© NIAB 2010
ISSN 1479-2621

Plant Genetic Resources: Characterization and Utilization (2011) 9(1); 86–96
doi:10.1017/S1479262110000420

Genetic diversity of Colombian landraces of common bean as detected through the use of silver-stained and fluorescently labelled microsatellites

Lucy M. Díaz[1], Héctor F. Buendía[1], Myriam C. Duque[1] and Matthew W. Blair[1,2]*

[1]*CIAT – International Center for Tropical Agriculture, AA6713, Cali, Colombia, South America and* [2]*Cornell University, Department of Plant Breeding, Ithaca, New York, USA*

Abstract

Colombia, situated at the northern end of the Andes mountains of South America and in proximity to Central America, is an important centre of diversity for common bean (*Phaseolus vulgaris* L.) that has a mix of cultivated germplasm from both major gene pools (Andean and Mesoamerican) for the species. Microsatellites are a useful marker system for analyzing genetic diversity of this crop and can be analyzed with manual (silver-stain) or automated (ABI) detection systems and using unlabelled or fluorescently labelled markers, respectively. The objectives of this research were to evaluate the genetic diversity of 92 Colombian landraces and gene pool controls with 36 fluorescent and 30 non-fluorescent microsatellite markers and to determine the extent of introgression between the Andean and Mesoamerican gene pools for this germplasm. A comparison of fluorescent *versus* non-fluorescent marker systems was performed with 14 loci, which were evaluated with both methods; the fluorescent markers were found to be more precise than the non-fluorescent markers in determining population structure. A combined analysis of 52 microsatellites using the 36 fluorescent markers and 16 non-overlapping, silver-stained markers produced an accurate population structure for the Andean gene pool that separated race Nueva Granada and race Peru genotypes and clearly identified introgression between these races and the gene pools. The results of this research are important for the application of microsatellite markers to diversity analysis in common bean and for the conservation of landraces in Colombia and neighbouring countries of Latin America, where similar germplasm exists and where gene pool or race mixtures also occur.

Keywords: genetic diversity indices; population structure; principal coordinate analysis; simple sequence repeats

Introduction

Common bean (*Phaseolus vulgaris* L.) is the third most important food legume in the world after soybean and peanut (Singh, 1999; Broughton *et al.*, 2003). Its high nutritional quality, elevated protein content and low cost make common bean a staple in the diet of many Latin Americans. Beans are also appealing for their diversity in seed form, colour and size; eating texture; and capacity to adapt to different climatic conditions (Kornegay and Cardona, 1991). Molecular markers indicate two principal gene pools of common bean, whose centres of origin and domestication are in the Andes and Mesoamerica/Central America (Gepts *et al.*, 1986; Becerra Velásquez and Gepts, 1994). Using morphological and biochemical characteristics, Singh *et al.* (1991a, b)

* Corresponding author. E-mail: m.blair@cgiar.org

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

subdivided the gene pools into six races: Nueva Granada (NG), Peru (P) and Chile (Andean gene pool), and Durango, Jalisco and Mesoamerica (Mesoamerican gene pool). With molecular markers, Beebe *et al.* (2000) also distinguished a fourth race in the Mesoamerican gene pool that they named Guatemala, which was later confirmed to be distinct from other Mesoamerican beans by Blair *et al.* (2009) using a larger group of genotypes. At least, three of these races (P, NG, Mesoamerica (M) and possibly Guatemala) overlap in northern South America, where a long history of bean cultivation has encouraged the creation of a diverse set of landraces (Gepts and Bliss, 1986).

Among the countries of northern South America or nestled in the Andes mountains, Colombia is the largest producer of common beans, with over 120,000 ha planted per year. Beans along with maize, cassava and several native root crops form the basis of the diet, and along with coffee and fruit species, they are of utmost importance to the small-farm economy, particularly in the medium altitudes of the three 'cordillera' mountain ranges (Ligareto, 1991). Given the wide area across which beans are grown in these mountains, Colombian landraces of beans are adapted to many altitudes and water regimes. The following departments of Colombia produce beans: Antioquia, Santander and Santander del Norte, Huila, Tolima, Boyacá, Cundinamarca and Nariño, in descending order of production (Ríos-Betancour and Quirós-Dávila, 2002). Some reduction in bean-growing area has been observed, although *per* hectare yields have generally increased. During 2002–2006, production fell to around 100,000 tons due to violence in the countryside but has been higher in subsequent years according to the Ministry of Agriculture and Rural Development (FAOSTATS).

Archaeological evidence of common beans in Colombia has been found among other crop remains especially in highland regions and intervening valleys (Morcote-Ríos, 2006). Pre-Colombian beans have been recorded in the Sabana de Bogotá in deposits near the University of Tunja dating from 1000 AC. Meanwhile, in the Valle del Cauca, beans may be even more ancient having been associated with the Yotocó and Calima cultures (100 BC to 1200 AC). Finally, other beans dating from 1115 AC were discovered in the Department of Risaralda. Iconography of common beans in archaeological remains is not as common as in other parts of South America where pre-Incan and Incan civilizations were predominant, but this may be due to the greater emphasis placed on animal *versus* plant iconography in the mega-diverse region of the Northern Andes compared to the drier and less diverse areas to the south.

Colombian landraces are highly heterogeneous for their seed form, colour and size, as well as for plant growth characteristics (Ríos-Betancour and Quirós-Dávila, 2002). This heterogeneity probably originated from farmers' preferences in different growing regions. Along with landraces, wild accessions of *P. vulgaris* are also reported from the Eastern Cordillera in the Departments of Santander, Norte de Santander, Boyacá and Cundinamarca (Gepts and Bliss, 1986; Tohme *et al.*, 1996). Hybrid swarms of wild and weedy types are known to exist in certain forest regions of the country (Beebe *et al.*, 1997). Tohme *et al.* (1996) analyzed the molecular structure of a core collection of wild beans using early generation molecular markers, and concluded that Colombia was a minor centre of diversity for the species along with centres in Mesoamerica, Ecuador and Northern Peru as well as in the Southern Andes. Debouck (1996) argued that genes from wild beans are assimilated into intermediate forms and landraces, and suggested that many landraces may have derived in part from wild sympatric populations. Colombian diversity within the Andean gene pool was reported to be high by Blair *et al.* (2007) with blurring between the NG and P races. Furthermore, because of its strategic location, Colombia has been a place of considerable exchange and hybridization between the two principal gene pools (Gepts and Bliss, 1986).

Microsatellite markers (simple sequence repeats, SSRs) have lately been used to show that both Andean and Mesoamerican beans possess high genetic diversity (Díaz and Blair, 2006; Blair *et al.*, 2007; Asfaw *et al.*, 2009; Blair *et al.*, 2009; Kwak and Gepts, 2009; Rossi *et al.*, 2009). SSRs have also been used to distinguish race Mesoamerica and the Durango–Jalisco complex of the Mesoamerican gene pool (Díaz and Blair, 2006), and races Chile, NG and P in the Andean pool (Blair *et al.*, 2007; Becerra *et al.*, 2010). SSR markers can determine genetic structure more precisely than other types of markers (Liu *et al.*, 2003). While some authors have used non-specific fluorescent dyes based on M13 labelling with common bean microsatellites (Kwak and Gepts, 2009; Oblessuc *et al.*, 2009), we have developed a group of specific, fluorescent dye-labelled microsatellites that have proved to be a highly efficient tool for analyzing allelic diversity in beans (Blair *et al.*, 2009), given that the estimate of allele sizes is not biased by the effect of the M13 label.

The objectives of this study, therefore, were (1) to better understand the variability present in Colombian landraces; (2) to determine the extent of inter-gene pool hybridization in the creation of Colombian landraces; (3) to evaluate the population structure found with fluorescently labelled microsatellite primers *versus* unlabelled primers such as those used in our previous studies of Andean or Mesoamerican genotypes from the northern Andes (Blair *et al.*, 2007) and (4) to compare allele diversity for 14 microsatellites evaluated with both labelled and unlabelled primers.

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

Materials and methods

Genotypes and DNA extraction

As a first step in the analysis of Colombian landraces, we selected 92 genotypes from the FAO collection of common bean genotypes in the genebank held under the responsibility of the CIAT genetic resource programme (GRP) along with four control genotypes Diacol Calima, DOR364, ICA Pijao and G19833, which had been used by Blair *et al.* (2007). These included all of the Andean genotypes evaluated by Blair *et al.* (2007) plus a set of Mesoamerican genotypes found at http://www.generationcp.org/gcp_central_registry excluding G4550, G7314, G8160, G12702, G14644, G19142B and G21227. Seeds for the analysis were multiplied by the bean programme from a genetic stock held by the GRP, while the GRP provided passport, morphological and phaseolin data on each accession analyzed (http://isa.ciat.cgiar.org/urg). For the DNA extraction, three seeds with the same seed morphology and colour within each accession were selected for each genotype and germinated on a square of germination paper (30 × 25 cm) laid over cotton wool that was previously wetted with distilled water. The seeds were incubated for 3 or 4 d, and young trifoliate leaves were removed from each seedling. Total DNA was extracted from this leaf tissue ground in liquid nitrogen and using the cetyl trimethylammonium bromide extraction method as described by Afanador *et al.* (1993). Agarose gel electrophoresis was then carried out to determine the quality of the extracted DNA. Concentrations were quantified in a DNA Quant™ 200 Fluorometer (Hoefer Pharmacia Biotech, Inc., San Francisco, CA, USA), and dilutions were made to 10 ng/ml for further use.

Fluorescent dye-labelled and unlabelled microsatellites

The allelic diversity of the genotypes was evaluated using 52 microsatellites specific to common beans, 36 of which came from a kit of fluorescent dye-labelled markers distributed in nine panels as described in Blair *et al.* (2009), while the other 16 microsatellites were unlabelled (Tables 1 and 2). Fourteen of the fluorescently labelled markers were also run as unlabelled markers for a total of 66 individual SSR evaluations (36 fluorescent and 30 non-fluorescent microsatellite markers). For the fluorescent markers, each SSR was placed into a fluorescent panel consisting of four markers each carrying a specific fluorescent tag of 6-FAM (blue), NED (yellow), PET (red) or VIC (green) dyes attached to the 5′ end of the forward primer according to the supplementary table in Blair *et al.* (2009). Of the 52 microsatellites, 31 were based on non-gene-coding sequences (genomic SSRs), and 21 were based on gene-coding regions (genic SSRs). All the microsatellites were previously selected according to their high indices of polymorphism information content (PIC), large number of alleles, even genetic diversity distribution across linkage groups in the genome and good amplification signal (Blair *et al.*, 2003, 2006, 2009). PCR conditions were similar for both labelled and unlabelled microsatellites with reactions including 1 × PCR buffer (10 mM Tris–HCl, pH 8.8; 50 mM KCl;

Table 1. Genetic diversity summary statistics for Colombian common bean landraces using 36 labelled microsatellites

Markers	Linkage group[a]	N_a	H_e[b]	H_o	PIC
AG01	3	4	0.662	0.082	0.590
BM139	2	12	0.832	0.024	0.816
BM140	4	10	0.723	0.071	0.701
BM141	9	13	0.875	0.047	0.863
BM143	2	22	0.919	0.047	0.914
BM149	ND	4	0.505	0.024	0.397
BM156	2	22	0.910	0.073	0.903
BM160	7	21	0.870	0.098	0.854
BM172	3	5	0.248	0.000	0.235
BM175	5	12	0.808	0.024	0.789
BM183	7	11	0.594	0.047	0.570
BM187	6	25	0.911	0.188	0.905
BM188-A	9	7	0.473	0.131	0.418
BM188-B	9	15	0.789	0.120	0.767
BM200	1	20	0.812	0.012	0.799
BM201	1	6	0.486	0.024	0.461
BM205	7	6	0.669	0.012	0.615
BMd01	3	9	0.809	0.333	0.780
BMd02	2	2	0.392	0.012	0.315
BMd08	ND	4	0.566	0.000	0.486
BMd15	4	4	0.551	0.247	0.479
BMd16	4	6	0.490	0.036	0.436
BMd17	2	3	0.457	0.012	0.405
BMd18	2	5	0.526	0.000	0.422
BMd20	5	7	0.638	0.063	0.597
BMd46	9	3	0.434	0.000	0.358
BMd47	2	4	0.472	0.024	0.390
BMd51	ND	2	0.024	0.000	0.023
BMd56	ND	3	0.069	0.024	0.068
GATs54	ND	2	0.189	0.000	0.171
GATs91	2	13	0.801	0.100	0.784
PV-ag003	1	5	0.461	0.012	0.396
PV-at001	4	41	0.946	0.094	0.943
PV-at003	4	2	0.499	0.155	0.375
PV-cct001	ND	4	0.360	0.000	0.320
PV-ctt001	4	7	0.753	0.073	0.730
Average		9.47	0.598	0.061	0.558

ND, linkage group not determined; N_a, number of alleles; N_o, number of observations made; H_o, observed heterozygosity; PIC, polymorphism information content.
[a] Linkage group according to Blair *et al.* (2003). [b] Gene diversity according to Nei and Li (1979).

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

Table 2. Genetic diversity summary statistics for Colombian common bean landraces using 16 unlabelled microsatellites to complement those analyzed in Table 1

Markers	Linkage group[a]	N_a	H_e[b]	H_o	PIC
BM142	2	4	0.418	0.000	0.348
BM152	2	16	0.869	0.024	0.855
BM155	5	3	0.319	0.000	0.290
BM164	2	7	0.783	0.000	0.754
BM167	2	13	0.824	0.048	0.803
BM170	6	10	0.690	0.047	0.669
BM184	11	5	0.525	0.000	0.489
BM189	8	6	0.712	0.000	0.662
BMc5	ND	2	0.463	0.024	0.356
BMd10	1	4	0.452	0.000	0.402
BMd12	6	5	0.551	0.000	0.513
BMd26	4	2	0.420	0.012	0.332
BMd33	11	5	0.529	0.024	0.417
GATS11B	10	3	0.461	0.035	0.365
PV-ag001	11	7	0.648	0.012	0.594
PV-gaat001	4	3	0.514	0.035	0.442
Average		5.94	0.574	0.016	0.518

ND, linkage group not determined; N_a, number of alleles; N_o, number of observations made; H_o, observed heterozygosity; PIC, polymorphism information content.
[a] Linkage group according to Blair *et al.* (2003). [b] Gene diversity according to Nei and Li (1979).

and 0.1% Triton X-100), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 nM of each primer and 1 U of Taq polymerase (Promega Corporation, Madison, WI, USA). However, for the unlabelled SSRs, we used 0.25 mM dNTPs.

The PCRs of the labelled SSRs were carried out individually for each pair of primers. The temperatures for the cycling profile were similar for the labelled and unlabelled microsatellites. For the unlabelled ones, the programme began with 92°C for 3 min; followed by 34 cycles of denaturation at 92°C for 30 s, annealing at 50–65°C (depending on the melting temperature for each primer pair) for 30 s and extension at 72°C for 45 s. For the labelled SSRs, only 28 cycles of denaturation at 95°C for 40 s, annealing at 55°C for 40 s and extension at 72°C for 1 min were used to moderate signal strength. To visualize the amplification products, two techniques were used: namely polyacrylamide gels at 4% and silver staining for unlabelled microsatellites according to Blair *et al.* (2006); and capillary separation on an ABI PRISM 3730 fragment analyzer (Applied Biosystems, Foster City, CA, USA) for fluorescently labelled microsatellites according to Blair *et al.* (2009). The PCR products of the labelled SSRs were run in the facilities of the Biotechnology Unit of Cornell University, while the unlabelled SSRs were evaluated in the Bean Genetic Characterization Laboratory of CIAT. For the labelled microsatellites, the molecular weight of the bands was estimated by comparing to an internal molecular weight standard GeneScan™ 500 LIZ® using GeneMapper® Software v.3.7 (both from Applied Biosystems), while for unlabelled markers, the allele sizes were determined through comparison to a 10-bp molecular weight standard (Invitrogen Corp., Carlsbad, CA, USA). Alleles from labelled microsatellites were binned using AlleloBin software from ICRISAT (http://www.icrisat.org/gt-bt/biometrics.htm), while unlabelled alleles were compared to those of controls Diacol Calima and G19833 (both Andean) and DOR364 and ICA Pijao (both Mesoamerican).

Data analysis

To assess the relationships between accessions, we first used a similarity index for shared alleles estimated with SAS v.9.1.3 software (SAS Institute, 1996), based on the formula: $PS = \sum PS_i/n$, where 'PS_i' equals the proportion of shared alleles at locus i and 'n' equals the total number of loci. A principal coordinate analysis (PCoA) was then carried out with SAS v.9.1.3 as well to graph the individuals in three-dimensional space and to determine the study samples' population structure comparing this to results from STRUCTURE v.2 software (Pritchard *et al.*, 2000) based on an 'admixture model', 50,000 burn-ins and 100,000 iterations. We also determined common parameters of genetic diversity such as the percentage of polymorphic loci; genetic diversity, which is also described as expected heterozygosity (H_e); proportion of heterozygous individuals observed in the population (H_o); and PIC. All these parameters were estimated using PowerMarker v.3.0 (Liu and Muse, 2005; http://www.powerMarker.net) software and determined for each group of genotypes discovered in the genetic similarity/PCoA. Wright statistics and analysis of molecular variance (AMOVA) were calculated with Arlequin v.3.11 (http://cmpg.unibe.ch/software/) software. Linkage disequilibrium (LD) between the markers was determined for the set of non-overlapping 52 polymorphic SSRs with the dedicated procedure found in TASSEL (http://www.maizegenetics.net/) software using 1,000 permutations.

Results

Allelic polymorphism for the complete dataset

Analysis of allele polymorphism was conducted for 36 fluorescent microsatellites and 16 non-fluorescent, non-overlapping microsatellites as described in Tables 1 and 2. Of these markers, seven loci have not been assigned to linkage groups. The other 45 markers represented every linkage group of the common bean

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

genetic map presented by Blair *et al.* (2003) with 4.1 loci/linkage group. In terms of diversity, this group of microsatellite loci revealed a total of 436 alleles with all markers polymorphic and the number of alleles ranging from 2 to 41 per locus with an average of 8.2. PIC values ranged from 0.023 to 0.914 with an average of 0.545. Meanwhile, H_e ranged from 0.024 to 0.946 with an average of 0.590, and observed heterozygosity was generally low (0.000 to 0.333 with an average of 0.0474).

As might be expected, the total number of alleles found with the automated system using the labelled SSR markers was higher than that reported for the manual system using the unlabelled SSR markers (341 *vs.* 95). Similarly, the average number of alleles/marker (9.47 *vs.* 5.94 for the two marker types, respectively), H_e (0.598 *vs.* 0.574) and PIC values (0.558 *vs.* 0.518) were also contrasting. This difference between the labelled markers and unlabelled marker results may be related to the number of markers used in each category or to the distribution of genomic and gene-based markers in each group of markers since both marker types were used in the study.

Among all the markers used, the highest diversity was found for PV-at001 with 41 alleles and an H_e of 0.946, confirming findings by Blair *et al.* (2009). Markers with intermediate to high diversity included those with a number of alleles equal to or higher than 20 including BM187 (25), BM143 and BM156 (22 alleles each), BM160 (21) and BM200 (20) each with an H_e that was equal or higher than 0.900. In contrast, none of the unlabelled SSRs presented H_e levels above 0.900, and the best unlabelled markers were BM152 with a PIC value of 0.869 and 16 alleles or BM167 with a PIC value of 0.824 and 13 alleles. Other unlabelled SSRs that stood out were BM170 (ten alleles) and PV-ag001 (seven alleles).

Markers with a low number of alleles (two) were found with both types of SSRs, for example BMc5 and BMd26 among the unlabelled markers, and BMd2, BMd51, GATs54 and PV-at003 among the labelled markers, with BMd51 being the least informative marker in terms of H_e ($H_e = 0.024$ and PIC value = 0.023). In total, 67% of the markers analyzed had an H_e level between 0.500 and 1.000. In terms of observed heterozygosity, both marker types generally detected low levels of heterogeneity or heterozygosity; however, labelled SSRs more often detected two bands from a genotype. This was especially the case for BMd01 (H_o of 0.333), BMd15 (0.247) and BM187 (0.188); where BMd01 has been observed before to be a multilocus marker (Blair *et al.*, 2006). In contrast, none of the unlabelled markers presented observed heterozygosity indices of more than 0.048.

Further comparison of the labelled versus *the unlabelled SSRs*

Among the 52 microsatellite loci that were analyzed, 14 were repeated with both techniques to establish the advantages and disadvantages of the two methodologies as indicated in Table 3. Compared to the labelled SSRs, the unlabelled SSRs presented a similar but slightly higher total (149 *vs.* 135) and average (10.6 *vs.* 9.6) number of alleles. H_e was similar for both repetitions of the same markers (0.70 and 0.66, respectively) as were PIC values (0.67 and 0.63, respectively). One apparent difference was that the labelled SSRs identified higher observed heterozygosity than the unlabelled SSRs (0.06 *vs.* 0.03). Despite these small differences, the correlation between the number of alleles was positive with a highly significant ($P < 0.000$) r-value (0.908). Missing data accounted for only 2.8% of genotype-by-marker combinations for labelled SSRs and only 0.2% for silver-stained alleles.

Among the individual markers, the unlabelled amplifications of BM143, BM175, BMd15, PV-ctt001, GATs54 and GATs91 presented a larger number of alleles than the corresponding labelled SSRs, probably because the interpretation of allele bands was more difficult when visualizing the alleles in silver-stained gels than with automated detection. In some cases, labelled SSRs showed a larger number of alleles than their unlabelled counterparts: as was found for BM140, BM160, BM183 and BM201. This could have been a result of the automated allele calling found in the ABI detection system, although all data from the capillary system were checked for allele differences.

Population structure and principal coordinates analysis

According to the results of the 52 non-overlapping SSRs and the PCoA shown in Fig. 1, the Colombian common bean landraces were classified into three principal groups. These three groups were confirmed by the dendrogram and STRUCTURE analyses at $K = 3$ shown in Figs 2 and 3 for both total and unlabelled *versus* labelled SSRs, respectively. A total of five subgroups were identified in the IML module of SAS v.9.1.3 explaining 82% of variation. The principal groups for the complete dataset included the Mesoamerican gene pool, represented by race M, and the Andean gene pool, represented by races NG and P. Both of the Andean races could be subdivided into two subgroups each (NG1, NG2, P1 and P2, respectively) resulting in a total of five subgroups, among which subgroup

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

Table 3. Comparison of genetic diversity parameters of labelled SSRs with unlabelled SSRs

	Unlabelled SSRs					Labelled SSRs				
Markers	N_a	N_o	H_e[a]	H_o	PIC	N_a	N_o	H_e	H_o	PIC
BM139	12	85	0.82	0.02	0.80	12	82	0.83	0.02	0.82
BM140	8	83	0.68	0.07	0.66	10	85	0.72	0.07	0.70
BM143	24	85	0.93	0.06	0.93	22	85	0.92	0.05	0.91
BM160	20	85	0.90	0.05	0.89	21	82	0.87	0.10	0.85
BM175	15	85	0.88	0.01	0.87	12	84	0.81	0.02	0.79
BM183	10	85	0.53	0.02	0.51	11	85	0.59	0.05	0.57
BM201	5	85	0.66	0.01	0.60	6	85	0.68	0.01	0.62
BMd15	6	85	0.60	0.07	0.55	4	85	0.55	0.25	0.48
BMd17	3	85	0.42	0.01	0.37	3	85	0.46	0.01	0.41
BMd20	7	85	0.60	0.02	0.57	7	80	0.64	0.06	0.60
PV-ag003	5	85	0.70	0.00	0.65	5	84	0.46	0.01	0.40
PV-ctt001	13	85	0.85	0.06	0.84	7	82	0.76	0.07	0.73
GATs54	3	85	0.33	0.00	0.30	2	85	0.19	0.00	0.17
GATs91	18	85	0.90	0.00	0.89	13	80	0.80	0.10	0.78
Total	149	–	–	–	–	135	–	–	–	–
Average[b]	10.64[C]	84.86[C]	0.70[C]	0.03[C]	0.67[C]	9.64[C]	83.50[D]	0.66[C]	0.06[C]	0.63[C]

N_a, number of alleles; N_o, number of observations made; H_o, observed heterozygosity; PIC, polymorphism information content.
[a] Gene diversity according to Nei and Li (1979). [b] *t*-test used with threshold of $P < 0.05$ for differences between the averages of labelled and unlabelled SSRs as distinguished by different letters for the same variable.

M contained 26 Mesoamerican accessions, while races NG and P had 21 and 38 genotypes, respectively.

The Mesoamerican genotypes had the characteristics typical of race M (Díaz and Blair, 2006), such as small (73.1%) and medium-sized (26.9%) seed; B (69.2%), S (23.1%), CH (3.9%) or CA (3.9%) phaseolin patterns and mostly type II (30.77%) and III (61.54%) growth habit with type I completely absent. There was no geographical focal point for Mesoamerican beans, as many were distributed in lower elevations sites

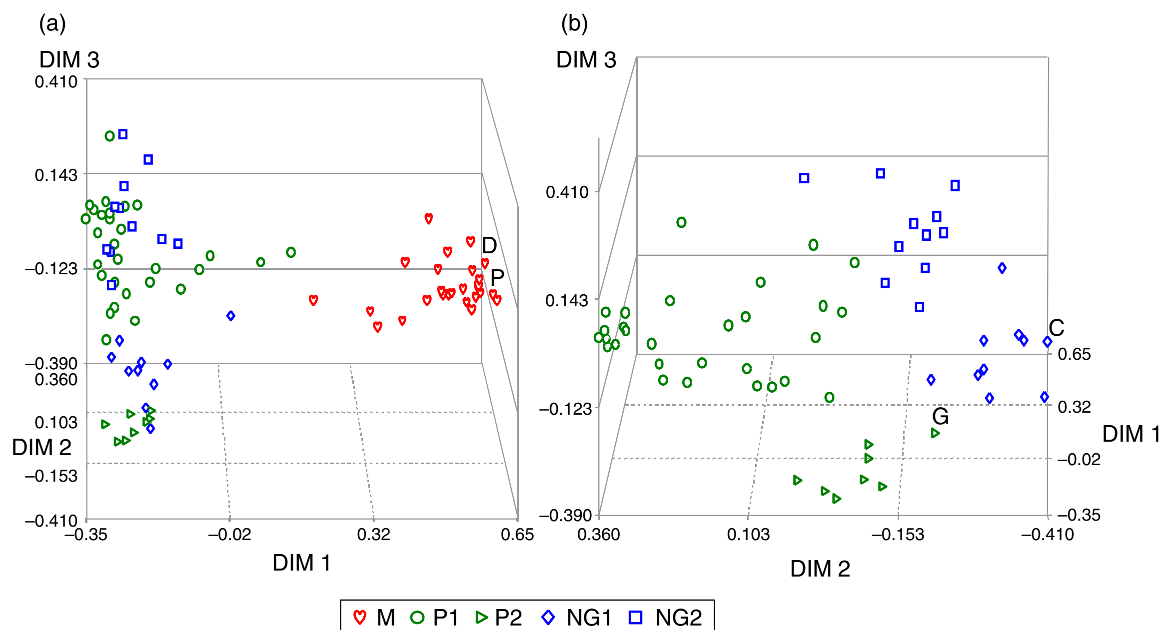

Fig. 1. Principal coordinate analysis of Colombian common bean landraces analyzed with 52 non-redundant SSRs and their division into gene pools and races. (a) Andean and Mesoamerican gene pools shown together and (b) only Andean gene pool shown alone. Identification of controls in study by initials: (G) G19833, (C) Diacol Calima, (D) DOR364 and (P) ICA Pijao. M, shown as red hearts; NG, subgroups 1 and 2 shown as blue diamonds and squares, respectively; P, subgroups 1 and 2 shown as dark green circles and isosceles triangles, respectively (A colour version of this figure can be found online at journals.cambridge.org/pgr).

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

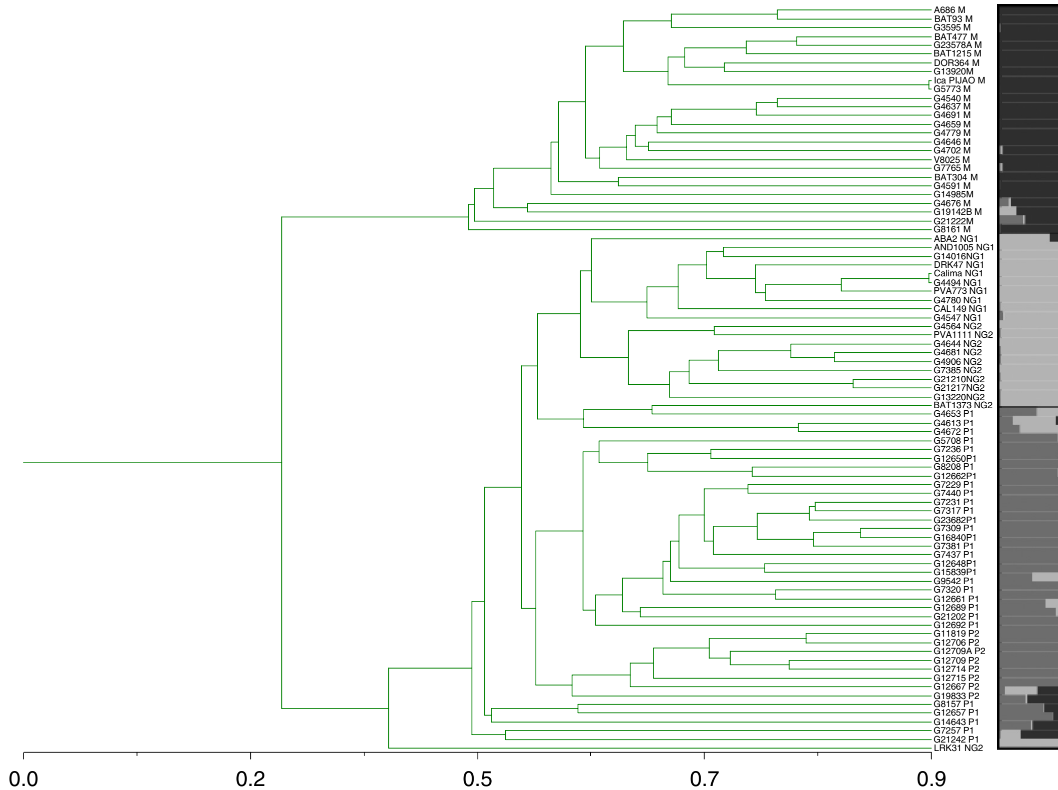

Fig. 2. Dendrogram and graph of population structure ($K = 3$) showing the relationship of Colombian common bean landraces. Different colours of races are shown as follows: M, top colour; NG, middle colour; and P, bottom dark colour with introgression found for certain individuals at the bottom of the dendrogram. Same colours are used for each race as indicated in Fig. 1 (A colour version of this figure can be found online at journals.cambridge.org/pgr).

around the country both in the northern coast (Guajira department) and mainly in the interior central and southern valleys (Departments of Antioquia, Boyaca, Cauca, Cundinamarca, Huila, Nariño, Risaralda, Tolima and Valle de Cauca).

Within the majority of Andean gene pool, we found that subgroup P1 had 30 accessions, mostly of T (40.00%) and H (36.67%) type phaseolin with a few showing introgression of S or C phaseolin. Almost 90% of the these genotypes possessed large seed size and type IV growth habit (83.33%), followed by type I (10.00%) or type III (6.67%) habits, with a total absence of type II growth habit. The accessions of this group originated mostly in the highland Departments of Antioquia, Cauca, Cundinamarca, Nariño and Boyacá, with only one genotype each from the Departments of Caldas and Huila (G7929 and G4653, respectively).

Subgroup P2, meanwhile, had eight accessions which were characterized by T (62.50%) and H (37.50%) phaseolin patterns and viny growth habits (87.50% type IV and 12.50% type III). These accessions were mainly from the Departments of Antioquia and Nariño and included G19833, which is a representative of race P (Blair *et al.*, 2007). Clustering along with the subgroup were genotypes such as G12715 (Sañudo 52), G14643 (Tunia 1), G7229 (Culateno), G7381

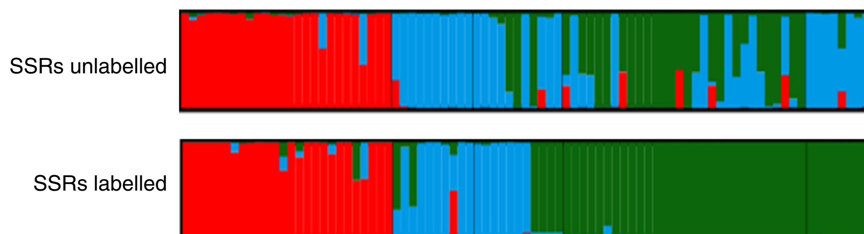

Fig. 3. Population structure of Colombian common bean landraces evaluated with 30 unlabelled SSRs and with 36 fluorescently labelled SSRs (A colour version of this figure can be found online at journals.cambridge.org/pgr).

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

(X15943), G7437 (Huevo de Pinché), G5708 and G7309 (both Sangretoro), and G7231 (Cargamento).

Two subgroups were also observed for race NG: with the NG1 subgroup having ten accessions, of which five were advanced lines of Andean origin developed at CIAT (ABA2, PVA 773, AND 1005, CAL 149 and DRK 47); and three were from the Departments of Antioquia or Cauca. This group was characterized by large (90%) or medium-sized (10%) seeds, T (90%) and H (10%) phaseolin patterns, and type I (60%), II (30%) and III (10%) growth habits. Also appearing in this group was Diacol Calima, a representative of race NG. Subgroup NG2 was composed of 11 genotypes, which presented Andean characteristics such as T (63.6%), CA (27.3%) and H (9.1%) phaseolin patterns, type I growth habit and large seed size. They originated from the Departments of Nariño (three), Antioquia (two), and Cundinamarca, Tolima and Valle (one accession each), while two accessions had no department ID. All the advanced lines with T phaseolin stood out for being located within race NG (both groups 1 and 2). Genotypes ABA2, PVA773, PVA1111 and LRK31 were previously classified by Voysest *et al.* (1994) within this race, although Blair *et al.* (2007) later reported that only 75% of the advanced lines analyzed were found within this race.

Diversity between groups

Groups M and P1 presented high and similar indices of H_e, number of alleles and PIC values; however, race P subgroup 1 presented higher observed heterozygosity (Table 4). Meanwhile, race M had the largest percentage of polymorphic loci (96.2% or 50 out of 52 loci). The lowest percentage of polymorphic loci was present in race P subgroup P2 (65.4%) with 18 of the 52 SSRs used monomorphic. Overall, marker PV-at001 provided high indices for all races. However, some SSRs were predominant for each given race. For example, for race P, high indices of diversity were observed for SSRs BM143, BM167 and BM156; while for race NG, GATs91, BM200 and BM141 were more highly polymorphic. For race M, the highest levels were seen for BM200 and BM187.

Large genetic differentiation was found among the races of genotypes of Colombian origin ($G_{ST} = 0.334$), with the highest effect of non-random pairing within the populations ($G_{IS} = 0.888$) and very low genetic flow among populations. Among the races, the greatest genetic differentiation was found between groups M and P2 ($G_{ST} = 0.311$), and the lowest between Andean groups P1 and NG2 ($G_{ST} = 0.125$). Gene flow was high between those groups having low genetic differentiations with at least one individual migrating per generation. In contrast, gene flow was very low between the Andean and Mesoamerican gene pools overall.

Genetic structure was confirmed through an AMOVA with the groups established for the Colombian landraces was found to be significantly different, and most of the variation was explained by interaction between individuals within populations (data not shown).

Analysis of linkage disequilibrium

Significant LD (at $P < 0.05$) was found for only 7.2% of the 1378 possible pairwise combinations of 52 SSR markers used in the diversity analysis. However, when comparing pairs of markers from the same linkage group, the percentage of significant LD values was 52.2%. The range in D' values between marker pairs ranged from 0.044 to 1.00, with an average of 0.580 and a range of r^2 between 0.000 and 0.810. LD was more highly significant for a larger number of marker comparisons in the Andean gene pool than in the Mesoamerican gene pool (Fig. 4).

Table 4. Genetic diversity parameters of groupings of Colombian landraces of common bean and genetic differentiation based on F_{ST} values

	Genetic diversity parameters						G_{ST} values				
Races	N	N_a	H_e[a]	H_o	PIC	P (%)	M	NG1	NG2	P1	P2
M	26	5	0.45	0.04	0.41	96.15	0.000				
NG1	10	3	0.31	0.04	0.28	75.00	0.469	0.000			
NG2	11	3	0.38	0.04	0.35	84.62	0.465	0.255	0.000		
P1	30	5	0.43	0.06	0.40	94.23	0.418	0.271	0.200	0.000	
P2	8	3	0.33	0.03	0.30	65.38	0.476	0.301	0.336	0.219	0.000

N, number of landraces; P (%), percentage of polymorphic loci; N_a, number of alleles; H_o, observed heterozygosity; PIC, polymorphism information content; P(%), percentage polymorphism.
[a] Gene diversity according to Nei and Li (1979).

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

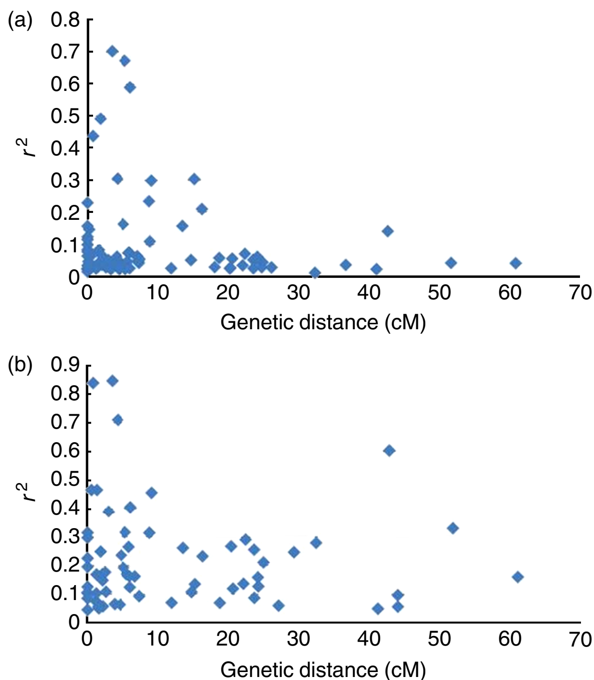

Fig. 4. Linkage disequilibrium among all pairs of 52 microsatellite loci in the (a) Andean and (b) Mesoamerican gene pools of Colombian common bean landrace (A colour version of this figure can be found online at journals.cambridge.org/pgr).

Discussion

Previous studies have reported the utility of fluorescently labelled microsatellite markers in detecting high levels of genetic diversity and well-defined population structure in common bean before both for a large set of cultivated genotypes (Blair *et al.*, 2009; Asfaw *et al.*, 2009; Blair *et al.*, 2010) and for wild *versus* domesticated genotypes (Kwak and Gepts, 2009). Advantages of microsatellite markers are self-evident in that they are highly polymorphic, co-dominant and widely distributed in the genome. This study, similar to our previous work, showed that the markers can be evaluated with various technologies giving similar results in terms of genetic differentiation (Blair *et al.*, 2006, 2009, 2010). An advantage of labelled microsatellites over unlabelled microsatellites was the potential for automatic detection and calling of alleles as well as the multiplexing of several markers together in a single run through the use of fluorescent panels as used here and in Blair *et al.* (2009). Care must be given with automated allele calling since we found that labelled SSRs were slightly higher in both observed and H_e, which can indicate superior sensitivity of the ABI3730x in reading of the alleles or perhaps overcalling of distinct alleles *versus* those that can be distinguished manually. Silver-stained detection of unlabelled SSR alleles has been found to be highly accurate, as long as samples are run on a single gel and repeat amplification and controls are used along with several molecular weight ladders to ensure accurate allele calling. So the issues of allele calling accuracy or lack of precision in that allele calling in either system must be carefully considered by researchers in the future. Meanwhile, the issue of single base pair allele differences is found with both detection systems.

Both marker types were found to be useful for the genetic diversity analysis of the Colombian landraces validating the fluorescent microsatellite panels and silver-stained microsatellites. This study includes the highest number of microsatellite markers (66 independent runs) evaluated to date on a single sample of diversity and thus was a rich source of interpretation of marker and genotype differences. In terms of genetic diversity, our study of both types of SSR markers confirmed that the Andean genotypes of Colombian origin are highly diverse as was found by Blair *et al.* (2007). Furthermore, the PCoA found clustering patterns that agreed with characteristics such as phaseolin, seed size and growth habit. These characteristics corresponded to races previously established by Singh *et al.* (1991a, b).

Compared to our previous study of Colombian landraces in Blair *et al.* (2007), here we also analyzed genotypes with Mesoamerican phaseolin patterns and found that the Mesoamerican group formed a compact cluster with characteristics of race M. The introduction of race M to Colombia may have originated from active genetic exchange through trade or hybridization with wild and weedy forms that exist in the country and may have generated genetic variability (Gepts and Bliss, 1986; Beebe *et al.*, 1997). Taking into account that phaseolin B was found in Colombian genotypes of the Mesoamerican gene pool and phaseolin T was found in genotypes of the Andean pool, we may assume that the former phaseolin type is of local origin and has probably been introgressed into landraces over long periods from a wild source.

A possible hypothesis is that the Colombian genotypes inherited ancestral characteristics such as phaseolin (shared by wild and cultivated forms in the same region without being altered by domestication), which then combined with modern characteristics (modified under local selection, whether intentionally or unintentionally). Furthermore, a single genotype (G4691 or 'Matahambre') had the phaseolin type CH, previously found in wild beans of Colombia (Koening *et al.*, 1990).

Similar results from Gepts and Bliss (1986) in their study of phaseolins of wild and cultivated common beans from Colombia suggested that domestication may also have occurred in Colombia, as well as in Mesoamerica and the southern Andes. Subsequent analysis with amplified fragment length polymorphism, α-amylase inhibitors and lectin confirmed the presence of unique alleles and atypical patterns for wild Colombian genotypes (Tohme *et al.*, 1996).

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

In addition, several other studies have found relationships between wild and domesticated common beans (Gepts and Bliss, 1986; Islam *et al.*, 2002) and common exchange of phaseolin alleles between gene pools (Blair *et al.*, 2007, 2009) from this region. For example, within each gene pool group for the Colombian landraces, we identified some genotypes with phaseolin type of the opposite gene pool. These included the Andean genotypes G19142B (Criollo), G4676 (Revoltura) and G21222 (Bola Maní Morado or Piojillo), which had B phaseolin patterns, and the Mesoamerican genotypes G14643, which had the CA phaseolin pattern.

With regards to racial identification, genotypes BAT93, BAT477, G4540, G4779, G13920, G23578A, A686, BAT1215, G3595, G4591, G4637, G14985 and V8025 clustered in the group belonging to race M, confirming the previous classification by Beebe *et al.* (2000) and Singh *et al.* (1991a, b). Meanwhile, genotypes G4659, G12650, G19142 and G21222, described by Beebe *et al.* (2000) as belonging to race Guatemala, did not cluster in our figures for either geographical origin or seed size, although they all belonged to the Mesoamerican group with the exception of G12650.

Introgression and admixture were also evident in the position and Q-coefficients for several genotypes including G7257 (Boyacá 82), G8157 (Revoltura), G12657 (Cauca 35) and G21242 (DGD-1399), all of which were intermediate between the gene pools in the PCoA. These genotypes have been identified as presenting between-gene pool introgression by previous authors as well (Beebe *et al.*, 2001; Islam *et al.*, 2004; Blair *et al.*, 2006).

The findings of introgressed materials in Colombia and that 39% of the genotypes studied presented phaseolin T strengthen the hypothesis that, in this area, the two principal gene pools of common beans came into contact, met and exchanged genes. Evidence was also present for gene flow from the southern part of the Andes, currently the countries of Argentina and Peru where CA and H phaseolin patterns are more common, respectively. In our study, the phaseolin CA was found equally in races NG and P, while phaseolin H was found almost exclusively in race P.

This study along with others showed the importance of using control genotypes for each gene pool and for the races. For example, the Andean controls Calima (G4494) and Chaucha Chuga (G19833) were found in different groups and represented races NG and P, respectively, as they had in Blair *et al.* (2007). Dorado (DOR364) and ICA Pijao meanwhile represented the same race within the Mesoamerican gene pool and were tightly clustered with that group of landraces. This situation has been consistent for the most part in other studies by Blair *et al.* (2009) and Becerra *et al.* (2010) except that when larger simples are used, the race structure and position of the control genotypes can be less clear. Therefore, a larger set of control genotypes should be developed with some specificity for each regional or national collection.

In conclusions, the joint use of labelled and unlabelled molecular markers (SSRs) enabled us to verify that Colombian landraces presented a clear and well-defined genetic structure. We confirmed the existence of genotypes that agreed with the race characteristics previously reported by Singh *et al.* (1991a, b) with type III growth habit standing out for the Mesoamerican genotypes, type I and II for race NG, and type I and IV for race P. Seed size presented the variability expected for Andean genotypes, and the pattern established was similar to that found for Andean races in Blair *et al.* (2007). Finally, given the high diversity of phaseolin patterns among the landraces, these could represent a centre of domestication that was later influenced by beans introduced from other centres of domestication with introgression between and within the gene pools playing a significant role in landrace development.

Acknowledgements

The authors thank Claritza Muñoz (CIAT) for DNA extraction and Sharon Mitchell (Cornell) for technical assistance. We also thank Daniel Debouck of the CIAT genetic resources programme for providing germplasm and Agobardo Hoysos of the Bean Project for seed multiplication and curation. This research was supported by the Generation Challenge Programme.

References

Afanador L, Hadley S and Kelly JD (1993) Adoption of a mini-prep DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L.). *Bean Improvement Cooperative* 36: 10–11.

Asfaw A, Blair MW and Almekinders C (2009) Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African Highlands. *Theoretical and Applied Genetics* 120: 1–12.

Becerra-Velásquez VL and Gepts P (1994) RFLP diversity of common bean (*Phaseolus vulgaris* L.) in its centres of origin. *Genome* 37: 256–263.

Becerra VL, Paredes M, Rojo C, Díaz LM and Blair MW (2010) Microsatellite marker characterization of Chilean common bean (*Phaseolus vulgaris* L.) germplasm. *Crop Science* 50: 1932–1941.

Beebe S, Toro O, Gonzalez AV, Chacón MI and Debouck D (1997) Wild–weed–crop complex of common bean (*Phaseolus vulgaris* L., Fabacea) in the Andes of Peru and Colombia, and their implications for conservation and breeding. *Genetic Resources and Crop Evolution* 44: 73–91.

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press

Beebe S, Renjifo J, Gaitán-Solís E, Duque MC and Tohme J (2001) Diversity and origin of Andean landraces of common bean. *Crop Science* 41: 854–862.

Beebe S, Skroch P, Tohme J, Duque MC, Pedraza F and Nienhhuis J (2000) Structure of genetic diversity among common bean landraces of middle American origin based on correspondence analysis of RAPD. *Crop Science* 40: 264–273.

Blair MW, Díaz LM, Buendia HF and Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 119: 955–973.

Blair MW, Gonzales LF, Kimani P and Butare L (2010) Intergenepool introgression, genetic diversity and nutritional quality of common bean (*Phaseolus vulgaris* L.) landraces from Central Africa. *Theoretical and Applied Genetics* 121: 237–248.

Blair MW, Diaz JM, Hidalgo R, Diaz LM and Duque MC (2007) Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 116: 29–43.

Blair MW, Giraldo MC, Buendia HF, Tovar E, Duque MC and Beebe S (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 113: 100–109.

Blair MW, Pedraza F, Buendia H, Gaitán-Solis E, Beebe S, Gepts P and Tohme J (2003) Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 107: 1362–1374.

Broughton WJ, Hernandez G, Blair MW, Beebe SE, Gepts P and Vanderleyden J (2003) Beans (*Phaseolus* spp.) – model food legumes. *Plant and Soil* 252: 55–128.

Debouck D (1996) Colombian common and lima beans; views on their and evolutionary significance. *Revista Corpoica* 1: 7–13.

Díaz LM and Blair MW (2006) Race structure within the Mesoamerican genepool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. *Theoretical and Applied Genetics* 114: 143–154.

Gepts P and Bliss F (1986) Phaseolin variability among wild and cultivated common bean (*Phaseolus vulgaris* L.) from Colombia. *Economy Botany* 40: 469–478.

Gepts P, Osborn T, Rashka K and Bliss F (1986) Phaseolin–protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris* L.): evidence for multiple centers of domestication. *Economy Botany* 40: 451–468.

Islam FM, Basford KE, Redden RJ, Gonzalez AV, Kroonenberg PM and Beebe SE (2002) Genetic variability in cultivated common bean beyond the two major gene pools. *Genetic Resources and Crop* 49: 271–283.

Islam FM, Beebe S, Muñoz M, Tohme J, Redden RJ and Basford KE (2004) Using molecular markers to assess the effect of introgression on quantitative attributes of common bean in the Andean gene pool. *Theoretical and Applied Genetics* 108: 243–252.

Kornegay J and Cardona C (1991) Breeding for insect resistance in beans. In: Schoonhoven A and Voysest O (eds) *Common Beans: Research for Crop Improvement*. Wallingford, UK: International CAB, pp. 619–648.

Koening R, Sing S and Gepts P (1990) Novel phaseolin types in wild and cultivated common bean *Phaseolus vulgaris*. *Economy Botany* 44: 50–60.

Kwak M and Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theoretical and Applied Genetics* 118: 979–992.

Ligareto GA (1991) Consideraciones generales sobre el cultivo de fríjol en Colombia. *ICA* 26: 1–25.

Liu K and Muse S (2005) PowerMarker: integrated analysis environment for genetic marker data. *Bioinformatics* 21: 2128–2129.

Liu K, Goodman M, Muse S, Smith JS, Buckler E and Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165: 2117–2128.

Morcote-Ríos G (2006) Tumbas y Plantas Antiguas del Surococcidente Colombiano. *Boletín Museo del Oro*. Bogotá: Banco de la República, p. 54.

Nei M and Li H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Procceding of the National Academy of Sciences of the United States of America* 76: 5269–5273.

Oblessuc PR, Campos T, Kupper Cardoso JM, Augusto Sforça D, Moro Baroni R, Pereira de Souza R and Benchimol LL (2009) Adaptation of fluorescent technique for genotyping with new microsatellite markers in common bean. *Pesquisa Agropecuária Brasileira* 44: 638–644.

Pritchard JK, Stehens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Ríos-Betancour MJ and Quirós-Dávila JE (2002) El fríjol (*Phaseolus vulgaris* L.): cultivo, beneficio y variedades. *Federacion Nacional de cultivadores de cereales (Fenalce)*: 141–151.

Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G and Papa R (2009) Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evolutionary Applications* 2: 504–522.

SAS Institute (1996) *SAS/STAT Users guide 6.11*. Cary, NC: SAS Institute.

Singh SP (1999) Production and utilization. In: Singh SP (ed.) *Common Bean Improvement in the Twenty-First Century*. London: Kluwer, pp. 1–24.

Singh S, Gepts P and Debouck D (1991a) Races of common bean (*Phaseolus vulgaris,* Fabaceae). *Economy Botany* 45: 379–396.

Singh S, Gutierrez A, Molina A, Urrea C and Gepts P (1991b) Genetic diversity in cultivated common bean. II. Marker-based analysis on morphological and agronomic traits. *Crop Science* 31: 23–29.

Tohme J, Gonzalez DO, Beebe S and Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. *Crop Science* 36: 1375–1384.

Voysest O, Valencia MC and Amezquita MC (1994) Genetic diversity among Latin American Andean and Mesoamerican common bean cultivars. *Crop Science* 34: 1100–1110.

https://doi.org/10.1017/S1479262110000420 Published online by Cambridge University Press