Theor Appl Genet (2010) 120:1381–1391
DOI 10.1007/s00122-010-1262-4

ORIGINAL PAPER

Extensive diversity and inter-genepool introgression in a world-wide collection of indeterminate snap bean accessions

Matthew W. Blair · Alejandro Chaves · Adriana Tofiño · Juan Felipe Calderón · Juan Diego Palacio

Received: 16 May 2009 / Accepted: 27 December 2009 / Published online: 20 January 2010

Abstract Common bean can be grown as a grain crop (dry beans) or as a fresh vegetable (snap beans/green beans), both items being important in nutritional terms for providing essential minerals and vitamins to the diet. Snap beans are thought to be derived predominantly from dry beans of the Andean genepool and to be of a recent European origin; however, the existence of Mesoamerican genepool characteristics especially in traditional indeterminate growth habit snap beans indicates a wider origin. The objective of this study was to evaluate genetic diversity within a set of 120 indeterminate (pole type) snap beans and 7 control genotypes representing each genepool using amplified fragment length polymorphism (AFLP) and simple sequence repeat or microsatellite (SSR) markers. The genotypes were predominantly from Asia, Europe and the United States but included some varieties from Latin America and Africa. AFLP polymorphism ranged from 53.2 to 67.7% while SSR polymorphism averaged 95.3% for the 32 fluorescent and 11 non-fluorescent markers evaluated and total expected heterozygosity was higher for SSR markers (0.521) than for AFLP markers (0.209). Both marker systems grouped the genotypes into two genepools with Andean and Mesoamerican controls, respectively, with the Mesoamerican group being predominant in terms of the number of genotypes assigned to this genepool. Phaseolin alleles were not tightly associated with genepool assignment indicating that introgression of this locus had occurred between the genepools, especially with phaseolin "S" in the Andean group (23.5%) and phaseolins "T" and "C" in the Mesoamerican group (12.2 and 8.2%, respectively). The implications of these results on the origin of pole type snap beans and on breeding strategies for this horticultural crop are discussed.

Communicated by H. T. Nguyen.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1262-4) contains supplementary material, which is available to authorized users.

M. W. Blair (✉) · A. Chaves · A. Tofiño
CIAT, International Center for Tropical Agriculture,
6713 Cali, Colombia
e-mail: m.blair@cgiar.org

Present Address:
A. Tofiño
CORPOICA-Motilonia, Km 5 via Bucaramanga,
Valledupar, Colombia

J. F. Calderón
Instituto Humboldt, Km 17.5 Recta Cali-Palmira,
Cali, Colombia

Present Address:
J. D. Palacio
ICA-Tibaitata, Km. 14 Via Mosquera,
Bogotá, Colombia

Introduction

Snap beans are one of the most important and commonly consumed horticultural products in the world (Myers and Baggett 1999). Snap beans and dry beans together make up the species known as common bean (*Phaseolus vulgaris* L.), but unlike dry beans, snap beans are consumed for their edible, whole pods rather than dry seeds. They, therefore, have been selected as a vegetable to have succulent pod walls and low pod wall fiber. In the case of "stringless" beans, very little fiber is found at the pod suture as well (Silbernagel 1986). Synonyms of snap beans are "French bean", "Garden bean" or "Haricot bean", while "String bean" refers to many older varieties that have fiber at the

pod suture but that are still consumed as a vegetable where the strings are removed manually before cooking. Snap beans are important sources of essential vitamins (A, B_{12} and C) and dietary fiber but are low in calories, which make them a healthy food.

Total world production of snap beans is around 9 million tons, with China, Turkey, India, Spain, France and the United States being among the biggest producers and consumers (FAOSTAT 2007). Marked preference and intense commercialization of snap beans occurs in developed countries of North America and Europe with many seed and food processing companies intensively involved in the product chain (Silbernagel 1986). Middle-income countries of Asia have important fresh market sales of snap beans though much of this is less formal than in Europe. Finally, the crop is of growing importance to other developing countries, both as an export crop and as a local product (Henry and Janssen 1992). In terms of export, trade between Central America and the United States or East Africa and Europe produce important income streams for countries like Guatemala and Kenya. Meanwhile, as wages have gone up in countries such as Colombia or India, the markets for snap beans have also increased (Pachico 1987).

The demand for a constant supply of fresh beans means that production in the northern hemisphere migrates with the seasons: winter or spring production is found in the very southern edges of Europe and North America (e.g. Florida, Mexico or Spain) followed by summer production for fresh market and processing further north in both hemispheres (Henry and Janssen 1992). The production system in the United States is particularly mobile with snap bean planters, harvesters and marketing systems moving sequentially north through various states from Florida to Georgia, Kentucky, New Jersey, New York, Michigan and Wisconsin as the season progresses in each location (Myers and Baggett 1999).

Like other sorts of common beans, the primary center of diversity for snap beans is believed to be in the Americas; however, snap beans have a wide distribution and are very diverse in various regions around the world especially in Europe (Metais et al. 2002) and Asia (Zhang et al. 2008). Snap beans are thought to have been mainly derived from dry beans after European (especially Spanish) colonization of the New World with rapid development of snap beans as a new crop in Europe (Myers and Baggett 1999), but perhaps this theory is based on the lack of an archeological record for green bean pods in the Americas. Preservation of fresh pods was rare, while dry seeds of common bean are found in early agricultural sites in a long arc from the United States through Mexico and Central America to the Andes (Kaplan and Kaplan 1988).

The concept of an American "genesis" of snap bean germplasm followed by predominantly European selection of snap bean traits (Myers and Baggett 1999) may also be re-enforced by the wide diversity and spread of snap beans in various European countries, by the historical record for early snap beans adoption by Europeans and by the integration of snap beans into European cuisine. Although less well documented the same characteristics would be evident for snap beans in Asia; and in both regions snap beans probably joined other legumes consumed for their fresh pods or immature seeds, which were easier to cook, faster to harvest and sometimes better tasting or less anti-nutritional than their dry seed counterparts (Zhang et al. 2008). However, this does not preclude that pre-Colombian societies of the Americas also consumed fresh green pods; however, this practice may have been uncommon.

Further to this issue of snap bean selections being from outside or inside the Americas, indigenous terms for snap beans exist in some Amerindian languages (Voysest 2000), which may show a long-term knowledge of snap beans that was not introduced or re-introduced as modified dry bean germplasm from outside the region. For example, in Quechua, the term Chaucha (interpreted as tender or light green/yellow) is used to refer to snap beans and show a probable original use of green pods in the Andean region. Snap beans in the temperate regions of North or South America as well as in the Middle East and Eastern or Southern Africa are assumed to be originally mainly from Europe or the United States.

In another possible misconception, various authors have generally assumed that all or at least most snap beans were derived from the Andean genepool of common beans, one of two major genepools of common bean and the one that originated in the Andes mountains of South America, given that early analysis showed that bush type snap beans generally had Andean type seed protein (phaseolin) patterns (Brown et al. 1982; Gepts et al. 1986). Recently, some studies have suggested that certain groups of snap beans, are actually from the Mesoamerican genepool, the other major genepool of common bean (Skroch and Nienhuis 1995; Cunha et al. 2004; Myers and Davis 2002).

One thing is certain, selection pressure for pod characteristics of snap beans has been intense and in parallel to the selection for seed types in dry beans, there has been selection of large phenotypic differences in snap bean types (Myers and Baggett 1999). Various classes of snap beans exist such as large-sieve green beans, wax types, flat-podded Romano types and fine-sieve types. Snap beans are especially diverse in Europe where pod types go from extra-fine to round with variability in pod color as well as other characteristics (Metais et al. 2002). In addition to their diversity for pod types, snap beans are very diverse in growth habit and plant ideotype, with type I erect bush beans to type IV (indeterminate) pole or climbing beans (Myers and Baggett 1999). In the tropics, snap bean

characteristics are somewhat less developed than in temperate production zones, and most genotypes are indeterminate climbing beans, except when export quality is required in which case imported bush bean varieties are produced.

Snap beans have been analyzed at the molecular level in four previous studies using molecular markers, three of these used RAPD markers to analyze various sets of snap beans, including, ten snap beans of predominantly Andean origin (Skroch and Nienhuis 1995), 77 mostly North American varieties (Cunha et al. 2004) and 75 varieties emphasizing Blue Lake-derived genotypes (Myers and Davis 2002); while the other study used 15 microsatellites to analyze 37 varieties from 6 European commercial classes (Metais et al. 2002). Among these studies, only Myers and Davis (2002) evaluated phaseolin seed storage protein patterns, which are useful for inferring genepool origin and introgression and all previous research has emphasized bush snap beans rather than climbing snap beans (pole types). The International Center for Tropical Agriculture has a collection of snap bean cultivars many of which are indeterminate types, which have not been analyzed before.

Our study was conducted to complement the previous studies by emphasizing climbing snap beans and using a mix of marker systems. Our specific objectives were (1) to evaluate a world-wide collection of publicly available snap bean landraces for their molecular diversity using amplified fragment length polymorphism (AFLP) and microsatellite markers and (2) to determine the extent of inter-genepool exchange through analysis of population structure and seed phaseolin allele distribution.

Materials and methods

Plant material

A total of 127 genotypes were analyzed for this study of which 120 were snap beans, 5 were standard dry beans controls for each genepool that our laboratory regularly uses in diversity studies (Calima, Miss Kelly, G19833 for the Andean genepool, ICA Pijao and DOR364 for the Mesoamerican genepool) and two were wild accessions from Colombia and Guatemala (G21117 and G23441, respectively). Genotypes were from throughout the world with three from Africa, 45 from the Americas, 39 from Asia and 40 from Europe and most were either type III or IV growth habit but varied in seed size (Supplementary Table 1). Heirloom snap bean varieties as listed by the Seed Savers Exchange at http://www.seedsavers.org/ included 'Blue Lake', Kentucky Wonder', 'Genuine Cornfield', Golden Gate Wax', 'Romano Pole', 'Romano Bush' and 'Tendergreen'. All seeds except for those of UNAPAL Milenio were provided by the Genetic Resources Unit of CIAT. Growth habit, seed size and phaseolin pattern data were from Tofiño et al. (2004).

DNA extraction

Ten seeds of each accession were scarified opposite the micropiles and germinated on germination paper in a 0.5 mM solution of calcium sulfate to prevent contamination. After 6 days, the young trifoliate leaves of each seedling were ground in liquid nitrogen using a small plastic pestle. DNA was then extracted with a Quiagen kit (Invitrogen Life Technologies, , Carlsbad, CA, USA) for the AFLP analysis and with a CTAB extraction buffer according to Afanador and Haley (1993) for microsatellite analysis. DNA quality was evaluated on a 1% agarose gel using ethidium bromide staining and DNA concentration was estimated using a Heofer fluorometer. DNA was then diluted to a working stock of 5 ng/ul.

AFLP reactions

AFLP reactions were carried out using the AFLP Analysis system I kit from Invitrogen Life Technologies using 300 ng of DNA per reaction digested for 2 h at 37°C with *Eco*RI and *Mse*I enzymes (5 U each) in one-phor-all 1× restriction buffer as provided by the manufacturer followed by ligation for 2 h at 20°C with kit adaptors and T4 DNA ligase (0.5 U). The digestion reaction was stopped by heat inactivation at 70°C for 15 min while the ligation reaction was diluted 1:5 in water and stored frozen until the pre-selective amplification step. Both reactions were in 12.5 ul final volumes. In a preliminary screening of the selective amplification step, 5 of the snap bean genotypes were tested with four primer combinations (E-AAG/M-CTC, E-ACC/M-CTT, E-ACA/M-CTT and E-AAG/M-CTT) and 2 replications. In each case and for each genotype 5 ul of a 1:10 dilution from the pre-selection step was used for selective primer amplification. PCR conditions were as recommended by Invitrogen and the products were run on 96-well, 4% polyacrylamide gels with silver staining as described in Muñoz et al. (2006). The best primer combination was selected to amplify DNA for 90 of the snap bean genotypes which were selected at random from the full set, thus leaving space on the 96 well gel for molecular weight standards.

Microsatellite analysis

A total of 47 microsatellites were used for the study, 36 of these with fluorescently labeled primers and automated detection as described in Blair et al. (2009) while 11 of these were non-fluorescent microsatellites detected with

silver staining (BM139, BM142, BM164, BM189, BM201, BMd07, BMd26, BMd32, BMd40, BMd54 and Pv-at001). PCR amplifications were carried out in 15 µl reaction volumes with 25 ng of DNA, 2.5 mM of $MgCl_2$, 1.5 µl of 10× PCR buffer [10 mM Tris–HCl (pH 8.0), 50 µM KCl, 0.1% TritonX-100] 10 µM of each primer, 20 mM of dNTPs and 1 U of *Taq* polymerase on PTC 100 or PTC200 MJ Research thermocyclers for both types of markers, but in the fluorescent microsatellite reactions a total of 0.15 µL of forward and reverse primers at 2 µM concentration was used with the appropriate dye-label. The thermocycle in each case consisted of denaturation for 3 min at 94°C, followed by 34 cycles of 92°C for 30 s, 50°C for 30 s and 72°C for 45 s and a final extension of 72°C for 5 min. After amplification, fluorescent microsatellites were mixed together in four color dye panels using 2 µl of each product and 10 µl of HPLC quality water, from which 0.5 µl was diluted 1:6 with water and submitted to an ABI3730 automatic fragment analyzer. Meanwhile, non-fluorescent markers were evaluated on silver-stained polyacrylamide gels as described in Blair et al. (2006) using Sequi-Gen GT electrophoresis units (Biorad, Hercules, CA, USA) run at 120 W for 1 h and Silver Sequence® DNA sequencing System (Promega, Madison, WI, USA).

Data analysis

Fluorescent microsatellite alleles were called with Gene Mapper v. 3.7 software (Applied Biosytems, Foster City, CA, USA), confirmed by manual observation of electropherogram peaks and binned with Allelobin software (http://www.icrisat.org/gt-bt/biometrics.htm) while non-fluorescent and AFLP alleles or bands were called by eye. Allele size estimates were aided by an internal size standard in the case of the fluorescent microsatellites and by a 10 bp molecular weight ladder (Invitrogen) in the case of the non-fluorescent microsatellites. AFLP alleles were determined by evaluating each band for presence and absence. The resulting presence/absence AFLP matrix was then used to calculate Jaccard similarity matrix and create a UPGMA dendrogram with NTSYS v. 2.02 (Rohlf 2002). For microsatellites, a matrix of allele sizes for all the successful genotype × marker combinations was converted to a presence/absence and used in a principal coordinate analysis with SAS statistical software (SAS Institute 1996). A Euclidean distance matrix was then derived from the principal coordinate analysis and used to create another UPGMA dendrogram also with NTSYS. Diversity analysis for both AFLP and microsatellite datasets were conducted with Popgene 1.3 (Yeh et al. 1997) to estimate geneflow (N_m), total heterozygosity (H_T), subpopulation heterozygosity (H_{ST}) and the resulting genetic differentiation (G_{ST}) coefficients based on the genepool divisions identified within the phylogenetic analysis. The same program was used to calculate Nei (1973) gene diversity, Shannon's Indices (I) as well as the number of observed alleles (N_a) and number of effective alleles (N_e) for all markers as well as observed and expected heterozygosity for SSR markers.

Results

Allelic diversity of AFLP and microsatellite markers

The AFLP analysis used in this study was successful at generating highly polymorphic fingerprints for the four primer combinations tested. Polymorphism values for the E-AAG/M-CTC, E-AAG/M-CTT, E-ACA/M-CTT and E-ACC/M-CTT primer combinations were 53.2, 58.2, 64.3 and 67.7% of bands, respectively. Given this evaluation, we determined that the E-ACC/M-CTT combination, the one with the highest polymorphism rate, was the best combination for further analysis of the snap bean genotypes and 90 genotypes were selected to run with this primer combination on a single 96 lane gel to avoid plate to plate PCR variability or gel to gel variation. The selected genotypes were from the same overall range of countries as the full set of 127 genotypes with two from Africa, 32 from the Americas (22 of these from the United States), 29 from Asia (16 of these from China) and 27 from Europe (17 from Western Europe and 10 from Eastern Europe). The selected primer combination generated a total of 74 distinguishable bands of which 50 bands were clearly polymorphic and these were read across the germplasm set so as to generate presence/absence and Jaccard similarity matrices for the genotypes.

Microsatellite analysis was also very successful, both for silver staining gels used for evaluation of 11 non-fluorescent markers and for the four color dye set and marker panels (PET, VIC, 6-FAM, NED) used for evaluation of 36 fluorescent ABI analysis. In both cases, the full set of 127 genotypes was evaluated, with comparable allele sizes estimated by means of molecular weight ladders and control genotypes on each polyacrylamide gel and internal size standards across all wells of the PCR plates analyzed on the ABI3730 automated fragment analyzer.

Allele sizes were within the expected range estimated for these microsatellites characterized by Blair et al. (2006) and in total, all 11 loci of the non-fluorescent microsatellites and 32 loci from the 36 fluorescent microsatellites had comparable alleles and complete allele calls and were therefore included in subsequent analysis, resulting in a data set of 43 microsatellite loci. Within this group of markers, 41 markers were polymorphic and two markers (BMd32 from silver staining and BMd51 from ABI

detection) were monomorphic across the snap bean germplasm. This is equivalent to a polymorphism rate for all markers evaluated of 95.4%.

In terms of allelic diversity, a total of 340 alleles and an average of 8.3 alleles per locus were discovered across all SSR markers with the fluorescent markers detecting a greater number of alleles (282) than the non-fluorescent (58) markers (Table 1). Overall average expected heterozygosity across all markers was 0.543 being similar across the two marker types. The most polymorphic loci were Pv-at001, BM200, BM187 and GATS91 for the fluorescent microsatellites and BM189 and BM201 for the non-fluorescent markers. Average observed heterozygosity was low in both the silver-stained markers (0.031) and the fluorescent markers (0.065) showing the inbred nature of the snap bean germplasm.

Observed heterozygosity was slightly higher for some of the fluorescent microsatellites possibly due to automated allele calls and sensitivity of the ABI detection system, which evaluates double peaks when similar amplification strength is found versus silver-stained gels where only the strongest bands were evaluated. However, for all the markers in both systems, the values were not out of the range expected for common bean given low out-crossing rates in this species.

Genetic diversity of the snap bean genotypes

In the UPGMA dendrogram generated from the Jaccard similarity matrix for the AFLP markers (Fig. 1a), the genotypes were divided into two major groups corresponding to the Andean and Mesoamerican genepools. A total of 17 genotypes were classified as Andean and 73 as Mesoamerican. Differentiation of the genepools by AFLP markers was moderate (Gst = 0.224) given the level of polymorphism and the number of monomorphic bands in the analysis. Nei's gene diversity was found to be 0.2199 overall. Geneflow (Nm) between the groups was found to be 1.733 with the genepools separating at a Jaccard similarity coefficient of 0.65.

Polymorphism was slightly higher for the Mesoamerican genepool than for the Andean genepool in terms of number of observed alleles, number of expected alleles, number of polymorphic loci, percentage polymorphism and average expected heterozygosity (Table 2). However, within the Mesoamerican genepool, comparisons of genotypes showed Jaccard similarity as low as 0.67 while within the Andean genepool Jaccard similarity was only as low as 0.75. This was due to the clustering of the wild Guatemalan genotype, G23441 and two more distant genotypes (G19093 and G16831) within the Mesoamerican group, which increased the diversity detectable in this genepool. G21117, the Colombian wild bean, meanwhile was also

Table 1 Expected and observed heterozygosity for fluorescent and non-fluorescent microsatellite markers evaluated on 120 snap bean genotypes and 7 common bean checks

Marker	Dye[a]	Panel	No alleles	Exp. Het.	Obs. Het.
Genomic					
Fluorescent SSRs					
GATS91	6-FAM	5A	15	0.875	0.021
BM172	PET	5A	7	0.412	0.086
BM188A	NED	5B	4	0.404	0.009
BM188B	NED	5B	11	0.826	Na
BM175	VIC	5B	13	0.629	0.087
BM200	6-FAM	5B	20	0.859	0.027
BM205	PET	5B	9	0.516	0.047
BM156	6-FAM	6A	9	0.722	0.059
BM160	NED	6A	12	0.641	0.038
AG01	VIC	7A	6	0.301	0.041
BM140	6-FAM	7A	10	0.666	0.130
GATS54	NED	7A	3	0.309	0.059
BM187	VIC	7B	15	0.795	0.179
BM183	NED	7B	11	0.851	0.105
BM143	PET	8A	12	0.714	0.197
BM149	6-FAM	8A	2	0.264	0.009
BM141	6-FAM	8B	10	0.720	0.056
BMd56	NED	9A	3	0.488	0.028
Average			9.6	0.611	0.065
Gene-based					
BMd08	NED	5A	5	0.317	0.016
BMd20	VIC	5A	7	0.495	0.058
Pv-ctt001	VIC	6A	7	0.708	0.151
BMd02	PET	6B	5	0.678	0.171
BMd16	VIC	6B	5	0.255	0.089
Pv-ag003	NED	6B	4	0.093	0.000
BMd17	PET	7B	5	0.758	Na
Pv-at001	6-FAM	7B	30	0.944	0.048
BMd15	NED	8B	6	0.261	0.028
BMd47	PET	8B	9	0.624	0.112
Pv-cct001	VIC	8B	15	0.041	0.008
BMd18	6-FAM	9A	4	0.064	0.016
Pv-at003	VIC	9A	8	0.534	0.145
Average			8.5	0.444	0.065
Overall average			9.1	0.541	0.065
Non-fluorescent SSRs					
BM139			6	0.507	0.032
BM142			5	0.462	0.000
BM164			7	0.516	0.000
BM189			10	0.827	0.066
BM201			9	0.778	0.032
Average			7.4	0.618	0.026
Gene-based					
BMd07			3	0.575	0.000
BMd26			5	0.567	0.079
BMd40			4	0.232	0.000
BMd46			3	0.510	0.042
BMd54			6	0.546	0.057
Average			4.2	0.486	0.036
Overall average			5.8	0.552	0.031

Data generated with PopGene (Yeh et al. 1997)

Exp. Het expected heterozygosity, *Obs. Het.* observed heterozygosity

[a] The fluorochromes represented the following dyes PET = Red, NED = Yellow, VIC = Green, 6-FAM = Blue

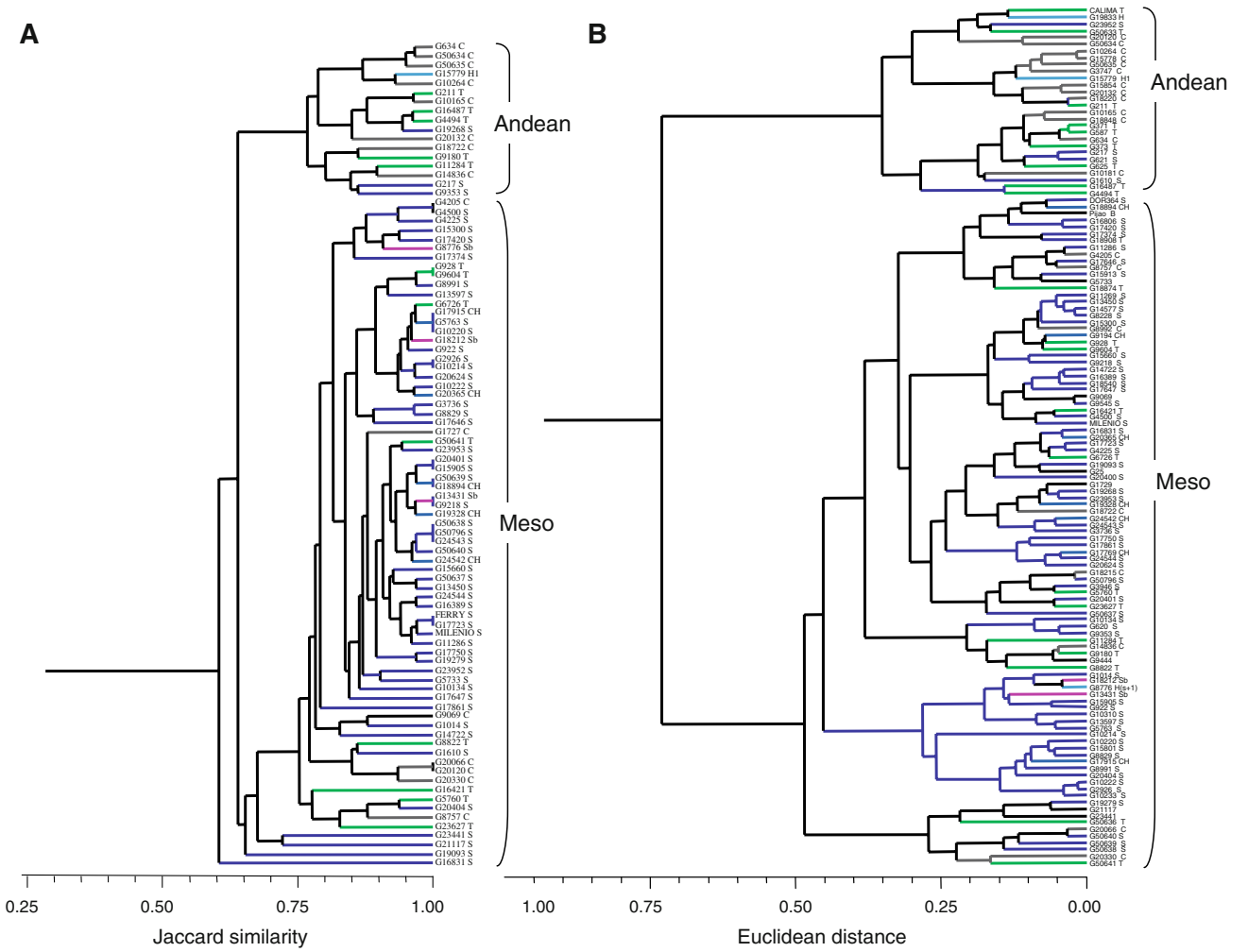

Fig. 1 UPGMA dendrograms for snap beans and common bean controls based on analysis of AFLP markers (**a**; 90 genotypes) with Jaccard similarity coefficient and SSR markers (**b**; 127 genotypes) with Euclidean distance. *Branch shading* indicate different phaseolin alleles (S, Sb, CH, B, C, T, H, H1) from each genepool (Andean or Mesoamerican)

associated with the Mesoamerican group but was distinct in banding pattern from the Guatemalan wild bean and also added to the diversity of this genepool.

Control genotypes were found in the Andean and Mesoamerican genepool groups with ICA Pijao in the lower part of the dendrogram (Mesoamerican group) and Calima (G4494) in the upper part of the dendrogram (Andean group). Some groupings within each genepool shown in the AFLP dendrogram were also of interest. For example, within the Mesoamerican group several Blue Lake genotypes clustered together with UNAPAL Milenio, which is derived from this snap bean variety. Some clustering within the Mesoamerican or Andean groups was also associated with phaseolin pattern with some intermediate genotypes having Andean type phaseolin such as "C" and "T" but grouping with the Mesoamerican genotypes having "S" and "CH" phaseolins.

The dendrogram based on SSR marker analysis (Fig. 1b) had a similar division of the germplasm, with 28 genotypes in the Andean group and 99 genotypes in the Mesoamerican group, but Nei's gene diversity was much greater with these highly polymorphic SSR markers than with the lower polymorphism AFLP markers (Table 2). This was reflected in the overall polymorphism rate within each genepool and for the entire germplasm set was 95.4% of the SSR loci. As we observed for the AFLP analysis, the number of observed alleles and the number of expected alleles was higher for the Mesoamerican genepool genotypes than for the Andean genepool genotypes. For the SSR-based dendrogram, the two genepools separated at a Euclidean genetic distance of 0.5–0.75. Genetic distance within the Andean genepool ranged up to 0.35 when compared with 0.50 within the Mesoamerican genepool. Genetic differentiation (Gst) between the genepools based

Table 2 Diversity parameters for amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR) markers evaluated across snap beans and controls represented by 90 and 127 genotypes, respectively

Analysis	N_a	N_e	Polymorphic loci	Polymorphism (%)	Nei's	I
AFLP markers						
Andean ($n = 17$)	1.40	1.1878	20	27.0	0.1147	0.1772
Mesoamerican ($n = 73$)	1.88	1.3229	44	67.5	0.2092	0.3330
Total ($n = 90$)	1.88	1.3484	44	67.5	0.2199	0.3461
SSR markers						
Andean ($n = 28$)	4.25	2.5076	39	90.70	0.4193	0.8514
Mesoamerican ($n = 99$)	6.97	2.8718	41	95.35	0.4698	1.0253
Total ($n = 127$)	7.83	3.1481	41	95.35	0.5206	1.1359

Diversity parameters include N_a number of observed alleles. N_e number of expected alleles (Kimura and Crow 1964), polymorphic loci and % polymorphism, *Nei's* gene diversity of Nei (1973), as well as I Shannon's Information index

on the SSR analysis was found to be 0.173 similar to the value for the AFLP analysis while geneflow (Nm) was found to be 1.192.

Distribution of phaseolin alleles was analyzed with respect to the groups identified by both AFLP and microsatellite analyses (Table 3). In the AFLP analysis, the Andean group showed a mixture of "C" (47.1%), "T" (23.5%) and "H1" (5.9%) phaseolin alleles typical of the Andean genepool and "S" (23.5%) phaseolin that is typical of the Mesoamerican genepool. Meanwhile the Mesoamerican group had "S" (65.8%), "CH" (6.8%) and "Sb" (4.1%) phaseolin that are from the genepool and a number of introgressed Andean phaseolin alleles, namely "T" (12.2%) and "C" (8.2%). The same situation existed for the dendrogram constructed with the microsatellite markers, where 14.2% of the Andean genotypes having "S" phaseolin and 24.2% of the Mesoamerican genotypes having either "T", "C" or "H" phaseolin types. One of the genotypes analyzed (G8776 from the United States) had a hybrid phaseolin pattern, "H1" and was found in the Mesoamerican group according to the microsatellite analysis. Growth habit was not very distinct between the genepools since mostly indeterminate climbing and semi-climbing beans were selected for this study (Table 3), but there was some tendency of type III beans to be clustered in the Andean group and type IV beans to be clustered in the Mesoamerican group.

Table 3 Phaseolin allele and growth habit type distribution among the Andean and Mesoamerican groups of snap beans analyzed in this study

Classification		AFLP		SSRs	
		Andean	Mesoamerican	Andean	Mesoamerican
Phaseolin	Allele[a]				
S	Mesoamerican	3	50	4	60
Sb	Mesoamerican	0	3	0	2
CH	Mesoamerican	0	5	0	7
B	Mesoamerican	0	0	0	1
C	Andean	8	7	13	10
T	Andean	5	8	9	13
H	Andean	0	0	1	1
H1	Andean	1	0	1	0
NA	A/M	0	0	0	5
Growth habit	Type[b]				
I	Andean	3	3	3	3
II	A/M	2	2	1	5
III	A/M	8	11	20	17
IV	A/M	4	56	4	74
NA	NA	0	1	0	0
Total		17	73	28	99

[a] Allele classification as Andean, Mesoamerican or both genepools (A/M) based on CIAT Genetic Resources Unit

[b] Assignment of growth habit types to genepools based on Singh (1989)

Discussion

Both marker systems used in this study for the analysis of snap beans were efficient; however, microsatellites were more polymorphic than AFLP markers. In the diversity analyses for both the AFLP and microsatellite markers, the genotypes were divided into two major groups corresponding to the Andean and Mesoamerican genepools. Both the AFLP and SSR analyses showed that a larger number of genotypes fell within the Mesoamerican genepool (almost two-thirds of the genotypes), showing that snap beans exist within each genepool. The AFLP fingerprint based on the E-ACC/M-CTT combination was moderately polymorphic and more useful than the combinations used by Beebe et al. (2001), which were of low polymorphism in Andean bean types. The selective primer combination we used was also used by Maciel et al. (2003) for AFLP analysis of Brazilian genotypes and therefore appears to be particularly useful in common bean genetic diversity studies.

Microsatellites were superior to AFLP analysis in terms of their very high polymorphism and ability to make gel to gel comparisons for the larger germplasm set. Expected heterozygosity for SSR analysis was about three to four times that of the AFLP analysis and also double previous analysis with isozymes by Tofiño et al. (2004), which was only 0.268. The detection of the fine scale divisions within each snap bean group was an advantage of the microsatellite evaluation that was also observed with snap bean germplasm by Metais et al. (2002).

The combination of fluorescent and non-fluorescent markers in the snap bean germplasm analysis was another aspect of this work. We found the combination advantageous since detection was rapid and required small PCR reaction volumes for the first type of marker, while the second technique was more economical. In addition, for both types of markers we targeted a mix of highly polymorphic and less-polymorphic markers giving an accurate picture of diversity. This was partly due to the specific selection of a wide range of polymorphic markers for the fluorescent panels as discussed in Blair et al. (2009). The application of fluorescent microsatellite marker technology to dry beans was also found to be useful by Masi et al. (2003) and is widely used in other crops, such as rice (Blair et al. 2002; Coburn et al. 2002).

As observed in several of our previous studies (Blair et al. 2006; Diaz and Blair 2006), non-coding, genomic microsatellites were more polymorphic than coding, gene-based microsatellites both in terms of number of alleles and expected heterozygosity. The highest expected heterozygosity was found with the same markers that had a large number of alleles and the two values were correlated for both genomic and gene-based markers evaluated with either system. Microsatellites, as codominant, multi-allelic and highly polymorphic markers that are amenable to high-throughput technologies like fluorescent detection, are proving to be very useful for determining commercial class identity in both dry and snap beans (Metais et al. 2002; Blair et al. 2009). Meanwhile, AFLP markers, although dominant and multi-copy, are useful for determining Andean versus Mesoamerican genepool identity (Tohme et al. 1996).

Several discoveries were made about the genetic diversity of the snap bean genotypes that were evaluated with the AFLP and microsatellite analyses. First, the pole type snap beans were found to be fairly diverse with AFLP analysis having similar levels heterozygosity as observed by Muñoz et al. (2006) for *P. vulgaris* as a whole. As expected, snap beans showed high diversity for microsatellite markers comparable with or higher than that detected by Metais et al. (2002) and Blair et al. (2006) for snap and dry beans, respectively. The high level of diversity in the snap beans we analyzed may be associated with their selection to be representative of pole types from many regions of the world.

A second observation was that the snap bean germplasm we analyzed was found to belong to both Andean and Mesoamerican genepools; and more importantly, the Mesoamerican background was predominant over the Andean background in the germplasm set we analyzed. In other words, a larger number of the indeterminate snap bean genotypes were found in the Mesoamerican group rather than in the expected Andean group. This result confirms that snap beans are not exclusively or always predominantly Andean as was suggested by early authors such as Brown et al. (1982) and Gepts et al. (1986). This seems to be especially true for climbing snap beans (often known as pole beans in the United States). These results expand on those of various authors who found some evidence for partial Mesoamerican ancestry in a some commercial snap beans mostly from the United States (Skroch and Nienhuis 1995; Cunha et al. 2004) but show that much of the indeterminate snap bean germplasm around the world from Asia to North America is actually mostly of Mesoamerican origin. Our AFLP results showed that 80% of the pole type snap bean genotypes were Mesoamerican while microsatellite analysis also validated the theory that snap beans can have both Andean and Mesoamerican genetic backgrounds. Myers and Davis (2002) found similar results with a specific group of snap beans from the Pacific Northwest of the United States, commonly known as Oregon Bush Blue Lake varieties, which were derived from climbing beans of uncertain European or early American origin.

Blue Lake snap beans are notable for their high processing quality, dark green pods and mild flavor (Myers and Baggett 1999). Therefore, six 'Blue Lake'-derived

varieties were analyzed in our study, one from Colombia released as 'UNAPAL Milenio' as a tropical representative of this snap bean germplasm group along with five other temperate type IV Blue Lake varieties from the CIAT genebank (G8228, G8992, G9545, G17646 and G17723). In the SSR analysis, four of these clustered together showing that the Blue Lake type is a subgroup within the Mesoamerican genepool. One other type III Blue Lake variety (G17646) was not as tightly clustered, but likewise was Mesoamerican and not Andean. Kentucky Wonder types (G922, G928, G3946, G4500, G5763, G6726, G8829, G8991 and G9218) were found in various clusters but all of the Mesoamerican genepool. The US heirloom varieties, G5760 (Golden Gate Wax), G8776 (Genuine Cornfield) and G9069 ('Burpee's Green Pod Stringless') were also found in the Mesoamerican genepool as were several French and German pole snap beans such as 'Enna', 'Fin de Montreux', 'Impuls' as well as 'Ejote' from Mexico. Andean group indeterminate snap beans such as 'Carnaval de Venecia' were mostly from Europe (Hungary, Italy, Portugal, Spain, Yugoslavia) but also from Asia (China and Turkey).

A third observation of our study was that phaseolin alleles typical of one genepool, known to be "T", "C" or "H" in the Andean group or "S" in the Mesoamerican group were introgressed from one genepool to the other in many of the snap bean accessions. In the SSR analysis, 4 out of 28 Andean group genotypes and 22 out of 99 Mesoamerican group genotypes had the phaseolin alleles characteristic of the opposite genepool. For example, accessions such as 'Golden Gate Wax' and Kentucky Wonder Wax' had the "T" phaseolin typical of the Andean genepool but grouped with the Mesoamerican genotypes. Introgression of phaseolin alleles from one genepool to another was also found in Caribbean germplasm, which shows many intermediate Andean/Mesoamerican morphological characteristics (Durán et al. 2005). Diversity analyses from Cunha et al. (2004) and Metais et al. (2002) also provides evidence of some inter-genepool introgression into North American and European commercial snap beans. Introgression and sampling of both Andean and Mesoamerican genepools may explain why we found the snap bean genotypes from Europe and Asia to be fairly diverse and may justify the statement by Singh (1989) that greatest variation for snap beans occurs "among cultivars from China, Europe, and the United States".

Inter-genepool introgression in our snap bean study was higher than in previous analysis of dry beans from the Mesoamerican or Andean genepool (Diaz and Blair 2006; Blair et al. 2007; Zhang et al. 2008) as evidenced by geneflow values, which were above 1 in this study and below 0.5 in those previous studies. Introgression, between the genepools in snap beans may be more likely than in dry beans since seed size and color constraints are different in snap beans than in dry beans. For example, large seed size is often selected against in snap beans so that pod shape is not affected by the developing seeds. This is especially true in the fine and extra-fine categories of snap beans where large or wide seed would cause lumpiness in the overall shape of the snap bean pods and where very narrow and thin cylindrical seed have been selected for as a way of maximizing the succulent green pod wall as a proportion of the total pod (Myers and Baggett 1999). As a result many small-sieve varieties have seed of light overall weight more typical of the Mesoamerican genepool. In addition, seed color in traditional snap beans can be highly variable since the pod is consumed before the seed color is fully developed. However, dark seed color is often selected against in modern snap beans and many snap beans today are white seeded so that the seed will be less noticeable in the harvested snap bean pod.

If we assume that selection pressure for seed size in snap beans are different than in dry beans, then phaseolin alleles, which are associated with seed size were set free to be introgressed from one genepool to another. Indeed this is what we found in both the AFLP and microsatellite analyses, where there was evidence of phaseolin allele swapping between the genepools. Some introgression of phaseolin may have been the result of direct selection and breeding efforts with the aim of producing an Andean type bean with narrower seed. In support to this, we found more Mesoamerican phaseolin alleles in the Andean genepool than vice versa. Introgression of Mesoamerican phaseolin types into an Andean background may also have been due to linkage with the *p* gene for white seed color which has been widely used in snap bean breeding and is linked to the *Phs* locus (Myers and Baggett 1999). In general, the distinction between the large-seeded Andean and small-seeded Mesoamerican types has been blurred in snap beans and the introgression of phaseolin alleles from one genepool to the other and the high seed size variability in the snap beans are symptoms of this.

In conclusion, we have discovered the important contribution that Mesoamerican beans have made to indeterminate snap bean germplasm and have uncovered high levels of introgression between the genepools that has occurred for these genotypes based on AFLP and microsatellite marker analysis, suggesting that selection of characteristics from both genepools has been important to the generation of snap bean traits. Evidence for snap bean ancestry having both Mesoamerican and Andean origins exists in seed size variability and has been observed before by Myers and Baggett (1999). Furthermore, isoenzymatic data for eight loci (ACP, DIAP, 6-PGDH, PGI, PRX, MDH, ME and SKDH) evaluated by Tofiño et al. (2004) on the same set of snap bean genotypes as this study support

the conclusion that inter-genepool introgression is common in indeterminate snap beans. On the other hand, diversity is often low within certain groups of snap beans such as the French bean types (Metais et al. 2002) due to selection for specific quality traits.

These results have important implications for breeding of snap beans since crosses between snap bean groups is likely to produce widely divergent recombinants especially where Andean and Mesoamerican background materials are used together which may increase genetic diversity but produce many off types in terms of fresh marker or processing quality. As an example of this, Myers and Davis (2002) suggested that crosses between Mesoamerican-derived Blue Lake varieties and Andean snap bean germplasm rarely results in usable segregants. Therefore, intra-genepool crosses will need to be guided by careful classification of snap bean germplasm and marker analysis, which can help determine which crosses are likely to succeed.

Acknowledgments We wish to thank Pilar Marquez and Jaime Eduardo Muñoz for advice on the study; and Myriam Cristina Duque and Lucy Díaz for help with diversity analyses. We also thank César Ocampo and Daniel Debouck for help with phaseolin analysis and seed supply.

References

Afanador L, Haley S (1993) Adoption of a "mini-prep" DNA extraction method for RAPD's marker analysis in common bean *Phaseolus vulgaris*. Bean Imp Coop 36:10–11

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

Blair M, Hedetale V, McCouch S (2002) Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). Theor Appl Genet 105:449–457

Blair M, Giraldo M, Buendía H, Duque MC, Beebe S (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 113:100–109

Blair MW, Díaz JM, Hidalgo R, Díaz LM, Duque MC (2007) Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 116:29–43

Blair MW, Díaz LM, Buendia HF, Duque MC (2010) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.) Theor Appl Genet (in press)

Brown JWS, McFerson JR, Bliss FA, Hall TC (1982) Genetic divergence among commercial classes of *Phaseolus vulgaris* in relation to phaseolin pattern. HortScience 17:752–754

Coburn J, Temnykh S, Paul E, McCouch S (2002) Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.). Crop Sci 42:2092–2099

Cunha C, Hintz T, Griffiths P (2004) Genetic diversity of snap bean cultivars determined using randomly amplified polymorphic DNA (RAPD) markers. HortScience 39(3):481–484

Diaz L, Blair M (2006) Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. Theor Appl Genet 114:143–154

Durán LA, Blair MW, Giraldo MC, Machiavelli R, Prophete E, Nin JC, Beaver JS (2005) Morphological and molecular characterization of common bean (*Phaseolus vulgaris* L.) landraces from the Caribbean. Crop Sci 45:1320–1328

FAOSTAT (2007) http://faostat.fao.org/. Accessed 17 Apr 2009

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

Henry G, Janssen W (1992) Snap beans in the developing word: proceedings of an international conference held in Cali, Colombia, p 366

Kaplan L, Kaplan LN (1988) *Phaseolus* in archeology. In: Gepts P (ed) Genetic resources of Phaseolus Benas. Kluwer, Dordrecht, pp 125–142

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Maciel FL, Echeverrigaray S, Sheng Gerald LT, Grazziotin FG (2003) Genetic relationships and diversity among Brazilian cultivars and landraces of common beans (*Phaseolus vulgaris* L.) revealed by AFLP markers. Gen Res Crop Evol 50:887–893

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

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

Muñoz C, Duque MC, Debouck D, Blair MW (2006) Taxonomy of tepary bean (*Phaseolus acutifolius*) and wild relatives as determined by amplified fragment length polymorphism (AFLP) markers. Crop Sci 46:1744–1754

Myers JR, Baggett JR (1999) Improvement of snap beans. In: Singh S (ed) Common bean improvement for the 21st century. Kluwer, Boston

Myers J, Davis J (2002) Molecular polygenetics of snap bean. Department of Horticulture, Oregon State University, Corvallis

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Nat Acad Sci 70:3321–3323

Pachico D (1987) The potencial of snap beans as a crop for small farmers in the tropics. El Mejoramiento Genético de la Habichuela en América Latina, CIAT. Cali- Colombia. pp 33–43

Rohlf F (2002) NTSYS pc: numerical taxonomy system. Exeter Publishing, Setauket

SAS Institute (1996) SAS/STAT user guide 6.11. SAS Institute, Cary

Silbernagel MJ (1986) Snap bean breeding. In: Basset MJ (ed) Breeding vegetable crops. Avi Publ. Co. Inc, West Connecticut, pp 243–282

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

Skroch P, Nienhuis J (1995) Qualitative and quantitative characterization of RADP variation among snap bean (*Phaseolus vulgaris*) genotypes. Theor Appl Genet 91:1078–1085

Tofiño AP, Ocampo C, Toro O (2004) Determination of genetic diversity of Snap beans *Phaesolus vulgaris* L. cultivated at secondary centers of domestication using morphological and biochemical descriptors. Ann Rep Bean Improv Coop 47:169–170

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 (2000) Mejoramiento genetico de frijol (*Phaseolus vulgaris* L.) Legado de las variedades de America Latina 1930–1999. CIAT, Cali-Colombia

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

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