Theor Appl Genet (2006) 114:143–154
DOI 10.1007/s00122-006-0417-9

ORIGINAL PAPER

Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers

L. M. Díaz · M. W. Blair

Received: 17 May 2006 / Accepted: 24 September 2006 / Published online: 18 October 2006
© Springer-Verlag 2006

Abstract Common bean (*Phaseolus vulgaris* L.) cultivars are distinguished morphologically, agronomically and ecologically into specific races within each of the two gene pools found for the species (Andean and Mesoamerican). The objective of this study was to describe the race structure of the Mesoamerican gene pool using microsatellite markers. A total of 60 genotypes previously described as pertaining to specific Mesoamerican races as well as two Andean control genotypes were analyzed with 52 markers. A total of 267 bands were generated with an average of 5.1 alleles per marker and 0.297 heterozygosity across all microsatellites. Correspondence analysis identified two major groups equivalent to the Mesoamerica race and a group containing both Durango and Jalisco race genotypes. Two outlying individuals were classified as potentially of the Guatemala race although this race does not have a defined structure and previously classified members of this race were classified with other races. Population structure analysis with $K = 1$–4 agreed with this classification. The genetic diversity based on Nei's index for the entire set of genotypes was 0.468 while this was highest for the Durango–Jalisco group (0.414), intermediate for race Mesoamerica (0.340) and low for race Guatemala (0.262). Genetic differentiation (G_{ST}) between the Mesoamerican races was 0.27 while genetic distance and identity showed race Durango and Jalisco individuals to be closely related with high gene flow (N_m) both between these two races (1.67) and between races Durango and Mesoamerica (1.58). Observed heterozygosity was low in all the races as would be expected for an inbreeding species. The analysis with microsatellite markers identified subgroups, which agreed well with commercial class divisions, and seed size was the main distinguishing factor between the two major groups identified.

Keywords Population structure · Genetic differentiation · Gene flow · Races Durango (D) · Jalisco (J) · Mesoamerica (M) · Simple sequence repeats (SSR)

Communicated by D. A. Hoisington.

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0417-9 and is accessible for authorized users.

L. M. Díaz · M. W. Blair (✉)
Centro Internacional de Agricultura Tropical (CIAT),
Apartado Aéreo 6713, Cali, Colombia, South America
e-mail: m.blair@cgiar.org

Introduction

Common bean is the third most important grain legume in the world and is the most important food legume, produced over an area of 18 million hectares with large amounts of production in developing countries of Latin America and Eastern and Southern Africa (Broughton et al. 2003). Cultivated common beans originated in two centers of diversity giving rise to two genepools: Mesoamerican from Central America and Mexico and Andean from the Andes mountains of South America (Gepts et al. 1986; Koenig and Gepts 1989; Singh et al. 1991a, b, c; Becerra et al. 1994; Tohme et al. 1996; Beebe et al. 2000, 2001). Some authors also refer to the Mesoamerican gene pool as the Middle American gene pool but both terms refer to the same set of characteristics. The differences between Mesoamerican and Andean gene pools of common bean include seed size, phaseolin

(seed storage protein) patterns, plant morphology, isozymes and RFLP, RAPD or AFLP markers. Cultivated bean gene pools have further been divided into races according to morphological criteria and agroecological adaptation (Singh et al. 1991a, b) where the term 'race' is used to denote a group of related landraces (Gepts 1988). Members of each race have distinctive and specific physiological, agronomic, biochemical and molecular characteristics and differ from other races in the allelic frequencies at specific loci (Singh et al. 1991a, b, c). Race structure has been analyzed by RAPD markers (Beebe et al. 2000) but less molecular evidence has been accumulated for within gene pool differences as compared to between gene pool differences. In this study we will concentrate on the Mesoamerican (or Middle American) races as defined by Singh et al. (1991a) and Beebe et al. (2000) and as discussed below.

Due to its worldwide importance the first race of interest within the Mesoamerican gene pool is race Mesoamerica (hereafter abbreviated as M). Of all the Mesoamerican races, race M is the most widely grown and is represented by small black, small red, small white, Rosinha and Carioca seed classes among others. Cultivars of race M have relatively small seeds and are adapted to a range of hot, humid to moderate climates in the tropics and subtropics but are also grown in high latitudes in the United States and Argentina. Greatest production is in Brazil, Mexico and Central America with additional production in other countries such as Ethiopia and Venezuela. The predominant phaseolin type is 'S', but 'Sb' and 'B' are also found (Singh et al. 1991a). Through RAPD analysis by Beebe et al. (2000), the race was subdivided into subraces, reflecting plant architecture and seed type; with subrace M_1 composed mostly of small black seeded beans with type II growth habit, and subrace M_2 composed of diverse seed color classes most with a prostrate type III growth habit (Beebe et al. 2000).

A second race, race Durango (D) is originally from the semi-arid mountains of northern and central Mexico and southwestern USA and includes the Great Northern, Pinto, Bayo, Garrapato y Ojo de Cabra seed classes among others (Singh et al. 1991a; Voysest et al. 1994; Becerra and Gepts 1994; Beebe et al. 2000). Most genotypes from this race have indeterminate prostrate or climbing growth habit and S or Sd phaseolin. A third race as proposed by Singh et al. (1991a) is race Jalisco (J) and includes the seed classes Flor de Mayo and Flor de Junio that are from the highland Mexican states of Jalisco, Guanajuato, Michoacán, Oaxaca and Puebla mostly with S phaseolin. This race overlaps with race D in terms of seed color, growth habit and geographical origin and in molecular studies they are similar as well (Singh et al. 1991c; Beebe et al. 2000; McClean et al. 2004; Rosales-Serna et al. 2005; Chacón et al. 2005; Pallottini et al. 2004). Meanwhile, a fourth race, not identified by Singh et al. (1991a) was proposed to exist by Beebe et al. (2000) for accessions from Guatemala and the neighbouring Mexican state of Chiapas. This race, termed race Guatemala (G) was characterized by genotypes mostly with indeterminate climbing growth habit and small seed size similar to race M.

Microsatellites are hypervariable, PCR-amplified genetic loci surrounding simple sequence repeats that vary in their central repeat motif (Powell et al. 1996). Microsatellite markers have been used with great success to effectively and rapidly estimate the genetic variability within and between populations, and between samples of accessions of cultivated or wild species (Mitchell et al. 1997). The ample genetic diversity uncovered by microsatellites is postulated to allow population structure to be revealed more accurately than in other types of markers (Liu et al. 2003). Microsatellites have been useful in determining population structure of various cereal and legume crop species (Li et al. 2001; Liu et al. 2003; Ferguson et al. 2004; Garris et al. 2005). In common bean, microsatellite evaluation of cultivars has concentrated mainly on snap bean fingerprinting (Metais et al. 2002), cross species amplification (Gaitán et al. 2002) and parental surveys with advanced breeding lines and released varieties (Blair et al. 2006). In the study by Blair et al. (2006), microsatellites detected lower genetic diversity in the Mesoamerican genepool compared to the Andean gene pool and while races were very evident in the Andean genepool they were not as distinct in the Mesoamerican genepool.

The objective of this study, therefore, was to identify whether Mesoamerican races could be identified using microsatellite markers and a set of landraces and varieties from the seed classes that are representative of the different races of this genepool. For this study we used 60 genotypes that had been identified by Singh et al. (1991a, b, c) or Beebe et al. (2000) as belonging to each race and analyzed them for allelic variation at a total of 52 microsatellite loci distributed throughout the genome. The use of microsatellites to identify common bean races is of interest to our laboratory, to bean researchers generally and to the curators and final users of international and national germplasm collections.

Materials and methods

Plant materials and DNA extraction

A total of 60 genotypes were used of which 35 were from Mexico, 8 from Guatemala, 7 from Brazil, 3 from

El Salvador, 2 from Colombia and 1 each from Costa Rica, Ecuador and the United States, all of these representing landraces or varieties from the Mesoamerican genepool; with two additional genotypes, 'Calima' from Colombia and 'G19833' from Peru, used as an Andean outgroup (Table 1). Genotypes were selected based on previous race designations (Beebe et al. 2000; Singh et al. 1991a, b, c) and the phaseolin pattern of each genotype was known to be typical of the Mesoamerican gene pool (S, Sb, Sd and M). The Mesoamerican genotypes 'ICA Pijao' and 'DOR364' from Colombia and El Salvador/CIAT, respectively, were considered control genotypes for the genepool since they had been evaluated previously (Blair et al. 2006). Genotypes were provided by the Genetic Resources Unit of CIAT and most represented landrace collections rather than bred varieties (http://www.ciat.cgiar.org/urg/beans.htm). For each accession, 10 seeds were selected at random from the original Genebank accessions and germinated on germination paper. Trifoliate leaf tissue was harvested from all of the 6-day old seedlings and mixed prior to grinding in liquid nitrogen. The bulk tissue was then used for DNA extraction with the method of Afanador et al. (1993). DNA quality was evaluated on 0.8% agarose gels followed by quantification on a Hoefer DyNA fluorometer (DNA Quant™ 200). DNA was diluted to 10 ng/ml for further experiments.

Microsatellite amplifications

In all, 52 microsatellites were used of which 22 were cDNA based and 30 were genomic (Yu et al. 2000; Gaitan et al. 2002; Blair et al. 2003). Microsatellites were selected based on their high polymorphism information content from Blair et al. (2006) and their even distribution around the genome based on Blair et al. (2003). Characteristics of the microsatellites selected are available as supplementary online material. PCR amplifications were conducted in 96-well plates using a PTC-100 (MJ Research) thermal cycler with conditions as given in Blair et al. (2006). PCR reaction volume was 12 μl and contained 50 ng genomic DNA and 0.16 μM of each primer (forward and reverse), 10 mM Tris–HCl (pH 7.2), 50 mM KCl, 1.5–2.5 mM Mg (depending on the primer), 0.2 mM dNTP and 1.0 unit *Taq* polymerase. All markers produced single bands except PV-ag004 which detected two loci both of which were scored. The PCR products were run on 4% polyacrylamide gels using Sequi-Gen® GT sequencing units (Bio-Rad Laboratories Inc., Hercules, CA, USA) at a constant 120 W for approximately 1.5 h which. Gels were stained with silver nitrate (Promega Inc.) and allele sizes were evaluated relative to a 10 bp molecular weight size standard (Invitrogen, Carlsbad, CA, USA). Alleles for the check genotypes (Calima, DOR364, ICA Pijao and G19833) were confirmed to be the same sizes as in Blair et al. (2006).

Data analysis

The allele information coded for band presence or absence was used in NTSYS-pc Version 2.10 (Rohlf 2002) for correspondence analysis and for creating the matrix of Euclidean distances between genotypes using the CORRESP and SIMINT subprograms, respectively. The distance matrix was then used in the SAHN subprogram for sequential agglomerative hierarchical nested cluster analysis to construct dendrograms based on UPGMA (Unweighted Pair Group Method with Arithmetic Mean) clustering method that were visualized with TREE PLOT. Neighbor joining analysis was also performed with Darwin software (Perrier et al. 2003). Within group genetic diversity was estimated for each group found within the dendrogram based on the genetic distance between all pairs of genotypes (Nei 1978) using the software program POPGENE Version 1.31 (Yeh et al. 1997). Other common parameters of genetic diversity (percentage polymorphic loci, allele frequencies, observed heterozygosity (H_o), indices of genetic differentiation (G_{ST}), and gene flow (N_m) were determined as were Wright's F-statistics computed to assess the degree of population differentiation among races. The number of populations (K) was confirmed with the software STRUCTURE (Pritchard et al. 2000) and visualized with the software DISTRUCT (Rosenberg 2002).

Results

Polymorphism in the population

A total of 267 alleles were identified in this study with an average of 5.1 alleles per marker. All the microsatellites were polymorphic presenting two or more alleles each. The maximum number of alleles found for a given locus in the full set of genotypes was 15 for the genomic microsatellite GATs91. Among the cDNA-based microsatellites, PV-ag004b had the highest number with 11 alleles. Microsatellites with only two alleles each included BM68, BM142, BM164, BM197, BMc5, BMd7, BMd12, BMd32, BMd41, BMd53, GATs11B and GATs54). Only six of the markers presented null alleles and these were in low proportion: BM170 and BMd10 each had two genotypes with null alleles

Table 1 Origin and identity of Mesoamerican race representatives and Andean controls used in the study

Entry number	CIAT entry	Genotype name	Country origin	Province/state origin	Phaseolin	Seed size[a]	Seed color[b]	Growth habit	Previous race designation[c]
1	G685	Moncure no. 12	Guatemala	Alto Verapaz	Sb	S	5, 6	4	G2
2	G855	Norvell no. 3583	Mexico	Oaxaca	Sb	M	3	4	J1
3	G1764	Gentry 21057 Apetito	Mexico	Mexico	S	L	3	5	J
4	G2026	Gentry 21394	Guatemala	Sacatepequez	Sb	M	8	4	G
5	G2268	Gentry 21236 Garrapato	Mexico	Guanajuato	Sb	M	2, 3	3	J
6	G2333	Gentry 21835 Colo.Teopisca	Mexico	Chiapas	S	S	6	4	G
7	G2379	Gentry 21926 Ojo De Liebre	Mexico	Oaxaca	Sb	M	2, 4	3	D1
8	G2402	Gentry 21955 Garrapato	Mexico	Sonora	Sd	M	2, 4	3	D
9	G2404	Gentry 21958 Pinto De Yaqui	Mexico	Sonora	S	M	2, 4	3	D
10	G2445	Chiapas 73 Rojo Oscuro	Mexico	Chiapas	S	S	6	3	M2
11	G2775	Gentry 22053 Ojo De Cabra	Mexico	Chihuahua	Sb	M	2, 4	3	D1
12	G2868	Gentry 22244 Azufrado	Mexico	Nayarit	S	S	3	3	D2
13	G2997	Rabia De Gato	Guatemala	Jutiapa	B	S	8	3	M
14	G3353	Puebla 152	Mexico	Puebla	S	M	8	3	J
15	G3545	Guerrero 90	Mexico	Guerrero	Sd	S	5, 2	2	M1
16	G3645	Jamapa	Mexico	Veracruz	Sb	S	8	2	M
17	G3807	Brasil 2 Pico de Oro	Brazil	NA	Sb	M	2	1	M1
18	G4017	Carioca	Brazil	NA	Sb	M	2, 4	3	M1
19	G4090	Rojo De Seda	El Salvador	NA	Sd	S	6	3	M2
20	G4206	36 Sal Rico De MG	Brazil	NA	S	S	8	2	M
21	G4495	Porrillo Sintetico	El Salvador	Cuscatlan	B	S	8	2	M1
22	G4497	Negro Jamapa	Mexico	Veracruz	Sb	S	8	2	M1
23	G4822	Rosinha G2	Brazil	NA	S	S	2	2	M1
24	G4830	Rio Tibagi	Brazil	Santa Catarina	Sb	S	8	2	M
25	G5036	Mulatinho	Brazil	NA	S	S	2	2	M
26	G5694	Cornell 49–242	United States	NA	B	S	8	2	M
27	G5711	Compuesto Chimaltenango 2	Guatemala	Chimaltenango	S	S	8	3	G
28	G5897	Flor De Mayo	Mexico	NA	S	S	2, 5	3	M
29	G7602	Mexico 222	Mexico	NA	S	L	2	3	D
30	G7932	Nahuizalco Rojo	El Salvador	Sonsonate	S	S	6	3	M2
31	G8897	Aguascalientes 29 Garrapato	Mexico	Aguascalientes	M	L	3	3	D2
32	G10850	Guatemala 1341	Guatemala	NA	S	S	8	4	G
33	G10912	Guatemala 1426	Guatemala	NA	S	M	3	4	J2
34	G10945	Flor De Mayo	Mexico	Zacatecas	S	M	5, 2	4	J
35	G10971	Bayo	Mexico	Zacatecas	Sd	M	2	3	D
36	G10982	Pinto	Mexico	Zacatecas	Sd	L	2, 4	3	D
37	G11010	Bayo Regional	Mexico	Durango	Sb	L	2	3	D1
38	G11012	Ojo De Liebre	Mexico	Durango	B	L	2, 4	4	D
39	G11404	Michoacan 16–2	Mexico	Michoacan	Sb	S	5	4	J2
40	G13614	De Celaya	Mexico	Guanajuato	B	L	6	3	J1, M
41	G13673	Mantequilla	Mexico	Puebla	S	L	3	3	M
42	G14241	M7742 Flor De Mayo	Mexico	Jalisco	S	M	5, 2	4	D1
43	G14914	Azufrado Amarillo 33	Mexico	Sinaloa	S	S	3	3	M
44	G15416	Catu	Brazil	NA	S	S	2	3	M
45	G17648	Pata De Zope	Guatemala	Jutiapa	Sd	S	8	3	M1
46	G17649	Chichicaste	Guatemala	El Progreso	S	S	8	3	G
47	G18440	Durango 222	Mexico	Durango	Sd	M	2	3	J
48	G18445	Bayo Zacatecas	Mexico	Zacatecas	Sd	L	3	3	D2
49	G18446	Bayo Madero	Mexico	NA	S	L	3	3	D
50	G19833	Chaucha Chuga[d]	Peru	Amazonas	H	L	3, 6	3	P

Table 1 continued

Entry number	CIAT entry	Genotype name	Country origin	Province/state origin	Phase olin	Seed size[a]	Seed color[b]	Growth habit	Previous race designation[c]
51	G22005	Bayo Criollo El Llano	Ecuador	NA	M	M	6, 1	4	D2
52	G22029	Conejo	Mexico	NA	B	M	2, 4	4	J
53	G22036	Flor De Mayo IV	Mexico	NA	S	M	5, 2	3	J
54	G22041	Garbancillo Zarco	Mexico	NA	M	M	3, 4	4	J
55	G22044	Garrapato	Mexico	NA	M	M	2, 4	3	D2
56	G22079	Ojo De Cabra 24	Mexico	Durango	S	M	4	3	D
57	G4494	DIACOL Calima[d]	Colombia	NA	T	L	6, 2	1	NG
58	G51105	DOR364[e]	El Salvador	NA	S	S	6	2	M
59	NA	Dos Meses	Costa Rica	NA	Sd	S	6	2	M2
60	G5773	ICA Pijao[e]	Colombia	Valle Del Cauca	B	S	8	2	M

[a] Seed size: *S* small, *M* medium, *L* large

[b] Seed color: primary and secondary color designations (separated by a comma) as *1* white, *2* cream, *3* yellow, *4* tan, *5* pink, *6* red, *8* black

[c] Growth habit as *I* determinate bush, *II* indeterminate bush, *III* indeterminate prostrate, *IV* indeterminate climbing beans

[d] Andean controls

[e] Mesoamerican controls

(3.3%) while the markers BM156, BM157, PV-ag001 and PV-ag004 each presented one genotype each with a null allele (1.7%). Repeat amplifications were used to confirm the null alleles and allele amplification was complete for the remainder of the genotype by marker combinations. Among the microsatellites, the genomic markers presented a higher average number of alleles with respect to the cDNA-based markers (3.59 versus 2.77). Average indices of diversity for genomic microsatellites were higher than those for the cDNA based microsatellites (0.51 versus 0.42), indicating that, depending on their origin the microsatellites contributed different levels of information to the study. The genomic microsatellites that presented the highest indices of diversity were BM143 and GATs91, whereas the least polymorphic were BM142, BM155, BM157 and BMd33. For the cDNA-based microsatellites, the most polymorphic was PV-ag004b and the least polymorphic was BMd32.

Genetic relationships

Relationships between the genotypes used in this study are shown in the UPGMA dendrogram in Fig. 1 where it was possible to distinguish two major groups within the Mesoamerican gene pool and obtain a clear separation of the Andean controls G19833 and ICA Calima from the Mesoamerican genotypes. The two gene pools separated at a Euclidean distance of 3.5 while the Mesoamerican groups separated from each other at a distance of 2.0. Two genotypes remained unassigned and were well separated at the maximum distance within the gene pool; as discussed below these may represent race G accessions and are tentatively named thus. Meanwhile within the remaining Mesoamerican genotypes two groups were found, one corresponding to race M and the other to a combination of race D and J, henceforth described as the D–J group. Race M separated at a Euclidean distance of slightly above 1.0 from the D–J group. Within both the D–J group and race M we observed subgroups that reflected phenotypic characteristics such as growth habit, seed size and geographical distribution as discussed below. Dendrogram results were confirmed to be the same with Neighbor Joining analysis.

Race Mesoamerica

A total of 26 genotypes, all with small to medium sized seed and including the Mesoamerican control genotypes DOR364 and ICA Pijao fell within our designation of race Mesoamerica (M) (Fig. 1). Although these race M accessions grouped at Euclidean distance of 0.75 or below, they were found to be fairly diverse in terms of observed and effective number of alleles as well as the percentage of polymorphic loci (Table 2). This race was subdivided into five subgroups (a–e), which were somewhat distinct in terms of growth habit, seed color and geographical origins. The first and second subgroups (a and b) were separated from the remaining groups at Euclidean distances of 0.5 and 0.4, respectively, while the latter three subgroups (c–e) clustered together at Euclidean distances below 0.3. The first two subgroups (a and b) contained a mix of genotypes previously designated as from other races or from subrace M_2 while the other subgroups (c–e) contained genotypes mainly identified as subrace M_1 with the exception of DOR364 considered to be from subrace M_2. The latter subgroups were from a range of countries while the first and second subgroups were

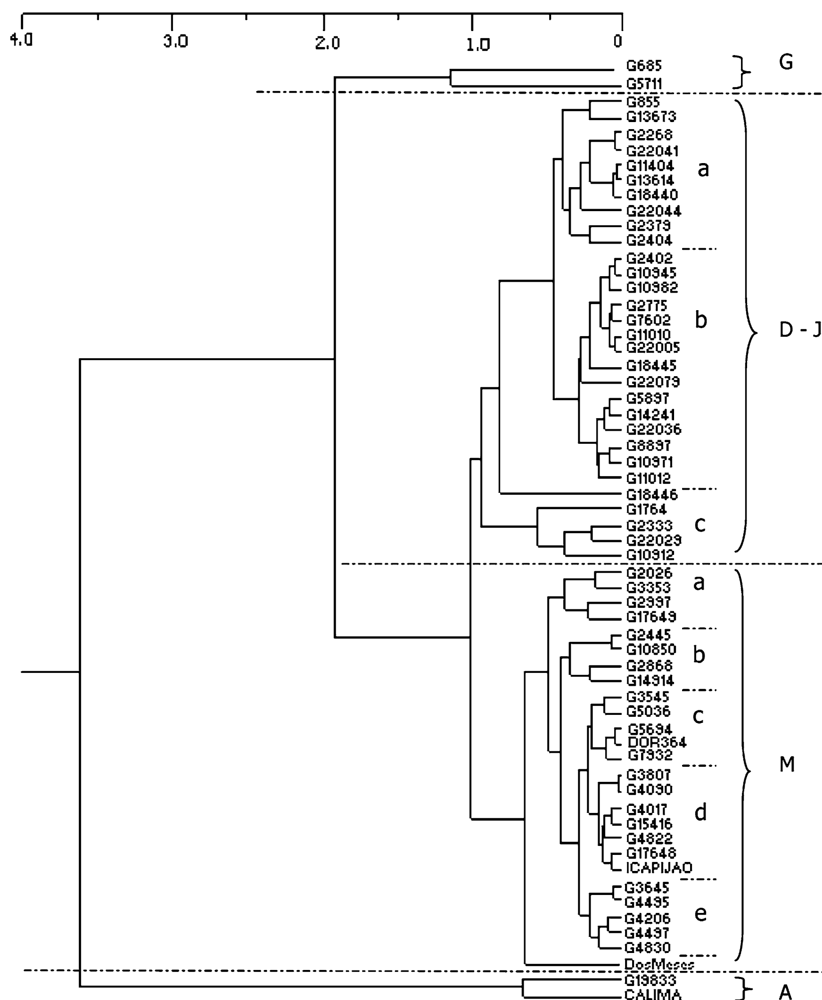

Fig. 1 Relationship between common bean genotypes representing Mesoamerican races (*D* Durango, *J* Jalisco, *G* Guatemala, *M* Mesoamerica) in a UPGMA dendrogram based on Euclidean distance analysis of data from 52 microsatellite markers. Andean control genotypes included for reference

composed mostly of accessions from Guatemala and Mexico. In terms of growth habit, the last three subgroups (c–e) possessed mainly bush indeterminate to prostrate architecture (52% type II and 38% type III) while accessions of the first and second subgroups (a and b) were all of type III (75%) or type IV (25%) growth habit. The last three subgroups (c–e) presented a range of seed colors including black, red, tan, carioca (tan with brown) and pink while the first subgroup (a) presented only black seeds. Meanwhile, the second subgroup (b) was more varied in seed color including yellow, red and black grain types. All the genotypes of this intermediate group had small seed size and S type phaseolin. This subgroup stood out within the race by falling between the first subgroup and the latter three subgroups. In terms of specific genotypes within each group, the controls ICA Pijao and DOR364 were found within two closely related subgroups (c and d); while the genotype 'dos meses' was separate from all of the above subgroups. It was notable that the genotypes in the closely related subgroups (c–e) were organized according to either growth habit or seed color. For example, all the accessions of the third group (G3545, G5036, G5694, G7932 and DOR364) had small seed size and four of five had type II growth habit; while genotypes of the fourth group (G3807, G4090, G4017, G14516, G4822, G17648 and ICA Pijao) were similar in terms of seed size but four of seven had type III growth habit. All the accessions of the last group (e) presented black seeds and type II growth habit. No clear distinctions were found for phaseolin pattern among the subgroups and all had the B, S and Sb alleles which are typical of the race although a few genotypes were found with Sd phaseolin which is not typical of the race.

Durango–Jalisco group

A group designated as Durango–Jalisco (D–J) consisted of 30 accessions (Fig. 1) almost all of medium to large seed size that had been previously identified as belonging to these races (Table 1). This group presented the largest number of observed and effective alleles as well as the highest number and percentage of

polymorphic loci although the number and percentage of polymorphic loci was similar to race M (Table 2). The D–J group could be divided at a Euclidean distance of approximately 0.5 into three subgroups (a–c) with 10, 16 and 4 genotypes, respectively, with one genotype (G18664) not located in any of these clusters. Genotypes in two subgroups (a and b) were similar in seed size both being predominantly medium-sized, however the second subgroup (b) contained a narrower range of seed colors mostly cream and brown, or pink and cream seed coats; while the first subgroup (a) included yellow to brown seed colors in addition to the cream and brown colors of the other subgroup. The majority of the genotypes in the subgroup a were previously designated as race J while the majority of the genotypes in the subgroup b were previously designated as race D, although the distinctions were not complete and some genotypes with other race designations were included in both groupings. Phaseolin types in these subgroups were mostly typical of races D and J, including S (40%) and Sd (17%), however a small proportion had atypical phaseolin patterns, including Sb (19%), M (15%) and B (7.6%). Both subgroups had types III and IV growth habits typical of races D and J. The second subgroup (b) could itself be subdivided into two clusters even though both of these were similar in terms of the growth habit and seed colors of the genotypes in them. The final subgroup (c) was represented by four genotypes, two of which had been classified as race J, one as race G and one of which was unassigned. The genotypes in this last group had yellow or cream-colored, small to medium-sized seed and were of type IV growth habit. All the genotypes originated in Mexico and Guatemala and had S type phaseolin except for G22029 with B type phaseolin. Further divisions were not evident within this subgroup, although G1764 was distinct from the other three accessions. This subgroup was intermediate in seed size compared to the rest of the group and to the genotypes of race M; from its placement in the dendrogram was more closely associated with the former compared to the latter group. For the purpose of calculating observed and effective alleles and percentage polymorphic loci (Table 2), subgroups a and c were considered to represent race J and subgroup b to represent race D, although the distinctiveness of races D and J from each other is low.

Table 2 Genetic diversity parameters for Mesoamerican race and Andean control genotypes

Groups	N	na	ne	H_o	Nei's	P	%
Race D	16	2.769	1.720	0.046	0.339	45	86.54
Race J	14	3.288	2.221	0.045	0.425	48	92.31
Total D–J	30	3.808	2.081	0.046	0.414	48	92.31
Race G	2	1.577	1.526	0.067	0.262	28	53.85
Race M	26	3.212	1.939	0.028	0.340	42	80.77
Total Mesoamerican	58	4.789	2.352	0.039	0.444	49	94.23
Andean checks	2	1.346	1.339	0.010	0.171	18	34.62
Grand total	60	5.077	2.447	0.038	0.468	52	100.0

Race abbreviations: *D* Durango, *J* Jalisco, *G* Guatemala, *M* Mesoamerica. Other abbreviations: number of genotypes (N), observed number of alleles (na), effective number of alleles (ne), observed heterozygosity (H_o), genetic diversity according to Nei (1978), number of polymorphic loci (P), percentage polymorphic loci (%)

Correspondence analysis

Correspondence analysis separated all Mesoamerican genotypes from the Andean genotypes in the first dimension (Fig. 2A and B). A second dimension separated two principal groups; supporting the assignment of the genotypes evaluated to the groups discussed above, namely race M versus the D–J race group (Fig. 2A and C). The separation of race J and race D was not supported as shown by the clustering of all the accessions within the D–J group and no separation of the genotypes assigned to one or the other race within that cluster.

The correspondence analysis also showed two genotypes separated in a third dimension but these were found associated in one case with the race M cluster (G5711) and in the other case with the D–J group (G685) (Fig. 2B and C). Both of these genotypes were of Guatemalan origin, have small seeds with phaseolin types S and Sb and growth habit types III and IV, respectively. These accessions could be designated as belonging to race G given their place of origin and inclusion in this classification in the study by Beebe et al. (2000), however as discussed below the existence of race G is called into question by the large number of accessions previously identified as of this race which cluster with the other three races in both the dendrogram and correspondence analysis. The observed and effective number of alleles as well as the percentage polymorphic loci were lower in race G than in other races (Table 2).

Genetic diversity

Genetic diversity as measured by Nei's index values was higher for the D–J group compared to races M and G (Table 2). Within the D–J group, race J genotypes had greater genetic diversity than race D genotypes. The genetic diversity of the entire set of genotypes was 0.468. Meanwhile, the observed heterozygosity was low for all the Mesoamerican races, reflecting the inbreeding

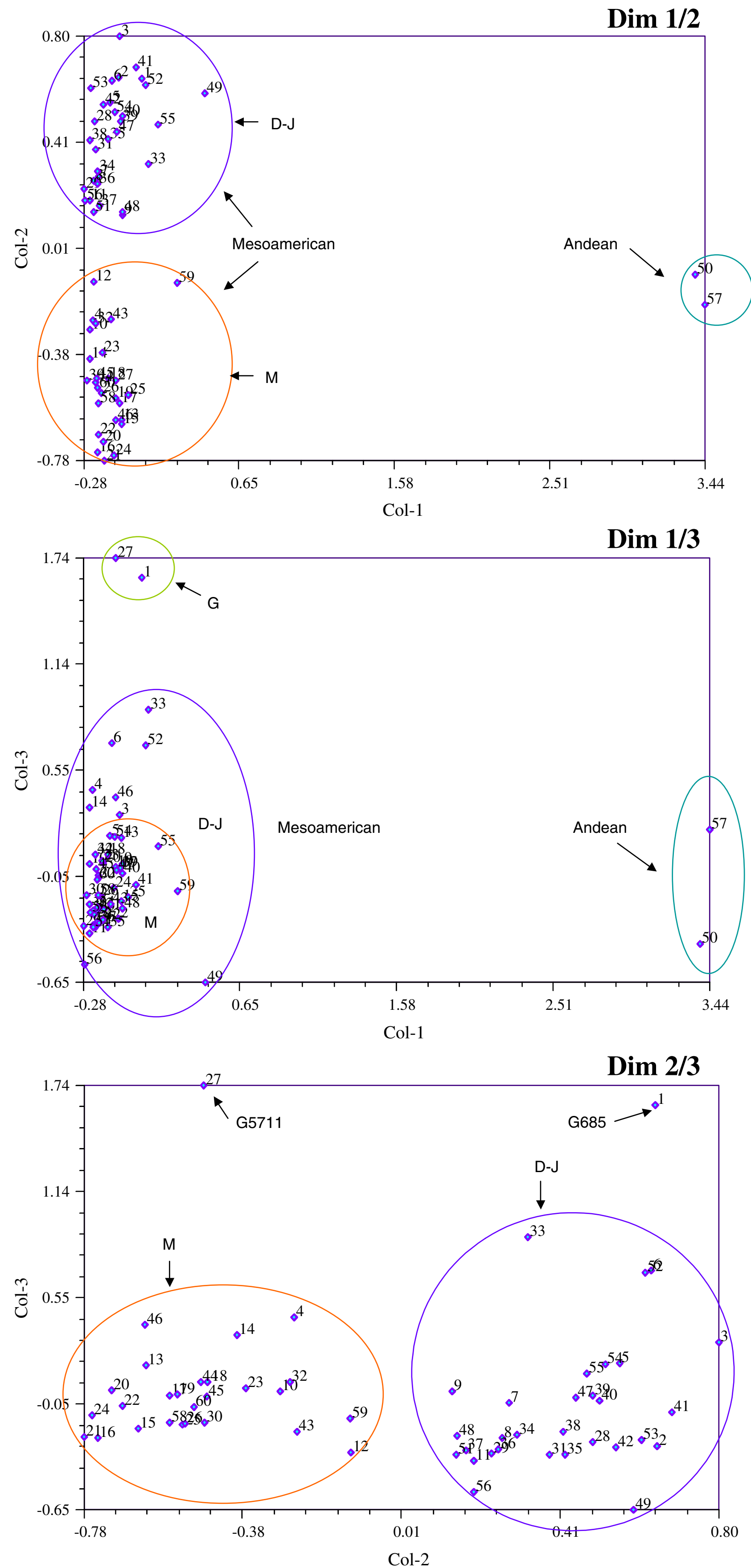

Fig. 2 Correspondence analysis showing the relationship of Mesoamerican races (*D* Durango, *J* Jalisco, *G* Guatemala, *M* Mesoamerica) and Andean controls based on data from 52 microsatellite markers. Genotype identification as listed in Table 1

nature of common beans. Slightly higher observed heterozygosity in race G may be explained by greater diversity and outcrossing in type IV growth habit beans from highland Chiapas–Guatemala (Beebe et al. 2000).

A high amount of population differentiation (G_{ST}) existed in the study with an overall average among populations of 0.434 although this was higher among gene pools (0.385) than among races (0.283). Correspondingly, gene flow (N_m) was low (<1) both between the gene pools (0.398) and the races (0.799). Gene flow was moderate (>1) between race J and race D and between race M and race D; but was low between race G and the other races (Table 3). Race G and the Andean control groups, both with small numbers of individuals, tended to create population structure among the analyzed accessions. In the first case, race G was highly differentiated ($G_{ST} = 0.22–0.24$) because of the distance between these two accessions and other Mesoamerican genotypes. In the second case, the Andean controls were very isolated from the Mesoamerican races ($G_{ST} = 0.43–0.57$). Differentiation between races M and J also was fairly high, while it was less high between races J and D or between races M and D. The highest indices of genetic identity were between race D and races J and M. Among the Mesoamerican races, D and J presented the lowest levels of genetic distance followed by races D and M, corroborating data obtained for genetic identity. While genetic identity was generally high for all the Mesoamerican genotypes (0.69–0.84), it was low between Mesoamerican races and Andean genotypes (0.24–0.28).

Population structure was confirmed for the Mesoamerican gene pool excluding the Andean controls with the software program Structure (Pritchard et al. 2000). Parameters used were a burn-in of 50,000 runs and range of K values from 1 to 4 (Table 4). Of the predetermined populations, race M and race D and J together were shown to be separate but with some admixture between them (Fig. 3). The possible race G individuals discussed above could also represent an additional population with admixture into the other races, however the low number of genotypes made it difficult to distinguish their contribution to population structure.

Discussion

The level of polymorphism in our study was higher than that previously reported for single copy markers such as RFLPs (Becerra and Gepts 1994), isozymes (Paredes and Gepts 1995) and chloroplast DNA (Chacón et al. 2005) or multi-copy markers such as RAPDs

Table 3 Genetic differentiation (G_{ST}), gene flow (N_m), genetic distance (GD) and genetic identity (I) among and between Mesoamerican races and Andean control genotypes analyzed with microsatellite markers

Race/gene pool	G_{ST}					I				
	M	D	J	G	A	M	D	J	G	A
Mesoamerica ($n = 26$)	****	0.14	0.22	0.22	0.44	****	0.8005	0.7813	0.7050	0.2442
Durango ($n = 16$)	1.58	****	0.13	0.22	0.43	0.2225	****	0.8884	0.6258	0.2159
Jalisco ($n = 14$)	0.88	1.67	****	0.24	0.51	0.2468	0.1184	****	0.7225	0.2869
Guatemala ($n = 2$)	0.90	0.87	0.80	****	0.57	0.3495	0.4687	0.3251	****	0.2550
Andean controls ($n = 2$)	0.32	0.33	0.24	0.19	****	1.4271	1.5329	1.2485	1.3664	****
	N_m					GD				

Race/gene pool abbreviations: *M* Mesoamerica, *D* Durango, *J* Jalisco, *G* Guatemala, *A* Andean controls. Genetic differentiation (G_{ST}) and genetic identity (I) in upper diagonals in left and right panels of the table, respectively. Gene flow estimated as $G_{ST} = 0.25(1 - G_{ST})/G_{ST}$ (N_m); genetic distance (*GD*) in lower diagonal of left and right panels, respectively

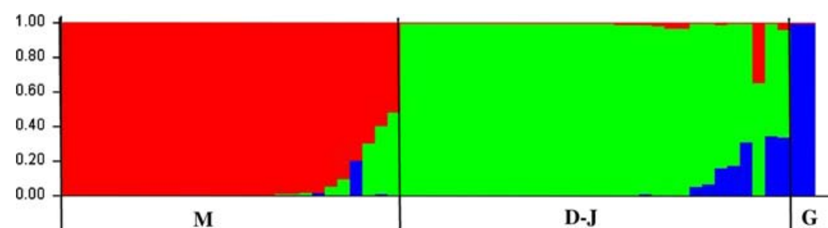

Fig. 3 Graph of the population structure ($K = 3$) for 58 Mesoamerican genotypes (x-axis) sorted by membership coefficients (y-axis) within clusters based on highest estimated probability among 50,000 runs using the software program Structure (Pritchard et al. 2000). Cluster/group names indicated below figure (*M* race Mesoamerica, *D–J* races Durango and Jalisco together, *G* race Guatemala)

(Duarte et al. 1999; Skroch et al. 1998; Beebe et al. 1995, 2000) but similar to those reported for microsatellites previously (Gaitán et al. 2002; Masi et al. 2003; Gomez et al. 2004; Blair et al. 2006). These differences in polymorphism can be explained by the sensitivity of microsatellites as molecular markers; as well as the fact that we selected a wide variety of divergent landrace genotypes from within the Mesoamerican gene pool and used a subset of very informative microsatellites from Blair et al. (2006). Observed heterozygosity was low despite the high polymorphism and average differentiation within populations ($G_{IS} = 0.88$) showing that a high degree of self-pollination probably occurs within each accession. Geographical and agroecological isolation, as well as selection by farmers or the germplasm banks from which the accessions were analyzed may have led to these results.

The race structure uncovered by our analysis of Mesoamerican genotypes agrees with previous studies (Beebe et al. 2000; Singh et al. 1991a, b, c) and expands on these. Several observations are worth mentioning. First, similar to the RAPD study of Beebe et al. (2000) as well as the AFLP study of Pallottini et al. (2004), we found that race M was well differentiated from race D and J together. The clustering of genotypes for race M on the one hand and races D–J on the other also agrees with morphological distinctions described by Singh et al. (1991a, b) with almost all the race D–J genotypes in our study producing medium to large seeds (27 out of 30); almost all of the race M genotypes producing small seeds (22 out of 26). Both groups of genotypes produced the seed colors typical of their respective races as described by previous authors (Singh et al. 1991a; Voysest et al. 1994). Commercial classes for the group D–J genotypes were Bayo, Flor de Mayo, Garbancillo, Mantequilla, Ojo de Cabra, Pinto; while for the race M genotypes they were black turtle, carioca, mulatinho, pink, rosinha and small red.

A second observation, as described above, was that race J was difficult to distinguish from race D. The close relatedness of these two races was also evident in previous studies (Beebe et al. 2000; McClean et al. 2004; Rosales-Serna et al. 2005; Pallottini et al. 2004)

Table 4 Inference of population number (K) in common bean genotypes representing Mesoamerican races

K	Ln $P(D)$	Var[Ln $P(D)$]	$P(K\|X)$
1	−5995.1	93.0	0
2	−5204.8	212.3	0
3	−4750.5	286.0	4.128E−123
4	−4468.7	344.3	1.0000

The *values* in the last column assume a uniform prior for K of four populations. D divergence between populations

and is expected given the similar geographical range from which they have originated in Central Mexico. Our results also agree with those of Chacón et al. (2005) who found shared chloroplast haplotypes between race D and J, although they identified one haplotype as more predominant in race J. The nature and number of RAPD loci sampled by Beebe et al. (2000) may explain why they observed further subdivisions of races D–J while we did not, however we include a large set of functional sequences since 22 out of 52 of the microsatellites were derived from expressed genes as described in Blair et al. (2003). The close relatedness but different agroecological adaptation of the genotypes within races D–J may make this group ideally suited for association mapping studies with single-copy markers such as the microsatellites used in this study.

Third, the existence of race G as a distinct entity as reported by Beebe et al. (2000) was not entirely supported by our study. We found many of the accessions identified as race G by Beebe et al. (2000) to be assigned to the other races and found only two accessions (G685 and G5711) that were grouped apart from races M and the D–J group and which could represent race G as G685 was classified by Beebe et al. (2000) as belonging to the race. However, it was not clear whether this group was coherent because of the small number of accessions grouped within it and because of the way they were located in the correspondence analysis where one of the genotypes was associated with race M and the other with D–J. The finding of diversity in these genotypes would contrast with the single chloroplast haplotype found for three accessions of race G by Chacón et al. (2005) but would agree with the results of Beebe et al. (2000) who found that cultivated Guatemalan climbing beans as whole were unusually diverse. Tohme et al. (1996) found wild beans from Guatemala to be distinct from other wild beans from Mexico. Differences between our study and these previous studies could be due to the maternal inheritance of the markers evaluated by Chacón et al. (2005) or the random, un-mapped and perhaps clustered nature of the RAPD and AFLP markers used by Beebe et al. (2000) and Tohme et al. (1996), compared to the well-distributed genomic and gene-based nuclear microsatellites we used. Microsatellites analysis of a larger set of genotypes might add information especially in regards to the Guatemala race and other diverse accessions.

Fourth, within the larger groupings of race M and races D–J we identified subgroups that were different in terms of seed classes found within them. For the D–J cluster, one group contained Garbancillo types

while another group contained all the Flor de Mayo genotypes, with Pinto and Bayos distributed between both. These results agreed with those of Rosales-Serna et al. (2005) who found division by seed type although their genotypes were mainly bred lines rather than landraces. In contrast to the findings of Beebe et al. (2000) we did not find clear divisions corresponding to subraces in race D. On the other hand, like Beebe et al. (2000) we found fairly clear separation in race M of two subgroups based on growth habit with two subgroups (c and e) containing most of the widely distributed short statured or type II commercial cultivars such as Brasil 2, DOR364, Jamapa, Porrillo Sintético, Rio Tibagi and Rojo de Seda, while types III and IV growth habit landraces were found in the other subgroups (b and d, especially). Similar to results from Beebe et al. (1995) we found that black beans tended to cluster apart from beans of other seed coat colors within race M. An exception to the separation between red and black beans was our check genotype, DOR364, a small red seeded variety from CIAT identified by Beebe et al. (2000) as belonging to subrace M_2, but which clustered with G5694 (Cornell 49–242), a black bean originally from Venezuela. The explanation for this may be found in the fact that G5694 was used in the pedigree of DOR364 which as an advanced line was derived from a breeding program that combined black and red seeded parents. Microsatellites could be very useful for assessing coefficient of parentage so additional SSR analysis of related genotypes within breeding programs would be of interest.

Fifth, the results of this study broadened information on the distribution of phaseolin among races. For example, in some genotypes of the D–J group we found phaseolin type M (G8897, G22005, G22041 and G22044) which has been observed to be more closely related to wild Mesoamerican ancestors (Singh et al. 1991c). An explanation for this may be that gene flow could have occurred for this phaseolin allele from the wild to the cultivated gene pool. In this light, the results of our study demonstrating two predominant populations for the genepool (race M and the D–J group) may indicate that two independent domestications of different wild populations may have occurred to give rise to this dichotomy although further studies would be needed to show this conclusively. Several domestications in Middle America have been postulated by Chacón et al (2005) who predicted potentially up to three independent domestications in the Mesoamerican genepool and also hypothesized that secondary domestications through gene flow between domesticated and wild beans or chloroplast capture through the formation of weedy intermediates could have been possible. The hypothesis of independent domestications might explain why race M and races D–J differ in terms of ecological adaptation, geographical range, agromorphological characteristics, isozymes and DNA markers and why races D and J share more of these characteristics than they do with race M (Singh et al. 1991a; Beebe et al. 2000). It is interesting to note that the major distinguishing agronomic characteristic for this division is not growth habit since in both groups there are types II, III and IV growth habits as much as it is seed size with small seeded genotypes pertaining to the race M portion of the genepool and medium seeded genotypes pertaining to the D–J complex.

In conclusion, the importance of this study lies in the application of a single copy, easy to use and highly reproducible marker type to the study of genetic diversity within the Mesoamerican gene pool The evaluation of population and especially race structure has important implications for genetic improvement of common bean in terms of predicting combining ability or searching for novel alleles (Singh et al. 1991a; Islam et al. 2004). The large difference between race M on the one hand and the D–J group on the other can be exploited in the design of crossing programs (Singh 1989) and has been used to improve architectural traits in common bean (Kelly 2001). Likewise hybridization between genotypes of different subgroups in each race could be explored (Beebe et al. 1995). Another important aspect of our work was that microsatellite diversity seems to be associated with useful agronomic variation as genotypes were separated fairly accurately according to seed size and color which together make up commercial classes. This may indicate that association mapping techniques could be applied to germplasm collections of common beans using a set of microsatellite markers such as the one used here.

Acknowledgments The authors wish to thank S. Beebe, R. Hidalgo, C. de Vicente and M.C. Duque for helpful suggestions during the study. We are also grateful to D. Debouck, O. Toro and A. Hoyos of the Genetic Resource Unit and Bean Projects of CIAT for germplasm, and to P. Zamorano for formatting. This research was part of a thesis conducted at Univ. Nacional de Colombia and was supported by the Generation Challenge Program and CIAT funds.

References

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

Becerra V, Gepts P (1994) RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. Genome 37:256–263

Beebe S, Ochoa I, Skroch P, Nienhuis J, Tivang J (1995) Genetic diversity among common bean breeding lines developed for Central America. Crop Sci 35:1178–1183

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

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

Beebe S, Toro O, González AV, Chacón MI, Debouck D (1997) Wild-weed–crop complex of common bean (*Phaseolus vulgaris* L., Fabaceae) in the Andes of Peru and Colombia, and their implications for conservation and breeding. Genet Resour Crop Evol 44:73–91

Blair MW, Pedraza F, Buendia H, Gaitan E, Beebe S, Gepts P, Tohme J (2003) Development of a genome wide anchored microsatellite for common bean (*Phaseolus vulgaris* L). Theor Appl Genet 107:1362–1374

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

Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Bean (*Phaseolus* spp.)—model food legumes. Plant Soil 252:55–128

Chacón MI, Pickersgill S, Debouck D (2005) Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of Mesoamerican and Andean cultivated races. Theor Appl Genet 110:432–444

Duarte J, Dos Santos J, Melo L (1999) Genetic divergence among common beans cultivars from different races based on RAPD markers. Genet Mol Biol 22:419–426

Ferguson ME, PJ Bramel, Chandra S (2004) Gene diversity among botanical varieties in peanut (*Arachis hypogaea* L.). Crop Sci 44:1847–1854

Gaitán E, Duque MC, Edwards K, Tohme J (2002) Microsatellite repeats in common bean (*Phaseolus vulgaris* L.): isolation, characterization, and cross-species amplification in *Phaseolus* spp. Crop Sci 42:2128–2136

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Gepts P (1988) Phaseolin as an evolutionary marker. In: Resources of *Phaseolus* beans. Kluwer, Dordtrecht, pp 215–241

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

Gomez O, Blair MW, Frankow-Lindberg B, Gullberg U (2004) Molecular and phenotypic diversity of common bean landraces from Nicaragua. Crop Sci 4:1412–1418

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

Kelly JD (2001) Remaking bean plant architecture for efficient production. Adv Agron 71:109–143

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

Li C-DF, Fatokun CA. Ubib B, Singh BB, Scoles GJ (2001) Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. Crop Sci 41:189–197

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

Masi P, Spagnoletti ZP, Donini P (2003) Development and analysis of multiplex microsatellite markers sets in common bean (*Phaseolus vulgaris* L.). Mol Breed 11:303–313

McClean P, Kami J, Gepts P (2004). Genomics and genetic diversity in common bean. In: Legume crop genomics. AOCS Press, Champaign, IL, pp. 60–82

Metais I, Hamon B, Jalouzot R, Peltier D (2002) Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. Theor Appl Genet 104:1346–1352

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

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

Pallottini L, Garcia E, Kami J, Barcaccia G, Gepts P (2004) The genetic anatomy of a patented yellow bean. Crop Sci 44:968–977

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

Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Enfield, Science Publishers, Montpellier pp 43–76

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

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

Rohlf F (2002) NTSYS pc. Numerical Taxonomy System Exeter Publishing, Setauket, NY

Rosales-Serna R, Hernández-Delgado S, González-Paz M, Acosta-Gallegos JA, Mayek-Pérez N (2005) Genetic relationships and diversity revealed by AFLP markers in Mexican common bean bred cultivars. Crop Sci 45:1951–1957

Rosenberg NA (2002) Distruct: a program for the graphical display of structure results. http://www.cmb.usc.edu/»noahr/distruct.html

Singh S (1989) Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ Bot 43:39–57

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

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

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

Skroch P, Nienhuis J, Beebe S, Tohme J, Pedraza F (1998) Comparison of Mexican common bean (*Phaseolus vulgaris* L.) core and reserve germplasm collections. Crop Sci 38:488–496

Tohme J, Gonzales D, Beebe S, Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. Crop Sci 36:1375–1384

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

Yeh FY, Boyle R, Ye T, Mao Z (1997) POPGENE, the user-friendly shareware for population genetic analysis, version 1.31. Molecular Biology and Biotechnology Centre, University of Alberta, Alberta

Yu K, Park J, 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