

*Domestication of small-seeded lima bean (Phaseolus lunatus L.) landraces in Mesoamerica: evidence from microsatellite markers*

**Rubén H. Andueza-Noh, Jaime Martínez-Castillo & María I. Chacón-Sánchez**

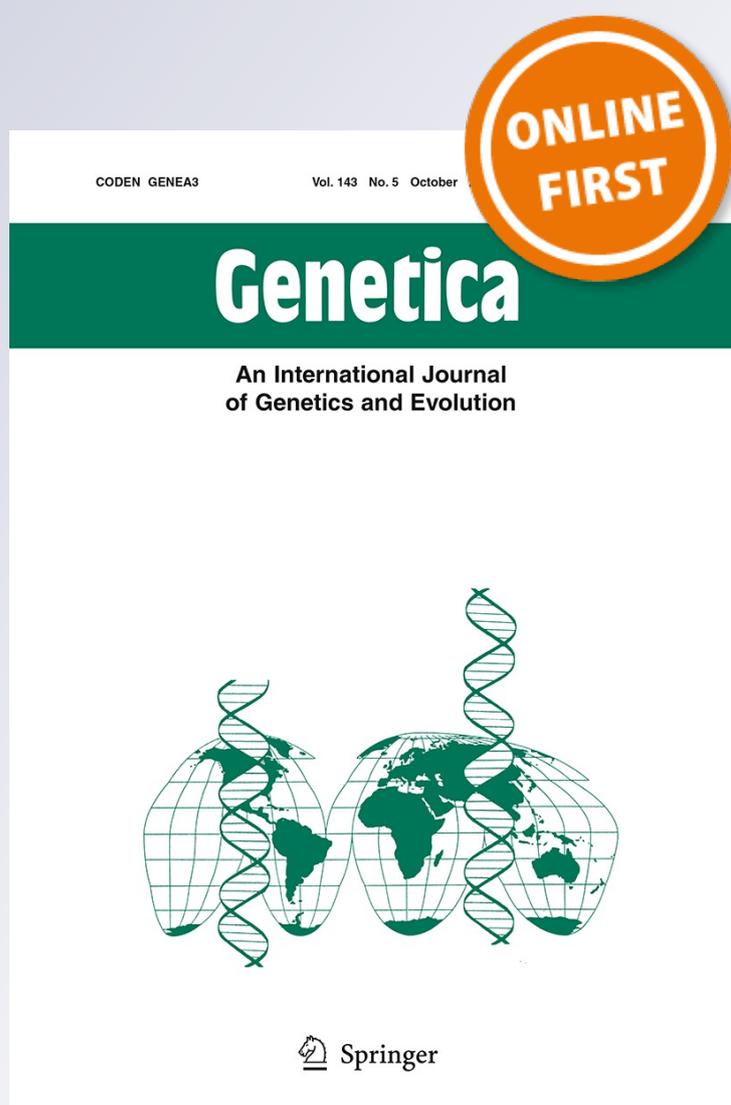
**Genetica**

An International Journal of Genetics and Evolution

ISSN 0016-6707

Genetica

DOI 10.1007/s10709-015-9863-0



**Your article is protected by copyright and all rights are held exclusively by Springer International Publishing Switzerland. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Domestication of small-seeded lima bean (*Phaseolus lunatus* L.) landraces in Mesoamerica: evidence from microsatellite markers

Rubén H. Andueza-Noh<sup>1</sup> · Jaime Martínez-Castillo<sup>2</sup> · María I. Chacón-Sánchez<sup>3</sup>

Received: 21 May 2015 / Accepted: 13 September 2015  
© Springer International Publishing Switzerland 2015

**Abstract** Previous studies have suggested that the Mesoamerican small-seeded landraces of Lima bean may have been domesticated more than once in Mesoamerica, once in central-western Mexico and another one in an area between Guatemala and Costa Rica. However, these findings were based on sequencing of only one locus from nuclear DNA, and additional confirmation was needed. Here we contribute with additional data on the origin of the Mesoamerican landraces and document the founder effect due to domestication. We characterized 62 domesticated, 87 wild and six weedy Lima bean accessions with ten microsatellite loci. Genetic relationships were analyzed using genetic distances and Bayesian clustering approaches. Domestication bottlenecks were documented using inter-population comparisons and  $M$  ratios. The results support at least one domestication event in the area of distribution of gene pool MI in central-western Mexico and also show that some landraces are genetically related to wild accessions of gene pool MII. Also, our data support founder effects due to domestication in Mesoamerican Lima bean landraces.

Although we could not establish more specifically the place of origin of the Mesoamerican Lima bean landraces, our results show that these are not a genetically homogeneous group, a finding that may be compatible with a scenario of more than one domestication event accompanied by gene flow. The complex genetic makeup of landraces that we found indicates that a more comprehensive geographic and genomic sampling is needed in order to establish how domestication processes and gene flow have shaped the current genetic structure of landraces.

**Keywords** Crop domestication · Bayesian clustering · Genetic distance · Founder effect · SSR markers · Wild ancestors

## Introduction

Mesoamerica has long been recognized as a domestication center of origin of many of the most important crops worldwide, among them maize, peppers and beans. The area of domestication of several of these crops has been pinpointed within Mexico, especially in central-western Mexico where wild relatives are distributed and where early agriculture activities took place (Pickersgill 1969; Matsuoka et al. 2002; Aguilar-Meléndez et al. 2009; Kwak and Gepts 2009; Zizumbo-Villarreal and Colunga-García-Marín 2010). One of the species that was domesticated in Mexico is Lima bean (*Phaseolus lunatus* L.). The oldest archeological remains of domesticated Lima beans in the Andes date back to 3500 years before present (YBP) from Guitarrero, Peru, and 5600 YBP from Chilca, Peru, both in a pre-ceramic context (Kaplan and Lynch 1999). In Mesoamerica, the oldest records date back to 1300 YBP from Dzibichaltún, Yucatán, Mexico (Kaplan 1965).

**Electronic supplementary material** The online version of this article (doi:10.1007/s10709-015-9863-0) contains supplementary material, which is available to authorized users.

✉ María I. Chacón-Sánchez  
michacons@unal.edu.co

<sup>1</sup> Instituto Tecnológico de Conkal, Av. Tecnológico S/N, Conkal, Yucatán, Mexico

<sup>2</sup> Centro de Investigación Científica de Yucatán, Calle 43 No. 130, Colonia Chuburná de Hidalgo, 97200 Mérida, Yucatán, Mexico

<sup>3</sup> Universidad Nacional de Colombia - Bogotá, Facultad de Ciencias Agrarias - Departamento de Agronomía, Carrera 30 No. 45-03 - Edificio 500, Bogotá, D.C. 111321, Colombia

Wild populations of Lima bean are widely distributed, from northern Mexico to northern Argentina (Freitag and Debouck 2002; Debouck 2008). Within Mesoamerica, wild populations are found in Mexico and all countries of Central America. Within Mexico, they are found along the Pacific side of the mountain systems of central-western Mexico, the coastal plains of the Gulf of Mexico, the Peninsula of Yucatan and in Chiapas. Wild populations have also been reported in Guatemala, Belize, El Salvador, Honduras, Nicaragua, Costa Rica and Panama, and the Caribbean islands of Cuba, Jamaica, Santo Domingo, Trinidad and Tobago, and Puerto Rico. In South America, they have been found in Colombia, Venezuela, Ecuador, Peru, Argentina, and probably in Bolivia and Brazil. Previous studies have suggested that wild Lima beans within Mesoamerica are not one genetically homogeneous group, but instead they are structured into two main gene pools, divergent geographically. These are called Mesoamerican I (MI) distributed mainly in central-western Mexico, and Mesoamerican II (MII) distributed along the Gulf of Mexico, Peninsula of Yucatan and Central America (Motta-Aldana et al. 2010; Serrano–Serrano et al. 2010). Recently, Martínez-Castillo et al. (2014) reported the possible existence of two further groups within the gene pool MI, called MIa and MIb, with no overlapping geographic distributions. While MIa is found along western Mexico from the state of Sinaloa to Oaxaca, MIb is restricted to the states of Jalisco, Colima and Morelos in central-western Mexico.

While Mesoamerica has been pinpointed as the place of origin of the Mesoamerican landraces characterized by having small seeds, the place of origin of the large-seeded landraces is the inter-Andean valleys on the western slope of the Andes of Ecuador and northern Peru (Debouck, et al. 1989; Gutiérrez-Salgado et al. 1995; Fofana et al. 2001; Martínez-Castillo et al. 2004; Motta-Aldana et al. 2010). Small-seeded landraces nowadays are found not only in Mesoamerica but also in several countries of South America, including Brazil, and the Caribbean. For the small-seeded landraces, recent studies have suggested that at least one domestication event took place in central-western Mexico, within the range of the MI gene pool (Motta-Aldana et al. 2010; Serrano–Serrano et al. 2012; Andueza-Noh et al. 2013). This result was somehow unexpected given that today the area of major cultivation activity and major landrace diversity in Mexico is the Peninsula of Yucatan, not central-western Mexico. In the Peninsula of Yucatan, Mayan communities play an important role in conserving landrace diversity, which is also increased by events of gene flow with local wild populations (Martínez-Castillo et al. 2004). A second domestication event for Mesoamerican landraces could have occurred in an area between Guatemala and Costa

Rica from the MII gene pool, on the basis of chloroplast DNA polymorphisms (Andueza-Noh et al. 2013).

The postulation of at least two domestication events for the Mesoamerican landraces, one from the gene pool MI and another one from the gene pool MII, was based on the sequencing of only one locus of the nuclear DNA [the internal transcribed spacer of the ribosomal DNA of the nuclear genome (ITS)] and two intergenic spacers of the chloroplast DNA (cpDNA). Therefore, further confirmation is needed from more loci in the nuclear DNA. The objective of the present study was to assess the hypothesis of two domestication events for the small-seeded Mesoamerican landraces. For doing this, a larger sample of wild and domesticated accessions was analyzed with a set of ten microsatellite loci that proved to be useful in previous studies (Martínez-Castillo et al. 2006, 2007).

## Materials and methods

### Plant material

A total of 155 Lima bean accessions was analyzed, 62 of them were domesticated, 87 were wild and six were weedy (Supplementary Table S1). Ninety of these accessions were obtained from the Lima bean collection held at the International Center for Tropical Agriculture—CIAT (G numbers in Supplementary Table S1), 59 accessions were collected in the field during 2009 and 2010 (JMC numbers in Supplementary Table S1) by Dr. Jaime Martínez (Centro de Investigación Científica de Yucatán—CICY), three accessions were obtained from Dr. Rogelio Lépiz (UdeG—Universidad de Guadalajara, Mexico), and three accessions were obtained from Dr. Acosta (INIFAP, Celaya, Mexico). The accessions were selected to represent the geographic range of wild and domesticated Lima beans in the Mesoamerican gene pool. The five wild accessions from Ecuador and Peru from the Andean gene pool were used in this study as outgroup. Five seeds per accession were analyzed to test intra-accession polymorphism, except for the outgroup accessions where only one or two seeds were used per accession, and in total 757 individuals were analyzed. Accessions from CIAT were obtained through a Material Transfer Agreement. The accessions collected in the field by Dr. Martínez (JMC numbers in Supplementary Table S1) did not require any specific permission since *P. lunatus* is not considered an endangered or protected species. Voucher specimens of the plants collected by Dr. Martínez are deposited in the Herbarium CICY (Martínez-Castillo et al. 2014). During the field trips, seeds were collected from 20 to 30 individuals per population only, taking care of sampling populations with high density of individuals. The geographic coordinates for all the

accessions used in this study can be seen in Supplementary Table S1.

### Molecular analyses

DNA was extracted from young leaflets of all the 757 individuals using the CTAB method (Doyle 1987). Ten SSR loci, which have proved useful in previous studies for Lima beans (Martínez-Castillo et al. 2006, 2007), were used. The names of the loci and the repeat motifs are shown in Table 1. As it can be seen, all of them contain dinucleotide repeat motifs, either in the form of perfect, compound or interrupted microsatellite sequences. Conditions for PCR amplification of SSR loci, polyacrylamide electrophoresis, primer sequences and expected allele sizes have been described elsewhere (Martínez-Castillo et al. 2014).

### Assumptions and rationale of data analyses

In a previous study, genetic structure analyses were conducted on wild populations of Lima beans from Mesoamerica (Martínez-Castillo et al. 2014). Therefore, in the present study we focused on establishing the geographic origin of Mesoamerican landraces and in documenting founder effects due to domestication. For doing this, we based our analyses on two main premises: (1) We assumed that the geographic origin of the landraces is in the area (s) where their wild ancestors are distributed, and (2) we assumed that the wild ancestors have not changed their distribution and genetic composition significantly since the time of domestication. As mentioned before, in Mesoamerica the oldest archeological remains of domesticated Lima beans come from the site known as

Dzibichaltún, Yucatán, with an age of 1300 YBP (Kaplan 1965). This age would indicate a minimum age for Lima bean domestication in Mesoamerica, and in comparison with the domestication of other crops such as maize and squash (Ladizinsky 1985; Pohl et al. 1996; Smith 1997; Piperno and Flannery 2001; Smith 2005; Piperno et al. 2009), this would be a relatively recent event. Under these assumptions, we have the following expectation. Given the relatively recent time for domestication and therefore very little time for post-domestication diversification, the wild ancestors would be those wild populations that are genetically more similar to landraces. However, as we are using microsatellite markers from the nuclear genome, we could have confounding effects of the original domestication patterns due to events of gene flow. With a small set of markers, it might be in some cases challenging to distinguish between domestication and gene flow patterns. Because of this, we will be using haplotype data generated from previous studies (Serrano-Serrano et al. 2012; Andueza-Noh et al. 2013) in this same set of accessions to compare our results.

It is widely believed that domestication events usually involve sampling a few numbers of individuals from wild populations to start new cultivated populations, a process that conveys reduction in the effective population size in the founding populations. This process will be likely reflected in loss of alleles and loss of genetic diversity in the domesticated populations by the stochastic effects of genetic drift. This effect, known as the founder effect of domestication (Ladizinsky 1985), can be measured by comparison of the ancestral wild populations and the domesticated populations looking for significant changes in allele frequencies and various measures of genetic diversity, with the implicit assumption that wild populations

**Table 1** Diversity indexes calculated for the ten SSR loci used in this study

| Locus   | Repeat motif                                | $N$   | $N_a$ | $N_e$ | $I$   | $H_O$ | $H_E$ | $F$   |
|---------|---|-------|-------|-------|-------|-------|-------|-------|
| GATS 91 | (GA) <sub>17</sub>                          | 757   | 10    | 4.193 | 1.634 | 0.042 | 0.762 | 0.944 |
| BM 140  | (GA) <sub>30</sub>                          | 757   | 11    | 6.609 | 2.063 | 0.079 | 0.849 | 0.907 |
| BM 156  | (CT) <sub>32</sub>                          | 757   | 7     | 2.921 | 1.288 | 0.020 | 0.658 | 0.970 |
| BM 164  | (GT) <sub>9</sub> (GA) <sub>21</sub>        | 757   | 8     | 1.899 | 1.079 | 0.030 | 0.473 | 0.936 |
| AG 1    | [(GA) <sub>8</sub> GGTA(GA) <sub>5</sub> ]  | 757   | 9     | 2.564 | 1.240 | 0.009 | 0.610 | 0.985 |
| BM 143  | (GA) <sub>35</sub>                          | 757   | 12    | 3.316 | 1.604 | 0.071 | 0.698 | 0.898 |
| BM 154  | (CT) <sub>17</sub>                          | 756   | 12    | 3.887 | 1.615 | 0.060 | 0.743 | 0.920 |
| BM 211  | (CT) <sub>16</sub>                          | 757   | 12    | 5.967 | 1.982 | 0.049 | 0.832 | 0.941 |
| BM 170  | [(CT) <sub>5</sub> CCTT(CT) <sub>12</sub> ] | 757   | 10    | 3.362 | 1.638 | 0.033 | 0.703 | 0.953 |
| BM 197  | (GT) <sub>8</sub>                           | 757   | 8     | 1.529 | 0.749 | 0.007 | 0.346 | 0.981 |
| Mean    |   | 757   | 9.9   | 3.625 | 1.489 | 0.040 | 0.667 | 0.943 |
| SE      |   | 0.100 | 0.586 | 0.516 | 0.127 | 0.008 | 0.050 | 0.009 |

$N$  sample size in number of individuals,  $N_a$  number of alleles,  $N_e$  effective number of alleles,  $I$  Shannon's information index,  $H_O$  observed heterozygosity,  $H_E$  expected heterozygosity,  $F$  fixation index,  $SE$  standard deviation

have been demographically and genetically stable since the time of domestication. However, wild populations may have also experienced reductions in population size since domestication, therefore making comparisons with their domesticated descendant not straightforward. In this study, we will measure founder effects due to domestication in two different ways: (1) by comparing genetic diversity indexes among wild and domesticated populations, and (2) by looking for evidence of population bottlenecks in both wild and domesticated populations.

### Genetic analyses of data

Bayesian clustering approaches, as implemented in the programs Structure (Pritchard et al. 2000) and Instruct (Gao et al. 2007), were applied to study the genetic structure of the sample of wild, weedy and landrace accessions. The approach in Structure was used to assign individuals to a number of  $K$  populations that are built so that Hardy–Weinberg (HW) and linkage disequilibria are minimized. A  $Q$ -matrix was obtained which shows, in terms of percentages, the global ancestry of each individual to each one of  $K$  populations. For each value of  $K$  (we evaluated  $K$  from 1 to 10), a total of 20 simulations were run. Each simulation consisted of a *burnin* period of one million and one million MCMC (Markov Chain Monte Carlo) steps after *burnin*. During the runs, the variation of the parameters was checked to assure convergence. A single  $Q$ -matrix for each  $K$  and from all the 20 independent simulation runs was obtained using the CLUMPP software (Jakobsson and Rosenberg 2007) and the optimal  $K$  was chosen according to Evanno et al. (2005) by using the STRUCTURE HARVESTER program (Earl 2012). Afterward, a single  $Q$ -matrix per accession was obtained by averaging the ancestry coefficients of each of the five individuals within accessions. Bar plots for the ancestry coefficients of accessions were drawn using the Distruct software (Rosenberg 2004). The simulations were carried out using the admixture model and correlated allele frequencies, a model appropriate for Lima beans that although predominantly autogamous, show a low percentage of outcrossing (about 10 %) (Hardy et al. 1997). Also, the analysis involves domesticated populations that probably have been moved from their area of origin and introduced to other regions where they can experience gene flow with local wild populations (Martínez-Castillo et al. 2007). The correlated allele frequencies model could be more appropriate for our data set because, as expressed above, we expect that domesticated populations and their wild ancestors show correlated allele frequencies due to ancestry.

Because Lima beans are predominantly autogamous, the assumption of HWE within each of the  $K$  populations in the

Structure methodology could not be appropriate; therefore, we also carried out analyses using the software Instruct that implements the approach of the software Structure but without the HWE assumption. In Instruct, we used 100,000 as period of *burnin* and 200,000 MCMC iterations, and five independent simulations for each  $K$  ( $K$  ranged from 1 to 10) for inferring simultaneously population structure and population selfing rates. The optimal  $K$  was chosen according to the deviance information criterion implemented in the software Instruct (Gao et al. 2011).

### Genetic relationships among the $K$ ancestral populations defined in the Bayesian clustering analyses

After clustering accessions into  $K$  ancestral populations, the genetic relationship among these populations was investigated by means of genetic distances. With the use of the program Microsatellite Analyzer, MSA (Dieringer and Schlötterer 2003), we built two different distance matrices based on the Nei's standard genetic distance (Nei 1987) and the chord genetic distance of Cavalli-Sforza and Edwards (Dc) (Cavalli-Sforza and Edwards 1967). These two distances were chosen because they make different assumptions about mutation processes; Nei's distance assumes the infinite allele model of mutation and Dc assumes no mutation processes and attributes all changes in allele frequencies to genetic drift. These distance matrices were used to build neighbor-joining (Saitou and Nei 1987) topologies as implemented in the software Neighbor of the Phylip package (Felsenstein 1989). The topologies were visualized with the software Figtree v. 1.4 (Rambaut and Drummond 2014). Bootstrap support for the clusters was established by means of 1000 permutations of data, using the program MSA and the program Consense of the Phylip package, under the majority rule criterion.

### Genetic diversity and founder effects

Genetic diversity was described and quantified in terms of percentage of polymorphic loci ( $P$ ), average number of alleles per locus or allele richness ( $A$ ), effective number of alleles ( $Ne$ ), average number of private alleles per locus ( $PA$ ), observed heterozygosity ( $H_O$ ) and expected heterozygosity ( $H_E$ ), as implemented in the software GenAlex (Peakall and Smouse 2012) and FSTAT (Goudet 1995).

Founder effects due to domestication were calculated as reduction in genetic diversity in landraces compared to wild ancestors. In order to compare various genetic diversity indexes, the program FSTAT (Goudet 1995) was used to carry out one-sided group comparison tests with one thousand permutations. The group comparisons were

as follows: (1) all wild accessions versus all domesticated accessions, (2) wild ancestors versus domesticated accessions for the gene pool MI and (3) wild ancestors versus domesticated accessions for the gene pool MII. The comparisons (2) and (3) are more appropriate to measure founder effects because they are comparing only those wild populations that are likely to be the source of landraces.

In addition, for detecting bottlenecks in wild and domesticated populations, we estimated the  $M$  ratio of allele number against the range in allele size as implemented in the software  $M$  for microsatellite data (Garza and Williamson 2001). This test is based on the prediction that in a population bottleneck, alleles in low frequency (for example, rare alleles) are likely to be lost in a stochastic manner independently of their size; therefore, it is expected that the total number of alleles will be reduced at a greater rate than the range in allele size, thus reducing the  $M$  ratio (total number of alleles/overall range in allele size). It is expected that the  $M$  ratio will be smaller for populations that have experienced more severe bottlenecks. For applying this test, we used the following parameters in the software  $M$  as recommended by the authors as conservative on the basis of population simulations (Garza and Williamson 2001): We assumed that 90 % of mutations are one-step mutations ( $p_s = 90$  %), an average size of 3.5 for non-one-step mutations ( $\Delta_g = 3.5$ ), and three different population mutation parameters of  $\theta$  [ $4 Ne \mu$ , with a mutation rate  $\mu$  of  $5 \times 10^{-4}$ /locus/generation (Hawley et al. 2006)] = 4, 10 and 25.

## Results

### SSR polymorphisms

A total of 99 alleles were observed in the whole sample and the ten loci analyzed, with an average number of alleles per locus of 9.9 and a range of 7–12 alleles per locus (Table 1). The locus with the highest information index ( $I$ ), highest effective number of alleles ( $Ne$ ) and highest expected heterozygosity ( $H_E$ ) was BM140 and that with the lowest  $I$ ,  $Ne$  and  $H_E$  was BM197. For the individual loci in the whole sample (Table 1), values of observed heterozygosity ( $H_O$ ) were relatively low (an average of 0.040, range from 0.007 to 0.079), values of expected heterozygosity ( $H_E$ ) were relatively high (an average of 0.667, range from 0.346 to 0.849), and fixation indexes were high (average of 0.943), as expected for autogamous species.

### Intra-accession polymorphism

In order to assess intra-accession polymorphism, five individuals of each of 150 accessions were analyzed and

genetic diversity indexes were calculated (Supplementary Table S1). The results show that most of the wild accessions were polymorphic for at least one of the loci analyzed (65 accessions out of 82), five of the six weedy accessions were also polymorphic, while a major proportion of the domesticated accessions were monomorphic for all the loci analyzed (40 out of 62 accessions).

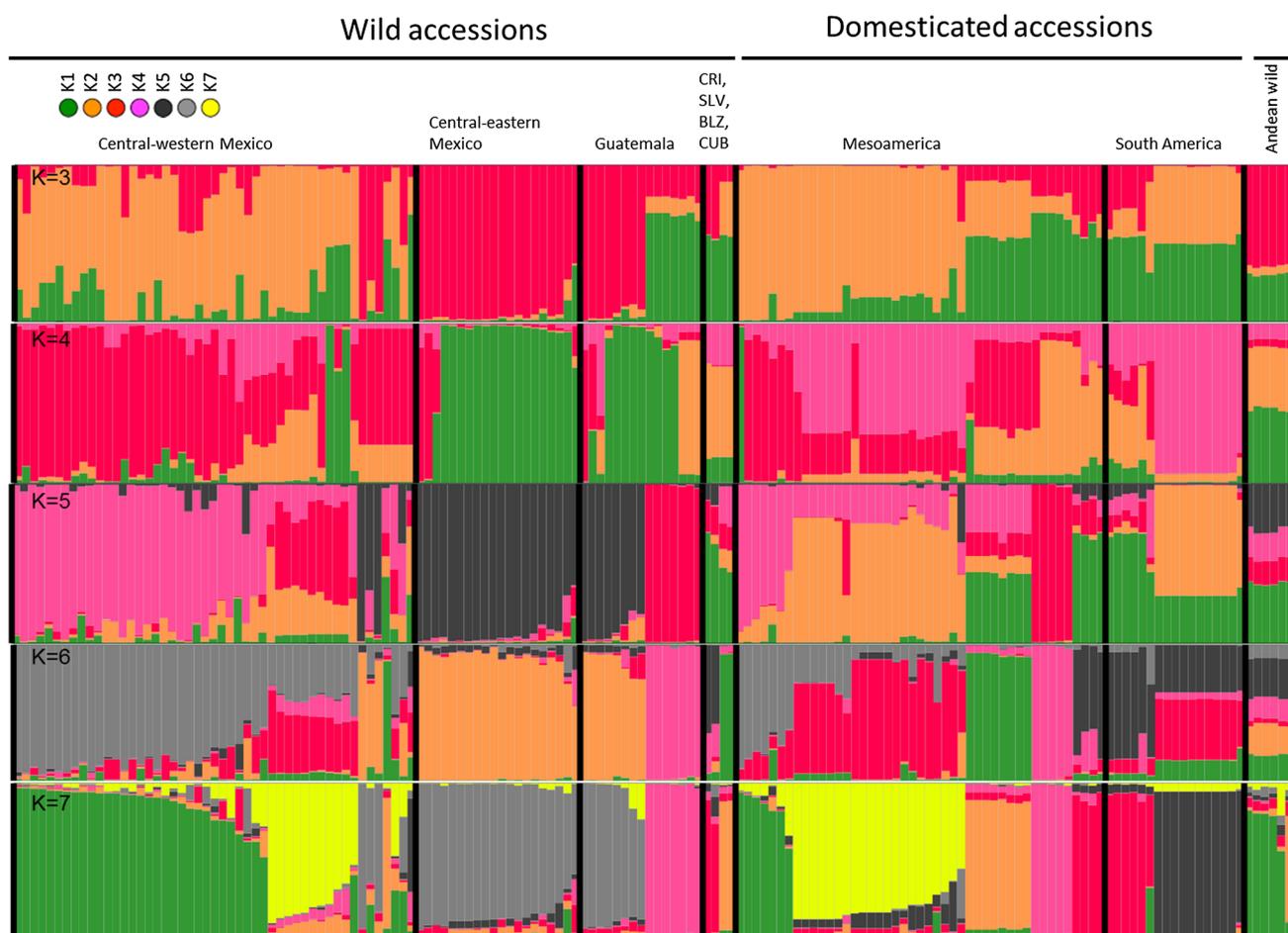
If we compare the accessions from field collections versus the accessions obtained from CIAT's genebank, we can see that in general a major proportion of the accessions collected in the field were polymorphic (75 %) compared to the proportion of accessions from CIAT's genebank (40 %). Also, wild accessions from field collections showed higher expected heterozygosity ( $H_E = 0.111$ ) than wild accessions from CIAT's genebank ( $H_E = 0.070$ ). This result is somehow expected given the bottleneck effect that operates in genebanks when accessions are multiplied, although this could also be related to the fact that a major part of wild accessions come from field collections (52 from field collections and 30 from CIAT's genebank). The opposite pattern was observed for domesticated accessions ( $H_E$  from field = 0.006,  $H_E$  from CIAT's genebank = 0.020), although this result could also be caused by the fact that most of the domesticated accessions come from CIAT's genebank (49 out of 62).

In relation to the biological status, in average wild accessions were more diverse ( $H_O = 0.066$ ,  $H_E = 0.096$ ,  $H_E$  ranging from 0 to 0.334) than weedy accessions ( $H_O = 0.023$ ,  $H_E = 0.066$ ,  $H_E$  ranging from 0 to 0.14) and domesticated accessions were less diverse ( $H_O = 0.008$ ,  $H_E = 0.017$ ,  $H_E$  ranging from 0 to 0.114). This result is expected for domesticated species due the founder effect, as it will be discussed below.

### Bayesian and clustering analyses

As shown above, not all individuals within accessions are genetically identical; therefore, in the Bayesian clustering analyses membership coefficients were first calculated for each of the five individuals within accessions, and then these coefficients were averaged to obtain a single coefficient per accession. Accessions were classified as belonging to one of the  $K$  populations if their membership coefficients were larger than 80 %; otherwise, they were classified as admixed (Supplementary Table S1, column "K").

The Bayesian approach indicated that the optimum  $K$  was 7. The bar graphs in Fig. 1 show the coefficients of membership for each of the 155 accessions analyzed, from  $K = 3$  to  $K = 7$ , and each accession was color-coded according to its percentage of membership. The results obtained with the software Structure and Instruct were very similar, with the Instruct software estimating selfing rates



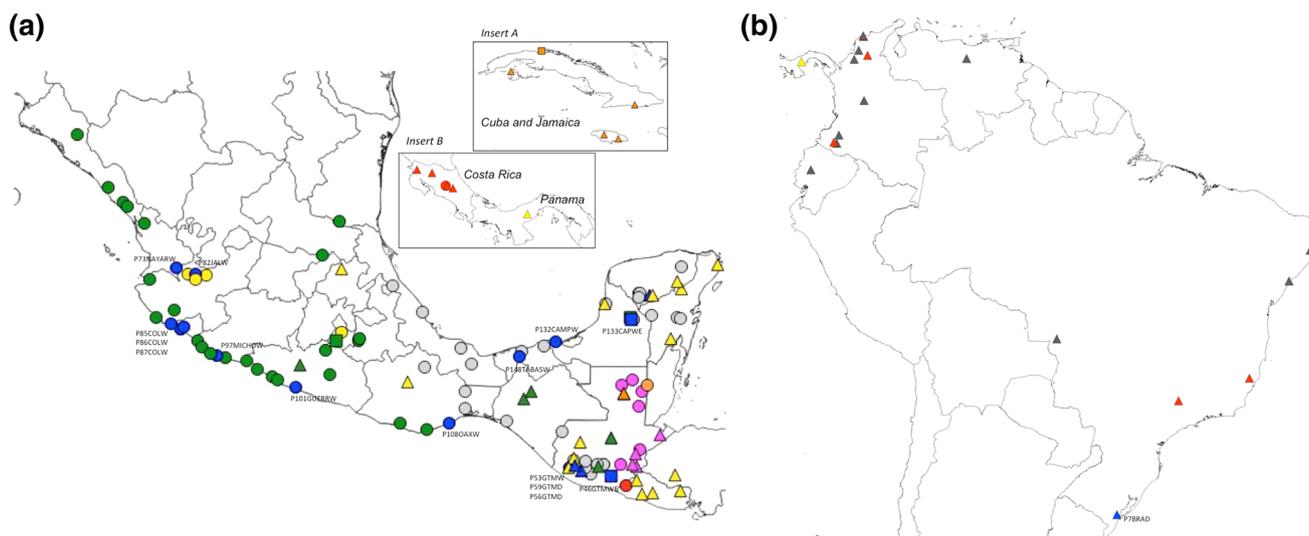
**Fig. 1** Bar plots of the results from Structure. In the figure, the global ancestry of each accession from each one of  $K$  populations is shown, from  $K = 3$  to  $K = 7$ . Optimum  $K$  was 7. The accessions were organized according to geographic origin from north to south as

shown at the *top* of the figures. *CRI* Costa Rica, *BLZ* Belize, *SLV* El Salvador, *CUB* Cuba. Each one of the  $K$  populations is shown by a different color:  $K1$  is green,  $K2$  is orange,  $K3$  is red,  $K4$  is pink,  $K5$  is black,  $K6$  is gray, and  $K7$  is yellow. (Color figure online)

of about 0.950. In Fig. 1, wild accessions are organized by geographic region, from north to south into four regions: (1) central-western Mexico (the Pacific range of the distribution in Mexico from the state of Sinaloa to Oaxaca, and the states of Morelos and Guanajuato), (2) central-eastern Mexico (the plains of the Gulf of Mexico in the states of Tamaulipas, Veracruz, Tabasco and the Peninsula of Yucatan), (3) Guatemala, and (4) other Central American countries and Caribbean islands. Domesticated accessions are divided into two regions: those from Mesoamerica and those from South America. Figure 2a, b shows the geographic distribution of the accessions, color coded according to the  $K$  population from which they derive their ancestry (for  $K = 7$ ). Admixed accessions (18 in total: 12 wild, 2 weedy and 4 domesticated) are shown in blue (see also Supplementary Table S1). The inserts A and B in Fig. 2a show the distribution of accessions in Costa Rica, Panama, Cuba and Jamaica. The NJ topology showing the relationship among the  $K$  populations (for

$K = 7$  and separated as wild, weedy and domesticated as applicable) is shown in Fig. 3 with the same color coding as in Figs. 1 and 2a, b.

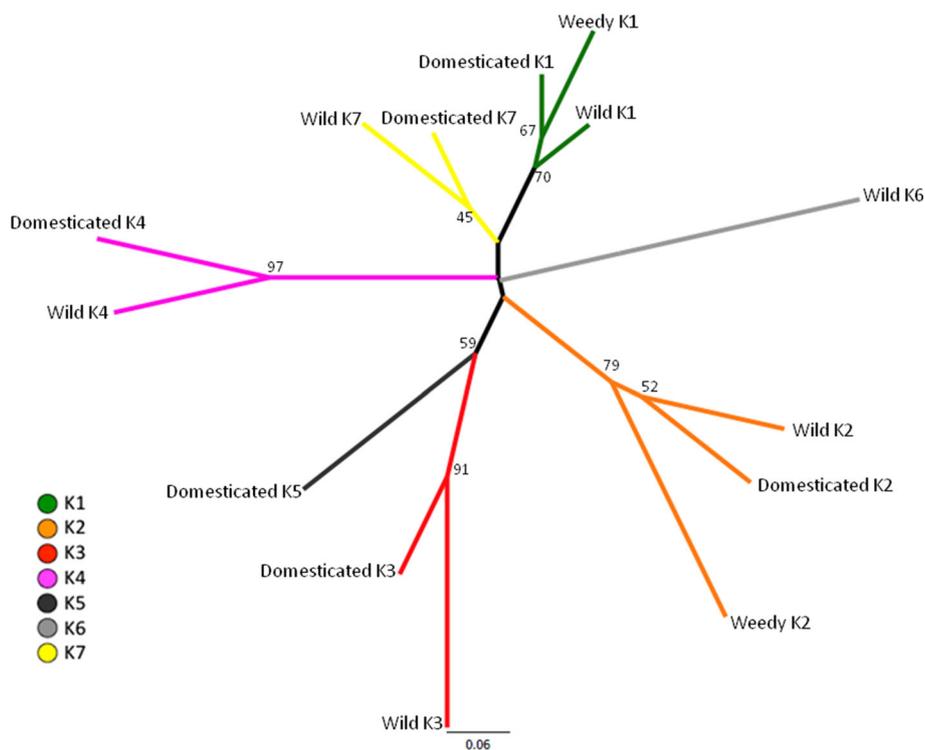
The results of the Bayesian clustering analyses (Fig. 1,  $K = 7$ ) indicate that most of the wild accessions from central-western Mexico derive their ancestry from  $K1$  or  $K7$ . In Fig. 2a, the  $K1$  wild accessions (green circles, 24 accessions) are distributed along central-western Mexico and Tamaulipas, while  $K7$  wild accessions (yellow circles, 8 accessions) come from the states of Jalisco, Colima and Morelos. It can also be seen in Fig. 1 that wild accessions from central-eastern Mexico derive their ancestry mainly from  $K6$  (light gray circles, 20 accessions in Fig. 2a). On the other hand, wild accessions from Guatemala derive their ancestry from  $K4$  (7 accessions) or  $K6$  (7 accessions), and it is interesting to see how accessions from the southwest of Guatemala (mostly  $K6$ ) differ from those of the southeast and north (mostly  $K4$ ). Wild accessions from Costa Rica and El Salvador belong to  $K3$  and the wild



**Fig. 2** Map showing the distribution of the wild (*circles*), weedy (*squares*) and domesticated (*triangles*) accessions. **a** Mesoamerica and **b** South America. The colors of the symbols match the color of the *K* populations recovered in the Structure analyses (see Fig. 1), and

the *blue* color refers to admixed accessions. Population IDs of admixed accessions are shown in the maps and complete information about their origin, biological status, gene pool and diversity indexes can be found in Supplementary Table S1. (Color figure online)

**Fig. 3** Neighbor-joining topology based on a Nei's genetic distance matrix showing the genetic relationships among the *K* populations defined in the Structure analysis (for *K* = 7). The colors of the *branches* match the color of the *K* populations (see Fig. 1). (Color figure online)



accession from Belize belongs to *K2*. Wild admixed accessions come from different places in Mexico and Guatemala (see blue circles in the map of Fig. 2a and also see Supplementary Table S1).

The results of the Bayesian clustering analyses (Fig. 1, *K* = 7) indicate that the landraces are not a homogeneous group but are rather a diverse group (in Fig. 2a, b, they are

shown as triangles). They derive their ancestry mostly from *K7* (19 accessions), *K5* (11 accessions), *K3* (9 accessions), *K2* (8 accessions), *K1* (6 accessions) and *K4* (5 accessions). Four domesticated accessions from Brazil, Guatemala and Mexico (Yucatan) were classified as admixed (see blue triangles in Fig. 2a, b). In the wild gene pool, the accessions that derive their ancestry from these same

*K* populations (*K*1, *K*2, *K*3, *K*4, *K*7, except *K*5) are widespread in an area comprising central-western Mexico and the area Guatemala–Costa Rica.

In the present study, we included six weedy accessions (shown as square symbols in Fig. 2a), three from Mexico (Campeche and Morelos), one from Guatemala and two from Cuba. Weedy accessions derive their ancestry from *K*1 (3 accessions), *K*2 (1 accession) and two were admixed. We found as unexpected that only two of the six weedy accessions were classified as admixed, given the fact that weedy accessions are the result of gene flow among wild and domesticated populations.

The neighbor-joining topology showing the relationships among the seven *K* populations and obtained from the matrix of Nei's genetic distance is shown in Fig. 3. For the chord distance, the results were similar. We can see that wild and domesticated accessions within the *K*1, *K*2, *K*3, *K*4 and *K*7 populations cluster together with relatively high bootstrap support most of the time (the weakest bootstrap values were for *K*2 and *K*7 clusters). We can also see that weedy accessions clustered together with their wild and domesticated counterparts within *K*1 and *K*2 with good bootstrap support, as expected. It is interesting to see how the *K*5 cluster, composed of domesticated accessions from South America, is more related to cluster *K*3, which also includes domesticated accessions from South America and wild accessions from Central America (El Salvador and Costa Rica).

Figure 4 shows in detail the membership coefficients of all the 18 admixed accessions found (in the maps in Fig. 2a, b, they are colored as blue and their population IDs are also labeled, except for accessions P124 and P141 which lack geographic coordinates). If we compare these coefficients and the geographic location of accessions, we can see that admixed accessions from central-western Mexico (from Nayarit to Guerrero) derive most of their ancestry from *K*1 and *K*7, and this is a region where wild and domesticated accessions from *K*1 and *K*7 are also found. On the other hand, admixed accessions located in central-eastern Mexico (from Tabasco to Chiapas, including the Peninsula of Yucatan), Oaxaca and southern Guatemala derive most of their ancestry from *K*1, *K*6 and *K*7, and in this region wild and domesticated accessions from these same *K* populations are also found. The admixed accession from Brazil derives its ancestry from *K*1 and *K*3, and in Brazil we also found other domesticated accessions from *K*3. These results suggest that these admixed accessions are the results of gene flow among accessions located in nearby regions.

In summary, our results depict a complex picture of ancestry for the wild and domesticated accessions, suggesting that the current makeup of these accessions may in part be a reflection of gene flow. The fact that domesticated

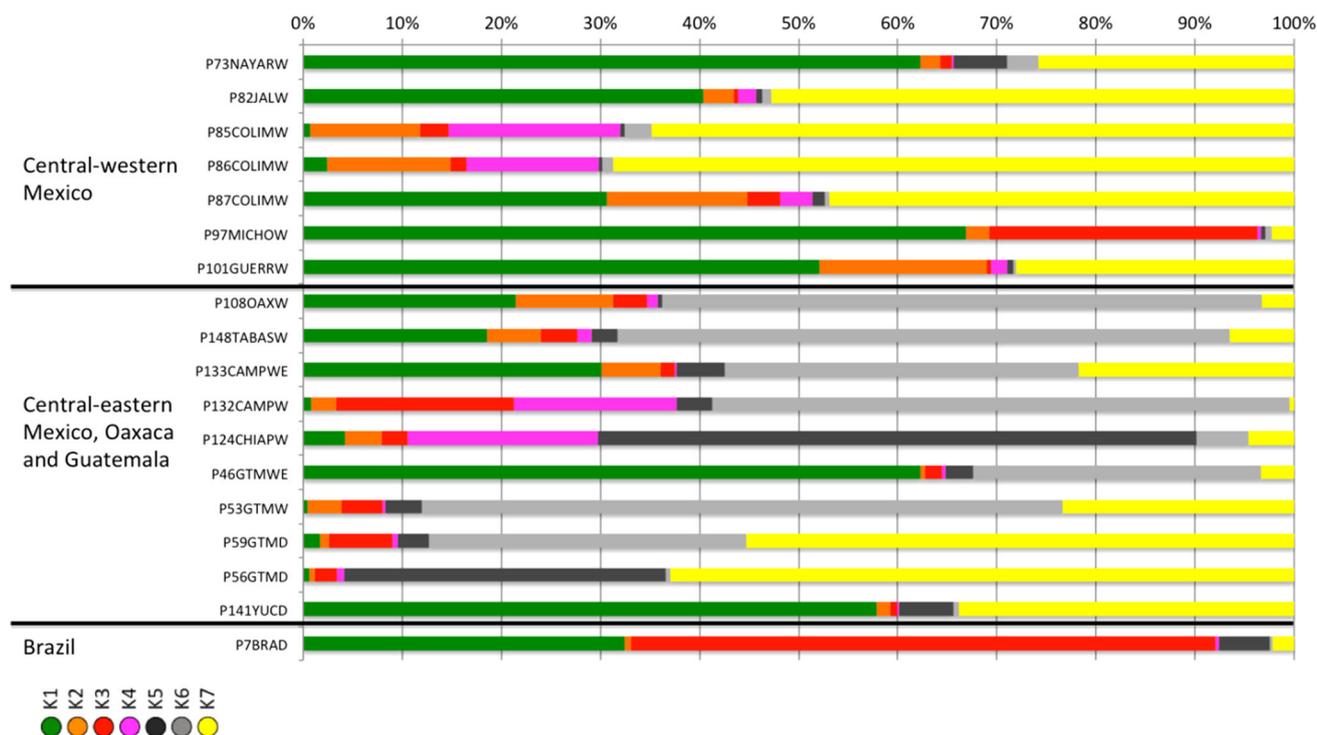
accessions are not a genetically homogeneous group would suggest a scenario of multiple domestications accompanied by gene flow. Our results however would not allow us to conclude about specific areas of domestication because putative wild ancestors are distributed in a wide area in central-western Mexico and Guatemala–Costa Rica.

### Founder effects

Founder effects due to domestication were measured as a reduction in allele richness ( $N_a$ ) and expected heterozygosity ( $H_E$ ) in inter-population comparisons, namely between landraces and their wild ancestors. In this study, we made three types of comparisons. First, we compared all wild and domesticated accessions, second we compared putative wild ancestors from gene pool MI (central-western Mexico) and related landrace accessions (mainly accessions within *K*1 and *K*7 clusters), and third we compared putative wild ancestors from gene pool MII (area from Guatemala–Costa Rica) and related landrace accessions (mainly accessions within *K*2, *K*3, *K*4 and *K*5 clusters).

In this first comparison, the landraces showed an average loss in allele richness of 25 % and in  $H_E$  of 17 % (Table 2). This reduction was significant according to the comparison tests carried out in FSTAT ( $p = 0.001$ ). For the domestication event within the gene pool MI, the founder event was much larger, with a statistically significant loss in allele richness of 45 % and in  $H_E$  of around 44 %. For the domestication event in MII, no reduction in allele richness (0 %) was observed and a significant although low reduction in  $H_E$  (1 %) was observed (Table 2). Another measure that gives us information about founder effects is %A, defined as the percentage of the total allelic diversity captured in the sample. In Table 2, it can be seen that, in general, a larger proportion of allelic diversity is found in the wild accessions (around 20 % more) as opposed to the domesticated accessions.

Founder effects were also measured in intra-population analyses of  $M$  ratios. When all wild and landrace populations were analyzed,  $M$  ratios were lower in domesticated (0.58) than in wild populations (0.62), and both of these populations showed significant reductions in  $M$  ratios with respect to hypothetical equilibrium populations; therefore, evidence of bottleneck was found and this was more drastic in landraces than in wild populations. When  $M$  ratios were calculated in wild and landrace populations within the MI and MII gene pools, wild populations showed lower  $M$  ratios than domesticated populations and significant reductions in  $M$  ratios were observed in MI wild populations, and in MII wild and domesticated populations, providing once more evidence of bottleneck, especially within the MII gene pool.



**Fig. 4** Coefficients of global ancestry for the 18 accessions that were classified as admixed in the Structure analyses. *Bar colors* are the same as in Fig. 1. Population IDs of admixed accessions are shown,

and complete information about their origin, biological status, gene pool and diversity indexes can be found in Supplementary Table S1. (Color figure online)

In summary, our data support founder effects due to domestication in Lima bean as measured as allele richness and  $H_E$ . The  $M$  ratios showed evidence of bottleneck for both wild and domesticated populations, but depending on the gene pool analyzed this bottleneck effect is not always more severe for domesticated populations.

## Discussion

### Domestication event in Mesoamerica

With the different methodologies used, most of the landraces from Mesoamerica clustered either with wild accessions located in central-western Mexico (25 out of 62 accessions), within the geographic range of wild gene pool MI (from Sinaloa to Oaxaca), or with wild accessions located in an area between Guatemala and Costa Rica (22 accessions), within the geographic range of wild gene pool MII, and not with wild accessions from central-eastern Mexico. These results then suggest at least one domestication event within gene pools MI and MII.

One of these domestication events may be located in central-western Mexico where wild populations (from K1 and K7) showed genetic affinity with around 40 % of the landrace accessions analyzed. Previous studies have also

pinpointed an area of domestication for Mesoamerican landraces of Lima bean in central-western Mexico within MI gene pool. Motta-Aldana et al. (2010) studied wild and domesticated accessions of Lima bean on the basis of chloroplast DNA (cpDNA) and the internal transcribed spacer of the ribosomal DNA (ITS) polymorphisms. The authors found that most of landraces carried a single cpDNA haplotype (haplotype G), which in the wild was widely distributed and was more abundant in wild populations from Mexico (Jalisco, Morelos, Puebla, Oaxaca, Campeche and Chiapas). On the other hand, the authors found that most of the landraces carried ITS haplotype L, which was only carried by wild populations from the Mexican states of Jalisco, Puebla and Oaxaca. Therefore, cpDNA and ITS data supported central-western Mexico as a possible domestication area but did not define more precisely where. Serrano-Serrano et al. (2012) also analyzed ITS polymorphisms in a larger sample of wild and Mesoamerican landraces and also found that most of landraces carried haplotype L, which in the wild was found most frequently in the areas of Jalisco-Nayarit and Guerrero-Oaxaca, showing again central-western Mexico as a possible place of origin for these landraces. Very recently, Andueza-Noh et al. (2013) studied cpDNA polymorphisms in a large sample of wild and domesticated accessions of Lima beans from Mesoamerica and also found that cpDNA

**Table 2** Diversity indexes and the founder effect in Lima bean

| Gene pools                  | <i>N</i> | <i>P</i> | <i>A</i> | <i>Na</i>    | <i>Ne</i> | <i>PA</i> | <i>H<sub>E</sub></i> | <i>r</i> ( <i>Na</i> ; <i>He</i> ) | <i>p</i> value <sup>2</sup><br>( <i>M</i> ratio<br>$\theta = 4$ ) | <i>p</i> value <sup>2</sup><br>( <i>M</i> ratio<br>$\theta = 10$ ) | <i>p</i> value <sup>2</sup><br>( <i>M</i> ratio<br>$\theta = 25$ ) | <i>M</i> ratio, $\theta$ total |
|-----------------------------|----------|----------|----------|--------------|-----------|-----------|----------------------|------------------------------------|---|--|--|--------------------------------|
| All                         |          |          |          |              |           |           |                      |                                    |   |  |  |                                |
| Mesoamerican wild           | 435      | 100      | 86       | 8.5          | 3.99      | 0.27      | 0.68                 |                                    | <b>0.01</b>   | <b>0.00</b>  | <b>0.00</b>  | 0.62                           |
| All landraces               | 315      | 100      | 65       | 6.4          | 2.66      | 0.07      | 0.56                 | (25; 17)                           | <b>0.00</b>   | <b>0.00</b>  | <b>0.00</b>  | 0.58                           |
| <i>p</i> value <sup>1</sup> | –        | –        | –        | <b>0.001</b> | –         | –         | <b>0.001</b>         |                                    |   |  |  |                                |
| Gene pool MI                |          |          |          |              |           |           |                      |                                    |   |  |  |                                |
| MI wild (ancestor)          | 176      | 100      | 56       | 5.5          | 2.92      | 1.2       | 0.53                 |                                    | <b>1.25</b>   | <b>1.15</b>  | <b>0.74</b>  | 0.71                           |
| MI landraces                | 120      | 80       | 30       | 3            | 1.6       | 0.1       | 0.30                 | (45; 44)                           | 8.54  | 11.07  | 11.47  | 0.76                           |
| <i>p</i> value <sup>1</sup> | –        | –        | –        | <b>0.001</b> | –         | –         | <b>0.001</b>         |                                    |   |  |  |                                |
| Gene pool MII               |          |          |          |              |           |           |                      |                                    |   |  |  |                                |
| MII wild (ancestor)         | 58       | 100      | 50       | 4.9          | 3.33      | 0.8       | 0.65                 |                                    | <b>0.00</b>   | <b>0.00</b>  | <b>0.00</b>  | 0.46                           |
| MII landraces               | 111      | 100      | 50       | 4.9          | 2.94      | 0.3       | 0.64                 | (0; 1)                             | <b>0.00</b>   | <b>0.00</b>  | <b>0.00</b>  | 0.51                           |
| <i>p</i> value <sup>1</sup> | –        | –        | –        | NS           | –         | –         | <b>0.001</b>         |                                    |   |  |  |                                |
| Andean wild                 | 7        | 70       | 18       | 1.8          | 1.56      | 0.6       | 0.29                 | –                                  | –   | –  | –  |                                |

*N* sample size in number of individuals; *P* percentage of polymorphic loci; *A*, % of the total allelic diversity captured in the sample; *Na* average number of alleles per locus or allele richness; *Ne* effective number of alleles; *PA* average number of private alleles per locus; *H<sub>E</sub>* expected heterozygosity; and *r*, % of loss of allele richness and genetic diversity due to founder effects

<sup>1</sup> *p* value of the inter-population comparison tests, significant *p* values are shown in bold, <sup>2</sup> *p* value of the intra-population tests (*M* ratios carried out with  $\mu = 90$  and  $\Delta g = 3.5$ , and  $\theta (4Ne\mu) = 5, 10$  and  $25$ , with  $\mu = 5 \times 10^{-4}$ ), *p* values supporting significant reductions in *M* ratios with respect to hypothetical equilibrium populations and indicative of bottleneck, are shown in bold

haplotype G was the most abundant among the landraces, and in the wild this haplotype was more frequent in the states of Sinaloa, Nayarit, Jalisco, Guerrero and Oaxaca, once more confirming central-western Mexico as a domestication area for Lima beans. In the present study, we were able to confirm a possible area of domestication within central-western Mexico as many of the Mesoamerican landraces shared ancestry with wild populations from this area in the Bayesian clustering analysis and also because in the NJ topology, wild and domesticated accessions in K1 clustered together with relatively high support. Central-western Mexico has also been suggested as a place of origin for other legume crops such as common bean (*P. vulgaris* L.) (Kwak et al. 2009) and tepary bean (*P. acutifolius* A. Gray) (Muñoz et al. 2006). These examples illustrate the importance of this area for legume domestication in Mesoamerica [see also Zizumbo-Villarreal and Colunga-GarcíaMarín (2010)].

Another possible area where domestication may have taken place is within the gene pool MII, in an area between Guatemala and Costa Rica, where we found wild populations genetically related to about 35 % of the landrace accessions analyzed. In the NJ topology, wild and domesticated accessions from this area grouped together in several clusters with high bootstrap support. Our results agree with previous studies that have also suggested Guatemala as an area of diversification and domestication of *P.*

*lunatus* (Sauer 1993; Gutiérrez-Salgado et al. 1995; Fofana et al. 2001). This area has also been pinpointed for domestication of two other *Phaseolus* species, namely *P. polyanthus* Greenm. in Guatemala (Schmit and Deboucq 1991) and *P. coccineus* L. in the area of Guatemala–Honduras (Spataro et al. 2011). In a recent study, Andueza-Noh et al. (2013) suggested a second domestication event for Lima beans within an area between Guatemala, Honduras and Costa Rica. The authors gave as evidence the presence of cpDNA haplotype T from the MII gene pool in seven accessions of domesticated Lima beans that were collected in these countries and the fact that this haplotype is particularly abundant in wild populations from these countries. However, this evidence may also support competing hypotheses, for example, local introgression of haplotype T from wild populations to landraces. Therefore, more evidence was needed to confirm a second domestication event in Mesoamerica.

If we compare previous ITS and chloroplast data for the landraces with results from this study, we see that the picture get very complex as the individual landrace accessions can be classified into one or another gene pool in different studies (see gene pool columns in Supplementary Table S1). In general, we can see that landrace accessions classified in the gene pool MI in this study, namely those that are within K1 or K7 populations, were also classified into gene pool MI in previous studies.

However, we can see that landrace accessions classified into gene pool MII in this study (those within populations K2, K3 and K4) were in many instances classified within gene pool MI in previous studies. One of the reasons for this could be gene flow and another reason could be that these SSR loci are not enough to sample adequately global ancestry of these accessions.

### Founder effects

During domestication, founder effects are mainly due to two different processes: The first one is the domestication process itself in which a small portion of the wild genetic reservoir is taken into cultivation, and this effect is also known in the literature as the domestication bottleneck, and second, during the spread of landraces outside the domestication area which may also convey a reduction in genetic diversity in the areas of introduction. For estimating the founder effects of domestication, we compared genetic diversity measures in three types of comparisons: one including all wild and domesticated accessions, a second one including wild and landraces within gene pool MI and a third one including wild and landraces within gene pool MII (Table 2).

In general, landrace accessions showed a reduction in genetic diversity as measured by expected heterozygosity ( $H_E$ ) (reduction of 17 %) and allele richness ( $N_a$ ) (25 %). When we look within each gene pool, a more drastic reduction in genetic diversity (44 % reduction in  $H_E$  and 45 % reduction in  $N_a$ ) is observed for gene pool MI than gene pool MII (only 1 % reduction in  $H_E$  and no reduction in  $N_a$ ), and in addition MII landraces contain more genetic diversity than MI landraces. These contrasting patterns might be explained by the current distribution of ancestral populations and landraces. The putative ancestral populations of MI landraces are found in a wide area in central-western Mexico, and the domestication bottleneck itself and subsequent introduction of these landraces to other areas of Mexico and Central America might have caused successive founder effects. MI landraces were found in our study in central and eastern Mexico, the Peninsula of Yucatan and El Salvador. On the other hand, MII wild ancestors are more restricted in distribution (Guatemala, El Salvador, Belize and Costa Rica mainly), and although harbor slightly more genetic diversity than MI ancestors ( $H_E = 0.65$  versus  $H_E = 0.53$ , Table 2), the MII landraces also have a restricted distribution in Guatemala and Costa Rica, a feature that may have limited the number of founder events due to dispersion. Another feature of MII landraces is that some of them are sympatric to wild populations in Guatemala and Costa Rica, raising the possibility of genetic interchange and therefore increase in allelic and genetic diversity. In fact, no reduction in

number of alleles was found among MII wild ancestors and MII landraces (Table 2).

In summary, these results indicate that landraces from the two gene pools may have been affected in a different way by the domestication process, but in general it can be seen that there is a reduction in genetic diversity due to domestication.

### Conclusions

With the results of the present study, we can conclude that the Mesoamerican landraces of Lima bean may have been domesticated at least once and that their domestication history may be more complex than previously thought. With the genetic markers used, we could not establish more specifically the place of origin of the Mesoamerican landraces within central-western Mexico; however, we could establish that the Mesoamerican landraces of Lima bean are not a genetically homogeneous group as it would be expected in a scenario of a single domestication without much introgression. Therefore, we believe that the possibility of more than one domestication event is still open and also the possibility of multiple introgression events as shown by the several cases of admixed accessions. In order to establish how domestication processes and gene flow have shaped the current genetic structure of landraces, a more complete sampling in many areas of Central America is needed, especially in Honduras, Nicaragua and Panama, and also in northern South America. Besides that, a more complete genome sampling is also necessary. Genomic analysis offers promise in identifying regions of the genome related to the domestication syndrome, and through the correct comparisons, this approach may help discriminating among different hypothesis of gene flow versus domestication.

**Acknowledgments** The first author thanks Ciencia Básica-CONACYT (Project No. 82642) and the National Geographic Society (Grant No. 8622-09) for the economic support to carry out this research. The authors thank the International Center for Tropical Agriculture, Drs. Rogelio Lépiz and Jorge Acosta for providing seed samples. The first author thanks Julian Coello Coello and Filogonio May Pat for technical support in the field and laboratory work.

**Conflict of interest** The authors declare that they have no conflict of interest.

### References

- Aguilar-Meléndez A, Morrell PL, Roose ML, Kim SC (2009) Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. *Am J Bot* 96:1190–1202

- Andueza-Noh RH, Serrano-Serrano ML, Chacón Sánchez MI, Sánchez del Pino I, Camacho-Pérez L, Coello-Coello J, Mijangos Cortes JO, Debouck DG, Martínez-Castillo J (2013) Multiple domestications of the Mesoamerican gene pool of lima bean (*Phaseolus lunatus* L.): evidence from chloroplast DNA sequences. *Genet Resour Crop Evol* 60:1069–1086
- Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* 19:233
- Debouck DG (2008) Notes sur les différents taxons de *Phaseolus* à partir des Herbiers-Section Paniculati. 2008. Disponible en: URL: <http://www.ciat.cgiar.org/urg:233>
- Debouck DG, Maquet A, Posso CE (1989) Biochemical evidence for two different gene pools in Lima beans, *Phaseolus lunatus* L. *Ann Rep Bean Improv Coop* 32:58–59
- Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* 3:167–169
- Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem bull* 19:11–15
- Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Felsenstein J (1989) PHYLIP - phylogeny inference package (version 3.2). *Cladistics* 5:164–166
- Fofana B, Du Jardin P, Baudoin JP (2001) Genetic diversity in the Lima bean (*Phaseolus lunatus* L.) as revealed by chloroplast DNA (cpDNA) variations. *Genet Resour Crop Evol* 48:437–445
- Freytag GF, Debouck DG (2002) Taxonomy, distribution, and ecology of the genus *Phaseolus* (Leguminosae–Papilionoideae) in North America, Mexico, and Central America. *Bot. Res. Inst., Fort Worth, TX. Taxonomy, distribution, and ecology of the genus Phaseolus*
- Gao H, Williamson S, Bustamante CD (2007) A Markov chain monte carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* 176:1635–1651
- Gao H, Bryc K, Bustamante CD (2011) On identifying the optimal number of population clusters via the deviance information criterion. *PLoS ONE* 6:e21014
- Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* 10:305–318
- Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *J Hered* 86:485–486
- Gutiérrez-Salgado A, Gepts P, Debouck DG (1995) Evidence for two gene pools of the Lima bean, *Phaseolus lunatus* L., in the Americas. *Genet Resour Crop Evol* 42:15–28
- Hardy O, Dubois S, Zoro-Bi AI, Baudoin JP, Wright S (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
- Hawley DM, Hanley D, Dhondt AA, Lovette IJ (2006) Molecular evidence for a founder effect in invasive house finch (*Carpodacus mexicanus*) populations experiencing an emergent disease epidemic. *Mol Ecol* 15:263–275
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806
- Kaplan L (1965) Archeology and domestication in American *Phaseolus* (beans). *Econ Bot* 19:358–368
- Kaplan L, Lynch TF (1999) *Phaseolus* (Fabaceae) in archaeology: AMS. *Econ Bot* 53:261–272
- Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor Appl Genet* 118:979–992
- Kwak M, Kami JA, Gepts P (2009) The putative Mesoamerican domestication center of *Phaseolus vulgaris* L. is located in the Lerma—Santiago Basin of Mexico. *Crop Sci* 49:554–563
- Ladizinsky G (1985) Founder effect in crop-plant evolution. *Econ Bot* 39:191–199
- Martínez-Castillo J, Zizumbo-Villarreal D, Perales-Rivera H, Colunga-GarcíaMarín P (2004) Intraspecific diversity and morpho-phenological variation in *Phaseolus lunatus* L. from the Yucatan peninsula. *Mex Econ Bot* 58:354–380
- Martínez-Castillo J, Zizumbo-Villarreal D, Gepts P, Delgado-Valerio P, Colunga-GarcíaMarín P (2006) Structure and genetic diversity of wild populations of Lima Bean (*Phaseolus lunatus* L.) from the Yucatan peninsula Mexico. *Crop Sci* 46:1071–1080
- Martínez-Castillo J, Zizumbo-Villarreal D, Gepts P, Colunga-GarcíaMarín P (2007) Gene flow and genetic structure in the wild-weedy-domesticated complex of Lima bean (*Phaseolus lunatus* L.) in its Mesoamerican center of domestication and diversity. *Crop Sci* 47:58–66
- Martínez-Castillo J, Camacho-Pérez L, Villanueva-Viramontes S, Andueza-Noh RH, Chacón Sánchez MI (2014) Genetic structure within the Mesoamerican gene pool of wild *Phaseolus lunatus* (Fabaceae) from Mexico as revealed by microsatellite markers: implications for conservation and the domestication of the species. *Am J Bot* 101:851–864
- Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci* 99:6080–6084
- Motta-Aldana JR, Serrano-Serrano ML, Hernández-Torres J, Castillo-Villamizar G, Debouck DG (2010) Multiple origins of Lima bean landraces in the Americas: evidence from chloroplast and nuclear DNA polymorphisms. *Crop Sci* 50:1773–1787
- Muñoz LC, Duque MC, Debouck DG, Blair MW (2006) Taxonomy of tepary bean and wild relatives as determined by amplified fragment length polymorphism (AFLP) markers. *Crop Sci* 46:1744–1754
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Pickersgill B (1969) The archaeological record of chili peppers (*Capsicum* spp.) and the sequence of plant domestication in Peru. *Am Antiq* 34(1):54–61
- Piperno DR, Flannery KV (2001) The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proc Natl Acad Sci* 98:2101–2103
- Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium BP maize from the central balsas river valley, Mexico. *Proc Natl Acad Sci* 106:5019–5024
- Pohl MD, Pope KO, Jones JG, Jacob JS, Piperno DR, deFrance SD, Lentz DL, Gifford JA, Danforth ME, Josserand JK (1996) Early agriculture in the maya lowlands. *Lat Am Antiq* 7(4):355–372
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rambaut A, Drummond AJ (2014) FigTree. Version 1.4. Available at <http://tree.bio.ed.ac.uk/software/figtree/>
- Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425

- Sauer JD (1993) Historical geography of crop plants. CRC, Boca Raton
- Schmit V, Debouck DG (1991) Observations on the origin of *Phaseolus polyanthus* Greenman. *Econ Bot* 45:345–364
- Serrano-Serrano ML, Hernández-Torres J, Castillo-Villamizar G, Debouck DG, Chacón Sánchez MI (2010) Gene pools in wild Lima bean (*Phaseolus lunatus* L.) from the Americas: evidences for an andean origin and past migrations. *Mol Phylogenet Evol* 54:76–87
- Serrano-Serrano ML, Andueza-Noh RH, Martínez-Castillo J, Debouck DG, Chacón Sánchez MI (2012) Evolution and domestication of Lima bean in Mexico: evidence from ribosomal DNA. *Crop Sci* 52:1698–1712
- Smith BD (1997) The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science* 276:932–934
- Smith BD (2005) Reassessing coxcatlan cave and the early history of domesticated plants in Mesoamerica. *Proc Natl Acad Sci USA* 102:9438–9445
- Spataro G, Tiranti B, Arcaleni P, Bellucci E, Attene G, Papa R, Zeuli PS, Negri V (2011) Genetic diversity and structure of a worldwide collection of *Phaseolus coccineus* L. *Theor Appl Genet* 122:1281–1291
- Zizumbo-Villarreal D, Colunga-GarcíaMarín P (2010) Origin of agriculture and plant domestication in West Mesoamerica. *Genet Resour Crop Evol* 57:813–825