

# Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes

Elena Bitocchi<sup>1</sup>, Elisa Bellucci<sup>1</sup>, Alessandro Giardini<sup>1</sup>, Domenico Rau<sup>2</sup>, Monica Rodriguez<sup>2</sup>, Eleonora Biagetti<sup>1</sup>, Rodolfo Santilocchi<sup>1</sup>, Pierluigi Spanoletti Zeuli<sup>3</sup>, Tania Gioia<sup>3</sup>, Giuseppina Logozzo<sup>3</sup>, Giovanna Attene<sup>2</sup>, Laura Nanni<sup>1</sup> and Roberto Papa<sup>1,4</sup>

<sup>1</sup>Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università Politecnica delle Marche, Via Brecce Bianche, 60131, Ancona, Italy; <sup>2</sup>Dipartimento di Agraria, Università degli Studi di Sassari, Via de Nicola, 07100, Sassari, Italy; <sup>3</sup>Scuola di Scienze Agrarie, Forestali, Alimentari ed Ambientali, Università degli Studi della Basilicata, via dell'Ateneo Lucano, 10, 85100, Potenza, Italy; <sup>4</sup>Cereal Research Centre, Agricultural Research Council (CRA-CER), S.S. 16, Km 675, 71122, Foggia, Italy

## Summary

Author for correspondence:

Roberto Papa

Tel. +39 0712204984

Email: r.papa@univpm.it

Received: 30 July 2012

Accepted: 7 September 2012

New Phytologist (2012)

doi: 10.1111/j.1469-8137.2012.04377.x

**Key words:** Andes, crop evolution, domestication, introgression, Mesoamerica, nucleotide data, *Phaseolus vulgaris*, sequential bottlenecks.

- We have studied the nucleotide diversity of common bean, *Phaseolus vulgaris*, which is characterized by two independent domestications in two geographically distinct areas: Mesoamerica and the Andes. This provides an important model, as domestication can be studied as a replicate experiment.
- We used nucleotide data from five gene fragments characterized by large introns to analyse 214 accessions (102 wild and 112 domesticated). The wild accessions represent a cross-section of the entire geographical distribution of *P. vulgaris*.
- A reduction in genetic diversity in both of these gene pools was found, which was three-fold greater in Mesoamerica compared with the Andes. This appears to be a result of a bottleneck that occurred before domestication in the Andes, which strongly impoverished this wild germplasm, leading to the minor effect of the subsequent domestication bottleneck (i.e. sequential bottleneck).
- These findings show the importance of considering the evolutionary history of crop species as a major factor that influences their current level and structure of genetic diversity. Furthermore, these data highlight a single domestication event within each gene pool. Although the findings should be interpreted with caution, this evidence indicates the Oaxaca valley in Mesoamerica, and southern Bolivia and northern Argentina in South America, as the origins of common bean domestication.

## Introduction

Plant domestication was a fundamental landmark in human history. This complex process made current crops completely dependent on humans for survival and marked the transition for humans from a hunter-gatherer lifestyle to life in settlements based on agriculture. Understanding the effects of domestication on genetic diversity of a crop is of crucial importance, not only for crop evolution but also for possible applications, such as the implementation of appropriate biodiversity conservation strategies, and the use of genetic variability in breeding programmes.

The main effect of domestication in crops is a reduction in the crop genetic diversity, relative to their wild progenitors. This arises from the reduction in population size, which causes a constriction in diversity at the genome-wide level, and from selection at target genomic regions. This selection also includes neutral loci that are strictly linked to these target genomic regions ('hitchhiking'). The disentangling of the roles of drift and selection in the level and organization of the genetic diversity of a crop is a major issue in investigating plant domestication processes.

In this context, the common bean (*Phaseolus vulgaris*) is characterized by a particular evolutionary history. This species originated in Mexico (Bitocchi *et al.*, 2012) and its expansion to South America resulted in the development of two major ecogeographically distinct gene pools with partial reproductive isolation (Gepts & Bliss, 1985; Koinange & Gepts, 1992): those of Mesoamerica and the Andes. Domestication took place after the formation of these gene pools, and thus their structure is evident in both the wild and the domesticated forms. This clear subdivision of the common bean germplasm is well documented, and it has been defined through several studies (see Papa *et al.*, 2006; Acosta-Gallegos *et al.*, 2007; Angioi *et al.*, 2009; McConnell *et al.*, 2010; Bitocchi *et al.*, 2012; for reviews). These were based on different markers, and they indicate that the common bean has undergone two independent domestication events: one for each gene pool. This evolutionary scenario renders *P. vulgaris* almost unique among crops, and therefore particularly useful to investigate plant domestication (Broughton *et al.*, 2003), as this process can be studied in the same species as a replicated experiment (i.e. in Mesoamerica and in the Andes).

Whether single or multiple domestications occurred within each common bean gene pool remains a matter of debate (Gepts *et al.*, 1986; Beebe *et al.*, 2000, 2001; Papa & Gepts, 2003; McClean *et al.*, 2004; Santalla *et al.*, 2004; Chacón *et al.*, 2005; Papa *et al.*, 2005; McClean & Lee, 2007; Kwak & Gepts, 2009; Kwak *et al.*, 2009; Rossi *et al.*, 2009; Mamidi *et al.*, 2011; Nanni *et al.*, 2011). Along with the number of times a crop was domesticated, a second fundamental question regarding crop evolution concerns the geographical origin of their domestication. Different locations have been suggested for various domestications in the two common bean gene pools (Beebe *et al.*, 2001; Chacón *et al.*, 2005; Kwak *et al.*, 2009). Mexico has been indicated as the cradle of common bean domestication in Mesoamerica, which occurred along with the domestication of other important crops, such as maize (*Zea mays*) and squash (*Cucurbita* spp.) (Gepts *et al.*, 1986, 1988; Smith, 1995, 1997; Piperno & Flannery, 2001; Matsuoka *et al.*, 2002; Doebley, 2004; Piperno *et al.*, 2009; Ranere *et al.*, 2009).

The main objectives of the present study were: first, to investigate and compare the effects of common bean domestication on genetic diversity in both the Mesoamerican and Andean gene pools; secondly, to establish the occurrence of single vs multiple domestication within each of these two gene pools; and finally, to pinpoint the geographical areas of common bean domestication.

To attain these objectives, nucleotide data from five gene fragments were used. Compared with multilocus molecular markers (on which most of the common bean literature has relied), nucleotide data are less inclined to homoplasy (e.g. Wright *et al.*, 2005; Haudry *et al.*, 2007; Morrell & Clegg, 2007). Furthermore, as the absence of recombination is an assumption for most phylogenetic reconstruction methods (Nei & Kumar, 2000), nucleotide data can be very useful for such analyses, especially if based on fragments of a few hundreds of base pairs.

## Materials and Methods

### Plant material

Here, 214 accessions of the common bean (*Phaseolus vulgaris* L.) were used: 102 wild and 112 domesticated accessions (mostly landraces), partially shared with the study of Bitocchi *et al.* (2012). A complete list of the accessions studied, along with their 'passport' information, is available in Supporting Information Table S1.

### PCR and sequencing

The sequences of five gene regions (from 500 to 900 bp) across the common bean genome were used (see Bitocchi *et al.*, 2012). The five gene fragments are all characterized by large introns: four are legume anchor (Leg) markers, developed by Hougaard *et al.* (2008), and one is a gene fragment, *PvSHP1*, developed by Nanni *et al.* (2011), that represents the homologue to the SHATTERPROOF (*SHP1*) gene, which is involved in the control of fruit shattering in *Arabidopsis thaliana*. Leg044 and *PvSHP1* are located on linkage group 6 (LG6), Leg100 and

Leg133 on LG11 (Hougaard *et al.*, 2008; Nanni *et al.*, 2011) and Leg223 on LG9 (P. McClean, pers. comm.); the distance between the loci in the same linkage group was larger than 25 cM (Hougaard *et al.*, 2008; Nanni *et al.*, 2011). The sequences of the 102 wild accessions were available from the study of Bitocchi *et al.* (2012), while those determined as part of the present study are accessible through GeneBank (GeneBank accession numbers JX532176–JX532683). For the *PvSHP1* gene fragment, the sequences of 40 and 35 wild and domesticated accessions, respectively, were obtained from the study of Nanni *et al.* (2011).

### Diversity analysis

Sequence alignment and editing were carried out using MUSCLE, version 3.7 (Edgar, 2004) and BIOEDIT, version 7.0.9.0 (Hall, 1999). Insertion/deletions (indels) were not included in the analyses. The following diversity estimates were computed for the five gene fragments and the concatenated sequences in the wild (W) and domesticated (D) populations within each gene pool (Mesoamerica and the Andes):  $V$  (number of variable sites),  $P_i$  (number of parsimony informative sites),  $S$  (number of singleton variable sites),  $H$  (number of haplotypes),  $H_d$  (haplotype diversity; Nei, 1987),  $\pi$  (Tajima, 1983),  $\theta$  (Watterson, 1975), and Tajima's  $D$  (Tajima, 1989). The number of shared and unique mutations between the wild and domesticated populations within each gene pool was also computed as a measure of divergence. The  $F_{ST}$  statistic (Hudson *et al.*, 1992) permutation test with 1000 replicates (Hudson, 2000) was computed to investigate the degree of differentiation between the wild and domesticated populations of the different gene pools. All of these estimates were obtained using DNASP, version 5.10.01 (Librado & Rozas, 2009). The loss of nucleotide diversity in the domesticated vs wild populations in Mesoamerica and in the Andes was computed using the statistic  $L\pi = 1 - (\pi_D/\pi_W)$  (Vigouroux *et al.*, 2002), where  $\pi_D$  and  $\pi_W$  are the nucleotide diversities in the domesticated and wild populations, respectively. Haplotype trees for each gene fragment were constructed using the median-joining network algorithm, implemented in NETWORK 4.5.1.6 (Bandelt *et al.*, 1999).

### Phylogeography and population structure analysis

There were 180 accessions of *P. vulgaris* for which high-quality sequence data for all of the five gene fragments were available. The relationships among these accessions were analysed using concatenated sequences, by developing an unrooted neighbour-joining (NJ) tree based on Kimura two-parameter distances. The relative support for each node was tested using the bootstrap method with 1000 replicates, in MEGA, version 4 (Tamura *et al.*, 2007).

The hidden genetic population structure of our collection was inferred using a Bayesian model-based approach, implemented in BAPS, version 5.3 (Corander *et al.*, 2003, 2008; Corander & Marttinen, 2006; Corander & Tang, 2007). A mixture analysis was performed to determine the most probable number of populations ( $K$ ) according to the data (Corander & Tang, 2007; Corander *et al.*, 2008). Indeed, BAPS allows the inclusion of  $K$  as a parameter to be estimated, and the best partition of the data into  $K$

clusters was identified as that with the highest marginal log-likelihoods. The 'clustering with linked loci' analysis was chosen, to account for the linkage between sites within aligned sequences; at the same time, the five loci were assumed to be independent. The multilocus sequence typing (MLST) data format was used. The procedures applied to determine the optimal number of genetically homogeneous groups and to estimate the individual admixture proportions were as described in Bitocchi *et al.* (2012).

## Results

### Nucleotide variation

Five gene fragments were sequenced in a large collection of 214 common bean accessions that included wild and domesticated forms. One hundred and two were wild accessions from the Mesoamerican ( $n=49$ ) and Andean ( $n=47$ ) major gene pools, and six were genotypes from northern Peru–Ecuador that are characterized by the ancestral type I phaseolin (Table S1). These accessions represent a cross-section of the entire geographical distribution of the wild form of *P. vulgaris*, from northern Mexico to northwestern Argentina. The remaining accessions (112) were domesticated and are representative of the two main gene pools of the species: the Mesoamerican ( $n=64$ ) and Andean ( $n=48$ ) (Table S1). Each sequenced gene fragment is located within the transcriptional unit and encompasses between 500 and 900 bp, which includes both introns and exons. Altogether, *c.* 3.4 kb per accession was sequenced.

The gene fragments used in the present study were shared with the study of Bitocchi *et al.* (2012), where the structure (identification of exons and introns) for each locus is clearly reported. The variation in the coding regions of the loci in all of the *P. vulgaris* accessions used in this study was very low; indeed, no additional substitutions were found in the coding regions compared with the study of Bitocchi *et al.* (2012), which was conducted only with the wild accessions. The Leg044, Leg100 and Leg223 loci did not show nucleotide substitutions. The Leg133 locus showed one synonymous substitution in three wild Mexican accessions and in all of the accessions characterized by phaseolin type I, and one nonsynonymous substitution in only one wild Andean accession from Argentina (a one-amino-acid replacement: Asp (D) for Ala (A)). Finally, *PvSHP1* showed a nonsynonymous substitution (a one-amino-acid replacement: Gln (Q) for His (H)) in two wild Mexican accessions (both from Nayarit state) and in four Andean landraces. For all the five gene fragments, the subsequent analyses were carried out considering the entire sequences (including both exons and introns).

The main aim of this study was to investigate the domestication process in *P. vulgaris*, and thus the two independent events that occurred in Mesoamerica and the Andes. For this reason we focused our attention on the wild and domesticated populations within each gene pool.

Considering the Mesoamerican gene pool, the genetic diversity estimates were computed for the Mesoamerican wild (MW) and domesticated (MD) populations (Table 1). The results obtained

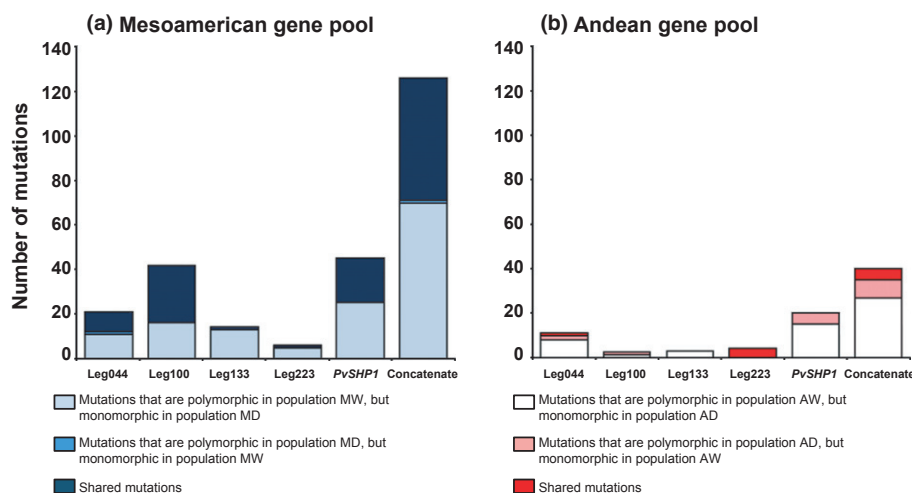
using the concatenated sequence showed that the MW population was characterized by a greater number of variable sites ( $V_{MW}=119$ ), number of haplotypes ( $H_{MW}=34$ ), haplotype diversity ( $Hd_{MW}=0.99$ ), nucleotide diversity  $\pi$  ( $\pi_{MW}=10.66 \times 10^{-3}$ ), and  $\theta$  ( $\theta_{MW}=8.77 \times 10^{-3}$ ) than the MD population ( $V_{MD}=56$ ;  $H_{MD}=19$ ;  $Hd_{MD}=0.81$ ;  $\pi_{MD}=3.00 \times 10^{-3}$ ;  $\theta_{MD}=3.75 \times 10^{-3}$ ). This trend is consistent with the diversity estimates computed individually for each gene fragment (Table 1). The reduction in diversity in the MD compared with the MW population ranged from 0.44 for Leg044 to 0.98 for Leg133, and is strong if we consider the concatenated sequence ( $L\pi=0.72$ ). Nonsignificant values of Tajima's  $D$  were found for the Mesoamerican populations, with the exception of the *PvSHP1* estimate for the MD accessions ( $D_{MD}=-1.83$ ;  $P<0.05$ ). The total number of shared mutations between the MW and MD groups was 55, ranging from one for Leg133 and Leg223, to 26 for Leg100 (Fig. 1a). The number of polymorphic mutations in the MW population that were monomorphic mutations in the MD population was 70, while the MD population showed only one polymorphic mutation, which was monomorphic in the MW accessions (Fig. 1a). The  $F_{ST}$  estimate between the MW and MD genotypes was significant, considering the concatenated sequence ( $F_{ST}=0.20$ ;  $P<0.001$ ; permutation test, Hudson, 2000) and the five loci separately ( $P<0.001$ ), varying from 0.10 (Leg044 and Leg100) to 0.36 (Leg223).

Table 2 gives the genetic diversity estimates for the Andean gene pool. The Andean wild (AW) population showed a greater number of variable sites ( $V_{AW}=32$ ), number of haplotypes ( $H_{AW}=18$ ), haplotype diversity ( $Hd_{AW}=0.86$ ), nucleotide diversity  $\pi$  ( $\pi_{AW}=1.02 \times 10^{-3}$ ), and ( $\pi_{AW}=2.27 \times 10^{-3}$ ) than the Andean domesticated (AD) population ( $V_{AD}=13$ ;  $H_{AD}=11$ ;  $Hd_{AD}=0.82$ ;  $\pi_{MD}=0.74 \times 10^{-3}$ ;  $\theta_{AD}=0.94 \times 10^{-3}$ ). The trend was maintained considering each gene fragment, although there were some exceptions: the Leg100 and Leg223 loci showed higher haplotype and nucleotide diversity for the AD population (Leg100:  $Hd_{AD}=0.08$  and  $\pi_{AD}=0.15 \times 10^{-3}$ ; Leg223:  $Hd_{AD}=0.60$  and  $\pi_{AD}=2.55 \times 10^{-3}$ ) than the AW population (Leg100:  $Hd_{AW}=0.04$  and  $\pi_{AW}=0.08 \times 10^{-3}$ ; Leg223:  $Hd_{AW}=0.45$  and  $\pi_{AW}=1.55 \times 10^{-3}$ ; Table 2). For the Andean gene pool, the total reduction in diversity ( $L\pi$ ) attributable to domestication was 0.27. However, discordant results were obtained considering each gene fragment; indeed, the Leg100 and Leg223 loci showed high nucleotide diversity for the AD population compared with the AW population, while Leg044, Leg133 and *PvSHP1* showed a loss of diversity for the AD population, varying from 0.51 to 1.0 (Table 2). All of the Tajima's  $D$  estimates were nonsignificant; however, it is interesting to note that, for all of the five loci, the Andean populations showed high negative Tajima's  $D$  estimates, ranging from  $-0.58$  (Leg223) to  $-1.79$  (*PvSHP1*), with the only exception being the AD population for Leg223. The total number of shared mutations between the AW and AD groups was five (Fig. 1b); the number of polymorphic mutations in the AW population that were monomorphic in the AD population was 27, while the domesticated population showed eight polymorphic mutations that were monomorphic in the wild accessions (Fig. 1b). The  $F_{ST}$  estimate

**Table 1** Population genetics statistics for the five gene fragments and the concatenated sequences in the Mesoamerican wild and domesticated populations of *Phaseolus vulgaris*

Locus	Pop	<i>n</i>	<i>V</i>	<i>Pi</i>	<i>S</i>	<i>H</i>	<i>Hd</i>	$\pi \times 10^{-3}$	$L\pi$	$\theta \times 10^{-3}$	<i>D</i>
Leg044	MW	44	20	10	10	9	0.82	4.32	0.44	5.58	−0.73ns
	MD	62	10	9	1	5	0.37	2.42		2.58	−0.18ns
Leg100	MW	47	40	39	1	10	0.83	22.93	0.54	16.09	1.46 ns
	MD	63	26	24	2	5	0.62	10.47		9.80	0.22 ns
Leg133	MW	48	13	12	1	6	0.75	6.23	0.98	5.06	0.70 ns
	MD	63	1	1	0	2	0.06	0.11		0.37	−0.90 ns
Leg223	MW	43	6	6	0	7	0.79	3.05	0.95	3.18	−0.11 ns
	MD	62	1	1	0	2	0.06	0.15		0.49	−0.89 ns
<i>PvSHP1</i>	MW	47	42	36	6	20	0.92	15.98	0.87	11.20	1.47 ns
	MD	61	20	18	2	4	0.13	2.06		5.03	−1.83*
Concatenate	MW	37	119	98	21	34	0.99	10.66	0.72	8.77	–
	MD	56	56	51	5	19	0.81	3.00		3.75	–

\*,  $P < 0.05$ ; Pop, population; MW, Mesoamerican wild; MD, Mesoamerican domesticated; *n*, number of sequences used for the analysis; *V*, variable sites; *S*, singleton variable sites; *Pi*, parsimony informative sites; *H*, number of haplotypes; *Hd*, haplotype diversity;  $\pi \times 10^{-3}$  (Tajima, 1983) and  $\theta \times 10^{-3}$  (Watterson, 1975), two measures of nucleotide diversity;  $L\pi$ , loss of nucleotide diversity in the Mesoamerican domesticated (MD) population vs Mesoamerican wild (MW) population; *D*, Tajima's (1989) *D* parameter for testing neutrality; ns, not significant; *PvSHP1*, the homologue to the SHATTERPROOF (*SHP1*) gene, which is involved in the control of fruit shattering in *Arabidopsis thaliana*.



**Fig. 1** Mutation patterns for the nucleotide data from five gene fragments across the two gene pools of common bean (*Phaseolus vulgaris*). (a) Shared and unique mutations between the Mesoamerican wild (MW) and domesticated (MD) populations. (b) Shared and unique mutations between the Andean wild (AW) and domesticated (AD) populations.

between the AW and AD genotypes was significant, considering the concatenated sequence ( $F_{ST} = 0.06$ ;  $P < 0.001$ ; permutation test, Hudson, 2000).

If we compare the Mesoamerican and Andean gene pools, there was a clear difference in the level of genetic diversity: the Andean gene pool showed a substantially lower level of genetic diversity than the Mesoamerican gene pool for all of the genetic diversity estimates (Tables 1, 2).

### Phylogeography and population structure analysis

The population structure analysis was performed considering 180 accessions (for which the sequences of all of the five loci were available), and this showed the best partition of our sample into six genetic groups (Fig. 2). The same partition was identified by Bitocchi *et al.* (2012), who performed the analysis only for the wild accessions.

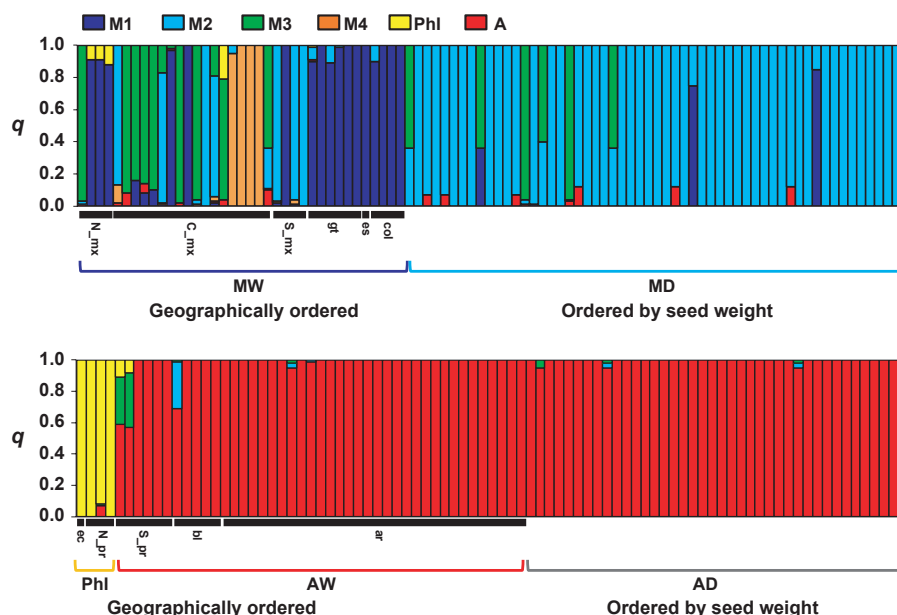
A percentage of membership ( $q$ ) for the six genetic groups was given for each accession (Fig. 2). We considered a threshold of  $q \geq 0.70$  to assign an individual to one of the groups identified. The Mesoamerican accessions were characterized by four different groups: M1, M2, M3 and M4. Nineteen accessions were assigned to group M1 ( $q_{M1} \geq 0.75$ ), 17 of which were MW genotypes distributed geographically throughout Mesoamerica (six from Mexico, six from Guatemala, four from Colombia, and one from El Salvador), while two were MD genotypes. Group M2 was characterized by 55 accessions ( $q_{M2} \geq 0.75$ ), and this group included most of the MD accessions ( $n = 48$ ) and seven MW genotypes from Mexico. The third group (M3) consisted of 10 accessions ( $q_{M3} \geq 0.75$ ): eight wild and two domesticated genotypes from Mexico. Four wild accessions from Mexico belonged to group M4 ( $q_{M4} \geq 0.95$ ). Five Mesoamerican accessions were admixed: one MW genotype (PI325677) from the Mexican state of Morelos that showed a high percentage of



**Table 2** Population genetics statistics for the five gene fragments and the concatenated sequences in the Andean wild and domesticated populations of *Phaseolus vulgaris*

Locus	Pop	<i>n</i>	<i>V</i>	<i>Pi</i>	<i>S</i>	<i>H</i>	<i>Hd</i>	$\pi \times 10^{-3}$	<i>L</i> $\pi$	$\theta \times 10^{-3}$	<i>D</i>
Leg044	AW	46	9	3	6	5	0.47	0.94	0.56	2.49	−1.77 ns
	AD	48	3	1	2	3	0.29	0.41		0.82	−1.04 ns
Leg100	AW	46	1	0	1	2	0.04	0.08	−0.47	0.40	−1.11 ns
	AD	47	1	1	0	2	0.08	0.15		0.40	−0.86 ns
Leg133	AW	47	3	1	2	4	0.20	0.42	1.00	1.17	−1.34 ns
	AD	47	0	0	0	1	0.00	0.00		0.00	–
Leg223	AW	47	4	3	1	4	0.45	1.55	−0.39	2.08	−0.58 ns
	AD	48	4	4	0	4	0.60	2.55		2.07	0.53 ns
<i>PvSHP1</i>	AW	45	15	15	0	4	0.32	1.72	0.51	4.04	−1.79 ns
	AD	42	5	5	0	4	0.18	0.85		1.37	−0.96 ns
Concatenate	AW	43	32	22	10	18	0.86	1.02	0.27	2.27	–
	AD	40	13	11	2	11	0.82	0.74		0.94	–

Pop, population; AW, Andean wild; AD, Andean domesticated; *n*, number of sequences used for the analysis; *V*, variable sites; *S*, singleton variable sites; *Pi*, parsimony informative sites; *H*, number of haplotypes; *Hd*, haplotype diversity;  $\pi \times 10^{-3}$  (Tajima, 1983) and  $\theta \times 10^{-3}$  (Watterson, 1975), two measures of nucleotide diversity; *L* $\pi$ , loss of nucleotide diversity in the Andean domesticated (AD) population vs Andean wild (AW) population; *D*, Tajima's *D* parameter (1989) for testing neutrality; ns, not significant; *PvSHP1*, the homologue to the SHATTERPROOF (*SHP1*) gene, which is involved in the control of fruit shattering in *Arabidopsis thaliana*.



**Fig. 2** Population structure, showing percentages of membership (*q*) for each of the clusters identified (M1, M2, M3, M4, PhI, and A; colour-coded as indicated). Each accession of common bean (*Phaseolus vulgaris*) is represented by a vertical line divided into coloured segments, the lengths of which indicate the proportions of the genome attributed to each cluster. The wild accessions are ordered according to latitude: for Mesoamerica, from northern Mexico to Colombia: N\_mx, northern Mexico; C\_mx, central Mexico; S\_mx, southern Mexico; gt, Guatemala; es, El Salvador; col, Colombia; for South America, from Ecuador to northern Argentina: ec, Ecuador; N\_pr, northern Peru; S\_pr, southern Peru; bl, Bolivia; ar, Argentina. The countries of origin are indicated by the horizontal black bar. The domesticated accessions are ordered by seed weight. MW, Mesoamerican wild; MD, Mesoamerican domesticated; AW, Andean wild; AD, Andean domesticated; PhI, type I phaseolin (northern Peru–Ecuador).

membership for group M3 ( $q_{M3} = 0.64$ ) followed by group M2 ( $q_{M2} = 0.25$ ); four domesticated genotypes that showed a high percentage of membership for group M3 ( $0.60 \leq q_{M3} \leq 0.64$ ), three of which showed high introgression from group M2 (PI309785, PI417760, G2571;  $0.36 \leq q_{M2} \leq 0.40$ ), and one from group M1 (PI313755;  $q_{M1} = 0.36$ ). Of the Andean accessions, 96.4% (40 genotypes for both the AW and AD forms) were clearly assigned to group A ( $q_A \geq 0.95$ ). Only three AW

accessions were admixed: one almost assigned to group A (W618826;  $q_A = 0.69$  and  $q_{M2} = 0.31$ ) and two AW genotypes from central Peru (the Junin and Huanuco regions) that showed a high percentage of membership for group A (G23420:  $q_A = 0.57$ ,  $q_{M3} = 0.35$ , and  $q_{PhI} = 0.08$ ; G12856:  $q_A = 0.59$ ,  $q_{M3} = 0.30$ , and  $q_{PhI} = 0.11$ ). All of the wild accessions characterized by phaseolin type I were assigned to a unique group (PhI;  $q_{PhI} \geq 0.92$ ).

Considering that no population structure was identified in the Andean gene pool, and that possible substructures might be hidden by the main population structure (Evanno *et al.*, 2005), we performed the same analysis only considering the Andean accessions. Four genetic groups were identified (Fig. 3; from A1 to A4). Group A1 included two AW accessions ( $q_{A1} = 1.00$ ), which were the two genotypes from central Peru that were identified as admixed in the previous analysis. Most of the accessions were assigned to group A2 (38 AW and 30 AD genotypes;  $q_{A2} \geq 0.90$ ). Group A3 included nine accessions ( $q_{A3} \geq 0.83$ ): three AW genotypes, two of which were from Argentina (the Jujuy Province) and one from Bolivia (the Chuquisaca Department), and six AD genotypes. Only AD accessions were included in group A4 ( $q_{A4} \geq 0.95$ ). One AD accession (PI207370) was admixed ( $q_{A2} = 0.51$  and  $q_{A4} = 0.49$ ).

The haplotype networks for each of the five loci are shown in Fig. 4. Table S1 gives the relative haplotype for each locus for all of the accessions. The number of haplotypes for each gene ranged from eight (Leg223) to 30 (*PvSHP1*). The number of MW haplotypes was the highest, varying from six (Leg223) to 20 (*PvSHP1*). These were scattered all along the tree, as previously observed by Bitocchi *et al.* (2012) when they analysed only the wild accessions. The MD genotypes showed a major haplotype (high frequency) for almost all of the loci, with the only exception being the Leg100 network tree, where there were two main haplotypes, although separated by only a few (four) mutational steps. These major haplotypes were always shared with the MW genotypes. A few domesticated genotypes showed minor haplotypes that were often shared with Andean and other MW genotypes.

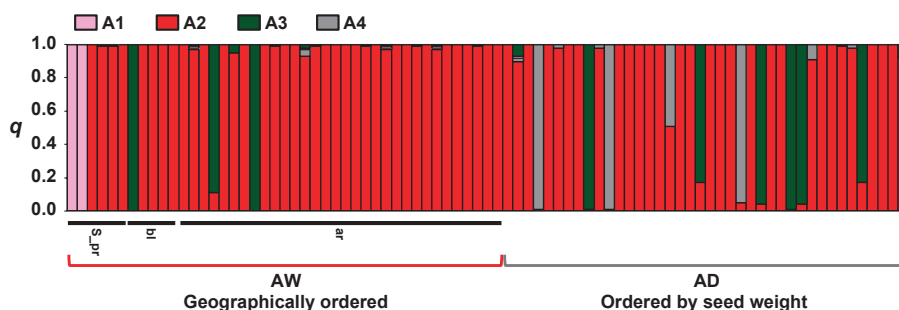
No clear distinction was seen between the AW and AD forms. For almost all of the five loci, they shared the major haplotype, which was characterized by a high frequency with few additional minor haplotypes.

Table S2 gives the relationships between the BAPS and haplotype results for all of the populations considered (MW, MD, AW, AD, and PhI) and for each locus. A detail from Table S2 is given in Table 3. In particular, for the Mesoamerican gene pool, we focused on the eight MD accessions that did not belong to the main M2 cluster (highest frequency of MD accessions) (see

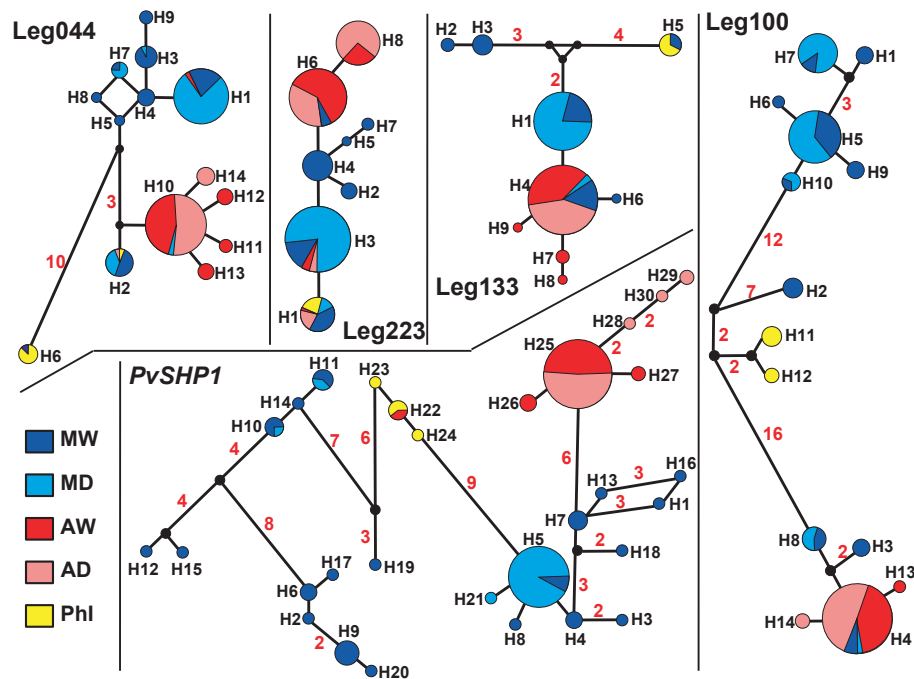
Fig. 2). Two MD accessions were assigned to cluster M1 and two to cluster M3, and four were introgressed, but with a high percentage of membership ( $0.60 \leq q_{M3} \leq 0.64$ ) for cluster M3. The two domesticated M1 and M3 accessions shared the same haplotype as the majority of the domesticated accessions (M2) for four out of the five loci, with *PvSHP1* and Leg100 being the exceptions, respectively (Table 3a). Similarly, the introgressed M3 domesticated accessions showed a haplotype that was not present in the M2 accessions only for one locus (Leg100). An exception here was the PI313755 accession, which showed Leg100 and *PvSHP1* haplotypes not shared with M2 genotypes (Table 3a). Among the four introgressed MD accessions, this accession was the only one that showed higher percentages of membership for clusters M1 and M3 (Fig. 2). A similar scenario was seen for the AD accessions (Table 3b), where the six A3 AD accessions shared the same haplotype as the majority of the A2 AD accessions for almost all of the loci, with the Leg223 locus being the only exception.

An NJ tree was constructed to investigate the relationships between the accessions, with the main aim being to identify the geographical locations of domestication (Fig. 5). However, considering that within each gene pool the recombination between domesticated and wild populations can affect the results, we also constructed an NJ tree by excluding all of the accessions that could have been derived from hybridization between forms after domestication (Fig. 6). This was done to take into account the occurrence of gene flow between the wild and domesticated forms.

Fig. 5 shows the relationships among all of the accessions studied. The structure of the NJ tree parallels that obtained using BAPS. In particular, two main clusters (A and B; Fig. 5) are seen, with the former (A) including all of the Andean and Mesoamerican M3 accessions, and the latter (B) characterized by the remaining groups (M1, M2, M4 and PhI). The B cluster was further subdivided into two groups (B1 and B2): B1 showed a clear relationship between the PhI and the Mesoamerican M4 accessions, and B2 grouped the Mesoamerican M1 and M2 accessions. These main relationships between the different genetic groups were consistent with those highlighted by Bitocchi *et al.* (2012).



**Fig. 3** Population structure in the Andean gene pool, showing percentages of membership ( $q$ ) for each of the clusters identified for the Andean accessions (A1, A2, A3 and A4; colour-coded as indicated). Each accession of common bean (*Phaseolus vulgaris*) is represented by a vertical line divided into coloured segments, the lengths of which indicate the proportions of the genome attributed to each cluster. The wild accessions are ordered according to latitude, from southern Peru to northern Argentina: S\_pr, southern Peru; bl, Bolivia; ar, Argentina. The countries of origin are indicated by the horizontal black bar. The domesticated accessions are ordered according to seed weight. AW, Andean wild; AD, Andean domesticated.



**Fig. 4** Haplotype networks of the five nuclear loci. Each circle represents a single different haplotype, with the circle size proportional to the number of individuals that carry the same haplotype. Black circles indicate missing intermediate haplotypes. The lengths of the lines are proportional to the number of mutational steps, as indicated when  $> 1$ . The colours show the genotypes of *Phaseolus vulgaris* belonging to the Mesoamerican wild (MW), Mesoamerican domesticated (MD), Andean wild (AW), Andean domesticated (AD) and type I phaseolin (PhI; northern Peru–Ecuador) populations. The haplotype label for each sample is also reported in Supporting Information Table S1.

Taking into consideration the Mesoamerican gene pool, we found that the MD accessions were assigned to three (M1, M2 and M3) of the four groups identified. However, excluding the four MD accessions that showed admixture ( $q \leq 0.70$ ), only two accessions were assigned to either the M1 or M3 group, while most of them (48) were assigned to group M2. In the NJ tree, six of the 48 M2 MD accessions were separated from the others (even if the split was not supported by a bootstrap value  $> 0.50$ ; Fig. 5). These accessions had a higher level of admixture ( $0.70 \leq q_{M2} < 0.95$ ) than the remaining M2 domesticated accessions ( $q_{M2} \geq 0.95$ ). Interestingly, a group of four wild M2 Mexican genotypes (two from Oaxaca, and one each from the Jalisco and Puebla States) represented the outgroup of the M2 domesticated accessions. The two MD accessions assigned to group M1 clustered with a set of MW M1 accessions from Mexico and Guatemala; however, these domesticated genotypes also showed a level of admixture  $> 5\%$ . Finally, the domesticated accessions of the M3 group clustered with a wild Mexican accession from Cihuahua State.

Considering the Andean accessions, there was an AW A3 genotype (W618826) from Bolivia (the Chuquisaca Department) that represented the outgroup of all of the other accessions. The remaining A3 genotypes (both AW and AD) were separated by the more numerous group that is characterized by all of the A1, A2 and A4 Andean accessions (Fig. 5).

Fig. 6 shows the NJ tree constructed excluding the admixed domesticated genotypes ( $q < 0.95$ ) and all of the weedy individuals (growing only within or around a cultivated field). As expected, the main structure of the tree was maintained, but there

were some interesting changes that were related, in particular, to the relationships between the wild and domesticated accessions within the two gene pools. Indeed, the M2 MD accessions were strictly related to two wild accessions from Mexico (from Oaxaca State). The remaining two MD accessions were those of the M3 group, which clustered with an MW accession from Mexico (Fig. 6). This association of two Oaxaca wild genotypes with the domesticated group was also observed in the results of the haplotype networks (data not shown).

The relationships between the Andean accessions were slightly modified, with the two AW accessions of the A1 group as the outgroup; these were both from Peru. The other accessions were split into two clusters: one characterized by the A3 genotypes (two AW from Argentina, Jujuy Province, one AW from Bolivia, Chuquisaca Department, and four AD accessions), and the other including all of the A2 and A4 accessions, and thus the majority of the AW and AD genotypes (Fig. 6).

## Discussion

### Domestication and genetic diversity

Our results show that the domestication of the common bean in Mesoamerica induced a severe reduction in genetic diversity. The domesticated populations showed an overall loss of diversity of 72%, and this was consistently reproduced for all of the five genes studied (range, 44–98%). This confirmed the results obtained by Nanni *et al.* (2011), with just one gene fragment (*PvSHP1*) that was shared with the present study.

**Table 3** Relationship between *Phaseolus vulgaris* haplotype and BAPS assignments; haplotypes of (a) the eight Mesoamerican domesticated (MD) accessions assigned by BAPS to clusters M1 and M3 and (b) the six Andean domesticated (AD) accessions assigned by BAPS to cluster A3

Haplotypes	BAPS cluster (no. of individuals)		
	M1 (2)	M3 (2)	M3 <sup>a</sup> (4)
(a) Mesoamerican gene pool			
Leg044 haplotypes			
H1 <sup>b</sup>	2		3
H2 <sup>c</sup>		2	
H7 <sup>c</sup>			1
Leg100 haplotypes			
H5 <sup>b</sup>	1		
H7 <sup>b</sup>	1		
H8		2	4
Leg133 haplotype			
H1 <sup>b</sup>	2	2	4
Leg223 haplotype			
H3 <sup>b</sup>	2	2	4
PvSHP1 haplotypes			
H5 <sup>b</sup>		2	3
H10	1		
H11	1		1
(b) Andean gene pool			
Leg044 haplotype			
H10 <sup>b</sup>		6	
Leg100 haplotype			
H4 <sup>b</sup>		6	
Leg133 haplotype			
H4 <sup>b</sup>		6	
Leg223 haplotypes			
H1		4	
H3		2	
PvSHP1 haplotype			
H25 <sup>b</sup>		6	

This table is a detail taken from Supporting Information Table S2, where the haplotype and BAPS assignment relationships are reported for all of the accessions.

<sup>a</sup>Four Mesoamerican wild (MW) accessions that are introgressed ( $q < 0.70$ ) were included in cluster M3 because of their relatively high ( $0.60 \leq q_{M3} \leq 0.64$ ) percentage of membership of this cluster.

<sup>b</sup>The major haplotypes (characterized by highest frequencies of domesticated accessions; see Fig. 2): the M2 and A2 clusters for the Mesoamerican and Andean gene pools, respectively;

<sup>c</sup>Haplotypes present also in seven M2 accessions; bold, accessions that show haplotypes different from those of the main clusters (M2 and A2). PvSHP1, the homologue to the SHATTERPROOF (SHP1) gene, which is involved in the control of fruit shattering in *Arabidopsis thaliana*.

A reduction was also detected using molecular markers, but at the nucleotide level we observed a much higher reduction. The reduction in Mesoamerica was 32% and 11% for AFLPs (amplified fragment length polymorphisms; Rossi *et al.*, 2009) and SSRs (simple sequence repeats; Kwak & Gepts, 2009), respectively. This pattern was associated with the marker type and was also confirmed in the Andes data (nucleotide data,  $L\pi = 27\%$ , present study; AFLPs, no reduction, Rossi *et al.*,

2009; SSRs,  $L\pi = 9\%$ , Kwak & Gepts, 2009). These differences among the markers can be explained as being attributable to different mutation rates. This is further evidence of the fundamental role of marker mutation rates in describing the diversity of plant populations (Thuillet *et al.*, 2005). Indeed, as suggested in previous studies (Glémin & Bataillon, 2009; Rossi *et al.*, 2009; Nanni *et al.*, 2011; Bitocchi *et al.*, 2012), in populations that have experienced a relatively recent bottleneck, the differences in loss of diversity estimates using different markers can be explained by their diverse mutation rates: markers that are characterized by high mutation rates, such as SSRs, recover the diversity lost after the domestication bottleneck more rapidly than markers with lower mutation rates, such as sequence data.

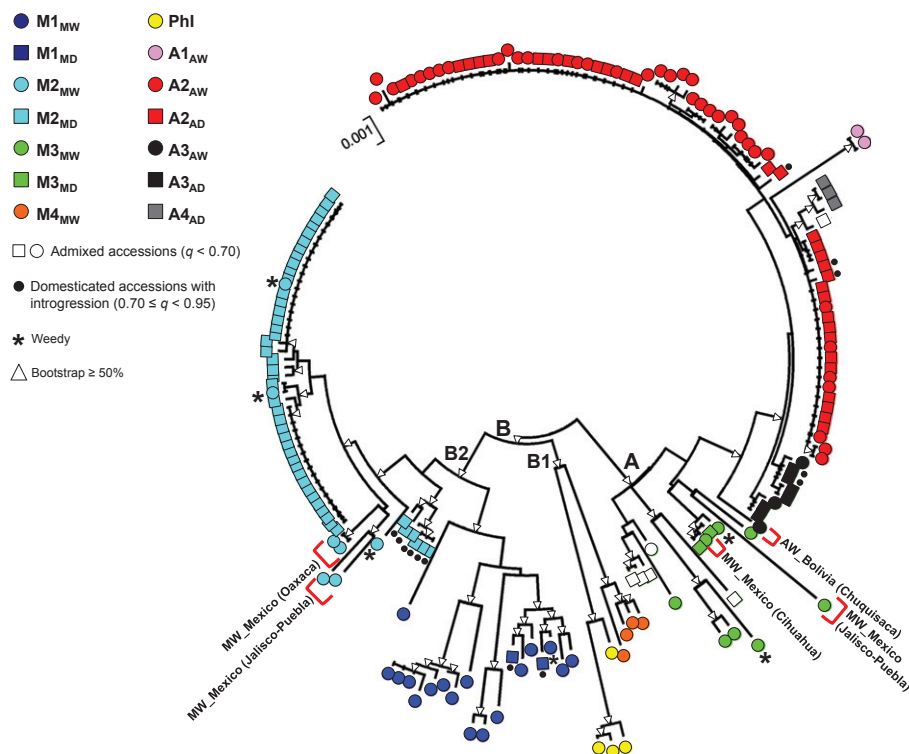
Our results show that, in both Mesoamerica and the Andes, the reductions in genetic diversity in the domesticated populations were in agreement with most of the published data (e.g. Papa & Gepts, 2003; Chacón *et al.*, 2005; Kwak & Gepts, 2009; Rossi *et al.*, 2009; Nanni *et al.*, 2011) and with theoretical expectations. Indeed, the reduction in genetic diversity in crops compared with their wild progenitors is a major effect of domestication that is imposed by founder effects; that is, a small initial population size that causes a constriction of the diversity at the genome-wide level.

Several studies have used nucleotide data to estimate the reduction in the genetic diversity of different major crops compared with their wild progenitors (Table 4). We found that the high reduction in genetic diversity attributable to domestication in *P. vulgaris* from the Mesoamerican gene pool was consistent with what has been seen for most inbreeding species, such as barley (*Hordeum vulgare*) ( $L\pi = 57\text{--}64\%$ ; Caldwell *et al.*, 2006; Kilian *et al.*, 2006) and rice (*Oryza sativa*) (62% and 69% for *indica* and *japonica* ssp., respectively; Caicedo *et al.*, 2007). The greatest loss of diversity appears to be characteristic of self-pollinating species (Table 4), as also suggested by Glémin & Bataillon (2009); this is probably because, when acting on domestication genes, selection affects a larger fraction of the genome in selfing species, because of the stronger genetic linkage (e.g. Caicedo *et al.*, 2007; Papa *et al.*, 2007), and restoration of genetic diversity after domestication through gene flow from wild to crop pollen is more likely in outcrossing than in selfing species (Glémin & Bataillon, 2009).

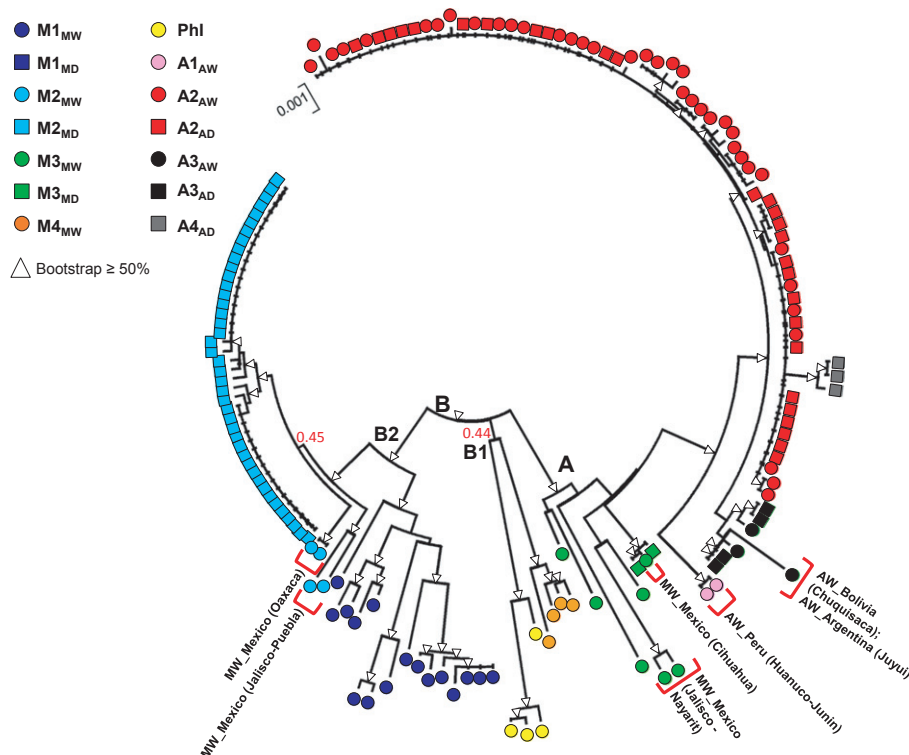
Compared with Mesoamerica, in the Andes there was a much smaller reduction in diversity associated with the process of domestication: three-fold lower. The most probable explanation for such different levels of reduction of diversity in the domesticated germplasm is that in the Andes domestication arose from wild germplasm which, in contrast to the Mesoamerican germplasm, was highly impoverished in genetic variability as a result of the bottleneck that occurred in the Andes before domestication (Rossi *et al.*, 2009; Nanni *et al.*, 2011; Bitocchi *et al.*, 2012; Mamidi *et al.*, 2012). Indeed, populations that have already lost most of their original genetic diversity might be unlikely to lose more during subsequent bottlenecks if they are already genetically depauperated, allowing also that the small effective population size that gave rise to the domesticated germplasm is sufficient to capture all of the genetic diversity that remains in a genetically depauperated species (Taylor & Jamieson, 2008). Thus, while



**Fig. 5** Phylogeographical relationships. The unrooted neighbour-joining (NJ) bootstrap tree was inferred from the concatenated sequence data (180 accessions). Each set of *Phaseolus vulgaris* accessions is represented by a coloured circle (wild accessions) or square (domesticated accessions), and each colour indicates membership of the BAPS group (as indicated). White circles and squares, admixed accessions ( $q < 0.70$ ); black circles, domesticated accessions with low introgression ( $0.70 \leq q < 0.95$ ); \*, weedy accessions. White triangles, nodes for which the bootstrap values are  $> 50\%$ .



**Fig. 6** Phylogeographical relationships without the admixed and introgressed *Phaseolus vulgaris* genotypes. The unrooted NJ bootstrap tree was inferred excluding the admixed ( $q < 0.70$ ) and introgressed ( $0.70 \leq q < 0.95$ ) domesticated accessions and weedy accessions. Each set of accessions is represented by a coloured circle (wild accessions) or square (domesticated accessions), and each colour indicates membership of the BAPS group (as indicated). White triangles are nodes for which bootstrap values are  $> 50\%$ , while bootstrap values close to 50% are shown in red.



the variable effect of domestication on the level of diversity of the domesticated pool might be associated with many factors, as described by Glémin & Bataillon (2009), it might also be explained by the history of the wild populations, and in

particular, the occurrence of one or more bottlenecks before domestication (sequential bottlenecks). This sequential bottleneck hypothesis can also explain the results obtained by Kilian *et al.* (2007) for the strictly autogamous einkorn wheat (*Triticum*

*monococcum*); indeed, they showed no loss of diversity in the domesticated forms compared with the wild race  $\beta$  from which domestication is considered to have taken place (Kilian *et al.*, 2007). However, when the comparison was made with the whole wild gene pool, a reduction in diversity of 37% was detected (Kilian *et al.*, 2007; Table 4).

The conflicting results that we obtained in the two gene pools for the same species is an example of the role of evolutionary history as the main factor that influences the level and structure of the current genetic diversity in crop species. Thus, our findings highlight the potential of the common bean as a model for the study of domestication (Gepts, 2002). This arises from the two geographically distinct and partially isolated gene pools, with independent domestications offering the almost unique opportunity to look at the domestication process as a replicate experiment.

For two genes in the Andes,  $\pi$  (but not  $\theta$ ) was slightly higher for the domesticated compared with the wild population. This was seen for genes with very low polymorphism, and may just be the result of a sampling effect. Alternatively, assuming neutrality, this may be a result of: (1) mutations that occurred after domestication in the cultivated gene pool; (2) the internal structure (i.e. races or varieties) of the diversity of the domesticated pool; (3) genetic erosion in the wild pool that is likely to be associated with asymmetric gene flow (higher from the domesticated to the wild, compared with the opposite direction; Papa & Gepts, 2003; Papa *et al.*, 2005). The first of these explanations might be valid only for Leg100, where one substitution was specific to AD, while for Leg223 all of the four substitutions were shared between AW and AD.

### Single vs multiple domestication events within both the Mesoamerican and Andean gene pools

To disentangle multiple vs single domestication is not an easy task, in relation to the role of post-domestication crop–wild gene flow. In most crops (Ellstrand *et al.*, 1999), and specifically in the common bean (Papa & Gepts, 2003; Papa *et al.*, 2005), the role of gene flow between wild and domesticated populations cannot be ignored. Gene flow tends to make populations more genetically homogeneous, and, as a consequence, this counteracts the population subdivisions resulting from potential genetically independent domestication events (Allaby *et al.*, 2008). Alternatively, because of gene flow, multiple independent introgression of genes from wild relatives to cultivated varieties can simulate multiple domestication events (Badr *et al.*, 2000; Kanazin *et al.*, 2002; Abdel-Ghani *et al.*, 2004). However, compared with multilocus data, our data appear to be useful for addressing this question, because of the use of single gene fragments with large introns (Hougaard *et al.*, 2008; Nanni *et al.*, 2011; Bitocchi *et al.*, 2012), and where intra-locus recombination events are rare or absent. Indeed, compared with molecular markers, sequence data are less prone to homoplasy (e.g. Wright *et al.*, 2005; Haudry *et al.*, 2007; Morrell & Clegg, 2007), and the assumption of no recombination (Nei & Kumar, 2000) is less likely to be violated; also, there is an acceptable discrimination power that is ensured by the degree of polymorphism found in large introns.

Our findings indicate that the subdivisions defined by the population structure and NJ tree analyses are not the result of multiple domestication events, but indicate multiple independent

**Table 4** Loss of nucleotide diversity estimates in domesticated populations compared with wild relatives for different species (modified from Haudry *et al.*, 2007)

Mating system	Crop	Sample size <sup>a</sup>		Nucleotide diversity <sup>b</sup>		No. loci	$L\pi$ (%) <sup>c</sup>	Reference
		W	D	$\pi_{\text{W}} (\times 10^{-3})$	$\pi_{\text{D}} (\times 10^{-3})$			
Outbreeding	Maize ( <i>Zea mays</i> )	16	14	$\pi_{\text{total}} = 6.4$	$\pi_{\text{total}} = 9.5$	774	33	Wright <i>et al.</i> (2005)
		16	25	$\pi_{\text{silent}} = 21.1$	$\pi_{\text{silent}} = 13.1$	12	38	Tenaillon <i>et al.</i> (2004)
Outbreeding	Alfalfa ( <i>Medicago sativa</i> )	19	31	$\pi_{\text{silent}} = 26.6$	$\pi_{\text{silent}} = 20.0$	2	25	Muller <i>et al.</i> (2006)
Outbreeding	Sunflower ( <i>Helianthus annuus</i> )	16	16	$\pi_{\text{total}} = 12.8$	$\pi_{\text{total}} = 5.6$	9	56	Liu & Burke (2006)
Inbreeding	Soybean ( <i>Glycine max</i> )	26	52	$\pi_{\text{total}} = 2.2$	$\pi_{\text{total}} = 1.4$	102	34	Hyten <i>et al.</i> (2006)
Inbreeding	Barley ( <i>Hordeum vulgare</i> )	34	15	$\pi_{\text{silent}} = 16.7$	$\pi_{\text{silent}} = 7.1$	5	57	Caldwell <i>et al.</i> (2006)
		25	20	$\pi_{\text{total}} = 7.7$	$\pi_{\text{total}} = 2.8$	7	64	Kilian <i>et al.</i> (2006)
Inbreeding	Rice ( <i>Oryza sativa</i> ssp. <i>indica</i> )	21	21	$\pi_{\text{total}} = 3.6$	$\pi_{\text{total}} = 1.4$	111	62	Caicedo <i>et al.</i> (2007)
Inbreeding	Rice ( <i>Oryza sativa</i> ssp. <i>japonica</i> )		39		$\pi_{\text{total}} = 1.1$	111	69	
Inbreeding	Foxtail millet ( <i>Setaria italica</i> )	34	50	$\pi_{\text{total}} = 4.0$	$\pi_{\text{total}} = 2.1$	9	48	Wang <i>et al.</i> (2010)
Inbreeding	Emmer wheat ( <i>Triticum turgidum</i> ssp. <i>dicoccum</i> )	28	12	$\pi_{\text{total}} = 2.7$	$\pi_{\text{total}} = 0.8$	21	70	Haudry <i>et al.</i> (2007)
Inbreeding	Einkorn wheat ( <i>Triticum monococcum</i> ) <sup>d</sup>	321	84	$\pi_{\text{total}} = 4.7$	$\pi_{\text{total}} = 3.0$	18	37	Kilian <i>et al.</i> (2007)
	Einkorn wheat ( <i>Triticum monococcum</i> ) <sup>e</sup>	42		$\pi_{\text{total}} = 1.6$		18	–47	
Inbreeding	Common bean ( <i>Phaseolus vulgaris</i> ) (Mesoamerica)	37	56	$\pi_{\text{total}} = 10.7$	$\pi_{\text{total}} = 3.00$	5	72	This study
	Common bean ( <i>Phaseolus vulgaris</i> ) (Andes)	43	40	$\pi_{\text{total}} = 1.0$	$\pi_{\text{total}} = 0.7$	5	27	

<sup>a</sup>W, wild; D, domesticated. Domesticated samples include landraces and/or modern varieties.

<sup>b</sup>For comparison across the data, the  $\pi_{\text{total}}$  estimates are given; when not available, the  $\pi_{\text{silent}}$  estimates are given.

<sup>c</sup>Loss of nucleotide diversity during domestication (%), calculated as  $L\pi = 1 - (\pi_{\text{D}}/\pi_{\text{W}})$ .

<sup>d</sup>Loss of diversity computed considering the whole *Triticum monococcum* ssp. *boeoticum* sample.

<sup>e</sup>Loss of diversity computed considering only the  $\beta$  form of the *Triticum monococcum* ssp. *boeoticum* sample.

introgressions of genes from wild relatives to cultivated varieties, and also between domesticated materials within and between the two gene pools. Our data indicate that the common bean underwent a single domestication event in both Mesoamerica and in South America.

**The Mesoamerican gene pool** The population structure from multilocus data shows that the MD accessions were mostly assigned to a single cluster (M2), with the remaining few MD accessions assigned with relatively high percentages of membership to different clusters that are also present in the MW populations. The problem arises when trying to explain this result, inasmuch as it might be a consequence of gene flow or it might indicate multiple domestications. Single locus analysis, where the assumption of no recombination might be more solid, was of value in resolving this issue. From the network tree analysis, we can see that, for all of the loci, the MD accessions are grouped into a major single haplotype, or in the case of Leg100, into two similar (four mutational steps) main haplotypes. In all cases, these haplotypes were associated with the M2 cluster (main domesticated group). This relationship confirms the observation obtained at the multilocus level, which indicates that cluster M2 originated from a single domestication. What about the MD accession assigned to a different cluster? We can consider the relationships between BAPS and haplotype assignment for each of the five gene fragments, which showed that the accessions that belong to minor clusters, such as M1 and M3, were included in the main haplotype (or, in the case of Leg044, they shared haplotypes with M2 accessions) for four out of the five loci, while they showed signs of introgression for just one locus (Table 3a). Only one accession (PI313755) showed membership to minor haplotypes for two loci (Leg100 and *PvSHPI*), which confirmed the admixed nature found in the population structure analysis.

**Andean gene pool** Similar to the findings for Mesoamerica, the AD accessions that belong to cluster A3 showed a different haplotype from that characteristic of the majority of AD accessions only for the Leg223 locus (Table 3b), while for four loci they had the main haplotype.

#### Putative geographical origins of common bean domestication in Mesoamerica and the Andes

One of the major issues related to crop domestication is to pinpoint the geographical regions where this process took place. The locations of the common bean domestications within each of the two gene pools remain a matter of debate.

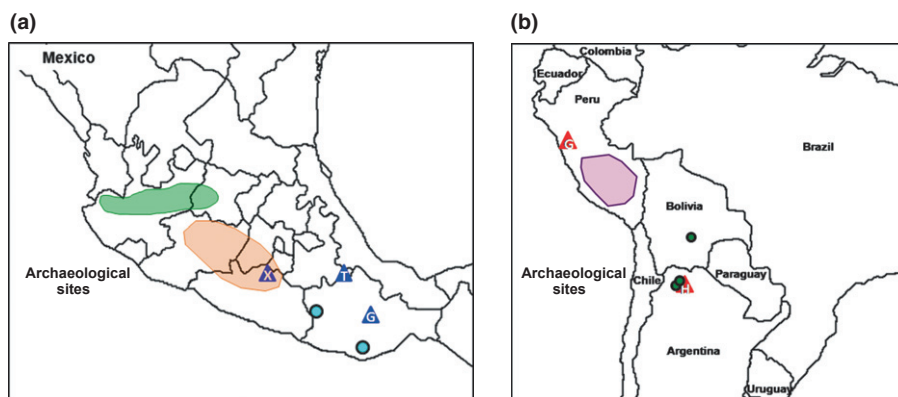
Considering the Mesoamerican gene pool, without a doubt, Mexico represents the cradle of domestication of important crops, such as maize, squash and the common bean, as is well documented in both archaeological and molecular data (Gepts *et al.*, 1986, 1988; Smith, 1995, 1997; Piperno & Flannery, 2001; Matsuoka *et al.*, 2002; Doebley, 2004; Piperno *et al.*, 2009; Ranere *et al.*, 2009). Recently, Kwak *et al.* (2009) used SSR data and proposed a restricted region in the Rio Lerma–Rio Grande de Santiago basin in west-central Mexico as the cradle of

domestication for the common bean in Mesoamerica (Fig. 7). This area does not overlap the Central River Balsas Valley, which instead represents the geographical area identified as the location of maize domestication using both archaeological and molecular evidence (Matsuoka *et al.*, 2002; Piperno *et al.*, 2009; van Heerwaarden *et al.*, 2011). Thus, Kwak *et al.* (2009) suggested that maize and the common bean were probably domesticated in different regions, and that they were reunited later in a single cropping system. Our data appear to suggest a different scenario: the NJ tree obtained when weedy and domesticated genotypes that presented signals of introgression were excluded (Fig. 6) indicated two MW genotypes from Oaxaca State as being most closely related to the MD forms. This area falls further to the south than the area proposed by Kwak *et al.* (2009) for *P. vulgaris*, and also when compared with the area characteristic of maize domestication (Fig. 7). However, it overlaps one of the first areas of the spread of maize through human migration, along the Mexican rivers (Zizumbo-Villarreal & Colunga-GarcíaMarín, 2010). Thus, we suggest that common bean domestication in Mesoamerica occurred in the geographical area of the Oaxaca Valley (Fig. 7).

Considering the Andean gene pool, different geographical areas have been indicated as the putative location of common bean domestication in South America. After analysing chloroplast DNA polymorphisms in a wide set of common bean accessions from the Americas, Chacón *et al.* (2005) suggested central-southern Peru as the geographical area where the common bean was domesticated in the Andes. In contrast, using AFLPs, Beebe *et al.* (2001) showed that wild beans from eastern Bolivia and northern Argentina group very closely with the Andean domesticated beans, and thus they suggested that this location might have been an important primary domestication site. Common bean archaeological remains have been found and dated from 8000 to 10 000 yr BP in the Andes (the Guitarrero cave, Ancash Department, Peru; Kaplan *et al.*, 1973; Huachichocana, Jujuy Province, Argentina; Tarrago, 1980). However, recently, these dating methods have been questioned by archaeologists, and thus the earliest record for the common bean in South America is 4300 yr BP, as revised using direct accelerator mass spectrometry dating (Kaplan & Lynch, 1999) on samples from the Guitarrero cave (Fig. 7).

Our data indicate that the majority of the AW and AD accessions (belonging mostly to the A2 BAPS group) are included in a single clade also in the NJ tree. Three AW and six AD A3 accessions were outgroups of this clade. As previously indicated, the AD A3 accessions were probably the result of introgression, and thus we focused our attention on the three AW genotypes from southern Bolivia (W618826; Chuquisaca Department) and northern Argentina (G21199 and G19888; Jujuy Province) (Fig. 7), that seem to be the wild materials at the basis of domestication. This result appears consistent with the suggestions of Beebe *et al.* (2001), and with the archaeological data from Huachichocana in the Jujuy Province of Argentina (Tarrago, 1980) which suggest that, for the Andean gene pool, common bean domestication occurred in southern Bolivia and northern Argentina.

However, our findings concerning these putative locations of common bean domestication must be considered with caution;



**Fig. 7** Putative geographical locations of common bean (*Phaseolus vulgaris*) domestications. (a) Mesoamerican gene pool. Light-blue circles, Mesoamerican wild (MW) accessions closest to the Mesoamerican domesticated (MD) population. Blue triangles, archaeological sites where the oldest archaeological common bean and maize remains were found: T, Tehuacán Valley (Puebla State), common bean macro-remains dated c. 2285 calibrated years before present (cal BP; Kaplan & Lynch, 1999); G, Guilá Naquitz Cave (Oaxaca State), common bean macro-remains dated c. 2098 cal BP (Kaplan & Lynch, 1999), and maize records dated c. 6300 cal BP (Piperno & Flannery, 2001); X, Xihuatoxtla Shelter (Guerrero State), oldest maize records dated c. 8700 cal BP (Piperno *et al.*, 2009). Green area, location of common bean domestication in Mesoamerica suggested by Kwak *et al.* (2009); orange area, location of maize domestication (Matsuoka *et al.*, 2002; Piperno *et al.*, 2009; van Heerwaarden *et al.*, 2011). (b) Andean gene pool. Green circles, Andean wild (AW) accessions that represent the outgroup of the Andean domesticated (AD) population. Red triangles, archaeological sites where the oldest (from 8000 to 10 000 cal BP) common bean archaeological remains were found: G, Guitarrero Cave, Ancash Department, Peru (Kaplan *et al.*, 1973); H, Huachichocana, Jujuy Province, Argentina (Tarrago, 1980). Violet area, location of common bean domestication in South America suggested by Chacón *et al.* (2005).

indeed, for example, gene flow between wild and domesticated common bean might be an alternative explanation. Further studies need to be carried out to obtain a more detailed picture. In particular, for the Andean gene pool, which is characterized by a very low level of genetic diversity, the identification of the domestication area might be more precisely achieved by analysing polymorphisms at loci identified as under selection during domestication.

## Acknowledgements

We would like to thank G. Bertorelle for critical reading of the manuscript and for valuable advice. This work was supported by grants from the Italian Government (MIUR; Grant number 20083PFSXA\_001, PRIN Project 2008), the Università Politecnica delle Marche (2008–2011) and the Marche Region (Grant number L.R.37/99art.2lett.I–PARDGR247/10-DDPF98/CSI10).

## References

- Abdel-Ghani AH, Parzies HK, Omary A, Geiger HH. 2004. Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. *Theoretical and Applied Genetics* **109**: 588–595.
- Acosta-Gallegos JA, Kelly JD, Gepts P. 2007. Pre-breeding in common bean and use of genetic diversity from wild germplasm. *Crop Science* **47**: 44–59.
- Allaby RG, Fuller DQ, Brown TA. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Science, USA* **105**: 13982–13986.
- Angioi SA, Desiderio F, Rau D, Bitocchi E, Attene G, Papa R. 2009. Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. *Plant Biology* **11**: 598–612.
- Badr A, Müller K, Schäfer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, Pozzi C, Rohde W, Salamini F. 2000. On the origin and domestication history of barley (*Hordeum vulgare*). *Molecular Biology and Evolution* **17**: 499–510.
- Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**: 37–48.
- Beebe S, Rengifo J, Gaitan E, Duque MC, Tohme J. 2001. Diversity and origin of Andean landraces of common bean. *Crop Science* **41**: 854–862.
- Beebe S, Skroch P, Tohme J, Duque M, Pedraza F, Nienhuis J. 2000. Structure of genetic diversity among common bean landraces of middle-American origin based on correspondence analysis of RAPD. *Crop Science* **40**: 264–273.
- Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Spagnoletti Zeuli P, Logozzo G, Stougaard J, McClean P, Attene G *et al.* 2012. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proceedings of the National Academy of Science, USA* **109**: E788–E796.
- Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J. 2003. Beans (*Phaseolus* spp.): model food legumes. *Plant and Soil* **252**: 55–128.
- Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR *et al.* 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* **3**: 1745–1756.
- Caldwell KS, Russell J, Langridge P, Powell W. 2006. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* **172**: 557–567.
- Chacón SMI, Pickersgill B, Debouck DG. 2005. Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theoretical and Applied Genetics* **110**: 432–444.
- Corander J, Marttinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology* **15**: 2833–2843.
- Corander J, Marttinen P, Sirén J, Tang J. 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**: 539.
- Corander J, Tang J. 2007. Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* **205**: 19–31.
- Corander J, Waldmann P, Sillanpää MJ. 2003. Bayesian analysis of genetic differentiation between populations. *Genetics* **163**: 367–374.
- Doebley J. 2004. The genetics of maize evolution. *Annual Review of Genetics* **38**: 37–59.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.



- Ellstrand NC, Prentice HC, Hancock JF. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology, Evolution, and Systematics* 30: 539–563.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.
- Gepts P. 2002. A comparison between crop domestication, classical plant breeding, and genetic engineering. *Crop Science* 42: 1780–1790.
- Gepts P, Bliss FA. 1985. F1 hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated bean germplasm. *Journal of Heredity* 76: 447–450.
- Gepts P, Kmiecik K, Pereira P, Bliss FA. 1988. Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability: I. The Americas. *Economic Botany* 42: 73–85.
- Gepts P, Osborn TC, Rashka K, Bliss FA. 1986. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. *Economic Botany* 40: 451–468.
- Glémin S, Bataillon T. 2009. A comparative view of the evolution of grasses under domestication. *New Phytologist* 183: 273–290.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.
- Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glémin S *et al.* 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution* 24: 1506–1517.
- van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Sanchez-Gonzalez JJ, Ross-Ibarra J. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Science, USA* 108: 1088–1092.
- Hougaard BK, Heegaard Madsen L, Sandal N, de Carvalho Moretzsohn M, Fredslund J, Schauer L, Nielsen AM, Rohde T, Sato S, Tabata S *et al.* 2008. Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. *Genetics* 179: 2299–2312.
- Hudson RR. 2000. A new statistic for detecting genetic differentiation. *Genetics* 155: 2011–2014.
- Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583–589.
- Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Science, USA* 103: 16666–16671.
- Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T. 2002. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). *Plant Molecular Biology* 48: 529–537.
- Kaplan L, Lynch T. 1999. *Phaseolus* (Fabaceae) in archaeology: AMS radiocarbon dates and their significance for pre-Columbian agriculture. *Economic Botany* 53: 261–272.
- Kaplan L, Lynch TF, Smith CE. 1973. Early cultivated beans (*Phaseolus vulgaris*) from an intermontane Peruvian valley. *Science* 179: 76–77.
- Kilian B, Ozkan H, Kohl J, von Haeseler A, Barale F, Deusch O, Brandolini A, Yucel C, Martin W, Salamini F. 2006. Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. *Molecular Genetics and Genomics* 276: 230–241.
- Kilian B, Özkan H, Walther A, Kohl J, Dagan T, Salamini F, Martin W. 2007. Molecular diversity at 18 loci in 321 wild and 92 domesticated lines reveals no reduction in nucleotide diversity during *Triticum monococcum* (Einkorn) domestication: implications for the origin of agriculture. *Molecular Biology and Evolution* 24: 2657–2668.
- Koinange EMK, Gepts P. 1992. Hybrid weakness in wild *Phaseolus vulgaris* L. *Journal of Heredity* 83: 135–139.
- Kwak M, Gepts P. 2009. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theoretical and Applied Genetics* 118: 979–992.
- Kwak M, Kami J, Gepts P. 2009. The putative Mesoamerican domestication center of *Phaseolus vulgaris* is located in the Lerma-Santiago Basin of Mexico. *Crop Science* 49: 554–563.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 173: 321–330.
- Mamidi S, Rossi M, Annam D, Moghaddam S, Lee R, Papa R, McClean P. 2011. Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Functional Plant Biology* 38: 953–967.
- Mamidi S, Rossi M, Moghaddam M, Annam D, Lee R, Papa R, McClean PE. 2012. Demographic factors shaped diversity in the two gene pools of wild common bean (*Phaseolus vulgaris* L.). *Heredity*, in press.
- Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Science, USA* 99: 6080–6084.
- McClean PE, Lee RK. 2007. Genetic architecture of chalcone isomerase non-coding regions in common bean (*Phaseolus vulgaris* L.). *Genome* 50: 203–214.
- McClean PE, Lee RK, Miklas PN. 2004. Sequence diversity analysis of dihydroflavonol 4-reductase intron1 in common bean. *Genome* 47: 266–280.
- McConnell M, Mamidi S, Lee R, Chikara S, Rossi M, Papa R, McClean P. 2010. Syntenic relationships among legumes revealed using a gene-based genetic linkage map of common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 121: 1103–1116.
- Morrell PL, Clegg MT. 2007. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Science, USA* 104: 3289–3294.
- Muller MH, Poncet C, Prosperi JM, Santoni S, Ronfort J. 2006. Domestication history in the *Medicago sativa* species complex: inferences from nuclear sequence polymorphism. *Molecular Ecology* 15: 1589–1602.
- Nanni L, Bitocchi E, Bellucci E, Rossi M, Rau D, Attene G, Gepts P, Papa R. 2011. Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (*PvSHP1*) in domesticated and wild common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 123: 1341–1357.
- Nei M. 1987. *Molecular evolutionary genetics*. New York, NY, USA: Columbia University Press.
- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. New York, NY, USA: Oxford University Press.
- Papa R, Acosta J, Delgado-Salinas A, Gepts P. 2005. A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theoretical and Applied Genetics* 111: 1147–1158.
- Papa R, Bellucci E, Rossi M, Leonardi S, Rau D, Gepts P, Nanni L, Giovanna A. 2007. Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Annals of Botany* 100: 1039–1051.
- Papa R, Gepts P. 2003. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical and Applied Genetics* 106: 239–250.
- Papa R, Nanni L, Sicard D, Rau D, Attene G. 2006. The evolution of genetic diversity in *Phaseolus vulgaris* L. In: Motley TJ, Zerega N, Cross H, eds. *Darwin's Harvest – new approaches to the origins, evolution and conservation of crops*. New York, NY, USA: Columbia University Press, 121–142.
- Piperno DR, Flannery K. 2001. The earliest archeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proceedings of the National Academy of Science, USA* 98: 2101–2103.
- Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R. 2009. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Science, USA* 106: 5019–5024.
- Ranere AJ, Piperno DR, Holst I, Dickau R, Iriarte J. 2009. The cultural and chronological context of early Holocene maize and squash domestication in the Central Balsas River Valley, Mexico. *Proceedings of the National Academy of Science, USA* 106: 5014–5018.
- Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, Papa R. 2009. Linkage disequilibrium and population structure in wild and

- domesticated populations of *Phaseolus vulgaris* L. *Evolutionary Applications* 2: 504–522.
- Santalla M, Menéndez-Sevillano MC, Monteagudo AB, De Ron AM. 2004. Genetic diversity of Argentinean common bean and its evolution during domestication. *Euphytica* 135: 75–87.
- Smith BD. 1995. *The emergence of agriculture*. New York, NY, USA: Scientific American Library.
- Smith BD. 1997. The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science* 276: 932–934.
- Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437–460.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software, version 4.0. *Molecular Biology and Evolution* 24: 1596–1599.
- Tarrago MN. 1980. El proceso de agriculturización en el Noroeste Argentino, zona Valliserrana. *Actas del V Congreso Nacional Arqueología Argentina*, vol. 1. San Juan, Argentina: Instituto de Investigaciones Arqueológicas y Museo, Universidad de San Juan, 181–217.
- Taylor SS, Jamieson IG. 2008. No evidence for loss of genetic variation following sequential translocations in extant populations of a genetically depauperate species. *Molecular Ecology* 17: 545–556.
- Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. *Molecular Biology and Evolution* 21: 1214–1225.
- Thuillet AC, Bataillon T, Poirier S, Santoni S, David JL. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. *Genetics* 169: 1589–1599.
- Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proceedings of the National Academy of Science, USA* 99: 9650–9655.
- Wang C, Chen J, Zhi H, Yang L, Li W, Wang Y, Li H, Zhao B, Chen M, Diao X. 2010. Population genetics of foxtail millet and its wild ancestor. *BMC Genetics* 11: 90.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7: 256–276.
- Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. *Science* 308: 1310–1314.
- Zizumbo-Villarreal D, Colunga-GarcíaMarín P. 2010. Origin of agriculture and plant domestication in West Mesoamerica. *Genetic Resources and Crop Evolution* 57: 813–825.

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

### Table S1 Plant material

### Table S2 Relationships between the haplotypes and the BAPS assignments

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



## New Phytologist Tansley Medal

For excellence in plant science

Full details, terms and conditions at:  
**[www.newphytologist.org](http://www.newphytologist.org)**

### Calling all early stage career scientists!

#### Deadline for submissions for 2013: 15 December 2012

Win £2000 (GBP) and have your work highlighted in *New Phytologist*, one of the world's leading plant science journals (2011 Impact Factor 6.645).

- The New Phytologist Tansley Medal is awarded annually in recognition of an outstanding contribution to research in plant science
- This is a global competition open to all plant scientists in the early stages of their career and includes both student and post-doctoral researchers with up to five years experience, excluding career breaks, since gaining/defending their PhD
- Selection is based on a two-stage process:
  - **Stage 1)** Submit your CV, a personal statement and reference: Deadline 15 December 2012
  - **Stage 2)** Submission of a single-authored minireview intended for publication: Deadline: 31 March 2013
- All competition articles that are accepted after peer review will be published in *New Phytologist* and the Tansley medal winner selected by the judges from these final papers.