Genetic Diversity in Cultivated Common Bean: I. Allozymes

Shree P. Singh, R. Nodari, and P. Gepts*

ABSTRACT

Previous studies using phaseolin seed protein as a marker have revealed that cultivated common bean (*Phaseolus vulgaris* L.) resulted from multiple domestications in Mesoamerica and in Andean South America. Because these studies were based on variation at a single locus, confirmation was sought by analyzing patterns of diversity at nine polymorphic allozyme loci, all unlinked to the phaseolin locus: ribulose bisphosphate carboxylase, shikimate dehydrogenase, cathodal peroxidase, malic enzyme, malate dehydrogenase (two loci), diaphorase (two loci), and leucine aminopeptidase. A total of 227 landraces representing geographical regions from Mexico to Argentina and Chile were analyzed for these enzyme systems. A crude homogenate of primary leaf or root tissue (depending on enzymes assayed) from five seedlings of each landrace grown in vermiculite was used for starch gel electrophoresis. A cluster analysis based on Nei's genetic distance (D) was performed according to the unweighted paired group method of Sneath and Sokal. Our results confirm the existence of two major groups in cultivated common bean, Mesoamerican vs. Andean American; provide indications of gene flow from wild to cultivated beans; and suggest at least five subgroups within Mesoamerican and four within Andean cultivar groups.

STUDIES IN CROP EVOLUTION have traditionally attempted to determine the crop's ancestry based on morphological similarities and the production of viable and fertile hybrids between wild ancestor and cultivated descendant. Recently, the use of molecular markers has helped identification of the actual ancestral populations in maize (*Zea mays* L.) (Doebley et al., 1987) and pea (*Pisum sativum* L.) (Palmer et al., 1985), and determination of the effect of domestication on genetic diversity in barley (*Hordeum vulgare* L.) (Clegg et al., 1984; Jana and Pietrzak, 1988) and pearl millet (*Pennisetum glaucum* [L.] R. Br.) (Gepts and Clegg, 1989).

The common bean is an annual, diploid ($2n = 2x = 22$) species that originated in the Americas and consists of wild and cultivated forms. The wild forms are distributed from northern Mexico to northern Argentina (Brücher, 1988; Delgado Salinas et al., 1988). They are morphologically similar to the cultivars and yield viable and fertile progenies when crossed with them (Brücher, 1988; Gentry, 1969). Because of the extensive distribution of the common bean's wild ancestor, the exact siting of its domestication has been subject to speculation. Phaseolin diversity data suggest that cultivated common bean arose from multiple domestications along this extended distribution (Gepts and Bliss, 1986; Gepts et al., 1986). In particular, two major domestications appear to have given rise to Mesoamerican and southern Andean cultivars, respectively. A third and minor domestication may have taken place in Colombia or Central America (Gepts and Bliss, 1986; Koenig et al., 1990).

Allozyme analyses of wild forms have confirmed the existence of these two major groups and clarified their geographic boundaries: the Mesoamerican forms include wild populations from northern Mexico to Colombia, whereas the Andean types include populations from Peru and Argentina (Koenig and Gepts, 1989b).

In this article, we report on allozyme analyses of 227 cultivated landraces representing a geographical distribution extending from Mexico to Argentina and Chile. The objectives were to determine whether cultivated common bean displays a geographic pattern of allozyme diversity similar to its wild ancestors and to identify subgroups within the large groups of Mesoamerican and Andean cultivars.

MATERIALS AND METHODS

The collection of 227 landraces analyzed in this study was obtained from the *Phaseolus* world collection at the International Center for Tropical Agriculture (CIAT), Cali, Colombia, and the Western Regional Plant Introduction Station, Pullman, WA. It is representative of major geographical and ecological regions in the area of origin of common bean, which extends from Mexico to Argentina. Among Mesoamerican and Caribbean materials, 66 originated in Mexico, 7 in Guatemala, 5 in El Salvador, 2 each in Nicaragua and Costa Rica, and 1 in the Dominican Republic. Among Andean South American materials, 54 landraces came from Colombia, 36 from Ecuador, 19 from Peru, 14 from Chile, 3 from Bolivia, and 1 from Argentina. From lowland South America, 17 Brazilian landraces were included. The wild *P. vulgaris* accessions included in the cluster analysis were those studied by Koenig and Gepts (1989b).

Allozyme analyses of a crude homogenate of primary leaf or root tissue (depending on the enzyme system assayed) from an average of five seedlings (10 d old) grown in vermiculite were performed as described by Koenig and Gepts (1989b). A total of nine enzyme systems showing polymorphism were assayed: ribulose bisphosphate carboxylase (small subunit; RBCS), shikimate dehydrogenase (SKDH), peroxidase (PRX), malic enzyme (ME), malate dehydrogenase (MDH), diaphorase (DIAP), and leucine aminopeptidase (LAP). The MDH and DIAP enzyme systems each had two independent loci. Seven additional enzyme systems did not reveal polymorphisms among cultivars: aspartate aminotransferase, fructokinase, glucose-6-phosphate dehydrogenase, glucose phosphate isomerase, peptidase, 6-phosphogluconate dehydrogenase, and triose phosphate isomerase. Allozyme loci and alleles were designated as in Koenig and Gepts (1989a). In the nine systems studied, the most common allele was designated 100 and all other allozymes were measured in millimeters from the standard.

In order to compare allozyme diversity in the wild ancestor and the cultivated descendant, data from cultivated landraces were analyzed jointly with data from 83 wild forms obtained earlier (Koenig and Gepts, 1989b). A dendrogram based on Nei's (1973) genetic distance (D) was constructed according to the unweighted paired group method of Sneath and Sokal (1973) with a statistical package provided by Dr. K. Ritland, Univ. of Toronto.

S.P. Singh, Centro Int. de Agric. Tropical (CIAT), Apartado Aereo 6713, Cali, Colombia; R. Nodari and P. Gepts, Dep. of Agron. and Range Sci., Univ of California, Davis, CA 95616. Received 15 Dec. 1989. *Corresponding author.

Published in Crop Sci. 31:19–23 (1991).

RESULTS AND DISCUSSION

The genetic control of the polymorphism for enzyme systems RBCS, SKDH, PRX, ME, MDH, DIAP, and LAP in common bean has been determined previously (Weeden, 1984, 1986; Koenig and Gepts, 1989a). In our study, polymorphism for five enzyme systems each originated at a single locus (the number of alleles is given in parentheses): *Rbcs* (2), *Skdh* (2), *Prx* (2), *Me* (3), and *Lap-3* (2); for the remaining two systems, two loci were involved: *Mdh-1* (2) and *Mdh-2* (2), and *Diap-1* (2) and *Diap-2* (2). The average number of alleles per polymorphic locus was therefore 2.1.

A total of 76 genotypes, each with its distinctive allele combination at the nine isozyme loci, were identified among the 83 wild forms (Koenig and Gepts, 1989b) and 227 cultivars studied (Fig. 1). Information on the composition of the allozyme genotypes (allelic combination, wild and cultivated accessions) can be

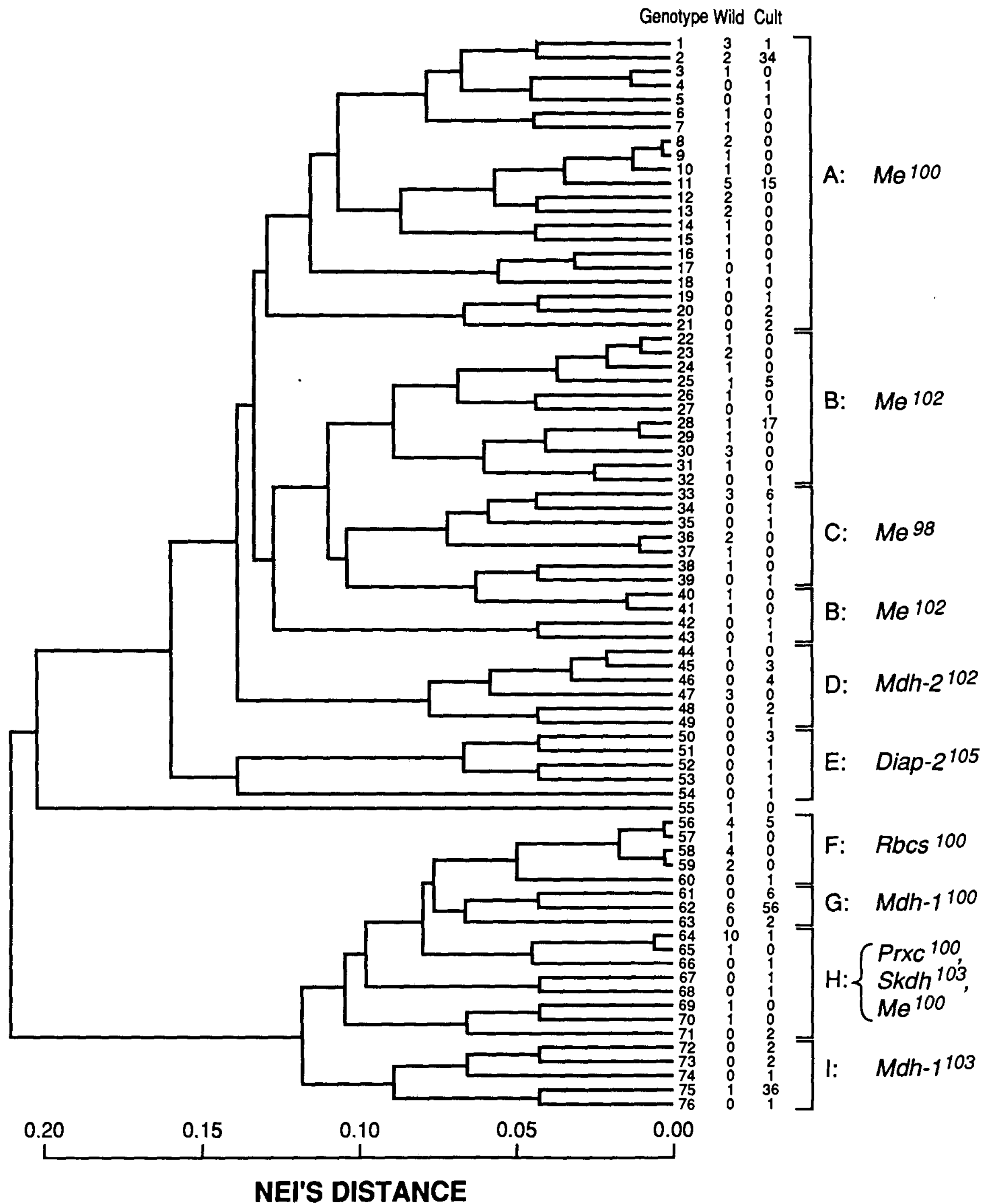

Fig. 1. Dendrogram of allozyme diversity in *Phaseolus vulgaris*. To the right of the consecutive genotype number is the number of wild and cultivated (cult) accessions for each genotype. Letters A to I indicate clusters of accessions sharing a common allozyme (as indicated to the right of the cluster letter).

obtained from the corresponding author. Because several allelic combinations were observed in more than one cultivar or wild population, only one cultivar and wild population per allelic combination were included in the cluster analysis (Fig. 1).

The cluster analysis based on Nei's (1973) genetic distance revealed the existence of two major groups of cultivars (Fig. 1). These two groups exhibited a differential geographic distribution. Most of the cultivars in the upper group (1 to 54) of Fig. 1 originated in Mesoamerica, while the cultivars of the lower group (56 to 76) originated in the Andes. Previous studies on the organization of genetic diversity at the molecular level in *P. vulgaris* had used phaseolin seed protein in wild populations and cultivars (Gepts and Bliss, 1985, 1986; Gepts et al., 1986; Koenig et al., 1990) and allozymes in wild populations (Koenig and Gepts, 1989b) and had provided evidence that the genetic diversity of *P. vulgaris* is organized into two major groups: Mesoamerican vs. Andean forms. As illustrated in Fig. 1, our allozyme data on cultivated landraces also distinguished two major groups: Mesoamerican vs. Andean. The allozyme profile of each of these two groups matched that of the corresponding wild populations in the same area (Koenig and Gepts, 1989b), providing additional evidence favoring independent origins for these two groups of cultivars. As had been observed in wild bean by Koenig and Gepts (1989b), cultivars with Type S, Sd, Sb, or B phaseolin had a Mesoamerican allozyme profile, whereas Type T, C, and H phaseolin cultivars had an Andean profile. The divergence between Mesoamerican and Andean landraces is therefore a consequence of the divergence between Mesoamerican and Andean wild ancestors, which predates the domestication process.

Table 1. Clusters of cultivated *Phaseolus vulgaris* from Latin America based on a characteristic common allozyme.

Cluster†	Genotype number†	Characteristic allozyme‡	Representative cultivars
		Mesoamerican	
A	1–21	Me^{100}	Puebla 152, Apetito, Flor de Mayo, Conejo, Cacahuate criollo, Frijola, Jalisco 20
B	22–32, 40–43	Me^{102}	G 2618, Pinto, Durango 222, Bayo Río Grande, Garrapato
C	33–39	Me^{98}	Rabia de Gato, Negro Argel, Pitouco
D	44–49	$Mdh\text{-}2^{102}$	Flor de Mayo, Rosa de Castilla, Naranja coral, Ecuador 299
E	50–54	$Diap\text{-}2^{105}$	Porillo Sintético, Rio Tibagi, H6 Mulatinho
		Andean	
F	56–60	$Rbcs^{100}$	Almidón, Balín d'Albenga, Bola 60 días
G	61–63	$Mdh\text{-}1^{100}$	Coscorrón, Frutilla, Burrito, Tórtola Los Angeles, Estrada Rosado, Cargamanto, Sangretoro, Algarrobo
H	64–66,	Prx^{100}	Blanco Sabanero
	67–68	$Skdh^{103}$	Pajaritos, Huevo de Pinche
	69–71	Me^{100}	Bagajo I, Huila 9
I	72–76	$Mdh\text{-}1^{103}$	Mortiño, Nuña, Bolón rojo, Bola roja, Bolón bayo

† Clustering and genotypes from Fig. 1.
‡ Diap = diaphorase, Me = malic enzyme; Mdh = malate dehydrogenase; Prx = peroxidase; Rbcs = ribulose bisphosphate carboxylase (small subunit); and Skdh = Shikimate dehydrogenase. Hyphenated numbers indicate locus; superscripted numbers indicate alleles.

At the molecular level, cultivated types of a major group (such as the Mesoamerican group) are more closely related to the wild ancestor of that same group than to cultivars of the other group (in this example, the Andean group).

Of 76 genotypes identified through analyses of allozyme data thus far in *P. vulgaris*, 34 were present only among wild populations and 32 were found among cultivars but not in wild forms; only 10 genotypes were found in both wild and cultivated germplasm (Fig. 1). Thus, the reduction in diversity observed upon domestication is not as pronounced for allozymes as for phaseolin (Gepts et al., 1986). This may be attributed to the lower level of total allozyme diversity present among wild bean at any given locus when compared to phaseolin diversity. For example, some 15 to 20 phaseolin patterns were observed among Mexican wild bean populations (Gepts et al., 1986), whereas only zero to four alleles per allozyme locus were observed among the same wild germplasm (Koenig and Gepts, 1989b).

Moreover, no definite trend could be established regarding increase or decrease in number of alleles per allozyme locus in wild vs. cultivated germplasm. For example, *Diap-1* displayed only two alleles ($Diap\text{-}1^{100}$ and $Diap\text{-}1^{95}$) in cultivars vs. four alleles ($Diap\text{-}1^{102}$, $Diap\text{-}1^{100}$, $Diap\text{-}1^{96}$, and $Diap\text{-}1^{95}$) in wild materials. On the other hand, *Diap-2* showed two alleles ($Diap\text{-}2^{105}$ and $Diap\text{-}2^{100}$) in cultivars and only one ($Diap\text{-}2^{100}$) in wild populations. The absence of the $Diap\text{-}2^{105}$ allele may be attributed to an insufficient sampling of the Mesoamerican wild populations, especially in Central America, from which most of the cultivated landraces carrying $Diap\text{-}2^{105}$ originated. If, however, the presence of $Diap\text{-}2^{105}$ among wild beans is confirmed, cultivars of Cluster E (see Fig. 1 and Table 1) would have evolved either as direct descendants or would represent introgressants from those wild beans. On the contrary, the $Diap\text{-}2^{105}$ allele could be the result of a mutation during or after domestication of this cultivar genotype. Allozymes of 34 genotypes found only in wild populations but absent in cultivars are likely due to inadequate sampling of the latter or else may suggest that all wild bean populations did not participate in the domestication and evolution of cultivars.

Landraces in Cluster D (Fig. 1 and Table 1) exhibited the $Mdh\text{-}2^{102}$ allele, which was observed previously only among Mesoamerican wild beans (Koenig and Gepts, 1989b). The absence of this allele among Andean cultivars and the most predominant clusters (A and B) of the Mesoamerican landraces suggests that it may have been introduced from Mesoamerican wild beans by occasional gene flow. Although common bean is a predominantly self-pollinated species, outcrossing rates of up to 70% have recently been reported (Wells et al., 1988). Outcrossing rates of much lower magnitude would, however, be sufficient to assure gene flow from wild to cultivated forms.

The Andean and Mesoamerican landraces exhibited contrasting alleles at the *Diap-1*, *Lap-3*, *Me*, *Prx*, *Rbcs*, and *Skdh* loci (Table 2). Alleles at the same or other loci distinguished subgroups within cultivars from the two major domestication centers. For example, approximately one-third of the Mesoamerican acces-

Table 2. Allozyme distribution in Mesoamerican and Andean cultivated *Phaseolus vulgaris*.

Geographic origin†	Diap-1		Diap-2		Lap-3		Me			Mdh-1		Mdh-2		Prx		Rbcs		Skdh	
	100	95	105	100	103	100	102	100	98	103	100	103	100	100	98	100	98	103	100
	%																		
Mesoamerican (83)	14	82	4	96	5	95	29	60	11	1	99	8	92	68	32	96	4	87	13
Andean (127)	94	6	0	100	98	2	1	3	96	30	70	0	100	11	89	16	84	11	89
Brazil (17)	29	71	6	94	29	71	12	47	41	0	100	0	100	59	41	71	29	71	29

Allozymes and alleles‡

† Number of genotypes is given in parentheses.
‡ Diap = diaphorase, Lap = leucine aminopeptidase; Me = malic enzyme; Mdh = malate dehydrogenase; Prx = peroxidase; Rbcs = ribulose bisphosphate carboxylase (small subunit); and Skdh = Shikimate dehydrogenase. Hyphenated numbers identify the locus; numbers below the allozyme identify the alleles at that locus.

sions carried the Me^{102} allele, whereas the other two-thirds exhibited the Me^{100} allele, and 96% of the Andean landraces had the Me^{98} allele. Similarly, one-third of the Andean accessions showed the $Mdh\text{-}1^{103}$ allele, whereas the other two-thirds of the Andean accessions and most of the Mesoamerican accessions showed the $Mdh\text{-}1^{100}$ allele. Thus, closer inspection of Fig. 1 reveals that, within the Mesoamerican and Andean cultivated germplasm, clusters of landraces can be identified that share a common allozyme and, therefore, can presumably be traced to a common ancestor (with exceptions due to independent mutations or outcrossing). The common ancestry hypothesis accounts for the inclusion of genotypes 40 to 43 in Group B (Fig. 1). Table 1 provides the summary of five cluster groups in Mesoamerican cultivars and four cluster groups in Andean cultivars, each with their characteristic allozyme and representative cultivars. For example, in Mesoamerica, Cluster A, comprising genotypes 1 to 21, is characterized by the Me^{100} allele. Similarly, Cluster B, characterized by the Me^{102} allele, includes genotypes 22 to 32 and 40 to 43. Clusters F and H, characterized by the $Rbcs^{100}$ allele and the Prx^{100}, $Skdh^{103}$, and Me^{100} alleles, respectively, were considered to be distinct from the G and I clusters because their characteristic alleles are representative of Mesoamerican genotypes (Table 2). Landraces belonging to Clusters F and H may therefore represent hybrids between the Andean and Mesoamerican groups. The relationship between these clusters and morphological and agronomic variation is investigated in the companion paper (Singh et al., 1991).

One striking feature of allozyme diversity among cultivars is the existence of a limited number of genotypes represented by a large number of landraces and which differ from other genotypes at one or more allozyme loci: e.g., genotypes 2, 62, and 75, with 34, 56, and 36 cultivated representatives, respectively (Fig. 1). Whereas cultivars with the same allozyme genotype show similarities for certain morphological traits (see companion paper, this issue), they can exhibit considerable diversity for other morphological traits such as seed type (size, shape, color, and color pattern; data not shown). This observation suggests that most of the cultivars with the same allozyme genotype, following their origination from a common ancestor, have undergone further diversification for morphological traits but not for molecular markers. Exceptions to this evolutionary pattern could be due to independent mutations or occasional outcrosses. In plants, it has been suggested that certain morphological traits, notably those involving the presence or absence of certain organs, are also coded by a small number of genes or even a single gene (Knight, 1948; Hilu, 1983; Gottlieb, 1984). In common bean, traits that distinguish wild vs. cultivated beans and cultivated beans among them, are also coded by few genes (with the exception of seed size, which is a multigenic trait) (Gepts, 1990). Major phenotypic changes can therefore, have a simple genetic control and occur without major divergence at the molecular level.

This contrast in diversity levels between morphological traits and biochemical markers raises the issue as to which trait should be used to assess genetic diversity. A higher plant contains several tens of thousands of genes (Kamalay and Goldberg, 1980). Most of these genes, including allozyme and phaseolin genes, do not have a strong effect on the phenotype and are most likely selectively neutral (Kimura, 1983); this distinguishes them from genes determining the morphology of cultivars (Gepts, 1990; Gepts and Debouck, 1991). Consequently, they are less likely to be subject to selection pressures; they could, however, be subject to genetic drift, which would ultimately lead to fixation. Because of their strong phenotypic effect and the high heritability of the traits they encode, genes for morphological traits constitute a sample of genes that is atypical of the large majority of genes in a genome. Their gene action, however, makes these traits highly responsive to selection during and after domestication, which accounts for their high levels of diversity in the primary cultivated gene pool as a whole as opposed to molecular markers. Allozymes and phaseolin offer the additional advantage that they can reveal genetic relationships among accessions. Based on these arguments, molecular markers may be more accurate descriptors of the overall levels of genetic diversity in a genome than morphological markers.

ACKNOWLEDGMENTS

We thank Dr. R. Kesseli for his assistance in the genetic diversity analysis and G. Arndt for useful comments. This research at the University of California, Davis, was supported by the International Center for Tropical Agriculture (CIAT), Cali, Colombia, and the U.S. Agency for International Development.

REFERENCES

Brücher, H. 1988. The wild ancestor of *Phaseolus vulgaris* in South America. p. 185–214. *In* P. Gepts (ed.) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands.
Clegg, M.T., A.H.D. Brown, and P.R. Whitfeld. 1984. Chloroplast

DNA diversity in wild and cultivated barley: Implications for genetic conservation. Genet. Res. 43:339–343.

Delgado Salinas, A., A. Bonet, and P. Gepts. 1988. The wild relative of *Phaseolus vulgaris* in Middle America. p 163–184. *In* P. Gepts (ed.) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, the Netherlands.

Doebley, J., W. Renfroe, and A. Blanton. 1987. Restriction site variation in the Zea chloroplast genome. Genetics 117:139–147.

Gentry, H.S. 1969. Origin of the common bean, *Phaseolus vulgaris*. Econ. Bot. 23:55–69.

Gepts, P. 1990. Biochemical evidence bearing on the domestication of *Phaseolus* beans. Econ. Bot. 44(3 Suppl.):28–38.

Gepts, P., and F.A. Bliss. 1985. F1 hybrid weakness in the common bean: Differential geographic origin suggests two gene pools in cultivated bean germplasm. J. Hered. 76:447–450.

Gepts, P., and F.A. Bliss. 1986. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. 40:469–478.

Gepts, P., and M.T. Clegg. 1989. Genetic diversity in pearl millet (*Pennisetum glaucum* [L.] R. Br.) at the DNA sequence level. J. Hered. 80:203–208.

Gepts, P.A., and D.G. Debouck. 1991. Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.). *In* A. van Schoonhoven and O. Voysest (ed.) Common beans: Research for crop improvement. CIAT, Cali, Colombia (in press).

Gepts, P.A., T.C. Osborn, K. Rashka, and F.A. Bliss. 1986. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): Evidence for multiple centers of domestication. Econ. Bot. 40:451–468.

Gottlieb, L.D. 1984. Genetics and morphological evolution in plants. Am. Nat. 123:681–709.

Hilu, K. 1983. The role of single-gene mutation in the evolution of flowering plants. *In* M.K. Hecht, B. Wallace, and G.T. Prance (ed.) Evol. Biol. 16:97–128.

Jana, S., and L.N. Pietrzak. 1988. Comparative assessment of genetic diversity in wild and primitive cultivated barley in a center of diversity. Genetics 119:981–990.

Kamalay, J.C., and R.B. Goldberg. 1980. Regulation of structural gene expression in tobacco. Cell 19:935–946.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge Univ. Press, Cambridge, England.

Knight, R.L. 1948. The role of major genes in the evolution of economic characters. J. Genet. 48:370–387.

Koenig, R., and P. Gepts. 1989a. Segregation and linkage of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J. Hered. 80:455–459.

Koenig, R., and P. Gepts. 1989b. Allozyme diversity in wild *Phaseolus vulgaris*: Further evidence for two major centers for genetic diversity. Theor. Appl. Genet. 78:809–817.

Koenig, R., S.P. Singh, and P. Gepts. 1990. Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ. Bot. 44:50–60.

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

Palmer, J.D., R.A. Jorgensen, and W.F. Thompson. 1985. Chloroplast DNA variation and evolution in *Pisum*: Patterns of change and phylogenetic analysis. Genetics 109:195–213.

Singh, S.P., J.A. Gutiérrez, A. Molina, C. Urrea, and P. Gepts. 1991. Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci. 31:23–29 (this issue).

Sneath, P.H.A., and R. Sokal. 1973. Numerical taxonomy. W.H. Freeman, San Francisco.

Weeden, N.F. 1984. Linkage between the gene coding the small subunit of ribulose biphosphate carboxylase and the gene coding malic enzyme in *Phaseolus vulgaris*. Annu. Rep. Bean Improv. Coop. 27:123–124.

Weeden, N.F. 1986. Genetic confirmation that the variation in the zymograms of 3 enzyme systems is produced by allelic polymorphism. Annu. Rep. Bean Improv. Coop. 29:117–118.

Wells, W.C., W.H. Isom, and J.G. Waines. 1988. Outcrossing rates of six common bean lines. Crop Sci. 28:177–178.