Theor Appl Genet (2003) 106:239–250
DOI 10.1007/s00122-002-1085-z

R. Papa · P. Gepts

Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica

Received: 17 September 2001 / Accepted: 2 May 2002 / Published online: 7 September 2002
© Springer-Verlag 2002

Abstract Using amplified fragment length polymorphisms (AFLPs), we analyzed the genetic structure of wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica at different geographical levels to test the hypothesis of asymmetric gene flow and investigate the origin of weedy populations. We showed both by phenetic and admixture population analyses that gene flow is about three- to four-fold higher from domesticated to wild populations than in the reverse direction. This result, combined with other work, points to a displacement of genetic diversity in wild populations due to gene flow from the domesticated populations. The weedy populations appear to be genetically intermediate between domesticated and wild populations, suggesting that they originated by hybridization between wild and domesticated types rather than by escape from cultivation. In addition, the domesticated bean races were genetically similar confirming a single domestication event for the Mesoamerican gene pool. Finally, the genetic diversity of the domesticated bean population showed a lower level of geographic structure in comparison to that of the wild populations.

Keywords Domestication · Population differentiation · Population admixture · Wild-weedy-crop complexes · Allopatry · Sympatry · AFLP · Spatial autocorrelation

Introduction

Crops and their wild relatives usually belong to the same biological species. As a rule, they cross freely and give rise to viable and fertile progenies (Harlan and de Wet 1971). Therefore, if wild and domesticated forms are sexually compatible, grow within pollinator flight distance (in the case of insect-pollinated species), and their flowering times overlap at least partially, gene flow may then occur between them. This gene flow will most likely affect the organization of genetic diversity in a crop's primary gene pool (Harlan and de Wet 1971). Possible outcomes of this gene flow have been discussed by Ellstrand et al. (1999). They include a reduction or an increase in the genetic diversity of the wild relatives, extinction of some populations, and the development of new and more aggressive weeds. More recently, the escape of transgenes through hybridization of crops with wild relatives in their centres of origin has been cited as one of the potential environmental risks of transgenic crops (Hails 2000). To what extent this escape will be a significant problem remains largely a matter of conjecture. However, a necessary condition for this escape is the existence of gene flow from domesticated to wild-types. While gene flow is likely to occur in allogamous species, such as maize (Doebley et al. 1990), its magnitude in autogamous species remains to be determined.

Common bean (*Phaseolus vulgaris* L., $2n = 2x = 22$) is considered a predominantly self-pollinated species. Although most data show outcrossing rates below 5%, recent information shows that occasionally higher rates can be achieved (Ibarra-Pérez et al. 1997). None of these estimates were obtained, however, in experiments involving sympatric wild and domesticated beans from the centers of origin. Debouck et al. (1993), Freyre et al. (1996) and Beebe et al. (1997) documented several examples of hybridization between wild and domesticated bean plants, based on morphological, biochemical and molecular marker analyses, and phenotypic observations in wild populations. Similar analyses among domesticated materials revealed a lower level of introgression possibly due to human selection against hybrids and suggested the occurrence of asymmetric gene flow (Singh et al. 1991; Gepts 1996). Many of these observations were related to the identification of weedy populations (Freyre

Communicated by F. Salamini

R. Papa (✉) · P. Gepts
Department of Agronomy and Range Science,
University of California, Davis, CA 95616-8515, USA
e-mail: rpapa@unian.it.
Fax: +39-071-2204-858

Permanent address:
R. Papa, Dipartimento di Biotecnologie Agrarie ed Ambientali,
Università degli Studi di Ancona, Ancona, Italy

et al. 1996; Beebe et al. 1997). Open to question is the origin of these populations, which could have originated by hybridization between wild and domesticated types or by escape from cultivation of seeds of domesticated types.

We have reported elsewhere (Papa et al., submitted) that, in spite of the selfing breeding system of common bean, gene flow is sufficient to prevent genetic isolation between wild and domesticated forms and that selection appears to be a major evolutionary factor maintaining the identity of wild and domesticated populations in sympatric situations. In regions of the genome where domestication genes and QTLs are located (D), the level of differentiation, measured as F_{ST}, was much higher than in other regions, where genes and QTLs not involved in the domestication syndrome (ND) or no known genes and QTLs (UN) have been located. The average size of genomic regions affected by linkage disequilibrium, inferred from the average map distance of linked markers, was quite large (14 cM) probably because of the autogamous breeding system of *P. vulgaris*.

The level of diversity for UN markers was similar between wild and domesticated populations and, in the wild, much lower than for D and ND markers. In domesticated populations, in contrast, the level of diversity was low and almost constant among all three classes of markers (Papa et al., submitted). This result could be due to the effect of purifying selection in UN genome areas or to the fact that genetic diversity in the wild populations was reduced by the effect of asymmetric gene flow, which would take place predominantly from the domesticated to the wild-types.

In this paper, we analyze the genetic structure of wild and domesticated common bean from Mexico at different geographical levels to test the hypothesis of asymmetric gene flow and investigate the origin of weedy populations. The same AFLP data set used to study the map-based analysis of introgression between wild and domesticated *P. vulgaris* (Papa et al., submitted) was employed in the present work for data analysis with, in addition, the results for nine populations not included in the study of Papa et al. (submitted).

Materials and methods

Plant materials

Analysis of germplasm accessions

This analysis included materials from germplasm banks and recent explorations. Germplasm materials were obtained from the Western Regional Plant Introduction Station of the United States Department of Agriculture (Pullman, Wash., USA) and the *Phaseolus* gene bank of the Centro Internacional de Agricultura Tropical (Cali, Colombia). In addition, one individual, randomly chosen in each of 20 populations collected in Mexico in 1996 and included in the second analysis (see below), was added to the sample (Table 1). The Mesoamerican sample was composed of 24 domesticated, 52 wild, and 10 weedy individuals including three wild accessions from Guatemala, one from Costa Rica, and one from Colombia. We define a weedy type as a material, which either shows morphological evidence of gene flow with domesticated types or grows only within or around a cultivated field. In addition to the Mesoamerican sample, 14 accessions from the Andean gene pool were also included as well as three wild accessions from the ancestral gene pool in northern Peru and Ecuador (Table 1). The Andean accession Bulk Cacahuate is cultivated in Mexico but was introduced at some time in the past from South America (Gepts et al. 1986).

Population analysis

A geographic sampling of wild, weedy, and domesticated populations from Chiapas and some other states of Mexico was performed. In Chiapas, different geographical scales were considered: among areas within the area of Teopisca, and in the part of this area (Rio Blanco) where cultivated and wild-types were growing in the same environment (sites E and I). A set of 382 individuals from populations of common bean collected in Mexico in 1996 was included in this analysis (Table 2). Eighteen wild and 13 domesticated populations were sampled. Pods were collected randomly from different plants in each population. The collection was primarily focused on the state of Chiapas, but two wild populations were also collected in areas where no domesticated beans were present in the states of Mexico and Morelos (Estado de Mexico, MXW; Morelos, MOW). In addition, four weedy populations (Durango, DURW; Jalisco, JW; Puebla, PUW; and Oaxaca, OXW) were collected. A small sample of domesticated populations from Jalisco, Puebla and Oaxaca was included as reference materials.

AFLP analysis

AFLP analysis was performed according to Vos et al. (1995). Selective amplifications were conducted using five *Eco*RI (A) and four *Mse*I (P) primers: A02 (5′-AAC-3′); A05 (5′-ACA-3′); A06 (5′-ACC-3′) A10 (5′-AGC-3′); A12 (5′-AGT-3′); P05 (5′-ACA-3′); P09 (5′-AGA-3′); P11 (5′-AGG-3′) and P15 (5′-ATG-3′). Four primer combinations were used for all the samples (A02/P11; A05/P05; A06/P15 and A12/P09) and an additional two were used only for the analysis of germplasm accessions (A10/P15 and A05/P09).

Statistical analyses

Diversity and cluster analysis

Because common bean is a predominantly selfing species (Ibarra-Pérez et al. 1997), all the individuals studied were considered to be homozygous. The data were therefore analyzed assuming a haploid genome, although results presented in this article suggest the assumption of complete autogamy may not be realistic because some outcrossing took place in those populations. On the other hand, the level of outcrossing needed in order to explain our result is very low and in agreement with the results obtained from direct measurements of outcrossing. Genetic distances between genotypes studied (Dice 1945; Nei and Li 1979) and unbiased genetic distances between populations (Nei 1978) were estimated by including both polymorphic and monomorphic markers. The resulting matrix was used for a UPGMA cluster analysis using NTSYS-pc version 1.70 (Rolf 1992) software. Unbiased gene diversity and its standard deviations (Nei 1987) were also estimated, using Arlequin 1.1 software (Schneider et al. 1997), in order to compare the diversity of different regions and populations. We analyzed the correlation between altitudes and the population frequencies of different AFLP markers. To minimize Type I errors, a level of $P < 0.01$ was chosen to test the significance of this correlation. Three AFLP markers were significantly correlated with altitude in wild but not domesticated types. These markers were excluded from AMOVA and spatial autocorrelation analyses because they possibly represent non-neutral markers or are linked to genes under selection.

Table 1 Domesticated and wild accessions included in the germplasm analysis

Label	Collector or donor	Accession code	Gene pool	Type	Country	State
WECD	CIAT	DGD2881	I	W	ECD	Loja
WPER2	CIAT	DGD2855	I	W	PER	Cajamarca
WPER3	CIAT	DGD1962	I	W	PER	Cajamarca
JALO	UCD	Jalo EEP558	A	D	BRA	Minas Gerais
CMxBCac	UCD	G13558	A	D	MEX	
BB	UCD	Brown Beauty	A	D	USA	California
CDRK	UCD	CDRK82	A	D	USA	California
CELRK	UCD	CELRK	A	D	USA	California
DRK	UCD	DRK	A	D	USA	California
LH	UCD	Linden	A	D	USA	California
MC	UCD	Montcalm	A	D	USA	California
REDK	UCD	Red Kidney	A	D	USA	California
WKID	UCD	White Kidney	A	D	USA	California
WARG1	Acosta	G19892	A	W	ARG	Salta
WARG2	CIAT	DGD643	A	W	ARG	Tucumán
WBOL	CIAT	DGD3025	A	W	BOL	Tarija
WPER1	CIAT	DGD2152	A	W	PER	Junín
BAT	UCD	BAT93	M	D	CLB	
CMxNTac	Acosta	DOI390	M	D	MEX	Mexico
CMxApe	UCD	UCD430	M	D	MEX	
CMxP152	Acosta	UCD1364/2	M	D	MEX	Puebla
CMxPuDP	Delgado	UCD1269A*	M	D	MEX	Puebla
CMxCHTc	UCD	CHCC3*	M	D	MEX	Chiapas
CMxCHTd	UCD	CHCD49*	M	D	MEX	Chiapas
CMxCHTh	UCD	CHCEH9*	M	D	MEX	Chiapas
CMxCHTe	UCD	CHCETE42*	M	D	MEX	Chiapas
CMxCHTf	UCD	CHCF5*	M	D	MEX	Chiapas
CMxCHTi	UCD	CHCI13*	M	D	MEX	Chiapas
CMxCHTl	UCD	CHCI2*	M	D	MEX	Chiapas
CMxCHLm	UCD	CHCM1*	M	D	MEX	Chiapas
CMxJa7	Delgado	JADC7(57)*	M	D	MEX	Jalisco
CMxBayR	UCD	G11010	M	D	MEX	
CMxBayC	UCD	UCD432	M	D	MEX	
CMxBayM	UCD	UCD435	M	D	MEX	
CMxDR22	UCD	UCD446	M	D	MEX	Durango
CMxAzuf	UCD	G14914	M	D	MEX	
CMxHua	UCD	G16077	M	D	MEX	
CMxCH7	UCD	UCD442	M	D	MEX	Chiapas
CMxAgua	UCD	G03255	M	D	MEX	Aguascalientes
CMxPu22	UCD	G03317	M	D	MEX	Puebla
CMxDCel	UCD	G13614	M	D	MEX	Guanajuato
WCLB	CIAT	Leroi Col-22	M	W	CLB	Cundinamarca
WCRA	Acosta	G23418	M	W	CRA	San José
WGTA2	CIAT	G19906	M	W	GUA	Sacatepequez
WGTA3	CIAT	G19907	M	W	GUA	Sacatepequez
WGTA4	CIAT	G19909	M	W	GUA	Sacatepequez
WMx	Acosta	UCD1386	M	W	MEX	Chiapas
WMxCHIx	Acosta	UCD1389/1	M	W	MEX	Chiapas
WMxCHL1	Acosta	UCD1378/1	M	W	MEX	Chiapas
WMxCHL2m	UCD	CHWM2*	M	W	MEX	Chiapas
WMxCHL2o	UCD	CHWO3*	M	W	MEX	Chiapas
WMxCHL3n	UCD	CHWN1*	M	W	MEX	Chiapas
WMxCHLT	Acosta	UCD1388/1	M	W	MEX	Chiapas
WMxCHT1	Acosta	UCD1383/1A	M	W	MEX	Chiapas
WMxCHT2	Acosta	UCD1384/1	M	W	MEX	Chiapas
WMxCHT3	Acosta	UCD1385/1	M	W	MEX	Chiapas
WMxCHTI	UCD	CHWI11*	M	W	MEX	Chiapas
WMxCHTxa	UCD	CHWA3*	M	W	MEX	Chiapas
WMxCHTxg	UCD	CHWG9*	M	W	MEX	Chiapas
WMxCHVr	Acosta	UCD1382/1	M	W	MEX	Chiapas
WMxCiu	CIAT	G23463	M	W	MEX	Chihuahua
WMxCiu2	Acosta	G22837	M	W	MEX	Chihuahua
WMxDr1	CIAT	G10022	M	W	MEX	Durango
WMxDr2	UCD	UCD700	M	W	MEX	Durango
WMxDr3	Acosta	G23556	M	W	MEX	Durango
WMxGua	CIAT	G12905	M	W	MEX	Guanajuato
WMxGue	Acosta	G12878	M	W	MEX	Guerrero
WMxJa1	CIAT	G11055	M	W	MEX	Jalisco

Table 1 (continued)

Label	Collector or donor	Accession code	Gene pool	Type	Country	State
WMxJa2	CIAT	G12866	M	W	MEX	Jalisco
WMxJa3	CIAT	G12884	M	W	MEX	Jalisco
WMxJa4	CIAT	G12925	M	W	MEX	Jalisco
WMxJa5	CIAT	G12949	M	W	MEX	Jalisco
WMxJa6	CIAT	G12865	M	W	MEX	Jalisco
WMxJa7	Acosta	G12979	M	W	MEX	Jalisco
WMxED0	Delgado	MXW10*	M	W	MEX	Mexico
WMxED2	Delgado	UCD1268/2A	M	W	MEX	Mexico
WMxMi	CIAT	G11050	M	W	MEX	Michoacán
WMxMo0	Delgado	MOW1*	M	W	MEX	Morelos
WMxMo2	CIAT	G12872A	M	W	MEX	Morelos
WMxMo3	CIAT	G12873	M	W	MEX	Morelos
WMxMo4	CIAT	G12877B	M	W	MEX	Morelos
WMxMo5	CIAT	G13018	M	W	MEX	Morelos
WMxMo6	CIAT	G13505	M	W	MEX	Morelos
WMxMo7	Acosta	G10012	M	W	MEX	Morelos
WMxOx1	Acosta	4002	M	W	MEX	Oaxaca
WMxOx2	Acosta	UCD1393/1A	M	W	MEX	Oaxaca
WMxOxSA	Acosta	UCD1391/1	M	W	MEX	Oaxaca
WMxOxSJ	Acosta	UCD1392/1	M	W	MEX	Oaxaca
WMxOxT	Acosta	UCD1390/1	M	W	MEX	Oaxaca
WMxPu1	Delgado	UCD1265/3	M	W	MEX	Puebla
WMxPu2	Acosta	G23429	M	W	MEX	Puebla
WMxPuX	Acosta	UCD1369/1	M	W	MEX	Puebla
WMxQUE	Acosta	G23415B	M	W	MEX	Querétaro
weMxCHTe	Acosta	UCD1387/1	M	Weed	MEX	Chiapas
weMxHid	Acosta	UCD1370	M	Weed	MEX	Hidalgo
weMxJa2	Acosta	UCD1395/C1	M	Weed	MEX	Jalisco
weMxJa96	Delgado	JADW96*	M	Weed	MEX	Jalisco
weMxJa11	Acosta	UCD1395/A1	M	Weed	MEX	Jalisco
weMxED1	Acosta	UCD1382	M	Weed	MEX	Mexico
weMxMi	Acosta	G12896A	M	Weed	MEX	Michoacán
weMxOxEl	Acosta	UCD1394	M	Weed	MEX	Oaxaca
weMxOx1	Acosta	UCD1396/E1	M	Weed	MEX	Oaxaca
weMxPuDP	Delgado	UCD1266(2)*	M	Weed	MEX	Puebla

* Genotype randomly extracted from the populations recently collected and used in the second experiment

AMOVA

To measure genetic structure and estimate pairwise and hierarchical F_{ST} (Wright 1965), we used the approach of Excoffier et al. (1992) for the analysis of the molecular variance, AMOVA, based on correlation between haplotypes at various levels of subdivision. Permutation tests were used to estimate the significance of variance components and F_{ST}. These analyses were implemented with the Arlequin software (Schneider et al. 1997). We also tested the isolation by distance model by spatial autocorrelation analysis using the program AIDA as proposed by Bertorelle and Barbujani (1995).

Admixture proportion

We initially used three estimators based on the assumption that the admixed population frequencies should be intermediate between the two parental populations (Cavalli-Sforza and Bodmer 1971; Bertorelle and Excoffier 1998): m_R (Roberts and Hiorns 1965), m_C (Chakraborty et al. 1992) and m_Y (Bertorelle and Excoffier 1998). Because the three estimators gave similar results only m_Y data are presented. The analysis was performed using ADMIX1_0 software developed by G. Bertorelle (http://www.unife.it/genetica/Giorgio/giorgio_soft.html), which also provides the estimator bootstrap average and its standard deviation. Our admixture model posits that the wild and domesticated populations each consist of two subpopulations: (1) "truly" wild and domesticated types without introgression from their domesticated and wild counterparts (P_W and P_D, respectively); and (2) admixture populations P_{hyW} and P_{hyD} with evidence of introgression (e.g., close-range sympatry; higher genetic diversity and intermediate gene frequencies), the first wild and the second domesticated. Each hybrid population consists of $N(1 - \mu)$ loci originating at random from one parental population and $N\mu$ from the other as follows: $p_{hyW} = \mu_2 p_D + (1 - \mu_2)p_W$ and $p_{hyD} = \mu_1 p_W + (1 - \mu_2)p_D$. We can thus compare μ_1 (P_W contribution to P_{hyD}) and μ_2(P_D contribution to P_{hyW}). For our analysis we pooled the allopatric and the medium-range wild (CHWJn, CHWJrd, CHWErd, and $CHWE_S$) and domesticated ($CHCJ_N$, CHCF, and CHCD) populations from Teopisca in the two parental populations P_W and P_D, respectively (Table 2), and the close-range sympatric wild ($CHWE_E$ and CHWI) and domesticated ($CHCE_E$, $CHCE_S$, and CHCI) populations in two hybrid population (Phy_w and Phy_D), respectively. The domesticated populations CHCC and $CHCE_{MAT}$ were not included in this analysis because of the different phenology from all the other domesticated populations from Teopisca. The analysis was performed using all the markers (78) that were polymorphic between the parental populations or markers selected to present significant differences between the frequencies of the two parental populations (35). In the presence of populations of different sizes, admixture coefficients will reflect both the migration rate and the relative population size.

Table 2 Germplasm accessions from Mexico in areas with sympatric wild and domesticated beans (population analysis)

Locations					Materials collected			
State	Area	Altitude (m.a.s.l.)	Sites		Wild		Domesticated	
					Population name	Number of individuals analyzed	Population name	Number of individuals analyzed
Chiapas	Tuxtla-Gutiérrez	1,300	A		CHWA	9	CHCA	2
		1,360	H		CHWH	12		
		1,450	G		CHWG	14		
	Las Rosas	1,550	M		CHWM	3	CHCM	8
		1,430	N		CHWN	9		
		1,180	O		CHWO	10		
	Teopisca	1,650	F				CHCF	8
		1,500	C				CHCC	11
		1,450	D				CHCD	8
		1,200	I		CHWI	12	CHCI	22
		1,200–1,050	E	South	$CHWE_S$	9	$CHCE_S$	20
				East	$CHWE_E$	18	$CHCE_E$	18
				Matita			$CHCE_{MAT}$	12
				Roads	$CHWE_{rd}$	8		
			J	North	$CHWJ_N$	13	$CHCJ_N$	12
				Roads	$CHWEJ_{rd}$	6		
Oaxaca	Santo Domingo de Albarradas	NA[a]			OXW	14	OXC	3
Durango	Durango	1,900		La Ferriera	DGW	20		
Jalisco	Jocotepec	1,740			JW	33	JC	6
México	Temascaltepec	2,000			MXW	17		
Morelos	Tepoztlán	1,830			MOW	15		
Puebla	Huapalejcan	1,280		Don Pascual	UW	26	PUC	4
					Total	248		134

[a] NA: not available

Results

In the germplasm analysis, accessions from the three gene pools of common bean (Mesoamerican, Andean, and Ancestral) were included. Out of 213 markers, 192 (90.1%) were reproducibly polymorphic. For the six primer combinations, the average number of markers scored per primer combination was 37, ranging from 21 (AO5/PO5) to 52 (AO2/P11) and the percentage of polymorphic markers ranged between 88.5% (A12/PO9) and 91.3% (AO6/P15). In the population analysis, where only Mesoamerican materials were studied, 145 markers were scored, the number of polymorphic markers was 110 (76%) ranging from 18 (AO5/PO5 and A12/PO9) to 39 markers (AO6/P15) depending on the AFLP primer combination.

Germplasm accessions analysis

The diversity for AFLP markers was analyzed in a set of samples from the Mesoamerican (n = 86; 191 markers), Andean (n = 14; 130 markers) and Ancestral (n = 3; 87 markers) gene pools including wild, weedy and domesticated accessions. The levels of polymorphism were 80, 25 and 5% for the Mesoamerican, Andean and Ancestral gene pools, respectively. All materials showed a unique haplotype. The highest gene diversity was found for the wild Mesoamerican accessions, which was 1.7-times higher than the diversity found in the domesticated accessions from Mesoamerica (Table 3). Wild Andean beans also showed a higher diversity compared to domesticates of the same area. Among wild accessions from Mesoamerica, the highest diversity was found in North-Central Mexico. The diversity was decreasing slightly towards the South with the lowest value in Central America. Northern Mexico also showed a slightly lower diversity in comparison to North-Central Mexico (Table 3).

A phenetic analysis was conducted to analyze relationship among individual accessions (Fig. 1). The Ancestral gene pool (wild accessions from northern Peru and Ecuador) showed a clear separation from the other two, derived gene pools (Andean and Mesoamerica). The latter gene pools were separated from each other at an average genetic distance of about 0.32. This structure of genetic diversity agrees with previous information on the evolution of *P. vulgaris* (Kami et al. 1995; Freyre et al. 1996). Within the Andean cluster, the structure of the kidney cultivar varieties group matched their pedigree information (McClean and Myers 1990). A correlation between the kinship coefficient and the Nei and Li (1979) distance among accessions was significant (r = 0.58, Mantel test, $P < 0.015$).

Within the Mesoamerican cluster, two subgroups were identified at a 0.28 distance (Fig. 1). The first sub-

Table 3 Genetic diversity of wild and domesticated common bean in Mexico

Type	*N*	No. of haplotypes (%)	No. of polymorphic markers/total no. of markers	Average gene diversity	Standard deviation
Mesoamerican wild	61	61	168/189	0.24	0.12
Central America (Guatemala, Costa Rica, Colombia)	5	5	65/132	0.16	0.10
South Mexico (Oaxaca, Chiapas)	21	21	124/159	0.20	0.11
Central Mexico (México, Morelos, Puebla, Guerrero)	17	17	126/166	0.22	0.11
Northcentral Mexico (Jalisco, Guanajuato, Querétaro, Hidalgo, Michoacán)	13	13	126/155	0.24	0.13
Northern Mexico (Chihuahua, Durango)	5	5	92/139	0.22	0.13
Mesoamerican domesticated	25	25	110/148	0.15	0.08
Andean wild	4	4	35/105	0.10	0.06
Andean domesticated	10	10	31/109	0.05	0.03
Ancestral wild	3	3	10/103	0.03	0.03

Table 4 Genetic diversity in sympatric wild and cultivated common bean populations from Mexico

Location and bean types	Population	No. of individuals	No. of haplotypes (%)	No. of polymorphic markers/total no. of markers	Average gene diversity	Standard deviation
Chiapas domesticated						
	Tuxtla	2	2 (100)	7/ 94	0.05	0.05
	Teopisca	111	96 (87)	83/129	0.13	0.07
	Las Rosas	8	8 (100)	24/102	0.06	0.03
Chiapas (domesticated)		121	106 (88)	86/129	0.13	0.06
Chiapas wild						
	Tuxtla	35	31 (89)	75/127	0.14	0.07
	Teopisca	66	55 (82)	80/128	0.12	0.06
	Las Rosas	22	18 (82)	43/108	0.09	0.05
Chiapas (wild)	123	104 (85)	89/135	0.16	0.08	
Other Mexican						
Durango	DGW	20	14 (70)	53/109	0.07	0.04
Jalisco	JC	6	6 (100)	29/96	0.08	0.05
	JW	33	31 (94)	83/125	0.19	0.10
Morelos	MOW	15	14 (93)	51/112	0.10	0.06
Mexico	MXW	17	17 (100)	64/116	0.15	0.08
Oaxaca	OXC	3	3 (100)	26/99	0.12	0.09
	OXW	14	14 (100)	57/113	0.13	0.07
Puebla	PUC	4	4 (100)	36/103	0.14	0.09
	PUW	26	26 (100)	79/123	0.15	0.08
Total wild		248	220 (89)	110/145	0.22	0.11
Total domesticated		134	119 (89)	92/132	0.14	0.07

group was composed of 52 accessions including all the domesticated accessions and 27 wild or weedy accessions. No geographical structure in this subgroup was evident from the dendrogram. The domesticated accessions were nested inside this cluster and separated from most of the wild accessions. The second group included 35 wild accessions structured according to their geographical origin. This second cluster included the following three groups: (1) 18 accessions from south Mexico and Central America; (2) nine accessions from Central Mexico; and (3) seven accessions from North-Central and Northern Mexico.

Population analysis

Among the 145 markers identified, 13 were private in wild beans. As in the analysis based on germplasm accessions, the average gene diversity for the wild was higher than for the domesticated types (Table 4). In Chiapas, the highest diversity was present in wild bean populations in the area of Tuxtla, followed by the areas of Teopisca and Las Rosas. The highest diversity was found in the wild populations in close-range sympatry (CHWA, 0.15; CHWI, 0.12; and $CHWE_E$ 0.11) and the lowest in $CHWE_S$ (0.01). Among domesticated populations, the lowest level of diversity was found in the determinate type $CHCE_{MAT}$ (0.01) and the highest in a population in close-range sympatry with wild-types, CHCI (0.11). The

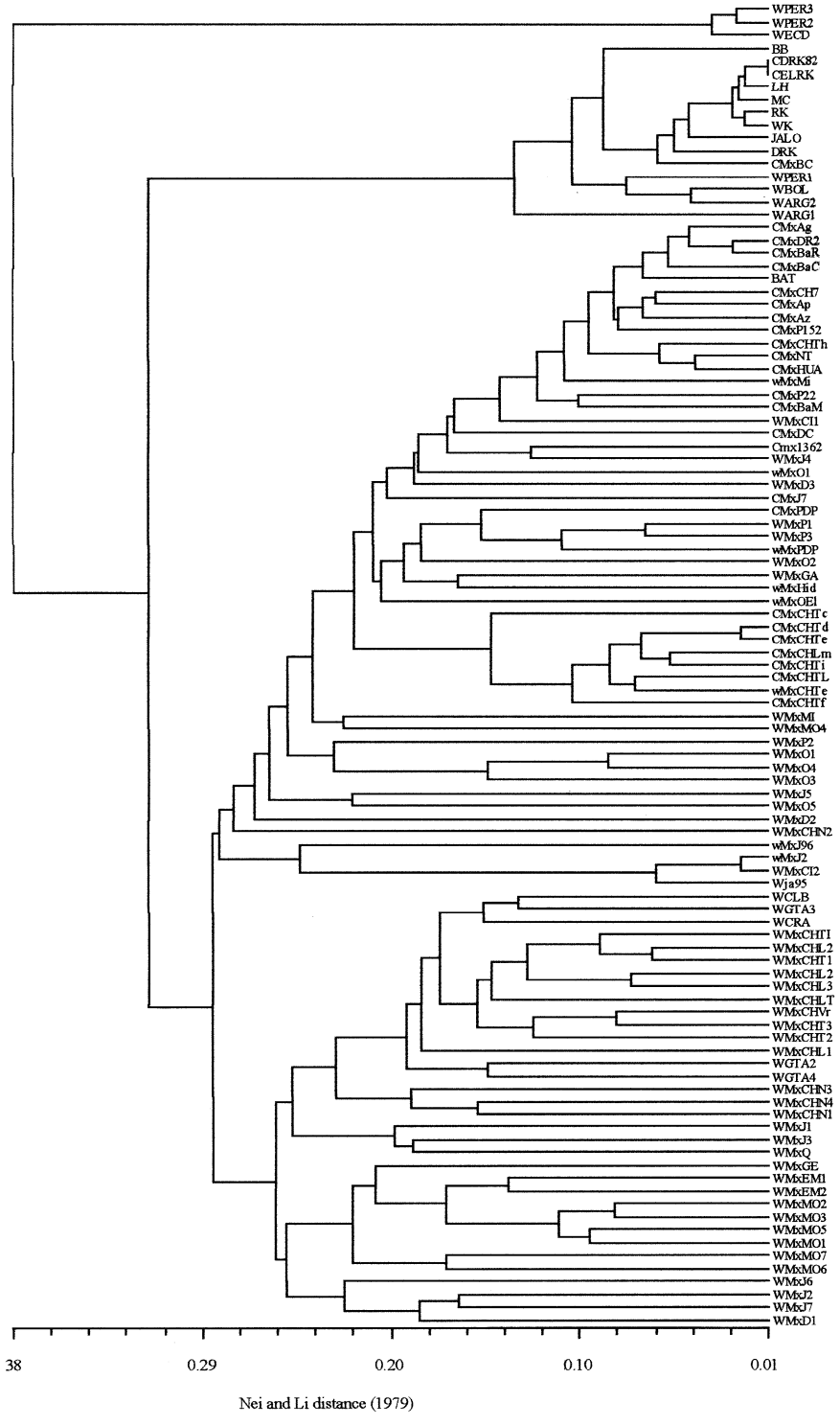

Fig. 1 UPGMA dendrogram based on Nei and Li distance (1979) matrix of wild and domesticated germplasm accessions of common bean

partitioning of the diversity among areas and populations in wild populations from Chiapas was studied using a molecular variance analysis (AMOVA). All the components of the AMOVA were found to be highly significant; 55.9% of the genetic variation was found to be among populations (37.7% between areas and 18.3% between populations within areas) and the within-population component was 44.1% of the total genetic variance.

Only one population of domesticated common bean was found in the area of Tuxtla and Las Rosas; therefore, only the structure of genetic diversity among populations was analyzed. Genetic variance components between and within domesticated populations of Chiapas were highly significant. The partitioning of the genetic variance of the domesticated population was 41.5% between populations and 58.5% within a population. As also sug-

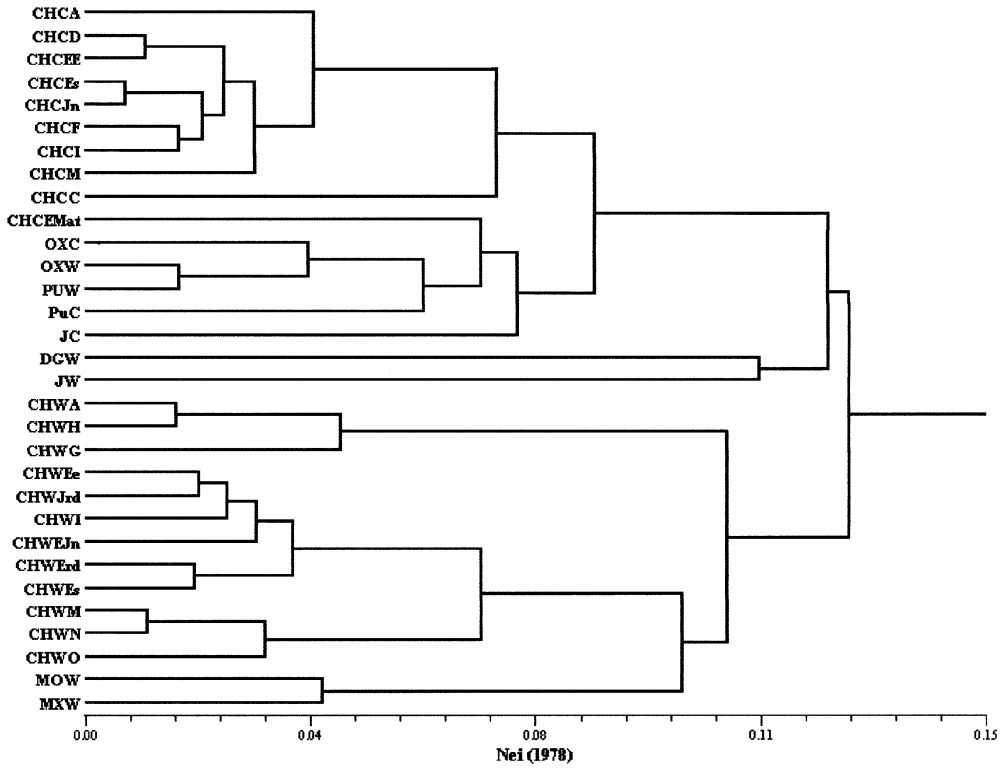

Fig. 2 UPGMA dendrogram based on Nei's genetic distance (1978) matrix of wild and domesticated populations of common bean from Mexico

gested by the spatial correlation analysis, most of the differentiation among populations was due to the $CHCE_{MAT}$ and CHCC populations, which presented a different phenology relative to the others. Indeed, removing these populations from the analysis gave a markedly different result: 22.75% between populations and 77.75% within a population.

The UPGMA clustering of populations based on Nei's unbiased genetic distance (1978) (Fig. 2) confirmed the results of the germplasm analysis. Two clusters were obtained at a distance of 0.12. The first cluster included all the domesticated and weedy populations. As among individual germplasm accessions, all the domesticated populations from Chiapas formed a single cluster with the exception of $CHCE_{MAT}$. With the exception of CHCC, the domesticated populations from Chiapas clustered according to their collection areas. The second cluster included only wild populations. The wild populations were subdivided between North Chiapas, South Chiapas and Central Mexico. The South Chiapas populations grouped according to the collection areas and sites.

Spatial autocorrelation

To obtain a more detailed picture of the spatial structure of the populations studied we performed an autocorrelation analysis according to Bertorelle and Barbujani (1995) (Fig. 3). For the three groups of populations (wild, weedy and domesticated), we chose the same distance classes with the exception of the wild materials (Fig. 3A), which did not include two distance classes (380 km, Chiapas-Oaxaca and 1,280 km North-South Mexico). For the wild populations (Fig. 3A), all the positive and negative autocorrelation coefficients (*II*), analogous to Moran's *I* (Sokal and Oden 1978), were highly significant ($P < 0.005$). Decreasing positive coefficients were found from the within-population class (0.0 km, minimum distance of the class) until the class "within South Chiapas and North Chiapas". The last two classes, North-South Chiapas and Morelos-Estado de Mexico (85.5 km) and Chiapas-Morelos and Chiapas-Estado de Mexico (380 km), showed a very similar negative autocorrelation coefficient. This pattern indicated that within each area of Chiapas, populations are distributed along a gradient in relation to drift and local gene flow. As distances increased and were enhanced by physical barriers like the high elevation that divides North from South Chiapas, gene flow became ineffective. When the weedy populations were included (Fig. 3B), the pattern was similar but after stabilization at intermediate distances (Central Mexico-South/North Chiapas), *II* was decreasing until the last class in which Northern and Southern Mexico populations were compared. This "long distance differentiation" pattern (Sokal et al. 1989; Barbujani et al. 1994) is regarded to be typical of ancient clines on which successive episodes of gene flow, drift and adaptation to local environment have been superimposed.

For the domesticated populations (Fig. 3C), we observed a lower genetic similarity at the population level and a much smaller differentiation between populations for all the correlograms. We also observed a depression

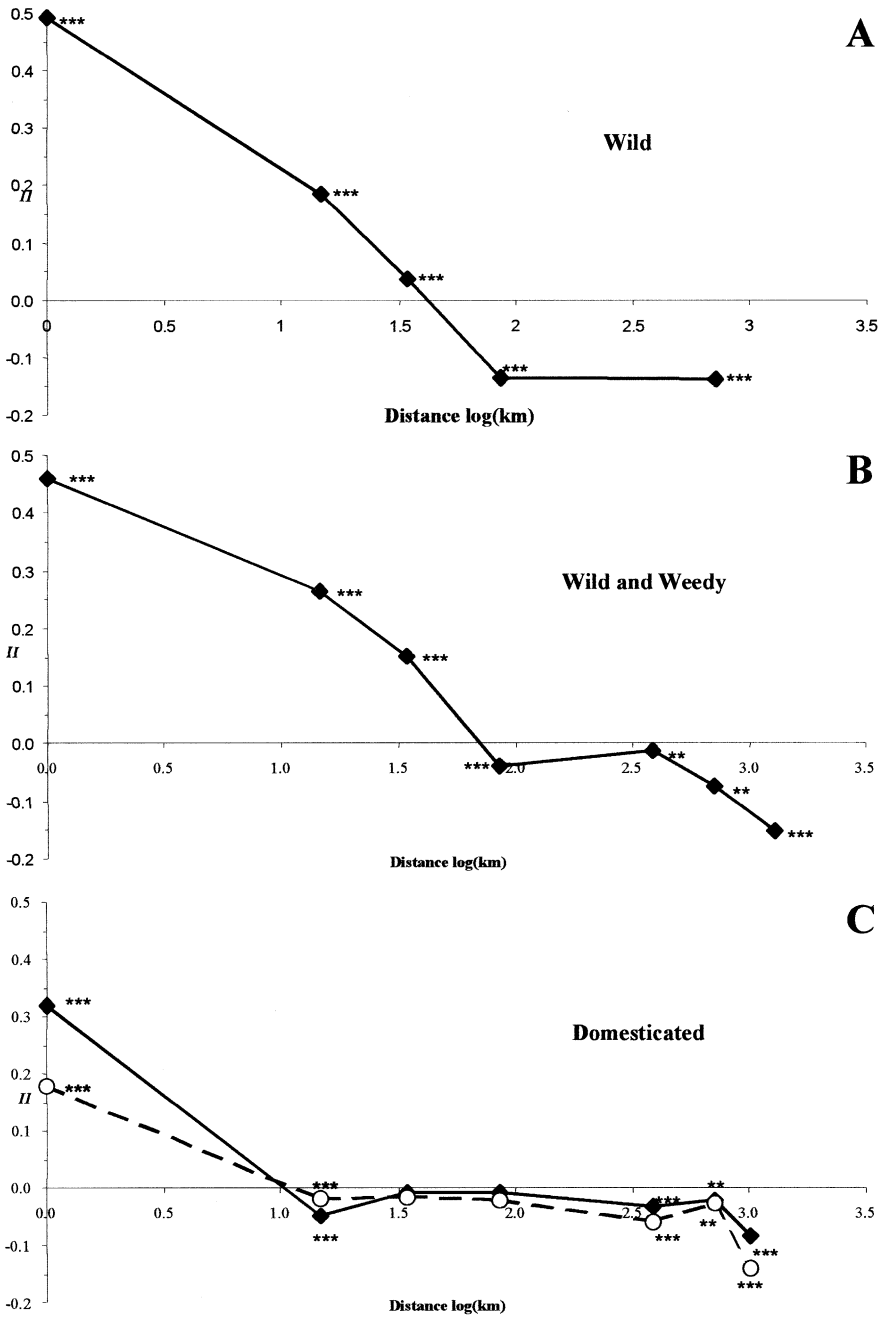

Fig. 3 Correlogram of AFLP variation in wild (**A**), wild and weedy (**B**), and domesticated (**C**) types. In **C**: all populations (solid line), without $CHCE_{MAT}$ and CHCC (dashed line). The first distance class (0 km) includes comparison between individuals belonging to the same population. The maximum distance for the other classes are: 14.5 km (within Chiapas area), 34 km (within south and north Chiapas), 85.5 (within Chiapas and between Estado de Mexico and Morelos), 380 km (between Chiapas and Oaxaca), 709 km and 1280 km (within Mexico). The 380 Km and the 1280 km classes were not present in wild population (**A**). *($P < 0.05$), **($P < 0.01$), ***($P < 0.005$) indicate significance level of the autocorrelation coefficient

at short distances followed by a non-significant coefficient for the intermediate class and negative autocorrelation coefficients at long distances. When two populations ($CHCE_{MAT}$ and CHCC) were removed, the depression for the second distance class disappeared and *II* of the within-population class decreased due to the low level of polymorphism of these two populations. Finally, lower coefficients were also obtained for two of the last three classes, Chiapas-Center Mexico and North-South Mexico, after removal of populations $CHCE_{MAT}$ and CHCC, suggesting that the two domesticated populations were introduced relatively recently in Chiapas.

Admixture

The relative contribution of the wild and domesticated parental populations to the composition of the two wild and domesticated admixture population were estimates based on the model presented in Materials and methods. The two parental populations differed significantly for 35 loci out of 78. For both sets of loci ($n = 78$ or 35), the estimate of the contribution of the domesticated parental population to the admixture wild population (m_{DW}) was significantly higher than the estimate of the contribution of the parental wild population to the domesticated admixture (m_{WD}) (Table 5).

Table 5 Analysis of admixture (m_Y) between domesticated and wild beans in close-range sympatry

Item	m_{DW}			m_{WD}			m_{DW}/m_{WD}	
	Estimated admixture coefficient	Bootstrap		Estimated admixture coefficient	Bootstrap		Estimated coefficient	Bootstrap average
		Average	Standard deviation		Average	Standard deviation		
All markers (78)	0.18	0.19	0.02	0.05	0.06	0.01	3.9	3.1
Differentiating markers (35)	0.19	0.19	0.02	0.06	0.06	0.01	3.4	3.1

Discussion

Genetic structure and diversity of wild and domesticated common bean

Our results showed a high level of differentiation between wild and domesticated common bean. Introgression from wild beans did not appear to have had a major influence on the overall organization of neutral genetic diversity of the domesticated gene pool. This contrasts with earlier claims of significant effects of gene flow on phenotypic traits in the domesticated gene pool (Gentry 1969). All the domesticated types from Mesoamerica, which belong to different races (Singh et al. 1991), were grouped in a single, large cluster (Figs. 2 and 3) suggesting – in agreement with electrophoretic data for phaseolin seed protein (Gepts et al. 1986) – that racial differentiation in the domesticated gene pool is mainly due to natural and farmer selection rather than multiple domestication events in Mesoamerica. The structure of genetic diversity in wild *P. vulgaris* was more complex compared to that of domesticated beans. In both dendrograms (Figs. 2 and 3), the clustering of most wild-types followed a clear geographical pattern. Other wild-weedy accessions or populations were clustered with the domesticated materials and contained markers typical of domesticated materials, suggesting they resulted from outcrosses in past generations with the latter materials. The absence of domesticated materials in the wild cluster is also indicative of the asymmetry of gene flow between the two forms.

The structure of genetic diversity appeared to be quite different between wild and domesticated types. A spatial autocorrelation analysis of population structure showed a clear geographical structure of the genetic diversity in the wild material as well as a strong differentiation between populations even at short distances, as expected in predominantly selfing natural populations. The genetic diversity of wild populations from Mexico seemed to be much more structured than that of wild populations from the Andean gene pool. Cattan-Touppance et al. (1998), using RAPDs, showed in fact a much higher within-population variance component in the wild bean population from Argentina (67.5%) than our estimate. Several studies (Koenig and Gepts 1989; Freyre et al. 1996) showed that the genetic diversity in wild material was much lower in the Andes than in Mesoamerica. Among the explanations for this observation, one should probably consider the more-marked geographic structure of genetic diversity that is present in the populations from Mesoamerica (present results). In contrast, the diversity of domesticated beans showed limited geographical structure and much less differentiation among populations and regions resulting in a much higher within-population component of genetic diversity. This observation can be attributed to the effect of seed exchange among farmers and homogeneous selection in different environments.

Our results confirm the reduction in genetic diversity induced by the "domestication bottleneck," which is very well documented in common bean (Gepts et al. 1986; Sonnante et al. 1994) and other plant species (Ladizinsky 1985; Doebley 1992). At the population level, the difference in genetic diversity between wild and domesticated types was attenuated. In the analysis of germplasm accessions, the genetic diversity of the domesticated populations represented 65% of the genetic diversity in wild forms. However, in Chiapas the corresponding value was 81%. This difference can be attributed to a lower level of diversity of the wild population from Chiapas in comparison to other wild populations from Mesoamerica as well as to the higher between-populations component of genetic variation showed by the wild populations.

The origin of weedy populations

The presence of weedy plants with intermediate traits between wild and domesticated plants (Freyre et al. 1996; Beebe et al. 1997) suggested either the existence of introgression from domesticated to wild forms or escapes from cultivation. Our results showed that weedy types were genetically intermediate for molecular marker variation suggesting that these weedy types were indeed derived from hybridization between wild and domesticated types. An alternative explanation could be that the wild vs weedy subdivision in the wild-gene pool was already present when common bean was domesticated in Mesoamerica and that the weedy types were domesticated rather than the wild-types. This hypothesis seems inconsistent with the absence of geographical structure of genetic diversity we observed in the weedy populations (Figs. 2 and 3). In addition, we observed

evidence of introgression even between wild and domesticated forms that are clearly separated such as in Chiapas.

Asymmetric gene flow

The admixture proportion estimates showed that m_{DW} was between 3.1 and 3.9-times higher than m_{WD} indicating that gene flow is asymmetric. Two, not-mutually exclusive explanations can account for this observation. First, pollen dispersal in common bean may be strongly affected by distance (i.e., because pollinators may have limited flight distances). In this situation, the number of wild individuals that can exchange pollen with domesticated individuals is to a large extent limited to those present inside or around the field, the number of which is usually much lower than the number of domesticated plants. With this scenario, asymmetric gene flow can be explained just in terms of differences in population size between wild and domesticated types.

Second, when most farmers select seeds for the next planting, the main traits they consider are seed size and color. Seeds of domesticated × wild F_1 hybrids generally have seeds of a different color than seeds of the female, domesticated parent. Usually they are also of intermediate size between those of the two parents. Hence, hybrids can be easily recognized and selected against by farmers, strongly reducing the possibility of introgression of wild germplasm into domesticated germplasm. Different selective forces are likely to operate against wild × domesticated F1 and later-generation hybrids. These hybrids, however, show marked hybrid vigor (Singh et al. 1995), which may favor the survival of genes introduced from the domesticated gene pool. It is important to distinguish between the selection against F1 hybrids and selection for adaptation to cultivated or natural environments. The first will act on the whole genome with an effect similar to a postzygotic reproductive barrier. This makes it difficult to distinguish, by indirect methods, if asymmetric gene flow is due to the effect differences in population size (domesticated > wild) or a stronger selection against F1 hybrids in domesticated populations compared to wild populations. Both factors result in a long-term reduction of average gene flow from wild to domesticated populations.

The lower genetic diversity observed for markers located in areas of the genome devoid of domestication genes can now be interpreted in light of the asymmetric gene flow from domesticated to wild populations. This gene flow may be responsible for displacing the genetic diversity that was present originally in these regions given the genetic diversity bottleneck characterizing the domesticated gene pool. In contrast, we hypothesize that in genomic regions where introgression from domesticated populations was inhibited by selection the original diversity of wild populations was maintained.

Acknowledgements This work was supported by the McKnight Foundation. R.P. was supported by CNR (Italy) for his stay in Davis and work at Ancona. We thank Alfonso Delgado and Jorge Acosta for providing some of the plant materials and assisting in the Chiapas exploration and Guido Barbujani for providing the AIDA software and for the critical reading of an early version of the manuscript.

References

Barbujani G, Pilastro A, De Domenico S, Renfrew C (1994) Genetic variation in North Africa and Eurasia: neolithic demic diffusion versus Paleolithic colonization. Am J Phys Anthropol 95:137–154

Beebe S, Toro O, González AV, Chacón MI, Debouck DG (1997) Wild-weed-crop complexes of common bean (*Phaseolus vulgaris*) in the Andes of Peru and Colombia, and their implications for conservation and breeding. Genet Res Crop Evol 44:73–91

Bertorelle B, Excoffier L (1998) Inferring admixture proportion from molecular data. Mol Biol Evol 15:1298–1311

Bertorelle G, Barbujani G (1995) Analysis of DNA diversity by spatial autocorrelation. Genetics 140:811–819

Cattan-Touppance I, Michalakis Y, Neema C (1998) Genetic structure of wild bean populations in their South-Andean centre of origin. Theor Appl Genet 96:844–851

Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations. W.H. Freeman, San Francisco

Chakraborty R (1986) Gene admixture in human populations: models and predictions. Yearbook of Phys Anthropol 29:1–43

Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P (1993) Genetic diversity and ecological distribution of *Phaseolus vulgaris* in northwestern South America. Econ Bot 47:408–423

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302

Doebley J (1992) Molecular systematics and crop evolution. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman Hall, New York pp 202–222

Doebley J, Stec A, Wendel J, Edwards M (1990) Genetic and morphological analysis of a maize-teosinte F_2 population: implications for the origin of maize. Proc Natl Acad Sci USA 87:9888–9892

Ellstrand N, Prentice, Hancock J (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annu Rev Ecol Syst 30:539–563

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Freyre R, Ríos R, Guzmán L, Debouck DG, Gepts P (1996) Ecogeographic distribution of *Phaseolus* spp. (Fabaceae) in Bolivia. Econ Bot 50:195–215

Gentry HS (1969) Origin of the common bean, *Phaseolus vulgaris*. Econ Bot 23:55–69

Gepts P (1996) Origin and evolution of cultivated *Phaseolus* species. In: Pickersgill B, Lock J (eds) Advances in Legume Systematics, Part 8. Legumes of economic importance. Kew Botanical Garden, Kew, UK, pp 65–74

Gepts P, Osborn TC, Rashka K, Bliss FA (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

Hails RS (2000) Genetically modified plants – the debate continues. Trends Ecol Evol 15:14–18

Harlan JR, Wef JMJ de (1971) Towards a rational classification of cultivated plants. Taxon 70:509–517

Ibarra-Pérez F, Ehdaie B, Waines G (1997) Estimation of outcrossing rate in common bean. Crop Sci 37:60–65

Kami J, Becerra Velásquez V, Debouck DG, Gepts P (1995) Identification of presumed ancestral DNA sequences of phaseolinin Phaseolus vulgaris. Proc Natl Acad Sci USA 92:1101–1104

Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor Appl Genet 78:809–817

Ladizinsky G (1985) Founder effect in crop-plant evolution. Econ Bot 39:191–198

McClean P, Myers J (1990) Pedigrees of dry bean cultivars, lines and PIs. Annu Rep Bean Improv Coop 33:25–30

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Li WH (1979) Mathematical models for studying genetic variation in terms of restriction endonuclease. Proc Natl Acad Sci USA 76:5269–5273

Roberts D, Hiorns R (1965) Methods of analysis of the genetic composition of hybrid populations. Human Biol 37:38–43

Rolf FJ (1992) NTSYS. Numerical taxonomy and multivariate analysis system, version 1.70. Exeter, Setauket, New York

Schneider S, Kueffer JM, Roessli D, Excoffier L (1997) Arlequin ver. 1.1: a software for population genetic data analysis. Genetic and Biometry Laboratory, University of Geneva, Switzerland

Singh SP, Gepts P, Debouck DG (1991) Races of common bean (*Phaseolus vulgaris* L. Fabaceae). Econ Bot 45:379–396

Singh SP, Molina A, Gepts P (1995) Potential of wild common bean for seed yield improvement of cultivars in the tropics. Can J Plant sci 75:807–813

Sokal RR, Oden NL (1978) Spatial autocorrelation analysis in biology. 1. Methodology. Biol J Linn Soc 10:199–228

Sokal RR, Harding RH, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. Am J Phys Anthropol 80:267–294

Sonnante G, Stockton T, Nodari RO, Becerra Velásquez VL, Gepts P (1994) Evolution of genetic diversity during the domestication of common-bean (*Phaseolus vulgaris* L.). Theor Appl Genet 89:629–635

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wright S (1965) The interpretation of population structure by *F*-statistics with special regard to systems of mating. Evolution 19:355–354