Theor Appl Genet (2011) 122:1281–1291
DOI 10.1007/s00122-011-1530-y

ORIGINAL PAPER

Genetic diversity and structure of a worldwide collection of *Phaseolus coccineus* L.

G. Spataro · B. Tiranti · P. Arcaleni · E. Bellucci · G. Attene · R. Papa · P. Spagnoletti Zeuli · V. Negri

Received: 23 March 2010 / Accepted: 6 January 2011 / Published online: 29 January 2011

Abstract *Phaseolus coccineus* L. is closely related to *P. vulgaris* and is the third most important cultivated *Phaseolus* species. Little is known about the patterns of its diversity. In this work, a representative collection of its worldwide diversity was initially developed. The collection includes 28 wild forms (WFs) and 52 landraces (LRs) from Mesoamerica (the crop domestication area), and 148 LRs from Europe (where the crop was introduced in the sixteenth century). The collection was studied by using 12 SSR molecular markers that were developed for the *P. vulgaris* genome. They were proved to be effective and reliable in *P. coccineus* in this work. Fourteen LRs of *P. dumosus* (previously identified as a subspecies of *P. coccineus*) were also studied. The genetic diversity, population structure and phylogenetic relationships were investigated. The results indicate that: (a) the European and Mesoamerican gene pools are clearly differentiated, (b) a certain reduction of diversity occurred with introduction into Europe, and (c) the Mesoamerican LRs (*P. dumosus* included) and WFs are closely related and are connected by a high gene flow. Inferences on the domestication process of *P. coccineus* are also presented. This study provides a picture of the genetic diversity distribution and outcomes with introduction into the Old World, which was not available before. It also underlines that the genetic diversity of both WFs and LRs is an important source for *Phaseolus* spp. breeding programs and deserves to be preserved in situ and ex situ.

Communicated by R. Varshney.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1530-y) contains supplementary material, which is available to authorized users.

G. Spataro · B. Tiranti · P. Arcaleni · V. Negri (✉)
Dipartimento di Biologia Applicata, Università degli Studi di Perugia (UNIPG), Borgo XX Giugno 74, 06121 Perugia, Italy
e-mail: vnegri@unipg.it

E. Bellucci · R. Papa
Dipartimento di Scienze Ambientali e delle Produzioni Vegetali, Università Politecnica delle Marche (UNIPM), Via Brecce Bianche, 60131 Ancona, Italy

G. Attene
Dipartimento di Scienze Agronomiche e Genetica Vegetale Agraria, Università degli Studi di Sassari (UNISS), 14 Via De Nicola, 07100 Sassari, Italy

R. Papa
Centre for Cereal Research CRA-CER, S.S. 16, km 675, 71122 Foggia, Italy

P. Spagnoletti Zeuli
Dipartimento di Biologia Difesa e Biotecnologie Agro-Forestali, Universita' degli Studi della Basilicata (UNIBAS), Viale dell'Ateneo Lucano 10, 85100 Potenza, Italy

Introduction

The genus *Phaseolus* (subtribe *Phaseolinae*, tribe *Phaseoleae*, family Leguminosae) includes over 70 species and five domesticated taxa, *P. vulgaris* L., *P. lunatus* L., *P. acutifolius* A. Gray., *P. coccineus* L. and *P. dumosus* Macfady, which have distinct geographical distributions, life histories and reproductive systems (Delgado-Salinas et al. 2006; Freytag and Debouck 2002; Gepts 1996; Maréchal et al. 1978).

P. coccineus is closely related to *P. vulgaris* (Chacón et al. 2007; Delgado-Salinas et al. 1999; Delgado-Salinas et al. 2006; Gepts 1996; Llaca et al. 1994). Phylogenetic analysis of the genus *Phaseolus,* based on ITS DNA and plastidial loci (Delgado-Salinas et al. 1999; Delgado-

Salinas et al. 2006), revealed that almost all species of *Phaseolus* belong to one of eight clades. *P. coccineus* belongs to the same clade as *P. vulgaris*, *P. dumosus*, *P. acutifolius*, *P. albescens* and *P. costaricensis.* However, within this clade, *P. coccineus*, *P. dumosus*, *P. albescens* and *P. costaricensis* are more closely related to each other than to *P. vulgaris* or *P. acutifolius* (Delgado-Salinas et al. 1999; 2006). This was confirmed by other cloroplast DNA investigations, which also placed the divergence of *P. vulgaris* from its sister taxa at 1.3 Ma (Chacón et al. 2007). Fertility relationships between species reflect the phylogenetic relationships, in particular, *P. coccineus* and *P. vulgaris* are cross fertile when *P. vulgaris* is the mother plant (Escalante et al. 1994; Llaca et al. 1994; Campion 1995).

From an economic point of view, *P. coccineus* (scarlet runner bean, $2n = 2x = 22$) is the third most important bean species, after *P. vulgaris* and *P. lunatus* (Santalla et al. 2004). It is a climbing perennial crop that is often grown as an annual for green pod production or to obtain dry seeds. It is also grown as an ornamental climber.

The centre of *P. coccineus* domestication is Mesoamerica where the wild forms grow (from Chihuahua to Panama, in the cool humid uplands, at altitudes from 1,500 to 3,000 m asl) and coexist with cultivated forms (Delgado-Salinas 1988). The few archaeological remains (Kaplan and Lynch 1999; Delgado-Salinas 1988) also indicate this area as the centre of origin and domestication. These data are consistent with cpDNA analysis, which grouped in a single cluster wild and cultivated accessions from Puebla, Mexico (Llaca et al. 1994).

In *P. coccineus*, an outbreeding species, gene flow between wild and domesticated forms living nearby probably makes the domestication bottleneck less restricted than that which occurred in *P. vulgaris* (Escalante et al. 1994; Llaca et al. 1994).

Similar to what happened to many other crops of American origin, *P. coccineus* was introduced into Europe by the Spaniards in the sixteenth century. The crop subsequently spread from Spain to Italy (where the presence of *Phaseolus* was already documented in 1533) and to other parts of the Old World (Papa et al. 2006; Santalla et al. 2004; Perale 2001; Zeven 1996).

Due to the lower temperature requirements for seed germination and plant growth (Rodiño et al. 2007), the species can grow under cooler environmental conditions than those required by *P. vulgaris.* Because of its tolerance to cooler spells, it has become an important crop in the northern part of Europe and in the mountainous areas of southern Europe where it is grown from spring to summer. It is also cultivated in the higher regions of Africa and Asia. Because of the nutritionist recommendations about fibre and vegetable protein intake in the human diet, the economic importance of *P. coccineus* is likely to increase in the future.

Moreover, runner bean is a useful source of diversity for *P. vulgaris* breeding. The development of new common bean varieties can benefit from interspecific crossing with *P. coccineus,* which was proven by the results of the breeding work for introducing resistance to pathogens (Beaver and Osorno 2009; Welsh and Grafton 2001; Singh 2001; Busogoro et al. 1999a, b). The incomplete reproductive isolation between *P. coccineus* and *P. vulgaris* may have led to hybridization between the two species in Europe and America, where they are often found in sympatry (Papa et al. 2006; Sicard et al. 2005; Negri and Tosti 2002; Delgado-Salinas 1988).

Since information about the level of *P. coccineus* diversity and its evolutionary history with domestication and introduction into other continents is scarce, the aims of this study were:

(a) to build a collection of *P. coccineus* wild forms (WFs) and Mesoamerican and European landraces (LRs) that is representative of the worldwide variation held in germplasm banks;
(b) to obtain information on its genetic diversity and structure, so as to investigate the relationships between accessions from America and Europe.

This is also relevant in estimating the potential of different accessions of *P. coccineus* for plant breeding.

Materials and methods

Collecting accessions and building up a core collection

In addition to the germplasm collections of the departments of the authors, all of the germplasm sources available on the Web or those that were known by the authors (i.e. farmers, amateurs, NGOs) were asked for germplasm. A total of 674 accessions of *P. coccineus* were collected (see "Acknowledgments"). All of the available records about the place of origin, form (cultivated or wild), biological *status* (landrace, cultivar, breeding line), geographic coordinates and altitude were initially collected for each accession from the germplasm bank databases. Geographic data that were not available from these sources were obtained from a number of web sites using the names of the collection sites.

Since the purpose of this study was to relate geographic and genetic diversity, commercial varieties, breeding lines and all materials of uncertain or unknown origin or *status* were excluded, while all of the WFs obtained and part of the LRs were included in the sample.

Regarding LRs, in particular, only those that were clearly declared to be LRs by the donors were maintained. However, the European LR accessions that were received

were not uniformly distributed across the geographical areas of cultivation: accessions from some countries (e.g. Italy, Austria) were overrepresented. In addition, other accessions were found to have been collected from the same collection area and were then, considered the breeding system of the species, suspected to be genetically similar.

Therefore, to avoid unbalanced sampling and to reduce genetic redundancy, a less redundant collection of the European gene pool was developed. Since no data on genetic diversity were available, the number of accessions per country and region within a country was reduced by relying on geographical diversity of the collection sites. Only one accession out of all those located on the same longitude, latitude and altitude in each country region was kept. However, some accessions with the same geographical data were retained in the final collection if the seed morphology was obviously different or if previous characterisation studies showed them to be different LRs (Sicard et al. 2005; Negri and Tosti 2002).

The stratification process described above reduced the number of accessions to a core collection including 228 accessions.

Fourteen LRs of *P. dumosus* Macfady were also included in the study because the species is so closely related to *P. coccineus* (see "Introduction") that it was previously identified as a subspecies of it (i.e. *P. coccineus* subsp. *darwinianus* Hernandez and Miranda, *P. coccineus* subsp. *polyanthus* Greenm., Maréchal et al. 1978).

DNA extraction

Seeds were grown in a glasshouse in pots containing compost. Young leaf material was collected from one individual plant per accession, frozen in liquid nitrogen and stored at −80°C. Plant tissue from a single plant from each accession was crushed with a Mixer Mill MM 300 (Qiagen) and the DNA was extracted using the DNAeasy 96 Plant Kit (Qiagen). To evaluate the quality and quantity of DNA, 1 µl of each DNA sample and 5 µl of Mass Ruler DNA Low Range (Fermentas) were loaded and run on 1.7% agarose.

SSR analysis

Since SSRs specifically developed for *P. coccineus* do not exist and a molecular map of the species is not available, ten SSR that span coding regions (Yu et al. 2000) and ten SSR that span non-coding regions (Blair et al. 2003), located on 9 of the 11 linkage groups of *P. vulgaris,* were initially tested on a sub-sample of 20 accessions (10 WFs and 10 LRs). Of the initially tested markers, only 12 SSR that gave clear, reliable amplifications in *P. coccineus* were used to assay genetic diversity within the collection. The primer details of the 12 SSR are reported in Table 1.

Polymerase chain reaction (PCR) amplifications were performed on 20 µl reaction volume with 25 ng of genomic DNA, 10 µM each of the fluorescence-labelled forward primer and of the reverse primer, 2.5 mM dNTP mixture, 5U of the Taq DNA polymerase (Invitrogen), 10× PCR buffer minus Mg++ (Invitrogen) and 50 mM $MgCl_2$ (Invitrogen). For the primers that span coding sequences, the PCR conditions were: an initial denaturation step at 95°C for 2 min, followed by 30 cycles at 94°C for 2 s (denaturation), 47–52°C (depending on the primer) for 10 s (annealing), 72°C for 2 min (elongation) and a final elongation step at 72°C for 30 min. For the primers that span non-coding sequences, the PCR conditions were the same except for the denaturation and annealing that were carried out for 1 min.

The 12 SSR loci were fluorescently labelled with different colours (FAM, NED, VIC) when their product size was different enough to be distinguished in the electrophoresis. DNA amplification was performed using a GeneAmp 9700 thermocycler (Applied Biosystems).

For electrophoresis, 2 µl of each of the differently coloured PCR products were mixed together in 10 µl of a 75:1 solution of formamide and GENESCAN-500 (ROX) size standard (Applied Biosystems) solution. Microsatellite polymorphisms were electrophoresed and analysed with a fluorescent detection method using an ABI 377 sequencer (Perkin-Elmer). Visualisation and sizing of the DNA fragments were performed using the GENESCAN 3.1 software (Applied Biosystems), which calculates the fragment dimensions by using the Local Southern equation (Southern 1979).

Genetic diversity analysis

Each allele detected was coded as integer size in base pairs (bp). *P. coccineus* accessions were first divided into three groups based on passport data: European LRs, Mesoamerican LRs and Mesoamerican WFs. Total number of alleles (N_a), average number of alleles per locus (N_{ma}), effective number of alleles (N_e), number of alleles with frequency greater than 0.05 (N_a *Freq* > *0.05*), number of private (i.e. unique to a single group) alleles, Nei's unbiased estimate of gene diversity (H_e, expected heterozygosity, Nei 1987), observed heterozygosity (H_o), inbreeding coefficient (f) and allelic richness (AR) (Hurlbert 1971; El Mousadik and Petit 1996; Petit et al. 1998) were worked out for each of them by using the GENEALEX 6 (Peakall and Smouse 2006) and the 9.3.2. FSTAT version (Goudet 2001) software. Because of the different sizes of the groups, to compare the levels of diversity between them, 200 samples with size equal to the smallest (i.e. the MesoamericanWF) sample were

Table 1 Characteristics of common bean microsatellite markers designed from six coding (Yu et al. 2000) and six non-coding (Blair et al. 2003) region sequences of *P. vulgaris* and the number of alleles detected in the *P. coccineus* world collection

Genebank entry	Description	Motif	Allele N	Forward/reverse primers	Ta (°C)	Linkage group (*P. vulgaris*)
X79722[a]	Sn-glycerol-3-phosphate acyltransferase (PVPLB)	$(CCT)_7$	12	ccaaccacattcttccctacgtc gcggaggcagttatctttaggagtg	49	B2
M75856[a]	Phathogenesis-related protein 3 (PHVPVPR3A)	$(CT)_{11}$	18	caatcctctctctcatttccaatc gaccttgaagtcggtgtcgttt	47	B11
X04660[a]	Phytohemagglutinin pseudogene (PVPDLEC1)	$(AG)_8$	14	*ttgatgacgtggatgcattgc* aaagggctaggagagtaagttgg	48	B4
X74919[a]	Endochitinase (PVGEC9)	$(AT)_5$	21	ccgttgcctgtatttccccat cgtgtgaagtcatctggagtggtc	50	B5
X80051[a]	NADP-dependent malic enzyme (PVME1G)	$(AT)_{12}$	29	agttaaattatacgaggttagcctaaatc cattccttcacacattcaccg	49	B9
J01263[a]	Beta-phaseolin (PHVBCSP)	$(ATCC)_3$ $(AG)_2$ $(TAC)_3$	7	atgcatgttccaaccacttctc ggagtggaaccctgctctctcatc	49	B7
AZ044945[b]	Bng225/R common bean genomic clone	$(AGC)_7$	12	catcaacaaggacagcctca gcagctggcgggtaaaacag	47	B6
AZ301561[b]	Bng91/R common bean genomic clone	$(ATT)_9$	12	cagtaaatattggcgtggatga tgaaagtgcagagtggtgga	47	B10
AZ301513[b]	Bng112/R common bean genomic clone	$(CCT)_5$	5	cagcatcaagaagacccaag cagcaccactatgggaggac	47	
AY298744[b]	RAPD clone isolated at CIAT	$(TA)_8$	28	cataacatcgaagcctcacagt acgtgcgtacgaatactcagtc	47	
AZ301573[b]	Bng125/F common bean genomic clone	$(AG)_5$	8	ggcagcttactaacccgaaa ttccttccccttcttctcc	47	B8
AZ301498[b]	Bng68/common bean genomic clone	$(TA)_5$	1	gaggcaaagaagctattgaa tttttaattcctgtgaattgttt	47	

[a] Yu et al. 2000; [b] Blair et al. 2003

resampled. Descriptive statistics per group were calculated on each of the 200 matrix and then averaged across samples. The Wilcoxon–Mann–Withney *U* test (Sokal and Rohlf 1995) was then used to assess differences between the European and Mesoamerican LR distributions of the descriptive statistics.

Investigating relationships between populations and accessions

Initially, pairwise distances between the European LRs, Mesoamerican LRs, Mesoamerican WFs and *P. dumosus* were calculated through the Nei's genetic distance coefficient (Nei 1972) by using GeneAlEx 6 (Peakall and Smouse 2006).

Subsequently, to evaluate if the pattern of introduction into Europe (from the Iberian peninsula to Italy and then to other countries) conditioned differentiation, Nei's genetic distances were calculated between accessions grouped according to country. A dendrogram based on this genetic distance matrix was constructed using the unweighted pair group procedure with an arithmetic mean (UPGMA) (Sneath and Sokal 1973).

Nei's genetic distances were also calculated between each single accession. A 242 × 242 distance matrix was then obtained and used to calculate a principal coordinate analysis (PCoA).

A geographic distance matrix between single accessions was also obtained using the following coefficient:

$$D = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

where x and y are the decimal logarithmic coordinates of sample i and j, respectively. In order to evaluate the degree of correlation between the genetic and geographic distance matrices, a Mantel test (Mantel 1967) was then performed. These analyses were worked out using the NTSYS.PC package Version 2.11Q 2000–2003 (Rohlf 1993).

Population structure was then examined on the whole data set using the assignment method implemented in the software STRUCTURE version 2.2 (Pritchard et al. 2000). The method infers the K number of clusters (i.e. populations) present in a sample, by comparing the posterior probability for different numbers of putative populations specified by the user, and assigns individuals to these clusters, giving the percentage of membership (q). Twenty independent runs for each K (from 1 to 20) were performed using 30,000 MCMC repetitions and 30,000 burn-in periods, using no prior information and assuming correlated allele frequencies and admixture. The number of clusters (K) was estimated by computing the ad hoc statistic ΔK, based on the rate of change in the log probability of the data between successive K values (Evanno et al. 2005, Supplementary Fig. 1). Based on the statistic ΔK, the number of clusters was set at 2

because this number maximised the ΔK parameter (Evanno et al. 2005), and a run at 800,000 MCMC repetitions and 200,000 burn-in periods was performed.

The output of the software gives the percentage of membership for the K clusters for each individual. The average percentage of membership (q) for each of the inferred K clusters of the two groups of accessions (European and Mesoamerican) were then computed. A genotype was assigned to a cluster when its percentage of membership was $\geq$0.70 and it was defined as an 'admixed' genotype when its percentage of membership was $\leq$0.69.

Possible substructures that might have been hidden by the main population structure were also explored using the methods described above. Two groups, one with 148 European individuals and the other with 94 Mesoamerican and 12 European individuals, were analysed separately.

Moreover, to test the phylogenetic relationships within the whole data set, the PowerMarker software v3.25 (Liu and Muse 2005) was used to create a UPGMA tree using C.S. Chord distance (Cavalli-Sforza and Edwards 1967), 100 bootstrapped, then, to display a coloured tree according to the STRUCTURE analysis results, the software Phylip v3.68 (Felsenstein 1989) was used to create a consensus tree, using the Majority Rule and the Dendroscope software (Huson et al. 2007).

Results

The developed core collection

Stratification reduced the collection from 678 to 228 accessions belonging to *P. coccineus* (Table 2). The collection includes 148 European LRs, 52 Mesoamerican LRs and 28 Mesoamerican WFs.

More detailed information, including information on the 14 *P. dumosus* LRs, is reported in Supplementary Table 1.

Genetic diversity among population groups

In the collection, 11 of the 12 loci studied were polymorphic (locus AZ301498 was monomorphic). A total of 167 alleles were detected in the *P. coccineus* samples (Table 1). The allelic number per locus ranged from 5 alleles in AZ301513 to 29 alleles in X80051 with an average number of 13.8 alleles per locus. In particular, three loci, X74919, AY298744, and X80051, had more than 20 alleles (Table 1).

Distributions of diversity parameters were always significantly different (U values always above 34473, $P \leq 0.001$) when the *P. coccineus* European was compared to the Mesoamerican LR gene pool (Table 3). Notably, private alleles characterised each group and, despite the outbreeding mating system, the inbreeding coefficient values (f) were high in all of the groups. This could be explained by an inbreeding quota in *P. coccineus* that was higher than presumed or by a low number of individuals per accession that was used during the propagation/regeneration procedures of the donor genebanks.

Relationships between populations and accessions

The greatest genetic distance was between the European LRs and the *P. dumosus* LRs ($D = 0.28$), followed by the distance between the European LRs and WFs ($D = 0.22$), and the European LRs and Mesoamerican LRs ($D = 0.20$), while the distance between the Mesoamerican LRs and WFs ($D = 0.08$) or *P. dumosus* ($D = 0.09$) was small.

Table 2 The *P. coccineus* collection obtained from the stratification process: number of landraces (LRs) and wild forms (WFs) for each geographic area and each country (total number of accessions: 228)

European country of origin	N of accessions studied	Mesoamerican country of origin	N of accessions studied
LRs		*LRs*	
Austria	9	Mexico	36
Georgia	11	Honduras	8
Germany	4	Costa Rica	8
Hungary	7		
Italy	41	Total	52
Poland	7		
Portugal	8		
Romania	10	*WFs*	
Slovak Republic	10		
Slovenia	5	Mexico	22
Spain	31	Honduras	2
The Netherlands	5	Guatemala	4
Total	148	Total	28

Table 3 Total number of alleles (N_a), average number of alleles per locus (N_{ma}), effective number of alleles (N_e), number of alleles with frequency greater than 0.05 (N_a Freq $\geq$ 0.05), number of private alleles, observed heterozygosis (H_o), expected heterozygosis (H_e), coefficient of inbreeding (f) and average allelic richness (AR), for each *P. coccineus* group of accessions (average values on 200 samples each made of 28 individuals, in brackets the values observed on the true European and Mesoamerican LR samples)

Accession group	European LRs	Mesoamerican LRs	Mesoamerican WFs
N_a	48 (91)	77 (101)	77
N_{ma}	4.4 (8.3)	7.0 (9.2)	7.0
N_e	25.3 (27)	40.7 (44)	45
N_a Freq. $\geq$ 5%	27.2 (27)	41.5 (39)	48
N private alleles	14 (36)	25 (28)	29
H_o	0.12 (0.12)	0.14 (0.14)	0.16
H_e	0.36 (0.37)	0.54 (0.55)	0.50
F	0.68 (0.69)	0.74 (0.74)	0.69
AR	3.9	6.2	6.2

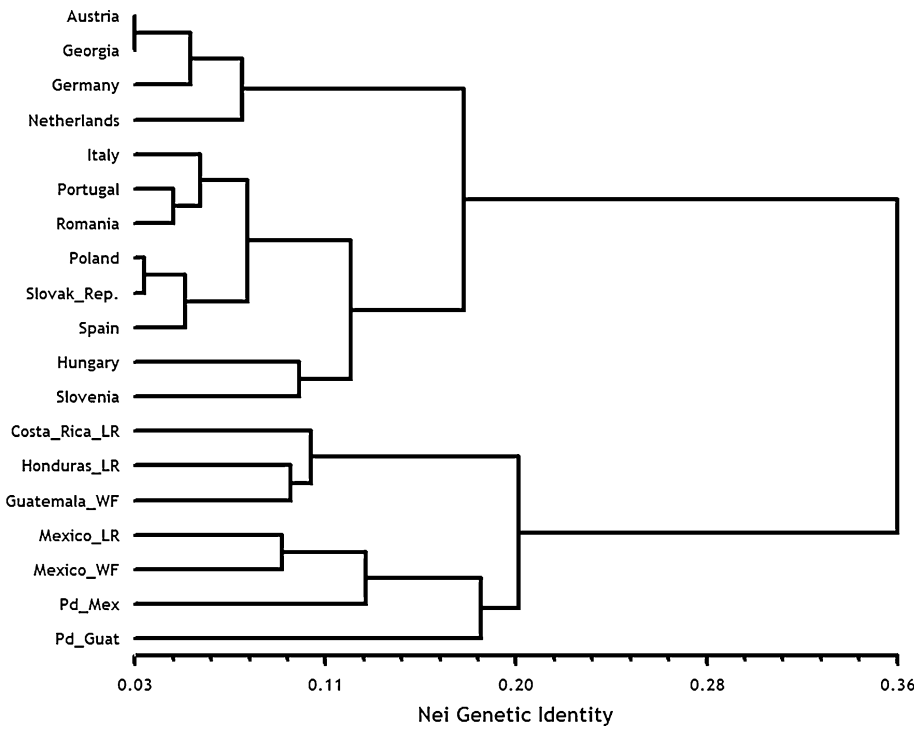

Fig. 1 Relationships (based on Nei's genetic distances) among accessions grouped according to country of origin. (*Pd P. dumosus*)

In addition to a clear distinction between European and Mesoamerican accessions and the lack of a clear distinction between *P. dumosus* and *P. coccineus* LRs and WFs, the cluster analysis based on Nei's genetic distance between groups of accessions from different countries (Fig. 1) also shows a subdivision of the European LRs cluster into sub-clusters. The Iberian, Italian, Romanian, Slovak and Polish accessions are grouped together and are separated from the Hungarian and Slovenian accessions, on one side, and from the Austrian, Georgian, German and Dutch accessions on the other. It is worth noting that Italian and Spanish accessions are located in different sub-sub-clusters. In addition, the PCoA (Supplementary Fig. 2) also shows that some Italian and Iberian accessions are mixed in with the Mesoamerican LRs and WFs.

The result of the Mantel test on the Nei's genetic distance matrix and the geographic distance matrix between accessions was highly significant ($r = 0.57$, $P < 0.01$) and showed a direct relationship between the genetic and geographic distances.

In the STRUCTURE analysis, the log_e *likelihood* values over 20 runs for K values (ranging from 1 to 20) increased progressively as K increased. The number of clusters (K) that maximised the ΔK parameter was $K = 2$ (Supplementary Fig. 1). On the basis of the method of Evanno et al. (2005), all of the analysed genotypes could then be

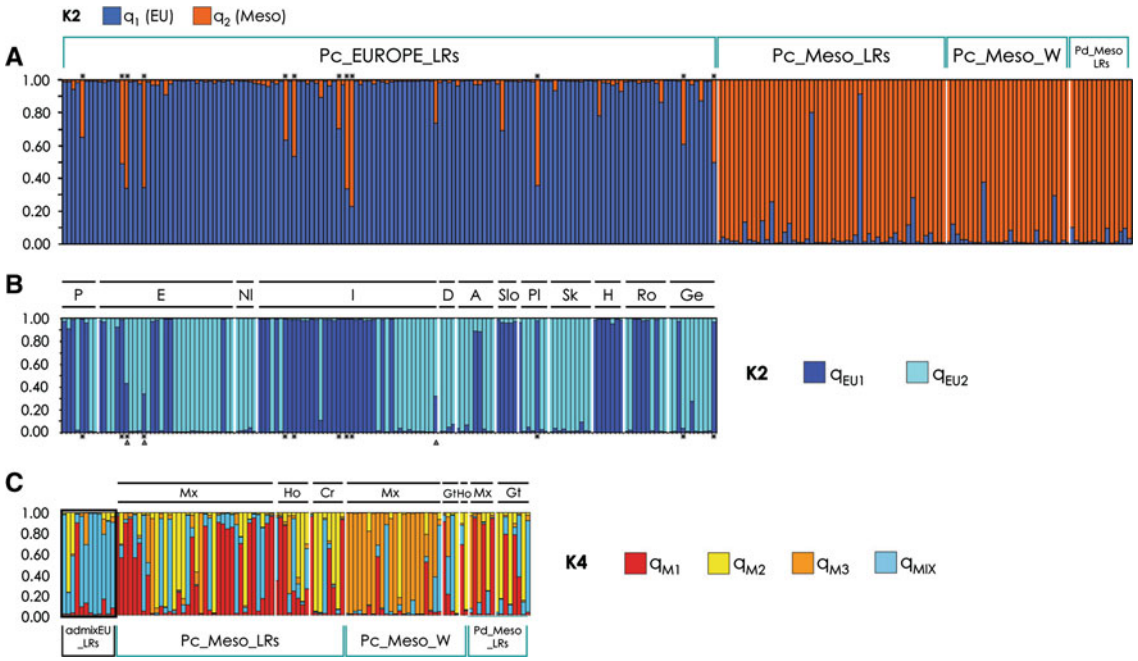

Fig. 2 Population structure as inferred by STRUCTURE v2.2 and 12 SSR markers data set; each individual is represented by a vertical line, partitioned into K coloured segments that represent the estimated membership of each individual. **a** Whole data set analysis (242 genotypes); black squares indicate European 'admixed' genotypes at $K = 2$. **b** Independent cluster analysis for the European genotypes (148); grey triangles indicate 'admixed' genotypes. **c** Independent cluster analysis for the Mesoamerican and 12 'admixed' European genotypes (106) previously identified at $K = 2$. Countries are abbreviated as follows: *A* Austria, *D* Germany, *E* Spain, *Ge* Georgia, *H* Hungary, *I* Italy, *Nl* The Netherlands, *P* Portugal, *Pl* Poland, *Ro* Romania, *Sk* Slovakia, *Slo* Slovenia, *Cr* Costa Rica, *Gt* Guatemala, *Ho* Honduras, *Mx* Mexico. Other abbreviations: *Pc P. coccineus*, *Pd P. dumosus*, *LRs* landraces, *W* wild, *EU* European, *Meso* Mesoamerican

split into $K = 2$ groups (hereafter referred to as clusters 1 and 2).

The average percentages of membership (q_1 for cluster 1, and q_2 for cluster 2) for both the European and Mesoamerican groups of accessions in each of the two clusters were computed. The European group (average $q_1 = 0.94$) was assigned to cluster 1 (EU) and the Mesoamerican group, including *P. dumosus* (average $q_2 = 0.94$), to cluster 2 (MESO). This analysis (Fig. 2a, see also Suppl. Table 2) clearly discriminated between the European and Mesoamerican groups (Wilcoxon–Kruskal–Wallis test, $P < 0.0001$). However, a few exceptions were found. When considering the percentages of membership for each European individual, 11 European genotypes, with q_1 values ranging from 0.34 to 0.65, were considered 'admixed' and one European individual with a q_1 value of 0.23 was assigned to cluster 2 (MESO). For the Mesoamerican individuals, it was possible to discriminate one 'admixed' genotype ($q_2 = 0.62$) and two individuals assigned to cluster 1 (EU), with a high q_1 value ($q_1 = 0.91$ and 0.80).

The analysis performed only on the 148 European LRs also highlighted a structure at $K = 2$ (hereafter referred to as clusters EU1 and EU2) (Fig. 2b).

The EU1 group was made up of 63 individuals (average $q_{EU1} = 0.98$) and the EU2 group of 82 individuals (average $q_{EU2} = 0.98$). Only three genotypes were 'admixed': one belonging to an Italian LR ($q_{EU1} = 0.32$) and two belonging to Spanish LRs ($q_{EU1} = 0.34$ and 0.43). No clear geographic structure was detected, confirming the UPMGA results: only the Hungarian and Slovenian LRs were completely assigned to cluster EU1 and the German, Dutch and Slovak LRs to cluster EU2. However, in these cases, the number of individuals analysed for each country was low, ranging from 4 to 10. For all the other countries, genotypes were assigned to both clusters (EU1 and EU2).

In the STRUCTURE analysis performed with the 94 Mesoamerican genotypes and the 12 European genotypes that were previously identified as 'admixed' or assigned to the cluster MESO, the number of clusters that maximised the ΔK parameter was $K = 4$ (M1, M2, M3 and MIX cluster, Fig. 2c). Cluster M1 consisted of 26 individuals (average $q_{M1} = 0.88$), mainly *P. coccineus* and *P. dumosus* LRs, including one wild genotype (from Guatemala) and one European genotype (from Spain). Cluster M2 consisted of 21 individuals (average $q_{M2} = 0.89$), mainly *P. coccineus* and *P. dumosus* LRs, with three wild genotypes (from Mexico, Guatemala and Honduras) and one European genotype (from Spain). Cluster M3 consisted of 21 individuals (average $q_{M3} = 0.93$), mainly wild *P. coccineus* genotypes (from Mexico), with three *P. coccineus* LRs

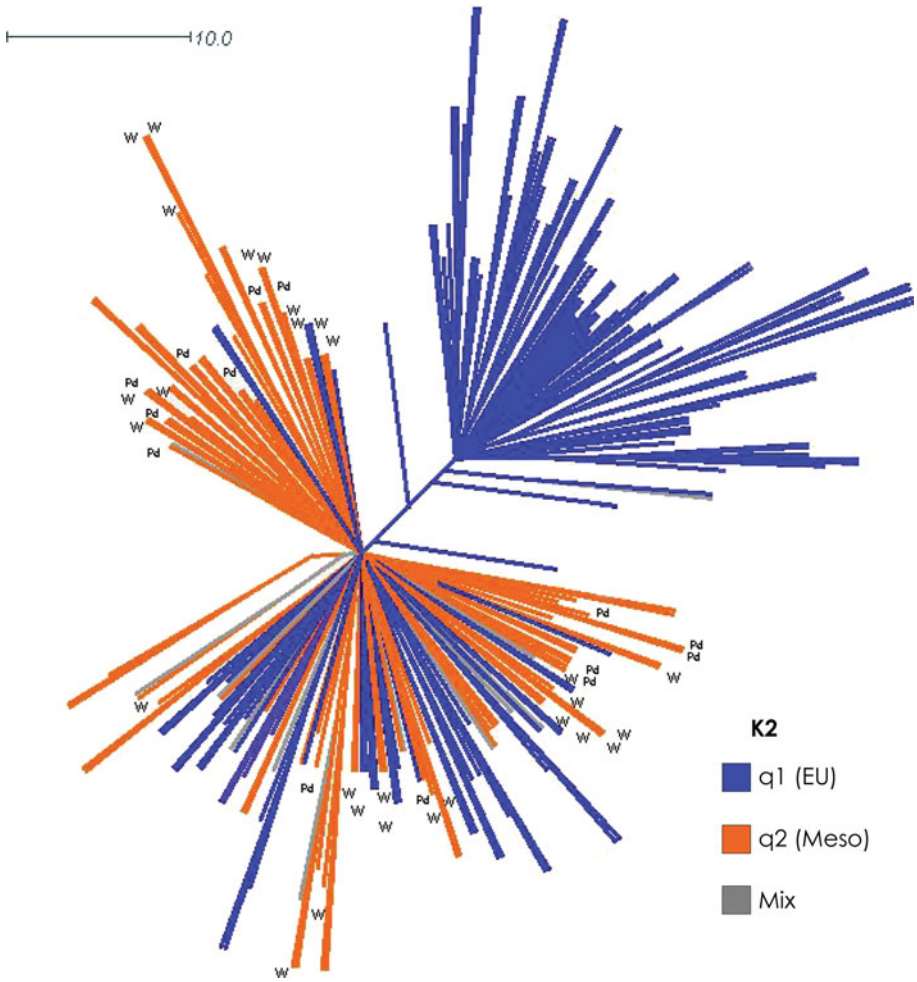

Fig. 3 UPGMA radial phylogram for all the *P. coccineus* and *P. dumosus* genotypes analysed in this study (data set of 12 SSR markers). The genotypes are coloured on the basis of the STRUCTURE analysis percentage of membership. *W* Mesoamerican Wild, *Pd P. dumosus*

(two from Mexico and one from Honduras) and two *P. dumosus* LRs (from Mexico). Cluster MIX consisted of 20 genotypes (average $q_{MIX} = 0.90$): 7 European LRs, 3 wild genotypes (two from Mexico and one from Guatemala) and 9 LRs (from Mexico, Honduras and Costa Rica) and one *P. dumosus* LR (from Guatemala). The remaining 18 individuals were 'admixed' (average $q_{M1} = 0.33$, $q_{M2} = 0.18$, $q_{M3} = 0.16$ and $q_{MIX} = 0.33$).

The phylogenetic relationships among the 242 accessions (12 SSR markers) are shown in Fig. 3 as a radial UPGMA phylogram. The genotypes are coloured based on the STRUCTURE assignments at $K = 2$.

Two main groups were distinguished. In the first group, only European LRs were present (most of the genotypes assigned to the EU population and one 'admixed' genotype from Georgia, $q_1 = 0.61$). In the second group, all the Mesoamerican (*P. dumosus* and *P. coccineus* cultivated and wild) forms were present, but some European and 'admixed' genotypes were also included. In this group, part of the *P. coccineus* WFs were closely associated with each other and with a number of LRs, while others were loosely connected with another group of numerous cultivated forms.

Discussion

This study developed and described a worldwide collection that includes accessions from the centre of origin and from the secondary introduction areas. As such, it provides a picture of the *P. coccineus* genetic diversity distribution and outcomes after its introduction into the Old Word, which was not available before. To the best of our knowledge, the genetic outcome after the introduction and dispersal of a crop from an American centre of origin into Europe has only been addressed in the case of maize (Rebourg et al. 2003) and *P. vulgaris* (Angioi et al. 2010).

The SSR molecular markers developed on *P. vulgaris* and used in this work were also able to amplify and produce polymorphic patterns in *P. coccineus*. Besides offering a number of reliable markers to be used in future studies, these results confirm that the species are closely related.

A high level of diversity was detected both in the Mesoamerican forms (even though a relatively limited sample of WFs and LRs was assessed in this study) and in the European LRs. Considering the importance that wild relatives of crops and LRs have with respect to facing the

challenges posed by current climatic and socio-economic changes and new breeding requirements, such diversity should be carefully preserved both in situ and ex situ.

The comparison of diversity level and STRUCTURE analysis suggest that a limited number of runner bean LRs were presumably introduced into Europe. However, the diversity level of European LRs is not so low as to claim a strong bottleneck effect. Similar conclusions were reached for *P. vulgaris* based on allozyme (Santalla et al. 2002) and cpSSR evidence (Angioi et al. 2010). The latter study showed no difference between America and Europe.

It should also be noted that, in contrast to what was observed using RFLP in maize, another allogamous crop of American origin (Rebourg et al. 2003), a relatively high number of private alleles was found in the European in comparison with the Mesoamerican LRs [14 (36) vs. 25 (28), while Rebourg et al. (2003) reported 21 vs. 102 private alleles for the two groups, respectively]. This indicates ongoing runner bean differentiation in Europe.

The obtained results also show that the Mesoamerican and European accessions clearly belong to distinct gene pools. The European continent could then be regarded as a secondary diversification centre for *P. coccineus*, similarly to what was recently suggested for *P. vulgaris* (Angioi et al. 2010). The differentiation of the European gene pool can be due to selective pressures that lead to adaptation to the new environment, genetic drift, differential gene flow with respect to Mesoamerica and lack of introgression from wild forms.

It can be noted that the areas where the crop was initially introduced (i.e. the Iberian and, immediately afterwards, the Italian peninsula; Perale 2001) have quite different LRs and that only a subset of them was probably later introduced into other countries (i.e. The Netherlands, Germany, Slovakia, Hungary and Slovenia). However, it was not possible to trace the crop movements across Europe by looking at genetic data so as to ascribe the structure observed in the European gene pool (subdivided in EU1 and EU2 subgene pools) to different times or routes of introduction into the Old Continent.

The results also clearly show that the cultivated and wild Mesoamerican gene pools are scarcely differentiated, i.e. LRs are closely related to WFs. These findings confirm previously obtained results (Llaca et al. 1994; Escalante et al. 1994) and indicate that the actual gene flow between domesticated and wild forms of *P. coccineus* in the centre of origin is high and continuous. *P. coccineus* WFs made a relevant contribution to the shaping of the cultivated Mesoamerican gene pool.

In the STRUCTURE analysis, most of the examined Mesoamerican LRs closely resemble wild genotypes from Guatemala and Honduras, while only a few resemble wild Mexican forms. This suggests that either *P. coccineus* domestication events took place in that area or two domestication events took place (in Guatemala–Honduras and Mexico, separately) followed by extensive hybridization with cultivated forms from Guatemala and Honduras. A clearer clue to the domestication area could have been obtained if a wider wild sample had been available for examination.

The examined cultivated forms of *P. dumosus* are not distinct from the *P. coccineus* gene pool. The two species, formerly identified as subtaxa of the same species (Chacón et al. 2007; Delgado-Salinas et al. 2006; Freytag and Debouck 2002; Delgado-Salinas et al. 1999; Schmit and Debouck 1991), are certainly closely related. Crosses between them are highly (although not totally) fertile (Maréchal et al. 1978) and natural hybrids are observed (Freytag and Debouck 2002). Then, introgression blurring differences between them could have occurred. In addition, the similarity between the nuclear DNA of *P. dumosus* and *P. coccineus* detected in this study supports the hypothesis that *P. dumosus* originated from a cross that involved *P. vulgaris* as the maternal parent, with successive backcrosses from *P. coccineus* as the paternal donor (Angioi et al. 2009; Delgado-Salinas et al. 1999; Llaca et al. 1994; Schmit et al. 1993; Piñero and Eguiarte 1988).

Acknowledgments G. Spataro and B. Tiranti made equal contributions to this study and so should be considered as joint first authors. This work was funded by the Italian Ministry for University and Scientific Research (project 2002075423, Beans in Europe: structure of genetic diversity in *P. coccineus* for SSR). Thanks are due to the following Institutions for kindly providing materials and/or information: the International Centre for Tropical Agriculture (CIAT), Colombia; the United States Department of Agriculture (USDA), USA; the Institute for Plant Genetics and Crop Plant Research (IPK), Germany; the Instituto de Investigaciones Agropecuarias (INIA), Chile; the Banco Portugues De Germoplasma Vegetal (BPGV), Portugal; the Suceava Genebank (SVGB), Romania; the Plant Breeding Department Pontevedra (INIA), Spain; the Agricultural Institute of Slovenia (AIS), Slovenia; the Research Institute of Crop Production, Czech Republic; the Plant Breeding and Acclimatization Institute, Poland; the Institute of Agrobotany (IA), Hungary; the Università degli Studi di Torino, Italy; the Università degli Studi di Udine, Italy; the Università degli Studi della Tuscia, Italy (UNITUS); the Agenzia Regionale per lo Sviluppo e Innovazione nel Settore Agricolo e Forestale (ARSIA), Florence, Italy. Thanks are also due to Dr. O. Porfiri and all the farmers who donated their LRs.

References

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

Angioi SA, Rau D, Attene G, Nanni L, Bellucci E, Logozzo G, Negri V, Spagnoletti Zeuli PL, Papa R (2010) Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. Theor Appl Genet 121:829–843. doi:10.1007/s00122-010-1353-2

Beaver JS, Osorno JM (2009) Achievements and limitations of contemporary common bean breeding using conventional and

molecular approaches. Euphytica 168:145–175. doi: 10.1007/s10681-009-9911-x
Blair MW, Pedraza F, Buendia HF, Gaitan-Solis E, Beebe SE, Gepts P, Tohme J (2003) Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 107:1362–1374
Busogoro JP, Jijakli MH, Lepoivre P (1999a) Identification of a novel source of resistance to angular leaf spot disease of common bean within the secondary gene pool. Plant Breed 118:417–423
Busogoro JP, Jijakli MH, Lepoivre P (1999b) Virulence variation and RAPD polymorphism in African isolates of *Phaeoisariospis griseola* (Sacc.) Ferr., the causal agent of angular leaf spot of common bean. Eur J Plant Path 105:559–569
Campion B (1995) "Venere" and "Alarico", new scarlet runner bean (*Phaseolus coccineus* L.) cultivars with determinate growth habit. HortScience 30:1483–1484
Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J Hum Genet 19:233–257
Chacón SMI, Pickersgill B, Debouck DG, Salvador Arias J (2007) Phylogeographic analysis of the chloroplast DNA variation in wild common bean (*Phaseolus vulgaris* L.) in the Americas. Plant Syst Evol 266:175–195
Delgado-Salinas A (1988) Variation, taxonomy, and germplasm potentialities in *Phaseolus coccineus*. In: Gepts P (ed) Genetic resources of *Phaseolus* beans. Kluwer, Dordrecht, pp 441–463
Delgado-Salinas A, Turley T, Richman A, Lavin M (1999) Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). Syst Bot 24:438–460
Delgado-Salinas A, Bibler R, Lavin M (2006) Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. Syst Bot 31:779–791
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
Escalante A, Coello G, Eguiarte LE, Piñero D (1994) Genetic structure and mating systems in wild and cultivated populations of *Phaseolus coccineus* and *P. vulgaris* (Fabaceae). Am J Bot 81:1096–1103
Evanno G, Reganut E, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620
Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166
Freytag GF, Debouck DG (2002) Taxonomy, distribution and ecology of the genus *Phaseolus* (Leguminosae-Papilionideae) in North America, Mexico and Central America. Sida, Botanical Miscellany, N.23, Botanical Research Institute of Texas, Fort Worth
Gepts P (1996) Origin and evolution of cultivated *Phaseolus* species. In: Pickersgill B, Lock JM (eds) Advances in legume systematics. Part 8: Legumes of economic importance. Royal Botanic Gardens, Kew, pp 65–74
Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Available from http://www2.unil.ch/popgen/softwares/fstat.htm. Accessed 1 March 2010
Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52:577–586
Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R (2007) Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatic 8: 460. Software freely available from http://www.dendroscope.org. Accessed 1 March 2010
Kaplan L, Lynch TF (1999) *Phaseolus* (Fabaceae) in archaeology: AMS radiocarbon dates and their significance for precolombian agriculture. Econ Bot 53:261–272
Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129
Llaca V, Delgado SA, Gepts P (1994) Chloroplast DNA as an evolutionary marker in the *Phaseolus vulgaris* complex. Theor Appl Genet 88:646–652
Mantel NA (1967) The detection of disease clustering and generalized regression approach. Cancer Res 27:209–220
Maréchal R, Mascherpa J, Stainier F (1978) Étude taxonomique d'un groupe complexe d'espèces des genres *Phaseolus* et *Vigna* (Papilionaceae) sur la base données morphologiques et polliniques, traitées par l'analyse informatique. Boissiera 28:1–273
Negri V, Tosti N (2002) *Phaseolus* genetic diversity maintained on farm in Central Italy. Gen Res Crop Evol 49:511–520
Nei M (1972) Genetic distance between population. Am Naturalist 106:283–292
Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
Papa R, Nanni L, Sicard D, Rau D, Attene G (2006) Evolution of genetic diversity in *Phaseolus vulgaris* L. In: Motley TJ, Zerega N, Cross H (eds) Darwin's harvest: new approaches to the origins, evolution and conservation of crops. Columbia University Press, New York, pp 121–142
Peakall R, Smouse PE (2006) GENEALEX: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295
Perale M (2001) Milacis cultus aperire paramus. "De milacis cultura" di Pierio Valeriano - il primo testo europeo dedicato al fagiolo, Momenti Aics, Belluno
Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12:844–855
Piñero D, Eguiarte L (1988) The origin and biosystematic status of *Phaseolus coccineus* subsp. polyanthus: electrophoretic evidence. Euphytica 37:199–203
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure from multilocus genotype data. Genetics 155:945–959
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. doi:10.1007/s00122-002-1140-9
Rodiño AP, Lema EM, Marlene PB, Santalla M, De Ron AM (2007) Assessment of runner bean (*Phaseolus coccineus* L.) germplasm for tolerance to low temperature during early seedling growth. Euphytica 155:63–70. doi:10.1007/s10681-006-9301-6
Rohlf FJ (1993) NTSYS.PC. Numerical Taxonomy and Multivariate Analysis System, Version 2.11Q. Applied BiostatisticsInc, New York
Santalla M, Rodino P, Ron De (2002) Allozyme evidence supporting southwestern Europe as a secondary centre of diversity for the common bean. Theor Appl Genet 104:934–944
Santalla M, Monteagudi AB, González AM, De Ron AM (2004) Agronomical and quality traits of runner bean germplasm and implications for breeding. Euphytica 135:205–215
Schmit V, Debouck DG (1991) Observations on the origin of *Phaseolus polyanthus*, Greenman. Econ Bot 45:345–364
Schmit V, du Jardin P, Baudoin JP, Debouck DG (1993) Use of cloroplast DNA polymorphisms for the phylogenetic study of seven *Phaseolus* taxa including *P. vulgaris* and *P.coccineus*. Theor Appl Genet 87:506–516
Sicard D, Nanni L, Porfiri O, Bulfon D, Papa R (2005) Genetic diversity of *Phaseolus vulgaris* L. and *Phaseolus 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

Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. WH Freeman, San Francisco

Sokal RR, Rohlf FG (1995) Biometry, The principles and practice of statistics in biological research. WH Freeman and Company, New York

Southern EM (1979) Measurement of DNA length by gel electrophoresis. Anal Biochem 100:319–323

Welsh MM, Grafton KF (2001) Resistance to common bacterial blight of bean introgressed from *Phaseolus coccineus*. HortScience 36:750–751

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

Zeven AC (1996) Results of activities to maintain landraces and other material in some European countries in situ before 1945 and what we may learn from them. Gen Res Crop Evol 43:337–341