Bot. Bull. Academia Sinica (1992) **33:** 111-122

Electrophoretic and morphological variation within and among natural populations of the wild soybean, *Glycine soja* Sieb. & Zucc.

C. J. Bult[1] and Y. T. Kiang[2]

Department of Plant Biology, Program in Genetics, University of New Hampshire, Durham, N.H. 03824, USA

(*Received October 1, 1991; Accepted January 4, 1992*)

Abstract. Isozymes and morphological traits were used to explore the organization of genetic variation within and among seven natural populations of the wild soybean, *Glycine soja* Sieb. and Zucc., from central Japan. A total of 111 individuals were scored for their genotype at 49 gene loci resolved from 20 enzyme systems and one seed protein. The average values for polymorphism (P; 99% criterion), number of alleles per locus (A), gene diversity (H_{EXP}) and proportion of heterozygotes observed (H_{OBS}) were 0.14, 1.1, 0.05 and 0, respectively. Over 60% of the isozyme variation [as measured by gene differentiation (G_{ST})] was distributed among populations. A canonical discriminant analysis using allele frequencies demonstrated that differentiation among populations was due largely to differences in allele frequency rather than to the presence or absence of specific alleles. Estimates of phenotypic variation were obtained by examining 31 traits in two common garden experiments. In both cases, approximately 30% of total phenotypic variation occurred among populations. Canonical discriminant analysis demonstrated that populations have diverged most for traits related to flower size, leaf shape and yield. The levels and distributions of variation were much different for biochemical and morphological traits; however, the two data types revealed similar patterns of population differentiation.

Key words: Genetic structure; Germplasm conservation; *Glycine soja*; Natural populations; Population differentiation.

Introduction

The wild soybean, *Glycine soja* (Sieb. & Zucc.), is the putative progenitor to the cultigen, *Glycine max* (L. Merrill) (Singh and Hymowitz, 1988). Although there is much interest in the evolutionary, genetic and ecological relationships between *G. soja* and *G. max*, only a handful of studies have explored basic aspects of the population biology and ecology of wild soybeans (Hu and Wang, 1985; Kiang *et al.*, 1987; Kiang and Chiang, 1989; Kiang and Chiang, 1990).

Natural populations of the wild soybean are valuable reservoirs of genetic diversity for breeding programs of *G. max*. Wild soybean germplasm has served as a source of genes for disease resistance (Ram *et al.*, 1984), small seed size (Fehr, 1987) and increased seed protein (Erickson *et al.*, 1981). A meager 30% of the estimated available genetic diversity in wild soybean populations worldwide has been collected and characterized for germplasm conservation efforts (Plucknett *et al.*, 1987). The need for further evaluation is critical for the wild soybean as many of its natural habitats are being destroyed due to expansion of urban areas and to modern agricultural practices (Hu and Wang, 1985; Plucknett *et al.*, 1987).

Glycine soja is a twining, annual legume that presumable originated in northeast China and is distributed throughout China, Japan (except Ryukyu Islands), the Korean peninsula, the eastern Siberian regions of

[1]Present address: Laboratory of Molecular Systematics, Smithsonian Institution, Washington, D. C. 20560, USA.

[2]To whom correspondence should be addressed.

the USSR, and Taiwan (Hymowitz, 1970). Populations of *G. soja* typically are small and distributed in patches. Wild soybean plants grow in fields, under hedgerows, and along roadsides and riverbanks, often colonizing sites which have been recently disturbed. The have vine -like stems and climb on and among neighboring plants. Their leaves are pinnately trifoliate, and the leaflets may be narrowly lanceolate, ovate or oblong -elliptic. Pubescence on pods, stems and leaves generally is tawny and is strigose to hirsute. Pods are short (1 to 4 cm), black, and contain an average of three small dark-brown to black oval seeds. The flowers are typically entomophilous, although outcrossing rates in general are low, ranging from 1-2% (Kiang *et al.*, 1987).

The few studies that have been published concerning genetic structure in populations of the wild soybean have focused on populations or accessions collected over relatively broad geographic distances (Hu and Wang, 1985; Kiang *et al.*, 1987; Kiang and Chiang, 1990). Populations must be sampled over several geographic scales to enhance the understanding of the factors which influence population structure (Loveless and Hamrick, 1984). In the research described, both biochemical and morphological traits were used to explore population structure within and among seven natural populations of *G. soja* from central Japan distributed on a local (1-3 km) scale. The objectives of this research were to estimate the level of genetic variation within the populations, to evaluate genetic relationships among populations, and to describe how variation is apportioned within and among populations.

Materials and Methods

Seed Source

Seven natural populations of *Glycine soja* from Mishima, Japan, were sampled (Fig. 1). Ten to thirty seeds were hand-collected from a total of 111 mature, individual wild soybean plants. Because the plants in the field were often entangled, care was taken to ensure that seeds were collected from different individual plants even though this limited the number of plants which could be sampled at a particular site (H. I. Oka, personal communication).

Electrophoresis

Twenty enzymes and one seed protein (trypsin inhibitor) representing 49 loci were examined electrophoretically (Table 1) using the techniques of Kiang and Gorman (1983) with some modifications (Bult *et al.*, 1989). Initially, seven seeds were screened for each plant to determine its genotype. If the bands were not sufficiently resolved or a heterozygote was detected, additional seeds (progenies of field-collected seed) were examined to confirm to genotype.

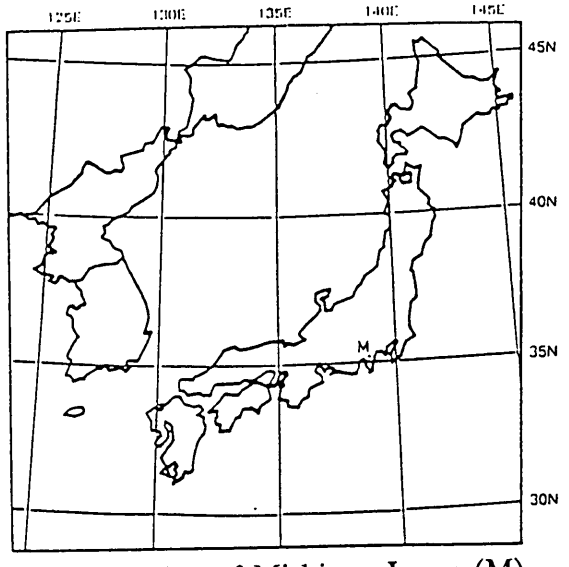
BIOSYS-1 (Swofford and Selander, 1981), a VAX -based FORTRAN computer program, was used to calculate estimates of within-population genetic diversity and population differentiation based on the isozyme data. Population structure was analyzed using Nei's (1973) gene diversity measures, taking into account the differences in population sizes. Patterns of population differentiation were explored using canonical discriminant analysis (CDA). CDA was performed using the computer program, SAS (SAS Institute, 1985). Frequencies for alleles from the variable isozyme loci were used as the data matrix.

Morphological Traits

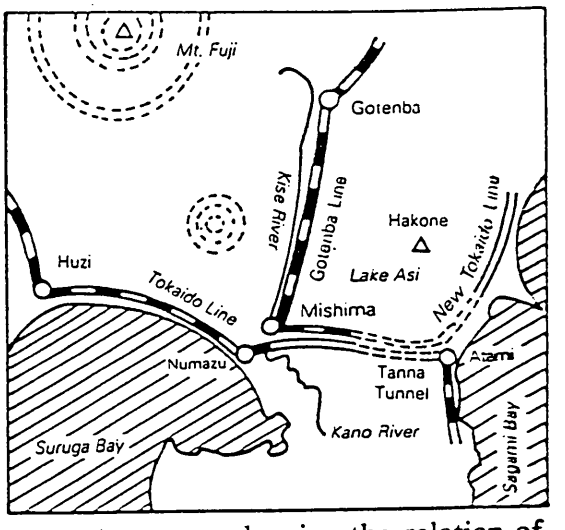
Thirty-one traits were examined for wild soybean plants grown in the greenhouse in two consecutive years. Fifteen plants per population were grown each year using a completely randomized design. Five of the 15 plants were sacrificed at 10 weeks to obtain root nodule and dry weight data. Two exceptions to this planting design were made for populations 6 (Asahigaoka-2) and 7 (Kakitagawa). Only 9 and 3 plants, respectively, were collected from these areas. For Asahigaoka-2, 14 plants were grown (5 were sacrificed at 10 weeks). For Kakitagawa, six plants were grown (3 were sacrificed at 10 weeks).

Seeds were scarified, inoculated with a commercial strain of *Rhizobium japonicum* and planted at a depth of 1 cm in 22 cm diameter clay pots. A 1:1 mixture of sterilized field soil and Promix (sphagnum moss and vermiculite) was used as the growth medium. Bamboo stakes were placed in the pots once the seedlings had germinated so that the plants could climb up the stakes and not become entangled.

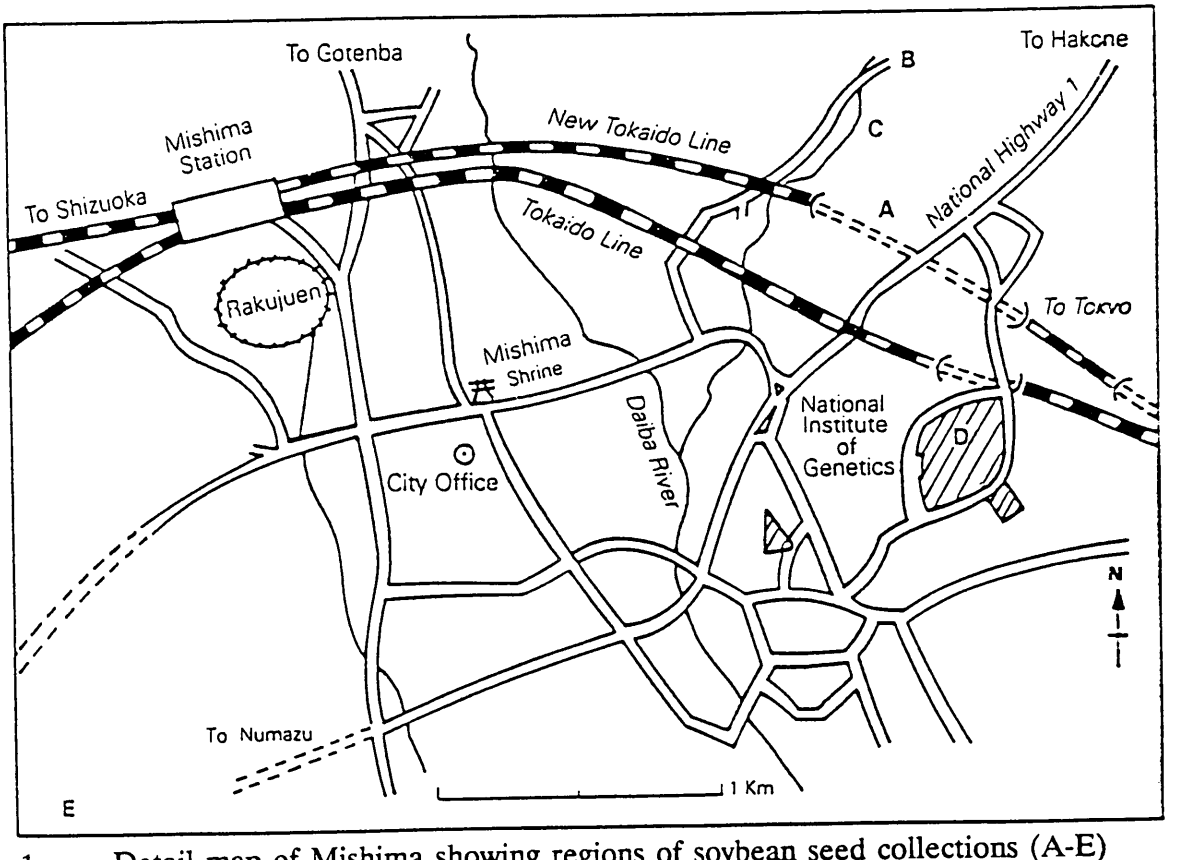
A one-way analysis of variance (ANOVA) for unbalanced data sets was performed on each of the morphological characters to test for differences among population means (SAS) Institute, 1985). Linearity, normality and homoscedasticity of error variances were assessed for all traits and data transformations were



1a. Location of Mishima, Japan (M)



1b. Area map showing the relation of Mishima to surrounding cities



1c. Detail map of Mishima showing regions of soybean seed collections (A-E)

Fig. 1a-c. Maps showing the location of Mishima, Japan and seed collection sites for the *Glycine soja* populations used in this study. Populations 1 and 5 were from region A; 2 and 6 were from region B; 3 was from region C; 4 was from region D; and 7 was from region E. 1a, Location of Mishima, Japan (M); 1b, Area map showing the relation of Mishima to surrounding cities; 1c, Detail map of Mishima showing regions of soybean seed collections (A-E).

Table 1. *Isozymes examined for* ***G. soja*** *seed*

Enzyme/protein	EC NO.	Type
1. Acid phosphatase (AP)	3.1.3.2	Hydrolase
2. Aconitase (ACO)	4.2.1.3	Lyase
3. Alcohol dehydrogenase (ADH)	1.1.1.1	Oxioreductase
4. Beta-amylase (AM)	3.2.1.2	Hydrolase
5. Diaphorase (DIA)	1.6.2.2	Oxioreductase
6. Endopeptidase (ENP)	3.4.?.?	Hydrolase
7. Esterase (EST)	3.1.1.1	Hydrolase
8. Fluorescent esterase (FLE)	3.1.1.2	Hydrolase
9. Glucose 6-phosphate dehydrogenase (GPD)	1.1.1.49	Oxioreductase
10. Glutamate oxaloacetic transaminase (GOT)	2.6.1.1	Transferase
11. Isocitrate dehydrogenase (NADP-active) (IDH)	1.1.1.42	Oxioreductase
12. Kunitz trypsin inhibitor (TI)		Seed protein
13. Leucine aminopeptidase (LAP)	3.4.11.1	Hydrolase
14. Malate dehydrogenase (MDH)	1.1.1.37	Oxioreductase
15. Mannose 6-phosphate isomerase (MPI)	5.3.1.8	Isomerase
16. Peroxidase (PER)	1.11.1.7	Oxioreductase
17. 6-Phosphogluconate dehydrogenase (PGD)	1.1.1.43	Oxioreductase
18. Phosphoglucose isomerase (PGI)	5.3.1.9	Isomerase
19. Phosphoglucomutase (PGM)	2.7.5.3	Transferase
20. Shikimate dehydrogenase (SDH)	1.1.1.25	Oxioreductase
21. Urease (EU)	3.5.1.5	Hydrolase

performed as necessary (Little and Hills, 1978). CDA was used to identify the suite of phenotypic and life-history characters which had the greatest capacity to discriminate among the soybean populations. Mahalanobis' distance (D^2) was used as a measure of the pairwise distance between groups in multivariate space based on the total covariance matrix of a data set (Tabachnik and Fidell, 1983).

Congruence Between Isozymes and Morphological Traits in Describing Population Differentiation

To test for the congruence between isozymes and morphological traits in describing population differentiation, the degree of association between Nei's (1972) genetic distance (D) and Mahalanobis' distance (D^2) was determined using Spearman rank correlation analysis (Sokal and Rohlf, 1969).

Patterns of Biochemical and Morphological Variation in Relation to Stand Density

Although data collected on environmental variables at each soybean seed collection site were limited, the percentage of the collection site occupied by *G. soja* relative to other co-habitating plant species (COVER) was estimated visually (H. I. Oka, personal communication). Relationships between COVER and the mean number of alleles per locus (A), polymorphism (P, 99% criterion), gene diversity (H_{EXP}), observed proportion of heterozygotes (H_{OBS}) and Mahalanobis' distance (D^2) were explored using Spearman rank correlation analysis (Sokal and Rohlf, 1969).

Results

Genetic Control of Isozyme Banding Patterns

The genetic bases for the electrophoretic mobility variants observed in the present study (Table 2) are understood and have been summarized elsewhere (Kiang and Gorman, 1983; Bult, 1989).

Allele Frequencies

Fifteen of the 49 isozyme loci examined were polymorphic. Thirty-one alleles were observed at the variable loci (Table 2). Within any one population, no locus had more than two alleles. While the most frequent allele at a particular locus was the same for most of the populations, some loci were marked by changes in the predominant allele occurring over short geo-

Table 2. *Frequencies for 31 alleles at 15 variable loci among seven natural populations of* ***G. soja***
The number in parantheses is the number of plants from which seeds were collected.

Locus[b]	Population[a]						
	1 (13)	2 (36)	3 (10)	4 (30)	5 (10)	6 (9)	7 (3)
Ap							
a	0	0	0	0	0.10	0	0
c	1.0	1.0	1.0	1.0	0.90	1.0	1.0
Aco2							
a	0.77	0.96	1.0	1.0	0.85	0.89	1.0
b	0.23	0.04	0	0	0.15	0.11	0
Aco4							
a	0.23	0.03	0	0	0	0	0
b	0	0	0	1.0	0	0	0
c	0.77	0.97	1.0	0	1.0	1.0	1.0
Am3							
a	0	0.78	1.0	1.0	0.80	0.61	1.0
b	1.0	0.22	0	0	0.20	0.39	0
Dial							
Dial	0.23	0.06	0.90	1.0	0.30	0.11	0.33
dial	0.77	0.94	0.10	0	0.70	0.89	0.67
Dia2							
a	0.23	0.67	1.0	1.0	0.80	0.56	1.0
b	0.77	0.33	0	0	0.20	0.44	0
Enp							
a	0.77	0.97	0	0	0.70	0.89	0.67
b	0.23	0.03	1.0	1.0	0.30	0.11	0.33
Est1							
a	0.23	0.86	0	0	0.50	0.83	0.67
b	0.77	0.14	1.0	1.0	0.50	0.17	0.33
Gpd							
Gpd	0.69	0.22	0.90	1.0	0.40	0.11	0.33
gpd	0.31	0.78	0.10	0	0.60	0.89	0.67
Idh2							
a	0	0	0	1.0	0.10	0.11	0
b	1.0	1.0	1.0	0	0.90	0.89	1.0
Lap1							
a	0.23	0.80	1.0	1.0	0.80	0.67	1.0
b	0.77	0.20	0	0	0.20	0.33	0
Mpi							
a	1.0	1.0	1.0	1.0	0.85	0.89	1.0
b	0	0	0	0	0.15	0.11	0
Pgd1							
b	0.77	1.0	1.0	1.0	0.80	0.67	1.0
c	0.23	0	0	0	0.20	0.33	0
Pgd3							
a	0	0	0	1.0	0	0	0
b	1.0	1.0	1.0	0	1.0	1.0	1.0
Pgm1							
a	1.0	1.0	1.0	0	1.0	1.0	1.0
b	0	0	0	1.0	0	0	0

[a]Populations: 1=Hatsunedai; 2=Kanodanchi; 3=Nishi-asahigaoka; 4=Yata; 5=Asahigaoka-1; 6=Asahigaoka-2; 7=Kakitagawa.
[b]Lower case letters indicate alleles at a locus are codominant; three letter designations indicate alleles are dominant/recessive.

graphic distances (Table 2). Heterogeneity contingency Chi-square tests (Workman and Niswander, 1970) demonstrated significant differences in allele frequencies at all variable loci among the seven wild soybean populations (data not shown).

Yata (population 4) clearly was different from all other populations in two respects. First, the plants from this area were monomorphic at all loci examined. Second, the alleles at several loci were either unique to the Yata population (i.e., *Aco4-b, Pgd3-a, Pgm1-b*) or were fixed in Yata, but occurred at frequencies of less than 0.50 in most of the other populations (e.g., *Idh2-a* and, to some extent, *Est1-b, Dial-b* and *Enp-b*).

Gene Diversity Within and Among Populations

The mean values of polymorphism (P) using the 99% criterion, mean number of alleles per locus (A), mean gene diversity (H_{EXP}) and proportion of heterozygotes observed (H_{OBS}) over all seven wild soybean populations were 0.14, 1.1, 0.05 and 0, respectively (Table 3). These values are consistent with the highly self-pollinating breeding system of soybeans.

Genetic Structure

Nei's gene diversity measures are given in Table 4. The mean value for total gene diversity (H_T) over fifteen variable loci was 0.31. The mean value for within-population genetic diversity (H_S) was 0.11. These two values are comparable to average population values of 0.33 and 0.15 reported for self-pollinating plant species (Hamrick and Godt, 1989). The average value for gene differentiation (G_{ST}) in this study was 0.63. G_{ST} measures the proportion of variation among populations relative to the total species' diversity (Nei, 1973). Thus, over 60% of the total genetic variation resided among populations. This value is slightly higher than the average population level G_{ST} value of 0.51 found among populations of selfing plant species (Hamrick and Godt, 1989). Loci which contributed most to G_{ST} had alleles unique to population 4.

Canonical Discriminant Analysis: Isozyme data

Four of the six canonical variates extracted from the 31-member isozyme allele data matrix were significant at the 1% level as indicated by Wilks' lambda. The eigenvalue for the first canonical variate was incalculable because the within-population variation for this variate was the value, zero. An eigenvalue is analagous to a ratio of between-class sums of squares to within-class sums of squares (Norusis, 1985). In general, a canonical variate with a large eigenvalue is one that accounts for a significant proportion of the total variance in the model attributable to between-class differences (Norusis, 1985). For the first canonical variate in the present study, all the variation was due to the between-class component. Thus, although an eigenvalue could not be computed, the variate was very important in differentiating among populations. The loci which contributed most to this function (as indicated by between canonical structure) were *Aco4, Idh2, Pgd3, Pgm1, Enp, Est1* and *Gpd*. These loci have alleles which were unique to population 4 (Yata) or which were fixed in Yata but occurred at frequencies below 50% in the other populations. The other three statistically significant canonical variates accounted for 55.7%, 37.0% and 4.8% of the total variation in the model, respectively. The loci which contributed most to the second canonical variate were *Aco2, Aco4, Am3, Lap1, Dia2* and *Enp*. The loci which contributed most to the third variate were *Aco4, Est1* and *Gpd*. Two loci, *Ap* and *Mpi*, contributed most to the fourth canonical variate.

A scattergram of the class means on the first two canonical variates is shown in Figure 2. The first canonical variate (x-axis) separated population 4 (Yata) from all other populations of wild soybean based largely on the presence of the alleles which were unique to Yata. The second canonical variate (y-axis) separated the remaining six populations based on differences in the frequencies of alleles among populations. Populations 2, 5, and 6 clustered together while 1, 3, and 7 were widely separated. As population 4 was

Table 3. *Within-population genetic diversity estimates for seven natural populations of wild soybean*
N is the number of plants from which seed were collected in each population.

Population	Sample size (N)	Mean number of alleles per locus (A)	Proportion polymorphic loci (P)	Heterozygosity[a]		Average distance to all other pops. (km)
				H_{EXP}	H_{OBS}	
1. Hatsunedai	3	1.2	0.19	0.07	0	1.1
2. Kanodanchi	36	1.2	0.19	0.04	0	1.3
3. Nishi-asahigaoka	10	1.0	0.04	0.01	0	1.1
4. Yata	30	1.0	0	0	0	1.5
5. Asahigaoka-1	10	1.3	0.26	0.09	0	1.1
6. Asahigaoka-2	9	1.3	0.23	0.07	0	1.3
7. Kakitagawa	3	1.1	0.08	0.04	0	3.4
Mean		1.1	0.140	0.05	0	1.5

[a] H_{EXP}=Nei's (1973) gene diversity; H_{OBS}=the proportion of individuals in a population that were heterozygous.

Table 4. *Partitioning of genetic variation within and among seven natural populations of wild soybean for fifteen variable loci*

Locus	Total diversity (H_T)	Within population (H_S)	Between population (D_{ST})	Gene differentiation (G_{ST})
Aco2	0.14	0.11	0.03	0.24
Aco4	0.31	0.06	0.25	0.80
Am3	0.39	0.18	0.20	0.53
Ap	0.03	0.02	0.01	0.39
Dia1	0.49	0.16	0.33	0.68
Dia2	0.37	0.25	0.12	0.32
Enp	0.49	0.13	0.36	0.74
Est1	0.49	0.16	0.33	0.67
Gpd	0.50	0.25	0.25	0.50
Idh2	0.29	0.03	0.25	0.89
Lap1	0.35	0.21	0.13	0.39
Mpi	0.07	0.04	0.03	0.45
Pgd1	0.19	0.11	0.08	0.43
Pgd3	0.24	0	0.24	1.00
Pgm1	0.24	0	0.24	1.00
Mean	0.31	0.11	0.19	0.63[a]

[a]Calculated as mean D_{ST} divided by mean H_T.

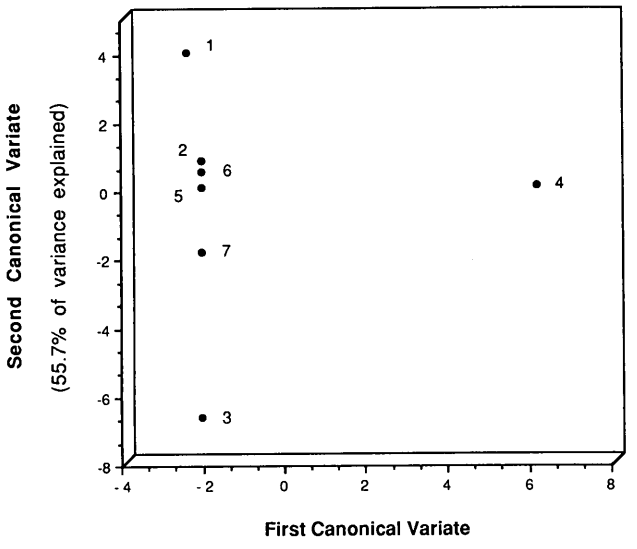


Fig. 2. Scattergram of population means on the first two canonical variates. Based on a canonical discriminant analysis of a 31-member isozyme allele frequency data matrix. The eigenvalue for the first canonical variate was incalculable because the within-population variation for this variate was the value, zero. Populations: 1=Hatsunedai; 2=Kanodanchi; 3=Nishi-asahigaoka; 4=Yata; 5=Asahigaoka-1; 6=Asahigaoka-2; 7=Kakitagawa.

completely monomorphic, its y-axis intercept was the value, zero.

Analysis of Variance (ANOVA)

Of the 31 traits examined, 21 were significantly different among *G. soja* populations in Year 1; eighteen were significantly different in Year 2 (Table 5). CTL (corolla tube length), LCL (length of central leaflet) and LLL (length of lateral leaflet) were significantly different in Year 1, but not in Year 2. As the methods of selecting and measuring morphological characters were the same for both years, the differences in the results of the ANOVA for the three traits may be due to year-to-year fluctuations in some, undetermined aspect(s) of the greenhouse environment.

The r^2 values ranged from 0.08 to 0.90 in Year 1, with an average of 0.32. In Year 2, the r^2 values ranged from 0.03 to 0.66, with an average of 0.29. Thus, about 30% of the total phenotypic variation was due to the among-population component in both years.

Canonical Discriminant Analysis: Morphological Traits

Relationships among the seven populations were explored further using canonical discriminant analysis. The final data matrices for Year 1 and Year 2

Table 5. *Results of a one-way ANOVA for 31 phenotypic and life-history traits among greenhouse-grown populations of wild soybean*

Dependent Variable[a]	Year 1		Year 2	
	r^{2b}	F-ratio	r^2	F-ratio
DTG-Days from sowing to germination	0.08	0.75	0.03	0.31
DTF-Days from germination to first flower	0.21	2.32*	0.45	7.32*
DFP-Days from first flower to first green pod	0.28	3.46**	0.26	3.17**
DDP-Days from first green pod to first dry pod	0.10	0.99	0.19	2.12
LDP-Days from first dry pod to last dry pod	0.05	0.48	0.19	2.11
LSN-Days from germination to last dry pod	0.11	1.08	0.07	0.70
B4W-No. branches at four weeks[c]	0.15	1.58	0.14	1.46
L4W-No. leaves at four weeks[c]	0.13	1.28	0.12	1.19
H4W-Height (cm) at four weeks[c]	0.27	3.33**	0.18	1.95
GH-Growth habit	0.35	4.68**	0.39	5.81**
L7W-No. of leaves at seven weeks[d]	0.14	1.40	0.10	1.01
ONO-Ovule number[e]	0.20	2.15	0.11	1.11
BPW-Banner petal width (cm)[f]	0.90	78.18**	0.51	9.36**
FL-Flower length (cm)[f]	0.79	32.90**	0.26	3.18**
CTL-Corolla tube length (cm)[f]	0.31	3.87**	0.12	1.17
LCL-Length of central leaflet (cm)[g]	0.34	4.49**	0.20	2.23
WCL-Width of central leaflet (cm)[g]	0.50	8.65**	0.55	11.06**
LRC-Ratio of length to width of central leaflet	0.51	8.83**	0.66	17.86**
LLL-Length of lateral leaflet (cm)[g]	0.34	4.49**	0.03	0.24
WLL-Width of lateral leaflet (cm)[g]	0.31	3.83**	0.33	4.46**
LRL-Ratio of length of width of lateral leaflet	0.46	7.51**	0.52	9.90**
AGW-Above-ground weight (gm)[h]	0.21	1.13	0.37	2.34
BGW-Below-ground weight (gm)[h]	0.10	0.48	0.25	1.33
NDW-Nodule dry weight (gm)[h]	0.25	1.47	0.35	2.18
NRN-No. root nodules[h]	0.19	1.04	0.11	0.48
PPP-No. pods per plant	0.40	5.95**	0.36	5.06**
P1SP-Proportion of one-seeded pods[i]	0.32	4.19**	0.32	4.28**
P2SP-Proportion of two-seeded pods	0.47	7.93**	0.43	7.34**
P3SP-Proportion of three-seeded pods	0.45	7.23**	0.49	6.93**
P4SP-Proportion of four-seeded pods	0.58	12.24**	0.39	5.82**
SWP-Seed weight per plant (gm)	0.18	1.91	0.14	1.53
SWH-Seed weight per 100 seeds (gm)	0.54	10.36**	0.53	10.06**
SPP-No. seeds per plant	0.34	5.08**	0.34	4.75**
ASP-Average no. of seeds per pod	0.33	4.43**	0.34	4.68**

[a]Plant origin (population) is the independent variable (d.f.=6).
[b]r^2 is the proportion of variance in the dependent variable which is accounted for by the model.
[c]Four weeks post-germination.
[d]Seven weeks post-germination.
[e]Averaged over 10 buds per plant.
[f]Averaged over 10 flowers per plant.
[g]Averaged over 3 leaves per plant.
[h]Ten weeks post-germination.
[i]P1SP-P4SP considered as one trait.
*and**: Difference significant at the 5% and 1% levels, respectively.

contained identical variables. Derived variables, such as leaf width/length ratios were not included in the data matrix to prevent multicollinearity and/or singularity. Variables which were measured for individuals sacrificed at 10 weeks (AGW; above ground weight, BGW; below ground weight, NDW; nodule dry weight, and NRN; number of root nodules) were eliminated due to the small number of cases for which data were available.

Year 1. Three of the canonical variates extracted from the 22-member phenotypic data set in Year 1 were significant at the 1% level as indicated by Wilks' lambda. These three variates accounted for 53.1%, 35.4% and 5.3% of the total phenotypic variation, respectively. Variables which contributed most (based on the between canonical structure) to the first canonical variate were LDP (days from first dry pod to last dry pod), WLL (width of lateral leaflet), ONO (average number of ovules), FL (flower length), CTL (corolla tube length), and SWH (seed weight per 100 seeds). Variables which contributed most to the second canonical variate were DTG (days from sowing to germination), DFP (days from first flower to first green pod), GH (growth habit), BPW (banner petal width), PPP (number of pods per plant), SPP (number of seeds per plant) and SWP (seed weight per plant). A scattergram of the class means on the first two canonical variates is shown in Figure 3.

Year 2. Two significant canonical variates were extracted from the morphological data matrix in Year 2. These variates accounted for 60.5% and 23.8% of the total phenotypic variation, respectively. The characters which contributed most to the first canonical variate were L7W (number of leaves at seven weeks), WCL (width of central leaflet), PPP (number of pods per plant), SPP (number of seeds per plant) and SWH (seed weight per 100 seeds). The traits which contributed most to the second canonical variate were BPW (banner petal width), FL (flower length) and CTL (corolla tube length). A scattergram of the class means on the first two canonical variates is shown in Figure 4.

The results for CDA for Year 1 and Year 2 are roughly similar. In both years, flower, leaf and yield characters were most effective at discriminating

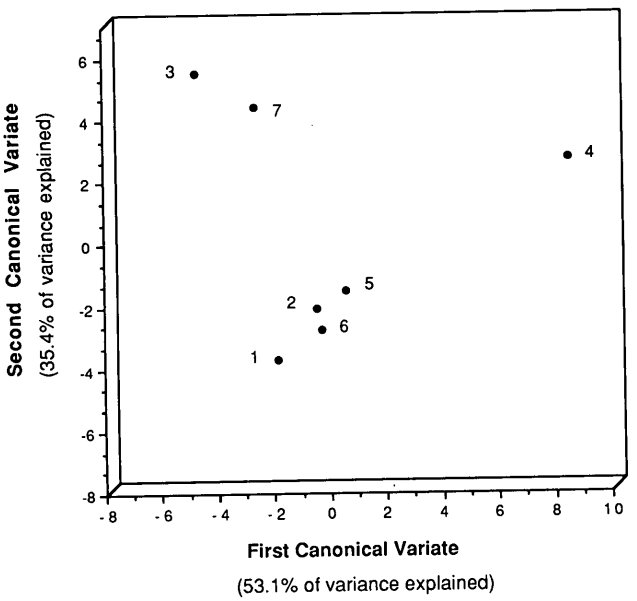


Fig. 3. Scattergram of population means on the first two canonical variates. Based on a canonical discriminant analysis of a 22-member morphological trait data matrix. Data collected in Year 1. Populations: 1=Hatsunedai; 2=Kanodanchi; 3=Nishi-asahigaoka; 4=Yata; 5=Asahigaoka-1; 6=Asahigaoka-2; 7=Kakitagawa.

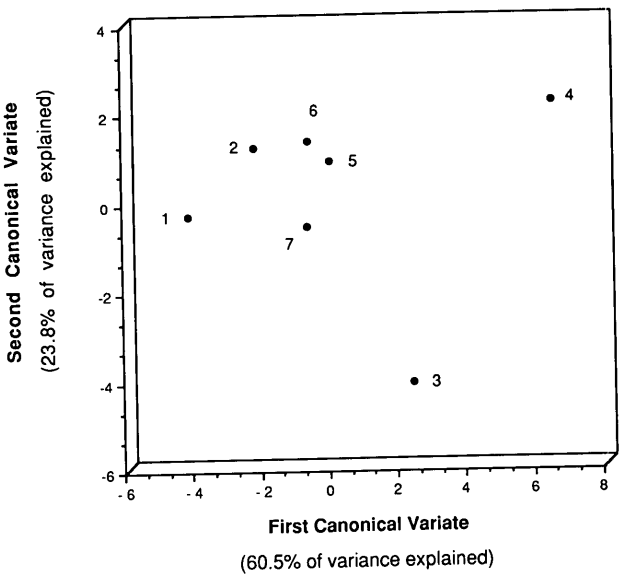


Fig. 4. Scattergram of population means on the first two canonical variates. Based on a canonical discriminant analysis of a 22-member morphological trait data matrix. Data collected in Year 2. Populations: 1=Hatsunedai; 2=Kanodanchi; 3=Nishi-asahigaoka; 4=Yata; 5=Asahigaoka-1; 6=Asahigaoka-2; 7=Kakitagawa.

among populations as indicated by their high loading values on the significant canonical variates. One major difference between the two years is that population 7 grouped with population 3 in Year 1, but with the cluster of populations 1, 2, 5, and 6 in Year 2 (Figures 3 and 4). Given that population 7 had only three representatives, the instability of its group assignment is not surprising. The year-to-year differences in overall group orientation on the two scattergrams are due to which characters loaded on the canonical variates in each year.

Congruence Between Isozymes and Morphological Traits

The degree of association between isozymes and morphological traits measured in Year 1 was relatively high, but not statistically significant. The correlation between D and D^2 was 0.61 ($p=0.14$). For Year 2, the degree of association was statistically significant ($r_s=0.85$; $p=0.02$).

Patterns of Isozyme and Morphological Variation and Stand Density

Evidence for ecotype formation in response to environmental variables was explored using Spearman rank correlation analyses. Significant (5% level) correlations were observed between COVER and A ($r_s=-0.83$), P ($r_s=-0.83$), H_{OBS} ($r_s=-0.88$) and D^2 ($r_s=0.87$ and 0.79 for Year 1 and Year 2, respectively).

Discussion

Plants within each of the seven natural populations of *G. soja* were highly uniform isoenzymatically, and differentiation among the populations was due primarily to differences in allele frequency rather than to the presence or absence of specific alleles. One exception was population 4 (Yata), in which several unique alleles were observed. The high degree of within-population isozyme uniformity and among-population differentiation is consistent with both the predominantly self-pollinating breeding system of this species and the severely restricted gene flow among its populations (Loveless and Hamrick, 1984; Hamrick and Godt, 1989).

The within-population estimates of genetic variation for the seven populations of *G. soja* examined in this study are lower than those reported in previous studies of natural populations and accessions (Kiang *et al.*, 1987; Kiang and Chiang, 1990). Kiang and Chiang (1990) reported P, A and H_{EXP} values of 0.38, 1.6 and 0.11 for wild soybeans from four natural populations collected along the Kitakami River in the northern region of the Honshu Island, Japan. Kiang and coworkers (1987) reported overall genetic diversity values of 0.62, 2.0 and 0.15, respectively, for wild soybean accessions. The differences between these two studies and the one reported here are indicative of the restricted sampling area and small sizes of the soybean populations examined in this study compared to the others. In the present study, three to thirty-six individuals per population from relatively small populations were collected, restricting the proportion of the total genetic variation which could be represented in any one sample. Also, the average distance between collection sites was only 1.5 km. In contrast, Kiang and Chiang (1990) collected seeds over a total distance of 120 km (north to south), and the accessions examined by Kiang *et al.* (1987) represented worldwide collections of *G. soja*. Sample size in both instances were larger than in the present study. Thus, these researchers had a greater chance of including more genetic variation (including rare alleles) in their germplasm collections. The pattern of genetic diversity decreasing as geographic scale decreases is consistent with the idea that populations of self-pollinating plants are structure on a fine spatial scale, allowing for storage of large amounts of genetic diversity in the species as a whole (Allard *et al.*, 1968).

In contrast to the uniformity in isozymes the soybean phenotype was highly variable within populations. Phenotypic plasticity is an important "buffering" mechanism for individual survival in the face of environmental variation when a population has a uniform gene pool (Moran and Marshall, 1978; Wu and Jain, 1978; Schlichting, 1986). Because of the limited number of seeds available, phenotypic plasticity could not be assessed quantitatively in the present study. As the choice of individuals from populations and arrangement in the common greenhouse environment was randomized, we have assumed that the measurements of morphological characters in the present study are representative. However, differences observed among populations grown in a common garden regime may be masked in the field if environmental heterogeneity is high (Schwaegerle *et al.*, 1986).

Although the levels and distributions of variation

differed for isozymes and morphological traits, the patterns of differentiation among the populations revealed by canonical discriminant analysis were similar for both data types (Figs. 2-4). It is puzzling that the concordance of the two data types was not consistent statistically across both years of the study. This likely was due to a number of factors including, (1) the high levels of within-population phenotypic plasticity in the wild soybean (Bult, 1989; Kiang and Chiang, 1990), (2) small sample sizes used for measuring morphological character variation in the present study, (3) the small number of loci represented by isozymes versus the number of loci represented by the morphological traits (Kiang and Chiang, 1990) and (4) differences in the levels of variation in biochemical and morphological traits (Lewontin, 1984).

The patterns of population differentiation indicated by the multivariate analyses of biochemical and morphological traits suggest that a collection scheme which emphasizes taking small samples from many populations of the wild soybean would be the most effective method to ensure a representative amount of genetic diversity in a germplasm collection (Brown, 1978).

Spearman rank correlation analyses revealed that the percentage of the collection site occupied by *G. soja* relative to other co-habitating plant species was significantly correlated with genetic diversity (A, P and H_{OBS}) and phenotypic diversity (D^2) estimates. Thus, inter- and intra-specific plant associations may have played a role in shaping the genetic structure of the seven populations. Stand composition and density have been indicated as a significant contributing factor to population structure in the annual grass, *Anthoxanthum* (Kiang, 1982; Antonovics *et al.*, 1988) and merit investigation in the wild soybean.

Acknowledgements. The authors thank Drs. H. I. Oka and H. Morishima for their generous donation of the wild soybean seed used in this study. This work was made possible in part by a UNH Dissertation Fellowship and a Central University Research Fund Grant (1000UFRCU6D) to C. J. B. Scientific Contribution No. 1698 from the New Hampshire Agricultural Experiment Station.

Literature Cited

Allard, R. W., S. K. Jain, and P. L. Workman. 1968. The genetics of inbreeding populations. Adv. Genet. **14**: 55-131.

Antonovics, J., N. C. Ellstrand, and R. N. Brandon. 1988. Genetic variation and environmental variation: expectations and experiments. *In* L. Gottlieb and S. K. Jain (eds.), Plant Evolutionary Biology. Chapman and Hall, New York, pp. 275-303.

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

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

Bult, C. J. 1989. Isozyme and quantitative trait variation within and among natural populations of the wild soybean, *Glycine soja* (Sieb. and Zucc.). Ph. D. Dissertation. University of New Hampshire, Durham.

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

Fehr, W. G. 1987. Soybean. *In* W. G. Fehr (ed.), Principles of Cultivar Development, vol. 2. MacMillan Publishers, New York, pp. 533-576.

Hamrick, J. L. and M. J. W. Godt. 1989. Allozyme diversity in plant species. *In* A. H. D. Brown, M. T. Clegg, A. L. Kahler, and B. S. Weir (eds.), Plant Population Genetics Breeding and Genetic Resources. Sinauer Associates, Sunderland, Massachusetts, pp. 3-63.

Hymowitz, T. A. 1970. On the domestication of the soybean. Econ. Bot. **24**: 408-421.

Kiang, Y. T. 1982. Local differentiation of *Anthoxanthum odoratum* L. populations on roadsides. Am. Midl. Nat. **107**: 340-350.

Kiang, Y. T. and M. B. Gorman. 1983. Soybean. *In* S. D. Tanksley and T. J. Orton (eds.), Isozymes in Plant Genetics and Breeding, part B. Elsevier Science Publishers, New York, pp. 295-238.

Kiang Y. T., Y. C. Chiang, H. J. Y. Doong, and M. B. Gorman. 1987. Genetic variation of soybean germplasm. *In* Crop Exploration and Utilization of Genetic Resources. Academia Sinica Press, Taipei, Taiwan, pp. 93-99.

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

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

Lewontin, R. C. 1984. Detecting population differences in quantitative characters as opposed to gene frequencies. Am. Nat. **123**: 115-124.

Little, T. M. and F. J. Hills. 1978. Agricultural Experimentation: Design and Analysis. J. Wiley and Sons, New York.

Loveless, M. D. and J. L. Hamrick. 1984. Ecological determinants

of genetic structure in plant populations. Ann. Rev. Ecol. Syst. **15**: 65-95.

Moran, G. G. and D. R. Marshall. 1978. Allozyme uniformity within and variation between races of the colonizing species *Xanthium strumarium* L. (Noogoura Burr.) Aust. J. Biol. Sci. **31**: 283-291.

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.

Norusis, M. J. 1985. SPSS-X: Advanced statistics guide. McGraw-Hill Book Co., New York.

Plucknett, D. L., N. J. H. Smith, J. T. Williams, and N. M. Anishetty. 1987. Gene Banks and the World's Food. Princeton University Press, Princeton.

Ram, H. H., K. S. Pushpendara, 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.

SAS Institute. 1985. SAS User's Guide: Statistics, version 5 edn. SAS Institute, Cary, North Carolina.

Schlichting, C. 1986. The evolution of phenotypic plasticity in plants. Ann. Rev. Ecol. Syst. **17**: 667-693.

Schwaegerle, K., K. Garbutt, and F. A. Bazzaz. 1986. Differentiation among nine populations of Phlox. I. Electrophoretic and quantitative variation. Evolution **40**: 506-517.

Singh, R. I. and T. Hymowitz. 1988. The genomic relationship between *Glycine max* (L.) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. Theor. Appl. Genet. **76**: 705-711.

Sokal, R. R. and F. J. Rohlf. 1969. Biometry: The Principles and Practice of Statistics in Biological Research. W. H. Freeman and Company, New York.

Swofford, D. L. and R. B. Selander. 1981. BIOSYS-1: A FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. J. Hered. **72**: 281-283.

Tabachnik, B. G. and L. S. Fidell. 1983. Using Multivariate Statistics. Harper and Row, Publishers, New York.

Workman, P. L. and J. D. Niswander. 1970. Population studies on Southwestern Indian tribes. II. Local genetic differentiation in the Papago. Amer. J. Human Genet. **22**: 24-49.

Wu, K. K. and S. K. Jain. 1978. Genetic and plastic responses in geographic differentiation of *Bromus rubens* populations. Can. J. Bot. **56**: 873-879.

Hu, Z. A. and H. X. Wang. 1985. Genetic structure of natural populations of wild soybean (*Glycine soja*) in Beijing region. Acta. Bot. Sin. **27**: 599-604.

野生大豆族群間和族群內的遺傳結構

C. J. Bult and 江永智

Department of Plant Biology, Program in Genetics,
University of New Hampshire, Durham, NH 03824, USA

使用同功異構酶和形態變異探討野生大豆族群內與族群間的遺傳結構。總共檢查 11 株，20 種同功異構酶，包括 49 基因座。異型結合體有 0.14，平均每基因座有 1.1 基因(alleles)，實際觀察異型結合體爲 0.05。平均多型態爲 0.14，每一基因座基因數爲 1.1。基因變異期望值爲 0.05，異型結合體爲零。高於 60%的異構酶的變異分佈於族群間。用 canonical discriminant 分析 allele 的頻率顯示族群間的分化大部份由於基因頻率的差異而非由於 allele 的存在與否。野生大豆的植株種在同一園地，以便觀察其形態變異度。分析顯示約有 30%的表型變異。這些族群的分歧大多和花的大小，葉型和種子產量有關。變異的分佈和水準與生化和形態特性有很大的差異，但是這兩種觀察的數據顯示這些族群具有相似的分化。