Published online January 22, 2007

Gene Flow and Genetic Structure in the Wild–Weedy–Domesticated Complex of *Phaseolus lunatus* L. in its Mesoamerican Center of Domestication and Diversity

Jaime Martínez-Castillo, Daniel Zizumbo-Villarreal, Paul Gepts, and Patricia Colunga-GarcíaMarín*

ABSTRACT

The role of gene flow in autogamous domesticated species diversity and their wild relatives is an issue that requires more field data. Using nine microsatellite loci, an analysis was done of the magnitude and direction of gene flow in the wild–weedy–domesticated complex of *Phaseolus lunatus* L. under traditional agricultural conditions in four regions on the Yucatan Peninsula, Mexico, its center of domestication and diversity in Mesoamerica. Two complementary methods were used. The Bayesian genotype assignment approach showed that recent gene flow was low at both the intraregional and interregional levels. The same was found with the frequency method for long-term gene flow (*Nm* intraregional from 0.31 to 0.51, and *Nm* interregional = 0.44). In addition, the gene flow from domesticate to wild populations was three times higher than in the opposite direction. This asymmetry can be explained by regional agricultural practices and seed selection criteria. Domesticate alleles were shown to be entering wild populations of different agricultural regions, suggesting exchange of domesticated seed between farmers of different regions. These results are very important because they show that *P. lunatus* on the Yucatan Peninsula has a predominantly domesticate to wild gene flow. This situation may lead to genetic assimilation of wild lima bean by its domesticated counterpart and may lead to the possible escape of transgenes in this center of origin and diversity.

In many areas where wild ancestors are distributed, landraces are still currently cultivated as an important element of traditional agroecosystems, thus potentially allowing gene flow and introgression from the wild to domesticate gene pool. The resulting new gene combinations of these events have played a vital role in the evolution of domesticated species (Harlan, 1965; Slatkin 1987; Stebbins, 1959) and continue to have a significant effect in augmenting genetic diversity in modern crops (Altieri and Montecinos, 1993; Jarvis and Hodgkin, 1999; Quirós et al., 1992). In contrast, when gene flow and introgression take place predominantly from domesticate toward wild populations, the consequences, due to the lower diversity of the domesticated forms, can be quite different, including a reduction of the genetic diversity of wild relatives, local extinction of wild populations, and development of more aggressive weedy varieties (Ellstrand et al., 1999; Gepts and Papa, 2003; Payró de la Cruz et al., 2005). With the development of transgenic crops, the potential environmental risk of transgenics escaping via hybridization between crops and wild relatives in centers of origin has emerged (Hails, 2000; Snow, 2002). The level of risk is still open to debate, though it is known that a necessary condition for this escape is the existence of gene flow from domesticate toward the wild gene pools (Gepts and Papa 2003). Gene flow in allogamous species has been extensively researched, but determining its magnitude in autogamous species requires further study (Papa and Gepts, 2003; Zizumbo-Villarreal et al., 2005).

Lima bean is an herbaceous species with an annual or short-cycle perennial lifecycle. Autogamy is favored by the synchronized maturity of pollen grains and stigma in this species, as well as their proximity within the bud (Webster et al., 1979). However, outcrossing rates of 0.02% up to 48% have been reported, depending on genotype, growth conditions, distance between plants, wind direction, and local pollinating insect populations (Baudoin et al., 1998; Zoro Bi et al., 2005). Lima bean consists of two subspecies: *P. lunatus* var. *lunatus*, which includes domesticated populations, and *P. lunatus* var. *silvester*, composed of wild populations (Baudet, 1977). It also has two main gene pools: Andean and Mesoamerican (Debouck et al., 1987; Gutiérrez-Salgado et al., 1995; Maquet et al., 1990).

The Yucatan Peninsula is within the putative area of domestication of the Mesoamerican gene pool of lima bean (Gutiérrez-Salgado et al., 1995). Traditional agriculture in this region has intensified over the last 30 years, leading to changes such as reduction in the richness of cultivated species and its landraces, reduction of the vegetation harboring wild species that surrounds croplands, shortening of fallow periods, increased agrochemical use, and greater incorporation of farmers into the external market system (Reyes and Aguilar, 1992; Lazos-Chavero, 1995; Ku-Naal, 1995; Remmers and Ucan, 1996). Despite these changes, lima bean remains the fourth most important crop among Mayan farmers in the region. This fact is reflected in high morphophenological diversity in landraces (Ballesteros, 1999; Debouck, 1979; Martínez-Castillo et al., 2004), making the Yucatan Peninsula the region with the highest richness of domesticated *P. lunatus* varieties in Mexico (Ballesteros, 1999).

Wild *P. lunatus* populations have been found on the Yucatan Peninsula growing sympatrically with domesticated populations. They include plants showing introgression with domesticated populations in their morphological characters, as in flowers, pods, and seeds

J. Martínez-Castillo, D. Zizumbo-Villarreal, and P. Colunga-García-Marín, Centro de Investigación Científica de Yucatán (CICY), Calle 43 No. 130, Col. Chuburná de Hidalgo, 97200 Mérida, Yucatán, México; P. Gepts, Dep. of Agronomy and Range Science, Univ. of California, Davis, CA. Received 13 Apr. 2006. *Corresponding author (pcolunga@cicy.mx).

Published in Crop Sci. 47:58–66 (2007).
Plant Genetic Resources
doi:10.2135/cropsci2006.04.0241

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: CEQROO, central eastern Quintana Roo; MCMC, Monte Carlo Markov Chain; NECAMP, northeast Campeche; SEYUC, southeast Yucatan; SSR, simple sequence repeat; SYUC, southern Yucatan.

(Martínez-Castillo et al., 2004). Martínez-Castillo et al. (2006) documented high microsatellite diversity levels in these wild populations in comparison with those reported for other areas in Mesoamerica using microsatellite and isozyme data (Baudoin et al., 2004; Ouédraogo and Baudoin, 2002; Zoro Bi et al., 2003). Weedy *P. lunatus* individuals have also been documented in the region and may be the product of gene flow between wild and domesticated types (Ballesteros, 1999; Debouck, 1979; Martínez-Castillo et al., 2004). For the purposes of the present study weedy individuals are defined as wild plants growing within cultivated *P. lunatus* areas—these are not sown or harvested because their seeds are inedible—and have an intermediate seed size between ones from wild plants and landraces. In fact, farmers growing *P. lunatus* refer to the wild and weedy plants using the same Mayan term: *ib cho*.

Considering that gene flow is a key evolutionary factor affecting the structure of the genetic diversity of domesticate species and wild ancestors in their centers of origin, and the importance of the Yucatan Peninsula to Mesoamerican gene pool of *P. lunatus*, the present study analyzed the data from nine microsatellite loci to define the magnitude and direction of recent and long-term gene flow and genetic structure in the lima bean (*P. lunatus*) wild–weedy–domesticated complex under traditional agricultural conditions on the Yucatan Peninsula, Mexico.

The microsatellite data of the 11 wild populations reported by Martínez-Castillo et al. (2006) were integrated with novel data to perform the analyses.

MATERIALS AND METHODS

Study Regions and Populations

The study was conducted in four traditional agricultural regions on the Yucatan Peninsula, where lima bean is an important crop and may be found at different degrees of agricultural intensification, defined by the number of fallow years, relative levels of agrochemical use, and degree of integration into external markets. A minimal level of intensification means that 20% of the farmers use agrochemicals and trade lima bean seeds out of their towns, moderate level means that 21 to 50% do it, and high level means that more of the 51% farmers do it. These regions are: (i) the central eastern portion of the state of Quintana Roo (CEQROO), where fallow periods are approximately 15 yr and agrochemical use and integration into markets are minimal; (ii) the southeastern portion of the state of Yucatan (SEYUC), where fallow periods are approximately 10 yr and agrochemical use and integration into markets are minimal; (iii) the northeastern portion of the state of Campeche (NECAMP), where fallow periods range from 4 to 5 yr and agrochemical use and market integration is moderate; and (iv) the southern portion of the state of Yucatan (SYUC), where fallow periods are from 2 to 3 yr long and agrochemical use and market integration are high (Fig. 1).

We studied a total of 24 lima bean populations belonging to the wild (11), weedy (1), and domesticate (12) gene pools. Most populations were at least 2 km apart, except for the weedy population that by definition was growing within a cultivated *P. lunatus* area, the wild population Itzinté that was 5 to 10 m apart from the cultivated population Bolonchén, and two pairs of wild populations (Itzinté-Bolonchén and Xohuayán1-Xohuayán2) that were 0.5 km apart and could be just one biological population. These populations had the following characteristics: (i) two wild populations that included one and two plants, respectively, showing morphological evidence of introgression with domesticate populations in their flower, pod, and seed characteristics; (ii) nine wild populations that did not show evidence of introgression; (iii) one weedy population with inedible seeds and intermediate seed size between wild and cultivated plants; (iv) two domesticate populations that included one and two plants showing morphological evidence of introgression with wild populations in their flower, pod, and seed characteristics, respectively; and (v) 10 domesticate populations that did not show evidence of introgression.

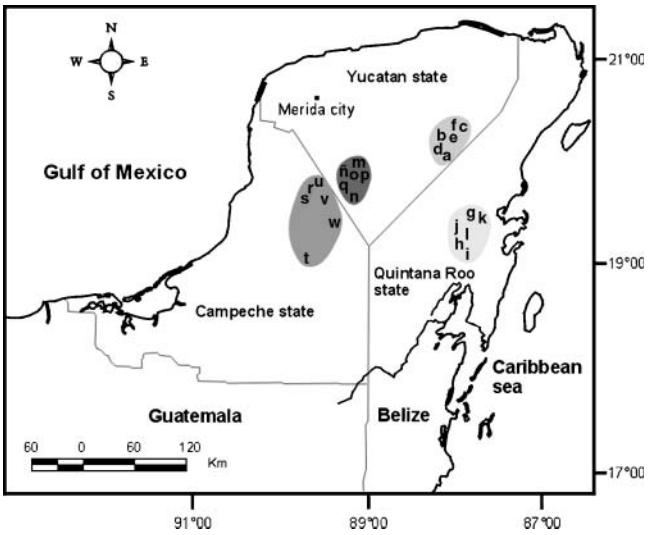

Fig. 1. Study regions (shaded circles) and populations (lowercase). SEYUC populations: San Fernando-1 (a), Weedy (b), Boje (c), San Fernando-2 (d), Marcos (e), and X-Uilub (f); CEQROO populations: Celestino (g), Domingo (h), Julián (i), Nohcá (j), Kik (k) and Holpat (l). SYUC populations: Xohuayán-1 (m), Nohcacab (n), Xohuayán-2 (ñ), Xohuayán-3 (o), Xohuayán-4 (p), and Rubén (q). NECAMP populations: Itzinté (r), Bolonchén-1 (s), Chunchintok (t), Bolonchén-2 (u), Elías (v), and Pascuál (w). Shady intensity indicates degree of agricultural intensification of the region defined by the number of fallow years, relative levels of agrochemical use and degree of integration into external markets.

Plant Material Collection and DNA Extraction

For DNA extraction, we used 20 plants per population, with the exception of the Chunchintok (19 plants), Nohca (14 plants), and the weedy (14 plants) populations. To obtain these plants, we collected seeds from 20 mother plants selected across the entire population distribution range. Each mother plant was separated at least 3 m from each other, to ensure that their seeds represented a distinct family. Five to 10 pods were taken from each mother plant, their seeds mixed, and 10 seeds were randomly selected to be germinated (soaked in liquid N to terminate dormancy). Only one germinated plant per mother plant was selected for DNA extraction. Genomic DNA was obtained from young leaves following the CTAB method (Doyle and Doyle, 1987).

Microsatellite Amplification and Electrophoresis

The microsatellite technique was applied following Gaitán-Solís et al. (2002) using nine pairs of primers reported as polymorphic in *P. lunatus* (Table 1). The microsatellite data of the 11 wild populations reported by Martínez-Castillo et al. (2006) were used in this study.

Table 1. Characteristics of the nine microsatellite loci used in analysis of gene flow in the *P. lunatus* wild–weedy–domesticated complex on the Yucatan Peninsula, Mexico.

Code	SSR sequence	5′ to 3′	Primer sequence	Tm†	NoA‡	RF§
GATS91	$(GA)_{17}$	Left	GAGTGCGGAAGCGAGTAGAG	53	5	218–231
		Right	TCCGTGTTCCTCTGTCTGTG			
AG1	$(GA)_8GGTA(GA)_5$	Left	CATGCAGAGGAAGCAGAGTG	52	7	147–155
		Right	GAGCGTCGTCGTTTCGAT			
BM140	$(GA)_{30}$	Left	TGCACAACACACATTTAGTGAC	55	7	162–173
		Right	CCTACCAAGATTGATTTATGGG			
BM156	$(CT)_{32}$	Left	CTTGTTCCACCTCCCATCATAGC	52	10	205–225
		Right	TGCTTGCATCTCAGCCAGAATC			
BM160	$(GA)_{15}(GAA)_5$	Left	CGTGCTTGGCGAATAGCTTTG	52	4	178–188
		Right	CGCGGTTCTGATCGTGACTTC			
BM164	$(GT)_9(GA)_{21}$	Left	CCACCACAAGGAGAAGCAAC	52	5	135–143
		Right	ACCATTCAGGCCGATACTCC			
BM183	$(TC)_{14}$	Left	CTCAAATCTATTCACTGGTCAGC	52	5	142–148
		Right	TCTTACAGCCTTGCAGACACT			
BM211	$(CT)_{16}$	Left	ATACCCACATGCACAAGTTTGG	52	16	194–219
		Right	CCACCATGTGCTCATGAAGAT			
BM212	$(CA)_{13}$	Left	AGGAAGGGATCCAAAGTCACTC	52	5	191–203
		Right	TGAACTTTCAGGTATTGATGAATGAAG			

† Tm, annealing temperature in °C.
‡ NoA, number of alleles per locus.
§ RF, range of fragment size found in base pairs.

Gene Flow Analysis

To analyze the magnitude and direction of wild–domesticate gene flow, we used two different approaches: genotype assignment methods to analyze recent gene flow and frequency methods to analyze long-term gene flow. The assignment methods are any of several related statistical methods that extract information about migration within the last few generations, from transient disequilibrium observed at individual multilocus genotypes of migrants or individuals recently descended from migrants. The frequency methods are statistical methods that test hypotheses about an event based on the expected frequency of that event happening over a large number of trials (frequency distribution) (Manel et al., 2005), they are based in simplified models of population structure that assumes constant population sizes, symmetrical migration at constant rates and population persistence for periods sufficient to achieve genetic equilibrium (Wright 1931, quoted by Wilson and Rannala, 2003). Assignment methods make fewer assumptions in comparison with indirect estimators for long-term gene flow but are informative only about recent patterns of migration (Wilson and Rannala, 2003). The two approaches are complementary, providing information about gene flow on different timescales.

Genotype Assignment Methods

We used Bayesian methods implemented in the Structure 2.1 program (Pritchard et al., 2000). This program uses a Bayesian clustering approach with Monte Carlo Markov Chain (MCMC) methods and assigns individual genotypes to a predefined number of populations (K) in a given sample (X) to achieve Hardy-Weinberg and linkage equilibriums. This method assumes a model with K populations, each characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned probabilistically to populations, or jointly to two populations if their genotypes indicate they are admixed. Gene flow magnitude and direction were based on the proportion of estimated ancestry of each individual (q) and each population (Q) as calculated by Structure.

Individuals were classified into two categories according to their biological status: wild or domesticated. Weedy individuals were classified as wild. The analysis was done on what were called the Peninsula, interregional, intraregional, and parcel levels. The Peninsula level included a simultaneous analysis of all studied individuals on the Yucatan Peninsula and grouping of them to calculate Q for the different populations where they were collected. The interregional level included simultaneous analysis of all studied populations and grouping of them to calculate Q for eight gene pools by biological status and region. The intraregional level involved separate analyses of populations in the same agricultural region and grouping to calculate Q in two gene pools per region according to biological status. The parcel level involved separate analyses of the populations in the Marcos parcel, where both weedy and domesticate populations grew, and grouping to calculate Q for the two gene pools by biological status. For the Yucatan Peninsula, intraregional and parcel levels, populations were assigned to $K = 2$ gene pools (i.e., wild and domesticate). Populations in the interregional level were assigned to $K = 8$ gene pools, that is, one wild and one domesticate per region. The model was applied using the previous data on the populations option; these data were their geographic location to determine which individuals in the sample were immigrants or had recent immigrant ancestors. Burn-in length was 10^4 and run length was 10^5 to allow the Markov chain to reach stationarity.

Frequency Methods

Two methods were used:

(1) Estimation of Nm [$Nm = 0.25\ (1 - G_{ST})/G_{ST}$] at the Peninsula and intraregional levels was done using the POPGENE 1.31 program (Yeh and Boyle, 1999).
(2) Estimation of mY, which is based on the average coalescence time of genes obtained from within and between parental and admixed populations. This estimator was initially described in Bertorelle and Excoffier (1998) and extended to any number of parental populations by Dupanloup and Bertorelle (2001). The analysis was performed using Admix 2.0 software developed by Dupanloup and Bertorelle http://web.unife.it/progetti/genetica/Isabelle/admix2_0.html (verified 23 Oct. 2006) with 1000 replicates. The admixture model used was based on Papa and Gepts (2003). It considers that both wild and domesticate populations consist of two subpopulations: "true" wild (P_W) and domesticated (P_D) types, without introgression from their domesticated or wild counterparts; and hybrid wild (P_{hyW}) and do-

mesticated (P_{hyD}) populations, with introgression. Each hybrid population consists of N $(1 - \mu)$ loci randomly obtained from a parent population and $N\mu$ loci from the other population: $P_{hyW} = \mu_2 P_D + (1 - \mu_2)P_W$ and $P_{hyD} = \mu_1 P_W + (1 - \mu_2)P_D$, where μ is the parent population's contribution to the hybrid population. This allows comparison of μ_1 (contribution of P_W to P_{hyD}) and μ_2 (contribution of P_D to P_{hyW}). The P_D population consisted of 57 individuals from the Marcos, X-Uilub, Xohuayán-3, and Rubén domesticated populations, excluding individuals manifesting introgression (one from Marcos and two from X-Uilub). The P_W population consisted of 38 individuals from the Holpat and Xohuayán-1 populations, excluding two individuals manifesting introgression in Xohuayán-1. The P_{hyD} consisted of 217 individuals from the San Fernando-2, Marcos, X-Uilub, Celestino, Domingo, Julián, Xohuayán-3, Xohuayán-4, Bolonchén-2, Elias, and Pascual domesticated populations. The P_{hyW} consisted of 208 individuals from the San Fernando-1, Boje, Nohca, Kik, Xohuayán-1, Xohuayán-2, Nohcacab, Itzinté, Bolonchén-1, and Chunchintok wild populations as well as the one weedy population. Selection of the individuals to create the P_{hyD} and P_{hyW} populations was done based on the Bayesian clustering results, the classification of the seeds by the Mayan farmers, and the morphological data collected in situ.

Genetic Structure

To analyze the differentiation among populations in the wild–weedy–domesticated complex, two statistical procedures were used: (i) The G_{ST} statistic ($G_{ST} = H_T - H_S/H_T$, where H_T is the total genetic diversity in the pooled populations and H_S is the diversity within each population; Nei, 1973) was estimated using POPGENE 1.31, and (ii) an analysis of molecular variance (AMOVA) (Excoffier et al., 1992) using ARLEQUIN ver. 2.0 program (Schneider et al., 2000). Both analyses were made considering three levels: the Peninsula (considering all wild, weedy, and domesticated populations), intraregional (considering only the wild, weedy—where available—and domesticated populations by each agricultural region), and parcel (weedy and Marcos domesticated populations).

RESULTS

Recent Gene Flow

At the Peninsula level, Bayesian clustering analysis showed that most of the wild populations were subjected to gene flow from the domesticate gene pool (Fig. 2, Table 2), with the highest Q values in the Bolonchén (0.513) and Itzinté (0.167) populations in NECAMP (Table 2). Neither of these populations had morphological evidence of introgression but they did grow a short distance from domesticated populations, the Itzinté population grew just 5 m from domesticated populations. The Chunchintok (0.035) and Boje (0.027) populations had midlevel Q *values* (Fig. 2, Table 2), the latter included two plants with morphological characteristics indicating introgression from the domesticate gene pool, with seeds very similar to those observed in the weedy population. After the Bolonchén population, the weedy population had the second highest Q value (0.370) (Table 2). This population was found growing together with a domesticated population in SEYUC and two types of seeds were collected from it: those with wild-type characteristics and others with domesticate-type characteristics (Fig. 3). Most of the domesticated populations had very low gene flow levels from the wild gene pool (Fig. 2, Table 2). The highest Q values were in Pascual (0.063) and Bolonchén (0.018), located in NECAMP, and Celestino (0.029) in CEQROO (Table 2). No morphological evidence of introgression was noted in the Pascual population and there were no wild populations nearby. Bolonchén was one of the domesticated populations growing next to the Itzinté wild population, but it did not manifest any morphological evidence of introgression. The farmer cultivating the Celestino population reported that wild plants had grown there in the last 3 yr, and one weedy plant with wild-type seeds similar to those in the weedy population was collected there.

The intraregional analysis showed gene flow levels even lower than at the Peninsula level (data not shown). An appreciable gene flow from the domesticate gene pool toward the wild gene pool was only observed in the SEYUC region. In the weedy population one individual showed a q-probability = 0.184 of having a domesticated parent and four individuals showed a q-probability = 0.094 of having a domesticated grandparent. Wild individuals from other regions showed q-probabilities ranging from 0.000 to 0.04 of having a domesticated grandparent. Gene flow was even lower from the wild gene pool toward the domesticate gene pool. In CEQROO, just one individual showed a q-probability = 0.720 of being a wild immigrant,

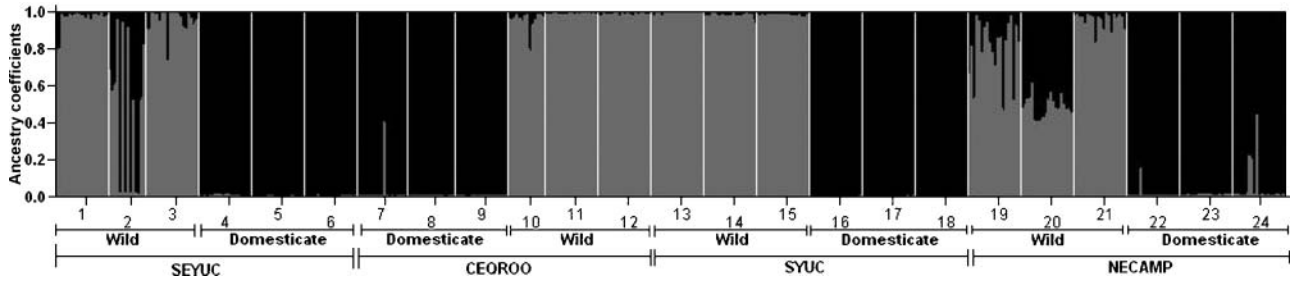

Fig. 2. Coefficients of estimated ancestry per individual (q), grouped by population, biological status and agricultural region. Each individual is represented by a single vertical line broken into two colored segments, with lengths proportional to the individual's estimated ancestry fraction from each of the two biological statuses: Wild (gray) and domesticate (black). Populations: San Fernando-1 (1), Weedy (2), Boje (3), San Fernando-2 (4), Marcos (5), X-Uilub (6), Celestino (7), Domingo (8), Julián (9), Nohca (10), Kik (11), Holpat (12), Xohuayán-1 (13), Nohcacab (14), Xohuayán-2 (15), Xohuayán-3 (16), Xohuayán-4 (17), Rubén (18), Itzinté (19), Bolonchén-1 (20), Chunchintok (21), Bolonchén-2 (22), Elías (23), and Pascual (24).

Table 2. Proportion of estimated ancestry (*Q*) of 24 wild and domesticated *P. lunatus* populations from four agricultural regions on the Yucatan Peninsula, Mexico.

Agricultural region†	Population‡	*Q* domesticate pool	*Q* wild pool
SEYUC	San Fernando (w)	0.006	0.994
	Marcos (we)	0.370	0.630
	Boje (w)	0.027	0.973
	San Fernando (d)	0.992	0.008
	Marcos (d)	0.995	0.005
	X-Uilub (d)	0.994	0.006
CEQROO	Nohca (w)	0.028	0.972
	Kik (w)	0.007	0.993
	Holpat (w)	0.007	0.993
	Celestino (d)	0.971	*0.029*
	Domingo (d)	0.993	0.007
	Julián (d)	0.993	0.007
SYUC	Xohuayán-1 (w)	0.006	0.994
	Nohcacab (w)	0.012	0.988
	Xohuayán-2 (w)	0.011	0.989
	Xohuayán-1 (d)	0.995	0.005
	Xohuayán-2 (d)	0.994	0.006
	Rubén (d)	0.995	0.005
NECAMP	Itzinté (w)	0.167	0.833
	Bolonchén (w)	0.513	0.487
	Chunchintok (w)	0.035	0.965
	Bolonchén (d)	0.982	0.018
	Elias (d)	0.987	0.013
	Pascual (d)	0.937	0.063

† SEYUC, southeastern Yucatan; CEQROO, central eastern Quintana Roo; SYUC, southern Yucatan; NECAMP, northeastern Campeche.
‡ d, domesticated; we, weedy; w, wild.

and in NECAMP only two domesticated individuals showed *q*-probabilities between 0.170 and 0.199 of having a wild grandparent. In SEYUC, one domesticated individual showed a *q*-probability = 0.147 of having a wild grandparent, while the remaining individuals had *q*-probabilities between 0.001 and 0.004. All domesticated individuals in SYUC had *q*-probabilities = 0.001 of having a wild grandparent.

The interregional analysis showed that gene flow does exist between agricultural regions, although at low levels. The SEYUC wild pool had a higher reception of domesticate genes from other agricultural regions, with the CEQROO (Q = 0.004) and NECAMP (Q = 0.003) domesticate gene pools being those contributing the most domesticate alleles (Table 3). One individual from the weedy population in SEYUC had a *q*-probability = 0.072 of having a domesticated parent from CEQROO, while four others had a *q*-probability = 0.057 of the same. In CEQROO, one wild individual showed a *q*-probability = 0.012 of having a domesticated grandparent from NECAMP, and in SYUC another individual showed a *q*-probability = 0.032 for the same. The region with the highest levels of wild gene infiltration toward the domesticate pool was CEQROO, where one domesticated individual had a *q*-probability = 0.022 of belonging to the SEYUC wild pool. In NECAMP, one domesticated individual from the Bolonchén population showed a *q*-probability = 0.072 of having a SEYUC wild grandparent and a *q*-probability = 0.059 of having a CEQROO wild grandparent.

The parcel level analysis also showed low gene flow levels between wild and domesticate gene pools. Just one individual manifested a *q*-probability = 0.006 for having a domesticate parent, while the remaining individuals showed *q*-probabilities ranging from 0.003 to 0.021 for having a domesticate grandparent. Domesticate individuals generally showed lower *q*-probabilities (from 0.004 to 0.005) for having a wild grandparent, though one did show a *q*-probability = 0.001 for having a wild parent and a *q*-probability = 0.064 for having a wild grandparent.

Long-Term Gene Flow

The *Nm* estimator showed relatively low gene flow values at the Peninsula level (Nm = 0.28), as well as at the intraregional level: NECAMP (Nm = 0.51), CEQROO (Nm = 0.42), SEYUC (Nm = 0.31), and SYUC (Nm = 0.31).

The admixture analysis showed the estimated contribution of Pw to P_{hyD} ($m_{WD} = 0.12 \pm 0.02$) to have been less than the estimated contribution of P_D to P_{hyW} ($m_{DW} = 0.34 \pm 0.04$). These values generate a ratio of $m_{DW}/m_{WD} = 2.83$, meaning there was an asymmetrical gene flow almost three times greater from the domesticate pool toward the wild pool.

Genetic Structure

At the Peninsula level, the G_{ST} value was of 0.47. This result was supported by an AMOVA which showed that 45.89% of the total variation was among populations (24.10% among gene pools and 21.79% among populations into each gene pool). At the intraregional level, the G_{ST} values were of 0.33 to NECAMP, 0.37 to CEQROO, and 0.45 to SEYUC and SYUC. These results were supported by an AMOVA which showed that in NECAMP the 37.72% of the total variation was among populations (27.3% among gene pools and 10.42% among populations into each gene pool); in CEQROO was of 46.54% (41.66% among gene pools

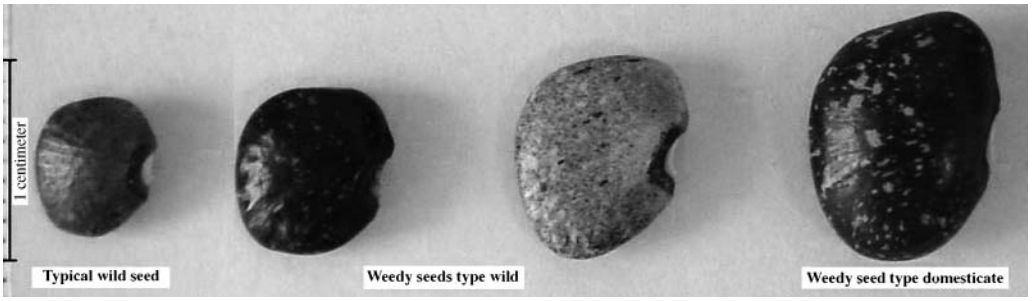

Fig. 3. Types of seed found into the weedy population.

Table 3. Proportion of estimated ancestry (*Q*) of eight wild and domesticate *P. lunatus* gene pools from four agricultural regions on the Yucatan Peninsula, Mexico.

		Q of agricultural region/biological status gene pool							
		SEYUC		CEQROO		SYUC		NECAMP	
Agricultural region†	Biological status‡	w	d	w	d	w	d	w	d
SEYUC	w	***0.988***	0.001	0.001	0.004	0.000	0.002	0.000	0.003
	d	0.000	***0.998***	0.000	0.000	0.000	0.001	0.000	0.001
CEQROO	w	0.000	0.000	***0.998***	0.000	0.000	0.000	0.000	0.000
	d	0.001	0.000	0.005	***0.994***	0.000	0.000	0.000	0.000
SYUC	w	0.001	0.000	0.001	0.000	***0.997***	0.000	0.000	0.000
	d	0.000	0.001	0.000	0.000	0.000	***0.998***	0.000	0.000
NECAMP	w	0.002	0.000	0.001	0.000	0.001	0.000	***0.995***	0.000
	d	0.001	0.000	0.001	0.001	0.000	0.000	0.000	***0.997***

† SEYUC, southeastern Yucatan; CEQROO, central eastern Quintana Roo; SYUC, southern Yucatan; NECAMP, northeastern Campeche.
‡ d, domesticate; w, wild.

and 4.54% among populations into each gene pool); in SEYUC was of 49.17% (43.8% among gene pools and 5.37% among populations into each gene pool); and in SYUC was of 51.49% (44.7% among gene pools and 6.79% among populations into each gene pool). At the parcel level, the G_{ST} value was of 0.29. It was supported by an AMOVA which showed that 38.33% of the total variation was among weedy and domesticated populations.

DISCUSSION

Recent Gene Flow

The highest magnitude of gene flow was observed at the Peninsula level. The wild populations with the highest *Q* values were those growing near or within domesticated populations. The domesticated populations with the highest *Q* values were those that included weedy plants or those growing very near a wild population. The estimated magnitude of gene flow was much less at the interregional, intraregional, and parcel levels, highlighting the importance of contributions from genes from extra-regional populations. This shows the significance of the movement of seeds between Mayan farmers from different areas and the risks of introduction of domesticate alleles to wild populations at the Peninsula level.

The low observed recent gene flow levels at the interregional, intraregional, and parcel levels were likely due to the limited outcrossing potential of *P. lunatus*. This is correlated with its short life cycle, the predominance of self-pollination and its limited ability for pollen and seed dispersal (Maquet et al., 1997). Though crossing rates of up to 48% have been reported, the synchronized ripening of pollen grains and the stigma, as well as their proximity in the bud, favor autogamy in *P. lunatus* (Baudoin et al., 1998). These authors reported that horizontal pollen and seed transference did not exceed 6 m. Neighborhood size in wild populations in the Central Valley of Costa Rica was 1.6 m.

The highly asymmetrical gene flow in the *P. lunatus* wild–weedy–domesticated complex, almost three times higher from the domesticate to the wild gene pool than vice versa, can be explained for the main characteristic of the Mayan traditional agriculture on the Yucatan Peninsula: the migratory-recurrent nature of the swidden system. This characteristic, combined with the existence of wild *P. lunatus* seed banks in the soil, can favor or limit genetic contact between wild and domesticated populations. Mayan farmers on the Peninsula cultivate their plots for 1 to 3 yr and then leave them fallow for 5 to 15 yr (Lazos-Chavero, 1995; Ku-Naal, 1995; Remmers and Ucan, 1996). If wild *P. lunatus* seeds are in the soil, they will germinate when farmers cut and burn the vegetation for a new agricultural cycle, leading to sympatric growth with domesticated populations and thus increasing the possibility of introgression between the two gene pools.

Papa and Gepts (2003) and Papa et al. (2005) suggest two other factors that may explain this asymmetry in *P. vulgaris* L.: (i) the smaller size or lower density of wild populations compared to domesticated populations; and (ii) the role of producers in seed selection. Though hand weeding is still common on the Yucatan Peninsula, increasing use of herbicides is leading to drastic reductions in the density of wild populations, a greater pollen production by the domesticate pool relative to that of the wild population, and consequently a higher pollen flow toward the wild pool. Seed selection also clearly favors the domesticate pool. Only 1 to 14 weedy individuals were found in agricultural parcels containing them, in contrast to the hundreds of domesticated individuals present there. In cultivated environments, farmers easily recognize and select against F_1 domesticate $\times$ wild hybrids because their seeds have generally an intermediate size between those of the parents and a different color from the domesticated maternal parent (Papa and Gepts 2003; Papa et al., 2005). This was observed in the three domesticated populations containing weedy plants. Both factors reduce the probability of wild gene introgression into the domesticate gene pool. Natural F_1 wild $\times$ domesticate hybrids, in contrast, can be better adapted due to their hybrid vigor and the overall dominance of wild-type traits (Singh et al., 1995; Papa et al., 2005), favoring later recombination and introgression of domesticate alleles into the wild pool (Papa and Gepts 2003; Papa et al., 2005).

In addition to morphological recognition of wild–domesticate *P. lunatus* hybrids, Mayan farmers can also distinguish and select against them based on seed flavor. Wild seeds contain high concentrations of linamarine, a cyanogenic compound that makes them inedible (Maquet, 1991). In cases of introgression, hybrid seeds

acquire a bitter taste that is easily detected, leading farmers to dispose of the harvest.

Another factor that may be further limiting the entrance of wild genes into the domesticate gene pool is selection for external markets. Regional markets on the Yucatan Peninsula currently prefer white-seed landraces, favoring elimination of hybrid seeds of different colors (Martínez-Castillo et al., 2004). This may be the case in SYUC, where a dominant selection criterion is focused on production of seed for sale. Farmers in this region report the intentional elimination of wild populations with herbicides to avoid mixing with their landraces and attain a better price (Martínez-Castillo et al., 2004). This may explain why SYUC domesticated populations have a lower degree of genetic infiltration (Fig. 2, Table 2).

Certain aspects of the Mayan traditional agriculture in the region, however, favor the entrance of wild genes into domesticated populations (Martínez-Castillo et al., 2004): (i) hand weeding (*lochepak* in Mayan) allows wild and weedy plants to reach the flowering stage at the same time as domesticated populations because it eliminates only the aerial part of the plant, allowing the subsequent recovery: (ii) wild populations growing near domesticated ones are tolerated when they do not affect the correct development of their crops, as was the case in NECAMP and; (iii) cultivation for subsistence purposes includes up to seven different types of landraces that, after hybridizing, create a wide variety of seed shapes, sizes, and colors that can hide the presence of weed seeds. This may be the case in the Pascual domesticated population, where molecular analysis showed genetic infiltration from wild genes, though no morphological evidence for introgression was observed. Fourth, women and children, who may not readily recognize weedy *P. lunatus* seeds, sometimes participate in agricultural activities, as is the case in the Domingo population. This would explain why seeds harvested as domesticated in this population included seeds very morphologically similar to weedy seeds, though no molecular evidence for introgression was observed there. This may also explain why reports for *P. vulgaris* in Costa Rica contrast with the present results in that gene flow appears to move from the wild toward the domesticate gene pool in that country, as was found by González-Torres (2004) using chloroplast DNA markers. Another explication about the findings of González-Torres (2004) in contrast with ours could be that chloroplast introgression occurs predominantly from wild to domesticate gene pool, whereas the introgression of nuclear genes, as our microsatellite data, is predominantly from the domesticate to wild gene pool (Papa and Gepts, 2003; Chacón et al., 2005).

Long-Term Gene Flow

Low levels of long-term gene flow were detected at the Peninsula and intraregional levels, which coincides with data reported for wild populations in Costa Rica (Hardy et al., 1997; Maquet et al., 2001; Ouédraogo and Baudoin, 2002). This may be explained by the joint action of limited recent gene flow and continuous selective pressure exercised by Mayan farmers against wild progenitor hybrids and retrocrosses; in other words, the adaptive disadvantages intrinsic to this agricultural system.

Genetic Structure

The genetic differentiation in the wild–weedy–domesticated complex of *P. lunatus* was high at the different levels analyzed, even at the parcel level (G_{ST} = 0.29) where the weedy and domesticated plants grew very close to each other (inclusive on the same maize [*Zea mays* L.] plant). It could be a result of the low levels of gene flow between the wild and domesticate gene pools, as it was indicated for the AMOVA analyses that showed higher levels of differentiation among gene pools than among populations from the same gene pools.

CONCLUSIONS

The findings reported here are very important for the conservation and biosafety of domesticated and wild *P. lunatus* populations within their Mesoamerican center of domestication and diversity. Even with the low levels of gene flow found in this study, the asymmetrical gene flow from the domesticate to the wild gene pool may create a drastic reduction in the genetic diversity of wild populations and even lead to local extinctions. This in turn could affect the genetic diversity of the domesticate gene pool and the availability of agriculturally interesting genes for plant breeders. In addition, many of the characteristics incorporated into domesticated plants using traditional improvement methods (e.g., lack of seed latency, dwarfing, and dependence on nutrient-rich soils) are maladaptive for wild plants (Ellstrand and Hoffman, 1990), meaning hybrids between domesticated forms and their wild parents may be poorly adapted to uncultivated environments, thus diminishing or even preventing transference of domesticate genes within natural populations (Doebley, 1992; National Research Council, 1989). However, the characteristics genetically transferred by genetic engineering (e.g., herbicide, pest, and disease resistance) may provide an adaptive advantage to wild plants (Gasser and Fraley, 1989). If these characteristics are introduced into this crop by genetic engineering, domesticate–weedy hybrids could threaten the current host–pest balance (Ellstrand and Hoffman, 1990; Rissler and Mellon, 1993). The problem becomes even more complex taking into account gene flow between pools from different agricultural regions through deliberate or accidental seed movement.

Finally, in contrast with González-Torres (2004), the results reported here and in other studies (Papa and Gepts, 2003; Zizumbo-Villarreal et al., 2005) highlight the importance of determining the magnitude and direction of gene flow between domesticate species and their wild parents for every species and environment before stating its possible implications.

ACKNOWLEDGMENTS

This research was done in the Plant Genetic Resources Diversity and Molecular Evolution Laboratory of the Depart-

ment of Natural Resources-CICY, as part of the first author's doctoral dissertation under the direction of Patricia Colunga-GarcíaMarín and with Daniel Zizumbo-Villarreal, Hugo Perales-Rivera, and Paul Gepts as advisors. The authors thank Filogonio May-Pat for assistance in the fieldwork, UC MEXUS-CONACYT and SINAREFI-SAGARPA for supplementary financial support, and the CONACYT for the first author's Ph.D. scholarship.

REFERENCES

Altieri, M.A., and C. Montecinos. 1993. Conserving crop genetic resources in Latin America through farmer's participation. p. 45–64. *In* S. Christopher et al. (ed.) Perspectives on biodiversity: Case studies of genetic resource conservation and development. American Association for the Advancement of Science, Washington, D.C.

Ballesteros, G.A. 1999. Contribuciones al conocimiento del frijol lima (*Phaseolus lunatus* L.) en América Tropical. Ph.D. diss. Colegio de Posgraduados. Montecillos, Estado de México, Mexico.

Baudet, J.C. 1977. The taxonomic status of the cultivated types of lima bean (*Phaseolus lunatus* L.). Tropical Grain Legume 7:29–30.

Baudoin, J.P., J. Degreef, O. Hardy, F. Janart, and I. Zoro Bi. 1998. Development of an *in situ* conservation strategy for wild lima bean (*Phaseolus lunatus* L.) populations in the central valley of Costa Rica. p. 417–426. *In* S.J. Owens and P.J. Rudall (ed.). Reproduction biology. Royal Botanic Garden Press, Kew, UK.

Baudoin, J.-P., O. Rocha, J. Degreef, A. Maquet, and L. Guarino. 2004. Ecogeography, demography, diversity and conservation of *Phaseolus lunatus* L. in the Central Valley of Costa Rica. Systematic and ecogeographic studies on crop genepools 12. International Plant Genetic Resources Institute, Rome, Italy.

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

Chacón, M.I., B. Pickersgill, and D.G. Debouck. 2005. Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated landraces. Theor. Appl. Genet. 110:432–444.

Debouck, D.G. 1979. Proyecto de recolección de germoplasma de Phaseolus en México. CIAT-INIA, Centro Internacional de Agricultura Tropical (CIAT), Colombia.

Debouck, D.G., J.H. Liñan Lara, A. Campana Sierra, and J.H. De la Cruz Rojas. 1987. Observations on the domestication of *Phaseolus lunatus* L. FAO/IBPGR Plant Genet. Resour. Newsl. 70:26–32.

Doebley, J. 1992. Molecular systematics and crop plant evolution. p. 202–222. *In* D.E. Soltis et al. (ed.) Plant molecular systematics. Chapman and Hall, New York.

Doyle, J., and J. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11–15.

Dupanloup, I., and G. Bertorelle. 2001. Inferring admixture proportions from molecular data: Extension to any number of parental populations. Mol. Biol. Evol. 18(4):672–675.

Ellstrand, N., and C. Hoffman. 1990. Hybridization as an avenue for the escape of engineered genes. Bioscience 40:438–442.

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

Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distance among DNA haplotypes: Applications to human mitochondrial DNA restriction data. Genetics 131:479–491.

Gaitán-Solís, E., M.C. Duque, K.J. Edwards, and J. Tohme. 2002. Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterization, and cross-species amplification in *Phaseolus* ssp. Crop Sci. 42:2128–2136.

Gasser, C.S., and R.T. Fraley. 1989. Genetically engineered plants for crop improvement. Science 244:1293–1299.

Gepts, P., and R. Papa. 2003. Possible effects of (trans) gene flow from crops on the genetic diversity from landraces and wild relatives. Environ. Biosafety Res. 2:89–103.

González-Torres, R.I. 2004. Estimación de flujo de genes en *Phaseolus vulgaris* L. mediante marcadores moleculares: Microsatélites y polimorfismo de ADN de cloroplasto. Master's thesis. Universidad Nacional de Colombia, Bogotá, Colombia.

Gutiérrez-Salgado, A., P. Gepts, and D.G. Debouck. 1995. Evidence for two gene pools of the lima beans, *Phaseolus lunatus* L., in the Americas. Genet. Resour. Crop Evol. 42:15–28.

Hails, R.S. 2000. Genetically modified plants—The debate continues. Trends Ecol. Evol. 15:14–18.

Hardy, O., S. Dubois, I. Zoro Bi, and J.-P. Baudoin. 1997. Gene dispersal and its consequences on the genetic structure of wild populations of lima bean (*Phaseolus lunatus*) in Costa Rica. Plant Genet. Resour. Newsl. 109:1–6.

Harlan, J.R. 1965. The possible role of weedy races in the evolution of cultivated plants. Euphytica 14:173–176.

Jarvis, D.I., and T. Hodgkin. 1999. Wild relatives and crop cultivars: Detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. Mol. Ecol. 8:S159–S173.

Ku-Naal, R. 1995. Cambios técnicos en la milpa bajo roza-tumba-quema en Yaxcabá, Yucatán. p. 401–418. *In* X.E. Hernández et al. (ed.) La milpa en Yucatán: Un sistema de producción agrícola tradicional. Colegio de Postgraduados, México.

Lazos-Chavero, E. 1995. La milpa en el sur de Yucatán: Dinámica y crisis. p. 35–86. *In* La milpa en Yucatán: Un sistema de producción agrícola tradicional. Colegio de Postgraduados, México.

Manel, S., O.E. Gaggiotti, and R.S. Maples. 2005. Assignment methods: Matching biological questions with appropriate techniques. Trends Ecol. Evol. 20(3):136–142.

Maquet, A. 1991. Lima bean (*Phaseolus lunatus* L.) catalogue. Working document No. 80. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.

Maquet, A., A. Gutiérrez, and D.G. Debouck. 1990. Further biochemical evidence for the existence of two gene pools in lima bean. Ann. Rep. Bean Improv. Coop. 33:128–129.

Maquet, A., B. Masumbuko, M. Ouedraogo, I. Zoro Bi, and J.P. Baudoin. 2001. Estimation of gene flow among wild populations of *Phaseolus lunatus* L. using isozyme markers. Ann. Rep. Bean Improv. Coop. 44:27–28.

Maquet, A., I. Zoro Bi, M. Delvaux, B. Wathelet, and J.P. Baudoin. 1997. Genetic structure of a lima bean base collection using allozyme markers. Theor. Appl. Genet. 95:980–991.

Martínez-Castillo, J., D. Zizumbo-Villarreal, H. Perales-Rivera, and P. Colunga-GarcíaMarín. 2004. Intraspecific diversity and morphophenological variation in *Phaseolus lunatus* L. from the Yucatan Peninsula, Mexico. Econ. Bot. 58:354–380.

Martínez-Castillo, J., D. Zizumbo-Villarreal, P. Gepts, P. Delgado-Valerio, and P. Colunga-GarcíaMarín. 2006. Structure and genetic diversity of wild populations of lima bean (*Phaseolus lunatus* L.) from the Yucatan Peninsula, Mexico. Crop Sci. 46:1071–1080.

National Research Council. 1989. Field testing genetically modified organisms: Framework for decisions. National Academy Press, Washington, DC.

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

Ouédraogo, M., and J.P. Baudoin. 2002. Comparative analysis of genetic structure and diversity in wild lima bean populations from the Central Valley of Costa Rica, using microsatellite and isozyme markers. Ann. Rep. Bean Improv. Coop. 45:240–241.

Papa, R., J. Acosta, A. Delgado-Salinas, and P. Gepts. 2005. A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. Theor. Appl. Genet. 111:1147–1158.

Papa, R., and P. Gepts. 2003. Asymmetry of gene flow and differential geographical structure of molecular diversity on wild and domesticated common bean (*Phaseolus lunatus* L.) from Mesoamerica. Theor. Appl. Genet. 106:239–250.

Payró de la Cruz, E., P. Gepts, P. Colunga-GarcíaMarín, and D. Zizumbo-Villarreal. 2005. Spatial distribution of genetic diversity in wild populations of *Phaseolus vulgaris* L. from Guanajuato and Michoacán, Mexico. Genet. Resour. Crop Evol. 52:589–599.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Quirós, C.F., R. Ortega, and L.W.D. Van Raamsdonk. 1992. Amplification of potato genetic resources in their center of diversity: The role of natural outcrossing and selection by the Andean farmer. Genet. Resour. Crop Evol. 39:107–113.

Remmers, G.G.A., and E. Ucan. 1996. La roza-tumba-quema maya:

Un sistema agroecológico tradicional frente al cambio tecnológico. Etnoecología 3:97–109.
Reyes, G.D., and G. Aguilar. 1992. Intensificación de la milpa en Yucatán. p. 347–358. *In* D. Zizumbo V. et al. (ed.) La modernización de la milpa en Yucatán: Utopía o realidad. CICY-DANIDA, Mérida, Mexico.
Rissler, J., and M. Mellon. 1993. Perils amidst the promise: Ecological risks of transgenic crops in a global market. Union of Concerned Scientists, Cambridge, MA.
Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
Singh, S.P., A. Molina, and P. Gepts. 1995. Potential of wild common bean for seed yield improvement of cultivars in the tropics. Can. J. Plant Sci. 75:807–813.
Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science 236:787–792.
Snow, A. 2002. Transgenic crops– why gene flow matters. Nat. Biotechnol. 20:542.
Stebbins, G.L. 1959. The role of hybridization in evolution. Proc. Am. Philos. Soc. 103:231–251.
Webster, B.D., S.P. Lynch, and C.L. Tucker. 1979. A morphological study of the development of reproductive structures of *Phaseolus lunatus* L. J. Am. Soc. Hortic. Sci. 104:240–243.
Wilson, G.A., and B. Rannala. 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163: 1177–1191.
Wright, S. 1931. Evolution in mendelian populations. Genetics 16: 97–159.
Yeh, F.C., and T.J.B. Boyle. 1999. Popgene version 1.31. Microsoft Windows-based freeware for population analysis. Univ. of Alberta and Centre for Int. Forestry Res., Edmonton, AB.
Zizumbo-Villarreal, D., P. Colunga-GarcíaMarín, E. Payró de la Cruz, P. Delgado-Valerio, and P. Gepts. 2005. Population structure and evolutionary dynamics of wild–weedy–domesticated complexes of common bean in a Mesoamerican region. Crop Sci. 45:1073–1083.
Zoro Bi, I., A. Maquet, and J.-P. Baudoin. 2003. Population genetic structure of wild *Phaseolus lunatus* (Fabaceae), with special reference to population sizes. Am. J. Bot. 90:897–904.
Zoro Bi, I., A. Maquet, and J.-P. Baudoin. 2005. Mating system of wild *Phaseolus lunatus* L. and its relationships with population size. Heredity 94:153–158.