Bot. Bull. Academia Sinica (1990) **31:** 129-142

Comparing differentiation of wild soybean (*Glycine soja* Sieb. & Zucc.) populations based on isozymes and quantitative traits

Y. T. Kiang and Y. C. Chiang

Department of Plant Biology, University of New Hampshire, Durham, New Hampshire 03824, U. S. A.

(*Received October 17, 1989; Accepted January 24, 1990*)

Abstract. Comparison of differentiation of 12 populations of wild soybean (*Glycine soja* Sieb. & Zucc.) collected from Japan and South Korea based on isozymes and quantitative traits were made. The protein and quantitative phenograms were compared for congruence by calculating distortion coefficients. In general, the concordance of population differentiation among the 12 populations was not high, but 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 phenological and agronomic characters. If many more protein loci were included, higher congruence may be observed.

Key words: Distortion coefficient; *Glycine soja* populations; Quantitative traits; Isozymes; Population distance; Populations.

Introduction

Electrophoretically distinguishable protein variants have been widely used to estimate the amount of genetic variation and differentiation of plant and animal populations. High degrees of correlation between quantitative traits and enzyme genotypes were found in *Avena barbata* (Hamrick and Allard, 1975), *Lycopersicon esculentum* L. and *Solanum pennelli* L. (Tanksley *et al.*, 1981). Linhart and Mitton reported that differential female cone production of ponderosa pine (*Pinus ponderosa* Laws) was associated with protein genotypes. Stuber *et al.* (1982) observed that selection based solely on allele frequencies at enzyme loci resulted in a significant increase in yield and ear number in maize. Price *et al.* (1984) examined enzyme polymorphism and quantitative traits in three inbreeding plant species (*Avena barbata, Hordeum jubatum, H. vulgare*) and one outcrosser (*Clarkia williamsonii*) and detected significant correlation between enzyme polymorphism and quantitative variation only in *A. barbata*. Nevo *et al.* (1979) observed positive correlations between isozyme variation and quantitative traits in *Hordeum spontaneum*. In general, multilocus associations are more common in selfing plants than in outcrossers (Brown 1979), and some degree of concordance is expected between the enzyme variation and quantitative characters. Since the quantitative traits and protein data are two pictures of a population, the degree of congruence between them in estimating population differentiation may determine to what extent information about one set of characters provides information about the other set of characters.

The objective of this paper is to compare the differentiation of wild soybean (*Glycine soja* Sieb. & Zucc.) populations based on variation of isozymes and quantitative traits. Because the wild soybean grows in natural conditions and has not been subjected to artificial selection, therefore, it is an ideal material for comparison of the degree of congruency of the estimate of differentiation based on protein variation and quantitative variation.

[1] Scientific contribution No. 1647 from the New Hampshire Agricultural Experiment Station.

Table 1. *The geographical location of the twelve wild soybean populations in South Korea and Japan*

Populations	Location	Latitude
K109 (PI487.430)	Hiratori, Hokkaido, Japan	40°32′ N
E4 (PI487.428)	Morioka, Iwate Pref., Japan	39°42′ N
K9 (PI407.192)	Chilcheon, Chunseong, South Korea	37°53′ N
K7 (PI407.181)	Maseogu, Yangju, South Korea	37°39′ N
K102 (PI407.278)	Yongin, Yongin, South Korea	37°17′ N
K28 (PI407.223)	Naecheon, Eumseong, South Korea	36°42′ N
K52 (PI407.233)	Sinam, Yeongi, South Korea	36°37′ N
K42 (PI407.262)	Changyeong, Changneong, S. Korea	35°32′ N
K101 (PI487.429)	Noborito, Kanagawa Pref., Japan	35°30′ N
K31 (PI407.252)	Milyang, Milyang, South Korea	35°30′ N
M (PI486.220)	Mishima, Sizuoka Pref., Japan	35°6′ N
K113 (PI487.431)	Ibusuki, Kagoshima Pref., Japan	31°12′ N

Materials and Methods

The wild soybean (*Glycine soja* Sieb. & Zucc.) is a predominantly selfing annual. It grows in roadsides, river banks and bottoms, and waste areas in Central, Northern and Northeastern China and adjacent areas of USSR, Korea, Japan, and Taiwan. The seed used in this study were the kind gift from the Asian Vegetable Research and Development Center in Taiwan, Crop Experiment Station at Suweon, South Korea, Dr. N. Kaizuma of the Iwate University, Japan, and Dr. H. I. Oka of the National Institute of Genetics, Japan. Seeds of the twelve populations used in this study were collected from natural habitats but the detailed method of collection is unknown. The geographical origins of the 12 populations is shown in Table 1 and Fig. 1.

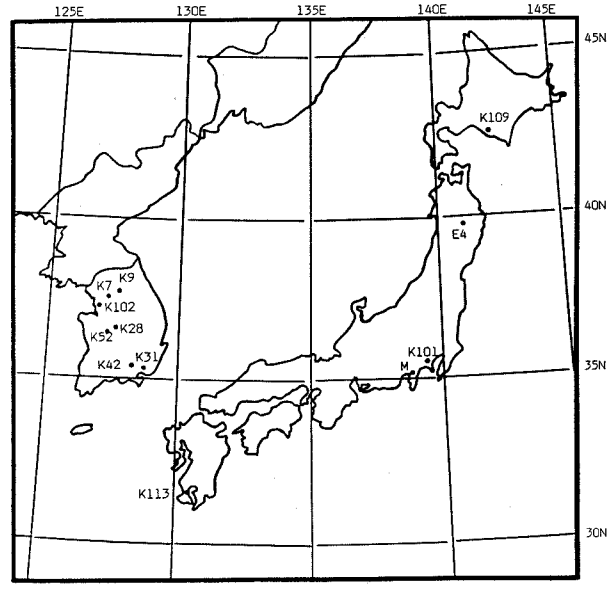

Fig. 1. The geopgraphic locations of the twelve *G. soja* populations of Japan and South Korea.

Genetic Variation Estimated by Electrophoresis

Ten or more seed from each population were examined by electroporesis for the following 17 enzymes and one non-enzymatic protein: Acid Phosphatase (AP, EC 3.1.3.2) Aconitase (ACO, EC 4.2.1.3), Alcohol dehydrogenase (ADH, EC 1.1.1.1.), Amylase (AMY, EC 3.2.1.2), Diaphorase (DIA, EC 1.6.2.2), Endopeptidase (ENP), Esterase (EST, EC 3.1.1.1., and EC 3.1.1.2), Glutamic dehydrogenase (GDH, EC 1.4.1.3), 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). A total of 42 loci were examined. The sample preparations, gels, electrophoretic procedures, and staining methods have been published elsewhere (Bult *et al.*, 1989).

Genetic diversity of each population was estimated

by calculating the allelic frequency, the average number of alleles per locus and the expected heterozygosity (Nei, 1973). Genetic distance and genetic identity were calculated for comparison among populations (Nei, 1972).

Recording Quantitative Data

Twenty plants from each population were grown in the greenhouse at the University of New Hampshire during the summers of 1982 and 1983. Seeds were scarified and inoculated with commercial Rhizobia inoculum before planting. One seed was sown in each of 15 cm diameter plastic pots with 45 cm spacing between pots. Steamed field soil mixed with Promix (a mixture of 60% peat moss, 20% perlite and 20% vermiculite) in a one to one ratio was used as growth medium. A 120 cm-long, 1.5 cm diameter bamboo cane was put up in each pot to support vines. The plants were watered twice a day. The ambient temperatures in the greenhouse were 30±2°C in the daytime and 25±2° at night. The following quantitative traits were recorded.

(1)Phenological characters:

a. number of days from sowing to germination
b. number of days from germination to the first flower
c. number of days from first flower to first fresh pod set
d. number of days from first fresh pod set to first mature pod
e. number of days from first dry pod to last dry pod
f. number of days from anthesis to seed maturity
g. life span-number of days between germination and last dry pod.

(2)Agronomical characters:

a. proportion 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
e. average weight per 100 seed
f. total seed weight per plant
g. harvest index [total seed weight/(total biomass-root)]
h. plant height at four weeks old
i. total dry weight, root dry weight and number of nodules at 10-weeks old

(3)Morphological traits:

a. number of branches at 4 weeks old
b. stem length from ground to the first branch
c. flower size-width of banner petal, longitudinal length of flower and flower tube
d. length and width of 3-seed-pod
e. pubescence length on four-week old green pod
f. pubescence length and density on mature leaf surface
g. length and width of the tenth leaf of main stem (emphasis on the size)
h. length and width of leaf picked randomly (emphasis on shape)

Genetic Distance Among Populations Based on Quantitative Variation

Dissimilarity between populations based on quantitative traits was measured as Euclidean distance (d^2). The formula for the Euclidean distance between two populations, P and q, in an n-dimensional space is $d^2pq=1/n\sum_{i=1}^{n}(x_{ip}-x_{iq})^2$ (Sneath and Sokal, 1973) where n=number of characters measured ($x_{ip}-x_{iq}$) is the difference between the measurements 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.

Genetic Distance Among Populations Based on Protein Variation

Besides Nei's genetic distance (D_n), the Euclidean distance based on allelle frequency was computed as described above. The protein data were further transformed into binary form (presence=1, absence=0) 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 P and absent in population q; C=number of attributes present in populations q and absent in population P.

The Congruency Between Quantitative Data and Protein Data in Estimate 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 quantitative and protein data to test their concordance.
b. The hierachical grouping method of Ward (1963) was applied to obtain phenograms based on quantitative data and protein data. Ward (1963) proposed a hierachical 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 allelic frequency and the binary present-absent transformed allelic data were used to construct phenograms. The congruence between phenograms was measured by mean coefficient of distortion (Farris, 1973).

Results

Genetic variation of the 12 wild soybean populations based on isozymic data are presented in Table 2. Genetic variation of these 12 populations was low and seven of them were monomorphic for all the enzymes examined. This observation reflects the fact that the wild soybean set seed by self-pollination. In addition, each of the 12 populations may have originated from a small number of seeds. The isozyme data were used to calculate genetic distance D_n and genetic identity (I_n) between the 12 populations (Nei, 1972). The amount of genetic identity among the populations varied among populations and no definite pattern associated with latitudinal locations of the populations was detected (Table 3).

Phenological data of the 12 populations are shown in Table 4. Analysis of variance showed that each of the traits examined was significantly different among the 12 populations. The phenological traits were significantly correlated with latitudinal locations of original seed collection (Kiang and Chiang, 1989). For example, plants of northern origin flowered earlier with a shorter life span than those of southern origin.

Many of the agronomic traits studied among the 12 populations were also associated with latitude (Table 5). For instance, the more northern a population origin, the less the seed yield per plant ($r=-0.74$) but more seed per pod ($r=0.80$).

Morphological data collected in the greenhouse are presented in Table 6. All the morphological traits examined were statistically different among the 12 populations. The principal component analysis was performed based on the ten morphological traits. The first two principal components accounted for 33.1% and 26.4%, respectively, of the total morphological variation observed. No significant correlation between morphological traits and latitude was detected.

All three sets of measurements, namely, phenological, agronomic, and morphological were pooled for principal component analysis. This set of pooled data is called quantitative data of 12 populations in this report. No clear group was clustered among the 12 populations on the plane formed by the first two principal components, which account for 37.2 and 18.0% of the total variation, respectively. A strong negative correlation ($r=-0.84$) between the first principal component and latitude indicated that 37.2% of the quantitative variation among the 12 populations was highly associated with latitudinal locations (Kiang and Chiang, 1989). 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 (Kiang and Chiang, 1989). However, no morphological variation is significantly correlated with latitude.

Comparison of Population Differentiation

The quantitative and protein data were used for comparison of differentiation of populations. Dissimilarities among the 12 populations were estimated by the Euclidean distance (Sneath and Sokal, 1973; Wishart, 1978) based on the following data set: (1) Phenological data (2) Agronomic data (3) Morphological data (4) Quantitative data (combination of 1, 2, and 3 (5) Isozyme data (allele frequency in numeric form) (6) Isozyme data (genotype coding in binary form, Presence=1, absence=0). The dissimilarity coefficient matrices for these five sets of data are listed in Table 8. All dissimilarity matrices were then used to cluster the 12 populations by Ward's error sum of squares method employing 'CLUSTAN' computer program (Wishart, 1978). The emerging phenograms are shown in Fig. 2. The congruence of a pair of phenograms refers to the concordance of their branching pattern. To assess the relative congruence of these phenograms, the cluster

Table 2. *Alleles and their frequencies of the 12* ***G. soja*** *populations.*

	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	1.00	0	0	1.00	0	0
Aco4-b	1.00	1.00	1.00	1.00	1.00	0.50	0	0	1.00	0	1.00	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	0	0
Aco5-b	0	0	0	0	0	0	0	0	0	0	1.00	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	0.50	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	1.00	1.00
Ap-c	0	0	1.00	0.50	0	0	0	0	1.00	1.00	0	0
Ap-d	0	0	0	0	0	0.50	0	0	0	0	0	0
Dia1-a	1.00	1.00	0	0.50	1.00	0.50	1.00	1.00	0	0	0	0
Dia1-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	1.00	0	0	0	0	0	0
Dia3-b	1.00	1.00	1.00	1.00	1.00	0	1.00	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
Est1-a	0	0	0	0.50	0	0.50	0	0	0	1.00	0	1.00
Est1-b	1.00	1.00	1.00	0.50	1.00	0.50	1.00	1.00	1.00	0	1.00	0
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
Lap1-a	0	0	0.50	0.50	0	0	0	0	1.00	0	1.00	0
Lap1-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
Pgd1-b	0	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00	1.00	1.00	1.00
Pgd1-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
Pgm1-a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0	1.00
Pgm1-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

*Monomorphic loci were not listed in this table.

Table 3. *Nei's measures of genetic distance (D_N) and identity (I_N) between twelve* ***G. soja*** *populations*

	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	I_N
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		

D_N

Table 4. *Mean number of days and standard deviation (in parenthesis) of phenological data of the 12* ***G. soja*** *populations (average of 1982 & 1983)*

Populations	Days to germination	Days to flower	First flower to first pod	First pod to last mature pod	First dry pod to last dry pod	Life Span	Anthesis to dry pod
K109	5.42 (0.83)	58.0 (2.15)	16.5 (2.34)	38.4 (2.03)	24.7 (6.90)	137.5	41.4 (5.3)
E4	5.72 (0.63)	64.4 (4.14)	14.1 (4.0)	39.3 (7.22)	27.2 (3.85)	144.9	44.2 (4.7)
K9	5.73 (0.66)	87.8 (3.47)	11.3 (2.70)	31.2 (1.87)	23.9 (2.00)	151.5	41.2 (3.6)
K7	5.73 (0.87)	86.5 (3.75)	9.6 (2.0)	32.6 (2.65)	21.6 (2.22)	150.1	43.5 (1.3)
K102	5.46 (0.86)	79.4 (4.86)	13.8 (3.05)	32.6 (2.01)	22.7 (2.53)	148.4	40.1 (4.3)
K28	6.01 (0.71)	86.8 (1.77)	10.4 (1.27)	31.6 (2.03)	22.5 (1.96)	151.1	42.0 (3.6)
K52	5.70 (1.23)	79.3 (4.71)	13.3 (3.80)	32.2 (3.04)	23.6 (1.78)	148.3	40.6 (1.1)
K42	5.23 (0.44)	88.2 (2.62)	9.1 (1.50)	30.5 (1.99)	22.5 (4.54)	150.2	39.5 (3.1)
K101	6.34 (0.97)	100.8 (1.99)	8.2 (2.04)	36.6 (1.90)	16.8 (1.92)	162.4	40.4 (2.3)
K31	4.48 (1.06)	93.0 (3.63)	9.7 (1.83)	34.5 (2.37)	20.4 (2.68)	157.5	41.5 (1.9)
M	6.21 (0.70)	105.2 (2.87)	6.9 (2.08)	36.1 (1.86)	19.5 (1.91)	167.6	42.6 (1.0)
K113	6.09 (0.94)	118.0 (2.29)	7.6 (1.89)	31.6 (1.52)	19.9 (2.52)	277.0	41.0 (0.9)

Table 5. *Means and standard deviation (in Parenthesis) of agronomic data of 12* ***G. soja*** *populations (Average of 1982 and 1983)*

	No. seed/ Pod	100 Seed/ weight (g)	No. pods/ Plant	Seed Wt (g)/ Plant	Harvest Index	No. Nodules /plant (10 wk-old)	Biomass (g) /plant (10 wk-old)	Root dry wt (g) (10 wk-old)	Plant height (cm) (10 wk-old)	Seed Packing (% Total Pods) 1 s-pod	2 s-pod	3 s-pod	4 s-pod
K109	2.72	2.93	142.9	11.5	0.39	56.3	9.05	2.04	27.17	5.3	24.5	64.2	5.9
	(0.14)	(0.17)	(28.9)	(2.37)	(0.02)	(37.9)	(2.57)	(0.06)	(10.95)	(3.8)	(7.6)	(8.0)	(4.6)
E4	2.51	1.98	226.3	11.7	0.36	29.4	9.91	1.97	35.83	12.2	29.1	58.2	0.5
	(0.10)	(0.11)	(51.8)	(2.78)	(0.02)	(28.3)	(3.29)	(0.22)	(18.37)	(4.2)	(6.8)	(7.7)	(0.6)
K9	2.42	2.01	331.0	16.3	0.39	16.0	10.99	2.37	17.52	15.3	31.7	52.7	0.3
	(0.14)	(0.15)	(46.7)	(1.98)	(0.02)	(13.6)	(3.24)	(0.14)	(3.67)	(7.2)	(5.9)	(9.4)	(0.5)
K7	2.40	2.09	297.6	15.3	0.37	18.3	12.75	3.32	12.06	10.7	43.5	45.8	0.1
	(0.12)	(0.21)	(41.2)	(2.62)	(0.03)	(7.6)	(3.42)	(0.11)	(4.61)	(4.7)	(5.7)	(8.8)	(0.2)
K102	2.39	2.22	197.8	10.6	0.34	15.8	10.40	2.60	12.13	15.0	39.7	45.2	0.2
	(0.10)	(0.26)	(38.3)	(2.4)	(0.03)	(30.2)	(2.77)	(0.11)	(4.56)	(5.5)	(5.0)	(7.8)	(0.4)
K28	2.43	2.06	320.7	16.6	0.37	30.2	9.45	2.38	7.49	11.2	37.0	50.7	0.9
	(0.10)	(0.24)	(50.7)	(2.6)	(0.01)	(19.6)	(3.98)	(0.19)	(0.83)	(4.5)	(4.3)	(5.6)	(1.1)
K52	2.04	2.33	319.4	14.5	0.35	8.4	8.20	2.55	5.88	30.2	42.9	26.9	0
	(0.25)	(0.19)	(65.5)	(3.4)	(0.01)	(5.9)	(2.16)	(0.11)	(1.58)	(13.5)	(5.5)	(13.7)	
K42	1.18	2.24	272.9	14.8	0.38	20.3	12.2	3.26	18.74	17.8	40.7	41.3	0
	(0.18)	(0.16)	(47.5)	(2.9)	(0.02)	(13.4)	(3.22)	(0.13)	(7.08)	(10.9)	(4.4)	(12.6)	
K101	2.16	1.62	412.7	14.2	0.31	15.5	7.25	1.70	11.40	16.5	41.8	41.6	0.1
	(0.11)	(0.11)	(62.7)	(2.1)	(0.02)	(6.9)	(2.83)	(0.16)	(5.08)	(3.8)	(4.5)	(7.0)	(0.2)
K31	2.05	2.07	378.2	16.4	0.39	28.8	7.45	1.85	5.57	25.2	40.8	33.7	0.2
	(0.16)	(0.17)	(72.3)	(3.3)	(0.03)	(23.8)	(2.11)	(0.17)	(0.91)	(9.9)	(4.5)	(9.4)	(0.6)
M	2.42	2.33	332.2	18.3	0.32	61.7	11.6	3.12	16.93	10.0	23.3	61.1	3.7
	(0.07)	(0.12)	(47.4)	(2.5)	(0.01)	(34.5)	(4.12)	(0.18)	(5.52)	(2.0)	(2.5)	(2.2)	(1.2)
K113	2.08	2.52	290.2	14.7	0.27	16.6	9.8	1.96	13.42	25.9	37.0	35.0	2.0
	(0.16)	(0.15)	(48.5)	(2.2)	(0.02)	(9.4)	(2.22)	(0.08)	(5.95)	(9.3)	(5.7)	(11.0)	(1.3)

Table 6. *Means of morphological data collected in the greenhouse in 1982 and 1983*

Populations	No. branches/ Plant	Stem Length (cm) ground to first node	Width (mm) banner petal	Flower Length (mm)	Flower Tube Length (mm)	3-seed Pod Length (cm)	Pod Pubscence Length (cm)	Leaf Pubscence density (no/mm)	10 th Central Leaflet (main stem) Length (cm)	10 th Central Leaflet (main stem) width (cm)
K109	0.40	6.81	5.63	6.78	2.94	3.16	1.13	19.50	7.00	3.00
	(0.67)	(5.88)	(0.50)	(0.41)	(0.17)	(0.11)	(0.12)	(1.60)	(0.69)	(0.28)
E4	0.45	10.25	4.97	6.78	3.03	2.46	1.47	26.21	6.70	2.69
	(0.56)	(2.48)	(0.43)	(0.48)	(6.39)	(0.13)	(0.16)	(5.41)	(0.78)	(0.41)
K9	2.88	2.36	5.06	6.31	2.50	2.48	1.50	17.00	5.14	2.57
	(1.10)	(1.22)	(0.44)	(0.48)	(0.00)	(0.10)	(0.13)	(2.60)	(0.53)	(0.25)
K7	3.10	1.79	4.97	6.75	2.38	2.80	1.50	16.2	5.80	2.63
	(1.00)	(0.65)	(0.34)	(0.45)	(0.22)	(0.08)	(0.13)	(2.20)	(0.45)	(0.32)
K102	2.77	2.58	5.03	6.22	2.94	2.60	1.44	7.40	5.45	3.07
	(1.20)	(0.72)	(0.34)	(0.31)	(0.17)	(0.09)	(0.17)	(0.90)	(0.54)	(0.32)
K28	2.79	1.71	5.72	0.69	2.50	2.98	1.00	14.70	5.15	2.90
	(1.20)	(0.65)	(0.55)	(0.40)	(0.00)	(0.09)	(0.09)	(2.10)	(0.82)	(0.43)
K52	3.17	1.24	4.34	5.56	2.34	2.64	1.47	11.01	5.14	1.92
	(0.88)	(0.75)	(0.35)	(0.40)	(0.24)	(0.09)	(0.16)	(3.62)	(0.41)	(0.19)
K42	1.58	2.65	4.81	6.56	2.47	2.60	1.51	13.30	5.69	3.11
	(1.08)	(0.75)	(0.36)	(0.44)	(0.13)	(0.12)	(0.14)	(1.70)	(0.66)	(0.41)
K101	1.98	1.57	4.47	5.69	2.06	2.34	1.71	21.51	4.90	2.12
	(0.86)	(0.71)	(0.39)	(0.25)	(0.17)	(0.10)	(0.16)	(3.10)	(0.45)	(0.21)
K31	3.70	0.99	4.47	6.00	2.25	2.41	1.50	18.20	4.11	1.99
	(1.70)	(0.35)	(0.22)	(0.0)	(0.26)	(0.09)	(0.13)	(2.30)	(0.59)	(0.29)
M	2.25	1.78	5.84	6.97	2.56	2.86	1.70	17.70	5.01	2.84
	(1.13)	(0.91)	(0.24)	(0.34)	(0.17)	(0.19)	(0.16)	(1.41)	(0.46)	(0.34)
K113	1.59	3.33	5.53	0.72	2.50	2.90	1.50	20.60	5.66	3.08
	(1.03)	(1.85)	(0.43)	(0.31)	(0.00)	(0.14)	(0.12)	(2.50)	(0.53)	(0.39)

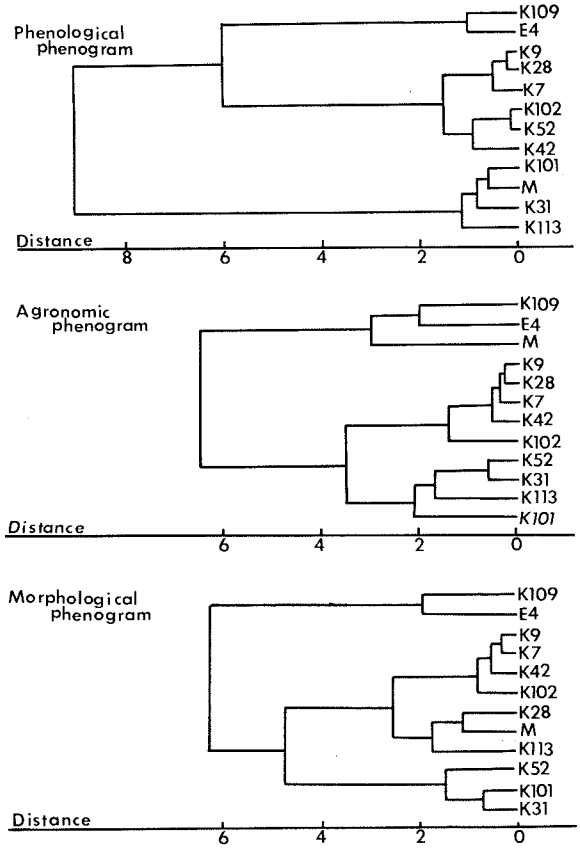

Fig. 2A. Phenograms of cluster analysis based on the phenological, agronomic and morphological data of 12 *G. soja* populations.

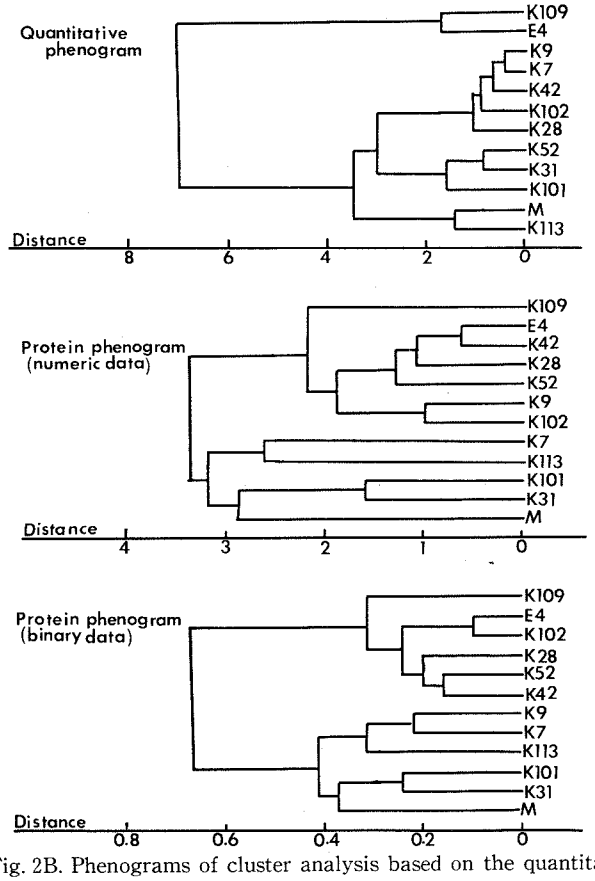

Fig. 2B. Phenograms of cluster analysis based on the quantitative, protein data of 12 *G. soja* populations.

distortion coefficient was calculated for each pair of phenograms (Farris, 1973). To calculate the coefficent each cluster of the first tree (phenogram) is taken as variable for which each population is scored as present or absent. These are referred to as tree variables. The mean coefficient of distortion is the average proportion of extra steps required for these tree variables 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.

The mean coefficients of distortion between the phenological phenogram and protein phenogram are similar; so are the mean distortion coefficients between agronomic phenogram and the protein phenogram (Table 7). However, the mean distortion coefficients between morphological phenogram and protein phenogram (numeric form) are larger than between morphological and protein phenograms (binary form). 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 natural 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 this is a phenetic, not phyletic study, the numeric form may show more precise genetics relationships among the populations than the binary form.

In general, the mean distortion coefficients

Table 7. *Mean distortion coefficients between phenograms of 12* ***G. soja*** *populations*

(A)	Based on protein (numeric form) phenogram		
	Phenogram	mean distortion coeff.	standard deviation
	Protein (binary)	0.52	0.44
	Phenological	0.72	0.39
	Agronomic	0.95	0.12
	Morphological	0.83	0.33
	Quantitative	0.98	0.08
(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	0.59	0.44
	Quantitative	0.81	0.33
(C)	Based on quantitative phenogram		
	Phenogram	mean distortion coeff.	standard deviation
	Protein (numeric)	0.82	0.20
	Protein (binary)	0.61	0.34
	Phenological	0.69	0.42
	Agronomic	0.49	0.46
	Morphological	0.25	0.41

between the two protein phenograms and among phenological, agronomic, morphological and quantitative phenograms are smaller than between protein and quantitative phenograms. The mean distortion coefficients observed in this study are large (Table 7) relative to a mean distortion coefficient of 0.0036 reported between morphometrics and allozymes phenograms of Menidia (Mickevich and Johnson, 1976).

Another way to examine the concordance between the estimates of population differentation 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 coefficients, Table 8 and Nei's D_N Table 3) of the 12 populations are listed in Table 9. The latitudinal distances among the 12 populations were also included in the correlation computation. The two methods of correlation computation produced similar results. Strong correlations were found between distance estimates based on the same protein data [D_N, Protein (numeric and binary forms)]. High correlations were also found between distance estimates based on quantitative characters (Table 9). However, the correlation coefficients between distance estimates of quantitative and protein data were relatively low and many of them were not significant. Distance estimates based on morphological characters especially showed discordance with distance estimates based on protein variations (Table 9).

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

Discussion

In general, the concordance of population differentiation among the 12 populations of *G. soja* measured by quantitative and protein variation is not high. The information collected from phenological, agronomic and morphological traits includes a large number of measurements, and each of the quantitative traits is controlled by many gene loci. In addition, phenotypes of quantitative traits are subject to environmental

Table 8. *Dissimilarity coefficient matrices of 12 populations*

(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	2.719	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

(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

Table 8. *Continued*

(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	2.470	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

Table 9. *Correlations between distances of quantitative, protein and latitude of 12* ***G. soja*** *populations*

(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.

influence. Furthermore, the wild soybean showed the plastic response to environmental variations. In contrast, the protein data were collected from qualitative traits, with trait controlled by a single gene locus. In this study only 42 protein loci were examined, which represent only a very small portion of the total soybean genome. In other words, quantitative variation included a broader representation of the *G. soja* genome than did qualitative isozyme variation. The additional explanation for the general low congruence between estimates based on protein and quantitative variations in *G. soja* populations may be: (1) Absence of, or weak association between the two sets of traits in respect to function or chromosomal position association. (2) Plastic responses of quantitative traits to some environmental variation, while the protein loci are qualitative traits and clearcut. (3) The mosaic nature of evolution in that the divergence which has occurred between populations in relation to one set of characters has not extended to other sets (Futuyma, 1986). Although the congruence in this study was generally low, 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 quantitative characters in the 12 populations (Table 9). Graef *et al.* (1989) examined the relation of isozyme genotypes to quantitative characters in the progeny of interspecific hybrid (*G. max* × *G. soja*) and found significant associations of allozyme genotypes and quantitative traits. If many more enzyme loci were included and a larger, more representative seed sample from natural populations were used in this study, the congruence between quantitative and protein data might be higher than that was observed in present study.

Acknowledgements. We would like to thank Bob Parker and Bill Given for taking care of the plants in the greenhouse. Our special thanks to Bob Parker for religiously controlling the pests with his wisdom. Without his efforts, it would not be possible to grow plants successfully. We also want to express our appreciation to Thelma Stolzenburg for typing the manuscript.

Literature Cited

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

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

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

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

Futuyma, D. J. 1986. Evolutionary Biology. 2nd edn. Sinaur Associates, Inc., Boston.

Graef, G. L., W. R. Fehr, and C. R. Cianzio. 1989. Relation of isozyme genotypes to quantitative characters in soybean. Crop Sci. **26**: 683-688.

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

Kiang, Y. T. and Y. C. Chiang. 1989. Geographic variability in the reproductive biology of wild soybean populations. *In* Jane H. Bock and Yan B. Linhart (eds.). The Evolutionary Ecology of Plants. Westview Press. Boulder, Colorado, pp. 469-489.

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.

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

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

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. U. S. A. **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.

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

Sneath, P. H. A. and R. R, Sokal. 1973. Numerical taxonomy: the principles and practice of numerical classification. *In* W. H. Freeman (ed.), & Co., San Francisco, CA.

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. Medina-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., Jr. 1963. Hierarchical grouping to optimize on objective function. J. Amer. Stat. Assoc. **58**: 236-244.

Wishart, D. 1978. CLUSTAN, User Manual. 3rd edn. Inter-University/Research Council Series, Report No. 47.

用同功酶和數量性狀的變異做野生大豆族群分化的比較

江永智　江月琴

美國新罕布什爾州立大學植物生物學系

本報告的目的是用日本和南韓的十二個野生大豆族群，根據它們的數量性狀和同功酶的變異做族群分化的比較。用數量分類化分別畫同功酶和數量性狀表型圖(Phenogram)以便比較它們的相合性。根據兩個表型圖的差異計算扭曲係數(Distortion Coefficient)。所得的扭曲係數顯示兩個表型圖的相合性不高。但值得一提的是使用如此小數目的同功酶基因座所估計的族群距離和數量性狀的距離具有明顯的相關。如增加同功酶基因座數目可能會提高相合性。對相合性不高的可能原因加以討論。