Winter 1985

GENETIC AND QUANTITATIVE VARIATION IN WILD SOYBEAN (GLYCINE SOJA) POPULATIONS

YUEH-CHIN CHIANG
University of New Hampshire, Durham

INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

***For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.**

Chiang, Yueh-Chin

GENETIC AND QUANTITATIVE VARIATION IN WILD SOYBEAN (GLYCINE SOJA) POPULATIONS

University of New Hampshire PH.D. 1985

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __✓__.

1. Glossy photographs or pages __✓__
2. Colored illustrations, paper or print _____
3. Photographs with dark background __✓__
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages __✓__
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received _____
16. Other _____

_____

_____

University
Microfilms
International

GENETIC AND QUANTITATIVE VARIATION IN WILD SOYBEAN (*GLYCINE SOJA*) POPULATIONS

BY

YUEH-CHIN CHIANG
B.S., Chung-Hsing University, 1972
M.S., University of New Hampshire, 1981

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, 1985

This dissertation has been examined and approved.

Yun-Tzu Kiang

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

Owen M. Rogers

Owen M. Rogers, Professor of Plant Science
and Genetics

James B. Loy

James B. Loy, Professor of Plant Science
and Genetics

Robert T. Eckert

Robert T. Eckert, Associate Professor of
Forestry and Genetics

Thomas D. Lee

Thomas D. Lee, Assistant Professor of Botany

Nov. 8, 1985

Date

ACKNOWLEDGEMENTS

I like to acknowledge Dr. Yun Tzu Kiang for his guidance and advice; Dr. Owen M. Rogers, Dr. James B. Loy, Dr. Robert T. Eckert and Dr. Thomas D. Lee for their advice and criticism of the manuscript.

I thank Mr. David Liao who helped with data computing. I also thank Hock-Ann and Bang-yen for their encouragements during this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

GENETIC AND QUANTITATIVE VARIATION
IN WILD SOYBEAN (*GLYCINE SOJA*) POPULATIONS

BY

YUEH-CHIN CHIANG
University of New Hampshire, December 1985

Glycine soja seeds collected from South Korea and Japan were studied for their genetic structure and quantitative variation. Seventy two *G. soja* seed accessions were examined for genetic variation by gel electrophoresis. Based on a total of 43 loci of 15 enzymes and one protein, an average of 67.4% polymorphism (at 99% level) was observed. The expected heterozygosity is 0.160, and the number of alleles per locus is 2.14. Genetic purity of these seed accessions are very high.

The four natural *G. soja* populations collected along the Kitakami river, Japan, showed 38.1% polymorphic loci, and the average number of alleles per locus is 1.55. Their overall expected heterozygosity is 0.114 which is much higher than the average observed heterozygosity (0.023). 19.8% of gene diversity of the four natural populations resides among populations and 81.2% within populations. The G_{ST} value calculated from isozyme and protein variation indicates that the four local populations are well differentiated isoenzymatically. The average Nei's genetic distance between each pair of these four populations is 0.044.

Four enzyme loci, *Ap*, *Ti*, *Lap1* and *Pgd2*, belong to Linkage Group 9 of soybeans. The gene order of the four loci is *Ap-Ti-Lap1-Pgd2*, and the recombination frequency between each pair of loci *Ap-Ti* is 9.41% $\pm$ 1.07%, *Ap-Lap1* 22.65% $\pm$ 0.73%, *Ap-Pgd2* 39.78% $\pm$ 0.98%, *Ti-Lap1* 17.96% $\pm$ 1.33%, *Ti-Pgd2* 36.85% $\pm$ 1.39%, and *Lap1-Pgd2* 20.62% $\pm$ 0.99%. Another gene pair, *Pgi1-Pgd2*, is linked with a recombination frequency of 15.34% $\pm$ 0.74%.

Twelve G. soja accessions were selected based on their latitudinal locations to study their genetic and quantitative variations. The measure of quantitative variation consistes of morphological, agronomic, and phenological traits. These 12 seed accessions are significantly different from one another in most of the quantitative traits measured. The principal component analysis shows about 60% of the phenological variation and about 43% of the agronomic variation among these 12 G. soja seed accessions are highly associated with their latitudinal locations. Among three reproduction components only the number of pods per plant is positively correlated with the total yield per plant. The average number of nodules per plant is positively correlated with the percent of 3-seed pods and 4-seed pods. The congruence of population differentiation based on genetic (protein) and quantitative variations of these 12 seed accessions is low.

I. INTRODUCTION

Genetic diversity is essential to success in cultivar development for high yield, wide adaptation, desirable quality, and pest and disease resistance. The naturally existing wild progenitors of crop plants are one potential source of such diversity (Harlan, 1976). An efficient and complete collection, evaluation and use of these wild germplasms would be greatly enhanced by information such as the amounts of genetic variation that exists in these wild relatives, how the variation is apportioned within and between populations, and whether the variability patterns are random. With gel electrophoresis and biochemical staining, enzymes — the gene products — can be directly separated and visualized. Electrophoretic surveys have hence been suggested as the most convenient mode for assessing genetic variation for wild genetic resources (Brown, 1978), and have recently been made in Lycopersicon pimpinellifolium (Rick et al., 1977), Hordeum spontaneum (Brown et al., 1978; Nevo et al., 1979) and Triticium dicoccoides (Nevo et al., 1982)

There are several reasons why wild soybean, Glycine soja Sieb & Zucc., was chosen for the present genetic diversity study. First, this species is the recognized

progenitor of modern cultivated soybeans, Glycine max (L.) Merrill. Both wild and cultivated species have a somatic chromosome number of 2n = 40, and hybrids between them are interfertile (Hymowitz, 1976). Soybean is one of the most important crops grown in the U.S. as well as the world. However, increase in seed yield in soybeans by breeding and selection has been slow due to its narrow genetic background (Johnson and Bernard, 1963). Since G. soja can be readily crossed with G. max, has higher DNA content (Yamamoto and Nagato, 1984), seed protein percentage (Kaizuma and Fukui, 1974) and total phosphorus, zinc and calcium content (Raboy et al., 1984) than the cultivated soybeans, it is a promising exotic genetic resource in soybean improvement. In fact, breeding experiments have been conducted and reported using G. soja as a resource of disease resistance (Ram et al., 1985) and producing soybean lines with higher seed protein (Erickson et al., 1981; Kaizuma et al., 1980).

Second, the genetics of several isozyme variants in G. max and G. soja has been extensively studied in our Lab as well as others. It is essential to have genetic interpretation of the number of loci and alleles involved in variation of zymograms observed for each enzyme before genetic diversity can be estimated with electrophoresis. The inheritance of some of the observed variants in soybeans for alcohol dehydrogenase, amylase, tetrazolium oxidase and acid phosphotase was reported by Gorman and Kiang (1978), Hildebrand et al. (1980), Hildebrand and Hymowitz (1980),

and Kiang (1981). The genetic basis of some variants observed for an additional nine enzymes was further worked out and reported (Gorman, 1983; Kiang and Gorman, 1983; Gorman et al., 1983; Kiang and Gorman, 1985; Kiang et al., 1985). Still there are zymogram variations observed in soybeans need genetic interpretation, and new enzyme systems need to be surveyed in order to obtain more accurate estimation of genetic diversity. Furthermore, these electrophoretic enzyme loci are valuable as markers in chromosome mapping, and the gene organization on chromosomes will be useful in the application of tissue culture and genetic engineering for soybean improvement in the future.

Third, G. soja grows in the wild, has not been subjected to artificial selection, and is therefore, an ideal material for the comparison of the amount of genetic variation within and between populations based on quantitative and biochemical (enzyme) variation. Since these two sets of characters merely represent two pictures of the same organism, the degree of their concordance may reveal the degree of association between the enzyme and the quantitative characters studied.

II. OBJECTIVES

1. To study the inheritance of observed electrophoretic variants of *Glycine soja* and *G. max*.

2. To study the genetic linkage of protein loci in soybeans.

3. To examine genetic structure of *G. soja* by gel electrophoresis.

4. To estimate quantitative variation between *G. soja* populations.

5. To compare the estimates of population differentiation based on quantitative and genetic (protein) variations in *G. soja*.

III. MATERIALS AND METHODS

1. THE PLANT

The wild soybean (*Glycine soja* Sieb & Zucc., formerly *G. ussuriensis* Regel & Maack) is an annual, twining vine with trifoliolate leaves, purple flowers (predominantly selfing), and small, hard, black to dark brown seeds. It grows wild in the Yangtze River Valley, the northern and northeastern provinces of China and adjacent areas of the USSR, Korea, Japan and Taiwan (Hymowitz, 1970).

2. SEED SOURCES

The *G. max* seeds used in this study were provided by Dr. R. L. Bernard, USDA, ARS and Agronomy, University of Illinois, and Dr. E. E. Hartwig, Delta Branch Exp. Stn., Stonville, Mississippi. The *G. soja* seeds were obtained from Dr. S. Shanmugasundaram of the Asian Vegetable Research and Development Center (AVRDC) in Taiwan, Dr. Young Soo Ham of Crop Experiment Station, Suweon, South Korea, Dr. N. Kaizuma of the Iwate University, Japan and Dr. H. I. Oka of the National Institute of Genetics, Japan. The accessions of G. soja and their PI (Plant Introduction) numbers are listed in Appendix I.

Among these seed materials, some of the G. soja seeds were collected from four natural populations near the Kitakami river in Iwate Prefecture, Northern part of Japan Honshu Island by Dr. N. Kaizuma.

Besides these four populations, the collection method for each seed accession examined in this study is unknown. Seeds in the same accession may be descended from a single or few wild plants and have been maintained in farms and gardens. Forty-one accessions of South Korean origin were obtained from AVRDC, Taiwan where they had been propagated. The rest of 43 South Korean seed accessions were provided directly from South Korea by Dr. Young Soo Ham. To separate these two batches of seed accessions, the AV#s' for those seeds received from AVRDC, Taiwan are also presented in the Appendix I. Many accessions from these two places have common PI numbers, and are descended from the same populations. Therefore, those seeds which were obtained from two different places but bear the same PI number were treated as one seed accession in the genetic structure study.

3. ELECTROPHORETIC PROCEDURES

All the methods described in this electrophoretic procedures are as described by Gorman and Kiang (1977), Kiang and Gorman (1983), Kiang and Gorman (1985) and Kiang and Chiang (1985) with some modifications.

(1). Sample preparation

Cotyledons of dry seed were used for the electrophoretic study. A piece of cotyledon (seed coat removed) was soaked in 2 - 3 drops of 0.005M L-histidine (HCl) buffer (pH 7.0) for 3 - 8 hrs. prior to grinding. A 1 X 1 cm square of lens paper was placed on top of the ground sample to serve as a filter, and appropriate size of bibulous paper was placed on lens paper as a sample wick. Each wick was inserted into one of 26 slots cut perpendicularly into the electrophoretic gel.

(2). Gel preparation

Five types of horizontal slab gels, differing in the concentrations of gelling agent and starch, were used. The gelling agent was prepared as 95% acrylamide and 5% N,N'-methylene-bis-acrylamide. L-histidine (HCl) 0.005M, pH 7.0 was used as gel buffer. The concentration of ammonium persulfate (AMP) and N,N,N',N'-tetramethylethylene diamine (TEMED) were 0.1% (W/V) and 0.2% (V/V) of the gel buffer solution, respectively. The gel molds were 18cm X 15.5cm with 3mm depth for staining one enzyme. Various depth of molds can be used depending on the number of enzymes to be examined. For example, if two enzymes are to be run in the same gel, a 5 - 6 mm deep gel mold can be used for making a gel which after electrophoretic run can be sliced into two gels for staining two different enzymes. The five types of gels used are:

(a) 7% (W/V) acrylamide gel —— 7% gelling agent + AMP +

TEMED. This gel was only used in amylase allozyme study. All the ingredients except for TEMED were mixed together with gel buffer and heated to 30 - 32 °C. The TEMED was added and mixed immediately prior to pouring the gel into the plexiglass gel mold.

(b) 9% (W/V) acrylamide gel —— 9% gelling agent + AMP + TEMED. The preparation procedure was same as described above for 7% acrylamide gel. This gel was only for soybean Kunitz trypsin inhibitor (TI) study.

(c) 7% (W/V) acrylamide + 2% (W/V) starch gel —— 7% gelling agent and AMP were dissolved in about 2/3 to 1/2 of the total amount of gel buffer, and maintained cool (4 - 5 °C). 2% starch was put in the rest amount of gel buffer, and heated to 78 - 80 °C, then, degassed, and added to the cold gelling agent buffer solution. The solution was immediately poured into gel mold after TEMED was added.

(d) 6% (W/V) acrylamide + 4% (W/V) starch gel —— This polyacrylamide and starch mixed gel contains 6% of gelling agent, 4% of starch and proper amount of AMP and TEMED. The preparation procedure was same as described for the type (c) gel.

(e) 12.5% (W/V) starch gel —— 12.5% starch in gel buffer was cooked in a water bath to 78 - 80 °C, degassed, then, poured into a gel mold.

All the gels were cooled in the refrigerater at least 8 hrs. before using.

(3). Electrophoresis

Tris-citrate buffer (0.13M, pH 7.0) was used as tray buffer. Gels were electrophoresed at 150 or 200 volts using a constant voltage power supply (ISCO apparatus). Gels were placed on and covered with ice packs during electrophoresis to maintain gel temperature around 4 °C. Electrophoretic runs last between 3 and 24 hours depending on the gel type, the gel thickness, the proteins to be examined, and the voltage selected. Methylene blue (1%) solution was used as dye marker for calculating the Rf value.

(4). Enzymes and protein staining

After electrophoresis, the gel was sliced horizontally into several slices depending on the thickness of the gel. Each slice was then stained for a specific enzyme or protein. A total of fifteen enzymes and one protein were studied. They are: aconitase (ACO, EC 4.2.1.3), alcohol dehydrogenase (ADH, EC 1.1.1.1), amylase (AM, EC 3.2.1.2), acid phosphotase (AP, EC 3.1.3.2), diaphorase (DIA, EC 1.6.2.2), endopeptidase (ENP), glutamate oxaloacetic transaminase (GOT, EC 2.6.1.1), isocitrate dehydrogenase (IDH, EC 1.1.1.42), leucine aminopeptidase (LAP, EC 3.4.11.1), malate dehydrogenase (MDH, EC 1.1.1.37), mannose-6-phosphate isomerase (MPI, EC 5.3.1.8), 6-phosphogluconate dehydrogenase (PGD, EC 1.1.1.44), phosphoglucose isomerase (PGI, EC 5.3.1.9), phosphoglucomutase (PGM, EC 2.7.5.1), shikimate dehydrogenase (SDH, EC 1.1.1.25), and Kunitz

trypsin inhibitor (TI). The staining methods of Shaw and Prasad (1969), Brewbaker et al. (1968), O'malley et al. (1980), Gorman and Kiang (1977, 1978) and Gorman (1983) were followed with some modifications, and methods of Kiang and Gorman (1985) for IDH, Kiang and Chiang (1986) for TI, Doong (personal communication) for ACO and ENP. Specific staining method and gel type used for each enzyme studied and for TI are listed in the Appendix II.

4. GENETIC INHERITANCE OF ENZYME VARIATION

Crosses between plants with different zymogram variants were made. The F_1, F_2 and F_3 seeds were examined electrophoretically for the zymograms and segregation ratios of the variants. The observed F_2 and F_3 segregation ratios were tested with the Chi square test against the hypothesized ratios. Since G. soja and G. max can be intercrossed and produce fully viable offsprings, the study of inheritance of variant zymograms were based on crosses between G. soja and G. soja, or G. soja and G. max, or G. max and G. max. The results of the genetic analysis from the progeny of intraspecific and interspecific crosses were designated as inheritance of soybean zymograms.

5. LINKAGE STUDY

The F_2 seeds segregating for two or more electrophoretic loci were used to test linkage of those loci. Homogeneity

tests were performed when data from two or more crosses were to be combined for estimating linkage strength. The frequency of recombination between two linked loci was calculated using the maximum likelihood method (Allard, 1956).

6. GENETIC STRUCTURE OF GLYCINE SOJA

(1). Genetic diversity of G. soja

Eight to more than fifty seeds per accession were examined for zymograms of the fifteen enzymes and TI (Appendix II). The genetic information of G. soja gathered in this study was pooled together based on the seed source to estimate genetic diversity of G. soja in South Korea and Japan, respectively. Data from the above two sources were also pooled to analyze the total genetic variation of the species. The following parameters of genetic variation were estimated:

(a). Allele frequency

(b). Proportion of polymorphic loci. Both 95% and 99% polymorphism level were tested.

(c). The average number of alleles per locus was calculated by dividing the total number of alleles observed by the total number of loci examined. This is a measure of allelic richness of a population.

(d). Expected heterozygosity. The expected heterozygosity at each locus was the theoretical Hardy-Weinberg heterozygosity of a population for that locus, and was calculated as

$$He = 1 - \sum x_i^2$$

where x_i is the frequency of the ith allele. The expected mean heterozygosity ($\overline{He}$) was then calculated by averaging the 'He' over all loci.

(2). Apportionment of genetic variation within and between G. soja populations

In order to understand the organization of the genetic diversity in the total population —— how it is apportioned within the between population —— Nei's (1973) genetic diversity analysis was performed.

In the analysis, seed accessions were grouped into South Korea and Japan, two geographic populations of G. soja. Nei's genetic diversity analysis was also applied to the four natural G. soja populations along the Kitakami river, Japan as well as to the twelve seed accessions of South Korea and Japan, selected on the basis of their latitudinal location as indicated in Figure 1 and Table 1.

(3). Genetic differentiation between G. soja populations

For comparison of genetic differentiation between G. soja populations, the genetic distance and identity of Nei (1972) were calculated. Nei's formula for genetic identity (1972) for a single locus with n alleles is

$$I_N = \frac{J_{xy}}{(J_x J_y)^{\frac{1}{2}}}$$

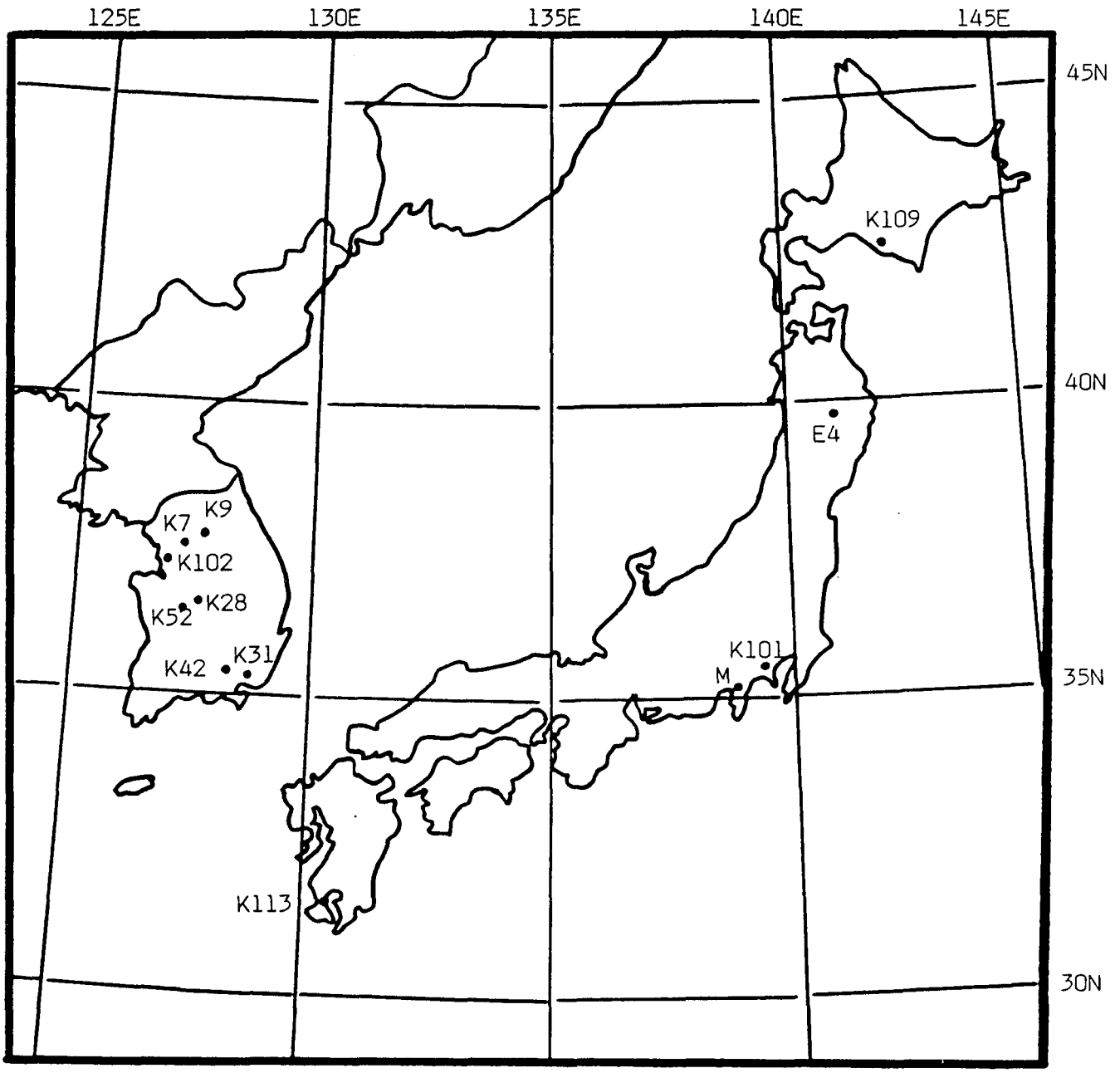


Figure 1. The geographic locations of the twelve G. soja seed accessions of South Korea and Japan.

Table 1. The geographical location of the twelve seed accessions of South Korea and Japan.

Seed Accessions	Collected Place	Latitude
K109 (PI487.430)	Hiratori, Hokkaido, Japan	42.60 N
E4 (PI487.428)	Morioka, Iwate Pref., Japan	39.70 N
K9 (PI407.192)	Chilcheon, Chunseong, South Korea	37.88 N
K7 (PI407.181)	Maseogu, Yangju, South Korea	37.65 N
K102 (PI407.278)	Yongin, Yongin, South Korea	37.28 N
K28 (PI407.223)	Naecheon, Eumseong, South Korea	36.70 N
K52 (PI407.233)	Sinam, Yeongi, South Korea	36.62 N
K42 (PI407.262)	Changyeong, Changneong, S. Korea	35.53 N
K101 (PI487.429)	Noborito, Kanagawa Pref., Japan	35.50 N
K31 (PI407.252)	Milyang, Milyang, South Korea	35.50 N
M (PI486.220)	Mishima, Sizuoka Pref., Japan	35.10 N
K113 (PI487.431)	Ibusuki, Kagoshima Pref., Japan	31.20 N

where $J_{xy} = \sum_{i=1}^{n} P_{i.x}P_{i.y}$

$J_x = \sum_{i=1}^{n} P^2_{i.x}$

$J_y = \sum_{i=1}^{n} P^2_{i.y}$

and $P_{i.x}$ and $P_{i.y}$ are the frequencies of the ith allele in populations X and Y, respectively. The genetic distance between the two populations is then defined as

$$D_N = - \ln (I_N)$$

For multiple loci, J_{xy}, J_x and J_y values are calculated as the arithmetic means over all loci including monomorphic loci.

7. QUANTITATIVE VARIATION OF GLYCINE SOJA POPULATIONS

(1). The quantitative variation of the 12 seed accessions

The 12 seed accessions of different latitudinal origins (Figure 1 and Table 1) were grown in the greenhouse and field to investigate the quantitative variation between accessions.

A completely randomized design with 12 seed accessions, and 20 plants per accession was conducted in the greenhouse at UNH during the summers of 1982 and 1983. Seeds were scarified and inoculated with commercial Rhizobia inoculum before planting. One seed was planted in each of 15 cm diameter plastic pots. Space between pots was 45 cm. A 120cm-long bamboo cane was put up in each pot to support vines. The ambient temperatures in the greenhouse were

about 30°C in the day time and 25°C at night. Steamed field soil mixed with Promix in a one to one ratio was used as growth medium.

A randomized block planting design with four replications, 12 seed accessions, 11 seeds per seed accession per replication was conducted in the Kingman Farm at UNH in the summers of 1982 and 1983. Seeds were also scarified and inoculated before planting. Space between two plants was 60 cm. Bamboo canes, 180cm long, were used to support plants.

The following information on phenological, agronomic and morphological characters were recorded.

(A). Phenological characters:

a. Number of days from sowing to germination.

b. Number of days from germination to the first flower.

c. Number of days between first flower and the first fresh pod set.

d. Number of days between first fresh pod set and the first mature pod.

e. Number of days from the first dry pod to the last dry pod harvested.

f. Number of days between anthesis and the seed maturity.

g. Life span, was defined as number of days between germination and the last dry pod harvested.

(B). Agronomic characters:

a. Percentage of 1-seed pod, 2-seed pod, 3-seed pod and 4-seed pod per plant.

b. Total number of pods per plant.

c. Total number of seed per plant.

d. Average number of seed per pod (total number of seed per plant / total number of pod per plant).

e. Average weight per 100 seed.

f. Total seed yield per plant (by weight).

g. Harvest index, was calculated as total seed weight/ (total seed weight + pod weight + above ground vegetative dry weight at the end of harvest season).

h. Plant height measured at four weeks after seed were planted in 1983 and at three weeks in 1982.

i. Total dry weight at 10 weeks old. Four plants per accession were pulled off, washed off the soil of the root, and dried to obtain the weight of dry matter. Root dry weight, number of nodules were also recorded.

j. Linear regression coefficient 'b' of regressing percent accumulation of number of pods harvested on two-day interval throughout harvesting duration.

(C). Morphological traits:

a. Number of branches recorded after one month (1983) and after two month (1982) from planting.

b. Stem length between ground and the first branch node.

c. Flower size —— width of the banner petal, longitudinal length of flower and length of flower tube.

d. Length and width of 3-seed pod.

e. Pubescence length of four-week-old green pod.

f. Angle of pubescence to pod surface.

g. Pubescence length, density (# pubescence / 1 mm^2) on

mature leaf surface.

h. Angle of pubescence to leaf surface.

i. Length and width of the 10th leaf on main stem (emphasized on the size — length and width).

j. Length and width of leaf, picked randomly (emphasized on the leaf shape — length/width ratio).

k. Seed coat luster (shiny or dull).

(2). Quantitative variation of the four Kitakami river populations

Two seeds per plant, and twelve plants per population were randomly picked and grown individually in 22cm diameter clay pots in the greenhouse at UNH in 1985. Soil and seed preparations were the same as in 1982 and 1983 experiments. Data were recorded about the same way as in 1982 and 1983, but only part of them were included in this report.

8. COMPARISON OF THE ESTIMATES OF POPULATION DIFFERENTIATION BASED ON PROTEIN VARIATION AND QUANTITATIVE VARIATION

(1). Genetic distance among populations based on quantitative variation

Dissimilarity between populations (seed accessions) based on the quantitative characters was measured as Euclidean distance (d^2). The formula for the Euclidean distance between two populations, q and p, in an n-dimensional space is

$$d^2_{pq} = \frac{1}{n} \sum_{i=1}^{n} (x_{ip} - x_{iq})^2$$ (Sneath and Sokal, 1973)

where, n = number of characters measured; x_{ip} and x_{iq} are the measurement of ith character of p and q populations, respectively.

All the data were standardized before d^2 was calculated. Two populations with a $d^2 = 0$ are not different from one another in the quantitative characters studied. The computer program 'CLUSTAN' (Wishart, 1978) was used to compute this dissimilarity coefficient.

(2). Genetic distance among populations based on protein variation

Besides Nei's genetic distance (D_N), the Euclidean distance based on allele frequency was computed as described above. The protein data were also transformed into binary (presence = 1, absence = 0) form to compute a distance between populations based on the formula (Wishart, 1978) below:

$$d = \frac{B + C}{M}$$

where, M = the number of attributes (alleles); B = number of attributes present in population i and absent in population k; C = number of attributes present in population k and absent in population i.

(3). The congruence between quantitative data and protein data in the estimates of population differentiation

(A). The product-moment correlation and Spearman's rank order correlation (Lindeman et al., 1980) were computed

between the population distances estimated by the protein and quantitative data to test their concordance.

(B). The hierarchical grouping method of Ward (1963) was applied to obtain phenograms based on quantitative data and protein data. Ward proposed a hierarchical method which combines those two clusters P and Q whose fusion yields the least increase in the error sum of squares. The error sum of squares is defined as the sum of distance from each individual to the centroid of its parent cluster. Both allele frequency and the binary present-absent transformed allelic data were used to construct the phenograms. The congruence between phenograms was measured using Farris' (1973) mean coefficient of distortion.

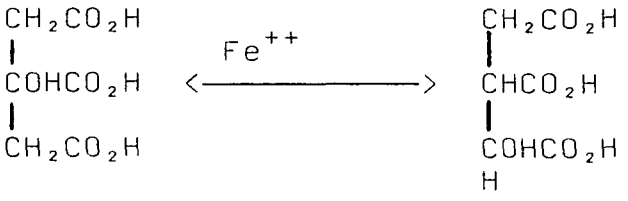
IV. RESULTS AND DISCUSSIONS

1. GENETIC CONTROL OF PROTEIN VARIATIONS

All the zymograms studied in this report were in the anodal regin. For each enzyme, in order for easy discussions, 'zymogram types' were assigned arbitrarily in 'number' for zymograms which had band patterns different from each other. Zymogram bands were also named in 'number', band 1 for the band closest to origin, band 2 the second closest, and so on. Rf values were shown on each zymogram figure as a reference of band migration rate by comparing the distance of bands migration with the distance of methylene blue migration on the same gels. The genetic symbols for loci and alleles were assigned following Rules for Genetic Symbols in Soybean Genetic Newsletter (1985).

Aconitase (ACO, EC 4.2.1.3)

Aconitase catalyzes the isomerization of citric acid and isocitric acid.



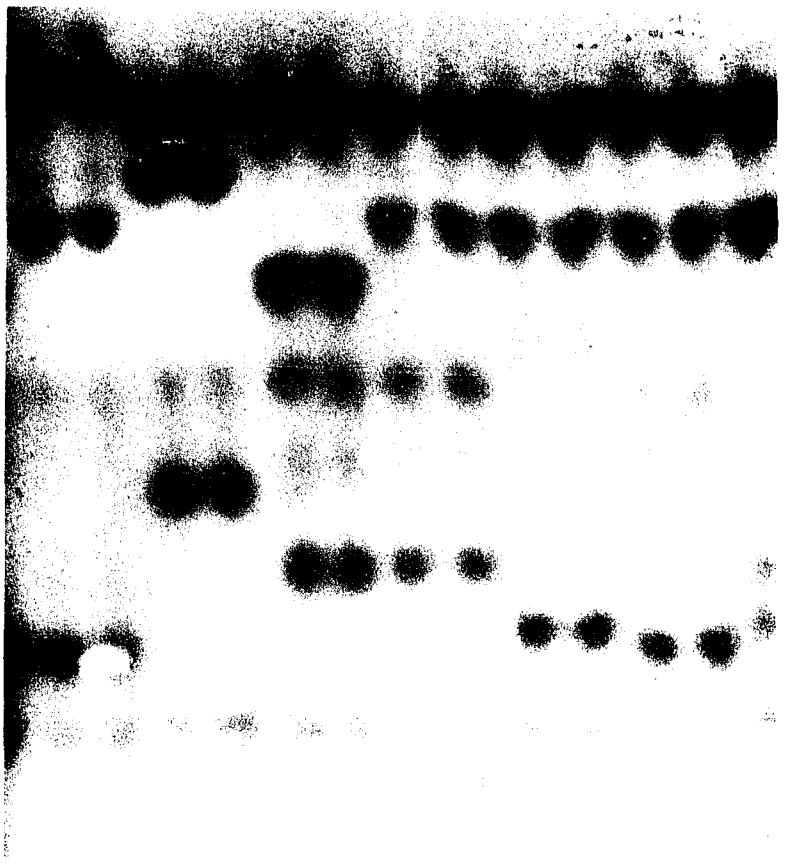
citric acid isocitric acid

It also catalyzes an isomerization between citric acid, isocitric acid, and a third acid, cis-aconitic acid, which is often indicated as an intermediate in the conversion of citric to isocitric acid (Williamson and Corkey, 1969). A monomeric enzyme structure was reported for the aconitase in Slow Lories while a polymeric structure in mouse tissue was suggested (Koen, 1969).

Aconitase zymograms in G. soja showed six anodal bands in mature dry seed (Figure 2). The different homozygous ACO zymograms observed in G. soja resulted from various combinations of five distinct electrophoretic variants. Only parental bands were observed in any hybrid seed, natural or by artificial crosses. Thus, a monomeric enzyme structure and loci with codominant alleles are suggested for the aconitase in G. soja. The inheritance models for ACO described below were hypothesized, and the symbols assigned for the loci and alleles were tentative because of insufficient genetic data. Our Lab is working on the detailed inheritance of these observed variant bands.

We hypothesized that the five electrophoretic variants found in ACO zymograms were controlled by five loci, designated as Aco1 to Aco5. Two variants of different mobility in band 1 (Figure 2), Rf 0.26 (Aco1-b) and Rf 0.21 (Aco1-a) were observed in Aco1. For all the G. soja seed examined, the Aco1-a was found only in Av3094 (PI407.198 - 199) accession.

The second ACO electrophoretic variants involved the



Aco5 —	b/b	a/a	a/a	a/a	a/a	a/a	a/a
Aco4 —	b/b	c/c	a/a	b/b	b/b	b/b	b/b
Aco3 —	a/a	a/a	a/a	a/a	b/b	a/a	b/b
Aco2 —	a/a	d/d	c/c	c/c	b/b	a/a	b/c
Aco1 —	b/b	b/b	b/b	a/a	b/b	b/b	b/b

Figure 2. Zymograms of Aconitase in G. soja mature seed.

mobility of band 2. Four homozygous zymograms with respect to this band were observed, and their Rf values were Aco2-a 0.32, Aco2-b 0.33, Aco2-c 0.37 and Aco2-d 0.43.

The third ACO electrophoretic variants involved the mobility of the third and fourth ACO bands. Two homozygous zymograms were observed with respect to these two bands. One zymogram had these two bands migrated at Rf 0.50 and 0.56 respectively, the other zymogram had them at Rf 0.45 and 0.50. One natural heterozygote showing three bands with Rf 0.45, 0.50 and 0.56 was observed in accession PI407.181. We hence hypothesized that bands 3 and 4 are the products of locus Aco3 with Aco3-a producing bands of Rfs' 0.45 and 0.50, and Aco3-b producing Rfs' 0.50 and 0.56. Since both alleles produce a band with Rf at 0.50, the heterozygote between them showed only three bands. The second hypothesis is that the band at Rf 0.50 is controlled by a separate locus, while at locus Aco3, Aco3-a produces a band at Rf 0.45 and Aco3-b produces a band at Rf 0.56.

Three migrants of different mobility in band 5 as well as two in band 6 were observed. They were hypothesized as the products of the following alleles of loci Aco4 and Aco5: Aco4-a for band at Rf 0.57, Aco4-b at Rf 0.65 and Aco4-c at Rf 0.69, and for band 6 Aco5-a at Rf 0.74 and Aco5-b at Rf 0.79. For all of the G. soja seed examined, the Aco5-b allele was found only in K113 (PI487.431) seed accession from Japan.

Alcohol dehydrogenase (ADH, EC 1.1.1.1)

The main biological reaction of ADH is to mediate the reversible formation of ethanol and NAD from acetaldehyde and NADH.

$$CH_3CH_2OH + NAD \longleftrightarrow CH_3CHO + NADH + H^+$$

Three homozygous ADH zymograms (Figure 3) were observed in G. soja of present study. The first zymogram type has five bands, while the type 2 lacks both bands 1 and 3. These two zymogram types were also observed in G. max. The third type zymogram shows band 1, 3, 4 and 5 with band 2 missing. This type 3 zymogram was observed only in some seeds of K31 (PI407.252 and PI407.253), a G. soja accession from South Korea.

We hypothesized that the bands 1, 3 and 5 were related in a homo-heterodimer relationship, bands 1 and 5 being the respective homodimers of loci designated Adh1 and Adh2, and band 3 being the heterodimer formed by the combination of monomers from Adh1 and Adh2. The band 3 of type 1 migrates approximately halfway between band 1 and 5, as expected for heterodimers in relation to their respective homodimers. The type 2 zymogram appears as a result of a recessive null allele, designated as adh1, being fixed at locus Adh1. Since both bands 1 and 3 were dependent on monomer units produced from Adh1, both bands were lost when the locus was homozygous for the null allele (adh1). The observed 3 to 1 segregation ratio in the F_2 from the reciprocal crosses between type 1 and 2 of G. max plants (Table 2) fitted the

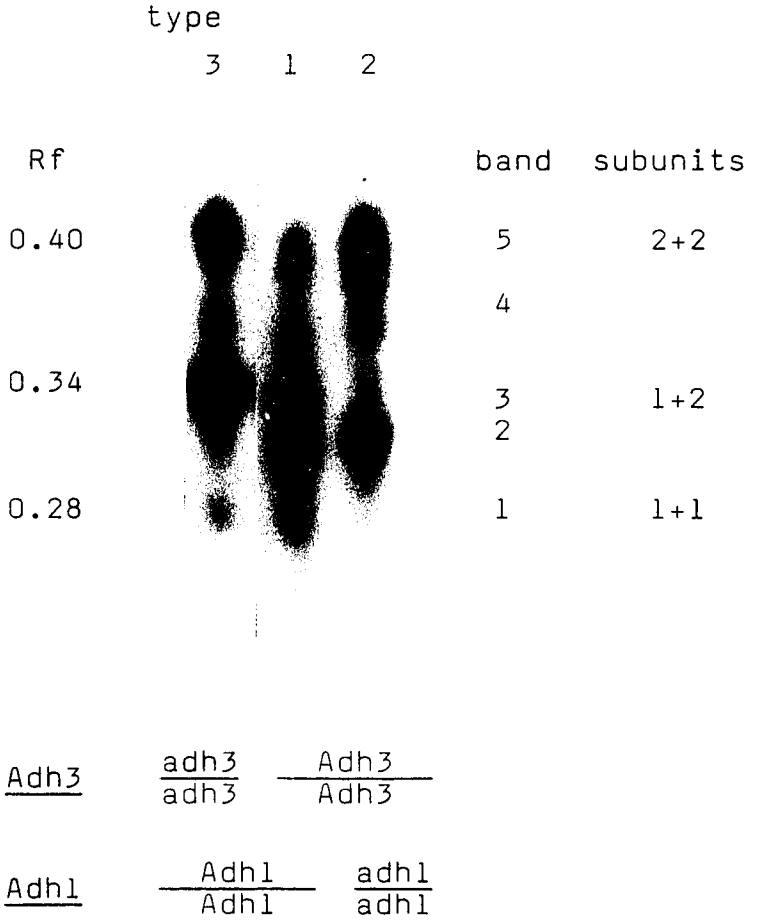


Figure 3. Zymograms of ADH in soybeans. The numbers under the subunits heading refer to the loci from which monomers originated.

Table 2. The F_2 segregation data at the ADH loci in soybeans

Cross	F_2 segregation	
Adh1/Adh1 X adh1/adh1	Adh1/-	adh1/adh1
Amsoy X Jefferson (type 1 X type 2)	308	113
x^2 (3 : 1)	0.76 (0.3 < P < 0.5)	
adh1/adh1 X Adh1/Adh1		
Jefferson X Amsoy (type 2 X type 1)	221	63
x^2 (3 : 1)	1.20 (0.2 < P < 0.3)	
Adh3/Adh3 X adh3/adh3	Adh3/-	adh3/adh3
PI407.181 X PI407.252 (type 1 X type 3)	173	57
x^2 (3 : 1)	0.005 (0.90 < P < 0.95)	

model of the null and functional alleles segregating at one locus. An electrophoretic five-band pattern was also observed in soybeans by Beremand (1975). Based on dissociation-reassociation and immunoelectrophoresis studies, Beremand proposed a similar hypothesis to ours for the homo-heterodimer relationship for bands 1, 3 and 5 of soybean ADH zymogram.

The ADH zymogram band 2 was hypothesized to be controlled by a third locus, designated as Adh3, having a recessive null (adh3) and a dominant functional (Adh3) alleles. The observed 3 : 1 segregation in F_2 generation of a cross between type 1 and type 3 of G. soja plants (Table 2) fitted a model of the null and functional alleles segregating at a single locus.

β-Amylase (AM, EC 3.2.1.2)

One locus (Am3) with four alleles, Am3-f, Am3-s, Am3-sw and Am3-n1 was reported in the genetic control of β-amylase allozyme variants in G. max (Kiang, 1981). Only the first two alleles, Am3-f and Am3-s were observed in present G. soja study.

Acid phosphatase (AP, EC 3.1.3.2)

Acid phosphatases catalyze the hydrolysis of monoesters of phosphoric acid, and are also called phosphomonoesterases (Fairley and Kilgour, 1966). Na-alpha-naphthyl acid

phosphate is an unnatural substrate which was used to visualize AP allozymes in this study.

Three electrophoretic zymograms differing in the mobility of an anodal band were observed in soybeans (Gorman and Kiang, 1977; Kiang et al., 1981). These variants were suggested by Gorman and Kiang (1977) to be inherited as three codominant alleles (Ap-a, Ap-b and Ap-c) at a single nuclear locus. The hypothesis was confirmed by Hildebrand et al. (1980) using a disc gel electrophoretic procedure. The same zymograms were observed in the present study, and were shown as zymogram type 1 (Ap-a), type 2 (Ap-b) and type 3 (Ap-c) in Figure 4. In addition, a new variant type 4, Rf 0.57) was observed in G. soja (Figure 4).

Reciprocal crosses between type 3 and type 4 plants of G. soja showed a 1 : 2 : 1 F_2 segregation ratio (Table 3). Heterozygotes showed both parental bands (Rf 0.53 of type 3; Rf 0.57 of type 4) with no intermediate band (H type in Figure 4). Only F_2 seed with a two band zymogram segregated in the F_3 with a ratio of 1 : 2 : 1 (Table 3). Thus, the results indicate that type 4 is the consequence of a fourth codominant allele at the Ap locus. This new allele is designated as Ap-d.

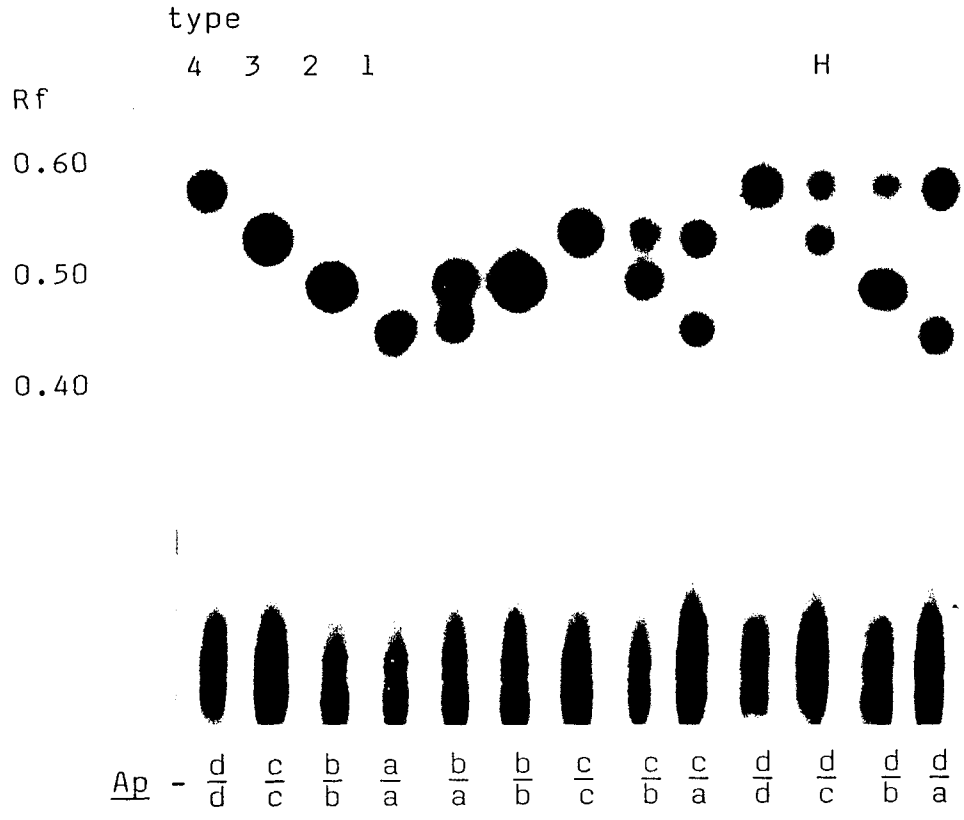


Figure 4. AP zymograms observed in soybean dry seeds.

Table 3. The F_2 and F_3 segregation data at the Ap locus in Glycine soja.

Cross	F_2 segregation			F_3 seeds from F_2 heterozygous plants		
	Ap-c/c	Ap-c/d	Ap-d/d	Ap-c/c	Ap-c/d	Ap-d/d
Ap-c/c X Ap-d/d						
PI407.192XPI407.223	89	194	97	30	51	23
(type 3 X type 4)						
x^2 (1 : 2 : 1)	0.50 (0.3 < P < 0.5)			0.99 (0.3 < P < 0.5)		
Ap-d/d X Ap-c/c						
PI407.223XPI407.192	133	289	133	40	86	46
(type 4 X type 3)						
x^2 (1 : 2 : 1)	0.96 (0.3 < P < 0.5)			0.65 (0.3 < P < 0.5)		

Diaphorase (DIA, EC 1.6.2.2)

Diaphorases are a class of enzymes capable of the oxidation of NADH with certain artificial dyes, such as 2,6-dichlorophenol indolphenol used in this DIA study. The function of DIA in plants is unknown.

DIA zymograms in cultivated and wild soybeans were found to have as many as twelve anodal bands in dry seed cotyledon, and were the products of five to seven DIA loci (Gorman et al., 1983). The first five observed bands are mitochondrion associated (Gorman, 1983). The monomers of two loci interact to form intra- and interlocus tetramers of band 1 to 5 as reported by Gorman et al. (1983). Among them, Dia1 was found to have two alleles (Dia1 and dia1) while the other locus was monomorphic. The same observation was made in G. soja in the present study.

The Dia2 was the locus designated to control the mobility of the seventh and eighth DIA bands. Two different mobility variants were reported to be controlled by two codominant alleles, Dia2-a and Dia2-b (Gorman et al., 1983). A third homozygous zymogram was observed in some seeds of K31 (PI407.252 and PI407.253), an accession of G. soja from South Korea. These seeds showed no DIA bands 7 and 8 (type 4, Figure 5). We hypothesized that a third allele, dia2, is recessive to Dia2-a and Dia2-b, and a homozygote for this null allele produces no bands 7 and 8. Crosses between Dia2-b and dia2 plants showed a 3 : 1 segregation ratio in

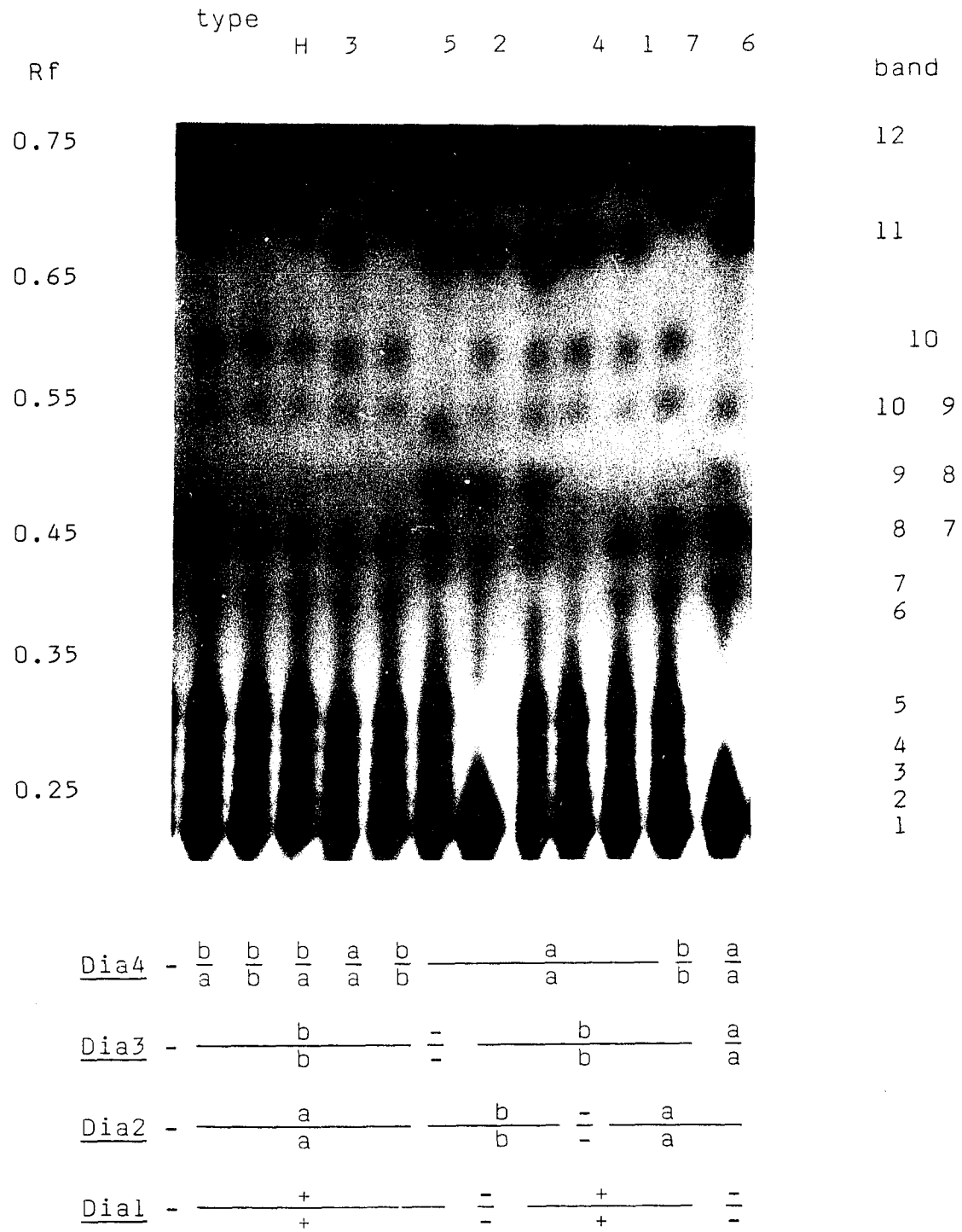


Figure 5. DIA zymograms of soybeans. '+' is the dominant allele, and '-' is the recessive allele.

F_2 seeds. There was no difference in reciprocal crosses (Table 4). Based on the F_2 plant progeny test, a segregation of 1 : 2 : 1 was observed for the F_2 seeds, while a 3 : 1 ratio was obtained in F_3 seeds from segregating F_2 plants (Table 4). All these results indicate that dia2 is recessive to Dia2-b and Dia2-a since Dia2-b and Dia2-a are codominant.

Gorman et al. (1983) reported that a single locus (Dia3) with variant dominant and recessive alleles (Dia3 and dia3) was responsible for the presence or absence of DIA band 10 (type 5, Figure 5). The DIA zymogram of G. soja K52 (PI407.233-407.235) was found to have slower mobility of bands 9 and 10 (type 6, Figure 5). The mobility of its band 9 is equal to the band 8 of type 1, and its band 10 equal to the band 9 of type 1. Thus, Dia3-a was tentatively designated for the allele controlling the slow movement of band 9 and 10 of K52 DIA zymogram and the Dia3-b as the former Dia3 (Gorman et al., 1983) allele in locus Dia3. More genetic data are needed to confirm this hypothesis.

All the G. soja seeds examined in the present study for DIA zymograms were found having bands 11 and 12 except K113 (PI487.431), an accession from southern Japan. In K113 only one band (type 7, Figure 5) was observed with the mobility between bands 11 and 12 of type 1 (Figure 5). Crosses between K113 (type 7) and PI486.220 (type 1) showed all three parental bands in F_1 (type H, Figure 5). In the F_2, three phenotypic classes were observed, the two parental

Table 4. The F_2 and F_3 segregation data at the Dia2 locus in soybeans.

Cross	F_2 seeds		F_2 genotype (by progeny test)			F_3 seeds from F_2 heterozygous plants	
	Dia2-b / -	dia2 / dia2	Dia2-b / Dia2-b	Dia2-b / dia2	dia2 / dia2	Dia2-b / -	dia2 / dia2
Dia2-b / Dia2-b X dia2 / dia2							
AV68* X PI407.252	123	45	46	77	45	218	98
(type 1 X type 4)							
x^2	0.28 (0.5<P<0.7)		1.17 (0.2<P<0.3)			0.76 (0.3<P<o.5)	
dia2 / dia2 X Dia2-b / Dia2-b							
PI407.252 X AV68	109	33					
(type 4 X type 1)							
x^2	0.24 (0.5<P<0.7)						

* AV68 is Glycine max obtained from AVRDC, Taiwan.

and the three-banded F_1. The number of progeny in each class yielded the 1 : 2 : 1 ratio expected for segregation of two codominant alleles at a single locus. There was no difference between reciprocal crosses. Only F_2 seeds with the three-banded pattern segregated in the F_3 generation with a ratio of 1 : 2 : 1 (Table 5). The results indicate that a nuclear locus with two codominant alleles was responsible for the DIA zymogram variant of bands 11 and 12 in G. soja. The locus was designated as Dia4 and the alleles, Dia4-a producing two bands with Rfs' 0.68 and 0.75, respectively, while Dia4-b producing one band with Rf 0.72.

These observations raise an interesting question. In type 1, a single locus (Dia4) produces two DIA bands, but type 7 only one band (Figure 5). The question is "why does type 7 produce only one band?" There are at least two possible explanations. (1) One of the (RNA) messages producing the two bands does not yield a functional enzyme or (2) Mutation occurred to one of the two messages so that the products of the two messages have the same migration rate. There may be some other explanations.

Table 5. The F_2 and F_3 segregation data at the Dia4 locus in soybeans.

Cross	F_2 segregation			F_3 seeds from F_2 heterozygous plants		
	Dia4-a/a	a/b	b/b	Dia4-a/a	a/b	b/b
Dia4-a/a X Dia4-b/b						
PI486.220 X PI487.431 (type 1 X type 7)	43	94	40	13	19	7
x^2 (1 : 2 : 1)	0.78 (0.3 < P < 0.5)			1.87 (0.1 < P < 0.2)		
Dia4-b/b X Dia4-a/a						
PI487.431 X PI486.220 (type 7 X type 1)	100	197	89	19	58	27
x^2 (1 : 2 : 1)	0.79 (0.3 < P < 0.5)			2.61 (0.1 < P < 0.2)		

Endopeptidases (ENP)

Pepsin, trypsin and chymotrypsin are termed endopeptidases in that they can catalyze the hydrolysis of peptide bonds in interior positions of a peptide chain (Fairley and Kilgour, 1966). The substrate used to visualize ENP isozymes in this study was alpha-benzoyl-DL-arginine-*β*-naphthylamide. Trypsin is most active toward peptide bonds involving carboxyl groups of this basic amino acids (Fairley and Kilgour, 1966).

Three homozygous zymogram types were observed in the G. soja studied (Figure 6). Types 1, 2 and 3 each had one anodal ENP band, but differed in their mobility. It was hypothesized that the difference among zymogram types 1, 2 and 3 was controlled by a single locus with three codominant alleles. The hybrid between plants of different types displayed both parental bands without intermediate band. The locus was tentatively designated as Enp locus with three alleles as Enp-a (Rf 0.33), Enp-b (Rf 0.36) and Enp-c (Rf 0.39). Further genetic data are needed to support the hypothesis.

Glutamate oxaloacetic transaminase (GOT, EC 2.6.1.1)

The GOT banding pattern of soybeans can be separated into two zones. The first banding zone contains the first three slow moving bands. No mobility difference was found with respect to these three bands in present G. soja

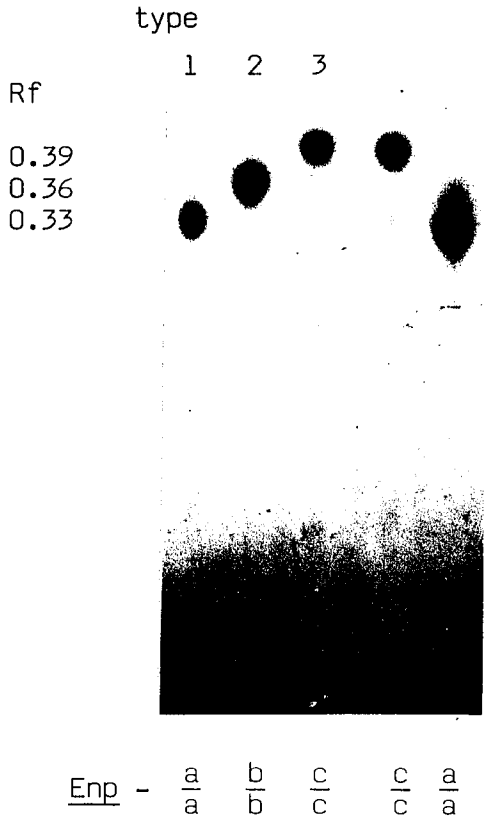


Figure 6. ENP zymograms of Glycine soja.

studied. We hypothesize that two interacting loci are responsible in producing these three bands.

Three mobility variants were observed in the fourth band of soybean GOT zymograms. One locus with three alleles is hypothesized to be responsible for the variation. Tentatively, the locus is designated as Got, and three alleles as Got-a producing band Rf 0.51, Got-b band Rf 0.53, and Got-c Rf 0.56.

NADP-active isocitrate dehydrogenase (IDH, EC 1.1.1.42)

NADP-active isocitric dehydrogenase isozymes catalyze the oxidation of isocitrate to α-ketoglutarate. Two electrophoretic band areas were observed in the cultivated and wild soybeans, and were reported as the products of four IDH loci in soybeans (Kiang and Gorman, 1985). Idh1 and Idh2 code for cytosol associated isozymes which are the top band area of IDH, with each having two codominant alleles (Idh1-a, Rf 0.49; Idh1-b, Rf 0.56; Idh2-a, Rf 0.56; Idh2-b, Rf 0.63). These two loci interact to form intra- and interlocus heterodimers. Since the mobility of bands produced by Idh1-b and Idh2-a is the same, the genotype homozygous for Idh1-b, Idh2-a displays one band pattern (type 5, Figure 7A).

A new variant (type 4 and 6 of Figure 7A) was observed in four plants of Shizukuishi population (S), Kitamkami river, Japan. With three-banded pattern as normal type 1, the top two bands of this new variant have slower mobilities

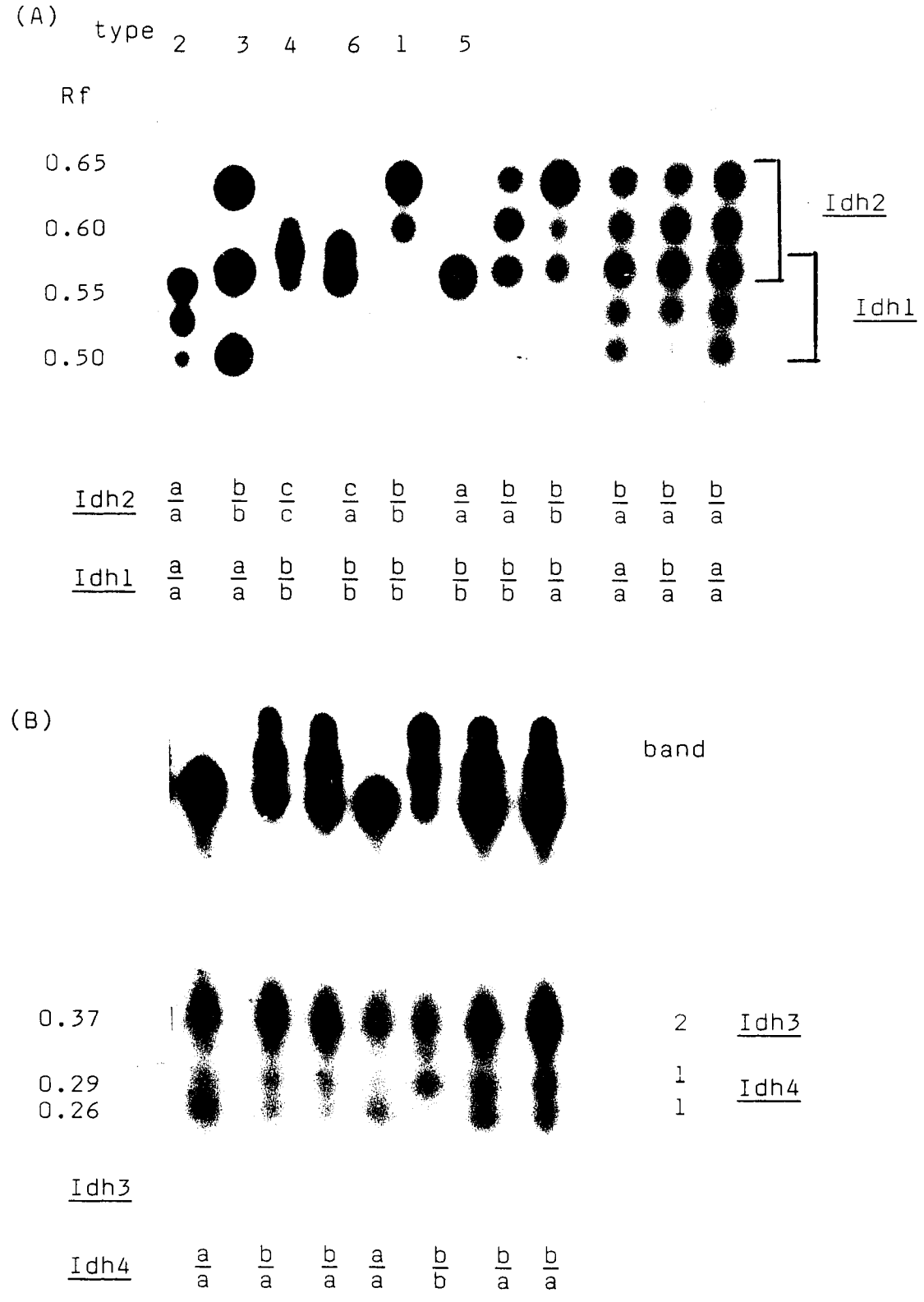


Figure 7. IDH zymograms of soybeans.

(Rf 0.58 and Rf 0.60). Three of these four plants bred true, and displayed type 4 zymogram (Figure 7A). Seeds from the fourth plant showed type 4 or 5 or 6 zymogram (Figure 7A). We hypothesize that the difference between type 1 and type 4 is controlled by two alleles at the Idh2 locus. One allele is Idh2-b, producing the farthest anodal band (Rf 0.63) of type 1. The other newly observed allele is tentatively designated as Idh2-c which produces the fastest migrating band (Rf 0.60) of type 4. The mid band of type 4 is the interlocus heterodimer produced by Idh1-b and Idh2-c. The genotype of type 6 (Figure 7A) at the Idh2 locus is heterozygous (Idh2-a/Idh2-c), so the intensity of its top band is much weaker when compared with the top band of type 4 which is homozygous for the Idh2-c allele.

At the bottom region of soybean IDH zymogram, three variants differing in the mobility of band 2 were reported as the products of the Idh3 locus with three alleles: Idh3-a (Rf 0.31), Idh3-b (Rf 0.37) and Idh3-c (Rf 0.41) (Kiang and Gorman, 1985). The same zymogram variants were also observed in the present study.

In this study two new variants differing in the band mobility (Rf's 0.29 and 0.26) (Figure 7B) were observed in the first band of soybean IDH zymograms. In the F_2, three phenotypic classes were observed, the two parental and the double-banded F_1. The number of progeny within each class yielded the 1 : 2 : 1 ratio expected for segregation of two codominant alleles at a single locus, designated as Idh4.

Only the F_2 seeds with two-banded pattern segregated in the F_3 with a ratio of 1 : 2 : 1 (Table 6). The two alleles were designated as Idh4-a (Rf 0.26) and Idh4-b (Rf 0.29).

The genotypes of PI487.429 and PI407.233 are Idh2-b/Idh2-b Idh4-b/Idh4-b and Idh2-a/Idh2-a Idh4-a/Idh4-a, respectively. The F_2 from dihybrid cross (PI487.429 X PI407.233) segregating at the Idh2 and Idh4 loci showed no recombinant types. Thus, Idh2 and Idh4 loci may be very tightly linked so that no recombinant between the two loci was observed in the F_2 generation. Crossing over may not occur under natural conditions since Idh2-b and Idh4-b always appeared together in all the seed examined, so did Idh2-a and Idh4-a. These observations cannot rule out the possibility that the bands are produced by a single locus.

Leucine aminopeptidase (LAP, EC 3.4.11.1)

Leucine aminopeptidases are nonspecific exopeptidases which hydrolyze the peptide bond adjacent to a free -amino group of peptides. The enzyme is called Leucine amino-peptidase because it reacts rapidly with leucine compounds, such as L-leucine- *6* -naphthyl-amide HCl, the substrate used to visualize LAP isozymes in soybeans.

Two electrophoretic variants of LAP have been observed in dry G. soja seeds. The two variants are inherited as codominant alleles at one locus, Lap1. The genetic symbol Lap1-a and Lap1-b are assigned to slow (Rf 0.49) and fast (Rf 0.53) migrating zymograms, respectively (Kiang et al.,

Table 6. The F_2 and F_3 segregation data at the Idh4 locus in Glycine soja.

Cross	F_2 segregation			F_3 seeds from F_2 heterozygous plants		
	Idh4-b/b	b/a	a/a	Idh4-b/b	b/a	a/a
Idh4-b/b X Idh4-a/a						
PI487.429 X PI407.233	146	282	119	23	49	32
x^2 (1 : 2 : 1)	3.19 (0.05 < P < 0.1)			1.90 (0.1 < P < 0.2)		

1985). The same zymograms were also observed in the seeds examined in this study.

Malate dehydrogenase (MDH, EC 1.1.1.37)

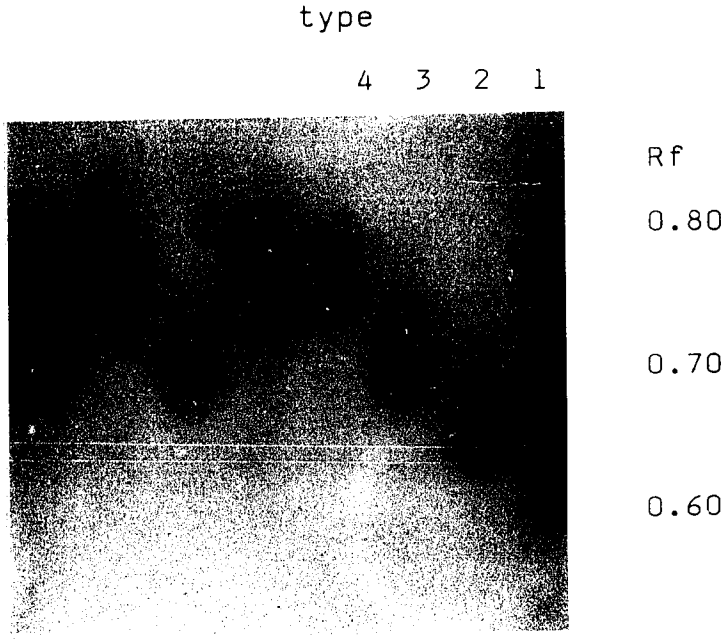
Malate dehydrogenase catalyzes the reaction of

$$\text{L-malate} + \text{NAD} \longrightarrow \text{oxaloacetate} + \text{NADH}_2$$

Only one MDH zymogram type with as many as six clear bands was observed in the G. soja seeds studied. Gorman (1983) observed a 10-banded zymogram in 180 G. max cultivars and 101 G. soja accessions. Based on band's mobility, subcellular distribution and coenzyme specificity, the six clear bands observed in the present study were hypothesized to be the products of at least four loci (Gorman, 1983). These four hypothesized MDH loci were used in the subsequent genetic structure study of G. soja.

Mannose-6-phosphate isomerase (MPI, EC 5.3.1.8)

Mannose-6-phosphate isomerase acts mainly to convert mannose-6-phosphate into fructose-6-phosphate (Lehninger, 1982). Four homozygous MPI zymograms (types 1, 2, 3 and 4, Figure 8) differing in the mobility of the two MPI anodal bands were observed in the present G. soja study. The difference of zymogram types 1, 2 and 3 was reported as a result of a single nuclear locus (Mpi) with three codominant alleles Mpi-a, Mpi-b and Mpi-c (Gorman et al., 1983; Kiang



Mpi - $\frac{d}{c}$ $\frac{d}{d}$ $\frac{c}{c}$ $\frac{d}{d}$ $\frac{d}{d}$ $\frac{c}{c}$ $\frac{b}{b}$ $\frac{a}{a}$

Figure 8. MPI zymograms of soybeans.

and Gorman, 1983). In the present study, we report that the fourth migrating variant (type 4, Figure 8) is the product of a fourth codominant allele at the Mpi locus, designated Mpi-d.

Data in Table 7 show that crosses between type 4 and type 3 plants as well as type 4 and type 2 plants resulted in a 1 : 2 : 1 segregation ratio in F_2 seeds. The heterozygous zymograms displayed both parental bands, and a 1 : 2 : 1 segregation ratio was found in the F_3 seeds harvested from heterozygous F_2 plants.

In summary, the four codominant alleles of Mpi locus and the Rf values of MPI zymogram they produce are as following: (1) Mpi-a (Rf 0.61 and 0.56 two bands of type 1 zymogram), (2) Mpi-b (Rf 0.56 and 0.70, type 2), (3) Mpi-c (Rf 0.71 and 0.75, type 3), and (4) Mpi-d (Rf 0.76 and 0.80, type 4) (Figure 8).

TAble 7. The F_2 and F_3 segregation data at the Mpi locus in soybeans.

Cross	F_2 segregation			F_3 seeds from F_2 heterozygous plants		
	Mpi-d/d	d/c	c/c	Mpi-d/d	d/c	c/c
Mpi-c/c X Mpi-d/d						
PI407.233 X PI407.192 (type 3 X type 4)	94	221	107	9	12	9
x^2 (1 : 2 : 1)	1.75 (0.1<P<0.2)			2.1 (0.1<P<0.2)		
	Mpi-d/d	d/b	b/b	Mpi-d/d	d/b	b/b
Mpi-b/b X Mpi-d/d						
PI407.223 X PI407.192 (type 2 X type 4)	127	277	151	46	79	43
x^2 (1 : 2 : 1)	2.07 (0.1<P<0.2)			0.70 (0.3<P<0.5)		
PI487.428 X PI407.192 (type 2 X type 4)	59	128	68			
x^2 (1 : 2 : 1)	0.64 (0.3<P<0.5)					
	Mpi-b/b	a/b	a/a	Mpi-b/b	a/b	a/a
Mpi-b/b X Mpi-a/a						
Amsoy X Wilson (type 2 X type 1)	53	102	52	27	54	30
x^2 (1 : 2 : 1)	0.05 (0.8<P<0.9)			0.24 (0.5<P<0.7)		

6-Phosphogluconate dehydrogenase (PGD, EC 1.1.1.44)

6-Phosphogluconate dehydrogenase functions specifically in the oxidative, phosphogluconate pentose phosphate pathway to oxidize and decarboxylate 6-phosphogluconate (Lehninger, 1982).

Four-banded PGD zymograms were observed in the present G. soja study (Figure 9). All of the observed zymograms resulting from homozygotes of different alleles are included. A model of two interacting nuclear loci (Pgd1 and Pgd2) producing bands 1, 2 and 3 with band 2 the interlocus heterodimer of band 1 and 3 was reported by Gorman et al. (1983) and Kiang and Gorman (1983).

Three migrants of different mobility (Rfs' 0.36, 0.42 and 0.50) were found in band 1 of soybean PGD zymograms. Two of these three variants were reported to be controlled by a nuclear locus Pgd1 with two codominant alleles, Pgd1-a (producing a band at Rf 0.36) and Pgd1-b (producing a band at Rf 0.42) (Gorman et al., 1983; Kiang and Gorman, 1983). The third variant is observed only in G. soja. A 1 : 2 : 1 segregation ratio was observed in the F_2 and the progeny of the F_2 heterozygote in F_3 generations of crosses involving these three band 1 variants (Table 8). The results indicate that a third allele, designated Pgd1-c, in locus Pgd1 is responsible for the third variant of band 1 (Rf 0.50).

There were also three different migrating rates of band 3 observed in the present G. soja study. The segregation

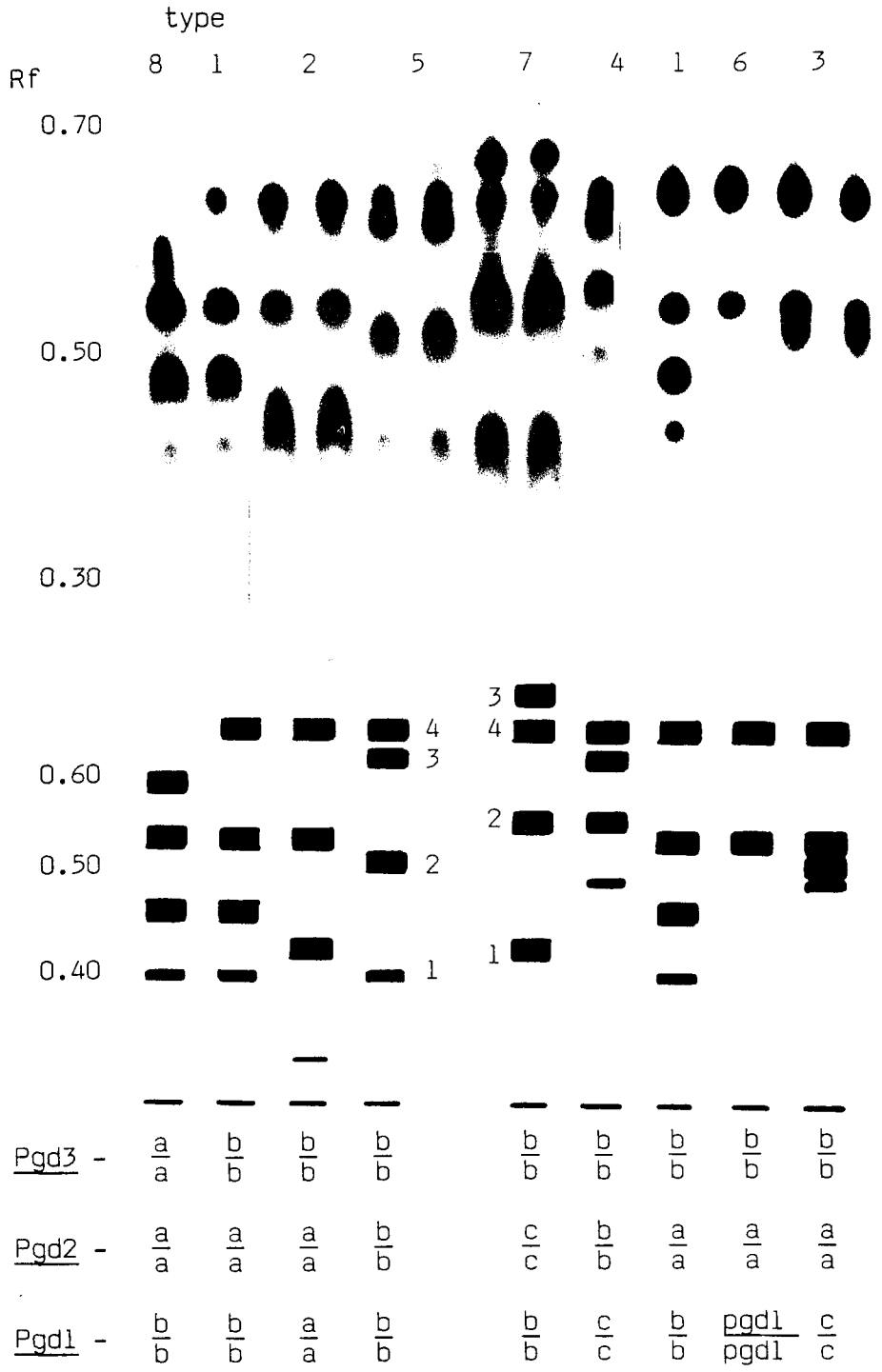


Figure 9. PGD zymograms of soybeans.

Table 8. The F_2 and F_3 segregation data of Pgd1 locus in soybeans

cross	F_2 generation			F_3 seeds from F_2 heterozygous plants		
	Pgd1-a/a	a/b	b/b	Pgd1-a/a	a/b	b/b
Pgd1-a/a x Pgd1-b/b						
AV 68 x PI407.252	36	75	45	29	48	25
x^2 (1 : 2 : 1)	1.27 (0.2 < P < 0.3)			0.67 (0.3 < P < 0.5)		
AV 68 x PI407.260	89	183	90	23	50	31
x^2 (1 : 2 : 1)	0.05 (0.8 < P < 0.9)			1.39 (0.2 < P < 0.3)		
Pgd1-b/b x Pgd1-a/a						
PI407.252 x AV 68	36	73	33	12	29	11
x^2 (1 : 2 : 1)	0.30 (0.5 < P < 0.7)			0.74 (0.3 < P < 0.5)		
AV 62 x PI424.032	210	408	230	31	43	30
x^2 (1 : 2 : 1)	2.15 (0.1 < P < 0.2)			3.14 (0.05< P < 0.1)		
	Pgd1-b/b	b/c	c/c	Pgd1-b/b	b/c	c/c
Pgd1-b/b x Pgd1-c/c						
AV 62 x PI407.160	84	165	72	48	113	47
x^2 (1 : 2 : 1)	1.13 (0.2 < P < 0.3)			1.57 (0.2 < P < 0.3)		
Pgd1-c/c x Pgd1-b/b						
PI407.160 x AV 62	177	365	186	36	94	39
x^2 (1 : 2 : 1)	0.23 (0.5 < P < 0.7)			2.24 (0.1 < P < 0.2)		
PI407.262 x PI407.233	29	69	38			
x^2 (1 : 2 : 1)	1.22 (0.2 < P < 0.3)					
	Pgd1-a/a	a/c	c/c	Pgd1-a/a	a/c	c/c
Pgd1-a/a x Pgd1-c/c						
PI424.032 x PI407.160	48	105	55	33	63	45
x^2 (1: 2 : 1)	0.49 (0.3 < P < 0.5)			3.64 (0.05< P < 0.1)		
Pgd1-c/c x Pgd1-a/a						
PI407.160 x PI424.032	126	219	122	38	73	31
x^2 (1 : 2 : 1)	1.87 (0.1 < P < 0.2)			0.81 (0.3 < P < 0.5)		
PI407.262 x PI424.045	15	40	15			
x^2 (1 : 2 : 1)	1.43 (0.2 < P < 0.3)					

data (Table 9) support our hypothesis that three codominant alleles are responsible for the observed zymogram difference. They were designated as Pgd2-a (Rf 0.54), Pgd2-b (Rf 0.62) and Pgd2-c (Rf 0.68), respectively.

A single-banded zymogram with respect to bands 1, 2 and 3 was observed in several G. soja seed accessions by Gorman (type 3 in Fig. 13 of Gorman, 1983). He proposed that the Pgd2-a/a with a recessive null allele at the Pgd1 locus was responsible for the observed zymogram. However, when these seed were examined with 6% acrylamide+4% starch gel, a three-distinct-band zymogram was observed (type 3, Figure 9). The 6% acrylamide+4% starch gel yields better resolution for the PGD zymogram than the 7% acrylamide+ 2% starch gel which was used by Gorman. The genotype of the seed (type 3, Figure 9) with respect to the Pgd1 and Pgd2 loci is Pgd1-c/c, Pgd2-a/a. The mobilities of bands produced by these two alleles (Rf 0.50, 0.54) and their interlocus heterodimer band (Rf 0.52) are not much different. In the present study, a one-band pattern with respect to band 1 to 3 (type 6, Figure 9) was found only in populations of Gandai (G) and Shizukuishi (S) of Japan. Seeds from one plant in the Shizukuishi population segregated for the three and one band phenotypes among eight seeds examined (6 three-band ; 2 one-band phenotypes). Thus, a recessive null allele is hypothesized to be responsible for the lack of PGD activity in bands 1 and 2 of type 6 zymogram (Figure 9). More inheritance data are

Table 9. The F_2 and F_3 segregation data of locus Pgd2 in soybeans

cross	F_2 generation			F_3 seeds from F_2 heterozygous plants		
	Pgd2-a/a	a/b	b/b	Pgd2-a/a	a/b	b/b
Pgd2-a/a x Pgd2-b/b						
PI487.428 x PI407.192	62	140	53			
x^2 (1 : 2 : 1)	3.09 (0.05< P <0.1)					
AV 62 x PI407.160	59	92	56	53	115	50
x^2 (1 : 2 : 1)	2.65 (0.4 < P < 0.5)			0.74 (0.3 < P < 0.5)		
PI424.032 x PI407.160	56	100	52	12	31	9
x^2 (1 : 2 : 1)	0.46 (0.3 < P < 0.5)			2.27 (0.4 < P < 0.5)		
PI407.233 x PI407.192	126	268	125			
x^2 (1 : 2 : 1)	0.56 (0.3 < P < 0.5)					
Pgd2-b/b x Pgd2-a/a						
PI407.160 x AV 62	170	359	199	23	63	31
x^2 (1 : 2 : 1)	2.45 (0.1 < P < 0.2)			1.79 (0.1 < P < 0.2)		
PI407.160 x PI424.032	120	242	105	10	30	12
x^2 (1 : 2 : 1)	1.58 (0.2 < P < 0.3)			1.39 (0.2 < P < 0.3)		
	Pgd2-a/a	a/c	c/c	Pgd2-a/a	a/c	c/c
Pgd2-a/a x Pgd2-c/c						
AV 68 x PI407.252	34	77	45	22	51	29
x^2 (1 : 2 : 1)	1.57 (0.2 < P < 0.3)			0.96 (0.3 < P < 0.5)		
Pgd2-c/c x Pgd2-a/a						
PI407.252 x AV 68	39	71	32	15	27	10
x^2 (1 : 2 : 1)	0.70 (0.3 < P < 0.5)			1.04 (0.3 < P < 0.5)		

needed to confirm this hypothesis.

Figure 10A shows a clear intralocus band formed in heterozygotes between Pgd2-a and Pgd2-b. When a cross was made between plants with different alleles in both Pgd1 and Pgd2 loci, nine zymogram phenotypes were obtained as demonstrated in Figure 11. The dihybrid segregants consist of a double heterozygous type, four single heterozygous types and four homozygous type. The dihybrid segregation data for Pgd1 and Pgd2 loci (Table 10) indicate that these two loci are not linked.

All of the G. soja seeds examined showed a PGD fourth band at Rf 0.64 except seeds of PI486.220 from Mishima, Japan. The mobility of the PGD fourth band in PI488.220 is at Rf 0.60 (type 8, Figure 9). Reciprocal crosses between PI486.220 and PI487.431 (type 1, Figure 9) displayed a three-banded pattern (two parental and one intermediate bands) in the F_1 seed (Figure 10B), and a 1 : 2 : 1 segregation ratio in the F_2 generation. Only progeny from three-banded seeds segregated in the F_3, in a 1 : 2 : 1 ratio (Table 11). All these results indicate a nuclear single locus, designated Pgd3, with two codominant alleles, Pgd3-a and Pgd3-b is responsible for the variant zymogram of PGD fourth band. The homodimer of Pgd3-a produces a band at Rf 0.60, and the Pgd3-b produces a band at Rf 0.64, and their heterodimer forms the intermediate band at Rf 0.62 (Figure 10B).

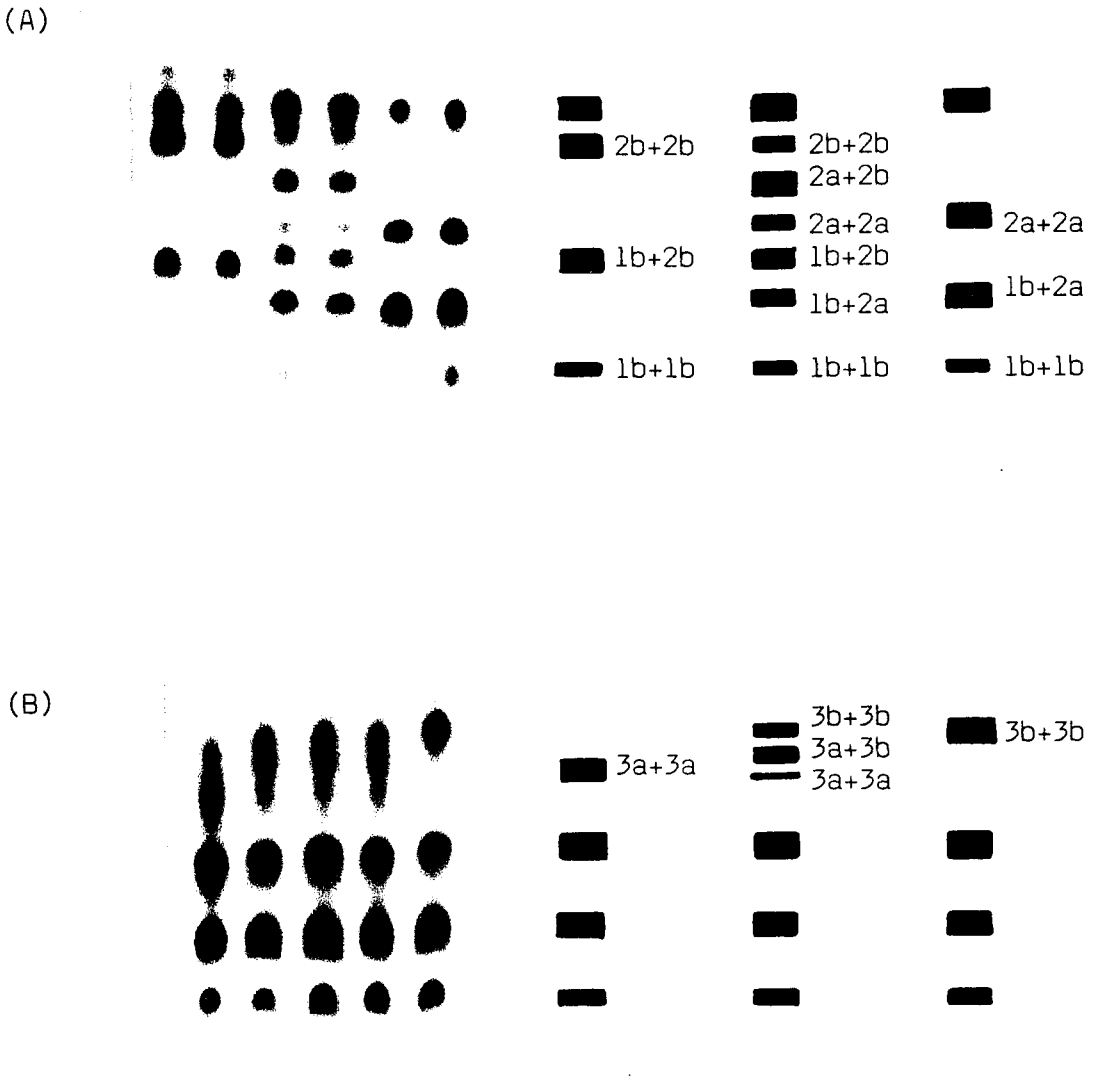


Figure 10. Three phenotypes of segregants from a cross between
(A) Pgd2-a/a X Pgd2-b/b
(B) Pgd3-a/a X Pgd3-b/b

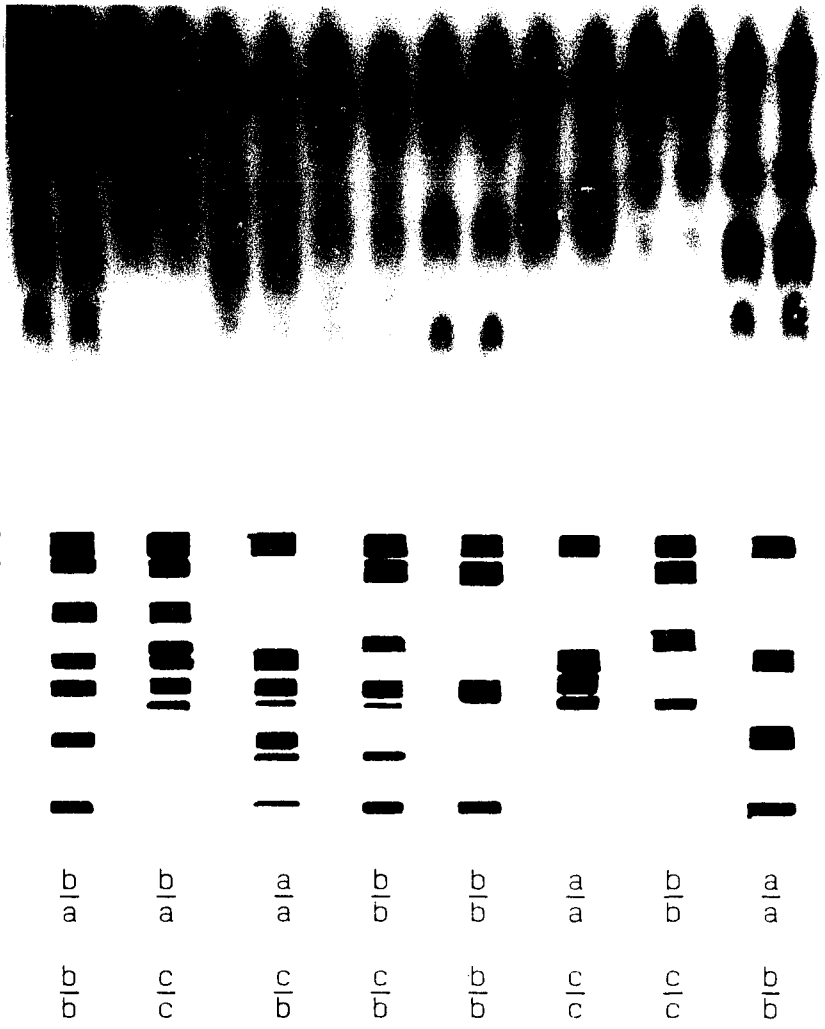


Figure 11. Nine phenotypes of dihybrid segregants from a cross between Pgd1-b/b, Pgd2-a/a x Pgd1-c/c, Pgd2-b/b.

Table 10. F_2 dihybrids segregation data for the gene pair of Pgd1 and Pgd2 in soybeans.

cross	Pgd1 Pgd2 / Pgd1 Pgd2	AB/AB	Ab/Ab	aB/aB	ab/ab	AB/Ab	aB/ab	AB/aB	Ab/ab	AB/ab + Ab/aB*	N	x^2	P
PI407.160 X PI424.032	c b / a a	33	28	33	32	61	61	45	54	120	467	5.03	0.7
PI424.032 X PI407.160	a a / c b	10	18	15	13	27	20	21	31	53	208	6.28	0.5
PI407.160 X AV62	c b / b a	55	47	57	45	84	74	87	78	201	728	12.7	0.1
AV62 X PI407.160	b a / c b	14	13	15	15	24	14	27	31	54	207	7.53	0.4

* A and B represent Pgd1 and Pgd2 loci, respectively.

Table 11. The F_2 and F_3 segregation data at the Pgd3 locus in soybeans.

Cross	F_2 generation			F_3 seeds from F_2 heterozygous plants		
	Pgd3-a/a	a/b	b/b	Pgd3-a/a	a/b	b/b
Pgd3-a/a X Pgd3-b/b						
PI486.220 X PI487.431	114	205	104	26	48	26
x^2 (1 : 2 : 1)	0.87 (0.3 < P < 0.5)			0.16 (0.5 < P < 0.7)		
Pgd3-b/b X Pgd3-a/a						
PI487.431 X PI486.220	77	155	78	28	68	36
x^2 (1 : 2 : 1)	0.01 (0.9 < P <0.95)			1.09 (0.2 < P < 0.3)		

Phosphoglucose isomerase (PGI, EC 5.3.1.9)

Phosphoglucose isomerase catalyzes the specific reversible reaction:

D-glucose-6-phosphate <———> D-fructose-6-phosphate (Lehninger, 1982).

The soybean PGI zymogram displays six anodal bands. Six homozygous zymogram types were observed in the present study (Figure 12). A cluster of three-band pattern (Rfs' 0.43, 0.44 and 0.45) at the lower band region was observed in types 3, 4 and 5, while only band 3 (Rf 0.45) appeared in type 1 zymogram. We hypothesize that two interacting loci produce this 3-band cluster, with band 1 and band 3 the homodimers produced by each of the two loci, and band 2 the interlocus heterodimer. A recessive null allele is further hypothesized to be responsible for the lack of bands 1 and 2 in zymogram type 1. A cross involving type 1 (AV68) and type 3 (PI424.032) resulted in a 3 : 1 [80 (3-band type 3) : 24 (1-band type 1)] segregation ratio in the F_2 seeds as expected in a dominant-recessive model. The locus with these two dominant and recessive alleles was designated as Pgi2, with Pgi2 for the dominant and pgi2 for the recessive allele. The locus which produces band 3 was designated as Pgi3. The activity of the Pgi2 locus is much weaker than the locus Pgi3.

A third PGI locus, Pgi1, was reported to produce monomers, and each two of the monomers combined to form

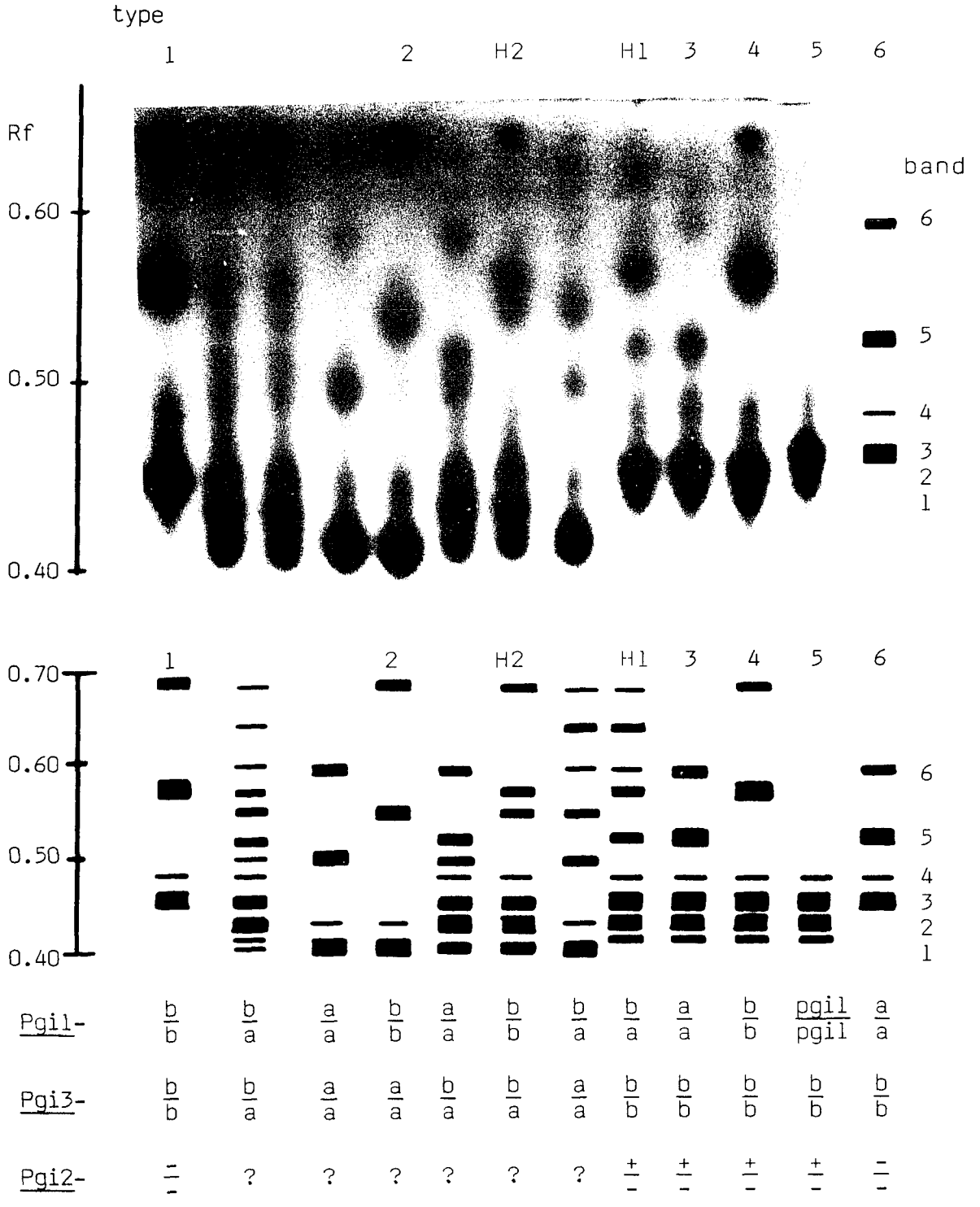


Figure 12. PGI zymograms of soybeans. $\frac{=}{=}$ is the homozygote of pgi2/pgi2, $\frac{\pm}{-}$ is the homozygote of Pgi2/Pgi2 or heterozygote of Pgi2/pgi2.

homodimer band 6. These monomer units also interact with the product of Pgi3 to form interlocus heterodimer band 5 (Gorman et al., 1983; Kiang and Gorman, 1983). Among the observed three variants with respect to bands 5 and 6, two differed in mobility while one lacked activity (Gorman et al., 1983; Kiang and Gorman, 1983). Similar zymogram variants are observed in the present G. soja study. The two mobility variants (type 1, Rf's of bands 5 and 6 are 0.57 and 0.67; type 3, Rf's of bands 5 and 6 are 0.52 and 0.58) were reported as a consequence of two codominant alleles Pgi1-b and Pgi1-a at the locus Pgi1 (Gorman et al., 1983; Kiang and Gorman, 1983). A cross between plants of type 1 and 3 (AV68 X PI424.032) was made in the present study. The heterozygote showed both parental bands and a new intralocus hybrid band at Rf 0.64 (type H1, Figure 12). A 1 : 2 : 1 (31 type 1 : 50 type H1 : 23 type 3, x^2 =1.39, P = 0.2) was obtained in the F_2 seeds, and only F_2 seeds with the H1 type zymogram segregated in the F_3 generation. Our results also support the homo-heterodimer relationship between loci Pgi1 and Pgi3. Zymogram type 5 (Figure 12) lacks both bands 5 and 6, and is likely produced by a recessive allele (tentatively designated pgi1) at the locus Pgi1.

Table 12 shows the genetic data of cross between plants with zymograms type 2 and type 4 (Figure 12). Both parent plants have bands 3, 4, 5 and 6, but differ in the mobility of bands 3, 4 and 5 (Rf's of 0.42, 0.45 and 0.55 in type 2;

Table 12. The F_2 and F_3 segregarion data at the Pgi3 locus of soybeans.

	F_2 generation			F_3seeds from F_2 heterozyqous plants		
Pgi3-b/b x Pgi3-a/a	Pgi3-a/a	a/b	b/b	Pgi3-a/a	a/b	b/b
AV 68 x PI407.260 (type 4 x type 2)	98	195	89	27	56	21
x^2 (1 : 2 : 1)	0.59 (0.3 < P < 0.5)			1.31 (0.2 < P < 0.3)		
Pgi3-a/a x Pgi3-b/b						
PI407.260 x AV 68 (type 2 x type 4)	68	142	72	23	49	32
x^2 (1 : 2 : 1)	0.13 (0.7 < P < 0.8)			1.80 (0.1 < P < 0.2)		

AV 68 is G. max obtained from AVRDC, Taiwan. PI407.260 is G. soja from South Korea.

Rf's of 0.45, 0.49 and 0.57 in type 4). Heterozygotes display both parental bands as well as a new intralocus hybrid dimer band at Rf 0.435 (type H2, Figure 12). A 1 : 2 : 1 F_2 segregation ratio was found in both reciprocal crosses, and only F_2 seeds with H2 zymogram type segregated in the F_3 generation. The results indicate that the zymogram differences between type 2 and type 4 with respect to bands 3, 4, 5 and 6 are the consequence of a single nuclear locus, Pgi3, with two codominant alleles, designated as Pgi3-a (Rf 0.42) and Pgi3-b (Rf 0.45). The hypothesis that Pgi1 and Pgi3 interact to produce a three-band, homo-heterodimer complex model is further confirmed by this result, since the different mobility in Pgi3-a and Pgi3-b alter the mobility of their interlocus band 5 with Pgi1-b.

Using cell fractionation technique, Gorman (1983) reported that in soybean PGI zymogram, bands 3, 5 and 6 were active in the cytosol and the invariant band 4 in plastids. Gorman (1983) thus hypothesized that band 4 is the product of a distinct PGI locus. However, because the mobility of band 4 was altered together with band 3 in present study, both bands may be controlled by a single locus. Because of the weak activity of band 4 and its overlaping migration rate with other PGI bands, it is difficult to study the inheritance of PGI band 4.

A cross between plants of type 2 and type 3 (Figure 12) was studied for the linkage analysis of Pgi1 and Pgi3. The genotypes of types 2 and 3 for these two loci are Pgi1-b,

Pgi3-a and Pgi1-a, Pgi3-b, respectively. Since Pgi1-a and Pgi1-b are codominant, so are Pgi3-a and Pgi3-b, nine genotypes as well as phenotypes among the dihybrid F_2 progeny can be distinguished. Linkage analysis showed that Pgi1 and Pgi3 assorted independently (Table 18).

Phosphoglucomutase (PGM, EC 2.7.5.1)

Phosphoglucomutase isozymges catalyze the reversible reaction: (Lehninger, 1982).

glucose-1-phosphate $\xleftrightarrow{Mg^{++}}$ glucose-6-phosphate

There are two zones of PGM banding pattern observed in soybeans. The first has a simple band, band 1. Two variants of band 1 differing in mobility were reported by Kiang and Gorman (1983) as a consequence of a nuclear locus, Pgm1, with two codominant alleles, Pgm1-a (producing the Rf 0.51 band) and Pgm1-b (producing the Rf 0.54 band). The same observations were made in the present study.

The second banding zone of PGM has five homozygous zymogram types (Figure 13). Zymogram types 1 and 2 have a weak band 3 (Rf 0.74) and a strong band 2, but differ in their band 2 mobility (Rf 0.69 and 0.64, respectively); type 5 has a weak band 2 and a strong band 3, with the same migration rates as those of type 1. Zymogram types 3 and 4 have only one strong band at Rf 0.69 and 0.74, respectively. We hypothesized that a single nuclear locus with three codominant alleles is responsible for the three different

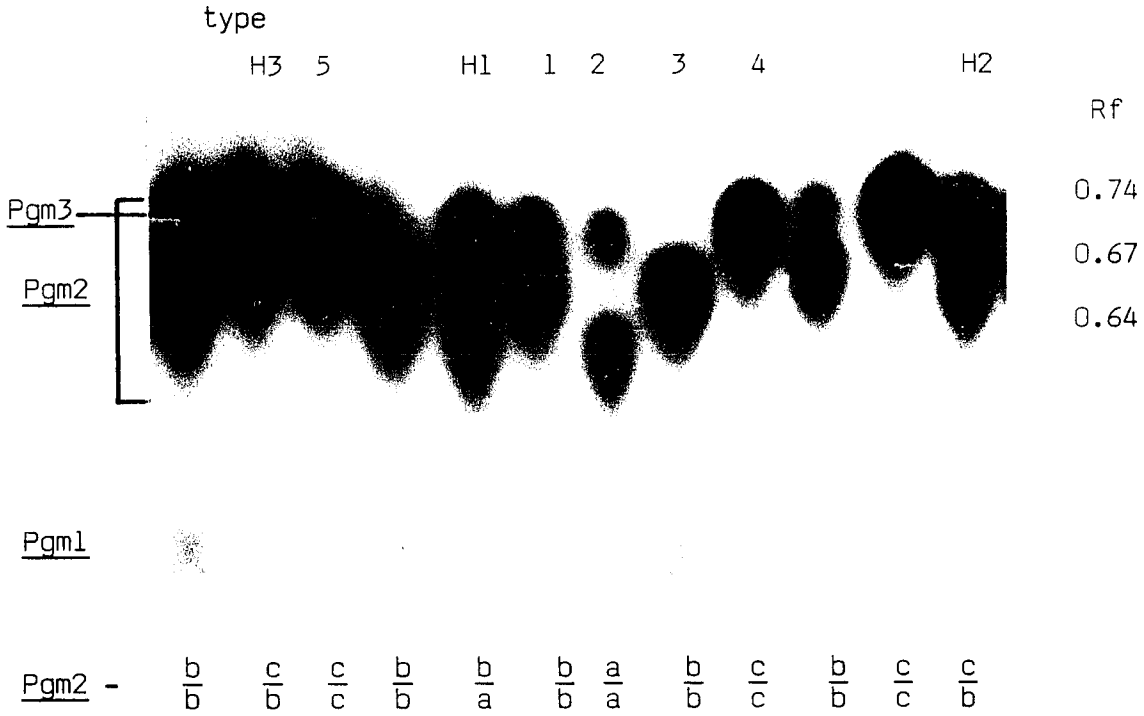


Figure 13. PGM zymogram of soybeans.

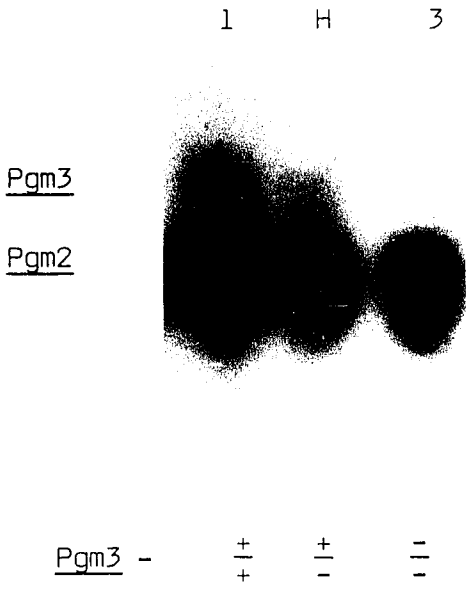


Figure 14. The three F_2 phenotypes of Pgm3 allozymes of cross between type 1 and type 3 plants.

migration rates of strong bands: Rf 0.64 of type 2, Rf 0.69 of types 1 and 3, and Rf 0.74 of types 4 and 5. The segregation data of the F_2 and F_3 of crosses between plants of types 1 and 2, types 1 and 4, types 3 and 5 are shown in Table 13. The heterozygous zymograms for these three crosses are H1, H2 and H3, respectively (Figure 13). They all showed both parental bands. The F_2 and the F_3 from heterozygous F_2 plants all displayed a 1 : 2 : 1 segregation ratio. There was no difference between reciprocal crosses, and only those F_2 seeds showing both parental bands segregated in the F_3. The results were same as predicted by the hypothesis. The single nuclear locus was designated as Pgm2, and three codominant alleles as Pgm2-a (Rf 0.64), Pgm2-b (Rf 0.67) and Pgm2-c (Rf 0.74).

Crosses were made between zymogram type 1 (with weak band 3) and type 3 (without weak band 3)(Figure 13 and 14) to study the genetic control of this weak band 3. A 1 : 2 : 1 segregation ratio was observed for type 1, H and 3 (Figure 14) in the F_2 seeds. There was no difference between reciprocal crosses (Table 14). The same segregation ratio was observed in the F_3 generation for those F_2 seeds with type H zymogram (Figure 14) that had band 3 even weaker than the type 1 (Figures 13 and 14). The results indicate a single nuclear locus, designated as Pgm3, with two incompletely dominant alleles (designated as Pgm3, pgm3) was responsible for the appearence of the weak band 3. The normal weak band 3 was the product of homozygous Pgm3/Pgm3.

Table 13. The F_2 and F_3 segregation data of the Pgm2 locus of soybeans

cross	F_2 generation			F_3 seeds from F_2 heterozygous plants		
Pgm2-b/b x Pgm2-a/a	Pgm2-b/b	b/a	a/a	Pgm2-b/b	b/a	a/a
AV 62 x PI407.265 (type 1 x type 2)	85	159	91	51	84	47
x^2 (1 : 2 : 1)	1.22 (0.2 < P < 0.3)			1.25 (0.2 < P < 0.3)		
Pgm2-a/a x Pgm2-b/b						
PI407.265 x AV 62 (type 2 x type 1)	56	139	73	11	20	9
x^2 (1 : 2 : 1)	2.54 (0.1 < P < 0.2)			0.20 (0.5 < P < 0.7)		
Pgm2-b/b x Pgm2-c/c	Pgm2-b/b	b/c	c/c	Pgm2-b/b	b/c	c/c
75002 x PI424.035 (type 3 x type 5)	65	119	48	42	99	40
x^2 (1 : 2 : 1)	2.65 (0.1 < P < 0.2)			1.64 (0.2 < P < 0.3)		
AV 62 x PI424.045 (type 1 x type 4)	63	112	59			
x^2 (1 : 2 : 1)	0.57 (0.3 < P < 0.5)					
Pgm2-c/c x Pgm2-b/b						
PI424.045 x AV 62 (type 4 x type 1)	18	40	20			
x^2 (1 : 2 : 1)	0.18 (0.5 < P < 0.7)					

AV 62 is G. max obtained from AVRDC, Taiwan. PI407.265, PI424.035, PI424.045 and 75002 are G. soja from South Korea.

Table 14. F_2 and F_3 segregation data of Pgm3 locus in soybeans

cross	F_2 generation			F_3 seeds from F_2 heterozygotes		
$\frac{Pgm3}{Pgm3} \times \frac{pgm3}{pgm3}$	$\frac{Pgm3}{Pgm3}$	$\frac{Pgm3}{pgm3}$	$\frac{pgm3}{pgm3}$	$\frac{Pgm3}{Pgm3}$	$\frac{Pgm3}{pgm3}$	$\frac{pgm3}{pgm3}$
AV215 x 75002 (type 1 x type 3)	25	46	20	39	76	45
x^2 (1 : 2 : 1)	0.56 (0.3 < P < 0.5)			0.85 (0.3 < P < 0.5)		
$\frac{pgm3}{pgm3} \times \frac{Pgm3}{Pgm3}$						
75002 x AV 215 (type 3 x type 1)	29	77	37	3	12	3
x^2 (1 : 2 : 1)	1.74 (0.1 < P < 0.2)			2.00 (0.1 < P < 0.2)		

AV 215 is G.max obtained from AVRDC, Taiwan. 75002 is G. soja from South Korea.

The genotype of the weaker activity in heterozygotes was Pgm3/pgm3, and pgm3/pgm3 was the genotype for the zymogram without any weak band 3.

Shikimic dehydrogenase (SDH, EC 1.1.1.25)

Shikimic dehydrogenase catalyzes the following reaction:

Shikimate + $NADP^+$ ————> 3-dehydroshikimate + NADPH

The zymogram of SDH in soybeans displayed six anodal bands with bands 3, 4 and 5 being most clear in dry seeds. No mobility differences were found for all these six bands among the seeds examined. However, some seed accessions showed consistently weak band 3 or band 5 (Figure 15). Since the mobility of the 4th band is about halfway between 3rd and 5th bands, and its band intensity is the strongest, we hypothesize that bands 3 and 5 are the products of two homozygous loci and band 4 is their heterodimer. More genetic study is needed to confirm this hypothesis.

Kunitz trypsin inhibitor (TI)

In soybeans, four trypsin inhibitor variants, designated as Ti-a (Rf 0.79), Ti-b (Rf 0.75), Ti-c (Rf 0.83) and ti are electrophoretically distinguishable, and are inherited as codominant or dominant-recessive alleles at a single locus (Singh et al., 1969; Hymowitz and Hadley, 1972; Hymowitz and Kaizuma, 1981). Only the Ti-a and Ti-b alleles were observed in the present G. soja study.

band	Rf
6	0.55
5	0.50
4	0.45
3	0.40
2	0.36
1	0.21

Figure 15. Zymogram of SDH in soybeans.

A summary of the genotypes of seed accessions for each allozyme and TI examined are listed in the Appendix III.

Table 15 lists all the protein loci and alleles which were confirmed or hypothesized from the zymogram studied. They were used in the subsequent analysis of genetic structure of G. soja.

Table 15. Protein loci and alleles confirmed or hypothesized in G. soja

enzyme symble	# of loci	#variable loci	alleles at the variable loci
ACO	5	5	Aco1-a, Aco1-b; Aco2-a, Aco2-b, Aco2-c, Aco2-d; Aco3-a, Aco3-b; Aco4-a, Aco4-b, Aco4-c; Aco5-a, Aco5-b
ADH	3	2	Adh1, adh1; Adh3, adh3
AM	3	1	Am3-s, Am3-f
AP	1	1	Ap-a, Ap-b, Ap-c, Ap-d
DIA	5	4	Dia1-a, Dia1-b; Dia2-a, Dia2-b, dia2; Dia3-a, Dia3-b, dia3; Dia4-a, Dia4-b
ENP	1	1	Enp-a, Enp-b, Enp-c
GOT	3	1	Got-a, Got-b, Got-c
IDH	4	4	Idh1-a, Idh1-b; Idh2-a, Idh2-b, Idh2-c; Idh3-a, Idh3-b, Idh3-c; Idh4-a, Idh4-b
LAP	1	1	Lap1-a, Lap1-b
MDH	4	0	
MPI	1	1	Mpi-a, Mpi-b, Mpi-c, Mpi-d
PGD	3	3	Pgd1-a, Pgd1-b, Pgd1-c, pgd1; Pgd2-a, Pgd2-b, Pgd2-c; Pgd3-a, Pgd3-b
PGI	3	3	Pgi1-a, Pgi1-b, pgi1; Pgi2, pgi2; Pgi3-a, Pgi3-b
PGM	3	3	Pgm1-a, Pgm1-b; Pgm2-a, Pgm2-b, Pgm2-c; Pgm3, pgm3
SDH	2	0	
TI	1	1	Ti-a, Ti-b
Total	43	31	

2. GENETIC LINKAGE OF PROTEIN LOCI

Very little information is available on genetic linkage in soybeans. The first seven small linkage groups of two or three loci each in soybeans were summarized by Bernard and Weiss (1973). Magenta flower color controlled by a mutant gene, *Wm*, is closely linked to *W1w1* (flower color) and together with *MS1ms1* (male sterility) are members of Linkage Group 8 (Buzzell *et al.*, 1977). Linkage between two chemical components of soybean seed controlled by the Kunitz trypsin inhibitor (TI) and the acid phosphatase (AP) loci, with a crossover frequency of 16.2% was reported as the Linkage Group 9 (Hildebrand *et al.*, 1980). A tenth linkage group was reported by Kilen and Barrentine (1983) between genes controlling reactions to phytophthora rot (*Rps1*) and metribuzin herbicide (*Hm*). The Linkage Group 11 was established between the *Rj1* locus, controlling restricted nodulation, and the *F* locus, controlling fasciated stem by Devine *et al.* (1983). The first linkage in soybeans between a biochemical locus (the α-amylase locus *Am3*) and a morphological locus (the pubescence color locus, *t*) was reported on Linkage Group 1 with 31.5% $\pm$ 2.2% recombination frequency (Kiang and Chiang, 1985). Kiang *et al.* (1985) reported that the *Lap1* locus is linked to the *Ap* locus with 19.88% $\pm$ 1.00% recombination frequency between the two loci in *G. soja* and suggested they may belong to Linkage Group 9. The *Lap1* and *Ti* loci in *G. max* were found to be

linked with a recombination frequency of 15.3% ± 0.9% on Linkage Group 9. Both the Lap1 and Ti loci are inherited independent of the flower color locus (W1) (Kiang and Chiang, 1986). In addition, the linkage of Ap, Lap1 and Ti three loci and the linkage between Adh1 and flower color W1 were established and the manuscripts are in preparation (Kiang, 1985).

In the present study, among the 60 isozyme gene pairs examined, only two linkage groups were found, involving six loci. One group has two loci, Pgd1 and Pgi1 with 15.34% ± 0.74% recombination frequency. Table 16 shows the F_2 data from two crosses, between two G. soja plants and between G. soja and G. max plants, segregating for the Pgd1 and Pgi1 gene pair. The x^2 test showed that the observed data significantly deviated from the 1 : 1 : 1 : 1 : 2 : 2 : 2 : 2 : 4 expected ratio for two genes which are inherited independently. Since the data of the two crosses were homogeneous, they were combined in calculating the frequency of recombination (Allard, 1956).

The other linkage group observed involved four enzyme loci, Ap, Ti, Lap1 and Pgd2. Table 17 shows the F_2 segregation data from the crosses involved plants with the four gene pairs. The data of the four crosses for Ap and Ti as well as Ti and Lap1 were not homogeneous. The recombination frequencies show in the linkage map in Figure 16 for these two pairs (Ap - Ti and Ti - Lap1) are the average of the four crosses. The data from crosses

Table 16. The F_2 data from soybean dihybrids segregating for the gene pair Pgdl and Pgil for linkage analysis.

Cross	genotypes*									N	x^2	P	Recombination frequency
	AB	Ab	aB	ab	AH	aH	HB	Hb	HH				
PI407.160 X PI424.032	97	6	6	107	48	41	40	21	205	571	394.2	<0.001	15.34% ± 0.74%
AV62 X PI424.032	169	2	8	157	55	50	53	51	303	848	693.5	<0.001	

* A and B represent Pgdl and Pgil loci, respectively.
H represents heterozygotes.

Table 17. The F_2 data from soybean dihybrids segregating for each pair of Ap, Ti, Lap1 and Pgd2 for linkage analysis.

Cross	genotypes*									N	x^2	P	Recombination frequency
	AB	Ab	aB	ab	AH	aH	HB	Hb	HH				
Ap - Ti													
PI407.160 X PI424.032	145	1	3	154	26	26	27	20	272	674	857.3	<0.001	8.24% ± 0.87%
PI424.032**	80	7	4	75	14	16	11	21	163	391	396.9	<0.001	11.40% ± 1.30%
AV68 X PI407.260	80	0	2	83	11	12	10	12	152	362	492.1	<0.001	7.00% ± 0.90%
AV62 X PI424.032	156	1	2	153	28	29	20	33	297	719	867.1	<0.001	11.00% ± 0.80%
Ap - Lap1													
PI407.160 X PI424.032	97	16	8	105	51	60	55	57	194	643	262.7	<0.001	22.65% ± 0.73%
PI424.032	68	5	7	66	34	37	26	39	139	421	196.5	<0.001	
AV68 X PI407.260	51	3	12	57	40	38	24	31	126	382	131.0	<0.001	
AV62 X PI424.032	111	11	13	113	61	71	61	63	223	719	301.9	<0.001	
Ap - Pgd2													
PI407.160 X PI424.032	60	32	38	59	81	85	67	72	180	674	23.5	<0.01	39.78% ± 0.98%
PI424.032	42	14	15	42	54	44	55	44	110	421	32.6	<0.01	
PI407.233 X PI407.192	46	18	21	52	73	61	57	57	134	519	31.3	<0.01	
PI407.160 X AV62	61	30	32	66	88	87	103	77	184	728	27.8	<0.01	

Table 17. (continued)

Cross	genotypes*									N	x^2	P	Recombination frequency
	AB	Ab	aB	ab	AH	aH	HB	Hb	HH				
Ti - *Lap1*													
PI407.160 X PI424.032	116	10	4	115	52	41	44	43	218	643	421.0	<0.001	17.65% ± 1.10%
PI424.032**	66	6	8	61	30	34	39	28	129	391	189.5	<0.001	21.30% ± 1.70%
AV68 X PI407.260	55	3	3	64	25	20	34	28	130	362	205.9	<0.001	18.10% ± 1.65%
AV62 X PI424.032	137	2	6	134	48	40	40	53	259	719	564.0	<0.001	14.80% ± 1.00%
Ti - *Pgd2*													
PI407.160 X PI424.032	68	34	27	73	73	75	68	58	198	674	63.7	<0.01	36.85% ± 1.39%
PI424.032	42	13	14	38	50	42	30	52	101	391	33.3	<0.01	
Lap1 - *Pgd2*													
PI107.160 X PI424.032	107	16	8	100	55	52	40	42	223	643	321.0	<0.001	20.62% ± 0.99%
PI424.032	65	9	3	72	33	36	30	33	140	421	209.4	<0.001	

* A and B each represents one locus of a gene pair studied.

** A natural hybrid segregating for all *Ap*, *Ti*, *Lap1* and *Pgd2* loci.

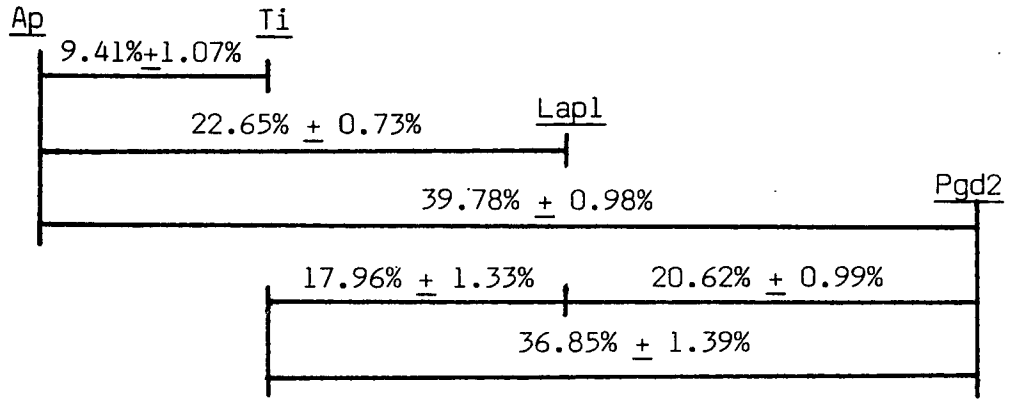


Figure 16. The relative locations of Ap, Ti, Lap1 and Pgd2 loci on Linkage Group 9 of soybeans.

involving the other gene pairs in Table 17 were homogeneous within each gene pair, and were combined in calculating the frequency of recombination. The linkage analysis (Allare, 1956) showed that the map units between Ap and Ti is 9.41 ± 1.07, Ap and Lap1 is 22.65 ± 0.73, Ap and Pgd2 39.78 ± 0.98, Ti and Lap1 17.96 ± 1.33, Ti and Pgd2 36.85 ± 1.39 and Lap1 and Pgd2 20.62 ± 0.99. The relative location of these four loci are shown in Figure 16. The Ti and Ap loci in G. max are in Linkage Group 9 (Hildebrand et al., 1980). Any pair of genes on the same chromosome in G. soja does not prove that the pair of genes belong to the same linkage group in G. max. However, in the present study, we conclude that Ti, Ap, Lap1 and Pgd2 belong to Linkage Group 9 in both G. max and G. soja because three of the crosses studied involved G. max and G. soja.

The recombination frequency observed in the present study between Ap and Lap1 (22.65% ± 0.73%) and Lap1 and Ti (17.96% ± 1.33%) are a lettle larger than those reported previously, 19.88% ± 1.00% of Ap - Lap1 (Kiang et al., 1985) and 15.3% ± 0.9% of Lap1 - Ti (Kiang and Chiang, 1986). But the recombination frequency observed between Ap and Ti in this study (9.41% ± 1.07%) in crosses between two G. soja plants and between G. soja and G. max plants is much smaller than 16.2% ± 1.5% reported by Hildebrand et al. (1980) in G. max based on 431 seeds. Different plants involved in the crosses in different growing seasons and different amount of seeds examined may explain these deviations.

Fluctuation in the recombination frequency in different crosses between the two loci in lima bean was suggested to be due to different growing seasons, different parental lines used in crosses, unequal recombination value between two sexes or a combination of these factors (Allard, 1956).

Table 18 lists the F_2 segregation data for 53 dihybrids of enzyme loci. All of the allozymes in this linkage study are codominant, therefore nine genotypes among the dihybrid F_2 progeny can be distinguished. However, some gene pairs involving Dia2, Dia3, Idh1 and Idh2 can only be distinguished into six or even four phenotypes in their F_2 progeny because some bands produced by different genes have the same mobility. All these 53 gene pairs were found to segregate independently in the expected 1 : 1 : 1 : 1 : 2 : 2 : 2 : 2 : 4 or 1 : 1 : 2 : 3 : 3 : 6 or 1 : 3 : 3 : 9 ratio.

Table 18 The F_2 segregation data from soybean dihybrids for linkage analysis

gene pair A - B	cross	expected ratio 1 : 1 : 1 : 1 : 2 : 2 : 2 : 2 : 4									N	x^2	P
		AB	Ab	aB	ab	AH	aH	HB	Hb	HH			
Am - Ap	PI407.233 X PI407.192	40	39	32	33	63	64	63	63	122	519	3.786	0.80
Am -Dial	PI407.233 X PI407.192	38	32	31	31	64	63	55	63	118	495	2.794	0.90
Am -Pgd2	PI407.233 X PI407.192	37	39	34	27	66	67	56	59	134	519	4.958	0.70
Am -Pgd3	PI487.431 X PI486.220	26	25	22	27	50	40	48	51	97	386	2.343	0.95
Ap -Dial	PI407.233 X PI407.192	36	29	26	34	64	68	61	65	112	495	4.012	0.80
Ap -Dia4	PI487.431 X PI486.220	24	25	36	21	59	57	56	51	111	440	5.517	0.70
Ap - Mpi	PI407.233 X PI407.192	33	33	27	30	85	70	73	70	134	555	6.400	0.50
Ap -Pgd1	AV68 X PI407.260	22	20	21	26	49	50	47	43	84	362	2.378	0.95
Ap -Pgd3	PI487.431 X PI486.220	39	47	49	53	93	95	89	89	166	720	4.156	0.80
Ap -Pgi1	AV62 X PI424.032	41	34	31	34	59	62	67	60	123	511	3.852	0.80
Ap -Pgi3	AV68 X PI407.260	19	31	25	18	48	46	50	59	86	382	8.360	0.30
Ap -Pgm2	AV62 X PI424.045	16	14	15	14	35	22	32	31	54	233	3.853	0.80
Dial- Mpi	PI407.233 X PI407.192	24	31	22	27	52	45	60	48	113	422	4.873	0.70
Dial-Pgd1	AV68 X PI407.260	19	20	22	24	41	40	48	49	99	362	3.264	0.90
Dial-Pgd2	PI407.233 X PI407.192	31	28	31	32	60	57	63	65	128	495	1.082	0.99
Dial-Pgi3	AV68 X PI407.260	20	21	21	23	57	45	44	48	102	381	4.036	0.80
Dial-Pgm2	AV62 X PI407.265	18	17	18	28	50	45	35	40	84	335	6.961	0.50
Dia4-Lap1	PI487.431 X PI486.220	29	39	23	25	63	64	53	65	126	487	6.946	0.50
Dia4-Pgd3	PI487.431 X PI486.220	19	21	22	26	62	51	51	54	107	413	5.630	0.60
Dia4-Pgm1	PI487.431 X PI486.220	23	16	20	20	47	48	48	42	98	362	3.780	0.80

Table 18 (continued)

gene pair A - B	cross	expected ratio 1 : 1 : 1 : 1 : 2 : 2 : 2 : 2 : 4 AB	Ab	aB	ab	AH	aH	HB	Hb	HH	N	x^2	P
Lap1-Pgd1	AV68 X PI407.260	17	20	26	25	46	36	47	43	102	362	6.016	0.50
Lap1-Pgd3	PI487.431 X PI486.220	17	27	26	26	45	59	53	49	84	386	7.452	0.40
Lap1-Pgi1	AV62 X PI424.032	38	35	36	37	57	63	60	66	119	511	4.420	0.80
Lap1-Pgi3	AV68 X PI407.260	19	23	22	18	45	51	57	47	99	381	4.937	0.70
Mpi -Pgd1	AV68 X PI407.252	6	15	10	13	18	20	20	17	37	156	5.930	0.60
Mpi -Pgd2	PI407.233 X PI407.192	22	29	22	25	43	60	53	48	120	422	7.005	0.50
Mpi -Pgd3	PI487.431 X PI486.220	28	36	33	30	65	65	65	64	104	490	4.972	0.70
Pgd1-Pgd2	PI407.160 X PI424.032	28	33	32	33	61	61	45	54	120	467	5.030	0.70
Pgd1-pgi3	AV68 X PI407.260	27	23	19	23	43	42	43	44	98	362	2.577	0.95
Pgd1-Pgm2	PI407.265 X Av62	11	8	11	8	20	18	23	22	35	156	2.430	0.95
Pgd1- Ti	Av68 X PI407.260	22	22	23	28	48	44	44	40	91	362	2.166	0.95
Pgd2-Pgi3	PI407.160 X PI424.032	42	43	48	30	78	87	75	90	181	674	7.240	0.40
Pgd3-Pgm1	PI487.431 X PI486.220	14	20	20	17	49	29	37	53	72	311	12.651	0.10
Pgi1- Ti	PI407.160 X PI424.032	23	23	32	34	66	62	60	54	113	467	5.400	0.70
Pgi3-pgi1	PI424.032 X PI407.260	11	11	15	12	21	25	29	29	55	208	2.866	0.90

		expected ratio 1 : 3 : 3 : 9 AB	A(b+H)	(a+H)B	(a+H)(b+H)			
Dia2-Dia3	PI407.233 X PI407.192	33	102	81	279	495	2.551	0.40

Table 18 (continued)

gene pair A - B	cross	expected ratio 1 AB : 1 aB : 2 HB : 3 A(b+H) : 3 a(b+H) : 6 H(b+H)	N	x^2	P
Ap -Idh1	AV68 X PI407.260	31 : 24 : 40 : 75 : 70 : 142	382	3.592	0.60
Ap -Idh2	AV68 X PI407.260	33 : 24 : 49 : 73 : 70 : 133	382	4.318	0.50
Ap -Idh3	AV68 X PI407.260	24 : 28 : 35 : 67 : 77 : 148	379	4.999	0.40
Dia1-Dia2	PI407.233 X PI407.192	34 : 39 : 63 : 92 : 84 : 183	495	3.305	0.60
Dia1-Dia3	PI407.233 X PI407.192	31 : 21 : 61 : 95 : 103 : 184	495	4.388	0.40
Dia1-Idh1	AV68 X PI407.260	23 : 19 : 53 : 69 : 67 : 151	382	2.418	0.70
Dia1-Idh2	AV68 X PI407.260	26 : 19 : 61 : 66 : 67 : 143	382	5.602	0.30
Dia2- Mpi	PI407.233 X PI407.192	29 : 25 : 58 : 65 : 82 : 163	422	3.625	0.50
Dia2-Pgd2	PI407.233 X PI407.192	30 : 30 : 76 : 90 : 88 : 181	495	3.730	0.50
Dia3- Mpi	PI407.233 X PI407.192	20 : 22 : 53 : 74 : 85 : 168	422	3.675	0.50
Dia3-Pgd2	PI407.233 X PI407.192	25 : 29 : 60 : 95 : 90 : 196	495	2.035	0.80
Idh1-Pgd1	AV68 X PI407.260	25 : 22 : 44 : 65 : 67 : 139	362	0.512	0.99
Idh1-Pgi3	AV68 X PI407.260	19 : 27 : 49 : 69 : 72 : 146	382	1.588	0.90
Idh2-Pgd1	AV68 X PI407.260	24 : 26 : 51 : 66 : 63 : 132	362	1.824	0.80
Idh2-Pgi3	AV68 X PI407.260	24 : 31 : 51 : 64 : 68 : 144	382	3.347	0.60
Idh3-Pgd1	AV68 X PI407.260	14 : 13 : 53 : 74 : 76 : 130	360	10.525	0.05
Idh3-Pgi3	AV68 X PI407.260	22 : 20 : 44 : 76 : 68 : 150	380	1.859	0.80

3. GENETIC STRUCTURE OF G. SOJA

(1). Genetic diversity of G. soja seed accessions from South Korea and Japan.

The allele frequencies of proteins examined for each seed accession are listed in the Appendix III. Table 19 lists the 30 variable loci examined, alleles observed at these loci, the frequency of each allele, and the number of G. soja accessions in which each allele was observed. Pgm3 locus is not included because genotypes of many accessions in this locus cannot be determined due to band overlapping with Pgm2-c band. The allele frequency (P) in Table 19 is based on allele frequencies listed in the Appendix III and weighted by number of accessions.

A total of 78 alleles was recorded with an average of 2.60 alleles per variable locus. When Pgm3 as well as the invariant loci of the two confirmed amylase and the ten hypothesized loci of MDH, SDH, GOT, ADH and DIA were included, the average number of alleles per locus was 2.14 in G. soja of South Korean and Japanese accessions. Hamrick (1979) in his review paper reported 1.80 (range 1.00 - 3.75) alleles per locus as an average of 39 annuals. While these data suggest that G. soja has above average of allelic richness, many alleles are rare, being found in three or fewer seed accessions (Table 19).

The expected heterozygosity ($\overline{H}exp$) for G. soja based on the total seed accessions examined is 0.237 for variable

Table 19. List of alleles observed, their frequencies (P) and number of seed accessions (N) in which they were observed.

	S. Korea		Japan		S. Korea + Japan	
	P	N	P	N	P	N
Aco1-a	0.02	1	0	0	0.01	1
Aco1-b	0.98	62	1.00	9	0.99	71
Aco2-a	0.03	4	0.37	5	0.07	9
Aco2-b	0.01	1	0	0	0.01	1
Aco2-c	0.91	60	0.63	7	0.88	67
Aco2-d	0.05	4	0	0	0.04	4
Aco3-a	0.84	54	1.00	9	0.86	63
Aco3-b	0.16	11	0	0	0.14	11
Aco4-a	0.11	7	0.22	4	0.12	11
Aco4-b	0.85	54	0.78	8	0.85	62
Aco4-c	0.04	3	0	0	0.03	3
Aco5-a	1.00	63	0.89	8	0.99	71
Aco5-b	0	0	0.11	1	0.01	1
Adh1	0.99	63	1.00	9	0.99	72
adh1	0.11	2	0	0	0.01	2
Adh3	0.99	63	1.00	9	0.99	72
adh3	0.01	1	0	0	0.01	1
Am3-s	0.33	30	0.81	8	0.39	38
Am3-f	0.67	48	0.19	4	0.61	52
Ap-a	0.39	36	0.52	7	0.46	43
Ap-b	0.16	15	0.20	3	0.16	18
Ap-c	0.42	37	0.28	5	0.35	42
Ap-d	0.03	3	0	0	0.03	3
Dia1-a	0.59	44	0.53	6	0.58	50
Dia1-b	0.41	37	0.47	7	0.42	44
Dia2-a	0.28	21	0.25	3	0.28	24
Dia2-b	0.71	49	0.75	7	0.71	56
dia2	0.01	1	0	0	0.01	1
Dia3-a	0.03	2	0	0	0.03	2
Dia3-b	0.91	60	1.00	9	0.91	69
dia3	0.06	5	0	0	0.06	5
Dia4-a	1.00	63	0.89	8	0.99	71
Dia4-b	0	0	0.11	1	0.01	1
Enp-a	0.17	13	0.24	5	0.18	18
Enp-b	0.82	54	0.69	8	0.80	62
Enp-c	0.01	1	0.07	2	0.02	3
Got-a	0.05	3	0	0	0.04	3
Got-b	0.93	59	1.00	9	0.95	68
Got-c	0.02	1	0	0	0.01	1
Idh1-a	<0.01	2	0	0	<0.01	2
Idh1-b	1.00	63	1.00	9	1.00	72

Table 19 (continued)

	S. Korea		Japan		S. Korea + Japan	
	P	N	P	N	P	N
Idh2-a	0.93	60	0.87	8	0.92	68
Idh2-b	0.07	10	0.12	2	0.08	12
Idh2-c	0	0	0.02	1	<0.01	1
Idh3-a	0.02	4	0.14	2	0.04	6
Idh3-b	0.87	56	0.78	8	0.86	64
Idh3-c	0.11	11	0.08	3	0.10	14
Idh4-a	0.93	60	0.87	8	0.92	68
Idh4-b	0.07	10	0.13	3	0.08	13
Lap1-a	0.20	18	0.33	5	0.21	23
Lap1-b	0.80	57	0.67	7	0.79	64
Mpi-a	0.02	3	0.12	1	0.03	4
Mpi-b	0.57	43	0.61	7	0.57	50
Mpi-c	0.41	33	0.27	5	0.41	38
Mpi-d	0.02	3	0	0	0.02	3
Pgd1-a	0.13	16	0.08	4	0.12	20
Pgd1-b	0.69	50	0.69	8	0.69	58
Pgd1-c	0.18	15	0.18	4	0.18	19
pgd1	0	0	0.06	2	0.01	2
Pgd2-a	0.84	56	1.00	9	0.85	65
Pgd2-b	0.14	12	0	0	0.13	12
Pgd2-c	0.02	2	0	0	0.02	2
Pgd3-a	0	0	0.11	1	0.01	1
Pgd3-b	1.00	63	0.89	8	0.99	71
Pgi1-a	0.04	4	0	0	0.04	4
Pgi1-b	0.94	61	1.00	9	0.95	70
pgi1	0.02	2	0	0	0.01	2
Pgi2	0.45	29	0.73	8	0.49	37
pgi2	0.55	36	0.27	6	0.51	42
Pgi3-a	0.02	2	0	0	0.02	2
Pgi3-b	0.98	62	1.00	9	0.98	71
Pgm1-a	0.98	62	0.82	8	0.96	70
Pgm1-b	0.02	2	0.18	4	0.04	6
Pgm2-a	0.05	5	0	0	0.04	5
Pgm2-b	0.69	49	0.79	8	0.70	57
Pgm2-c	0.26	23	0.21	5	0.26	28
Ti-a	0.95	61	1.00	9	0.96	70
Ti-b	0.05	4	0	0	0.04	4
Total # alleles	73		56		78	
Av. # alleles/locus	2.43		1.87		2.60	

loci, and 0.160 for total loci studied (Table 20). This is about the average (0.154, range 0.000 - 0.414) reported by Hamrick (1979) for 39 annuals. The most variable locus in G. soja is Ap with four alleles and Hexp = 0.639 (Table 20).

The proportion of polymorphic loci for G. soja in this study is 67.4% at the 99% polymorphism level and 46.5% at the 95% level. An average value of 46.2% was calculated including 44 annuals by Hamrick (1979).

The results indicate that the genetic diversity of G. soja is about average for annuals.

(2). Comparison of two geographic populations: South Korea and Japan.

(A) Genetic diversity

When seed accessions were grouped into two geographic populations, South Korea and Japan, they were significantly different in the number of alleles per variable locus by paired t-test (2.43 and 1.87 for South Korea and Japan, respectively, Table 19). The proportion of polymorphic loci at the 99% polymorphism level is 60.46% in South Korea and 44.19% in Japan. The results may be biased by the fact that distinctly different number of seed accessions were included in the study from the two areas (63 from South Korea and 9 from Japan).

The mean expected heterozygosities in South Korea and Japan geographic populations are 0.155 and 0.162, respectively. They are not statistically different using

Table 20. The expected heterozygosity of G. soja based on seed accessions from S. Korea and Japan

	S. Korea		Japan		S. Korea + Japan	
Loci	#alleles	Hexp	#alleles	Hexp	#alleles	Hexp
Aco1	2	0.0392	1	0.0000	2	0.0198
Aco2	4	0.1684	2	0.4662	4	0.2190
Aco3	2	0.2688	1	0.0000	2	0.2408
Aco4	3	0.2638	2	0.3432	3	0.2622
Aco5	1	0.0000	2	0.1958	2	0.0198
Adh1	2	0.0198	1	0.0000	2	0.0198
Adh3	2	0.0198	1	0.0000	2	0.0198
Am3	2	0.4422	2	0.3078	2	0.4758
Ap	4	0.6450	3	0.6112	4	0.6394
Dia1	2	0.4838	2	0.4982	2	0.5564
Dia2	3	0.4174	2	0.3750	3	0.4174
Dia3	3	0.1674	1	0.0000	3	0.1674
Dia4	1	0.0000	2	0.1958	2	0.0198
Enp	3	0.2986	3	0.4614	3	0.3272
Got	3	0.1322	1	0.0000	3	0.0958
Idh2	2	0.1302	3	0.2283	3	0.1472
Idh3	3	0.2306	3	0.3656	3	0.2488
Idh4	2	0.1302	2	0.2262	2	0.1472
Lap1	2	0.3200	2	0.4422	2	0.3318
Mpi	4	0.5062	3	0.5429	4	0.5057
Pgd1	3	0.4746	4	0.4815	4	0.4770
Pgd2	3	0.2744	1	0.0000	3	0.2602
Pgd3	1	0.0000	2	0.1958	2	0.0198
Pgi1	3	0.1144	1	0.0000	3	0.0958
Pgi2	2	0.4950	2	0.3942	2	0.4998
Pgi3	2	0.0392	1	0.0000	2	0.0392
Pgm1	2	0.0392	2	0.2952	2	0.0768
Pgm2	3	0.4538	2	0.3318	3	0.4408
Ti	2	0.0950	1	0.0000	2	0.0768
Mean		0.2300		0.2399		0.2368
Mean*		0.1551		0.1618		0.1597

* Mean $\overline{H}exp$ based on all (43) loci examined.

paired t-test. In fact, these two populations are similar in their commonest alleles for most of loci examined (Table 20).

(B) Apportion of gene diversity within and among the two geographic populations.

According to Nei (1973) gene diversity in a total population can be divided into gene diversitied within and between populations. The apportionment of gene diversities within and between populations was determined by the gene diversity analysis (Nei, 1973). The total gene diversity H_T was obtained by calculating the weighted average allele frequencies of the polymorphic loci over all populations ($H_T = 1 - \sum \bar{x}_i^2$). The mean gene diversity within populations (H_S) at polymorphic loci was the weighted average over all populations of the values of $1 - \sum x_i^2$ for each locus. D_{ST} is the gene diversity among populations. These three parameters are related as $H_T = H_S + D_{ST}$. The coefficient of gene differentiation G_{ST} measures the amount of genetic variation in the whole population that is attributable to genetic differentiation among populations. It is given by $G_{ST} = D_{ST} / H_T$. When $G_{ST} = 0$ there is no genetic differentiation among populations. The polymorphic loci were determined at the 99% polymorphism level. The exclusion of the invariant loci in the calculations of H_T, H_S, D_{ST} and G_{ST} was for emphasizing on partitioning the variation detected.

Coefficient of gene differentiation (G_{ST}) between two

(South Korean and Japanese) geographic populations of G. soja was 0.0459 (Table 21). It indicates that only 4.59% of the total genetic variation occurs between these two geographic populations, and 95.41% of the total genetic variation resides within each of the two geographic populations. The amount of genetic variation within each geographic population is greatly increased because of the large geographic areas, and each geographic population consists of many populations (or subpopulations).

(3). Comparison of 12 seed accessions of G. soja

Table 22 lists the alleles and their frequencies observed in the 12 G. soja seed accessions selected for study based on their latitudinal locations (Table 1). Only those variable loci among these 12 accessions are listed. At least 50 seeds per accession were examined. High genetic uniformity was found in each of the accessions, K109, E4, K102, K52, K101 M and K113, which showed monomorphism for all 43 loci scored. Seeds of each of the five accessions, K7, K9, K28, K31 and K42 were received in two or three batches. High genetic uniformity was found within each batch. Thus, the allele frequency for each of these five accessions was calculated and weighted by number of batches. Based on these allele frequencies, the relative measures of genetic distance and identity (Nei, 1972) between 12 accessions were calculated (Table 23).

The mean value of genetic distance (D_N) for these 12

Table 21. Apportion of gene diversity within and between two geographic populations (S. Korean and Japanese) of G. soja.

Loci	H_T	H_S	D_{ST}	G_{ST}
Aco1	0.0198	0.0196	0.0002	0.0101
Aco2	0.3665	0.3173	0.0492	0.1341
Aco3	0.1472	0.1344	0.0128	0.0870
Aco4	0.3150	0.3043	0.0107	0.0340
Aco5	0.1045	0.0979	0.0066	0.0630
Adh1	0.0100	0.0099	0.0001	0.0100
Adh3	0.0100	0.0099	0.0001	0.0100
Am3	0.4902	0.3750	0.1152	0.2350
Ap	0.6379	0.6281	0.0098	0.0153
Dia1	0.4928	0.4910	0.0018	0.0037
Dia2	0.3969	0.3962	0.0007	0.0018
Dia3	0.0869	0.0837	0.0032	0.0368
Dia4	0.1045	0.0979	0.0066	0.0630
Enp	0.3864	0.3800	0.0064	0.0166
Got	0.0681	0.0661	0.0020	0.0294
Idh2	0.1809	0.1793	0.0016	0.0088
Idh3	0.3040	0.2981	0.0059	0.0194
Idh4	0.1800	0.1782	0.0018	0.0100
Lap1	0.3900	0.3811	0.0089	0.0228
Mpi	0.5320	0.5246	0.0074	0.0139
Pgd1	0.4785	0.4781	0.0004	0.0008
Pgd2	0.1486	0.1372	0.0114	0.0767
Pgd3	0.1045	0.0979	0.0066	0.0630
Pgi1	0.0586	0.0572	0.0014	0.0239
Pgi2	0.4838	0.4446	0.0392	0.0810
Pgi3	0.0198	0.0196	0.0002	0.0101
Pgm1	0.1800	0.1672	0.0128	0.0711
Pgm2	0.3966	0.3928	0.0038	0.0095
Ti	0.0488	0.0475	0.0013	0.0266
Mean	0.2463	0.2350	0.0113	0.0459*

* calculated as $\frac{0.2463 - 0.2350}{0.2463} = 0.0459$

Table 22. Alleles and their frequencies of the 12 G. soja accessions.

	K109	E4	K9	K7	K102	K28	K52	K42	K101	K31	M	K113
Aco2-a	0	0	0.50	0.50	0	0	0	0	0	0	1.00	1.00
Aco2-b	0	0	0	0.50	0	0	0	0	0	0	0	0
Aco2-c	1.00	1.00	0.50	0	1.00	0.50	1.00	1.00	1.00	0.50	0	0
Aco2-d	0	0	0	0	0	0.50	0	0	0	0.50	0	0
Aco3-a	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco3-b	0	0	0	0.50	0	0	0	0	0	0	0	0
Aco4-a	0	0	0	0	0	0	0	1.00	0	0	1.00	0
Aco4-b	1.00	1.00	1.00	1.00	1.00	0.50	1.00	0	1.00	1.00	0	1.00
Aco4-c	0	0	0	0	0	0.50	0	0	0	0	0	0
Aco5-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0
Aco5-b	0	0	0	0	0	0	0	0	0	0	0	1.00
Adh3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00
adh3	0	0	0	0	0	0	0	0	0	0.50	0	0
Am3-s	1.00	1.00	0	0	0	1.00	1.00	1.00	1.00	0.50	1.00	0
Am3-f	0	0	1.00	1.00	1.00	0	0	0	0	0	0	1.00
Ap-a	1.00	1.00	0	0.50	1.00	0.50	1.00	1.00	1.00	0	0	0
Ap-b	0	0	0	0	0	0	0	0	0	0	0	1.00
Ap-c	0	0	1.00	0.50	0	0	0	0	0	1.00	1.00	0
Ap-d	0	0	0	0	0	0.50	0	0	0	0	0	0
Dial-a	1.00	1.00	0	0.50	1.00	0.50	1.00	1.00	0	0	0	0
Dial-b	0	0	1.00	0.50	0	0.50	0	0	1.00	1.00	1.00	1.00
Dia2-a	0	0	0	0	0	1.00	1.00	1.00	0	0	1.00	1.00
Dia2-b	1.00	1.00	1.00	1.00	1.00	0	0	0	1.00	0.50	0	0
dia2	0	0	0	0	0	0	0	0	0	0.50	0	0
Dia3-a	0	0	0	0	0	0	1.00	0	0	0	0	0
Dia3-b	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	1.00
Dia4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0
Dia4-b	0	0	0	0	0	0	0	0	0	0	0	1.00
Enp-a	1.00	0	0	0	0	0	1.00	0.50	0	0	0	0
Enp-b	0	1.00	1.00	1.00	1.00	1.00	0	0.50	1.00	1.00	1.00	1.00
Idh2-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	0.50	1.00	1.00
Idh2-b	0	0	0	0	0	0	0	0	1.00	0.50	0	0
Idh3-a	1.00	0	0	0.50	0	0	0	0	0	0	0	0
Idh3-b	0	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	0.50	1.00	1.00
Idh4-b	0	0	0	0	0	0	0	0	1.00	0.50	0	0

Table 22 (continued)

	K109	E4	K9	K7	K102	K28	K52	K42	K101	K31	M	K113
Lapl-a	0	0	0.50	0.50	0	0	0	0	1.00	0	1.00	0
Lapl-b	1.00	1.00	0.50	0.50	1.00	1.00	1.00	1.00	0	1.00	0	1.00
Mpi-a	1.00	0	0	0	0	0	0	0	0	0	0	0
Mpi-b	0	1.00	0	0.50	1.00	0.50	0	1.00	1.00	1.00	1.00	0
Mpi-c	0	0	0.50	0.50	0	0.50	1.00	0	0	0	0	1.00
Mpi-d	0	0	0.50	0	0	0	0	0	0	0	0	0
Pgdl-b	0	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00
Pgdl-c	1.00	0	0	0	0	0	0	1.00	0	0	0	0
Pgd2-a	1.00	1.00	0.50	1.00	0	1.00	1.00	1.00	1.00	0.50	1.00	1.00
Pgd2-b	0	0	0.50	0	1.00	0	0	0	0	0	0	0
Pgd2-c	0	0	0	0	0	0	0	0	0	0.50	0	0
Pgd3-a	0	0	0	0	0	0	0	0	0	0	1.00	0
Pgd3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00
Pgil-b	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
pgil	0	0	0	0.50	0	0	0	0	0	0	0	0
Pgi2	1.00	0	0	0.50	0	0	0	0	1.00	0	1.00	1.00
pgi2	0	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0	1.00	0	0
Pgml-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00
Pgml-b	0	0	0	0	0	0	0	0	0	0	1.00	0
Pgm2-b	0	1.00	1.00	0.50	0	0.50	1.00	1.00	1.00	1.00	1.00	1.00
Pgm2-c	1.00	0	0	0.50	1.00	0.50	0	0	0	0	0	0

Table 23. Nei's measures of genetic distance (D_N) and identity (I_N) between twelve G. soja seed accessions of South Korean and Japan.

	K109	E4	K9	K7	K102	K28	K52	K42	K101	K31	M	K113
K109		0.857	0.769	0.837	0.833	0.828	0.833	0.850	0.786	0.758	0.690	0.714
E4	0.154		0.890	0.900	0.929	0.939	0.905	0.923	0.881	0.908	0.786	0.786
K9	0.263	0.116		0.936	0.890	0.886	0.842	0.810	0.842	0.930	0.818	0.854
K7	0.177	0.105	0.067		0.900	0.875	0.824	0.817	0.850	0.888	0.812	0.875
K102	0.182	0.074	0.116	0.105		0.889	0.833	0.850	0.810	0.870	0.714	0.762
K28	0.189	0.063	0.121	0.134	0.117		0.914	0.914	0.840	0.903	0.828	0.840
K52	0.182	0.100	0.172	0.194	0.182	0.090		0.898	0.786	0.821	0.738	0.786
K42	0.162	0.081	0.210	0.202	0.162	0.090	0.107		0.803	0.838	0.803	0.755
K101	0.241	0.127	0.172	0.163	0.211	0.174	0.241	0.220		0.883	0.810	0.762
K31	0.277	0.097	0.073	0.119	0.139	0.102	0.197	0.176	0.125		0.807	0.807
M	0.370	0.241	0.201	0.201	0.336	0.189	0.304	0.220	0.211	0.214		0.786
K113	0.336	0.241	0.158	0.133	0.272	0.174	0.241	0.282	0.272	0.214	0.241	

I_N

D_N

accessions is 0.179 (range 0.063 - 0.370), which is similar to the level that Ayala and Kiger (1980) calculated for populations in either the first or the second stage of speciation, which are subspecies or incipient species. Postzygotic RIMs (Reproductive Isolating Mechanisms) in the form of hybrid sterility usually exhibit between subspecies and between incipient species. But no hybrid sterility was observed in any of the crosses made between these 12 seed accessions. Since seeds of all the 12 accessions were increased in greenhouse and / or farm, the number of original plants sampled from the field and contributed to the seed which were examined in this study is unknown. The high seed purity of each accession suggests that the seeds used in the present study may have been descended from a small number of plants. Because of the sampling error (Founder's principle, random genetic drift), the seed accessions examined in this study may represent only a portion of genetic variation existing in the original populations from which the seeds were sampled. In addition to the sampling error, the highly selfing breeding system may have resulted in most seeds in each accession being fixed for one allele in all or most loci examined. The high genetic uniformity within accessions, and the occasional loci fixed by chance for different alleles among accessions will greatly enlarge the measure of genetic distance (D_N). However, if the genotypes observed for each seed accession represent the most common ones in their wild populations, the D_N obtained

can still be used as a measure for relative genetic distance between seed accessions. The data collected from E4 and natural population S seem to support the idea that the observed genotypes represent the most common alleles in the wild populations. Both E4 and population S were sampled from Morioka, Japan. All of the alleles fixed in E4 were the most common alleles found in each locus of population S (Appendix III).

The D_Ns' of the following accession pairs were small and less than 0.080: E4 - K28 (0.063), K9 - K7 (0.067), K9 - K31 (0.073) and E4 - K102 (0.074) (Table 23). Although the geographical distance between E4 and K7, K9, K28, K31 and K102 is large (Figure 1), the latitudinal difference among their sampling sites are relatively small. The results suggest that latitudinal difference has more influence on the genetic differentiation of populations than the geographical distance. The high genetic distances found between K109 and M (D_N = 0.370), and K109 and K113 (D_N = 0.336) seem to support this suggestion. The latitudinal differences between K109 and K113, and between K109 and M are the two largest found among all the accession pairs examined.

Principal components analysis based on 57 allele frequencies (Table 22) was applied to visualize the relationships among these 12 seed accessions by projection of accession positions in 57-dimensional space onto a plane defined by the first two principal components. The input

data matrix consisted of allele frequencies for 57 alleles in 12 accessions. From this matrix, a variance-covariance matrix was derived from which principal components were extracted. The principal components represent least-square lines through the multidimentional space defined by the 57 alleles. The computer program 'CLUSTAN' (Wishart, 1978) was used. The positions of the 12 accessions on the first two principal component axes based on allele frequency is graphically illustrated in Figure 17. The first two components accounted for 19.50% and 14.92% of the total variance, respectively. No distinctive group of accessions was formed by these two components. A graph of the first principal component against latitudinal location for the 12 accessions is shown in Figure 18. A strong correlation ($r = -0.695$, $P = 0.05$) was observed between these two axes. The correlation between latitudinal location and the remaining principal components were not significantly different from zero. Thus, there is about 20% of the total genetic variation of these 12 G. soja seed accessions associated significantly with latitude. Alleles of Aco2-a, Aco2-c, Ap-a, Dial-a and Dial-b contribute the most to the first principal component.

(4). Four wild populations collected along the Kitakami river

Seeds were collected from four natural populations near the Kitakami river, Iwate Prefecture, nothern Japan, Honshu Island. Seeds from each individual plant were

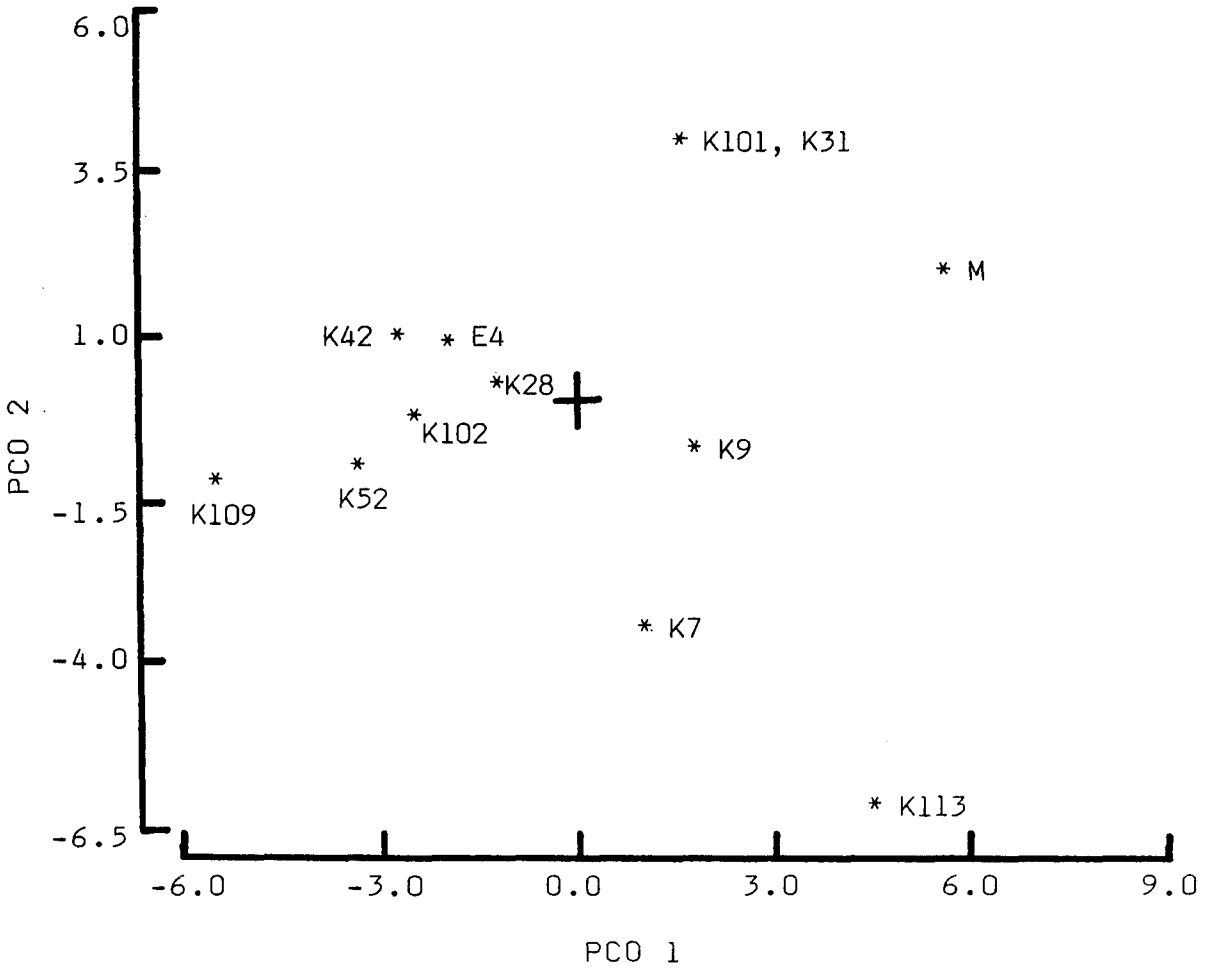


Figure 17. A plot of 12 G. soja seed accessions on the first two principal component axes based on allele frequencies. The first principal component (PCO 1) accounts for 19.50% and the second principal component (PCO 2) accounts for 14.92% of the total variance.

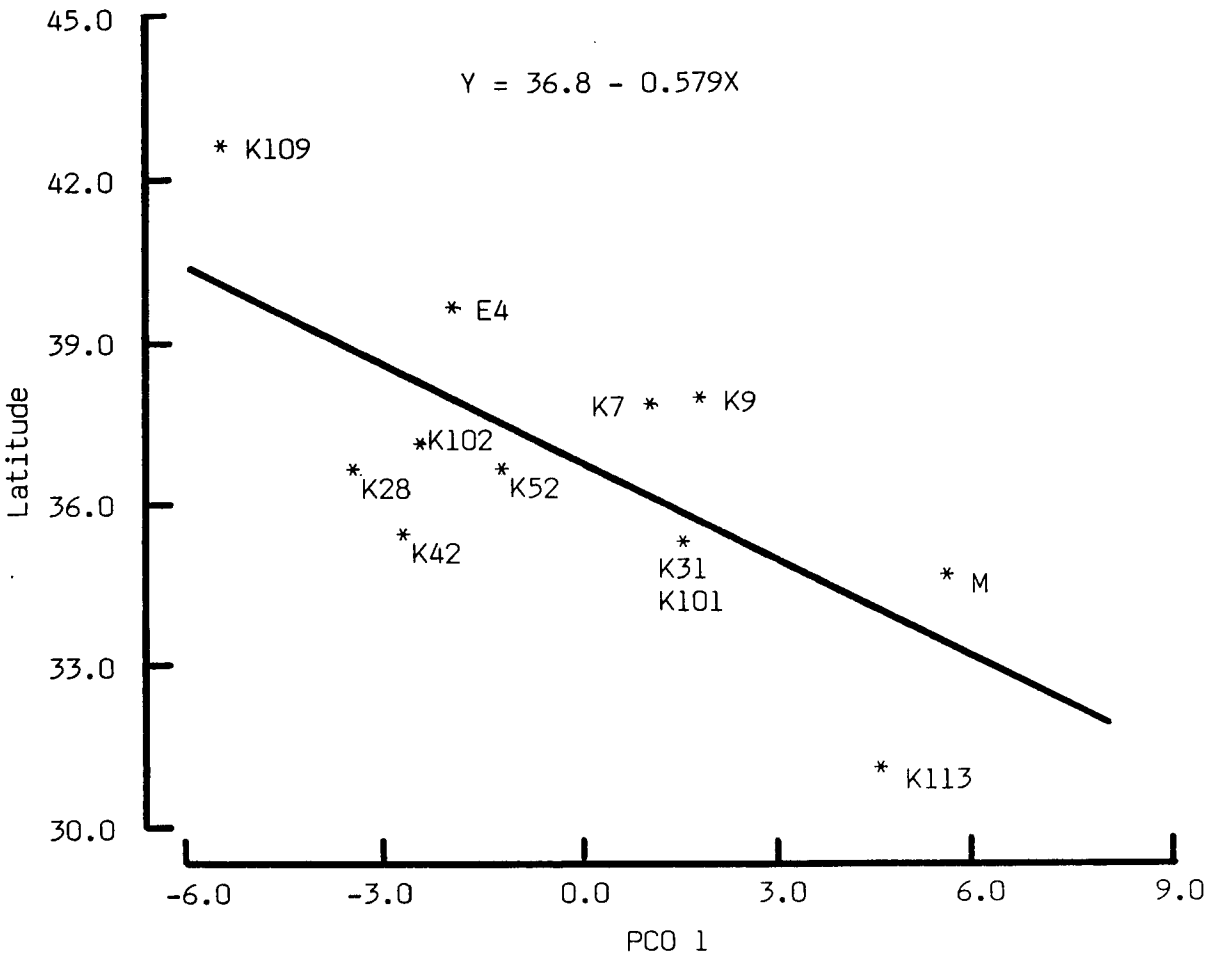


Figure 18. A graph of the first principal component (PCO 1) (based on allele frequencies) against latitudinal location of the 12 G. soja seed accessions.

sampled separately. Since these wild soybean populations were growing in 5 - 10 meters width along the river banks intermingling with other plant species, it was difficult to separate each individual plant in a colony. Twenty one to twenty five vines each with four or more mature pods from separate sites at least 1 - 2 meters apart each other were collected for each population. Each of these 21 to 25 vines was treated as a different plant in this study. The Kitakami river originates at the north of Morioka city, flows in the midst of Iwate Prefecture and terminates in nearby Ishinomaki city, Miyagi Prefecture (Figure 19). The river runs from north to south and connects many small rivers (the Shizukuishi river, Iwaigawa river and so on). The collection sites of these four populations from north to south were:

(a) The Gandai population (G) at latitude 39.96 N — 21 plants (G1 - G21) were collected at the Iwate University campus, Morioka. This place is a terrace land formed by the Kitakami river, 300 - 400 meters apart from the present stream.

(b) The Shizukuishi population (S) at latitude 39.70 N — 25 plants (S1 - S25) collected at the Shizukuishi river shore, Morioka, 3 kms from the Iwate Univ. This river connects to the Kitakami river at south of Morioka city.

(c) The Kitakami population (K) at latitude 39.30 N — 25 plants (K1 -K25) were collected at the Kitakami

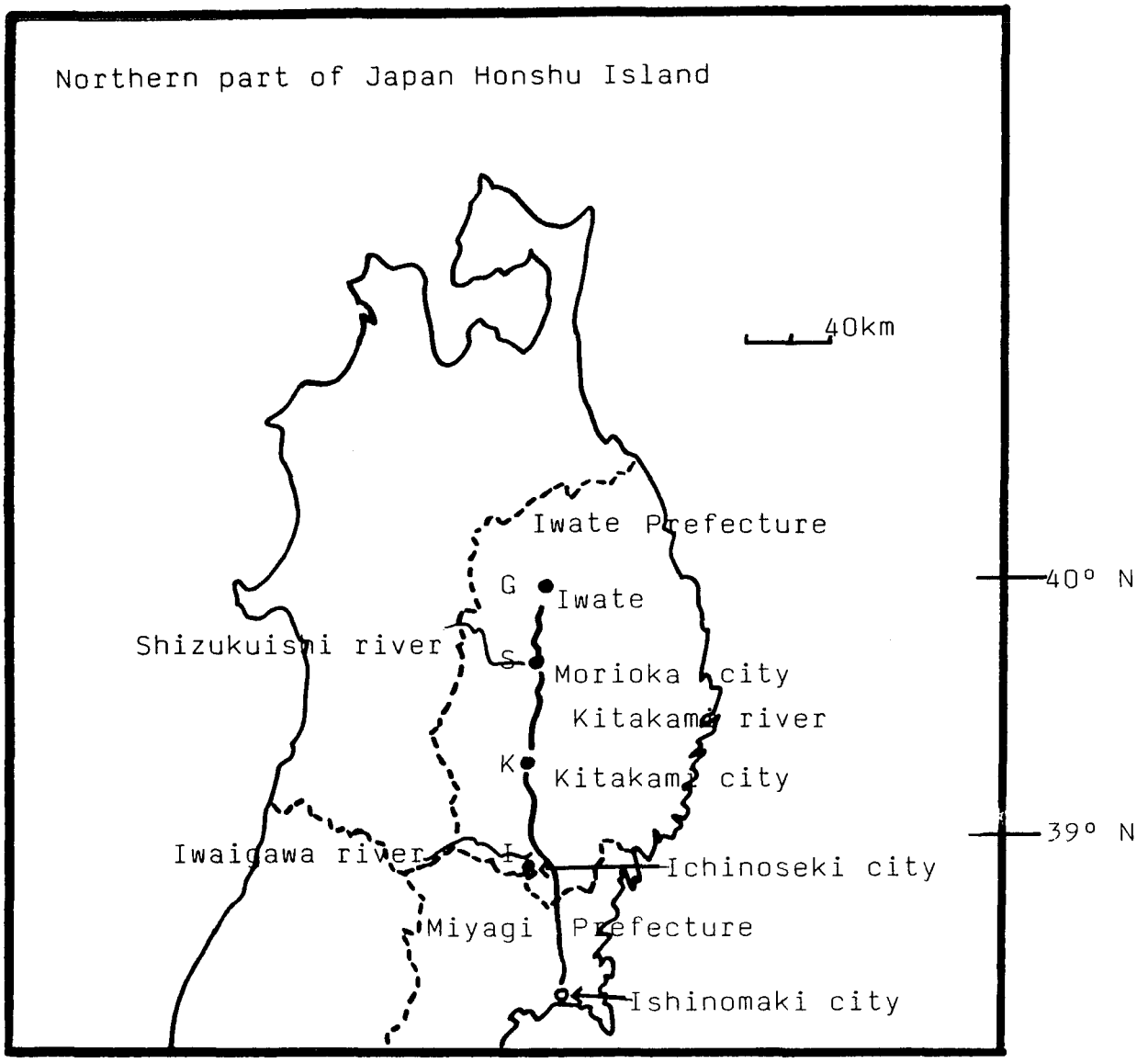


Fig 19 The collecting sites (●) of four populations (G, S, K and I) in northern part of Japan Honshu Island.

river shore in Kitakami city.

(d) The Iwaigawa population (I) at latitude 38.92 — 25 plants (I1 - I25) were collected at the Iwaigawa river shore in Ichinoseki city. Iwaigawa river connects to the Kitakami river at north of Ichinoseki city.

The total area distance between north G population to south I population is about 120 kms.

(A) Genetic diversity

Tables 24 and 25 are the lists of alleles observed in variable loci and their frequencies in these four natural populations of G. soja, based on the genotypes of plants sampled and seeds collected, respectively. A total of 39 alleles was observed in 16 variable loci of enzymes examined. The Pgi2 locus was excluded in Table 25 because of the dominant-recessive relationship between the Pgi2 and pgi2 alleles. The genotypes of seeds in this locus cannot be determined directly from the phenotypes observed. The allelic frequencies based on plants and their offsprings are similar (Tables 24 and 25). A number of significant differences in allele frequency among populations were seen at individual loci. However, no consistent north - south pattern of variation was observed.

Among the 16 variable enzyme loci, population G had six loci fixed for one allele, S had two, K had five, while population I was variant in all 16 loci. Consequently, popilation I had a higher number of alleles per locus,

Table 24. A list of alleles observed and their frequencies of the four natural G. soja populations along Kitakami river, Japan. (based on the genotypes of the plants that sampled)

	Population				
	G	S	K	I	Mean
Aco2-a	0.43	0.40	0	0.48	0.3275
Aco2-c	0.57	0.60	1.00	0.52	0.6725
Aco4-a	0.43	0	0.20	0.32	0.2375
Aco4-b	0.57	1.00	0.80	0.68	0.7625
Am3-s	1.00	0.66	0.96	0.68	0.8250
Am3-f	0	0.34	0.04	0.32	0.1750
Ap-a	0.48	0.80	0.40	0.04	0.4300
Ap-b	0.28	0	0	0.48	0.1900
Ap-c	0.24	0.20	0.60	0.48	0.3800
Dial-a	0.86	0.90	0.96	0.04	0.6900
Dial-b	0.14	0.10	0.04	0.96	0.3100
Dia2-a	0	0	0	0.28	0.0700
Dia2-b	1.00	1.00	1.00	0.72	0.9300
Enp-a	0.10	0.26	0.40	0.40	0.2900
Enp-b	0.61	0.74	0.24	0.60	0.5475
Enp-c	0.29	0	0.36	0	0.1625
Idh2-a	1.00	0.86	1.00	0.96	0.9550
Idh2-b	0	0	0	0.04	0.0100
Idh2-c	0	0.14	0	0	0.0350
Idh3-a	0	0	0	0.28	0.0700
Idh3-b	1.00	0.92	0.44	0.66	0.7550
Idh3-c	0	0.08	0.56	0.06	0.1750
Idh4-a	1.00	0.86	1.00	0.96	0.9550
Idh4-b	0	0.14	0	0.04	0.0450
Lapl-a	0	0.14	0.20	0.66	0.2500
Lapl-b	1.00	0.86	0.80	0.34	0.7500
Mpi-a	0	0	0.04	0	0.0100
Mpi-b	0.57	0.84	0.50	0.60	0.6275
Mpi-c	0.43	0.16	0.46	0.40	0.3625
Pgdl-a	0.14	0.18	0.04	0.34	0.1750
Pgdl-b	0.48	0.36	0.72	0.66	0.5550
Pgdl-c	0.10	0.24	0.24	0	0.1450
pgdl	0.28	0.22	0	0	0.1250
Pgi2	0.67	0.36	0.76	0.80	0.6475
pgi2	0.33	0.64	0.24	0.20	0.3525
Pgml-a	0.62	0.82	1.00	0.96	0.8500
Pgml-b	0.38	0.18	0	0.04	0.1500
Pgm2-a	0.86	0.92	0.78	0.56	0.7800
Pgm2-b	0.14	0.08	0.22	0.44	0.2200

Table 25. A list of alleles observed and their frequencies of the four natural G. soja populations along Kitakami river, Japan. (based on the genotypes of seeds collected)

	Population				
	G	S	K	I	Mean
Aco2-a	0.430	0.347	0	0.480	0.3143
Aco2-c	0.570	0.653	1.000	0.520	0.6856
Aco4-a	0.430	0	0.203	0.320	0.2383
Aco4-b	0.570	1.000	0.797	0.680	0.7617
Am3-s	1.000	0.656	0.962	0.680	0.8245
Am3-f	0	0.344	0.038	0.320	0.1755
Ap-a	0.480	0.807	0.377	0.040	0.4260
Ap-b	0.280	0	0	0.480	0.1900
Ap-c	0.240	0.193	0.623	0.480	0.3840
Dial-a	0.860	0.890	0.962	0.040	0.6880
Dial-b	0.140	0.110	0.038	0.960	0.3120
Dia2-a	0	0	0	0.280	0.0700
Dia2-b	1.00	1.00	1.00	0.720	0.9300
Enp-a	0.100	0.240	0.413	0.400	0.2883
Enp-b	0.610	0.760	0.227	0.600	0.5493
Enp-c	0.290	0	0.360	0	0.1625
Idh2-a	1.000	0.820	1.000	0.960	0.9450
Idh2-b	0	0	0	0.040	0.0100
Idh2-c	0	0.180	0	0	0.0450
Idh3-a	0	0	0	0.280	0.0700
Idh3-b	1.000	0.930	0.453	0.660	0.7608
Idh3-c	0	0.070	0.547	0.060	0.1693
Idh4-a	1.000	0.811	1.000	0.960	0.9428
Idh4-b	0	0.189	0	0.040	0.0572
Lapl-a	0	0.167	0.222	0.645	0.2585
Lapl-b	1.000	0.833	0.778	0.355	0.7415
Mpi-a	0	0	0.038	0	0.0095
Mpi-b	0.570	0.846	0.523	0.600	0.6348
Mpi-c	0.430	0.154	0.439	0.400	0.3558
Pgdl-a	0.140	0.171	0.038	0.335	0.1710
Pgdl-b	0.480	0.368	0.698	0.665	0.5528
Pgdl-c	0.100	0.215	0.264	0	0.1448
pgdl	0.280	0.246	0	0	0.1315
Pgml-a	0.620	0.781	1.000	0.960	0.8403
Pgml-b	0.380	0.219	0	0.040	0.1597
Pgm2-a	0.860	0.930	0.769	0.560	0.7798
Pgm2-b	0.140	0.070	0.231	0.440	0.2202

polymorphism (99% level) and expected heterozygosity than the other three populations (Table 26). There were five alleles, Ap-a, Dia1-a, Idh2-b, Idh4-b and Pgm1-b, each of them was shown only in one plant of population I. Among them the alleles Ap-a and Pgm1-b were exhibited by a single individual. These rare alleles resulted in a large reduction of polymorphism in population I at 95% level versus the 99% level (Table 26). The appearance of a unique individual in a predominantly selfing plant population, such as G. soja, is more likely through seed flow instead of pollen flow. The downstream position of population I may increase the gene flow by seed depositing from the upstream populations.

The overall number of alleles per locus for these four populations was 2.44 for variable loci, and 1.55 for all the loci examined. The number of alleles per locus for each population of G, S, K and I were 1.33, 1.38, 1.33 and 1.43, respectively (Table 26). They are not significantly different (F-test) and are within the range (1.00 - 3.75) reported by Hamrick (1979) based on 39 annuals. The low end of this range is a totally monomorphic population.

The overall polymorphism (99% level) estimated for the four populations was 38.1% (16 polymorphic loci / 42 loci examined). This value is higher than an average polymorphism of 23.1% reported by Ayala (1982) based on 12 self-pollinating plant species (an average of 15 loci per species was studied).

The overall expected heterozygosity of the four

Table 26. Genetic diversity of the four natural populations (G, S, K and I) along the Kitakami river, Japan.

	# of alleles		% polymorphism		Expected Heterozygosity	
	variable loci	total loci	99% level	95% level	Hexp total loci	Hexp vari. loci
G	1.88	1.33	24	24	0.112	0.293
S	2.00	1.38	33	33	0.110	0.285
K	1.88	1.33	26	21	0.101	0.255
I	2.13	1.43	38	29	0.138	0.361
pool population	2.44	1.55	38.1	38.1	0.114	0.299

populations was 0.114 (Table 26) which was within the range (0.000 - 0.414) that Hamrick (1979) compiled for 39 annual plants.

The expected (Hexp) and observed (Hobs) heterozygosities of the four natural populations are listed in Tables 27 and 28 based on the plants sampled and their offsprings collected, respectively. Although there are no significant difference among the four mean Hexps, indicating that the four populations are not statistically different in the amount of total genetic variation, they do show variation in Hexp of several loci (Tables 27 and 28). The most variable locus in population G was Pgd1, containing four alleles, with a Hexp = 0.662. In population S the most variable locus was also Pgd1, with four alleles and Hexp = 0.732. Locus Enp with three alleles showed the highest Hexp (0.653) among the 16 variable loci examined in population K. And locus Ap with three alleles and Hexp = 0.538 in population I (Table 27). The four populations were also different in the number of heterozygotes observed. The two middle populations (S and K) had higher Hobs than G and I which were the most northern and southern populations. The average of Hobs in these natural populations was 2.3% (Table 27), much higher than 0.06% and 0.11% reported by Chiang (1981) and Gorman (1983), respectively for cultivated soybeans. The remarkable decrease in Hobs from the theoretical Hexp is due to the predominantly self-fertilization of G. soja. The lesser amount of observed heterozygosity in the offspring

Table 27. The expected (Hexp) and observed (Hobs) heterozygosities of the four natural *G. soja* populations along the Kitakami river, japan*.

Variable loci	G			S			K			I		
	# of alleles	Hexp	Hobs	# of alleles	Hexp	Hobs	# of alleles	Hexp	Hobs	# of alleles	Hexp	Hobs
Aco2	2	0.490	0	2	0.480	0.080	1	0	0	2	0.499	0
Aco4	2	0.490	0	1	0	0	2	0.320	0.080	2	0.435	0
Am3	1	0	0	2	0.449	0.040	2	0.077	0	2	0.435	0
Ap	3	0.634	0	2	0.320	0.080	2	0.480	0	3	0.538	0
Dia1	2	0.241	0	2	0.180	0.040	2	0.077	0	2	0.077	0
Dia2	1	0	0	1	0	0	1	0	0	2	0.403	0
Enp	3	0.534	0	2	0.385	0.120	3	0.653	0.080	2	0.480	0
Idh2	1	0	0	2	0.241	0.040	1	0	0	2	0.077	0
Idh3	1	0	0	2	0.147	0	2	0.493	0.080	3	0.482	0.040
Idh4	1	0	0	2	0.241	0.040	1	0	0	2	0.077	0
Lap1	1	0	0	2	0.241	0.040	2	0.320	0.080	2	0.449	0.040
Mpi	2	0.490	0	2	0.241	0.080	3	0.537	0.200	2	0.480	0
Pgd1	4	0.662	0	4	0.732	0.080	3	0.422	0	2	0.449	0.040
Pgi2	2	0.442	0	2	0.461	0	2	0.365	0	2	0.320	0
Pgm1	2	0.471	0	2	0.295	0.040	1	0	0	2	0.077	0
Pgm2	2	0.241	0	2	0.147	0	2	0.343	0.120	2	0.493	0
Mean**	1.88	0.293	0	2.00	0.285	0.043	1.88	0.255	0.040	2.13	0.361	0.008

* The allele frequencies are based on the genotypes of the plants sampled

** The overall mean of Hexp = 0.299 (for variable loci); Hobs = 0.023 (for variable loci);

$\overline{Hexp}$ = 0.114 (for all loci examined)

Table 28. The expected (Hexp) and observed (Hobs) heterozygosities of the four natural G. soja populations along the Kitakami river, Japan. (Calculation based on the genotypes of seeds collected)

	G			S			K			I		
Variable loci	# of alleles	Hexp	Hobs	# of alleles	Hexp	Hobs	# of alleles	Hexp	Hobs	# of alleles	Hexp	Hobs
Aco2	2	0.490	0	2	0.453	0.028	1	0	0	2	0.499	0
Aco3	2	0.490	0	1	0	0	2	0.324	0.085	2	0.435	0
Am3	1	0	0	2	0.451	0.009	2	0.073	0	2	0.435	0
Ap	3	0.634	0	2	0.312	0.018	2	0.470	0	3	0.538	0
Dia1	2	0.241	0	2	0.196	0.009	2	0.073	0	2	0.077	0
Dia2	1	0	0	1	0	0	1	0	0	2	0.403	0
Enp	3	0.534	0	2	0.365	0.027	3	0.648	0.053	2	0.480	0
Idh2	1	0	0	2	0.295	0.026	1	0	0	2	0.077	0
Idh3	1	0	0	2	0.130	0	2	0.496	0.066	3	0.482	0
Idh4	1	0	0	2	0.307	0.026	1	0	0	2	0.077	0
Lap1	1	0	0	2	0.278	0.018	2	0.345	0.028	2	0.458	0.01
Mpi	2	0.490	0	2	0.261	0.018	3	0.532	0.198	2	0.480	0
Pgd1	4	0.662	0	4	0.729	0.009	3	0.442	0	2	0.446	0.01
Pgm1	2	0.471	0	2	0.342	0.026	1	0	0	2	0.077	0
Pgm2	2	0.241	0	2	0.130	0	2	0.355	0.047	2	0.493	0
Mean*	1.88	0.293	0	2.00	0.283	0.040	1.88	0.251	0.032	2.13	0.364	0.001

* The overall mean $\bar{H}exp = 0.298$; $\bar{H}obs = 0.012$

population (Table 28) than the parental population (Table 27) may also be the consequence of self-fertilization, or the post emergence. However, whether the 2.3% heterozygosity represents an average through years is unknown. It would require several yearly data collection to answer the question.

(B) Apportionmemt of gene diversity

The within population gene diversity (H_S) of the four G. soja natural populations varied from 0.08 in Idh2 and Idh4 loci to 0.566 in Pgd1, with an average of 0.299 for all the variable loci examined (Table 29). The mean gene diversity within populations in autogamous plant species was 0.128, compiled from 39 studies by Loveless and Hamrick (1984). Brown (1979) also reported an average of 0.126 (range 0.022 - 0.239) for H_S on a geographic scale for 13 inbred plant species, and a range from 0.15 to 0.29 on a microgeographic scale for four inbred plant species. These comparisons indicate that the four G. soja natural populations have high within population gene diversity as an inbreeder.

G_{ST} measures the proportion of variation among populations relative to the total species' diversity, or the proportion of variation among population subdivisions relative to the total population's diversity. The single locus estimates of G_{ST} among the four G. soja populations vary from 0.070 in Idh4 to 0.664 in Dial (Table 29). The overall loci G_{ST} among these four populations is 0.198,

Table 29. Apportionment of gene diversity within and among four natural populations (G, S, K and I) of G. soja.

	H_T	H_S	D_{ST}	G_{ST} *	H_T	H_S	D_{ST}	G_{ST} **
Aco2	0.440	0.367	0.073	0.166	0.431	0.361	0.070	0.162
Aco4	0.362	0.311	0.051	0.141	0.363	0.312	0.051	0.140
Am3	0.289	0.240	0.049	0.171	0.289	0.240	0.049	0.171
Ap	0.635	0.493	0.142	0.224	0.635	0.489	0.146	0.230
Dia1	0.428	0.144	0.284	0.664	0.429	0.147	0.282	0.658
Dia2	0.130	0.101	0.029	0.223	0.130	0.101	0.029	0.225
Enp	0.590	0.513	0.077	0.131	0.589	0.507	0.082	0.139
Idh2	0.087	0.080	0.007	0.080	0.105	0.093	0.012	0.113
Idh3	0.394	0.281	0.113	0.287	0.388	0.277	0.111	0.285
Idh4	0.086	0.080	0.006	0.070	0.108	0.096	0.012	0.110
Lap1	0.375	0.253	0.122	0.325	0.383	0.270	0.113	0.296
Mpi	0.475	0.437	0.038	0.080	0.470	0.441	0.029	0.062
Pgd1	0.625	0.566	0.059	0.094	0.627	0.570	0.057	0.091
Pgi2	0.456	0.397	0.059	0.129	---	---	---	---***
Pgm1	0.255	0.211	0.044	0.173	0.268	0.223	0.045	0.169
Pgm2	0.343	0.306	0.037	0.108	0.343	0.305	0.038	0.112

* based on the genotypes of the plant sampled

** based on the genotypes of seeds collected

*** genotype can not be determined directly from seed examined because of the dominant - recessive relationship of the two alleles in this locus.

indicating that 19.8% of the total genetic variation occurs among populations. This G_{ST} value is in the range (0.029 - 0.372, with average 0.220) that was reviewed by Loveless and Hamrick (1984) for G_{ST} among population subdivisions of four autogamous species. However, when comparing with outbreeding species which have a mean G_{ST} of 0.071 among populations and 0.041 among population subdivisions (Loveless and Hamrick, 1984), these four G. soja populations have a relatively high G_{ST} value. It was argued by Loveless and Hamrick (1984) that characters such as outbreeding promote gene flow and increase effective population size (Ne). Gene flow in turn decreases population differentiation. Thus, outbreeders are expected to have a smaller G_{ST} than inbreeders which have restricted gene migration.

Qualitatively speaking, according to Hartl (1980), the range of 0.05 to 0.15 for F_{ST}. a similar statistic to G_{ST}, may be considered to indicate moderate differentiation, 0.15 to 0.25 to indicate great differentiation, and above 0.25 to indicate very great differentiation. Thus, based on Hartl's criteria these four natural G. soja populations can be said to be well differentiated isoenzymatically.

(C) Genetic distance and identity

The observed values of genetic distance (D_N) and identity (I_N) of the four natural populations are listed in Table 30. They are similar to the average (D_N 0.035; I_N 0.966) calculated by Ayala and Kiger (1980) for local populations of plants.

Table 30. Nei's measures of genetic distance (D_N) and identity (I_N) between four populations of G. soja along Kitakami river, Japan.

		Population			
		G	S	K	I
D_N	G		0.980	0.972	0.947
	S	0.020 (0.26)*		0.964	0.938
	K	0.029 (0.66)	0.036 (0.40)		0.944
	I	0.054 (1.04)	0.064 (0.78)	0.058 (0.38)	

I_N

* Latitudinal difference between populations

Table 31. The mean genetic distance ($\bar{D}_N$) and ($\bar{I}_N$) of the four populations of G. soja along Kitakami river, Japan.

	Population				
	G	S	K	I	average
$\bar{D}_N$	0.034	0.040	0.041	0.059	0.044
$\bar{I}_N$	0.966	0.961	0.960	0.943	0.958

The most southern population (I) had consistently high D_N values with the other three populations (Table 30). This can also be seen by comparing the mean D_N of these four populations (Table 31). The population I had the highest mean genetic distance ($\overline{D}_N$). Since these four populations are located along Kitakami river which flows almost in a straight line from north to south on 141.10E longitudinally, the corresponding latitudinal range of each population was thus taken to calculate the relative geographic distance between populations. The correlation coefficient between D_N and this geographic distance is 0.529, which is not significantly different from zero. However, a pattern of increasing D_N corresponding with increasing geographic distance was found between population G and the other three populations as well as in population S and the other three populations (Table 30).

Principal component analysis based on protein allele frequencies is further used to assess relationship among these four populations. The first component (accounting for 44.7% of the total genetic variation) divided the four populations into two groups: G, S and K, I. The second component (accounting for 34.8% of the total genetic variation) separated S, I from G, K(Figure 20). Consequently, these four natural populations were clearly separated by the first two principal components of protein variation observed. When ordination was plotted with first principal component and latitude of these four populations, a correlation coefficient $r = -0.841$ was obtained (Figure 21). The correlation

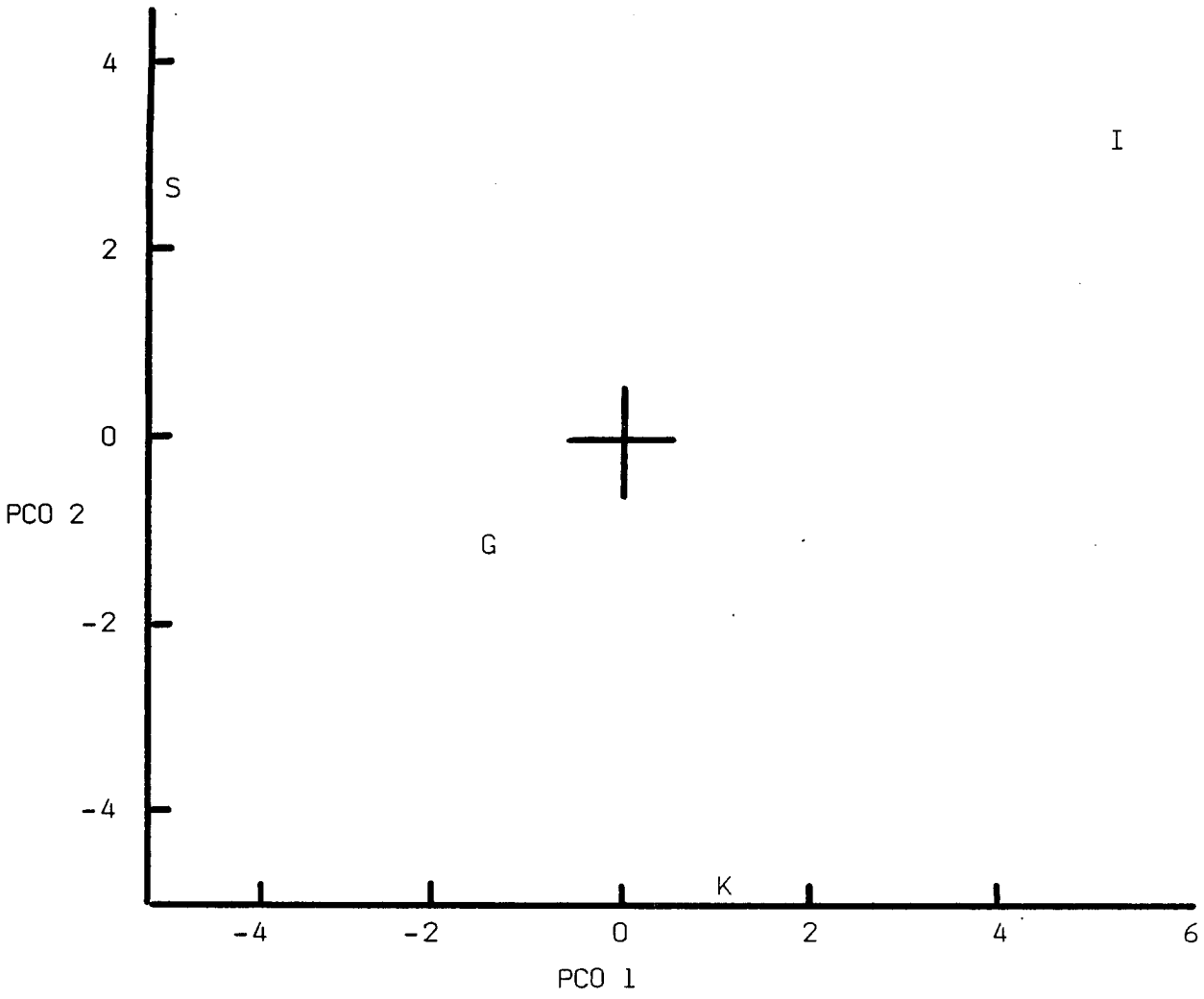


Figure 20. Ordination of the four natural G. soja populations (G, S, K and I) against the first (PCO 1) and second (PCO 2) principal components of protein gnentic variation.

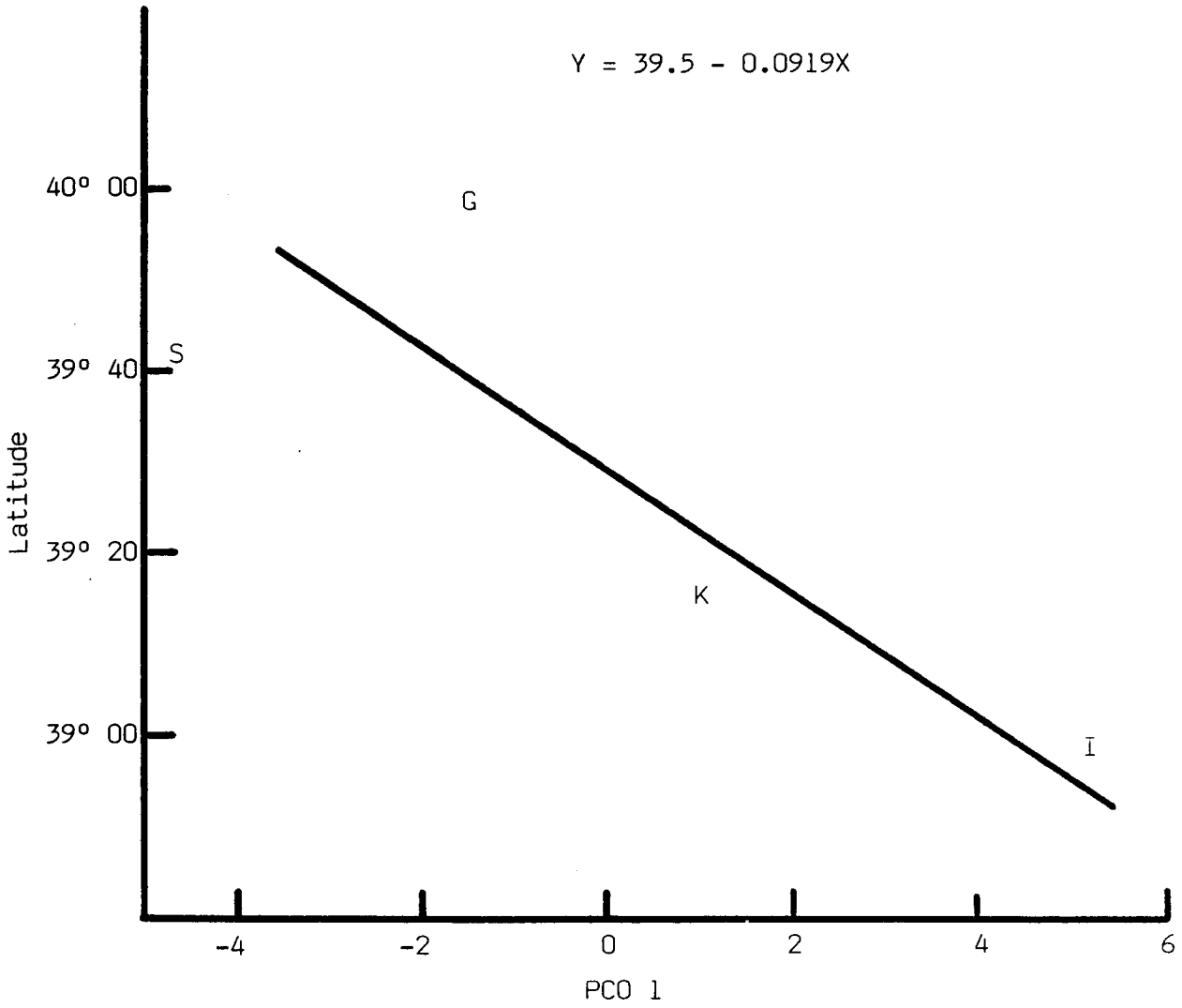


Figure 21. Ordination of the four natural populations (G, S, K and I) of G. soja against the latitude and the first principal component of protein genetic variation.

coefficient was r = -0.187 between the second principal component of protein variation and latitude. Although both coefficients are not significantly different from zero, the value of -0.841 is relatively high. The insignificance may be due to the small sample size. Alleles which contribute the most to the first principal component of protein variation are Ap-a, Pgm2-b, Pgm2-c, Lap1-a, Lap1-b, Pgi1 and pgi1. The results reveal that about 45% of the protein variation among the four populations may be accounted for by latitude or spatial distance represented by the relative latitudinal scales.

4. QUANTITATIVE VARIATION OF G. SOJA POPULATIONS

(1). The twelve seed accessions selected based on their latitudinal locations.

(A) Phenological data collected in 1982 and 1983 from the greenhouse experiments are presented in Table 32. The data include the number of days from sowing to seed germination, to the date of first flower appearance, to first fresh pod set, to first dry pod formation and to last dry pod harvested. The life span was obtained as days between germination and the end of harvesting. Each phenological trait measured was significantly different among these 12 seed accessions. A significant correlation was found between the number of days from germination to the first flower and latitude (r = -0.955 in 1982; r = -0.917

Table 32. The means (number of days) and standard deviations (numbers in parentheses) of phenological data of 12 G. soja seed accessions. Data collected from the greenhouse in 1982 and 1983.

	sowed - germinated		germinated - 1st flower		1st flower - 1st pod	
	1982	1983	1982	1983	1982	1983
K109	6.95 (1.32)	3.88 (0.34)	62.8 (1.94)	53.1 (2.36)	13.9 (2.36)	19.0 (2.31)
E4	6.91 (0.44)	4.53 (0.82)	66.9 (3.19)	61.8 (5.09)	16.3 (3.96)	11.8 (4.03)
K9	7.19 (0.87)	4.26 (0.44)	85.9 (3.02)	83.6 (3.92)	11.9 (2.23)	10.6 (3.16)
K7	6.78 (0.90)	4.67 (0.83)	86.0 (3.81)	86.9 (3.68)	10.2 (1.86)	8.9 (2.14)
K102	6.60 (1.00)	4.32 (0.72)	81.9 (6.10)	76.8 (3.62)	11.3 (2.46)	16.3 (3.63)
K28	7.38 (0.92)	4.63 (0.49)	88.5 (1.03)	85.0 (2.50)	10.3 (1.58)	10.4 (0.96)
K52	6.76 (1.67)	4.64 (0.79)	80.5 (4.16)	78.0 (5.25)	11.6 (1.86)	14.9 (5.73)
K42	6.18 (0.41)	4.28 (0.46)	92.5 (2.02)	83.8 (3.21)	8.4 (1.01)	9.8 (1.98)
K101	7.67 (1.20)	5.00 (0.74)	100.7 (3.05)	100.9 (0.93)	7.2 (2.48)	9.2 (1.60)
K31	7.05 (1.28)	4.91 (0.84)	96.4 (4.41)	89.5 (2.85)	8.8 (1.37)	10.6 (2.28)
M	7.25 (0.64)	5.17 (0.75)	100.9 (3.24)	109.4 (2.50)	7.4 (1.84)	6.4 (2.22)
K113	7.24 (1.04)	4.93 (0.83)	119.1 (2.67)	116.8 (1.91)	7.9 (1.53)	7.3 (2.24)
F test	2.47*	8.31**	64.2**	454.3**	23.7**	25.4**

* significant at 5% level

** significant at 1% level

Table 32. (continued)

	1st pod set - 1st pod dry		1st pod dry - last pod dry		life span		anthesis - seed matured
	1982	1983	1982	1983	1982	1983	
K109	37.9 (2.33)	38.9 (1.73)	16.3 (2.87)	33.0 (10.92)	131.0	144.0	41.4 (5.3)
E4	37.8 (3.49)	40.8 (3.73)	16.3 (3.67)	38.1 (4.02)	137.3	152.5	44.2 (4.7)
K9	29.2 (1.72)	33.2 (2.01)	20.9 (2.09)	26.9 (1.91)	147.9	154.3	41.2 (3.6)
K7	30.4 (1.89)	34.7 (3.40)	19.4 (2.24)	23.7 (2.19)	146.0	154.2	43.5 (1.3)
K102	29.3 (1.44)	35.8 (2.75)	20.1 (2.76)	25.2 (2.29)	142.6	154.1	40.1 (4.3)
K28	29.7 (1.96)	33.4 (2.10)	19.5 (1.86)	25.4 (2.06)	148.0	154.2	42.0 (3.6)
K52	30.1 (1.31)	34.3 (4.76)	21.6 (1.93)	25.6 (1.63)	143.8	152.8	40.6 (1.1)
K42	27.9 (1.97)	33.0 (2.00)	18.5 (5.87)	26.4 (3.20)	147.3	153.0	39.5 (3.1)
K101	31.7 (2.21)	41.4 (1.59)	21.6 (1.93)	12.0 (1.90)	161.2	163.5	40.4 (2.3)
K31	29.3 (1.39)	39.6 (3.35)	18.7 (2.44)	22.1 (2.91)	153.2	161.8	41.5 (1.9)
M	31.5 (1.89)	40.6 (1.82)	20.2 (1.52)	18.8 (2.30)	160.0	175.2	42.6 (1.0)
K113	27.8 (1.33)	35.4 (1.71)	19.9 (2.83)	19.8 (2.20)	174.7	179.3	41.0 (0.9)
F test	43.6**	22.3**	5.78**	38.1**	34.3**	24.5**	2.70**

in 1983). The life span and the latitude was also highly correlated (r = -0.925 in 1982; r = -0.842 in 1983). The more northern the origin of the seed accession, the earlier the first flower appeared, and the shorter the life span. The high correlation between life span and latitudinal habitat was also reported in the wild soybean strains of Siberia, Northeastern China, South Korea and Japan, and the length of the frost free period of the habitat was suggested to be the decisive factor (Fukui et al., 1978). Earlier flowering and shorter life span are most likely the results of the response of plant populations to the short growing season of their habitats.

An additional significant correlation was observed between latitude and the number of days between first flower appeared and first fresh pod set (r = 0.802 in 1982; r = 0.724 in 1983). A lot of flowers aborted before pod set was observed in those northern accessions, especially in K109 and E4. In K109, the duration between flowering and the first pod set was 45 days when seeds were sowed on 4/25/83; 19 days when sowed on 5/24/83; and 8 days when sowed on 6/20/83. The dates of first pod set of K109 on these three sowing dates were around 7/23/83, 8/8/83 and 8/25/83, respectively. That is, K109 started to grow pods no earlier than late July in 1983, no matter how early it started to flower. The cause of long duration of flower abortion is unknown. The environmental factors, such as quantity and quality of sun light, temperature etc., may

have some effect.

At least 10 flowers per accession were tagged to obtain the number of days between anthesis and seed maturity. Although the duration from anthesis to seed maturity was different statistically among these 12 seed accessions (Table 32), no clearcut north-south pattern could be followed. The difference between E4 and K42 was about 4.7 days, the largest among the 12 accessions. Thus, in general, the difference in the length of the seed filling period was insignificant among the 12 accessions. The main difference in life span is mostly attributable to the duration from germination to prior to flowering.

The correlation between latitude and the first principal component which accounted for 59.6% of the total phenological variation among the 12 accessions, is $r = -0.912$ (Figure 22). It is significantly different from zero at the 1% level. The correlation between the latitude and the second principal component is $r = 0.171$ (not significant). The result reveals that around 60% of the phenological variation among these 12 accessions in highly associated with their latitudinal locations.

(B) Agronomic data of greenhouse experiments in 1982 and 1983 are presented in Table 33. The differences among the 12 seed accessions for each of the 11 items were statistically significant (F-test).

The yield of 1983 was about two to three times higher than that of 1982 due to a large increment in the number of

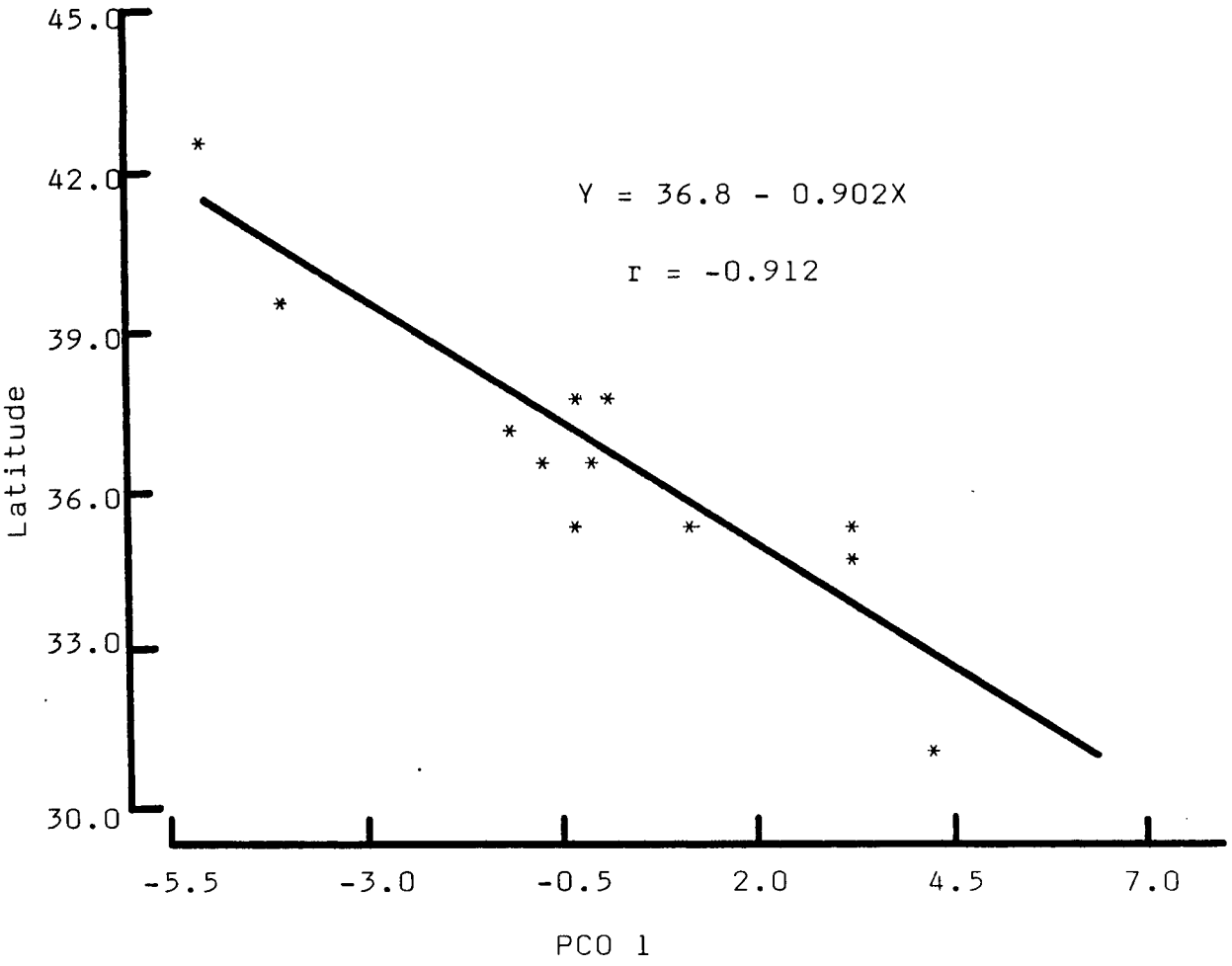


Figure 22. Plot of latitudinal location on the first principal component (PCO 1) axis which accounts for 59.6% of phenological variation among the 12 G. soja seed accessions.

Table 33. Agronomic data of 12 G. soja accessions in 1982 and 1983 greenhouse experiments (The number in parenthesis is the standard deviation of the accession mean).

	Av. # seed/pod		100 seed wt. (g)		Av. # pod/plant		Yield/plant(g)		Harvest index		reg. coeff. b*	
	1982	1983	1982	1983	1982	1983	1982	1983	1982	1983	1982	1983
K109	2.71 (0.13)	2.72 (0.15)	2.82 (0.21)	3.03 (0.12)	70.9 (24.97)	214.8 (32.73)	5.35 (1.77)	17.68 (2.96)	0.393 (0.024)	0.394 (0.017)	0.075	0.048
E4	2.47 (0.11)	2.55 (0.09)	1.84 (0.11)	2.12 (0.11)	126.3 (31.11)	326.3 (72.39)	5.75 (1.49)	17.65 (4.08)	0.354 (0.023)	0.359 (0.010)	0.074	0.041
K9	2.38 (0.16)	2.45 (0.11)	1.93 (0.15)	2.09 (0.14)	204.2 (45.62)	457.8 (47.71)	9.21 (1.78)	23.35 (2.18)	0.384 (0.017)	0.401 (0.022)	0.073	0.055
K7	2.35 (0.13)	2.45 (0.10)	1.95 (0.17)	2.22 (0.24)	198.1 (34.39)	397.1 (48.01)	9.00 (1.33)	21.61 (3.90)	0.369 (0.025)	0.372 (0.025)	0.074	0.058
K102	2.31 (0.12)	2.46 (0.08)	2.17 (0.23)	2.26 (0.28)	150.6 (45.47)	244.9 (31.15)	7.44 (1.93)	13.66 (2.82)	0.369 (0.031)	0.318 (0.036)	0.067	0.050
K28	2.41 (0.09)	2.44 (0.11)	1.84 (0.26)	2.27 (0.21)	204.7 (37.47)	436.6 (64.07)	9.10 (1.55)	24.09 (3.58)	0.361 (0.014)	0.386 (0.015)	0.082	0.069
K52	1.97 (0.27)	2.10 (0.24)	2.32 (0.17)	2.34 (0.21)	196.1 (42.97)	442.7 (88.07)	8.80 (1.81)	20.14 (5.07)	0.347 (0.014)	0.356 (0.016)	0.073	0.053
K42	2.24 (0.24)	2.47 (0.12)	2.18 (0.21)	2.29 (0.13)	159.4 (36.50)	386.4 (58.40)	7.77 (2.07)	21.85 (3.70)	0.386 (0.023)	0.382 (0.016)	0.082	0.065
K101	2.25 (0.10)	2.07 (0.11)	1.60 (0.11)	1.64 (0.11)	256.0 (53.73)	569.4 (71.60)	9.20 (2.07)	19.14 (2.13)	0.330 (0.028)	0.285 (0.016)	0.076	0.117
K31	2.09 (0.19)	2.00 (0.12)	1.85 (0.08)	2.29 (0.25)	252.5 (60.55)	503.9 (84.09)	9.67 (2.02)	23.12 (4.61)	0.403 (0.022)	0.374 (0.028)	0.064	0.053
M	2.62 (0.05)	2.22 (0.08)	2.31 (0.10)	2.34 (0.13)	228.3 (36.93)	436.1 (57.96)	13.89 (2.43)	22.64 (2.66)	0.361 (0.015)	0.285 (0.014)	0.066	0.071
K113	2.14 (0.22)	2.02 (0.10)	2.53 (0.13)	2.50 (0.16)	215.0 (48.41)	365.3 (48.66)	11.08 (2.19)	18.36 (2.14)	0.308 (0.017)	0.234 (0.020)	0.070	0.071
F test	27.4**	3.6**	13.8**	47.1**	23.4**	43.6**	21.7**	12.7**	21.8**	81.2**	0.830	9.14**

* regression coeff. b of % accumulated number of dry pod vs. two days harvesting interval

** statistically significant at 1% level

Table 33. (continued)

	Seed packing (%/total pod)				#nodule/plant***		total dry wt(g)***		root/total dry wt***		height(cm)§	
	1s/pod	2s/pod	3s/pod	4s/pod	1982	1983	1982	1983	1982	1983	1982	1983
K109	5.3 (3.80)	24.5 (7.62)	64.2 (8.06)	5.9 (4.6)	79.6 (61.4)	33.0 (14.3)	6.1 (2.34)	11.5 (2.72)	0.307 (0.022)	0.192 (0.060)	12.49 (5.55)	41.85 (16.34)
E4	12.2 (4.18)	29.1 (6.79)	58.2 (7.68)	0.5 (0.64)	24.2 (31.4)	34.5 (25.1)	6.6 (3.38)	12.8 (3.04)	0.226 (0.032)	0.191 (0.113)	24.02 (11.22)	47.63 (25.51)
K9	15.3 (7.22)	31.7 (5.88)	52.7 (9.37)	0.3 (0.45)	16.2 (10.3)	15.8 (17.0)	7.5 (3.72)	14.0 (2.68)	0.286 (0.044)	0.185 (0.042)	8.86 (2.97)	26.18 (4.36)
K7	10.7 (4.72)	43.5 (5.74)	45.8 (8.83)	0.1 (0.20)	23.2 (10.7)	13.5 (4.4)	9.6 (3.30)	15.9 (3.53)	0.307 (0.027)	0.232 (0.037)	6.98 (4.34)	17.13 (4.89)
K102	15.0 (5.52)	39.7 (4.96)	45.2 (7.75)	0.2 (0.42)	15.8 (30.2)	0	8.4 (3.29)	12.4 (2.25)	0.318 (0.025)	0.203 (0.061)	6.61 (1.90)	17.65 (7.22)
K28	11.2 (4.47)	37.0 (4.34)	50.7 (5.64)	0.9 (1.06)	34.4 (22.9)	26.0 (16.2)	7.0 (2.86)	11.9 (5.10)	0.307 (0.050)	0.219 (0.048)	5.08 (0.75)	9.90 (0.90)
K52	30.2 (13.5)	42.9 (5.48)	26.9 (13.7)	0	12.0 (9.67)	4.8 (2.1)	6.1 (2.79)	10.3 (1.54)	0.346 (0.048)	0.289 (0.059)	3.98 (0.85)	7.78 (2.30)
K42	17.8 (10.9)	40.7 (4.42)	41.3 (12.6)	0	----	40.5 (13.4)	---	12.2 (3.22)	---	0.267 (0.039)	8.18 (4.17)	29.3 (10.0)
K101	16.5 (3.84)	41.8 (4.45)	41.6 (7.00)	0.1 (0.19)	16.4 (8.20)	14.5 (5.7)	6.3 (3.64)	8.2 (2.01)	0.292 (0.068)	0.189 (0.032)	4.95 (1.20)	17.85 (8.95)
K31	25.2 (9.87)	40.8 (4.52)	33.7 (9.41)	0.2 (0.62)	36.6 (37.7)	21.0 (9.8)	5.6 (2.66)	9.3 (1.56)	0.336 (0.086)	0.195 (0.072)	4.06 (0.84)	7.08 (0.99)
M	10.0 (2.02)	23.3 (2.47)	61.1 (2.20)	3.7 (1.16)	107.2 (57.2)	16.3 (11.8)	13.5 (5.17)	9.7 (3.07)	0.287 (0.022)	0.243 (0.081)	6.68 (1.24)	27.18 (4.28)
K113	25.9 (9.28)	37.0 (5.72)	35.0 (11.0)	2.0 (1.30)	16.7 (10.9)	16.5 (7.9)	6.6 (3.04)	13.0 (1.4)	0.262 (0.058)	0.168 (0.054)	4.36 (0.69)	22.48 (11.2)
F test	15.9**	25.4**	25.2**	29.7**	4.5**	3.6**	2.25**	2.28**	2.44**	1.411	36.8**	5.76**

*** 10 weeks after sown

§ Plants were 3-week old in 1982 and 4-week old in 1983

pod per plant. Among the three yield components (#seed/pod, 100 seed weight and #pod/plant) only the number of pod per plant was found to be significantly positively correlated with the total yield per plant. The other two components were weakly and negatively correlated with the total yield. Thus, increase in flower number and pod set per plant would be the most efficient way for yield improvement of wild soybeans.

High and relatively synchronous seed setting is a desired agronomic character for seed crops. However, a tight synchrony in flowering and seed setting for a plant in natural condition may have effects on the potential of outcrossing and the ability to avoid serious damage by seed and fruit predators (Augsurger, 1981). The linear regression coefficient 'b' (Table 33) was obtained by linear regression of the cumulative number of matured pods in percentage to two-day harvesting interval. The larger the 'b' value is, the more synchronous seed set of a plant. The values of 'b' among the 12 accessions were not statistically different in 1982, but significantly different in 1983 with K101 having a very steep 'b', and K109 and E4 having average low 'b' values. Also the 'b' values between two years were significantly different. A difference in the length of suitable flowering season, photoperiod as well as greenhouse managements, such as fertilizing, watering could have contributed to this yearly variation.

The 12 seed accessions were very different in the way

of seed packing. For example, K52 had 42.9% of total pods with 2 seed, 30.2% with 1 seed and 26.9% with 3 seed; while K109 had 64.2% of total pods with 3 seed, 24.5% with 2 seed, 5.9% with 4 seed and 5.3% with 1 seed (Table 33). The differences are further demonstrated in Figure 23. The seed-packing patterns of K109 and M were similar. So were K7 and K102, K52 and K31, and K101 and K42. In order to see whether the difference in seed-packing pattern resulted from the difference in the number of ovules per ovary, 36 flower buds per seed accession were sampled randomly from four plants, and examined under a dissecting microscope. Most of the flower buds examined consisted of three ovules. An average of three ovules per ovary was observed in all 12 accessions. Since there is no difference in the number of ovules per ovary among these 12 accessions, the difference in seed packing was due to ovule and seed abortion. The different seed packing patterns may have resulted from ecological differentiation, such as resource limitation, unpredictable seasonal changes, seed and pod predators, etc. More information on the ecological factors of the populations from which these seed were collected are needed for understanding the observed seed packing patterns.

Among many of the quantitative characters studied we found some interesting relationships (Table 34). The average number of nodules per plant positively correlated with the percents of 3-seed pods and 4-seed pods, but negatively with 2-seed pods and 1-seed pods (Table 34).

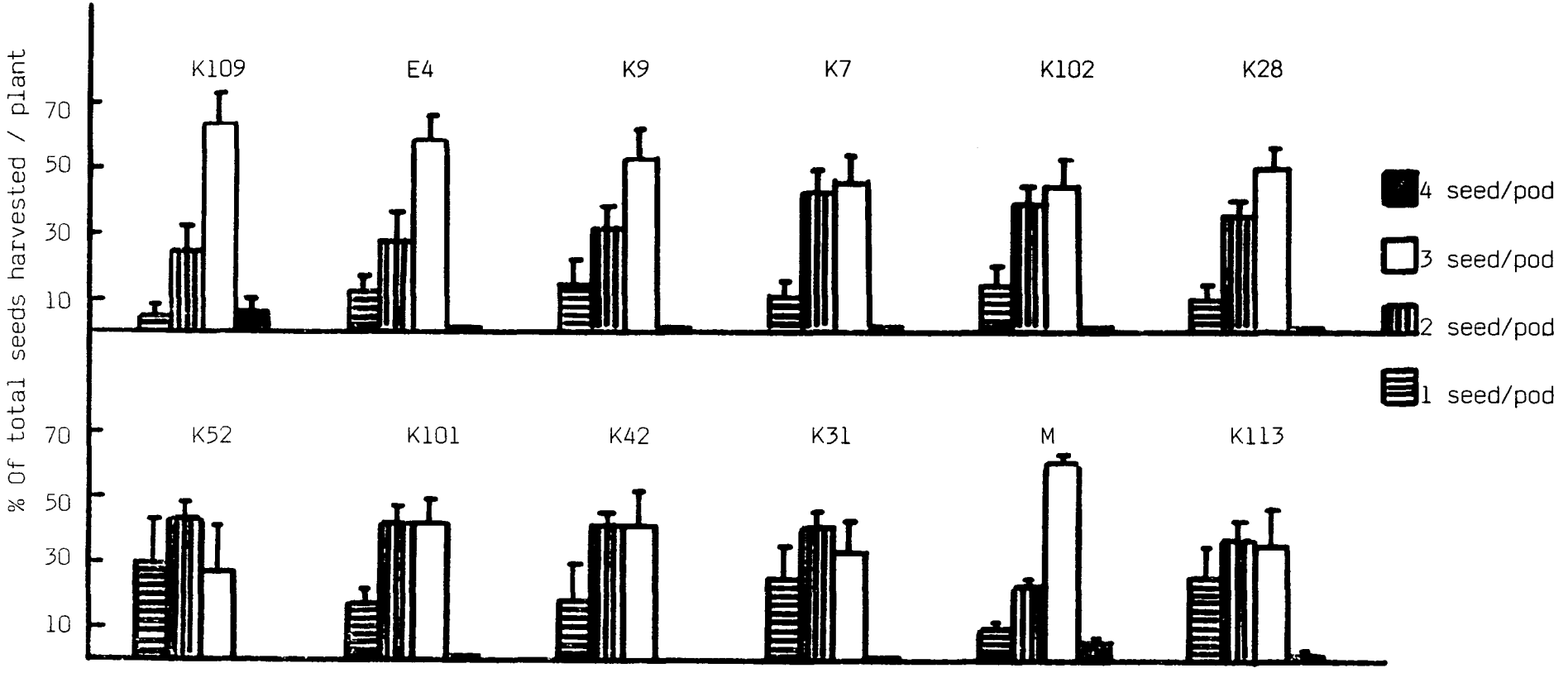


Figure 23. Seed packings of 12 seed accessions in 1982 greenhouse experiment.

Table 34. Correlation coefficients between several quantitative characters studied of 12 G. soja accessions.

	percents of			
	1-seed pod	2-seed pod	3-seed pod	4-seed pod
Av. #nodule/plant	-0.560	-0.745**	0.666*	0.827**
width banner petal	-0.565	-0.776**	0.690*	0.825**
flower length	-0.648*	-0.620*	0.700*	0.557
flower tube length	-0.510	-0.594	0.630*	0.362

* significant at 5% level

** significant at 1% level

That is, among these 12 accessions, the higher the percents of 3-seed pod or 4-seed pod an accession has, the more number of nodules per plant it has. It seems to suggest that nitrogen nutrition is important for increasing the number of seed per pod in wild soybean. Another interesting observation is the positive correlation between the flower size and the percents of 3-seed pod and 4-seed pod, and the negative correlation between flower size and the percents of 2-seed pod and 1-seed pod. If the size of the flower parts is proportional to the flower size, then the larger flower will have a larger ovary. And the larger ovary in turn may have more room or nutrition supply for seed develoment, and reduce seed abortion. The observation that the positive correlation between the flower tube length and the percent of the 3-seed pod seems to support the idea.

Many variations of agronomic characters studied among the 12 G. soja accessions were statistically associated with the latitude. For instance, in 1982, the more northern an accession origins, the less the yield per plant ($r = -0.738$), and the number of pods per plant ($r = -0.744$), but more seed per pod ($r = 0.595$). In 1983, significant correlations of latitude with the average number of seed per pod ($r = 0.799$), and with the harvest index ($r = 0.695$) were observed. The plot of 12 accessions with their respective latitudinal locations and the first principal component (accounting for 43.1% of the total variation) of agronomic variation is presented in Figure 24. Correlation between two axes is

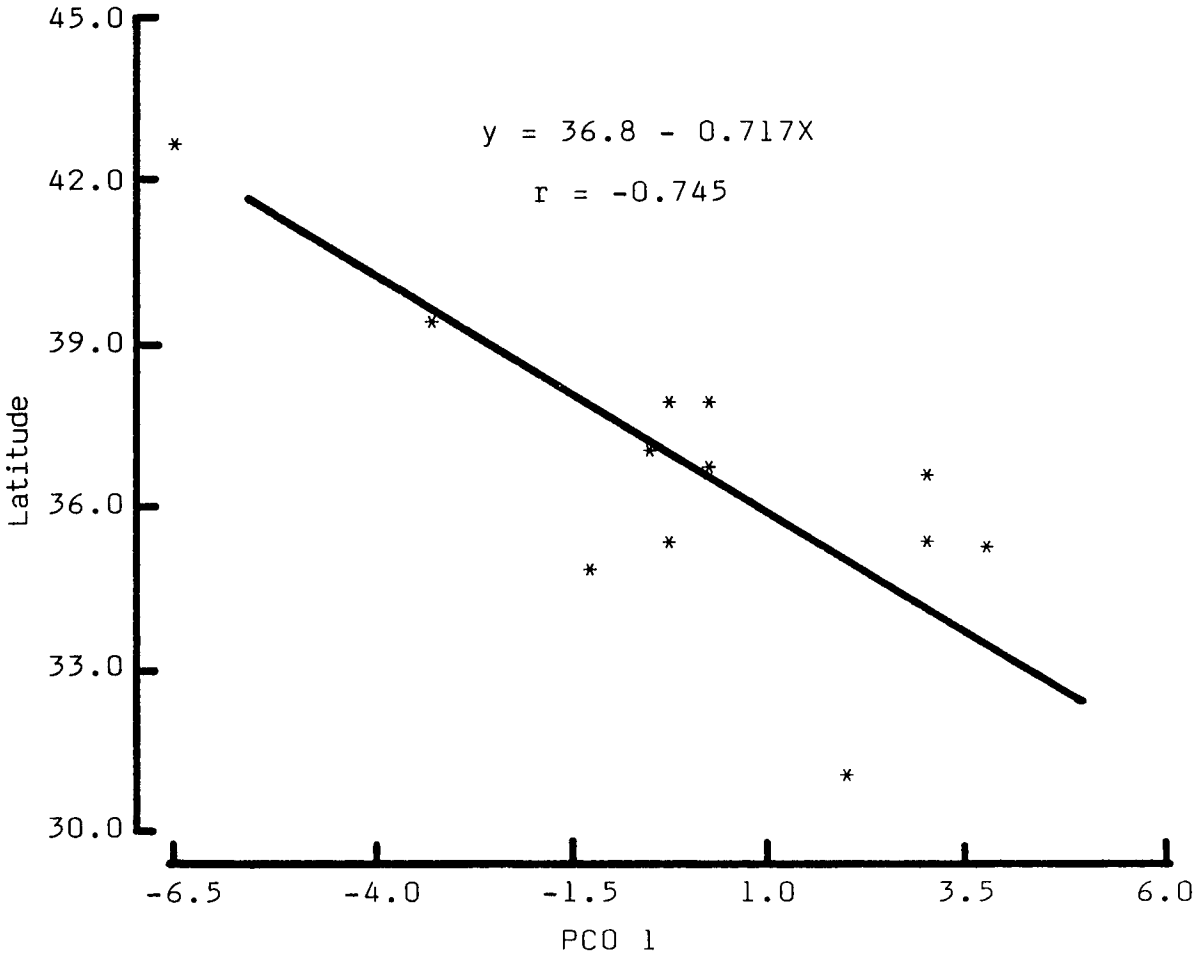


Figure 24. A plot of 12 G. soja accessions with their latitudinal locations against the first principal component (PCO 1) of the agronomic variation. The PCO 1 accounts for 43.1% of the total agronomic variation observed.

significant (r = -0.745). The correlation coefficient r = -0.326 is between latitude and the second principal component which accounted for 17.6% of the total agronomic variation observed.

(C) Morphological data of greenhouse experiments are presented in Table 35. These data included the number of branches, stem length from ground surface to the first node, flower size, 3-seed-pod size, length and angle of pubescence on pod, length, angle and density of pubescences on leaf surface, leaflet size (length and width of the widest area) and length/width ratio of the leaflet. Besides the angles of pubescence to pod and to leaf surfaces, and the length of pubescences on leaf, the morphological measurements listed in Table 35 are statistically different among the 12 seed accessions.

The principal component analysis was performed based on the 18 morphological measurements (Table 35) and the luster of seed coat. The luster of seed coat of E4 is shiny and the rest of the 11 accessions are dull. The first two principal components accounted for 33.1% and 26.4%, respectively, of the total morphological variation observed. Both components were not significantly correlated with latitude.

(D) Quantitative data When all three sets of measurements (phenological, agronomic and morphological) were pooled for principal component analysis, no clear group was clustered

Table 35. The means and standard deviations (numbers in parentheses) of morpholocical data of 12 G. soja seed accessions. Data collected from the greenhouse in 1982 and 1983.

	# of branches* 1983	# of branches* 1982	stem L.cm ground-ist node	W.(mm) banner petal	flower Length (mm)	Flower tube L.(mm)	3-seed pod length(cm)	3-seed pod width(cm)	Pubescence of pod length(cm)	angle
K109	0.25 (0.44)	0.55 (0.89)	6.81 (5.88)	5.63 (0.50)	6.78 (0.41)	2.94 (0.17)	3.16 (0.11)	0.555 (0.029)	1.13 (0.126)	45
E4	0.75 (0.79)	0.14 (0.36)	10.25 (2.48)	4.97 (0.43)	6.78 (0.48)	3.03 (0.39)	2.46 (0.13)	0.488 (0.025)	1.47 (0.166)	60
K9	2.00 (0.73)	3.76 (1.48)	2.36 (1.22)	5.06 (0.44)	6.31 (0.48)	2.50 (0.00)	2.48 (0.10)	0.464 (0.034)	1.50 (0.130)	90
K7	2.00 (0.92)	4.17 (1.07)	1.79 (0.65)	4.97 (0.34)	6.75 (0.45)	2.38 (0.22)	2.80 (0.08)	0.538 (0.031)	1.50 (0.135)	90
K102	1.90 (0.45)	3.63 (1.74)	2.58 (0.72)	5.03 (0.34)	6.22 (0.31)	2.94 (0.17)	2.60 (0.09)	0.527 (0.028)	1.44 (0.169)	60
K28	1.95 (0.61)	4.48 (1.78)	1.71 (0.65)	5.72 (0.55)	6.94 (0.40)	2.50 (0.00)	2.98 (0.09)	0.552 (0.030)	1.00 (0.09)	5
K52	1.85 (0.49)	3.67 (1.28)	1.24 (0.59)	4.34 (0.35)	5.56 (0.40)	2.34 (0.24)	2.64 (0.09)	0.586 (0.023)	1.47 (0.159)	60
K42	1.60 (1.23)	1.55 (0.93)	2.65 (0.75)	4.81 (0.36)	6.56 (0.44)	2.47 (0.13)	2.60 (0.12)	0.492 (0.018)	1.51 (0.136)	90
K101	1.85 (0.49)	2.10 (1.22)	1.57 (0.71)	4.47 (0.39)	5.69 (0.25)	2.06 (0.17)	2.34 (0.10)	0.473 (0.031)	1.71 (0.161)	30
K31	2.65 (1.27)	4.75 (2.12)	0.99 (0.35)	4.47 (0.22)	6.00 (0.00)	2.25 (0.26)	2.41 (0.09)	0.491 (0.024)	1.50 (0.130)	60
M	1.80 (0.62)	2.70 (1.63)	1.78 (0.91)	5.84 (0.24)	6.97 (0.34)	2.56 (0.17)	2.86 (0.19)	0.502 (0.020)	1.70 (0.160)	60
K113	0.80 (0.90)	2.38 (1.16)	3.33 (1.85)	5.53 (0.43)	6.72 (0.31)	2.50 (0.00)	2.90 (0.14)	0.605 (0.015)	1.50 (0.120)	60
F test	14.4**	23.8**	17.7**	26.9**	25.5**	16.8**	151.9**	91.9**	28.8**	

* Plants were 4-week old in 1983 and 8-week old in 1982

** Significant at 1% level

Table 35. (continued)

	Pubescence of leaf			10th central leaflet of main stem			average length/width of leaflet	
	length(mm)	density(#/1mm²)	angle	length(cm)	width(cm)	L/W ratio	central	lateral
K109	0.4	19.5 (1.64)	5	7.00 (0.69)	3.00 (0.28)	2.24 (0.42)	3.27 (0.32)	2.28 (0.64)
E4	0.3	26.2 (5.44)	5	6.70 (0.78)	2.69 (0.41)	2.52 (0.24)	3.26 (0.53)	2.43 (0.55)
K9	0.3	17.0 (2.58)	5	5.14 (0.53)	2.57 (0.25)	2.00 (0.11)	2.29 (0.29)	1.75 (0.20)
K7	0.3	16.0 (2.08)	5	5.80 (0.45)	2.63 (0.32)	2.22 (0.17)	2.35 (0.27)	1.85 (0.17)
K102	0.3	7.4 (0.88)	5	5.45 (0.54)	3.07 (0.32)	1.78 (0.09)	2.05 (0.24)	1.82 (0.20)
K28	0.5	14.7 (2.06)	5	5.15 (0.82)	2.90 (0.43)	1.78 (0.22)	1.70 (0.15)	1.45 (0.09)
K52	0.4	11.0 (3.57)	5	5.14 (0.41)	1.92 (0.19)	2.69 (0.21)	3.45 (0.55)	2.54 (0.33)
K42	0.4	13.3 (1.66)	5	5.69 (0.66)	3.11 (0.41)	1.84 (0.11)	2.07 (0.26)	1.71 (0.15)
K101	0.4	21.5 (3.10)	5	4.90 (0.45)	2.12 (0.21)	2.32 (0.18)	2.39 (0.30)	1.87 (0.18)
K31	0.3	18.2 (2.29)	5	4.11 (0.59)	1.99 (0.29)	2.07 (0.14)	2.81 (0.44)	2.08 (0.23)
M	0.5	17.7 (1.42)	5	5.01 (0.46)	2.84 (0.34)	1.77 (0.11)	1.81 (0.15)	1.64 (0.13)
K113	0.5	20.6 (2.48)	15	5.66 (0.53)	3.08 (0.39)	1.80 (0.09)	1.80 (0.43)	1.64 (0.12)
F test		68.7**		35.3**	34.8**	29.3**	40.3**	12.73**

among the 12 seed accessions on the plane formed by the first two principal components (accounting for 37.2% and 18.0% of the total variation, respectively)(Figure 25). This set of pooled data is called quantitative data of 12 seed accessions in this report. A strong correlation ($r = -0.841$) between the first principal component and latitude is shown in Figure 26. The result indicates that 37.2% of the quantitative variation measured among the 12 seed accessions is highly associated with their latitudinal locations. This association may be mainly contributed by the phenological and agronomic variation because 59.6% of the phenological variation and 43.1% of the agronomic variation are highly associated with latitude, and no morphological variation is significantly correlated with latitude. Since most of phenological and agronomic characters are relative to reproduction of a plant, whether a plant can exist and reproduce in an area is mainly determined by environmental factors, such as day length, temperature, water, and growing season. These environmental factors change latitudinally. Many morphological characters are generally believed to be the result of adaptation of plants for growth and competitive survival. For instance, leaf size, shape as well as the density and length of pubescence affect heat exchange and photosynthesis (Givnish, 1979, Barbour et al., 1980). The inflorescence size and flower shape affect the pollination (Waser, 1983). However, the environmental factors which affect the morphological

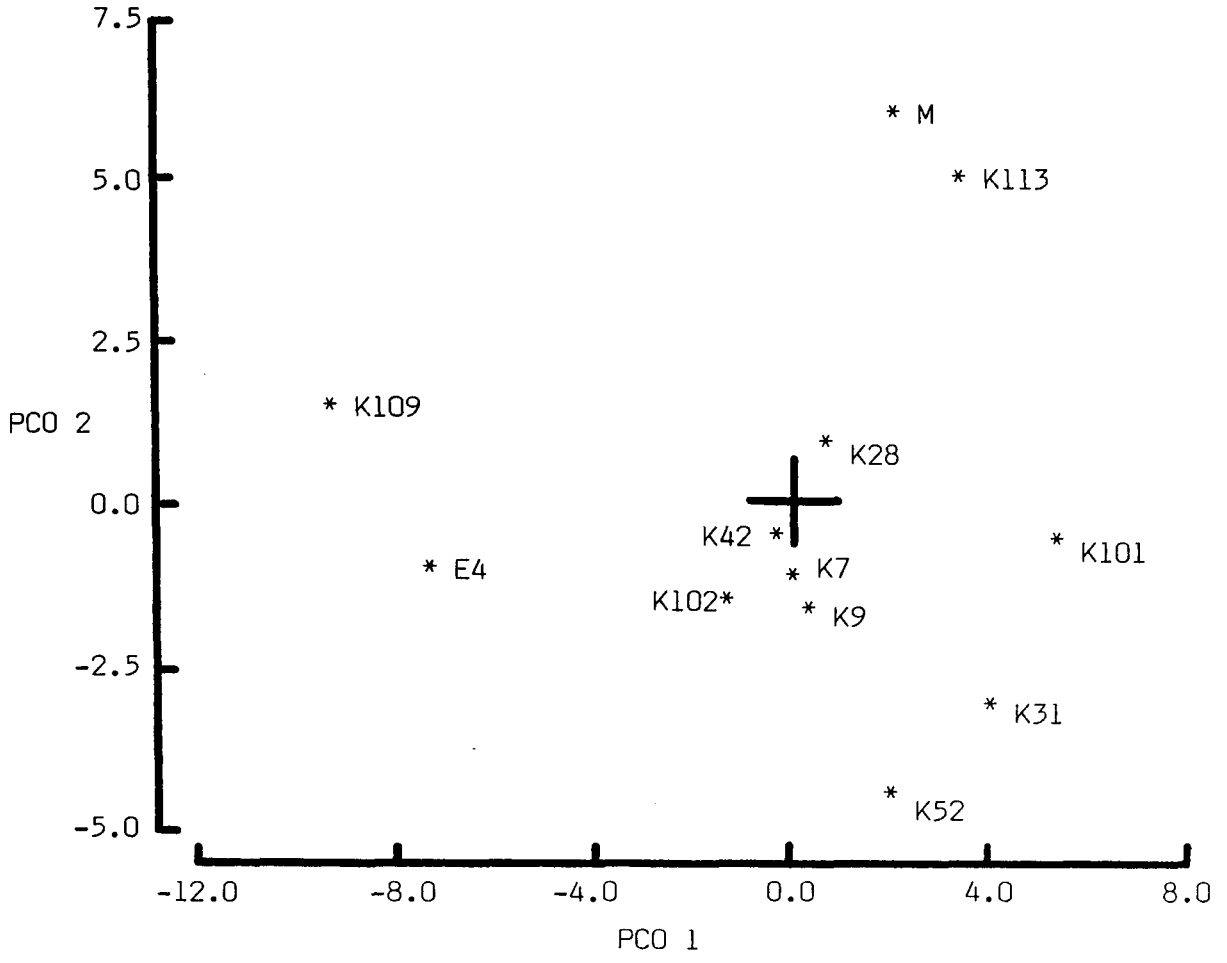


figure 25. Ordination of the 12 G. soja accessions against the first (PCO 1) and second (PCO 2) principal components of the quantitative variation observed.

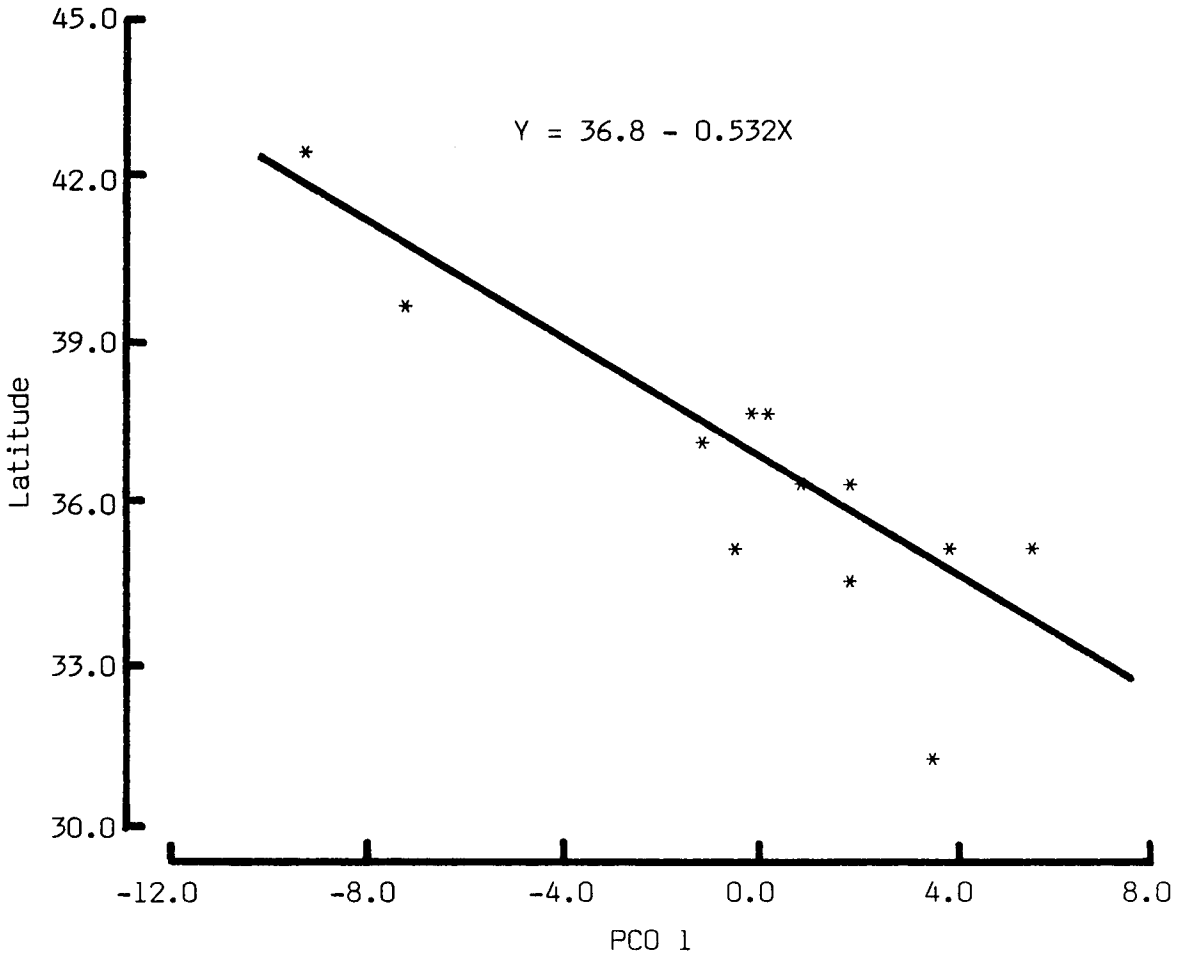


Figure 26. Ordination of the 12 G. soja accessions against the latitude and the first principal component (PCO 1) which accounting 37.2% of the total quantitative variation observed.

characters may vary locally, such as the canopy, the neighborhood species, and the pollinator.

(E) Quantitative measurements of 12 accessions in field experiment. Tables 36 and 37 are lists of several quantitative measurements for the 12 seed accessions in 1982 and 1983 field experiments, respectively. It took longer for seed to germinate and for plants to flower in field than in the greenhouse. K113 and M even didn't flower before frost damaged the plants in 1982. However, the flowering time still correlated highly with latitude as it did in greenhouse experiment. Higher within accession variation was observed in field than in greenhouse experiments for all of the quantitative characters studied. In addition to the more heterogeneous environment in field than in greenhouse, the higher within-accession variation observed in the field may be due to various degree of insecticide and herbicide injuries among plants. Many plants died or damaged seriously after insecticide and herbicide application to control cut worms and weeds.

Table 36. Quantitative data of 12 G. soja accessions in 1982 field experiment. (The number in the parenthesis is the standard deviation of the accession mean.)

	# days (sown - germinated)	height (cm) 9 wks old	# branch 9 wks old	# days(germinated - flowering)	10th leaflet length/width
K109	10.2 (2.51)	29.1 (13.9)	3.4 (2.57)	104.4 (2.57)	2.2 (0.28)
E4	12.6 (3.95)	37.7 (8.39)	5.6 (2.07)	103.6 (3.03)	2.2 (0.14)
K9	9.5 (1.61)	24.0 (11.1)	5.4 (2.48)	111.4 (4.38)	1.6 (0.20)
K7	9.9 (1.63)	22.0 (10.1)	4.7 (1.84)	106.1 (1.94)	1.7 (0.21)
K102	9.5 (1.57)	17.7 (8.52)	4.1 (1.76)	109.8 (7.28)	1.6 (0.12)
K28	10.6 (1.42)	10.5 (4.83)	4.3 (2.32)	111.3 (9.64)	1.5 (0.21)
K52	10.0 (1.66)	6.8 (1.81)	6.6 (2.22)	106.4 (3.47)	2.0 (0.23)
K42	7.8 (1.36)	14.4 (9.84)	4.4 (2.97)	101.0 (7.81)	1.5 (0.13)
K101	10.9 (2.73)	8.6 (4.51)	4.7 (2.16)	118.8 (5.46)	1.9 (0.17)
K31	10.6 (1.27)	5.9 (2.50)	4.5 (2.58)	111.7 (5.61)	1.7 (0.15)
M	9.0 (1.37)	15.3 (5.55)	4.6 (1.45)	---	1.7 (0.16)
K113	11.5 (1.44)	19.5 (9.78)	4.5 (2.23)	---	1.6 (0.15)
F test	8.1**	8.9**	1.77	9.2**	17.3**

** significant at 1% level

Table 37. Quantitative data of 12 G. soja accessions in 1983 field experiment. (The number in parenthesis is the standard deviation of the accession mean).

	# branches 10 wks old	height(cm) 10 wks old	above ground dry wt.(g) 10 wks old	root/total wt. 10 wks old	# nodules 10 wks old	# days(germinnated - flowering)
K109	9.7 (2.9)	27.2 (12.8)	3.5 (2.4)	10.6 (6.2)	9.3 (6.9)	83.8 (7.1)
E4	12.5 (4.7)	59.9 (32.2)	9.2 (7.1)	8.8 (4.5)	8.4 (7.4)	91.5 (6.9)
K9	10.7 (3.3)	26.9 (14.9)	7.3 (5.4)	9.7 (6.1)	11.3 (6.6)	108.1 (3.9)
K7	9.5 (4.7)	42.6 (30.6)	8.1 (8.0)	12.5 (8.5)	10.2 (6.3)	104.3 (2.6)
K102	10.5 (3.5)	40.5 (30.9)	11.0 (8.6)	8.5 (2.9)	7.9 (16.6)	103.7 (5.6)
K28	11.3 (3.0)	24.9 (14.3)	6.3 (6.6)	14.2 (4.6)	28.0 (17.7)	105.3 (3.4)
K52	11.2 (2.7)	21.2 (12.5)	9.5 (5.2)	8.8 (1.9)	18.3 (11.2)	103.6 (2.8)
K42	10.0 (1.4)	18.4 (4.8)	4.2 (2.5)	13.0 (1.4)	26.3 (7.0)	102.3 (1.3)
K101	11.0 (4.8)	31.9 (23.4)	12.2 (8.1)	12.3 (7.5)	17.1 (10.0)	117.0 (1.9)
K31	11.2 (2.3)	13.9 (6.9)	8.6 (5.6)	12.6 (3.2)	18.1 (11.2)	107.4 (3.2)
M	12.9 (1.9)	36.3 (15.7)	9.9 (6.2)	8.8 (1.7)	12.6 (8.2)	129.9 (3.7)
K113	14.0 (3.5)	49.2 (25.9)	16.2 (12.4)	6.8 (2.2)	14.5 (7.3)	138.2 (4.7)
F test	1.59	3.09**	0.24	2.16*	0.31	287.7**

* significant at 5% level; ** significant at 1% level

(2). Quantitative variation of the four natural G. soja populations along the Kitakami river, Japan.

Table 38 is the list of population means and standard deviations of 16 quantitative measurements for G, S, K and I, the four populations along the Kitakami river. Among these 16 measured items, only three (length and width of 3-seed pods, 100 seed weight) were significantly different among four populations at the 1% level, and two (total dry weight of 6-week old samplings and pubescence angle to the undersurface of leaf) were different at the 5% level. The rest of the items were not different interpopulationally. The relative positions of these four populations on the first two principal component axes, which accounted for 42.4% and 34.3% of total quantitative variation, respectively, are shown in Figure 27. Although populations G and I are difficult to distinguish on the first component axis, and populations S and K are difficult to distinguish on the second component axis, the four populations can be separated clearly by the first two quantitative principal components (Figure 27). The correlation coefficient between latitude and the first principal component is $r = -0.166$, and between latitude and the second principal component is $r = -0.959$ (Figure 28). This result indicates that 34.3% of the total quantitative variation observed among the four natural populations is significantly associated with latitude. The quantitative characters which contribute the most to the second principal component are length and width of 3-seed

Table 38. The means and standard deviations (number in parenthesis) of the quantitative measurements of the four natural G. soja populations along the Kitakami river, Japan.

	(days) sowed-germinated	#branches 4 wks old	#branches 6 wks old	height(cm) 4 wks old	# of leave 5 wks old	total dry wt. (g) 6 wks old	root/total 6 wks old	#nodules 6 wks old
G	4.867 (0.370)	0.167 (0.389)	2.833 (0.835)	33.5 (8.6)	8.00 (1.37)	3.043 (0.617)	13.10 (3.06)	2.08 (3.60)
S	4.850 (0.243)	0.167 (0.389)	2.833 (0.718)	42.4 (12.6)	9.04 (2.20)	3.657 (0.845)	10.92 (2.31)	0.42 (0.67)
K	4.933 (0.472)	0.417 (0.669)	2.250 (0.754)	37.0 (16.3)	8.67 (2.26)	2.729 (1.010)	11.67 (1.70)	0.58 (1.00)
I	4.750 (0.358)	0.333 (0.651)	3.000 (0.953)	40.0 (10.7)	8.25 (1.20)	3.766 (0.922)	11.73 (1.43)	1.58 (2.35)
F test	0.50	0.64	1.93	1.14	0.77	3.97*	2.03	1.54

	3-seed pod length(cm)	3-seed pod width (cm)	100 seed wt. (g)	length/width leaflet central	lateral	pubescence of leaf undersurface density(/mm²)	angle	length(mm)
G	2.542 (0.116)	0.500 (0)	1.837 (0.355)	1.983 (0.272)	1.742 (0.188)	17.08 (3.23)	13.33 (12.85)	0.367 (0.123)
S	2.480 (0.189)	0.506 (0.049)	2.104 (0.334)	1.942 (0.261)	1.733 (0.192)	17.58 (4.17)	9.58 (10.10)	0.292 (0.131)
K	2.700 (0.295)	0.538 (0.048)	2.067 (0.445)	1.858 (0.156)	1.625 (0.160)	19.83 (5.44)	5.42 (4.98)	0.242 (0.090)
I	2.850 (0.215)	0.558 (0.047)	2.439 (0.391)	1.800 (0.160)	1.683 (0.119)	16.58 (3.45)	16.67 (9.37)	0.350 (0.138)
F test	7.26**	5.24**	5.02**	1.69	1.24	1.42	2.97*	2.63

* significant at 5% level
** significant at 1% level

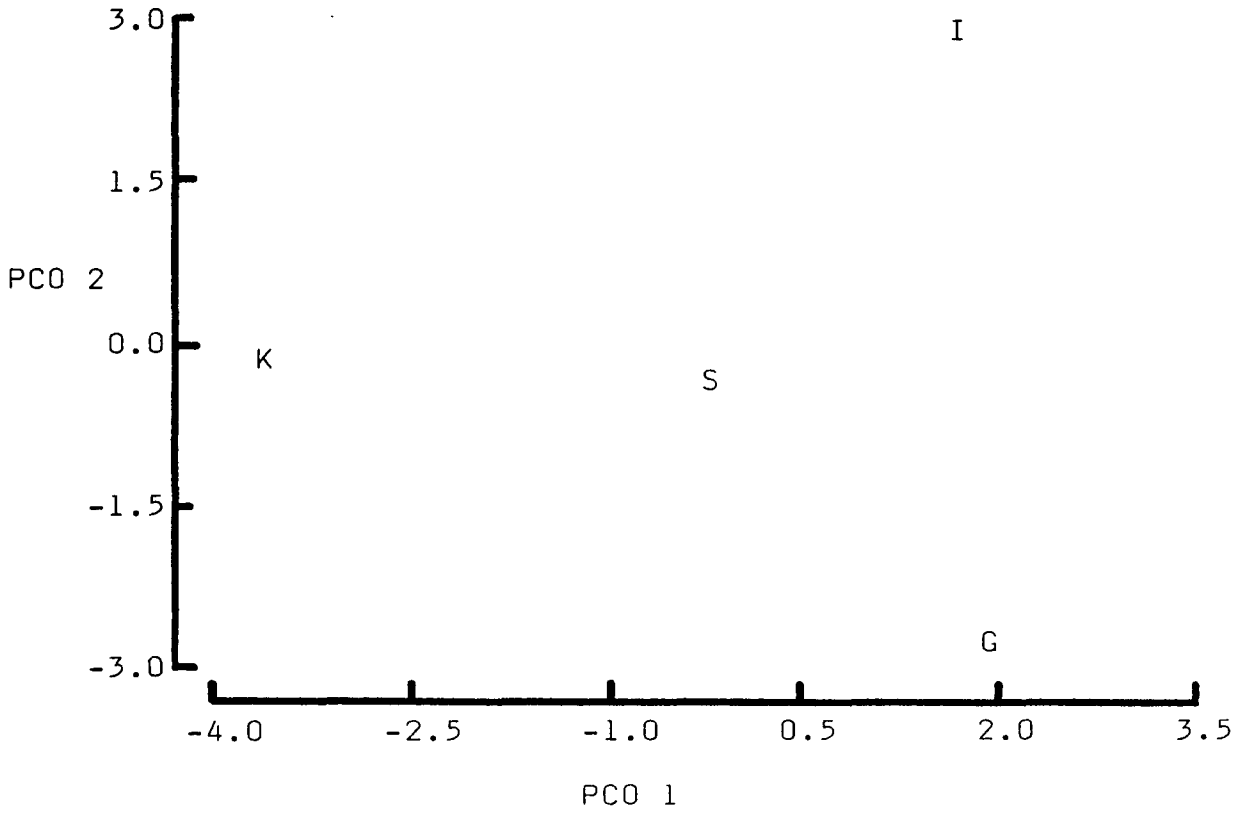


Figure 27. Ordination of the four natural populations (G, S, K and I) against the first (PCO 1) and second (PCO 2) principal components of quantitative variation observed.

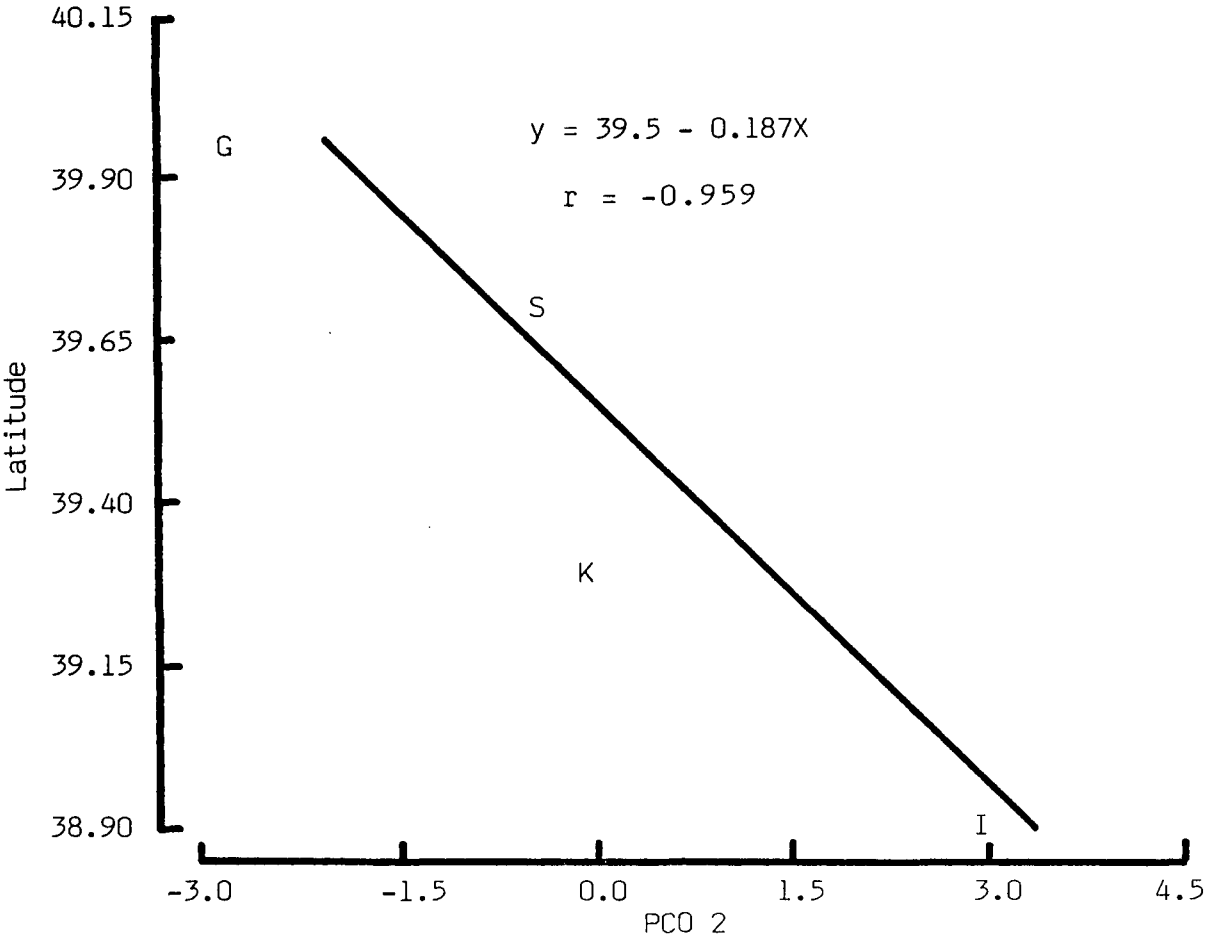


Figure 28. Ordination of the four G. soja populations (G, S, K and I) against the latitude and the second principal component (PCO 2) of the quantitative variation observed.

pod, 100 seed weight and length/width of central leaflet.

5. COMPARISON OF THE POPULATION DIFFERENTIATION BASED ON PROTEIN VARIATION AND QUANTITATIVE VARIATION.

Electrophoretically distinguishable protein variants have been widely used to estimate the amount of genetic variation within and among populations of plants (Brown, 1979). High associations between enzyme loci and quantitative characters were found in Avena barbata (Hamrick and Allard, 1975), and Lycopersicon esculentum and Solanum pennelli (Tanksley et al., 1981). Linhart and Mitton (1985) reported that differential female cone production of ponderosa pine was associated with protein genotype, and Stuber et al. (1982) reported that selection based solely allele frequencies at enzyme loci resulted in a significant increase in yield and ear number of maize. In the present study, many correlations between enzyme genotype and quantitative traits were found statistically different from zero as shown in the Appendix IV for the four natural populations. Since the quantitative measurements and protein data are two pictures of a population, the degree of the congruence between them in the estimate of population differentiation may determine to what extent that information about one set of characters provides information about the other set of characters.

(1). The 12 G. soja seed accessions

The Euclidean distance (Sneath and Sokal, 1973; Wishart, 1978) was used to estimate dissimilarities among the 12 seed accessions based on the following data sets:

a. Phenological data of greenhouse
b. Agronomic data of greenhouse
c. Morphological data of greenhouse
d. Quantitative data (combination of a, b, and c)
e. Protein data (allele frequency in numeric form)
f. Protein data (genotype coding in binary form, presence = 1, absence = 0)

The dissimilarity coefficient matrices for these six sets of data of the 12 accessions are listed in Appendix V. All dissimilarity matrices were then used to cluster the 12 seed accessions by Ward's error sum of squares method using the 'CLUSTAN' computer program (Wishart, 1978). The resulting phenograms are presented in Figure 29. The congruence of a pair of phenograms refers to the concordance of their branching topologies. To assess the relative congruence of these phenograms, Farris' cluster distortion coefficient (Farris, 1973) was calculated for each pair of phenograms and listed in Table 39. To calculate this coefficient, each cluster of the first tree (phenogram) is taken as a variable for which each population (accession) is scored as present or absent. These are referred to as tree variables. The mean coefficient of distortion is the mean proportion of extra steps required for these tree variables

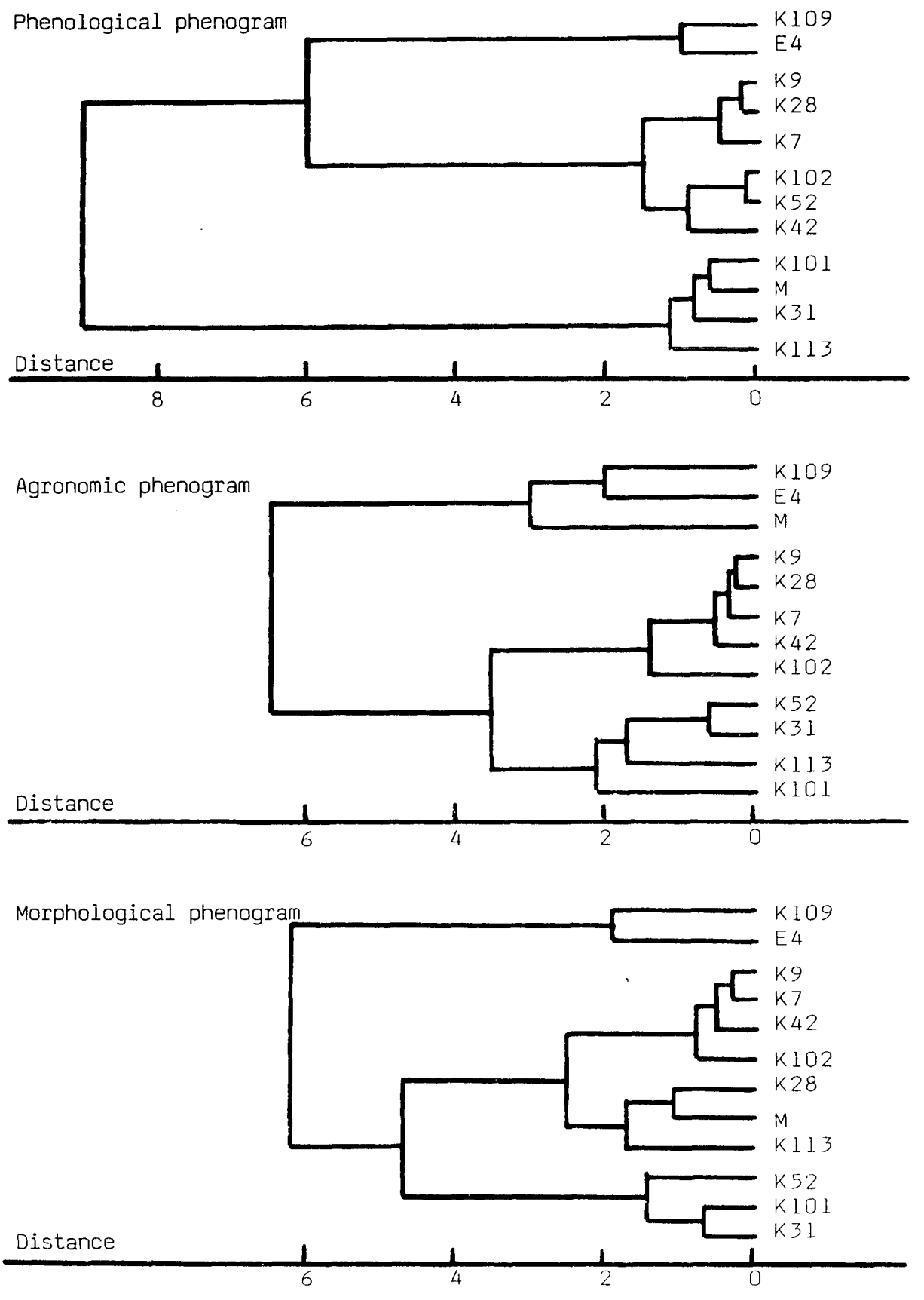


Figure 29. Phenograms of cluster analyses, based on the quantitative and protein data of 12 G. soja seed accessions.

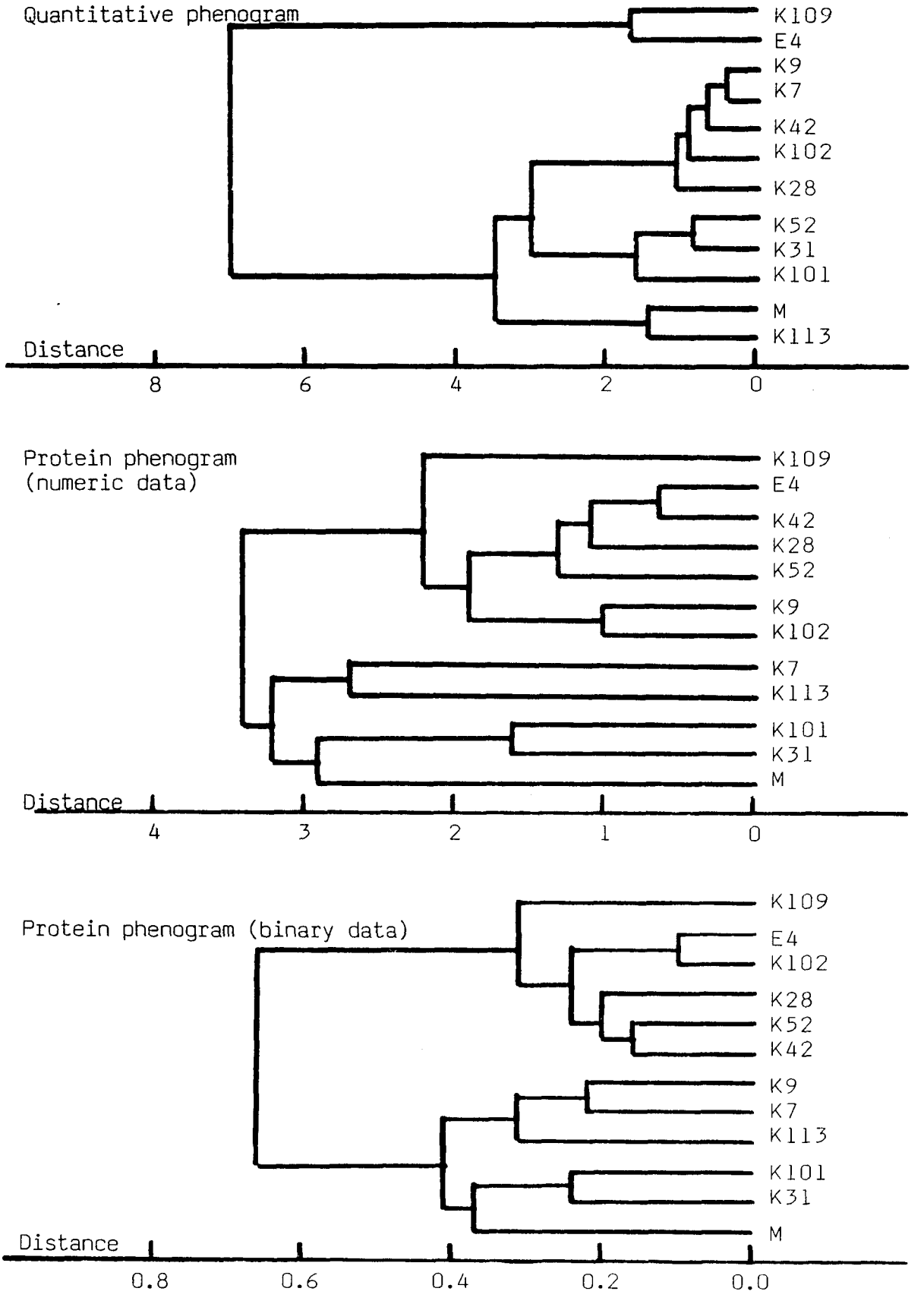


Figure 29. (continued)

Table 39. Mean distortion coefficients between phenograms of 12 G. soja accessions.

(A) Based on protein (numeric form) phenogram

Phenogram	mean distortion coeff.	standard deviation
Protein (binary)	0.52	0.44
Phenological	0.70	0.40
Agronomic	0.95	0.16
Morphological	0.81	0.33
Quantitative	0.96	0.09

(B) Based on protein (binary form) phenogram

Phenogram	mean distortion coeff.	standard deviation
Protein(numeric)	0.54	0.43
Phenological	0.73	0.37
Agronomic	0.98	0.06
Morphological	0.64	0.39
Quantitative	0.79	0.33

(C) Based on quantitative phenogram

Phenogram	mean distortion coeff.	standard deviation
Protein(numeric)	0.81	0.22
Protein(binary)	0.61	0.34
Phenological	0.69	0.42
Agronomic	0.49	0.46
Morphological	0.25	0.41

on the second tree, and represents the mean percentage of possible distortion. Perfect congruence yields a mean distortion coefficient of 0, and complete distortion, a value of 1.0.

The mean distortion coefficients between the phenological phenogram and both protein (numeric and binary forms) phenograms are similar. So are the mean distortion coefficients between agronomic phenogram and the two protein phenograms. However, the mean distortion coefficient between morphological phenogram and protein (numeric form) phenogram is larger than between morphological and protein (binary form) phenograms. A similar pattern is found between the quantitative and both protein phenograms. In a review of the application of electrophoretic data in systematic studies, Buth (1984) argued that electrophoretic data transformation, coding and method of analysis usually affect the results of comparative studies. Using numeric form or binary form of allozyme data is especially at issue. In a cladistic study of Menidia, Mickevich and Johnson (1976) suggested 'although it might seem that frequency coding would yield greater precision, presence-absence coding measures an evolutionarily more significant variable. Since selection can alter the frequencies of only those alleles that are present, acquisition of an allele, for cladistics, may be more important than subsequent modification of a frequency.' Since the present study is a phenetic, not phyletic study, the numeric form may show more

precise genetic relationships among seed accessions than the binary form. But the high genetic purity within accession may reduce the difference in using these two forms to calculate the distance between accessions.

In general, the mean distortion coefficients between the two protein (numeric and binary) phenograms and between the four quantitative phenograms (phenological, agronomic, morphological and quantitative) are smaller than between protein and quantitative phenograms. And the mean distortion coefficients observed in the present study are relatively large. A mean distortion coefficient 0.0036 was reported between morphometrics and allozymes phenograms of Menidia (Mickevich and Johnson, 1976).

Another way to see the concordance between the estimates of population differentiation based on quantitative and protein variations is to compute the correlation between the population distances which are estimated from the quantitative and protein data. The product-moment correlation and the Spearman's rank order correlation (Lindeman et al., 1980) between quantitative and protein distances (dissimilarities in Appendix V and Nei's D_N in Table 23) of the 12 seed accessions are listed in Table 40. The latitudinal distances among these 12 accessions were also included in the correlation computation. The results based on both correlation computing methods were similar. Strong correlations were found between distance estimates based on the same protein data [D_N, protein (numeric form) and

Table 40. Correlations between distances of quantitative, protein and latitude of 12 G. soja seed accessions.

(A) Product-moment correlation

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(a) Latitude							
(b) D_N	0.438**						
(c) Protein(numeric)	0.340**	0.797**					
(d) Protein(binary)	0.308*	0.851**	0.931**				
(e) Phenological	0.747**	0.457**	0.204	0.217			
(f) Agronomic	0.616**	0.443**	0.311*	0.255*	0.666**		
(g) Morphological	0.502**	0.080	-0.069	-0.135	0.523**	0.564**	
(h) Quantitative	0.730**	0.392**	0.185	0.142	0.854**	0.889**	0.807**

(B) Spearman's ranking order correlation

	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(a) Latitude							
(b) D_N	0.333**						
(c) Protein(numeric)	0.292*	0.789**					
(d) Protein(binary)	0.277*	0.840**	0.920**				
(e) Phenological	0.667**	0.387**	0.181	0.199			
(f) Agronomic	0.564**	0.459**	0.341**	0.268*	0.648**		
(g) Morphological	0.536**	0.087	-0.024	-0.080	0.533**	0.532**	
(h) Quantitative	0.691**	0.349**	0.186	0.133	0.834**	0.834**	0.838**

* Correlation coefficient is significantly different from zero at 5% level.

** Correlation coefficient is significantly different from zero at 1% level.

protein (binary form)]. High correlations were also found between distance estimates based on quantitative characters (phenological, agronomic, morphological and quantitative data). However, the correlation coefficients between distance estimates of quantitative and protein characters were relatively low and many of them were not significantly different from zero. Distance estimates based on morphological characters especially showed disconcordant with distance estimates based on protein variations.

Both distances of D_N and protein (numeric form) were computed based on protein allele frequency. However, in the calculation of D_N, the unit character was 'locus', and all the loci (polymorphic and monomorphic) were included. While in the calculation of protein (numeric form) distance, the unit character was 'allele', and only those variable loci among the 12 accessions were included. Thus, D_N includes more genetic information than protein (numeric form) distance. This may be the reason that among the three protein distance estimates, D_N showed better correlation with all quantitative estimates.

(2). The four natural G. soja populations along the Kitakami river, Japan.

The Euclidean distance (Sneath and Sokal, 1973; Wishart, 1978) was used to estimate dissimilarities among these four populations based on the following data sets:

a. Protein data (numeric form, allele frequency in Table 24)

b. Protein data (binary form, presence = 1, absence = 0, based on Table 24)

c. Quantitative data (Table 38)

Thr dissimilarity coefficient matrics obtained are listed in Appendix VI. The product-moment correlation and the Spearman's rank order correlation (Lindeman et al., 1980) were calculated among these three sets of dissimilarity coefficient matrices and Nei's genetic distance (D_N, Table 30). The results are listed in Table 41. There was no difference in the results computed by the two correlation computing methods. All of the correlation coefficients obtained were not statistically different from zero. The concordance between D_N and quantitative distance, again, was higher than between protein (numeric form) and quantitative distance. The correlation coefficient of the latter was negative (Table 41). Among the three protein distance estimates, the distance based on the protein data in binary form showed the highest concordance with the quantitative estimate (Table 41).

Brown (1979) summarized data on plants and concluded that inbreeding plant species show more intense multilocus association than outbreeders. Price et al. (1984) compared the estimates of population differentiation based on genotypes at enzyme loci and measurement characters for three predominantely selfing and one outcrossing plant species. They found that associations between the estimates were statistically significant in the three selfing, but not

Table 41. The (A) product-moment correlation and (B) Spearman's rank order correlation between population distance estimates by protein and quantitative data of the four natural G. soja populations along the Kitakami river, Japan.

(A)

		(1)	(2)	(3)	(4)
(1)	Latitude				
(2)	D_N	0.529			
(3)	Protein(numeric)	0.253	0.791		
(4)	Protein(binary)	0.325	0.548	0.115	
(5)	Quantitative	0.188	0.173	-0.338	0.489

(B)

		(1)	(2)	(3)	(4)
(1)	Latitude				
(2)	D_N	0.486			
(3)	Protein(numeric)	0.200	0.771		
(4)	Protein(binary)	0.207	0.414	0	
(5)	Quantitative	0.371	0.257	-0.314	0.414

in outcrossing species. Price et al. (1984) further suggested that enzyme marker loci may be useful tools for identifying desirable characteristics in predominantly self-fertilizing plant species. In general, the concordance of population differentiation in G. soja measured by quantitative and protein variations is moderately low in the present study. The information collected from phenology, agronomy and morphology include a large number of measurements. Consequently, the quantitative information represents a broader sample of genetic variation than the information collected from 43 protein loci. In other words, quantitative measurements include a better representation of genetic information (gene pool) because each of the traits may be controlled by many genes. While the total protein loci examined is only 40 - 50 loci which is a small sample of the total genome of G. soja. If many more loci are included in the electrophoretic study, the congruence between quantitative and the protein data may be higher than what was observed in the current study. It is remarkable that the population distance estimated on the basis of such a small number of protein loci showed a significant correlation with the distance estimated by the quantitative characters (phenological and agronomic) in the 12 seed accessions (TAble 40). This significant correlation may be mainly attributable to the portion of variation in both protein and quantitative characters that associated with latitude. Since the first principal components (PCO 1) of both

variations strongly associated with latitude (Figures 18 and 26). A graph of these two PCO1 which account for 19.5% and 37.2% of the total protein and quantitative variations, respectively, is presented in Figure 30.

The additional explanation for the general low congruence between estimates based on protein and quantitative characters in G. soja populations may be:

i. None or weak association between these two sets of characters, either function association or position (chromosomal linkage) association.

ii. Plasticity and environmental variation of some of the quantitative traits while the protein characters are qualitative and often clearcut.

iii. The mosaic nature of evolution (Futuyma, 1979) — that the divergence which has occurred between populations in relation to one set of characters has not extended to the other set.

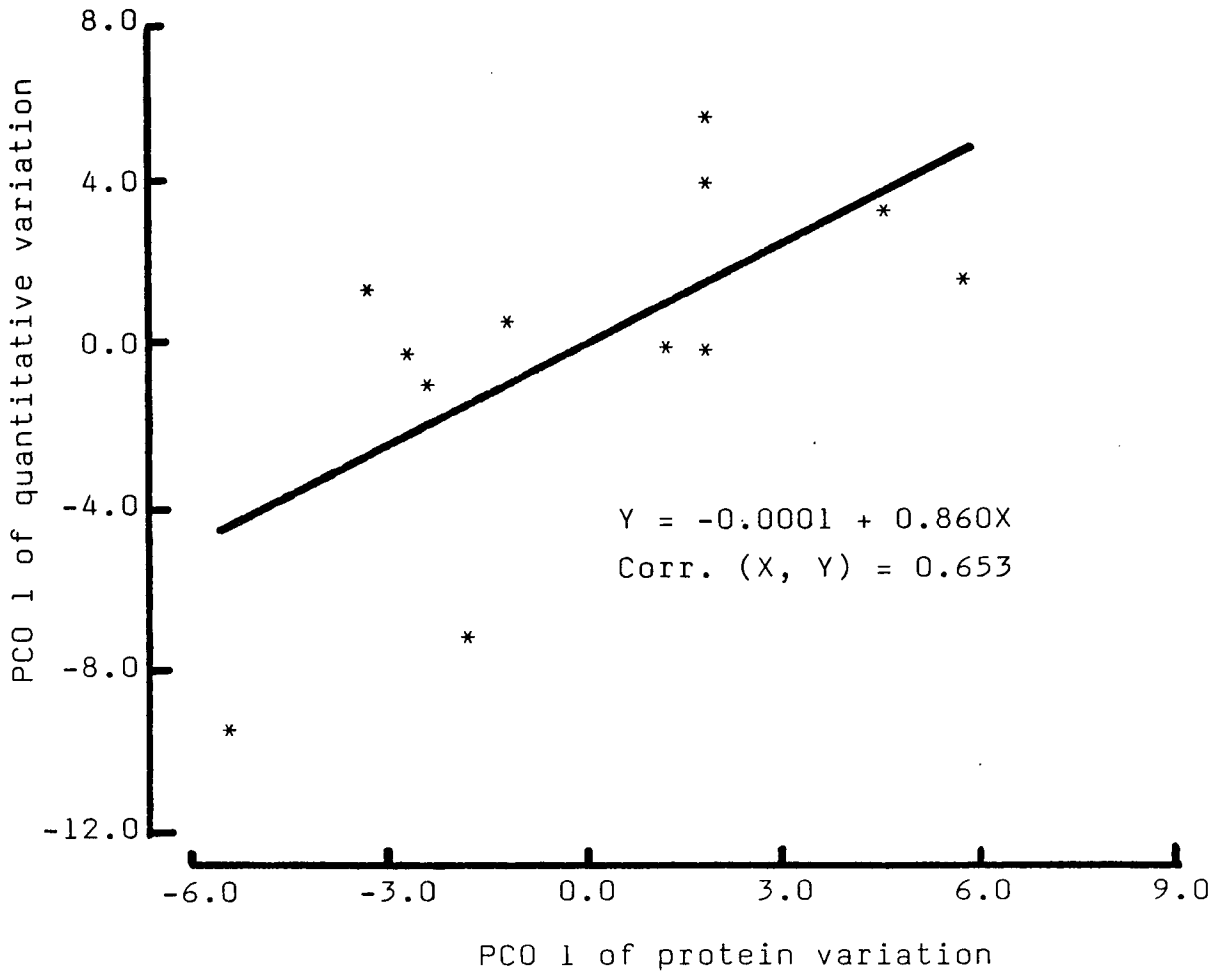


Figure 30. A graph of the 12 G. soja accessions on the first principal compoents (PCO 1) of protein and quantitative variation.

V. CONCLUSIONS

The amount of genetic diversity in four natural populations of G. soja (G, S, K and I) along the Kitakami river shore, Japan, is similar to the average that was found for selfing annual plant species (Hamrick, 1979; Loveless and Hamrick, 1984; Ayala and Kiger, 1980). Based on 42 loci of 15 enzymes and one protein systems, the average polymorphism (P) of these four populations is 38% at the 99% polymorphism level. The expected heterozygosity ($\bar{H}$exp) is 0.114, and the average number of alleles per locus is 1.55. These three measurements are similar to those reported in G. max (Gorman, 1983). The physical distance from north to south among the four G. soja populations is about 120 kms. The results indicate that the genetic diversity of G. soja within 120 kms small area is equal to that of G. max collected from a large geographic area including China, Russia, Japan, South Korea and Taiwan. When the 72 G. soja accessions from South Korea and Japan examined in this study were pooled together, significantly higher genetic diversity was found in G. soja (P 67.4%, $\bar{H}$exp 0.160, no. alleles/locus 2.14) than in G. max (P 34%, $\bar{H}$exp 0.103, No. alleles/locus 1.51). The comparisons show that G. soja has more genetic diversity and is a promising source of germplasm to broaden genetic diversity of cultivated soybeans.

The observed heterozygosity varied greatly among the four G. soja natural populations, ranging from 0 to 4.3%. Except for the observed heterozygosity, the amount of genetic diversity was similar among the four populations. When the four populations are treated as four local populations of the whole Kitakami river population, 19.8% (G_{ST} = 0.198) of the observed total genetic variation occurred among the four local populations whereas 81.2% resided within local populations. The G_{ST} value indicates that the four local G. soja populations are well differentiated isoenzymatically. Their average mean genetic distance $\overline{DN}$ is 0.044. Since Glycine soja is predominantly selfing, low gene flow could be the main cause for the observed genetic differentiation, but ecological differentiation may also play a role.

Several unique alleles were found in G. soja when compared with G. max. These unique alleles can be used for mapping soybean chromosomes. For instance, the linkage of Pgd2 (monomorphic in G. max) with Lap1, Ap and Ti was detected through G. soja because of the polymorphism of Pgd2. Ap, Ti, Lap1 and Pgd2 four loci belong to Linkage Group 9 of soybeans. The gene order of the four loci is Ap-Ti-Lap1-Pgd2, and the recombination frequency between Ap-Ti is 9.41% ± 1.07%, Ap-Lap1 22.65% ± 0.73%, Ap-Pgd2 39.78% ± 0.98%, Ti-Lap1 17.96% ± 1.33%, Ti-Pgd2 36.85% ± 1.39%, and Lap1-Pgd2 20.62% ± 0.99%. Another gene pair, Pgi1-Pgd1 was found to be linked with a recombination frequency of 15.34% ± 0.74%.

Although high genetic purity in each accession was observed in most of G. soja seed accessions, the genetic variation among accessions was significantly different. The high genetic purity within accession may due to sampling error and to the predominantly selfing breeding system in G. soja. The first principal component which accountes for 19.5% of the total genetic variation among 12 seed accessions was significantly associated with latitude ($r = -0.695$).

The 12 G. soja seed accessions were significantly different in most of the quantitative characters examined. Among the phenological characters studied, the durations between seed germination and flowering, and between seed germination and last seed matured were highly correlated with latitudinal locations of the seed accessions. The more northern the origin of the seed accession, the earlier the first flower appeared, and the shorter the life span was. The first principal component which accounted for 59.6% of the total phenological variation among 12 accessions was highly correlated with latitude ($r = -0.912$).

The first principal component which accounted for 43.1% of the total agronomic variation among the 12 accessions was also significantly associated with latitude ($r = -0.745$). This significant association was mainly contributed by the significant correlation between latitude and yield per plant ($r = -0.738$), and number of pods per plant ($r = -0.744$), and average number of seed per pod ($r = 0.595$) in 1982, and harvest index ($r = 0.695$) and average number of seeds per

pod (r = 0.799) in the 1983 greenhouse experiment. None of the principal components of morphological variation among the 12 seed accessions was significantly associated with latitude. When three sets of measurements (phenological, agronomic and morphological) were pooled (as quantitative) for principal component analysis, the first principal component which accounted for the 37.2% of the total quantitative variation among 12 accessions was strongly correlated with latitude (r = -0.841).

Some interesting associations among the quantitative characters were observed. Among the 12 seed accessions, the average number of nodules per plant was positively correlated with the percents of 3-seed pods and 4-seed pods, but negatively correlated with percents of 2-seed pods and 1-seed pods. This seems to suggest that nitrogen is important for increasing the number of seeds per pod in G. soja. Another interesting observation is the positive correlation between the flower size and the percents of 3-seed pods and 4-seed pods, and the negative correlation between flower size and the percents of 2-seed pods and 1-seed pods.

Among the three calculations of protein distance, Nei's D_N had better concordance with quantitative distance in estimating the distance among the 12 seed accessions. In general, the congruence between quantitative and protein variation in estimating population differentiation in G. soja was moderately low. Including more protein loci

in the study of genetic distance is expected to result in better concordance between genetic distance and quantitative distance. The concordance detected in the present study among the genetic and quantitative distances of 12 accessions may be due to the contribution of those characters which were associated with latitude of the seed accessions.

VI. REFERENCES

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

Augspurger, C. K. 1981. Reproductive synchrony of a tropical shrub: Experimental studies on effects of pollinators and seed predators on Hybanthus prunifolius (Violaceae). Ecology 62: 775-788.

Ayala, F. J. 1982. Population and Evolutionary Genetics. Benjamin/ Cummings Pub. Co. Inc., Menlo Park, Ca.

Ayala, F. J.and J. A. Kiger, Jr. 1980. Modern Genetics. Benjamin/ Cummings Pub. Co. Inc., Ca.

Barbour, M. G., J. H. Burk and W. D. Pitts. 1980. Terrestrial Plant Ecology. Benjamin/Cummings Pub. Co. Inc., Mass.

Beremand, M. H. 1975. Abst. Int. Cong. Genet. Genetics 80: s12.

Bernard, R. L. and M. G. Weiss. 1973. Qualitative genetics. In Caldwell, B. E. (ed). Soybeans: Improvement, Production, and Uses. p. 117-154. Am. Soc. of Agronomy, Inc., Madison, Wisconsin.

Brewbaker, J. L., M. D. Upadhaya, Y. Makinen and T. Macdonald. 1968. Isozyme polymorphism in flowering plants. III. Gel electrophoretic methods and applications. Physiol. Plantarium. 21: 930-940.

Brown, A. H. D. 1978. Isozymes, plant population genetic structure and genetic conservation. Theor. Appl. Genet. 52: 145-157.

Brwon, A. H. D. 1979. Enzyme polymorphism in plant populations. Theor. Pop. Biol. 15: 1-42.

Brown, A. H. D., E. Nevo, D. Zohary and O. Dagan. 1978. Genetic variation in natural populations of wild barley (Hordeum spontaneum). Genetica 49: 97-108.

Buth, D. G. 1984. The application of electrophoretic data in systematic studies. Ann. Rev. Ecol. Syst. 15: 501-522.

Buzzell, R. I., B. R. Buttery and R. L. Bernard. 1977. Inheritance and linkage of a magenta flower gene in soybeans. Can. J. Genet. & Cytol. 19: 749-751.

Chiang, Y. C. 1981. The effects of planting arrangement on honey bee visiting pattern and outcrossing rate and yield of soybeans. M.S. thesis. University of New Hampshire.

Devine, T. E., R. G. Palmer and R. I. Buzzell. 1983. Analysis of genetic linkage in the soybean. J. Heredity 74: 457-460.

Erickson, L. R., H. D. Boldeng and W. D. Beversdorf. 1981. Early generation selection for protein in Glycine max x Glycine soja crosses. Can. J. Plant Sci. 61: 901-908.

Fairlet, J. L. and G. L. Kilgour. 1966. Essentials of Biological Chemistry. Reinhold Pub. Co.

Farris, J. S. 1973. On comparing the shapes of taxonomic trees. Syst. Zoology 22: 50-54.

Fukui, J., S. Sunaga and N. Kaizuma. 1978. Comparative investigation on interstrain variation in the growing periods of Siberian (USSR), Northeastern Chinese, South Korean and Japanese strains of wild soybean, Glycine soja Sieb. and Zucc. J. Faculty Agriculture (Iwate Univ.) 14: 71-79.

Futuyma, D. J. 1979. Evolutionary Biology. Sinauer Assoc., Sunderland.

Givinish, T. 1979. On the adaptive significance of leaf form. In Solbrig, O. T. (ed). Topics in Plant Population Biology. Columbia University Press, N.Y.

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, N.H.

Gorman, M. B. and Y. T. Kiang. 1977. Variety specific electrophoretic variants of four soybean enzymes. Crop Sci. 17: 963-965.

Gorman, M. B. and Y. T. Kiang. 1978. Models for the inheritance of several variant soybean electrophoretic zymograms. J. Heredity 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.

Hamrick, J. L. 1979. Genetic variation and longevity. In Solbrig, O. T. (ed). Topics in Plant Population Biology. p. 84-113. Columbia University Press, N.Y.

Hamrick, J. L. and R. W. Allard. 1975. Correlation between quantitative characters and enzyme genotypes in Avena barbata. Evolution 29: 438-443.

Harlan, J. R. 1976. Genetic resources in wild relatives of crops. Crop Sci. 16: 329-333.

Hartl, D. L. 1980. Principles of Population Genetics. Sinauer Associates, Inc., Sunderland, Mass.

Hildebrand, D. F., J. H. Orf and T. Hymowitz. 1980. Inheritance of an acid phosphatase and its linkage with the Kunitz-trypsin inhibitor in seed protein of soybeans. Crop Sci. 20: 83-85.

Hildebrand, D. F. and T. Hymowitz. 1980. Inheritance of beta-amylase nulls in soybean seed. Crop Sci. 20: 727-730.

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

Hymowitz, T. 1976. Soybeans. In Simmonds, N. W. (ed). Evolution of Crop Plants. Longman, London.

Hymowitz, T. and H. H. Hadley. 1972. Inheritance of a trypsin inhibitor variant in seed protein of soybeans. Crop Sci. 12: 197-198.

Hymowitz, T. and N. Kaizuma. 1981. Soybean seed protein electrophoresis profiles from 15 Asian countries or regions: Hypotheses on paths of dissemination of soybeans from China. Econo. Bot. 35: 10-23.

Johnson, H. W. and R. L. Bernard. 1963. Soybean genetics and breeding. In Norman, A. G. (ed). The Soybean. p. 1-73. Academic Press, N.Y.

Kaizuma, N. and J. Fukui. 1974. Intraspecific differences in contents of seed protein and sulphur-containing aminoacids in soya bean (Glycine soja Sieb. and Zucc.) and their significance for plant breeding. Japanese J. Breeding (Ikushugaku Zasshi) 24: 65-72.

Kaizuma, N., Y. Kiuch and F. Ono. 1980. Breeding of high protein soybeans from a species cross between soybeans Glycine max and wild soybeans Glycine soja necessity of repeated backcrossing method. J. Fac. Agric. Iwate Univ. 15: 11-28.

Kiang, Y. T. 1981. Inheritance and variation of amylase in cultivated and wild soybeans and their wild relatives. J. Heredity 72: 382-386.

Kiang, Y. T. and M. B. Gorman. 1983. Soybean. In Tanksley, S. D. and T. J. Orton (eds). Isozymes in Plant Genetics and Breeding, Part B. p. 295-328.

Kiang, Y. T. and M. B. Gorman. 1985. Inheritance of NADP-active isocitrate dehydrogenase isozymes in soybeans. J. Heredity 76: 279-284.

Kiang, Y. T., M. B. Gorman and Y. C. Chiang. 1981. Amylase and acid phosphatase genotypes of Glycine max, Glycine soja and Neonotonia wightii. Soybean Genetics Newsletter 8: 93-104.

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

Kiang, Y. T. and Y. C. Chiang. 1985. Genetic linkage of biochemical and morphological traits in soybean. Crop Sci in press.

Kiang, Y. T. and Y. C. Chiang. 1986. Genetic linkage of a Leucine Aminopeptidase locus (Lap1) with the Kunitz trypsin inhibitor (Ti) in soybeans. J. Heredity. in press.

Kilen, T. C. and W. L. Barrentine. 1983. Linkage relationships in soybeans between genes controlling reactions to phytophthora rot and metribuzin. Crop Sci. 23: 894-896.

Koen, A. L. and M. Goodman. 1969. Aconitate hydratase isozymes: subcellular location, tissue distribution and possible subunit structure. Biochim. Biophys. Acta. 191: 698-701.

Lehninger, A. L. 1982. Principles of Biochemistry. Worth Inc., New York, N.Y.

Lindeman, R. H., P. F. Merenda and R. Z. Gold. 1980. Introduction to Bivariate and Multivariate Analysis. Scott, Foresman and Co., USA.

Linhart, Y. B. and J. B. Mitton. 1985. Relationships among reproduction, growth rates, and protein heterozygosity in ponderosa pine. Am. J. Bot. 72: 181-184.

Loveless, M. D. and J. L. Hamrick. 1984. Ecological determinants of 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. An. Nat. 106: 283-292.

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

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.

Nevo, E., E. Golenberg and A. Beiles. 1982. Genetic diversity and environmental associations of wild wheat, Triticum dicoccoides, in Israel. Theor. Appl. Genet. 62: 241-254.

O'malley, D., N. C. Wheeler and R. P. Guries. 1980. A manual for starch gel electrophoresis. Univ. Wisconsin Madison Staff paper # 11. Dept. of Natural Resources.

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. Heredity 75: 141-142.

Raboy, V., D. B. Dickinson and F. E. Below. 1984. Variation in seed total phosphorus, phytic acid, zinc, calcium, magnesium, and protein among lines of Glycine max and G. soja. Crop Sci. 24: 431-434.

Ram, H. H., Pushpendara, K. Singh and V. D. Verma. 1984. New breeding lines of soybean having a gene for resistance to yellow-mosaic virus from Glycine soja. Indian J. Agric. Sci. 54: 1027-1029.

Rick, C. M., J. F. Fobes and M. Holle. 1977. Genetic variation in Lycopersicon pimpinellifolium: evidence of evolutionary change in mating systems. Plant Syst. and Evol. 127: 139-170.

Rules for Genetic Symbols 1985. Soybean Genetics Newsletter 12: 6-9.

Shaw, C. R. and R. Prasad. 1969. Starch gel electrophoresis of enzymes ---- A compilation of recipes. Biochem. Gen. 4: 297-320.

Singh, L., E. M. Wilson and H. H. Hadley. 1969. Genetic differences in soybean trypsin inhibitors separated by disc electrophoresis. Crop Sci. 9: 489-490.

Sneath, P. H. A. and Sokal, R. R. 1973. Numerical Taxonomy: the principles and practice of numerical classification. W. H. Freeman & Co., San Francisco.

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.

Tanksley, S. D., H. Mdeina-Filho and C. M. Rick. 1981. The effect of isozyme selection on metric characters in an interspecific backcross of tomato - basis of an early screening procedure. TAG 60: 291-296.

Ward, J. H. 1963. J. Amer. Stat. Ass. 58: 236

Waser, N. M. 1983. The adaptive nature of floral traits: ideas and evidence. In Pollination Biology. p. 242-277. Academic Press, Inc.

Williamson, J. R. and B. E. Corkey. 1969. Assays of intermediates of the citric acid cycle and related compounds by fluorometric enzyme methods. In Lowenstein, J. M. (ed). Methods in Enzymology. XIII. Citric acid cycle. p. 434-509.

Wishart, D. 1978. CLUSTAN, user manual. 3rd edition. Inter-University/ Reserach Councils Series, Report No 47.

Yamamoto, K. and Y. Nagato. 1984. Variation of DNA content in the genus Glycine. JPN J. Breed 34: 163-170.

APPENDIX I

Plant Introduction (PI) number of *Glycine soja* seed accessions used in this study.

Accession #	PI #	
South Korean accessions		
1	PI 407.160	(AV 3080)
2	PI 407.161	(AV 3081)
3	PI 407.167	(AV 3083)
4	PI 407.168	(AV 3084; K2c-2,3)
5	PI 407.172	(AV 3085)
6	PI 407.174-177	(AV 3086; K4-2)
7	PI 407.178	(AV 3087; K5-2)
8	PI 407.179-180	(K6-3)
9	PI 407.181	(AV 3089; K7-1,3)
10	PI 407.184	(AV 3110; K29-1,2)
11	PI 407.185-187	(AV 3111; K30-1)
12	PI 407.192	(K9)
13	PI 407.193	(AV 3091; K10-2)
14	PI 407.194	(AV 3092)
15	PI 407.195-197	(K12-2)
16	PI 407.198-199	(AV 3094)
17	PI 407.203-204	(AV 3097)
18	PI 407.205	(K17-1)
19	PI 407.206	(K18-1,2)
20	PI 407.208	(AV 3101)
21	PI 407.211-213	(AV 3102)
22	PI 407.214-215	(AV 3103)
23	PI 407.216	(AV 3104)
24	PI 407.217	(AV 3105)
25	PI 407.219-220	(AV 3106)
26	PI 407.221	(AV 3107)
27	PI 407.222	(AV 3108)
28	PI 407.223-224	(AV 3109; K28)
29	PI 407.225-228	(AV 3128)

APPENDIX I (continued)

Accession #	PI #	
30	PI 407.232	(AV 3130)
31	PI 407.233-235	(AV 3133; K52)
32	PI 407.236-237	(K53-1,3)
33	PI 407.242-244	(K36-2)
34	PI 407.249	(AV 3120)
35	PI 407.252-253	(AV 3112; K31-1,2,3)
36	PI 407.254-255	(AV 3113)
37	PI 407.256-257	(AV 3114)
38	PI 407.260	(AV 3122; K40)
39	PI 407.261	(AV 3123)
40	PI 407.262-264	(AV 3124; K42)
41	PI 407.265	(AV 3125)
42	PI 407.266-268	(AV 3126; K44-3)
43	PI 407.269-270	(K45-1)
44	PI 407.271	(K49-2)
45	PI 407.274	(AV 3132)
46	PI 407.278	(AV 3136; K102)
47	PI 424.016	(74014)
48	PI 424.019	(74017)
49		(74028)
50	PI 424.031	(74034)
51	PI 424.032	(74035)
52	PI 424.035	(74038)
53		(74041)
54	PI 424.045	(74060)
55	PI 424.047	(74062)
56	PI 424.048	(74063)
57	PI 424.053	(74068)
58		(75002)
59		(75042)
60		(9055)
61		(9056)

APPENDIX I (continued)

Accession #	PI #	
62		(9057)
63		(9058)

Japanese accessions

1	PI 486.220	(Misima)
2	PI 487.428	(E4)
3	PI 487.429	(K101)
4	PI 487.430	(K109)
5	PI 487.431	(K113)
6		(G)
7		(S)
8		(K)
9		(I)

APPENDIX II

Gel types and staining methods of proteins

The staining solution listed below is based on 50 ml per one slice gel.

Enzyme	Gel*	Staining solution and method
ACO	5	Buffer(1)** + 15 mg Nicotinamide adenine dinucleotide phosphate (NADP) + 10 mg $MgCl_2$ + 8 ml 10% cis-aconic acid (pH 7.0) + 15 mg 3-(4,5-dimethylthiazo-2yl)-2,5-diphenyl tetrazolium bromide (MTT) + 1 mg Phenazine methosulfate (PMS), 37 °C dark, 3 hrs.
ADH	3+NAD (15 mg)	0.05M sodium phosphate buffer (pH 5.0) + 30 mg Nicotinamide adenine dinucleotide (NAD) + 15 mg MTT + 2 mg PMS + 10 ml 95% ethanol, room temp., dark, 3 hrs.
AM	1	0.2M acetate buffer (pH 5.0) + 1% potato starch, incubated at room temp. for 15 - 30 min., then rinsed with distilled water and stained with 10 ml of 0.1% iodine in 0.5% KI solution.
AP	3 or 4	0.2M acetate buffer (pH 5.0) + 40 mg Na-alpha-naphthyl acid phosphate + 40 mg black K salt, room temp., 1 hr.
DIA	3	Buffer(1) + 1 mg 2,6-dichlorophenol indolphenol + 20 mg nicotinamide adenine dinucleotide reduced form (NADH) + 10 mg MTT, 37 °C, dark, 1 hr.
ENP	3	0.1M Tris-0.1M Maleic acid buffer (pH 5.5) + 20 mg black K salt + 10 mg $MgCl_2$ + 20 mg N-alpha-benzoyl-DL-arginine-beta-naphthyl amide (HCl), 37 °C, dark, 1 hr.
GOT	3	0.1M Tris(HCl) buffer (pH 8.0) + 25 mg pyridoxal-5'-phosphate + 112 mg fast blue BB salt + 272 mg L-aspartic acid + 36 mg Keto-glutaric acid, 37 °C, dark, 1hr.

APPENDIX II (continued)

Enzyme	Gel	Staining solution and method
IDH	3	Buffer(1) + 200 mg DL-isocitric acid + 10 mg NADP + 10 mg MTT + 20 mg $MgCl_2$ + 1 mg PMS, 37°C, dark, 3 hrs.
LAP	3	0.025M Tris maleate buffer (pH 5.2) + 10 mg L-leucine- - naphthyl amide (HCl), incubated 1 hr, 37 °C, dark, then stained with 25 mg black K salt in the same buffer.
MDH	3	0.05M L-malic acid in 0.026M Tris (NaOH) buffer (pH 7.0) + 15 mg MTT + 30 mg NAD + 2 mg PMS, 37 °C, dark, 1 hr.
MPI	3	Buffer(1) + 20 mg mannose-6-phosphate + 10 mg MTT + 15 mg. NAD + 1 mg PMS + 50 units NAD active glucose-6-phosphate dehydrogenase + 50 units phosphoglucose isomerase, 37 °C, dark, 1hr.
PGD	4	Buffer(1) + 15 mg 6-phosphogluconate + 10 mg MTT + 10 mg NADP + 20 mg $MgCl_2$ + 1 mg PMS, 37 °C, dark, 1 hr.
PGI	3	Buffer(1) + 10 mg NAD + 10 mg MTT + 20 mg $MgCl_2$ + 80 mg fructose-6-phosphate (barium salt anhydrous) + 50 units NAD active glucose-6-phosphate dehydrogenase, 37 °C, dark, 1 hr.
PGM	3	Buffer(1) + 15 mg NAD + 10 mg MTT + 100 mg $MgCl_2$ + 1 mg PMS + 100 mg D-glucose-1-phosphate + 50 units NAD active glucose-6-phosphate dehydrogenase, 37 °C, dark, 2hrs.
SDH	3 or 4	Buffer(1) + 15 mg MTT + 15 mg NADP + 2 mg PMS + 15 mg Shikimic acid, 37 °C, dark, 3 hrs.
TI	2	Stained with 0.7% in 10% acetic acid 1-2 min., then incubated with 7% acetic acid, room temp., over night.

---

* Gel type: 1 = 7% acrylamide gel; 2 = 9% acrylamide gel; 3 = 7% acrylamide + 2% starch gel; 4 = 6% acrylamide + 4% starch gel; 5 = 12.5% starch gel.

** 0.2M Tris buffer (pH 8.0)

APPENDIX III

The allele frequency of Glycine soja seed accessions. (The PI number for each accession is listed in the APPENDIX I.)

	Korean Accession #										
	1	2	3	4	5	6	7	8	9	10	11
Aco1-a	0	0	0	0	0	0	0	0	0	0	0
Aco1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco2-a	0	0	0	0	0	0	0	0	0.50	0	0
Aco2-b	0	0	0	0	0	0	0	0	0.50	0	0
Aco2-c	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00
Aco2-d	0	0	0	0	0	0	0	0	0	0	0
Aco3-a	1.00	1.00	1.00	1.00	0	0	0	1.00	0.50	0	0
Aco3-b	0	0	0	0	1.00	1.00	1.00	0	0.50	1.00	1.00
Aco4-a	0	0	0	0	0	0	0	0	0	0	0
Aco4-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco4-c	0	0	0	0	0	0	0	0	0	0	0
Aco5-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco5-b	0	0	0	0	0	0	0	0	0	0	0
Adh1	1.00	1.00	1.00	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh1	0	0	0	0.33	0	0	0	0	0	0	0
Adh3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh3	0	0	0	0	0	0	0	0	0	0	0
Am3-s	0	0	0	0.33	1.00	1.00	0.03	0.15	0	0	0
Am3-f	1.00	1.00	1.00	0.67	0	0	0.97	0.85	1.00	1.00	1.00
Ap-a	1.00	1.00	0.08	0.67	0	0	0.97	0.02	0.50	0	0
Ap-b	0	0	0.04	0	1.00	1.00	0	0.10	0	1.00	0
Ap-c	0	0	0.88	0.33	0	0	0	0.88	0.50	0	1.00
Ap-d	0	0	0	0	0	0	0.03	0	0	0	0
Dia1-a	1.00	1.00	0	0.96	0	0	0.97	0	0.50	1.00	1.00
Dia1-b	0	0	1.00	0.04	1.00	1.00	0.03	1.00	0.50	0	0
Dia2-a	0	0	0	0	0	0	0.06	0.21	0	0	1.00
Dia2-b	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.79	1.00	1.00	0
dia2	0	0	0	0	0	0	0	0	0	0	0
Dia3-a	0	0	0	0	0	0	0	0	0	0	0
Dia3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
dia3	0	0	0	0	0	0	0	0	0	0	0
Dia4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Dia4-b	0	0	0	0	0	0	0	0	0	0	0
Enp-a	0	1.00	0	0.33	1.00	1.00	0	1.00	0	0	0
Enp-b	1.00	0	1.00	0.67	0	0	1.00	0	1.00	1.00	1.00
Enp-c	0	0	0	0	0	0	0	0	0	0	0
Got-a	0	0	0	0	0	0	0	0	0	0	0
Got-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00
Got-c	0	0	0	0	0	0	0	0	0	1.00	0

APPENDIX III (continued)

	Korean Accession #										
	1	2	3	4	5	6	7	8	9	10	11
Idh1-a	0	0	0.08	0	0	0	0	0	0	0	0
Idh1-b	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh2-a	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh2-b	0	0	0.08	0	0	0	0	0	0	0	0
Idh2-c	0	0	0	0	0	0	0	0	0	0	0
Idh3-a	0	0	0	0	0	0	0	0	0.50	0	0
Idh3-b	0	1.00	1.00	1.00	1.00	1.00	0.95	1.00	0.50	1.00	1.00
Idh3-c	1.00	0	0	0	0	0	0.05	0	0	0	0
Idh4-a	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh4-b	0	0	0.08	0	0	0	0	0	0	0	0
Lap1-a	0	1.00	0.17	0	1.00	1.00	0	0	0.50	0	0
Lap1-b	1.00	0	0.83	1.00	0	0	1.00	1.00	0.50	1.00	1.00
Mpi-a	0	0	0	0	0	0	0	0	0	0	0
Mpi-b	1.00	0	1.00	1.00	0	0	0.94	0.21	0.50	0	1.00
Mpi-c	0	1.00	0	0	1.00	1.00	0.06	0.03	0.50	1.00	0
Mpi-d	0	0	0	0	0	0	0	0.76	0	0	0
Pgd1-a	0	0	0	0	0	0	0	0.03	0	0	0
Pgd1-b	0	1.00	1.00	1.00	0	0	0.98	0.97	1.00	1.00	1.00
Pgd1-c	1.00	0	0	0	1.00	1.00	0.02	0	0	0	0
pgd1	0	0	0	0	0	0	0	0	0	0	0
Pgd2-a	0	1.00	1.00	1.00	1.00	1.00	0.10	0.35	1.00	1.00	1.00
Pgd2-b	1.00	0	0	0	0	0	0.90	0.65	0	0	0
Pgd2-c	0	0	0	0	0	0	0	0	0	0	0
Pgd3-a	0	0	0	0	0	0	0	0	0	0	0
Pgd3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgi1-a	0	0	0	0	1.00	1.00	0	0	0	0	0
Pgi1-b	1.00	1.00	1.00	1.00	0	0	1.00	1.00	0.50	1.00	1.00
pgi1	0	0	0	0	0	0	0	0	0.50	0	0
Pgi2	0	0	1.00	0	0	0	0	0.50	0.50	1.00	0
pgi2	1.00	1.00	0	1.00	1.00	1.00	1.00	0.50	0.50	0	1.00
Pgi3-a	0	0	0	0	0	0	0	0	0	0	0
Pgi3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-a	1.00	1.00	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-b	0	0	0.06	0	0	0	0	0	0	0	0
Pgm2-a	0	0	0.06	0	0	0	0	0	0	0	0
Pgm2-b	1.00	1.00	0.94	1.00	0	0	0	1.00	0.50	1.00	0
Pgm2-c	0	0	0	0	1.00	1.00	1.00	0	0.50	0	1.00
Ti-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ti-b	0	0	0	0	0	0	0	0	0	0	0

APPENDIX III (continued)

	Korean Accession #										
	12	13	14	15	16	17	18	19	20	21	22
Aco1-a	0	0	0	0	1.00	0	0	0	0	0	0
Aco1-b	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	1.00	1.00
Aco2-a	0.50	0	0	0	0	0	0	0.50	0	0	0
Aco2-b	0	0	0	0	0	0	0	0	0	0	0
Aco2-c	0.50	0	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00
Aco2-d	0	1.00	0	0	0	0	0	0	0	0	0
Aco3-a	1.00	0	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00
Aco3-b	0	1.00	0	0	0	0	0	0.50	0	0	0
Aco4-a	0	0	0	0	0	0	1.00	0	0	0	0
Aco4-b	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	0
Aco4-c	0	0	0	0	0	0	0	0	0	0	1.00
Aco5-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco5-b	0	0	0	0	0	0	0	0	0	0	0
Adh1	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh1	0	0	0	0.50	0	0	0	0	0	0	0
Adh3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh3	0	0	0	0	0	0	0	0	0	0	0
Am3-s	0	0	1.00	1.00	0	0.33	0	0.50	0	0	0
Am3-f	1.00	1.00	0	0	1.00	0.67	1.00	0.50	1.00	1.00	1.00
Ap-a	0	0.92	0	0	0	0.33	0	0.50	1.00	1.00	0.25
Ap-b	0	0	1.00	0	0	0	0.05	0	0	0	0
Ap-c	1.00	0.08	0	1.00	1.00	0.67	0.95	0.50	0	0	0.75
Ap-d	0	0	0	0	0	0	0	0	0	0	0
Dia1-a	0	0.92	0	1.00	1.00	1.00	1.00	0.50	0.83	0.50	0.67
Dia1-b	1.00	0.08	1.00	0	0	0	0	0.50	0.17	0.50	0.33
Dia2-a	0	0	0	0	0	1.00	0	0	0	0	1.00
Dia2-b	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	0
dia2	0	0	0	0	0	0	0	0	0	0	0
Dia3-a	0	0	0	0	0	0	0	0	0	0	0
Dia3-b	1.00	1.00	1.00	1.00	1.00	1.00	0.13	1.00	0.17	0.50	1.00
dia3	0	0	0	0	0	0	0.87	0	0.83	0.50	0
Dia4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Dia4-b	0	0	0	0	0	0	0	0	0	0	0
Enp-a	0	0	0	0	0	0	0	0	0	0	0
Enp-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Enp-c	0	0	0	0	0	0	0	0	0	0	0
Got-a	0	0	1.00	0	0	0	0	0	0	0	0
Got-b	1.00	1.00	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Got-c	0	0	0	0	0	0	0	0	0	0	0

APPENDIX III (continued)

	Korean Accession #										
	12	13	14	15	16	17	18	19	20	21	22
Idh1-a	0	0	0	0	0	0	0.03	0	0	0	0
Idh1-b	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00
Idh2-a	1.00	1.00	1.00	1.00	1.00	1.00	0.87	1.00	1.00	1.00	1.00
Idh2-b	0	0	0	0	0	0	0.13	0	0	0	0
Idh2-c	0	0	0	0	0	0	0	0	0	0	0
Idh3-a	0	0	0	0	0	0	0.03	0	0	0	0
Idh3-b	1.00	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00
Idh3-c	0	0	0	0	0	0	0	0	0	0	0
Idh4-a	1.00	1.00	1.00	1.00	1.00	1.00	0.87	1.00	1.00	1.00	1.00
Idh4-b	0	0	0	0	0	0	0.13	0	0	0	0
Lap1-a	0.50	0.89	0	1.00	1.00	0	0	0	0.87	0.50	0
Lap1-b	0.50	0.11	1.00	0	0	1.00	1.00	1.00	0.13	0.50	1.00
Mpi-a	0	0	0	0	0	0.67	0	0.50	0	0	0
Mpi-b	0	0.92	1.00	1.00	1.00	0	0.84	0.50	1.00	1.00	1.00
Mpi-c	0.50	0	0	0	0	0.33	0.16	0	0	0	0
Mpi-d	0.50	0.08	0	0	0	0	0	0	0	0	0
Pgd1-a	0	0.50	0.08	0	0.92	0	0.06	0	0	0	0
Pgd1-b	1.00	0.50	0.92	1.00	0.08	1.00	0.94	1.00	1.00	1.00	1.00
Pgd1-c	0	0	0	0	0	0	0	0	0	0	0
pgd1	0	0	0	0	0	0	0	0	0	0	0
Pgd2-a	0.50	0.50	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00
Pgd2-b	0.50	0.50	0	0	0	0	1.00	0	0	0	0
Pgd2-c	0	0	0	0	0	0	0	0	0	0	0
Pgd3-a	0	0	0	0	0	0	0	0	0	0	0
Pgd3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgi1-a	0	0	0	0	0	0	0	0	0	0	0
Pgi1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00
pgi1	0	0	0	0	0	0	0	0.50	0	0	0
Pgi2	0	1.00	0	1.00	0	0	1.00	0.50	1.00	1.00	1.00
pgi2	1.00	0	1.00	0	1.00	1.00	0	0.50	0	0	0
Pgi3-a	0	0	0	0	0	0	0	0	0	0	0
Pgi3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-b	0	0	0	0	0	0	0	0	0	0	0
Pgm2-a	0	0	0	0	0	0	0	0	0	0	0
Pgm2-b	1.00	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	1.00	0
Pgm2-c	0	0	0	0	0	0.33	0	0	0	0	1.00
Ti-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ti-b	0	0	0	0	0	0	0	0	0	0	0

APPENDIX III (continued)

	Korean Accession #										
	23	24	25	26	27	28	29	30	31	32	33
Aco1-a	0	0	0	0	0	0	0	0	0	0	0
Aco1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco2-a	0	0	0	0	0	0	0	0	0	0	0
Aco2-b	0	0	0	0	0	0	0	0	0	0	0
Aco2-c	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00
Aco2-d	0	0	0	0	0	0.50	0	0	0	0	0
Aco3-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco3-b	0	0	0	0	0	0	0	0	0	0	0
Aco4-a	0	0	0	0	0	0	0	0	0	0	0
Aco4-b	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	1.00
Aco4-c	0	0	0	0	0	0.50	0	0	0	0	0
Aco5-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco5-b	0	0	0	0	0	0	0	0	0	0	0
Adh1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh1	0	0	0	0	0	0	0	0	0	0	0
Adh3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh3	0	0	0	0	0	0	0	0	0	0	0
Am3-s	0.25	0.83	0	1.00	1.00	1.00	0	1.00	1.00	0	1.00
Am3-f	0.75	0.17	1.00	0	0	0	1.00	0	0	1.00	0
Ap-a	0.58	0.17	0	0	0	0.50	0	0	1.00	0.50	0
Ap-b	0	0	0	0	0	0	0	0	0	0	0
Ap-c	0.42	0.83	1.00	1.00	0	0	1.00	1.00	0	0.50	1.00
Ap-d	0	0	0	0	1.00	0.50	0	0	0	0	0
Dia1-a	0.17	0	0.92	0	0	0.50	0	1.00	1.00	1.00	1.00
Dia1-b	0.83	1.00	0.08	1.00	1.00	0.50	1.00	0	0	0	0
Dia2-a	0	0	0	0	1.00	1.00	0	1.00	1.00	0.50	0
Dia2-b	1.00	1.00	1.00	1.00	0	0	1.00	0	0	0.50	1.00
dia2	0	0	0	0	0	0	0	0	0	0	0
Dia3-a	0	0	0	0	0	0	0	0	1.00	0	0
Dia3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00
dia3	0	0	0	0	0	0	0	0	0	0	0
Dia4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Dia4-b	0	0	0	0	0	0	0	0	0	0	0
Enp-a	0	0	0	0	0	0	0	1.00	1.00	0.50	1.00
Enp-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	0	0.50	0
Enp-c	0	0	0	0	0	0	0	0	0	0	0
Got-a	0	0	0	0	0	0	0	0	0	0	0
Got-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Got-c	0	0	0	0	0	0	0	0	0	0	0

APPENDIX III (continued)

	Korean Accession #										
	23	24	25	26	27	28	29	30	31	32	33
Idh1-a	0	0	0	0	0	0	0	0	0	0	0
Idh1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh2-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh2-b	0	0	0	0	0	0	0	0	0	0	0
Idh2-c	0	0	0	0	0	0	0	0	0	0	0
Idh3-a	0	0	0	0	0	0	0	0	0	0	1.00
Idh3-b	0.58	0.83	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0
Idh3-c	0.42	0.17	1.00	0	0	0	0	0	0	0	0
Idh4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh4-b	0	0	0	0	0	0	0	0	0	0	0
Lap1-a	0	0	0	0	0	0	0	1.00	0	0	0
Lap1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00
Mpi-a	0	0	0	0	0	0	0	0	0	0	0
Mpi-b	1.00	1.00	1.00	1.00	1.00	0.50	1.00	0	0	0	1.00
Mpi-c	0	0	0	0	0	0.50	0	1.00	1.00	1.00	0
Mpi-d	0	0	0	0	0	0	0	0	0	0	0
Pgd1-a	0	0	0	0	0	0	0	0	0	0.50	0
Pgd1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	0
Pgd1-c	0	0	0	0	0	0	0	0	0	0.50	1.00
pgd1	0	0	0	0	0	0	0	0	0	0	0
Pgd2-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0
Pgd2-b	0	0	0	0	0	0	0	0	0	0	0
Pgd2-c	0	0	0	0	0	0	0	0	0	0	1.00
Pgd3-a	0	0	0	0	0	0	0	0	0	0	0
Pgd3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgi1-a	0	0	0	0	0	0	0	0	0	0	0
Pgi1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
pgi1	0	0	0	0	0	0	0	0	0	0	0
Pgi2	0	0	0	0	0	0	1.00	1.00	0	0.50	1.00
pgi2	1.00	1.00	1.00	1.00	1.00	1.00	0	0	1.00	0.50	0
Pgi3-a	0	0	0	0	0	0	0	0	0	0	0
Pgi3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-a	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00
Pgm1-b	0	0	0	0	0	0	1.00	0	0	0	0
Pgm2-a	0	0	0	0	0	0	0	0	0	0	0
Pgm2-b	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	0.50	1.00
Pgm2-c	0	0	0	0	0	0.50	0	0	0	0.50	0
Ti-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ti-b	0	0	0	0	0	0	0	0	0	0	0

APPENDIX III (continued)

	Korean Accession #										
	34	35	36	37	38	39	40	41	42	43	44
Aco1-a	0	0	0	0	0	0	0	0	0	0	0
Aco1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco2-a	0	0	0	0	0	0	0.50	0	0	0	0
Aco2-b	0	0	0	0	0	0	0	0	0	0	0
Aco2-c	1.00	0.50	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	0
Aco2-d	0	0.50	0	0	0	0	0	0	0	0	1.00
Aco3-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco3-b	0	0	0	0	0	0	0	0	0	0	0
Aco4-a	0	0	0	0	0	0	1.00	0	0	0	0
Aco4-b	1.00	1.00	1.00	1.00	1.00	0	0	1.00	1.00	1.00	1.00
Aco4-c	0	0	0	0	0	1.00	0	0	0	0	0
Aco5-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco5-b	0	0	0	0	0	0	0	0	0	0	0
Adh1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh1	0	0	0	0	0	0	0	0	0	0	0
Adh3	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh3	0	0.50	0	0	0	0	0	0	0	0	0
Am3-s	0	0.50	1.00	0	0	0	0.50	0	0.50	0	0
Am3-f	1.00	0.50	0	1.00	1.00	1.00	0.50	1.00	0.50	1.00	1.00
Ap-a	0	0	0	0	0.99	0	0.50	0	0	1.00	1.00
Ap-b	0	0	0	1.00	0.01	0.83	0	0	0	0	0
Ap-c	1.00	1.00	1.00	0	0	0.17	0.50	1.00	1.00	0	0
Ap-d	0	0	0	0	0	0	0	0	0	0	0
Dia1-a	1.00	0	0	0	0.98	0	1.00	1.00	0.50	1.00	1.00
Dia1-b	0	1.00	1.00	1.00	0.02	1.00	0	0	0.50	0	0
Dia2-a	0	0	0	0	0	0	1.00	0	0	0	1.00
Dia2-b	1.00	0.50	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	0
dia2	0	0.50	0	0	0	0	0	0	0	0	0
Dia3-a	0	0	0	0	0	0	0	0	0	0	0
Dia3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
dia3	0	0	0	0	0	0	0	0	0	0	0
Dia4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Dia4-b	0	0	0	0	0	0	0	0	0	0	0
Enp-a	0	0	0	0	0.67	1.00	0.50	0	0	0	0
Enp-b	1.00	1.00	1.00	1.00	0.33	0	0.50	1.00	1.00	1.00	1.00
Enp-c	0	0	0	0	0	0	0	0	0	0	0
Got-a	0	0	0	0	0	0	0	0	0	1.00	0
Got-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00
Got-c	0	0	0	0	0	0	0	0	0	0	0

APPENDIX III (continued)

	Korean Accession #										
	34	35	36	37	38	39	40	41	42	43	44
Idh1-a	0	0	0	0	0	0	0	0	0	0	0
Idh1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh2-a	1.00	0.50	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0
Idh2-b	0	0.50	0	0	0	0	0.33	0	0	0	1.00
Idh2-c	0	0	0	0	0	0	0	0	0	0	0
Idh3-a	0	0	0	0	0	0	0	0	0	0	0
Idh3-b	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	1.00	0	0
Idh3-c	0	0	0	1.00	0	0	0	0	0	1.00	1.00
Idh4-a	1.00	0.50	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0
Idh4-b	0	0.50	0	0	0	0	0.33	0	0	0	1.00
Lap1-a	0	0	0	0	0.50	0	0.33	0	0	0	0
Lap1-b	1.00	1.00	1.00	1.00	0.50	1.00	0.67	1.00	1.00	1.00	1.00
Mpi-a	0	0	0	0	0.02	0	0	0	0	0	0
Mpi-b	1.00	1.00	1.00	0	0.15	0.04	0.50	1.00	0.50	1.00	0
Mpi-c	0	0	0	1.00	0.83	0.96	0.50	0	0.50	0	1.00
Mpi-d	0	0	0	0	0	0	0	0	0	0	0
Pgd1-a	0.50	0	0	0	0	0	0	1.00	0	0	0
Pgd1-b	0.50	1.00	1.00	1.00	1.00	1.00	0.50	0	1.00	0	1.00
Pgd1-c	0	0	0	0	0	0	0.50	0	0	1.00	0
pgd1	0	0	0	0	0	0	0	0	0	0	0
Pgd2-a	1.00	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgd2-b	0	0	0	0	0	0	0	0	0	0	0
Pgd2-c	0	0.50	0	0	0	0	0	0	0	0	0
Pgd3-a	0	0	0	0	0	0	0	0	0	0	0
Pgd3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgi1-a	0	0	0	0	0	0	0	0	0	0	0
Pgi1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
pgi1	0	0	0	0	0	0	0	0	0	0	0
Pgi2	0	0	0	0	?	1.00	?	1.00	0	1.00	1.00
pgi2	1.00	1.00	1.00	1.00	?	0	?	0	1.00	0	0
Pgi3-a	0	0	0	0	1.00	0	0.50	0	0	0	0
Pgi3-b	1.00	1.00	1.00	1.00	0	1.00	0.50	1.00	1.00	1.00	1.00
Pgm1-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-b	0	0	0	0	0	0	0	0	0	0	0
Pgm2-a	1.00	0	0	0	0	0	0	1.00	0.50	0	0
Pgm2-b	0	1.00	1.00	0	1.00	1.00	1.00	0	0.50	0	1.00
Pgm2-c	0	0	0	1.00	0	0	0	0	0	1.00	0
Ti-a	1.00	1.00	1.00	0	0.50	1.00	0.50	1.00	1.00	1.00	1.00
Ti-b	0	0	0	1.00	0.50	0	0.50	0	0	0	0

APPENDIX III (continued)

	Korean Accession #										
	45	46	47	48	49	50	51	52	53	54	55
Aco1-a	0	0	0	0	0	0	0	0	0	0	0
Aco1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco2-a	0	0	0	0	0	0	0	0	0	0	0
Aco2-b	0	0	0	0	0	0	0	0	0	0	0
Aco2-c	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco2-d	0	0	0	0	0	0	0	0	0	0	0
Aco3-a	1.00	1.00	0	1.00	1.00	1.00	1.00	0	0	1.00	1.00
Aco3-b	0	0	1.00	0	0	0	0	1.00	1.00	0	0
Aco4-a	0	0	0	0	1.00	0	0	0	0	1.00	0
Aco4-b	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	0	1.00
Aco4-c	0	0	0	0	0	0	0	0	0	0	0
Aco5-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco5-b	0	0	0	0	0	0	0	0	0	0	0
Adh1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh1	0	0	0	0	0	0	0	0	0	0	0
Adh3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh3	0	0	0	0	0	0	0	0	0	0	0
Am3-s	0.17	0	0	0	1.00	0	0.42	1.00	1.00	0.06	0.45
Am3-f	0.83	1.00	1.00	1.00	0	1.00	0.58	0	0	0.94	0.55
Ap-a	0.92	1.00	0	0	1.00	0	0.05	1.00	1.00	0.06	0.59
Ap-b	0	0	0	1.00	0	0.08	0	0	0	0	0
Ap-c	0.08	0	1.00	0	0	0.92	0.95	0	0	0.94	0.41
Ap-d	0	0	0	0	0	0	0	0	0	0	0
Dia1-a	1.00	1.00	0	1.00	1.00	1.00	0.37	0	0	0	0.57
Dia1-b	0	0	1.00	0	0	0	0.63	1.00	1.00	1.00	0.43
Dia2-a	1.00	0	0	1.00	1.00	0	0	0	0	0	0.57
Dai2-b	0	1.00	1.00	0	0	1.00	1.00	1.00	1.00	1.00	0.43
dia2	0	0	0	0	0	0	0	0	0	0	0
Dia3-a	1.00	0	0	0	0	0	0	0	0	0	0
Dia3-b	0	1.00	1.00	1.00	1.00	0	0.17	1.00	1.00	1.00	1.00
dia3	0	0	0	0	0	1.00	0.83	0	0	0	0
Dia4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Dia4-b	0	0	0	0	0	0	0	0	0	0	0
Enp-a	0	0	0	0	0	0	0	0	0	0	1.00
Enp-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0
Enp-c	0	0	0	0	0	0	0	0	0	0	0
Got-a	0	0	0	0	0	0	0	0	0	0	0
Got-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Got-c	0	0	0	0	0	0	0	0	0	0	0

APPENDIX III (continued)

	Korean Accession #										
	45	46	47	48	49	50	51	52	53	54	55
Idh1-a	0	0	0	0	0	0	0	0	0	0	0
Idh1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh2-a	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	1.00	1.00	0.90
Idh2-b	0	0	0	1.00	0	0	0	0	0	0	0.10
Idh2-c	0	0	0	0	0	0	0	0	0	0	0
Idh3-a	0	0	0	0	0	0	0	0	0	0	0.03
Idh3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.87
Idh3-c	0	0	0	0	0	0	0	0	0	0	0.10
Idh4-a	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	1.00	1.00	0.90
Idh4-b	0	0	0	1.00	0	0	0	0	0	0	0.10
Lap1-a	0.17	0	0	0	0	0.80	0.77	0	0	0	0
Lap1-b	0.83	1.00	1.00	1.00	1.00	0.20	0.23	1.00	1.00	1.00	1.00
Mpi-a	0	0	0	0	0	0	0	0	0	0	0
Mpi-b	0	1.00	0	0	1.00	0	0.29	0	0	1.00	0.92
Mpi-c	1.00	0	1.00	1.00	0	1.00	0.71	1.00	1.00	0	0.08
Mpi-d	0	0	0	0	0	0	0	0	0	0	0
Pgd1-a	0.71	0	0	0	0	0.94	0.33	0	0	0.94	0.33
Pgd1-b	0.29	1.00	1.00	1.00	0	0.06	0.67	1.00	1.00	0	0.19
Pgd1-c	0	0	0	0	1.00	0	0	0	0	0.06	0.48
pgd1	0	0	0	0	0	0	0	0	0	0	0
Pgd2-a	1.00	0	0	1.00	1.00	1.00	0.53	1.00	1.00	1.00	1.00
Pgd2-b	0	1.00	1.00	0	0	0	0.47	0	0	0	0
Pgd2-c	0	0	0	0	0	0	0	0	0	0	0
Pgd3-a	0	0	0	0	0	0	0	0	0	0	0
Pgd3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgi1-a	0	0	0	0	0	0	0.52	0	0	0	0
Pgi1-b	1.00	1.00	1.00	1.00	1.00	1.00	0.48	1.00	1.00	1.00	1.00
pgi1	0	0	0	0	0	0	0	0	0	0	0
Pgi2	1.00	0	1.00	1.00	0	1.00	1.00	1.00	1.00	1.00	0
pgi2	0	1.00	0	0	1.00	0	0	0	0	0	1.00
Pgi3-a	0	0	0	0	0	0	0	0	0	0	0
Pgi3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-b	0	0	0	0	0	0	0	0	0	0	0
Pgm2-a	0	0	0	0	0	0	0	0	0	0	0
Pgm2-b	1.00	0	1.00	0	1.00	0.08	0.62	0	0	0.09	0.56
Pgm2-c	0	1.00	0	1.00	0	0.92	0.38	1.00	1.00	0.91	0.44
Ti-a	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00
Ti-b	0	0	0	0	0	0	1.00	0	0	0	0

APPENDIX III (continued)

	Korean Accession #								
	56	57	58	59	60	61	62	63	average
Aco1-a	0	0	0	0	0	0	0	0	0.02
Aco1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
Aco2-a	0	0	0	0	0	0	0	0	0.03
Aco2-b	0	0	0	0	0	0	0	0	0.01
Aco2-c	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91
Aco2-d	0	0	0	0	0	0	0	0	0.05
Aco3-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.84
Aco3-b	0	0	0	0	0	0	0	0	0.16
Aco4-a	1.00	1.00	1.00	0	0	0	0	0	0.11
Aco4-b	0	0	0	1.00	1.00	1.00	1.00	1.00	0.85
Aco4-c	0	0	0	0	0	0	0	0	0.04
Aco5-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco5-b	0	0	0	0	0	0	0	0	0.00
Adh1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
adh1	0	0	0	0	0	0	0	0	0.01
Adh3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
adh3	0	0	0	0	0	0	0	0	0.01
Am3-s	0.82	1.00	0	0	0	0	0	0	0.33
Am3-f	0.18	0	1.00	1.00	1.00	1.00	1.00	1.00	0.67
Ap-a	1.00	1.00	0.55	1.00	0	0.06	0	1.00	0.39
Ap-b	0	0	0	0	1.00	0.91	1.00	0	0.16
Ap-c	0	0	0.45	0	0	0.03	0	0	0.42
Ap-d	0	0	0	0	0	0	0	0	0.03
Dia1-a	0.90	1.00	0.43	1.00	0	0.03	1.00	1.00	0.59
Dia1-b	0.10	0	0.57	0	1.00	0.97	0	0	0.41
Dia2-a	0.89	1.00	0	0	1.00	0.89	0.64	0	0.28
Dia2-b	0.11	0	1.00	1.00	0	0.11	0.36	1.00	0.71
dia2	0	0	0	0	0	0	0	0	0.01
Dia3-a	0	0	0	0	0	0	0	0	0.03
Dia3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91
dia3	0	0	0	0	0	0	0	0	0.06
Dia4-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Dia4-b	0	0	0	0	0	0	0	0	0.00
Enp-a	0	0	0	0	0	0	0	0	0.17
Enp-b	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	0.82
Enp-c	0	0	0	0	0	0.50	0	0	0.01
Got-a	0	0	0	1.00	0	0	0	0	0.05
Got-b	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	0.93
Got-c	0	0	0	0	0	0	0	0	0.02

APPENDIX III (continued)

	Korean Accession #								
	56	57	58	59	60	61	62	63	average
Idh1-a	0	0	0	0	0	0	0	0	< 0.01
Idh1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh2-a	0.93	1.00	1.00	1.00	1.00	0.72	0	1.00	0.93
Idh2-b	0.07	0	0	0	0	0.28	1.00	0	0.07
Idh2-c	0	0	0	0	0	0	0	0	0
Idh3-a	0	0	0	0	0	0	0	0	0.02
Idh3-b	0.93	1.00	1.00	0	1.00	1.00	1.00	1.00	0.87
Idh3-c	0.07	0	0	1.00	0	0	0	0	0.11
Idh4-a	0.93	1.00	1.00	1.00	1.00	0.72	0	1.00	0.93
Idh4-b	0.07	0	0	0	0	0.28	1.00	0	0.07
Lap1-a	0	0	0.45	0	0	0	0	0	0.20
Lap1-b	1.00	1.00	0.55	1.00	1.00	1.00	1.00	1.00	0.80
Mpi-a	0	0	0	0	0	0	0	0	0.02
Mpi-b	0.90	1.00	0	1.00	0	0.17	0	1.00	0.57
Mpi-c	0.10	0	1.00	0	1.00	0.83	1.00	0	0.41
Mpi-d	0	0	0	0	0	0	0	0	0.02
Pgd1-a	0.10	0	0.25	0	0	0	1.00	0	0.13
Pgd1-b	0.13	0	0.75	0	1.00	1.00	0	0	0.69
Pgd1-c	0.77	1.00	0	1.00	0	0	0	1.00	0.18
pgd1	0	0	0	0	0	0	0	0	0.00
Pgd2-a	1.00	1.00	1.00	1.00	1.00	0.94	0	0	0.84
Pgd2-b	0	0	0	0	0	0.06	1.00	1.00	0.14
Pgd2-c	0	0	0	0	0	0	0	0	0.02
Pgd3-a	0	0	0	0	0	0	0	0	0.00
Pgd3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgi1-a	0.07	0	0	0	0	0	0	0	0.04
Pgi1-b	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94
pgi1	0	0	0	0	0	0	0	0	0.02
Pgi2	0	0	1.00	1.00	0	0	0	0	0.45
pgi2	1.00	1.00	0	0	1.00	1.00	1.00	1.00	0.55
Pgi3-a	0	0	0	0	0	0	0	0	0.02
Pgi3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
Pgm1-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
Pgm1-b	0	0	0	0	0	0	0	0	0.02
Pgm2-a	0	0	0.43	0	0	0	0	0	0.05
Pgm2-b	0.93	1.00	0.54	0	1.00	0.94	1.00	1.00	0.69
Pgm2-c	0.07	0	0.03	1.00	0	0.06	0	0	0.26
Ti-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95
Ti-b	0	0	0	0	0	0	0	0	0.05

APPENDIX III (continued)

	Japanese Accession #									
	1	2	3	4	5	6	7	8	9	average
Aco1-a	0	0	0	0	0	0	0	0	0	0.00
Aco1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco2-a	1.00	0	0	0	1.00	0.43	0.40	0	0.48	0.37
Aco2-b	0	0	0	0	0	0	0	0	0	0.00
Aco2-c	0	1.00	1.00	1.00	0	0.57	0.60	1.00	0.52	0.63
Aco2-d	0	0	0	0	0	0	0	0	0	0.00
Aco3-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Aco3-b	0	0	0	0	0	0	0	0	0	0.00
Aco4-a	1.00	0	0	0	0	0.43	0	0.20	0.32	0.22
Aco4-b	0	1.00	1.00	1.00	1.00	0.57	1.00	0.80	0.68	0.78
Aco4-c	0	0	0	0	0	0	0	0	0	0.00
Aco5-a	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	0.89
Aco5-b	0	0	0	0	1.00	0	0	0	0	0.11
Adh1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh1	0	0	0	0	0	0	0	0	0	0.00
Adh3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
adh3	0	0	0	0	0	0	0	0	0	0.00
Am3-s	1.00	1.00	1.00	1.00	0	1.00	0.66	0.96	0.68	0.81
Am3-f	0	0	0	0	1.00	0	0.34	0.04	0.32	0.19
Ap-a	0	1.00	1.00	1.00	0	0.48	0.80	0.40	0.04	0.52
Ap-b	0	0	0	0	1.00	0.28	0	0	0.48	0.20
Ap-c	1.00	0	0	0	0	0.24	0.20	0.60	0.48	0.28
Ap-d	0	0	0	0	0	0	0	0	0	0.00
Dia1-a	0	1.00	0	1.00	0	0.86	0.90	0.96	0.04	0.53
Dia1-b	1.00	0	1.00	0	1.00	0.14	0.10	0.04	0.96	0.47
Dia2-a	1.00	0	0	0	1.00	0	0	0	0.28	0.25
Dia2-b	0	1.00	1.00	1.00	0	1.00	1.00	1.00	0.72	0.75
dia2	0	0	0	0	0	0	0	0	0	0.00
Dia3-a	0	0	0	0	0	0	0	0	0	0.00
Dia3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
dia3	0	0	0	0	0	0	0	0	0	0.00
Dia4-a	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00	0.89
Dia4-b	0	0	0	0	1.00	0	0	0	0	0.11
Enp-a	0	0	0	1.00	0	0.10	0.26	0.40	0.40	0.24
Enp-b	1.00	1.00	1.00	0	1.00	0.61	0.74	0.24	0.60	0.69
Enp-c	0	0	0	0	0	0.29	0	0.36	0	0.07
Got-a	0	0	0	0	0	0	0	0	0	0.00
Got-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Got-c	0	0	0	0	0	0	0	0	0	0.00

APPENDIX III (continued)

	Japanese Accession #									
	1	2	3	4	5	6	7	8	9	average
Idh1-a	0	0	0	0	0	0	0	0	0	0.00
Idh1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Idh2-a	1.00	1.00	0	1.00	1.00	1.00	0.86	1.00	0.96	0.87
Idh2-b	0	0	1.00	0	0	0	0	0	0.04	0.12
Idh2-c	0	0	0	0	0	0	0.14	0	0	0.02
Idh3-a	0	0	0	1.00	0	0	0	0	0.28	0.14
Idh3-b	1.00	1.00	1.00	0	1.00	1.00	0.92	0.44	0.66	0.78
Idh3-c	0	0	0	0	0	0	0.08	0.56	0.06	0.08
Idh4-a	1.00	1.00	0	1.00	1.00	1.00	0.86	1.00	0.96	0.87
Idh4-b	0	0	1.00	0	0	0	0.14	0	0.04	0.13
Lap1-a	1.00	0	1.00	0	0	0	0.14	0.20	0.66	0.33
Lap1-b	0	1.00	0	1.00	1.00	1.00	0.86	0.80	0.34	0.67
Mpi-a	0	0	0	1.00	0	0	0	0.04	0	0.12
Mpi-b	1.00	1.00	1.00	0	0	0.57	0.84	0.50	0.60	0.61
Mpi-c	0	0	0	0	1.00	0.43	0.16	0.46	0.40	0.27
Mpi-d	0	0	0	0	0	0	0	0	0	0.00
Pgd1-a	0	0	0	0	0	0.14	0.18	0.04	0.34	0.08
Pgd1-b	1.00	1.00	1.00	0	1.00	0.48	0.36	0.72	0.66	0.69
Pgd1-c	0	0	0	1.00	0	0.10	0.24	0.24	0	0.18
pgd1	0	0	0	0	0	0.28	0.22	0	0	0.06
Pgd2-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgd2-b	0	0	0	0	0	0	0	0	0	0.00
Pgd2-c	0	0	0	0	0	0	0	0	0	0.00
Pgd3-a	1.00	0	0	0	0	0	0	0	0	0.11
Pgd3-b	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89
Pgi1-a	0	0	0	0	0	0	0	0	0	0.00
Pgi1-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
pgi1	0	0	0	0	0	0	0	0	0	0.00
Pgi2	1.00	0	1.00	1.00	1.00	0.67	0.36	0.76	0.80	0.73
pgi2	0	1.00	0	0	0	0.33	0.64	0.24	0.20	0.27
Pgi3-a	0	0	0	0	0	0	0	0	0	0.00
Pgi3-b	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pgm1-a	0	1.00	1.00	1.00	1.00	0.62	0.82	1.00	0.96	0.82
Pgm1-b	1.00	0	0	0	0	0.38	0.18	0	0.04	0.18
Pgm2-a	0	0	0	0	0	0	0	0	0	0.00
Pgm2-b	1.00	1.00	1.00	0	1.00	0.86	0.92	0.78	0.56	0.79
Pgm2-c	0	0	0	1.00	0	0.14	0.08	0.22	0.44	0.21
Ti-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ti-b	0	0	0	0	0	0	0	0	0	0.00

APPENDIX IV.

The correlation coefficients between protein genotypes and quantitative traits of the four natural G. soja populations along the Kitakami river, Japan.

	Quantitative traits§							
	1	2	3	4	5	6	7	8
Aco2-a	-0.11	-0.28	0.31	-0.06	0.02	0.13	-0.01	0.19
Aco2-c	0.11	0.24	-0.17	0.03	-0.07	-0.08	0.08	0.28
Aco4-a	0.01	-0.19	0.13	0.13	0.32	-0.14	-0.10	0.30
Aco4-b	-0.01	0.25	-0.03	-0.02	-0.22	0.09	0.10	-0.39*
Am3-s	-0.08	0.14	-0.36*	-0.42*	-0.02	-0.02	-0.34*	-0.27
Am3-f	0.10	-0.17	0.39*	0.38*	-0.05	0.05	0.30	0.19
Ap-a	0.08	-0.37*	0.22	0.24	-0.25	-0.04	-0.18	0.45*
Ap-b	-0.35*	-0.04	0.02	-0.01	-0.13	0.08	0.13	-0.26
Ap-c	0.18	0.39*	-0.20	-0.20	0.35*	-0.05	0.07	-0.25
Dia1-b	-0.09	0.08	-0.04	-0.12	0.34*	0.48**	0.34*	-0.15
Dia1-a	0.09	-0.08	0.04	0.12	-0.34*	-0.48**	-0.34*	0.15
Dia2-b	-0.06	0.15	-0.30	-0.11	-0.31	-0.16	-0.21	-0.11
Dia2-a	0.06	-0.15	0.30	0.11	0.31	0.16	0.21	0.11
Enp-a	0.20	0.06	0.08	-0.08	-0.04	-0.12	0.01	-0.09
Enp-b	0.08	0.09	0.17	0.11	0.24	0.03	0.17	0.20
Enp-c	-0.26	-0.07	-0.28	0	-0.14	0.20	-0.16	-0.16
Idh2-a	-0.06	0.07	-0.07	-0.08	0.14	0.08	-0.16	0.19
Idh2-b	0.06	-0.07	0.07	0.08	-0.14	-0.08	0.16	-0.19
Idh2-c	0	-0.11	0.34*	-0.06	-0.12	0.11	0.17	-0.18
Idh3-a	0.04	-0.13	0.22	0.10	-0.01	0.04	0.27	-0.03
Idh3-b	-0.08	0.28	-0.12	-0.10	-0.18	-0.34*	-0.32	-0.06
Idh3-c	0.02	-0.09	0.07	0.10	0.32	0.34*	0.28	-0.06
Idh4-a	-0.06	0.07	-0.07	-0.08	0.14	0.08	-0.16	0.19
Idh4-b	0.04	-0.13	0.32	0	-0.18	0.04	0.24	-0.27
Lap1-a	-0.18	0.13	-0.02	-0.13	0.55**	0.27	0.29	-0.33
Lap1-b	0.32	-0.19	0.11	0.09	-0.56**	-0.29	-0.34*	0.26
Mpi-a	-0.22	-0.07	0.16	0.25	0.09	0.23	0.07	-0.06
Mpi-b	-0.03	0.34*	-0.18	-0.17	-0.08	-0.16	0.03	-0.26
Mpi-c	0.17	-0.23	0.29	0.16	0.13	0.04	-0.04	0.11
Pgd1-a	-0.15	0.20	-0.15	0.08	-0.62**	-0.19	-0.06	-0.18
Pgd1-b	0.11	0.10	-0.03	0.20	0.09	0.04	0.08	0.29
Pgd1-c	-0.08	-0.12	0.08	-0.22	-0.17	0	-0.11	-0.34*
pgd1	0.15	-0.20	0.15	-0.08	0.62**	0.19	0.06	0.18
Pgi2	-0.16	0	0	0.20	-0.30	-0.24	-0.14	0.10
pgi2	0.02	0.04	0	-0.10	0.35*	0.17	0.10	-0.16
Pgm1-a	-0.14	-0.07	0.07	-0.14	-0.34*	0	-0.07	-0.36*
Pgm1-b	0.22	-0.12	0.22	0.05	0.16	0.09	0.34*	-0.02
Pgm2-b	-0.08	0.05	0.10	0.09	-0.04	0.10	-0.21	0.10
Pgm2-c	0.03	0.09	0.09	-0.08	0.13	-0.20	-0.02	0.13

APPENDIX IV (continued)

Quantitative traits§

	9	10	11	12	13	14	15	16
Aco2-a	0.04	-0.06	0.41*	0.38*	-0.15	0.24	0.09	-0.04
Aco2-b	-0.18	0	-0.37*	-0.37*	0.13	-0.19	-0.04	0.01
Aco4-a	0.04	0.17	0.15	0.26	-0.22	-0.19	0.16	-0.24
Aco4-b	-0.20	-0.01	-0.25	-0.37*	0.18	0.20	-0.19	0.23
Am3-s	-0.24	-0.23	0.06	-0.04	0.15	-0.30	-0.15	-0.22
Am3-f	0.10	0.18	0.04	0.11	-0.17	0.28	0.20	0.17
Ap-a	0.30	-0.11	0.16	0.08	-0.31	0.18	0.03	0.08
Ap-b	-0.12	0.12	-0.11	0.12	-0.04	0.08	0.09	-0.04
Ap-c	-0.13	0.05	-0.10	-0.21	0.32	-0.25	-0.12	-0.10
Dia1-b	0.04	-0.20	0.20	0.08	0.39*	0.16	-0.06	0.16
Dia1-a	-0.04	0.20	-0.20	-0.08	-0.39*	-0.16	0.06	-0.16
Dia2-b	-0.14	-0.03	-0.11	-0.33	0.09	-0.35*	-0.10	-0.10
Dia2-a	0.14	0.03	0.11	0.33	-0.09	0.35*	0.10	0.10
Enp-a	0.02	0.08	-0.22	0.10	0.14	-0.01	-0.06	0.02
Enp-b	0.11	0.04	0.24	-0.19	-0.13	-0.03	-0.20	0.09
Enp-c	0.18	-0.18	-0.20	-0.04	-0.06	0	0.13	-0.10
Idh2-a	0.08	0.10	-0.05	-0.22	0.08	0.01	-0.05	-0.22
Idh2-b	-0.08	-0.10	0.05	0.22	-0.08	-0.01	0.05	0.22
Idh2-c	-0.31	-0.14	0.13	0.06	-0.06	0.13	0.13	-0.06
Idh3-a	0.06	0.13	0.09	0.38*	-0.13	-0.01	0.10	-0.02
Idh3-b	-0.07	0.04	0.09	-0.14	0.17	0.16	0.12	0.18
Idh3-c	-0.05	-0.04	-0.19	-0.21	-0.17	-0.11	0.03	-0.12
Idh4-a	0.08	0.10	-0.05	-0.22	0.08	0.01	-0.06	-0.22
Idh4-b	-0.30	-0.17	0.14	0.18	-0.10	0.10	0.14	0.08
Lap1-a	-0.35*	-0.02	0.13	-0.13	0.10	-0.08	-0.11	-0.08
Lap1-b	0.19	0.12	-0.10	0.17	-0.16	0.05	0.14	0
Mpi-a	-0.08	-0.34*	0.05	-0.01	-0.08	-0.02	-0.12	-0.13
Mpi-b	0.29	0.25	-0.27	-0.47**	0.29	-0.28	-0.26	0
Mpi-c	0.05	-0.04	0.22	0.40*	-0.34*	0.29	0.30	-0.01
Pgd1-a	-0.22	0.17	-0.30	-0.14	-0.17	0.06	-0.07	0.14
Pgd1-b	0.27	-0.01	0.11	0.04	-0.15	-0.22	-0.41*	-0.20
Pgd1-c	-0.39*	-0.04	0	0.04	0.12	0.19	0.46*	0.15
pgd1	0.22	-0.17	0.30	0.14	0.17	-0.06	0.07	-0.14
Pgi2	-0.07	0.04	-0.13	0.21	-0.51**	0.11	-0.09	-0.06
pgi2	-0.07	0	0.04	-0.28	0.46*	-0.21	0.13	-0.03
Pgm1-a	-0.44*	-0.02	-0.11	0.09	-0.03	0.12	0.30	0.03
Pgm1-b	0.21	0.01	0.11	0.31	-0.06	0.08	0.24	0.01
Pgm2-b	0.03	0.03	0.17	-0.08	-0.14	0.04	-0.09	0.02
Pgm2-c	0.01	-0.08	-0.14	-0.04	0.24	-0.26	-0.32	-0.06

APPENDIX IV (continued)

§ Quantitative traits:

1. Days between seed sown and germination
2. Number of branches of 4-week old plant
3. Height (cm) of 4-week old plant
4. Number of leave of 5-week old plant
5. Length (cm) of 3-seed pod
6. Width (cm) of 3-seed pod
7. 100 seed weight (g)
8. Length/width of central leaflet
9. Length/width of lateral leaflet
10. Pubescence density of leaf
11. Pubescence angle of leaf
12. Pubescence length of leaf
13. Number of nodules of 6-week old plant
14. Total dry weight of 6-week old plant
15. Root/total dry weight of 6-week old plant
16. Number of branches of 6-week old plant

* significant at 5% level

** significant at 1% level

APPENDIX V

Dissimilarity coefficient matrices of 12 seed accessions

(1) Coefficients matrix based on phenological data

	K109	E4	K9	K7	K102	K28	K52	K101	K42	K31	M
E4	1.046										
K9	2.529	2.591									
K7	2.813	2.030	0.538								
K102	1.773	2.577	0.507	0.957							
K28	2.656	2.411	0.179	0.314	0.800						
K52	2.238	2.644	0.318	0.725	0.166	0.560					
K101	5.582	5.184	1.899	1.941	2.444	1.537	2.102				
K42	3.151	3.675	0.923	1.063	0.618	1.114	0.863	2.939			
K31	3.290	2.731	0.883	0.568	1.013	0.542	0.970	0.809	1.206		
M	5.845	4.286	2.055	1.357	2.766	1.498	2.382	0.593	2.814	0.610	
K113	7.181	6.119	2.104	2.051	2.968	1.760	2.713	1.132	2.507	1.125	0.715

(2) Coefficients matrix based on agronomic data

	K109	E4	K9	K7	K102	K28	K52	K101	K42	K31	M
E4	2.024										
K9	2.990	1.256									
K7	3.278	1.620	0.334								
K102	2.620	1.390	1.185	0.804							
K28	3.089	1.845	0.331	0.379	1.232						
K52	4.994	3.139	1.503	1.334	1.283	1.312					
K101	6.579	3.540	2.070	2.081	2.409	1.552	1.780				
K42	2.441	1.364	0.470	0.472	0.839	0.511	1.051	2.075			
K31	5.152	3.351	1.178	1.313	1.881	0.950	0.575	1.638	0.974		
M	2.719	2.845	1.812	2.324	2.800	1.496	3.234	2.775	2.022	2.524	
K113	4.532	3.061	2.033	1.682	1.447	1.848	0.965	1.904	1.729	1.863	2.195

(3) Coefficients matrix based on morphological data

	K109	E4	K9	K7	K102	K28	K52	K101	K42	K31	M
E4	1.900										
K9	2.754	2.766									
K7	2.101	2.807	0.383								
K102	2.285	3.231	0.650	0.740							
K28	2.505	4.361	1.546	1.541	1.144						
K52	3.136	3.621	1.797	1.436	2.002	2.934					
K101	3.711	3.238	0.984	1.456	1.962	2.109	1.355				
K42	2.116	2.729	0.431	0.552	0.573	1.357	2.165	1.388			
K31	4.077	3.773	0.613	0.985	1.463	2.029	1.083	0.659	1.492		
M	2.349	3.541	0.978	0.978	1.185	1.144	2.994	1.850	0.722	2.022	
K113	2.129	3.863	2.205	1.739	1.997	1.894	3.451	2.919	1.559	3.328	1.216

APPENDIX V (continued)

(4) Coefficients matrix based on quantitative data

	K109	E4	K9	K7	K102	K28	K52	K101	K42	K31	M
E4	1.729										
K9	2.785	2.159									
K7	2.721	2.167	0.404								
K102	2.279	2.378	0.813	0.819							
K28	2.761	2.926	0.745	0.795	1.089						
K52	3.599	3.192	1.311	1.217	1.266	1.725					
K101	5.256	3.847	1.622	1.812	2.252	1.756	1.704				
K42	2.501	2.462	0.571	0.653	0.684	0.980	1.418	2.039			
K31	4.277	3.350	0.892	1.001	1.504	1.248	0.865	1.062	1.226		
M	3.378	3.471	1.563	1.576	2.190	1.365	2.927	1.874	1.740	1.849	
K113	4.312	4.139	2.115	1.797	2.040	1.843	2.337	2.085	1.864	2.220	1.453

(5) Coefficients matrix based on Protein data (numeric form)

	K109	E4	K9	K7	K102	K28	K52	K101	K42	K31	M
E4	1.468										
K9	2.301	0.995									
K7	2.519	1.746	1.711								
K102	1.750	0.731	1.067	1.946							
K28	2.045	0.882	1.535	2.244	1.389						
K52	1.841	0.990	1.709	2.541	1.721	1.399					
K101	2.436	1.274	1.731	2.485	2.005	1.999	2.264				
K42	1.433	0.663	1.659	2.409	1.394	1.153	1.147	1.938			
K31	2.858	1.390	1.456	2.565	1.799	1.759	2.308	1.610	1.981		
M	3.297	2.135	2.055	2.878	2.866	2.373	2.846	2.340	2.057	2.714	
K113	3.153	2.154	1.891	2.708	2.578	2.310	2.475	2.807	2.538	2.755	2.821

(6) Coefficients matrix based on protein data (binary form)

	K109	E4	K9	K7	K102	K28	K52	K101	K42	K31	M
E4	0.218										
K9	0.400	0.218									
K7	0.364	0.255	0.218								
K102	0.255	0.109	0.218	0.255							
K28	0.327	0.145	0.291	0.291	0.218						
K52	0.255	0.145	0.291	0.364	0.255	0.182					
K101	0.327	0.182	0.291	0.327	0.291	0.291	0.327				
K42	0.236	0.127	0.345	0.382	0.236	0.200	0.164	0.309			
K31	0.418	0.200	0.236	0.345	0.273	0.273	0.345	0.236	0.327		
M	0.473	0.327	0.327	0.364	0.436	0.364	0.400	0.291	0.309	0.382	
K113	0.436	0.327	0.255	0.327	0.364	0.327	0.327	0.364	0.382	0.382	0.327

APPENDIX VI

The dissimilarity coefficient matrices of the four natural *G. soja* populations along the Kitakami river, Japan.

(1) Coefficient matrix based on protein (numeric form) data

population	G	S	K
S	1.539		
K	1.455	2.287	
I	2.027	2.589	2.103

(2) Coefficient matrix based on protein (binary form) data

population	G	S	K
S	0.500		
K	0.500	0.500	
I	0.625	0.500	0.625

(3) Coefficient matrix based on quantitative data

population	G	S	K
S	1.644		
K	2.439	1.717	
I	2.051	1.757	2.391