Winter 1992

Genetic and morphological variation and differentiation of South Korean natural populations of wild soybean, Glycine soja Sieb and Zucc

Hongrun Yu
University of New Hampshire, Durham

Order Number 9307361

Genetic and morphological variation and differentiation of South Korean natural populations of wild soybean, *Glycine soja* Sieb. and Zucc.

Yu, Hongrun, Ph.D.

University of New Hampshire, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

GENETIC AND MORPHOLOGICAL VARIATION AND DIFFERENTIATION OF SOUTH KOREAN NATURAL POPULATIONS OF WILD SOYBEAN, GLYCINE SOJA SIEB. & ZUCC.

BY

HONGRUN YU
B.S., Shenyang Agricultural College, China, 1981
M.S., Kansas State University, 1988

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Plant Science

December, 1992

This dissertation has been examined and approved.

Yun-Tzu Kiang

Dissertation director, Yun-Tzu Kiang,
Professor of Plant Biology and Genetics

J. Brent Loy, Professor of Plant
Biology and Genetics

Lincoln C. Peirce, Professor of Plant
Biology and Genetics

Anita S. Klein, Associate Professor
of Biochemistry and Molecular Biology
and Genetics

Robert T. Eckert

Robert T. Eckert, Associate Professor
of Natural Resources and Genetics

August 14, 1992

Date

ACKNOWLEDGEMENTS

I would like to thank my major advisor, Dr. Yun-Tze Kiang, for his advice and guidance in the last four years. Drs. Bob Eckert, Anita Klein, Brent Loy and Lincoln Peirce served on my supervisory committee. I am grateful to all of you for your suggestions and criticism on my research and on this dissertation. I will remember you as my mentors. I feel indebted to Dr. Peirce, who has corrected my numerous writing including the initial research proposal and the CURF grant applications. I wish him a wonderful retirement.

Bill Given and Bob Parker took care of my plants. Their hard work and prompt attendance of my work requests are acknowledged. Consultations with Petr Brym on use of the Biosys-1 computer program are appreciated.

I wish to thank my family for their support and patience during the studies. Mao Mao's desire to follow my footsteps was an encouragement for me.

This research was partly supported by the university CURF grants (3012 UFR CUG7 in 1989 and 1312 UAB CUJJ in 1990-1991).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

GENETIC AND MORPHOLOGICAL VARIATION AND DIFFERENTIATION OF SOUTH KOREAN NATURAL POPULATIONS OF WILD SOYBEAN, GLYCINE SOJA SIEB. & ZUCC.

by

Hongrun Yu
University of New Hampshire, December, 1992

Genetic variation is the basis of crop improvement. As genetic background narrows in the cultivated germplasm, genes need to be introduced from new sources. Glycine soja is a wild relative of the cultivated soybean, Glycine max (L.) Merr. It can be used in soybean breeding. Evaluation of wild soybean populations is not only necessary for use in cultivar improvement, will also provide information about origin, migration, evolution and natural selection of this species. Seeds were collected from six natural populations in South Korea to study the genetic variation and differentiation of wild soybean.

The study was divided into two parts: (a) lab assay for 17 isozymes and one protein involving 35 loci; and (b) two-year greenhouse experiment, during which data for morphological traits were recorded.

The average number of alleles per locus, 99% polymorphism and the expected heterozygosity in the total popula-

tion were 2.1, 77.1% and 0.215, respectively. Nei's gene differentiation (G_{ST}) was 0.383. The average Nei's genetic distance was 0.117.

Populations were not significantly different in mean CV (coefficients of variation) for both years. However, 27 of the 33 individual morphological traits examined in 1989 and 31 of the 39 in 1990 differed significantly among populations. The average among-population variation per trait per degree of freedom was 84.38% in 1989 and 83.11% in 1990.

Although there were no significant differences among populations in mean CV, those populations with high genetic variation also had high mean CV for morphological traits. There was no congruence between the isozyme and morphological data in terms of population relationships.

Multilocus association analysis provided significant insight into the genetic structure of these natural populations. The analysis indicated that within each population, there were two to three dominant multilocus genotypes. The multilocus genotypes were "biotypes" at the morphological level. These different types might be the basic genetic division of mosaic self-pollinated plant populations, and the basic units in natural selection and evolution. The analysis also suggested that there was no migration among these six populations in recent history.

The numbers of loci different between individuals were used as measures of genetic variation and genetic distance.

INTRODUCTION

GENERAL BACKGROUND INFORMATION

RATIONALE

Soybean (Glycine max [L.] Merr.) is one of the most important crops in modern agriculture. Soybean seeds contain nearly 40% protein and 20% oil, and are also rich in polyunsaturated fatty acids, which help reduce cholesterol levels in the human body. Thus, soybean is ideal for human consumption. As people become more aware of nutritional and healthy diet, soybean will play an increasing role in our daily life.

In the last half century, remarkable progress has been made in soybean breeding and production in the world. For example, in the northern U.S., cultivars released in 1940's and 1950's yielded 26% more than the selections made from the plant introductions in 1920's (Gai, 1985). The second wave of cultivars developed before and during 1970's gave another yield boost of 16%. Meanwhile, 20% of genetic gain in yield was made in the southern U.S. during the same period from 1940's through 1970's. Significant progress in soybean breeding was also made in China, Brazil, Argentina and other major soybean producing countries around the world. Soybean production in the U.S. increased 2.23

times from 1963 to 1983 (Smith and Huyser, 1987). In Brazil, it was increased 28.37 fold from 1965 to 1983. Argentina beginning from a nearly negligible level of soybean production joined the rank of major soybean producing countries in the same two decades. China, where soybean originated and has been cultivated for thousands of years, also saw a moderate increase of soybean production.

However, research in soybean genetics has lagged far behind. The situation is not comparable to other crops, such as corn, wheat and barley. By 1987, about 200 gene symbols had been assigned to soybean (Palmer and Kilen, 1987). Among these, there was much redundancy resulting from the same trait studied by different people and given different gene symbols. In 1990, a soybean genetic map consisting of 17 linkage groups was published (Palmer and Kiang, 1990). Most of these linkage groups consisted of only two to three loci. Later, it was found that linkage group 5 and linkage group 16 were linked (Kiang, 1990a). Thus, there were actually 16 linkage groups found in soybean. Some of these also may belong to the same linkage group. Soybean has 40 chromosomes and should have 20 linkage groups. At least four linkage groups remain to be found. Astonishingly, none of the present linkage groups has been associated with a specific chromosome. Only one locus outside the current linkage groups, Dia1, has been assigned to a chromosome (cited from Kiang and Chiang, 1987a).

The lack of progress in soybean genetics can be attributed to three factors. First, the nature of self-pollination and small size of soybean flowers hamper successful hybridization. Pollination occurs before the flower opens. Secondly, soybean chromosomes are numerous and very small. They are difficult to distinguish under a microscope. Lastly, there are not enough genetic markers in the uniform crop of soybean.

Genetic Markers.

One of the objectives of population genetic studies involving isozymes is to find new isozyme genetic markers. The work in our laboratory has demonstrated that this is quite a successful approach. By screening the cultivated soybean germplasm and wild soybean populations, our laboratory has found many isozyme variants (Bult, 1989; Chiang, 1985; Doong, 1986; and Gorman, 1983). Many isozyme loci have been assigned to them and 12 of these loci have been mapped in linkage groups by our laboratory (Palmer and Kiang, 1990).

Recently, RFLP (Restriction Fragment Length Polymorphism), ASP (Amplified Sequence Polymorphism obtained by PCR, or Polymerase Chain Reaction) and RAPD (Random Amplified Polymorphic DNA) have been employed as genetic markers in genetic mapping of qualitative genes and quantitative trait loci (QTL's), and in the genetic diversity studies

(Apuya et al., 1988; Keim et al., 1989; Keim et al., 1990; and Williams et al., 1990). A soybean RFLP map was also established (Keim et al., 1990). But, like the morphological genetic markers, the usefulness of isozyme genetic markers has not been diminished. Compared with RFLP amd RAPD methods, the isozyme technique "is relatively simple and rapid, and, as a consequence, it permits the screening of large samples" (Brown et al., 1990. pp. 98-115). Because of large samples, it has more statistical power, which is important in population genetic studies. From economic consideration, isozyme assays are cheaper than RFLP and RAPD assays. Therefore, isozymes will remain a powerful tool in genetic mapping and population genetic studies.

In addition to their usefulness in genetic mapping and population genetic studies, isozymes, as genetic markers, are also important in other applications. Isozymes can be used in parental line selection to maximize genetic differences in breeding populations (Kiang and Gorman, 1983), cultivar fingerprinting (Doong, 1986), hybrid seed identification, cross pollination studies (Chiang and Kiang, 1987a), and many other aspects of biological research.

Genetic Variation.

Three terms, "genetic variation", "genetic variability" and "genetic diversity" are used interchangeably in population genetic studies. They refer to the existence of genetically different identities within a population, within

a species or among closely related species. "Biodiversity", on the other hand, refers to the existence of different kinds of organisms within a geographic region.

According to Fisher's fundamental theorem of natural selection, "the increase in average fitness in one generation of natural selection in a population with nonoverlapping generations equals the additive genetic variance in fitness divided by the average fitness." (Hartl, 1988. pp. 245). Therefore, genetic variation is essential for the improvement in fitness of populations. This theorem explains why genetic diversity is important, not only for natural populations or wild species, but also for artificial populations or the cultivated species in which the improvement of yield potential and other traits is the goal of human selection.

Genetic variation is also important for population or organism stability. Those populations or organisms without genetic diversity will sooner or later be eliminated by natural selection in the face of changing environments or conditions. Those with genetic diversity will probably maintain the constant state by internal adjustment of their genetic composition. Therefore, plant breeders advocate diversifying commercial cultivars and broadening their genetic base so as to achieve stable production.

However, genetic diversity is low and genetic back-

ground is narrow in the cultivated germplasm. In 1972, the Committee on Genetic Vulnerability at the National Academy of Sciences in Washington D.C. issued a report pointing out that the maternal ancestors and their combined frequencies of occurrence in parentage of the U.S. northern and southern soybean cultivars were: Mandarin 51%, Illini 23%, Tokyo 11%, Dunfield 8%, Mukden 4% and Roanoke 4%. Four of these maternal parents (accounting for 86% of parentage) were introduced from Northeast China. Another parentage study found that the germplasm in all 158 U.S. and Canadian public cultivars of hybrid origin released in 1970's could be traced to as few as 50 plant introductions (Delanny et al., 1983). In the northern cultivars of North America, 50% of the germplasm was contributed by four introductions and 80% by ten introductions. A single introduction, "Mandarin", accounted for more than 30% of the genes. It was even worse in the south, with 50% of the genes contributed by two introductions and 70% by seven introductions. Specht and Williams (1984) reported that 88% of the germplasm in the 136 US and Canadian soybean cultivars of Maturity Groups 00 to IV released between 1939 and 1981 was contributed from 12 ancestors. Revealed in a survey conducted by Duvick (1984), 56% of the US soybean acreage was dedicated to six cultivars in 1970 and nine cultivars in 1980. Also based on other crops, Duvick (1984) concluded that an improvement was made in diversifying genetic background of major crops in this

period, thanks to the effort of plant breeders. Progress in diversifying commercial soybean cultivars was also indicated by Delanny et al. (1983).

Narrow genetic background increases vulnerability to epidemic pests and diseases and limits the potential of further yield increase of the current cultivars. There must be no complacency. Broadening genetic background in the cultivated soybean is still an important task.

Wild soybean, Glycine soja Sieb. & Zucc. is believed to be the progenitor of the cultigen, G. max (Hymowitz and Singh, 1987). These two species have the same chromosome number, can be intercrossed freely and produce fertile offspring. The two species together form the soybean gene pool (Kiang et al., 1987). Therefore, wild soybean is the best candidate for broadening the genetic background of the cultivated soybean. Some of the characteristics in wild soybean, such as pest and disease resistance, high protein content, and small seed size for special uses, can be and have been, in some instances, transferred to the cultivated soybean (Gai, 1985 and LeRoy et al., 1991a and 1991b).

However, wild soybean is not widely used in the breeding programs. This situation can be attributed to several factors. First, there are many characteristics associated with G. soja undesirable for cultivation, such as vining of stem, non-abscission of leaves, shattering of pods, dark

color of seed coat and small size of seeds. Therefore, it takes more time to recover the plants that resemble the cultivated soybean in interspecific crosses. Based on their experimental results, Ertl and Fehr (1985) concluded that the introgression of G. soja germplasm into the cultivated soybean was not an effective method for increasing yield potential of soybean cultivars. Above all, soybean breeders do not feel that the cultivated soybean germplasm is threatened by genetic disasters at the present time. Therefore, it seems to be a realistic approach that wild soybean be used to form populations in breeding programs, such as recurrent selections. Evaluation of genetic variation in wild soybean populations, which is another objective of the present study, can enhance our understanding of G. soja and is an indispensable process in its use in breeding programs.

Genetic variation in the cultivated and wild soybean has been evaluated with RFLPs of mitochondrial DNA (mtDNA) (Sisson et al., 1978), chloroplast DNA (cpDNA) (Shoemaker et al., 1986), random genomic DNA (Keim et al., 1989), the tandemly repeated multigene family encoding the 18S and 25S ribosomal RNA (18S and 25S rDNA) (Doyle and Beachy, 1985) and the 5S ribosomal RNA genes (5S rDNA) (Doyle, 1988), with isozymes (Bult, 1989; Chiang, 1985; Doong, 1986; Gorman, 1983; Hu and Wang, 1985; Hymowitz and Kaizuma, 1979 and 1981; Kiang and Chiang, 1990; Kiang and Gorman, 1983; and Kiang et al., 1987) and other biochemical, such as flavo-

noids, and morphological genetic markers (Broich and Palmer, 1981), and with morphological traits (Broich and Palmer, 1980; Bult, 1989; Chiang, 1985; Kiang and Chiang, 1989; and Kiang and Chiang, 1990). While most of these studies used seed accessions available in the USDA Soybean Germplasm Collection, a few of them employed natural populations of wild soybean (Bult, 1989; Chiang, 1985; and Hu and Wang, 1985).

To summarize these studies, a few observations can be made. First, there is little variation in mtDNA, cpDNA, 18S and 25S rDNA, and 5S rDNA revealed by RFLP analysis of the soybean gene pool, especially for the maternal parents of the current commercial cultivars of North America. This supports the common belief that the soybean genetic background is narrow. On the other hand, considerable genetic variation can be found with random genomic RFLP and isozymes. The number of alleles per locus, polymorphism and the expected heterozygosity are comparable to those of other crops. Second, G. soja has consistently shown higher genetic variation than the cultivated G. max. The hypothesis that G. soja is the progenitor of G. max seems to be supported by this result. The reduction in genetic variation in G. max is proposed to be caused by founder effect and the artificial selection pressure applied on the cultigen. Third, it appears that the Korean peninsula and southern Japan are the

centers of genetic variation for both G. max and G. soja. Fourth, there is low congruence between genetic information and morphological information in terms of relationships among populations, at least in the wild soybean that has been studied.

Origin, Dissemination and Evolution.

Although a center of diversity is not the same as a center of origin for a plant species, it is still useful to classify the variation within the species and to plot the geographic distribution of variation (Harlan, 1971). Coupled with historical, archeological and linguistic evidence, the geographic distribution of genetic variation can provide significant information about origin, dissemination and evolution of a plant species. The eastern half of northern China, i.e. the winter wheat-kaoliang region, was postulated to be the region where soybean began to be domesticated around 11th century B.C. during the Chou Dynasty (Hymowitz, 1970 and Hymowitz and Newell, 1981). Hymowitz (1970) also considered the eastern half of northern China to be the primary gene center and Northeast China to be a secondary center.

Hymowitz and Kaizuma (1979 and 1981) studied 477 soybean cultivars from Japan and 1603 soybean accessions from 15 other Asian countries or regions in the USDA Soybean Germplasm Collection for geographic distribution of Ti and Sp1 (same as Am3) alleles. For Japanese cultivars, the Sp1

locus had no clear geographic pattern of distribution, but the less frequent b allele at the Ti locus had higher frequencies in the Southern Kyushu Short-season Crop Region and the least frequent c allele carried by six cultivars was from the Tohoku District. Among the 15 countries or regions, South Korea had the highest frequencies of the less frequent alleles at Ti and Sp1 loci, Ti-b and Sp1-a, for the 1603 soybean accessions examined. A null allele at Ti locus was found in two South Korean accessions. Based on the geographic differences in the frequencies of Ti and Sp1 alleles and other evidence, the soybean grown in Asia was divided into seven soybean germplasm pools. The authors also proposed seven possible paths of dissemination of soybean from the eastern half of northern China to the other parts of Asia. The studies clearly indicated that the Korean peninsula and southern Japan are the diversity centers for Ti and Sp1.

Germplasm Conservation.

Germplasm conservation is vital for our future success in plant breeding programs and production as a whole. Population genetic studies, such as the present one, also serve the purpose of germplasm conservation. If unique genotypic variants or unique types of plants are found, they can be sent directly to the USDA Soybean Germplasm Collection and conserved. The T (Genotype Collection) and PI (Plant Intro-

duction) strains, which are usually more morphologically diverse than cultivars, also tend to have higher genetic variation by isozyme analysis (Kiang et al., 1987. Table 2). Those plants possessing rare biochemically alleles may also have higher probability of being agronomically desirable for use in cultivar improvement. In addition, geographic mapping of genetic variation may provide significant information about where to go for collecting expeditions and where to protect in the face of diminishing natural habitats. Perhaps, it may become necessary in the future to establish wild soybean reserves in the areas of gene centers.

OBJECTIVES

This study was designed:

1) to find new isozyme variants in the South Korean natural populations of wild soybean and to study the inheritance of the new isozyme variants and linkage relations of isozyme loci;

2) to examine the organization of genetic variation (synonym of isozyme variation) and differentiation in the South Korean natural populations of wild soybean;

3) to examine the organization of morphological variation and differentiation in the South Korean natural populations of wild soybean;

4) to determine the relationship between the genetic information obtained from isozymes and the information about

morphological traits in the South Korean natural populations of wild soybean.

SEED MATERIALS

Seeds of six wild soybean populations were collected in early October, 1986 from South Korea by Dr. Y.T. Kiang. The populations were located basically along the 127°E latitudinal line. The geographic location for each population is given in Fig. 1 and Table 1. The estimated geographic distances among populations are given in Table 2. The distance from the most northern population A to the most southern population F was about 194 kilometers. The average distance between two neighboring populations was 39 kilometers.

Eighteen to 41 natural plants were sampled for each population. The distance between plants was at least 3 meters in each population. Nine to 90 seeds were collected from each natural plant.

APPROACH

This study was divided into two parts: (a) laboratory assay of the isozyme genotypes; and (b) greenhouse experiments for morphological traits.

Electrophoresis on horizontal slab gels made of various concentrations of acrylamide and starch was used to determine genotypes of the natural plants. For each natural plant, two sources of seeds were used: five original seeds

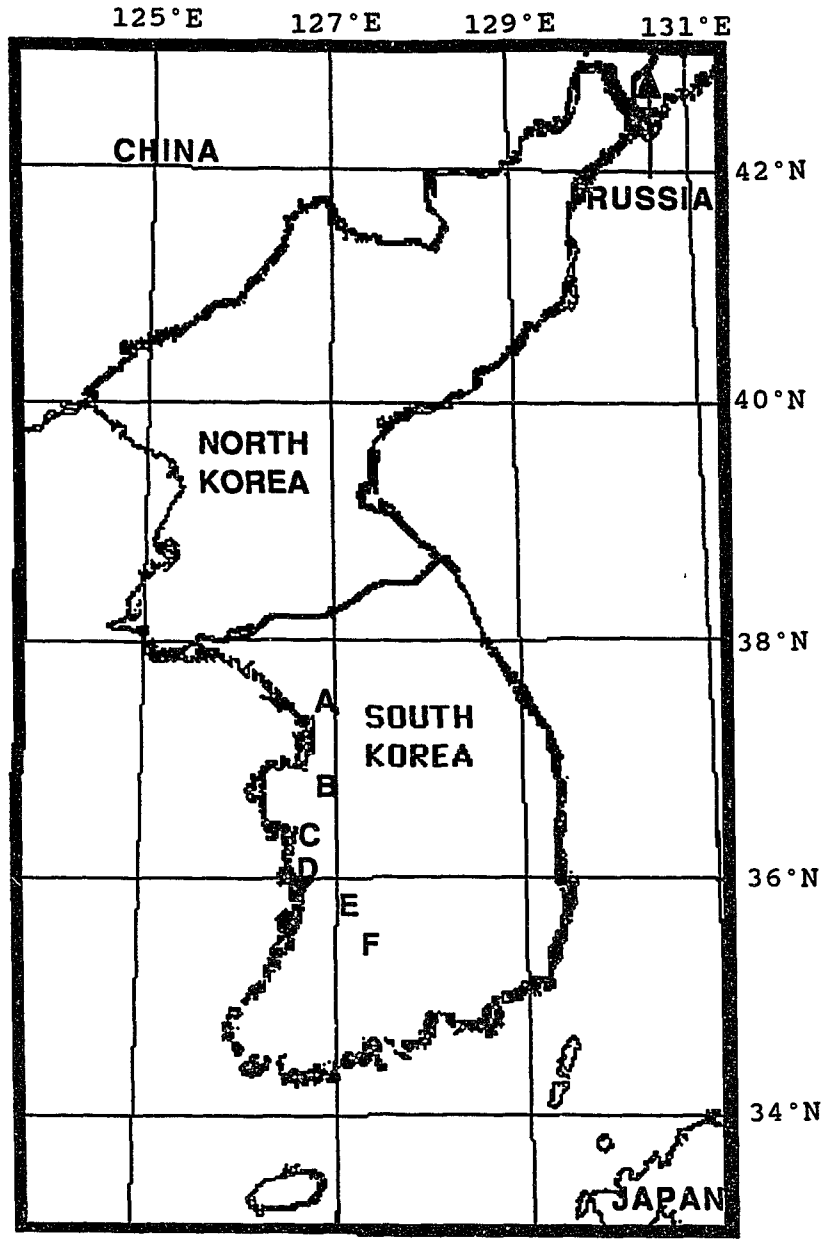


Fig. 1. Geographic locations of six South Korean natural populations of wild soybean.

Table 1. Geographic locations of six South Korean natural populations of wild soybean.

Population	Number of plants	Location	Latitude	Longitude
A	27	Wang Shium Ri, Bong Dam Myeon, Gyeon Gi Do	37°14'N	126°56'E
B	30	Gook-Kyeo River, Yeum Chi Myeon, A San Gun, Chung Ch'ong Nam Do	36°51'N	126°56'E
C	30	Worl Gae River, Dae Gyo Ri, Hong-Sun Gup, Chung Ch'ong Nam Do	36°34'N	126°41'E
D	18	Chang Am Ri, Jusam Myeon, Bo Lung Gun, Chung Ch'ong Nam Do	36°11'N	126°34'E
E	41	Saeg Chang River, Nam Gu Dong, Chonju City, Cholla Buk Do	35°49'N	127°07'E
F	26	Osu Ri, Cholla Buk Do	35°32'N	127°20'E

Table 2. Estimated geographic distances (km) among the six populations.

Population	A	B	C	D	E	F
B	44					
C	77	37				
D	119	77	44			
E	157	114	92	64		
F	118	149	129	99	36	
Mean	118	84	76	81	93	121

and six seeds from one greenhouse-grown plant. Thus, at least 11 seeds representing six original seeds were assayed. The genotypes of five original seeds and six greenhouse-harvested seeds must have matched for each natural plant across all isozyme loci, except for heterozygotes. More seeds were assayed for any discrepancy until the genotypes of each natural plant at all the isozyme loci could be assured. This method greatly reduced errors during the process of data collection. It also allowed the detection of heterozygotes, not only in the 172 natural plants, but also in the 172 natural seeds that gave rise to the 172 greenhouse-grown plants. Six seeds would determine genotypes of the plant at all isozyme loci.

Seventeen enzymes and one protein were examined. They were Aconitase (Aco), Alcohol dehydrogenase (Adh), Beta-Amylase (Am3), Acid phosphatase (Ap), Diaphorase (Dia), Endopeptidase (Enp), Esterase (Est), Urease (Eu), Fluorescent esterase (Fle), Glutamate oxaloacetic transaminase (Got), NADP-active isocitrate dehydrogenase (Idh), Leucine aminopeptidase (Lap), Mannose-6-phosphate isomerase (Mpi), 6-phosphogluconate dehydrogenase (Pgd), Phosphoglucose isomerase (Pgi), Phosphoglucomutase (Pgm), Shikimate dehydrogenase (Sdh), and Kunitz trypsin inhibitor (Ti).

For the morphological trait part of this study, because the plants were entangled in their natural habitats, they could not be studied under wild conditions. Thus, a "common

garden" approach was employed. The experiments had to be conducted in the greenhouse due to the short growing season in New Hampshire. Light was not provided so that conditions in the greenhouse would be as close to field conditions as possible. Heat was provided in the late growing season when the outside temperature was low. Because of limited greenhouse space, only about 120 plants could be grown each year. Therefore, 20 single natural plant seed sources were taken at random from each of the six populations, and one seed for each seed source was sown in the first year. In the second year, 10 seeds representing 10 random natural plants from each population were sown, along with one seed from each of the remaining 52 natural plants that were not represented in the first year. Some natural plants were represented once and the others twice in the two-year experiments. The planting dates were in mid-May, similar to the field planting time. The detailed growing conditions and the morphological traits examined will be discussed in the materials and methods section of Chapter III.

CHAPTER I

INHERITANCE AND LINKAGES OF ISOZYME LOCI

INTRODUCTION

Genetic research in soybean has lagged behind other major crops, such as corn. By 1990, only 16 linkage groups had been found out of 20 (2n=40 in soybean) (Palmer and Kiang, 1990). Most of them consisted of only a few loci. Later, it was found that the linkage groups 5 and 16 were linked (Kiang, 1990a). It is also likely that other linkage groups are on one chromosome. Astonishingly, the chromosomal placement has not been determined for any of the present linkage groups. Only one locus, Dia1, outside the current linkage groups has been associated with a chromosome (cited from Kiang and Chiang, 1987a).

The lack of progress in soybean genetic research can be attributed to several factors. First, the flowers of soybean (including the cultivated, Glycine max [L.] Merr. and wild, Glycine soja Sieb. & Zucc.) are small, especially for wild soybean. Pollination has already occurred before the flower opens. Hybridization thus is more difficult than in other crops. Cytologically, soybean chromosomes are small, and difficult to manipulate. Lastly, there are not enough genetic markers for the morphologically quite uniform soybean.

Isozyme loci have proven to be effective genetic markers. Nearly half of the mapped loci in soybean are isozyme loci (Palmer and Kiang, 1990). Isozyme loci are also important genetic markers for studies of population genetics, ecology and plant breeding.

Seeds of six populations of wild soybean composed of 172 natural plants were collected from South Korea and screened for isozyme variants. This chapter reports the inheritance of new isozyme variants and linkage relationship among most gene pairs of 24 isozyme loci.

MATERIALS AND METHODS

Seed Materials.

Dry seeds were used in electrophoresis. Seeds of the 172 natural plants of the six populations, which were described in the Introduction chapter, were screened for new electrophoretic variants. New variants of wild soybean were used as parents in making crosses. One variety of the cultivated soybean, AV68, was also used as a parent. AV68 originally was from the Asian Vegetable Research and Development Center in Taiwan. Hybridization was carried out in the greenhouse. Plants were grown in pots placed on greenhouse benches. A mixture of steam-sterilized soil and Promix (1:1, v/v) was used as the growth medium. For wild soybean, a bamboo stake for training was placed in each pot to maintain separation of the plants. For regular growing seasons, i.e. from May to November, no supplemental light was provided. In the late growing season, when the outside temperature was low, the pipe heating system provided the temperature of 30°C during the day and 25°C at night. Plants were also grown in other seasons for making crosses and advancing generations. For other growing seasons, natural light was supplemented by 10 hours of incandescent light. Soybean is a short-day species. Wild soybean and some cultivars are sensitive to the length of light. In order for wild soybean

to flower in winter months, the length of artificial light was shortened one hour every 10-15 days when the plants reached optimum sizes (depending on how many seeds were desired). This flower induction method was quite effective.

Electrophoresis.

Methods of electrophoresis were basically the same as those described by Bult (1989), Bult et al. (1989) and Chiang (1985) with slight modifications. The recipes of the gels and staining solutions are given in Appendix I along with the buffers, stock solutions and preparations used in electrophoresis.

1) Enzymes.

The 17 enzymes and one protein assayed in this study are listed in Table 3.

2) Gels.

The gels were made from either starch or acrylamide or both. There were 5 different types of gels in terms of concentrations of starch and acrylamide: A. 12.5% (w/v) starch; B. 7% (w/v) acrylamide; C. 9% (w/v) acrylamide; D. 7% (w/v) acrylamide + 2% (w/v) starch; and E. 6% (w/v) acrylamide + 4% (w/v) starch. For gels containing acrylamide, 5% (w/w) of this gelling agent was N,N'-methylene-bis-acrylamide (Bis) and the other 95% was ordinary acrylamide. Ammonium persulfate (APS) and N,N,N',N'-tetramethylethylenediamine (TEMED) were used as polymerization cata-

Table 3. Enzymes and protein examined.

Symbol	Enzyme (type of enzyme)	EC number
Aco	Aconitase (lyase),	EC 4.2.1.3
Adh	Alcohol dehydrogenase (oxioreductase),	EC 1.1.1.1
Am	Beta-Amylase (hydrolase),	EC 3.2.1.2
Ap	Acid phosphatase (hydrolase),	EC 3.1.3.2
Dia	Diaphorase (oxioreductase),	EC 1.6.2.2
Enp	Endopeptidase (hydrolase),	EC 3.4.?.?
Est	Esterase (hydrolase),	EC 3.1.1.1
Eu	Urease(hydrolase),	EC 3.5.1.5
Fle	Fluorescent esterase (hydrolase),	EC 3.1.1.2
Got	Glutamate oxaloacetic transaminase (transferase),	EC 2.6.1.1
Idh	NADP-active isocitrate dehydrogenase (oxioreductase),	EC 1.1.1.42
Lap	Leucine aminopeptidase (hydrolase),	EC 3.4.11.1
Mpi	Mannose-6-phosphate isomerase (isomerase),	EC 5.3.1.8
Pgd	6-phosphogluconate dehydrogenase (oxioreductase),	EC 1.1.1.44
Pgi	Phosphoglucose isomerase (isomerase),	EC 5.3.1.9
Pgm	Phosphoglucomutase (transferase),	EC 2.7.5.1
Sdh	Shikimate dehydrogenase (oxioreductase),	EC 1.1.1.25
Ti	Kunitz trypsin inhibitor (protein)	

lysts. Their concentrations were 0.1% (w/v) and 0.2% (v/v), respectively. The gel buffer was 0.005 M L-histidine (pH 7.0).

The gels were made in gel molds. Both single layer and multiple layer gels were used. The gel molds were made from PVC (polyvinylchlorine) board. The inner measurements of gel molds were 18.0 cm long and 15.5 cm wide, and 0.3 cm deep for single layer, 0.6 cm deep for 2 layer, 0.9 cm deep for three layer, 1.2 cm deep for four layer molds. The capacities of single, two, three and four layer gel molds were 150, 210, 280 and 350 ml, respectively. The cutting board used to slice gels was 18.5 cm long, 17.5 cm wide and 0.2 cm deep.

3) Sample Preparation.

A small piece of cotyledon was cut off each seed. The seed chips were placed in small wells on a polyvinylchlorine board, one chip per well. Then, two to three drops of 0.005 M L-histidine buffer were added into each well. The seeds were soaked for 24 to 36 hours at room temperature before they were ground to a paste with smooth and round-ended glass rods. A small piece (1 X 1 cm) of lens paper was placed on the homogenate as a filter. Wicks made from bibulous paper were put on top of the filter paper to take up the filtrate. These wicks were 0.3-0.4 cm wide and had various lengths corresponding to the thickness of the gel used. Then, 30 slots, 0.3-0.4 cm wide, were made on the gel

with a screw driver-like thin metal stick. These slots were parallel to and 3 cm from one of the longer sides of the gel mold. The wet wicks were inserted into these slots with forceps.

4) Running.

The gel was then placed on a 7 cm thick block of ice packs held in a large plastic tray. These ice packs were made with three to four plastic bags filled with reusable blue ice. The gel was positioned on the ice packs so that the side with wicks was connected to the cathode of a power supply. This was because nearly all enzymes carry negative charges and move from cathode to anode in the electromagnetic field of the gel, except for Est, which showed a cathodal band. The gel was connected to the power supply by two cellulose sponge pads and two trays of electrode buffer. The sponge pads covered 1 cm gel surface on both sides of the gel. The other halves of the sponges were immersed in about 250 ml 0.065 M tris-citrate electrode buffer (pH 7.0) contained in each of two plastic trays on both sides of the ice packs. There was a platinum wire serving as an electrode at bottom of the buffer trays. The electrode wires were connected to the power supply through small plugs at one end of the buffer tray. The DC power source was an ISCO model 493 electrophoresis apparatus. Methylene blue (1%) was dipped into the gel on the anodal side to serve as a dye marker for calculating RF values of isozyme bands. Then,

the gel was covered with a piece of plastic wrap. Thirty minutes later, wicks were removed with forceps. A PVC board was put on top of the gel, and two or three ice packs covered the PVC board to keep the gel at about 4°C.

For gels used for staining only one enzyme, the gel type, voltage and running time are given in Table 4. To run more enzymes on one gel, multiple layer gels were usually used. The gel types specified in the Table 4 could be changed. For example, one single layer C gel was cut into two pieces and used for both Am and Ti. The top layer was stained for Am, and the bottom for Ti. Usually, the top layer was discarded. Using both top and bottom layers for Am and Ti worked quite well and saved materials. Many combinations of enzymes were tried on one gel. Table 5 gives one example of combining enzymes so as to be run on multiple layer gels for the 17 enzymes and one protein, which seemed to work well. This method was also suitable for overnight running.

5) Staining.

After the gels were run, they were sliced with a bow-shaped cheese cutter with a guitar string in it. Each slice was put in a plastic tray, which was slightly larger than the gel. Staining solution was then poured over the gel. The gel was incubated either at 37°C in a dark chamber or at room temperature, being covered with a PVC board.

Table 4. Gel type, voltage and running time required for each enzyme run on single layer gels.

Enzyme	Gel type	Voltage	Running time
Aco	A	150	8–10 hr
Adh	D + 30 mg NAD	200	8–10 hr
Am	B	200	5–10 hr
Ap	D	200	5–10 hr
Dia	D	200	8–10 hr
Enp	D	200	8–12 hr
Est	E	200	5–8 hr
Eu	D	200	8–12 hr
Fle	D	200	8–12 hr
Got	D	200	8–12 hr
Idh	D	200	8–12 hr
Lap	D	200	8–12 hr
Mpi	D	200	8–12 hr
Pgd	E	175	10–14 hr
Pgi	D	200	8–12 hr
Pgm	D	200	8–12 hr
Sdh	E + 15 mg NADP	200	12–14 hr
Ti	C	200	5–7 hr

Table 5. Gel type, voltage and running time required for multiple layer gels[a].

Gel type	Enzymes	Voltage	Running time
Single layer C	Am (top layer), Ti (bottom layer)	110	8–11 hr
Two-layer A	Aco (middle layer)	150	8–11 hr
Single layer D + 30 mg NAD	Lap (top layer), Adh (bottom layer)	200	9–12 hr
Four-layer D	Ap (top layer), Dia, Pgm, Mpi, Pgi (bottom layer)	200	10–14 hr
Four-layer D	Enp (top layer), Idh, Sdh[b], Got, Fle (bottom layer)	200	12–16 hr
Single layer E	Eu (top layer), Pgd (bottom layer)	180	9–12 hr
Single layer E	Est (bottom layer, 2 lane loading)	160	8–11 hr

[a] A single layer gel was cut into two pieces, a two-layer gel into three, a three-layer gel into four, etc.

[b] Although Sdh required E gel with NADP, Sdh bands could be read on D gels.

6) Remarks.

Some of the enzymes were relatively easy to assay. They included Am, Ap, Enp, Est, Eu, Got, Pgi, Pgm, Sdh, and Ti. Aco was run under 150 volts without ice packs to cover the gels. Running at higher voltage caused difficulty in slicing the gel, double or shadow bands at band 3 and poor contrast when the gel was stained. The starch gels for Aco were freshly made the previous day and refrigerated overnight before use. The optimum running temperature for Adh was higher than 4°C. Otherwise, the bands were light. Dia, Fle, Idh, Mpi, and Pgm band 1, especially Mpi, were sensitive to temperature, and, in order to get good results, they were run at the room temperature lower than 60°F, plus ice packs on the gel to achieve about 4°C temperature around the gel. For better resolution of bands 3 and 4 of Idh, seeds were soaked in the seed buffer at a higher-than-room-temperature, such as 35°C in an incubation chamber. To achieve a good assay on the band of Pgi2 locus, the gel was incubated in a refrigerator for one hour and at room temperature for another. In the past, we were unable to get good results for Pgd. Incidentally, I found that the gels kept in a refrigerator for at least two weeks gave better separation of the bands. Lap was another enzyme difficult to assay sometimes. Using the gels with NAD seemed to give improved results.

Data Analysis.

The contingency Chi-square was used to test significant associations of two loci in dihybrid segregation populations. The maximum likelihood method was used to calculate recombination frequencies (Allard, 1956). Actual computation of the contingency Chi-square and the recombination frequencies was carried out with a customized version of Linkage-1, a Pascal computer program provided by Dr. Karl Suiter at Duke University (Suiter et al., 1983) except crosses 7 and 8 (Table 6), which were analyzed with a hand calculator.

RESULTS

Among the 172 natural plants, isozyme mobility variants were found for alcohol dehydrogenase (Adh), leucine aminopeptidase (Lap), mannose-6-phosphate isomerase (Mpi), phosphoglucomutase (Pgm) and shikimate dehydrogenase (Sdh). A rare zymogram type was also observed for diaphorase (Dia).

Most soybean varieties have five well-resolved bands for alcohol dehydrogenase (Adh) (Bult, 1989; Chiang, 1985; Gorman and Kiang, 1978; and Kiang and Gorman, 1983). Bands 1 and 3 are missing in some varieties. It is believed that bands 1 and 5 are homodimers, whereas band 3 is their heterodimeric product, because the mobility of band 3 is between bands 1 and 5. When band 1 is missing, so is band 3. Band 2 is controlled by a separate locus. The Soybean Genetics Committee has approved two gene symbols to explain the inheritance of bands 1 and 2 (Palmer and Kilen, 1987 and Palmer et al., 1987). Adh1 controls the first band. The absence of homodimeric band 1 and its heterodimeric band 3 is caused by the recessive or null adh1 allele. The absence of the second band is controlled by the recessive or null adh2 allele at Adh2 locus (called Adh3 by Bult, 1989 and Chiang, 1985).

Kiang and Chiang (1987a) found that Adh1 was linked to W1 (a flower color gene) in linkage group 8. Later, the

position of Adh1 in linkage group 8 was determined (Kiang, 1990b).

In this study, it was found that Adh band 4 was missing in seeds of several natural plants of the South Korean populations (Fig. 2). Reciprocal crosses between plants grown from seeds of KA12 without band 4 and KC1 with band 4 were made (crosses 1 and 2, Table 6). The numbers of plants with and without band 4 were not significantly different from a 3:1 ratio in F_2 populations of both crosses (Table 8). F_3 progeny testing of eight F_2 plants with band 4 indicated the homozygotes and heterozygotes were in a 1:2 ratio (Table 9). The F_3 plants derived from F_2 heterozygotes continued to segregate in the 3:1 ratio for the presence and absence of band 4 (Table 9). Those F_2 plants without band 4 did not segregate in F_3. Thus, band 4 is controlled by one gene. Adh3 and adh3 were assigned for the presence and absence of band 4.

Linkage tests showed that Adh3 segregated independently of Ap, Dia1, Mpi, Pgm1 and Pgm2 (Table 10).

Another enzyme studied was mannose-6-phosphate isomerase (Mpi). There is only one banding zone for this enzyme with two bands. Four mobility variants have been found, which are controlled by one locus, Mpi (Chiang and Kiang, 1988). Alleles Mpi-a, Mpi-b, Mpi-c and Mpi-d are responsible for the slowest, slow, fast and fastest mobility variants, respectively. Another allele, mpi, is a null

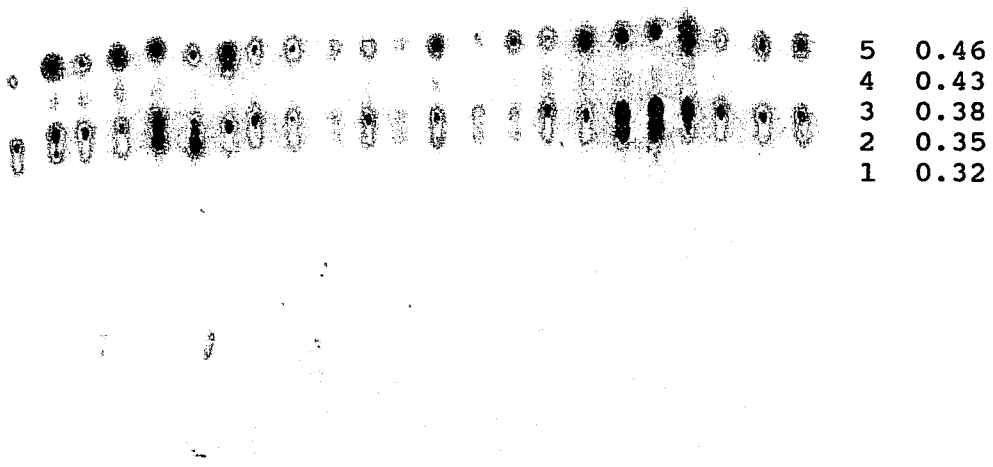


Fig. 2. Zymograms of soybean Adh.
Lanes 1 to 6, 11 to 19: KC1, Adh3/Adh3.
Lanes 7 to 10, 20 to 23: KA12, adh3/adh3.

Table 6. Crosses used in the inheritance and linkage studies[a].

Cross number	Female parent	Male parent
1	KA12	KC1
2	KC1	KA12
3	KA5	KB20
4	KB20	KA5
5	KA3	KD14
6	KD14	KA3
7	AV68	KC13
8	KC13	AV68
9	AV68	KE29

[a] All the varieties used were G. soja except AV68, which was G. max.

Table 7. Genotypes of the crossing parents at the loci studied[a].

	KA3	KA5	KA12	KB20	KC1	KC13	KD14	KE29	AV68[b]
Aco1	bb	bb	bb	bb	bb	bb	aa	bb	bb
Aco2	bb	bb	aa	bb	bb	bb	bb	bb	aa
Aco3	aa	aa	aa	aa	aa	aa	bb	aa	aa
Aco5	aa	aa	aa	aa	aa	aa	aa	--	aa
Adh3	++	++	--	--	++	++	++	++	++
Ap	aa	cc	aa	aa	cc	cc	cc	cc	bb
Dia1	--	++	--	--	++	--	++	--	++
Dia2	aa	aa	bb	aa	bb	bb	bb	aa	bb
Dia3	aa	aa	bb	bb	bb	bb	bb	aa	bb
Enp	bb	bb	bb	bb	bb	aa	bb	bb	bb
Est1	bb	bb	bb	bb	bb	bb	bb	aa	bb
Eu	aa	aa	aa	aa	aa	aa	aa	aa	bb
Fle	--	--	--	--	--	--	--	--	++
Idh1	bb	bb	bb	bb	bb	bb	aa	bb	aa
Idh2	aa	bb	aa	aa	aa	aa	aa	aa	bb
Lap1	bb	bb	bb	bb	bb	aa	bb	bb	bb
Mpi	bb	bb	bb	bb	ee	cc	aa	bb	cc
Pgd1	bb	bb	bb	aa	bb	bb	bb	bb	aa
Pgd2	aa	bb	aa	aa	aa	aa	bb	bb	aa
Pgi1	aa	aa	bb	--	bb	bb	bb	bb	bb
Pgm1	aa	aa	bb	aa	aa	aa	aa	aa	aa
Pgm2	bb	bb	bb	bb	cc	dd	bb	cc	bb
Sdh	aa	aa	aa	aa	aa	aa	bb	aa	aa
Ti	aa	bb	aa	aa	aa	bb	aa	aa	aa

[a] + and - denote dominant and recessive (null) alleles, respectively.

[b] AV68 was G. max. All others were G. soja.

Table 8. F_2 allele segregation at individual loci for all the crosses used in the inheritance and linkage studies.

Locus	Cross[a]	Segregation ratio	Genotypes & observed frequencies[b]	n	x^2	df	p
Aco1	6	1:2:1	aa 124 ab 219 bb 113	456	1.24	2	0.538
Aco2	7 & 8	1:2:1	aa 69 ab 159 bb 72	300	1.14	2	>0.500
	9	1:2:1	aa 96 ab 170 bb 84	350	1.11	2	0.574
Aco3	6	1:2:1	aa 102 ab 169 bb 97	368	2.58	2	0.275
Aco5	9	3:1	aa & a- 285 -- 87	372	0.52	1	0.472
Adh3	1	3:1	++ & +- 233 -- 89	322	1.20	1	0.274
	2	3:1	++ & +- 293 -- 97	390	0.00	1	0.953
Ap	1	1:2:1	aa 52 ac 120 cc 60	232	0.83	2	0.661
	2	1:2:1	aa 85 ac 211 cc 93	389	3.13	2	0.209
	3	1:2:1	aa 84 ac 172 cc 80	336	0.29	2	0.867
	4	1:2:1	aa 84 ac 161 cc 84	329	0.15	2	0.928
	6	1:2:1	aa 110 ac 239 cc 116	465	0.52	2	0.772
	7 & 8	1:2:1	bb 74 bc 145 cc 81	300	0.66	2	>0.500
	9	1:2:1	bb 130 bc 273 cc 135	538	0.21	2	0.899
Dia1	1	1:2:1	++ 93 +- 147 -- 81	321	3.17	2	0.205
	2	1:2:1	++ 87 +- 201 -- 102	390	1.52	2	0.467
	3	1:2:1	++ 24 +- 66 -- 30	120	1.80	2	0.407
	4	1:2:1	++ 73 +- 176 -- 81	330	1.85	2	0.396
	6	1:2:1	++ 89 +- 202 -- 107	398	1.72	2	0.423
	7 & 8	1:2:1	++ 74 +- 156 -- 70	300	0.59	2	>0.500
	9	1:2:1	++ 136 +- 271 -- 131	538	0.12	2	0.941
Dia2	6	1:2:1	aa 84 ab 160 bb 70	314	1.36	2	0.506
	9	1:2:1	aa 127 ab 287 bb 124	538	2.44	2	0.295
Dia3	3	3:1	bb & ba 275 aa 90	365	0.02	1	0.880
	4	3:1	bb & ba 248 aa 82	330	0.00	1	0.949
	6	3:1	bb & ba 348 aa 86	434	6.22	1	0.013
	9	3:1	bb & ba 403 aa 135	538	0.00	1	0.960
Enp	7 & 8	1:2:1	aa 74 ab 157 bb 69	300	0.82	2	>0.500
Est1	9	1:2:1	aa 136 ab 283 bb 118	537	2.77	2	0.250
Eu	7 & 8	3:1	bb & ba 237 aa 63	300	2.56	1	>0.100
	9	3:1	bb & ba 401 aa 137	538	0.06	1	0.803
Fle	7 & 8	3:1	++ & +- 221 -- 79	300	0.28	1	>0.500
	9	3:1	++ & +- 406 -- 132	538	0.06	1	0.803
Idh1	6	1:2:1	aa 117 ab 231 bb 116	464	0.01	2	0.994
	7 & 8	1:2:1	aa 59 ab 161 bb 80	300	4.55	2	>0.100
	9	1:2:1	aa 138 ab 265 bb 105	508	5.24	2	0.073
Idh2	3	1:2:1	aa 91 ab 193 bb 88	372	0.58	2	0.750
	4	1:2:1	aa 82 ab 166 bb 82	330	0.01	2	0.994
	7 & 8	1:2:1	aa 74 ab 150 bb 76	300	0.03	2	>0.975
	9	1:2:1	aa 105 ab 278 bb 125	508	6.11	2	0.047
Lap1	7 & 8	1:2:1	aa 69 ab 163 bb 68	300	2.26	2	>0.250
Mpi	1	1:2:1	bb 80 be 154 ee 83	317	0.31	2	0.855
	2	1:2:1	bb 90 be 209 ee 88	387	2.50	2	0.286
	9	1:2:1	bb 112 bc 210 cc 96	418	1.23	2	0.539

Table 8. F2 allele segregation at individual loci for all the crosses used in the inheritance and linkage studies (continued).

Locus	Cross[a]	Segregation ratio	Genotypes & observed frequencies[b]					n	x^2	df	p
Pgd1	7 & 8	3:1	bb & ba	221	aa	79		300	0.28	1	>0.500
Pgd2	6	1:2:1	aa	99	ab	176	bb 77	352	2.75	2	0.253
Pgi1	3	3:1	aa & a-	196	--	56		252	1.04	1	0.309
	4	3:1	aa & a-	226	--	104		330	7.47	1	0.006
	6	1:2:1	aa	111	ab	230	bb 123	464	0.66	2	0.721
Pgm1	1	1:2:1	aa	13	ab	27	bb 10	50	0.68	2	0.712
	2	1:2:1	aa	82	ab	172	bb 84	338	0.13	2	0.937
Pgm2	1	1:2:1	bb	66	bc	156	cc 89	311	3.41	2	0.182
	2	1:2:1	bb	83	bc	188	cc 93	364	0.95	2	0.623
	7	1:2:1	bb	50	bd	83	dd 47	180	1.19	2	>0.500
	8	1:2:1	bb	63	bd	153	dd 84	300	3.06	2	>0.100
	9	1:2:1	bb	162	bc	256	cc 120	538	7.81	2	0.020
Sdh	5	1:2:1	aa	85	ab	175	bb 100	360	1.53	2	>0.250
	6	1:2:1	aa	126	ab	210	bb 125	461	3.65	2	0.161
Ti	3	1:2:1	aa	78	ab	183	bb 81	342	1.74	2	0.420
	4	1:2:1	aa	56	ab	113	bb 67	236	1.45	2	0.485

[a] Cross 7 and cross 8 were reciprocal crosses. Due to small sample sizes, they were combined except for Pgm2. The combined sample size was 300 for all the loci involved, consisting of 180 individuals of cross 7 and 120 of cross 8.

[b] + and - denote dominant and recessive (null) alleles, respectively.

Table 9. F_3 allele segregation of F_2 heterozygotes for the new alleles or loci found in this study.

Locus	Cross	Segregation ratio	Genotypes & observed frequencies[a]	n	x^2	df	p
Adh3	1	1:2[b]	++ 2 +- 2	4	0.51	1	>0.250
		3:1	++ & +- 13 -- 3	16	0.33	1	>0.500
	2	1:2[b]	++ 1 +- 3	4	0.12	1	>0.500
		3:1	++ & +- 27 -- 11	38	0.32	1	>0.500
Mpi	1	1:2:1	bb 22 be 36 ee 12	70	2.91	2	>0.100
	2	1:2:1	bb 9 be 16 ee 7	32	0.25	2	>0.750
Pgm2	7	1:2:1	bb 7 bd 17 dd 6	30	0.60	2	>0.500
	8	1:2:1	bb 9 bd 14 dd 7	30	0.40	2	>0.750
Sdh	6	1:2:1	aa 7 ab 15 bb 8	30	0.67	2	>0.500

[a] + and - denote dominant and recessive (null) alleles, respectively.

[b] F_3 progeny testing of F_2 plants with Adh band 4.

Table 10. F_2 linkage tests of 24 isozyme loci.

Gene pair	Cross[a]	Phase[b]	Segregation ratio[c]	n	x^2[d]	df	p
Aco1-Aco3	6		1	364	0.56	4	0.967
Aco1-Ap	6		1	455	1.15	4	0.886
Aco1-Dia1	6		1	390	2.18	4	0.703
Aco1-Dia2	6		1	304	4.21	4	0.378
Aco1-Idh1	6		1	455	0.50	4	0.974
Aco1-Pgd2	6		1	345	1.90	4	0.755
Aco1-Pgi1	6		1	455	0.26	4	0.992
Aco1-Sdh	6		1	453	2.99	4	0.559
Aco2-Ap	7 & 8		1	300	1.61	4	0.900-0.750
	9		1	350	1.24	4	0.871
Aco2-Dia1	7 & 8		1	300	3.97	4	0.500-0.250
	9		1	350	1.54	4	0.819
Aco2-Dia2	9		1	350	7.53	4	0.110
Aco2-Enp	7 & 8		1	300	0.07	4	1.000-0.995
Aco2-Est1	9		1	349	9.31	4	0.054
Aco2-Idh1	7 & 8		1	300	0.55	4	0.975-0.950
	9		1	350	8.87	4	0.064
Aco2-Idh2	7 & 8		1	300	4.53	4	0.500-0.250
	9		1	350	1.29	4	0.862
Aco2-Lap1	7 & 8		1	300	1.66	4	0.900-0.750
Aco2-Mpi	9		1	348	0.89	4	0.926
Aco2-Pgm2	7 & 8		1	300	4.10	4	0.500-0.250
	9		1	350	6.98	4	0.137
Aco3-Ap	6		1	367	4.89	4	0.299
Aco3-Dia1	6		1	361	7.02	4	0.134
Aco3-Dia2	6		1	250	8.72	4	0.068
Aco3-Idh1	6		1	367	5.34	4	0.254
Aco3-Pgd2	6		1	294	1.97	4	0.741
Aco3-Pgi1	6		1	367	3.39	4	0.495
Aco3-Sdh	6		1	365	3.04	4	0.552
Aco5-Aco2	9		2	335	1.27	2	0.531
Aco5-Ap	9		2	372	0.30	2	0.860
Aco5-Dia1	9		2	372	1.96	2	0.376
Aco5-Dia2	9		2	372	3.01	2	0.221
Aco5-Dia3	9	C	3	372	2.64	1	0.104
Aco5-Est1	9		2	372	0.95	2	0.622
Aco5-Eu	9	C	3	372	0.25	1	0.617
Aco5-Fle	9	C	3	372	0.19	1	0.665
Aco5-Idh1	9		2	372	1.23	2	0.542
Aco5-Idh2	9		2	372	1.70	2	0.427
Aco5-Mpi	9		2	342	3.34	2	0.188
Aco5-Pgm2	9		2	372	3.10	2	0.212
Adh3-Ap	1		2	232	2.27	2	0.322
	2		2	389	1.31	2	0.521
Adh3-Dia1	1		2	321	1.13	2	0.568

Table 10. F_2 linkage tests of 24 isozyme loci (continued).

Gene pair	Cross[a]	Phase[b]	Segregation ratio[c]	n	x^2[d]	df	p
Adh3-Dia1	2	2		390	2.45	2	0.294
Adh3-Mpi	1	2		317	0.42	2	0.812
	2	2		387	0.65	2	0.723
Adh3-Pgm1	1	2		50	1.43	2	0.490
	2	2		338	0.75	2	0.688
Adh3-Pgm2	1	2		311	0.31	2	0.858
	2	2		364	1.01	2	0.604
Ap-Dia1	1	1		232	2.13	4	0.712
	2	1		389	1.73	4	0.785
	3	1		120	4.32	4	0.364
	4	1		329	1.44	4	0.838
	6	1		397	5.84	4	0.212
	7 & 8	1		300	3.79	4	0.500-0.250
	9	1		538	9.39	4	0.052
Ap-Dia2	6	1		314	1.85	4	0.764
	9	1		538	9.45	4	0.051
Ap-Enp	7 & 8	1		300	0.63	4	0.975-0.950
Ap-Est1	9	1		537	1.59	4	0.810
Ap-Idh1	6	1		463	6.46	4	0.168
	7 & 8	1		300	0.50	4	0.975-0.950
	9	1		508	8.75	4	0.068
Ap-Idh2	3	1		336	3.31	4	0.507
	4	1		329	7.01	4	0.136
	7 & 8	1		300	0.59	4	0.975-0.950
	9	1		508	2.22	4	0.695
Ap-Lap1	7 & 8	1		300	102.81	4	0.005-0.000
Ap-Mpi	1	1		232	2.89	4	0.577
	2	1		387	3.82	4	0.431
	9	1		418	2.35	4	0.672
Ap-Pgd2	6	1		352	16.71	4	0.002
Ap-Pgi1	6	1		463	1.41	4	0.843
Ap-Pgm1	1	1		50	5.43	4	0.246
	2	1		337	5.22	4	0.265
Ap-Pgm2	1	1		221	4.66	4	0.324
	2	1		363	3.77	4	0.437
	7 & 8	1		300	6.28	4	0.250-0.100
	9	1		538	1.57	4	0.814
Ap-Sdh	6	1		460	3.63	4	0.459
Ap-Ti	3	1		306	283.31	4	0.000
	4	1		235	335.65	4	0.000
Dia1-Dia2	6	1		285	4.56	4	0.336
	9	1		538	5.90	4	0.207
Dia1-Enp	7 & 8	1		300	5.90	4	0.250-0.100
Dia1-Est1	9	1		537	5.15	4	0.272

Table 10. F_2 linkage tests of 24 isozyme loci (continued).

Gene pair	Cross[a]	Phase[b]	Segregation ratio[c]	n	x^2[d]	df	p
Dia1-Idh1	6		1	397	1.41	4	0.843
	7 & 8		1	300	9.11	4	0.100-0.050
	9		1	508	0.68	4	0.954
Dia1-Idh2	3		1	120	1.11	4	0.893
	4		1	330	[illegible].54	4	0.820
	7 & 8		1	300	8.91	4	0.100-0.050
	9		1	508	4.17	4	0.383
Dia1-Lap1	7 & 8		1	300	[illegible]72	4	0.250-0.100
Dia1-Mpi	1		1	316	[illegible]59	4	0.159
	2		1	387	6.32	4	0.176
	9		1	418	2.14	4	0.709
Dia1-Pgd2	6		1	292	[illegible]9	4	0.647
Dia1-Pgi1	6		1	396	[illegible]	4	0.062
Dia1-Pgm1	1		1	50	[illegible]8	4	0.545
	2		1	338	.95	4	0.567
Dia1-Pgm2	1		1	310	0.58	4	0.966
	2		1	364	0.71	4	0.950
	7 & 8		1	300	5.10	4	0.500-0.250
	9		1	538	9.92	4	0.042
Dia1-Sdh	6		1	394	2.87	4	0.580
Dia1-Ti	3		1	90	1.69	4	0.793
	4		1	236	1.94	4	0.747
Dia2-Est1	9		1	53[illegible]	0.89	4	0.926
Dia2-Idh1	6		1	31[illegible]	4.10	4	0.393
	9		1	50[illegible]	2.53	4	0.639
Dia2-Idh2	9		1	50[illegible]	0.71	4	0.950
Dia2-Mpi	9		1	413	7.35	4	0.118
Dia2-Pgd2	6		1	20[illegible]	2.74	4	0.602
Dia2-Pgi1	6		1	[illegible]	5.68	4	0.224
Dia2-Pgm2	9		1	[illegible]	6.79	4	0.147
Dia2-Sdh	6		1	31[illegible]	1.63	4	0.804
Dia3-Aco1	6		2	424	1.81	2	0.404
Dia3-Aco2	9		2	350	3.09	2	0.214
Dia3-Aco3	6		2	367	0.68	2	0.711
Dia3-Ap	3		2	331	3.59	2	0.167
	4		2	329	1.61	2	0.447
	6		2	433	1.01	2	0.605
	9		2	538	0.38	2	0.829
Dia3-Dia1	3		2	119	0.17	2	0.917
	4		2	330	7.10	2	0.029
	6		2	397	0.80	2	0.669
	9		2	538	1.56	2	0.458
Dia3-Dia2	6		2	312	1.50	2	0.472
	9		2	538	0.22	2	0.896

Table 10. F_2 linkage tests of 24 isozyme loci (continued).

Gene pair	Cross[a]	Phase[b]	Segregation ratio[c]	n	x^2[d]	df	p
Dia3-Est1	9		2	537	0.39	2	0.821
Dia3-Eu	9	C	3	538	0.02	1	0.887
Dia3-Fle	9	C	3	538	3.52	1	0.060
Dia3-Idh1	6		2	432	2.35	2	0.309
	9		2	508	0.48	2	0.788
Dia3-Idh2	3		2	365	4.42	2	0.110
	4		2	330	0.82	2	0.665
	9		2	508	1.26	2	0.534
Dia3-Mpi	9		2	418	0.77	2	0.679
Dia3-Pgd2	6		2	323	1.58	2	0.454
Dia3-Pgi1	3	R	3	245	0.07	1	0.787
	4	R	3	330	0.05	1	0.817
	6		2	432	1.41	2	0.494
Dia3-Pgm2	9		2	538	0.62	2	0.733
Dia3-Sdh	6		2	429	2.05	2	0.359
Dia3-Ti	3		2	335	8.16	2	0.017
	4		2	236	0.39	2	0.824
Enp-Idh1	7 & 8		1	300	3.91	4	0.500-0.250
Enp-Idh2	7 & 8		1	300	4.13	4	0.500-0.250
Enp-Lap1	7 & 8		1	300	5.80	4	0.250-0.100
Enp-Pgm2	7 & 8		1	300	3.73	4	0.500-0.250
Est1-Idh1	9		1	507	3.23	4	0.521
Est1-Idh2	9		1	507	4.83	4	0.304
Est1-Mpi	9		1	417	11.22	4	0.024
Est1-Pgm2	9		1	537	2.69	4	0.610
Eu-Aco2	7 & 8		2	300	1.46	2	0.500-0.250
	9		2	350	0.44	2	0.803
Eu-Ap	7 & 8		2	300	4.57	2	0.250-0.100
	9		2	538	0.54	2	0.765
Eu-Dia1	7 & 8		2	300	0.67	2	0.750-0.500
	9		2	538	0.14	2	0.932
Eu-Dia2	9		2	538	0.11	2	0.945
Eu-Enp	7 & 8		2	300	0.72	2	0.750-0.500
Eu-Est1	9		2	537	3.76	2	0.152
Eu-Fle	7 & 8	C	3	300	0.60	1	0.500-0.250
	9	C	3	538	0.69	1	0.406
Eu-Idh1	7 & 8		2	300	0.77	2	0.750-0.500
	9		2	508	1.41	2	0.494
Eu-Idh2	7 & 8		2	300	0.29	2	0.900-0.750
	9		2	508	0.07	2	0.967
Eu-Lap1	7 & 8		2	300	2.55	2	0.900-0.750
Eu-Mpi	9		2	418	2.55	2	0.279
Eu-Pgd1	7 & 8	R	3	300	3.03	1	0.100-0.050
Eu-Pgm2	7 & 8		2	300	4.41	2	0.250-0.100

Table 10. F_2 linkage tests of 24 isozyme loci (continued).

Gene pair	Cross[a]	Phase[b]	Segregation ratio[c]	n	x^2[d]	df	p
Eu-Pgm2	9		2	538	1.57	2	0.456
Fle-Aco2	7 & 8		2	300	1.43	2	0.500-0.250
	9		2	350	3.81	2	0.149
Fle-Ap	7 & 8		2	300	2.74	2	0.500-0.250
	9		2	538	3.51	2	0.173
Fle-Dia1	7 & 8		2	300	0.31	2	0.900-0.750
	9		2	538	4.25	2	0.119
Fle-Dia2	9		2	538	305.88	2	0.000
Fle-Enp	7 & 8		2	300	0.57	2	0.900-0.750
Fle-Est1	9		2	537	1.15	2	0.561
Fle-Idh1	7 & 8		2	300	2.36	2	0.500-0.250
	9		2	508	3.19	2	0.203
Fle-Idh2	7 & 8		2	300	2.74	2	0.500-0.250
	9		2	508	0.66	2	0.720
Fle-Lap1	7 & 8		2	300	3.86	2	0.250-0.100
Fle-Mpi	9		2	418	3.04	2	0.219
Fle-Pgd1	7 & 8	R	3	300	4.10	1	0.050-0.025
Fle-Pgm2	7 & 8		2	300	4.33	2	0.250-0.100
	9		2	538	4.19	2	0.123
Idh1-Idh2	7 & 8		1	300	9.98	4	0.050-0.025
	9		1	508	5.45	4	0.244
Idh1-Lap1	7 & 8		1	300	3.06	4	0.750-0.500
Idh1-Mpi	9		1	418	2.93	4	0.569
Idh1-Pgd2	6		1	351	1.73	4	0.785
Idh1-Pgi1	6		1	462	3.77	4	0.438
Idh1-Pgm2	7 & 8		1	300	3.55	4	0.500-0.250
	9		1	508	1.39	4	0.846
Idh1-Sdh	6		1	461	11.45	4	0.022
Idh2-Lap1	7 & 8		1	300	2.55	4	0.750-0.500
Idh2-Mpi	9		1	418	2.43	4	0.657
Idh2-Pgm2	7 & 8		1	300	1.90	4	0.900-0.750
	9		1	508	2.19	4	0.701
Idh2-Ti	3		1	342	4.73	4	0.316
	4		1	236	7.98	4	0.092
Lap1-Pgm2	7 & 8		1	300	4.55	4	0.500-0.250
Mpi-Pgm1	1		1	50	5.80	4	0.215
	2		1	335	0.67	4	0.955
Mpi-Pgm2	1		1	306	1.27	4	0.866
	2		1	361	1.36	4	0.852
	9		1	418	4.00	4	0.406
Pgd1-Aco2	7 & 8		2	300	4.02	2	0.250-0.100
Pgd1-Ap	7 & 8		2	300	7.54	2	0.025-0.010
Pgd1-Dia1	7 & 8		2	300	0.08	2	0.975-0.950
Pgd1-Enp	7 & 8		2	300	0.27	2	0.900-0.750

Table 10. F_2 linkage tests of 24 isozyme loci (continued).

Gene pair	Cross[a]	Phase[b]	Segregation ratio[c]	n	x^2[d]	df	p
Pgd1-Idh1	7 & 8		2	300	4.56	2	0.250-0.100
Pgd1-Idh2	7 & 8		2	300	1.05	2	0.750-0.500
Pgd1-Lap1	7 & 8		2	300	0.10	2	0.950-0.900
Pgd1-Pgm2	7 & 8		2	300	0.73	2	0.750-0.500
Pgd2-Pgi1	6		1	351	7.93	4	0.094
Pgd2-Sdh	6		1	349	0.98	4	0.913
Pgi1-Ap	3		2	216	1.40	2	0.498
	4		2	329	1.12	2	0.570
Pgi1-Dia1	3		2	120	1.00	2	0.608
	4		2	330	3.24	2	0.198
Pgi1-Idh2	3		2	252	2.29	2	0.318
	4		2	330	1.94	2	0.379
Pgi1-Sdh	6		1	459	2.88	4	0.578
Pgi1-Ti	3		2	222	0.34	2	0.844
	4		2	236	0.79	2	0.673
Pgm1-Pgm2	1		1	50	5.06	4	0.281
	2		1	338	1.71	4	0.789

[a] Cross 7 and cross 8 were reciprocal crosses. Due to small sample sizes, they were combined. The combined sample size was 300 for all gene pairs involved, consisting of 180 individuals from cross 7 and 120 from cross 8.

[b] Phase is only given for the four category segregation data. Other types of data could be organized into both coupling or repulsion phase. C: coupling; R: repulsion.

[c] 1: Both loci segregated at 1:2:1 ratio. The theoretical segregation ratio of dihybrids was 1:2:2:4:1:2:1:2:1.
2: The first locus segregated at 3:1 ratio. The second segregated at 1:2:1 ratio. The theoretical segregation ratio of dihybrids was 3:6:3:1:2:1.
3: Both loci segregated at 3:1 ratio. The theoretical segregation ratio of dihybrids was 9:3:3:1.

[d] The contingency Chi-square.

Table 11. Recombination frequencies of the gene pairs significant at p=0.01[a].

Gene pair[b]	Cross[c]	Observed frequencies[d]									Recombination frequency
		e	f	g	h+i	j	k	l	m	n	
Ap-Lap1	7 & 8	1	23	35	97	38	25	45	31	5	24±2.1%
Ap-Pgd2	6	18	46	40	99	28	38	35	37	11	39.5±2.5%
Ap-Ti	3	1	12	15	130	57	14	55	20	2	11.7±1.4%
	4	0	7	2	104	58	7	49	6	2	5.7±1.1%
Fle-Dia2	9	--22--			--262--			--122-	105	25	2	9.8±1.3%

[a] For sample sizes, x^2 and probabilities, see Table 10.
[b] All gene pairs were in simulated repulsion phase.
[c] Cross 7 and cross 8 were reciprocal crosses and combined due to small sizes.
[d] Genotypic classifications e to n were per Allard (1956).

allele for a leaky type.

A new mobility variant was found in seeds of KC1 of the South Korean populations. The mobility of the two Mpi bands of this variant was located between mobilities of the variants controlled by alleles Mpi-b and Mpi-c (Fig. 3). A new allele symbol Mpi-e was assigned. The inheritance study with reciprocal crosses (crosses 1 and 2, Table 6) using seeds of this variant and KA12 (Mpi-bb) showed a 1:2:1 segregation ratio for the plants with parental banding patterns and their heterozygotes in F_2 (Table 8). The parental types did not segregate in F_3, while the heterozygotes segregated in the 1:2:1 ratio as in F_2 (Table 9). This result suggested that Mpi-e is another allele at Mpi locus.

Phosphoglucomutase (Pgm) exhibits three bands. At the first band near the origin of gels, there are two variants, slow and fast, controlled by Pgm1 locus with Pgm1-a and Pgm1-b alleles (Kiang and Gorman, 1983 and Chiang, 1985). Most soybean varieties have Pgm1-aa genotype and slow band 1. A few have Pgm1-bb genotype and are the fast type. The second and third bands form a separate banding zone. There are three mobility variants, slow, medium and fast at the second band controlled by Pgm2-a, Pgm2-b and Pgm2-c alleles, respectively (Chiang, 1985). The third band is an invariant band for most soybean varieties. However, Chiang (1985) observed a null type on the plants with Pgm2-b allele at Pgm2 locus. She studied the inheritance of this variant and

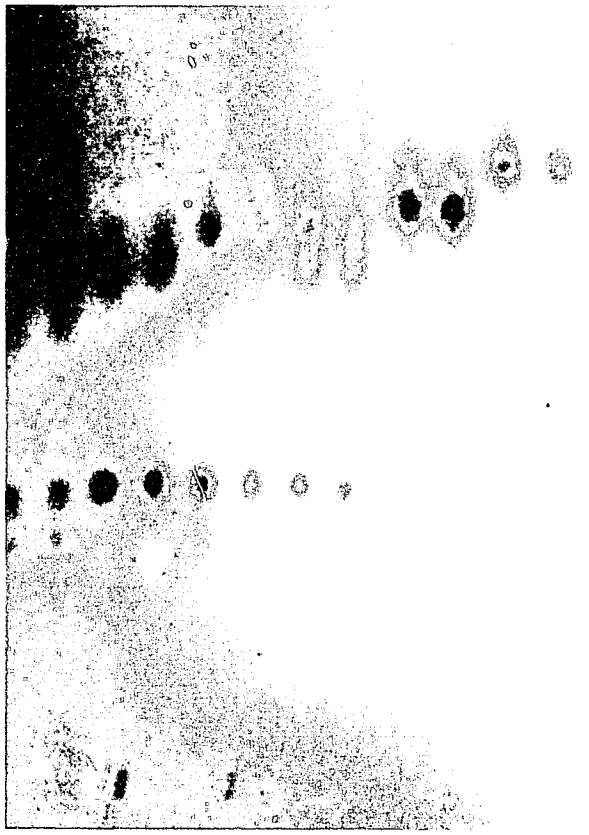


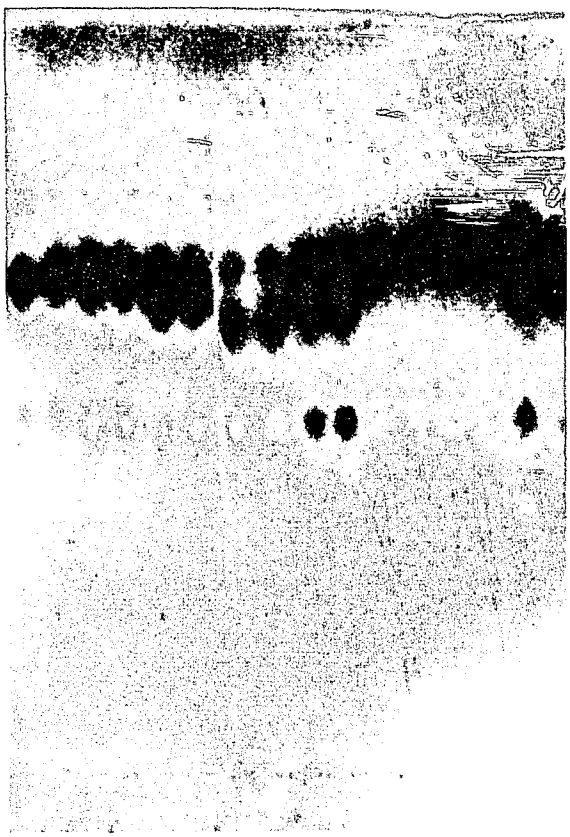
Fig. 3. Zymograms of soybean Mpi.
Lanes 1 & 2: KD14, Mpi-aa.
Lanes 3 & 4: KA12, Mpi-bb.
Lanes 5 & 6: KC1, Mpi-ee.
Lanes 7 & 8: KA12 x KC1, Mpi-be.
Lanes 9 & 10: KC13, Mpi-cc.
Lanes 11 & 12: PI407192, Mpi-dd.

assigned Pgm3 as the gene symbol, with Pgm3 and pgm3 alleles responsible for the presence and absence of the third band. It should be noted that for the plants with Pgm2-cc genotype, there is no way to know whether the band is overlapping bands 2 and 3 or only one of them is present while the other is null.

A fourth variant at band 2 was found in the South Korean populations, in which band 2 was faster than band 3 (Fig. 4). Pgm2-d allele symbol was assigned. Reciprocal crosses (crosses 7 and 8, Table 6) with seeds of KC13 (Pgm2-dd) and variety AV68 (Pgm2-bb) were made. The F_2 segregated in a 1:2:1 ratio for the parental types and their heterozygotes (Table 8). The parental types did not segregate further in F_3. The heterozygotes segregated in the 1:2:1 ratio as in F_2 (Table 9). Therefore, Pgm2-d is a new allele at Pgm2 locus.

Shikimate dehydrogenase (Sdh) has been studied by Bult (1989) and Chiang (1985). They did not observe any variation. All the varieties showed the same banding pattern.

In this study, seeds of a few natural plants of the South Korean populations were found to have a different Sdh banding pattern (Fig. 5). Sdh has three major bands. In the mutant type, all the three major bands were one band faster than the wild type. Reciprocal crosses (crosses 5 and 6, Table 6) were made between the mutant KD14 and the wild type KA3. F_2 and F_3 of the F_2 heterozygotes all segregated



Band	RF
2	0.79
3	0.75
2	0.71
2	0.67
1	0.55
1	0.52

Lane 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Fig. 4. Zymograms of soybean Pgm.
Lanes 1 & 2: KA12 x KC1, Pgm1-bb, Pgm2-cc.
Lanes 3 & 4: KA12 x KC1, Pgm1-ab, Pgm2-cc.
Lanes 5 & 6: KA12 x KC1, Pgm1-aa, Pgm2-bb, Pgm3-++.
Lanes 7 & 8: PI407265, Pgm1-aa, Pgm2-aa, Pgm3-++.
Lanes 9 & 10: AV68, Pgm1-aa, Pgm2-bb, Pgm3-++.
Lanes 11 & 12: KC1, Pgm1-aa, Pgm2-cc.
Lanes 13 & 14: KC13, Pgm1-aa, Pgm2-dd, Pgm3-++.
Lanes 15 & 16: AV68 x KC13, Pgm1-aa, Pgm2-bd, Pgm3-++.

Band Rf

6 0.41

6 0.38
5 0.34
4 0.31
3 0.27
2 0.24
1 0.19

Lane 1 2 3 4 5 6 7 8

Fig. 5. Zymograms of soybean Sdh.
Lanes 1 to 4: KA3, Sdh-aa.
Lanes 5 & 6: KD14, Sdh-bb.
Lanes 7 & 8: KD14 x KA3, Sdh-ab.

in a 1:2:1 ratio for the parental types and their heterozygotes (Tables 8 and 9). The F_2 parental types did not segregate in F_3. Therefore, Sdh is controlled by one gene. The Sdh gene symbol with Sdh-a and Sdh-b alleles was assigned. Sdh-aa is the genotype of wild type, whereas Sdh-bb is the mutant.

Linkage tests indicated that Sdh was not linked to Aco1, Aco3, Ap, Dia1, Dia2, Dia3, Pgd2, and Pgi1 (Table 10). If the 95% significance level was used, Sdh was linked to Idh1 (Table 10). The recombination frequency was 43.1±2.3%. If the 99% significance level was used, there was no association confirmed between these two loci. Further study is needed to verify their relationship.

Linkage tests were also performed on other loci. A total of 9 crosses were used for all inheritance and linkage studies (Table 6). Genotypes of the parents at all the loci studied are shown in Table 7. Table 8 gives the monogenic segregation data. Most loci did not deviate significantly at the 95% probability level from theoretical segregation ratios, except for Dia3 in cross 6, Idh2 in cross 9, Pgi1 in cross 4 and Pgm2 in cross 9. The linkage test results of 157 combinations involving 24 loci are given in Table 10.

Some of the 157 gene pairs had never been tested. Besides Adh3 and Sdh, which has been discussed above, Aco5, Dia3, Eu and Fle were also not previously tested for linkages with most other loci. If the 95% probability level was

used, 12 gene pairs showed significant associations (Table 10). The possible linkages of Dia1-Pgm2 in cross 9, Dia3-Dia1 in cross 4, Dia3-Ti in cross 3, and Idh1-Idh2 and Pgd1-Ap in crosses 7 and 8 were quickly eliminated, because either they did not have significant associations in other crosses tested or we already knew they were not linked. Ap-Lap1, Ap-Pgd2, Ap-Ti were known to be linked. The newly found linkage was between Fle and Dia2. Because of very high Chi-square, it was definite that Fle and Dia2 were linked. Although Est1-Mpi showed statistical significance at the 95% probability level, the recombination value was 46.6±2.4%. Thus, there was basically no linkage in the pair. The other doubtful linkages were Fle-Pgd1 and Idh1-Sdh. The recombination frequencies for the two pairs were 41±4.7% and 43.1±2.3%, respectively. Further studies are needed to determine whether they are linked.

After examining the above 12 gene pairs, the 99% probability level seemed to be a good criterion for linkages. If there was association between two loci at the 99% significance level, it was certain that there was a linkage. The linked loci detected at the 99% probability level and their recombination frequencies are given in Table 11.

Ap, Lap1, Pgd2 and Ti are in linkage group 9. Their recombination frequencies in Table 11 were very similar to those previously reported (Chiang and Kiang, 1987b and

Kiang, 1987). For the new linkage between Fle and Dia2 found in this study, the recombination frequency was 9.8±1.3%. I assign Fle-Dia2 to linkage group 16. Previously, linkage group 16 was assigned to linked Pgi1-Pgd1 (Chiang et al., 1987). However, it was later found that Pgi1 and Pgd1 were in linkage group 5 (Kiang, 1990a).

DISCUSSION

There are as many as 12 bands for diaphorase (Chiang, 1985; Gorman et al., 1983; and Kiang and Gorman, 1983). The first five bands near the origin of gels are controlled by Dia1 locus. For Dia1/Dia1 plants, all five bands are present. Only band 1 is intense when plants have dia1/dia1 genotype. In the heterozygotes, usually the first three bands show strong intensity. There are slow and fast variants for synchronous bands 7 and 8. They are controlled by Dia2-a and Dia2-b alleles, respectively, at Dia2 locus. Gorman et al. (1983) observed absence of band 10 in a variant and named the Dia3 locus. The presence and absence of band 10 were hypothesized to be controlled by Dia3 and dia3 alleles. However, Chiang (1985) observed a slow type for bands 7, 8, 9 and 10. She proposed that bands 9 and 10 were just like bands 7 and 8, and controlled by Dia3 locus with Dia3-a and Dia3-b alleles for the slow and fast variants at these two bands. The plants without band 10 observed by Gorman et al. (1983) actually had the genotypes of Dia2-bb and Dia3-aa. Bands 8 and 9 thus were overlapping, and band 10 was mistaken for band 9. Therefore, the original alleles Dia3 and dia3 became Dia3-b and Dia3-a, respectively. But, Chiang (1985) did not make crosses to confirm the hypothesis.

The slow type for bands 7, 8, 9 and 10 was also observed in these South Korean populations. The hypothesis proposed by Chiang (1985) was proved. Reciprocal crosses (crosses 3 and 4, Table 6) were made using seeds of KA5 (Dia2-aa and Dia3-aa), the slow type for bands 7, 8, 9 and 10, and KB20 (Dia2-aa and Dia3-bb) with slow bands 7 and 8, and fast bands 9 and 10. The result indicated that the slow and fast types of bands 9 and 10 segregated in a single gene manner (Table 8). However, the fast type and the heterozygotes were sometimes difficult to distinguish. They were combined and a 3:1 ratio was tested in the reciprocal crosses. Both crosses 3 and 4 did not deviate significantly from the 3:1 ratio. In the other two crosses also segregating at Dia3 locus, crosses 6 and 9, only cross 9 significantly deviated from the 3:1 ratio at the 95% probability level (Table 8).

Another enzyme to be discussed in this section is leucine aminopeptidase (Lap). There are two anodal bands for this enzyme. Two alleles, Lap1-a and Lap1-b at Lap1 locus control the slow and fast types, respectively, at band 1 near the origin of gels (Kiang et al., 1985). The second band is apparent only in the germinated green cotyledons. Three variants, slow, fast and null at the second band, are controlled by Lap2-a, Lap2-b and lap2 alleles, respectively.

A third variant, even slower than the slow variant at band 1 was found in KF18 and KF25 of the South Korean popu-

lations. Since the plants grown from KF18 and KF25 seeds had a very long growing season and flowered very late, the flowering time was difficult to match with that of other plants, and hybridization was unsuccessful. No genetic data were obtained. In order to analyze data for the population genetics part of this study, the Lap1-c allele symbol was assigned.

CHAPTER II

GENETIC VARIATION AND DIFFERENTIATION

INTRODUCTION

Genetic variation is important in the improvement of crop cultivars. However, commercial cultivars of the major crops are developed from a limited number of ancestral lines. In soybean (Glycine max [L.] Merr.), a parentage study revealed that 50% of the North American germplasm was contributed by four introductions in the northern cultivars and two introductions in the southern cultivars released during 1971-1981 (Delannay et al., 1983). Furthermore, a single introduction, "Mandarin", accounted for 30% of the genes in the northern gene pool. The issue of narrowing genetic background is not discussed today in the academic community as much as in the 1970's and 1980's, mainly because of improvements by plant breeders in diversifying the genetic background of the commercial cultivars. Diversification of genetic background, however, is still a very important factor in plant breeding programs. Wild soybean (Glycine soja Sieb. & Zucc.) is believed to be the progenitor of the cultivated soybean (Hymowitz and Singh, 1987). The two species can be intercrossed freely and produce fertile offspring. They together form the soybean gene pool (Kiang

et al., 1987). Therefore, wild soybean is the best candidate for broadening the genetic background of the cultivated soybean germplasm, especially for characters, such as protein content, disease and pest resistance. Evaluation of genetic variation in wild soybean can enhance our understanding about wild soybean and facilitate its use in breeding programs.

Genetic variation in wild soybean as well as in the cultivated soybean has been studied by a few authors. The results indicate that there is little variation in mtDNA, cpDNA, the 18S and 25S rDNA, and the 5S rDNA in the soybean gene pool assayed by RFLP (Restriction Fragment Length Polymorphism) technique (Doyle, 1988; Doyle and Beachy, 1985; Shoemaker et al., 1986; and Sisson et al., 1978). On the other hand, considerable genetic variation is found with random genomic DNA RFLP, isozymes and other biochemical and morphological genetic markers (Broich and Palmer, 1981; Bult, 1989; Chiang, 1985; Doong, 1986; Gorman, 1983; Hu and Wang, 1985; Hymowitz and Kaizuma, 1979 and 1981; Keim et al., 1989; Kiang and Chiang, 1990; Kiang and Gorman, 1983; and Kiang et al., 1987). This genetic variation is comparable to that in other autogamous plant species. Wild soybean has consistently shown higher genetic variation than domestic cultivars (Gorman, 1983; Kiang and Gorman, 1983, and Kiang et al., 1987). It appears that the Korean peninsula and southern Japan are the centers of gene diversity for

both the cultivated and wild soybeans (Broich and Palmer, 1981; Chiang, 1985; Gorman, 1983; Hymowitz and Kaizuma, 1979 and 1981; Kiang and Gorman, 1983; and Kiang et al., 1987).

Previous studies on genetic variation of South Korea cultivated and wild soybeans all used accessions from the USDA Soybean Germplasm Collection. More information would be obtained to use natural populations to study the genetic variation in South Korean wild soybean. The objectives of this study were to determine how the genetic variation in six South Korean natural populations of wild soybean was organized within and among populations and how differentiated these populations were. I was also interested in comparing the results of this study with the results of other studies on South Korean wild soybean accessions obtained from the USDA Soybean Germplasm Collection and other natural populations of wild soybean, such as those collected along the Kitakami River and Mishima city of Japan (Bult, 1989 and Chiang, 1983).

MATERIALS AND METHODS

Seed Materials.

The seed source of six natural populations, consisting of 172 natural plants, was described in the Introduction chapter. Seeds were collected from South Korea in 1986 by Dr. Y.T. Kiang and kept in a lab freezer before use. At least five original seeds of each natural plant were examined for all the enzymes involved in this study. One seed from each natural plant was sown in the greenhouse. Six more seeds from this plant were also examined for all enzymes. Thus, a total of 11 seeds representing six original seeds were examined for each natural plant. The results from the two batches of seeds were compared. If any discrepancies were found, more original seeds as well as seeds from the greenhouse plant were examined until the genotypes across all the isozyme loci were correctly identified for each natural plant. This procedure helped reduce errors, also allowed detection of heterozygotes, not only in the 172 natural plants, but also in the greenhouse-grown plants, which were derived from original seeds.

Electrophoresis.

The enzymes and protein assayed and methods of electrophoresis were described in Chapter I. The recipes of gels and staining solutions are given in Appendix I. Usually, a

seed was cut into three pieces, one for testing half number of the enzymes, one for the other half and the third piece as a backup for a repeat of the assay.

Statistical Analysis.

The traditional population genetics measures were calculated with Biosys-1, a Fortran computer program (Swofford and Selander, 1981 and 1989). A total number of 35 loci was used throughout the analysis. The following measures were calculated:

1). Mean number of alleles per locus.

2). 95% and 99% polymorphism. Polymorphism is the percentage of polymorphic loci in the total number of loci examined. A locus is considered polymorphic when the most frequent allele is less than 95% or 99%.

3). Expected heterozygosity based on the Hardy-Weinberg equilibrium. For a particular population at one locus, the expected heterozygosity can be estimated with the standard formula:

$$H_{exp}=1-\Sigma X_i^2,$$

where X_i is the frequency of the ith allele. Since the population sizes were small, the unbiased estimate by

$$H_{exp}'=2n(1-\Sigma X_i^2)/(2n-1),$$

where n is population size, was used (Nei, 1978). The values of the unbiased estimates were larger than the biased estimates in this study, but the differences were small and no larger than 0.003. The mean expected heterozygosity ($\bar{H}_{exp}'$)

is the average across all loci examined.

4). Partitioning of the total gene diversity. Since the expected heterozygosity is never reached in reality in inbreeding species, Nei (1973) called it gene diversity and showed that the gene diversity in the total population consisting of several populations could be partitioned into the gene diversities within and between populations, i.e. $H_T=H_S+D_{ST}$, where H_T is the total gene diversity, H_S is the average within-population gene diversity and D_{ST} is the gene diversity between populations. The proportion of the between-population gene diversity in the total gene diversity is termed gene differentiation, expressed as G_{ST}. The Biosys-1 program did not directly give these terms. G_{ST} was given as F_{ST} (the fixation index) in the F statistics section of the output, since G_{ST} is the same as F_{ST} (Hartl, 1988; Hedrick, 1983; and Swofford, 1989). H_T was given as the total limiting variance. H_S was re-calculated from the data as explained by Hartl (1988. pp. 79-81). D_{ST} is the difference between H_T and H_S.

5). Nei's genetic identity and distance. Nei's genetic distance is calculated by the formula,

$$D=-\ln(I),$$

where I is the gene identity.

$$I=J_{XY}/\sqrt{(J_X J_Y)},$$

where J_{XY}, J_X and J_Y are the averages of j_{XY}, j_X and j_Y

across all the polymorphic loci examined.

$$j_{XY}=\Sigma X_i Y_i,$$

$$j_X=\Sigma X_i^2,$$

$$j_Y=\Sigma Y_i^2,$$

where X_i and Y_i are the frequencies of ith allele at a particular locus in populations X and Y, respectively.

Nei (1978) later developed a formula for calculating unbiased estimates of genetic identity and distance for populations of small size. The unbiased estimates are obtained by substituting $(2n_XJ_X-1)/(2n_X-1)$ for J_X and $(2n_YJ_Y-1)/(2n_Y-1)$ for J_Y in the previous formula, where n_X and n_Y are the sizes for populations X and Y, respectively. In this study, the unbiased genetic identity was larger than the standard genetic identity, and the differences were about 0.003. The unbiased genetic distance was smaller than the standard estimate, with differences around 0.02.

Multilocus association was analyzed with hand calculators and will be discussed in the results section. Regression of the average number of loci different for each multilocus genotype with multilocus genotypes in other populations on that in the same population was conducted with PLOT procedure of SPSS computer program (Norusis, 1990).

The CLUSTER procedure of SPSS computer program was used to do cluster analysis. Three cluster analyses were performed for populations using: a) Nei's unbiased genetic

distance obtained from Biosys-1, b) the squared Euclidean distance available in SPSS, and c) the number of loci different between individuals of different populations calculated with hand calculators. One cluster analysis was performed for multilocus genotypes using the number of loci different between them. The average linkage between groups method (UPGMA) in CLUSTER was chosen for combining clusters in all four cluster analyses.

For the cluster analysis for populations using the squared Euclidean distance, the allele frequency data obtained from Biosys-1 were converted to Z score data with the DESCRIPTIVES procedure of SPSS before CLUSTER procedure was applied. The squared Euclidean distance and dendrogram were requested on the output.

There were 64 allele frequency variables for the 27 polymorphic loci in the data set that could be used for clustering the six populations. At each locus, one allele could be deleted from the analysis. Just as the degrees of freedom are always one less than the sample size, the information of this allele was totally contained in other alleles at that locus, because the combined frequency for all alleles at one locus was 1. Thus, the possibility arose that several analyses could be done. Three analyses were carried out, one with all 64 alleles in the data, one with 37 alleles excluding the most frequent allele at each locus and

the other with 37 alleles excluding the least frequent allele at each locus. The results of the first two were similar. The last one was rather different. Since differences among populations are usually reflected by rare and less frequent alleles in isozyme studies, especially for self-pollinated plant species, and the most frequent alleles can mask such subtle differences among populations, the analysis with 37 alleles excluding the most frequent alleles was appropriate and is presented.

For the other three cluster analyses, the data were in matrix forms. The MATRIX DATA and PROXIMITIES procedures were used first before CLUSTER procedure was applied.

RESULTS

A variant without Adh band 4 and a variant with fast mobilities for all Sdh bands were found. Two new loci, Adh3 and Sdh, were assigned. For the total 35 loci examined including the two new loci, 27 were polymorphic. The 172 natural plants all had the same genotypes at eight other loci, which were Aco4-bb at Aco4, Adh1/Adh1 at Adh1, Adh2/Adh2 at Adh2, fle/fle at Fle, Got-bb at Got, Pgd3-bb at Pgd3, Pgi3-bb at Pgi3 and Pgm3/Pgm3 at Pgm3. Although the genotypes of some natural plants were not identified at Pgm3 locus due to overlapping bands, those that were determined showed no variation. Pgm3 was considered monomorphic.

Several new or rare zymogram types were found in the six populations: a rare zymogram type with slow bands for Dia2 and Dia3 loci, an even slower band than the slow band for the first locus of Lap enzyme, a new allele producing a Mpi band between those produced by Mpi-b and Mpi-c alleles, and a variant having an even faster second band than the fastest mobility band controlled by Pgm2-c allele. Their inheritance was studied, and Lap1-c, Mpi-e and Pgm2-d allele symbols were assigned. There was a total of 92 confirmed alleles for the 35 loci, which included 87 alleles reported in literature and seven new alleles, Adh3, adh3,

Mpi-e, Lap1-c, Pgm2-d, Sdh and sdh assigned in this study. Seventy two of them were present in these six populations. The allele frequencies for the 27 polymorphic loci are presented in Table 12. The maximum numbers of alleles per locus were 4 at Ap and Mpi.

The mean number of alleles per locus, 99% polymorphism and the expected heterozygosity were 1.4, 37.2% and 0.134, respectively, averaged over the six populations, and 2.1, 77.1% and 0.215, respectively, in the total population (Table 13). Population F had the highest genetic variation by all measures, followed by population A (Table 13). Populations E and C were the third and fourth, respectively. Compared with population D, population B also had more genetic variation, except as measured by the 95% polymorphism. Population C had the highest observed heterozygosity per locus per individual.

The contingency Chi-square tests indicated that the populations were significantly different for allele distribution at each polymorphic locus (Table 14). Partitioning of the total gene diversity at the polymorphic loci showed that G_{ST} (gene differentiation) varied from 0.077 for Aco5 and Dia3 to 0.675 for Pgm2 (Table 15). The mean G_{ST} across all polymorphic loci was 0.383, which means that 38.3% of the total gene diversity existed between populations.

Nei's genetic identity and distance were calculated for each pair of populations (Table 16). Population D and popu-

Table 12. Allele frequencies for the 27 polymorphic loci.

	Population (plant number in parenthesis)						
Locus	A(27)	B(30)	C(30)	D(18)	E(41)	F(26)	Mean(29)
Aco1							
a	0.000	0.000	0.000	0.222	0.000	0.231	0.076
b	1.000	1.000	1.000	0.778	1.000	0.769	0.925
Aco2							
a	0.556	0.033	0.000	0.000	0.000	0.308	0.150
b	0.444	0.967	1.000	1.000	1.000	0.500	0.819
c	0.000	0.000	0.000	0.000	0.000	0.192	0.032
Aco3							
a	1.000	1.000	1.000	0.556	0.622	0.346	0.754
b	0.000	0.000	0.000	0.444	0.378	0.654	0.246
Aco5							
a	1.000	1.000	1.000	1.000	0.927	0.885	0.969
b	0.000	0.000	0.000	0.000	0.000	0.115	0.019
-	0.000	0.000	0.000	0.000	0.073	0.000	0.012
Adh3*							
+	0.333	0.533	1.000	1.000	0.927	0.654	0.74.
-	0.667	0.467	0.000	0.000	0.073	0.346	0.259
Am3[a]							
a	0.000	0.000	0.000	0.000	0.378	0.269	0.108
b	1.000	1.000	1.000	1.000	0.622	0.731	0.892
Ap							
a	0.704	0.467	0.700	0.778	0.000	0.385	0.506
b	0.000	0.500	0.000	0.000	0.000	0.000	0.083
c	0.037	0.033	0.300	0.222	1.000	0.615	0.368
d	0.259	0.000	0.000	0.000	0.000	0.000	0.043
Dia1							
+	0.481	0.467	0.433	0.444	0.000	0.000	0.304
-	0.519	0.533	0.567	0.556	1.000	1.000	0.696
Dia2							
a	0.259	0.467	0.000	0.778	0.524	0.000	0.338
b	0.741	0.533	1.000	0.222	0.476	1.000	0.662
Dia3							
a	0.074	0.033	0.000	0.000	0.146	0.192	0.074
b	0.926	0.967	1.000	1.000	0.854	0.808	0.926
Enp							
a	0.000	0.000	0.450	0.000	0.354	0.000	0.134
b	1.000	1.000	0.550	1.000	0.646	1.000	0.866

Table 12. Allele frequencies for the 27 polymorphic loci (continued).

	Population (plant number in parenthesis)						
Locus	A(27)	B(30)	C(30)	D(18)	E(41)	F(26)	Mean(29)
Est1							
a	0.000	0.000	0.417	0.000	0.073	0.192	0.114
b	1.000	1.000	0.583	1.000	0.927	0.808	0.886
Eu[b]							
a	1.000	1.000	1.000	1.000	1.000	0.885	0.981
-	0.000	0.000	0.000	0.000	0.000	0.115	0.019
Idh3							
a	1.000	0.500	0.533	1.000	1.000	1.000	0.839
b	0.000	0.500	0.467	0.000	0.000	0.000	0.161
Idh4							
a	0.778	1.000	0.450	1.000	1.000	1.000	0.871
b	0.222	0.000	0.550	0.000	0.000	0.000	0.129
Lap1							
a	0.000	0.467	0.033	0.000	0.354	0.000	0.142
b	1.000	0.533	0.967	1.000	0.646	0.923	0.845
c*	0.000	0.000	0.000	0.000	0.000	0.077	0.013
Mpi							
a	0.407	0.000	0.000	0.778	0.427	0.000	0.269
b	0.556	1.000	0.000	0.000	0.073	0.654	0.381
c	0.037	0.000	0.967	0.222	0.500	0.346	0.345
e*	0.000	0.000	0.033	0.000	0.000	0.000	0.006
Pgd1							
a	0.407	0.933	0.000	0.000	0.000	0.115	0.243
b	0.593	0.033	1.000	1.000	0.805	0.423	0.642
c	0.000	0.033	0.000	0.000	0.195	0.462	0.115
Pgd2							
a	0.870	1.000	1.000	0.778	0.354	0.808	0.802
b	0.130	0.000	0.000	0.222	0.646	0.192	0.198
Pgi1							
a	0.074	0.467	0.000	0.000	0.000	0.000	0.090
b	0.926	0.067	1.000	1.000	1.000	1.000	0.832
-	0.000	0.467	0.000	0.000	0.000	0.000	0.078
Pgi2							
+	0.519	1.000	0.933	1.000	0.927	0.115	0.749
-	0.481	0.000	0.067	0.000	0.073	0.885	0.251
Pgm1							
a	0.778	1.000	1.000	1.000	1.000	1.000	0.963
b	0.222	0.000	0.000	0.000	0.000	0.000	0.037

Table 12. Allele frequencies for the 27 polymorphic loci (continued).

	Population (plant number in parenthesis)						
Locus	A(27)	B(30)	C(30)	D(18)	E(41)	F(26)	Mean(29)
Pgm2							
b	1.000	1.000	0.000	1.000	0.927	0.885	0.802
c	0.000	0.000	0.817	0.000	0.073	0.115	0.168
d*	0.000	0.000	0.183	0.000	0.000	0.000	0.031
Sdh*							
a	1.000	1.000	1.000	0.778	1.000	1.000	0.963
b	0.000	0.000	0.000	0.222	0.000	0.000	0.037
Ti							
a	0.927	0.967	0.433	1.000	1.000	1.000	0.888
b	0.074	0.033	0.567	0.000	0.000	0.000	0.112

* Loci or alleles which were assigned in this study.
[a] Same as Sp1.
[b] A separate locus was proposed for the null type (Kloth et al., 1987), but needs to be independently confirmed.

Table 13. Genetic variation as measured by the number of alleles per locus, polymorphism and the expected heterozygosity[a].

Population	Sample size	Mean number of alleles per locus	Polymorphism 95%	Polymorphism 99%	Heterozygosity H_{exp}[b]	Heterozygosity H_{obs}[c]
A	27	1.5 (0.1)	42.9	42.9	0.158 (0.035)	0.001 (0.001)
B	30	1.4 (0.1)	22.9	31.4	0.113 (0.035)	0.000 (0.000)
C	30	1.3 (0.1)	28.6	34.3	0.129 (0.035)	0.016 (0.006)
D	18	1.3 (0.1)	25.7	25.7	0.100 (0.030)	0.000 (0.000)
E	41	1.4 (0.1)	40.0	40.0	0.133 (0.033)	0.005 (0.002)
F	26	1.5 (0.1)	45.7	48.6	0.168 (0.035)	0.000 (0.000)
Mean	28.7	1.4	34.3	37.2	0.134	0.004
Total pop	172	2.1 (0.1)	62.9	77.1	0.215 (0.032)	0.004 (0.001)

[a] Numbers in parentheses are standard errors.
[b] Nei's unbiased estimate of the expected heterozygosity based on the Hardy-Weinberg equilibrium.
[c] The observed heterozygosity per locus per individual.

Table 14. Contingency Chi-square tests of the heterogeneity of allelic distributions among populations at each polymorphic locus[a].

Locus	Number of alleles	X^2	DF
Aco1	2	61.80	5
Aco2	3	187.36	10
Aco3	2	122.15	5
Aco5	3	53.60	10
Adh3	2	111.49	5
Am3	2	84.45	5
Ap	4	395.20	15
Dia1	2	85.35	5
Dia2	2	108.37	5
Dia3	2	23.58	5
Enp	2	97.51	5
Est1	2	75.18	5
Eu	2	34.29	5
Idh1	2	55.41	5
Idh2	2	131.31	5
Idh3	2	130.46	5
Idh4	2	131.31	5
Lap1	3	114.72	10
Mpi	4	344.48	15
Pgd1	3	306.45	10
Pgd2	2	120.79	5
Pgi1	3	295.18	10
Pgi2	2	189.31	5
Pgm1	2	66.77	5
Pgm2	3	281.37	10
Sdh	2	70.07	5
Ti	2	145.76	5

[a] $p<0.001$ for all loci.

Table 15. Partitioning of the gene diversity within and among populations at the polymorphic loci[a].

Locus	H_T	H_S	D_{ST}	G_{ST}
Aco1	0.140	0.117	0.023	0.163
Aco2	0.307	0.196	0.111	0.361
Aco3	0.371	0.236	0.135	0.364
Aco5	0.061	0.056	0.005	0.077
Adh3	0.384	0.255	0.129	0.335
Am3	0.192	0.144	0.048	0.252
Ap	0.600	0.368	0.232	0.387
Dia1	0.423	0.330	0.093	0.220
Dia2	0.448	0.288	0.160	0.357
Dia3	0.138	0.127	0.011	0.077
Enp	0.232	0.159	0.073	0.316
Est1	0.201	0.155	0.046	0.229
Eu	0.038	0.034	0.004	0.098
Idh1	0.083	0.070	0.013	0.159
Idh2	0.224	0.140	0.084	0.375
Idh3	0.270	0.166	0.104	0.385
Idh4	0.224	0.140	0.084	0.375
Lap1	0.266	0.194	0.072	0.272
Mpi	0.664	0.325	0.339	0.511
Pgd1	0.515	0.253	0.262	0.509
Pgd2	0.318	0.223	0.095	0.298
Pgi1	0.293	0.116	0.177	0.604
Pgi2	0.376	0.161	0.215	0.573
Pgm1	0.071	0.058	0.013	0.192
Pgm2	0.328	0.107	0.221	0.675
Sdh	0.071	0.058	0.013	0.192
Ti	0.199	0.115	0.084	0.421
Mean	0.275	0.170	0.105	0.383

[a] H_T, H_S and D_{ST} are the total, within- and between-population gene diversities, respectively. G_{ST} is gene differentiation.

Table 16. Nei's unbiased estimates of genetic identity (above diagonal line) and genetic distance (below diagonal line).

Population	A	B	C	D	E	F	Mean
A	***	0.922	0.874	0.931	0.896	0.936	0.912
B	0.081	***	0.825	0.882	0.864	0.867	0.872
C	0.135	0.193	***	0.875	0.870	0.847	0.858
D	0.072	0.125	0.133	***	0.945	0.899	0.906
E	0.110	0.147	0.140	0.057	***	0.924	0.900
F	0.066	0.143	0.166	0.106	0.079	***	0.895
							***
Mean	0.093	0.138	0.153	0.099	0.107	0.112	0.890 *** 0.117

lation E were the closest, indicated by the highest genetic identity and smallest genetic distance. The smallest genetic identity occurred between population B and population C (0.825), but, based on the genetic distance, they were not the most different populations. Instead, populations C and F showed the largest genetic distance (0.166). However, the mean genetic identities and the mean genetic distances corresponded quite well. Population C had the smallest genetic identities and largest genetic distances with other populations, and was the most distinct in these six populations. Population A showed the largest genetic identities and smallest genetic distances with other populations. The overall mean genetic identity and genetic distance were 0.890 and 0.117, respectively.

Multilocus Association Analysis.

It was found during the study that there were correlations among isozyme loci in these populations. For instance in population D, plant KD16 had allele b at Aco2 locus and allele a at Ap locus. Other plants, such as KD17, KD18 and KD19 also tended to have allele a at Ap locus if it had allele b at Aco2 locus. Consequently, all plants in population D could be classified into three types, although there were nine polymorphic loci in this population. This phenomenon is called "linkage disequilibrium" or "multilocus association" in population genetics. If this multilocus association was analyzed, a great deal of new information could be

obtained about the genetic structure of natural populations. This led me to classify all the 172 plants in these six populations into multilocus genotypes (Table 17).

First, the plants in all six populations were examined for the first locus Aco1. Most had Aco1-bb genotype. Only a few had Aco1-aa genotype. The plants with the least frequent Aco1-a allele were further examined at Aco2, Aco3, and so on. At Dia1 locus, a difference was found. They were therefore divided into two groups based on their genotypes at Dia1. Examination of later loci did not reveal any further division within each group. Thus, multilocus genotypes 1 and 2 were assigned. Table 18 shows that multilocus genotypes 1 and 2 consisted of four plants from population D, and six plants from population F, respectively.

Since all the other plants had Aco1-bb genotype at Aco1, the second locus, Aco2, was examined for the rest of plants. At this locus, Aco2-cc genotype was the least frequent genotype. Examination of the plants with Aco2-cc for genotypes at other loci indicated that they could not be classified. This group, consisting of five plants from population F, was named multilocus genotype 3. The next least frequent genotype was Aco2-aa at Aco2 locus. They were further divided into ten multilocus genotypes based on such sequential examination of the genotypes at other loci.

Since the rest of plants had bb genotype at Aco2 locus,

Table 17. Multilocus genotypes at the polymorphic loci[a].

MLG[b]	Aco1	Aco2	Aco3	Aco5	Adh3	Am3	Ap	Dia1	Dia2
1	aa	bb	bb	aa	++	bb	cc	++	bb
2	aa	bb	bb	aa	++	bb	cc	--	bb
3	bb	cc	aa	aa	++	aa	cc	--	bb
4	bb	aa	bb	aa	--	bb	aa	--	bb
5	bb	aa	aa	aa	++	bb	aa	--	bb
6	bb	aa	aa	aa	++	bb	bb	--	bb
7	bb	aa	aa	aa	++	bb	dd	--	bb
8	bb	aa	aa	aa	--	bb	dd	++	bb
9	bb	aa	aa	aa	--	bb	aa	++	bb
10	bb	aa	aa	aa	--	bb	aa	--	bb
11	bb	aa	aa	aa	--	bb	aa	--	bb
12	bb	aa	aa	aa	--	bb	aa	--	bb
(11x12)	bb	aa	aa	aa	--	bb	aa	--	bb
13	bb	bb	bb	bb	++	aa	aa	--	bb
14	bb	bb	bb	aa	--	aa	cc	--	bb
15	bb	bb	bb	aa	--	bb	cc	--	bb
16	bb	bb	bb	aa	++	bb	aa	++	aa
17	bb	bb	bb	aa	++	aa	cc	--	aa
18	bb	bb	bb	aa	++	aa	cc	--	bb
19	bb	bb	aa	bb	++	bb	cc	--	bb
20	bb	bb	aa	--	++	bb	cc	--	aa
21	bb	bb	aa	aa	--	bb	aa	--	aa
22	bb	bb	aa	aa	--	bb	aa	++	aa
23	bb	bb	aa	aa	--	bb	aa	++	aa
24	bb	bb	aa	aa	++	bb	dd	++	bb
25	bb	bb	aa	aa	++	bb	bb	++	bb
26	bb	bb	aa	aa	++	bb	aa	--	bb
27	bb	bb	aa	aa	++	bb	aa	--	aa
28	bb	bb	aa	aa	++	bb	aa	--	aa
29	bb	bb	aa	aa	++	bb	aa	++	bb
30	bb	bb	aa	aa	++	bb	aa	++	bb
(26x30)#1	bb	bb	aa	aa	++	bb	aa	++	bb
(26x30)#2	bb	bb	aa	aa	++	bb	aa	+-	bb
31	bb	bb	aa	aa	++	bb	cc	++	aa
32	bb	bb	aa	aa	++	bb	cc	++	bb
33	bb	bb	aa	aa	++	bb	cc	++	bb
34	bb	bb	aa	aa	++	bb	cc	--	aa
35	bb	bb	aa	aa	++	bb	cc	--	bb
36	bb	bb	aa	aa	++	bb	cc	--	bb
37	bb	bb	aa	aa	++	bb	cc	--	bb
38	bb	bb	aa	aa	++	bb	cc	--	bb
(26x?)	bb	bb	aa	aa	++	bb	cc	+-	bb
(18x34)	bb	bb	ab	aa	++	ab	cc	--	ab

Table 17. Multilocus genotypes at the polymorphic loci (continued)[a].

MLG[b]	Dia3	Enp	Est1	Eu	Idh1	Idh2	Idh3	Idh4	Lap1
1	bb	bb	bb	aa	aa	aa	aa	aa	bb
2	bb	bb	bb	aa	bb	aa	aa	aa	bb
3	aa	bb	aa	aa	bb	aa	aa	aa	bb
4	bb	bb	bb	aa	bb	aa	aa	aa	bb
5	bb	bb	bb	aa	bb	aa	aa	aa	bb
6	bb	bb	bb	aa	bb	aa	bb	aa	bb
7	bb	bb	bb	aa	bb	aa	aa	aa	bb
8	bb	bb	bb	aa	bb	aa	aa	aa	bb
9	bb	bb	bb	aa	bb	aa	aa	aa	bb
10	bb	bb	bb	aa	bb	aa	aa	aa	bb
11	bb	bb	bb	aa	bb	aa	aa	aa	bb
12	bb	bb	bb	aa	bb	aa	aa	aa	bb
(11x12)	bb	bb	bb	aa	bb	aa	aa	aa	bb
13	bb	bb	bb	aa	bb	aa	aa	aa	cc
14	aa	bb	bb	aa	bb	aa	aa	aa	bb
15	bb	bb	bb	aa	bb	aa	aa	aa	bb
16	bb	bb	bb	aa	bb	aa	aa	aa	bb
17	bb	bb	bb	aa	bb	aa	aa	aa	bb
18	bb	bb	bb	aa	bb	aa	aa	aa	bb
19	bb	bb	bb	aa	aa	aa	aa	aa	bb
20	aa	bb	aa	aa	bb	aa	aa	aa	bb
21	bb	bb	bb	aa	bb	aa	bb	aa	bb
22	bb	bb	bb	aa	bb	bb	aa	bb	bb
23	bb	bb	bb	aa	bb	bb	aa	bb	bb
24	bb	bb	bb	aa	bb	aa	aa	aa	bb
25	bb	bb	bb	aa	bb	aa	aa	aa	aa
26	bb	bb	bb	aa	bb	bb	aa	bb	bb
27	aa	bb	bb	aa	bb	aa	aa	aa	bb
28	bb	bb	bb	aa	bb	aa	aa	aa	bb
29	bb	aa	aa	aa	bb	aa	bb	aa	bb
30	bb	aa	aa	aa	bb	aa	bb	aa	bb
(26x30)#1	bb	aa	aa	aa	bb	ab	bb	ab	bb
(26x30)#2	bb	aa	aa	aa	bb	ab	ab	ab	bb
31	aa	bb	bb	aa	bb	bb	aa	bb	bb
32	bb	aa	aa	aa	bb	aa	bb	aa	bb
33	bb	bb	bb	aa	bb	aa	bb	aa	bb
34	bb	aa	bb	aa	bb	aa	aa	aa	aa
35	aa	bb	bb	aa	bb	aa	aa	aa	bb
36	bb	aa	bb	aa	bb	aa	bb	aa	aa
37	bb	bb	bb	--	bb	aa	aa	aa	bb
38	bb	bb	bb	aa	bb	aa	aa	aa	bb
(26x?)	bb	ab	ab	aa	bb	ab	ab	ab	bb
(18x34)	bb	ab	bb	aa	bb	aa	aa	aa	ab

Table 17. Multilocus genotypes at the polymorphic loci (continued)[a].

MLG[b]	Mpi	Pgd1	Pgd2	Pgi1	Pgi2	Pgm1	Pgm2	Sdh	Ti
1	aa	bb	bb	bb	++	aa	bb	bb	aa
2	bb	cc	aa	bb	--	aa	bb	aa	aa
3	bb	cc	aa	bb	--	aa	bb	aa	aa
4	cc	bb	aa	bb	--	aa	bb	aa	aa
5	cc	bb	aa	bb	++	aa	bb	aa	aa
6	bb	bb	aa	bb	++	aa	bb	aa	bb
7	aa	aa	aa	bb	++	aa	bb	aa	aa
8	aa	aa	aa	bb	--	aa	bb	aa	aa
9	bb	bb	aa	bb	--	aa	bb	aa	aa
10	bb	bb	aa	bb	--	bb	bb	aa	aa
11	aa	aa	aa	bb	--	aa	bb	aa	aa
12	aa	aa	bb	bb	--	aa	bb	aa	aa
(11x12)	aa	aa	ab	bb	--	aa	bb	aa	aa
13	bb	bb	bb	bb	++	aa	cc	aa	aa
14	aa	bb	bb	bb	++	aa	bb	aa	aa
15	cc	bb	aa	bb	--	aa	bb	aa	aa
16	cc	bb	aa	bb	++	aa	bb	aa	aa
17	cc	bb	bb	bb	++	aa	bb	aa	aa
18	cc	bb	bb	bb	++	aa	bb	aa	aa
19	bb	cc	aa	bb	++	aa	cc	aa	aa
20	bb	bb	bb	bb	--	aa	cc	aa	aa
21	bb	aa	aa	--	++	aa	bb	aa	aa
22	bb	bb	aa	bb	++	bb	bb	aa	aa
23	bb	bb	bb	bb	++	aa	bb	aa	bb
24	bb	bb	aa	bb	++	aa	bb	aa	aa
25	bb	aa	aa	aa	++	aa	bb	aa	aa
26	cc	bb	aa	bb	++	aa	cc	aa	bb
27	bb	bb	aa	aa	++	aa	bb	aa	aa
28	aa	bb	aa	bb	++	aa	bb	aa	aa
29	cc	bb	aa	bb	++	aa	cc	aa	aa
30	cc	bb	aa	bb	++	aa	dd	aa	aa
(26x30)#1	cc	bb	aa	bb	++	aa	cd	aa	ab
(26x30)#2	cc	bb	aa	bb	++	aa	cd	aa	ab
31	bb	bb	bb	aa	++	aa	bb	aa	bb
32	cc	bb	aa	bb	++	aa	cc	aa	aa
33	ee	bb	aa	bb	--	aa	cc	aa	aa
34	aa	bb	aa	bb	++	aa	bb	aa	aa
35	bb	cc	aa	bb	++	aa	bb	aa	aa
36	cc	bb	aa	bb	--	aa	dd	aa	bb
37	bb	aa	bb	bb	--	aa	bb	aa	aa
38	cc	cc	bb	bb	++	aa	bb	aa	aa
(26x?)	cc	bb	aa	bb	++	aa	cd	aa	aa
(18x34)	ac	bb	ab	bb	++	aa	bb	aa	aa

[a] + and - denote dominant and recessive (null) alleles, respectively.

[b] MLG: Multilocus genotype.

Table 18. Natural plants in each multilocus genotype.

MLG[a]	Pop	Natural plants	No. of plants
5	A	KA28	1
7	A	KA20	1
8	A	KA18	1
9	A	KA11	1
10	A	KA9, KA12	2
11	A	KA13, KA16, KA19, KA23, KA25, KA26, KA27	7
12	A	KA14	1
22	A	KA2, KA4, KA6, KA7	4
23	A	KA1	1
24	A	KA8, KA10, KA15, KA21, KA22	5
27	A	KA3	1
31	A	KA5	1
(11x12)	A	KA17 (F_1)	1
6	B	KB28	1
21	B	KB10, KB11, KB12, KB13, KB14, KB15, KB16, KB17, KB18, KB19, KB20, KB21, KB22, KB23	14
25	B	KB1, KB2, KB3, KB4, KB5, KB6, KB7, KB9, KB24, KB25, KB26, KB27, KB29, KB30	14
35	B	KB8	1
26	C	KC5, KC6, KC7, KC8, KC10, KC12, KC14, KC15, KC16, KC17, KC18, KC19, KC21, KC22, KC25	15
29	C	KC29	1
30	C	KC26, KC28, KC30	3
32	C	KC2, KC3, KC20, KC23, KC24, KC27	6
33	C	KC1	1
36	C	KC13	1
(26x30)#1	C	KC9 ($F_{>1}$)	1
(26x30)#2	C	KC4 ($F_{>1}$)	1
(26x?)	C	KC11 ($F_?$)	1
1	D	KD1, KD2, KD13, KD14	4
16	D	KD16, KD17, KD18, KD19	4
28	D	KD3, KD4, KD5, KD6, KD7, KD8, KD9, KD10, KD11, KD12	10
14	E	KE9, KE10, KE11	3
17	E	KE16, KE19, KE20, KE27	4
18	E	KE22, KE25, KE33, KE34, KE35, KE36, KE37, KE39	8
20	E	KE29, KE30, KE41	3
34	E	KE12, KE13, KE14, KE15, KE17, KE18, KE21, KE23, KE24, KE26, KE28, KE32, KE38, KE40	14
38	E	KE1, KE2, KE3, KE4, KE5, KE6, KE7, KE8	8
(18x34)	E	KE31 (F_1)	1
2	F	KF1, KF2, KF3, KF4, KF20, KF26	6
3	F	KF13, KF14, KF15, KF16, KF17	5
4	F	KF7, KF8, KF9, KF10, KF11, KF12, KF21, KF23	8

Table 18. Natural plants in each multilocus genotype (continued).

MLG[a]	Pop	Natural plants	No. of plants
13	F	KF18, KF25	2
15	F	KF6	1
19	F	KF24	1
37	F	KF5, KF19, KF22	3

[a] MLG: Multilocus genotype.

Aco3 was examined. This classification process was carried out until all the natural plants were classified into multilocus genotypes. A total of 38 multilocus genotypes were obtained for the 172 natural plants except for five heterozygotes (Table 18).

Parentage for the five heterozygotes in the 172 plants was also inferred (Table 18). Three criteria were used. First, the two parental multilocus genotypes must have had different genotypes at the heterozygous loci so that the heterozygotes could be produced. Second, the two parental multilocus genotypes together must have been able to provide all the alleles at the homozygous loci of the heterozygotes. Third, the parental multilocus genotypes must have existed in the same population, since cross pollination between populations is highly improbable.

KA17 was heterozygous at Pgd2 locus. Careful examination according to the above criteria showed that its parental multilocus genotypes must have been 11 and 12. Any other two multilocus genotypes would have given rise to heterozygosities at other loci. If it had been derived from an outcrossing between other multilocus genotypes and fixation had occurred so that only one locus was heterozygous, there would have more heterozygous plants fixed at other loci in population A, which was not the case.

This plant was also probably F_1. Since the probabilities for one locus to be heterozygous F_1: 1.0, F_2: 0.5,

F_3: 0.25 and so on, the total probability of this plant to be F_1 was 50% compared with 25% to be F_2. Another reasoning would be again that there would have been more heterozygous plants in population A if KA17 had been F_2 or a higher generation. One cross pollination event would have produced many heterozygotes if the F_1 heterozygote had survived.

For the three heterozygotes KC4, KC9 and KC11 in population C, it is possible that a single cross pollination occurred between multilocus genotype 26 and another unknown or not sampled multilocus genotype several years ago. Because they were heterozygous at different loci, fixation probably had occurred during inbreeding. Since multilocus genotype 30 could also produce KC9 and KC4 heterozygotes with multilocus 26, for later calculation purposes, multilocus genotypes 26 and 30 were inferred as parents for KC9 and KC4. KC11 was included as a half individual in the multilocus genotype analysis because the parentage of the other half was unknown. In this study, only one natural seed from KC11 plant was found to be heterozygous, since the greenhouse plant derived from this seed was heterozygous for Ti, at which KC11 itself was homozygous.

KE31, which was heterozygous at seven loci, must have been derived from a hybridization between multilocus genotypes 18 and 34, since only these two could produce a heterozygous plant with the genotypes exactly the same as KE31.

It also must have been F_1. The probability for it to be F_2 would have been 0.5^7. The probabilities for it to be a higher generation were even more minute.

There were a total of 5 heterozygous plants in the 172 natural plants of the six populations. The outcrossing rates would be 5/172=0.029. However, if the time each heterozygote was produced was taken into account, it seemed that the outcrossing rate of wild soybean could not be determined through such a study.

No single multilocus genotype had plants from more than one population, i.e. no multilocus genotypes existed in more than one population (Table 18). In each population, there were always two to three dominant multilocus genotypes. For example in population A, although there were 12 multilocus genotypes, only 11, 22 and 24 were dominant if the 27 plants represented a random sample of that population. At least one parent of the heterozygotes was the dominant multilocus genotype.

The numbers of loci different in all comparisons of the 38 multilocus genotypes were tabulated in Table 19. They varied from one to the maximum of 14 for the 27 polymorphic loci in the total number of 35 examined. The average numbers of loci different for each multilocus genotype with the multilocus genotypes in the same population and in another population were also calculated (Table 20). Some multilocus genotypes had a closer relation to the multilocus genotypes

Table 19. Number of loci different between multilocus genotypes.

Pop	MLG[a]	N[b]	A 5 1	A 7 1	A 8 1	A 9 1	A 10 2	A 11 7½	A 12 1½	A 22 4	A 23 1	A 24 5	A 27 1	A 31 1
A	7	1	3											
A	8	1	6	3										
A	9	1	4	6	3									
A	10	2	4	6	5	2								
A	11	7½	4	3	2	3	3							
A	12	1½	5	4	3	4	4	1						
A	22	4	8	10	9	6	6	9	10					
A	23	1	9	11	10	7	9	10	9	3				
A	24	5	4	4	5	4	6	7	8	6	7			
A	27	1	5	7	10	7	7	8	9	7	8	5		
A	31	1	11	12	13	11	13	14	13	7	4	8	6	
B	6	1	4	5	8	6	6	7	8	10	9	5	7	10
B	21	14	7	7	8	7	7	6	7	7	8	7	5	11
B	25	14	7	6	7	7	9	8	9	9	10	4	6	9
B	35	1	5	5	8	7	7	7	8	9	10	4	4	8
C	26	16½	5	8	11	9	9	9	10	7	6	7	8	8
C	29	1	6	9	10	8	10	10	11	10	11	6	9	13
C	30	4	6	9	10	8	10	10	11	10	11	6	9	13
C	32	6	7	9	10	9	11	11	12	11	12	6	10	12
C	33	1	7	8	7	6	8	8	9	10	11	5	9	11
C	36	1	8	10	11	10	10	9	11	14	13	9	11	13
D	1	4	9	9	10	10	12	11	10	12	11	7	11	11
D	16	4	4	7	8	6	8	8	9	6	7	4	5	9
D	28	10	3	4	7	6	6	5	6	6	7	4	3	9
E	14	3	8	8	9	9	9	8	7	11	9	8	8	10
E	17	4	6	8	11	10	10	10	9	10	9	7	7	9
E	18	8½	5	7	10	9	9	9	8	11	10	6	8	10
E	20	3	10	11	12	10	10	11	10	12	11	9	7	9
E	34	14½	6	6	9	9	9	8	9	9	10	6	6	10
E	38	8	4	5	8	8	8	7	6	10	9	5	7	9
F	2	6	7	7	8	7	7	7	7	11	12	6	8	12
F	3	5	8	8	9	8	8	8	9	13	14	8	8	12
F	4	8	3	6	5	3	3	3	4	9	10	7	8	14
F	13	2	8	10	13	10	10	11	10	12	11	8	9	13
F	15	1	5	7	6	5	5	5	6	9	10	6	8	12
F	19	1	7	7	10	9	9	9	10	11	12	6	8	12
F	37	3	7	6	7	7	7	6	5	11	10	6	8	10

Table 19. Number of loci different between multilocus genotypes (continued).

Pop	MLG[a]	N[b]	B 6 1	B 21 14	B 25 14	B 35 1	C 26 16½	C 29 1	C 30 4	C 32 5	C 33 1	C 36 1	D 1 4	D 16 4
B	21	14	7											
B	25	14	7	7										
B	35	1	6	7	6									
C	26	16½	7	10	10	8								
C	29	1	8	9	9	9	7							
C	30	4	8	9	9	9	8	1						
C	32	6	8	10	9	8	8	1	2					
C	33	1	7	9	8	7	8	5	6	4				
C	36	1	7	11	10	9	8	7	6	6	6			
D	1	4	11	13	10	9	12	11	11	10	9	13		
D	16	4	8	7	7	7	7	6	6	7	7	10	7	
D	28	10	6	5	7	5	6	7	7	8	7	9	8	3
E	14	3	10	10	11	6	11	12	12	11	10	12	7	8
E	17	4	9	9	10	7	9	10	10	9	9	10	7	4
E	18	8½	8	10	9	6	8	9	9	8	8	9	6	5
E	20	3	11	11	12	7	11	10	11	9	8	11	12	10
E	34	14½	8	8	7	6	9	8	8	7	8	6	9	6
E	38	8	7	8	7	3	7	8	8	7	7	8	7	6
F	2	6	8	9	8	4	10	11	11	10	7	9	7	7
F	3	5	9	11	10	4	12	11	11	10	9	11	13	11
F	4	8	7	8	10	8	8	9	9	10	8	9	10	5
F	13	2	10	11	10	9	9	10	11	11	10	12	10	8
F	15	1	8	8	9	6	8	9	9	8	6	7	8	5
F	19	1	8	9	8	4	8	9	10	8	7	10	9	9
F	37	3	8	8	7	5	10	11	11	10	7	9	9	9

Table 19. Number of loci different between multilocus genotypes (continued).

Pop	MLG[a]	N[b]	D 28 10	E 14 3	E 17 4	E 18 8½	E 20 3	E 34 14½	E 38 8	F 2 6	F 3 5	F 4 8	F 13 2	F 15 1	F 19 1
E	14	3	7												
E	17	4	5	4											
E	18	8½	6	3	1										
E	20	3	8	9	8	9									
E	34	14½	3	8	6	7	9								
E	38	8	5	6	4	3	8	6							
F	2	6	7	8	7	6	9	8	5						
F	3	5	9	8	9	8	7	10	7	6					
F	4	8	6	7	7	6	11	9	7	6	9				
F	13	2	8	7	6	5	9	10	8	9	11	9			
F	15	1	6	5	5	4	9	7	5	4	8	2	9		
F	19	1	7	10	9	8	8	8	5	6	8	10	7	8	
F	37	3	7	8	7	6	7	8	4	5	7	8	9	6	7

[a] MLG: Multilocus genotypes.
[b] Number of plants. Heterozygotes were counted half towards each of their parental multilocus genotypes.

Table 20. Average number of loci different between each multilocus genotype and the multilocus genotypes in the same or in another population.

MLG[a]	Pop	N[b]	Within pop	Between population						
				A	B	C	D	E	F	Mean
5	A	1	5.73		5.75	6.50	5.33	6.50	6.43	6.23
7	A	1	6.27		5.75	8.83	6.67	7.50	7.29	7.38
8	A	1	6.27		7.75	9.83	8.33	9.83	8.29	8.92
9	A	1	5.18		6.75	8.33	7.33	9.17	7.00	7.81
10	A	2	5.91		7.25	9.67	8.67	9.17	7.00	8.35
11	A	7.5	5.82		7.00	9.50	8.00	8.83	7.00	8.12
12	A	1.5	6.36		8.00	10.67	8.33	8.17	7.29	8.50
22	A	4	7.36		8.75	10.33	8.00	10.50	10.86	10.00
23	A	1	7.91		9.25	10.67	8.33	9.67	11.29	10.12
24	A	5	5.82		5.00	6.50	5.00	6.83	6.71	6.23
27	A	1	7.18		5.50	9.33	6.33	7.17	8.14	7.58
31	A	1	10.18		9.50	11.67	9.67	9.50	12.14	10.73
6	B	1	6.67	7.08		7.50	8.33	8.83	8.29	7.82
21	B	14	7.00	7.25		9.67	8.33	9.33	9.14	8.53
25	B	14	6.67	7.58		9.17	8.00	9.33	8.86	8.47
35	B	1	6.33	6.83		8.33	7.00	5.83	5.71	6.71
26	C	16.5	7.80	8.08	8.75		8.33	9.17	9.28	8.66
29	C	1	4.20	9.42	8.75		8.00	9.50	10.00	9.34
30	C	4	4.60	9.42	8.75		8.00	9.67	10.29	9.44
32	C	6	4.20	10.00	8.75		8.33	8.50	9.57	9.31
33	C	1	5.80	8.25	7.75		7.67	8.33	7.71	8.03
36	C	1	6.60	10.75	9.25		10.67	9.33	9.57	10.03
1	D	4	7.50	10.25	10.75	11.00		8.00	9.43	9.89
16	D	4	5.00	6.75	7.25	7.17		6.50	7.71	7.03
28	D	10	5.50	5.50	5.75	7.33		5.67	7.14	6.20
14	E	3	6.00	8.67	9.25	11.33	7.33		7.57	8.88
17	E	4	4.60	8.83	8.75	9.50	5.33		7.14	8.25
18	E	8.5	4.60	8.50	8.25	8.50	5.67		6.14	7.69
20	E	3	8.60	10.17	10.25	10.00	10.00		8.57	9.78
34	E	14.5	7.20	8.08	7.25	7.67	6.00		8.57	7.81
38	E	8	5.40	7.17	6.25	7.50	6.00		5.86	6.72
2	F	6	6.00	8.25	7.25	9.67	7.00	7.17		8.06
3	F	5	8.17	9.42	8.50	10.67	11.00	8.17		9.45
4	F	8	7.33	6.25	8.25	8.83	7.00	7.83		7.39
13	F	2	9.00	10.42	10.00	10.50	8.67	7.50		9.65
15	F	1	6.17	7.00	7.75	7.83	6.33	5.83		6.97
19	F	1	7.67	9.17	7.25	8.67	8.33	8.00		8.52
37	F	3	7.00	7.50	7.00	9.67	8.33	6.67		7.77

[a] MLG: Multilocus genotype.
[b] Number of plants. Heterozygotes were counted half toward each of their parental multilocus genotypes.

in another population than to those in the same population. For instance, multilocus genotype 24 in population A differed, on the average, by 5.82 loci from the multilocus genotypes in its own population, but only by 5.00 loci from the multilocus genotypes in populations B and D. However, the average number of loci different between each multilocus genotype and other multilocus genotypes in the same population was always smaller than the average number of loci different between that multilocus genotype and the multilocus genotypes in all other populations (Table 20).

The average number of loci different between each multilocus genotype and the multilocus genotypes in the same population had a significant correlation of 0.452 with the average number of loci different between that multilocus genotype and the multilocus genotypes in all other populations. Regression of the latter on the former showed R^2=0.204, which was significant at the 99% probability level (Fig. 6). Thus, the average number of loci different between each multilocus genotype and other multilocus genotypes within a population could predict, to a certain extent, the average number of loci different between this multilocus genotype and those in other populations.

The average numbers of loci different between multilocus genotypes within and between populations are presented in Table 21. All the numbers in Table 21, including the weighted means and grand mean, were calculated directly

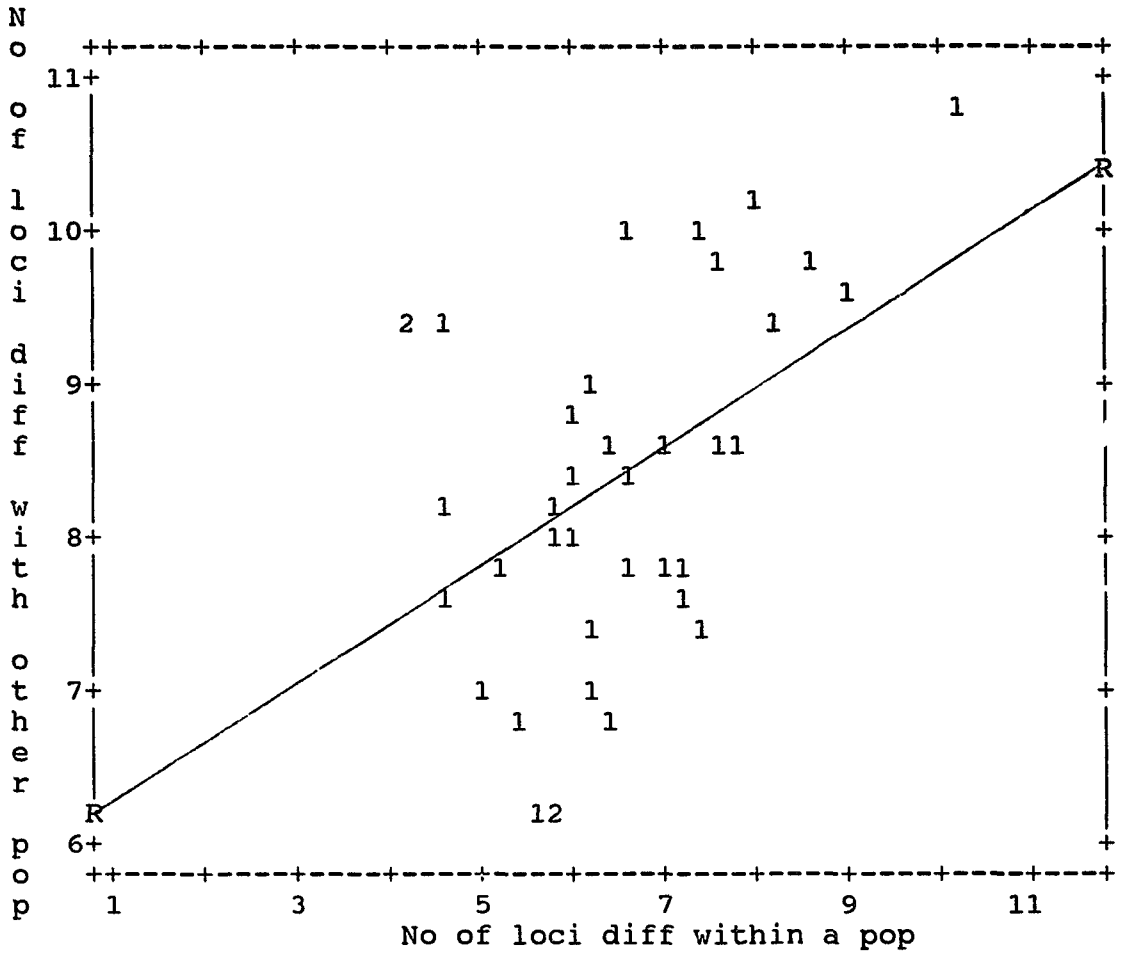


Fig. 6. Plot of the mean numbers of loci different of each multilocus genotype with the multilocus genotypes in other populations and in the same population. Data points in the plot represent the numbers of multilocus genotypes. Regression: $Y=5.73'+0.401X$, $R^2= 0.204$ significant at the 99% probability level.

Table 21. Average number of isozyme loci and average percentage (%) of genome different between multilocus genotypes within and between populations.

Pop	Number of MLG[a]	Within pop	Between populations					
			B	C	D	E	F	Mean
A	12	6.67	7.19	9.32	7.50	8.57	8.29	8.33
		19.05	20.54	26.63	21.43	24.48	23.67	23.80
B	4	6.67		8.67	7.92	8.33	8.00	7.88
		19.05		24.76	22.62	23.81	22.86	22.52
C	6	5.53			8.50	9.08	9.40	9.14
		15.81			24.29	22.95	26.87	26.10
D	3	6.00				6.72	8.10	7.70
		17.14				19.21	23.13	22.01
E	6	6.07					7.31	8.19
		17.33					20.88	23.59
F	7	7.33						8.26
		20.95						23.59
Mean	6.33	6.56						8.32
		18.73						23.76

[a] MLG: Multilocus genotypes.

from Table 19. A multilocus genotype in population A, for example, was, on the average, 6.67 loci different from other multilocus genotypes in population A, 7.19 loci different from the multilocus genotypes in population B, and 8.33 loci different from the multilocus genotypes in all other populations. Populations F and C had the largest (7.33) and smallest (5.53) numbers of loci different between multilocus genotypes within populations, respectively. Populations C and D had the largest (9.14) and the smallest (7.70) numbers of loci different, respectively, when the multilocus genotypes in the two populations were compared with those in all other populations. On the average, multilocus genotypes differed by 6.56 loci within populations and by 8.32 loci between populations.

The numbers of loci different between multilocus genotypes could be divided by 35, the total number of loci assayed to get the percentages of isozymes loci assayed different. They could also be interpreted as the percentage of genome different if these 35 loci represented a random sample of the wild soybean genome. The overall within- and between-population multilocus genotypic differences were 18.73% and 23.76% of the genome, respectively.

From Table 21, it is apparent that the numbers of loci different could be used as measures of genetic variation and genetic distance of populations. But, basing the comparisons on multilocus genotypes was not appropriate, be-

cause the number of plants in each multilocus genotype was not the same. Therefore, the number of loci different between multilocus genotypes had to be converted to the number loci different between individuals. The following formulae were developed:

$$L_X = \frac{\sum_{i \neq j} l_{ij} n_i n_j}{\frac{N_X(N_X-1)}{2}} \qquad (1)$$

$$L_{XY} = \frac{\sum l_{xy} n_x n_y}{N_X N_Y} \qquad (2)$$

L_X is the average number of loci different between individuals in population X. l_{ij} is the number of loci different between the ith and jth multilocus genotypes in population X. n_i and n_j are the numbers of plants in the ith and jth multilocus genotypes, respectively. N_X is the sample size of population X. L_{XY} is the average number of loci different between individuals of population X and individuals of population Y. l_{xy} is the number of loci different between the xth multilocus genotype in population X and the yth multilocus genotype in population Y. n_x and n_y are the numbers of plants in the xth and yth multilocus genotypes, respectively. N_X and N_Y are the sample sizes of populations X and Y, respectively. The number of plants for each multilocus genotype is given in Table 19. Population sample sizes are listed in Table 22.

For example, the average number of loci different between individuals in population D and the average number of loci different between individuals of population B and individuals of D were calculated as follows:

$$L_D = \frac{7x4x4+8x4x10+3x4x10}{\frac{18(18-1)}{2}} = 3.61,$$

$$L_{BD} = \frac{11x1x4+13x14x4\cdots5x1x10}{30x18} = 7.41.$$

In order to show the mean number of loci different between individuals of one population from those in all other populations, the grand means of the numbers of loci different between individuals within and between populations were calculated, let

$$D_X = \sum_{i \neq j} l_{ij} n_i n_j,$$

$$C_X = \frac{N_X(N_X-1)}{2},$$

$$D_{XY} = \sum l_{xy} n_x n_y,$$

$$C_{XY} = N_X N_Y.$$

D_X and D_{XY} are the sums of products of the numbers of loci different between multilocus genotypes and the numbers of comparisons they represented within population X and between populations X and Y, respectively. C_X and C_{XY} are the total numbers of non-redundant individual-to-individual compari-

sons within population X and between populations X and Y, respectively. Eq. 1 and 2 can be rewritten as:

$$L_X = \frac{D_X}{C_X} \qquad (3)$$

$$L_{XY} = \frac{D_{XY}}{C_{XY}} \qquad (4)$$

The overall mean number of loci difference between individuals within population X is obtained by:

$$\bar{L}_X = \frac{\Sigma\, D_X}{\Sigma\, C_X} \qquad (5)$$

For between population comparisons,

$$\bar{L}_{XY} = \frac{\Sigma_{X \neq Y}\, D_{XY}}{\Sigma_{X \neq Y}\, D_{XY}} \qquad (6)$$

When X=I, and Y varies from A to F, this formula gives the mean number of loci different between individuals of Ith population and individuals of all other populations. When X and Y both vary from A to F, it gives the grand mean of the number of loci different between individuals for between-population comparisons.

The relationships among populations in terms of the within-population genetic variation expressed by the average number of loci different corresponded by rank exactly to those by the expected heterozygosity (Tables 13 and 22). Populations F and D had the highest and lowest expected

Table 22. Average number of isozyme loci and average percentage (%) of genome different between individuals within and between populations.

Pop	Number of plants	Within pop	Between populations					
			B	C	D	E	F	Mean
A	27	5.65	7.11	8.70	6.66	8.23	7.57	7.78
		16.14	20.32	24.86	19.02	23.52	21.62	22.23
B	30	4.04		9.57	7.41	8.51	8.93	8.40
		11.53		27.33	21.17	24.31	25.52	24.00
C	29.5	4.48			7.79	8.52	9.68	8.90
		12.80			22.27	24.34	27.66	25.42
D	18	3.61				5.79	7.76	6.98
		10.31				16.55	22.16	19.94
E	41	4.73					7.54	7.89
		13.52					21.54	22.53
F	26	5.99						8.29
		17.11						23.69
Mean	28.58	4.79						8.09
		13.69						23.12

heterozygosities, respectively. They also had the largest (5.99) and smallest (3.61) average numbers of loci different between individuals within populations, respectively. There was also some degree of correlations between Nei's genetic distance and the number of loci different between individuals for between-population comparisons. For example, population C had both the largest mean Nei's genetic distance and the largest mean number (8.90) of loci different between individuals in comparisons with all other populations. The mean number of loci different between individuals in comparisons with all other populations showed that population D (6.98) was the most indistinguishable from other populations. Population D also had a small mean Nei's genetic distance (0.99).

The overall mean number of loci and the percentage of genome different between individuals within populations were 4.79 and 13.69%, respectively, compared with those of 8.09 and 23.12%, respectively when the comparisons were made between populations. Therefore, nearly twice as much genetic difference existed between individuals of different populations compared with that in the same population.

Cluster Analysis.

Three separate cluster analyses were performed for the six populations using Nei's genetic distance, the squared Euclidean distance (Appendix II) and the number of loci different between individuals of different populations (Fig.

7). In all three sets of clusters, population C was the most distinct population. Populations D and E always clustered together. Populations A and B clustered together, then clustering with population F in both cluster analyses using the squared Euclidean distance and the number of loci different between individuals. In the analysis using Nei's genetic distance, population F clustered with population A.

Cluster analysis was also performed on the 38 multilocus genotypes using the numbers of loci different (Fig. 8). There was no guarantee that multilocus genotypes within a population always clustered together. For example, multilocus genotype 1 from population D clustered with those of populations E and F rather than with multilocus genotypes 16 and 28 in population D.

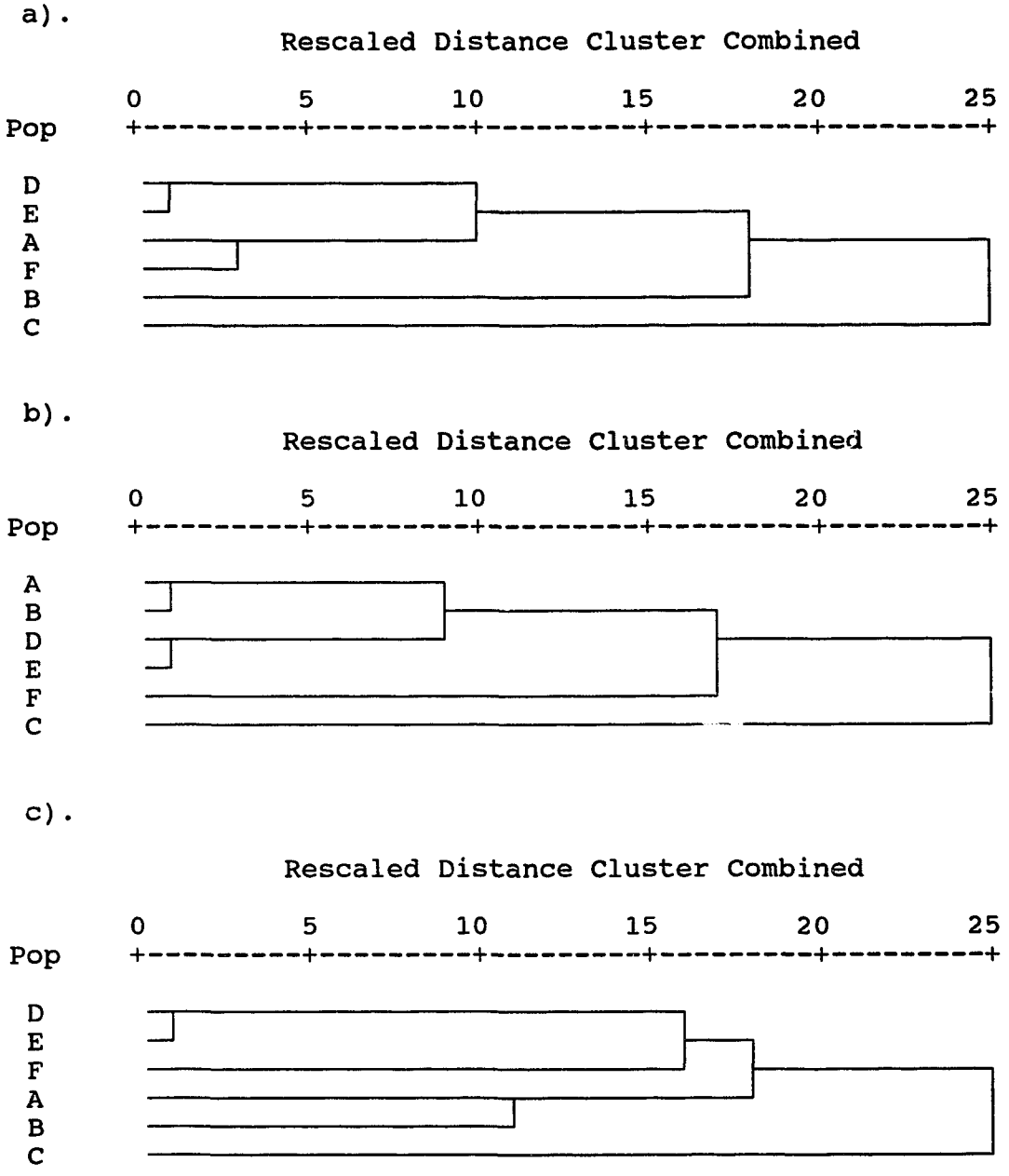


Fig. 7. Dendrograms obtained from cluster analysis for the six populations using the average linkage between groups method (UPGMA) based on a) Nei's unbiased genetic distances, b) the squared Euclidean distance, c) the number of loci different between individuals of different populations for isozyme data.

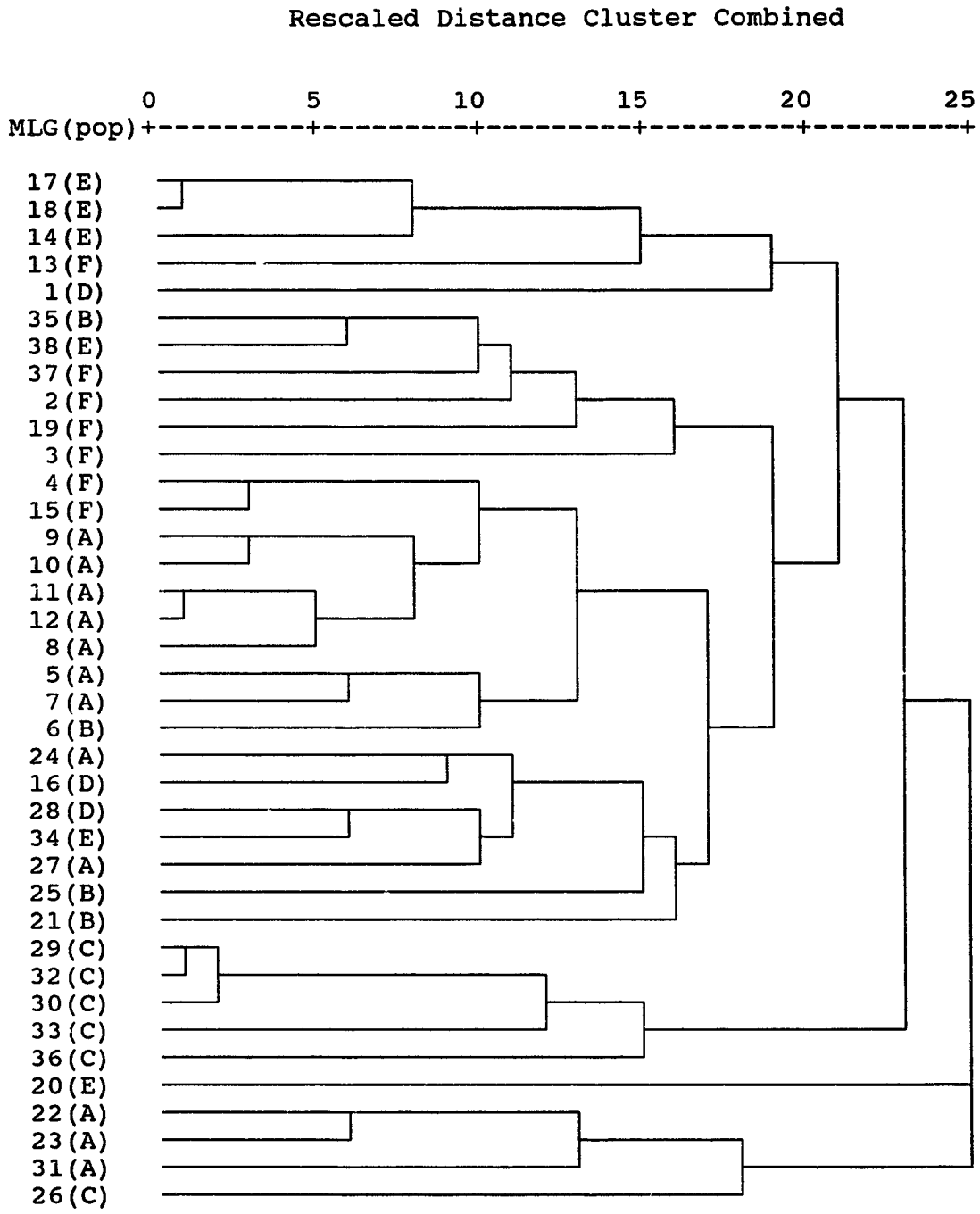


Fig. 8. Dendrogram obtained from cluster analysis for the 38 multilocus genotypes (MLG) using the average linkage between groups method (UPGMA) based on the number of loci different between multilocus genotypes.

DISCUSSION

Genetic Variation of Populations.

The most used measures for comparing different species and populations regarding the amount of genetic variation are the mean number of alleles per locus (A) and no-criterion polymorphism (P). Hamrick et al. (1979) reported that A=1.69, P=0.368, averaged over 113 taxa of plants, among which the 33 primarily selfed species have the average A of 1.27 and P of 0.179. In natural populations of 15 plant species, P=0.259 per population (Nevo, 1978). In this study, A=1.4 and 2.1, and P=0.372 and 0.771 per population and in the total population, respectively. Therefore, the genetic variation in these populations was high.

In soybean, Kiang et al. (1987) summarized the results of several studies on genetic variation. Although all these studies used basically the same kinds of enzymes, the numbers of loci included in the analyses were different. In previous studies, those loci hypothesized and not genetically confirmed were included in the analyses. Since the number of loci controlling the enzymes is not known until their inheritance is studied, only the genetically-studied loci were included in analysis of the present study. In order to make more meaningful comparisons, the results of this study were revised. Information in Tables 3 and 4 of

Kiang et al. (1987), the results of seven local natural populations of Mishima, Japan (Bult, 1989), and the revised results of this study based on the total locus numbers of 49 for populations means and 46 for the total population were combined in Table 23.

The genetic variation in these six South Korean natural populations of wild soybean was much higher than that based on 857 accessions of cultivated soybean (Table 23). This result was in agreement with the previous finding that G. soja had higher genetic variation than G. max (Kiang et al., 1987). When the results of the study based on 66 wild soybean accessions from South Korea and the present study based on 172 natural plants were compared, the genetic variations in these two total populations were about the same (Table 23). This agreement reflected the same geographic origin of these two total populations and the consistency of isozyme studies. These six populations of wild soybean also had much higher genetic variation than G. soja populations from other geographic areas (Table 23). Hymowitz and Kaizuma (1979 and 1981) and Kiang et al. (1987) reported that South Korean cultivated soybean also had higher genetic variation. The higher genetic variation in South Korean soybean populations is suspected to be caused by the favorable peninsular climate pattern of South Korean geography.

It should be noted, however, that it is difficult to compare the results of different studies. The geographic

Table 23. Comparisons of these six populations with G. max and other wild soybean populations in terms of genetic variation.

Population	Size[a]	No.of loci	No.of alleles per locus		Polymorphism (99%)		H_{exp}[b]	
			Mean	Total	Mean	Total	Mean	Total
G. max	857a	46	---	1.43	---	0.457	---	0.140
China	21a	46	---	1.37	---	0.237	---	0.110
Japan	41a	46	---	1.58	---	0.467	---	0.168
S. Korea-1[c]	66a	46	---	1.93	---	0.578	---	0.149
USSR	20a	46	---	1.21	---	0.158	---	0.057
M. Japan[d]	111p	49	1.14	---	0.140	---	0.046	---
S. Korea-2[e]	172p	--[f]	1.29	1.83	0.266	0.587	0.094	0.164

[a] "a": denoting accessions from USDA Soybean Germplasm Collection;
"p": denoting natural plants.
[b] Nei's biased estimate of the expected heterozygosity based on the Hardy-Weinberg equilibrium.
[c] South Korea population based on 66 accessions.
[d] Seven local natural populations from Mishima City, Japan (Bult, 1989).
[e] The present study.
[f] 49 loci for the population means and 46 loci for the total population for all genetic diversity measures.

range, the sample size for populations, the numbers of isozyme loci examined and used in analysis, the number of populations, the nature of populations (accessions from the USDA Soybean Germplasm Collection versus natural populations) and the measures of genetic variation (for example, only population means were reported by Bult (1989), whereas the genetic variation was reported as in the total populations in all other studies) were all different. Although South Korea seems to have higher genetic variation in cultivated soybean than other regions, the conclusion that South Korea also has higher genetic variation for wild soybean should not be made now, pending the studies of populations from other geographic origins with larger sample sizes, especially those from peninsular regions of Northeast China.

Genetic variation of populations was not only measured by the mean number of alleles per locus, polymorphism and the expected heterozygosity, but also by the number of loci different between multilocus genotypes and between individuals within populations in this study.

These six populations were basically along the 127°E longitudinal line. Examination of the above parameters did not reveal any latitudinal influence on genetic variation in any direction. However, in comparing these six populations, there seemed to be a slight increase in genetic variation in two directions, north from population D to population B to population A, and south from population D to population E

to population F, with population C as an exception (Tables 13, 21 and 22).

Genetic Differentiation of Populations.

Genetic differentiation of populations can be measured by gene differentiation (the proportion of gene diversity among populations in the total gene diversity, denoted by G_{ST}), Nei's genetic distance, and in this study, by the number of loci different between multilocus genotypes and between individuals of different populations. While gene differentiation gives a general idea about how populations are differentiated, Nei's genetic distance is actually a measure of population-to-population relationships in terms of divergence from each other. Hartl (1988) considered little differentiation for populations with F_{ST} (the fixation index, exactly the same as G_{ST}) of 0 to 0.05, moderate differentiation for 0.05 to 0.15 F_{ST}, great differentiation for 0.15 to 0.25 F_{ST}, and very great differentiation for F_{ST} more than 0.25. Gene differentiation is the highest for selfing among all mating systems, and G_{ST}=0.560 for 31 selfing annuals (Loveless and Hamrick, 1984).

In wild soybean, 0.198 for G_{ST} and 0.044 for average Nei's genetic distance in four natural populations along the Kitakami River of Japan (Chiang, 1985), and 0.063 for average Nei's genetic distance in seven local natural populations in Mishima City of Japan were reported (Bult, 1989).

In this study, G_{ST} and average Nei's genetic distance were 0.383 and 0.117, respectively. Therefore, these six populations were well differentiated, perhaps also reflecting a larger geographic area for these six populations than for those along the Kitakami River and in Mishima City of Japan.

Although only 38.3% of the total gene diversity (G_{ST}=0.383) existed among populations, individuals of different populations were, on the average, 8.09 loci or 23.12% of the genome different (Table 22). The average multilocus genotypic difference between populations was even higher, 8.32 loci or 23.76% of the genome (Table 21).

As for the relationships among these populations with respect to the average divergence of each population from the total population, the general trend of divergence seemed to increase in two directions, north from population D to population B and population A, and south from population D to population E and population F, with population C as an exception (Tables 16, 21 and 22). This trend agreed with the changes of within-population genetic variation. The only discrepancy was population A, which had smaller mean Nei's genetic distance (0.93) and mean number (7.78) loci different between individuals of different populations than population B (Tables 16, 21 and 22).

Relationship between Genetic Variation and Genetic divergence.

The correlation between genetic variation and genetic

divergence might be an important phenomenon and warranted further examination. The least distinct population among the six was population D, as indicated by the numbers of loci different between multilocus genotypes (Table 21) and between individuals (Table 22) of different populations, and to a lesser extent, by the mean Nei's genetic distance (Table 16). Population D also had the lowest genetic variation, as determined by the mean number of alleles per locus, polymorphism, the expected heterozygosity and the number of loci different between individuals within populations (Tables 13 and 22). On the other hand, population C had higher genetic variation than population D, as measured by polymorphism, the expected heterozygosity and the number of loci different between individuals of different populations (Tables 13 and 22).

Genetic variation should also be determined by dynamics of the within-population genetic variation. In population C, the number of loci between multilocus genotypes varied from one to eight, which might indicate active progress of the gene recombination process in population C. In population D, the numbers of loci different between multilocus genotypes were six to seven. The observed outcrossing rates were different. In population C, three heterozygotes were found compared with none in D, and none or one in other populations. In this study, only one natural seed was found to be

a F_1 heterozygote and happened to be on the natural plant KC11 in population C. The observed heterozygosity is important because it ultimately determines whether the gene recombination of individuals in a population can occur or not. All these factors might be important in explaining the most divergent nature of population C among these six populations.

In brief, the within-population genetic variation and dynamics of the within-population genetic variation seemed to determine the divergence or the potential to diverge of a population from other populations.

The amount of genetic variation in the founding population, the rates of mutation and migration and the rates of cross-pollination may determine the within-population genetic variation of a population. The small amount of outcrossing in wild soybean may be different from population to population, depending on the climate of geographic region of the population. The rationale for this reasoning is that in the circumstances of artificial hybridization, wind and dry weather can reduce the success rates of hybridization. Those populations with high outcrossing rates recombine genes quickly. Some genes are saved by the recombination process from elimination by natural selection when they are present in less fit gene combinations. Therefore, these populations maintain higher within-population genetic variation. During the process of gene recombination, many

new gene combinations (multilocus genotypes) are created. Those gene combinations that are more competitive under natural selection may be different, not only from those in the same population, but also from those in other populations. This relationship was demonstrated by the significant regression of the average number of loci different for each multilocus genotype with those in other populations on the average number of loci different for that multilocus genotype with those in the same population (Fig. 6). It is the accumulation of those different new gene combinations that are presumably more fit and elimination of old less fit gene combinations by natural selection that make one population different from other populations. Genetic drift may make one mosaic population look like one of its components if no gene recombination occurs, but probably will not make it divergent. Therefore, the outcrossing rate may be a very important factor in maintaining the within-population genetic variation and creating new gene combinations, and may ultimately decide the fate of a population of highly self-pollinated plant species. In highly outcrossing species, outcrossing rates may not be a limiting factor of gene recombination.

Multilocus Association and Natural Selection.

Multilocus association is widespread, especially in inbreeding plant species (Brown, 1979). Clegg et al. (1972)

studied the gametic phase disequilibrium (linkage disequilibrium) in two experimental populations of barley composite crosses and found that linkage disequilibrium increased with time of selfing. Only a few complementary multilocus genotypes in the gametes at three linked and one unlinked esterase loci were favored in the advanced generations. In this study, the sample sizes were small for the purpose of studying linkage disequilibrium. The linkage disequilibrium parameter and the deviations from the expected multilocus genotypic frequencies based on allele frequencies were not calculated. Since the number of multilocus genotypes in each population was always lower than the number of polymorphic loci, the linkage disequilibrium was presumably very high.

Just as in the artificial breeding systems, some multilocus genotypes might be homozygous for all their genes and act like pure lines, whereas others might not be stable with possible further segregation at some un-assayed loci, depending on how long ago they were created by an accidental natural outcrossing event. Those multilocus genotypes in population B might belong to the former category. Two lines of evidence supported this observation: the numbers of loci different between these multilocus genotypes, which were all six to seven, and the fact that there were no heterozygotes detected in this population, which means there had been no cross-pollination events in recent years. They might be similar to the pure lines of the artificial breeding systems

and could not be further divided genetically even when the non-isozyme loci had been examined. The multilocus genotypes 5, 6, 7, 8, 9, 10, 11 and 12 in population A, which clustered together in cluster analysis (Fig. 8), might belong to the second category. The numbers of loci different varied from one to five. In addition, there was one heterozygote detected between multilocus genotypes 11 and 12 (Tables 17 and 18). It seemed that these multilocus genotypes might have had the same parentage origin, i.e. from a single outcrossing event.

In artificial breeding systems, lines are tested as early as in F_2 and sometimes show significant differences. It might also be true in natural conditions that the multilocus genotypes show their distinct genetic features whether they are stable or segregating. Therefore, the multilocus genotypes as long as they are revealed by a sufficient large number of isozyme loci might be important basic units of population genetic structure and the evolutionary process in mosaic populations, especially the stable multilocus genotypes.

A natural selection experiment was conducted on barley by Harlan and Martini in 1938 (Briggs and Knowles, 1977. pp. 151). A mixture of an equal number of seeds of 11 varieties was grown at ten experiment stations in different parts of the United States for a number of years. The final census

showed that only a few varieties dominated the mixture at each location. Such a trend was also seen in this study. In each population, only two to three multilocus genotypes were dominant. This might have resulted from elimination of those less fit multilocus genotypes by natural selection.

The existence of multilocus genotypes may have very important implications in studying ecology and evolution. If we can map the micro-geographic distribution of multilocus genotypes in a particular population and monitor the changes of the occupying area of each multilocus genotype, we are literally watching the process of natural selection.

Migration and Dissemination.

No single multilocus genotype existed in more than one population. The lowest number of loci different between multilocus genotypes of different populations was three, which means that the plants in different populations were at least three loci different from each other for the 35 examined. The outcrossing rate in wild soybean is only 2% to 3%. If an immigrant individual had survived the new environment after it had migrated from another population, the chance for the genes in each of its germinating seeds the next season to be recombined with the genes of native individuals was very small. Therefore, there was no migration between these six populations in recent history, probably for several hundreds of years.

On the other hand, as can be seen from cluster analy-

sis, some multilocus genotypes in one population were closer to the multilocus genotypes in another population than to those in the same population, perhaps indicating that even 35 loci were not enough for distinguishing these populations, or that these populations wei mosaic themselves and related to each other during their origins.

Interesting results were obtained from a study involving four natural populations of wild soybean along the Kitakami River in Japan (Chiang, 1985). The average Nei's genetic distance for each population with all other populations increased in the direction of river flow, with the most upstream population having the lowest and the most downstream the highest Nei's genetic distances. This trend seemed to suggest that the most upstream population was most related to other populations and might have served as the source of migration. In this study, the mean genetic distance of each population to all other populations, as measured by the mean Nei's genetic distance, the numbers of loci different between multiloc is genotypes and between individuals of different populations, generally increased from population D north to popula ion B and population A, and south to population E and population F, with C population as an exception (Table 16, Tables 21 and 22). If the genetic relationships among populations can reflect migration paths and patterns, population D in this study might have been the

source population that migrated north to establish populations B and A and south to establish populations E and F. Population C probably had a different origin. The within-population genetic variation also changed in the same pattern, which implied that the secondary gene centers had higher genetic variation. This appeared to be supported by the fact that South Korea has higher genetic variation in the cultivated soybean than China where the cultivated soybean was first domesticated.

The Number of Loci Different between Multilocus Genotypes or between Individuals Serving as Measures for both Genetic Variation and Genetic Distance.

As demonstrated earlier, the number of loci different could be used as measures of the within-population genetic variation and the between-population genetic distance. In this study, the number of loci different between individuals within populations changed in the same way as the expected heterozygosity among the six populations, with population F having the highest and population D the lowest genetic variation (Tables 13 and 22). The number of loci different between multilocus genotypes also had a similar pattern of change among the six populations (Tables 13 and 21). There was also a high correlation between the number of loci different between individuals of different populations and Nei's genetic distance (0.893 for both Pearson r and Spearman r).

Compared with the expected heterozygosity and Nei's genetic distance, the number of loci different has units, makes sense biologically and is easily understood by non-population biologists. If it is divided by the total number of loci assayed, the number of loci different between individuals can be converted to the percentage of loci assayed different, which, in turn, can be interpreted as the percentage of genome different if the assayed loci represent a random sample of a genome.

Since the number of loci different between individuals is based on a large number of comparisons, statistics can be applied. For example, a standard deviation can be calculated for the number of loci different between individuals, and a t test can be performed on two sets of the numbers of loci different between individuals. Based on the results of t tests, it can be determined if one population has significantly higher genetic variation than another if the number of loci different between individuals is used as a measure of within-population variation, or if the genetic distance between two populations is significantly different from that between another two populations.

The number of loci different between individuals of different populations had better correlations than Nei's genetic distance with the squared Euclidean distance (a multivariate measure of genetic distance) for isozyme data,

geographic distance and morphological traits of these six populations (Chapter IV).

Although complicated formulae were used in this study to calculate the number of loci different between individuals, it might not be necessary with computers. The total numbers of individual-to-individual comparisons within a population and between two populations are $N(N-1)/2$ and N_1N_2, respectively. Therefore, the awkward way to calculate the numbers of loci different between individuals within a population and between two populations is to make $N(N-1)/2$ and N_1N_2 comparisons, respectively, and obtain averages for the numbers of loci different. For instance, for two populations with 50 individuals each, the total numbers of individual-to-individual comparisons within each population and between the two populations will be $50(50-1)/2=1225$ and $50^2=2500$, respectively. With aid of computers, 1225 and 2500 comparisons and the means of the numbers of loci different based on these comparisons are not difficult to calculate at all. This method can also be applied to cross-pollinated species and animals, which do not have as much multilocus association as in this study of the self-pollinated wild soybean.

CHAPTER III

MORPHOLOGICAL VARIATION AND DIFFERENTIATION

INTRODUCTION

Population genetic studies involving isozymes are useful in revealing discrete, qualitative genetic differences among populations. However, the most visible differences among populations are in the morphological characteristics. The final goal of most population genetic studies is to address these morphological population differences. In addition, many morphological traits are controlled not by single genes, but by multiple genes having small effects, which enable plants to adapt to different environments, a property termed "plasticity". Thus, "no study of population genetics is sufficient without consideration of traits influenced by multiple genetic and environmental factors" (Hartl, 1988). Genetics and morphology alone provide separate pictures about natural populations. Therefore, a study of genetic variation coupled with a study of morphological variation would provide more information about populations.

Many morphological studies of natural plant populations have focused on the effect of mating systems. The general conclusion from these studies is that outcrossers have high within-population variation and low differentiation among

populations, whereas self-fertilizers have a high degree of differentiation among populations and low within-population variation (Carey, 1982 and Wolff, 1991).

Wild soybean in its natural habitat is not disturbed by human activity. It is not dependent on humans for survival. Therefore, natural populations of wild soybean are an ideal material for studies of morphological traits as they relate to variation, differentiation, migration and natural selection. Wild soybean is a typical self-pollinated species, with a natural out-crossing rate around 2% (Chapter II and Kiang and Chiang, 1989).

A cluster analysis of soybean morphological traits indicated that Glycine max and Glycine soja are different entities, and the intermediate type or the so-called Glycine gracilis resemble the cultivated soybean rather than wild soybean (Broich and Palmer, 1980).

Most morphological traits in wild soybean are significantly different among populations, which means that most of the morphological variation exists among populations (Bult, 1989; Chiang, 1985 and Kiang and Chiang, 1989). Latitude is correlated mostly with developmental stage characteristics (Chiang, 1985 and Kiang and Chiang, 1989). Bult (1989) suggested that plants from populations with dense coverage of vegetation tend to have a vine-like growth habit, which allows individuals to climb on neighboring plants. They also tend to have fewer, larger seeds than those in less dense

vegetation. These traits ensure that individuals can compete in competitive environments.

The objectives of this study were to investigate morphological variation within and among populations, morphological differentiation of populations and to compare the morphological variation to genetic variation as determined by isozyme studies in South Korean natural populations of wild soybean.

MATERIALS AND METHODS

In their natural habitats, wild soybean plants are entangled with each other, making it difficult to measure morphological traits. Therefore, seeds collected from the natural populations were used for planting in experimental conditions so that the measurements could be made. Since the short New Hampshire growing season would not allow wild soybeans to mature in the field conditions (Chiang, 1985), "common garden approach" experiments in the greenhouse were used to study morphological variation. The experiments were carried out for two years in 1989 and 1990.

Seed Materials.

The original seeds of six natural populations of wild soybean (KA to KF) were collected from South Korea by Dr. Y. T. Kiang in 1986. The geographic locations and distances of the six populations are given in the Introduction chapter. The numbers of natural plants represented by the original seed collection were KA, 27; KB, 30; KC, 30; KD, 18; KE, 41; and KF, 26.

In 1989, 20 single plant seed sources were taken at random from each population. One seed from each of these seed sources was planted. Two seeds of KD1 and KD13 were used, since there were only 18 natural plants in population D. In 1990, 10 seeds, one each from 10 random natural

plants, were planted along with one seed from each of those natural plants of each population, except E, that were not represented in the 1989 experiment. Population E had enough natural plants so that only seeds of half natural plants were planted each year. In other populations, some of the natural plants were represented twice in the two year experiment.

Growth Conditions.

The planting dates were May 13 in 1989 and May 11 in 1990. After removing a small piece of seed coat, the seeds were inoculated with a commercial nodulating bacterium, Rhizobium japonicum, and sown in 22 cm diameter clay pots in the greenhouse. The pots contained a mixture of steam-sterilized soil and Promix (1:1, v/v). A single seed was planted at a depth of 1 cm in each pot. The pots were arranged in a completely randomized design on four benches, two rows of plants per bench.

A bamboo stake for training was placed in each pot to maintain separation with the plants from other pots. The plants received normal water supply and insect and disease control. No supplemental light was supplied and the maximum temperature in the greenhouse was set at 30°C during the day and 25°C at night in the late growing season when the weather was cold.

Traits Examined.

Based on previous studies (Bult, 1989 and Chiang,

1985), 33 and 39 characters were chosen to be examined, respectively, in 1989 and 1990. These characters could be classified into three categories, developmental stage characters, vegetative characters and reproductive characters. The symbols and names of these characters are listed below, along with the methods with which they were recorded.

A. Developmental stage characters:

S1. Vegetative growth: number of days from sowing to the first flower.

S2. Flower initiation: number of days from the first flower to the first pod (reaching 5 mm).

S3. Pod initiation: number of days from the first pod to the first seed (reaching 3 mm).

S4. Seed initiation: number of days from the first seed to the first mature pod (brown color).

S5. Seed development: number of days from the first mature pod to 95% mature pods (1990 data only).

S6. Life span: number of days from sowing to 95% mature pods (1990 data only).

B. Vegetative characters:

V7. Early plant height: height (cm) of plant at five weeks after sowing.

V8. Early node number: number of nodes at five weeks after sowing (1990 data only).

V9. Early branch number: number of branches at five

weeks after sowing.

V10. Base leaf width: average width (cm) of the terminal leaflets of two leaves taken at random from the base of the plant.

V11. Base leaf length: average length (cm) of the terminal leaflets of two leaves taken at random from the base of the plant.

V12. Base leaf shape: ratio of the base leaf length to the base leaf width (=V11/V10).

V13. Base petiole length: average length (cm) of the petioles of two leaves taken at random from the base of the plant.

V14. Leaf width: average width (cm) of the terminal leaflets of three leaves taken at random from the middle or upper parts of the plant.

V15. Leaf length: average length (cm) of the terminal leaflets of three leaves taken at random from the middle or upper parts of the plant.

V16. Leaf shape: ratio of leaf length to leaf width (=V15/V14).

V17. Petiole length: average length (cm) of the petioles of three leaves taken at random from the middle or upper parts of the plant.

V18. Lower stem width: width (mm) of the main stem at the cotyledonary node.

V19. Upper stem width: width (mm) of the main stem at 5

cm above the cotyledonary node.

V20. Node intensity: number of nodes on 5 cm of the main stem above the cotyledonary node.

V21. Final branch number: number of branches including the main stem counted at 5 cm above the cotyledonary node at harvest.

V22. Shoot dry weight: dry weight (g) of the harvested shoots.

C. Reproductive characters:

R23. Flower width: average width (mm) of the banner petals of ten random flowers (1990 data only).

R24. Flower length: average length (mm) of the banner petals of ten random flowers (1990 data only).

R25. Ovule number per ovary: average number of ovules per ovary recorded by dissecting ten random flowers under a dissecting scope (1990 data only).

R26. Flower number per inflorescence: average number of flowers on ten or more tagged inflorescences.

R27. Pod number per inflorescence: average number of mature pods set on the tagged inflorescences.

R28. Flower & pod abortion = 100 x (R26 - R27) / R26.

R29. Ovule & seed abortion: percentage of aborting ovules and seeds derived from counting empty seed sacks in the pods harvested from the tagged inflorescences.

R30. Pod number per plant: total number of pods harvested

from the plant.

R31. Percentage of 4-seed pods: number of 4-seed pods in a sample of 100 pods.

R32. Percentage of 3-seed pods: number of 3-seed pods in the sample.

R33. Percentage of 2-seed pods: number of 2-seed pods in the sample.

R34. Percentage of 1-seed pods: number of 1-seed pods in the sample.

R35. Seed number per pod = (4 x R31 + 3 x R32 + 2 x R33 + R34) / 100.

R36. Pod length: average length (cm) of three to eight 3-seed pods.

R37. Seed weight: dry weight (g) of all seeds harvested from the plant.

R38. Harvest index = R37 / (V22 + dry weight of empty pods + R37).

R39. 100 seed weight: weight (g) of a sample of 100 dry seeds.

The stage of development descriptions for the cultivated soybean (Fehr and Caviness, 1981 and Fehr et al., 1971) were used to record developmental stage characters.

At five weeks after planting, the early plant height, early node and branch numbers (V7, V8 and V9) were recorded. Flower characters (R23 to R26) were examined shortly after the first flower. Leaf characters (V10 to V17), stem and

shoot characters (V18 to V20) were examined within three weeks after the first flower. Separate measurements were made for the leaves at base of the plant, since they were observed to be different from those in the other parts of the plants. Other characters (V21, V22, R27 to R39) were recorded either at or after harvest.

Pods were harvested every other day at maturity to avoid shattering. When 95% or more pods were mature, all pods were harvested. Seeds were dried at room temperature until seed weight had stabilized. The plant was cut at 5 cm above the cotyledonary node and the final branch number was counted. The plant was then put in a brown paper bag and dried in an oven at about 50°C until a constant dry weight was achieved.

Statistical Analysis.

No data were transformed for all univariate analyses. GLM procedure of SAS (SAS Institute, 1985) was used for the analysis of variance for data from each year and also for the combined data of both years. Character means and standard errors for the combined data of both years were obtained with LSMEAN and PDIFF statements of GLM. Type 1 test was used to test the significance of effects. Character coefficients of variation (CV) were obtained with SAS MEANS procedure. Variables R29 and R31 were deleted from the CV analysis, because they had too many zeros. The CVs were further

analyzed by SAS ANOVA procedure to see if they differed significantly among populations.

Two multivariate analyses, canonical discriminant analysis with SAS CANDISC procedure and cluster analysis with CLUSTER procedure of SPSS (Norusis, 1990), were performed. First, the variables that were calculated from other variables and the variables that were not significant at 95% probability level in univariate F tests were excluded from the multivariate analyses. R31 (percentage of 4-seed pods) was also excluded, because it was totally predictable from R32, R33 and R34 (the combined frequency was 100). Thus, only 22 variables in 1989 data and 26 variables in 1990 data were used in the multivariate analyses. Normalities and outliers were checked with FREQUENCIES procedure of SPSS. Finally, transformations necessary for some variables were made according to Tabachnick and Fidell (1989. pp. 85). Outliers were transformed with RECODE procedure of SPSS in cluster analysis so that they were closer to other values.

The listwise deletions were used for missing data in canonical discriminant analysis. The scattergrams of canonical variables were generated with SAS PLOT procedure immediately following CANDISC procedure. Mahalanobis distances among populations were requested in CANDISC procedure. For cluster analysis with SPSS, the data were first converted to Z scores with DESCRIPTIVES procedure. Then, the mean Z scores of populations were calculated with AGGREGATE proce-

dure. Finally, CLUSTER procedure was applied to the six populations. There were no deletions resulting from missing data, because mean Z scores were used. The squared Euclidean distances among populations were requested in the CLUSTER procedure.

RESULTS

Univariate Analysis.

The population and year means, and F values for the 39 characters are given in Table 24. Thirty seven characters differed significantly among populations at the 95% probability levels. Only pod initiation and early node number were not significant. Of 33 characters examined in both 1989 and 1990, 23 were significant between years. The population x year interaction was significant only for 10 characters, indicating most characters were consistent across years.

Although the populations were significantly different from each other for nearly all characters examined, the differences were generally small (Table 24). The maximum difference for life span was only 10 days. Population C had the earliest maturity, whereas population E was the latest. It seemed that the time from the first flower to the first mature pod (including flower initiation, pod initiation and seed initiation) was relatively constant. The difference between the early and late maturing populations seemed to reside in the vegetative growth and seed development. Early maturing populations, such as C, flowered early and filled pods quickly, whereas the later population D flowered late and took a longer time to fill the pods. Some plants in population C were observed to fill pods very quickly.

Table 24. Means and standard errors (in parentheses) of populations and years on morphological traits[a].

Pop or year	Vegetative growth (days)	Flower initiation (days)	Pod initiation (days)	Seed initiation (days)	Seed development (days)	Life span (days)
A	111.26	9.63	6.76	20.96	15.29	165.41
	(0.54)	(0.31)	(0.27)	(0.33)	(0.66)	(1.04)
B	113.58	8.03	6.88	19.58	14.70	164.50
	(0.52)	(0.29)	(0.26)	(0.32)	(0.61)	(0.96)
C	106.45	10.03	7.35	21.58	11.85	159.50
	(0.52)	(0.29)	(0.26)	(0.32)	(0.61)	(0.96)
D	113.85	8.73	7.23	20.13	13.90	165.20
	(0.64)	(0.36)	(0.32)	(0.39)	(0.87)	(1.36)
E	114.15	8.19	6.84	22.28	15.52	169.14
	(0.51)	(0.29)	(0.25)	(0.32)	(0.60)	(0.94)
F	110.16	9.11	7.49	21.44	14.63	163.88
	(0.55)	(0.31)	(0.27)	(0.34)	(0.69)	(1.08)
1989	110.93	8.56	5.69	21.75	-	-
	(0.30)	(0.17)	(0.15)	(0.19)	-	-
1990	112.22	9.34	8.49	20.24	14.33	164.60
	(0.33)	(0.19)	(0.17)	(0.21)	-	-
Mean	111.50	8.89	6.98	21.11	14.33	164.60
Pop F	31.19**	7.28**	1.10	8.55**	4.65**	10.53**
Year F	9.59**	8.37**	161.25**	28.18**	-	-
P*Y F	2.47*	1.78	1.45	0.65	-	-

Table 24. Means and standard errors (in parentheses) of populations and years on morphological traits (continued).

Pop or year	Early plant height (cm)	Early node number	Early branch number	Lower stem width (mm)	Upper stem width (mm)	Node intensity	Final branch number	Shoot dry weight (g)
A	15.43	6.92	4.32	7.11	2.41	6.56	18.18	26.79
	(2.04)	(0.41)	(0.26)	(0.14)	(0.10)	(0.26)	(1.24)	(1.12)
B	5.78	7.23	4.31	6.91	2.02	7.35	23.78	28.97
	(2.04)	(0.41)	(0.26)	(0.13)	(0.10)	(0.25)	(1.18)	(1.06)
C	4.26	7.00	4.68	6.74	1.80	8.43	28.88	21.94
	(1.86)	(0.35)	(0.24)	(0.14)	(0.10)	(0.25)	(1.18)	(1.06)
D	19.75	7.25	5.54	6.89	2.33	6.23	17.88	32.33
	(2.39)	(0.53)	(0.30)	(0.17)	(0.11)	(0.31)	(1.44)	(1.30)
E	16.95	7.11	4.27	6.47	2.64	5.66	12.81	30.28
	(1.88)	(0.35)	(0.24)	(0.13)	(0.10)	(0.25)	(1.16)	(1.05)
F	10.84	7.71	5.00	6.66	2.51	7.07	17.91	28.71
	(1.99)	(0.40)	(0.25)	(0.14)	(0.10)	(0.27)	(1.25)	(1.12)
1989	14.02	-	4.38	7.06	2.28	6.94	18.96	27.57
	(1.05)	-	(0.13)	(0.08)	(0.05)	(0.14)	(0.68)	(0.61)
1990	10.32	7.19	4.99	6.53	2.28	6.82	20.85	28.77
	(1.30)	-	(0.16)	(0.09)	(0.06)	(0.16)	(0.75)	(0.68)
Mean	12.44	7.19	4.62	6.79	2.28	6.91	20.01	27.94
Pop F	12.22**	0.50	3.74**	2.79*	11.63**	14.68**	22.61**	10.05**
Year F	3.35	-	9.72**	19.80**	0.00	0.38	3.94*	1.63
P*Y F	1.52	-	2.58*	1.87	0.69	0.48	1.51	1.31

Table 24. Means and standard errors (in parentheses) of populations and years on morphological traits (continued).

Pop or year	Base leaf width (cm)	Base leaf length (cm)	Base leaf shape	Base leaf petiole length (cm)	Leaf width (cm)	Leaf length (cm)	Leaf shape	Leaf petiole length (cm)
A	2.20	4.57	2.11	2.91	2.82	6.02	2.14	4.07
	(0.05)	(0.10)	(0.04)	(0.10)	(0.06)	(0.13)	(0.04)	(0.16)
B	2.33	4.34	1.86	3.07	3.12	5.91	1.91	4.12
	(0.05)	(0.10)	(0.04)	(0.10)	(0.06)	(0.12)	(0.04)	(0.15)
C	2.44	4.98	2.07	3.22	2.87	6.32	2.22	4.04
	(0.05)	(0.10)	(0.04)	(0.10)	(0.06)	(0.12)	(0.04)	(0.15)
D	2.22	4.43	2.01	2.61	2.90	5.92	2.05	4.07
	(0.06)	(0.12)	(0.05)	(0.12)	(0.08)	(0.15)	(0.05)	(0.19)
E	2.41	4.62	1.93	3.07	3.35	6.45	1.94	4.77
	(0.05)	(0.10)	(0.04)	(0.09)	(0.06)	(0.12)	(0.04)	(0.15)
F	2.65	4.56	1.73	3.31	3.51	6.10	1.75	4.70
	(0.05)	(0.10)	(0.04)	(0.10)	(0.07)	(0.13)	(0.04)	(0.16)
1989	2.49	4.77	1.93	3.26	3.06	6.05	2.00	4.43
	(0.03)	(0.05)	(0.02)	(0.05)	(0.04)	(0.07)	(0.02)	(0.09)
1990	2.26	4.40	1.97	2.80	3.14	6.19	2.00	4.16
	(0.03)	(0.06)	(0.03)	(0.06)	(0.04)	(0.08)	(0.02)	(0.10)
Mean	2.38	4.60	1.95	3.06	3.10	6.13	2.00	4.32
Pop F	10.41**	4.97**	11.61**	4.70**	19.06**	3.28**	19.44**	4.46**
Year F	34.83**	22.90**	1.63	31.50**	2.38	1.92	0.04	3.45
P*Y F	3.05*	1.57	1.58	0.73	1.82	1.02	0.94	0.59

Table 24. Means and standard errors (in parentheses) of populations and years on morphological traits (continued).

Pop or year	Flower width (mm)	Flower length (mm)	Ovule number per ovary	Flower number per infloresc.	Pod number per infloresc.	Flower & pod abortion (%)	Ovule & seed abortion (%)	Pod length (cm)
A	4.69	5.26	2.39	4.87	1.94	59.00	1.48	2.64
	(0.10)	(0.10)	(0.05)	(0.28)	(0.10)	(1.98)	(1.09)	(0.03)
B	4.73	5.38	3.03	7.06	2.26	66.60	4.27	2.68
	(0.10)	(0.10)	(0.05)	(0.27)	(0.09)	(1.89)	(1.04)	(0.03)
C	4.73	5.59	2.86	6.32	1.82	70.40	5.44	2.75
	(0.10)	(0.09)	(0.05)	(0.26)	(0.09)	(1.84)	(1.02)	(0.03)
D	4.65	5.46	2.74	6.52	2.44	60.89	3.81	2.60
	(0.13)	(0.12)	(0.06)	(0.31)	(0.10)	(2.21)	(1.21)	(0.03)
E	4.87	5.47	2.91	7.57	2.34	66.99	4.07	2.78
	(0.11)	(0.10)	(0.05)	(0.26)	(0.09)	(1.83)	(1.00)	(0.03)
F	5.14	5.83	2.79	7.46	2.16	67.85	7.09	2.75
	(0.11)	(0.10)	(0.05)	(0.28)	(0.09)	(1.94)	(1.07)	(0.03)
1989	-	-	-	5.98	2.09	62.90	3.33	2.67
	-	-	-	(0.16)	(0.05)	(1.15)	(0.63)	(0.02)
1990	4.81	5.50	2.79	7.29	2.23	67.68	5.39	2.73
	-	-	-	(0.16)	(0.05)	(1.11)	(0.61)	(0.02)
Mean	4.81	5.50	2.79	6.70	2.16	65.62	4.46	2.70
Pop F	2.88*	4.18**	21.17**	14.55**	6.53**	5.04**	3.10*	6.33**
Year F	-	-	-	33.31**	3.63	7.80**	5.58*	5.47*
P*Y F	-	-	-	3.08*	1.51	3.19**	0.93	1.93

Table 24. Means and standard errors (in parentheses) of populations and years on morphological traits (continued).

Pop or yr	Pod number per plant	Proportions of 1-seed pods (%)	2-seed pods (%)	3-seed pods (%)	4-seed pods (%)	Seed number per pod	Seed weight per plant (g)	Harvest index	100 Seed weight (g)
A	560.20	10.07	56.73	33.11	0.03	2.22	24.02	0.34	2.10
	(23.19)	(0.92)	(1.61)	(2.01)	(0.20)	(0.03)	(0.86)	(0.00)	(0.04)
B	586.78	10.43	38.11	50.34	1.12	2.42	24.67	0.33	2.01
	(21.96)	(0.88)	(1.54)	(1.93)	(0.19)	(0.03)	(0.82)	(0.00)	(0.04)
C	532.63	7.03	39.13	53.24	0.58	2.47	23.29	0.36	2.03
	(21.96)	(0.88)	(1.54)	(1.93)	(0.19)	(0.03)	(0.82)	(0.00)	(0.04)
D	672.73	9.63	41.40	48.43	0.55	2.40	27.78	0.33	1.86
	(26.90)	(1.06)	(1.86)	(2.34)	(0.23)	(0.03)	(1.00)	(0.01)	(0.05)
E	583.78	9.13	38.65	51.21	1.01	2.44	27.35	0.34	2.15
	(21.70)	(0.86)	(1.50)	(1.88)	(0.19)	(0.03)	(0.81)	(0.00)	(0.04)
F	556.07	7.94	43.61	48.11	0.34	2.41	26.60	0.35	2.28
	(23.30)	(0.92)	(1.61)	(2.02)	(0.20)	(0.03)	(0.87)	(0.00)	(0.04)
1989	556.38	9.82	44.77	45.18	0.21	2.36	22.21	0.32	2.01
	(12.74)	(0.51)	(0.89)	(1.12)	(0.11)	(0.02)	(0.47)	(0.00)	(0.02)
1990	607.68	8.25	41.11	49.64	1.00	2.43	29.03	0.36	2.13
	(14.06)	(0.56)	(0.97)	(1.22)	(0.12)	(0.02)	(0.52)	(0.00)	(0.03)
Mean	576.37	9.03	42.86	47.51	0.60	2.40	25.30	0.34	2.07
Pop F	3.40**	2.30*	20.40**	13.86**	4.74**	9.92**	3.87**	5.07**	9.79**
Year F	7.01**	4.89*	8.34**	8.07**	23.31**	12.02**	95.05**	127.71**	9.40**
P*Y F	1.47	3.58**	1.71	2.91*	0.97	2.70*	2.65*	3.56**	0.28

*,**: Statistically significant at 95% and 99% probability levels, respectively.

a Refer to the Materials and Methods for unit and recording method for each trait.

Early plant height, early node and branch numbers were indicators of the early growth rate of wild soybean (Table 24). These characters might be important in intra- and inter-specific competitions in natural habitats. Population D grew the fastest among the six populations during the early season. The thicker the stem widths in a 5 cm length of the main stem above the cotyledonary node, the fewer the number of nodes in this region, and, as a consequence, the fewer the final number of branches (similar to the cultivated soybean). These characters thus reflect the degree to which the main stem is distinct from other branches in the plants of a population. In this sense, population C had thinnest upper stem width, the highest number of nodes in the 5-cm long main stem and the highest final branch number, and was the most different from the cultivated soybean. In contrast, population E had the thickest upper stem width, the lowest number of nodes in the 5-cm long main stem and the lowest number of final branch number, and was the most advanced form in the sense of domestication. Populations C and E also had the lowest and the highest shoot dry weight, respectively, which means that the main stem contributed more to shoot dry weight than all other branches combined.

Leaves at the base of a plant were much smaller than those in other parts of the plant (Table 24). Plants in population F had the largest leaf width, long petioles, and

the most oval leaf shape as indicated by the smallest length to width ratio. However, the leaf characteristics of these six populations were, in general, not very different from each other.

Population F had the largest flowers (Table 24), population A had the smallest flower length, population B had the highest number of ovules per ovary, and population E had the highest number of flowers per inflorescence. Since the flower and pod abortion rates were different among populations, the number of pods harvested from each inflorescence did not correspond to the number of flowers initiated on each inflorescence. Population C had the highest flower and pod abortion rate and also had the lowest number of pods per inflorescence. If more pods per inflorescence characterizes the cultivated soybean, population C would be the most different from the cultivated soybean. The average flower and pod abortion rate was 65.62%, compared to 73.5% reported by Bult (1989). The average ovule and seed abortion rate was 4.46%, nearly identical to 4.5% reported by Bult (1989). Population E had the longest pods.

Population D had the highest number of pods per plant (Table 24). The number of seeds per pod seemed to be determined mainly by the percentages of two- and three-seed pods. Population C had the highest number of three-seed pods and the highest number of seeds per pod. Population A, on the

other hand, had the highest number of two-seed pods and the lowest number of seeds per pod. The seed weight per plant was determined mostly by the number of pods per plant. Populations D and C had the highest and lowest number of pods per plant and also the highest and lowest seed yields, respectively. Although population D had the highest seed yield per plant, it had smallest seeds as expressed by 100 seed weight. Population F had the largest seeds.

From the above results, it can concluded that, in general, population C was most different from other populations, especially from populations D and E. If characters resembling those of the cultivated soybean were considered the most advanced characters in terms of domestication, population C would be more primitive than other populations, and populations D and E would be the most advanced among these six populations. However, population D had small seeds, in contrast to the large seeds of the cultivated soybean. Therefore, natural selection may favor different combinations of traits, whereas artificial selection would be only directed towards high seed yield. In nature, wild soybean may be exposed to far more complicated environments.

Morphological variation can be expressed by coefficients of variation (CV). The mean CV's of 22 variables of 1989 data and 26 variables of 1990 data were listed in Table 25. The differences among populations in CV were not significant at 95% probability level based on ANOVA. However,

Table 25. Analysis of variance on coefficients of variation (CV) of both 1989 and 1990 morphological data[a].

Population	1989 Mean	1989 Std Dv	1990 Mean	1990 Std Dv
A	21.25	(16.35)	20.64	(20.92)
B	19.80	(11.66)	15.99	(11.07)
C	19.18	(15.14)	17.48	(11.27)
D	19.85	(14.65)	18.13	(13.63)
E	20.43	(14.28)	18.23	(15.18)
F	24.72	(24.95)	20.53	(15.73)
F value	1.85		2.12	
p	0.11		0.07	

[a] 31 variables for 1989 and 37 for 1990.

populations A and F tended to have higher morphological variation than other populations in both years.

The within- and among-population morphological variation were analyzed by F tests on the 33 variables for 1989 and 39 variables for 1990 (Tables 26 and 27). Populations explained 5.84% up to 47.54% and 4.90% up to 54.60% of the sums of squares, respectively, in the individual traits of 1989 and 1990 data. The average percentages of the sums of squares attributable to population in 1989 and 1990 were 19.86% and 20.57%, respectively. F values varied from 1.14 to 26.66 in 1989 and from 1.01 to 21.17 in 1990. The average F was 5.40 and 4.92, respectively, in 1989 and 1990.

There is another way to interpret these F tests. The percentages of sums of squares explained by populations could be considered as the among-population variation. The percentages of sums of squares explained by the error term could be considered as the within-population variation caused by the plants. Therefore, the average among- and within-population variation for 1989 and 1990 data were 19.86% and 80.14%, and 20.57% and 79.43%, respectively. However, the degrees of freedom were different for populations and error terms, a factor that must be taken into consideration if a comparison between the among- and within-population morphological variation is to be on an equal basis. In this sense, F values would be explained as the ratio of the among-population variation to the within-

Table 26. F tests of population effect on 33 variables of 1989 morphological data.

Variable	% of SS		DF		F[a]
	Pop	Err	Pop	Err	
S1	47.54	52.46	5	114	26.66**
S2	23.59	76.41	5	114	7.04**
S3	7.80	92.20	5	114	1.93
S4	17.95	82.05	5	114	4.99**
V7	26.01	73.99	5	113	7.94**
V9	17.16	82.84	5	113	4.68**
V10	19.30	80.70	5	114	5.45**
V11	12.23	87.77	5	114	3.18*
V12	14.77	85.23	5	114	3.95**
V13	11.66	88.34	5	114	3.01*
V14	23.75	76.25	5	114	7.10**
V15	5.84	94.16	5	114	1.14
V16	23.52	76.48	5	114	7.01**
V17	8.83	91.17	5	114	2.21
V18	11.66	88.34	5	106	2.80*
V19	17.83	82.17	5	114	4.95**
V20	27.63	72.37	5	114	8.70**
V21	30.86	69.14	5	113	10.09**
V22	19.79	80.21	5	113	5.57**
R26	27.51	72.49	5	83	6.30**
R27	22.32	77.68	5	83	4.77**
R28	25.76	74.24	5	83	5.76**
R29	11.80	88.20	5	83	2.22
R30	6.97	93.03	5	113	1.69
R31	8.28	91.72	5	111	2.00
R32	34.29	65.71	5	111	11.58**
R33	37.69	62.31	5	111	13.43**
R34	15.31	84.69	5	111	4.01**
R35	28.39	71.61	5	111	8.80**
R36	15.46	84.54	5	113	4.13**
R37	10.16	89.84	5	113	2.56*
R38	24.93	75.07	5	113	7.50**
R39	18.90	81.10	5	113	5.27**
Mean	19.86	80.14	5	109	5.40**

*,**: 95% and 99% significant levels, respectively.

[a] Mean F=(19.86/5)/(80.14/109).

Table 27. F tests of population effect on 39 variables of 1990 morphological data.

Variable	% of SS		DF		F[a]
	Pop	Err	Pop	Err	
S1	39.75	60.25	5	98	12.93**
S2	12.22	87.78	5	98	2.73*
S3	4.90	95.10	5	98	1.01
S4	19.00	81.00	5	98	4.60**
S5	19.18	80.82	5	98	4.65**
S6	34.95	65.05	5	98	10.53**
V7	25.53	74.47	5	78	5.35**
V8	8.24	91.76	5	78	1.40
V9	10.20	89.80	5	78	1.77
V10	30.87	69.13	5	98	8.75**
V11	16.99	83.01	5	98	4.01**
V12	32.34	67.66	5	98	9.37**
V13	15.12	84.88	5	98	3.49**
V14	39.55	60.45	5	98	12.83**
V15	12.35	87.65	5	98	2.76*
V16	41.15	58.85	5	98	13.71**
V17	12.90	87.10	5	98	2.90*
V18	7.99	92.01	5	95	1.65
V19	26.46	73.54	5	98	7.05**
V20	24.72	75.28	5	98	6.44**
V21	39.26	60.74	5	98	12.67**
V22	22.64	77.36	5	98	5.74**
R23	14.05	85.95	5	88	2.88*
R24	19.18	80.82	5	88	4.18**
R25	54.60	45.40	5	88	21.17**
R26	34.59	65.41	5	95	10.05**
R27	16.49	83.51	5	95	3.75**
R28	7.25	92.75	5	95	1.49
R29	9.79	90.21	5	95	2.06
R30	16.04	83.51	5	98	3.74**
R31	12.43	87.57	5	98	2.78*
R32	17.14	82.86	5	98	4.06**
R33	29.11	70.89	5	98	8.05**
R34	5.67	94.33	5	98	1.18
R35	12.54	87.46	5	98	2.81*
R36	15.75	84.25	5	98	3.66**
R37	18.05	81.95	5	98	4.32**
R38	5.01	94.99	5	98	1.03
R39	18.04	81.96	5	98	4.31**
Mean	20.57	79.43	5	95	4.92**

*,**: 95% and 99% significant levels, respectively.

a Mean F=(20.57/5)/(79.43/95).

population variation on the basis of equal degrees of freedom. The significance in F tests would test whether the among-population variation is significantly higher than the within-population variation. Therefore, a 5.40- and 4.92-fold variation existed among populations, compared with the within-population variation in 1989 and 1990 experiments, respectively. The differences between the among-population variation and the within-population variation were highly significant. The percentages of the among-population variation would be 5.40 / (5.40 + 1) = 84.38% in 1989 and 4.92 / (4.92 + 1) = 83.11% in 1990. The percentages of the within-population variation would be 1 / (5.40 + 1) = 15.63% in 1989 and 1 / (4.92 + 1) = 16.89%. It thus can be seen that the partitioning of morphological variation by this method was consistent in the two years.

Multivariate Analysis.

For 1989 data, multivariate statistics in the CANDISC procedure indicated significant differences among populations at 0.0001 probability levels with Wilks' lambda and all other multivariate statistics. A total of five canonical variables were extracted, four of which were significant at 0.01 probability level and one was significant at p=0.05 by F statistics based on Rao's approximation to the distribution of the likelihood ratio (Appendix III). The proportions of variance were 34.9%, 23.3%, 18.2%, 13.9% and 9.7% explained by the first to the fifth canonical variables,

respectively, with the first two accounting for 58.201% of the total variance.

The first canonical variable separated population A from C, and also separated populations A and C from the all other populations (Fig. 9). The second canonical variable seemed to separate populations B and D from populations E and F, and separate them from populations A and C.

Examining the pooled within canonical structure in the output revealed that S1, V7, V20, R26, R32 and R33 loaded high on the first canonical variable, which means that the correlations between these original variables and the first canonical variable were high, more than 0.30 in this instance (Appendix III). In other words, they contributed most to this canonical variable, which separated population A from C, and populations A and C from populations B, D, E and F. The second canonical variable, which separated mainly populations B and D from populations E and F, was contributed most by S1, S2 and R33 ($r>0.30$). There were no abstractions that could be made about what kind of morphological characters each of the five canonical variables represented.

All multivariate statistics, including Wilks' lambda, for 1990 data also indicated significant differences among populations at $p=0.0001$. Five canonical variables were extracted. Four were significant at $p=0.01$ and one was not significant at $p=0.05$ according to F statistics based on

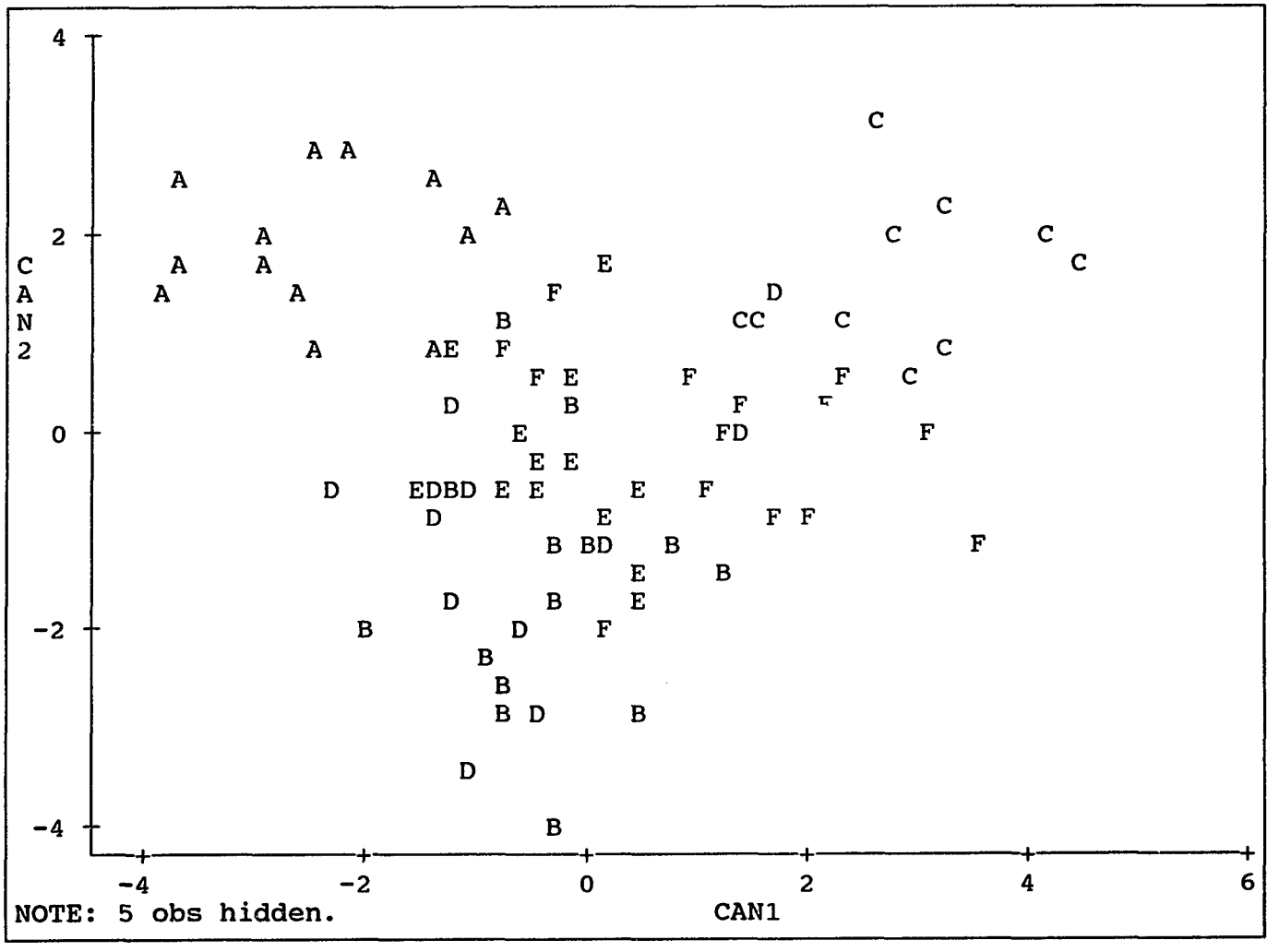


Fig. 9. Scattergram of individuals in the six populations on the first two canonical variables based on the canonical discriminant analysis for 1989 morphological data.

Rao's approximation to the distribution of the likelihood ratio (Appendix III). The proportions of variance explained by the first to the fourth significant canonical variables were 40.7%, 25.1%, 17.9% and 9.7%, respectively. The first and the second accounted for 65.8% of the total variance.

The first canonical variable discriminated populations A and C from populations B and D and also discriminated them from populations E and F (Fig. 10). Populations E and F could also be distinguished from each other by the first canonical variable. The second canonical variable mainly separated three groups: population A, population C, and populations E and F. Populations B and D seemed to be lumped with both A and C in this dimension.

The pooled within canonical structure showed that S1, R25 and R26 loaded high on the first canonical variable ($r>0.25$) and were more responsible for the differentiation of three groups: populations A and C, populations B and D, and populations E and F, than other variables (Appendix III). V7, V10, V14, V19 and V21 had high loading on the second canonical variable, which separated basically three groups: population A, population C, and populations E and F. The representation of each significant canonical variable with respect to the original variables could not be determined.

Although the numbers of variables were different in the canonical discriminant analyses for both years, the results

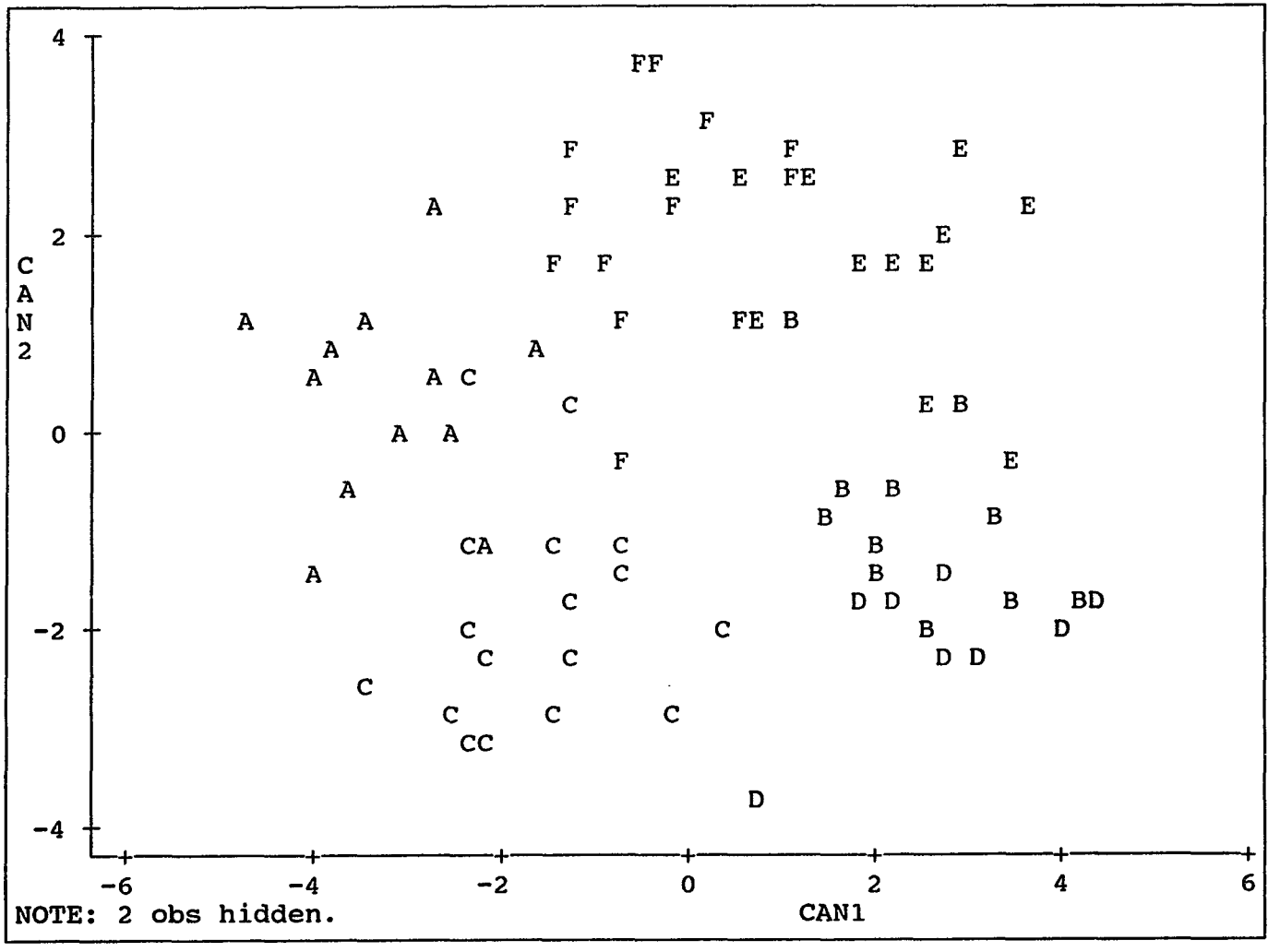


Fig. 10. Scattergram of individuals in the six populations on the first two canonical variables based on the canonical discriminant analysis for 1990 morphological data.

were relatively constant. Populations A and C were always different from each other and also different from the other populations. Populations B and D were associated in both years. So were, to a lesser degree, populations E and F.

The results of cluster analysis for both years were also consistent (Fig. 11). Populations B and D always clustered together, as did populations E and F. Populations A and C formed different clusters. The main difference between the two-year data was that the distance between populations A and C was shorter in 1990 than in 1989 so that they clustered for 1990 data.

It is obvious that the results of both the canonical discriminant analysis with SAS and the cluster analysis with SPSS were also very similar.

The Mahalanobis distances and squared Euclidean distances based on the morphological traits of the two-year study generated from the canonical discriminant analysis and the cluster analysis are presented in Table 28. No clear trends for the average values of both distance measures were observed among populations, except for population C which had the highest values of the average squared Euclidean distance for both years and the highest value of the average Mahalanobis distance for 1989 data, but not for 1990 data. The average values of genetic distances increased from D to B to A and from D to E to F. This pattern was not observed for morphological distances among populations.

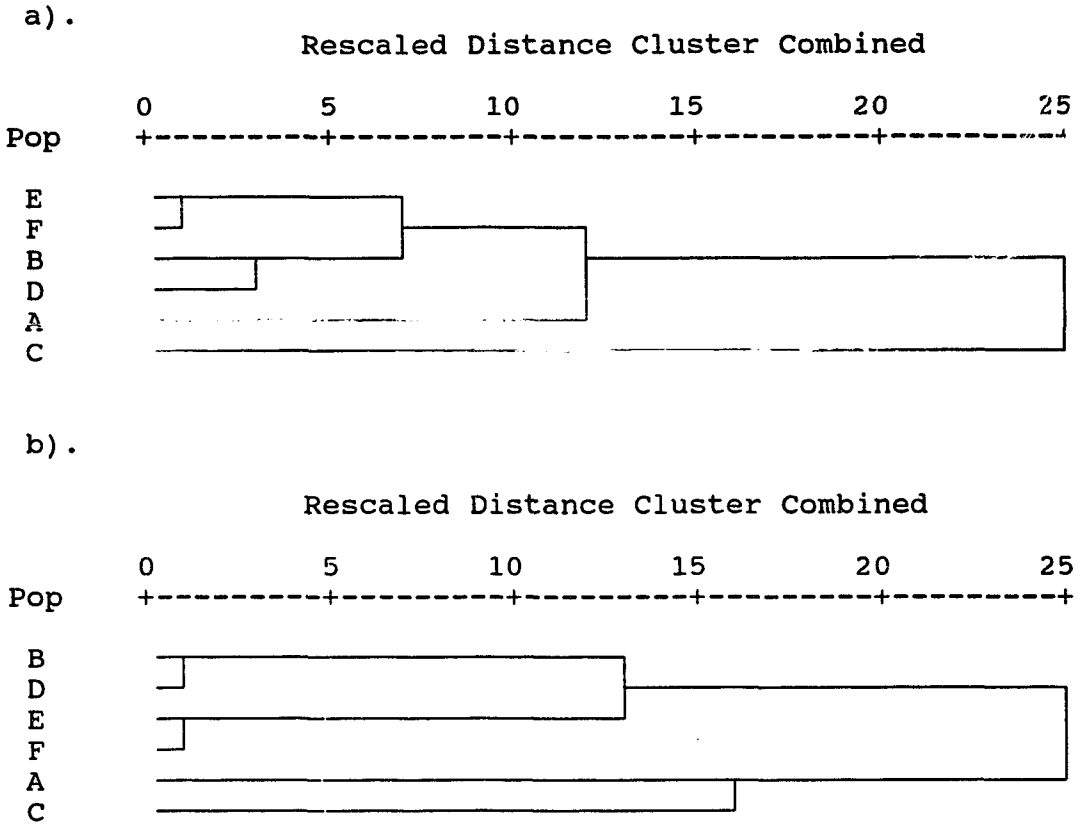


Fig.11. Dendrograms obtained from cluster analysis for the six populations using the average linkage between groups method (UPGMA) based on the squared Euclidean distance for a) 1989 and b) 1990 morphological data.

Table 28. Mahalanobis and the squared Euclidean distances among populations based 1989 and 1990 morphological data.

	A	B	C	D	E	F
(1)	Mahalanobis distance for 1989 morphological data					
B	18.06					
C	30.32	23.03				
D	17.55	10.43	22.26			
E	15.53	10.78	20.61	11.23		
F	20.61	11.58	15.68	15.75	9.83	
Mean	20.41	14.78	22.38	15.44	13.60	14.69
(2)	Mahalanobis distance for 1990 morphological data					
B	39.52					
C	22.07	24.44				
D	44.29	16.10	34.61			
E	35.88	16.81	30.82	27.61		
F	24.35	22.87	22.39	37.95	17.15	
Mean	33.22	23.95	26.87	32.11	25.65	24.94
(3)	Squared Euclidean distance for 1989 morphological data					
B	12.04					
C	22.49	12.21				
D	10.58	6.52	22.02			
E	10.54	7.84	21.17	6.85		
F	10.19	6.99	11.16	11.01	5.08	
Mean	13.17	9.12	17.81	11.40	10.30	8.89
(4)	Squared Euclidean distance for 1990 morphological data					
B	16.10					
C	13.57	12.30				
D	11.56	6.26	18.77			
E	23.28	10.76	24.46	11.83		
F	19.65	10.68	16.18	14.03	6.28	
Mean	16.83	11.22	17.06	12.49	15.32	13.36

DISCUSSION

There were no significant differences among populations in the amount of morphological variation expressed by CV. However, in both years populations A and F had higher variation than other populations. Perhaps, more sensitive tests, such as the paired t test, might reveal significant differences in CV for populations A and F with other populations. This trend of the within-population morphological variation agreed with that of the within-population genetic variation (Chapter II). Populations A and F also had higher genetic variation than other populations.

In Chapter II, population C was proposed to be a distinct population with an origin different from other populations. This hypothesis was supported by the studies of the morphological traits. The canonical discriminant analysis and the cluster analysis on morphological traits all showed population C to be different from other populations. Population C had the highest average distance from other populations in morphological traits. The pattern of relationship between the within- and among-population genetic and morphological variation was also different for population C. Generally, other populations which had higher genetic and morphological divergence also had higher genetic and morphological variation. Population C did not have higher genetic

or morphological variation than other populations, although it had higher divergence.

In Chapter II, I also proposed that population D served as a migration source, and individuals migrated north to establish populations B and A and south to establish populations E and F. Alternatively, populations B, A, E and F might not have been established from population D, but there might have been massive migrations from population D to other populations so that the genetic and morphological distances between D and others were smaller than comparisons among other populations.

The hypothesis that there were migrations from population D north to B to A and south from population D to E to F is supported by the following evidence. First, the general trend of the average genetic distance for each population with all other populations is that D was smaller than B, B smaller than A, and D was smaller than E, E smaller than F. Second, the within-population genetic variation increased from D to B to A and from D to E to F. Populations A and F also had higher morphological variation than other populations. The premise for this reasoning is that newly established populations or populations with new immigrants coming in tend to have higher genetic and morphological variation than stable populations. It is often found that secondary gene centers have higher variation than the primary gene centers. Third, the cluster analysis with the

squared Euclidean distance and the number of loci different between individuals of different populations based on isozyme data all indicated that populations B and A clustered together, as did populations D and E. The cluster analysis on morphological data indicated that populations D and B clustered together, and E and F clustered together. Thus, a genetic and morphological connection formed from populations A and B through population D to populations E and F.

Since only six populations along the 127°E longitudial line were involved in this study and the effects of latitude and populations were confounded, the effect of latitude was not analyzable. Visual inspection did not reveal strong correlations between latitude and any morphological trait. But, the most northern population A and the most southern population F had higher, but not significant CV values. The average genetic distance for each population with all other populations generally increased from D to B to A and from D to E to F. But, this trend did not exist for the distance measures of morphological traits.

CHAPTER IV

RELATIONSHIP BETWEEN ISOZYME DATA AND MORPHOLOGICAL TRAITS

INTRODUCTION

Electrophoresis has been widely used in genetic studies in the past few decades. The natural question that has arisen is whether results are congruent between isozyme and morphological studies. Answering this question can provide justification for the use of electrophoresis in genetic studies. It will also have implications in population genetic studies of other molecular genetic markers, such as RFLP and RAPD markers, which have recently gained popularity.

Results of the past research on this subject are somewhat confusing. Mickevich and Johnson (1976) studied populations of several fish species. The evolutionary trees generated from the isozyme data and the morphological data were remarkably similar. A strong association was also found between quantitative characters and isozyme genotypes of the slender wild oat species, Avena barbata (Hamrick and Allard, 1975). In a "common garden" study on this wild oat, four of the five characters examined were significantly different between two multilocus genotypes that were homozygous for different alleles at five isozyme loci and apparently adapt-

ed to two different environments, mesic and xeric. In studying the genetic and morphological variation of face fly, Bryant (1984) found significant correlations between the number of isozyme alleles per locus and the average CV of 14 morphological characters in several US and European populations. Stuber et al. (1982) selected individuals from a corn population to form a new population in such a way so that the isozyme allele frequency composition of the new population was similar to that of a high yielding population. The result indicated that yield and ear number were improved over the parental population.

However, several studies show no definite relationships between isozyme and morphological data. In some cases, there is considerable variation in morphological traits among populations (Giles, 1983; Jain et al., 1980 and Ryman et al., 1984), but, there is little genetic variation revealed by isozymes, or the relationships among populations reflected by isozymes and morphological traits are different. Even within the same study, results are somewhat inconsistent. Price et al. (1984) found a significant correlation between genetic and morphological distances among populations in one self-fertilizing plant species, Avena barbata, but not in the other two self-fertilizing species, Hordeum jubatum, Hordeum vulgare, and an outcrossing species, Clarkia williamsonii. Similar results were obtained by Nevo et al. (1979). Their study of Israeli populations of

Hordeum spontaneum showed correlations only between spikelet morphological variation and two isozyme loci. Wolff (1991) also found that different populations of three Plantago species, P. major, P. coronopus and P. lanceolata, had different degrees of concordance between isozyme genetic variation and morphological variation.

In soybean, Graef et al. (1989) made crosses between the cultivated soybean, Glycine max, and wild soybean, G. soja, and compared the BC_2F_4 lines having various numbers of alleles from the wild soybean at five isozyme loci. Those lines with alleles of wild soybean at all the five loci were more similar to the wild parent in five morphological traits examined, whereas those with alleles all from the cultivated soybean at these five loci were morphologically closer to the cultivated soybean. But, there were no associations between individual isozyme loci and morphological traits. Congruence between genetic relationships among wild soybean populations and morphological relationships was found only for certain kinds of morphological characters (Chiang, 1985 and Kiang and Chiang, 1990) and in certain years (Bult, 1989).

The objective of this study was to continue to examine the congruence between isozyme data and morphological data with six natural populations of wild soybean from South Korea.

MATERIALS AND METHODS

Seed source, procedure of electrophoresis, enzymes assayed, growing conditions of the greenhouse experiments and morphological traits examined were described in the Introduction chapter, and Chapters I, II and III. Gel and staining solution recipes are given in Appendix I. Isozyme data were analyzed with a Fortran computer program, Biosys-1. Univariate analysis of morphological traits was performed with GLM procedure of SAS. Multivariate analyses were conducted with both CANDISC procedure of SAS and CLUSTER procedure of SPSS. These statistical procedures were also described in Chapters II and III.

The Pearson product-moment and Spearman rank correlation coefficients based on the 15 coefficients of genetic and morphological distance measures were computed with CORRELATION and NONPAR CORR procedures of SPSS, respectively. The means of morphological traits for multilocus genotypes were generated from LSMEAN statement of SAS GLM procedure. Multiple comparisons of the means were based on t tests given by PDIFF statement in SAS GLM.

RESULTS

As pointed out in Chapters II and III, population C was the most different population among the six. If C is considered separately, then, all measures of genetic and morphological variation had similar patterns among the populations; i.e. both types of variation increased from D to B to A and from D to E to F. The only exception was B, which had a smaller CV than D in 1990 morphological data (Table 29). It should be mentioned that there were no significant differences among populations for CV in both years, but, it seemed that there was a general trend of congruence between genetic variation and morphological variation.

Another important aspect in the question of congruence between isozyme data and morphological traits that needs to be looked at is the partitioning of both genetic and morphological variation within and among populations. When Nei's gene diversity was partitioned into the components within and between populations, the between-population portion was only 0.618 of the within-population portion (Table 30). The average number of loci different between multilocus genotypes of different populations was 1.27 times that between those of the same population (Table 30). When two individuals of different populations were compared, the average

Table 29. Comparison of genetic variation measured by the number of alleles per locus (A), 99% polymorphism (P) and the expected heterozygosity with morphological variation measured by CV.

Pop	A	P	H_{exp}	1989CV	1990CV
A	1.5	42.9	0.158	21.25	20.64
B	1.4	31.4	0.113	19.80	15.99
C	1.3	34.3	0.129	19.18	17.48
D	1.3	25.7	0.100	19.85	18.13
E	1.4	40.0	0.133	20.43	18.23
F	1.5	48.6	0.168	24.72	20.53

Table 30. Comparison of genetic variation and morphological variation in partitioning into within- and between-population components.

	Nei's gene differentiation	Multilocus genotypic difference	Individual difference	F test 1989 data	F test 1990 data
Within	0.170 (H_S)	6.56 loci	4.79 loci	15.63%	16.89%
Between	0.105 (D_{ST})	8.32 loci	8.09 loci	84.38%	83.11%
Between/Within	0.618 (D_{ST}/H_S)	1.27	1.69	5.40 (F)	4.92 (F)

number of loci different was 1.69 times that of the comparisons of the same population (Table 30). As for morphological traits, the mean sums of squares per trait per degree of freedom among populations were 5.40 and 4.92 times those within populations, respectively, for 1989 and 1990 morphological data. Thus, genetically, more variation existed within populations, whereas, morphologically, more variation existed among populations.

The congruence between isozyme data and morphological data has been studied mostly in terms of relationships among populations. In this study, cluster analysis of both isozyme and morphological data all indicated that C was a distinct population (Chapters II and III). However, correlation analysis with genetic and morphological distance measures showed there were no significant correlations between these two sets of distance measures, although there were significant correlations within each set (Table 31). Cluster analysis and canonical discriminant analysis also gave different results for isozyme and morphological data of the populations other than C. For isozyme data, A clustered with B, and D with E. But, for morphological traits, B clustered with D, and E with F. Thus, it can be concluded that there was congruence if there was a drastic difference, but, generally, there was no congruence between isozyme data and morphological traits in terms of population relationships.

Among the three measures of genetic distance, the

Table 31. Correlation coefficients among all genetic, morphological and geographic distance measures[a].

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(1) Pearson Product-Moment Correlation Coefficients							
(a) Nei's							
(b) Loc.df	.893**						
(c) Euc.is	.789**	.868**					
(d) Euc.89	.312	.268	.329				
(e) Euc.90	.049	.181	.266	.593*			
(f) Mah.89	.253	.301	.266	.852**	.475		
(g) Mah.90	-.350	-.364	-.303	.297	.481	.264	
(h) Geo.ds	-.100	.162	.245	-.200	.353	-.102	.023
(2) Spearman Rank Correlation Coefficients							
(a) Nei's							
(b) Loc.df	.893**						
(c) Euc.is	.761**	.896**					
(d) Euc.89	.389	.436	.432				
(e) Euc.90	.018	.236	.279	.636**			
(f) Mah.89	.195	.324	.404	.899**	.604**		
(g) Mah.90	-.368	-.314	-.243	.404	.511*	.406	
(h) Geo.ds	.007	.209	.240	-.140	.284	-.077	.061

*,**: 95% and 99% significant levels, respectively.

[a] (a) Nei's unbiased genetic distance.
(b) the number of loci different between individuals.
(c) the squared Euclidean distance of isozyme data.
(d) the squared Euclidean distance of 1989 morphological data.
(e) the squared Euclidean distance of 1990 morphological data.
(f) Mahalanobis distance of 1989 morphological data.
(g) Mahalanobis distance of 1990 morphological data.
(h) geographic distance.

squared Euclidean distance had the highest correlations with morphological distances and geographic distance (Table 31). The number of loci different between individuals also had better correlations with all measures of morphological distance and geographic distance than Nei's genetic distance. Between the two measures of morphological distances, the squared Euclidean distance correlated better with all the genetic distance measures than Mahalanobis distance. Geographic distance did not have significant correlations with any genetic or morphological distance measure.

Since there was no congruence between isozyme data and morphological traits with respect to population relationships, did the isozyme genotypes or the genetic division within populations revealed by the isozyme markers affect the morphological traits in any way? In Chapter II, all the natural plants were classified into 38 multilocus genotypes plus their heterozygotes. An analysis of variance was performed on the morphological traits using populations and multilocus genotypes within populations as effects. Some multilocus genotypes had only one plant, or more than one plant, only one of which was planted each year. In doing the analysis, these plants were included as miscellaneous multilocus genotypes within respective populations. This conservative approach ensured that each multilocus genotype in the analysis had at least two plants so that one degree of freedom was spared for the error term. Thus, only 25 and

23 multilocus genotypes were included in the analysis of variance, respectively, for 1989 and 1990 morphological data, although there were 38 multilocus genotypes. The analysis was carried out individually for each trait. For some traits, the number of multilocus genotypes in each year's data was fewer than the numbers mentioned above due to missing data.

For the 33 characters of 1989 data, only five and six were not significant by F tests, respectively, for the effects of populations and multilocus genotypes within populations (Table 32). None of the F values for both populations and multilocus genotypes within populations was less than 1. The average F values of populations and multilocus genotypes within populations were 7.47 and 3.32, respectively. Both were highly significant. For the 39 characters of 1990 data, six and four were not significant for populations and multilocus genotypes within populations, respectively (Table 33). Again none of the F values for populations or multilocus genotypes within populations was below 1. The average F values were 6.95 and 3.25, respectively for populations and multilocus genotypes within populations, respectively. Both were highly significant.

In 1989, populations, multilocus genotypes within populations and error term explained 19.87%, 31.74% and 48.40% of variation (i.e. sums of squares) per character,

Table 32. F tests of effects of populations and multilocus genotypes within populations for 1989 morphological data.

	% of SS			DF			F[a]	
Variab	Pop	MLG[b](pop)	Err	Pop	MLG[b](pop)	Err	Pop	MLG[b](pop)
S1	47.54	25.55	26.91	5	19	95	33.57**	4.75**
S2	23.59	21.24	55.16	5	19	95	8.13**	1.93*
S3	7.80	22.68	69.53	5	19	95	2.13	1.63
S4	17.95	28.26	53.78	5	19	95	6.34**	2.63**
V7	26.01	44.88	29.11	5	19	94	16.80**	7.63**
V9	17.16	19.42	63.42	5	19	94	5.09**	1.52
V10	19.31	30.99	49.69	5	19	95	7.38**	3.12**
V11	12.23	44.99	42.77	5	19	95	5.43**	5.26**
V12	14.82	36.99	48.19	5	19	95	5.82**	3.83**
V13	11.66	21.10	67.24	5	19	95	3.30**	1.57
V14	23.74	26.09	50.17	5	19	95	8.99**	2.60**
V15	5.84	42.66	51.50	5	19	95	2.15	4.14**
V16	23.50	43.55	32.95	5	19	95	13.56**	6.60**
V17	8.83	32.50	58.66	5	19	95	2.86*	2.77**
V18	11.66	25.99	62.35	5	17	89	3.33**	2.18**
V19	17.82	41.68	40.49	5	19	95	8.36**	5.15**
V20	27.63	46.46	25.91	5	19	95	20.26**	8.97**
V21	30.86	44.56	24.58	5	19	94	23.60**	8.97**
V22	19.79	23.26	56.95	5	19	94	6.53**	2.02*
R26	27.51	20.30	52.19	5	15	68	7.17**	1.76
R27	22.33	37.84	39.82	5	15	68	7.63**	4.31**
R28	25.76	36.33	37.90	5	15	68	9.24**	4.35**
R29	11.80	19.18	69.01	5	15	68	2.33	1.26
R30	6.97	36.17	56.87	5	19	94	2.30	3.15**
R31	8.28	19.54	72.18	5	18	93	2.13	1.40
R32	34.29	21.70	44.01	5	18	93	14.49**	2.56**
R33	37.69	24.37	37.94	5	18	93	18.48**	3.32**
R34	15.31	26.12	58.57	5	18	93	4.86**	2.30**
R35	28.40	21.75	49.85	5	18	93	10.59**	2.25**
R36	15.46	46.12	38.42	5	19	94	7.57**	5.94**
R37	10.16	28.65	61.19	5	19	94	3.12*	2.32**
R38	24.93	33.50	41.57	5	19	94	11.27**	3.99**
R39	18.90	52.81	28.29	5	19	94	12.56**	9.24**
Mean	19.87	31.74	48.40	5	18	91	7.47**	3.32**

*,**: 95% and 99% significant levels, respectively.
a Mean F calculated using mean % of SS and mean DF.
b MLG: Multilocus genotype.

Table 33. F tests of effects of populations and multilocus genotypes within populations for 1990 morphological data.

	% of SS			DF			F[a]	
Variab	Pop	MLG[b](pop)	Err	Pop	MLG[b](pop)	Err	Pop	MLG[b](pop)
S1	39.75	25.99	34.25	5	17	81	18.80**	3.62**
S2	12.22	25.17	62.61	5	17	81	3.16*	1.91*
S3	4.89	24.16	70.94	5	17	81	1.12	1.62
S4	19.00	34.62	46.37	5	17	81	6.64**	3.56**
S5	19.18	35.85	44.97	5	17	81	6.91**	3.80**
S6	34.95	27.67	37.37	5	17	81	15.15**	3.53**
V7	25.53	45.47	29.00	5	14	64	11.27**	7.17**
V8	8.24	45.13	46.62	5	14	64	2.26	4.43**
V9	10.20	39.48	50.33	5	14	64	2.59*	3.59**
V10	30.87	16.08	53.05	5	17	81	9.43**	1.44
V11	16.99	35.51	47.51	5	17	81	5.79**	3.56**
V12	32.34	24.13	43.53	5	17	81	12.04**	2.64**
V13	15.12	31.77	53.10	5	17	81	4.61**	2.85**
V14	39.56	16.28	44.16	5	17	81	14.51**	1.76*
V15	12.35	41.12	46.53	5	17	81	4.30**	4.21**
V16	41.15	29.44	29.41	5	17	81	22.67**	4.77**
V17	12.89	30.58	56.53	5	17	81	3.70**	2.58**
V18	7.99	13.65	78.37	5	16	79	1.61	0.86
V19	26.46	44.39	29.15	5	17	81	14.71**	7.26**
V20	24.72	34.46	40.82	5	17	81	9.81**	4.02**
V21	39.26	19.28	41.46	5	17	81	15.34**	2.22**
V22	22.64	20.97	56.39	5	17	81	6.51**	1.77*
R23	14.04	59.77	26.19	5	16	72	7.72**	10.27**
R24	19.18	56.46	24.36	5	16	72	11.34**	10.43**
R25	54.60	11.33	34.07	5	16	72	23.08**	1.50
R26	34.59	30.66	34.75	5	17	78	15.53**	4.05**
R27	16.49	23.15	60.36	5	17	78	4.26**	1.76*
R28	7.25	32.12	60.62	5	17	78	1.87	2.43**
R29	9.79	41.42	48.79	5	17	78	3.13*	3.89**
R30	16.04	28.28	55.68	5	17	81	4.67**	2.42**
R31	12.43	32.45	55.12	5	17	81	3.65**	2.80**
R32	17.14	34.00	48.86	5	17	81	5.68**	3.32**
R33	29.11	31.95	38.95	5	17	81	12.11**	3.91**
R34	5.67	32.12	62.21	5	17	81	1.48	2.46**
R35	12.54	36.56	50.90	5	17	81	3.99**	3.42**
R36	15.75	46.68	37.57	5	17	81	6.79**	5.92**
R37	18.05	25.51	56.44	5	17	81	5.18**	2.15*
R38	5.00	43.37	51.63	5	17	81	1.57	4.00**
R39	18.00	48.74	33.26	5	17	81	8.78**	6.98**
Mean	20.56	32.72	46.72	5	17	79	6.95**	3.25**

*,**: 95% and 99% significant levels, respectively.
[a] Mean F calculated using mean % of SS and mean DF.
[b] MLG: Multilocus genotype.

respectively (Table 32). In 1990, they accounted for 20.56%, 32.72% and 46.72%, respectively (Table 33). These numbers cannot be used unless the respective degrees of freedom are taken into consideration. Thus, F values should be examined, because they are the unit-degree-of-freedom ratios of the between- and within-population sums of squares. The average morphological variation explained by populations and multilocus genotypes within populations were 7.47 and 3.32 times the variation attributed to error term for 1989 data. They were 6.95 and 3.25, respectively, for 1990 data. Alternatively, the total variation can be partitioned into those components explained by populations, multilocus genotypes within populations and plants within multilocus genotypes (i.e. error term). For 1989 data,

$$\text{Populations} = \frac{7.47}{(7.47+3.32+1)} \times 100\% = 63.34\%,$$

$$\text{Multilocus genotypes} = \frac{3.32}{(7.47+3.32+1)} \times 100\% = 28.16\%,$$

$$\text{Plants} = \frac{1}{(7.47+3.32+1)} \times 100\% = 8.48\%.$$

For 1990 data, calculated similarly, they were 62.05%, 29.02% and 8.93%, respectively. It can be seen that these numbers were consistent across years. The variation explained by populations and multilocus genotypes within populations were significantly larger than that explained by plants within multilocus genotypes, which were indicated by

significant F values for both years. Thus, multilocus genotypes within populations explained nearly 30% of the total variation per trait per degree of freedom.

To demonstrate the effect of multilocus genotypes on morphological traits, multiple comparisons were made among the means of multilocus genotypes within each population for a few selected characters of each year's data. Twenty and 19 multilocus genotypes were included, respectively, for 1989 and 1990 data in the multiple comparisons. The others that were represented by a single plant were excluded from the multiple comparisons. The characters selected were those in which multilocus genotypes explained more than 40% of the sums of squares in both year's experiments.

There were many significant differences between the means of multilocus genotypes within populations for all the selected morphological traits (Tables 34 and 35). For example in both years, multilocus genotype 11 in population A and 34 in E had the largest early plant height and grew the fastest during the early season in their respective populations. It is also surprising that the maximum differences among populations in seed size expressed by 100 seed weight were 2.09-1.81=0.28 gram in 1989 and 2.40-1.94=0.46 gram in 1990. However, the 100 seed weight of multilocus genotypes varied from 1.52 to 2.61, with 1.09 gram difference in 1989 for population F. That difference in population

Table 34. Comparisons among means of multilocus genotypes within each population for a few selected morphological traits of 1989 data[a].

Pop	MLG[b]	N[c]	Early plant height	Leaf length	Upper stem width	Pod length	100 seed weight
			cm	cm	mm	cm	
A			15.44	6.24	2.47	2.76	2.09
	10	2	6.00b	6.90a	1.75b	2.80a	2.25a
	11	6	29.83a	5.30b	2.92a	2.58b	2.00b
	22	2	10.50b	6.54a	2.75a	2.90a	2.03ab
B			7.10	5.76	1.96	2.62	1.96
	21	8	9.06a	5.29b	2.06a	2.79a	2.19a
	25	10	5.15a	6.22a	1.85a	2.45b	1.74b
C			4.85	6.52	1.97	2.76	1.98
	26	10	4.05a	6.20b	1.75a	2.92a	2.17a
	30	3	6.17a	6.24b	2.17a	2.73b	1.83b
	32	3	4.33a	7.14a	2.00a	2.63b	1.95b
D			24.80	6.03	2.44	2.57	1.81
	1	5	44.80a	6.59a	3.10a	2.64a	1.79a
	16	4	3.50c	5.95ab	2.13b	2.60ab	1.85a
	28	11	26.09b	5.54b	2.09b	2.48b	1.78a
E			13.15	6.29	2.29	2.70	2.01
	14	2	6.50bc	6.85a	1.75bc	2.80ab	2.04bc
	17	2	4.00c	6.55ab	1.50c	2.50d	1.81bc
	18	5	12.60bc	5.87bc	2.30b	2.60cd	2.01b
	34	4	25.25a	6.77a	3.50a	2.88a	2.39a
	38	4	17.38ab	5.43c	2.38b	2.70bc	1.82c
F			6.49	5.80	2.32	2.68	2.04
	2	5	9.80a	6.74a	2.50a	2.70b	1.99b
	3	4	8.00a	5.44b	2.38ab	2.90a	2.06b
	4	7	5.14a	5.70b	2.64a	2.70b	2.61a
	13	2	3.00a	5.32b	1.75b	2.40c	1.52c

[a] Means followed by the same letter within each population were not different at 95% significant level by t tests.
[b] MLG: Multilocus genotype.
[c] Number of plants.

Table 35. Comparisons among means of multilocus genotypes within each population for a few selected morphological traits of 1990 data[a].

Pop	MLG[b]	N[c]	Early plant height	Leaf length	Upper stem width	Pod length	100 seed weight
			cm	cm	mm	cm	g
A			17.28	6.07	2.66	2.65	2.14
	11	3	41.00a	4.80c	3.23a	2.60b	2.07a
	22	4	7.67b	7.09a	2.95a	2.77a	2.21a
	24	5	3.20b	6.33b	1.80b	2.59b	2.15a
B			4.83	5.87	2.04	2.75	2.07
	21	10	5.58a	5.73a	2.08a	2.89a	2.23a
	25	10	4.07a	6.00a	1.99a	2.60b	1.91b
C			3.48	6.48	1.73	2.69	1.94
	26	11	2.95a	6.18a	1.58a	2.85a	2.30a
	30	2	3.75a	6.53a	1.65a	2.72a	1.83b
	32	4	3.75a	6.74a	1.95a	2.50b	1.70b
D			14.50	6.11	2.21	2.67	1.94
	1	2	-	7.04a	2.70a	2.77a	2.10a
	16	2	-	5.57b	1.50b	2.62a	1.84a
	28	6	14.50	5.73b	2.43a	2.64a	1.88a
E			19.08	6.48	2.54	2.85	2.19
	18	3	7.83b	6.92a	2.26b	2.84a	2.25ab
	20	2	-	6.17ab	2.10b	2.85a	2.10bc
	34	9	27.56a	6.86a	3.70a	2.85a	2.39a
	38	4	21.83a	5.96b	2.08b	2.86a	2.00c
F			10.63	6.22	2.50	2.80	2.40
	2	4	8.50b	6.88a	2.55a	2.63c	1.99c
	3	3	6.00b	5.12c	2.60a	2.78b	2.18c
	4	4	4.50b	5.89b	2.38a	2.61c	2.58b
	37	3	23.50a	7.00a	2.47a	3.18a	2.86a

[a] Means followed by the same letter within each population were not different at 95% significant level by t tests.
[b] MLG: Multilocus genotype.
[c] Number of plants.

F was 2.86 - 1.99 = 0.87 gram in 1990. Although there were not enough plants in the analysis for some multilocus genotypes, those with more plants showed significant and consistent differences across years, such as 21 and 25 in population B for pod length and 100 seed weight. In addition, the differences among multilocus genotypes were also consistent across traits. For instance, multilocus genotype 13 in population F of the 1989 planting had the smallest pod length and also the smallest 100 seed weight. Although only two plants were included in the analysis, the consistency across traits confirmed that this multilocus genotype was significantly different from others in population F, since the two measurements were obtained from different pod samples.

Multilocus genotype 37 was planted only in 1990. It had the fastest early growth, long leaves, longest pods and largest seed size in population F (Table 35). In 1990 data, multilocus genotypes within populations explained nearly 60% of the total variation for flower width and length (variables R23 and R24 in Table 33). This proportion of variance was mainly attributed to the significantly larger flowers of multilocus genotype 37 relative to other plants in the six populations (data not shown). There were three natural plants collected for this multilocus genotype, KF5, KF19 and KF20. The greenhouse plants grown from their seeds were all very large compared with other plants. KF5 was re-planted in

the winter of 1992 along with F_1 hybrids between G. max and G. soja. KF5 was morphologically very similar to the hybrid plants. Probably, these three plants were the offspring of a rare outcrossing event between the cultivated and wild soybean.

DISCUSSION

Comparisons among Measures of Genetic and Morphological Distances.

Among the three measures of genetic distance, the squared Euclidean distance had the highest correlations with morphological and geographic distances. The number of loci different between individuals also had much higher correlations with morphological and geographic distances than Nei's genetic distance.

The difference between the squared Euclidean distance and Nei's genetic distance in correlations with morphological distance measures for populations may lie in the ways they are calculated. In calculating the squared Euclidean distance, each allozyme frequency variable was converted to Z score,

$$Z_i = \frac{X_i - \bar{X}_i}{S_{di}},$$

where Z_i and X_i were Z score and the frequency of a particular population, respectively, at ith allele, and $\bar{X}_i$ and S_{di} were the mean frequency and standard deviation of that allele among populations (see Ott, 1984. pp. 73-77). This conversion ensured that all allele frequency variables were on an equal basis and their absolute values had no effects

on the calculation. Then, the squared Euclidean distance between populations X and Y was calculated as:

$$\text{Distance}(X,Y) = \sum_{i=1}^{n} (Z_{Xi} - Z_{Yi})^2,$$

where n was the total number of alleles, Z_{Xi} and Z_{Yi} were the Z scores of populations X and Y, respectively at ith allele (see Norusis, 1990. pp. 203). As in calculating Nei's genetic distance, only polymorphic loci were included in the calculation. The most frequent allele at each polymorphic locus was also excluded, because it was totally predictable from other alleles at that locus (the combined frequency was always 1).

Thus, it can be seen that if a population possessed a rare allele, which did not exist in other populations, this population had very high Z score for this allele. If a population possessed more rare alleles than any other populations, it would have very high Z scores for all these allele variables. In calculating the squared Euclidean distance, the difference between populations in Z scores was squared, which means that the difference in possessing rare alleles was further magnified. Also, the most frequent allele at each polymorphic locus was excluded. Therefore, the squared Euclidean distance basically reflected the population relationships in terms of possession of rare alleles. The higher correlations between the squared Euclidean distance and the measures of morphological distance might indicate that the

morphological differences among populations were caused by the rare genotypes revealed by rare isozyme alleles and that the squared Euclidean distance best described such genetic differences among populations.

On the contrary, in calculating Nei's genetic distance, multiplication with the more frequent alleles and squaring probably diminish much of the contribution of rare alleles to the genetic distance, just in the same way as the number of digits retained and rounding can affect calculation results.

Between the measures of morphological distances among populations, the squared Euclidean distance seemed better than Mahalanobis distance in correlations with genetic distances. This might be due to the inclusion of more information in the squared Euclidean distance. In calculating the squared Euclidean distance for morphological traits, each morphological variable was first converted to Z scores, as with the isozyme data. Then, it was averaged among populations. Because the mean Z scores were used to calculate the squared Euclidean distance, no observations were deleted due to missing values. The difference was that the mean Z scores for some populations were based on fewer observations than those for other populations. Mahalanobis distance, on the other hand, was based on the listwise deletion. If an observation had a missing value even for one variable, it was

deleted from the calculation. Thus, the squared Euclidean distance might contain more information about morphological differences among populations.

Congruence between Isozyme Data and Morphological Traits

The question of congruence between isozyme genotypes and the morphological traits has been treated as a general question. There may be different answers when it is asked in different situations. The question may be asked in the following situations: 1). studying the taxonomic relationships among different taxa or evolutionary relationships among different populations of the same species or taxa which can exchange genes freely; 2). studying the levels of genetic variation of different taxa or different populations of the same species or taxa which can exchange genes freely; 3). studying the genomic contribution by parents in a hybridization population; and 4). studying the association between individual isozyme markers and individual morphological traits in QTL (quantitative trait loci) mapping.

In the first situation, when isozymes and morphological traits are used to study the taxonomic relationships among different taxa, especially when there is no gene flow, the congruence between these two types of data should be high. This assertion is supported by the study of Mickevich and Johnson (1976).

However, as for the evolutionary relationships among populations of the same species or taxa which can exchange

genes freely, the congruence is low. Morphological traits often give a different picture from that of isozymes in terms of population relationships. Several studies, including the present one, all indicate that there are no general correlations between genetic and morphological distances for populations (Bult, 1989; Chiang, 1985, and Kiang and Chiang, 1990). It seems that both genetic and morphological distance measures need to be examined in order to understand evolutionary relationships among populations. The genetic bases of isozymes and morphological traits may be quite different. Isozymes involve a smaller number of loci and are controlled in restricted genomic areas. Morphological traits, on the other hand, are often controlled by a large number of genes (polygenes) involving widespread genomic regions. The chance to detect differences among populations is thus much greater for morphological traits. Besides, there are interactions between alleles and between loci (dominance and epistasis) controlling morphological traits. Plants also have plasticity, which defines a plant's capacity to change morphologically to adapt to different environments without changing genetically. When grown in the same controlled environment, populations should reveal genetic differences in morphological traits. However, these genetic differences among populations may change in another controlled environment. In a greenhouse study, the morphological relationships among

seven Japanese local populations of wild soybean in the first year were very different from those of the second year (Bult, 1989). This differences might be caused by differential plastic response to different environments in the populations or by the differential environmental factors in the plastic response of different populations.

Another problem is the measure of genetic and morphological distances. Different measures give different results. By the currently used measures, most of the genetic variation of this study existed within populations, whereas most of morphological variation existed between populations (Table 30).

Although isozymes do not reflect the same relationship among populations as morphological traits, isozymes have been and are still a good tool for assaying the levels of genetic variation for different species or different populations of the same species or taxa which can exchange genes freely. This is the second situation mentioned previously. There is a general agreement between the genetic variation assayed by isozymes and the morphological variation if there are variabilities in both types of data and a sufficient number of isozymes are used. In this study, CV values for 1989 and 1990 data generally changed among populations in a similar pattern as genetic variation (Table 29). In populations of inbreeding species, the positive association between genetic and morphological variation probably result

from the founder's events and restricted gene flow as suggested by other authors (Bryant, 1984 and Bult, 1989).

In the third situation of using isozyme markers in selection programs, the directional change in isozyme allele frequencies would result in the directional change of morphological traits if a large number of isozyme loci are used (Stuber et al., 1982 and Graef et al., 1989).

In the fourth situation, the objective is to find the associations of individual isozyme markers and individual morphological traits. The chance of finding such associations depends on whether there are any genetic linkages between them.

Multilocus Genotypes are Morphological "Biotypes".

In discussing the variation of self-pollinated species, Stebbins (1957) called different morphological types in natural plant populations "biotypes". Layton and Ganders (1984) referred to them as "inbred lines". Tables 34 and 35 showed that multilocus genotypes were morphologically distinguishable. Therefore, the multilocus genotypes corresponded to the "biotypes" at morphological level.

The multilocus genotypes and the "biotype" they represented at morphological level could be best described by Stebbins (1957) in summarizing the previous studies on the morphological variation of natural populations of self-pollinated species:

> ...The natural population consists of several morphological types, each of which is represented by many similar or identical individuals. These types can be recognized by distinctive morphological characteristics, so that the population appears like a cluster of closely related but separate microspecies. In addition to these dominant morphological types, there may exist also several other recognizable types which are represented by only one or two individuals.
>
> If seeds are collected from a single plant of one of the dominant types, the progeny are most often found to be entirely uniform, and breed true for an indefinite number of generations, indicating that the original plant was completely homozygous. In some instances, as in certain families of Senecio vulgaris, the variation decreased slightly in the second and third generations of automatic selfing under artificial isolation, indicating a slight degree of heterozygosity for the original plant.

Layton and Ganders (1984) also made similar remarks about the variation in natural populations of self-pollinated species.

These authors all point out the absolute importance of the very small outcrossing rates in the natural populations of self-pollinated species. It is these rare accidental outcrossing events that recombine genes and create new multilocus genotypes. Those multilocus genotypes that prove to be superior in natural selection will finally replace the existing ones. This process of repeated creation of new gene combinations by outcrossing and elimination of old or less fit gene combinations by natural selection is certainly the most important feature of self-pollinated species. The genetic variation of any population is totally determined by the number of such multilocus genotypes or the morphological biotypes. Under extreme environmental conditions, there may

exist one or very few of them. For example, one multilocus genotype was found for xeric environment and one for mesic environment in Californian natural populations of Avena barbata, and they were morphologically distinct from each other (Clegg and Allard, 1972 and Hamrick and Allard, 1975).

It was also mentioned that these multilocus genotypes or biotypes basically grow sympatrically (Stebbins, 1957 and Layton and Ganders, 1984). Thus, they can not be explained by Wright's shifting balance theory, which divides a population into small semi-isolated demes (Wright, 1977. pp. 443-473 and Hartl, 1988). The shifting balance theory explains a phenomenon on a larger geographic scale.

Evidence of natural selection can be seen in this study. For example, multilocus genotype 11 in population A, 34 and 38 in population E all grew quickly during the early growing season, which was indicated by the largest early plant height in both populations for both years (Tables 34 and 35). They were also represented in the largest proportions in the samples of their populations (Table 18. Chapter II). There were seven plants belonging to multilocus genotype 11 out of a total number of 30 plants in population A. There were 22 plants for multilocus genotypes 34 and 38 in population E, accounting half of the sample of population E. If the samples of this study represented random samples

of respective populations, large proportions for these multilocus genotypes would mean that they occupied large geographic areas. Apparently, quick early growth gave these plants an advantage in competition with other multilocus genotypes for sunlight and nutrients in the natural environment. Consequently, they expanded at the expense of declining of other multilocus genotypes in populations A and E. Multilocus genotype 26, which accounted for 16 plants of 30 in population C (Table 18. Chapter II), had significantly larger seeds than other multilocus genotypes, as indicated by the highest 100 seed weight in the data of both years (Tables 34 and 35). This large seed trait might provide an advantages in K-selection as suggested by Bult (1989). If such speculations are true, then multilocus genotype 1 in population D and 37 in F, which had the largest early plant height, would probably expand very quickly in their populations in the next few years or decades. Multilocus genotype 37 also had the largest seeds in population F.

In conclusion, although they gave a different picture about population relationships from morphological traits, isozymes, as genetic markers, effectively revealed the basic genetic division in the mosaic self-pollinated plant populations of wild soybean.

REFERENCES

Allard, R.W. 1956. Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235-278.

Apuya, N.R., B.L. Frazier, P. Keim, E.J. Roth, and K.G. Lark. 1988. Restriction fragment length polymorphisms as genetic markers in soybean, Glycine max (L.) Merrill. Theor. Appl. Genet. 75:786-792.

Briggs, F.N. and P.F. Knowles. 1977. Introduction to Plant Breeding. Reinhold Publishing Corporation.

Broich, S.L. and R.G. Palmer. 1980. A cluster analysis of wild and domesticated soybean phenotypes. Euphytica 29:23-32.

Broich, S.L. and R.G. Palmer. 1981. Evolutionary studies of the soybean: the frequency and distribution of alleles among collections of Glycine max and G. soja of various origin. Euphytica 30:55-64.

Brown, A.H.D. 1979. Enzyme polymorphism in plant populations. Theoretical Population Biology 15:1-42.

Bryant, E.H. 1984. A comparison of electrophoretic and morphological variability in the face fly, Musca autumnalis. Evolution 38:455-458.

Bult, C.J. 1989. Isozyme and Quantitative Trait Variation within and among Natural Populations of the Wild Soybean, Glycine soja (Sieb. & Zucc.). Ph.D. Dissertation, University of New Hampshire, Durham, New Hampshire.

Bult, C.J., Y.T. Kiang, Y.C. Chiang, J.Y.H. Doong, and M.B. Gorman. 1989. Electrophoretic methods for soybean genetics studies. Soybean Genet. Newsl. 16:175-187.

Carey, K. 1982. Breeding system, genetic variability, and response to selection in Plectritis (Valerianaceae). Evolution 37:947-956.

Chiang, Y.C. 1985. Genetic and Quantitative Variation in Wild Soybean (Glycine soja) Populations. Ph. D. Dissertation, University of New Hampshire, Durham, New Hampshire.

Chiang, Y.C. and Y.T. Kiang. 1987a. Geometric position of genotypes, honeybee foraging patterns and outcrossing in soybean. Botanical Bull. Acad. Sinica 28:1-11.

Chiang, Y.C. and Y.T. Kiang. 1987b. Inheritance and linkage relationships of 6-phosphogluconate dehydrogenase isozymes in soybean. Genome 29:786-792.

Chiang, Y.C. and Y.T. Kiang. 1988. Genetic analysis of mannose-6-phosphate isomerase in soybeans. Genome 30:808-811.

Chiang, Y.C., M.B. Gorman, and Y.T. Kiang. 1987. Inheritance and linkage analysis of phosphoglucose isomerase isozymes in soybeans. Biochemical Genetics 25:893-900.

Clegg, M.T. 1990. Molecular diversity in plant populations. In: A.H.D. Brown, M.T. Clegg, A.L. Kahler and B.S. Weir (eds.). Plant Population Genetics, Breeding, and Genetic Resources. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.

Clegg, M.T. and R.W. Allard. 1972. Patterns of genetic differentiation in the slender wild oat species Avena barbata. Proc. Nat. Acad. Sci. 69:1820-1824.

Clegg, M.T., R.W. Allard, and A.L. Kahler. 1972. Is the gene the unit of selection? Evidence from two experimental plant populations. Proc. Nat. Acad. Sci. USA 69:2474-2478.

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

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

Doong, J-Y, H. 1986. An Electrophoretic Study on Soybean Isozymes. Master Thesis, University of New Hampshire, Durham, New Hampshire.

Doyle, J.J. 1988. 5S ribosomal gene variation in the soybean and its progenitor. Theor. Appl. Genet. 75:621-624.

Doyle, J.J. and R.N. Beachy. 1985. Ribosomal gene variation in soybean (Glycine) and its relatives. Theor. Appl. Genet. 70:369-376.

Duvick, D. N. 1984. Genetic diversity in major farm crops on

the farm and in reserve. Economic Botany 38:161-178.

Ertl, D. S. and W. F. Fehr. 1985. Agronomic performance of soybean genotypes from Glycine max X G. soja crosses. Crop Sci. 25:589-592.

Fehr, W.R. and C.E. Caviness. 1981. Stages of soybean development. Special Report 80, Cooperative Extension Service, Iowa State University, Ames, Iowa.

Fehr, W.R., C.E. Caviness, D.T. Burmood, and J.S. Pennington. 1971. Stage of development descriptions for soybeans, Glycine max (L.) Merrill. Crop Sci. 11:929-931.

Gai, Junyi. 1985. Utilizing the genetic diversity of annual soybean species. In: R. Shibles (ed.). World Soybean Research Conference III: Proceedings. Westview Press, Inc., Boulder, Colorado. pp. 353-360.

Giles, B. 1984. A comparison between quantitative and biochemical variation in the wild barley Hordeum murinum. Evolution 38:34-41.

Gorman, M.B. 1983. An Electrophoretic Analysis of the Genetic Variation in the Wild and Cultivated Soybean Germplasm. Ph.D. Dissertation, University of New Hampshire, Durham, New Hampshire.

Gorman, M.B. and Y.T. Kiang. 1978. Models for the inheritance of several variant soybean electrophoretic zymograms. J. Hered. 69:255-258.

Gorman, M.B., Y.T. Kiang, R.G. Palmer, and Y.C. Chiang. 1983. Inheritance of soybean electrophoretic variants. Soybean Genet. Newsl. 10:67-84.

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

Hamrick, J.L. and R.W. Allard. 1975. Correlation between quantitative characters and enzyme genotypes in Avena barbata. Evolution 29:438-442.

Hamrick, J.L., Y.B. Linhart, and J.B. Mitton. 1979. Relation between life history characteristics and electrophoretically detectable genetic variation in plants. Ann. Rev. Ecol. Syst. 10:173-200.

Harlan, J.R. 1971. Agricultural origins: centers and noncenters. Science 174:468-474.

Hartl, D.L. 1988. A Primer of Population Genetics. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts.

Hedrick, P.W. 1983. Genetics of Populations. Science Books International, Boston.

Hu, Zhi-ang and Hong-xin Wang, 1985. Genetic structure of natural populations of wild soybean (Glycine soja) in Beijing region. Acta Bot. Sin. 27:599-604.

Hymowitz, T. 1970. On the domestication of the soybean. Economic Botany 24:408-421.

Hymowitz, T. and N. Kaizuma. 1979. Dissemination of soybeans (Glycine max): seed protein electrophoresis profiles among Japanese cultivars. Economic Botany 33:311-319.

Hymowitz, T. and N. Kaizuma. 1981. Soybean seed protein electrophoresis profiles from 15 Asian countries or regions: hypotheses on paths of dissemination of soybean from China. Economic Botany 35:10-23.

Hymowitz, T. and C.A. Newell. 1981. Taxonomy of the genus Glycine, domestication and uses of soybeans. Economic Botany 35:272-288.

Hymowitz, T and R. J. Singh. 1987. Taxonomy and speciation. In: J.R. Wilcox (ed.). Soybeans: Improvement, Production and Uses, 2nd Ed. American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., Madison, Wisconsin. pp. 23-48.

Jain, S.K., L. Wu, and K.R. Vaidya. 1980. Levels of morphological and allozyme variation in Indian amaranths: a striking contrast. J. Hered. 71:283-285.

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

Keim, P., B.W. Diers, T.C. Olson, and R.C. Shoemaker. 1990. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Kiang, Y.T. 1987. Mapping three protein loci on a soybean chrosmosome. Crop Sci. 27:44-46.

Kiang, Y.T. 1990a. Linkage analysis of Pgd1, Pgi1, pod color

(L1), and determinant stem (dt1) loci on soybean linkage group 5. J. Hered. 81:402-404.

Kiang, Y.T. 1990b. Mapping the alcohol dehydrogenase locus (Adh1) in soybean linkage group 8. J. Hered. 81:488-489.

Kiang, Y.T. and Y.C. Chiang. 1987a. Genetic linkage of the Adh1 and W1 loci in soybean. Genome 29:582-583.

Kiang, Y.T. and Y.C. Chiang. 1987b. Genetic analysis of the Lap2 locus in soybeans and tests of linkage with other loci. J. Hered. 78:131-132.

Kiang, Y.T. and Y.C. Chiang. 1989. Latitudinal variation and evolution in wild soybean (Glycine soja Sieb. & Zucc.) populations. In: J.H. Bock and Y.B. Linhart (eds.). The Evolutionary Ecology of Plants. Westview Press, Boulder, Colorado. pp. 469-489.

Kiang, Y.T. and Y.C. Chiang. 1990. Comparing differentiation of wild soybean (Glycine soja Sieb. & Zucc.) populations based on isozymes and quantitative traits. Bot. Bull. Academia Sinica 31:129-142.

Kiang, Y.T. and M.B. Gorman. 1983. Soybean. In: S.D. Tanksley and T.J. Orton (eds.). Isozymes in Plant Genetics and Breeding, Part B. Elsevier Science Publishing B. V., Amsterdam. pp. 295-328.

Kiang, Y.T., M.B. Gorman, and Y.C. Chiang. 1985. Genetic and linkage analysis of a leucine aminopeptidase in wild and cultivated soybean. Crop Sci. 25:319-321.

Kiang, Y.T., Y.C. Chiang, J.Y.H. Doong, and M.B. Gorman. 1987. Genetic variation of soybean germplasm. In: "Crop Exploration and Utilization of Genetic Resources" --- Proceedings of an International Symposium held at Changhua, Taiwan, R.O.C. December 6-12, 1986. Published by Taichung District Agriculture Improvement Station.

Kloth, R.H., J.C. Polacco, and T. Hymowitz. 1987. The inheritance of a urease-null trait in soybean. Theor. Appl. Genet. 73:410-418.

Layton, C.R. and F.R. Ganders. 1984. The genetic consequences of contrasting breeding systems in Plectritis (Valerianaceae). Evolution 38:1308-1325.

LeRoy, A.R., W.R. Fehr, and S.R. Cianzio. 1991a. Introgression of genes for small seed size from Glycine soja

into G. max. Crop Sci. 31:693-697.

LeRoy, A.R., S.R. Cianzio, and W.R. Fehr. 1991b. Direct and indirect selection for small seed size in temperate and tropical environments. Crop Sci. 31:697-699.

Loveless, M.D. and J.L. Hamrick. 1984. Ecological determinants of genetic structure in plant populations. Ann. Rev. Ecol. Syst. 15:65-95.

Mickevich, M.F. and M.S. Johnson. 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. Syst. Zool. 25:260-270.

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

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. USA 70:3321-3323.

Nei, M. 1978. Estimate of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Nevo, E. 1978. Genetic variation in natural populations: patterns and theory. Theoretical Population Biology 13:121-177.

Nevo, E., D. Zohary, A.H.D. Brown, and M. Haber. 1979. Genetic diversity and environmental associations of wild barley, Hordeum spontaneum, in Israel. Evolution 33:815-833.

Norusis, M.J. 1990. SPSS Advanced Statistics Student Guide. SPSS, Inc., Chicago, Illinois.

Palmer, R. G. and Y.T. Kiang. 1990. Linkage map of soybean (Glycine max L. Merr.). In: S.J. O'Brien (ed.). Genetic Maps---Locus Maps of Complex Genomes, 5th Ed. Cold Spring Harbor Laboratory, New York. pp. 6.68-6.92.

Palmer, R.G. and T.C. Kilen. 1987. Qualitative genetics and cytogenetics. In: J.R. Wilcox (ed.). Soybeans: Improvement, Production and Uses, 2nd Ed. American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., Madison, Wisconsin. pp. 135-209.

Palmer, R.G., R.C. Shoemaker, and B. Rennie. 1987. Approved soybean gene symbols. Soybean Genet. Newsl. 14:41-58.

Price, S.C., K.M. Shumaker, A.L. Kahler, R.W. Allard, and J.E. Hill. 1984. Estimates of population differentiation obtained from enzyme polymorphisms and quantitative characters. J. Hered. 75:141-142.

Ryman, N, V. Lagercrantz, L. Andersson, R. Chakeraborty, and R. Rosenberg. 1984. Lack of correspondence between genetic and morphologic variability patterns in Atlantic herring (Clupea harengus). Heredity 53:687-704.

SAS Institute. 1985. SAS User Guide: Statistics, Version 5 Edition. SAS Institute, Cary, North Carolina.

Shoemaker, R.C., P.M. Hatfield, R.G. Palmer, and A.G. Atherly. 1986. Chloroplast DNA variation in the genus Glycine subgenus Soja. J. Hered. 77:26-30.

Sisson, V.A., C.A. Brim, and C.S. Leving, III. 1978. Characterization of cytoplasmic diversity in soybeans by restriction endonuclease analysis. Crop Sci. 18:991-996.

Smith, K.J. and W. Huyser. 1987. World distribution and significance of soybean. In: J.R. Wilcox (ed.). Soybeans: Improvement, Production and Uses, 2nd Ed. American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., Madison, Wisconsin. pp. 1-22.

Specht, J.E. and J.H. Williams. 1984. Contribution of genetic technology to soybean productivity---retrospect and prospect. In: Fehr (ed.). Genetic contributions to yield gains of five major crop plants. Spec. Pub. 7. Crop Science Society of America, Madison, WI.

Stebbins, G.L. 1957. Self fertilization and population variability in the higher plants. Am. Nat. 861:337-354.

Stuber, C.W., M.M. Goodman, and R.H. Moll. 1982. Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. Crop Sci. 22:737-740.

Suiter, K.A., J.F. Wendel, and J.S. Case. 1983. Linkage-1: a PASCAL computer program for detection and analysis of genetic linkage. J. Hered. 74:203-204.

Swofford, D.L. and R.B. Selander. 1981. BIOSYS-1: a Fortran program for the comprehensive analysis of electrophoretic data in population genetics and systematics. J.

Hered. 72:281-283.

Swofford, D.L. and R.B. Selander. 1989. Biosys-1 User Manual, Release 1.7. Illinois Natural History Survey, Champaign, Illinois.

Tabachnick, B.G. and L.S. Fidell. 1989. Using Multivariate Statistics, 2nd Ed. Harper & Row, Publisher, New York.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalki, and S.V. Tingey. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535.

Wolff, K. 1991. Genetic analysis of morphological variability in three Plantago species with different mating systems. Theor. Appl. Genet. 81:111-118.

Wright, S. 1977. Evolution and the Genetics of Populations, Volume III: Experimental Results and Evolutionary Deductions. The University of Chicago Press, Chicago.

APPENDICES

APPENDIX I

PROTOCOLS FOR ISOZYME GELS AND STAINING SOLUTIONS

ABBREVIATIONS OF CHEMICALS

APS: Ammonium persulfate
Bis: N,N'-methylene-bis-acrylamide
dH_2O: Distilled water
DTT: Dithiothreitol
EDTA: Ethylenediaminetetraacetic acid (disodium salt and dihydrate)
NAD: Nicotinamide adenine dinucleotide (sodium salt)
NADH: Nicotinamide adenine dinucleotide reduced (disodium salt)
NADP: Nicotinamide adenine dinucleotide phosphate (monosodium salt)
MTT: 3-(4,5-dimethyl thiazoyl-2)-2,5-diphenyl tetrazolium bromide
PMS: Phenazine methosulfate
Starch: Hydrolyzed potato starch for electrophoresis unless specified
TEMED: N,N,N',N'-tetramethyl-ethylenediamine
Tris: Tris(hydroxymethyl)aminothane

BUFFERS, STOCK SOLUTIONS AND PREPARATIONS

Gel and Electrode Buffers.

A. 0.005 M L-Histidine buffer (pH7.0, gel & seed buffer)

dH_2O	1.0	l	4.0	l
L-Histidine(HCl)	1.048	g	4.192	g

Adjust pH with 4 M NaOH.

B. 0.065 M Tris-citrate buffer (pH7.0, tray buffer)

dH_2O	2.0	l	4.0	l
Tris	15.74	g	31.48	g
Citric acid	7.6[illegible]	g	15.36	g

Adjust pH with either concentrated HCl or 4 M NaOH.

Staining Buffers.

A. 0.1 M Acetate buffer (pH5.0)

0.2 M Acetic acid (11.55 ml in 1.0 l dH_2O)		
0.2 M Sodium acetate (16.4 g in 1.0 l dH_2O)		

Add 148 ml 0.2 M acetic acid and 352 ml 0.2 M sodium acetate in a gradual cylinder and bring up to 1.0 l with dH_2O.

B. 0.05 M Sodium phosphate buffer (pH7.0, for Adh only)

0.05 Monobasic sodium phosphate		
dH_2O	1.0	l
NaH_2PO_4	6.9	g
0.05 Dibasic sodium phosphate		
dH_2O	1.0	l
Na_2HPO_4	7.1	g

Add 39 ml 0.05 M monosodium phosphate to 61 ml 0.05 M disodium phosphate just before use.

C. 0.1 M Sodium phosphate buffer (pH6.4, for Est only)

dH_2O	1.0	l
Monobasic sodium phosphate	13.9	g
Dibasic sodium phosphate	5.3	g

Adjust pH with concentrated HCl or 4 M NaOH.

D. 0.02 M Tris-HCl buffer (pH8.0, for Pgm only)

dH_2O	1.0	l
Tris	2.422	g

Adjust pH with concentrated HCl.

E. 0.2 M Tris-HCl buffer (pH8.5)

dH_2O	1.0	l
Tris	24.22	g

Adjust pH with concentrated HCl.

F. 0.025 M Tris-maleate buffer (for Lap only)

dH_2O	1.0	l
Tris	12.11	g
Maleic anhydride	6.698	g

Adjust pH with 4M NaOH to pH5.2. Add 25 mls of this stock buffer to 75 ml dH_2O to make 0.025 M staining buffer before staining.

G. 0.1 M Tris-maleate buffer (pH5.5, for Enp only)

dH_2O	1.0	l
Tris	24.22	g
Maleic acid	23.22	g

Adjust pH to 3.7 with HCl or NaOH. Use this as stock solution. The final staining buffer can be made of 5 parts of the stock solution, 3 parts of dH_2O and 2 parts of 0.2 M NaOH just before use. The pH of the final solution is around 5.5.

Stock Solutions.

A. 7% Acetic acid (for Ti only)

dH_2O	930	ml
Acetic acid	70	ml

B. 1% Alpha-naphthyl acetate (for Est only)

Acetone	50	ml
Alpha-naphthyl acetate	1.0	g
dH_2O	50	ml

Dissolve alpha-naphthyl acetate in acetone and add dH_2O to the solution. Put the solution in a sealed container to avoid precipitation.

C. 0.05% Aniline blue solution (for Ti only)

7% Acetic acid	100	ml
Aniline blue	0.05	mg

D. 1% cis-Aconitic acid (pH7.5, for Aco only)

dH_2O	100	ml
cis-Aconitic acid	1.0	g

Adjust pH with 4 M NaOH.

E. Potassium iodine solution (for Am only)

dH_2O	100	ml
Iodine (0.1%, w/v)	0.10	g

Potassium iodine (0.5%)	0.50 g

The chemicals can be dissolved by stirring in an aluminum-foil-covered beaker for several hours. The solution should be stored in an amber bottle.

Preparations of Enzymes.

Three enzymes are used in the staining solutions: Glucose-6-phosphate dehydrogenase (Gpd, NAD active) from Leuconostoc mesenteroids, phosphoglucose isomerase (Pgi) from Bakers yeast and isocitric dehydrogenase (Idh) from porcine heart. Upon arrival, these enzymes are suspended in refrigerated dH_2O to make 40 units/ml solutions. Then one ml of the solutions is pipetted into each of small plastic vials. The vials are stored in a freezer for use.

GELS

A. 12.5% Starch (2 layer, for Aco only)

0.005 M L-Histidine buffer	240	ml
Starch	30	g
NADP	15	mg

Heat the starch solution in a side-armed flask in a water bath on a hot plate with a large magnetic bar stirring constantly. When the thick solution reaches 78-80°C, usually after the starch becomes clear, NADP is added and mixed well. Then, the solution is degassed for 40 seconds via the side arm of the flask by a vacuum aspiration apparatus of tap water and poured into a gel mold. The gel is covered by a 20 X 18 X 0.5 cm glass.

B. 7% Acrylamide (single layer, for Am only)

0.005 M L-Histidine buffer	150	ml
Acrylamide	9.975	g
Bis	0.525	g
APS	0.15	g
TEMED	0.30	ml

Combine the buffer and all the chemicals except for TEMED, and heat the solution to 30°C on a hot plate, again with a magnetic bar stirring constantly. Then, add TEMED with a pipetman and pour the gel solution into a gel mold. Be cautious in handling of APS, which causes burning and smoking when water drops and other

chemicals are also present. Add APS reasonably shortly before heating. Otherwise, the solution will solidify if the beaker sits there too long.

C. 9% Acrylamide (single layer, for Ti only)

0.005 M L-Histidine buffer	150	ml
Acrylamide	12.82	g
Bis	0.675	g
APS	0.15	g
TEMED	0.30	ml

Follow the same procedure as for gel B.

D. 7% Acrylamide + 2% starch

Number of layers	single	two	three	four	
0.005 M L-Histidine buffer	150	210	280	350	ml
Acrylamide	9.975	13.965	18.62	23.275	g
Bis	0.525	0.735	0.98	1.225	g
APS	0.15	0.21	0.28	0.35	g
Starch	3.0	4.2	5.6	7.0	g
TEMED	0.30	0.42	0.56	0.70	ml

Add acrylamide, Bis and APS and half amount of the gel buffer in a beaker, and stir the chemicals with a magnetic bar on a stirring plate to dissolve them. Add the starch and the other half amount of the gel buffer into a side-armed flask and heat the starch to 78-80°C as for gel A. After the starch is degassed, it is poured into the beaker. Then the solution is mixed well, with TEMED being added at the same time. Pour the gel into a gel mold. For Adh requiring NAD in the gel, 30 mg NAD is added in the beaker with acrylamides.

E. 6% acrylamide + 4% starch (single layer)

0.005 L-Histidine buffer	150	ml
Acrylamide	8.55	g
Bis	0.45	g
APS	0.15	g
Starch	6.0	g
TEMED	0.30	ml

Follow the same procedure as for gel D. For Sdh gel, 15 mg NADP is added in the beaker with acrylamides, the same way as for Adh gel.

Allow the gels to cool to room temperature and to dry

on the surface. The gels are wrapped in plastic wrap and seasoned in a refrigerator for several hours before use.

STAINING

Aco

0.2 M Tris-HCl buffer	50	ml
MTT	10	mg
NADP	10	mg
$MgCl_2$	10	mg
1% cis-Aconitic acid	8	ml
Idh enzyme	40	units
PMS	1	mg

Incubate at 37°C in the dark for 3 hr. Seeds soaked at high temperature seem to give better resolution for bands 1 and 3.

Adh

0.05 M Sodium phosphate buffer	50	ml
MTT	15	mg
NAD	15	mg
95% ethanol	5	ml
PMS	2	mg

Incubate at 37°C in the dark for 3 hr.

Am

Heat the solution of 1% (w/v) soluble potato starch in 0.1 M acetic acid on a hot plate with a magnetic bar stirring until the starch gets dissolved, i.e. clear. Cool the solution to 30°C, pour it on the gel and incubate the gel for 15-30 minutes at 37°C in the dark. Then, rinse the gel with dH_2O. Add 10 ml of the potassium iodine solution on the surface of the gel. Bands show up immediately.

Ap

0.1 M Acetate buffer	50	ml
Sodium alpha-napthyl acid phosphate	40	mg
Fast black K salt	40	mg

Incubate at room temperature in the dark for 2 hr.

Dia

0.2 M Tris-HCl buffer	50	ml
MTT	10	mg
NADH	10	mg
2,6 Dichlorophenol indophenol	2	mg

Incubate at room temperature in the dark for 2 hr.

Enp

0.1 M Tris-maleate buffer	50	ml
$MgCl_2$	10	mg
Fast black K salt	20	mg
N-alpha-benzoyl-DL-arginine-beta-naphylamide hydrochloride (BANA)	20	mg

Incubate at 37°C in the dark for 3 hr.

Est

0.1 M Sodium phosphate buffer	50	ml
Fast blue RR salt	0.1	g
Alpha-naphthyl butyrate	0.05	cc
1% Alpha-naphthyl acetate	1	ml

Add alpha-naphthyl butyrate with a syringe to the fast blue RR salt in a beaker and drop a few drops of acetone over them. Then add 50 ml of 0.1 M sodium phosphate buffer. Heat the beaker with a magnetic bar stirring vigorously. Meanwhile 1% alpha-naphthyl acetate is added. Pour the solution to the gel through a cheesecloth. Incubate at 37°C in the dark for 2 hr.

Eu

0.1 M Acetate buffer	50	ml
Cresol red (sodium salt) (0.1%, w/v)	50	mg

Incubate the gel in this cresol red solution at room temperature in the dark for 30 minutes or more.

dH_2O	50	ml
Urea (333 mM)	1	g
EDTA (0.1%, w/v)	50	mg
Cresol red (sodium salt) (0.1%, w/v)	50	mg

Replace the incubation solution with this solution and re-incubate the gel at room temperature in the dark for 30 minutes or more.

Fle

0.1 M Acetate buffer	40	ml
4-methyl umbelliferyl acetate	15	mg

First dissolve 4-methyl umbelliferyl acetate in 10 ml acetone in a beaker. Then 0.1 M aceate buffer is added to the beaker. Use either Kim wipes or paper towels to cover the gel. Pour the solution over the Kim wipes or paper towels. Let the gel be stained for 15 minutes. Examine the gel in the dark under

a UV light (366 nm) source. Wear UV light protective glasses when examining Fle bands

Got

0.2 Tris-HCl buffer	50	ml
Pyridoxal-5'-phosphate	25	mg
L-Aspartic acid (monosodium salt)	272	mg
Keto-glutaric acid	36	mg
Fast blue BB salt	112	mg

Incubate at 37°C in the dark for 3 hr.

Idh

0.2 M Tris-HCl buffer	50	ml
MTT	10	mg
NADP	10	mg
$MgCl_2$	120	mg
DL-Isocitric acid (trisodium salt)	200	mg
PMS	1	mg

Incubate at 37°C in the dark for 3 hr. Seeds soaked at high temperature and gels run at low temperature may be necessary for clear bands 3 and 4.

Lap

0.025 M Tris-maleate buffer	100	ml
L-Leucine-beta-naphthyl-amide	20	mg

First dissolve L-leucine-beta-naththyl-amide in 1 ml of ethanol. Add this to the buffer. Incubate the gel at 37°C in the dark for two hours. Then add 50 mg fast black K salt to the solution and re-incubate the gel at room temperature in the dark for 1 hr. The top layer of D gel containing NAD seems somehow to give better bands.

Mpi

0.2 M Tris-HCl buffer	50	ml
MTT	10	mg
NAD	15	mg
D-Mannose-6-phosphate (barium salt)	20	mg
Gpd enzyme (NAD active)	40	units
Pgi enzyme	40	units
PMS	1	mg

Incubate at 37°C in the dark for 3 hr. The middle layer of the multiple layer gels is needed for Mpi.

Pgd

0.2 M Tris-HCl buffer	50	ml
MTT	10	mg

	NADP	10	mg
	$MgCl_2$	20	mg
	6-Phosphogluconic acid (trisodium salt)	15	mg
	PMS	1	mg

Incubate at 37°C in the dark for 1 hr.

Pgi

0.2 M Tris-HCl buffer	50	ml
MTT	10	mg
NAD	10	mg
$MgCl_2$	20	mg
D-Fructose-6-phosphate (disodium salt)	30	mg
Gpd enzyme (NAD active)	40	units
PMS	1	mg

Incubate at room temperature in the dark for 1 hour. To get the best results concerning Pgi2 gene, first incubate the gel in a refrigerator for one hour.

Pgm

0.02 M Tris-HCl buffer	50	ml
MTT	10	mg
NAD	10	mg
$MgCl_2$	20	mg
alpha-D-Glucose-1-phosphate (disodium salt)	125	mg
Gpd enzyme (NAD active)	40	units
PMS	1	mg

Incubate at 37°C in the dark for 2 hr. For clear band 1, run the gel at low temperature and use the middle layer of the multiple layer gels.

Sdh

0.2 M Tris-HCl buffer	50	ml
MTT	15	mg
NADP	15	mg
(-) Shikimic acid	15	mg
PMS	2	mg

Incubate at 37°C in the dark for 3 hr.

Ti

Rinse the gel in 0.05% aniline blue solution for about 5 minutes until the light purple bands show up at lower part of the gel. Then, replace this solution, and de-stain with 7% acetic acid for 5 hours or more for clear Ti bands on upper part of the gel.

APPENDIX II

THE SQUARED EUCLIDEAN DISTANCE FOR ISOZYME DATA

Table 1. The squared Euclidean distance for isozyme data[a].

Population	A	B	C	D	E	F	Mean
B	51.03						
C	77.93	88.70					
D	51.76	62.90	86.88				
E	68.36	72.92	84.24	51.28			
F	75.17	95.91	106.89	74.82	61.21		
Mean	64.85	74.29	88.93	65.53	67.60	82.80	74.00

[a] Not divided by the number of allele variables.

APPENDIX III

CANONICAL DISCRIMINANT ANALYSIS FOR MORPHOLOGICAL DATA

Table 1. Canonical variables.

Canonical variable	Canonical correlation	Eigen-value	Proportion of variance	Approx. F	Pr>F
(1) 1989 morphological data					
CAN1	0.846	2.518	0.349	3.448	0.0001
CAN2	0.792	1.679	0.233	3.043	0.0001
CAN3	0.753	1.313	0.182	2.789	0.0001
CAN4	0.707	0.999	0.139	2.532	0.0001
CAN5	0.642	0.701	0.097	2.260	0.0101
(2) 1990 morphological data					
CAN1	0.912	4.954	0.407	3.731	0.0001
CAN2	0.868	3.057	0.251	3.063	0.0001
CAN3	0.828	2.178	0.179	2.553	0.0001
CAN4	0.736	1.179	0.097	2.011	0.0022
CAN5	0.668	0.806	0.066	1.759	0.0516

Table 2. Total canonical structure[a].

	CAN1		CAN2		CAN3		CAN4	
Var[b]	1989	1990	1989	1990	1989	1990	1989	1990
S1	0.494	0.587	0.593	-0.016	0.035	-0.394	-0.162	-0.003
S2	0.298	-0.162	0.462	0.060	-0.077	-0.089	-0.163	-0.104
S4	0.196	-0.106	0.297	0.280	0.266	0.284	0.496	-0.334
S5	-	0.055	-	0.334	-	-0.297	-	0.050
V7	0.582	-0.415	0.065	-0.396	0.016	0.349	-0.497	0.037
V9	0.249	-	-0.139	-	-0.459	-	-0.065	-
V10	0.423	0.204	0.338	0.425	0.428	0.386	-0.108	0.189
V11	0.149	-0.219	0.415	0.143	-0.086	0.232	0.075	-0.291
V13	0.211	-0.028	0.197	0.306	0.234	0.339	-0.181	-0.041
V14	0.328	0.322	-0.151	0.638	0.532	0.173	0.111	0.127
V15	-	-0.040	-	0.238	-	0.090	-	-0.340
V17	-	0.141	-	0.328	-	0.125	-	0.039
V18	-0.164	-	0.163	-	-0.226	-	-0.061	-
V19	-0.306	0.260	0.113	0.442	0.406	-0.231	0.314	-0.111
V20	0.510	-0.353	0.042	-0.243	-0.107	0.325	-0.482	0.299
V21	0.412	-0.226	-0.084	-0.586	-0.301	0.289	-0.590	0.185
V22	-0.465	0.433	-0.373	0.221	0.080	-0.232	0.313	0.234
R23	-	0.070	-	0.346	-	0.201	-	0.197
R24	-	0.024	-	0.195	-	0.354	-	0.237
R25	-	0.548	-	-0.078	-	0.587	-	-0.150
R26	0.490	0.509	-0.310	0.390	0.150	0.247	0.127	-0.009
R27	-0.297	0.429	-0.361	-0.033	0.046	0.015	0.357	0.123
R30	-	0.354	-	0.017	-	-0.284	-	0.193
R32	0.531	0.328	-0.376	0.070	-0.331	0.459	0.037	-0.214
R33	-0.445	-0.462	0.494	-0.049	0.313	-0.541	-0.080	0.094
R34	-0.446	-	-0.065	-	0.149	-	0.047	-
R36	0.297	0.260	0.372	0.199	0.222	0.246	-0.008	-0.161
R37	-0.040	0.342	-0.037	0.225	0.049	-0.168	0.633	0.229
R39	0.149	-0.055	0.199	0.418	0.561	0.148	-0.107	0.087

[a] Twenty two variables of 1989 data and 26 of 1990 were included in the canonical discriminant analyses.

[b] The following transformations were made (* represents the original variable): S1(1989)=SQRT(119-*); V7(1989)=1/*; V7(1990)=1/*; V21(1990)=LOG10(*); R34(1989)=SQRT(*).

Table 3. Pooled within canonical structure[a].

	CAN1		CAN2		CAN3		CAN4	
Var[b]	1989	1990	1989	1990	1989	1990	1989	1990
S1	0.344	0.311	0.473	-0.010	0.030	-0.285	-0.149	-0.003
S2	0.179	-0.068	0.319	0.030	-0.057	-0.051	-0.130	-0.072
S4	0.120	-0.048	0.209	0.154	0.201	0.177	0.404	-0.251
S5	-	0.025	-	0.183	-	-0.184	-	0.038
V7	0.392	-0.217	0.050	-0.251	0.013	0.250	-0.444	0.032
V9	0.155	-	-0.099	-	-0.352	-	-0.053	-
V10	0.271	0.100	0.249	0.254	0.339	0.260	-0.092	0.154
V11	0.086	-0.099	0.274	0.078	-0.061	0.143	0.058	-0.216
V13	0.120	-0.013	0.128	0.165	0.164	0.207	-0.136	-0.030
V14	0.204	0.176	-0.107	0.423	0.408	0.130	0.092	0.115
V15	-	-0.018	-	0.127	-	0.054	-	-0.247
V17	-	0.061	-	0.172	-	0.074	-	0.028
V18	-0.091	-	0.103	-	-0.154	-	-0.045	-
V19	-0.188	0.124	0.080	0.256	0.308	-0.151	0.256	-0.088
V20	0.328	-0.173	0.031	-0.145	-0.085	0.219	-0.410	0.243
V21	0.275	-0.119	-0.065	-0.375	-0.248	0.209	-0.523	0.161
V22	-0.296	0.207	-0.272	0.128	0.063	-0.152	0.264	0.184
R23	-	0.031	-	0.187	-	0.123	-	0.145
R24	-	0.011	-	0.108	-	0.221	-	0.179
R25	-	0.324	-	-0.056	-	0.476	-	-0.146
R26	0.309	0.267	-0.224	0.247	0.117	0.177	0.107	-0.008
R27	-0.184	0.194	-0.257	-0.018	0.035	0.010	0.293	0.092
R30	-	0.160	-	0.009	-	-0.176	-	0.145
R32	0.353	0.157	-0.286	0.041	-0.272	0.300	0.033	-0.169
R33	-0.302	-0.242	0.384	-0.031	0.262	-0.388	-0.072	0.081
R34	-0.262	-	-0.044	-	0.108	-	0.036	-
R36	0.176	0.115	0.253	0.107	0.162	0.149	-0.006	-0.117
R37	-0.024	0.155	-0.025	0.124	0.037	-0.104	0.503	0.172
R39	0.090	-0.025	0.139	0.226	0.420	0.090	-0.086	0.064

[a] Twenty two variables of 1989 data and 26 of 1990 were included in the canonical discriminant analyses.

[b] The following transformations were made (* represents the original variable): S1(1989)=SQRT(119-*); V7(1989)=1/*; V7(1990)=1/*; V21(1990)=LOG10(*); R34(1989)=SQRT(*).