits, pooled across sets. Each group includes three populations, and each population includes approximately 99 inbred soybean lines

| Source | Mean square | | | |
|---|---|---|---|---|
| | Yield | Plant height | Physiological maturity | Lodging |
| Genotype | | | | |
| Group 1[a] | 249 238 | 105.02 | 37.32 | 0.20 |
| POP 1[b] | 217 354 | 82.02 | 29.47 | 0.07 |
| POP 2 | 167 723 | 94.16 | 23.70 | 0.23 |
| POP 3 | 360 785 | 138.63 | 58.97 | 0.31 |
| Group 2 | 384 178 | 138.05 | 49.78 | 0.66 |
| POP 4 | 390 416 | 138.23 | 59.54 | 0.72 |
| POP 5 | 260 732 | 126.81 | 40.04 | 0.70 |
| POP 6 | 495 639 | 148.59 | 49.43 | 0.56 |
| *G* × *E* | | | | |
| Group 1 | 122 118 | 42.46 | 5.99 | 0.11 |
| POP 1 | 119 301 | 21.91 | 4.74 | 0.07 |
| POP 2 | 121 189 | 43.00 | 5.80 | 0.11 |
| POP 3 | 125 934 | 63.13 | 7.46 | 0.14 |
| Group 2 | 152 188 | 33.57 | 8.13 | 0.25 |
| POP 4 | 130 261 | 33.33 | 7.73 | 0.26 |
| POP 5 | 141 980 | 36.62 | 8.99 | 0.25 |
| POP 6 | 183 246 | 30.93 | 7.72 | 0.25 |

[a] Group 1 includes populations 1, 2, and 3; group 2 includes populations 4, 5, and 6

[b] Population 1 is 'Ozzie'*'Wilkin', population 2 is 'Evans'*'Wilkin', populattion 3 is 'Ozzie'*ND867, population 4 is 'Evans'*PI 417511, population 5 is 'Pioneer 0877'*PI 261475, population 6 is 'Pioneer 9061'*PI 238924

**Table 4** Genetic variance component estimates, standard errors (SE), and number of environments evaluated for yield, plant height, physiological maturity, and lodging for six soybean populations

| Trait | Population | *Genetic variance* Component | SE | Number of environments |
|---|---|---|---|---|
| *Yield* | | | | |
| | Ozzie*Wilkin | 24 500 | 8 600 | 4 |
| | Evans*Wilkin | 11 600 | 6 900 | 4 |
| | Ozzie*ND867 | 58 700 | 13 800 | 4 |
| | Evans*PI 417511 | 65 000 | 15 000 | 4 |
| | Pioneer 0877*PI 261475 | 29 700 | 10 500 | 4 |
| | Pioneer 9061*PI 238924 | 78 100 | 19 100 | 4 |
| *Plant height* | | | | |
| | Ozzie*Wilkin | 15.0 | 3.1 | 4 |
| | Evans*Wilkin | 12.8 | 3.7 | 4 |
| | Ozzie*ND867 | 18.9 | 5.4 | 4 |
| | Evans*PI 417511 | 26.2 | 5.2 | 4 |
| | Pioneer 0877*PI 261475 | 22.6 | 5.0 | 4 |
| | Pioneer 9061*PI 238924 | 29.4 | 5.6 | 4 |
| *Physiological maturity* | | | | |
| | Ozzie*Wilkin | 8.2 | 1.5 | 3 |
| | Evans*Wilkin | 6.0 | 1.2 | 3 |
| | Ozzie*ND867 | 17.2 | 3.0 | 3 |
| | Evans*PI 417511 | 17.3 | 3.0 | 3 |
| | Pioneer 0877*PI 261475 | 10.4 | 2.1 | 3 |
| | Pioneer 9061*PI 238924 | 13.9 | 2.5 | 3 |
| *Lodging* | | | | |
| | Ozzie*Wilkin | 0.00 | 0.01 | 2 |
| | Evans*Wilkin | 0.06 | 0.02 | 2 |
| | Ozzie*ND867 | 0.08 | 0.03 | 2 |
| | Evans*PI 417511 | 0.23 | 0.06 | 2 |
| | Pioneer 0877*PI 261475 | 0.22 | 0.06 | 2 |
| | Pioneer 9061*PI 238924 | 0.16 | 0.05 | 2 |

variance for yield among these six populations. Also, there is no relationship between genetic distance and genetic variance for plant height, physiological maturity, or lodging among these six populations.

If the coefficient of parentage was useful for predicting the magnitude of the genetic variance component, genetic variance components of each of the three populations derived from crossing closely related parents would be expected to be of similar magnitude. However, a comparison of the genetic variance estimates for yield among the three populations derived from crossing closely related parents showed that this was not the case (Table 4). There was five-fold increase in genetic variance for yield between the 'Evans' × Wilkin population as compared to the 'Ozzie' × ND867 population. Therefore, coefficient of parentage was not a good predictor of genetic variance for a specific population derived from the cross of two parents.

Genetic variance estimates for plant height within each of the three populations derived from crossing closely related parents are similar in magnitude (Table 4). However, for physiological maturity and lodging, there are significant differences in genetic variance estimates among the three populations derived from crossing closely related parents. When we compare the three populations derived from crossing closely related parents, there is no relationship between the coefficient of parentage and the genetic variance estimate for yield, physiological maturity, and lodging.

The apparent lack of relationship between coefficient of parentage and genetic variance is limited to these populations, and a closer association could occur among another set of populations. Additional research that includes different populations will be required to determine the overall association between the coefficient of parentage and genetic variance estimates. Due to the amount of resources required to estimate genetic variance components with sufficient precision for comparing populations, this experiment was not intended to provide a global answer to this question. However, it is apparent that the coefficient of parentage was not useful as a predictor of the magnitude of the genetic variance for a specific population derived from crossing two parents.

The explanation for the lack of relationship between the coefficient of parentage and the genetic variance is that the coefficient of parentage is a measure of the degree to which two genotypes (parents) are identical by

descent. The genetic variance is a measure of the degree to which two genotypes (parents) are alike in state. The coefficient of parentage estimates the minimum proportion of alleles that would be expected to be identical by descent and therefore alike in state between two parents. Also, of the alleles that are not identical by descent, the proportion of alleles that are alike in state is unknown. Another explanation for the lack of relationship between coefficient of parentage and genetic variance could be inaccurate pedigree information. Although pedigree relationships may not always be accurately reported, the applied plant breeder must use the pedigree information that is available.

Among these six populations, the genetic variance component for yield in the 'Ozzie' × ND867 population was larger than the genetic variance component for yield of the 'Pioneer 0877' × PI 261475 population (Table 4). The genetic variance for plant height within the 'Pioneer 0877' × PI 261475 population was not significantly different than that of the 'Ozzie' × ND867 population. This result provides further evidence that the genetic variance within some populations derived from closely related parents are greater or equal to that of populations derived from unrelated parents. The genetic variance for physiological maturity of the 'Ozzie' × ND867 population is not significantly different from that of the 'Evans' × PI 417511 population, and the magnitudes of these estimates are very similar.

A comparison of the genetic variance for lodging among closely related and unrelated crosses provides a completely different conclusion than for the traits yield, plant height and physiological maturity (Table 4). The genetic variance for lodging of all three populations derived from closely related parents is smaller than that of any of the unrelated populations. The adapted × PI populations had a greater lodging susceptibility than the adapted × adapted populations (Table 1). The larger genetic variance component for lodging in the adapted × PI populations was probably due to the addition of lodging susceptible genes from the PI parents.

The genetic variance for yield, pooled across populations within each group, was significantly different between groups (Table 5). The pooled genetic variance for populations developed from closely related crosses (group one) is smaller than that of populations developed from unrelated crosses (group two) for each of the four traits evaluated.

These results suggest that although breeders cannot predict the magnitude of the genetic variance component for any specific population from the coefficient of parentage, crosses among unrelated parents will, on average, result in a larger genetic variance than crosses among related parents. If the mean of two different populations comprised of inbred progenies are equal, the population with the larger genetic variance would be expected to produce a greater proportion of high yielding inbred lines. It will require further research, evaluating different populations, to determine whether this result is specific to this sample of populations.

Souza and Sorrells (1991) found that as the coefficient of parentage increased, the genetic variance estimate decreased for one trait of the six traits they evaluated. They found that crosses between more distantly related parents produced larger genetic variance estimates than crosses between closely related parents for plant biomass. However, crosses between closely related parents resulted in larger genetic variances than crosses among distantly related parents for grain yield, test weight, heading date, grain filling period, and maturity date. Because the crosses to the unadapted parents matured too late for complete expression of yield, the genetic variance estimates for yield were inconclusive.

The results of this experiment provide evidence that there was no relationship between the coefficient of parentage and genetic distance estimated from RAPD data. Also, there was no relationship between the genetic distance and the genetic component of variance estimate. The RAPD markers we used may not have been linked to any of the traits that were evaluated. In the future, it may be necessary to identify a large number of marker loci closely linked to quantitative trait loci (QTL) and then compare genetic distance, based on these linked markers. The use of markers linked to QTLs might result in an association between marker-based estimates of genetic distance and the magnitude of the genetic variance.

**Table 5** Pooled genetic variance component estimates, standard errors (SE), and number of environments for yield, plant height, physiological maturity, and lodging for two groups of soybean populations. Group one includes three populationss resulting from crossing closely related parents, and group two includes three populations resulting from crossing parents that are assumed unrelated based on coefficient of parentage estimates. Populations were nested within groups

| Trait | Group | Genetic variance | | Number of environments |
|---|---|---|---|---|
| | | Component | SE | |
| Yield | One | 31 780 | 5700 | 4 |
| Yield | Two | 58 000 | 8700 | 4 |
| Plant height | One | 15.63 | 2.35 | 4 |
| Plant height | Two | 26.12 | 3.06 | 4 |
| Physiological maturity | One | 10.44 | 1.10 | 3 |
| Physiological maturity | Two | 13.88 | 1.47 | 3 |
| Lodging | One | 0.048 | 0.01 | 2 |
| Lodging | Two | 0.203 | 0.03 | 2 |

The relationship between the genetic variance estimate and the coefficient of parentage does not appear to be of sufficient predictive value for any specific combination of two parents. The average genetic variance of populations derived from closely related parents was smaller than the average genetic variance of populations derived from crosses among PI × adapted parents.

**Acknowledgements** We would like to thank M. A. Halvorson and P. J. Schaus for collecting field data and two anonymous reviewers.

## References

Bernard RL (1964) Hawkeye 63, Harosoy 63, and Chippewa 64 soybeans. Crop Sci 4:663–664

Bernard RL, Juvik GA, Nelson RL (1989) USDA soybean germplasm collection inventory. Int Agric Publ INSTOY Ser. No. 31 University of Illinois, Urbana, Ill.

Bush RH, Janke JC, Frohberg RC (1974) Evaluation of crosses among high and low yielding parents of spring wheat (*Triticum aestivum* L.) and bulk prediction of line performance. Crop Sci 14:47–50

Cowen NM, Frey KJ (1987) Relationship between genealogical distance and breeding behavior in oats (*Avena sativa* L.). Euphytica 36:413–424

Cox TS, Kiang YT, Gorman MB, Rodgers DM (1985a) Relationship between coefficient of parentage and genetic similarity indicies in the soybean. Crop Sci 25:529–532

Cox TS, Lookhart GL, Walker DE, Harrell LG, Albers LD, Rodgers DM (1985b) Genetic relationships among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polyacrylamide gel electophoretic patterns. Crop Sci 25:1058–1063

Cox TS, Murphy JP, Rodgers DM (1986) Changes in genetic diversity in the red winter wheat regions of the United States. Proc Natl Acad Sci USA 83:5583–5586

Doyle JJ, Doyle JF (1990) Isolation of plant DNA from fresh material. Focus 12:13–15

Goldberg RB (1978) DNA sequence organization in the soybean plant. Biochem Genet 16:45–68

Goldberg RB, Hoschek G, Kamaley JC (1978) Sequence complexity of nuclear and polysomal RNA in leaves of the tobacco plant. Cell 14:123–131

Hallauer AR, Miranda JB (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames, Iowa

Hammond JJ (1993) SAP statistical program (unpublished)

Johnson HW (1958) Registration of soybean varieties, VI. Agron J 50:690–691

Johnson HW (1960) Registration of soybean varieties, VII. Agron J 52:659–660

Keim P, Beavis W, Schupp J, Freestone R (1992) Evaluation of soybean RFLP maker diversity in adapted germ plasm. Theor Appl Genet 85:205–212

Kempthorne O (1969) An introduction to genetic statistics. Iowa State University Press, Ames, Iowa

Lambert JW, Kennedy BW (1973) Registratioon of Wilkin soybean. Crop Sci 13:583

Lambert JW, Kennedy BW (1975) Registration of Evans and Hodgson soybeans. Crop Sci 15:735

Martin GB, Williams JGK, Tanksley SD (1991) Rapid identification of markers linked to a *Psedomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proc Natl Acad Sci USA 88:2336–2340

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Orf JH, Lambert JW, Kennedy BW (1985) Registration of Ozzie soybean. Crop Sci 25:366

Souza E, Sorrells ME (1991) Prediction of progeny variation in oat from parental genetic relationships. Theor Appl Genet 82: 233–241

Weber CR, Fehr WR (1970a) Registration of Wirth and Rampage soybeans. Crop Sci 10:729

Weber CR, and Fehr WR (1970b) Registration of Corsoy soybeans. Crop Sci 10:729

Weiss MG, Stevenson TM (1955) Registration of soybean varieties, V. Agron J 47:541–543

Wilcox JR, Athow KL, Laviolette FA, Abney TS, Probst AH, Martin RJH (1973) Registration of Wells soybean. Crop Sci 13:583