Mol Breeding (2009) 23:413–429
DOI 10.1007/s11032-008-9245-8

Nuclear and chloroplast microsatellite diversity in *Phaseolus vulgaris* L. from Sardinia (Italy)

S. A. Angioi · D. Rau · M. Rodriguez · G. Logozzo · F. Desiderio · R. Papa · G. Attene

Received: 6 May 2008 / Accepted: 24 November 2008 / Published online: 12 December 2008

Abstract Studies of the level and the structure of the genetic diversity of local varieties of *Phaseolus vulgaris* are of fundamental importance, both for the management of genetic resources and to improve our understanding of the pathways of dissemination and the evolution of this species in Europe. We have here characterized 73 local bean populations from Sardinia (Italy) using seed traits and molecular markers (phaseolins, nuSSRs and cpSSRs). American landraces and commercial varieties were also included for comparison. We see that: (a) the Sardinian material is distinct from the commercial varieties considered; (b) the variation in the seed traits is high and it mostly occurs among populations (95%); (c) compared to the American sample and the commercial varieties, the Sardinian collection has a low level of diversity; (d) the majority (>95%) of the Sardinian individuals belong to the Andean gene pool; (e) the Sardinian material shows a strong genetic structure, both for cpSSRs and nuSSRs; (f) the nuSSRs and cpSSRs concur in differentiating between gene pools, but a lack of congruence between nuclear and chloroplast has been observed within gene pools; and (g) there are three putative hybrids between the Andean and Mesoamerican gene pools. Despite the relatively low level of diversity, which is probably due to a strong founder effect, the Sardinian landraces are worth being conserved and studied further because of their distinctiveness and because hybridization within and between the gene pools could generate variation that will be useful for breeding.

Keywords *Phaseolus vulgaris* · Landraces · Genetic diversity · Simple sequence repeats · Phaseolins

Electronic supplementary material The online version of this article (doi:10.1007/s11032-008-9245-8) contains supplementary material, which is available to authorized users.

S. A. Angioi · D. Rau · G. Attene (✉)
Dipartimento di Scienze Agronomiche e Genetica Vegetale Agraria, Center for Biotechnology Development and Biodiversity Research, Università degli Studi di Sassari, Via de Nicola, 07100 Sassari, Italy
e-mail: attene@uniss.it

M. Rodriguez · G. Attene
Centro per La Conservazione e Valorizzazione della Biodiversità Vegetale, Università degli Studi di Sassari, Surigheddu, Alghero, Italy

M. Rodriguez
Center for Biotechnology Development and Biodiversity Research, Università degli Studi di Sassari, Via de Nicola, 07100 Sassari, Italy

G. Logozzo
Centro Interdipartimentale per la Salvaguardia delle Risorse Genetiche Vegetali "Pierino Iannelli", Università degli Studi della Basilicata, Viale dell'Ateneo Lucano 10, 85100 Potenza, Italy

F. Desiderio · R. Papa
Dipartimento di Scienze ambientali e delle produzioni animali, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy

Introduction

Little is known about the pathways of dissemination and the evolution of *Phaseolus vulgaris* L. in Europe. This species was introduced after the voyages of Columbus passing from Spain and Portugal into the whole of the European Continent (Simmonds 1976) from both centres of domestication (Andes and Mesoamerica) (Gepts and Bliss 1988; Zeven 1997).

It has been suggested that crop expansion from America into Europe resulted in a reduction in the diversity of the European common bean because of strong founding effects and adaptation to new environments, and because of consumer preferences (Gepts 1999). Some more recent studies have suggested that this reduction might not have been as strong as previously suspected (Santalla et al. 2002; Papa et al. 2006).

Electrophoretic analysis of the phaseolins (the major seed storage proteins of the common bean) is a useful and extensively used tool for establishing the genetic relationships between the two gene pools of the common bean (Gepts et al. 1986; Singh et al. 1991). Recently, the phaseolins have been useful for the study of the genetic diversity at different levels: from the continental (Gepts and Bliss 1988; Lioi 1989; Rodiño et al. 2001, 2003; Ocampo et al. 2002; Logozzo et al. 2007) to the regional level (e.g., Galicia in Spain, Escribano et al. 1998; Abruzzo and Lazio in central Italy, Piergiovanni et al. 2000a, b, 2006; and Basilicata in southern Italy, Limongelli et al. 1996).

Based on the phaseolins and on morphological traits, common beans from both of the American gene pools have often been identified in Europe within the same geographical areas or landraces (e.g., Lioi 1989; Gepts and Bliss 1988; Santalla et al. 2002; Logozzo et al. 2007). Moreover, recent studies have suggested that hybridization between the Andean and Mesoamerican genotypes could have taken place in Europe (Santalla et al. 2002; Sicard et al. 2005; Ocampo et al. 2005; Papa et al. 2006; Angioi et al. in preparation).

However, with only a few exceptions (Negri and Tosti 2002; Escribano et al. 1998; Sicard et al. 2005; Lioi et al. 2005; Tiranti and Negri 2007; Sánchez et al. 2007), the European common bean has been analysed only using phaseolin markers. However, different markers that cover different regions of the plant genome and that differ in their modes of inheritance are needed (biparental nuclear microsatellites [nuSSRs] versus uniparental chloroplast microsatellites [cpSSRs]; Provan et al. 2001). Due to their high mutation rates, nuSSR markers are ideal for the characterization of genetic diversity in crop species at the inter-specific, inter-subspecific, inter-varietal and even intra-varietal levels (Lee 1995; Mitchell et al. 1997; Matus and Hayes 2002). SSR markers have been developed for the common bean (Yu et al. 2000; Gaitán-Solís et al. 2002; Blair et al. 2003; Yaish and Pérez de la Vega 2003; Hanai et al. 2007) and they have been used to construct a PCR-based genetic map (Yu et al. 2000; Blair et al. 2003) to evaluate intra-specific specific diversity within the genus (Gaitán-Solís et al. 2002) and to fingerprint genetic diversity in commercial varieties of the common bean from Europe (Métais et al. 2002; Masi et al. 2003) and from Nicaragua (Gomez et al. 2004). However, they have not been extensively tested in other regional collections or used for association mapping. Recently Blair et al. (2006), used 129 SSR markers on a set of 44 genotypes of *P. vulgaris* and showed that it is possible to distinguish between the two gene pools and among races in the Andean gene pool.

Due to their relatively high levels of polymorphism and because they are generally maternally inherited in angiosperms, chloroplast microsatellites (cpSSRs) represent another useful tool in the study of genetic variation and evolution of plants (Olmstead and Palmer 1994). The flanking regions of the cpSSRs loci are conserved among different related species (Olmstead and Palmer 1994). Studies that have used flanking PCR primers have shown that cpSSRs are polymorphic among different species and accessions of *Glycine* (Powell et al. 1996; Xu et al. 2002), *Hordeum* (Provan et al. 1999), *Oryza* (Provan et al. 1996; Ishii and McCouch 2000), *Pinus* (Cuenca et al. 2003), *Solanum* (Bryan et al. 1999; Sukhotu et al. 2006), *Vitis* (Arroyo-García et al. 2002), *Anthyllis* (Nanni et al. 2004) and *Phaseolus* spp. (Sicard et al. 2005). In addition, Angioi et al. (in preparation) recently tested 39 primer pairs (available in the literature or developed de novo in their study) and selected a set of markers highly informative for the study of the diversity and the evolution of the *Phaseolus* species, and in particular of *P. vulgaris* and *P. coccineus.*

In the present study, we have characterized a *P. vulgaris* L. collection from Sardinia (Italy) using molecular and morphological traits. The genetic diversity was assessed with three main aims: to

clarify the origin of the Sardinian bean germplasm; to compare the local accessions with commercial varieties introduced in the last decades in Sardinia and with plant materials from the Americas; and to estimate the level and the pattern of genetic diversity of *P. vulgaris* on an island (Sardinia) of the European continent. This information will be important to elucidate the colonization process of *P. vulgaris* in Europe. Moreover, the present study is useful for the definition of an appropriate management of the local genetic resources, particularly for breeding purposes.

Materials and methods

Plant materials

The accessions of common bean (*Phaseolus vulgaris* L.) analysed in this study were collected in Sardinia in 2003 and are currently conserved at the "Centro per la Conservazione e la Valorizzazione della Biodiversità Vegetale" of the University of Sassari, Sardinia. Each seed lot was obtained from a single farmer (and from a single field) and was considered as a single local population when (according to the information given by the farmer) it had been cultivated locally in the same farm no less than 30 years (Louette 2000).

The molecular analysis included: 73 Sardinian individuals (one per population) of *P. vulgaris* (Table S1); 21 American individuals of the cultivated bean (Table S2), of which eight were from Central America (including BAT93) and 13 from South America (including Jalo EPP558); and 15 commercial varieties chosen from among those most widespread in Sardinia over the past decades. All of the American accessions were kindly provided by Centro Internacional de Agricoltura Tropical (CIAT), United States Department of Agriculture (USDA) and University of Davis, California (UCDavis). Nine of the commercial varieties were kindly provided by the Istituto Sperimentale per le Colture Industriali (ISCI) of the Ministero delle Politiche Agrarie e Forestali (MIPAF); the remaining six were obtained from local Sardinian markets (Table S3). One wild genotype from Argentina (Andean origin) (Table S2) and two genotypes of runner bean (*Phaseolus coccineus* L.), one from Sardinia (Table S1) and one from Peru (Table S2), were also added as controls. The Sardinian accessions were mainly from the central-eastern area of the region; a few were from the North of Sardinia, while there were no accessions from the South.

Morpho-phenotypic analysis

Morpho-phenotypic analyses of the seeds were carried out for the 73 Sardinian populations of *P. vulgaris* (on average 18 seeds per population, for a total of 1,306 seeds) (Table S4) and for 10 of the 15 commercial varieties (Table S5), as the seeds of five varieties were sufficient for DNA extraction only. The following characters were recorded, according to the International Plant Genetic Resources Institute descriptors list (IPGRI 1982) for *P. vulgaris*: seed coat (pattern, darker colour and lighter colour) and seed shape.

However, according to IPGRI descriptors, two seeds can have the same colour but different patterns and be classified as identical. Similarly, the colour of the stripes of one seed might be equal to the background colour of the second seed, and vice versa. To avoid these ambiguities in this study, we introduced a new character that we have called 'prevalent', for which three states are possible: (1) Lighter colour as background and darker colour as stripes. (2) Darker colour as background and lighter colour as stripes. (3) Darker colour and lighter colour equally distributed.

The morpho-phenotypic characteristics of the seeds of the American genotypes were not recorded because the amounts of seed were sufficient for DNA extraction only.

Phaseolin analysis

Ten seeds were individually analysed for each accession. The seed coats were manually removed before grinding. Phaseolin was extracted by resuspending flour in 0.5 M NaCl (Limongelli et al. 1996). One-dimensional SDS-PAGE of the extracts was performed according to Ma and Bliss (1978), using 17% (w/v) polyacrylamide slab gels.

Molecular analyses (nuSSRs and cpSSRs)

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 1U Taq polymerase (Invitrogen). The amplifications were carried out with a Perkin–Elmer 9700 thermocycler (Applied Biosystems) using different conditions, depending on the primer pair (Table 1). The DNA fragments were separated in 6% 8 M denaturing acrylamide:bisacrylamide (19:1) gels 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). Controls were added to each PCR reaction and gel to verify the repeatability and accuracy of the method, and to allow comparisons within and across gels.

We analysed the whole sample with 10 nuclear microsatellite (nuSSR) primer pairs (Table 1). Five of these were developed from expressed sequence tags and were located on the consensus genetic map of *P. vulgaris* constructed using a RIL population obtained from the cross between the Mesoamerican cultivar BAT93 and the Andean cultivar Jalo EPP558 (Xu et al. 2002). The other five were developed by Gaitán-Solís et al. (2002) from genomic sequences of *P. vulgaris*. The molecular analyses were completed by using 11 universal chloroplast microsatellite (cpSSR) primer pairs (Weising and Gardner 1999; Chung and Staub 2003) and three primer pairs (cp1, cp2, cp3) designed by Angioi et al. (in preparation) specifically for *P. vulgaris* (Table 1) and tested in a set of Leguminosae (comprising *P. vulgaris* and *P. coccineus*).

Statistical analyses

The genetic diversity (for the nuclear and chloroplast markers, separately) within each of the analysed groups of individuals (Sardinian, American and commercial varieties) was studied by calculating the number of observed alleles (n_a), the number of haplotypes, the average gene diversity ($\overline{H}_E$; Nei 1978) as:

$$\bar{H}_E = \left(\frac{n}{n-1}\right)\left(1 - \sum_{j=1}^{r}\sum_{i=1}^{k} p_{ij}^2 \Big/ r\right)$$

where p_{ij} is the frequency of the *i*th variants at the *j*th locus and r is the number of loci, and genotypic diversity (I_{nor}, Shannon and Weaver 1949; Hutcheson 1970; Kumar et al. 1999) as:

$$I = -\sum_{i=1}^{n} p_i \ln(p_i)$$

where p_i is the frequency of the *i*th genotype and n is the number of individuals in the sample, normalized as $I_{nor} = I/\ln(n)$where n is the number of individuals in the sample.

The statistics of diversity were calculated using all of the loci (monomorphic and polymorphic).

To compare number of alleles for the groups of individuals of different sample size, we calculated the

Table 1 NuSSR and cpSSR primer pairs, annealing temperatures (T_a), PCR programs, and references

SSR	Primer name	T_a (°C)	PCR program[a]	References
nuSSR	BM138, BM184, BM199, BM210, BM211 PH3B4, PH10B11, PH5B5, PH7B3, PH9B2	50	1	Gaitán-Solís et al. (2002), Yu et al. (2000)
cpSSR	Ccmp2[b], ccSSR2[b], ccSSR9, ccSSR16, ccSSR7, ccSSR18, ccSSR12, ccSSR19, ccSSR20			Weising and Gardner (1999), Chung and Staub (2003)
cpSSR	Ccmp3	45	2	Weising and Gardner (1999)
ccSSR	ccSSR11	43	3	Chung and Staub (2003)
cpSSR	Cp1	50	4	Angioi et al. (in preparation)
cpSSR	Cp2, Cp3	48	5	Angioi et al. (in preparation)

[a] PCR programs used: (1) 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%; (2) 5 min at 94°C; 20 cycles of: 1 min at 94°C, 1 min at 45°C, 1 min at 72°C; 7 min at 72°C; 4°C ∞. Touch down 53–45°C with −1°C every 2 cycles; (3) 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 ∞. Ramping: 75%; (4) 5 min at 94°C; 30 cycles of: 30 s at 94°C, 30 s at 50°C, 30 s at 72°C; 35 min at 72°C; 4°C ∞; (5) 5 min at 94°C; 30 cycles of: 30 s at 94°C, 30 s at 48°C, 30 s at 72°C; 35 min at 72°C; 4°C ∞

[b] These two primer pairs were designed on the same locus around the same microsatellite

R_S statistic (allelic richness) (El Mousadik and Petit 1996) as:

$$R_S = \sum_{i=1}^{R}\left[1 - \frac{\left(\frac{2N-N_i}{2n}\right)}{\left(\frac{2N}{2n}\right)}\right]$$

where N_i is the number of alleles of i type among the $2N$ genes, with the FSTAT software version 2.9.3.2 (Goudet 2002).

To identify the genetic groups within the Sardinian collection and and to investigate their relationships with the Mesoamerican, Andean and commercial varieties samples the software STRUCTURE version 2.0 (Pritchard and Wen 2003) was used. The estimate of the most likely number of genetic groups (K_s) was performed following the procedure of Evanno et al. (2005) that proposed an ad hoc statistic, ΔK. The range of possible K_s tested was from 1 to 7, i.e. from 1 to 4 (the number of groups considered a priori), plus 3. We chose the admixture model and the option of correlated allele frequencies between populations that are considered the best in cases of subtle population structures (Evanno et al. 2005; Falush et al. 2003). Alpha was inferred from the data and lambda was set to 1 (Pritchard and Wen 2003; Evanno et al. 2005). For each K, 20 runs (100,000 burn-in, 100,000 Marchov Chain Monte Carlo) were carried out. Afterwards, a comparison was performed between the K groups and the groups previously defined according to other criteria (e.g., geographic, morphological, genetic).

The pairwise genetic distances between individuals based on nuSSR and cpSSR data were estimated separately using the Nei and Li (1979) genetic distances matrix. A dendrogram was then obtained through the UPGMA method. The levels of support for the nodes were evaluated by bootstrap analysis (1,000 replicates) carried out using the TreeCon software ver 1.13 (Van de Peer and De Watcher 1994). For these analyses, each allele (30 for cpSSR and 44 for nuSSR, see "Result") was considered as a single locus with two possible states: 0 (band absent) or 1 (band present).

All of the polymorphic recorded traits (morpho-phenotypic, nuSSRs and cpSSRs) were simultaneously used to determine the distances between the Sardinian local accessions and the commercial varieties. CpSSR and nuSSR were used as previously described. Each morphological character state was treated as a single locus, for which two states are possible: 0 (trait absent) or 1 (trait present). The distances were calculated based on the simple matching method using the TreeCon software ver 1.13 (Van de Peer and De Watcher 1994).

The genetic divergence among populations (according to the morpho-phenotypic seed traits) was calculated using the F_{ST} statistic (Wrigth 1951), after the grouping of the individuals by origin, name of the accession, name of the farmer, and seed type. The groups (14 locations, 9 groups of names, 14 groups of farmers and 13 morphotype groups) with ≥2 individuals were considered as populations. All of the F_{ST} statistics were estimated according to Weir and Cockerham (1984)and implemented in the TFPGA software (Miller and Kapuscinsky 1997). The F_{ST} significance (>0) was calculated by adopting the bootstrap resampling technique (1,000 replicates) implemented in the TFPGA software.

Results

Seed traits

Combining all of the variants for all of the traits detected, 29 different morphotypes were seen for the 1,306 seeds analysed (Fig. 1; Table S4). The frequency of each of the morphotypes ranged from 23 to 0.1%. Six out of the 29 morphotypes (21%) showed a frequency ≥5%, and represented 59% of the collected seeds ('white oval' 23%, 'maroon black striped reniform' 10%, 'maroon cuboid' 8%, 'maroon black striped cuboid' 8%, 'pale cream to puff cuboid' 5%, and 'brown, pale to dark reniform' 5%; Fig. 2).

According to the seed trait analysis, 62 out of 73 (85%) populations of *P. vulgaris* were homogeneous, and 11 (15%) were heterogeneous, e.g., contained more than one morphotype (Table S1). These heterogeneous populations contained from two to four seed morphotypes (mean = 2.7); in all of these the predominant morphotype showed a frequency >50%, except for the two populations of '77-Iscrittu-Gavoi' (50%) and '11-Murra-Tonara' (35%).

The ten commercial varieties were divided into six different morphotypes (Table S5). Four of these 'commercial' morphotypes comprised 8.4% of the Sardinian collected seeds. Five out of the 73 populations analysed (7%) were mostly constituted by seeds that were indistinguishable from at least one of the commercial varieties used as controls. In

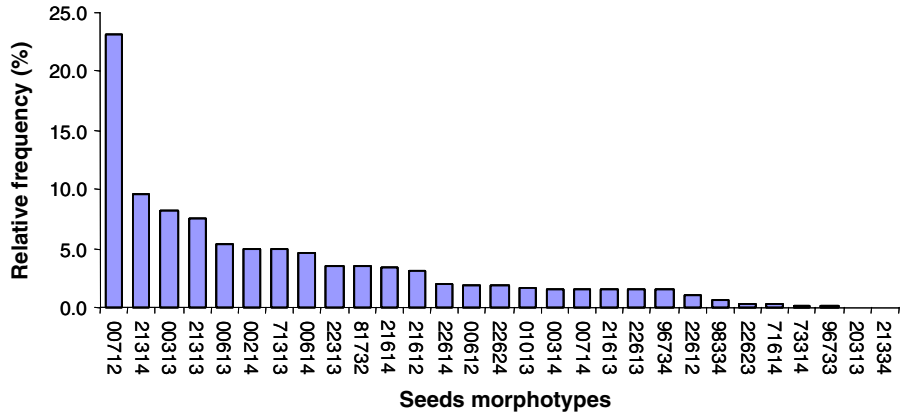


Fig. 1 Frequency distribution of the seed morphotypes within the Sardinian bean collection

Fig. 2 The six out of the 29 most frequent morphotypes 00712 = *white oval*; 21314 = *maroon black striped reniform*; 00313 = *maroon cuboid*; 21313 = *maroon black striped cuboid*; 00613 = *pale cream to puff cuboid*; 00214 = *brown, pale to dark reniform*

particular, 95% of the seeds of the '45-Santa Teresa-Talana' and '3-Faigedda Murra-Tonara' populations, and 30% of the '11-Murra-Tonara' population were indistinguishable from the commercial varieties 'Borlotto Dracula', 'Borlotto lingua di fuoco' and 'Stragonta', respectively. All of the '87-Asolu nied-du-Pattada' seeds and 15% of the '7-Murra-Tonara' seeds were indistinguishable from the 'Meraviglia di Venezia' and 'Marconi' varieties. Finally, all of the '10-Pisu balla-Tonara' seeds were indistinguishable from 'Milagrow' and 'Blue Lake S.B.', and all of the '63-Sordadina/Pintone-Ollolai' seeds were indistinguishable from the 'Lingua di fuoco 2'.

The differentiation level among the Sardinian populations according to the seed traits analysis was very high ($F_{ST} = 0.952$).

Because of these results, we analysed a single individual for each population (the unique morphotype for the homogenous populations, or the most frequent morphotype for the heterogeneous ones).

Phaseolin

The phaseolin pattern was analysed in all of the three groups of materials (Sardinian, American and commercial varieties). There were three kinds of phaseolin seen: types S (typical of the Mesoamerican gene pool), and C and T (typical of the Andean gene pool), with phaseolins C and T always prevalent. In particular, phaseolins C and T represented 99% of the Sardinian sample (only the genotype '10-Pisu balla-Tonara' had phaseolin S), 55% of the American sample, and in 80% of the commercial varieties sample (Tables S1, S2, S3).

Genetic diversity for nuSSRs and cpSSRs

NuSSR genetic diversity: nine out of 10 nuSSR loci were polymorphic in *P. vulgaris;* the BM199 locus, which was monomorphic in *P. vulgaris* (Table 2a), was nonetheless able to discriminate between *P. vulgaris* and *P. coccineus* (data not shown). The BM210 and PH7B3 loci were the most polymorphic ones (with seven alleles each) while the PH9B2 and PH5B5 loci were the least polymorphic (with three alleles each) (Table 2a). Overall, 44 alleles were identified (mean = 4.4 per locus) (Table 2a). An equal number of alleles (22) were observed for the genomic and genic SSRs. The genomic loci were on average more polymorphic (+14%) than the genic loci (5.3 versus 4.6 alleles per locus) when only polymorphic loci were considered.

The Sardinian germplasm and the South American group showed three private alleles each (i.e., three alleles are exclusive of the Sardinian germplasm, and three alleles of the South American group), while the Central American group showed six private alleles; two private alleles were found in the commercial varieties (Table 2a). Overall here, 28 alleles were identified in the Sardinian materials, 36 in the American (29 for South America and 23 for Central America), and 29 in the commercial varieties group (Table 2a). Thus, the Sardinian collection showed a lower number of alleles (n_a) as compared to the American and the commercial varieties groups. This was more evident when the allelic richness (R_S) was considered (Table 3). The Sardinian allelic richness was 65% of the richness observed in the American and 75% of that observed for the commercial varieties.

No heterozygotes were found in any of the analysed groups ($H_o = 0$). The H_E of the Sardinian collection was 57% of that for the American sample, and 68 and 73% of the South and Central American groups, respectively. Eighty-two haplotypes out of 109 genotypes (75%) were detected, 49 in the Sardinian, 20 in the American (13 in South and eight in Central American) and 14 in the commercial varieties (Table 3).

The private haplotypes were distributed across the groups as follows: 48 for the Sardinian (98% of the Sardinian haplotypes), 12 for the South American (92%), seven for the Central American (88%) and 13 for the commercial varieties (93%) (Table 3).

Differences were less evident when considering the multilocus diversity (I_{nor}). The Sardinian genotypic diversity was 89% of the American sample and 90% of the commercial varieties group.

CpSSR genetic diversity: of the 14 primer pairs analysed in *P. vulgaris*, 10 were polymorphic and four (ccSSR7, ccSSR12, ccSSR18 and ccmp3) were monomorphic (Table 2b). Overall, 30 alleles were detected (mean = 2.14 per locus). Six loci were highly polymorphic (ccmp2, ccSSR2, ccSSR11, ccSSR9, ccSSR19, ccSSR20), showing three alleles each, while four loci (ccSSR16, cp1, cp2, cp3) showed only two alleles (Table 2b). The Sardinian materials showed two private alleles, with one detected in the commercial varieties. The American materials did not show any private alleles. The number of alleles observed in the Sardinian collection (27) was similar to that in the American sample (26), but higher than that for the commercial varieties (17) (Table 2b). The Sardinian group showed an allelic richness (R_S) that was 85% of that of the American sample (75 and 101% for South and Central America, respectively), and 132% of the commercial variety group (Table 3).

The gene diversity (H_E, Table 3) of the Sardinian collection was 41% of that for the American sample, which was 47 and 76% for the South and Central American groups, respectively and 133% of the commercial varieties.

Thirty-five haplotypes were detected out of 109 individuals (32%): 18 in the Sardinian, 20 in the American (13 for South and 7 for Central American) and 6 in the commercial varieties (Table 3). The Sardinian collection showed 11 private haplotypes (61% of the total Sardinian haplotypes). The most frequent Sardinian haplotype accounted for 25% of

Table 2 Allele numbers for each locus and for each group of accessions of *P. vulgaris* for both genomic and genic nuSSRs (a) and for cpSSR (b)

Groups	Sample size	Genomic						Genic						
		BM138	BM184	BM210	BM211	BM199	n_a	PH3B4	PH9B2	PH10B11	PH5B5	PH7B3	n_a	Total
(a) Genic nuSSRs														
Sardinia	73	3	3	3 (1)	3 (1)	1	13	3	1	4	2	5 (1)	15	28
America	21	5	3	5	2	1	16	5	3	4	3	5	20	36
South America	13	5	3	3 (1)	1	1	13	4	1	4	2	5 (2)	16	29
Central America	8	4	3	3 (1)	2 (1)	1	13	2 (1)	3 (2)	1	2 (1)	2	10	23
Varieties	15	4	4 (1)	4	3 (1)	1	16	2	1	4	2	4	13	29
Total	109	5	4	7	5	1	22	5	3	4	3	7	22	44

Groups	Sample size	Ccmp2	ccSSR2	ccSSR11	ccSSR16	ccSSR9	ccSSR19	ccSSR20	cp1	cp2	cp3	ccSS7	ccSSR12	ccSSR18	Ccmp3	Total
(b) cpSSR																
Sardinia	73	3	3	3	2	2	2	2	2	2 (1)	2 (1)	1	1	1	1	27
America	21	3	3	3	2	2	3	2	2	1	1	1	1	1	1	26
South America	13	3	3	3	2	2	3	2	2	1	1	1	1	1	1	26
Central America	8	2	2	2	2	2	3	2	1	1	1	1	1	1	1	22
Varieties	15	1	1	1	1	2 (1)	1	2	2	1	1	1	1	1	1	17
Total	109	3	3	3	2	3	3	3	2	2	2	1	1	1	1	30

In parentheses, the numbers of private alleles

Table 3 NuSSR and cpSSR genetic diversity

Groups	Sample size	nuSSR					cpSSR				
		n_a	R_S	H_E	No. of haplotypes	I_{nor}	n_a	R_S	H_E	No. of haplotypes	I_{nor}
Sardinia	73	28	22.3 (20.3)	0.277	49 (48)	0.866	27	18.5 (16.1)	0.120	18 (11)	0.510
Sardinia 2	72 (71)	26	[19.8]	0.271	48 (47)	0.856	25	[15.4]	0.102	17 (11)	0.497
America	21	36	34.4	0.486	20 (20)	0.978	26	21.8	0.290	20 (15)	0.978
South America	13	29	(27.2)	0.405	13 (12)	1.000	26	(21.4)	0.257	13 (8)	1.000
South America 2	11	25	[23.7]	0.331	11 (11)	1.000	23	[18.9]	0.265	11 (7)	1.000
Central America	8	23	(22.0)	0.380	8 (7)	1.000	20	(16.0)	0.157	7 (7)	0.917
Central America 2	10	24	[22.7]	0.377	9 (9)	0.940	21	[17.0]	0.128	9 (9)	0.940
Varieties	15	29	28.0 (26.9)	0.428	14 (13)	0.966	18	14.0 (13.4)	0.090	6 (2)	0.597
Total	109	44	30.6	0.392	82	0.909	30	22.3	0.179	35	0.583

The accessions were analyzed according to their geographical origins and (indicated by 2) according to their more likely gene pools of origin (Mesoamerican or Andean), as inferred by the molecular analyses (nuSSRs performed by STRUCTURE software, and cpSSRs) and seed trait analysis

n_a number of alleles; R_S allelic richness; H_E genetic diversity; Number of haplotypes, private haplotypes between parentheses; I_{nor} normalized Shannon–Weaver index. R_S (allelic richness) calculated over different groups: Sardinian, America, varieties and total sample, over Sardinia, Central America, South America and varieties (within parentheses) and over Sardinia 2, South America 2, Central America 2 (within square brackets)

The sample 'Sardinia 2' for the nuSSRs indicates the diversity of the Andean gene pool in Sardinia because it does not include one genotype ('10-Pisu balla-Tonara') that according to the molecular analysis is more likely to be attributable to the Mesoamerican gene pool. The sample 'Sardinia 2' for cpSSRs is composed of 71 individuals because two Sardinian genotypes are attributable to the Mesoamerican gene pool on the basis of the cpSSR analysis ('30-Granino-Tempio' and '40-Pisu-Ulassai')

the individuals, with the three most common haplotypes accounting for 69% of the individuals.

The genotypic diversity (I_{nor}) of the Sardinian collection was lower than that observed in the other two groups, representing 52% of the American group and 85% of the commercial varieties group (Table 3).

In comparing the two molecular markers for all of the statistics of richness and evenness that were detected, it is of note that the nuSSRs were always more informative than the cpSSRs (see Table 3). Finally here, when we considered the groups obtained by the genotype-to-population assignment (Sardinia 2, Central America 2 and South America 2) performed through the STRUCTURE software, for both of the molecular analyses (nuSSRs and cpSSRs) the pattern of diversity was substantially unchanged (see Table 3).

Number and structure of the genetic groups

The preliminary analysis to establish the more likely number of genetic groups (K) within the collection was performed according to the nuSSRs, and showed a clear peak for $K = 5$. To better understand the distribution of all of the individuals across the five groups detected by STRUCTURE (Table 4), we can first consider the American individuals:

1. All of the genotypes collected in Mesoamerica were assigned to group I; this group also includes BAT93, a Mesoamerican control.
2. Six out of the 13 genotypes (46%) collected in South America were assigned to group III. This group also includes Jalo EPP558, an Andean control. Four out of the 13 genotypes (31%) collected in South America were assigned to group IV, two to group I and one to group II.

Thus, group I was considered to be representative of the Mesoamerican gene pool (all of the individuals have phaseolin S), and groups II, III and IV to be representative of the Andean gene pool (all of the individuals have phaseolin T or C) (Table 4). Therefore, the two genotypes collected in South America (PI313879 and PI300668) that were assigned to Group I can be considered as two migrants from Central to South America. This hypothesis is

Table 4 Distribution of the individuals of the Sardinian, Central American, South American and commercial varieties across the five genetic groups identified by the STRUCTURE software

Original groups	Genetic groups based on the nuSSR polymorphism					
	Group I	Group II	Group III	Group IV	Group V	Total
Sardinia	1 (S)	17 (11C-6T)	7 (4C-3T)	18 (15C-3T)	30 (21C-9T)	73
Central America	8 (S)	0	0	0	0	8
South America	2 (S)	1 (T)	6 (3C-2T)	4 (4C)	0	13
Commercial varieties	2 (S)	1 (T)	7 (5T)	5 (3T)	0	15
Total	13	19	20	27	30	109

In parentheses, type of phaseolin detected

supported by the cpSSR analysis (Fig. 3a), by the phaseolin analysis (Table S2; Fig. 3c) and by the 100-seed weight measurements (Table S2).

Seeds belonging to the Mesoamerican gene pool have a 100-seed weight lower than those belonging to the Andean gene pool (Gepts 1988). The seed weights were measured for 11 out of 13 South American individuals and for seven out of eight individuals from Central America. The 11 South American accessions showed a mean 100-seed weight of 48 g, while the seven Central American accessions showed a mean 100-seed weight of 31 g (Table S2). When the two putative migrant genotypes from South America to Mesoamerica (PI313879 and PI300668) were assigned to the Mesoamerican gene pool (Table 3), the difference in seed weight between the two gene pools (Andean, 55 g versus Mesoamerican, 28 g) increased (+97%) and became highly significant (*t* test: $P = 0.0018$).

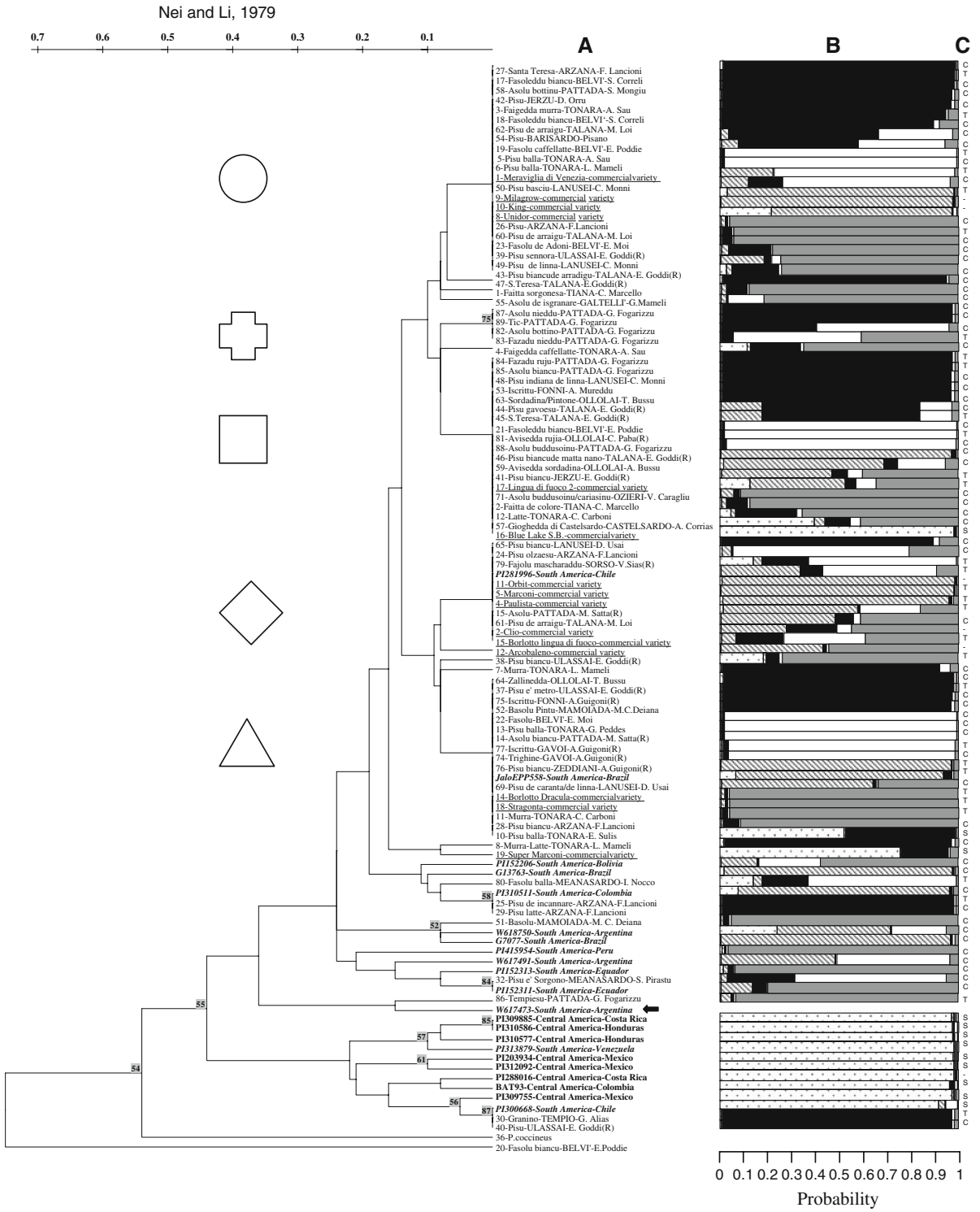
Group V was comprised of only Sardinian individuals (Table 4) and can be considered as an Andean group. Indeed, all of the genotypes attributed to Group V showed phaseolin T or C, which was typical of the Andean group (Fig. 3b, c; Table S1). Moreover, according to the nuSSR genetic distances (Fig. S1), the genotypes assigned to this group were more related to the genotypes belonging to groups IV, III and II (the Andean gene pool), and they showed the highest genetic distance from the individuals belonging to group I (the Mesoamerican gene pool). At a genetic distance of 0.35 (Nei and Li 1979), the cpSSR cluster analysis (Fig. 3a) distinguished two groups: the first comprised all of the South American genotypes (Phaseolin T or C), and the second, all of the Mesoamerican ones (phaseolin S).

Fig. 3 Comparisons between cpSSRs, nuSSRs and phaseolins.▶ **a** In the dendrogram: *bold italics* genotypes from South America; *bold* genotypes from Central America; *underlined* commercial varieties. The *arrow* indicates the wild genotype from South America; the *circle*, *cross*, *square*, *rhombus* and *triangle* indicate the most frequent (>5%) haplotypes. **b** In the histogram: Group 1 *white pointed*; Group 2 *white*; Group 3 *white dashed*; Group 4 *gray*; Group 5 *black*. **c** C and T phaseolin types are from the Andes, S type is from Mesoamerica

Thus, divergence between the two gene pools is in general detected by both the nuclear and the chloroplast analyses, despite incongruence between the two marker systems within the Sardinian Andean group also being detectable. Indeed, individuals that shared the same cpSSR haplotype were often attributed to different nuSSR groups (Fig. 3a, b).

However, it is interesting to note that two commercial varieties ('Blu Lake S. B.' and 'Super Marconi', with 100-seed weights of 27.6 and 50.4 g, respectively) and '10-Pisu balla-Tonara' (56.4 g) were assigned to the Mesoamerican gene pool by the nuSSR analysis (Fig. 3b), while according to the cpSSR analysis, they were assigned to the Andean gene pool (Fig. 3a). These three individuals might therefore be regarded as putative Andean × Mesoamerican hybrids. Conversely, two Sardinian individuals ('30-Granino-Tempio' and '40-Pisu-Ulassai', at 57.8 and 46.7 g, respectively) were attributed to the Andean gene pool for the nuSSR analysis and phaseolin, and to the Mesoamerican gene pool for the cpSSR analysis (Fig. 3a). These two individuals can be regarded as putative Mesoamerican × Andean hybrids.

On the basis of the nuSSR analysis, 1.4% of the Sardinian individuals ('10-Pisu balla-Tonara') were assigned to the Mesoamerican gene pool. On the basis of the cpSSR analysis, 2.7% of the Sardinian



individuals ('30-Granino-Tempio' and '40-Pisu-Ulassai') were assigned to the Mesoamerican gene pool. Thus, according to one or the other of the molecular analysis systems, a maximum of 4.1% of the individuals can be attributed to the Mesoamerican gene pool.

Genetic differentiation among location, local name, farmer and morphotype

For both the nuSSR and cpSSR analyses, the genetic divergence among the groups inferred on the basis of different locations, local names, farmers and morphotypes was low ($F_{ST} = 0.03$–0.16), and in only a few cases significant (CI 95% > 0) (Table 5). Thus, the grouping of the Sardinian material according to these sources of variation was not particularly relevant towards explaining the genetic variability among the accessions.

The groups obtained by the sources of variation of 'location' and 'farmer' showed F_{ST} values for the cpSSRs (both significant) that were higher than those for the nuSSRs (both not significant) (Table 5). Thus, no strong geographical structure was evident, as also shown in Fig. 4. Nonetheless, both the nuSSR ($F_{ST} = 0.12$) and cpSSR ($F_{ST} = 0.16$) for 'local name' explained a higher proportion of molecular variance among the accessions; a lower proportion was explained by 'morphotype' ($F_{ST} = 0.06$ and 0.03, respectively).

Considering the 22 groups composed of nuclear and chloroplast information (as indicated in Fig. 4) we obtained from two to nine morphotypes for each group and a $F_{ST} = 0.37$. When we considered only the four groups with $n \geq 5$, the F_{ST} decreased (0.17).

When the genetic diversity of the Sardinian materials was investigated by using all of the 24 polimorphic traits (five phenotypic, 9 nuSSR and 10 cpSSR) simultaneously, only four cases of identity were observed ('74-Trighine-Gavoi' = '77-Iscrittu-Gavoi'; '5-Pisu balla-Tonara' = '6-Pisu balla-Tonara'; '13-Pisu balla-Tonara' = '14-Asolu biancu-Pattada'; '14-Borlotto Dracula' = '18-Stragonta'). The Sardinian individuals were always distinguishable from the varieties (Fig. S2).

The originality of these materials was also confirmed when we assumed that two individuals must differ for two or more of the 24 traits to be considered different.

Discussion

The collection and characterization of local germplasm are of primary importance for the prevention of their loss and to allow the exploitation of materials that are potentially useful for future purposes. Furthermore, the characterization of collections of local germplasm can also help in the clarification of some aspects concerning the diffusion and evolution of crop plants; several observations at a "micro" scale level can help to elucidate trends at the "macro" scale. Therefore, a characterization of local varieties of the common bean from different European regions can help us to understand the pathways of dissemination and evolution of this species after its introduction into the Old World. Accordingly, in the present study, a Sardinian collection of 73 *P. vulgaris* accessions was analysed using four different types of markers, including morphological (seed) traits, phaseolins, and nuclear and chloroplast microsatellites.

The combined use of different types of markers that are likely to be unevenly affected by major

Table 5 Genetic divergence (F_{ST}) between groups of accessions (locations, local names, farmers or morphotypes)

Source of variation	Number of groups	Mean number of individuals per group	Polymorphic loci	$F_{ST} \pm$ SD	CI 95%
nuSSR					
Location	14	4.7	8	0.08 ± 0.06	−0.04 ÷ 0.17
Farmer	14	3.0	6	0.05 ± 0.10	−0.10 ÷ 0.22
Local name	9	3.2	8	0.12 ± 0.06	0.00 ÷ 0.24
Morphotype	13	4.7	7	0.06 ± 0.06	−0.05 ÷ 0.14
cpSSR					
Location	14	4.7	10	0.10 ± 0.03	0.03 ÷ 0.17
Farmer	14	3.0	9	0.14 ± 0.04	0.00 ÷ 0.19
Local name	9	3.2	7	0.16 ± 0.06	−0.04 ÷ 0.22
Morphotype	13	4.7	10	0.03 ± 0.03	−0.00 ÷ 0.11

F_{ST} standard deviations calculated by jack-knifing over loci; the lower and upper limits of the confidence intervals (CI) were calculated by bootstrapping over loci

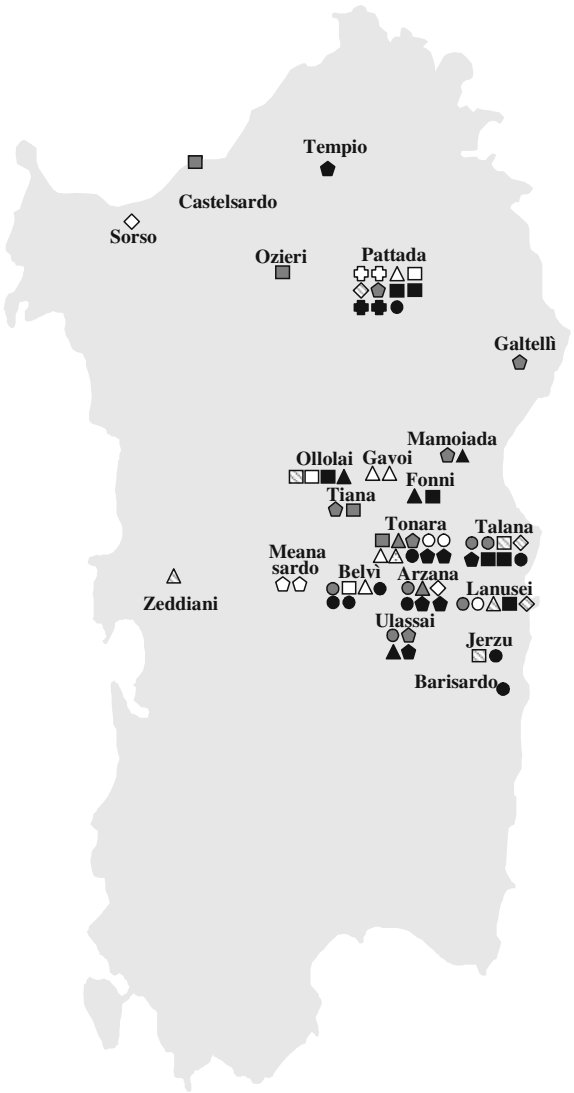


Fig. 4 Geographical distribution of the 73 Sardinian individuals analyzed. *Different colours* indicate different genetic groups, as identified by nuSSR analysis. The *circles, squares, triangles, rhombi* and *crosses* indicate the most frequent (>5%) haplotypes from the cpSSR analysis (see Fig. 3a), the *pentagons* indicate the rare haplotypes (frequency <5%)

evolutionary forces, such as migration, selection, mutation, genetic drift and recombination, offer the opportunity to make inferences from the patterns of genetic diversity, and thus to define a detailed evolutionary scenario and to fully describe the genetic makeup of different groups of germplasm. Our results show that this approach is very useful and informative. For instance, the use of markers showing biparental vs uniparental transmission allowed us to detect the role of hybridization (between and within gene pools) in the shaping of the structure of genetic diversity in the common bean, even in presence of a very low outcrossing rate, as suggested by the absence of heterozygosis found using co-dominant nuclear SSRs. Similarly, we were able to observe that the reduction in genetic diversity in improved varieties was very high for the cytoplasmic (chloroplast) genome compared to the nuclear one, which reflects the use of a few improved genotypes as maternal parent in modern plant breeding, and which outlines the potential usefulness of landraces. Finally, the very low level of variation within accessions for selected traits like seed morphological variants, together with the low level of divergence among different morphotypes, suggested that farmer's deliberate selection played an important role for the conservation of landraces. Overall, our results argue for a coordinated multiple marker approach for germplasm characterization, conservation, and evaluation.

The level of diversity of the Sardinian collection was always lower than that of the American sample, when considering both the statistics of richness (n_a and number of haplotypes/individuals) and evenness (H_E and I_{nor}). This was also true, although to a lesser extent, when the gene pools were considered separately. Moreover, the Sardinian germplasm was almost exclusively of Andean origin (>95%). This suggests that the low variability of the Sardinian collection versus the American sample is not only due to the very low presence of the Mesoamerican plant material in Sardinia, but also to a limited introduction (or to a low persistence of some Andean types) of the Andean material into Sardinia. Thus, a two-step process can be hypothesized: (1) the introduction of germplasm prevalently from one gene pool (Andean) and (2) the low diversity of the Andean germplasm introduced. These might both be attributable to founder effects, but could also be due to the effects of selection (human or natural).

The contribution of the Mesoamerican gene pool to the Sardinian germplasm collected appears not to be greater than 4–5%. This result is remarkably low when compared to other results obtained in Italy at a regional level. For example, when analyzing a collection of *P. vulgaris* from the Marche region using similar markers (nuSSRs, cpSSRs and ISSRs), Sicard et al. (2005) assigned 29% of the individuals analysed to the Mesoamerican gene pool. Previously,

Piergiovanni et al. (2000a) combining data for 36 landraces from Basilicata with 20 landraces from the same region analysed by Limongelli et al. (1996), reported that 12% of the landraces were of Mesoamerican origin. Then in another study based on four populations of *P. vulgaris* from the Abruzzo region Piergiovanni et al. (2000b) assigned 13% of the landraces analysed to the Mesoamerican gene pool.

In line with other regions of Italy, the present study has revealed that in Sardinia the Andean gene pool is also prevalent; however, the contribution of the Mesoamerican gene pool is from 3 to 7 times lower than in other regions. The relative isolation of Sardinia as an island might be the explanation for this "amplified" bottleneck effect when compared to the other regions in Europe. However, a strong selection against Mesoamerican types compared to other parts of Italy cannot be excluded a priori.

It was possible to clearly distinguish the Sardinian landraces from the commercial varieties. Indeed, when all of the traits were considered simultaneously (seeds traits, and nuSSR and cpSSR polymorphisms), the Sardinian landraces were always different from the commercial varieties tested. However, when compared to the commercial varieties, the Sardinian germplasm showed a lower nuSSR variation and a higher cpSSR variation. With regard to this, breeding might have induced a loss of diversity of the chloroplast genome of the commercial varieties. This highlights the importance of Sardinian materials for breeding.

Finally, it is also relevant to note that none of the commercial varieties tested can be considered indistinguishable from individuals belonging to the Sardinian germplasm. This again suggests that Sardinian landraces might be a precious source of genes that will be useful for the breeding of *P. vulgaris*. Nevertheless, it is conceivable that varieties other that the ones included in this study are related to the Sardinian material.

The morphological analysis of the seeds showed a low variation within populations (<5%) and a high differentiation among populations (>95%), with 29 morphotypes detected. This could reflect the tendency of the farmers to select against off-types. Based on the nuSSR analysis, the Sardinian Andean germplasm was subdivided into four different groups. In three of these groups, American individuals with a different country of origin were observed. This suggests that different Andean genetic groups from South America contributed to the composition of the Sardinian gene pool. Based on the cpSSR analysis, the Sardinian germplasm showed a strong structure. This was revealed by the three haplotypes (represented by a similar number of individuals) that accounted for 69% of the total number of individuals. However, no association between nuSSRs and cpSSRs was detected. This could be due to hybridization. Regarding this, it is of note that even if a low frequency of Sardinian individuals is attributable to the Mesoamerican gene pool, one individual can be regarded as an Andean × Mesoamerican putative hybrid, and two individuals can be regarded as Mesoamerican × Andean putative hybrids. This suggests that despite *P. vulgaris* being an autogamous species, hybridization in our environment might not be uncommon, as has been reported for other European, and Italian, regions (Santalla et al. 2002; Ocampo et al. 2005; Sicard et al. 2005).

In the present study, the absence of heterozygosity is in agreement with previous observations that showed the common bean as a strict autogamous species (Park et al. 1996; Wells et al. 1988; Ibarra-Pérez et al. 1997) where, however, hybridization is possible and may induce a significant change in the structure of the genetic diversity (Papa and Gepts 2003; Papa et al. 2005, 2007). Hybridization might depend on the simultaneous cultivation of several landraces from the same farmer in fields that are close together. Together with the exchange of seeds among neighbouring farms, this condition may have resulted in occasional outcrossing and pollen flow.

In considering the best sampling strategy to capture the genetic diversity here, in our case the morphotypes allowed the capturing of low levels of diversity, as indicated by the low levels of molecular variation among the accessions for this trait. Moreover, the high levels of genetic variability for the nuSSRs and cpSSRs present within morphotypes (>95%) suggests that they are useful for the analysis of genetic structure within populations even if they are only composed of one or a few seed morphotypes. This analysis may actually provide information about the frequency of hybridization.

The local name was the best indicator of the genetic divergence among accessions (with 12–16% of the total variance explained). This might be because the local names could represent a 'synthetic

indicator' of the differences among the accessions. Indeed, the local names of the landraces often refer to the seed traits (e.g., *Pisu balla* = round bean; *Iscrittu* = striped; *Latte* = *caffellatte* = café latte colour; *Fajolu mascharaddu* = masked bean), to the putative geographical origins (*Pisu e' Sorgono* = bean from Sorgono; *Pisu e' Meana* = bean from Meana Sardo; *Pisu olzaesu* = bean from Olzai; *Santa Teresa* = from Santa Teresa di Gallura), to the cultivation systems (*Pisu de linna* = *Pisu de arraigu* = *Pisu de incannare* = that need to be supported because of its indeterminate growth habit), or to the shapes and origins simultaneously (*Gioghedda de Castelsardo* = literally snail (shell of) from Castelsardo). An approach "local name" based was also recently proposed by Jarvis et al. (2008). The present study has thus suggested that a stratified sampling might help to capture the diversity, choosing accessions with different names from different farmers and from different locations.

Conclusions

We have described here the status of the genetic resources of *P. vulgaris*, an important species for feeding humans, in a typical environment of the Mediterranean Basin. Theoretical and practical objectives have been achieved. Knowledge of the evolution of *P. vulgaris* in a region far from its centres of origin has been obtained. A detailed database of the materials collected in Sardinia that is based on different types of information has been set up and is available to others upon request. This should help and support future field experiments that are aimed at the exploitation and evaluation of these materials.

Acknowledgments This study was supported by the "Fondazione Banco di Sardegna" for the projects "Interventi di tutela e valorizzazione del patrimonio genetico locale di leguminose da granella in Sardegna" (grant no. #27/4721-02/01/2003) and "Studio della diversità genetica al *locus* faseolinico in fagiolo (*Phaseolus vulgaris* L.): implicazioni nel miglioramento genetico per caratteri di importanza agronomica ed economica e per la conservazione e valorizzazione di varietà locali" (grant no. #389/2005-0112). This work constituted a part of the PhD thesis of S. A. Angioi. We would like to thank the farmers for their attempts towards the prevention of the loss of biodiversity, and the referents, who were fundamental in the obtaining of our seed collection.

References

Angioi SA, Desiderio F, Rau D et al (in preparation) Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. Plant Biol

Arroyo-García R, Lefort F, de Andrés F et al (2002) Chloroplast microsatellite polymorphisms in *Vitis* species. Genome 45:1142–1149. doi:10.1139/g02-087

Bassam BJ, Caetano-Anolles G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80–83. doi:10.1016/0003-2697(91)90120-I

Blair MW, Pedraza F, Buendia HF et al (2003) Development of a genome-wide anchored microsatellite map of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 107:1362–1374. doi:10.1007/s00122-003-1398-6

Blair MW, Giraldo MC, Buendía HF et al (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 113:100–109. doi:10.1007/s00122-006-0276-4

Bryan GJ, McNicoll J, Ramsay G et al (1999) Polymorphic simple sequence repeats markers in chloroplast genomes of *Solanaceous* plants. Theor Appl Genet 99:859–867. doi:10.1007/s001220051306

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. doi:10.1007/s00122-003-1311-3

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. doi:10.1046/j.1365-294X.2003.01890.x

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. doi:10.1007/BF00221895

Escribano , Santalla M, Casquero PA et al (1998) Patterns of genetic diversity in landraces of common bean (*Phaseolus vulgaris* L.) from Galicia. Plant Breed 117:49–56. doi:10.1111/j.1439-0523.1998.tb01447.x

Evanno G, Regnaut S, Goudet J (2005) How to use the software structure: a simulation study. Mol Ecol 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Gaitán-Solís E, Duque MC, Edwards KJ et al (2002) Microsatellite repeats in common bean (*Phaseolus vulgaris*): isolation, characterization, and cross-species amplification in *Phaseolus* ssp. Crop Sci 97:847–856

Gepts P (1988) A Middle American and Andean gene pool. In: Gepts P (ed) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, pp 375–390

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 (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, Osborne TC, Rashka K et al (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

Gomez OJ, Blair MW, Frankow-Lindberg BE et al (2004) Molecular and phenotypic diversity of common bean landraces from Nicaragua. Crop Sci 44:1412–1418

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

Hanai LR, de Campos T, Camargo LEA et al (2007) Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. Genome 50:266–277. doi:10.1139/G07-007

Hutcheson K (1970) A test for comparing diversities based on Shannon formula. J Theor Biol 26:151–154. doi:10.1016/0022-5193(70)90124-4

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

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. doi:10.1007/s001220051432

Jarvis D, Brown AHD, Cuong PH et al (2008) A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. Proc Natl Acad Sci USA 105:5326–5331. doi:10.1073/pnas.0800607105

Kumar J, Nelson RJ, Zeigler RS (1999) Population structure and dynamics of Magnaporthe grisea in the Indian Himalayas. Genetics 152:971–984

Lee M (1995) DNA markers and plant breeding programs. Adv Agron 55:265–344. doi:10.1016/S0065-2113(08)60542-8

Limongelli G, Laghetti G, Perrino P et al (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

Lioi L, Piergiovanni AR, Pignone D et al (2005) Genetic diversity of some surviving on-farm Italian common bean (*Phaseolus vulgaris* L.) landraces. Plant Breed 124:576–581. doi:10.1111/j.1439-0523.2005.01153.x

Logozzo G, Donnoli R, Macaluso L et al (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. doi:10.1007/s10722-006-9185-2

Louette D (2000) Traditional management of seed and genetic diversity: what is a landrace? In: Brush SB (ed) Genes in the field: on-farm conservation of crop diversity. Lewis, Boca Raton, pp 109–142

Ma Y, Bliss FA (1978) Seed proteins in bean. Crop Sci 18:431–437

Masi P, Spagnoletti Zeuli PL, Donini P (2003) Development and analysis of multiplex microsatellite markers sets in common bean (*Phaseolus vulgaris* L.). Mol Breed 11:303–313. doi:10.1023/A:1023443109985

Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome 45:1095–1106. doi:10.1139/g02-071

Métais I, Hamon B, Jalouzot R et al (2002) Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. Theor Appl Genet 104:1346–1352. doi:10.1007/s00122-002-0901-9

Miller LM, Kapuscinsky AR (1997) Historycal analysis of genetic variation reveals low effective size in a northern pike (*Exon lucius*) population. Genetics 147:1249–1258

Mitchell SE, Kresovich S, Jester CA et al (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. Crop Sci 37:617–624

Nanni L, Ferradini N, Taffetani F et al (2004) Molecular phylogeny of *Anthyllis* spp. Plant Biol 6:454–464. doi:10.1055/s-2004-820968

Negri V, Tosti N (2002) Phaseolus genetic diversity maintained on-farm in central Italy. Genet Resour Crop Evol 49:511–520. doi:10.1023/A:1020902508166

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

Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273. doi:10.1073/pnas.76.10.5269

Ocampo CH, Martin JP, Ortiz JM et al (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

Ocampo CH, Martin JP, Sánchez-Yélamo MD et al (2005) Tracing the origin of Spanish common bean cultivars using biochemical and molecular markers. Genet Resour Crop Evol 52:33–40. doi:10.1007/s10722-005-1931-3

Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 81:1205–1224. doi:10.2307/2445483

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, Acosta J, Delgado-Salinas A et al (2005) A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. Theor Appl Genet 111:1147–1158. doi:10.1007/s00122-005-0045-9

Papa R, Nanni L, Sicard D et al (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

Papa R, Bellucci E, Rossi M et al (2007) Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. Ann Bot (Lond) 100:1039–1051. doi:10.1093/aob/mcm151

Park SJ, Michaels TE, Myers JR et al (1996) Outcrossing rates of common beans grow in Ontario and Idaho. Annu Rep Bean Improv Coop 39:90–91

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. doi:10.1023/A:1008719105895

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. doi:10.1023/A:1008777119601

Piergiovanni AR, Taranto G, Losavio FP et al (2006) Common bean (*Phaseolus vulgaris* L.) landraces from Abruzzo and Lazio regions (Central Italy). Genet Resour Crop Evol 53:313–322. doi:10.1007/s10722-004-6144-7

Powell W, Morgante M, Doyle JJ et al (1996) Genepool variation in the genus glycine subgenus soja revealed by polymorphic nuclear and chloroplast microsatellites. Genetics 144:792–803

Pritchard JK, Wen W (2003) Documentation for Structure Software Version 2. http://pritch.bsd.uchicago.edu

Provan J, Corbett G, Waugh R et al (1996) DNA fingerprints of rice (*Oryza sativa*) obtained from hypervariable chloroplast simple-sequence repeats. Proc R Soc Lond 263: 1275–1281. doi:10.1098/rspb.1996.0187

Provan J, Russell JR, Booth A et al (1999) Polymorphic chloroplast simple-sequence repeat primers for systematic and population studies in the genus Hordeum. Mol Ecol 8:505–511. doi:10.1046/j.1365-294X.1999.00545.x

Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol Evol 16:142–147. doi:10.1016/S0169-5347(00)02097-8

Rodiño AP, Santalla M, Montero I et al (2001) Diversity in common bean (*Phaseolus vulgaris* L.) germplasm from Portugal. Genet Resour Crop Evol 48:409–417. doi:10.1023/A:1012248002436

Rodiño AP, Santalla M, De Ron AM et al (2003) A core collection of common bean from the Iberian Peninsula. Euphytica 131:165–175. doi:10.1023/A:1023973309788

Sánchez E, Sifres A, Casañas F et al (2007) The endangered future of organoleptically prestigious European landraces: ganxet bean (*Phaseolus vulgaris* L.) as an example of a crop originating in the Americas. Genet Resour Crop Evol 55:45–52. doi:10.1007/s10722-007-9213-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. doi:10.1007/s00122-001-0844-6

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

Sicard D, Nanni L, Porfiri O et al (2005) Genetic diversity of *Phaseolus vulgaris* L and *P. coccineus L. landraces* in central Italy. Plant Breed 124:464–472. doi:10.1111/j.1439-0523.2005.01137.x

Simmonds NW (1976) Evolution of crop plants. Longman, London

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. doi:10.1007/s10722-004-0573-1

Tiranti B, Negri V (2007) Selective microenvironmental effects play a role in shaping genetic diversity and structure in a *Phaseolus vulgaris* L landrace: implications for on-farm conservation. Mol Ecol 16:4942–4955. doi:10.1111/j.1365-294X.2007.03566.x

Van de Peer Y, De Watcher R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Appl Biosci 10:569–570

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evol Int J Org Evol 38:1358–1370. doi:10.2307/2408641

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19. doi:10.1139/gen-42-1-9

Wells WC, Isom WH, Waines JG (1988) Outcrossing rates of six common bean lines. Crop Sci 28:177–178

Wrigth S (1951) The genetical structure of populations. Ann Eugen 15:323–354

Xu DH, Abe J, Gai JY et al (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theor Appl Genet 105:645–653. doi:10.1007/s00122-002-0972-7

Yaish MWF, Pérez de la Vega M (2003) Isolation of (GA)n microsatellite sequences and description of a predicted MADS-box sequence isolated from common bean (*Phaseolus vulgaris* L.). Genet Mol Biol 26:337–342. doi:10.1590/S1415-47572003000300019

Yu K, Park SJ, Poysa V et al (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L). J Hered 91:429–434. doi:10.1093/jhered/91.6.429

Zeven AC (1997) The introduction of the common bean (*Phaseolus vulgaris* L.) into Western Europe and the phenotypic variation of dry bean collected in the Netherlands in 1946. Euphytica 94:319–328. doi:10.1023/A:1002940220241