Theor Appl Genet (2010) 121:829–843
DOI 10.1007/s00122-010-1353-2

ORIGINAL PAPER

Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L.

S. A. Angioi · D. Rau · G. Attene · L. Nanni · E. Bellucci · G. Logozzo · V. Negri · P. L. Spagnoletti Zeuli · R. Papa

Received: 31 December 2009 / Accepted: 1 May 2010 / Published online: 21 May 2010
© Springer-Verlag 2010

Abstract This study focuses on the expansion of *Phaseolus vulgaris* in Europe. The pathways of distribution of beans into and across Europe were very complex, with several introductions from the New World that were combined with direct exchanges between European and other Mediterranean countries. We have analyzed here six chloroplast microsatellite (cpSSR) loci and two unlinked nuclear loci (for phaseolin types and *Pv-shatterproof1*). We have assessed the genetic structure and level of diversity of a large collection of European landraces of *P. vulgaris* (307) in comparison to 94 genotypes from the Americas that are representative of the Andean and Mesoamerican gene pools. First, we show that most of the European common bean landraces (67%) are of Andean origin, and that there are no strong differences across European regions for the proportions of the Andean and Mesoamerican gene pools. Moreover, cytoplasmic diversity is evenly distributed across European regions. Secondly, the cytoplasmic bottleneck that was due to the introduction of *P. vulgaris* into the Old World was very weak or nearly absent. This is in contrast to evidence from nuclear analyses that have suggested a bottleneck of greater intensity. Finally, we estimate that a relatively high proportion of the European bean germplasm (about 44%) was derived from hybridization between the Andean and Mesoamerican gene pools. Moreover, although hybrids are present everywhere in Europe, they show an uneven distribution, with high frequencies in central Europe, and low frequencies in Spain and Italy. On the basis of these data, we suggest that the entire European continent and not only some of the countries therein can be regarded as a secondary diversification center for *P. vulgaris*. Finally, we outline the relevance of these inter-gene pool hybrids for plant breeding.

Communicated by A. Bervillé.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1353-2) contains supplementary material, which is available to authorized users.

S. A. Angioi · D. Rau · G. Attene
Dipartimento di Scienze Agronomiche e Genetica Vegetale Agraria, Università degli Studi di Sassari, Via de Nicola, 07100 Sassari, Italy

L. Nanni · E. Bellucci · R. Papa (✉)
Dipartimento di Scienze Ambientali e delle Produzioni Vegetali, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy
e-mail: r.papa@univpm.it

G. Logozzo · P. L. Spagnoletti Zeuli
Dipartimento di Biologia Difesa e Biotecnologie Agro-Forestali, Università degli Studi della Basilicata, Viale dell'Ateneo Lucano 10, 85100 Potenza, Italy

V. Negri
Dipartimento di Biologia Applicata (DBA), Università degli Studi di Perugia, Borgo XX Giugno 74, 06121 Perugia, Italy

Introduction

The common bean (*Phaseolus vulgaris* L.) is a species of great agronomical interest worldwide, and it represents a major protein source in the human diet (Beebe et al. 2000). Domestication of *P. vulgaris* occurred independently in the Mesoamerican and Andean areas, which gave rise to two highly differentiated gene pools (Gepts and Debouck 1991; Gepts 1998) that are characterized by geographical and partial reproductive isolation (Gepts and Bliss 1985; Paredes and Gepts 1995). The two gene pools can be distinguished according to seed size and other differences

(Gepts et al. 1986). The Mesoamerican germplasm shows as either small (<25 g 100 seed weight^{-1}) or medium (25–40 g 100 seed weight^{-1}) seed, with phaseolin (the major seed-storage proteins of the common bean) patterns as either the S or B types; these characteristics are different from those of the large seeded (>40 g 100 seed weight^{-1}) Andean counterpart, with T, C, H and A phaseolin types (Gepts et al. 1986; Singh et al. 1991). The occurrence of two clearly distinct gene pools that originated from independent domestications of *P. vulgaris* has been recognized by several authors, using various marker systems (Koenig and Gepts 1989; Beebe et al. 2000, 2001; Papa and Gepts 2003; McClean et al. 2004; Rossi et al. 2009; Kwak and Gepts 2009; Nanni et al. submitted).

The Mesoamerican common bean probably arrived in Europe through Spain and Portugal in 1506 (Ortwin-Sauer 1966), and the Andean in the same way in 1528, after the exploration of Peru by Pizarro (Berglund-Brücher and Brücher 1976). The pathways of dissemination of the common bean into and across Europe were very complex, with several introductions from America combined with direct exchanges between European and other Mediterranean countries (see Papa et al. 2006, for a review).

In recent years, combinations of phaseolins and various morphological traits have been intensively used to analyze the structure of the European common bean and the presence of two gene pools. A prevalence of the Andean type (76%) was first detected by Gepts and Bliss (1988), and was then confirmed by Lioi (1989) in an analysis of a large collection from Italy, Greece and Cyprus (66%), by Logozzo et al. (2007) for a broad European collection (76%), and by others, for Portuguese and Spanish samples (Rodiño et al. 2001, 2003; Ocampo et al. 2002). This trend has also been confirmed at a regional scale using phaseolin and other molecular markers (Limongelli et al. 1996; Escribano et al. 1998; Piergiovanni et al. 2000a, b; Sicard et al. 2005; Angioi et al. 2009a). Moreover, recent studies have focused on hybridization between the Andean and Mesoamerican gene pools in Europe. Evidence of this phenomenon has been detected using phaseolins, allozymes and morphological data (Santalla et al. 2002; Rodiño et al. 2006), and ISSRs and SSRs from both the chloroplast and nuclear genomes (Sicard et al. 2005; Angioi et al. 2009a).

Crop expansion from America to Europe has been suggested to have reduced the diversity of the European common bean because of strong founder effect, adaptation to a new environment and consumer preference (Gepts 1999). In comparing the levels of variation at isozyme loci obtained in studies that have analyzed samples from both America (Singh et al. 1991) and the Iberian peninsula (Santalla et al. 2002), Papa et al. (2006) estimated a loss of diversity in Europe of 30% and low differentiation (G_{ST}) between the gene pool in Europe (18%) when compared with the differences in the Americas (47%). These can be explained either by the combined actions of greater gene flow between the two gene pools in Europe, or by convergent evolution for adaptation to European environments. These studies suggested the need for further investigations into the genetic diversity of European bean populations to test these hypotheses; in particular, they called for a comprehensive study including the domesticated common bean from both the Americas and Europe. Indeed, better knowledge of the genetic diversity of *P. vulgaris* in Europe is needed to identify the most appropriate material for the choice of parental lines in breeding programs (Gepts 1998; Gepts et al. 1999; Acosta-Gallegos et al. 2007), and to develop new and more productive varieties with high-quality grains (Gepts et al. 1999; Singh 2001; González et al. 2009).

Chloroplast microsatellites (cpSSRs) are highly polymorphic sequences (Powell et al. 1995; Vendramin et al. 1996; Petit et al. 2005; Elbert and Peakall 2009). In Angiosperms, the chloroplast genome is maternally inherited, which allows clarification of the contributions of seed flow to the genetic structure of populations through a comparison of nuclear markers (Ennos et al. 1999; Angioi et al. 2009a). Moreover, cpDNA markers have proven particularly useful in identifying hybridization events (Elbert and Peakall 2009), and their relatively high mutation rates mean that they are suitable for intraspecific analyses (Hale et al. 2004). CpSSRs are polymorphic among different species and among accessions of *Glycine* (Xu et al. 2002), *Hordeum* (Provan et al. 1999a), *Oryza* (Ishii and McCouch 2000), *Pinus* (Cuenca et al. 2003), *Solanum* (Sukhotu et al. 2006), *Vitis* (Arroyo-Garcia et al. 2002) and *Phaseolus* spp. (Sicard et al. 2005). Recently, Angioi et al. (2009b) selected a set of markers that were highly informative for the study of the diversity and evolution of *P. vulgaris,* in which they were able to identify the two gene pools.

Nanni et al. (submitted) identified and characterized *Pv-shatterproof1* in *P. vulgaris,* a sequence of about 1,200 bp that is homologous to SHATTERPROOF (SHP1) of *Arabidopsis thaliana*. This *Pv-shatterproof1* sequence is highly polymorphic, and it has been particularly useful in investigations into the origin and domestication history of *P. vulgari*s. Using two molecular markers spanning indels, *Pv-shatterproof1* has been genetically mapped in chromosome 6, close to the *V* locus for flower color (Nodari et al. 1993; McClean et al. 2002) and in the same region as a QTL for days to flowering (Blair et al. 2006) and for resistance to common bacterial blight (Miklas et al. 2006). Moreover, the combined information relating to these two markers was shown to unambiguously discriminate between the Andean and Mesoamerican origins. Thus, these two markers are very informative for the tracing of the distribution of the domesticated Andean and Mesoamerican gene pools (Nanni et al. submitted).

The focus of the present study is to investigate the effects of expansion of *P. vulgaris* into and through Europe. For this, we first compared domesticated accessions of *P. vulgaris* from the Americas and Europe (Mesoamerican and Andean gene pools) using cpSSRs. Secondly, we coupled the cpSSR data to nuclear and morphological data. Thus, we have been able to estimate the degree of the cytoplasmic bottleneck in *P. vulgaris* following its expansion in Europe, and the frequency of inter-gene pool hybridization. This information is relevant both in understanding of the mode of evolution of this crop far from its center of origin, and to estimate the potential of European accessions of *P. vulgaris* for plant breeding.

Materials and methods

Plant materials

We analyzed 307 individual genotypes, each of which represented a single accession of European landraces of the common bean. Most of these (279) were derived from a core collection developed by Logozzo et al. (2007), along with some individuals (28) from countries that were less represented in this core collection (e.g. France) (Supplementary Materials, Table A1). The core collection was obtained from an original collection of 544 accessions that covered all of the European countries, in which the Andean phaseolin types 'T' (45.6%) and 'C' (30.7%) prevailed over the Mesoamerican phaseolin type 'S' (23.7%). Moreover, accessions with a cuboid seed shape (39.9%), a darker maroon seed coat color (44.3%), and a uniform seed color (69.6%) were the most frequent. The core collection was developed using a sampling approach that was based on the information available in the GenBank databases and on the phaseolin patterns; it was validated using seven morphological seed traits. The sampling strategy by phaseolin type resulted in a core collection with the following overall phaseolin-type frequencies: 'C' type, 33.0%; 'T' type, 35.7%; and 'S' type, 31.3%.

The European genotypes were compared with 94 domesticated accessions of *P. vulgaris* from the Americas that had been previously assigned to the Andean or Mesoamerican gene pools using morphological traits and phaseolin types. Thus, 44 of these were classified as Andean, and 50 as Mesoamerican.

All of the accessions were provided by the Institute of Plant Genetics and Crop Plant Research (IPK), Germany; Centro Internacional de Agricoltura Tropical (CIAT), Colombia; the United States Department of Agriculture (USDA), USA; the Department of 'Scienze Ambientali e delle Produzioni Vegetali (SAPROV) of the Università Politecnica delle Marche' (Univ. P.M.), Italy; and the Department of Plant Science, UC Davis, USA (Supplementary Materials, Table A2).

DNA extraction and genotyping

For cpSSR and *Pv-shatterproof1* locus analyses (see below), DNA was extracted from an individual plant per accession. Young leaves were harvested for DNA extraction, which was carried out on a single-plant basis using the DNeasy 96 Plant Kit and an MM300 Mixer Mill (Qiagen GmbH, Hilden, Germany). PCR was carried out in a 25 μl volume, containing 25 ng template DNA, 10 pmol of each primer, 20 μM dNTPs, PCR buffer 1× (200 mM Tris–HCl, pH 8.4, 500 mM KCl), 50 mM $MgCl_2$ and 1 U *Taq* DNA polymerase (Invitrogen).

The genotyping of the 401 *P. vulgaris* individuals (94 from the Americas and 307 from Europe) was carried out with six universal chloroplast microsatellite primer pairs (Chung and Staub 2003), with a Perkin-Elmer 9700 thermocycler (Applied Biosystems), using the relevant parameters for the specific primer pairs. For ccSSR2, ccSSR9, ccSSR16, ccSSR19 and ccSSR20, we set the following parameters: 5 min at 94°C; 35 cycles of 1 min at 94°C, 1 min at 50°C, 1 min at 72°C; 7 min at 72°C, 4°C ∞, ramping: 75%. For ccSSR11, we set the following parameters: 5 min at 94°C; 35 cycles of 1 min at 94°C, 1 min at 43°C, 1 min at 72°C; 7 min at 72°C, 4°C ∞. Touch down 53–45°C, 1°C every 2 cycles, ramping: 75%.

DNA fragments were separated in 6% 8 M denaturing acrylamide:bisacrylamide (19:1) gels that were run at 70 W for 3 h in a vertical cell (Biorad Laboratories, Milan, Italy), and the fragments were visualized using the silver staining method (Bassam et al. 1991).

Analysis of the two unlinked nuclear loci (phaseolin and *Pv-shatterproof1* locus) and morphological traits

These analyses were performed to allow better interpretation of the patterns of variation observed in *P. vulgaris* for the cpSSRs. In particular, we investigated the variations in the phaseolin patterns that are broadly used to identify the gene pools of *P. vulgaris,* along with new markers developed by Nanni et al. (submitted) from a genomic sequence of 1,200 bp located on chromosome 6 of *P. vulgaris*, which is highly similar to the SHATTERPROOF (*SHP1*) gene of *Arabidopsis thaliana*. Nanni et al. (submitted) identified the strict relationship between each haplotype obtained using a combination of two STS (sequence-tagged sites) markers and the gene pool of origin by sequencing 91 representative wild and domesticated individuals from their centers of origins.

For the phaseolin protein patterns, we analyzed 369 individuals (89 from the Americas and 279 from Europe) of *P. vulgaris* (Supplementary Materials, Tables A1, A2). One seed per accession was analyzed, using the half without the embryo. The extractions were performed according to Romero et al. (1975) and Sun and Hall (1975) (see Logozzo et al. 2007, for modifications). Protein extracts were analyzed by SDS-PAGE (Logozzo et al. 2007).

The *Pv-shatterproof1* variation was analyzed for 354 individuals (93 from the Americas and 261 from Europe) of *P. vulgaris* (Supplementary Materials, Tables A1, A2) and the PCR products obtained were separated by electrophoresis and visualized by the *Genomyx* Sc system (Beckman, Palo Alto, CA, USA).

Data were also collected for coat pattern and seed size, color and shape (according to the Biodiversity International) for 260 European individuals (Logozzo et al. 2007), and the differences among classes (Andean, Mesoamerican and hybrids) were tested using the Tukey–Kramer test (JMP, SAS Institute Inc. 2007). To summarize the information related to coat pattern and seed color and shape, the variables were subjected to principal coordinate analyses (PCOOA) on morphotypes using the NTSYS-pc 2.02i program (Rohlf 2000).

Statistical analysis

The gene diversity for the cpSSRs and the *Pv-shatterproof1* locus were estimated over loci using Nei's unbiased gene diversity (Nei 1978). Multilocus diversity was estimated using the Shannon diversity index (Shannon and Weaver 1949). The observed number of alleles in a sample is highly dependent on the sample size. To avoid this problem, El Mousadik and Petit (1996) suggested the adaption of the rarefaction index of Hurlbert (1971) to population genetics (see also Petit et al. 1998). We calculated this statistic (R_s) with the FSTAT software, version 2.9.3.2 (Goudet 2002).

To detect the relationships among all of the cpSSR individuals analyzed, we performed PCOOA, using NTSYS-pc, version 2.02i (Rohlf 2000), using the covariance matrix. The genetic divergences between groups of individuals (i.e. continents, gene pools between continents) were estimated calculating the F_{ST} (Weir and Cockerham 1984) statistics, using FSTAT (Goudet 2002). The significance was assessed by randomizing genotypes among samples (10,000 randomizations) not assuming random mating within samples, using the log likelihood G (Goudet et al. 1996).

To measure the bottleneck that occurred during the distribution of the beans from the Americas to Europe, we estimated the relative deficit of a given statistic (number of alleles, number of rare alleles, allelic richness, gene diversity and multilocus diversity) in Europe versus America, with the method defined by Vigouroux et al. (2005) in *Zea mays*: to detect the deficit for gene diversity (GD), we used the parameter $\Delta GD = 1 - (H_E/H_A)$, where H_E and H_A are the genetic diversity in Europe and America, respectively, and $H_A > H_E$. If $H_E > H_A$, we calculated this parameter as $\Delta GD = -[1 - (H_A/H_E)]$. The ΔGD varied between -1 and 1: when ΔGD is positive, the diversity is higher in America, and when negative, the reverse. Because of the different size of the European sample, to test the loss of total alleles, heterozygosity and multilocus diversity in the European sample, we resampled 200 samples with a size equal to the sample from the Americas, following the method of Luikart et al. (1998a, b). We then calculated different statistics across the 200 European samples. This value was then compared with the value of the samples from the Americas, to calculate the χ^2 statistic. However, the χ^2 statistic cannot follow a χ^2 distribution, because the same loci are measured for both the pre-bottleneck and post-bottleneck samples; hence, the data are not independent. Furthermore, the cell count expected values were often too small to justify the large sample χ^2 approximation. Consequently, we estimated the distribution of χ^2 test statistic for samples from pre-bottlenecked populations (i.e. we tested for a bottleneck effect in two samples from the same pre-bottleneck allele frequency distribution). The distribution was obtained by re-sampling 100 times. The *P* value of the test is then taken as the proportion of the randomized dataset giving a larger χ^2 than the observed χ^2.

Results

The Mesoamerican and the Andean gene pools in Europe

Overall, we identified 101 cpSSR haplotypes out of 401 individuals of *P. vulgaris* from the Americas and Europe that we analyzed. Figure 1 summarizes their relationships as depicted according to PCOOA, and also shows the frequency and geographical origin of each haplotype.

A strong population genetic structure for cpSSRs was seen, as the two coordinates cumulatively explained 80.9% of the total cpSSR genetic variance. Moreover, axis 1 explains quite a high proportion of the total genetic variance (67.2%) and splits all of the haplotypes into two clearly defined groups (Groups 1 and 2), while axis 2 explains only 13.7% of the total variance.

From now on, we here classify Groups 1 and 2 as representative of the Andean and Mesoamerican cpSSR gene

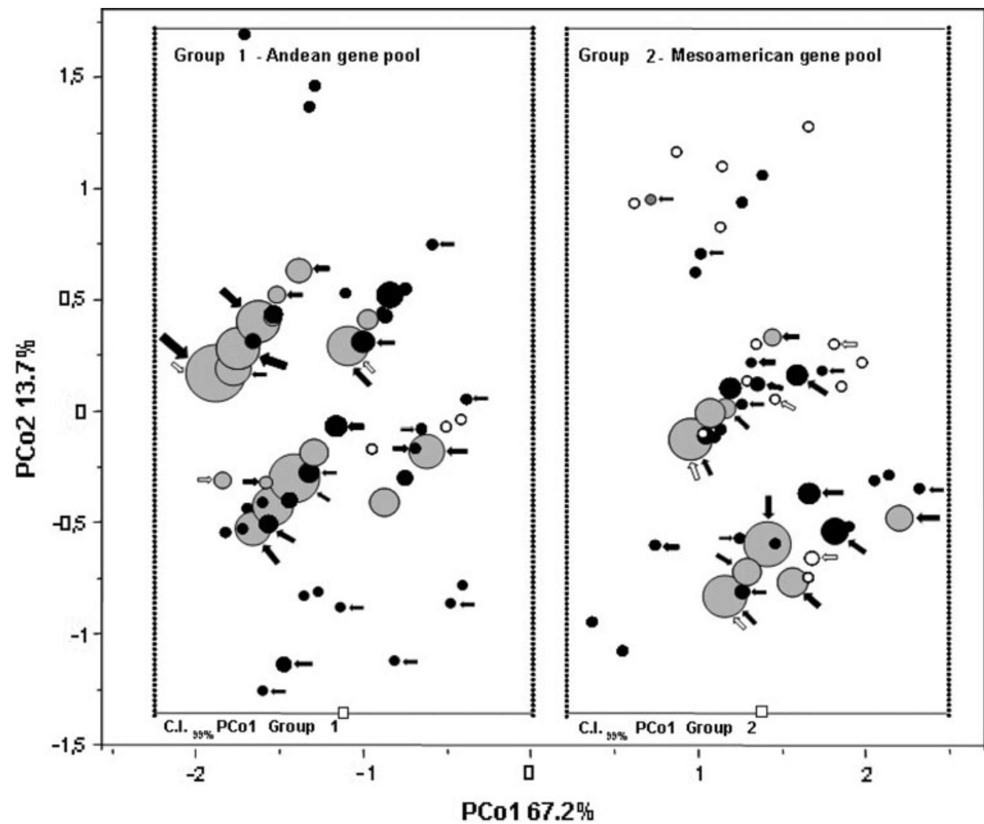

Fig. 1 Principal coordinate analysis of all of the cpSSR haplotypes from the Americas and Europe. The *circle size* is proportional to the haplotype frequency (min size, one individual; max size, 32 individuals). *Open circles* private haplotypes from the Americas; *filled black circles* European private haplotypes, *filled gray circles* shared haplotypes between Americas and Europe. Confidence interval (CI) for PCo1 as obtained for the two groups separately and considering only haplotypes from the Americas. *Arrows* haplotypes with at least one putative hybrid individual (*black arrows* hybrid individuals are European, *open arrows* hybrid individuals are from the Americas) (for the definition of the hybrid please refer to "Hybridization between Mesoamerican and Andean gene pool"). The size of each *arrow* is proportional to the number of hybrids (min size, 1 hybrid; max size, 9 hybrids)

pools of *P. vulgaris*, respectively. Several lines of evidence support this interpretation (see Table 1):

- first, almost all of the individuals from the Americas with typical Andean phaseolin types (C or T, 98%) and *Pv-shatterproof1* Andean-specific haplotypes (93%) are included in Group 1, while the vast majority of the individuals from the Americas with the typical Mesoamerican phaseolin type (S, 90%) and *Pv-shatterproof1* Mesoamerican-specific haplotypes (92%) belong to Group 2. The two groups from the Americas are different in their proportions of phaseolin (Pearson $\chi^2 = 68.2$, $df = 1$, $P < 0.0001$) and *Pv-shatterproof1* (Pearson $\chi^2 = 67.0$, $df = 1$, $P < 0.0001$) types.
- Second, within the European germplasm, the distribution of the phaseolin types across the two cpSSR groups parallel that observed for the individuals from the Americas. Indeed, within Group 1, the European individuals with phaseolin types 'C' or 'T' (Andean gene pool) are clearly predominant (cumulative frequency 86%). Conversely, a prevalence of individuals of phaseolin type 'S' (Mesoamerican gene pool) was seen (64% of the total) within Group 2. This difference in the proportions of the phaseolin types between the

Table 1 Distribution of the phaseolin types and of *Pv-shatterproof1* gene-pool-specific haplotypes between the two cpSSR groups (Groups 1 and 2) identified by principal coordinate analysis

	America						Europe					
	Phaseolin type			*Pv-shatterproof1* gene-pool-specific haplotype			Phaseolin type			*Pv-shatterproof1* gene-pool-specific haplotype		
	C	T	S	Andean	Mesoamerican	Sample size	C	T	S	Andean	Mesoamerican	Sample size
Distribution (%)												
Group 1	60	38	2	93	7	43/44	39	47	14	74	26	188/172
Group 2	4	6	90	8	92	48/49	27	9	64	25	75	92/90

Sample size is given both for phaseolin and *Pv-shatterproof1* analysis. Data are shown for individuals from the Americas and Europe separately

two European groups is highly significant (Pearson $\chi^2 = 68.9$, $df = 1$, $P < 0.0001$). Moreover, within Group 1, 74% of the European individuals have a *Pv-shatterproof1* haplotype that is specific to the Andean gene pool, according to the independent classification of Nanni et al. (submitted). In contrast, within Group 2, 75% of the European individuals have a *Pv-shatterproof1* haplotype that is specific to the Mesoamerican gene pool. In addition, this difference between these two groups is significant (Pearson $\chi^2 = 54.7$, $df = 1$, $P < 0.0001$).

Figure 2 shows the distribution of cpSSRs for the Andean and Mesoamerican gene pools in Europe. It is evident that the two gene pools co-occurred in all of the European regions considered. However, the majority (67%) of European individuals has an Andean cpSSR haplotype. The prevalence of Andean types was seen in all regions, although the incidence of the Andean gene pool is particularly high for the Iberian and Italian peninsulas and for central-northern Europe, while in eastern and south-eastern Europe, the Mesoamerican cpSSR types tend to increase in frequency. The picture obtained by phaseolin analysis is overall relatively close to that depicted by the cpSSR markers. For *Pv-shatterproof1*, the proportion of Andean types in Europe is lower (56%) than for the cpSSR and phaseolin analyses. Moreover, in eastern and south-eastern Europe, the Mesoamerican gene pool was at an equal or higher frequency than the Andean one (Fig. 2).

Comparing chloroplast genetic diversity in Europe and America

To compare the levels of chloroplast diversity between the germplasm from Europe and the Americas, several summary statistics were calculated, both overall and for each gene pool separately (Table 2). A total of 24 cpSSR alleles were identified (Table 2). A similar number of alleles were detected in the two continents even if, in Europe, there were two more alleles than in America (23 vs. 21, respectively; Table 2) ($\Delta n_a = -0.09$). A similar number of rare alleles ($n_{a<0.1}$) was detected in America and Europe (5 vs. 4; $\Delta n_{a<0.1} = 0.20$). Three alleles were exclusive to Europe, and one allele was exclusive to America. Thus, Europe and America shared a large number of alleles (20 out of 24 alleles detected, 83%).

The allelic richness in Europe ($R_s = 22.0$) was slightly higher than that in America ($R_s = 21.0$) ($\Delta R_s = -0.05$). A high proportion of exclusive haplotypes was observed both in Europe (67.9%) and America (38.6%). However, the 27 haplotypes shared between these continents account for 68.1% (209/307) of the individuals in Europe and

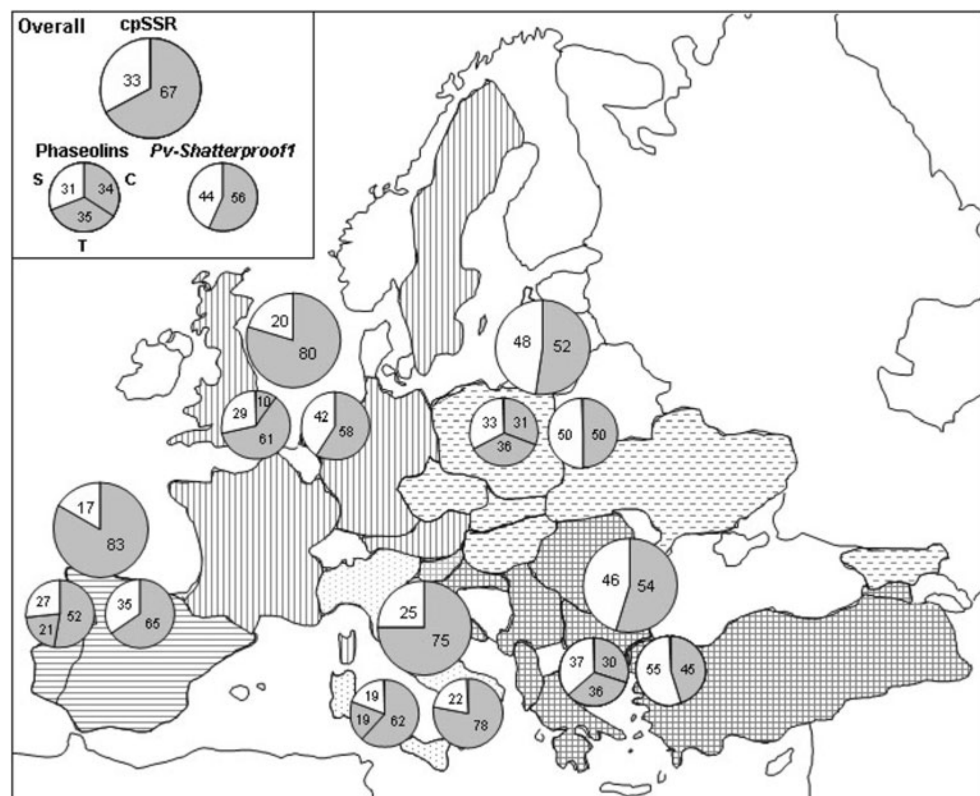

Fig. 2 Distribution of Andean (*gray*) and Mesoamerican (*white*) gene pools of *P. vulgaris* in five European regions (Iberian peninsula, Italy, central-northern Europe, eastern Europe, south-eastern Europe) based on cpSSR, phaseolin and *Pv-shatterproof1* polymorphism analyses. For the phaseolin pie charts (*bottom left*), the *gray area* is subdivided into the '*C*' and '*T*' types (first and second segment, respectively, clockwise) [see Table 3 for the sample sizes for cpSSRs and *Pv-shatterproof1;* for the phaseolins, the sample sizes were: Iberian peninsula (52), Italy (31), central-northern Europe (49), eastern Europe (68), south-eastern Europe (79)]

Table 2 CpSSR genetic diversity summary statistics for the samples from the Americas and Europe of *P. vulgaris*

Accessions	Sample size	Number of alleles		R_s	Number of haplotypes		Hap./Ind.	H_E	I_{nor}
		n_a	$n_{a\leq 0.1}$		Total	Private			
America	94	21	5	(21.0)	44	17	0.47	0.46	0.76
Andean gene pool	44	15	3	[15.0]	20	3	0.45	0.32	0.72
Mesoamerican gene pool	50	18	6	{18.0}	24	14	0.48	0.34	0.72
Europe	307	23	4	(22.0)	84	57	0.27	0.45	0.68
Andean gene pool	206	18	3	[17.0]	51	33	0.25	0.36	0.63
Mesoamerican gene pool	101	20	6	{19.1}	35	24	0.35	0.34	0.67
Total	401	24	4	–	101	–	0.25	0.45	0.72

n_a number of alleles (total and rare), R_s allelic richness, number of total and private haplotypes; haplotype/individuals ratio, H_E Nei's (1978) gene diversity, I_{nor} normalized Shannon–Weaver index of haplotype diversity. Note: within the R_s columns, comparisons should be made between the values enclosed between the same bracket types

for 80.8% (75/94) in America. In contrast, with allelic richness, the gene and haplotype diversities were slightly higher for America ($H_E = 0.46$ and $I_{nor} = 0.76$, respectively) than for Europe ($H_E = 0.45$ and $I_{nor} = 0.68$, respectively) (Table 2). This leads to positive but small values for both ΔH_E (0.02) and ΔI_{nor} (0.11).

Moreover, when the two gene pools are considered separately, the differences in the number of rare alleles between these continents disappear ($\Delta n_{a<0.1} = 0.00$ for both the Andean and Mesoamerican gene pools). For the other four statistics, the variations between these continents were never significant.

The cpSSR genetic divergence between Europe and America

The cpSSR genetic differentiation between these continents was very low ($F_{ST} = 0.02$), but it was significant ($P = 0.004$). Moreover, when Europe and America are compared for each gene pool separately, the shift of allelic frequencies becomes negligible and statistically not significant for both the Andean ($F_{ST} = 0.01$, $P = 0.266$) and the Mesoamerican ($F_{ST} = 0.01$, $P = 0.153$) gene pools.

Distribution of genetic diversity in Europe

No strong differences in the cpSSR genetic diversity (as number of alleles, and haplotype and gene and genotypic diversities) were seen across the five different European regions (Table 3).

Pv-shatterproof1 polymorphism analysis confirmed the substantial uniformity of the European regions for the levels of diversity (Table 3). In this case, Italy showed the lowest H_E and I_{nor} values, while eastern Europe showed the highest.

Hybridization between the Mesoamerican and Andean gene pools in Europe

When combining the information provided by both chloroplast (SSR) and nuclear (phaseolin and *Pv-shatterproof1*) markers, we have coded the genotypes of each individual with a three-letter string using the notation X[XX]. The first letter refers to the result of the chloroplast analysis, and the second and the third letters refer to the outcome of the nuclear analyses. For each of the three analyses, X = A if an individual is attributed to the Andean gene pool, or

Table 3 Diversity in the five regions of the European continent based on cpSSR and *Pv-shatterproof1* analyses

	cpSSR					*Pv-shatterproof1*				
	Sample size	n_a	Number of haplotypes	H_E	I_{nor}	Sample size	n_a	Number of haplotypes	H_E	I_{nor}
Iberian peninsula	53	19	35	0.44	0.85	53	6	4	0.52	0.26
Italy	32	17	20	0.45	0.82	32	7	4	0.41	0.22
Central-northern Europe	74	19	40	0.44	0.80	48	7	8	0.67	0.45
Eastern Europe	69	18	32	0.44	0.76	54	7	8	0.70	0.45
South-eastern Europe	79	19	39	0.45	0.80	77	7	8	0.63	0.35

n_a number of alleles, H_E gene diversity, I_{nor} genotypic diversity. Note: Iberian peninsula: Spain and Portugal; central-northern Europe: Austria, France, The Netherlands, Germany, England, Sweden; eastern Europe: Czech Republic, Poland, Slovakia, Georgia, Russia, Ukraine; south-eastern Europe: Albania, Croatia, Slovenia, Yugoslavia, Romania, Hungary, Bulgaria, Greece, Turkey

X = M if it is attributed to the Mesoamerican gene pool. Using this notation, which does not consider the occurrence of heterozygosity (i.e. no heterozygous genotype was found in our study), we have categorized all of the individuals into six categories. Thus, the two categories that are coded as A[AA] and M[MM] contain individuals that are attributed to the same gene pool by all three of these marker types, as either Andean (119 individuals) or Mesoamerican (55 individuals), respectively (Table 4).

The remaining four categories identify different putative hybrid types that probably have different degrees of introgression between the two gene pools. In particular, these are

A[MM] = 22 individuals with a maternal Andean chlorotype and a paternal nuclear DNA of Mesoamerican origin, both for the phaseolin and *Pv-shatterproof1* analyses.

A[AM or MA] = 28 individuals with Andean chlorotype and paternal nuclear DNA attributed to both the Mesoamerican and Andean gene pools (of which three are due to discordance between the chloroplast and phaseolin analyses [MA], and 25 to that between the chloroplast and *Pv-shatterproof1* analyses [AM]).

M[AM or MA] = 15 individuals with Mesoamerican chlorotype and paternal nuclear DNA attributed to both the Mesoamerican and Andean pools (of which 13 are due to discordance between the chloroplast and phaseolin analyses [AM], and two to that between the chloroplast and *Pv-shatterproof1* analyses [MA]).

M[AA] = 21 individuals with Mesoamerican chlorotype and paternal nuclear DNA that was always attributed to the Andean gene pools.

These four classes of putative hybrids are relatively similar in abundance. Overall, among the 260 European individuals for which the analyses with three different markers systems were available, 86 were putative hybrids, which is a relatively high frequency (33.1%). Indeed, this is almost fourfold higher than that seen in the Americas (8.8%), a difference between the two continents that is statistically significant (Pearson $\chi^2 = 19.9$, $df = 1$, $P < 0.0001$).

Considering the distributions of these hybrids, two interesting additional results can be noted. First, as shown in Fig. 1, the putative hybrids do not have any specific chlorotypes, i.e. hybridization is independent of the specific chloroplast haplotype carried by the mother plant. Secondly, in Europe, the hybrids are present in all of the regions considered, although the distribution is not homogeneous across this continent (Pearson $\chi^2 = 17.5$, $df = 4$, $P = 0.0015$) (Table 4). This is essentially true also when each of the four classes of hybrids are considered separately (Table 4), with the M[AM or MA] type completely absent in the Iberian peninsula and in Italy.

We also searched for validation of hybrid identification from the evaluation of seed traits. We tested the six genotypic categories by ANOVA, for differences in their seed traits: namely, seed size, shape and coat characteristics (Figs. 3, 4).

The six genotypic classes explained 24.9% of the variance in seed size ($F = 16.8$, $df = 5$, $P < 0.0001$). Seed size (which has a nuclear genetic bases) clearly shows a decreasing trend in passing from the [AA] individuals with an Andean nucleus, to the [AM] or [MA] individuals with a hybrid nucleus, to the [MM] individuals with a Mesoamerican nucleus; however, there is only a small effect of the chlorotype (Fig. 3a). In particular, individuals belonging to the Andean A[AA] and the Mesoamerican M[MM] classes are significantly different according to average seed size (0.75 ± 0.02 vs. 0.50 ± 0.03 cm^3). The M[AA] and A[MM] hybrids are also different from each other (0.83 ± 0.04 vs. 0.51 ± 0.04 cm^3), and are not statistically different from the A[AA] and M[MM] classes, respectively. The two categories containing individuals with a hybrid nucleus have intermediate seed sizes, as A[AM] (0.57 ± 0.04 cm^3) and M[AM] (0.60 ± 0.05 cm^3), and they are not statistically different from each other.

The average difference (0.10 cm^3) between the European-Andean cpSSR gene pool (Group 1: 0.70 cm^3, se = 0.02) and the European-Mesoamerican gene pool (Group 2: 0.60 cm^3, se = 0.02) is highly significant ($F = 11.5$; $df = 1$; $P = 0.0008$) (Fig. 3b). However, after removing the hybrids, the difference between the European 'pure' Andean A[AA] and 'pure' Mesoamerican M[MM]

Table 4 Distribution of the four hybrids classes of putative hybrids (from chloroplast and nuclear polymorphism comparisons) across the five European regions

	A[MM]	A[AM or MA][a]	M[AM or MA][b]	M[AA]	Total
Iberian peninsula	7 (32%)	3 (11%)	0	1 (5%)	11
Italy	1 (4%)	1 (4%)	0	3 (14%)	5
Central-northern Europe	6 (27%)	13 (46%)	3 (20%)	4 (19%)	26
Eastern Europe	3 (23%)	2 (7%)	5 (47%)	6 (29%)	16
South-eastern Europe	5 (14%)	9 (32%)	7 (33%)	7 (33%)	28
Total	22	28	15	21	86

[a] [AM] = 25 individuals; [MA] = 3 individuals

[b] [AM] = 2 individuals; [MA] = 13 individuals

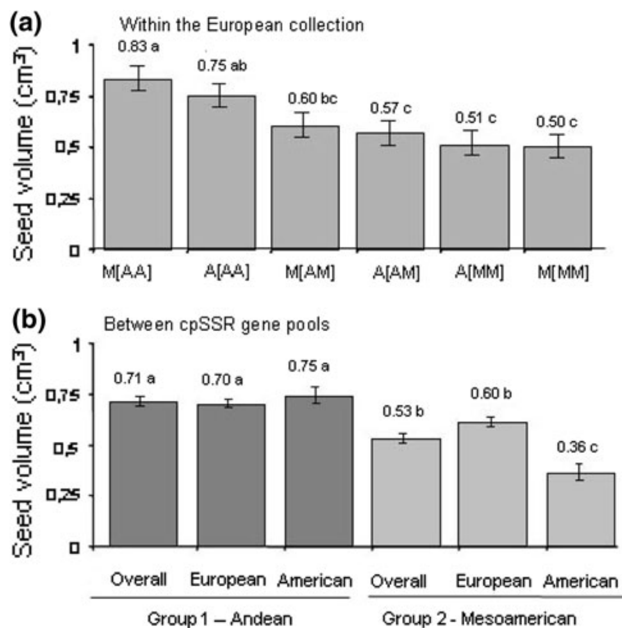

Fig. 3 Seed sizes (as volume in cm^3) of European individuals by genotypic class (**a**), and by cpSSR gene pools (Groups 1 and 2) overall, and for the European and American collections as a whole (**b**). The classes are defined based on both chloroplast and nuclear polymorphism analyses. Differences among the classes were tested using Tukey–Kramer test. In **a**, the classes that do not share the *same letter* are significantly different ($P \leq 0.05$). In **b**, the means with the same *bracket type* should be compared. Note in **b** the average seed is presented including the hybrids. However, if the hybrids are excluded: Group 1, Andean American = 0.70 ± 0.04, Group 2, Mesoamerican American = 0.35 ± 0.04; $F = 41.1$, $df = 1$, $P < 0.0001$, Group 1 Andean European = 0.75 ± 0.02, Group 2, Mesoamerican European = 0.50 ± 0.03; $F = 52.8$; $df = 1$, $P < 0.0001$)

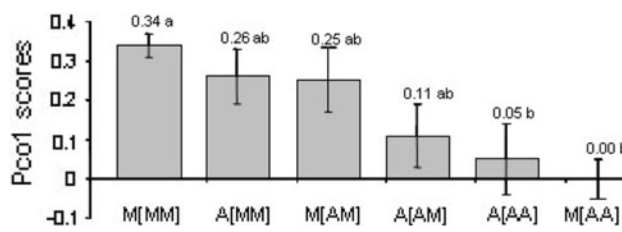

Fig. 4 Differences among the six genotypic classes for coat pattern (absent, bicolor, rhomboid, spotted bicolor, striped and other), darker seed color (black, brown, chlorophyll, gray, maroon, pink, whitish, yellow, pure white and purple) and lighter seed color (absent, black, brown, maroon, pure white and whitish, purple) and seed shape (cuboid, oval, round, truncate fastigiated and rhomboid). To summarize the information, the variables were subjected to principal coordinate analyses on morphotypes. The average of each class was obtained by averaging the PCo1 scores of each individual

increases up to 0.25 cm^3 ($F = 54.6$, $df = 1$, $P < 0.0001$) and although this is lower, it tends to be more similar to that seen between the Andean and Mesoamerican gene pools in America (0.39 cm^3; $F = 65.8$, $df = 1$, $P < 0.0001$).

The histogram presented in Fig. 4 compares the average values of the first principal coordinate (PCo1), which summarizes seed shape and seed coat characteristics, for the six genotypic classes. The means are clearly separated ($F = 6.45$, $df = 5$, $P = 0.0001$), and the genotypic classes explain 22.1% of the PCo1. The ranking of the six categories based on their average PCo1 scores is quite close to that seen for seed size (Fig. 4). However, in this case, the differences among the classes are less clear cut, with a significant contrast between the A[AA] and M[MM] genotypes, and with the hybrids not significantly different from each other.

It is also interesting to consider the variation of each single seed trait. Indeed, A[AA] genotypes are prevalently maroon (62.1%), while the M[MM] genotypes are prevalently pure white (60.3%). Moreover, almost all of the M[MM] genotypes are monochrome (94.8%), while 48.5% of the A[AA] genotypes have a colored pattern.

Differences in seed shape between the two gene pools were also observed: the A[AA] genotypes are 34.1% kidney, 31.8% oval, 23.5% cuboid, 6.8% round and 3.8% truncate fastigiated; in contrast, the M[MM] genotypes are 53.5% cuboid, 22.4% kidney, 13.8% oval, 10.3% truncate fastigiated and 0% round.

Interestingly, the pool of all of the hybrid types has an intermediate composition: 29.2% of the genotypes are maroon, and 28.1% are white; also 19.9% of these genotypes showed a colored pattern and 42.7% are cuboids, 25.8% oval, 22.5% kidney, 6.7% round and 2.2% truncate fastigiated.

Finally, the European individuals for which seed traits were available (278) could be classified into 70 different morphotypes (=combinations of single traits). Six of these have a frequency ≥5% (maximum frequency of 10%) and account for 44.6% of the individuals analyzed. The morphotype diversity of the hybrid pool is the highest. Indeed, the ratio of haplotypes to individuals was 0.29 for M[MM], 0.37 for A[AA] and 0.43 for the hybrids. The same pattern was observed for I_{nor} (0.57, 0.66 and 0.75, respectively).

Estimating the frequency of the 'hidden' hybrids

Through the comparison between the cytoplasmic and nuclear markers, we revealed direct evidence of hybridization between the gene pools. However, several individuals that were observed as 'pure' (Andean or Mesoamerican for all of the molecular markers used) might have derived because of recombination from crosses between parents belonging to the two different gene pools. Thus, the number of individuals that had a progenitor that originated from hybridization between gene pools is expected to be larger than the frequency of the individuals that show direct evidence of hybridization. Most of these

Table 5 Distribution of individuals among the 'pure' and hybrid classes

	Observed			Estimated[a]		
	Pure[b]	Hybrids[c]	Total	Pure[b]	Hybrids[c]	Total
Individuals with Andean chloroplast	119	50	169	102	67	169
Individuals with Mesoamerican chloroplast	55	36	91	43	48	91
Total	174	86	260	145	115	260

[a] Based on maximum likelihood approach as
$L = \{N[a + (1-a)/4]\}^p \times \{3N(1-a)/4\}^q$, setting $\frac{dL}{a} = p\ln[a + (1-a)/4] + q\ln(1-a)/4 + \text{costant} = 0$, and solving for a, where N is the sample size of the group of individuals with Andean (or Mesoamerican) chlorotype, a is the estimated fraction of individuals with Andean (or Mesoamerican) chloroplast that is 'pure' (i.e. A[AA] or M[MM] individuals that do not derive from hybridization and recombination), p is the observed number of pure Andean (or Mesoamerican) individuals, q is the observed number of hybrids with Andean (or Mesoamerican) chloroplast

[b] A[AA] or M[MM] individuals

[c] A[MM], A[AM], M[AA], M[AM] individuals

individuals could be identified if more nuclear markers that discriminate between the two gene pools were available. However, using a maximum likelihood approach (see also notes for Table 5), we estimated the frequency of these 'hidden' hybrids, by assuming an absence of selection and effects due to nuclear–cytoplasm interactions. Thus, we estimated that 60.6% of individuals that have an Andean chloroplast are 'pure' (i.e. were never involved in hybridization events), as also for 47.0% of individuals that have a Mesoamerican chloroplast (Table 5). Thus, the observed number of hybrids (86) should be increased by the addition of 29 expected 'hidden' hybrids (17 Andean × Mesoamerican, and 12 Mesoamerican × Andean) (Table 5). Thus, the overall frequency of individuals that have derived from hybridization between the Andean and Mesoamerican gene pools is 33.1% (observed) + 11.2% (estimated 'hidden'), or 44.3% of the whole sample (Table 5).

Discussion

To study the effects of the introduction of *P. vulgaris* into and across Europe, we have compared the genetic variations at chloroplast microsatellite loci (cpSSR) in two germplasm collections, one from the 'New World' and the other from the 'Old World'. Moreover, the results obtained using cpSSRs were complemented by those collected using two types of nuclear markers: phaseolin seed-storage proteins and an STS (sequence-tagged sites) PCR-based assay for length polymorphism at the *Pv-shatterproof1* locus (Nanni et al. submitted). Indeed, the comparison between these uniparentally inherited cytoplasmic markers (cpSSRs) and bi-parental nuclear markers allows a more complete description of the level and structure of the genetic diversity, and, most importantly, the high-precision and resolution in the identification of hybridization events (Elbert and Peakall 2009).

The use of phaseolin types as a marker is a well-established tool and it has been used to conduct several informative analyses on different geographic scales for the European landraces of *P. vulgaris* (see Papa et al. 2006, for a review). The STS targeting the polymorphism at the *Pv-shatterproof1* locus has proven to be a very useful tool to track the gene pool structure in *P. vulgaris* (Nanni et al. submitted), and we have used it in the present study to assess the genetic variation of the European *P. vulgaris* germplasm. CpSSR markers are a powerful tool to study plant genetic variations, given the high cpSSR polymorphism and their ability to mark bottleneck events (Provan et al. 1999b). Moreover, it has been shown that cpSSRs can significantly improve the power for deciphering the population genetics structure of *Phaseolus* spp. (Angioi et al. 2009b) and of local populations of *P. vulgaris* (Sicard et al. 2005; Angioi et al. 2009a). To our knowledge, the present study represents by far the largest analysis of cytoplasmic diversity within the landrace gene pool of *P. vulgaris.*

Using this composite 'marker system', we were able not only to estimate the level of hybridization between gene pools in Europe, but also the "direction" of the introgression events. Overall, we have investigated three aspects of the genetic make-up of the European germplasm of *P. vulgaris* and have determined their relevance for both the evolution and the breeding of this crop.

First: the contribution of the Andean and Mesoamerican gene pools to the European germplasm

Based on cpSSRs, we show here that the European common bean arose from both of the gene pools from the Americas, but that the majority (67%) of the European accessions analyzed were attributable to the Andean gene pool. This is quite similar to our phaseolin analysis here (69%), although it is higher than that obtained here by our *Pv-shatterproof1* analysis (56%). The lack of strict

concordance among these three marker systems, which implies recombination subsequent to hybridization, might be due to drift or to a departure from neutrality for these markers (either directly, or indirectly through linkage).

Other studies based on phaseolin analyses have revealed that in the European common bean, the Andean gene pool is always prevalent, with percentages between 66 (Lioi 1989) and 76% (Gepts and Bliss 1988; Logozzo et al. 2007). The slightly higher incidence of the Mesoamerican gene pool seen in the present study is probably due to the sampling strategy that was adopted to build our collection (see Logozzo et al. 2007 and "Results"). Moreover, we also show that the prevalence of the Andean gene pool is consistent across all of the main regions of Europe considered here. When regions within countries are considered, the prevalence of the Andean gene pool is often confirmed (Limongelli et al. 1996; Escribano et al. 1998; Piergiovanni et al. 2000a, b; Sicard et al. 2005; Angioi et al. 2009a). Overall, this suggests high gene flow among the different regions of Europe and/or homogenizing selection (anthropic or 'natural').

In contrast to Europe, a greater presence (75%) of the Mesoamerican gene pool was shown for a collection of 299 Chinese accessions (Zhang et al. 2008). Thus, different regions of the World appear to have different proportions of the two major gene pools of the common bean, which suggests that the distribution of the Andean and Mesoamerican gene pools outside the primary centers of diversity need to be determined empirically.

Second: diversity of gene pools in Europe and divergence between Europe and America

We also show here that the Andean and the Mesoamerican gene pools in Europe have similar levels of polymorphism for cpSSRs ($H = 0.36$ and 0.34, respectively). Despite the differences in the absolute values, in a broad study, Santalla et al. (2002) obtained similar results when working on Iberian landraces with isozyme markers ($H = 0.21$ and $H = 0.20$, respectively). Interestingly, in the center of origin, the Mesoamerican gene pool has a higher level of diversity than the Andean one as also found by several studies (for review, see Papa et al. 2006; Kwak and Gepts 2009), even if the intensity of this difference depends on the mutation rate of the type of molecular marker used (Rossi et al. 2009; Nanni et al. submitted). Thus, the similar level of diversity between the two gene pools seen in Europe suggest that in this continent selection might have reduced the variation of the Mesoamerican germplasm, as was suggested by Logozzo et al. (2007) for seed size. Alternatively, the diversity of the Mesoamerican samples introduced into Europe might have already had a reduced diversity when compared with the Mesoamerican gene pool.

Our data indicate here that the genetic bottleneck that followed the introduction of *P. vulgaris* into Europe was very weak. Indeed, the variation between continents (Δ) was not statistically significant for all of the genetic diversity estimators used and also when the two gene pools were considered as a whole or separately.

Moreover, genetic differentiation between these two continents is very low ($F_{ST} = 0.02$). However, this already small shift in the allelic frequencies becomes negligible (and not significant) when these gene pools are considered singly ($F_{ST} = 0.01$), both for those Andean and those Mesoamerican. Thus, the genetic divergence between these two continents is mainly due to the different Andean: Mesoamerican gene pool ratio between the two collections.

Gepts (1999) suggested that crop expansion from the Americas to Europe resulted in a reduction in the diversity of the European common bean because of strong founder effects, adaptation to a new environment and consumer preferences. Isozyme loci have been used to characterize the domesticated common bean both from the Americas (Singh et al. 1991) and from the Iberian peninsula (Santalla et al. 2002). Recalculation of the diversity values using the eight isozymes that have loci in common between these two studies reveals that the bottleneck effect of the introduction of the common bean into Europe might not have been as strong (loss of diversity of about 30%) as was previously suspected (Papa et al. 2006). However, of note, this loss of genetic diversity (H) is much higher than that measured in the present study (about 0%), and also higher as compared to our (not significant) reduction in the cpSSR genotypic diversity (I_{nor}). This could be due to several reasons, such as the type of markers used (isozymes and cpSSRs) or sampling effects, i.e. in the study of Santalla et al. (2002), only the Iberian peninsula was sampled, rather than the entire European continent, as in the present study. However, it is also possible that the plastidial and nuclear genomes have undergone genetic bottlenecks of different intensities and/or with different effects of selection. As an example, Provan et al. (1999b) observed that an extreme cytoplasmic bottleneck in the modern European cultivated potato (*Solanum tuberosum*) was not reflected in decreased levels of nuclear diversity. When compared with this study of Provan et al. (1999b), we observed the reverse situation. However, if we assume that in the original gene pools from the Americas the total number of plastidial variants (alleles and genotypes) was lower than that for the nuclear genome, it is conceivable that the repeated and subsequent sampling of *P. vulgaris* individuals from the Americas to Europe during the last 500 years after the first travels of Columbus were sufficient to saturate (capture) the greater part of existing variation in the plastidial genome in the founding populations, but not that in the nuclear genome. If we assume that cpSSRs are neutral, or nearly

so, an alternative hypothesis could explain the reduction in diversity at the nuclear level as due to direct or indirect (hitchhiking) effects of selection for adaptation acting in the nuclear genome. In this regard, it is interesting to note that our results obtained here using the markers tagging *Pv-shatterproof1* showed a different trend when compared with the cpSSRs. Indeed, for the *Pv-shatterproof1* markers, the loss of diversity in the European germplasm was higher than for the cpSSRs also considering the gene pool separately (data not shown). Interestingly, *Pv-shatterproof1* is located in chromosome 6, and it is strictly linked to the flower color gene *V* (Nanni et al. submitted), where a QTL for flowering time is also present (Blair et al. 2006). Moreover, in using the results of neutrality tests (e.g. Tajima), Nanni et al. (submitted) suggested that this locus could have been the target of selection during the domestication of the common bean in Mesoamerica.

To the best of our knowledge, the introduction and dispersal of a crop from an American center of origin into the Old World has only been addressed in the case of *Zea mays*. This was done using nuclear RFLP molecular markers (Rebourg et al. 2003). Allelic richness and genetic diversity were both higher in America ($H = 0.587$) than in Europe ($H = 0.550$) ($\Delta H = 0.06$), and the number of private alleles was clearly lower in Europe (21) than in the Americas (102) ($\Delta n_a = 0.88$). We have also re-analyzed the data presented in Table 3 of Rebourg et al. (2003). For maize, the two continents were weakly differentiated ($F_{ST} = 0.03$), this differentiation is 1.5 fold higher to that calculated for *P. vulgaris* using cpSSRs at the continent level ($F_{ST} = 0.02$). Thus, the overall shift of the nuclear allelic frequencies for *Zea mays* appears slightly higher than that for cpSSRs in *P. vulgaris*. However, a much stronger loss of allelic richness is observed in *Zea mays*.

In conclusion, based on cpSSRs, it appears that the founding populations were highly representative of the diversity present in the gene pools from the Americas. This might be because there were several different introductions from the Americas, or because the attractiveness of various types of seed color and shape has favored the capture of different alleles and genotypes (Papa et al. 2006). However, it cannot be excluded a priori that heterogeneous selection for different uses and local adaptation to a wide range of environments and agronomic practices in Europe might also have counteracted the effects of drift and homogeneous selection for adaptation to European environmental conditions (Papa et al. 2006).

Third: hybridization between the Andean and Mesoamerican gene pools in Europe

A large percentage (44.2%) of individuals in the European sample can be regarded as having derived from progenitors that were hybrids between the two gene pools. Notably, this proportion is fivefold (and significantly) higher than that seen in the sample from the Americas (8.8%). It is important to note that in addition to the molecular results, the individuals identified as hybrids also show evidence of hybridization from the analysis of seed traits. Indeed, seed size and coat traits tend to vary with the level of introgression between the two gene pools, with relatively good agreement. In particular, in agreement with the expected nuclear determinism of seed size, we found that hybrids tend to have an intermediate seed size in comparison with the 'pure' Andean or Mesoamerican seeds, with the Andean × Mesoamerican smaller than the 'pure' Andean (from Europe or the Americas) and the Mesoamerican × Andean larger than the 'pure' Mesoamerican (European or American). Logozzo et al. (2007) first noted that European accessions with phaseolin 'S' (Mesoamerican) showed a significantly larger average seed size as compared to those of the same phaseolin class from America, while those with phaseolins 'T' and 'C' did not. This might also explain the larger frequency of Andean phaseolins in the Andean European materials (as detected by chloroplast SSRs; 86%) as compared to the frequency of the 'S' type in the European genotypes of Mesoamerican origin (64%), probably due to the selection for larger seed size, given that the phaseolin locus co-segregates with a QTL for seed size (Gepts et al. 1988). Moreover, González et al. (2009) analyzed the progenies from crosses within and between gene pools in the common bean. In particular, they observed that Andean × Andean crosses had an average seed volume of 0.88 cm^3 (recalculated from Table 3 of González et al. 2009). This is quite close to our volume for the M[AA] (0.83 cm^3) and A[AA] (0.75 cm^3) genotypes. In addition, the progenies from Mesoamerican × Andean crosses produced an average seed volume of 0.55 cm^3 (Table 3 of González et al. 2009). This is very close to the average volume of our hybrid classes, with 0.60 cm^3 for M[AM] and 0.57 cm^3 for A[AM].

To date, only a few studies with DNA markers have shown the occurrence of hybridization between gene pools in Europe (e.g. in the Marche region in central Italy; Sicard et al. 2005), and in Sardinia (Angioi et al. 2009a). Moreover, using isozymes, introgression was seen in the Iberian peninsula with 25% of putative recombinants (Santalla et al. 2002). These studies suggested that hybridization could be explained because landraces are cultivated in small gardens, in proximity to each other. Together with the exchange of seeds among neighboring farms, these conditions might have resulted in occasional outcrossing and gene flow. Outcrossing rates are usually below 10% (Park et al. 1996). In our analysis, no heterozygous genotype was found using the *Pv-shatterproof1* co-dominant markers. However, this is not necessarily related to the

level of outcrossing, because we mainly used accessions derived from gene banks that are likely to be subject to strong inbreeding during seed regeneration.

In 400–500 generations (i.e. the number of years of cultivation from the introduction of *P. vulgaris* into Europe, according to historical records), and assuming a neutral island model, the probability of an individual to be derived without hybridization between gene pools is expected to vary between q^{400} and q^{500} where $q = [1 -$ frequency of hybridization between the gene pools per year]. Our estimated frequency of non-hybrids of 55.8% (100 − 44.2%) will be obtained assuming a frequency of hybridization between gene pools per year ranging from 0.12% (500 generations) to 0.15% (400 generations) that are conceivable values for an inbreeder, such as *P. vulgaris*.

Overall, the existence of a relatively high frequency of inter-gene pool hybridization in Europe appears to have had a significant impact on the structure of the genetic and genotypic diversity in the nuclear genome. This is consistent with the breakdown of geographical isolation between the two gene pools in Europe.

Conclusions

Santalla et al. (2002) proposed Spain and the Iberian peninsula as a likely secondary center of diversification for *P. vulgaris*. This hypothesis is possible when the historic context is taken into consideration. Indeed, the introduction of the common bean into the Old World probably took place through the Iberian peninsula from where it spread to the rest of Europe and other parts of the World (Gepts and Bliss 1988).

Moreover, a high level of variation and the presence of frequent hybridization that led to new variants that have not been seen in the Americas have been documented (Santalla et al. 2002). This was further suggested by Ocampo et al. (2002). However, we show here that Spain alone is not a hot spot for either genetic diversity or hybridization events. Indeed, for cpSSRs, we show that the Iberian peninsula has high gene diversity and genotypic diversity, but this is similar (or often lower) when compared with the other European countries. This trend was also confirmed with the *Pv-shatterproof1* analysis, where the genetic diversity of the Iberian peninsula was often equal to or lower than the other groups. Moreover, hybridization between gene pools is a widespread phenomenon in Europe, and we also saw an uneven distribution across the different regions, with higher frequencies in Central Europe, and low levels in Italy and in the Iberian peninsula. Thus, we strongly suggest that the entire European continent can be regarded as a secondary diversification center for the *P. vulgaris* species. Moreover, even if our study is limited to the chloroplast genome, two nuclear loci and seed traits, it suggests that recombination has not been limited, as would be expected from the knowledge of the partial isolation between the gene pools. For these reasons, the European germplasm of the common bean appear to be of great importance for breeding, which often aims to recombine Andean and Mesoamerican traits (Johnson and Gepts 1999, 2002). Indeed, the European landraces described here (about 44% generated by spontaneous introgression) are of great value for plant breeding because with their selection by farmers, they have themselves overcome the difficulties of transferring traits (e.g., seed size, resistance to pathogens) between gene pools. For the same reasons, the European germplasm represents a perfect model to search for the signature of selection and for association mapping.

Acknowledgments S.A. Angioi and D. Rau made equal contributions to this study, and so should be considered joint first authors. We are grateful to A.H.D. Brown for his suggestions on the statistical analyses of this study and for critical reading of the manuscript. This study was supported by the Italian Government (MIUR) Grant no. # 2005071310, Project PRIN 2005 and constituted part of the PhD thesis of S.A. Angioi.

References

Acosta-Gallegos JA, Kelly JD, Gepts P (2007) Prebreeding in common bean and use of genetic diversity from wild germplasm. Crop Sci 47(S3):S44–S59

Angioi SA, Rau D, Rodriguez M, Logozzo G, Desiderio F, Papa R, Attene G (2009a) Nuclear and chloroplast microsatellite diversity in *Phaseolus vulgaris* L. from Sardinia (Italy). Mol Breed 23:413–429

Angioi SA, Desiderio F, Rau D, Bitocchi E, Attene G, Papa R (2009b) Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. Plant Biol 11:598–612

Arroyo-Garcia R, Lefort F, de Andrés F, Ibanez J, Borrego J, Jouve N, Cabello F, Martìnez-Zapater JM (2002) Chloroplast microsatellite polymorphisms in *Vitis* species. Genome 45:1142–1149

Bassam BJ, Caetano-Anolles G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80–83

Beebe S, Skroch PW, Tohme J, Duque MC, Pedraza F, Nienhuis J (2000) Structure of genetic diversity between common bean landraces of Middle American origin based on correspondence analysis of RAPD. Crop Sci 40:264–273

Beebe S, Rengifo J, Gaitan E, Duque MC, Tohme J (2001) Diversity and origin of Andean landraces of common bean. Crop Sci 41:854–862

Berglund-Brücher B, Brücher H (1976) The South American wild bean (*Phaseolus aborigenus* Burk.) as an ancestor of the common bean. Econ Bot 30:257–272

Blair MW, Iriarte G, Beebe S (2006) QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (*Phaseolus vulgaris* L.) cross. Theor Appl Genet 112:1149–1163

Chung S, Staub JE (2003) The development and evaluation of consensus chloroplast primer pairs that possess highly variable

sequence regions in a diverse array of plant taxa. Theor Appl Genet 107:757–767

Cuenca A, Escalante E, Pinero D (2003) Long-distance colonization, isolation by distance, and historical demography in a relictual Mexican pinyon pine (*Pinus nelsonii* Shaw) as revealed by paternally inherited genetic markers (cpSSRs). Mol Ecol 12:2087–2097

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theor Appl Genet 92:832–839

Elbert D, Peakall R (2009) Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour 9:673–690

Ennos RA, Sinclair WT, Hu XS, Langdon A (1999) Using organelle markers to elucidate the history, ecology and evolution of plant populations. In: Hollingsworth PM, Bateman RM, Gornall RJ (eds) Molecular systematics and plant evolution. Taylor & Francis, London, pp 1–19

Escribano MR, Santalla M, Casquero PA, De Ron AR (1998) Patterns of genetic diversity in landraces of common bean (*Phaseolus vulgaris* L.) from Galicia. Plant Breed 117:49–56

Gepts P (1998) Origin and evolution of common bean: past events and recent trends. Hort Sci 33:1121–1130

Gepts P (1999) Development of an integrated genetic linkage map in common bean (*Phaseolus vulgaris* L.) and its use. In: Singh S (ed) Bean breeding for the 21st century, 53–91. Kluwer, Dordrecht, pp 389–400

Gepts P, Bliss FA (1985) F1 hybrid weakness in the common bean: differential geographic origin suggests two gene pools in cultivated bean germplasm. J Hered 76(6):447–450

Gepts P, Bliss FA (1988) Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. II Europe and Africa. Econ Bot 42:86–104

Gepts P, Debouck DG (1991) Origin, domestication and evolution of common bean, *Phaseolus vulgaris*. In: Schoonhoven A, Voysest O (eds) Common beans: research for crop improvement. CAB International, Wallingford, UK, pp 4–54

Gepts P, Osborne TC, Rashka K, Bliss FA (1986) Electrophoretic analysis of phaseolin protein variability in wild forms and landraces of the common bean *Phaseolus vulgaris* L.: evidence for two centers of domestications. Econ Bot 40:451–468

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. Econ Bot 42:86–104

Gepts P, Papa R, Gonzales Mejia J, Acosta J, Delgado Salinas A (1999) Human effects on *Phaseolus* adaptation during and after domestication. In: van Raamsdonk L, den Njis J (eds) Proceeding VIIth international IOPB symposium. Amsterdam, 10–15 August

González AM, Rodiño AP, Santalla M, De Ron AM (2009) Genetics of intra-gene pool and inter-gene pool hybridization for seed traits in common bean (*Phaseolus vulgaris* L.) germplasm from Europe. Field Crop Res 112:66–76

Goudet J (2002) FSTAT: a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). http://www.unil.ch/izea/softwares/fstat.html

Goudet J, Raymond M, de-Meeus T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144:1933–1940

Hale ML, Borland AM, Gustafsson MH, Wolff K (2004) Causes of size homoplasy among chloroplast microsatellites in closely related *Clusia* species. J Mol Evol 58:182–190

Hurlbert SH (1971) The non-concept of species diversity: a critique and alternative parameters. Ecology 52:577–586

Ishii T, McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. Theor Appl Genet 100:1257–1266

Johnson WC, Gepts P (1999) Segregation for performance in recombinant inbred populations resulting from inter-gene pool crosses of common bean (*Phaseolus vulgaris* L.). Euphytica 106:45–56

Johnson WC, Gepts P (2002) The role of epistasis in controlling seed yield and other agronomic traits in an Andean × Mesoamerican cross of common bean (*Phaseolus vulgaris* L.). Euphytica 125:69–79

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

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

Limongelli G, Laghetti G, Perrino P, Piergiovanni AR (1996) Variation of seed storage protein in landraces of common bean (*Phaseolus vulgaris* L.) from Basilicata, southern Italy. Plant Breed 119:513–516

Lioi L (1989) Geographical variation of phaseolin patterns in an old world collection of *Phaseolus vulgaris*. Seed Sci Technol 17:317–324

Logozzo G, Donnoli R, Macaluso L, Papa R, Knupffer H, Spagnoletti Zeuli PL (2007) Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. Genet Resour Crop Evol 54:1763–1779

Luikart G, Sherwin WB, Steele BM, Allendorf FW (1998a) Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. Mol Ecol 7:963–974

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998b) Distortion of allele frequency distributions provide a test for recent population bottlenecks. J Hered 89:238–247

McClean PE, Lee RK, Otto C, Gepts P, Bassett MJ (2002) Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). J Hered 93:148–152

McClean PE, Lee RK, Miklas PN (2004) Sequence diversity analysis of dihydroflavonol 4-reductase intron 1 in common bean. Genome 47:266–280

Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. Euphytica 147:106–131

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

Nodari RO, Tsai SM, Gilbertson RL, Gepts P (1993) Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor Appl Genet 85:513–520

Ocampo CH, Martin JP, Ortiz JM, Sanchez-Yelamo MD, Toro O, Debouck D (2002) Possible origins of common bean (*Phaseolus vulgaris* L.) cultivated in Spain in relation to the wild genetic pools of Americas. Annu Rep Bean Improv Coop 45:236–237

Ortwin-Sauer C (1966) The early Spanish man. University of California Press, Berkeley, pp 51–298

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. Theor Appl Genet 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) New approaches to the origins, evolution and conservation of crops. Darwin's harvest. Columbia University Press, USA

Paredes OM, Gepts P (1995) Extensive introgression of Middle American germplasm into Chilean common bean cultivars. Genet Resour Crop Evol 42:29–41

Park S, Michaels T, Myers J, Hunt D, Stewart-Williams K (1996) Outcrossing rates of common beans grow in Ontario and Idaho. Annu Rep Bean Improv Coop 39:90–91

Petit R, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12:844–855

Petit RJ, Dumilin J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14:689–701

Piergiovanni AR, Cerbino D, Brandi M (2000a) The common bean populations from Basilicata (southern Italy). An evaluation of their variation. Genet Resour Crop Evol 47:489–495

Piergiovanni AR, Taranto G, Pignone D (2000b) Diversity among common bean populations from the Abruzzo region (central Italy): a preliminary inquiry. Genet Resour Crop Evol 47:467–470

Powell W, Morgante M, McDevitt R, Vendramin G, Rafalski J (1995) Polymorphic simple-sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–7763

Provan J, Russell JR, Booth A, Powell W (1999a) Polymorphic chloroplast simple-sequence repeat primers for systematic and population studies in the genus *Hordeum*. Mol Ecol 8:505–511

Provan J, Powell W, Dewar H, Bryan G, Machray GC, Waugh R (1999b) An extreme cytoplasmic bottleneck in the modern European cultivated potato (*Solanum tuberosum*) is not reflected in decreased levels of nuclear diversity. Proc Biol Sci 266:633–639

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903

Rodiño AP, Santalla M, Montero I, Casquero PA, De Ron AM (2001) Diversity in common bean (*Phaseolus vulgaris* L.) germplasm from Portugal. Genet Resour Crop Evol 48:409–417

Rodiño AP, Santalla M, De Ron AM, Singh SP (2003) A core collection of common bean from the Iberian peninsula. Euphytica 131:165–175

Rodiño AP, Santalla M, Gonzàlez AM, De Ron AM, Singh SP (2006) Novel genetic variation in common bean from the Iberian peninsula. Crop Sci 46:2540–2546

Rohlf F (2000) NTSYS pc. Numerical taxonomy system exeter publishing, Setauket

Romero J, Sun SM, McLeester RC, Bliss FA, Hall TC (1975) Heritable variation in a polypeptide subunit of the major storage protein of the bean, *Phaseolus vulgaris* L. Plant Physiol 56:776–779

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. Published Online July 2009. doi:10.1111/j.1752-4571.2009.00082.x

Santalla M, Rodiño AP, De Ron AM (2002) Allozyme evidence supporting southwester Europe as a secondary center of genetic diversity for common bean. Theor Appl Genet 104:934–944

SAS Institute Inc. (2007) JMP Software ver. 7

Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana

Sicard D, Nanni L, Porfiri O, Bulfon D, Papa R (2005) Genetic diversity of *Phaseolus vulgaris* L and *P*. coccineus L. landraces in central Italy. Plant Breed 124:464–472

Singh SP (2001) Broadening the genetic base of common bean cultivars: a review. Crop Sci 41:1659–1675

Singh SP, Nodari R, Gepts P (1991) Genetic diversity in cultivated common bean. I. allozymes. Crop Sci 31:19–23

Sukhotu T, Kamijima O, Hosaka K (2006) Chloroplast DNA variation in the most primitive cultivated diploid potato species *Solanum stenotomum* Juz. et Buk. and its putative wild ancestral species using high-resolution markers. Genet Resour Crop Evol 53:53–63

Sun SM, Hall TC (1975) Solubility characteristics of globulins of *Phaseolus* seeds in regard to their isolation and characterisation. J Agric Food Chem 23:184–189

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. Mol Ecol 5:595–598

Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley J (2005) An analysis of genetic diversity across the maize genome using microsatellites genetics. Genetics 169:1617–1630

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Xu DH, Abe J, Gai JY, Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theor Appl Genet 105:645–653

Zhang X, Blair MW, Wang S (2008) Genetic diversity of Chinese common bean *Phaseolus vulgaris* L.) landraces assessed with simple sequence repeats markers. Theor Appl Genet 117:629–640