ESTIMATION OF THE GENETIC STRUCTURE OF AN ELITE SOYBEAN POPULATION AND IT'S APPLICATION TO A SOYBEAN BREEDING PROGRAM

By

Clay Hurd Sneller

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 Science

1991

ABSTRACT

ESTIMATION OF THE GENETIC STRUCTURE OF AN ELITE SOYBEAN POPULATION AND IT'S APPLICATION TO A SOYBEAN BREEDING PROGRAM

By

Clay Hurd Sneller

The source and frequency of genes, and the genetic distance between lines, was estimated with coefficient of parentage (CP) and a similarity index (SI) based on marker loci for a set of 62 elite soybean lines. The genetic base of this population derived from 27 ancestral parents with "Lincoln" and "Mandarin Ottawa" contributing 30.6% and 17.7% of the parentage respectively. The average SI value between the lines was 0.64 and the average CP was 0.26. While the average relationships suggested a narrow genetic base, pairs of lines could be identified that theoretically shared few genes and certain lines were identified as outliers. There was a poor correlation between the SI and CP values. Selection of lines based on maturity and yield criteria appeared to result in a more inbred population than a general set of public cultivars with a broader maturity range. The restriction in the elite population appeared to be due to selection for similar maturity versus selection for similar yield potential in a narrowly defined environment.

Genetically similar and distance elite lines were crossed and approximately 55 F2:3 or F4:5 families were derived from each cross and field tested for seed yield, plant height, and date of maturity. The generalized variance, and the genetic variance for individual traits were estimated for each population. A third parental distance (PCD) was derived from a principal component analysis of CP and SI data. The variance parameters of a population increased as the parental genetic distance increased, regardless of the method of estimating the distance. There was a stronger association of the measures with the generalized variances than with the genetic variances. No measure was significantly associated with yield genetic variance while all seemed predictive of the genetic variances for maturity and height. The utility of the distance measures appeared to be limited to identifying which populations would have an increased likelihood of having an above average genetic variance for the individual traits. The PCD appeared to be the best predictor of the variance parameters.

Table of Contents

List of Tables

Estimation of the Genetic Structure of an Elite Soybean Population

Application of Genetic Distances to a Soybean Breeding Program

List of Figures

Estimation of the Genetic Structure of an Elite Soybean Population

Application of Genetic Distances to a Soybean Breeding Program

Literature Review

I. Genetic Distance and Plant Breeding

The existence of genetic diversity is the cornerstone of improving plants to meet human needs. Our present crop species were selected from among the vast array of species in the plant kingdom and for centuries farmers selected among existing diverse genotypes to improve these species. Later plant breeders began crossing different genotypes to create further diversity by recombining the genes of one parent with the different genes of the other parent. The success of crossing depends on the assumption that the two parents possess different genes (i.e. that there is genetic distance between them). Furthermore, breeders make crosses with the purpose of generating new and improved genotypes for a particular trait and therefore require that the crossed parents are genetically distant in regards to trait to be improved.

The necessity of genetic distance between parents to obtain breeding progress can be seen in the gain from selection equation.

$$\Delta G = S\,h^2 = S\,\sigma^2_a/\sigma^2_p = S\,(\sigma^2_a/\sigma^2_a + \sigma^2_d + \sigma^2_{na} + \sigma^2_e + \sigma^2_{ge})$$

ΔG = gain from one cycle of mass selection
S = selection differential
h^2 = narrow sense heritability
σ^2_a = additive genetic variance = $\Sigma\,(2p_iq_i\alpha_i^2)$
α_i = average effect of an allele substitution, $[a_i+d_i(q_i-p_i)]$
p_i, q_i = frequency of the two alleles at the i^{th} locus
a_i = absolute deviation of the homozygous genotypes at the i^{th} locus fromtheir mean
d_i = deviation of the i^{th} heterozygote from the mean of the two homozygous genotypes
σ^2_p = phenotypic variance
σ^2_{na} = non-additive genetic variance
σ^2_e = environmental variance
σ^2_{ge} = genotype x environment variance

Gain from selection should increase as additive genetic variance increases assuming that the population mean and phenotypic variance remain constant. In a segregating population that is generated from crossing two individuals, additive genetic variance at the i^{th} locus is maximized when $p_i = q_i = 0.5$, a situation that occurs when each parent possess different alleles. Additive genetic variance is the sum of the additive effects from all loci that affect the measured trait and will increase as the genetic distance between two parents increases as this should increase the number of segregating loci that affect the trait increases in the population.

The objective of plant breeding for metric traits is to select a genotype or a population of individuals that has a more desirable value for the selected trait than the parents. The mean of the progeny (u') derived from selected individuals can be expressed in terms from the gain from selection equation as

$$\Delta G = u' - u = S\,h^2 = (u_s - u)\,h^2$$
$$u' = (u_s - u)\,h^2 + u$$

u = mean of the parental population
u_s = mean of the selected individuals

The heritable portion of the selection differential is added to the mean of the parental population to produce u'. The ideal situation for obtaining progeny with a high mean is when the parents have a high mean (elite parents) and are genetically diverse. A high parental mean can be assured by thorough testing of the parents in the desired environment but apriori knowledge of the genetic diversity between the parents can only be estimated or inferred.

The amount of genetic variability in a population is one parameter that can be measured in a hybrid population that is influenced by the genetic distance of the parents (Cowen and Frey, 1987a). A second parameter is the amount of mid-parent heterosis (H_m) exhibited by the hybrid population (Falconer, 1981):

$$H_m = \sum_{i=1}^{n} d_i y_i^2$$

y_i = p - p'
p = frequency of one of the two alleles at the i^{th} of n loci in parent one
p' = frequency of the same allele in parent two

It is apparent from this equation that heterosis will increase in proportion to the degree of genetic divergence of the parents (y) when a directional dominance exists over all loci (i.e. $\Sigma d_i = 0$). The genetic variance and heterosis exhibited in a hybrid population are functions of genetic effects (a_i and d_i respectively) as well as parental genetic distance such that the amount of either may not always be proportional to parental genetic distance.

A third parameter associated parental genetic distance is the number of trangressive segregants in a hybrid population. Assuming additive or dominant gene action for a given trait, then a high transgressive segregant has accumulated more alleles for increased expression of the trait than the high parent (or decreasing alleles than the low parent). The frequency of such progeny from parents with equal phenotypes should increase as parents become more diverse at loci that affect the trait. The frequency of transgressive segregants is also dependent on the distribution of the increasing and decreasing alleles between the parents: a parent with all decreasing alleles is quite distant from a parent with all increasing alleles but a cross of these parents would not produce any transgressive segregants. Thus these three parameters measured of hybrid populations that are affected by parental genetic distance are also affected by other factors and may not be directly proportional to the genetic distance of the parents.

The accuracy of the estimate of genetic distance between the parents will vary with the estimation method. One approach is to compare the parents on the basis of various heritable traits and attempts to measure genetic differences through phenotypes. This approach has been used with both quantitative traits such as morphological traits and with qualitative traits such as polymorphisms at isozyme and restriction fragment length polymorphism (RFLP) loci. Another approach is based on the examination of pedigree data under a set of assumptions concerning ancestral genealogy, ancestral genotypes and the genetic transmission between parents and offspring. These approaches, their assumptions, biases, and their applicability to a breeding situation are discussed below by the type of data that is used to determine the genetic distance. While there is a large body of research where inferred genetic distances are related to breeding behavior or where quantified genetic distances are used only to infer relationships among genotypes, this author has only attempted to review methods of estimating genetic distance and research where quantified genetic distances between parents were related to some form of breeding behavior in the hybrid population. The terms genotype and population are used synominously

II. Estimating Genetic Distance

A. Quantitative Data

Morphological data can be used to estimate the relationship between two populations by obtaining a reliable estimate of the phenotype of the populations for one (univariate) or more (multivariate) traits and statistically comparing them. This approach actually results in a statistical distance between the populations and inferences of a genetic distance are made under the assumption that that phenotypic differences reveal underlying genotypic difference. A large phenotypic difference will result in a large distance regardless of the extent of underlying genetic differences. Accurate inferences are more likely to be made when highly heritable traits are used and as more traits are examined since this will assay for genetic differences at more loci.

There are several statistical methods for estimating the genetic distance between two populations with phenotypic data. One of the simplest approaches is to calculate the Euclidian distance (ED) between the populations (Goodman, 1972) where the distance between the i^{th} and j^{th} population is

$$ED_{ij} = [\sum_{k=1}^{n} (X_{ik} - X_{jk})^2]^{1/2}$$

X_{ik}, X_{jk} = mean of the k^{th} of n traits for the i^{th} and j^{th} populations respectively that have been standardized by dividing by the standard deviation of k^{th} trait.

Each population can be visualized as a point in n dimensions corresponding to the n traits where the mean for the kth trait for the population is the coordinate for that point along the kth axis. ED_{ij} is the distance between the ith and jth point in this multidimensional space. ED_{ij} is also equivalent to the distance between the principal components derived from an analysis of the correlation matrix between the traits (Goodman, 1972) where

$$ED_{ij} = [\sum_{k=1}^{n} (Y_{ik} - Y_{jk})^2]^{1/2}$$

Y_{ik}, Y_{jk} = coordinates of the kth of n principal components for the ith and jth populations respectively.

In a principal component analysis the original n axes that corresponded to each trait are rotated in space to correspond to n new variates that are uncorrelated to each other. The population points remain in the same position as before but their locations in the n dimensional space are now defined by coordinates for the new axes. The new coordinates for the ith population are it's principal component scores.

The Euclidian distances such as Pearson's (1926) coefficient of racial likeness or Sokal's (1961) distance can be used when all the measured traits are uncorrelated (Goodman, 1972). This would seem to be an unlikely scenario especially when many traits are compared and ignoring the correlations could result in an exaggerated genetic distance when the correlations have a genetic basis and would be similar to giving some gene differences more weight than others in the distance measure.

Mahalanobis (1936) first addressed the problem of intercorrelated traits by calculating a generalized distance (D) where the Euclidian distance is adjusted by the common within-population correlation matrix such that the distance between the ith and jth population is

$$D_{ij} = [(\mathbf{X}_i - \mathbf{X}_j)' \mathbf{R}^{-1} (\mathbf{X}_i - \mathbf{X}_j)]^{1/2}$$

$$D_{ij} = [\sum_{k=1}^{n} \{(Y_{ik} - Y_{jk})^2 / \lambda_k\}]^{1/2}$$

$\mathbf{X}_i, \mathbf{X}_j$ = vectors of the standardized means of the n traits for the i^{th} and j^{th} populations
$\mathbf{R}^{-1}$ = inverse of the correlation matrix of the n traits
λ_k = eigenvalue of the k^{th} principal component

D is the multivariate generalization of "Student's" t test (Hotteling, 1954) which is commonly used to test the equality of two means and D^2 can be used to calculate Hotteling's T^2 test of the similarity of two mean vectors in conjunction with a multivariate analysis of variance. All eigenvalues equal unity when all traits are uncorrelated and D then equals ED. A problem with D is that principal components with very small eigenvalues and which therefore account for very little of the overall variability and may not be biologically significant, can have an inflated contribution to D, negating the contribution from more important principal components (Goodman, 1972).

An alternative is to include only some of the standardized principal components in the distance (Goodman, 1972). This approach projects the population point in a space with fewer but hopefully more biologically significant dimensions. Goodman (1972) suggested using the k^{th} principal component in the distance estimate only when $\lambda_k >= k$. A less conservative approach is to use all principal components where $\lambda_k >= 1$.

Calculating and interpreting these distances becomes more complex when there is heterogeneity between the covariance matrices of the populations being compared as this can result in D_{ij} differing from D_{ji} (Atchley *et al.*, 1982). The heterogeneity of covariance matrices reflects different relationships of the traits within different populations and may be due to biological differences brought about by genetic and developmental differences and the interaction of these factors with the environments where the traits were measured (Atchley *et al.* 1988). The interaction of the populations and the traits with the environment requires that all populations and traits be tested simultaneously in multiple environments and new populations could not be added to this data base and compared to the others without remeasuring the old populations.

When a breeder tries to improve a particular trait by crossing parents with high means for that trait then the parents will not show much phenotypic variability for that trait. In this situation the breeder would have to measure phenotypic variability in other morphological traits and hope that it is predictive of genetic variability for the desired trait. Narrow breeding populations may lack sufficient variability for other morphological traits for successful application of this approach especially if the other traits are correlated to the trait that is being improved. The use of quantitative traits has the potential to assay genetic distance at many loci as there are perhaps hundreds of gene differences between two populations that reside at the extreme ends of the expression range with multiple morphological traits, though it seems unlikely that this extreme range would exist in an elite breeding population.

Genetic distances are pairwise comparisons between populations but they can also be used to elucidate a broader relationship that may exist among all genotypes by coupling the resulting distance matrix with a clustering technique. This is a common practice in taxonomy and classification work that also can be applied to a breeder's parental populations.

There are numerous reports of the use of quantitative trait distance techniques in taxonomy, classification, and evolution studies but there are relatively few studies that have related these parental distances to any form of breeding behavior exhibited by a hybrid population. The morphological distance between parents has been positively correlated to the mid-parent heterosis for grain yield exhibited by hybrids in rapeseed (*Brassica napus* L.)(Lefort-Buson *et al.*, 1987), soybeans (*Glycine max* L. Merr.)(Chauhan and Singh, 1982), wheat (*Triticum aestivum* L.) (Shamsuddin, 1985; Cox and Murphy, 1990), dry beans (*Phaseolus vulgaris* L.) (Ghaderi *et al.*, 1984), peanut (*Arachis hypogaea* L.)(Arunachalam, 1984; Isleib and Wynne, 1983), maize (*Zea mays* L.) (Prasad and Singh, 1986) and tomato (*Lycopersicum esculentum* Mill.)(Maluf et al., 1983) while no relationship was found in faba bean (*Vicia faba* L.) (Ghaderi *et al.*, 1984) or oats (*Avena sativa* L.) (Cowen and Frey, 1987a). Cox and Murphy (1990) found that the value of their morphological data in predicting heterosis in F_2 wheat populations depended on the environment in which the data was collected.

There are few studies relating distances based on morphological data to genetic variance and the number of transgressive segregants in segregating populations despite the importance of these parameters to the plant breeder. The Euclidian distance between oat parents based on morphological data was negatively correlated with the number of transgressive segregants and the generalized variance (Cowen and Frey, 1987a) which is a measure of the overall genetic variance in a population comprised of the variation of all measured traits (Sokal, 1965; Goodman, 1968).

B. Qualitative Data

Actual genetic distances between parents can be estimated by using allele frequency data from simply inherited polymorphisms. In plants, suitable data can be derived from isozyme, RFLP, disease resistance, and some qualitative morphological trait loci, hereafter referred to collectively as marker loci (ML). This data has an advantage over quantitative morphological variation which can only infer genetic difference at an unknown number of loci as it directly assays for differences at known loci and provides a quantified estimate of actual genetic differences. The ML phenotypes are generally not influenced by the environment, thus eliminating a major limitation of morphological data and allowing new populations to be easily added to an existing data base of ML profiles.

The general approach to calculating a genetic distance from ML data is to determine the frequency of all ML alleles in the parental populations and comparing these ML profiles using a distance equation. An early statistic for measuring the distance between populations calculated the correlation (F_{st}) between random gametes within the two populations relative to the correlation of the gametes if the populations were combined. This was first proposed by Wright (1965) and modified by Nei (1965) for multiple alleles:

$$F_{st} = (- \sigma p_j p_k) / (p_j p_k)$$

$\sigma p_j p_k$ = covariance of the frequencies of the j^{th} and k^{th} alleles
p_j, p_k = mean frequencies of the j^{th} and k^{th} alleles for the two populations

A drawback to this statistic is that F_{st} can be negative.

Sokal and Sneath (1963) presented a formula for calculating the probability that a randomly sampled allele from one population is identical by state (IBS) to a randomly sampled allele from the same locus in the other population, averaged over all sampled loci (I_s):

$$I_s = (1/n) \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ija} \, p_{ijb}$$

p_{ija}, p_{ijb} = the frequency of the i^{th} of m alleles at the j^{th} of n loci in populations A and B respectively

I_s measures the probability of gametic similarity between populations and ranges from zero (no alleles in common) to unity when each population is fixed for the same alleles at all loci. Two populations which have equal p_{ij} values but where $0<p_{ij}<1$ will not produce I_s values of one and some populations with identical gene frequencies may appear more distant from one another by this measure than two populations which have different gene frequencies (Spiess, 1977). Nei's (1972) measure of the genetic identity (I_N) and distance (D_N) are commonly used to compare populations and will result in a identity score of unity (or a distance score of zero) when two populations have identical allele frequencies regardsless of the value of p_{ij}:

$$I_N = I_s / [\{ (\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ija}^2)(\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ijb}^2) \} / n^2]^{1/2}$$

$$D_N = -\log I_N$$

From a breeding stand point it is important to note that loci where $p_{ija}=p_{ijb}$ and $0<p_{ija}<1$ will still contribute to the heterozygosity of a hybrid between the two populations despite producing an I_N value of unity. If the parental populations are randomly mating then the heterozygosity of the hybrid population will be equal to that of it's parents and the I_N value of one accurately indicates that the hybrid population will be the same as the parents. When the the parental populations are inbred, an I_N value of unity for such loci will not reflect the heterozygosity that will occur in the hybrid population in proportion to allelic heterogeneity at the j^{th} locus. Upon inbreeding, the hybrid population will return to the genotype frequency at the j^{th} locus of the parents but there will have been an opportunity for recombination to produce new genetic variation when I_N at other loci is less than unity or when other loci with an unity I_N are heterogeneous. Sokal and Sneath's I_s measure of distance is predictive of hybrid heterozygosity and therefore would seem to be more relevant to plant breeders.

Hedrick (1970) proposed a measure called the probability of genotypic identity (IH_{jk}) which is based on genotype frequencies instead of allele frequencies.

$$IH_{jk} = \left(\sum_{i=1}^{n} p_{ij}p_{ik} \right) / \left[1/2 \left(\sum_{i=1}^{n} p_{ij}^2 + \sum_{i=1}^{n} p_{ik}^2 \right) \right]$$

p_{ij}, p_{ik} = frequencies of the i^{th} of n genotypes in the j^{th} and k^{th} populations

This statistic has certain appeal in evolutionary studies but seems less applicable to breeding situations where gamete frequency and thus allele frequency is a primary interest.

Principal component analysis of the correlation matrix of allele frequencies in conjunction with Goodman's approach can be used to produce a distance measure where the variables are uncorrelated. As with quantitative data, all of these distance measures are pairwise comparisons between two populations but the resulting distance matrices can be coupled with a clustering technique to ascertain an overall structure between all populations.

As with quantitative data, the effectiveness of this approach depends on sampling and the accuracy of estimates of genetic distance will increase when genomic diversity is assayed with more ML. The development of isozyme and RFLP technology has made this approach feasible in species where few polymorphic ML were previously available.

ML are generally regarded to be neutral in their effect on other traits and differences ML must accurately predict differences at the loci that affect the trait being improved to be of value to a breeder. The ability to predict heterogeneity at loci controling the object trait depends on the number and chromosome distribution of the loci controlling the trait, the number and distribution of the ML, and the degree of linkage disequilibrium between the two sets of loci. If the assayed ML are randomly sampled then the ML differences between populations should reflect their overall genomic diversity. The calculations of Chakraborty (1981) indicate that an unrealistically large number must be assayed to obtain an accurate estimate of genetic distance when genomes contain thousands of independent genes but with the organization of genes into linkage groups located on chromosomes, one ML may be able to assay for differences between other genes within it's linkage group. The use of RFLPs and saturated genetic maps it is now possible to use ML to assay for genetic diversity across virtually all chromosome segments. ML differences will more accurately assay differences at other loci when a linkage disequilibrium exists between the loci: a situation that is more likely to occur in autogamous species and in populations that have a narrow genetic base that further increases inbreeding.

There is accumulating evidence (Frei *et al.*, 1986; Graef *et al.*, 1989; Kahler and Wehrhahn, 1986; Nienhuis *et al.*, 1987; Osborn *et al.*, 1987, Stuber *et al.*, 1987) that ML heterogeneity in hybrid populations is frequently associated with quantitative trait variation, indicating that a linkage disequilibrium between ML and loci that affect the a quantitative trait can be considerable. As chromosome segments are identified that contain genes that affect the breeder's object trait then ML linked to those segments can be selected and assayed for polymorphisms. This approach will be more cost efficient and will result in a distance measure that is unbiased by data generated by ML not associated with relevant genes.

There have been several attempts, with mixed results, to correlate the performance of hybrids between inbred maize lines, which is generally dependent on the extent of heterosis, to a ML derived genetic distance between the parents. Price *et al.* (1986) and Lamkey *et al.* (1987) found a significant but low positive correlation between hybrid yield and inbred parental distances based on isozymes indicating that the diversity estimate had little value in predicting hybrid performance. Other researchers using isozymes (Hunter and Kannenberg, 1971; Hadjinov *et al.* 1982) and RFLPs (Godshalk *et al.*, 1990) have found insignificant correlations between hybrid yield and parental distances. Frei *et al.* (1986) found the ML distance between parents to be predictive of hybrid performance only when there was a known, pedigree relationship between the parents; a situation that increases inbreeding and the probability of linkage disequilibrium between ML and other genes. Lee *et al* (1989), using RFLPs to estimate parental diversity within and across chromosomes, found a significant correlation between parental diversity and hybrid yield and that the diversity of some chromosomes was more predictive of yield than other chromosomes. Their analysis of specific combining ability (SCA) produced similar results while others have reported no association of SCA and parental diversity based on isozymes (Hunter and Kannenberg, 1971; Heidrich-Sobrinho and Cordeiro, 1975; Hadjinov *et al.*, 1982).

A general conclusion from this work is that ML appeared to have some utility in assigning inbreds to general heterotic groups but that they had little value in predicting parameters such as heterosis, GCA, and SCA which are so important in maize. The work of Lee *et al.* (1989) suggests this approach will be more successful when only the diversity of relevant chromosome segments is used. No research has been performed relating ML derived distances to the amount of genetic variance or the number of transgressive segregants in hybrid populations.

C. Pedigree Data

The known genealogical relationship between individuals and populations can be used to infer the probability that they share common alleles at a locus. Wright (1920) used the complete pedigrees of two population to calculate a coefficient of probable relationship which not only described the genetic similarity between two populations, but also the inbreeding coefficient of their progeny. It is the probability that a randomly sampled allele from one individual (or population) is identical by descent (IBD) to an allele from the same locus from the other individual. Malecot (1948) called this value the kinship coefficient, Kempthorne (1969) refered to it as the coefficient of parentage (CP) while others have called it the coefficient of coancestry. (1-CP) is the theoretical distance between populations. The CP between populations A and B can be expressed as

$$r_{AB} = \sum_{i=1}^{k} (1/2)^{n+p+1} (1+fC_i)$$

where C_i, whose inbreeding coefficient is fC_i, is a common ancestor of A and B to the n_i and p_i degrees respectively and where there are k different exclusive ways of relating A and B to their common ancestor(s). Kempthorne (1969) presented a working formula of

$$r_{AB} = 1/2\ (r_{AX} + r_{AY})$$

where r_{AX} and r_{AY} are the CPs between A and the parents of B, X and Y. A CP value of one indicates the populations are identical while a CP of zero indicates that the populations theoretically share no IBD genes. The CP is easy to calculate and encompasses all of the genome thus eliminating the problem of sampling genes that exists with the other approaches.

A CP is based on the assumptions that (*i*) all ancestral parents are completely unrelated to one another (i.e. no genes in common), (*ii*) an estimate of the level of inbreeding in each ancestral parent and (*iii*) that each parent contributes equally to all offspring. The accuracy of the estimate depends on the validity of the assumptions. It is quite unlikely that the first assumption is true for all loci as it dictates that each ancestral parent posses an unique allele at each locus. The level of ancestral parent inbreeding can be assured in inbred lines though gametic heterogeneity can still occur if the ancestral parent is not a pure line, as is often the case with plant introductions and landraces.

The ancestral parent inbreeding can only be inferred for any locus in an ancestral parent where cross pollination occurs. It seems certain that the assumption of equal biparental contributions to all progeny must be violated to some extent in cultivars developed through many cycles of intense selection and particularly for loci that have a large and highly heritable effect on a selected trait. St. Martin (1982) estimated that in soybeans (2n=40), 88% of the lines derived from a biparental cross will obtain from 40 to 60% of their genes from one parent (assumed percentage is 50%) and that a breeder would be unlikely to select a line with even 70% of it's genes from one parent. These calculations indicate, at least for genes with a small effect on a selected trait, that the assumption of equal parental contribution may not be seriously violated. The validity of the assumptions will vary from trait to trait and from gene to gene. The violations may cancel out when averaged over all genes that affect a trait (unless there is a directional bias due to selection) suggesting that this approach may produce more accurate estimates of distance for traits that are controlled by many genes.

While the coefficient of parentage is usually calculated between the two populations being compared, one can also calculate the coefficient of parentage between each population and all the ancestral parents that appear in the pedigree of the populations (Souza and Sorrels, 1989). This method describes each population with a set of variables that estimate the probability that the population carries an allele the is IBD to that of an ancestral parent. Multivariate statistic techniques such as principal components and clustering can then be applied to this data set.

There has been very little work relating the CP between parents to the breeding behavior of their hybrid populations. A low, though significant, positive correlation was reported in rapeseed (Lefort-Buson *et al.*, 1987) while the association was not significant in oats (Cowen and Frey, 1987b) and wheat (Cox and Murphy, 1990). When predicting the heterosis in F_2 wheat populations, Cox and Murphy (1990) suggested using the CP to select diverse parents after the parental candidates had already been selected for their performance and phenotypic divergence. Cowen and Frey (1987b) found a significant association between the CP of the parents of segregating oat populations and the number of transgressive segregants for plant height but not for bundle weight, grain yield, straw yield, heading date, and harvest index.. They did find a significant association between the CP of the parents and the generalized genetic variance of the segregating population.

III. Application of Genetic Distances to Soybean Breeding

The genetic base of commercial soybeans is known to be quite narrow. The Committee on Genetic Vulnerability of Major Crops (1972) analyzed the pedigree of 62 northern cultivars and found that 88% of the parentage derived from only 12 plant introductions (PIs). Delanney *et al.* (1983) analyzed the pedigree of 158 US and Canadian northern cultivars (maturity group 00-IV) released from 1971 to 1980 and found that 50 PIs accounted for all the parentage with 10 of these accounting for 80% of the parentage. Two PIs, "Mandarin" and "Mandarin Ottawa" accounted for 30% of the parentage and while these were assumed to be genetically distinct, there is evidence that the may genetically similar (Cox *et al.*, 1985; Kiem *et al.*, 1989). From 1951 to 1980 the was an overall reinforcement of the predominance of certain major PIs in the adapted gene pool and little change in the constitution of the genetic base. The gene pool is even more restricted in southern adapted lines where 82% of the germplasm released from 1971-1980 can be traced to seven PIs. St. Martin (1982) calculated an average CP of 0.25 (i.e. equivalent to that expected among inbred half-sibs) among 25 cultivars released from 1976 to 1980 under the assumption that "Mandarin " and "Mandarin Ottawa" are identical while Cox *et al.* (1985) reported an average CP of 0.19 among 39 cultivar released from 1971 to 1981. The CPs represent the minimum distance between lines assuming that the PIs are completely unrelated, an assumption that may not be entirely valid in soybeans (Cox *et al.*, 1985; Delanney, 1983; Kiem *et al.*, 1989). Cox *et al.* (1985) also calculated a similarity index value between the cultivars in their study using qualitative data from isozyme and morphological loci and reported an average similarity index value of 0.63 for lines released from 1971-1981.

Ultimately the genetic base of commercial soybeans must be broadened by utilizing new PIs to continue to achieve breeding progress. At present only 50 of 8000 PIs have been extensively used and there are some PIs with acceptable agronomic merit. The initial use of a PI in a breeding program generally results in progeny that are unacceptable and further breeding is necessary to develop a PI-derived progeny that is agronomically desirable. In the meantime the breeder must continue to recombine the genes in the adapted gene pool to produce better cultivars. The narrow genetic base presents a problem as it is apparent from the reviewed research that a cross between two randomly selected adapted lines will result in considerable inbreeding in the resulting F_1 and this would reduce the number of segregating loci in subsequent selfed populations. It would be very useful to have a simple and accurate method of estimating the genetic distance between adapted soybean lines that would allow a breeder to maximize the use of the genetic diversity that does exist in this narrow gene pool.

Coefficient of parentage can be easily applied to adapted soybeans as there is extensive pedigree data available on most high yielding adapted lines. The major shortcoming with this technique in soybeans is the high probability that some PIs are related, a violation of one of the CP assumptions. A distance measure based on qualitative data could also be used with soybeans as the adapted soybeans exhibit a moderate amount of variability at a number of morphological, isozyme, and RFLP loci (Gorman, 1983; Doong and Kiang, 1987; Kiem *et al.*, 1989). The narrow genetic base and self pollinated nature of soybeans would increase the linkage disequilibrium between ML and associated quantitative trait loci such that dissimilarity at ML will be more likely to indicate dissimilarity at these linked loci than in a broad based random mating population.

List of References

List of References

Arunachalam, V., A. Bandyppadhyay, S.N. Nigam, and R.W. Gibbons. 1984. Heterosis in relation to genetic divergence and specific combining ability in groundnut (*Arachis hypogaea* L.). Euphytica 33:33-39.

Atchley, W.R., E.V. Nordheim, F.C. Gunsett, and P.L. Crump. 1982. Geometric and probabilistic aspects of statistical distance functions. Syst. Zool. 31:445-460.

Atchley, W.R., S. Newman, and D.E. Cowley. 1988. Genetic divergence in mandible form in relation to molecular divergence in inbred mouse strains. Genetics: 120:239-253.

Chakraborty, R.. 1981. The distribution of the number of heterozygous loci in an individual in natural populations. Genet. 98:461-466.

Chauhan, V.S., and B.B. Singh. 1982. Heterosis and genetic variability in relation to genetic divergence in soybean. Indian J. Genet. 42:324-328.

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

Cowen, N.M., and K.J. Frey. 1987b. Relationship between genealogical distance and breeding behavior in oats. Euphytica 36:413-424.

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.

Cox, T.S., and J.P. Murphy. 1990. The effect of parental divergence on F_2 heterosis in winter wheat crosses. Theor. Appl. Genet. 79:241-250.

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

Doong, J.Y.H., and Y.T. Kiang. 1987. Cultivar identification by isozyme analysis. Soy. Genet. Newsl. 14:189-225.

Falconer, D.S.. 1981. *Introduction to quantitative genetics*. Longman House, New York.

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.

Ghaderi, A., M.W. Adams, and A.M. Nassib. 1984. Relationship between genetic distance and heterosis for yield and morphological traits in dry bean and faba bean. Crop Sci. 24:37-42.

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

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

Goodman, M.M.. 1972. Distance analysis in biology. Syst. Zool. 21:174-186.

Gorman, M.B. 1983. An electrophoretic analysis of the genetic variationin the wild and cultivated soybean germplasm. Ph.D. dissertation, Univ. of New Hampshire.

Gorman, M.M., Y.T. Kiang, Y.C. Chiang, and R.G. Palmer. 1983. Electrophoretic classification of the early maturity groups of named soybean cultivars. Soybean Genet. Newsl. 9:143-156.

Graef, G.L., W.R. Fehr, and S.R. Cianzio. 1989. Relation of isozyme genotypes to quantitative characters in soybeans. Crop Sci. 29:683-688.

Hadjinov, M.I., V.S. Scherbak, N.I. Benko, V.P. Gusev, T.B. Sukhorzhevskaya, and L.P. Voronova. 1982. Interrelationship between isozymic diversity and combining ability in maize lines. Maydica 27:135-149.

Hedrick, P.W.. 1971. A new approach to measuring genetic similarity. Evol. 25:276-280.

Heidrich-Sobrinho, E., and A.R. Cordeiro. 1975. Codominant isoenzymic alleles as markers of genetic diversity correlated with heterosis in maize (*Zea mays*). Theor. Appl. Genet. 46:197-199.

Hotelling, H.. 1954 Multivariate analysis. P 67-80 *in* O. Kempthorne, T.A. Bancroft, J.W. Gowan, and J.L. Lush (eds.). *Statistics and Mathematics in Biology*. Iowa State College Press, Ames.

Hunter, R.B., and L.W. Kannenberg. 1971. Isozyme characterization of corn (*Zea mays*) inbreds and it's relationship to single cross hybrid performance. Can. J. Gent. Cytol. 13:649-655.

Isleib, T.G., and J.C. Wynne. 1983. Heterosis in testcrosses of 27 exotic peanut cultivars. Crop Sci. 23:832-841.

Kahler, A.L., and C.F. Wehrhahn. 1986. Associations between quantitative traits and enzyme loci in the F_2 population of a maize hybrid. Theor. Appl. Genet. 72:15-26.

Kempthorne, O.. 1969. *An Introduction to Genetic Statistics*. Iowa State University Press, Ames.

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

Lamkey, K.R., A.R. Hallauer, and A.L. Kahler. 1987. Allelic differences at enzyme loci and hybrid performance in maize. J. Hered. 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. Crp Sci. 29:1067-1071.

Lefort-Buson, M., Y. Datee, and B. Guillot-Lemoine. 1987. Heterosis and genetic distance in rapeseed (*Brassica napus* L.): use of kinship coefficient. Genome 29:11-18.

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

Malecot, G.. 1948. *Les Mathematiques de L'heredite*. Masson and Cie, Paris

Maluf, W.R., P.E. Ferreira, and J.E.C. Miranda. 1983. Genetic diversity in tomatoes and it's relationship with heterosis for yield in F_1 hybrids. Rev. Brasil Genet. 6:453-460.

Nei, M.. 1965. Variations and covariation of gene frequencies in subdivided populations. Evol. 19:256-258

Nei. M.. 1972. Genetic distance between populations. Am. Nat. 106:283-292.

Nienhuis, J., T. Helentjaris, M. Slocum, B. Ruggero, and A. Schaefer. 1987. Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. Crop Sci. 27:797-803.

Osborn, T.C., D.C. Alexander, and J.F. Fobes. 1987. Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor. Appl. Genet. 73:350-356.

Pearson, K.. 1926. On the coefficient of racial likeness. Biometrika 18:105-117.

Prasad, S.K., and T.P. Singh. 1986. Heterosis in relation to genetic divergence in maize (*Zea mays* L.). Euphytica 35:919-924.

Price, S.C., A.L. Kahler, A.R. Hallauer, P. Charmley, and D.A. Giegel. 1986. Relationship between performance and multilocus heterozygosity at enzyme loci in single-cross hybrids of maize. J. Hered. 77:341-344.

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

Shamsuddin, A.K.M.. 1985. Genetic diversity in relation to heterosis and combining ability in spring wheat. Theor. Appl. Genet. 70:306-308.

Sokal, R.R.. 1961. Distance as a measure of taxonomic similarity. Syst. Zool. 10:70-79.

Sokal, R.R.. 1965. Statistical methods in systematics. Bio. Rev. Camb. Phil. Soc. 40:337-341.

Sokal,R.R., and P.H.A. Sneath. 1963. *Principles of Numerical Taxonomy*. Freeman, San Francisco.

Souza, E., and M.E. Sorrels. 1989. Pedigree analysis of North American oat cultivars released from 1951-1985. Crop Sci. 29:595-601.

Spiess, E.B. 1977. *Genes in Populations*. John Wiley and Sons, New York.

Stuber, C.W., M.D. Edwards, and J.F. Wendel. 1987. Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and it's components. Crop Sci. 27:639-648.

Tanksley, S.D., H. Medina-Filho, and C.M. Rick. 1982. Use of naturally occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Hered. 49:11-25.

Vallejos, C.E., and S.D. Tanksley. 1983. Segregation of isozyme markers and cold tolerance in an interspecific backcross of tomato. Theor. Appl. Genet. 66:241-247.

20

Wright, S. 1917. Coefficients of inbreeding and relationship. Am. Nat. 51:545-559.

Estimation of the Genetic Structure of an Elite Soybean Population

Abstract

The source and frequency of genes, as well as the genetic distance between lines, was estimated with marker loci data and coefficient of parentage calculations for a set of 62 elite public and private soybean cultivars and breeding lines. Through CP and correlation analysis it was determined that genetic base of the elite population was derived from a total of 27 ancestral parents with the parents "Lincoln", "Mandarin Ottawa", "A.K. Harrow", and "Richland" contributing 30.6%, 17.7%, 10.6%, and 7.9% of the parentage respectively The marker loci analysis revealed no more than two alleles per locus within this population and the average probability of sampling an allele from one line that is identical by state to an allele sampled from another line was 0.64 while the average CP between any two elite lines was 0.26. While the data on average suggested a very narrow genetic base, pairs of elite lines could be identified that theoretically shared few genes and certain lines were identified as outliers. There was a poor correlation between the marker loci and CP estimates of pairwise distances and the overall relationship among the elite lines. Each method has it's own biases and it would be valuable to evaluate their accuracy in estimating the actual genetic distance between individuals in conjunction with unbiased estimates of the same. Selection of lines for inclusion in a breeding population based on a narrow range of maturity and yield performance criteria appeared to result in a more inbred population when compared to a general set of public cultivars released during the same time frame as the elite lines that had a broader maturity range and an average CP of 0.19 . Selection of lines for the elite population based on similar maturity could explain the restricted nature of the elite population versus the general population, rather than selection of lines for their similar high yield potential in Michigan.

Introduction

Breeders of autogamous crops tend to use parental lines that already have an acceptable phenotype for the trait to be improved. The mean of the progeny (u') of lines selected from a segregating population expressed in terms of gain from mass selection is equal to the mean of that population (u) plus the heritable portion of the selection differential.

$$u' = (u_s - u)\,h^2 + u$$

The use of superior parents, termed elite lines, generally assures a high u. A problem can arise with this approach when the selection of elite parental lines with similar phenotypes, for example high regional yield and adapted maturity, results in the selection of similar genotypes. This can reduce the amount of genetic variability that will be generated in segregating populations derived by crossing among elite lines and consequently reduce heritability and gain from selection.

This problem may be more pronounced in crops such as soybeans (*Glycine max* L. Merr.) where the adapted gene pool from which elite lines are derived has a narrow genetic base. Delanney *et al.* (1983) analyzed the pedigrees of 110 United States and Canadian cultivars released from 1971 to 1980 from maturity groups OO to IV and found that 50 plant introductions accounted for all the parentage, 10 of these plant introductions accounted for 88% of the pool, and two of these, "Mandarin" and "Mandarin Ottawa", made up 30% of the parentage. The southern gene pool had an even more restricted genetic base. St. Martin (1982) and Cox *et al.* (1985a) calculated an average coefficient of parentage among northern cultivars released in the 1980s of .25 and .19 respectively again illustrating the high degree of relationship among the adapted northern cultivars. It is likely that this gene pool is even narrower than this data indicates due to probable genetic relationships among the plant introductions as suggested by their similar geographic origins (Delanney *et al.*, 1983) and similar isozymes (Cox *et al.*, 1985) and restriction fragment length polymorphism patterns (Kiem *et al.*, 1989).

While soybean breeders have recently been using new plant introductions in their breeding programs to broaden the genetic base, the incorporation of such germplasm into the elite gene pool often takes multiple cycles of breeding and a breeder must continue to work within the existing elite population to develop agronomically superior cultivars. Thus it would be useful to estimate the genetic diversity within the elite soybean population and the genetic distance between it's members to maximize the use of the existing genetic diversity of this population and to hopefully maximize the breeding progress attained from this population. These parameters together can be termed the genetic structure of a population.

There are several ways to estimate the genetic distance between genotypes, each with it's own biases and assumptions. There are extensive pedigree records for adapted soybeans which can be used to calculate the coefficient of parentage (CP) (Kempthorne, 1969) between lines; a probability of genetic identity derived from assumptions of ancestral genealogy, inbreeding, and gene transmission. Another approach to estimating genetic distance is to use highly heritable phenotypes, such as isozymes, restriction fragment length polymorphisms, and certain morphological traits, as qualitative markers to assay for differences between genotypes at the DNA level. These polymorphisms occur with a moderate frequency in adapted soybeans (Apuya *et al.*, 1988; Doong and Kiang, 1987; Gorman, 1983; Kiem *et al.* 1989) and the accuracy of the distance estimate will depend on obtaining a representative sample of potentially polymorphic loci. These two approaches can also estimate allele and genotype frequencies within a population. A third approach infers a genetic distance from a statistical distance that is based on quantitative trait differences (Goodman, 1972) would seem less applicable in the narrowly defined elite soybean population where the range of phenotypes is quite restricted. Furthermore, each quantitative phenotype assays an unknown number of genetic differences that may not be proportional to the phenotypic difference and the effect of the environment on these traits requires that the phenotype of each individual be measured in multiple locations and years.

This research was undertaken to (i) see whether selection of elite parental lines from the northern adapted soybean gene pool based on regional performance and maturity resulted in a further restriction of the gene pool and (ii) to determine and compare the genetic structure of a set of elite lines using the CP and marker loci approaches.

Materials and Methods

Coefficient of Parentage Data

Sixty-two public breeding lines, public cultivars, and private cultivars (Table 1) were selected for inclusion in this study based on their performance over multiple years and locations in Michigan State University yield trials and were collectively termed the elite population. Forty-seven of the elite lines were in maturity group II, 14 were maturity group I and one was a group III maturity. Another set of lines termed the general population (Table 2) was formed from all maturity group 00 to group IV cultivars that were released after 1979 or that were used as a check variety in the USDA Northern Uniform Trials conducted from 1980 to 1990. This set was used to represent a broader population of adapted germplasm developed during the same time frame as the elite lines. All the elite public cultivars were also in this set.

The CP between two genotypes, defined as the probability that a randomly sampled allele from one genotype is identical by descent to a randomly sampled allele from the same locus from the other genotype, was calculated according to the formula of Kempthorne (1969)

$$r_{X,Y} = 1/2\,(r_{X,A} + r_{Y,B})$$

where $r_{X,Y}$ is CP between genotypes X and Y, genotypes A and B are the parents of genotype Y, and $r_{X,A}$ and $r_{X,B}$ are the CPs between genotype X and genotype Y's parents. The genotypes "Elgin" and "Kenwood" were derived from Iowa State University soybean population AP6 that was formed by an essentially random mating of 40 cultivars (Fehr and Ortiz, 1975). The CP between any genotype and one of these two lines was calculated as

$$r_{X,AP6} = 1/40\,(\sum_{i=1}^{40} r_{x,pi})$$

where $r_{X,AP6}$ is the CP between genotype X and a genotype derived from AP6 and $r_{x,pi}$ is the CP between X and the ith parent of the AP6 population. Furthermore, two of the 40 parents of AP6 were derived from a University of Maryland population formed by a random mating of eight parents and the CP between these two lines and any other lines was calculated in a fashion

Table 1. Summary of the source and name of the 62 elite lines, the average, minimum, maximum, standard deviation, and correlation of the coefficient of parentage (CP) and similarity index (SI) values for each line as compared to the other 61 lines.

		CP				SI				
Source	Name	Avg.	Min.	Max.	Std.	Avg.	Min.	Max.	Std.	Corr.
Pub. cultivars	BSR 101	.27	.14	.52	.07	.65	.45	.87	.08	.01
	BSR 201	.30	.09	.60	.09	.68	.37	.84	.10	.15
	BURLISON	.21	.11	.59	.09	.62	.37	.87	.11	.33**
	CENTURY	.29	.12	.72	.14	.68	.40	.95	.13	.37**
	CHAPMAN	.30	.10	.47	.09	.69	.45	.95	.11	-.05
	CONRAD	.21	.09	.38	.06	.66	.40	.87	.09	.43**
	CORSOY	.25	.01	.94	.16	.65	.50	.97	.09	.54**
	ELGIN	.24	.11	.64	.09	.70	.53	.95	.10	.45**
	HACK	.31	.04	.71	.12	.58	.42	.74	.08	.18
	HARDIN	.25	.04	.94	.15	.64	.47	.97	.10	.54**
	KENWOOD	.28	.08	.64	.11	.70	.48	.95	.11	.09
	PELLA	.32	.07	.63	.13	.64	.34	.90	.13	.51**
	SIBLEY	.21	.08	.43	.08	.55	.37	.76	.09	.23
	STURDY	.25	.12	.60	.09	.67	.53	.87	.08	.27*
	ZANE	.31	.10	.63	.08	.69	.47	.95	.09	.33**
Ill. A.E.S.	LN82-296	.29	.14	.70	.11	.68	.45	.95	.11	.46**
Ind. A.E.S	C1664	.24	.13	.56	.08	.67	.45	.90	.10	.40**
Iowa A.E.S	A80-147002	.32	.09	.63	.10	.61	.42	.82	.10	.42**
	A82-161035	.29	.12	.50	.07	.61	.50	.90	.08	.14
	A83-271010	.28	.12	.56	.07	.63	.45	.90	.09	.05
	A84-185032	.22	.14	.42	.05	.62	.42	.87	.11	-.25*
	A85-192034	.31	.08	.68	.10	.67	.45	.87	.09	.36**
	A85-291010	.30	.08	.73	.13	.69	.45	.90	.09	.14
	A85-292023	.28	.09	.58	.10	.58	.37	.79	.08	.02
	A85-293033	.28	.13	.45	.08	.58	.37	.90	.11	.08
	A86-102004	.28	.10	.63	.10	.66	.47	.92	.11	.20
	A86-103017	.24	.12	.60	.07	.65	.42	.84	.10	.14
	A86-103027	.31	.06	.71	.13	.61	.42	.76	.08	.29*
	A86-202026	.24	.12	.60	.07	.64	.42	.82	.10	.22
	A87-198005	.32	.09	.56	.09	.64	.40	.87	.09	.34**
	A87-296011	.31	.08	.56	.10	.66	.47	.87	.07	.06
	A87-297015	.19	.11	.42	.07	.61	.42	.76	.08	.03
Mich. A.E.S	E84098	.22	.09	.51	.09	.65	.42	.90	.11	.25*
	E84150	.24	.11	.54	.08	.67	.50	.84	.09	.32**
	E84159	.23	.07	.59	.08	.60	.37	.79	.09	.36**
	E84165	.23	.07	.59	.08	.64	.40	.82	.09	.34**
	E85100	.25	.06	.45	.09	.60	.40	.87	.09	.47**
	E85166	.29	.15	.61	.08	.61	.40	.79	.10	.49**
	E85168	.29	.15	.61	.08	.65	.40	.87	.09	.28*
	E86339	.26	.12	.62	.08	.67	.50	.87	.09	.38**
	E86348	.29	.12	.62	.09	.68	.45	.87	.09	.24
	E87223	.24	.09	.59	.09	.63	.42	.76	.08	.22
	E88080	.28	.14	.63	.09	.68	.47	.84	.09	.21
Minn. A.E.S.	M82-946	.23	.13	.38	.05	.56	.34	.76	.09	.25*

Table 1. Continued

		CP				SI				
Source	Name	Avg.	Min.	Max.	Std.	Avg.	Min.	Max.	Std.	Corr.
	M84-916	.25	.10	.37	.06	.62	.40	.82	.10	.39**
Ohio A.E.S.	HW8008	.26	.13	.72	.12	.65	.45	1.00	.12	.37**
Callahan Seeds	1250	.23	.14	.42	.06	.61	.37	.90	.12	.16
	7260	.26	.13	.40	.06	.64	.37	.87	.09	.03
	7299	.15	.09	.35	.06	.61	.42	.82	.09	.17
	8252	.28	.13	.61	.08	.60	.37	.79	.10	.26*
Pioneer	9271	.24	.09	.50	.07	.64	.42	.92	.10	.26*
	9292	.24	.09	.50	.07	.66	.40	.92	.10	.24
Asgrow Seeds	A1937	.30	.06	.73	.14	.67	.47	.92	.10	.28*
	A2234	.30	.01	.61	.13	.65	.45	1.00	.12	.18
	A2943	.19	.10	.35	.04	.60	.37	.74	.08	.04
Agripro	AP 1989	.26	.10	.57	.09	.59	.45	.87	.09	.37**
Dairyland Seed	DSR-262	.24	.06	.60	.09	.67	.40	.84	.11	.29*
Jacques Seed	J-231	.26	.09	.66	.10	.68	.53	.90	.08	.24
Northrup King	s1884	.30	.08	.59	.10	.61	.42	.82	.08	.51**
	s19-90	.18	.04	.51	.09	.60	.37	.79	.11	.31*
	s23-12	.13	.06	.31	.05	.59	.37	.79	.10	-.02
	s2596	.22	.09	.31	.05	.68	.47	.90	.11	.28*
Over all comparisons		.26	.01	.94	.08	.64	.34	1.00	.10	.27**

Corr. = correlation between the coefficient of parentage and similarity index for that elite line.

*, ** denote significant correlations at the 0.05 and 0.01 levels of probability respectively.with 59 df.

Table 2. Summary soybean cultivars included in the general population.

Maturity Group					
IV	III	II	I	0	00
Avery	Bass	Amcor	BSR 101	Aries	Bicentennial
Douglas	BSR 302	Amsoy	Hardin	Chico	Maple Amber
Eygptian	Cartter	Beeson	Hodgson	Clay	Maple Arrow
Flyer	Chamberlain	Bell	Kasota	Dassel	Maple Isle
Franklin	Cumberland	BSR 201	Kato	Dawson	Maple Ridge
Hamilton	Edison	Burlison	Lakota	Evans	McCall
Lawerence	Fayette	Century	M82-106	Glenwood	Portage
LN83-2356	Fremont	Chapman	M83-899	Libra	Proto
Morgan	Harper	CN210	Sibley	Maple Donovan	Scorpio
Pennyrile	Hayes	CN290	Sturdy	Minatto	
Pershing	Jack	Conrad	Weber	Ozzie	
Pyramid	Logan	Corsoy		Piceses	
Regal	Mead	Elgin		Simpson	
Sparks	Pella	E86339			
Spencer	Resnik	Hack			
Stafford	Sherman	Kenwood			
Union	Williams	Marcus			
	Winchester	Miami			
	Zane	Newton			
		Platte			
		Preston			
		Vinton			
		Wells			
17[a]	19	23	11	13	9

[a] number of genotypes listed in the column.

similar to that of the AP6 derived lines. The following assumptions were made in the CP calculations: (i) all plant introductions are completely unrelated, including lines selected from other plant introductions, (ii) all plant introductions were completely inbred, and (iii) that each parent of a biparental cross contributed equally to all progeny derived from the cross. A genotype derived by five or more backcrosses was considered to be genetically equivalent to the recurrent parent. The pedigree information used in the CP calculations was obtained from public sources and in a few instances, the private companies that developed the genotype.

The CP between an elite line and an ancestral parent is the probability that the elite line carries an allele from that ancestral parent at any particular locus but it can also be thought of as the percentage of the elite line's genes that derive from that ancestral parent. The CP between an elite line and an ancestral parent, hereafter referred to as the ancestral parent contribution, were calculated between each elite line and each ancestral parent found in the pedigrees of the 62 elite lines, producing a multivariate data set describing each elite line with p variables where p is the number of ancestral parents. This data set was used to estimate the genetic constitution of the elite population and to examine the interrelationship of the elite lines through principal component analysis of the variance-covariance matrix and cluster analyses. The data was also analyzed for correlations between the ancestral parent contributions. All CPs were calculated using computer programs written by the author in the SAS Interactive Matrix Language. All other analyses were performed using SAS software.

Marker Loci Data

Ten seeds of each of the 62 elite lines were tested for their genotypes at the following thirteen isozyme loci: diaphorase (*Dia*$_1$ locus; EC 1.6.4.3), endopeptidase (*Enp* locus), isocitrate dehydrogenase (*Idh*$_1$, *Idh*$_2$ loci; EC 1.1.1.42), mannose phosphate isomerase (*Mpi* locus; EC5.10.11), phosphoglucosemutase (*Pgm*$_1$ locus; EC 5.3.1.9), and superoxide dimutase (*Sod* locus; EC 1.15.1.1) using the D buffer system of (Cardy and Beversdorf, 1984) and acid phosphatase (*Acp* locus; Ec 3.1.3.2), aconitase (*Aco*$_2$, *Aco*$_4$ loci; EC 4.2.1.3), fluorescent esterase (*Fle* locus), 6-phosphogluconate dehydrogenase (*Pgd_1* locus; EC 1.1.1.44), and phosphoglucose isomerase (*Pgi* locus; EC) using a Tris-Citrate system (.2 M Tris, .062M citric acid, pH 6.7 electrode buffer/ .0075 M Tris, .002 M citric acid, pH 6.7 gel buffer). All gels were 11.5 % (w/v) starch. An additional ten seeds were analyzed if heterogeneity was found within a line.

To prepare electrophoretic samples, seeds were imbibed in distilled water for 24 hours after which half of one cotyledon was removed and ground in 0.25 ml of grinding buffer (16.6% (w/v) sucrose, 8.3% ascorbic acid, pH 7.4). D system gels were run for 6 hours at 275 volts/50 mAmps while the Tis-citrate gels were run for six hours at 200 volts/50 mAmps. The stains and procedures of (Cardy and Beversdorf, 1984) were used for all isozymes except florescent esterase which was resolved with a stain using 40 ml of 0.1 M sodium acetate-HCl buffer (pH 5.2) and 15 mg of 4-methylumbelliferyl acetate, dissolved in 10 ml of acetone. *Fle* Bands were then visualized under longwave UV light.

The genotype was determined for each elite line for the *W, T, L, I,* and *R* loci that control the color of the flower, pubescence, pod, hilum and hilum respectively. The level of peroxidase activity was assayed using the method of Buttery and Buzzel (1968) under the assumption of single gene inheritance of this polymorphism (Buttery and Buzzel, 1968).

No more than two alleles were found for any locus among the 62 elite lines such that the frequency of one arbitrarily chosen allele (p_i) could be used to represent the genotype of a line for the j^{th} locus in the marker loci data set . Loci that exhibited heterogeneity within a line were assigned a p_i of 0.5 under the assumption that the heterogeneity was a result of heterozygosity at that locus in the single plant from which the line was bulked and that assaying a larger sample would have produced an actual p_i of 0.5 (Appendix A). A similarity index (SI) equivalent to the probability that an allele randomly sampled from one genotype is identical by state to an allele randomly sampled from the same locus from the other genotype was calculated between all elite lines. The SI with this data was

$$SI = (1/n) \sum_{i=1}^{2} \sum_{j=1}^{n} p_{ijx}\, p_{ijy}$$

where p_{ijx} and p_{ijy} are the frequencies of the i^{th} allele at the j^{th} of n loci in genotype X and Y respectively. The SI is equivalent to the I_s measure reported by Sokal and Sneath (1963) and is equivalent to the numerator of Nei's index before normalization (Nei, 1972). The allele frequency data was also used in principal component, principal coordinate, and cluster analyses. The SI and principal coordinate analyses were performed using a programs written by the author in SAS Interactive Matrix Language while principal component and cluster analyses were performed with standard SAS software.

Results

Coefficient of Parentage

The average CP between the elite lines was .26 (Table 1) (Appendix B) and pairwise CPs ranged from a low of 0.01 between "Corsoy" and A2234 to a high of 0.94 between "Hardin" and it's recurrent parent in three backcrosses, Corsoy. Some lines were more related to the elite population as a whole than others as shown by the average CP for individual elite lines that ranged from 0.13 for Northrup King s23-12 to 0.32 for A87-198005 and "Pella" (Table 1). Every elite line could be paired with at least one other elite line to produce a CP of less than 0.16 . Removing the five elite lines that appeared to be least related to the population from the data set increased the average CP between the remaining elite lines to 0.27 (Table 3).

The average CP between the general set of group 00-IV maturity lines was 0.19 (Table 3). Subdividing this population by maturity groups showed that the the average CP between lines within maturity groups was about equal to that between the elite lines and that the bulk of the diversity among these lines was distributed between lines with different maturities (Table 3).

The use of multivariate statistic techniques facilitated the investigation of the overall relationship between the elite lines. While cluster analysis could have been done on the 62 x 62 matrix of CPs between the elite lines, a more informative analysis was performed by using the multivariate ancestral parent contribution profile of the elite line as this allowed an analysis of the genetic base of the elite population and the source of the variation between the elite lines.

This analysis showed that 34 plant introductions contributed genes to the elite population (excluding the donor parents of lines derived from five

Table 3. The average coefficient of parentage (CP) between lines of the general population and it's maturity subdivisions and between the elite lines with or without a subset of the elite lines that appeared to be outliers.

Population		CP	n[a]
General	00-IV	.19	92
	0-III	.22	66
	I-III	.25	53
	I-II	.24	34
	00	.15	9
	0	.24	13
	I	.27	11
	II	.24	23
	III	.35	19
	IV	.25	17
Between general lines of same maturity		.27	
Between general lines of different maturity		.17	
Elite lines		.26	62
Elite lines - outliers[b]		.27	56

[a] number of lines in the population or subdivision.

[b] outlier were the elite lines s19-90, s23-12, 7299, A87-297015, Corsoy, and Hardin.

or more backcrosses) with 15 of these contributing at least 1% of the parentage (Figure 1) while the other 19 contributed a total of only 3.8% of the parentage. The first twelve plant introductions accounted for 91.4% of the parentage. There were many significant positive correlations (Table 4) between the plant introduction contributions to the elite population that could be traced to the correlated plant introductions being used primarily in unison as parents in only one cross. This means that the genes from the correlated ancestors shared a common pedigree as a route of introduction to the elite gene pool. Coupled with progeny inbreeding, this represents a further restriction on the elite gene pool as an inbred genotype can pass on to it's progeny an allele from only one of it's parents. For example, the contributions from "Mandarin" and "Manchu" were perfectly correlated and each was used only once in forming the elite population, as the parents of the cultivar "Lincoln". As an inbred line, Lincoln can pass on to subsequent progeny and elite lines only one type of allele, derived from either Mandarin or Manchu. Thus the contributions from Mandarin and Manchu to the elite population can be summed and appropiately termed the contribution from Lincoln in recognition of Lincoln as the sole source of these genes.

All ancestral parent contributions with correlations of unity were pooled as were the contributions from "Illini" and "Dunfield" ($r = 0.99$) and termed the contributions from the primary derived progeny (Table 4). While not pooled, the contributions from "No. 171" and "A.K. Harrow" also followed a restricted entry route into the elite population. Both were parents of "Capital" which is a parent of Corsoy, while A.K. Harrow (along with "Mandarin Ottawa") was also a parent of "Harosoy" which in turn was the other parent of Corsoy. The major source of entry of Harosoy genes to the elite population is through the extensive use of Corsoy as a parent indicating a common entry route for the genes of No. 171 and A.K. Harrow.

Figure 2 shows the genetic constitution of the elite population after the pooling of correlated ancestral contributions which resulted in the parentage of the elite population being effectively derived from only 27 ancestral parents or their derived progeny with Lincoln accounting for 30.3% of the parentage and the first nine genotypes accounting for 91.4% of the elite population's parentage.

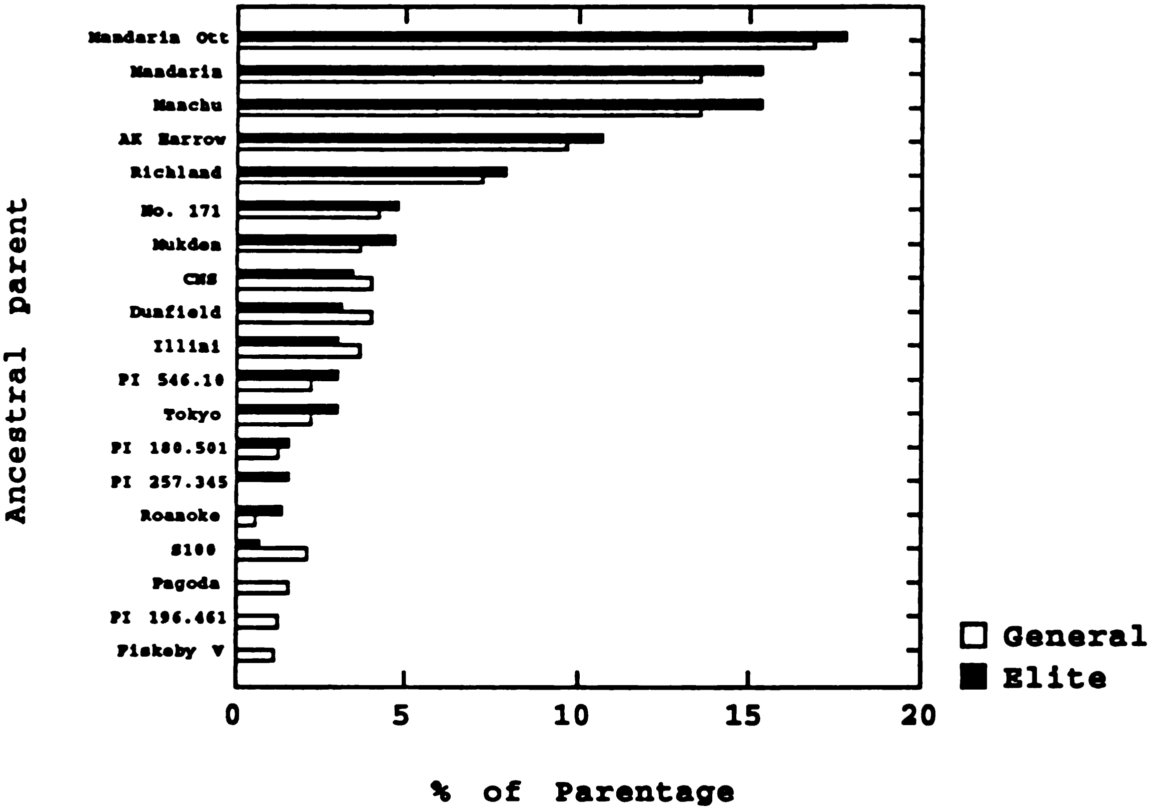

Figure 1. The percentage of the elite and general populations's parentage that derived from the ancestral parents that contributed at least 1.0% of the parentage.

Table 4. Summary of significant correlations between the contributions from ancestral parents (APs) to the elite population (r_e) and the broader population ($r_{00\text{-}IV}$), their contributions to the elite population (APC) along with the primary progeny line derived from the correlated ancestral parents.

AP1		AP2	r_e	$r_{00\text{-}IV}$	APC1[a]	APC2	Total	Progeny Line
Mandarin	with	Manchu	1.00	0.00	15.30	15.30	30.60	Lincoln
Tokyo	with	PI 546.10	1.00	0.00	2.95	2.95	5.90	Ogden
Patoka	with	PI 840.41	1.00	0.00	.40	.40	.80	Perry
FC31745	with	PI 171.442	1.00	0.00	.13	.06	.19	Tracy
FC33243	with	C143	1.00	0.00	.09	.04	.13	
FC33243	with	Mansoy	1.00	0.00	.09	.04	.13	
Mansoy	with	C143	1.00	0.00	.04	.04	.08	
Flambeau	with	Pagoda	1.00	0.00	.04	.03	.07	
Illini	with	Dunfield	.99	0.00	3.02	3.12	6.14	Adams
A.K. Harrow	with	No. 171	.94	0.00	10.55	4.81	15.36	Capital
Roanoke	with	PI 257.345	.93	0.00	1.26	1.61	2.87	NKs1346
S100	with	Patoka	.89	0.00	.72	.40		
S100	with	PI 84041	.89	0.00	.72	.40		
Richland	with	Mukden	.86	0.00	7.88	4.74		
Haberlandt	with	FC 31745	.84	0.00	.10	.13		
Haberlandt	with	PI 171.442	.84	0.00	.10	.06		
Mandarin Ottawa	with	A.K. Harrow	.74	0.00	17.67	10.55		Harosoy
Ralsoy	with	Patoka	.72	0.00	.05	.40		
Ralsoy	with	PI 840.41	.72	0.00	.05	.40		
Mandarin Ottawa	with	Manchu	-.71	0.00	17.66	15.30		
Mandarin Ottawa	with	Mandarin	-.71	0.00	17.66	15.30		

[a] APC1 and APC2 are the ancestral parent contribution to the elite population from ancestral parents 1 and 2 respectively.

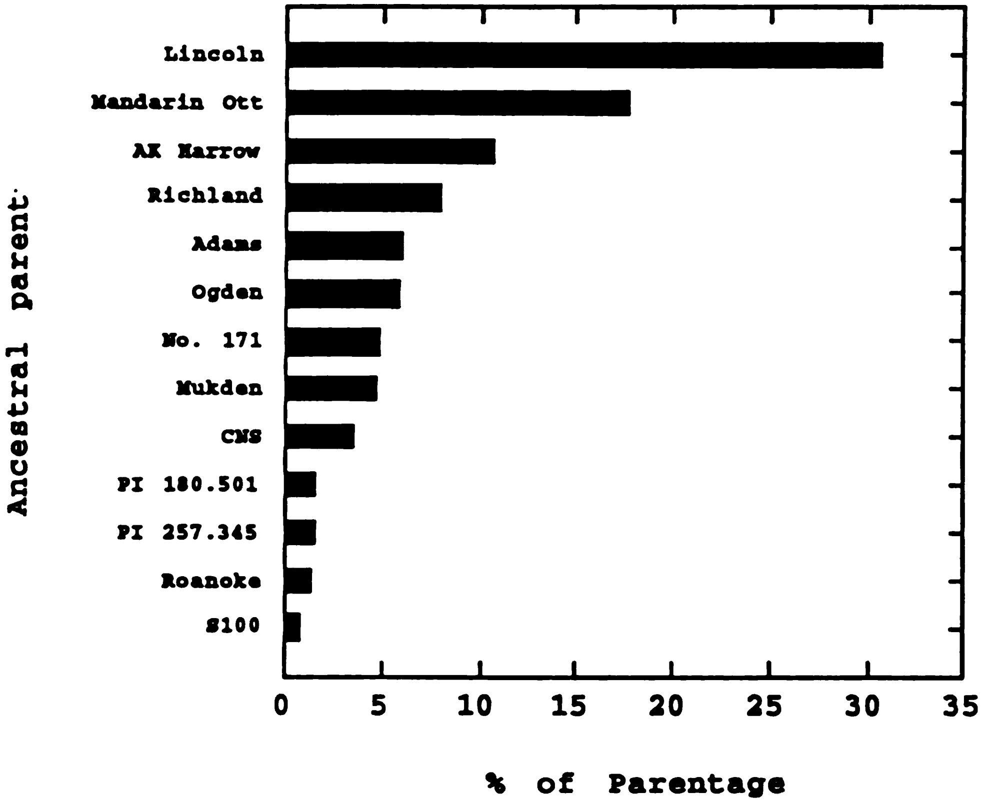

Figure 2. The percentage of the elite population's parentage that derived from the ancestral parents that contributed at least 1.0% of the parentage.

A principal component analysis was performed on the variance-covariance matrix of the ancestral contribution data set to summarize the source of the variability among the elite lines. The first three principal components accounted for 86.4% of the variability among the elite lines and the position of the elite lines relative to these coordinates is shown in Figure 3. The first principal component accounted for 56.9% of the variability and the correlation of the original ancestral parent contributions with the scores for this principal component (Table 5) indicated that it primarily seperated the elite lines based on what proportion of their genome originated from Lincoln or from A.K. Harrow, Mandarin Ottawa, and No. 171 which are the parents of Harosoy and Corsoy. This showed that a major source variation among the elite lines can be attributed to what proportion of their parentage derived from these four ancestral parents that combined accounted for 58.6% of the elite parentage. The second principal component separated lines primarily based on their genomic content from the ancestral parents "Roanoke" and PI 257.345, or from No. 171 and A.K. Harrow while principal component three separated the elite lines primarily based on the proportion of their genome derived from "Ogden", "Richland", and "Mukden" or from Roanoke and PI 257.345. PI 257.345 was only found in the pedigree of Northrup King s1346 (along with Roanoke) which in turn is one of the parents of only four elite lines.

Cluster analysis can incorporate all dimensions of a multivariate data set in it's resulting dendogram while principal component analysis can only graphically present three dimensions at best. A cluster analysis of the ancestral parent contribution data set was performed using various linking methods using only those variables found to significantly contribute to the variability of the data set through the principal component analysis. The results showed that while cluster membership was fairly constant across linking methods, this being especially true for tight groupings, there was considerable variation regarding the position of the tight groups and outlying genotypes relative to one another. Thus the association between many of the elite lines became dependent on the biases of the linking method and it appeared to be a very subjective method of determining the genetic distance between any two elite lines. The cluster analysis seemed to have more utility in graphically presenting an overall relationship between all the elite lines and only the clustering produced using Ward's linking

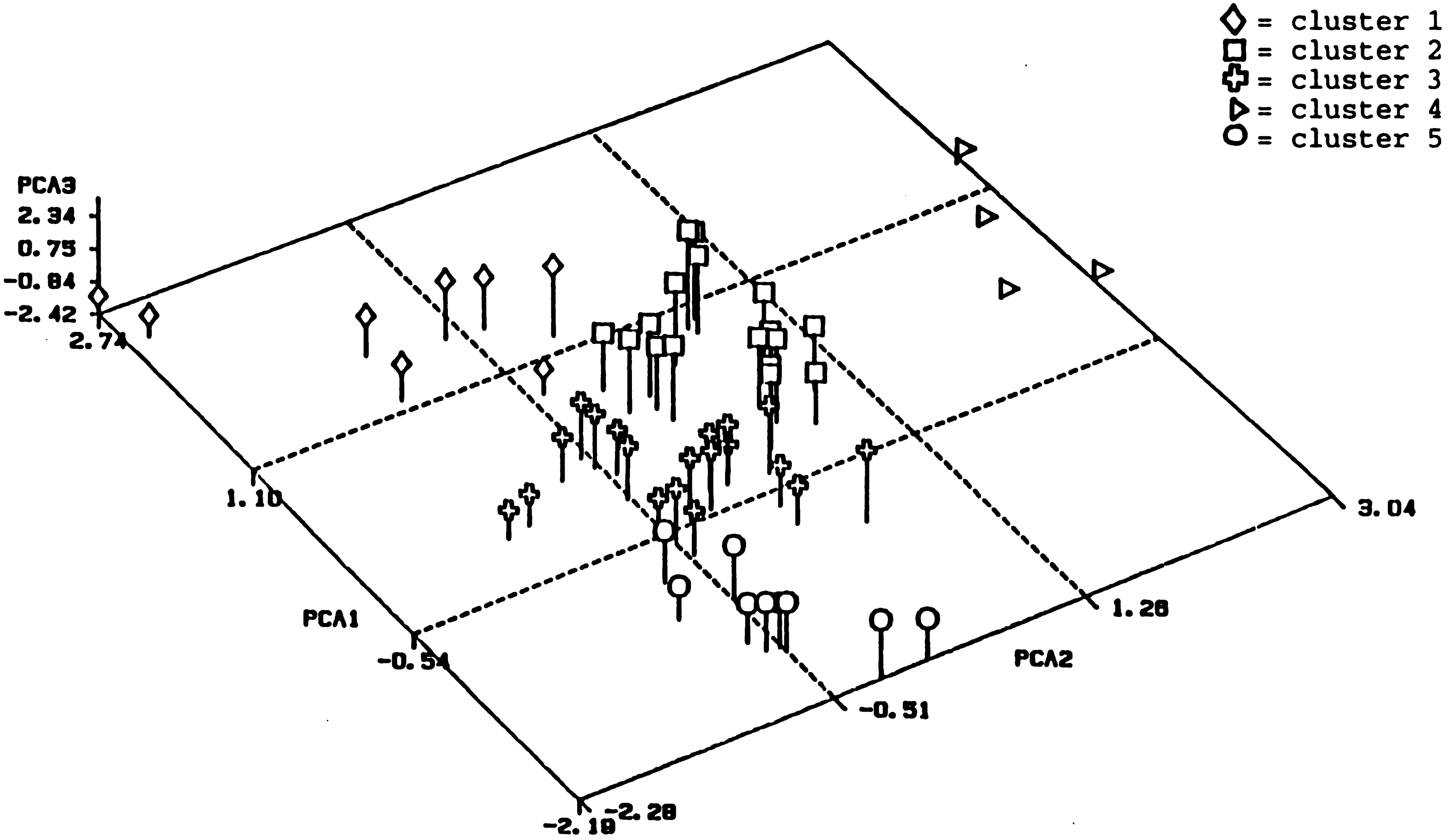

Figure 3. Plot of the first three principal components (PCA1, PCA2, and PCA3) from the principal component analysis of the ancestral parent contribution data of the 62 elite lines. Shapes correspond to the five clusters derived from hierarchical clustering of the elite lines using the same data and Ward's clustering criterion (see Figure 4).

Table 5. Correlations of the contributions from ancestral parents that were a source of at least 1% of the genes in the elite population with the principal component scores from the analysis of this data set along with the the percentage of the variation in this data set that is accounted for by the first three principal components.

Ancestral Parent	APC[a]	Correlation with Principal Component 1	2	3
Lincoln	30.6%	-.97**	-.22	.03
Mandarin Ottawa	17.7%	.84**	-.34**	.36**
A.K. Harrow	10.6%	.71**	-.65**	-.21
Richland	7.9%	-.14	.52**	.52**
Adams	6.1%	-.36**	.22	-.00
Ogden	5.9%	.14	.36**	.69**
No. 171	4.8%	.52**	-.70**	-.46**
Mukden	4.7%	.36**	.58**	.53**
CNS	3.4%	-.56**	-.09	-.21
PI 180.501	1.6%	-.01	.14	.13
PI 257.345	1.6%	.17	.75**	-.54**
Roanoke	1.3%	.16	.78**	-.56**
% of total variation accounted for by the principal component		56.87%	18.91%	10.33%

[a] Ancestral parent contribution to the elite population.

** Denotes a significant correlation at the 0.01 probability level and 60 df.

method which minimizes within cluster variation while maximizing intercluster variation is presented (Figure 4) as this method seemed to produce the most logical association between the elite lines based on obvious pedigree relationships. For example, Corsoy figured prominantly in the pedigrees of E87223, "Sturdy", "Sibley", J-231, and AP 1989 and Ward's linking algorthm placed them in the same cluster while other methods positioned Corsoy as an outlier from all elite lines. The validity of this structure is also suggested by Figure 3 where members of the same cluster were grouped together in the graph of the first three principal components and by an analysis of the ancestral parent contribution profile of each cluster (Table 6) that showed that the two clusters that were most divergent from the the rest of the lines were at the extremes in terms of the amount of their genes that derive from the ancestral parents that were highly correlated with the first principal component.

Marker Loci Data

The average SI value among all pairwise comparisons of the elite lines was 0.64 (Table 1)` (Appendix C) indicating that an F_1 obtained by mating two randomly selected elite lines would be homozygous at 64% of it's loci. The SI values ranged from a low of 0.34 between Pella and M82-946 to a high of 1.00 (identical, homogeneous genotypes) between HW8008 and A2234 (Table 1). The average SI value for individual elite lines compared to all others ranged from a low of 0.55 for Sibley to a high of 0.70 for Kenwood and Elgin (Table 1). Some of Sibley's uniqueness could be attributed to the fact that it is the only elite line that carries the *Sod* allele while the similarity of Elgin and it's progeny, Kenwood, to many elite lines may be due to their derivation from the broad based AP6 soybean population. Every elite line could be paired with at least one other to produce a SI value of 0.53 or less.

No more than two alleles were found among the elite lines for any locus. The frequency of the allele choosen to represent each locus in the marker loci data set is presented in Table 7 along with the number of elite lines that were heterogeneous at the associated locus. Forty-eight percent of the elite lines were heterogeneous at at least one locus. Fifty-five percent of the breeding lines were heteogeneous at an average of 1.9 loci per line while 42% of the released cultivars were heterogeneous at an average of 1.1

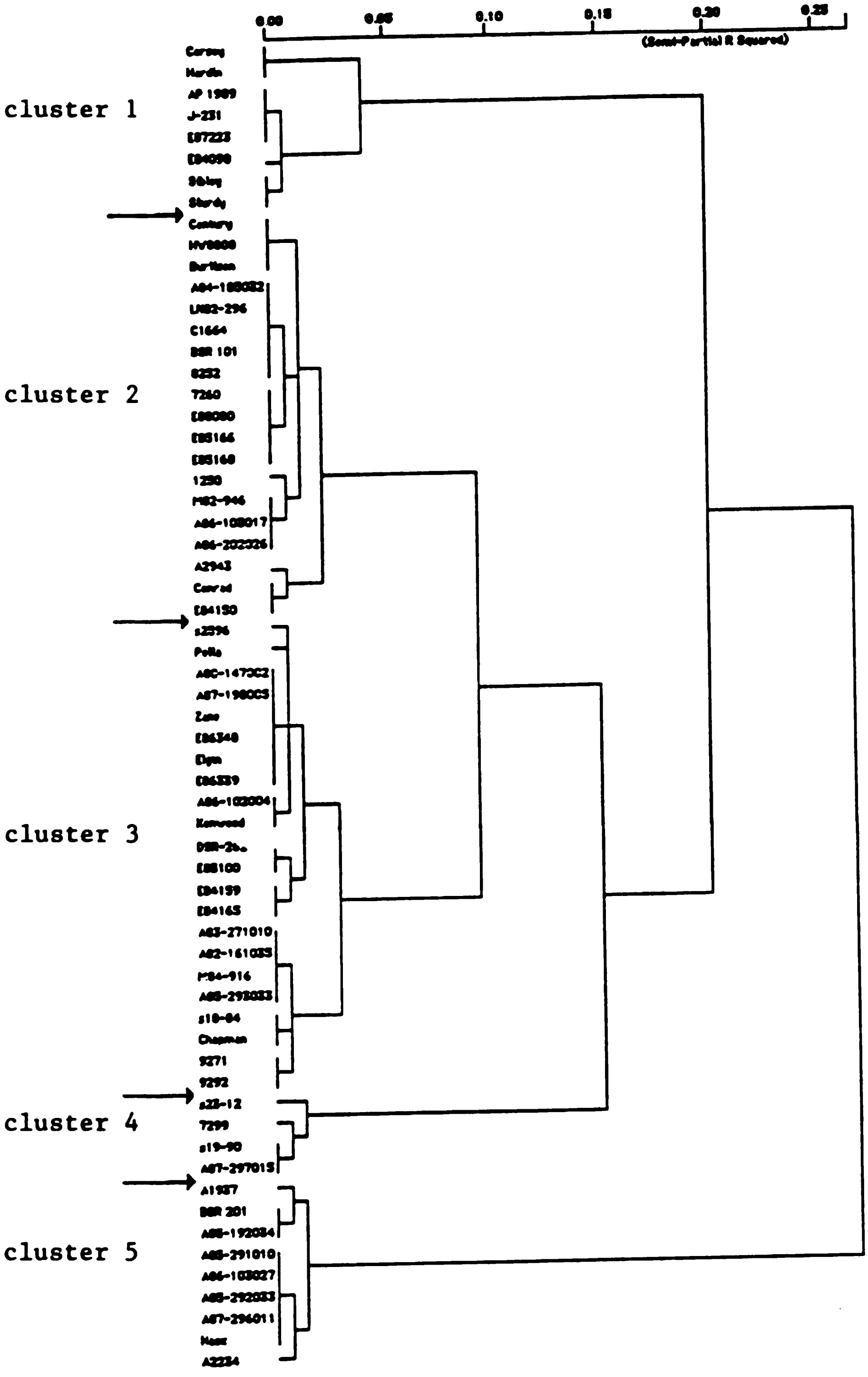

Figure 4. Results of clustering the 62 elite lines using the ancestral parent contribution data set after pooling the highly correlated contributions and using Ward's minimum variance clustering criterion.

Table 6. Ancestral parent contributions to the five clusters of elite lines obtained by Ward's clustering method.

	Cluster				
Ancestral Parent	1	2	3	4	5
Lincoln	11.17[a]	24.81	35.66	14.21	50.36
Mandarin Ottawa	28.70	21.65	15.35	9.96	10.61
AK-Harrow	24.29	8.53	10.28	4.88	6.32
Richland	4.74	10.22	6.42	10.21	8.63
Adams	1.28	5.88	8.39	5.47	6.51
Ogden	4.49	9.97	4.37	4.69	3.58
Mukden	3.78	8.39	2.44	8.98	2.67
No. 171	14.75	1.32	5.58	1.56	2.78
PI 257.345	0.00	0.00	0.00	25.00	0.00
CNS	0.96	2.94	4.32	0.98	5.08
Roanoke	0.49	0.70	0.50	12.50	0.00
PI 180.501	3.16	0.99	0.91	1.56	3.47

[a] Ancestral parent contributions are expressed as percentage of the cluster's genome.

Table 7. Summary of allele frequencies and heterogeneity in the elite population for nineteen marker loci and the correlation of the frequencies with the first three principal components.

Trait	Allele	Allele frequency	n[a]	Principal Component 1	2	3
Flower color	W_1	.73	4	-.17	-.55**	.45**
Pubescence color	t_1	.45	0	.67**	.24	.16
Pod color	*br*	.67	3	.61**	-.12	-.11
Hylum color	*i*	.74	0	-.50**	.18	-.27*
Hylum color	*r*	.52	0	.42**	.19	-.37**
	Idh_1	.75	5	-.21	.17	.02
	Idh_2	.41	2	-.47**	.05	-.01
	Aco_4	.60	1	.09	-.17	-.11
	Aco_2	.95	0	-.21	-.06	-.06
	Mpi_1	.35	4	.25*	.58**	.12
	Pgm_1	.40	5	.18	-.02	-.77**
	Dia_1	.65	3	-.34**	-.55**	.33**
	Enp	1.00	0	.00	.00	.00
	Per	.35	3	.64**	-.10	.33**
	Sod	.02	0	.09	.24	-.15
	Fle	.36	12	.46**	-.16	.33**
	Pgi	.53	4	-.19	.76**	.33**
	Acp	1.00	0	.00	.00	.00
	Pgd_1	.84	2	.17	-.17	.06

*, ** denote significant correlations at the 0.05 and 0.01 probability levels respectively with 60 df.

[a] number of elite lines that were heterogeneous at the locus.

loci per line. The high frequency of heterogeneous loci in these supposedly inbred lines may be attributed to the fact that half are breeding lines that have not yet gone through a final purification and that even then, biochemical homogeneity is rarely ascertained nor selected for. *Enp* and *Acp* were monomorphic in the elite population while *Sod* and Aco_2, were represented by an alternate allele in only one and two elite lines respectively.

The variance-covariance matrix of the allele frequency data was used in a principal component analysis that resulted in the first three principal components accounting for only 39.0% of the variation, indicating that graphing the observations in these three dimensions would not produce a reliable picture of their actual relationship. The correlations of the marker loci to the principal componets (Table 7) did not appear to produce a clear interpretation of the new axses. To try and achieve a better data summary, a principal coordinate analysis was performed on a 62 x 62 distance matrix between the elite lines with 1 - SI being the distance measure. This multivariate technique attempts to produce a low-dimensional plot in Euclidian space where the proximity of the points approximates their original distance. The first three principal coordinates accounted for 59.1% of the variation and the postion of the elite lines relative to these coordinates are shown in Figure 5. The elite lines did not appear to form mutiple groups in this space, rather they were simply dispersed around the centroid with a few outlying elite lines. Cluster analysis of this data also failed to produce distinct groupings within the elite population that was independent of what linking method was used. The dendogram that resulted from the average linking of elite lines is shown in Figure 6. It's validity seemed to be supported by a similar association of points in the principal coordinate graph (Figure 5).

The correlation of the CP and SI values between pairs of elite lines was 0.27 (Table 1) and while this is highly significant with 1879 df, it indicated that one measure was not very predictive of the other suggesting that each measured genetic distance independently. For individual elite lines compared to all others, the correlations ranged from -0.25 for A84-185032 to 0.54 for Corsoy and Hardin (Table 1). The reason for the variation is not apparent. More evidence for the independence of these distance measures came from representing the points in the principal component

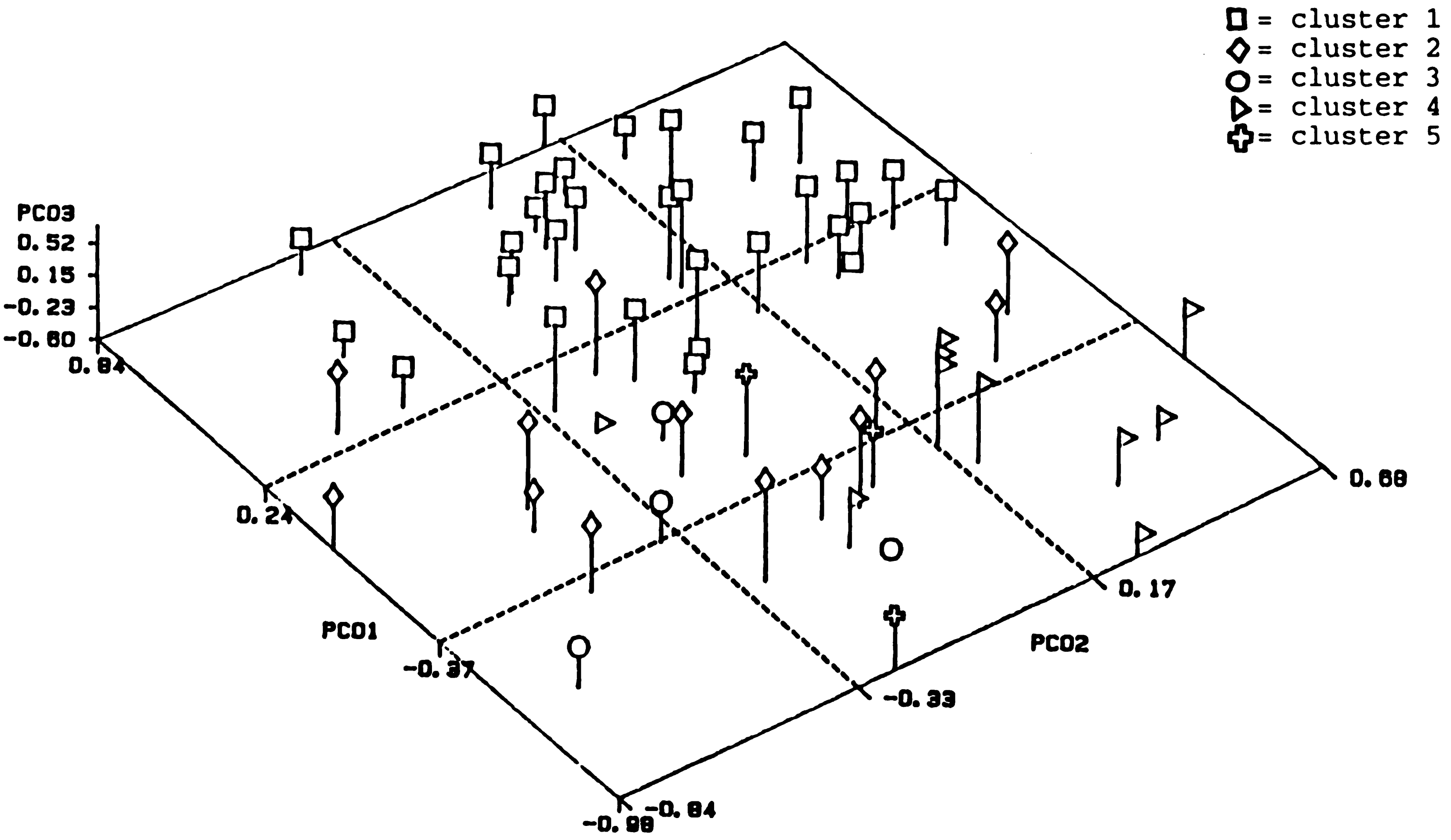

Figure 5. Plot of the first three principal coordinates (PCO1, PCO2, and PCO3) from the principal coordinate analysis of the marker loci data of the 62 elite lines. Shapes correspond to the five clusters derived from hierarchical clustering of the elite lines using the same data set and the average clustering criterion (see Figure 6).

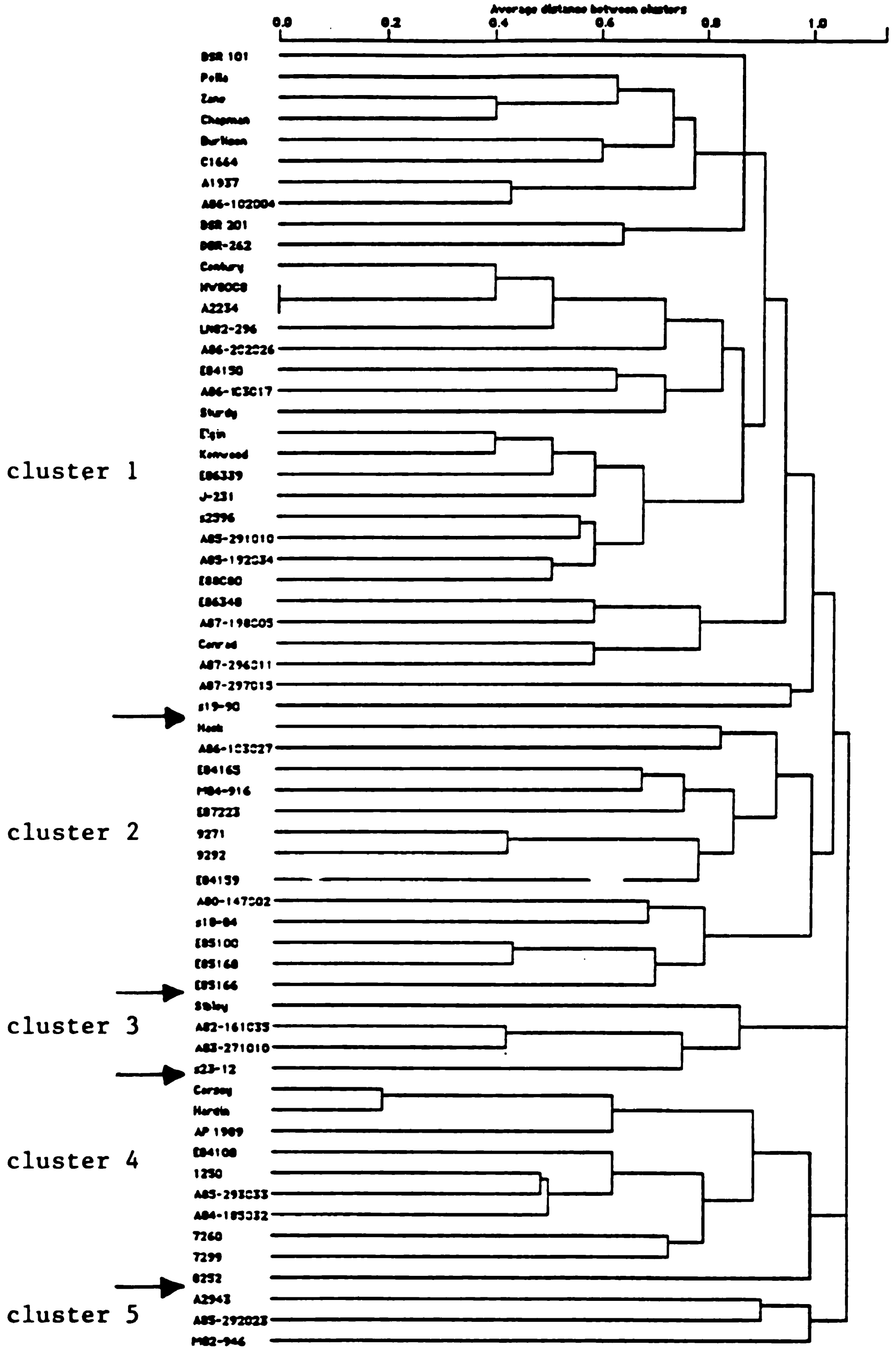

Figure 6. Results of clustering the 62 elite lines using the marker loci data set and the average clustering criterion.

and coordinate graphs of the ancestral parent contribution and marker loci data respectively according to the clustering pattern produced by the other data set. Figure 7 shows the members of a marker loci cluster dispersed across the space defined by the first three principal components of the ancestral parent contribution analysis while Figure 8 shows the memebers of the ancestral parent contribution clusters dispersed across the space defined by the first three principal coordinates of the marker loci analysis indicating that the groupings achieved by one data set differed from those produced by the other data set.

Discussion

The elite lines can be thought of as a selected portion of the soybean germplasm developed from 1980 to 1990 and the CP data suggested that the elite population was more inbred (CP = 0.26) (Table 1.) than a general group of cultivars ranging in maturity from group 00 to group IV (CP = 0.19) (Table 3) that were developed during the same time frame. Cox *et al.* (1985a) reported an average CP of 0.19 between maturity group 00-IV soybeans released from 1971 to 1980 while St. Martin (1982) calculated an average CP of 0.25 between lines of similar maturity released from 1976-1980 under the assumption that the ancestral parents Mandarin and Mandarin Ottawa were indentical. The increased inbreeding of the elite population appeared to result from the narrow maturity range represented in the elite population, and not from selection for high yield in Michigan per se as the average CP between the elite lines was equivalent to that between the unselected maturity group I-III or group I-II lines (Table 3). This analysis indicated that the genetic diversity among adapted northern soybeans was primarily distributed between, rather than within maturity groups and that many unselected genotypes could be paired to a completely unrelated line outside of their own maturity group while this occurred only once when comparisons were made within maturity groups.

The elite soybean population consisted of a set of highly related lines as shown by an average CP among the elite lines of 0.26 (Table 1), or slightly higher than that expected between half-sibs derived from unrelated inbred parents. This means that under the assumptions of the CP that the average F_1 obtained by mating two randomly selected elite lines would be

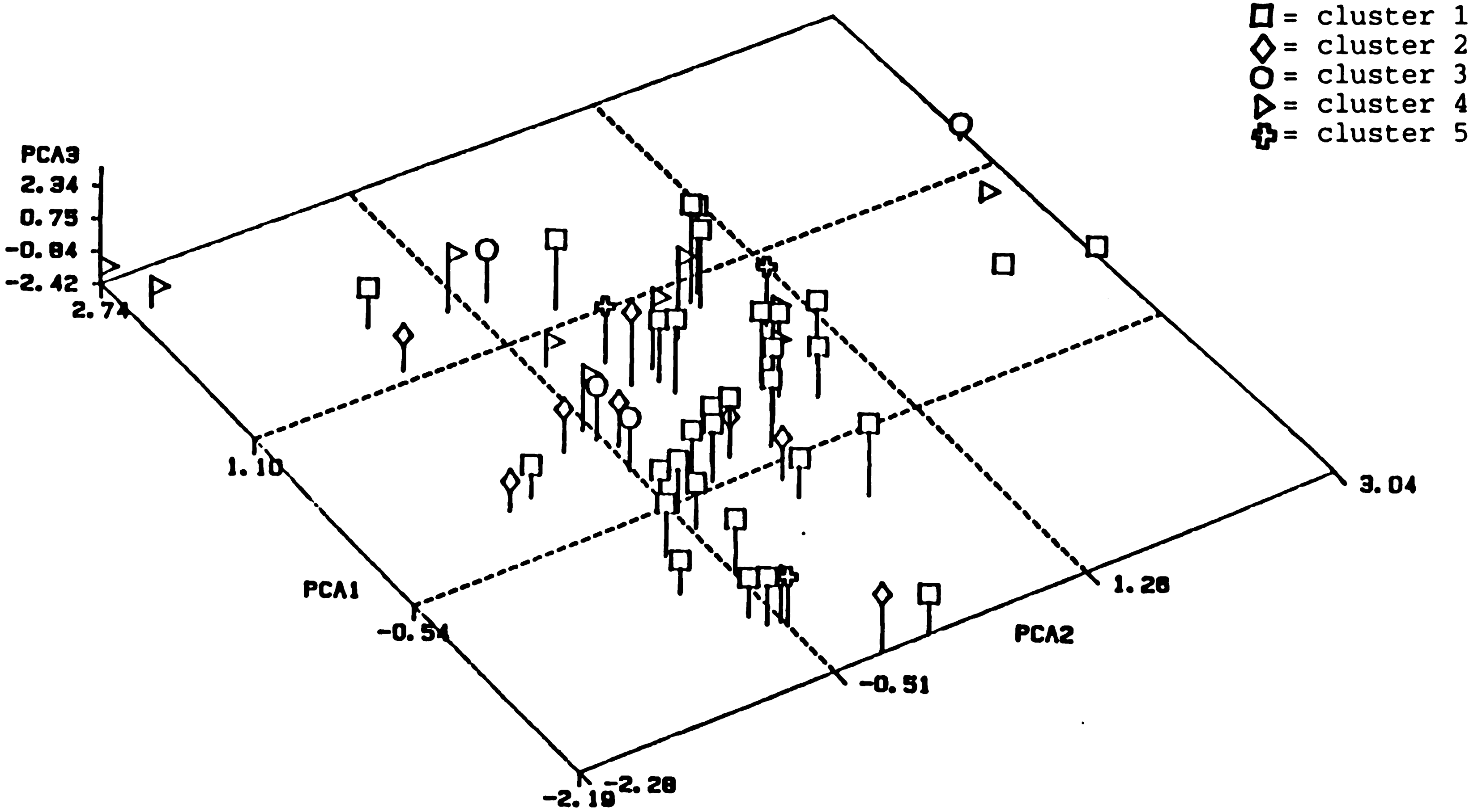

Figure 7. Plot of the first three principal components (PCA1, PCA2, and PCA3) from the principal component analysis of the ancestral parent contribution data of the 62 elite lines. Shapes correspond to the five clusters derived from hierarchical clustering of the elite lines using the marker loci data set and the average clustering criterion (see Figure 6).

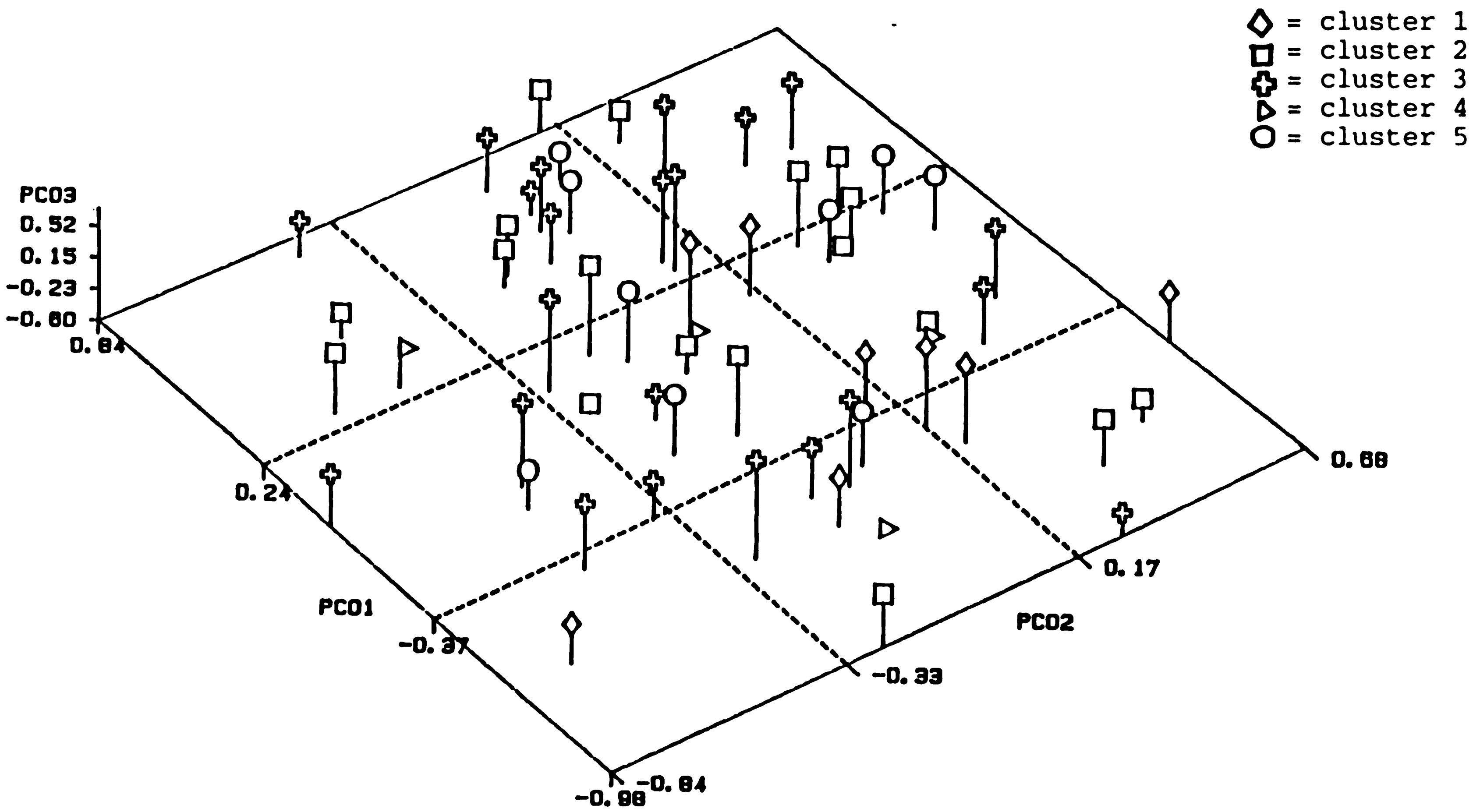

Figure 8. Plot of the first three principal coordinates (PCO1, PCO2, and PCO3) from the principal coordinate analysis of the marker loci data of the 62 elite lines. Shapes correspond to the five clusters derived from hierarchical clustering of the elite lines using the ancestral parent contribution data set and Ward's clustering criterion (see Figure 4).

homozygous at 26% of it's loci, leaving the remaining 74% to generate the genetic variability in the succeeding generations of selfed progeny. The probable genetic relationships among the ancestral parents (Delanney *et al.*, 1983; Cox *et al.*, 1985a; Kiem *et al.*, 1989) violate the CP assumptions and would result in a actual genetic similarity between elite lines being greater than this estimate, perhaps approaching the 0.64 value estimated by the marker loci based SI which makes no such assumptions (Table 1).

The restricted nature of this genetic base was also illustrated by the analysis of the ancestral parent contributions that showed that almost 67% of the parentage derives from only four ancestral lines with Lincoln contributing 30.6% of the parentage (Figure 2). The significant correlations between unpooled ancestral parent contributions (Table 4) indicated that the genetic base is effectively narrower than suggested by the profile shown in Figure 2 . This profile was remarkably similar to those reported by Delanney *et al.* (1983) for maturity group 00 to IV varieties released from 1971 to 1981 and from 1961-1970 indicating that any efforts in the last ten years to broaden the genetic base of northern adapted soybeans have yet to appear in the parentage of this elite population. The CP data only reflects the frequency that breeders have used the ancestral lines or their derived progeny as parents in crosses and does not reflect the effects of the breeder's selection among the progeny of a cross on the genetic makeup of the elite population. All CP calculations assumed no selection or sampling and St. Martin (1982) indicated that when averaged over all genes that the deviations from the expected value would be small. Yet there certainly must be some biases toward the genes of one parent over the other, especially for genes with major effects on a selected trait and these biases coupled over successive cycles of selection could significantly alter the genetic constitution of the elite population from that predicted by CP analysis. Indeed if the ancestral parent profiles have remained relatively unchanged over the past thirty years then the yield of adapted soybeans would have remained unchanged without effective progeny selection.

While the average values from all analyses of CP and SI data showed a narrow genetic base, each analysis was also able to detect considerable variation between the elite lines. Each elite line could be paired with another elite line to produce a CP and SI value that were considerably lower than the means (Table 1) and the low average CP and SI of some elite lines compared to all others suggested that certain lines were quite diverse from the main body of the population. The graphs of the principal component (Figure 3) and coordinate analyses (Figure 5) performed on the ancestral parent contribution and marker loci data sets respectively also illustrated this diversity and showed the existance of outlying genotypes.

The principal component analysis of the ancestral parent contributions to the elite popualtion suggested that much of the diversity among the lines is due to what proportion of their parentage derived from the ancestral parents that were the major sources of genes, particularly Lincoln and Mandarin Ottawa, A.K. Harrow, and No.171 which are the parents of Harosoy and it's progeny Corsoy (Table 5). This is illustrated by the differences in the ancestral parent contribution profiles of the groupings obtained from Ward's clustering of this data, where the percentage of parentage derived from Lincoln or from Mandarin Ottawa, A.K. Harrow and No. 171 are 11.2% and 67.7% respectively for cluster one and 50.4% and 19.7% respectively for cluster five (Table 6). The isozyme analysis of the genetic diversity of the elite population found the minimum amount of allelic diversity indicating that the genetic diversity within this population had an interlocus versus an intralocus source (Table 7). These results indicated that apparently diverse genotypes can perform well in Michigan and that it may be possible to generate variable progeny populations with high mean yields from which better progeny could be selected by crossing selected elite lines.

Both the CP and marker loci indicated that gentic diverisity existed in the elite population yet both appeared to measure the diversity between individuals independently as shown by the low correlation between the CP and SI measures (r = 0.25) (Table 1.) and by Figures 7 and 8 where the grouping of the lines by one data set fails to correspond to the grouping obtained with the other. Cox *et al.* (1985a) reported a correlation of 0.63 between the CP and a similarity index based on 20 marker loci from a set of soybean cultivars released from 1971 to 1981 though the correlation was only 0.48 among lines that were released from 1961 to 1970 and only 0.24 among lines released before 1961. A low though significant correlation has also been found between such estimates of genetic similarity in cultivated *Triticum aestivum* L. (Cox *et al.*, 1985b). The independence of these estimates undoubtably arises from the different assumptions and type of data used in the calculations. The CP is a probability based solely on assumptions and is only as valid as those assumptions, while the SI is a probability based on actual genetic differences and will be as accurate as the data and it's assumptions allows. An illustration of the differences between these measures is provided by the fact that twenty-seven ancestral parents contributed genes to the elite population (after the pooling of contributions suggested by the correlation analysis) and the CP assumed that each contributed an unique allele to the population at every locus while the marker loci analysis discovered no more than two alleles at any locus within the elite population (Table 7). Other researchers using isozymes (Chiang, 1985; Doong and Kiang, 1987; Gorman, 1983; Doong and Kiang, 1987; Chiang, 1985) and RFLPs (Apuya *et al.*, 1988; Kiem *et al.*, 1989) have also failed to find any allelic diversity approaching the level assumed by the CP even across *Glycine* species. The SI on the other hand assumed that alleles in different individuals that produced the same phenotype are identical though there can be cyptic variation between them (Ramshaw *et al.*, 1979) that would cause the SI to overestimate the genetic similarity between lines. The CP assumed that all progeny from a cross are genetically equivalent while the SI can reflect the effect of selection and random sampling on the genotype of the derived lines though it is not known whether the allele frequencies for the elite population reported in Table 6 reflect these effects or not. The accuracy of the SI estimate of genetic distance is also dependent on sampling considerations such as the number of genes in the genome, the number and distribution of the assayed loci, the size of linkage blocks, and the extent of the linkage disequilibrium between ML and the loci within these blocks. The SI accuracy could be improved by assaying more loci and it's relation to genetic diversity for a particular trait could be further improved by assaying marker loci that are known to be linked to chromosome segments that affect that trait (Lee *et al.*, 1989).

It is possible that a index of similarity that utilizes both types of data could provide a more accurate estimate of genetic similarity than either alone as suggested by Cox *et al.* (1985a) who noted that the biases of these measures would tend to cancel out each other. There are a number of ways that these measures could be combined, ranging from simply adding the CP and SI estimate to combining the multivariate ancestral parent contribution and marker loci data sets and calculating a distance based on a principal component analysis of the data (Goodman, 1972). It would be useful to attempt to integrate the two data types in light of some measurement of actual genetic distance between individuals to see what measure is most predictive.

List of References

List of References

Apuya, N.R., B. Frazier, P. Kiem, E.J. Roth, anf K.G. Lark. 1988. Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* L. Merrill.. Theor. Appl. Genet. 75:889-901.

Buttery, B.R., and R.I. Buzzell. 1968. Peroxidase activity in seeds of soybean varieties. Crop Sci. 8:722-725.

Cardy, B.J. and W.D. Beversdorf. 1984. A procedure for starch gel electrophoretic detection of isozymes of soybean (*Glycine max* L. Merr.). Departemnt of Crop Science Technical Bulletin 119/8401. University of Guelph, Guelph Ontario, Canada, N1G 2W1.

Chiang, Y.C.. 1986. Genetic and quantitative variation in wild soybean (*Glycine soja*) population. Ph.D. dissertation, Univ. of New Hampshire.

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

Cox, T.S., G.L. Lookhart, D.E. Walker, L.G. Harrel, L.D. Albers, and D.M. Rodgers. 1985b. Genetic relationships among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polyacrylamide gel electrophoretic patterns. Crop Sci. 25:1058-1063.

Delanney, X., D.M. Rodgers, and R.G. Palmer. 1983. Realtive contribution among ancestral lines to North American soybean cultivars. Crop Sci. 23:944-949.

Doong, J.Y.H., and Y.T. Kiang. 1987. Cultivar identification by isozyme analysis. Soy. Genet. Newsl. 14:189-225.

Fehr, W.R., and L.B. Ortiz. 1975. Registration of a soybean germplasm population. Crop Sci 15:739.

Goodman, M.M.. 1972. Distance analysis in biology. Syst. Zool. 21:174-186.

Gorman, M.B. 1983. An electrophoretic analysis of the genetic variation in the wild and cultivated soybean germplasm. Ph.D. dissertation, Univ. of New Hampshire.

Kempthorne, O.. 1969. *An Introduction to Genetic Statistics.* Iowa State University Press, Ames.

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

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. Crp Sci. 29:1067-1071.

Malecot, G.. 1948. *Les Mathematiques de L'heredite.* Masson and Cie, Paris

Nei. M.. 1972. Genetic distance between populations. Am. Nat. 106:283-292.

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

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

Sokal,R.R., and P.H.A. Sneath. 1963. *Principles of Numerical Taxonomy*. Freeman, San Francisco.

Appendices

Appendix A. Summary of the frequency of the monitored allele in the 62 elite lines at the 19 marker loci.

Name	W_1	t_1	*br*	*i*	*r*	Idh_1	Idh_2	Aco_4	Aco_2	Mpi_1	Pgm_1	Dia_1	*Enp*	*Per*	*Sod*	*Fle*	*Pgi*	*Acp*	Pgd_1
A80-147002	0.0	1.0	0.0	1.0	0.5	1.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
A82-161035	0.0	1.0	1.0	1.0	0.5	1.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	0.5	0.0	0.0	1.0
A83-271010	0.0	1.0	1.0	1.0	0.5	1.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
A84-185032	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	0.5	0.5	0.5	1.0	1.0	0.0	1.0	0.0	0.0	1.0
A85-192034	0.5	0.0	1.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
A85-291010	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
A85-292023	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0
A85-293033	0.5	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.5	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0
A86-102004	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.5	0.0	1.0
A86-103017	1.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.5	1.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0
A86-103027	0.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
A86-202026	1.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.5	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
A87-198005	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.5	1.0	0.0	0.0
A87-296011	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0
A87-297015	1.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0
AP 1989	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0
A1937	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
A2234	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
A2943	1.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.5	0.5	0.0	1.0
BSR 101	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	0.5	1.0	0.0	0.0	0.0	1.0	0.0	1.0
BSR 201	1.0	0.0	1.0	1.0	0.5	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.5	0.0	1.0	0.0	0.0	1.0
BURLISON	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.5	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0
C1664	0.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
1250	1.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0
7260	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.5
7299	1.0	1.0	1.0	0.0	0.0	0.5	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0
8252	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0
CENTURY	1.0	0.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
CHAPMAN	1.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
CONRAD	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.5	1.0	0.0	1.0
CORSOY	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0
DSR-262	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0
E84108	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0
E84150	1.0	0.0	1.0	1.0	0.0	1.0	0.5	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0

Appendix A. Continued

Name	W_1	t_1	*br*	*i*	*r*	Idh_1	Idh_2	Aco_4	Aco_2	Mpi_1	Pgm_1	Dia_1	*Enp*	*Per*	*Sod*	*Fle*	*Pgi*	*Acp*	Pgd_1
E84159	0.0	1.0	0.0	1.0	0.5	1.0	0.0	0.0	1.0	0.0	0.5	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0
E84165	0.0	1.0	1.0	1.0	0.5	1.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0
E85100	0.5	0.0	0.0	1.0	0.0	0.5	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.5	1.0	0.0	1.0
E85166	0.0	0.0	0.0	1.0	0.0	0.5	0.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.5
E85168P	1.0	0.0	0.5	1.0	0.0	1.0	0.0	0.0	1.0	0.5	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0
E86339	1.0	0.0	0.5	1.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.5	0.5	0.0	1.0
E86348	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0
E87223	1.0	1.0	0.5	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.5	0.0	1.0	1.0	0.0	0.5	0.5	0.0	1.0
E88080	0.5	0.0	1.0	1.0	0.5	0.5	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.5	0.0	0.0	0.0	0.0	1.0
ELGIN	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
HACK	0.0	1.0	0.0	1.0	0.5	0.0	0.0	1.0	1.0	0.0	1.0	0.5	1.0	0.0	0.0	0.0	1.0	0.0	0.0
HARDIN	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.5	1.0	0.0	1.0
HW8008	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
J-231	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
KENWOOD	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.1	0.0	0.0	1.0
LN82-296	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
M82-946	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.5	0.0	0.0	1.0	0.0	1.0
M84-916	0.0	1.0	1.0	1.0	0.5	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0
s23-12	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.5	0.0	0.0	1.0
s2596	1.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
s1884	0.0	1.0	1.0	1.0	0.5	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
s19-90	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0
PELLA	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0
9271	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0
9292	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	0.5	0.0	0.0	1.0
SIBLEY	0.0	1.0	1.0	0.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0
STURDY	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.5	1.0	0.0	1.0
ZANE	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0

The frequencies are for the alleles reported in Table 7 and were determined as described in the materials and methods. Most lines were homogenous and therefore had allele frequencies of either 1.0 or 0.0 while other lines were heterogeneous and received a frequency of 0.5.

Appendix B. Coefficient of parentage between the 62 elite lines.

	CORSOY	CENTURY	AP 1989	HW8008	s2596	A1937	BURLISON	s23-12	SIBLEY
CORSOY	1.000	.211	.570	.211	.141	.250	.176	.141	.426
CENTURY	.211	1.000	.387	.719	.201	.154	.594	.188	.219
AP 1989	.570	.387	1.000	.356	.162	.209	.260	.141	.301
HW8008	.211	.719	.356	1.000	.272	.137	.439	.207	.203
s2596	.141	.201	.162	.272	1.000	.229	.180	.182	.144
A1937	.250	.154	.209	.137	.229	1.000	.129	.057	.348
BURLISON	.176	.594	.260	.439	.180	.129	1.000	.142	.166
s23-12	.141	.188	.141	.207	.182	.057	.142	1.000	.128
SIBLEY	.426	.219	.301	.203	.144	.348	.166	.128	1.000
A83-271010	.338	.311	.325	.298	.227	.299	.218	.116	.221
BSR 201	.285	.247	.309	.216	.247	.384	.184	.088	.207
7299	.160	.194	.182	.172	.086	.195	.125	.309	.204
J-231	.660	.315	.475	.297	.154	.316	.221	.126	.405
PELLA	.070	.406	.263	.379	.244	.298	.256	.099	.129
9271	.297	.150	.218	.155	.211	.289	.127	.087	.183
9292	.297	.150	.218	.155	.211	.289	.127	.087	.183
STURDY	.336	.603	.349	.457	.173	.176	.377	.155	.293
A 2934	.141	.221	.163	.196	.166	.169	.177	.152	.149
A80-147002	.285	.295	.300	.275	.233	.361	.206	.093	.197
A82-161035	.320	.245	.318	.242	.258	.337	.185	.122	.215
A84-185032	.141	.301	.264	.286	.231	.210	.208	.140	.162
A85-291010	.160	.160	.169	.164	.291	.732	.136	.080	.238
1250	.211	.290	.225	.252	.197	.227	.221	.169	.217
CONRAD	.088	.252	.179	.248	.193	.200	.177	.092	.110
C1664	.141	.561	.243	.440	.235	.198	.349	.143	.156
DSR-262	.143	.157	.177	.144	.202	.304	.129	.063	.124
E84150	.106	.538	.218	.398	.181	.188	.336	.114	.130
E84159	.250	.121	.188	.125	.192	.280	.109	.070	.148
E84165	.250	.121	.188	.125	.192	.280	.109	.070	.148
HACK	.035	.287	.187	.260	.288	.429	.198	.068	.121
LN82-296	.141	.703	.325	.549	.223	.226	.425	.143	.174
M82-946	.292	.262	.267	.271	.230	.301	.190	.144	.378
s1884	.469	.173	.316	.161	.210	.397	.161	.082	.249
s19-90	.035	.219	.139	.201	.124	.153	.136	.304	.077
ZANE	.285	.287	.294	.276	.233	.319	.199	.099	.192
A85-192034	.178	.369	.229	.283	.248	.679	.248	.084	.257
A85-292023	.088	.200	.149	.192	.309	.425	.163	.105	.140
A86-102004	.285	.158	.221	.171	.254	.626	.132	.099	.272
A86-103017	.266	.226	.229	.228	.239	.239	.178	.155	.207
A86-103027	.143	.221	.198	.198	.258	.715	.164	.062	.234
A86-202026	.266	.226	.229	.228	.239	.239	.178	.155	.207
A87-198005	.285	.295	.300	.275	.233	.361	.206	.093	.197
A87-296011	.160	.234	.211	.223	.293	.425	.174	.082	.161
A87-297015	.143	.163	.158	.149	.119	.184	.111	.300	.111
BSR 101	.141	.402	.301	.519	.300	.242	.263	.158	.166
7260	.293	.404	.353	.358	.185	.303	.269	.131	.302
8252	.141	.348	.244	.315	.221	.263	.239	.134	.173
E85100	.143	.179	.170	.171	.193	.288	.130	.061	.119
E85166	.283	.398	.314	.341	.226	.251	.290	.164	.229
E85168	.283	.398	.314	.341	.226	.251	.290	.164	.229
E86348	.213	.324	.273	.294	.227	.308	.224	.116	.179
HARDIN	.938	.211	.542	.208	.163	.259	.177	.135	.405
M84-916	.356	.192	.265	.178	.229	.307	.157	.096	.271
CHAPMAN	.455	.176	.309	.182	.245	.375	.148	.103	.246
E88080	.248	.629	.341	.475	.203	.317	.388	.136	.293
ELGIN	.169	.216	.186	.202	.234	.270	.169	.108	.169
A85-293032	.137	.400	.217	.316	.250	.313	.268	.110	.167
E84098	.469	.143	.296	.143	.162	.240	.127	.088	.223
E86339	.155	.285	.217	.258	.227	.262	.206	.123	.164
KENWOOD	.210	.185	.198	.169	.231	.635	.149	.082	.258
A2234	.013	.161	.098	.164	.303	.411	.127	.067	.090
A85-293033	.356	.237	.281	.211	.210	.325	.188	.128	.241
E87223	.585	.214	.378	.207	.187	.260	.172	.124	.297

Appendix B. continued

	A83-271010	BSR 201	7299	J-231	PELLA	9271	9292	STURDY	A 2934
CORSOY	.338	.285	.160	.660	.070	.297	.297	.336	.141
CENTURY	.311	.247	.194	.315	.406	.150	.150	.603	.221
AP 1989	.325	.309	.182	.475	.263	.218	.218	.349	.163
HW8008	.298	.216	.172	.297	.379	.155	.155	.457	.196
s2596	.227	.247	.086	.154	.244	.211	.211	.173	.166
A1937	.299	.384	.195	.316	.298	.289	.289	.176	.169
BURLISON	.218	.184	.125	.221	.256	.127	.127	.377	.177
s23-12	.116	.088	.309	.126	.099	.087	.087	.155	.152
SIBLEY	.221	.207	.204	.405	.129	.183	.183	.293	.149
A83-271010	1.000	.317	.145	.324	.321	.244	.244	.269	.175
BSR 201	.317	1.000	.140	.275	.327	.281	.281	.229	.188
7299	.145	.140	1.000	.210	.159	.093	.093	.158	.100
J-231	.324	.275	.210	1.000	.215	.239	.239	.330	.146
PELLA	.321	.327	.159	.215	1.000	.247	.247	.263	.197
9271	.244	.281	.093	.239	.247	1.000	.500	.170	.199
9292	.244	.281	.093	.239	.247	.500	1.000	.170	.199
STURDY	.269	.229	.158	.330	.263	.170	.170	1.000	.182
A 2934	.175	.188	.100	.146	.197	.199	.199	.182	1.000
A80-147002	.458	.362	.146	.298	.632	.286	.286	.248	.188
A82-161035	.304	.499	.131	.285	.287	.264	.264	.232	.183
A84-185032	.242	.290	.148	.186	.277	.173	.173	.227	.174
A85-291010	.307	.396	.145	.221	.336	.299	.299	.161	.186
1250	.236	.248	.138	.207	.224	.192	.192	.248	.286
CONRAD	.262	.218	.105	.155	.302	.209	.209	.179	.342
C1664	.249	.236	.129	.206	.313	.167	.167	.355	.176
DSR-262	.229	.599	.096	.158	.277	.268	.268	.141	.351
E84150	.226	.223	.125	.179	.318	.202	.202	.330	.200
E84159	.222	.267	.089	.206	.241	.289	.289	.138	.176
E84165	.222	.267	.089	.206	.241	.289	.289	.138	.176
HACK	.331	.420	.136	.161	.481	.331	.331	.201	.223
LN82-296	.316	.287	.177	.265	.457	.235	.235	.433	.227
M82-946	.328	.239	.163	.308	.232	.199	.199	.263	.164
s1884	.559	.373	.125	.356	.248	.306	.306	.218	.171
s19-90	.166	.171	.334	.112	.510	.126	.126	.144	.123
ZANE	.312	.331	.140	.292	.626	.323	.323	.241	.204
A85-192034	.313	.382	.174	.263	.364	.268	.268	.274	.193
A85-292023	.297	.393	.105	.143	.356	.284	.284	.169	.212
A86-102004	.271	.320	.148	.285	.257	.286	.286	.180	.173
A86-103017	.239	.252	.120	.230	.220	.238	.238	.216	.232
A86-103027	.315	.402	.166	.238	.389	.310	.310	.189	.196
A86-202026	.239	.252	.120	.230	.220	.238	.238	.216	.232
A87-198005	.396	.424	.146	.298	.527	.286	.286	.248	.188
A87-296011	.326	.471	.127	.209	.415	.302	.302	.198	.195
A87-297015	.235	.188	.327	.153	.326	.145	.145	.136	.119
BSR 101	.306	.304	.155	.239	.404	.200	.200	.280	.185
7260	.292	.278	.349	.360	.325	.186	.186	.313	.177
8252	.278	.288	.148	.211	.612	.219	.219	.256	.241
E85100	.296	.277	.094	.168	.425	.268	.268	.150	.165
E85166	.352	.292	.149	.288	.431	.223	.223	.312	.223
E85168	.352	.292	.149	.288	.431	.223	.223	.312	.223
E86348	.302	.325	.142	.256	.621	.248	.248	.252	.200
HARDIN	.333	.292	.157	.625	.089	.294	.294	.324	.143
M84-916	.265	.364	.118	.290	.218	.243	.243	.293	.160
CHAPMAN	.318	.467	.122	.347	.262	.350	.350	.218	.195
E88080	.298	.294	.208	.340	.351	.191	.191	.420	.190
ELGIN	.235	.274	.105	.181	.258	.218	.218	.190	.177
A85-293032	.280	.327	.129	.196	.344	.232	.232	.281	.197
E84098	.237	.246	.106	.334	.159	.273	.273	.191	.160
E86339	.253	.281	.121	.197	.435	.214	.214	.223	.195
KENWOOD	.267	.329	.150	.249	.278	.253	.253	.183	.173
A2234	.269	.373	.089	.092	.431	.478	.478	.124	.276
A85-293033	.291	.448	.135	.293	.244	.259	.259	.241	.233
E87223	.287	.280	.133	.421	.164	.257	.257	.263	.159

Appendix B. continued.

	A80-147002	A82-161035	A84-185032	A85-291010	1250	CONRAD	C1664	DSR-262	E84150
CORSOY	.285	.320	.141	.160	.211	.088	.141	.143	.106
CENTURY	.295	.245	.301	.160	.290	.252	.561	.157	.538
AP 1989	.300	.318	.264	.169	.225	.179	.243	.177	.218
HW8008	.275	.242	.286	.164	.252	.248	.440	.144	.398
s2596	.233	.258	.231	.291	.197	.193	.235	.202	.181
A1937	.361	.337	.210	.732	.227	.200	.198	.304	.188
BURLISON	.206	.185	.208	.136	.221	.177	.349	.129	.336
s23-12	.093	.122	.140	.080	.169	.092	.143	.063	.114
SIBLEY	.197	.215	.162	.238	.217	.110	.156	.124	.130
A83-271010	.458	.304	.242	.307	.236	.262	.249	.229	.226
BSR 201	.362	.499	.290	.396	.248	.218	.236	.599	.223
7299	.146	.131	.148	.145	.138	.105	.129	.096	.125
J-231	.298	.285	.186	.221	.207	.155	.206	.158	.179
PELLA	.632	.287	.277	.336	.224	.302	.313	.277	.318
9271	.286	.264	.173	.299	.192	.209	.167	.268	.202
9292	.286	.264	.173	.299	.192	.209	.167	.268	.202
STURDY	.248	.232	.227	.161	.248	.179	.355	.141	.330
A 2934	.188	.183	.174	.186	.286	.342	.176	.351	.200
A80-147002	1.000	.330	.244	.376	.237	.246	.257	.284	.251
A82-161035	.330	1.000	.416	.357	.244	.202	.233	.334	.207
A84-185032	.244	.416	1.000	.238	.228	.197	.237	.206	.214
A85-291010	.376	.357	.238	1.000	.245	.228	.226	.328	.208
1250	.237	.244	.228	.245	1.000	.166	.222	.179	.200
CONRAD	.246	.202	.197	.228	.166	1.000	.208	.383	.232
C1664	.257	.233	.237	.226	.222	.208	1.000	.177	.341
DSR-262	.284	.334	.206	.328	.179	.383	.177	1.000	.210
E84150	.251	.207	.214	.208	.200	.232	.341	.210	1.000
E84159	.273	.245	.156	.292	.159	.196	.151	.260	.356
E84165	.273	.245	.156	.292	.159	.196	.151	.260	.356
HACK	.420	.343	.280	.479	.248	.314	.287	.373	.307
LN82-296	.341	.266	.289	.248	.257	.295	.437	.235	.446
M82-946	.246	.248	.212	.272	.223	.194	.214	.169	.180
s1884	.593	.351	.198	.392	.235	.180	.190	.276	.174
s19-90	.323	.150	.155	.172	.148	.161	.165	.147	.173
ZANE	.486	.304	.231	.332	.217	.259	.246	.286	.264
A85-192034	.368	.329	.251	.562	.254	.235	.304	.300	.293
A85-292023	.366	.344	.256	.575	.259	.239	.247	.327	.229
A86-102004	.309	.301	.199	.503	.213	.192	.191	.262	.180
A86-103017	.247	.255	.208	.260	.391	.175	.204	.199	.186
A86-103027	.391	.340	.245	.606	.237	.257	.242	.339	.247
A86-202026	.247	.255	.208	.260	.391	.175	.204	.199	.186
A87-198005	.517	.393	.244	.376	.237	.246	.257	.315	.251
A87-296011	.405	.413	.259	.465	.247	.257	.256	.365	.245
A87-297015	.323	.171	.139	.192	.155	.132	.137	.151	.139
BSR 101	.320	.299	.304	.284	.234	.262	.311	.226	.277
7260	.293	.265	.289	.255	.235	.208	.270	.185	.251
8252	.434	.266	.252	.291	.404	.234	.268	.228	.259
E85100	.414	.246	.179	.312	.170	.209	.183	.254	.204
E85166	.422	.283	.260	.260	.288	.209	.282	.209	.263
E85168	.422	.283	.260	.260	.288	.209	.282	.209	.263
E86348	.468	.293	.243	.326	.233	.237	.262	.254	.255
HARDIN	.288	.318	.147	.177	.211	.097	.147	.154	.114
M84-916	.283	.342	.189	.300	.218	.159	.182	.250	.164
CHAPMAN	.354	.470	.220	.378	.231	.205	.197	.344	.197
E88080	.305	.273	.269	.277	.250	.223	.389	.201	.371
ELGIN	.262	.249	.205	.292	.228	.182	.210	.217	.188
A85-293032	.319	.285	.249	.342	.254	.226	.310	.256	.293
E84098	.247	.244	.137	.216	.160	.154	.134	.208	.326
E86339	.345	.252	.224	.284	.229	.205	.238	.221	.224
KENWOOD	.311	.293	.207	.512	.228	.191	.204	.261	.188
A2234	.387	.316	.231	.474	.211	.339	.231	.428	.320
A85-293033	.309	.434	.219	.331	.417	.177	.212	.298	.192
E87223	.274	.284	.173	.226	.220	.135	.175	.180	.147

Appendix B. continued.

	E84159	E84165	HACK	LN82-296	M82-946	s1884	s19-90	ZANE	A85-192034
CORSOY	.250	.250	.035	.141	.292	.469	.035	.285	.178
CENTURY	.121	.121	.287	.703	.262	.173	.219	.287	.369
AP 1989	.188	.188	.187	.325	.267	.316	.139	.294	.229
HW8008	.125	.125	.260	.549	.271	.161	.201	.276	.283
s2596	.192	.192	.288	.223	.230	.210	.124	.233	.248
A1937	.280	.280	.429	.226	.301	.397	.153	.319	.679
BURLISON	.109	.109	.198	.425	.190	.161	.136	.199	.248
s23-12	.070	.070	.068	.143	.144	.082	.304	.099	.084
SIBLEY	.148	.148	.121	.174	.378	.249	.077	.192	.257
A83-271010	.222	.222	.331	.316	.328	.559	.166	.312	.313
BSR 201	.267	.267	.420	.287	.239	.373	.171	.331	.382
7299	.089	.089	.136	.177	.163	.125	.334	.140	.174
J-231	.206	.206	.161	.265	.308	.356	.112	.292	.263
PELLA	.241	.241	.481	.457	.232	.248	.510	.626	.364
9271	.289	.289	.331	.235	.199	.306	.126	.323	.268
9292	.289	.289	.331	.235	.199	.306	.126	.323	.268
STURDY	.138	.138	.201	.433	.263	.218	.144	.241	.274
A 2934	.176	.176	.223	.227	.164	.171	.123	.204	.193
A80-147002	.273	.273	.420	.341	.246	.593	.323	.486	.368
A82-161035	.245	.245	.343	.266	.248	.351	.150	.304	.329
A84-185032	.156	.156	.280	.289	.212	.198	.155	.231	.251
A85-291010	.292	.292	.479	.248	.272	.392	.172	.332	.562
1250	.159	.159	.248	.257	.223	.235	.148	.217	.254
CONRAD	.196	.196	.314	.295	.194	.180	.161	.259	.235
C1664	.151	.151	.287	.437	.214	.190	.165	.246	.304
DSR-262	.260	.260	.373	.235	.169	.276	.147	.286	.300
E84150	.356	.356	.307	.446	.180	.174	.173	.264	.293
E84159	1.000	.590	.312	.199	.170	.288	.128	.290	.257
E84165	.590	1.000	.312	.199	.170	.288	.128	.290	.257
HACK	.312	.312	1.000	.402	.229	.339	.249	.407	.457
LN82-296	.199	.199	.402	1.000	.247	.211	.241	.352	.367
M82-946	.170	.170	.229	.247	1.000	.245	.125	.237	.273
s1884	.288	.288	.339	.211	.245	1.000	.127	.324	.349
s19-90	.128	.128	.249	.241	.125	.127	1.000	.320	.190
ZANE	.290	.290	.407	.352	.237	.324	.320	1.000	.331
A85-192034	.257	.257	.457	.367	.273	.349	.190	.331	1.000
A85-292023	.276	.276	.497	.278	.227	.354	.184	.325	.427
A86-102004	.261	.261	.347	.217	.262	.339	.135	.361	.457
A86-103017	.201	.201	.257	.232	.223	.258	.133	.310	.245
A86-103027	.296	.296	.715	.314	.265	.368	.201	.363	.568
A86-202026	.201	.201	.257	.232	.223	.258	.133	.310	.245
A87-198005	.273	.273	.420	.341	.246	.476	.270	.433	.368
A87-296011	.293	.293	.509	.315	.239	.371	.214	.372	.427
A87-297015	.144	.144	.219	.183	.132	.299	.416	.249	.192
BSR 101	.182	.182	.371	.403	.257	.222	.211	.308	.306
7260	.165	.165	.279	.365	.278	.244	.173	.279	.311
8252	.200	.200	.364	.357	.228	.241	.329	.421	.309
E85100	.244	.244	.369	.258	.171	.380	.217	.360	.288
E85166	.195	.195	.297	.353	.248	.388	.227	.346	.296
E85168	.195	.195	.297	.353	.248	.388	.227	.346	.296
E86348	.235	.235	.391	.350	.230	.295	.319	.452	.337
HARDIN	.250	.250	.065	.150	.285	.457	.045	.286	.192
M84-916	.222	.222	.278	.205	.243	.327	.115	.264	.286
CHAPMAN	.313	.313	.351	.235	.252	.419	.134	.461	.333
E88080	.170	.170	.315	.490	.283	.242	.187	.289	.385
ELGIN	.199	.199	.312	.237	.202	.250	.139	.248	.280
A85-293032	.216	.216	.399	.372	.223	.276	.183	.295	.479
E84098	.420	.420	.195	.169	.192	.317	.086	.264	.204
E86339	.198	.198	.337	.298	.208	.241	.227	.334	.293
KENWOOD	.240	.240	.371	.232	.251	.324	.146	.284	.479
A2234	.411	.411	.610	.365	.195	.322	.221	.455	.408
A85-293033	.232	.232	.304	.240	.242	.352	.142	.281	.313
E87223	.225	.225	.174	.189	.247	.359	.087	.267	.229

Appendix B. continued.

	A85-292023	A86-102004	A86-103017	A86-103027	A86-202026	A87-198005	A87-296011	A87-297015	BSR 101
CORSOY	.088	.285	.266	.143	.266	.285	.160	.143	.141
CENTURY	.200	.158	.226	.221	.226	.295	.234	.163	.402
AP 1989	.149	.221	.229	.198	.229	.300	.211	.158	.301
HW8008	.192	.171	.228	.198	.228	.275	.223	.149	.519
s2596	.309	.254	.239	.258	.239	.233	.293	.119	.300
A1937	.425	.626	.239	.715	.239	.361	.425	.184	.242
BURLISON	.163	.132	.178	.164	.178	.206	.174	.111	.263
s23-12	.105	.099	.155	.062	.155	.093	.082	.300	.158
SIBLEY	.140	.272	.207	.234	.207	.197	.161	.111	.166
A83-271010	.297	.271	.239	.315	.239	.396	.326	.235	.306
BSR 201	.393	.320	.252	.402	.252	.424	.471	.188	.304
7299	.105	.148	.120	.166	.120	.146	.127	.327	.155
J-231	.143	.285	.230	.238	.230	.298	.209	.153	.239
PELLA	.356	.257	.220	.389	.220	.527	.415	.326	.404
9271	.284	.286	.238	.310	.238	.286	.302	.145	.200
9292	.284	.286	.238	.310	.238	.286	.302	.145	.200
STURDY	.169	.180	.216	.189	.216	.248	.198	.136	.280
A 2934	.212	.173	.232	.196	.232	.188	.195	.119	.185
A80-147002	.366	.309	.247	.391	.247	.517	.405	.323	.320
A82-161035	.344	.301	.255	.340	.255	.393	.413	.171	.299
A84-185032	.256	.199	.208	.245	.208	.244	.259	.139	.304
A85-291010	.575	.503	.260	.606	.260	.376	.465	.192	.284
1250	.259	.213	.391	.237	.391	.237	.247	.155	.234
CONRAD	.239	.192	.175	.257	.175	.246	.257	.132	.262
C1664	.247	.191	.204	.242	.204	.257	.256	.137	.311
DSR-262	.327	.262	.199	.339	.199	.315	.365	.151	.226
E84150	.229	.180	.186	.247	.186	.251	.245	.139	.277
E84159	.276	.261	.201	.296	.201	.273	.293	.144	.182
E84165	.276	.261	.201	.296	.201	.273	.293	.144	.182
HACK	.497	.347	.257	.715	.257	.420	.509	.219	.371
LN82-296	.278	.217	.232	.314	.232	.341	.315	.183	.403
M82-946	.227	.262	.223	.265	.223	.246	.239	.132	.257
s1884	.354	.339	.258	.368	.258	.476	.371	.299	.222
s19-90	.184	.135	.133	.201	.133	.270	.214	.416	.211
ZANE	.325	.361	.310	.363	.310	.433	.372	.249	.308
A85-192034	.427	.457	.245	.568	.245	.368	.427	.192	.306
A85-292023	1.000	.343	.260	.461	.260	.366	.467	.189	.304
A86-102004	.343	1.000	.419	.487	.419	.309	.348	.161	.236
A86-103017	.260	.419	1.000	.248	.599	.247	.259	.146	.232
A86-103027	.461	.487	.248	1.000	.248	.391	.467	.202	.306
A86-202026	.260	.419	.599	.248	1.000	.247	.259	.146	.232
A87-198005	.366	.309	.247	.391	.247	1.000	.555	.357	.320
A87-296011	.467	.348	.259	.467	.259	.555	1.000	.209	.327
A87-297015	.189	.161	.146	.202	.146	.357	.209	1.000	.170
BSR 101	.304	.236	.232	.306	.232	.320	.327	.170	1.000
7260	.222	.246	.212	.291	.212	.293	.258	.157	.317
8252	.307	.235	.306	.313	.306	.382	.331	.241	.319
E85100	.308	.253	.195	.328	.195	.449	.327	.304	.231
E85166	.278	.234	.253	.274	.253	.457	.289	.314	.300
E85168	.278	.234	.253	.274	.253	.457	.289	.314	.300
E86348	.378	.267	.230	.349	.230	.415	.366	.242	.317
HARDIN	.111	.286	.262	.162	.262	.288	.179	.144	.150
M84-916	.283	.271	.226	.292	.226	.314	.328	.148	.210
CHAPMAN	.345	.402	.330	.363	.330	.409	.419	.180	.237
E88080	.252	.254	.220	.316	.220	.305	.283	.163	.398
ELGIN	.304	.243	.222	.291	.222	.262	.302	.141	.236
A85-293032	.367	.266	.236	.356	.236	.319	.365	.170	.303
E84098	.184	.244	.204	.218	.204	.247	.218	.130	.151
E86339	.347	.234	.217	.300	.217	.319	.314	.183	.275
KENWOOD	.364	.434	.231	.503	.231	.311	.364	.163	.239
A2234	.481	.380	.281	.510	.281	.387	.482	.199	.292
A85-293033	.318	.287	.437	.315	.437	.372	.383	.174	.235
E87223	.196	.264	.244	.217	.244	.274	.231	.142	.188

Appendix B. continued.

	7260	8252	E85100	E85166	E85168	E86348	HARDIN	M84-916	CHAPMAN
CORSOY	.293	.141	.143	.283	.283	.213	.938	.356	.455
CENTURY	.404	.348	.179	.398	.398	.324	.211	.192	.176
AP 1989	.353	.244	.170	.314	.314	.273	.542	.265	.309
HW8008	.358	.315	.171	.341	.341	.294	.208	.178	.182
s2596	.185	.221	.193	.226	.226	.227	.163	.229	.245
A1937	.303	.263	.288	.251	.251	.308	.259	.307	.375
BURLISON	.269	.239	.130	.290	.290	.224	.177	.157	.148
s23-12	.131	.134	.061	.164	.164	.116	.135	.096	.103
SIBLEY	.302	.173	.119	.229	.229	.179	.405	.271	.246
A83-271010	.292	.278	.296	.352	.352	.302	.333	.265	.318
BSR 201	.278	.288	.277	.292	.292	.325	.292	.364	.467
7299	.349	.148	.094	.149	.149	.142	.157	.118	.122
J-231	.360	.211	.168	.288	.288	.256	.625	.290	.347
PELLA	.325	.612	.425	.431	.431	.621	.089	.218	.262
9271	.186	.219	.268	.223	.223	.248	.294	.243	.350
9292	.186	.219	.268	.223	.223	.248	.294	.243	.350
STURDY	.313	.256	.150	.312	.312	.252	.324	.293	.218
A 2934	.177	.241	.165	.223	.223	.200	.143	.160	.195
A80-147002	.293	.434	.414	.422	.422	.468	.288	.283	.354
A82-161035	.265	.266	.246	.283	.283	.293	.318	.342	.470
A84-185032	.289	.252	.179	.260	.260	.243	.147	.189	.220
A85-291010	.255	.291	.312	.260	.260	.326	.177	.300	.378
1250	.235	.404	.170	.288	.288	.233	.211	.218	.231
CONRAD	.208	.234	.209	.209	.209	.237	.097	.159	.205
C1664	.270	.268	.183	.282	.282	.262	.147	.182	.197
DSR-262	.185	.228	.254	.209	.209	.254	.154	.250	.344
E84150	.251	.259	.204	.263	.263	.255	.114	.164	.197
E84159	.165	.200	.244	.195	.195	.235	.250	.222	.313
E84165	.165	.200	.244	.195	.195	.235	.250	.222	.313
HACK	.279	.364	.369	.297	.297	.391	.065	.278	.351
LN82-296	.365	.357	.258	.353	.353	.350	.150	.205	.235
M82-946	.278	.228	.171	.248	.248	.230	.285	.243	.252
s1884	.244	.241	.380	.388	.388	.295	.457	.327	.419
s19-90	.173	.329	.217	.227	.227	.319	.045	.115	.134
ZANE	.279	.421	.360	.346	.346	.452	.286	.264	.461
A85-192034	.311	.309	.288	.296	.296	.337	.192	.286	.333
A85-292023	.222	.307	.308	.278	.278	.378	.111	.283	.345
A86-102004	.246	.235	.253	.234	.234	.267	.286	.271	.402
A86-103017	.212	.306	.195	.253	.253	.230	.262	.226	.330
A86-103027	.291	.313	.328	.274	.274	.349	.162	.292	.363
A86-202026	.212	.306	.195	.253	.253	.230	.262	.226	.330
A87-198005	.293	.382	.449	.457	.457	.415	.288	.314	.409
A87-296011	.258	.331	.327	.289	.289	.366	.179	.328	.419
A87-297015	.157	.241	.304	.314	.314	.242	.144	.148	.180
BSR 101	.317	.319	.231	.300	.300	.317	.150	.210	.237
7260	1.000	.280	.190	.336	.336	.288	.289	.225	.239
8252	.280	1.000	.298	.360	.360	.427	.150	.218	.247
E85100	.190	.298	1.000	.363	.363	.322	.152	.207	.283
E85166	.336	.360	.363	1.000	.606	.360	.282	.246	.271
E85168	.336	.360	.363	.606	1.000	.360	.282	.246	.271
E86348	.288	.427	.322	.360	.360	1.000	.220	.249	.296
HARDIN	.289	.150	.152	.282	.282	.220	1.000	.352	.444
M84-916	.225	.218	.207	.246	.246	.249	.352	1.000	.373
CHAPMAN	.239	.247	.283	.271	.271	.296	.444	.373	1.000
E88080	.378	.301	.204	.330	.330	.305	.249	.228	.238
ELGIN	.206	.243	.208	.238	.238	.251	.178	.223	.254
A85-293032	.262	.299	.248	.290	.290	.308	.152	.244	.273
E84098	.193	.160	.191	.205	.205	.203	.508	.244	.331
E86339	.244	.332	.250	.299	.299	.483	.164	.218	.246
KENWOOD	.255	.253	.248	.244	.244	.279	.218	.265	.315
A2234	.185	.321	.430	.246	.246	.350	.041	.257	.410
A85-293033	.247	.330	.225	.289	.289	.274	.349	.345	.448
E87223	.249	.192	.175	.261	.261	.232	.558	.289	.355

Appendix B. continued.

	E88080	ELGIN	A85-293032	E84098	E86339	KENWOOD	A2234	A85-293033	E87223
CORSOY	.248	.169	.137	.469	.155	.210	.013	.356	.585
CENTURY	.629	.216	.400	.143	.285	.185	.161	.237	.214
AP 1989	.341	.186	.217	.296	.217	.198	.098	.281	.378
HW8008	.475	.202	.316	.143	.258	.169	.164	.211	.207
s2596	.203	.234	.250	.162	.227	.231	.303	.210	.187
A1937	.317	.270	.313	.240	.262	.635	.411	.325	.260
BURLISON	.388	.169	.268	.127	.206	.149	.127	.188	.172
s23-12	.136	.108	.110	.088	.123	.082	.067	.128	.124
SIBLEY	.293	.169	.167	.223	.164	.258	.090	.241	.297
A83-271010	.298	.235	.280	.237	.253	.267	.269	.291	.287
BSR 201	.294	.274	.327	.246	.281	.329	.373	.448	.280
7299	.208	.105	.129	.106	.121	.150	.089	.135	.133
J-231	.340	.181	.196	.334	.197	.249	.092	.293	.421
PELLA	.351	.258	.344	.159	.435	.278	.431	.244	.164
9271	.191	.218	.232	.273	.214	.253	.478	.259	.257
9292	.191	.218	.232	.273	.214	.253	.478	.259	.257
STURDY	.420	.190	.281	.191	.223	.183	.124	.241	.263
A 2934	.190	.177	.197	.160	.195	.173	.276	.233	.159
A80-147002	.305	.262	.319	.247	.345	.311	.387	.309	.274
A82-161035	.273	.249	.285	.244	.252	.293	.316	.434	.284
A84-185032	.269	.205	.249	.137	.224	.207	.231	.219	.173
A85-291010	.277	.292	.342	.216	.284	.512	.474	.331	.226
1250	.250	.228	.254	.160	.229	.228	.211	.417	.220
CONRAD	.223	.182	.226	.154	.205	.191	.339	.177	.135
C1664	.389	.210	.310	.134	.238	.204	.231	.212	.175
DSR-262	.201	.217	.256	.208	.221	.261	.428	.298	.180
E84150	.371	.188	.293	.326	.224	.188	.320	.192	.147
E84159	.170	.199	.216	.420	.198	.240	.411	.232	.225
E84165	.170	.199	.216	.420	.198	.240	.411	.232	.225
HACK	.315	.312	.399	.195	.337	.371	.610	.304	.174
LN82-296	.490	.237	.372	.169	.298	.232	.365	.240	.189
M82-946	.283	.202	.223	.192	.208	.251	.195	.242	.247
s1884	.242	.250	.276	.317	.241	.324	.322	.352	.359
s19-90	.187	.139	.183	.086	.227	.146	.221	.142	.087
ZANE	.289	.248	.295	.264	.334	.284	.455	.281	.267
A85-192034	.385	.280	.479	.204	.293	.479	.408	.313	.229
A85-292023	.252	.304	.367	.184	.347	.364	.481	.318	.196
A86-102004	.254	.243	.266	.244	.234	.434	.380	.287	.264
A86-103017	.220	.222	.236	.204	.217	.231	.281	.437	.244
A86-103027	.316	.291	.356	.218	.300	.503	.510	.315	.217
A86-202026	.220	.222	.236	.204	.217	.231	.281	.437	.244
A87-198005	.305	.262	.319	.247	.319	.311	.387	.372	.274
A87-296011	.283	.302	.365	.218	.314	.364	.482	.383	.231
A87-297015	.163	.141	.170	.130	.183	.163	.199	.174	.142
BSR 101	.398	.236	.303	.151	.275	.239	.292	.235	.188
7260	.378	.206	.262	.193	.244	.255	.185	.247	.249
8252	.301	.243	.299	.160	.332	.253	.321	.330	.192
E85100	.204	.208	.248	.191	.250	.248	.430	.225	.175
E85166	.330	.238	.290	.205	.299	.244	.246	.289	.261
E85168	.330	.238	.290	.205	.299	.244	.246	.289	.261
E86348	.305	.251	.308	.203	.483	.279	.350	.274	.232
HARDIN	.249	.178	.152	.508	.164	.218	.041	.349	.558
M84-916	.228	.223	.244	.244	.218	.265	.257	.345	.289
CHAPMAN	.238	.254	.273	.331	.246	.315	.410	.448	.355
E88080	1.000	.223	.338	.181	.264	.270	.222	.254	.236
ELGIN	.223	1.000	.646	.169	.620	.635	.301	.247	.203
A85-293032	.338	.646	1.000	.169	.472	.479	.354	.274	.200
E84098	.181	.169	.169	1.000	.163	.204	.261	.248	.319
E86339	.264	.620	.472	.163	1.000	.441	.307	.242	.196
KENWOOD	.270	.635	.479	.204	.441	1.000	.356	.286	.231
A2234	.222	.301	.354	.261	.307	.356	1.000	.276	.157
A85-293033	.254	.247	.274	.248	.242	.286	.276	1.000	.301
E87223	.236	.203	.200	.319	.196	.231	.157	.301	1.000

Appendix C. Similarity Index (SI) between the 62 elite lines

	CORSOY	CENTURY	AP 1989	HW8008	s2596	A1937	BURLISON	s23-12	SIBLEY
CORSOY	1.000	.579	.842	.526	.632	.632	.553	.553	.658
CENTURY	.579	1.000	.526	.947	.842	.737	.763	.553	.500
AP 1989	.842	.526	1.000	.474	.474	.474	.605	.553	.605
HW8008	.526	.947	.474	1.000	.790	.684	.711	.605	.447
s2596	.632	.842	.474	.790	1.000	.895	.711	.605	.500
A1937	.632	.737	.474	.684	.895	1.000	.763	.500	.500
BURLISON	.553	.763	.605	.711	.711	.763	.974	.526	.526
s23-12	.553	.553	.553	.605	.605	.500	.526	.974	.737
SIBLEY	.658	.500	.605	.447	.500	.500	.526	.737	.974
A83-271010	.500	.658	.447	.605	.658	.658	.632	.790	.711
BSR 201	.632	.737	.579	.684	.737	.737	.605	.500	.368
7299	.658	.658	.711	.658	.605	.500	.526	.632	.474
J-231	.790	.684	.632	.632	.737	.737	.553	.553	.711
PELLA	.526	.842	.579	.790	.684	.790	.868	.395	.447
9271	.579	.579	.526	.526	.737	.737	.500	.605	.447
9292	.658	.658	.553	.605	.816	.711	.526	.658	.526
STURDY	.763	.763	.763	.711	.605	.605	.632	.553	.579
A2943	.737	.684	.632	.632	.632	.579	.500	.368	.447
A80-147002	.605	.658	.553	.605	.553	.658	.737	.579	.711
A82-161035	.526	.579	.526	.526	.579	.579	.553	.737	.737
A84-185032	.763	.500	.711	.447	.605	.605	.421	.632	.526
A85-291010	.737	.737	.579	.684	.895	.790	.605	.711	.605
1250	.737	.526	.684	.579	.579	.579	.447	.658	.395
CONRAD	.605	.816	.605	.763	.763	.763	.790	.605	.579
C1664	.632	.737	.579	.684	.684	.790	.868	.605	.605
DSR-262	.579	.790	.526	.737	.737	.842	.711	.395	.395
E84150	.658	.816	.553	.763	.711	.605	.579	.579	.526
E84159	.526	.579	.474	.526	.632	.632	.553	.605	.474
E84165	.711	.553	.658	.500	.553	.553	.579	.632	.605
HACK	.526	.579	.474	.632	.684	.684	.711	.658	.579
LN82-296	.632	.947	.579	.895	.895	.790	.816	.605	.500
M82-946	.763	.395	.605	.447	.553	.553	.368	.684	.684
s1884	.658	.500	.500	.553	.500	.605	.579	.632	.658
s19-90	.632	.737	.684	.684	.684	.684	.711	.500	.658
ZANE	.684	.790	.632	.737	.632	.737	.711	.553	.605
A85-192034	.658	.711	.500	.711	.868	.763	.632	.684	.579
A85-292023	.684	.579	.526	.632	.632	.526	.553	.553	.447
A86-102004	.553	.816	.500	.763	.868	.921	.790	.579	.526
A86-103017	.711	.816	.658	.763	.711	.605	.579	.632	.579
A86-103027	.526	.526	.579	.474	.684	.684	.711	.658	.605
A86-202026	.658	.763	.605	.816	.763	.658	.632	.790	.579
A87-198005	.658	.658	.658	.605	.605	.605	.632	.447	.579
A87-296011	.684	.684	.737	.632	.632	.632	.658	.605	.658
A87-297015	.526	.632	.684	.579	.684	.684	.658	.553	.500
BSR 101	.763	.658	.711	.605	.605	.711	.684	.684	.684
7260	.605	.605	.605	.658	.658	.658	.579	.632	.368
8252	.579	.684	.632	.737	.526	.632	.605	.605	.500
E85100	.605	.605	.500	.605	.553	.658	.579	.395	.474
E85166	.579	.684	.474	.684	.632	.737	.763	.395	.500
E85168	.684	.684	.526	.632	.632	.737	.605	.395	.500
E86348	.737	.737	.684	.684	.684	.684	.605	.447	.553
HARDIN	.974	.553	.868	.500	.605	.605	.526	.553	.632
M84-916	.763	.500	.605	.447	.605	.605	.526	.632	.763
CHAPMAN	.632	.842	.579	.790	.684	.790	.763	.500	.553
E88080	.684	.790	.526	.790	.790	.684	.658	.605	.474
ELGIN	.684	.790	.526	.737	.842	.842	.658	.553	.605
E84098	.842	.632	.790	.579	.684	.579	.500	.658	.500
E86339	.658	.763	.553	.711	.816	.763	.579	.500	.579
KENWOOD	.629	.840	.476	.787	.892	.787	.603	.605	.550
A2234	.526	.947	.474	1.000	.790	.684	.711	.605	.447
A85-293033	.684	.421	.632	.474	.579	.579	.395	.711	.500
E87223	.632	.632	.526	.684	.632	.579	.447	.632	.500

Appendix C. continued

	A83-271010	BSR 201	7299	J-231	PELLA	9271	9292	STURDY	A2943
CORSOY	.500	.632	.658	.790	.526	.579	.658	.763	.737
CENTURY	.658	.737	.658	.684	.842	.579	.658	.763	.684
AP 1989	.447	.579	.711	.632	.579	.526	.553	.763	.632
HW8008	.605	.684	.658	.632	.790	.526	.605	.711	.632
s2596	.658	.737	.605	.737	.684	.737	.816	.605	.632
A1937	.658	.737	.500	.737	.790	.737	.711	.605	.579
BURLISON	.632	.605	.526	.553	.868	.500	.526	.632	.500
s23-12	.790	.500	.632	.553	.395	.605	.658	.553	.368
SIBLEY	.711	.368	.474	.711	.447	.447	.526	.579	.447
A83-271010	.974	.632	.526	.658	.605	.711	.684	.632	.447
BSR 201	.632	.947	.711	.632	.684	.842	.763	.711	.632
7299	.526	.711	.974	.658	.605	.553	.579	.684	.605
J-231	.658	.632	.658	1.000	.632	.579	.658	.763	.737
PELLA	.605	.684	.605	.632	1.000	.526	.500	.711	.579
9271	.711	.842	.553	.579	.526	1.000	.921	.605	.526
9292	.684	.763	.579	.658	.500	.921	.974	.658	.579
STURDY	.632	.711	.684	.763	.711	.605	.658	.974	.737
A2943	.447	.632	.605	.737	.579	.526	.579	.737	.947
A80-147002	.763	.526	.421	.658	.711	.500	.474	.632	.605
A82-161035	.895	.605	.500	.684	.526	.684	.632	.579	.500
A84-185032	.632	.763	.684	.605	.447	.816	.737	.632	.605
A85-291010	.763	.737	.711	.842	.579	.737	.816	.711	.632
1250	.553	.790	.763	.526	.474	.737	.658	.658	.579
CONRAD	.632	.658	.526	.605	.763	.658	.711	.711	.579
C1664	.763	.737	.500	.632	.790	.632	.605	.711	.474
DSR-262	.553	.842	.658	.684	.842	.684	.605	.711	.684
E84150	.684	.763	.632	.658	.658	.763	.842	.790	.658
E84159	.737	.711	.553	.526	.526	.790	.711	.605	.632
E84165	.658	.684	.632	.658	.500	.711	.737	.790	.658
HACK	.632	.447	.605	.684	.632	.474	.553	.553	.526
LN82-296	.658	.737	.605	.632	.790	.632	.711	.711	.632
M82-946	.526	.421	.526	.711	.342	.553	.632	.526	.605
s1884	.711	.579	.526	.711	.553	.553	.526	.579	.553
s19-90	.500	.474	.605	.737	.790	.421	.500	.605	.579
ZANE	.763	.737	.658	.790	.842	.579	.553	.868	.632
A85-192034	.684	.711	.605	.763	.553	.711	.790	.579	.553
A85-292023	.500	.632	.500	.579	.421	.579	.658	.553	.684
A86-102004	.737	.711	.474	.658	.816	.711	.684	.632	.526
A86-103017	.632	.711	.684	.658	.658	.658	.737	.790	.605
A86-103027	.658	.632	.605	.632	.579	.737	.763	.553	.421
A86-202026	.632	.658	.684	.605	.605	.605	.684	.632	.447
A87-198005	.526	.658	.684	.763	.711	.605	.658	.658	.632
A87-296011	.605	.684	.553	.684	.632	.684	.711	.711	.579
A87-297015	.658	.737	.605	.632	.684	.737	.658	.553	.474
BSR 101	.737	.658	.632	.658	.711	.605	.579	.737	.500
7260	.632	.816	.790	.605	.605	.711	.632	.632	.553
8252	.553	.632	.711	.579	.737	.474	.395	.711	.579
E85100	.526	.658	.447	.605	.658	.658	.605	.605	.684
E85166	.553	.632	.526	.684	.790	.474	.447	.553	.632
E85168	.605	.737	.500	.684	.737	.737	.711	.711	.684
E86348	.500	.790	.711	.842	.684	.632	.658	.763	.737
HARDIN	.474	.658	.684	.763	.500	.605	.658	.763	.737
M84-916	.711	.526	.474	.816	.447	.658	.737	.684	.658
CHAPMAN	.711	.790	.605	.737	.895	.632	.605	.816	.579
E88080	.632	.790	.684	.684	.632	.684	.763	.711	.632
ELGIN	.658	.737	.658	.895	.737	.684	.763	.763	.632
E84098	.553	.790	.816	.632	.474	.737	.763	.763	.684
E86339	.579	.711	.632	.868	.658	.658	.711	.658	.711
KENWOOD	.708	.792	.713	.840	.682	.740	.816	.711	.632
A2234	.605	.684	.658	.632	.790	.526	.605	.711	.632
A85-293033	.605	.684	.658	.579	.368	.737	.658	.553	.526
E87223	.605	.632	.658	.684	.526	.684	.737	.737	.684

Appendix C. continued

	A80-147002	A82-161035	A84-185032	A85-291010	1250	CONRAD	C1664	DSR-262	E84150
CORSOY	.605	.526	.763	.737	.737	.605	.632	.579	.658
CENTURY	.658	.579	.500	.737	.526	.816	.737	.790	.816
AP 1989	.553	.526	.711	.579	.684	.605	.579	.526	.553
HW8008	.605	.526	.447	.684	.579	.763	.684	.737	.763
s2596	.553	.579	.605	.895	.579	.763	.684	.737	.711
A1937	.658	.579	.605	.790	.579	.763	.790	.842	.605
BURLISON	.737	.553	.421	.605	.447	.790	.868	.711	.579
s23-12	.579	.737	.632	.711	.658	.605	.605	.395	.579
SIBLEY	.711	.737	.526	.605	.395	.579	.605	.395	.526
A83-271010	.763	.895	.632	.763	.553	.632	.763	.553	.684
BSR 201	.526	.605	.763	.737	.790	.658	.737	.842	.763
7299	.421	.500	.684	.711	.763	.526	.500	.658	.632
J-231	.658	.684	.605	.842	.526	.605	.632	.684	.658
PELLA	.711	.526	.447	.579	.474	.763	.790	.842	.658
9271	.500	.684	.816	.737	.737	.658	.632	.684	.763
9292	.474	.632	.737	.816	.658	.711	.605	.605	.842
STURDY	.632	.579	.632	.711	.658	.711	.711	.711	.790
A2943	.605	.500	.605	.632	.579	.579	.474	.684	.658
A80-147002	.974	.790	.526	.553	.447	.737	.763	.658	.579
A82-161035	.790	.947	.658	.684	.526	.579	.684	.526	.605
A84-185032	.526	.658	.921	.711	.868	.526	.553	.605	.632
A85-291010	.553	.684	.711	1.000	.684	.658	.684	.632	.711
1250	.447	.526	.868	.684	1.000	.500	.579	.632	.605
CONRAD	.737	.579	.526	.658	.500	.974	.763	.763	.737
C1664	.763	.684	.553	.684	.579	.763	1.000	.737	.658
DSR-262	.658	.526	.605	.632	.632	.763	.737	1.000	.658
E84150	.579	.605	.632	.711	.605	.737	.658	.658	.974
E84159	.684	.711	.737	.632	.684	.605	.579	.684	.658
E84165	.658	.632	.684	.658	.658	.684	.658	.658	.737
HACK	.632	.553	.421	.684	.474	.605	.579	.579	.500
LN82-296	.658	.579	.553	.790	.579	.868	.790	.737	.763
M82-946	.526	.553	.632	.658	.605	.526	.447	.395	.526
s1884	.816	.737	.579	.605	.605	.579	.711	.605	.579
s19-90	.605	.526	.447	.579	.368	.658	.579	.632	.553
ZANE	.763	.684	.605	.737	.632	.711	.842	.790	.711
A85-192034	.526	.658	.605	.868	.605	.632	.711	.605	.684
A85-292023	.500	.526	.553	.632	.632	.500	.632	.474	.658
A86-102004	.684	.658	.579	.763	.553	.790	.816	.763	.632
A86-103017	.474	.553	.658	.711	.658	.632	.658	.605	.842
A86-103027	.553	.632	.553	.684	.474	.711	.684	.632	.605
A86-202026	.474	.553	.605	.763	.711	.684	.711	.553	.684
A87-198005	.632	.579	.526	.605	.447	.711	.605	.711	.737
A87-296011	.711	.684	.605	.632	.526	.868	.737	.684	.711
A87-297015	.553	.737	.658	.684	.579	.605	.684	.632	.553
BSR 101	.737	.658	.711	.711	.711	.737	.816	.658	.632
7260	.526	.605	.737	.763	.868	.579	.658	.711	.579
8252	.658	.526	.605	.526	.737	.658	.632	.790	.500
E85100	.737	.579	.579	.447	.553	.658	.605	.763	.684
E85166	.763	.579	.447	.526	.474	.605	.737	.790	.553
E85168	.711	.579	.632	.579	.579	.711	.684	.790	.816
E86348	.605	.579	.605	.684	.579	.711	.684	.842	.711
HARDIN	.579	.526	.790	.711	.763	.605	.605	.605	.632
M84-916	.711	.737	.632	.711	.500	.579	.605	.500	.684
CHAPMAN	.711	.632	.553	.684	.579	.763	.895	.842	.763
E88080	.526	.553	.605	.790	.684	.658	.737	.684	.816
ELGIN	.553	.579	.553	.842	.526	.711	.737	.790	.763
E84098	.447	.526	.868	.790	.895	.605	.579	.632	.711
E86339	.579	.632	.579	.763	.500	.658	.605	.763	.684
KENWOOD	.497	.632	.608	.892	.582	.658	.682	.740	.813
A2234	.605	.526	.447	.684	.579	.763	.684	.737	.763
A85-293033	.447	.632	.842	.684	.895	.395	.526	.526	.500
E87223	.553	.553	.632	.684	.684	.632	.474	.632	.763

Appendix C. continued

	E84159	E84165	HACK	LN82-296	M82-946	s1884	s19-90	ZANE	A85-192034
CORSOY	.526	.711	.526	.632	.763	.658	.632	.684	.658
CENTURY	.579	.553	.579	.947	.395	.500	.737	.790	.711
AP 1989	.474	.658	.474	.579	.605	.500	.684	.632	.500
HW8008	.526	.500	.632	.895	.447	.553	.684	.737	.711
s2596	.632	.553	.684	.895	.553	.500	.684	.632	.868
A1937	.632	.553	.684	.790	.553	.605	.684	.737	.763
BURLISON	.553	.579	.711	.816	.368	.579	.711	.711	.632
s23-12	.605	.632	.658	.605	.684	.632	.500	.553	.684
SIBLEY	.474	.605	.579	.500	.684	.658	.658	.605	.579
A83-271010	.737	.658	.632	.658	.526	.711	.500	.763	.684
BSR 201	.711	.684	.447	.737	.421	.579	.474	.737	.711
7299	.553	.632	.605	.605	.526	.526	.605	.658	.605
J-231	.526	.658	.684	.632	.711	.711	.737	.790	.763
PELLA	.526	.500	.632	.790	.342	.553	.790	.842	.553
9271	.790	.711	.474	.632	.553	.553	.421	.579	.711
9292	.711	.737	.553	.711	.632	.526	.500	.553	.790
STURDY	.605	.790	.553	.711	.526	.579	.605	.868	.579
A2943	.632	.658	.526	.632	.605	.553	.579	.632	.553
A80-147002	.684	.658	.632	.658	.526	.816	.605	.763	.526
A82-161035	.711	.632	.553	.579	.553	.737	.526	.684	.658
A84-185032	.737	.684	.421	.553	.632	.579	.447	.605	.605
A85-291010	.632	.658	.684	.790	.658	.605	.579	.737	.868
1250	.684	.658	.474	.579	.605	.605	.368	.632	.605
CONRAD	.605	.684	.605	.868	.526	.579	.658	.711	.632
C1664	.579	.658	.579	.790	.447	.711	.579	.842	.711
DSR-262	.684	.658	.579	.737	.395	.605	.632	.790	.605
E84150	.658	.737	.500	.763	.526	.579	.553	.711	.684
E84159	.947	.790	.605	.579	.447	.632	.368	.579	.553
E84165	.790	.974	.632	.553	.579	.711	.395	.658	.579
HACK	.605	.632	.947	.579	.605	.684	.632	.579	.658
LN82-296	.579	.553	.579	1.000	.447	.500	.684	.737	.763
M82-946	.447	.579	.605	.447	.974	.684	.553	.500	.632
s1884	.632	.711	.684	.500	.684	.974	.447	.711	.632
s19-90	.368	.395	.632	.684	.553	.447	1.000	.632	.605
ZANE	.579	.658	.579	.737	.500	.711	.632	1.000	.605
A85-192034	.553	.579	.658	.763	.632	.632	.605	.605	.921
A85-292023	.526	.605	.526	.632	.658	.658	.421	.474	.763
A86-102004	.605	.474	.605	.868	.474	.526	.711	.763	.737
A86-103017	.526	.579	.447	.763	.526	.421	.711	.711	.684
A86-103027	.632	.763	.737	.579	.553	.605	.579	.526	.711
A86-202026	.421	.474	.553	.816	.632	.526	.658	.658	.790
A87-198005	.500	.684	.605	.605	.579	.684	.711	.658	.632
A87-296011	.526	.711	.474	.737	.605	.658	.632	.684	.658
A87-297015	.526	.447	.474	.684	.447	.500	.684	.632	.711
BSR 101	.553	.632	.526	.711	.632	.684	.605	.868	.579
7260	.658	.632	.605	.658	.526	.684	.395	.711	.684
8252	.579	.553	.579	.632	.500	.605	.632	.790	.447
E85100	.711	.684	.553	.553	.526	.737	.553	.605	.526
E85166	.579	.553	.684	.632	.447	.763	.684	.684	.632
E85168	.684	.711	.526	.632	.553	.711	.579	.737	.579
E86348	.526	.711	.526	.684	.553	.658	.684	.737	.711
HARDIN	.553	.737	.500	.605	.737	.632	.605	.658	.632
M84-916	.684	.816	.684	.500	.737	.763	.553	.605	.684
CHAPMAN	.526	.605	.526	.790	.447	.658	.684	.947	.658
E88080	.605	.684	.605	.790	.526	.632	.526	.684	.816
ELGIN	.526	.658	.684	.737	.605	.605	.737	.790	.816
E84098	.684	.763	.474	.684	.605	.500	.474	.632	.658
E86339	.553	.579	.605	.711	.579	.579	.763	.658	.790
KENWOOD	.582	.608	.629	.787	.550	.550	.682	.734	.866
A2234	.526	.500	.632	.895	.447	.553	.684	.737	.711
A85-293033	.684	.605	.526	.474	.658	.605	.421	.526	.658
E87223	.711	.763	.684	.579	.658	.658	.474	.632	.605

Appendix C. Continued

	A85-292023	A86-102004	A86-103017	A86-103027	A86-202026	A87-198005	A87-296011	A87-297015	BSR 101
CORSOY	.684	.553	.711	.526	.658	.658	.684	.526	.763
CENTURY	.579	.816	.816	.526	.763	.658	.684	.632	.658
AP 1989	.526	.500	.658	.579	.605	.658	.737	.684	.711
HW8008	.632	.763	.763	.474	.816	.605	.632	.579	.605
s2596	.632	.868	.711	.684	.763	.605	.632	.684	.605
A1937	.526	.921	.605	.684	.658	.605	.632	.684	.711
BURLISON	.553	.790	.579	.711	.632	.632	.658	.658	.684
s23-12	.553	.579	.632	.658	.790	.447	.605	.553	.684
SIBLEY	.447	.526	.579	.605	.579	.579	.658	.500	.684
A83-271010	.500	.737	.632	.658	.632	.526	.605	.658	.737
BSR 201	.632	.711	.711	.632	.658	.658	.684	.737	.658
7299	.500	.474	.684	.605	.684	.684	.553	.605	.632
J-231	.579	.658	.658	.632	.605	.763	.684	.632	.658
PELLA	.421	.816	.658	.579	.605	.711	.632	.684	.711
9271	.579	.711	.658	.737	.605	.605	.684	.737	.605
9292	.658	.684	.737	.763	.684	.658	.711	.658	.579
STURDY	.553	.632	.790	.553	.632	.658	.711	.553	.737
A2943	.684	.526	.605	.421	.447	.632	.579	.474	.500
A80-147002	.500	.684	.474	.553	.474	.632	.711	.553	.737
A82-161035	.526	.658	.553	.632	.553	.579	.684	.737	.658
A84-185032	.553	.579	.658	.553	.605	.526	.605	.658	.711
A85-291010	.632	.763	.711	.684	.763	.605	.632	.684	.711
1250	.632	.553	.658	.474	.711	.447	.526	.579	.711
CONRAD	.500	.790	.632	.711	.684	.711	.868	.605	.737
C1664	.632	.816	.658	.684	.711	.605	.737	.684	.816
DSR-262	.474	.763	.605	.632	.553	.711	.684	.632	.658
E84150	.658	.632	.842	.605	.684	.737	.711	.553	.632
E84159	.526	.605	.526	.632	.421	.500	.526	.526	.553
E84165	.605	.474	.579	.763	.474	.684	.711	.447	.632
HACK	.526	.605	.447	.737	.553	.605	.474	.474	.526
LN82-296	.632	.868	.763	.579	.816	.605	.737	.684	.711
M82-946	.658	.474	.526	.553	.632	.579	.605	.447	.632
s1884	.658	.526	.421	.605	.526	.684	.658	.500	.684
s19-90	.421	.711	.711	.579	.658	.711	.632	.684	.605
ZANE	.474	.763	.711	.526	.658	.658	.684	.632	.868
A85-192034	.763	.737	.684	.711	.790	.632	.658	.711	.579
A85-292023	1.000	.500	.605	.526	.658	.553	.579	.526	.447
A86-102004	.500	.974	.684	.605	.737	.526	.658	.763	.737
A86-103017	.605	.684	.974	.500	.816	.579	.605	.605	.684
A86-103027	.526	.605	.500	1.000	.553	.763	.737	.684	.553
A86-202026	.658	.737	.816	.553	.974	.526	.658	.658	.737
A87-198005	.553	.526	.579	.763	.526	.974	.816	.658	.579
A87-296011	.579	.658	.605	.737	.658	.816	1.000	.737	.711
A87-297015	.526	.763	.605	.684	.658	.658	.737	1.000	.605
BSR 101	.447	.737	.684	.553	.737	.579	.711	.605	.974
7260	.605	.632	.526	.605	.684	.579	.605	.711	.684
8252	.368	.658	.605	.421	.658	.500	.579	.526	.763
E85100	.605	.579	.526	.553	.421	.711	.658	.500	.526
E85166	.632	.658	.500	.579	.500	.711	.579	.579	.553
E85168	.579	.658	.658	.579	.500	.763	.684	.526	.658
E86348	.632	.605	.658	.684	.605	.868	.842	.684	.605
HARDIN	.658	.526	.684	.553	.632	.658	.711	.553	.737
M84-916	.658	.526	.579	.711	.474	.684	.658	.500	.579
CHAPMAN	.526	.816	.763	.579	.711	.711	.737	.684	.816
E88080	.790	.658	.763	.632	.763	.658	.632	.579	.605
ELGIN	.579	.763	.763	.737	.711	.763	.684	.632	.658
E84098	.632	.553	.763	.579	.711	.553	.632	.579	.711
E86339	.605	.711	.684	.658	.632	.763	.711	.711	.526
KENWOOD	.629	.761	.813	.687	.761	.711	.634	.687	.603
A2234	.632	.763	.763	.474	.816	.605	.632	.579	.605
A85-293033	.632	.553	.605	.526	.658	.395	.474	.632	.605
E87223	.579	.526	.632	.579	.579	.632	.579	.421	.553

Appendix C. continued

	7260	8252	E85100	E85166	E85168	E86348	HARDIN	M84-916	CHAPMAN
CORSOY	.605	.579	.605	.579	.684	.737	.974	.763	.632
CENTURY	.605	.684	.605	.684	.684	.737	.553	.500	.842
AP 1989	.605	.632	.500	.474	.526	.684	.868	.605	.579
HW8008	.658	.737	.605	.684	.632	.684	.500	.447	.790
s2596	.658	.526	.553	.632	.632	.684	.605	.605	.684
A1937	.658	.632	.658	.737	.737	.684	.605	.605	.790
BURLISON	.579	.605	.579	.763	.605	.605	.526	.526	.763
s23-12	.632	.605	.395	.395	.395	.447	.553	.632	.500
SIBLEY	.368	.500	.474	.500	.500	.553	.632	.763	.553
A83-271010	.632	.553	.526	.553	.605	.500	.474	.711	.711
BSR 201	.816	.632	.658	.632	.737	.790	.658	.526	.790
7299	.790	.711	.447	.526	.500	.711	.684	.474	.605
J-231	.605	.579	.605	.684	.684	.842	.763	.816	.737
PELLA	.605	.737	.658	.790	.737	.684	.500	.447	.895
9271	.711	.474	.658	.474	.737	.632	.605	.658	.632
9292	.632	.395	.605	.447	.711	.658	.658	.737	.605
STURDY	.632	.711	.605	.553	.711	.763	.763	.684	.816
A2943	.553	.579	.684	.632	.684	.737	.737	.658	.579
A80-147002	.526	.658	.737	.763	.711	.605	.579	.711	.711
A82-161035	.605	.526	.579	.579	.579	.579	.526	.737	.632
A84-185032	.737	.605	.579	.447	.632	.605	.790	.632	.553
A85-291010	.763	.526	.447	.526	.579	.684	.711	.711	.684
1250	.868	.737	.553	.474	.579	.579	.763	.500	.579
CONRAD	.579	.658	.658	.605	.711	.711	.605	.579	.763
C1664	.658	.632	.605	.737	.684	.684	.605	.605	.895
DSR-262	.711	.790	.763	.790	.790	.842	.605	.500	.842
E84150	.579	.500	.684	.553	.816	.711	.632	.684	.763
E84159	.658	.579	.711	.579	.684	.526	.553	.684	.526
E84165	.632	.553	.684	.553	.711	.711	.737	.816	.605
HACK	.605	.579	.553	.684	.526	.526	.500	.684	.526
LN82-296	.658	.632	.553	.632	.632	.684	.605	.500	.790
M82-946	.526	.500	.526	.447	.553	.553	.737	.737	.447
s1884	.684	.605	.737	.763	.711	.658	.632	.763	.658
s19-90	.395	.632	.553	.684	.579	.684	.605	.553	.684
ZANE	.711	.790	.605	.684	.737	.737	.658	.605	.947
A85-192034	.684	.447	.526	.632	.579	.711	.632	.684	.658
A85-292023	.605	.368	.605	.632	.579	.632	.658	.658	.526
A86-102004	.632	.658	.579	.658	.658	.605	.526	.526	.816
A86-103017	.526	.605	.526	.500	.658	.658	.684	.579	.763
A86-103027	.605	.421	.553	.579	.579	.684	.553	.711	.579
A86-202026	.684	.658	.421	.500	.500	.605	.632	.474	.711
A87-198005	.579	.500	.711	.711	.763	.868	.658	.684	.711
A87-296011	.605	.579	.658	.579	.684	.842	.711	.658	.737
A87-297015	.711	.526	.500	.579	.526	.684	.553	.500	.684
BSR 101	.684	.763	.526	.553	.658	.605	.737	.579	.816
7260	.974	.711	.526	.579	.553	.658	.632	.474	.658
8252	.711	1.000	.605	.632	.579	.632	.605	.395	.737
E85100	.526	.605	.921	.790	.868	.711	.605	.684	.658
E85166	.579	.632	.790	.947	.737	.737	.553	.605	.737
E85168	.553	.579	.868	.737	.947	.737	.658	.711	.790
E86348	.658	.632	.711	.737	.737	1.000	.763	.658	.790
HARDIN	.632	.605	.605	.553	.658	.763	.974	.737	.605
M84-916	.474	.395	.684	.605	.711	.658	.737	.974	.553
CHAPMAN	.658	.737	.658	.737	.790	.790	.605	.553	1.000
E88080	.711	.526	.605	.658	.684	.737	.658	.632	.737
ELGIN	.605	.579	.605	.684	.737	.842	.658	.711	.842
E84098	.763	.632	.500	.421	.579	.684	.868	.605	.579
E86339	.579	.553	.658	.711	.684	.868	.658	.684	.711
KENWOOD	.661	.529	.553	.629	.682	.792	.605	.655	.787
A2234	.658	.737	.605	.684	.632	.684	.500	.447	.790
A85-293033	.763	.632	.526	.474	.500	.526	.711	.605	.474
E87223	.658	.632	.684	.526	.711	.632	.632	.711	.579

Appendix C. continued

	E88080	ELGIN	E84098	E86339	KENWOOD	A2234	A85-293033	E87223
CORSOY	.684	.684	.842	.658	.629	.526	.684	.632
CENTURY	.790	.790	.632	.763	.840	.947	.421	.632
AP 1989	.526	.526	.790	.553	.476	.474	.632	.526
HW8008	.790	.737	.579	.711	.787	1.000	.474	.684
s2596	.790	.842	.684	.816	.892	.790	.579	.632
A1937	.684	.842	.579	.763	.787	.684	.579	.579
BURLISON	.658	.658	.500	.579	.603	.711	.395	.447
s23-12	.605	.553	.658	.500	.605	.605	.711	.632
SIBLEY	.474	.605	.500	.579	.550	.447	.500	.500
A83-271010	.632	.658	.553	.579	.708	.605	.605	.605
BSR 201	.790	.737	.790	.711	.792	.684	.684	.632
7299	.684	.658	.816	.632	.713	.658	.658	.658
J-231	.684	.895	.632	.868	.840	.632	.579	.684
PELLA	.632	.737	.474	.658	.682	.790	.368	.526
9271	.684	.684	.737	.658	.740	.526	.737	.684
9292	.763	.763	.763	.711	.816	.605	.658	.737
STURDY	.711	.763	.763	.658	.711	.711	.553	.737
A2943	.632	.632	.684	.711	.632	.632	.526	.684
A80-147002	.526	.553	.447	.579	.497	.605	.447	.553
A82-161035	.553	.579	.526	.632	.632	.526	.632	.553
A84-185032	.605	.553	.868	.579	.608	.447	.842	.632
A85-291010	.790	.842	.790	.763	.892	.684	.684	.684
1250	.684	.526	.895	.500	.582	.579	.895	.684
CONRAD	.658	.711	.605	.658	.658	.763	.395	.632
C1664	.737	.737	.579	.605	.682	.684	.526	.474
DSR-262	.684	.790	.632	.763	.740	.737	.526	.632
E84150	.816	.763	.711	.684	.813	.763	.500	.763
E84159	.605	.526	.684	.553	.582	.526	.684	.711
E84165	.684	.658	.763	.579	.608	.500	.605	.763
HACK	.605	.684	.474	.605	.629	.632	.526	.684
LN82-296	.790	.737	.684	.711	.787	.895	.474	.579
M82-946	.526	.605	.605	.579	.550	.447	.658	.658
s1884	.632	.605	.500	.579	.550	.553	.605	.658
s19-90	.526	.737	.474	.763	.682	.684	.421	.474
ZANE	.684	.790	.632	.658	.734	.737	.526	.632
A85-192034	.816	.816	.658	.790	.866	.711	.658	.605
A85-292023	.790	.579	.632	.605	.629	.632	.632	.579
A86-102004	.658	.763	.553	.711	.761	.763	.553	.526
A86-103017	.763	.763	.763	.684	.813	.763	.605	.632
A86-103027	.632	.737	.579	.658	.687	.474	.526	.579
A86-202026	.763	.711	.711	.632	.761	.816	.658	.579
A87-198005	.658	.763	.553	.763	.711	.605	.395	.632
A87-296011	.632	.684	.632	.711	.634	.632	.474	.579
A87-297015	.579	.632	.579	.711	.687	.579	.632	.421
BSR 101	.605	.658	.711	.526	.603	.605	.605	.553
7260	.711	.605	.763	.579	.661	.658	.763	.658
8252	.526	.579	.632	.553	.529	.737	.632	.632
E85100	.605	.605	.500	.658	.553	.605	.526	.684
E85166	.658	.684	.421	.711	.629	.684	.474	.526
E85168	.684	.737	.579	.684	.682	.632	.500	.711
E86348	.737	.842	.684	.868	.792	.684	.526	.632
HARDIN	.658	.658	.868	.658	.605	.500	.711	.632
M84-916	.632	.711	.605	.684	.655	.447	.605	.711
CHAPMAN	.737	.842	.579	.711	.787	.790	.474	.579
E88080	.895	.790	.737	.711	.840	.790	.605	.684
ELGIN	.790	1.000	.632	.868	.945	.737	.526	.684
E84098	.737	.632	1.000	.605	.687	.579	.790	.684
E86339	.711	.868	.605	.921	.868	.711	.553	.632
KENWOOD	.840	.945	.687	.868	.995	.787	.582	.684
A2234	.790	.737	.579	.711	.787	1.000	.474	.684
A85-293033	.605	.526	.790	.553	.582	.474	.947	.632
E87223	.684	.684	.684	.632	.684	.684	.632	.895

Application of Genetic Distances to a Soybean Breeding Program

Abstract

The genetic distance was estimated between elite soybean lines using the coefficient of parentage (CP), a similarity index (SI) based on marker loci, and a distance value (PCD) derived from a principal component analysis of CP and SI data. Genetically similar and distant parents were crossed based on the SI values and approximately 55 F2:3 or F4:5 families were derived from each cross and then field tested for seed yield, plant height, and date of maturity. The generalized variance, and the genetic and progeny variance for individual traits were estimated for each population. In general the variance parameters of a population increased as the genetic distance between the parents of the population increased, regardless of the method of estimating the distance. There was a stronger association of the measures with the generalized variances than with the genetic variances of the individual traits of the populations. No measure was significantly associated with the yield genetic variance of a population while all seemed predictive of the genetic variances for the maturity and height traits. The utility of the measures appeared to be limited to identifying which populations would have a higher likelihood of having an above average genetic variance for the individual traits. The PCD appeared to be the best predictor of the variance parameters of the populations.

Introduction

The objective of breeding soybeans (*Glycine max* L. Merr.) for a metric trait such as yield is to develop progeny from a segregating population that are better than the parents of the population. The mean of the progeny derived from selected individuals (u') can be expressed in terms derived from the gain from mass selection equation (Falconer, 1981)

$$u' = (u_s - u)\, h^2 + u$$

where u is the mean of the parental population, $(u_s - u)$ is the selection differential, u_s is the mean of the selected individuals, and h^2 is the ratio of additive genetic variance to phenotypic variance. Assuming equal selection differentials then u' will increase as u and h^2 increase. A high u can be assured by crossing elite parents that have an acceptable level of expression for the trait to be improved and h^2 can be improved in a population by crossing parents that each posses different alleles at the loci that control the trait. Elite parents can be identified through testing and their use as parents is a common practice in cultivar development programs. A quantification of the actual amount of allelic differences between two parents, defined as their genetic distance, can be estimated but is rarely used in parent selection as breeders usually cross lines that they instinctively deem as distant.

There are various methods of estimating the genetic distance between two individuals and each has it's own assumptions and biases that could influence it's ability to accurately estimate the true genetic distance. Genetic distance can be estimated from pedigree data by the coefficient of parentage (CP) (Kempthorne, 1969) which is the probability of sampling an allele from one individual that is identical by descent to an allele sampled from the same locus in the other individual. The value (1 - CP) estimates the genetic distance and the ability of this value to predict the true genetic distance between individuals is entirely dependent on the validity of the assumptions that are made during it's calculation. There are extensive pedigree records for adapted soybeans that would allow this approach to be used with an elite population. Another approach is to use highly heritable phenotypes, such as isozymes, restriction fragment length polymorphisms (RFLPs), and certain morphological traits, as qualitative markers to assay for differences between parents at the DNA level. The loci the control such polymorphisms are termed marker loci (ML). The frequency of each polymorphism at each ML in each parent is determined and a comparison equation is employed to estimate the genetic distance between parents. The accuracy of the estimate will depend on obtaining a representative sample of potentially polymorphic loci and on the assumption that alleles that produce identical phenotypes are themselves identical. Suitable polymorphisms occur with a moderate frequency in adapted soybeans (Doong and Kiang, 1987; Gorman *et al.* 1983; Kiem *et al.* 1989) making this method a practical option. A third approach infers genetic distance between parents from a statistical distance based on quantitative trait differences (Goodman, 1972) and can assay for allelic differences at many loci when truly polygenic traits are measured. Some limitations of this approach are that each quantitative phenotype assays an unknown number of genetic differences that may not be proportional to the observed phenotypic differences and the effect of the environment on the traits requires that each trait and genotype be assayed in multiple locations and years. There is a restricted range of phenotypes among the elite soybean lines that might limit the utility of this approach within this population

There has been extensive use of genetic distances to estimate the relationship among individuals and yet there has been relatively little research investigating whether estimates of parental genetic distance relate to the parameters of a hybrid or segregating population that would be expected to be influenced by this distance and that concern a breeder. Such parameters include the amount of heterosis, the number of transgressive segregants, or the amount of genetic variance (Cowen and Frey, 1987). A low, though significant, positive correlation between parental genetic distances estimated with CP and hybrid mid-parent heterosis for grain yield was reported in rapeseed (Lefort-Buson *et al.*, 1987) while the association was not significant in oats (Cowen and Frey, 1987) or wheat though Cox and Murphy (1990) noted the value of using the CP to select diverse wheat parents after the parental candidates had already been selected for their performance and phenotypic divergence. Cowen and Frey (1987) found a significant association between the CP of the parents of a segregating oat population and the number of transgressive segregants for plant height but not for the traits bundle weight, grain yield, straw yield, heading date, and harvest index.. They also reported a significant association between the CP of the parents and the generalized genetic variance of the segregating population.

There have been several attempts, with mixed results, to correlate a ML derived parental genetic distance with the performance of hybrids between inbred maize lines, which is generally dependent on the extent of heterosis. Price *et al.* (1986) and Lamkey *et al.* (1987) found a significant but low positive correlation between hybrid grain yield and inbred parental genetic distances based on isozymes indicating that the diversity estimate had little value in predicting hybrid performance. Other researchers using isozymes (Hunter and Kannenberg, 1971; Hadjinov *et al.* 1982) and RFLPs (Godshalk *et al.*, 1990) have found insignificant correlations between hybrid grain yield of maize and parental genetic distances. Frei *et al.* (1986) found the ML genetic distance between parents to be predictive of hybrid performance only when there was a known, pedigree relationship between the parents; a situation that increases inbreeding and the probability of linkage disequilibrium between ML and other genes. Lee *et al* (1989), using RFLPs to estimate parental diversity within and across chromosomes, found a significant correlation between parental genetic distance and hybrid grain yield of maize and that the diversity of some chromosomes was more predictive of yield than other chromosomes suggesting that this approach would be more successful when only the diversity of relevant chromosome segments is used. No research has been performed relating ML derived genetic distances to the amount of genetic variance or the number of transgressive segregants in segregating populations.

The purpose of the reported research was to investigate if a CP or a ML based estimate of genetic distance between parents could be used to to predict which elite soybean lines, when crossed, would produce the greatest amount of useful genetic variability in the resulting segregating population.

Materials and Methods

Sixty-two public breeding lines and public and private cultivars were selected as elite parents on the basis of their high yield and adapted maturity in performance trials conducted over several years by Michigan State University. The genotype of each elite line was determined for the following five morphological and thirteen isozyme loci: the *W, T, L, I,* and *R* loci that control the color of the flower, pubescence, pod, hylum, and hylum respectively, diaphorase (Dia_1 locus; EC 1.6.4.3), endopeptidase (*Enp* locus), isocitrate dehydrogenase (Idh_1, Idh_2 loci; EC 1.1.1.42), mannose phosphate isomerase (*Mpi* locus; EC 5.10.11), phosphoglucosemutase (Pgm_1 locus; EC 5.3.1.9), and superoxide dimutase (*Sod* locus; EC 1.15.1.1) acid phosphatase (*Acp* locus; EC 3.1.3.2), aconitase (Aco_2, Aco_4 loci; EC 4.2.1.3), fluorescent esterase (*Fle* locus), 6-phosphogluconate dehydrogenase (Pgd_1 locus; EC 1.1.1.44), and phosphoglucose isomerase (*Pgi* locus; EC). The isozyme methodology and results reported in Sneller (1991). The level of peroxidase activity was also ascertained according to the method of Buttery and Buzzel (1968). The data from these loci were used to calculate a similarity index (SI) that is the probability of sampling an allele from one individual that is identical by state to an allele sampled from the same locus in the other individual, averaged over all loci (Sokal and Sneath, 1963). The SI value can range from unity for two homozygous, homogeneous lines that contain the same alleles to zero for lines that have no alleles in common. The SI was calculated between all pairs of the 62 elite lines using a program written in SAS Interactive Matrix Language by the author and was used to select similar and distant parents which were then crossed to produce the segregating populations that were analyzed in this study (Table 1).

The CP was calculated between the 62 elite lines according to the formula of Kempthorne (1969). Derivations of this formula (Sneller, 1991) were used to calculated the CP between any genotype and "Elgin" and "Kenwood" which were derived from the AP6 soybean population (Fehr and Ortiz, 1978) that was created by an essentially random mating of 40 genotypes. The following assumptions were made in all CP calculations: (i) all the ancestral parents were completely unrelated, including lines selected from other ancestral parents, (ii) all ancestral parents were completely homogeneous and inbred, and (iii) that each parent of a biparental cross contributed equally to all progeny derived from the cross. A genotype derived by five or more backcrosses was considered to be genetically equivalent to the recurrent parent. The CP between the parents that were crossed to generate the populations used in this study are listed in Table 1. All CPs were calculated using programs written by the author in SAS Interactive Matrix Language and a complete summary of the CPs among the elite lines can be found in Sneller (1991)

A distance estimate was calculated that encompassed both the CP and marker loci data. The CP was calculated between each of the 62 elite lines and each of the n ancestral parents that contributed genes to the elite population resulting in a data set where each elite line was described by n variables. This data set was combined with the marker loci data set and a principal component analysis was performed on the correlation matrix. The principal component distance (PCD) was calculated between all elite lines according to the formula

$$PCD_{ij} = [\ \Sigma\ \{(Y_{ik} - Y_{jk})^2\ /\ \lambda_k\}\ ]^{1/2}$$

where Y_{ik} and Y_{jk} are the scores of the k^th^ principal component for the i^th^ and j^th^ elite lines and λ_k is the eigenvalue of the k^th^ principal component (Goodman, 1972). Only principal components where λ_k was greater than one were used in the calculation. The PCD between the parents used to generate the populations analyzed in this study are listed in Table 1. The PCDs were calculated using a program written

Table 1. Summary of the parentage, coefficient of parentage (CP), similarity index (SI), and the principal component distance (PCD) between the parents for all crosses used to generate the segregating populations along with the number of families in each population and the years tested.

Cross	Parent 1	Parent 2	CP	SI	PCD	Number of families	Year(s) tested
1	Hack	Pella	.48	.63	2.40	54	1990, 1989
2	Hardin	A80-147002	.29	.58	3.01	69	1990, 1989
3	A80-147002	A82-161035	.33	.90	1.88	60	1990, 1989
4	A83-271010	s2596	.23	.66	3.02	58	1990, 1989
5	Burlison	Century	.59	.76	2.77	50	1990, 1989
6	C1664	Century	.56	.74	2.19	66	1990, 1989
7	HW8008	Century	.72	.95	2.23	55	1990, 1989
8	HW8008	LN82-296	.55	.90	1.70	54	1990, 1989
9	A2934	s2596	.17	.63	5.02	58	1990, 1989
10	A80-147002	E84098	.25	.45	3.64	47	1990, 1989
11	s2596	A82-161035	.26	.58	3.34	63	1990, 1989
12	A1937	E85100	.29	.66	1.64	51	1990
13	A83-271010	s18-84	.56	.71	2.79	54	1990
14	E85168	A84-185032	.26	.63	1.97	55	1990
15	C1664	7260	.27	.66	2.38	54	1990
16	Hack	E84150	.31	.50	3.79	53	1990
17	s23-12	Pella	.10	.40	4.12	44	1990
18	Pella	E85166	.43	.79	1.66	54	1990
19	s23-12	1250	.17	.66	4.99	54	1990
20	1250	E85166	.29	.47	3.82	52	1990
21	s23-12	E85166	.16	.40	4.32	53	1990
22	Burlison	M82-946	.19	.37	4.57	52	1990
23	E84165	Century	.12	.55	3.14	60	1990
24	A2934	A83-271010	.18	.45	4.98	43	1990

in SAS Interactive Matrix Language by the author while the principal component analysis was performed with standard SAS software.

The progeny from eleven different crosses were field tested in both 1989 and 1990 (Table 1). The crosses were made during the summer of 1987, the hybrids were grown in a greenhouse during the winter of 1987-88 and the F_2 populations were grown in the field during the summer of 1988. An average of 59 $F_{2:3}$ families were generated from each cross (Table 1). Each family of each cross was field tested for seed yield, plant height, and date of maturity in single row, 1.2 M long, hill plots, planted in 76.2 cm rows in a randomized complete block design experiment with two replications. The seeding rate was 27 seeds per meter of row. The experiment was planted and harvested in Lenawee county, Michigan, in 1989 and in Ingham county, Michigan, in 1990. An additional 13 crosses, each represented by an average of 52 $F_{4:5}$ families, plus the parents of all crosses were similarly tested in 1990 only (Table 1) at the Ingham county location. These crosses were made in 1987, the hybrids grown during the summer of 1988, and then advanced through two generations of single seed descent during the winter of 1988-89 with the $F_{4:5}$ families being harvested from the 1989 summer planting of F_4 seed. These crosses were added to the study to provide a broader range of SI and CP values for the 1990 data analysis.

The variance among the progeny families of a cross for a particular trait was estimated from a separate ANOVA of that trait for each cross from the 1989, 1990, and the combined 1989-90 data sets. The progeny variances ($\sigma^2_{progeny}$) from the single years and from the analysis combined over both years were interpreted as follows:

$$\text{Single year } \sigma^2_{progeny} = \sigma^2_g + \sigma^2_{gxe}$$

$$\text{Two year } \sigma^2_{progeny} = \sigma^2_g$$

where σ^2_g and σ^2_{gxe} are the genetic and genotype by environment variances of the population respectively. The generalized variance of each progeny population was estimated by taking the determinate of the variance-covariance matrix of the trait means from each family member for the 1989, 1990 and 1989-1990 data sets. The natural log of the generalized variance (GV) was used as an estimate of the overall variability of each population (Goodman, 1968; Sokal, 1965). All statistics were performed with SAS software.

Results

The generalized variances and progeny variances for the traits seed yield, maturity, and height of the populations estimated in 1989, 1990 and 1989-90 are presented in Table 2. The estimate of the progeny variance was negative in some cases and these values were assumed to be zero in any further analyses. The CP, SI, and PCD estimates of the genetic distance between the parents of the crosses were significantly correlated to one another (Table 3) and each was significantly correlated to the GVs of the populations derived from these crosses regardless of the year in which the GVs were estimated. The CP and SI measure the genetic similarity between two parents and the significant negative correlations of these values with the GVs means that the overall variability of a population decreased as the estimated genetic similarity of it's parents increased. The PCD estimates parental genetic distance, not similarity, and this measure was positively correlated with the GV values of the segregating populations. Only linear effects were significant in the regressions of the GVs of the populations on CP, SI or PCD values between the parents (Table 4). Each measure of parental genetic distance accounted for a significant amount of the variation among the GVs of the populations and all were approximately equal in their ability to predict the GV of a population. The regressions of 1990 GVs on the CP, 1989 GVs on the SI, and the 1989-90 GVs on the PCDs are shown in Figures 1-3 respectively.

The GV of a population encompasses the genetic variation for all traits, the genetic covariance between all traits, all non-genetic sources of variation and covariation as well as the interactions between these effects. The 1989 estimates of the GVs were significantly correlated (r = 0.93) with the 1990 estimates suggesting that the GV

Table 2. Summary of the generalized variances (GV) and the maturity, height, and seed yield progeny variances (PV) estimated in 1989, 1990 and 1989-90 for each population.

	GV			Maturity PV			Height PV			Yield PV		
Cross	1989	1990	1989-90	1989	1990	1989-90[a]	1989	1990	1989-90	1989	1990	1989-90
1	12.1	10.7	9.9	15.4	1.6	3.6	2.0	0.5	2.2	2409.4	1684.1	0.0
2	13.7	13.0	12.5	36.8	7.2	16.8	16.1	11.0	13.4	7535.1	44.4	1659.1
3	11.0	11.0	9.8	4.4	1.5	1.9	1.0	2.8	1.3	1086.6	766.4	354.6
4	12.3	10.9	10.6	10.0	1.0	2.0	6.0	4.4	3.0	2280.6	1360.9	410.6
5	11.1	10.4	9.7	6.0	1.5	2.8	2.2	1.5	1.3	633.1	235.0	1096.2
6	10.6	10.3	9.2	4.7	1.6	2.1	0.6	0.1	0.0	1035.1	0.0	224.9
7	11.1	10.6	9.6	4.9	1.1	1.0	2.5	1.2	1.4	453.6	1333.8	306.5
8	11.1	10.2	9.3	5.1	1.2	2.2	0.7	0.0	1.7	495.3	0.0	26.1
9	13.4	12.4	12.0	33.0	9.2	15.6	6.9	8.3	9.0	6930.6	595.3	1214.6
10	13.7	12.9	12.0	23.4	10.1	8.8	10.0	5.3	5.2	5383.7	1307.1	12.4
11	12.1	11.3	10.7	8.1	1.6	2.3	5.3	5.5	2.9	516.1	806.3	574.9
12	.	10.8	.	.	0.0	.	.	7.1	.	.	2113.0	.
13	.	10.8	.	.	0.3	.	.	3.5	.	.	1122.5	.
14	.	10.3	.	.	0.2	.	.	5.6	.	.	1201.3	.
15	.	11.1	.	.	1.3	.	.	3.2	.	.	1298.6	.
16	.	11.4	.	.	2.7	.	.	5.4	.	.	1105.5	.
17	.	11.8	.	.	3.5	.	.	6.5	.	.	486.9	.
18	.	11.3	.	.	2.5	.	.	1.0	.	.	0.0	.
19	.	13.5	.	.	8.4	.	.	7.9	.	.	2314.5	.
20	.	13.5	.	.	13.3	.	.	13.5	.	.	66.3	.
21	.	11.7	.	.	3.7	.	.	6.4	.	.	42.8	.
22	.	11.5	.	.	2.8	.	.	6.1	.	.	978.0	.
23	.	12.3	.	.	3.0	.	.	8.8	.	.	1696.9	.
24	.	13.4	.	.	15.4	.	.	8.9	.	.	2053.5	.

[a] The 1989-90 progeny variance estimates the genetic variation within a population for a particular trait

Table 3. Spearman's rank correlations of the generalized variances (GV) and the maturity, height, and seed yield progeny variances (PV) of the segregating populations with the coefficient of parentage (CP), similarity index (SI), and the principal component distance (PCD) values of the parents of the populations, summarized by year(s).

	Year(s)	Distance measure		
		CP	SI	PCD
GV	1989	-.72**	-.65***	.81***
GV	1990	-.67***	-.79***	.73***
GV	1989-1990	-.79***	-.81***	.78***
Maturity PV	1989	-.65**	-.80***	.77***
Maturity PV	1990	-.44	-.78***	.56*
Maturity PV	1989-1990[a]	-.45	-.77***	.60**
Height PV	1989	-.65**	-.68**	.84***
Height PV	1990	-.73***	-.67***	.80***
Height PV	1989-1990	-.77***	-.73**	.75***
Yield PV	1989	-.67**	-.69**	.55*
Yield PV	1990	-.24	-.17	.33
Yield PV	1989-1990	-.28	-.13	.41

[a] The 1989-90 progeny variance estimates the genetic variation within a population for a particular trait.

*, **, ***, denote significance at the a = 0.10, 0.05, and 0.01 levels respectively. The 1990 correlations were estimated with 22 df while the 1989 and 1989-1990 correlations were estimated with 9 df.

Table 4. Summary of the R^2 values and their significance from the regressions of the GV and maturity, height and seed yield progeny variances (PV) of the segregating populations on the coefficient of parentage (CP), similarity (SI) and principal component distance (PCD) values of the parents of the populations, summarized by year(s).

		Distance measure		
	Year(s)	CP	SI	PCD
GV	1989	.56***	.66***	.58***
GV	1990	.37***	.32***	.53***
GV	1989-90	.59***	.57***	.61***
Maturity.PV	1989	.37**	.45**	.50**
Maturity.PV	1990	.17**	.23**	.47***
Maturity PV	1989-90[a]	.32*	.31*	.48**
Height PV	1989	.36*	.47**	.31*
Height PV	1990	.50***	.42***	.38***
Height PV	1989-90	.34*	.29*	.37**
Yield PV	1989	.41**	.42**	.49**
Yield PV	1990	.06	.01	.03
Yield PV	1989-90	.09	.05	.24

[a] The 1989-90 progeny variance estimates the genetic variance within a population for a particular trait.

*, **, ***, denote significance of the linear regression model at the a = 0.10, 0.05, and 0.01 levels respectively. The error of the 1990 regressions had 22 df while the 1989 and 1989-1990 regression errors had 9 df.

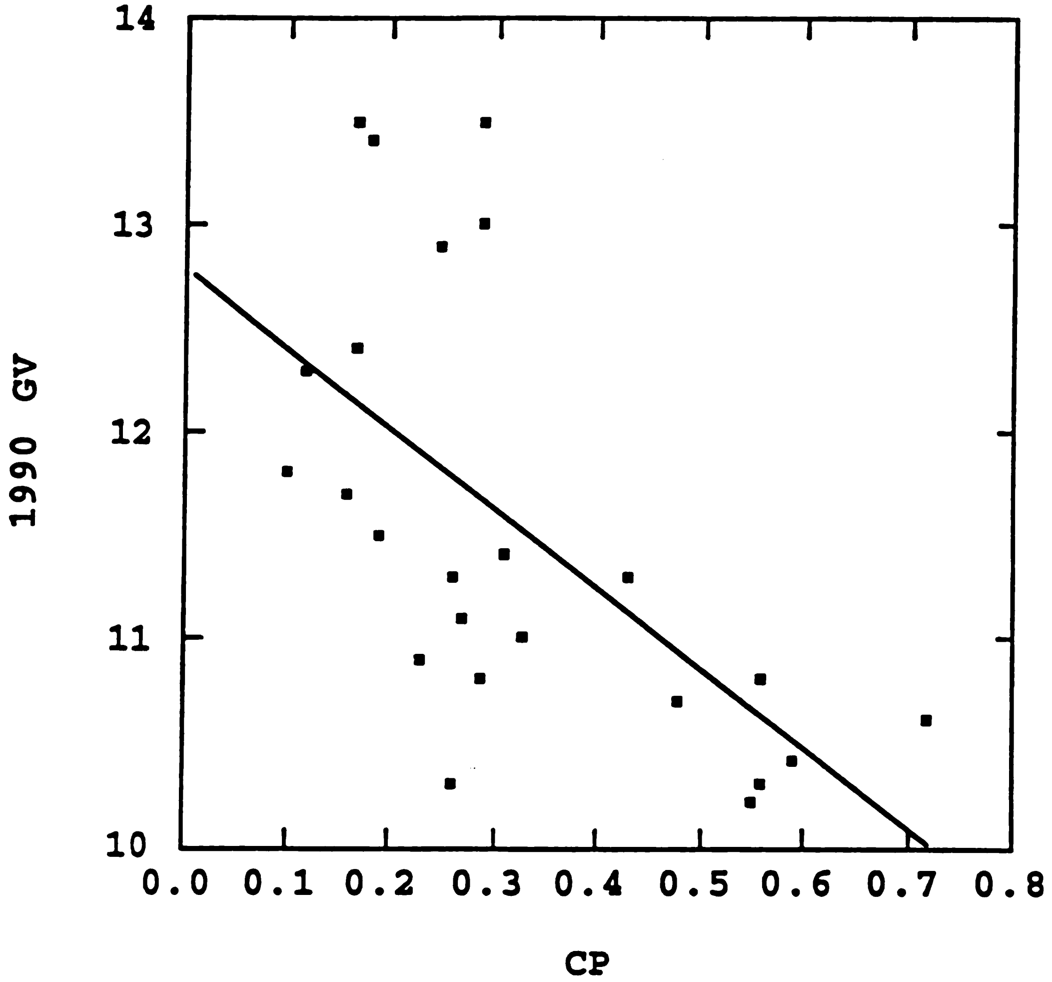

Figure 1. Results of the linear regression of the 1990 estimates of the generalized variances (GV) of the segregating populations on the coefficient of parentage (CP) of the parents of the populations.

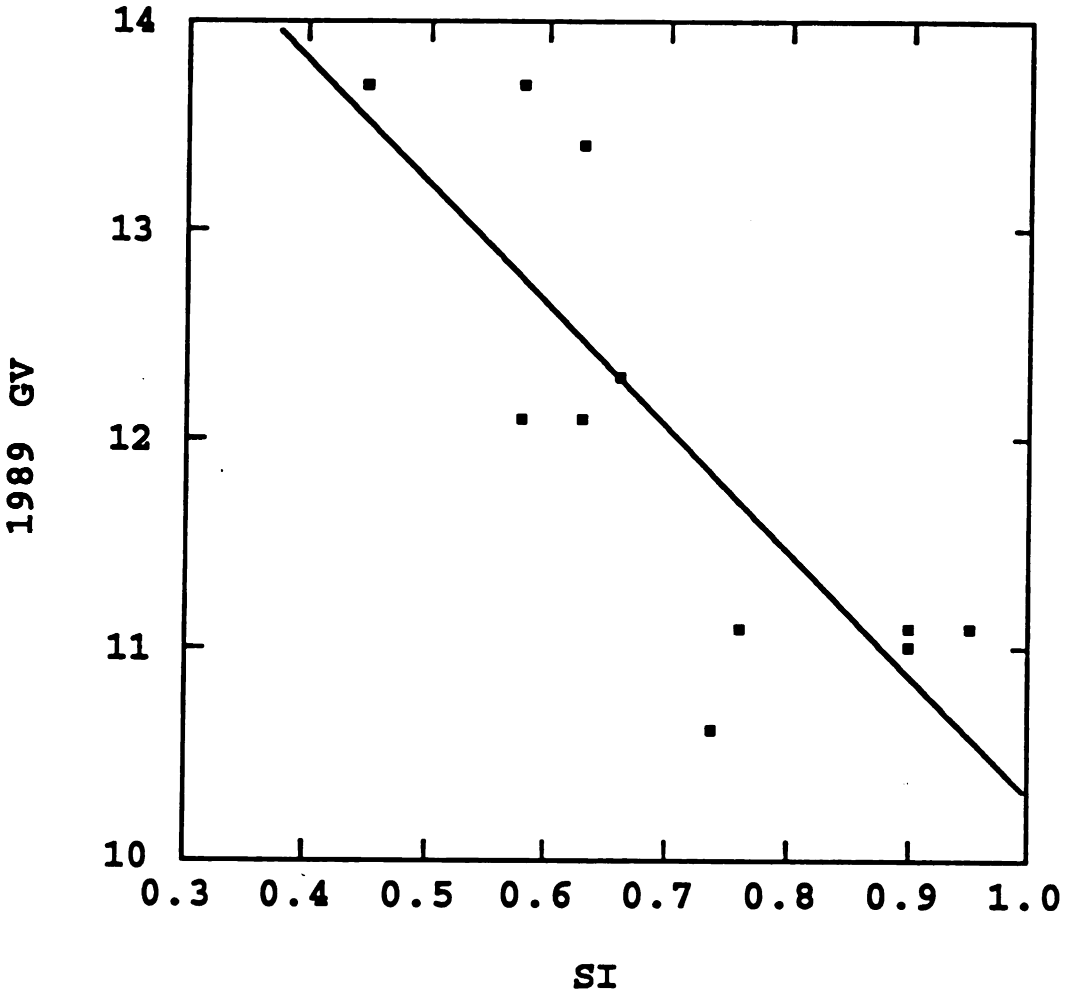

Figure 2. Results of the linear regression of the 1989 estimates of the generalized variances (GV) of the segregating populations on the similarity index (SI) of the parents of the populations.

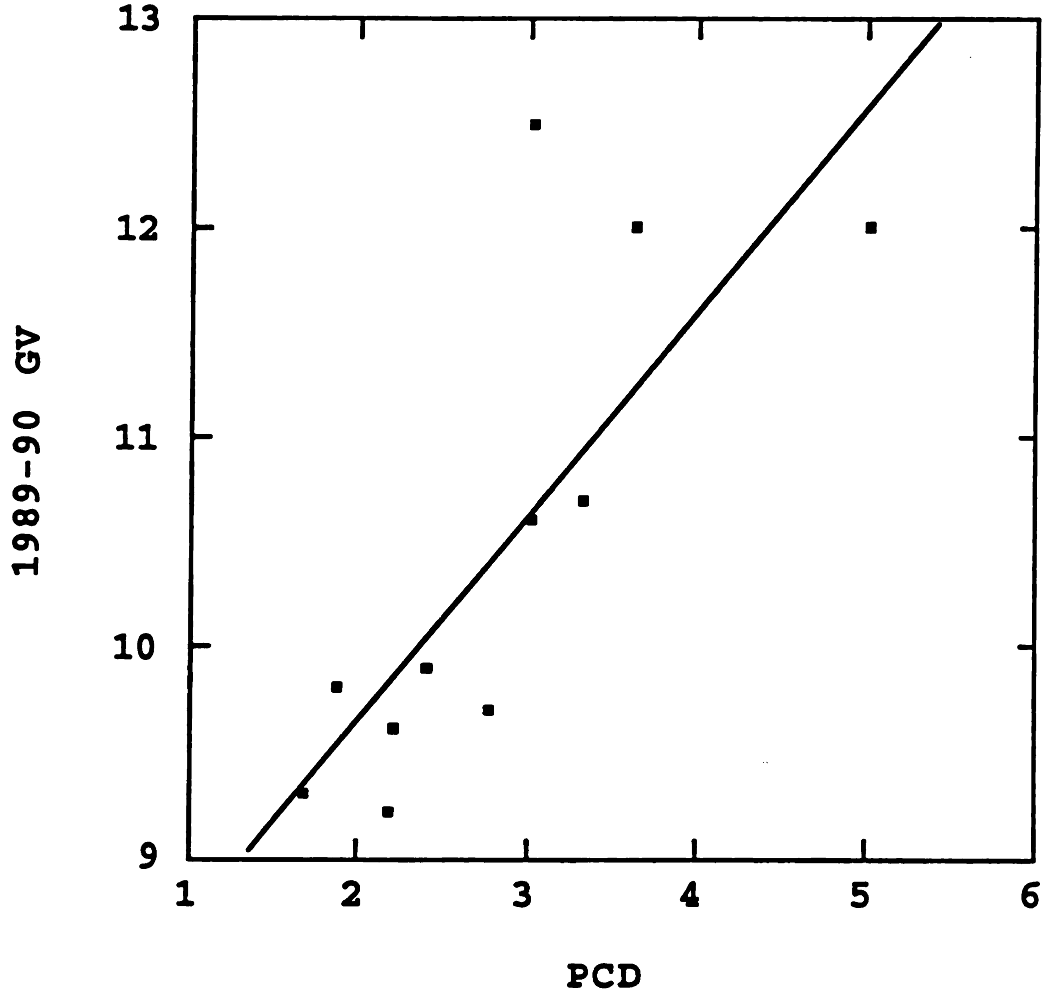

Figure 3. Results of the linear regression of the 1989-90 estimates of the generalized variances (GV) of the segregating populations on the principal component distance (PCD) between the parents of the populations.

was a repeatable characteristic of a population and was not an artifact of random error or genotype by environment variances and covariances. In general, the progeny variances for individual traits from a particular year(s) were significantly correlated to the GVs estimated in that year(s), the only exceptions being the 1990 yield progeny variances with all GV estimates and the 1989-90 yield progeny variances with the 1989 and 1990 GV estimates. An analysis of the partial correlations between the progeny variances and the GVs showed that the correlations of the yield progeny variances with the GVs were lower when either the maturity or height progeny variance was held constant, suggesting that yield progeny variance was probably not important in determining the magnitude of the GV estimates.

Table 3 shows the correlations of the measurements of parental genetic distance with the progeny variances for the individual traits. All correlations possessed the expected sign under the hypothesis of greater progeny variance with greater parental genetic distance. All correlations were significant between the maturity progeny variances of the populations and the CP, SI and PCD of the parents except for those between the CP and the 1990 and 1989-90 maturity progeny variances. The linear regressions of maturity progeny variance on the CP, SI and PCD values were all significant (Table 4) though the R^2 values were lower than the R^2 values obtained by regressing the GVs on these values indicating that the CP, SI, and PCD values were less predictive of the maturity progeny variance of a population than of the GV of a population. The 1989 and 1990 progeny variances contains genetic and genotype by environment variance while the 1989-90 estimate contains only the genetic variance of a population. The significant regressions of the genetic variance on all three measures of parental genetic distance showed that they all were able to predict to some degree the genetic variation for maturity of the populations. The PCD measure of parental genetic distance was able to account for more of the variability among the maturity genetic variances of the populations than either the CP or SI could and Figure 4 shows a plot of the regression of the genetic variances of the populations on the PCD values. The maturity of the parents of each population was determined in 1990 and the difference between the parents of a population was

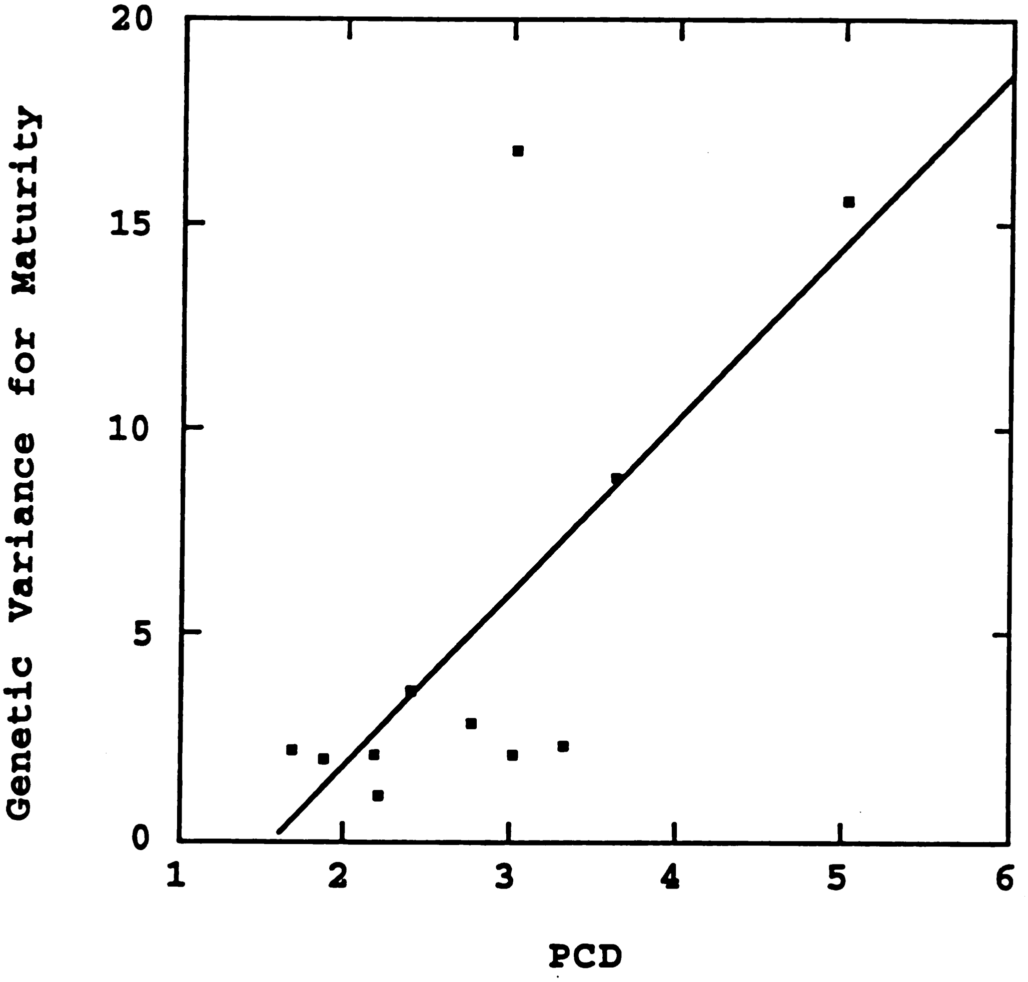

Figure 4. Results of the linear regression of the maturity genetic variances of the segregating populations on the principal component distance (PCD) between the parents of the populations.

not correlated to the amount of progeny variance for maturity in a population indicating that the magnitude of these variances were not influenced by the phenotypic differences of the parents.

All three parental genetic distance measures were significantly correlated to the progeny variances for plant height (Table 3). All linear regressions of the height progeny variances of the populations on either the CP, SI or PCD measures of parental distance were significant but resulted in lower R^2 values than the regressions of the GVs of the populations on these measures, indicating that they had less value in predicting the height progeny variance than the GV of a population (Table 4). The CP and PCD measures had about equal value in predicting the genetic variance for height of a population and were able to account for more of the variation among these variances than the SI measure could. Figure 5 shows the regression of the height genetic variances of the populations on the PCD of the parents of the populations. The height the parents of each population was determined in 1990 and the difference between the parents of a particular population was significantly correlated to the amount of progeny and genetic variance for height in that population suggesting that the larger phenotypic differences of some of the parents may have contributed to the larger height progeny variance of those populations.

While all three parental genetic distance measures were significantly correlated to the 1989 yield progeny variances of the populations, this was not repeated with the 1990 or 1989-90 data (Table 3). The regression analyses produced similar results (Table 4) (Figures 6 and 7). The R^2 values of the genetic variance for yield with all three measures were low, indicating that none were predictive of this parameter. Table 5 summarizes the average yield progeny variances for the populations whose parents had either a higher or a lower CP, SI or PCD value than the mean of these measures. This summary suggests that while none of the measures appeared to be predictive of yield progeny variance and genetic variance, that in general, populations that were derived from genetically distant parents higher yield progeny variances and genetic variances than the populations derived from parents that were similar. The yield difference of the parents of each population was determined in 1990

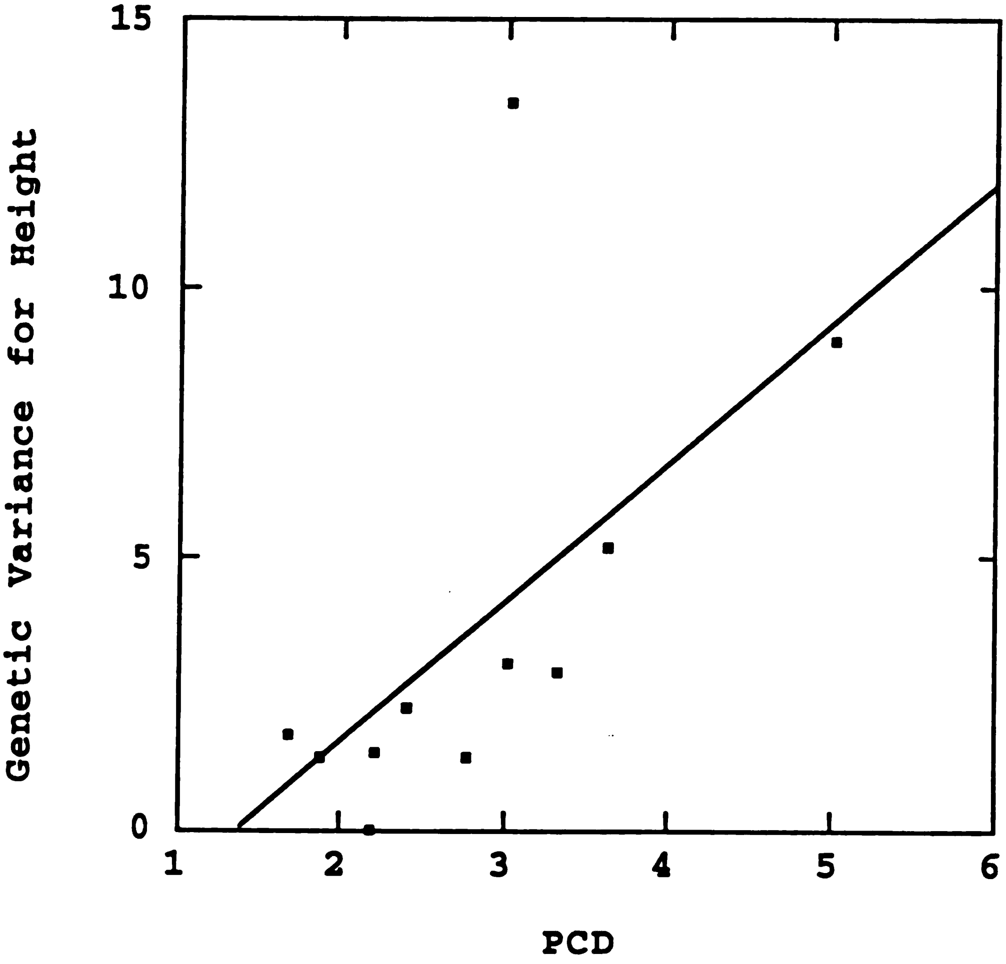

Figure 5. Results of the linear regression of the height genetic variances of the segregating populations on the principal component distance (PCD) between the parents of the populations.

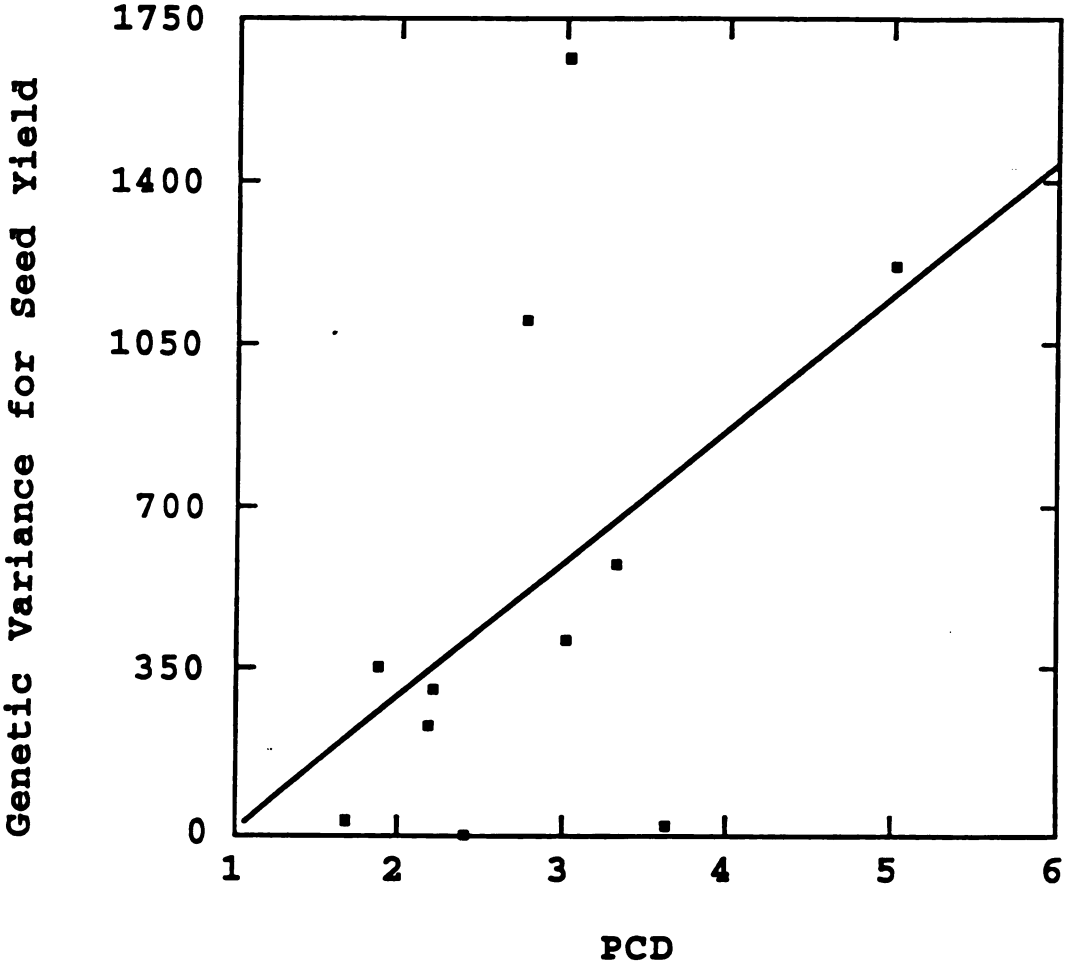

Figure 6. Results of the linear regression of the seed yield genetic variances of the segregating populations on the principal component distance (PCD) between the parents of the populations.

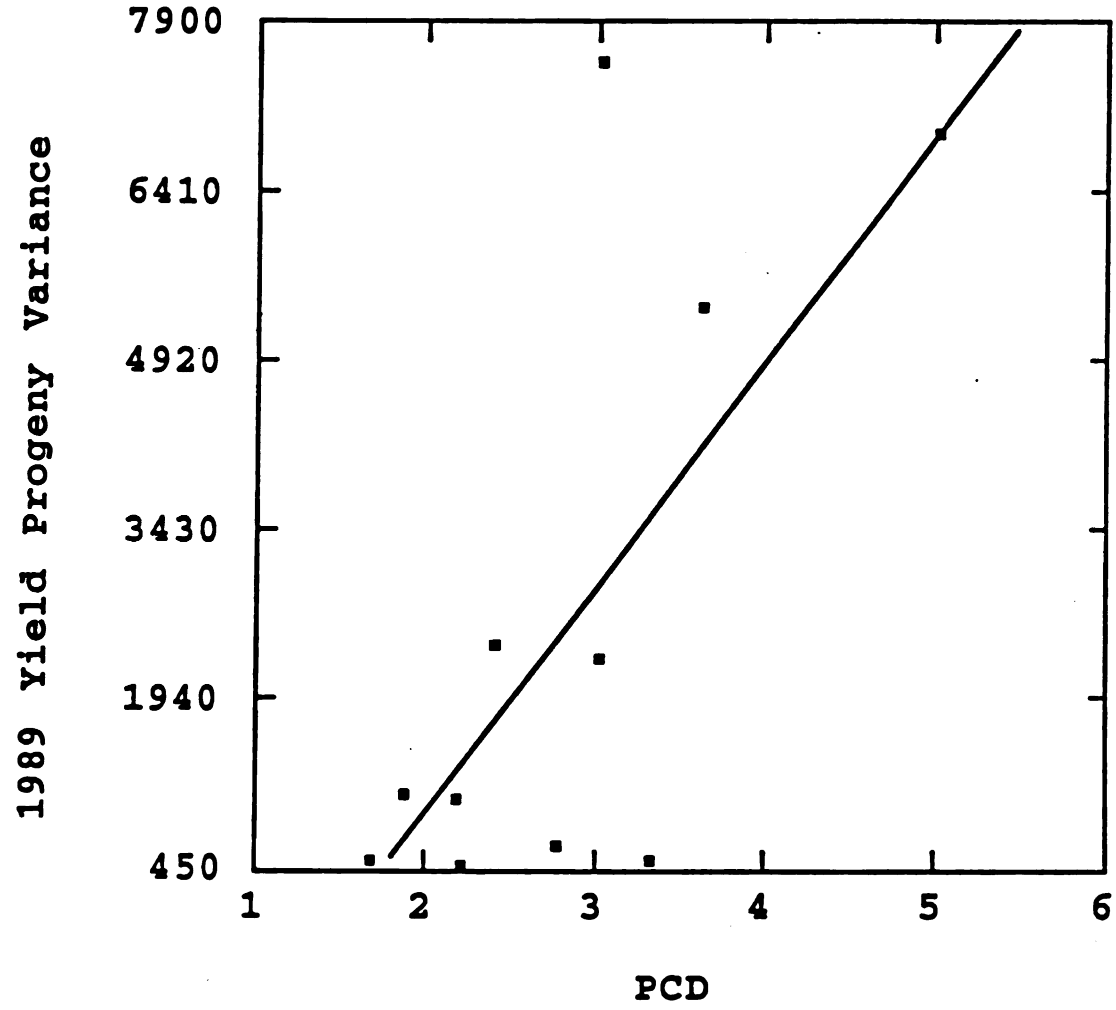

Figure 7. Results of the linear regression of the 1989 estimates of the seed yield progeny variances of the segregating populations on the principal component distance (PCD) between the parents of the populations.

Table 5. The average seed yield progeny variances of the populations whose parents had coefficient of parentage (CP), similarity index (SI) or principal component distance (PCD) values that were either higher (High) or lower (Low) than the mean of these measures, summarized by years.

Distance measure	High or low	Year(s) 1989	1990	1989-90[a]
CP	Low	4529.2	1092.0	774.3
CP	High	1018.9	503.2	309.3
SI	Low	4555.0	928.3	661.7
SI	High	997.4	857.1	403.2
PCD	Low	1991.1	838.6	490.7
PCD	High	4276.8	975.6	600.6

[a] The 1989-90 progeny variance estimates the genetic variance for seed yield within a population

and found not to be correlated to the amount of progeny variance for yield.

Discussion

The amount of genetic variation in a segregating population should increase as the genetic distance of the parents of the population increases. The CP and SI estimate the percentage of the loci that will be homozygous in a hybrid of two parents. As the CP and SI values decrease, F_1 heterozygosity should increase and the progeny populations derived by selfing the F_1 should become more variable under the assumption of equal genetic effects at all loci. The PCD measures the genetic distance between individuals and the heterozygosity of an F_1 should increase as the PCD of the parents increase. All three measures of parental genetic distance were significantly associated with the amount of overall variability in the segregating populations analyzed in this study, suggesting that each was predictive, to some degree, of the true genetic distance between the parents. Each measure was able to account for approximately 60% of the variation among the GVs of the populations estimated from the 1989-90 data set (Table 4) and therefore each appeared to be a useful tool in selecting elite soybean parents that would produce variable progeny populations from which a breeder could select diverse phenotypes. Cowen and Frey (1987) found a similar association between the CP of the parents and the GV of segregating oat populations.

The ability of the measures of parental genetic distance to predict the amount of variation among the F_2 or F_4 derived progeny of a cross for individual traits was lower than their ability to predict the GV of a population (Table 4). This may be explained by the fact that all three measures estimate the genetic distance between lines, averaged across the whole genome. As such they should be more predictive of the genetic distance for traits that are controlled by many loci dispersed throughout the genome than for traits that are controlled by relatively few loci as the violations of the assumptions of each measure that occur at individual loci would tend to cancel each other out when averaged over a larger number of loci. For example the CP assumes that each inbred parent contributes equally to each locus in all progeny of a cross when in reality an inbred progeny carries an allele from only one of the parents. The discrepancies at individual loci will average out as more loci are "sampled" by a trait such this assumption will in effect be true when averaged across many loci and when the other CP assumptions are not violated. The GV of a segregating population is affected by the genetic distance of the parents at the loci that control all of the measured traits such that it effectively samples more loci than the individual traits and therefore the measures of genetic distance may be more accurate for the loci influencing the GV than for the loci contributing to the progeny variance or the genetic variance of the individual traits.

None of the measures of parental genetic distance had a repeatable association with the seed yield progeny variances and all were unable to predict the seed yield genetic variance of a population that results from crossing two elite soybean lines (Table 4). While this may limit the utility of these measures in a soybean cultivar development program, the data presented in Table 5 and Figure 6 do suggest some advantage in using these measures to decide which elite lines to cross to produce segregating populations which will have a higher likelihood of having an above average genetic variance for seed yield. The measures of parental genetic distance were significantly better at predicting the maturity and height progeny variance and genetic variance of a population, though their ability to predict these parameters were still relatively low (Table 4) suggesting that their utility in a breeding program for these traits would be similar to that described above for seed yield. The progeny variances estimated in a single year contain genotype by environment variance that the measures of parental genetic distance can not predict. Yet the CP, SI and PCD values were still significantly associated with these estimate of progeny variance for the traits maturity and height and this association was repeated in both years of the study even when additional populations were evaluated in 1990. This suggests that the genotype by environment component of the single year progeny variances for these traits was small or proportional to the genetic variance. Seed yield generally has the lowest heritability of all the measured traits (Brim, 1973) and it is possible that the single year yield progeny variances contained proportionately larger contributions of random or population specific genotype by environment effects than the other traits. This may explain why the significant association of the measures of parental distance with the 1989 estimates could not be repeated in 1990. The low heritability of seed yield also makes it harder to obtain an accurate estimate of the genetic variance of a population and may explain why the measures of parental genetic distance were unable to predict the magnitude of this parameter.

It is not clear which measure of parental genetic distance was the better predictor of the GVs and the progeny or genetic variance of a population as all were highly correlated to one another and produced fairly similar results. The PCD distance was able to account for more of the variability among the 1989-90 GVs and the genetic variances of the individual traits of the populations than either the CP or SI measures could (Table 4) and on this basis it appeared to be the better measure. The PCD measure has the advantage of combining the purely theoretical CP distance with actual data that evaluates genetic distances at the DNA level and this may allow the shortcomings of one measure to offset the shortcomings of the other. Cox *et al.* (1985) noted that the biases of the CP and SI measures would tend to cancel each other out. The SI data can monitor and adjust for violations of the CP assumptions that all ancestral parents are completely unrelated and that each parent contributes equally to all progeny while the CP can estimate the probability of genetic distance at loci that were not effectively assayed with linked ML. The PCD estimate of genetic distance can be improved by evaluating more ML that can assay for genetic differences in additional regions of the genome, while the CP estimate of genetic distance can not be improved. The research of Lee *et al.* (1989) indicates that the predictive value of a ML measure of genetic distance for a particular trait may be improved by assaying for genetic differences only in those regions of the genome that affect the trait of interest thereby eliminating useless contributions to the genetic distance estimate from irrelevant portions of the genome. Pertinent chromosome segments could be identified through ML/quantitative trait loci studies using mapped ML.

In conclusion, all three measures of parental genetic distance appeared to have utility in selecting elite soybean parents that when crossed would have a high likelihood of producing segregating populations with above average genetic variability. The use of such measures could allow a breeder to eliminate crossing certain parents and allow cultivar development resources to be allocated to populations with a greater potential of producing outstanding progeny.

List of References

List of References

Brim, C.A.. 1973. Quantitative genetics and breeding: in *Soybeans: Improvement, Production and Uses.* ed B. Caldwell. American Society of Agronomy, Madison.

Buttery, B.R., and R.I. Buzzell. 1968. Peroxidase activity in seeds of soybean varieties. Crop Sci. 8:722-725.

Cardy, B.J. and W.D. Beversdorf. 1984. A procedure for starch gel electrophoretic detection of isozymes of soybean (*Glycine max* L. Merr.). Departemnt of Crop Science Technical Bulletin 119/8401. University of Guelph, Guelph Ontario, Canada, N1G 2W1.

Cowen, N.M., and K.J. Frey. 1987. Relationship between genealogical distance and breeding behavior in oats. Euphytica 36:413-424.

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.

Cox, T.S., and J.P. Murphy. 1990. The effect of parental divergence on F_2 heterosis in winter wheat crosses. Theor. Appl. Genet. 79:241-250.

Doong, J.Y.H., and Y.T. Kiang. 1987. Cultivar identification by isozyme analysis. Soy. Genet. Newsl. 14:189-225.

Falconer, D.S.. 1981. *Introduction to quantitative genetics.* Longman House, New York.

Fehr, W.R., and L.B. Ortiz. 1975. Registration of a soybean germplasm population. Crop Sci 15:739.

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.

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

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

Goodman, M.M.. 1972. Distance analysis in biology. Syst. Zool. 21:174-186.

Gorman, M.M., Y.T. Kiang, Y.C. Chiang, and R.G. Palmer. 1983. Electrophoretic classification of the early maturity groups of named soybean cultivars. Soybean Genet. Newsl. 9:143-156.

Hadjinov, M.I., V.S. Scherbak, N.I. Benko, V.P. Gusev, T.B. Sukhorzhevskaya, and L.P. Voronova. 1982. Interrelationship between isozymic diversity and combining ability in maize lines. Maydica 27:135-149.

Hunter, R.B., and L.W. Kannenberg. 1971. Isozyme characterization of corn (*Zea mays*) inbreds and it's relationship to single cross hybrid performance. Can. J. Gent. Cytol. 13:649-655.

Kempthorne, O.. 1969. *An Introduction to Genetic Statistics*. Iowa State University Press, Ames.

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

Lamkey, K.R., A.R. Hallauer, and A.L. Kahler. 1987. Allelic differences at enzyme loci and hybrid performance in maize. J. Hered. 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. Crp Sci. 29:1067-1071.

Lefort-Buson, M., Y. Datee, and B. Guillot-Lemoine. 1987. Heterosis and genetic distance in rapeseed (*Brassica napus* L.): use of kinship coefficient. Genome 29:11-18.

Sneller, C.. 1991. Estimation of the genetic structure of an elite soybean population and it's application to a soybean breeding program. Ph D dissertation, Michigan State University.

Sokal, R.R.. 1965. Statistical methods in systematics. Bio. Rev. Camb. Phil. Soc. 40:337-341.

Sokal,R.R., and P.H.A. Sneath. 1963. *Principles of Numerical Taxonomy*. Freeman, San Francisco.