ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2012) **10**(2); 108–118
doi:10.1017/S1479262112000081

Microsatellite markers reveal promising genetic diversity and seed trait associations in common bean landraces (*Phaseolus vulgaris* L.) from Nicaragua

O. R. Jiménez[1,2] and H. Korpelainen[2]*

[1]*National Center of Agricultural Research and Biotechnology (CNIAB), Nicaraguan Institute of Agricultural Technology (INTA), Km 14.1 North Highway, Managua, Nicaragua and* [2]*Department of Agricultural Sciences, University of Helsinki, PO Box 27 (Latokartanonkaari 5), FI-00014 Helsinki, Finland*

Received 7 November 2011; Accepted 30 March 2012

Abstract

Nicaragua is located in the Mesoamerican diversity centre for common beans (*Phaseolus vulgaris* L.). Yet, there is insufficient knowledge of the molecular characteristics of most common bean landraces in Nicaragua. The objectives of the present study were to investigate the genetic diversity of common bean landraces and to identify promising sources of genetic variation for breeding purposes. Two cultivars and 40 landraces never studied before were selected from a collection based on the geographic origin, seed coloration and information provided by farmers. Fourteen microsatellite markers distributed in different linkage groups were analysed. The study revealed that there is a high genetic diversity (mean 8.9 alleles per locus). The populations showed structuring into three groups where seed weight had a strong relationship with population clustering. At least 20% of the populations hold promising allelic variation and potential for good market acceptance that could be maximized in breeding programmes. Additionally, four markers revealed a high correlation with seed length, width and weight, suggesting that marker-assisted selection for these yield-determinant traits could be straightforward. Nonetheless, more marker–trait associations should be addressed in order to enforce this practice.

Keywords: genetic diversity; landraces; microsatellite markers; *Phaseolus vulgaris*; plant breeding

Introduction

Evidence obtained during recent years suggests that the common bean (*Phaseolus vulgaris* L.) was domesticated in Mesoamerica and the Andes, but its secondary centre of genetic diversity probably extended to Brazil, China and Europe (Chacón *et al*., 2005; Zhang *et al*., 2008; Burle *et al*., 2010; Santalla *et al*., 2010). After domestication, this species has become one of the most important crop plants in developing countries, as it is an economical source of important nutritional components (Santalla *et al*., 1999).

Nicaragua is located in the Mesoamerican centre of genetic diversity for *Phaseolus* species, where a high genetic diversity is expected to prosper in diverse environmental conditions. Some studies have aimed to describe the genetic diversity content in a small number of Nicaraguan landraces (Gómez *et al*., 2004, 2005). However, most of the genetic diversity remains without any estimation and is undervalued as a potential source of

* Corresponding author. E-mail: helena.korpelainen@helsinki.fi

genetic variation. On the other hand, many landraces and old cultivars that were quite popular some years ago are in danger to become extinct and some of them, according to recent expeditions, have already been lost (Oswalt R. Jiménez, personal observation). Thus, a proper estimation of the genetic diversity of on-farm conserved landraces is urgently needed. They have evolved along with farmers' preferences, as their subsistence relies on the beans' good adaptation capacity and high culinary quality, traits that are appreciated in national and international markets. Nicaraguan landraces produce very low yields (Gómez and Frankow-Lindberg, 2005) when compared with improved lines and cultivars, and this characteristic is the main disadvantage for producing these beans. However, despite yield being influenced by the environment, this trait can be improved if proper levels of genetic variation are present in breeding programmes (Falconer and Mackay, 1996; Acquaah, 2007).

The main bean cultivars that are produced in Nicaragua nowadays were improved by regional bean breeding programmes using germplasm from different sources (Jiménez, 2009). These programmes produce new cultivars for the Central American region where the preferences and environmental conditions vary among different countries. After validating the promising genetic material in Nicaraguan conditions, most advanced lines do not meet the requirements (distinctness, uniformity and stability) and market acceptance, and only a very few lines are finally released as new cultivars.

The identification of novel sources of genetic variation and their use in local breeding programmes can justify and further enhance the conservation of locally adapted bean genetic resources in countries where a robust conservation strategy is still missing. The objectives of the present study were to investigate the genetic diversity of Nicaraguan common bean landraces using microsatellite markers and to identify promising sources of genetic variation for breeding purposes.

Materials and methods

Germplasm collection and population selection

Between March and May 2010, four expeditions were carried out to different regions of Nicaragua with the aim to establish a seed bank. When visiting the farms, small amounts of seeds were requested from the farmers and passport data were recorded. From the information obtained, the reasons why farmers prefer and produce those populations were especially valued in the registers. We sampled each bean population by taking at least 300 seeds from different bean bags, respectively, in order to have a representative sample. The amount of seeds collected from each population varied from 300 to 1000 g (depending on seed availability). The seeds were cleaned and their physiological qualities were tested at the National Center of Agricultural Research and Biotechnology (CNIAB). The 100-seed weights, among other seed traits, were determined following the ISTA rules (ISTA, 2004). Seed length and width were determined as in Blair *et al.* (2009). The accessions were then conserved in a cold room (12°C) at the CNIAB. Geographic origin, diverse seed coloration and high level of acceptance by farmers were the criteria for selection. Forty out of 200 bean accessions collected during the expeditions, including three Tepary bean (*Phaseolus acutifolius* A. Gray) populations, were chosen for this study (Table 1). According to farmers' statements and an information review, all chosen populations have never been studied before. Thus, they represent a novel source of information regarding Nicaraguan bean genetic resources. Additionally, breeder's seeds from two cultivars, 'INTA ROJO' and 'INTA FUERTE SEQUIA', were included as reference populations.

DNA extraction

DNA extraction from individual seedlings (germinated in sterile sand), ten randomly sampled individuals per accession, was carried out at the Biotechnology Laboratory at the CNIAB in Nicaragua. The mini-preparation protocol (Dellaporta *et al.*, 1983) was modified to be used in common beans. About 15 mg of leaf tissue was taken from each individual and placed into a sterile Eppendorf tube. Thereafter, 200 µl of cold miniprep II extraction buffer (containing 100 mM Tris–HCl, 50 mM EDTA, 500 mM NaCl and 20 mM 2-mercaptoethanol) was added, and the tissue was macerated and homogenized using a plastic pestle. Then, 26 µl of sodium dodecyl sulphate (10%) was added and mixed. Subsequently, the samples were placed into a water bath (65°C) for 10 min. Then, 120 µl of 3 M potassium acetate was added, and the samples were incubated on ice for 25 min and centrifuged at full speed (14,800 ***g***) for 15 min. Then, 200 µl of the supernatant was removed carefully and placed into a clean Eppendorf tube avoiding the debris. Afterwards, 120 µl of extra-pure isopropanol was added, and the samples were kept at −24°C for 30 min and centrifuged at full speed (14,800 ***g***) for 15 min. Then, isopropanol was removed, and the pellet was washed twice with ethanol (70%) and left to dry. The pellet was dissolved in 100 µl of TE (containing 10 mM Tris–HCl, pH 8.0 and 1 mM EDTA). Finally, the samples (ten individual DNA samples per accession, 420 in total) were placed into a freezer at −24°C.

https://doi.org/10.1017/S1479262112000081 Published online by Cambridge University Press

Table 1. Forty bean landraces selected from 200 accessions in an *ex situ* collection, their seed features and origin in Nicaragua

No.	Code	Common name	Seed coloration	100-Seed weight (g)[a]	Place of origin
1	PV0001	SRC 1 12 182[b]	Red	15.7	Siuna
2	PV0002	Chile pálido (1)	Light red	16.8	Yahoya, Siuna
3	PV0003	Chile pálido (2)	Light red	16.0	Yahoya, Siuna
4	PV0004	Chile pálido (3)	Light red	21.6	Mongallo, Siuna
5	PV0005	Chile pálido (4)	Light red	23.3	Coperna, Siuna
6	PV0006	Rojo pálido	Light red	23.2	Negro Was, Siuna
7	PV0007	Chile pálido (5)	Light red	24.1	El Hormigero, Siuna
8	PV0008	Chile pálido (6)	Light red	17.5	Wanní, Siuna
9	PV0009	Chile pálido (7)	Light red	17.2	El Hormiguero, Siuna
10	PV0010	Chile pálido (8)	Light red	19.3	Siuna
11	PV0011	Rojo claro	Light red	18.4	Río Blanco, Matagalpa
12	PV0012	Frijol pinto	Red and cream	20.6	Matagalpa
13	PV0013	Chile claro	Light red	20.8	Matagalpa
14	PV0014	Frijol cuarenteño (1)	Red	19.2	Diria, Granada
15	PV0015	Frijol mono	Brown	20.3	Diriomo, Granada
16	PV0016	Canelo[b]	Dark red	20.7	La Florida, Boaco
17	PV0017	Congo	Red	21.7	Boaco
18	PV0018	Frijol guaba	Dark red	21.7	Muelle de los bueyes
19	PV0019	Sangre de toro	Light red	16.6	La Libertad, Chontales
20	PV0020	Chilamatillo	Pale reddish	18.9	Nueva Guinea
21	PV0021	Frijol cuarenteño (2)	Red	22.6	Santa Lucia, Boaco
22	PV0022	Gualiceño	Cream	22.5	Masaya
23	PV0023	Rojo ligero	Light red	18.7	El Sauce, León
24	PV0024	Achis negro	Dark red	19.7	El Sauce, León
25	PV0025	Rojo seda	Light red	21.1	Matagalpa
26	PV0026	Rojo criollo (1)	Red	21.1	Jinotega
27	PV0027	Rojo criollo (2)	Red	26.7	Matagalpa
28	PV0028	Rojo criollo (3)	Red	16.7	Las Flores, Masaya
29	PV0029	Rojo criollo (4)	Red	20.5	Rancho Grande, Matagalpa
30	PV0030	Rojo criollo (5)	Red	19.0	Wiwili, Jinotega
31	PV0031	Rojo criollo (6)	Red	22.4	Condega, Estelí
32	PV0032	Rojo criollo (7)	Red	23.3	Condega, Estelí
33	PV0033	Rojo criollo (8)	Red	20.3	Condega, Estelí
34	PV0034	Rojo criollo (9)	Light red	16.6	Santa Teresa, Carazo
35	PV0035	Rojo criollo (10)	Light red	13.5	Santa Teresa, Carazo
36	PV0036	Rojo criollo (11)	Red	14.8	La Conquista, Carazo
37	PV0037	Frijol pinto[c]	Cream and red	36.7	Condega, Estelí
38	PA0001	Frijol blanco (1)[d]	White	12.6	Masaya
39	PA0002	Frijol blanco (2)[d]	White	12.1	Belén, Rivas
40	PA0003	Frijol pinto[d]	Black and white	12.0	Belén, Rivas

[a] The highest 100-seed weights were found in cultivars 'INTA ROJO' (24.5 g) and 'INTA FUERTE SEQUIA' (27.3 g). Seed lengths and widths averaged 1.04 and 0.61 cm, respectively. [b] Apparently an old cultivar. [c] Possibly a species different from *Phaseolus vulgaris* L. [d] Tepary bean (*Phaseolus acutifolius* A. Gray).

Microsatellite genotyping

Microsatellite genotyping was carried out in the laboratory of the Department of Agricultural Sciences, University of Helsinki, Finland. The following 14 microsatellite markers were selected for this study: BM205, AG1, BM154, BM156, BM184, BM189, BM210, BM212, BM114 (Gaitán *et al.*, 2002), PVag001, PVag004 (Yu *et al.*, 2000), BMd8, BMd53 (Blair *et al.*, 2003) and ATA10 (Blair *et al.*, 2008). The marker selection was mainly based on the high level of polymorphisms reported in previous studies and the markers' wide distribution in the common bean genome. Also, these markers have been associated with important QTLs (Quantitative trait loci) for yield components when assayed in materials from different genetic backgrounds (Blair *et al.*, 2006b, 2009, 2010; Rodrigues *et al.*, 2007; Pereira *et al.*, 2008; Torga *et al.*, 2010).

Polymerase chain reactions (PCRs) were carried out in 10 µl volumes by mixing the following components: 1 µl of 10 × buffer, 0.2 µl of dNTPs (10 mM each), 6 µl of MQ water (Millipore Quality water), 1 µl of each primer

https://doi.org/10.1017/S1479262112000081 Published online by Cambridge University Press

(5 pmol, forward primers fluorescently labelled with FAM (6-carboxyfluorescein), TET (Tetrachlorofluorescein) or HEX (Hexachlorofluorescein) labels), 0.3 μl of DNA polymerase (Dynazyme, 2U/μl) and 0.5 μl DNA template (about 30 ng). The PCRs were carried out as follows: DNA denaturation at 94°C for 4 min followed by 30 cycles of denaturation at 94°C for 45 s, annealing at 46–58°C (depending on the primer pair) for 45 s, and elongation at 72°C for 1 min, with a final elongation at 72°C for 10 min. After amplification, the PCR products were diluted with MQ water at the 1:200 ratio. Of this solution, 0.5 μl of each PCR product was mixed with 20 μl of HiDi-formamide and 0.15 μl of size standard (GeneScan 500 ROX). After mixing, the samples were denatured for 5 min at 95°C. Finally, DNA fragments were analysed using a capillary electrophoresis system (3730 DNA Analyzer; Applied Biosystems) in the Sequencing Laboratory of the Institute of Biotechnology, University of Helsinki, Finland. The allele sizes were determined using the software Peak Scanner version 1.0 (Applied Biosystems).

Statistical analyses

For all populations, the distribution of genetic variation was revealed by an analysis of molecular variance (AMOVA), and the observed and expected heterozygosities (H_{obs} and H_{exp}) were determined using the software Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010). Genetic diversity parameters were estimated for a set of populations (total number of alleles, alleles per locus, number of genotypes and F_{IS} index), and allele frequencies were calculated and compared using Fisher's exact test with software Genepop version 4.1 (Rousset, 2008). To estimate pairwise differences among populations, F_{ST} values were calculated using the software Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010). Then, the pairwise matrix was clustered by the unweighted pair group method of arithmetic averages (UPGMA), and the phylogenetic tree was plotted using the software MEGA version 5.1 (Tamura *et al.*, 2011).

Subsequently, we conducted a Bayesian analysis of population structure with the software Structure version 2.3.3 (Pritchard *et al.*, 2000). We applied an admixture and correlated allele frequencies model where individuals may have mixed ancestry and allele frequency at each locus is correlated along the populations. To determine the number of clusters (K value), we first tested a continuous series of Ks (1–12) in five independent runs for each K value, with a length of burn-in period of 10,000 iterations followed by 100,000 MCMC (Markov chain Monte Carlo) iterations. The five log-likelihood values for each K were then charted to infer the K values around a plateau of the likelihood values. The identified candidate K values were further tested in ten independent runs for each K with a burn-in period of 100,000 iterations followed by 1,000,000 iterations. The most likely K value was determined by analysing ΔK values using the method proposed by Evanno *et al.* (2005). Populations with a proportion of membership less than 0.8 were considered putative hybrids (Santalla *et al.*, 2010).

Additionally, considering that seed features (length, width and weight) are the only quantitative traits available from the collection, we implemented a germplasm-regression-combined marker–trait association to test specific genotype associations with these traits as proposed in Ruan *et al.* (2009) and Ruan (2010). The Tepary bean accessions and the population PV0037 were excluded from these analyses, because the former represents another species and the latter possibly belongs to a different genetic background (based on seed features). Basically, quantitative traits, seed length, width and weight, were treated as dependent variables as implemented by Virk *et al.* (1996). All microsatellite genotypes were scored into a binary dataset (1 for presence or 0 for absence) and then considered as independent variables. Stepwise multiple regression analysis (MRA) was based on the following model:

$$Y = a + b_1 m_1 + b_2 m_2 + , \ldots, b_j m_j, \ldots, b_n m_n + d + e,$$

where Y represents quantitative traits; m_j represents marker genotypes; b_j represents partial regression coefficients that specify empirical relationships between Y and m_j; d represents population residuals; and e is the random error of Y that includes environmental variation (Virk *et al.*, 1996). F values with P values between 0.045 and 0.099 were employed to enter and remove independent variables from the regression equation, respectively (Ruan *et al.*, 2009; Ruan, 2010). R^2 denotes the coefficient of determination. Multiple regression analyses were carried out using the software SPSS version 16.0 (SPSS, Chicago, IL, USA; http://www.spss.com). After the regression analyses, the selected marker genotypes were tested with linear curve fitting using linear models to confirm the significance of Beta statistics (β) for each genotype identified by the MRA.

Results

Allelic and genotypic diversity

The microsatellite markers produced good PCR products, except for the loci BM205 and ATA10 that did not produce any PCR product in Tepary bean accessions. The loci BM205 and PVag004 produced a multi-banding pattern (two and between four and six DNA fragments, respectively). As the locus PVag004 produced a complex

https://doi.org/10.1017/S1479262112000081 Published online by Cambridge University Press

banding pattern, it was excluded from the genetic analyses. One of the amplified fragments produced by the locus BM205 was monomorphic in the whole array of populations and, consequently, also excluded from the data analyses. AMOVA showed that most genetic variation (64.3%) was present among the populations, while 35.7% of the variation occurred within them. The thirteen microsatellite markers revealed 115 different alleles in total. The average of alleles per locus was 8.9, ranging from 2 (BMd53) to 27 (BM154). A total of 134 genotypes in all populations were identified (Table 2). The highest number of genotypes was displayed at the locus BM154 (29 genotypes). The locus PVag004 producing multibanding patterns amplified ten different DNA fragments (96, 98, 104, 198, 200, 202, 204, 238, 240 and 242 bp). When the allele frequencies of all populations were compared, Fisher's exact test revealed highly significant differences in frequencies at all loci ($P = 0.000$). Once the average H_{exp} and H_{obs} values were contrasted, it was evident that the observed heterozygosities were very low (0.034) compared with the expected heterozygosities (0.423). The average F_{ST} value was 0.625 and the F_{IS} value was estimated to be 0.914. Overall, when the number of alleles detected in this study was compared with the results obtained when the markers were reported for the first time, five out of 13 markers displayed a higher number of alleles (Table 2).

When populations were assessed individually, landraces PV0006, PV0013, PV0028, PV0023 and PV0024 were found to possess the highest amount of allelic variation, excluding Tepary bean populations and PV0037, ranging from 27 to 31 alleles (mean 2.4 alleles per polymorphic locus). When the four most polymorphic markers (BM156, BM154, BM184 and BM114) were scored, populations PV0006, PV0013, PV0023, PV0024 and PV0027 averaged from three to four alleles per locus. On the other hand, landraces PV0005, PV0031 and PV0035 proved unrelated to other populations, i.e. they possessed allele combinations not shared with other populations.

Genetic structure

The UPGMA tree derived from the pairwise F_{ST} values showed that there is moderate genetic differentiation among the populations. The Tepary bean populations, PA0001, PA0003 and PA0002, were clearly plotted as an outgroup in the tree. In the same way, the population PV0037 displayed a different microsatellite profile when compared with the other populations (Fig. 1). The genetic distances suggest that the landrace PV0006 is closely related to the cultivar 'INTA FUERTE SEQUIA'. In general, the phylogenetic clustering of the whole set of populations does not reflect the origin of the populations, as populations from different agroecological regions could be located in the same phylogenetic branch of the tree.

The Bayesian analyses confirmed the phylogenetic tree clustering, capturing most of the genetic diversity into three groups ($K = 3$). The cluster membership was assigned as follows: cluster 1 involved Tepary bean populations (PA0001, PA0002 and PA0003) and the population PV0037. Clusters 2 and 3 contained common bean populations, 15 out of 38 populations belong to cluster 1 (39.5%) and 23 out of 38 to cluster 3 (60.5%). Cluster 2 comprised a slightly higher allelic diversity (on average

Table 2. Summary of allele and genotype data obtained from the 14 microsatellite loci in 42 bean populations

Marker	Linkage group	Allelic range (bp)	Number of alleles	Number of alleles reported[a]	Number of genotypes
BM205	b11	131–300	5	7	8
AG1	b03	134–354	4	8	5
BM156	b02	206–274	11	10	14
BMd8	b04	174–178	3	3	4
BM212	b10	196–228	7	7	8
BM189	b08	102–110	5	8	8
BM154	b09	206–284	27	8	29
BM210	b07	143–189	10	12	11
BM184	b11	149–349	12	6	14
ATA10	b06	104–300	7	5	7
PVag001	b11	136–278	10	2	11
PVag004[b]	b04	96–242	–	–	–
BM114	b09	224–370	13	8	14
BMd53	b05	103–106	2	5	2

[a] Alleles reported in Yu *et al.* (2000), Gaitán *et al.* (2002) and Blair *et al.* (2003).
[b] Complex banding pattern, thus excluded from the analyses.

https://doi.org/10.1017/S1479262112000081 Published online by Cambridge University Press

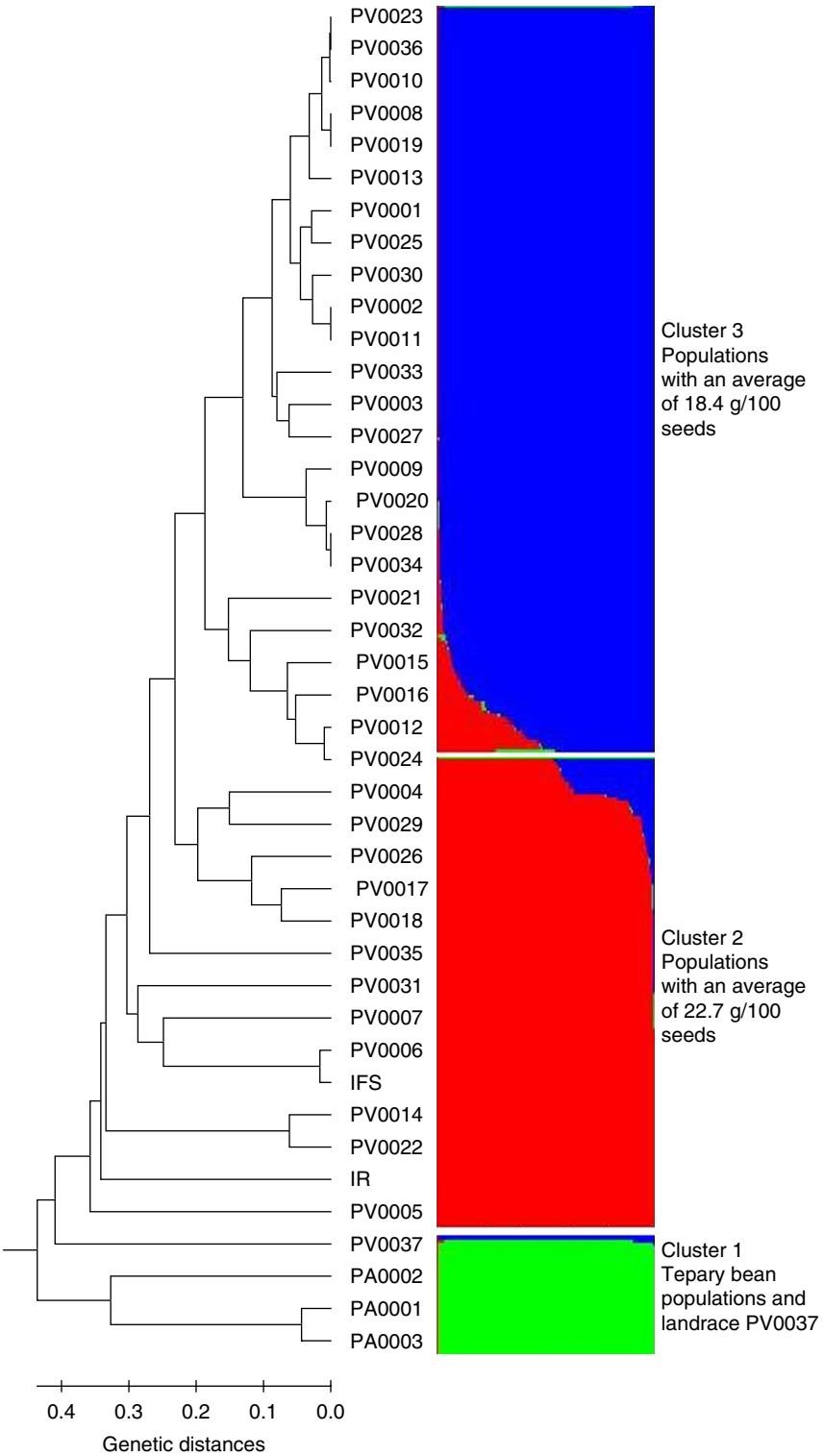

Fig. 1. UPGMA tree describing the genetic relatedness among the 40 common bean landraces and two cultivars from Nicaragua (left), and estimated population structure ($K = 3$) sorted by the membership coefficient (Q value) in correspondence to phylogenetic inference (right). Cluster 1 contains *Phaseolus acutifolius* populations and landrace PV0037, cluster 2 contains populations with relatively large seeds and cluster 3 contains populations with relatively small seeds on average. IR = 'INTA ROJO'; IFS = 'INTA FUERTE SEQUIA'. A colour version of this figure can be found online at http://www.journals.cambridge.org/pgr.

https://doi.org/10.1017/S1479262112000081 Published online by Cambridge University Press

22.5 alleles per population across the 13 loci) compared with cluster 3 (21.4 alleles). Most genetic variation in these clusters is ascribed to within-population variation (69.7%) with a F_{ST} value of 0.321, as revealed by AMOVA. Cluster 2 included two reference cultivars, 'INTA ROJO', 'INTA FUERTE SEQUIA' and the landraces PV0004, PV0005, PV0006, PV0007, PV0016, PV0017, PV0018, PV0021, PV0024, PV0026, PV0027, PV0029 and PV0031. Finally, cluster 3 encompassed landraces PV0001, PV0002, PV0003, PV0008, PV0009, PV0010, PV0011, PV0012, PV0013, PV0014, PV0015, PV0019, PV0020, PV0022, PV0023, PV0025, PV0028, PV0030, PV0032, PV0033, PV0034, PV0035 and PV0036 (Fig. 1). The populations PV0015, PV0020, PV0024, PV0028, PV0030, PV0034 and PV0036 possessed a coefficient of membership less than 0.8 and were considered as hybrid populations under admixture structure. The proportion of non-hybrid populations was 83.3%. As common bean populations were inferred into two clusters, we conducted an AMOVA omitting populations from cluster 1. These results showed that 47.7% of the variation represents within-population variation and 53.3% among-population variation.

Association of microsatellite markers with seed features

Stepwise MRA was conducted to outline the correlation of 107 microsatellite genotypes with the seed length, width and weight of 38 common bean populations. Three stepwise runs were programmed in order to get the best independent variables that explain the variation. After discarding unsuitable and hybrid genotypes, six alleles from four microsatellite markers (BM205, AG1, BM156 and PVag001) explained most of the phenotypic variation (Table 3). The associations of alleles with seed weight were also tested with a curve-fitting programme, which confirmed a linear relationship.

For seed length, allele *138* from the marker AG1 (called as $AG1_{138}$), allele *224* from the marker BM156 (called as $BM156_{224}$) and allele *278* from the marker PVag001 (called as $PVag001_{278}$) showed a correlation with seed length (Table 3). Genotypes $AG1_{138}$ and $PVag001_{278}$ showed negative correlations and genotype $BM156_{224}$ showed a positive correlation. Genotype $PVag001_{278}$ showed the highest ($R^2 = 0.214$) significant ($P = 0.003$, $t = -3.133$) correlation with a high standardized β value of -0.463. When the genotypes $BM156_{224}$ and $AG1_{138}$ were added to the model, the correlation increased ($R^2 = 0.339$ and 0.423, respectively). For seed width, only allele *222* from the marker BM156 (called as $BM156_{222}$) showed a positive ($R^2 = 0.175$) significant ($P = 0.009$, $t = 2.767$) correlation with this trait, with a standardized β value of 0.419.

For seed weight, two alleles from the marker BM205, alleles *131* and *133* (called $BM205_{131}$ and $BM205_{133}$) showed significant correlations with seed weight (Table 3; Fig. 2). The genotype $BM205_{131}$ showed the greatest ($R^2 = 0.340$) highly significant ($P = 0.000$, $t = 4.304$) positive correlation with seed weight. The standardized β coefficient was also high (0.583). When both genotypes were included in the model, the correlation increased ($R^2 = 0.429$). Information concerning correlations, regressions and ANOVAs for these four markers is presented in Table 3.

Discussion

In the present study, we investigated the genetic diversity of 40 common bean landraces from Nicaragua and two

Table 3. Results of stepwise multiple regression analyses conducted for microsatellite genotypes associated with seed features

			Stepwise MRA				ANOVA	
Trait[a]	Correlation	Microsatellite genotype[b]	R^2 [c]	R^2 change[d]	F change[e]	P value of F change	F value	P value
Seed length	Negative	$PVag001_{278}$	0.214	0.214	9.817	0.003	9.817	0.003
	Positive	$BM156_{224}$	0.339	0.125	6.614	0.015	8.981	0.001
	Negative	$AG1_{138}$	0.423	0.084	4.924	0.033	8.300	0.000
Seed width	Positive	$BM156_{222}$	0.175	0.175	7.655	0.009	7.655	0.009
Seed weight	Positive	$BM205_{131}$	0.340	0.340	18.527	0.000	18.527	0.000
	Negative	$BM205_{133}$	0.429	0.089	5.445	0.025	13.130	0.000

MRA, multiple regression analysis.
[a] Within each trait, different parameters were estimated when genotypes were included in succeeding MRA steps. [b] Subscripts correspond to fragment sizes (bp) for each locus. [c] R^2 is the coefficient of determination, which expresses the amount of variation explained by the independent variable. [d] R^2 change is the change in R^2 statistics. [e] F change is the change in F statistic that are produced when an independent variable is added or deleted.

https://doi.org/10.1017/S1479262112000081 Published online by Cambridge University Press

cultivars using microsatellite markers, and identified promising populations for breeding purposes. The detected level of genetic variation (mean 8.9 alleles per locus) was higher than that previously reported for small red-seeded landraces (the same market class as studied here) in Nicaragua, 5.7 and 4.3 alleles per locus (Gómez *et al*., 2004, 2005), and higher than that reported in most other studies on common beans where microsatellite markers have been used (Blair *et al*., 2006a; Díaz and Blair, 2006; Benchimol *et al*., 2007; Zhang *et al*., 2008; Díaz *et al*., 2010; Santalla *et al*., 2010; Cabral *et al*., 2011; García *et al*., 2011; Avila *et al*., 2012). Typically, lower levels of genetic variation have been detected in common bean populations when compared with other self-pollinated species (Santalla *et al*., 2010). A very high degree of genetic diversity has been discovered by Blair *et al*. (2009), on average 18.4 alleles per locus. However, this study was conducted on a core collection holding accessions from different species, centres of origin and races. In more standard microsatellite comparisons, averages ranging from 2.8 to 7.8 have been found (Blair *et al*., 2006a; Díaz and Blair, 2006; Benchimol *et al*., 2007; Zhang *et al*., 2008). Thus, it is reasonable to suggest that the bean populations studied here contain a high amount of genetic variation. Additionally, the level of genetic variation detected for five markers (BM156, BM154, BM184, PVag001 and BM114) exceeds that reported for these markers in previous literature (Yu *et al*., 2000; Gaitán *et al*., 2002).

On the other hand, the locus PVag004, presumably associated with the arcelin, phytohaemagglutinin and α-amylase inhibitor gene family, exhibited a multi-banding pattern and, consequently, was excluded from the analyses. However, its alleles *184*, *195* and *207 bp*, presumably associated with resistance to bruchids (*Zabrotes subfasciatus* (Boheman) and *Acanthoscelides obtectus* (Say)) and previously found in advanced bean lines of a Mesoamerican origin (Blair *et al*., 2010a), were not discovered in this study. This finding supports the general idea that there is no resistance for bruchids in most of the landraces and old cultivars produced in Nicaragua. Alleles for bruchid resistance may remain in wild relatives, which have not been subjected to domestication processes.

Populations PV0006, PV0013, PV0020, PV0021, PV0023, PV0024, PV0027, PV0029 and PV0031, in addition to having a high level of allelic variation, are classified in national and international markets as '*rojo chile*' or '*rojo nacional*' because of their good culinary quality and light red seed coloration of some of these populations (similar to colours 2.5R4/10, 5R3/8 and 5R3/10 in the Munsell colour charts for plant tissue, 1977). Thus, there are sufficient arguments to suggest that at least 20% of the landraces presented here are promising sources of variation and also have a high market potential.

There was a good correspondence between the genetic clustering pattern displayed in the UPGMA tree and the grouping by structure analysis, even though they use different genetic parameters to infer population structures. Structure analysis is a Bayesian model-based method that uses genotypes from unlinked markers, demonstrating the presence of a population structure, identifying distinct genetic populations, assigning individuals to populations, and identifying migrants and admixed individuals (Pritchard *et al*., 2000). In contrast, the UPGMA tree was inferred based on pairwise genetic distances among populations (Tamura *et al*., 2011). Nonetheless, both approaches brought similar results for genetic clustering. Only six out of 42 populations mismatched these two clustering proposals (PV0014, PV0016, PV0021, PV0022, PV0027 and PV0035).

The number of clusters detected in the structure analysis was less than expected ($K = 3$). Basically, common bean populations were structured into two main groups. AMOVA analysis omitting populations from cluster 1 showed that the within-population variation is 47.7%, i.e. about half of the variation represents the variation among populations within clusters. The geographic origin of these landraces was a criterion for selection in this study. We tried to cover most Nicaraguan regions where bean production takes place. However, the region of origin does not influence the genetic structure, as populations from different agroecological regions can be found in the same cluster. On the contrary, other studies have found a good congruence between the genetic structure and geographic location (Santalla *et al*., 2010). In addition, it was surprising that concerning seed weights, eight out the ten populations with the biggest seeds were plotted into cluster 2 and the ten populations with the smallest seeds were plotted into cluster 3. This tendency proposes that seed weights are connected with the population structure. Such marker–trait relationship agrees with other studies that have found that seed weight is the main factor influencing the genetic structure (Díaz and Blair, 2006; Santalla *et al*., 2010).

Farmers who perform on-farm conservation modify the genetic structure of landraces by selection in response to their preferences and interests (Negri and Tiranti, 2010). Similarly, when African landraces (introduced from Mesoamerican and Andean centres) were analysed, there were clues to ascribe many changes in genetic structure to farming conditions and preferences for specific types of beans, expressed in seed sizes and colours (Asfaw *et al*., 2009; Blair *et al*., 2010b). Under the perspective that seed weight is a trait related to quality of sowing material and particular food preferences, it is a trait probably affected strongly by Nicaraguan farmers' selection.

https://doi.org/10.1017/S1479262112000081 Published online by Cambridge University Press

Another interesting observation was obtained from marker BM114. This marker possessed 13 different alleles in total, the same number found by Blair *et al.* (2006a) when working on almost the same number of populations (43). The allele *248 bp* had a frequency above 0.8 exclusively in the populations with a known shorter period of time to start flowering. These populations are quite popular among Nicaraguan farmers and they are called '*frijoles cuarenteños*', the name suggesting that the populations start to flower earlier than improved cultivars. Precisely, this marker was also previously associated with QTLs for days to flowering in populations mapped by Blair *et al.* (2006b).

Finally, the structure inference and the genetic parameters estimated from the studied populations suggest that most genetic diversity described for common bean landraces could be more efficiently captured by selecting a few numbers of populations in equal numbers from both clusters using an appropriate number of individuals. MRA showed that genotypes $AG1_{138}$, $BM156_{224}$, $BM156_{222}$, $PVag001_{278}$, $BM205_{131}$ and $BM205_{133}$ are correlated with seed features (seed length, width and weight). The first two traits influence the final seed weight. As an example, when five populations with biggest seeds and five populations with smallest seeds are plotted, it is noticeable that the model based on genotype $BM205_{131}$ explains a good proportion of the observed variation (Fig. 2). An increase in the frequency of this genotype predicts gains in seed weights. Markers AG1, BM205 and BM156 have been previously mapped in Mesoamerican populations and highly significant associations with QTLs for seed weight have been found ($P < 0.001$; Blair *et al.*, 2006b, 2009). In the same way, marker BM156 has been reported to be associated with high grain productivity when identifying QTLs for high yield in beans of the type Carioca (Rodrigues *et al.*, 2007; Pereira *et al.*, 2008; Torga *et al.*, 2010).

Similar approaches to reveal germplasm-regression-combined marker–trait associations as used here have been reported with successful results in many other species and traits, such as rice (*Oryza sativa* L.), alfalfa (*Medicago sativa*), oat (*Avena sativa* L.) and sea buckthorn (*Hippophae* L.) (Virk *et al.*, 1996; Maureira-Butler *et al.*, 2007; Achleitner *et al.*, 2008; Ruan *et al.*, 2009). The validation of the linkage between important traits

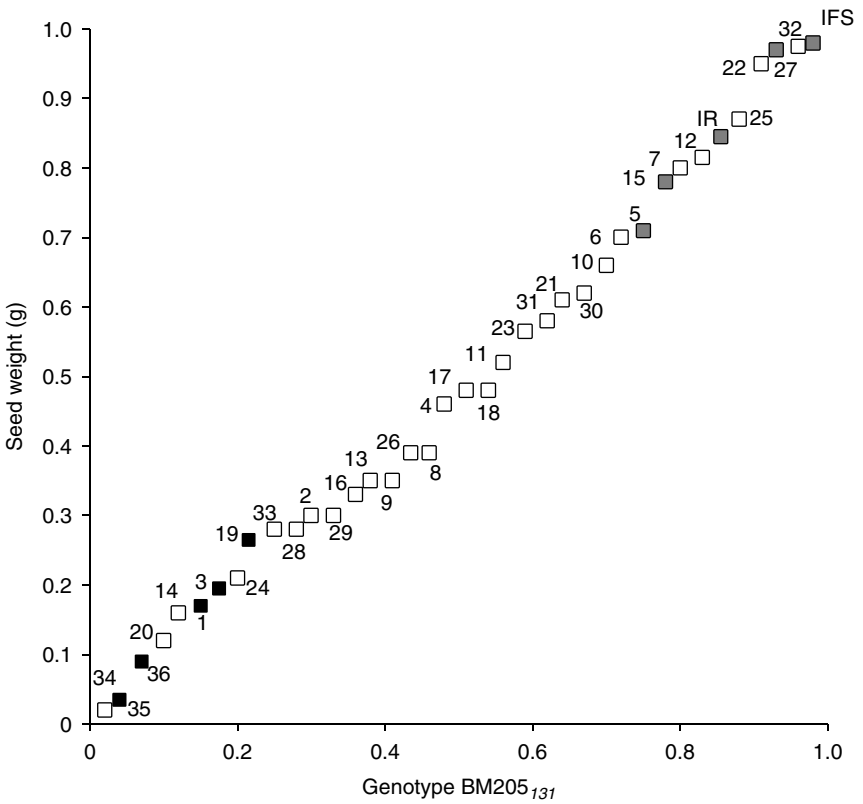

Fig. 2. Regression ($P = 0.000$) plotted for seed weight and the frequency of the microsatellite genotype $BM205_{131}$ in 38 common bean populations (see Table 1 for population information). Black squares represent the five populations possessing the smallest seeds (36, 35, 1, 3 and 19), and grey squares represent the five populations with the largest seeds (38, 37, 27, 7 and 5). IR = 'INTA ROJO'; IFS = 'INTA FUERTE SEQUIA'.

https://doi.org/10.1017/S1479262112000081 Published online by Cambridge University Press

and molecular markers is crucial not only for plant breeding proposes, but also for the characterization of potential germplasm. The high correlations found in this study in common beans suggest that four of the markers examined here could be important for marker-assisted selection for seed weight in segregating populations derived from our collection. Of course, it is necessary to test a higher number of markers and to include additional quantitative traits (pods per plant, seeds per pod, seeds per plant, for instance) in order to have a better estimation of breeding potential. Yet, an analysis of the available trait and marker information could be advantageous before performing extensive phenotypic testing. Seed weight is an important yield-determinant component in common beans (White and Izquierdo, 1991; Dalla-Corte *et al.*, 2010). Even though there are many factors influencing the yield, a high diversity in seed weight is an important indicator to consider when selecting potential genetic variation.

In conclusion, there is promising genetic diversity in the common bean collection described in this study. This genetic variation is higher than that reported in most previous studies, which highlights the importance to conserve these materials. A special attention should be paid to those diverse populations that are attractive in national and international markets and are present in both discovered clusters. This diversity could be maximized in breeding programmes as long as this molecular information is included as criteria for germplasm selection. The significant correlation of four microsatellite markers with seed features suggests that marker-assisted selection for these quantitative traits could be straightforward. Yet, additional marker–trait associations with other yield components should be addressed in order to enforce this practice.

Acknowledgements

We would like to thank the authorities from the Nicaragua–Finland Agrobiotechnology Program (NIFAPRO), the Finnish Ministry of Foreign Affairs, the University of Helsinki and the Embassy of Finland in Nicaragua for supporting this study. We are grateful to the National Agrarian University, Nicaragua for collaborating with us during germplasm collection. Finally, we extend our gratitude to bean farmers who kindly contributed small amounts of seeds for this study.

References

Achleitner A, Tinker NA, Zechner E and Buerstmayr H (2008) Genetic diversity among oat varieties of worldwide origin and associations of AFLP markers with quantitative traits. *Theoretical and Applied Genetics* 117: 1041–1053.

Acquaah G (2007) *Principles of Plant Genetics and Breeding*. London: Blackwell Publishing, p. 569.

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

Avila T, Blair MW, Reyes X and Bertin P (2012) Genetic diversity of bean (*Phaseolus*) landraces and wild relatives from the primary centre of origin of the Southern Andes. *Plant Genetic Resources: Characterization and Utilization* 10: 83–92.

Benchimol LL, De Campos T, Morais SA, Colombo CA, Chioratto AF, Fernandes E, Lima LR and Pereira A (2007) Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. *Genetic Resources and Crop Evolution* 54: 1747–1762.

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

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

Blair MW, Iriarte G and Beebe S (2006*b*) QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (*Phaseolus vulgaris* L.) cross. *Theoretical and Applied Genetics* 112: 1149–1163.

Blair MW, Buendía HF, Giraldo MC, Métais I and Peltier D (2008) Characterization of AT-rich microsatellites in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 118: 91–103.

Blair MW, Diaz LM, Buendia HF and Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 119: 955–972.

Blair MW, Prieto S, Diaz LM, Buendia HF and Cardona C (2010*a*) Linkage disequilibrium at the APA insecticidal seed protein locus of common bean (*Phaseolus vulgaris* L.). *BMC Plant Biology* 10: 79.

Blair MW, Gonzales LF, Kimani PM and Butare L (2010*b*) Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. *Theoretical and Applied Genetics* 121: 237–248.

Burle ML, Fonseca JR, Kami JA and Gepts P (2010) Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. *Theoretical and Applied Genetics* 121: 801–813.

Cabral PDS, Soares TCB, Lima ABP, De Miranda FD, Souza FB and Gonçalves LSA (2011) Genetic diversity in local and commercial dry bean (*Phaseolus vulgaris*) accessions based on microsatellite markers. *Genetics and Molecular Research* 10: 140–149.

Chacón MI, Pickersgill B and Debouk DG (2005) Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theoretical and Applied Genetics* 110: 432–444.

Dalla-Corte A, Moda-Cirino V, Arias CAA, De Toledo JFF and Destro D (2010) Genetic analysis of seed morphological traits and its correlations with grain yield in common

bean. *Brazilian Archives of Biology and Technology* 53: 27–34.

Dellaporta SJ, Wood J and Hicks JB (1983) A plant DNA minipreparation: version II. *Plant Molecular Reporter* 1: 19–21.

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

Díaz LM, Buendía HF, Duque MC and Blair MW (2010) Genetic diversity of Colombian landraces of common bean as detected through the use of silver-stained and fluorescently labelled microsatellites. *Plant Genetic Resources: Characterization and Utilization* 9: 86–96.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.

Excoffier L and Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Falconer DS and Mackay TFC (1996) *Introduction to Quantitative Genetics*. Malaysia: Prentice Hall, p. 464.

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

García RAV, Rangel PN, Brondani C, Martins WS, Melo LC, Carneiro MS, Borba TCO and Brondani RPV (2011) The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*. *BMC Genetics* 12: 41.

Gómez OJ and Frankow-Lindberg BE (2005) Yield formation in Nicaraguan landraces of common bean compared to bred cultivars. *Journal of Agricultural Science* 143: 369–375.

Gómez OJ, Blair MW, Frankow-Lindberg BE and Gullberg U (2004) Molecular and phenotypic diversity of common bean landraces from Nicaragua. *Crop Science* 44: 1412–1418.

Gómez OJ, Blair MW, Frankow-Lindberg BE and Gullberg U (2005) Comparative study of common bean (*Phaseolus vulgaris* L.) landraces conserved *ex situ* in genebanks and *in situ* by farmers. *Genetic Resources and Crop Evolution* 52: 371–380.

ISTA (2004) *International Rules for Seed Testing*. Glattbrugg: Seed Science and Technology.

Jiménez OR (2009) Genetic purity of the common bean (*Phaseolus vulgaris* L. cv. 'INTA ROJO') during seed production in Nicaragua Master's Thesis, University of Helsinki, Finland.

Maureira-Butler IJ, Udall JA and Osborn TC (2007) Analyses of a multi-parent population derived from two diverse alfalfa germplasm: test cross evaluations and phenotype-DNA associations. *Theoretical and Applied Genetics* 115: 859–867.

Munsell Color (1977) *Munsell Color Charts for Plant Tissues*. 2nd edn revised. Baltimore, MD: Macbeth Divisions of Kollmorgen Corporation.

Negri V and Tiranti B (2010) Effectiveness of *in situ* and *ex situ* conservation of crop diversity. What a *Phaseolus vulgaris* L. landrace case study can tell us? *Genetica* 138: 985–998.

Pereira HS, Dos Santos JB, De Souza TP and Lima IA (2008) Phenotypic and marker-assisted selection of common bean families with high grain yield. *Pesquisa Agropecuaria Brasileira* 43: 1551–1558 (In Portuguese).

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

Rodrigues TB, Dos Santos JB, Ramalho MAP, Amorim EP and Silva NO (2007) QTLs identification in common bean through SSR markers affected by natural selection. *Ciência e Agrotecnologia* 31: 1351–1357 (In Portuguese).

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.

Ruan CJ (2010) Germplasm-regression-combined marker–trait association identification in plants. *African Journal of Biotechnology* 9: 573–580.

Ruan CJ, Li H and Mopper S (2009) Characterization and identification of ISSR markers associated with resistance to dried-shrink disease in sea buckthorn. *Molecular Breeding* 24: 255–268.

Santalla M, Fueyo MA, Rodino A, Montero I and De Ron A (1999) Breeding for culinary and nutritional quality of common bean (*Phaseolus vulgaris* L.) in intercropping systems with maize (*Zea mays* L.). *Biotechnology, Agronomy, Society and Environment* 3: 225–229.

Santalla M, De Ron AM and De La Fuente M (2010) Integration of genome and phenotypic scanning gives evidence of genetic structure in Mesoamerican common bean (*Phaseolus vulgaris* L.) landraces from the southwest of Europe. *Theoretical and Applied Genetics* 120: 1635–1651.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using likelihood, distance, and parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.

Torga PP, Dos Santos JB, Pereira HS, Furtado D and Leite ME (2010) Selection of common bean families based on grain type and yield, and QTLs information. *Ciência e Agrotecnologia* 34: 95–100 (In Portuguese).

Virk PS, Ford-Lloyd BV, Jackson MT, Pooni HS, Clemeno TP and Newbury HJ (1996) Predicting quantitative variation within rice germplasm using molecular markers. *Heredity* 76: 296–304.

White JW and Izquierdo J (1991) Physiology of the yield potential and stress tolerance. In: van Schoonhoven A and Voysest O (eds) *Common Beans: Research for Crop Improvement*. Wallingford/Cali, Colombia: CABI/CIAT, pp. 287–383.

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

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