



Developing a common bean core collection suitable for association mapping studies

Juliana Morini Küpper Cardoso Perseguini^{1,2}, Gliciane Micaele Borges Silva², João Ricardo Bachega Feijó Rosa³, Rodrigo Gazaffi⁴, Jéssica Fernanda Marçal², Sérgio Augusto Morais Carbonell⁵, Alisson Fernando Chiorato⁵, Maria Imaculada Zucchi⁶, Antonio Augusto Franco Garcia³, and Luciana Lasry Benchimol-Reis^{1,2}

¹*Departamento de Genética e Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.*

²*Centro de Recursos Genéticos Vegetais, Instituto Agronômico de Campinas, Campinas, SP, Brazil.*

³*Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP, Brazil.*

⁴*Departamento de Biotecnologia Vegetal, Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras, SP, Brazil.*

⁵*Centro de Grãos e Fibras, Instituto Agronômico de Campinas, Campinas, SP, Brazil.*

⁶*Agência Paulista de Tecnologia dos Agronegócios, Centro de Desenvolvimento Tecnológico, Instituto Agronômico, Campinas, SP, Brazil.*

Abstract

Because of the continuous introduction of germplasm from abroad, some collections have a high number of accessions, making it difficult to explore the genetic variability present in a germplasm bank for conservation and breeding purposes. Therefore, the aim of this study was to quantify and analyze the structure of genetic variability among 500 common bean accessions to construct a core collection. A total of 58 SSRs were used for this purpose. The polymorphism information content (PIC) in the 180 common bean accessions selected to compose the core collection ranged from 0.17 to 0.86, and the discriminatory power (DP) ranged from 0.21 to 0.90. The 500 accessions were clustered into 15 distinct groups and the 180 accessions into four distinct groups in the Structure analysis. According to analysis of molecular variance, the most divergent accessions comprised 97.2% of the observed genetic variability present within the base collection, confirming the efficiency of the selection criterion. The 180 selected accessions will be used for association mapping in future studies and could be potentially used by breeders to direct new crosses and generate elite cultivars that meet current and future global market needs.

Keywords: molecular markers, genetic diversity, genetic structure, microsatellites, *Phaseolus vulgaris* L.

Received: April 17, 2014; Accepted: October 8, 2014.

Introduction

Common bean (*Phaseolus vulgaris* L.) is a species of great agronomic interest, as it is an important grain legume for human consumption worldwide (Angioi *et al.*, 2010). This species was domesticated by Middle American and South American Andean cultures (Gepts *et al.*, 1986a; Gepts, 1998) and has progressively dispersed worldwide (Angioi *et al.*, 2010; Asfaw *et al.*, 2009). Bitocchi *et al.* (2012) suggested a Mesoamerican origin of the common bean.

Burle *et al.* (2010) pointed out Brazil as a secondary center of common bean diversity. In Brazil, the common bean most likely came from at least two different routes, as indicated by the occurrence of both small and large beans (Gepts, 1998). Nonetheless, beans of Mesoamerican genetic origin are preferred by most of the population, and this preference is shown by the dominance of carioca and black bean types in their diets.

The narrow genetic base of modern crop cultivars is a serious obstacle to sustaining and improving crop productivity due to the vulnerability of genetically uniform cultivars to potentially new biotic and abiotic stresses (Abdurakhmonov and Abdugarimov, 2008). Plant germplasm resources worldwide, including wild plant species, modern cultivars, and their wild crop relatives, are important reservoirs of natural genetic variations. The Common Bean Germplasm Bank of the Agronomic Institute (IAC, Campi-

nas, S.P. Brazil) holds more than 1800 accessions representing the two principal centers of origin (Andean and Mesoamerican) and includes ecotypes from different South American countries and a large number of lines from both Brazilian and international genetic improvement programs (Chiorato *et al.*, 2006).

Association mapping, also known as linkage disequilibrium (LD)-based association mapping (Mackay and Powell, 2007; Zhu *et al.*, 2008; Myles *et al.*, 2009), has been proposed as an alternative to quantitative trait locus (QTL)-mapping. The LD associates single DNA sequence changes with traits of interest using collections of unrelated individuals. It is rapid and cost effective as many alleles may be assessed simultaneously, resulting in higher resolution mapping. It uses most of the recombination events that occur over time, while avoiding the need to expensively conduct crossing of populations. Field evaluation and use of large germplasm collections for associative mapping are mostly constrained by problems related to accession redundancy, economic cost, and time. Assessment of genetic resources, thus, could be more rational if focused on a subset of accessions, or the so-called core collection, which includes the maximum variability of the base collection with the minimal possible size (Frankel and Brown, 1984; Spagnoletti-Zeuli and Qualset, 1993; van Hintum *et al.*, 2000).

A core collection is formed by selecting a small percentage of the original collection that will represent most of the total genetic variation with minimum redundancy (Brown, 1995). The principal steps to establish a core collection are as follows: (a) determine the size of the core subset; (b) divide the collection into distinct groups; and (c) select entries in each group to form the core collection. The complexity of establishing a core subset is a function of the available data and applied sampling procedure (Brown, 1989a,b; Brown and Spillane, 1999). The established core collection must be validated to ensure its adequacy and usefulness by assessing whether the characteristics and variability of the entire collection have been maintained. Comparison of the entire and the core collection properties is accomplished using mean, variance, frequency, and distribution data of several morphological traits or molecular markers.

Understanding the genetic diversity and population structure of a core collection is also an important step since unaccounted population structure can lead to spurious associations (Pritchard *et al.*, 2000a,b). Logozzo *et al.* (2007) developed a core collection for European common bean germplasm with 544 accessions by using sampling methods based on the information available in the GenBank database and phaseolin pattern.

Accessions with similar phenotypes may not necessarily have close genetic relationships (Marita *et al.*, 2000) because of the polygenic properties of most traits and the effect of the environment on the expression of the analyzed

trait. Hence, applying molecular marker information reflecting the DNA polymorphism pattern is a powerful tool in core collection development.

Microsatellites (simple sequence repeats - SSRs, Tautz, 1989) have a high level of polymorphism, which allows the discrimination of cultivars and closely related common bean breeding lines, providing a reliable and efficient tool for germplasm characterization, conservation, and management (Blair *et al.*, 2006, 2007, 2009; Benchi-mol *et al.*, 2007; Perseguini *et al.*, 2011). Blair *et al.* (2009) and McClean *et al.* (2012) assessed the genetic diversity of common bean core collection by using SSRs and found a significant population structure that can be used for association studies.

The aim of the present study was to access the diversity level and genetic structure of 500 accessions from the IAC Common Bean Germplasm Bank and select 180 accessions that represent most of the variability in order to use this core collection in association mapping studies.

Materials and Methods

Plant material and DNA extraction

Five hundred genotypes from the IAC Common Bean Germplasm Bank (Campinas, S.P., Brazil) were used (Table S1). These 500 genotypes were selected from among more than 1800 accessions from the genebank accessions because they already had information of important agronomic traits for these accessions. Among the agronomical traits considered were resistance to anthracnose, angular leaf spot, rust, fusarium wilt, bacterial blight, a gold mosaic virus, tolerance to water deficit, grain size and tegument color. Total genomic DNA for all recombinant inbred lines was isolated from bulked young leaves of 10 plants per genotype using the CTAB extraction method as described in Hoisington *et al.* (1994).

SSR analysis

A total of 58 microsatellites (Table 1) were selected for their broad genomic distribution and high polymorphism information content. From these, 43 were EST-SSRs (Hanai *et al.*, 2007) and 15 were genomic-SSRs previously mapped (Campos *et al.*, 2011). The PCR amplifications were performed in a 25 μ L final volume containing 50 ng DNA, 1x buffer, 0.2 μ M of each forward and reverse primer, 100 μ M of each dNTP, 2.0 mM MgCl₂, 10 mM Tris-HCl (pH 8.0), 50 mM KCl, and 0.5 U of *Taq*-DNA polymerase. The following conditions were used for amplification: 1 min at 94 °C, followed by 30 cycles of 1 min at 94 °C, 1 min at annealing temperature specific for each SSR and 1 min at 72 °C, with a final extension of 5 min at 72 °C. The PCR products were viewed on a 3% agarose gel. Amplicons were separated by 6% denaturing polyacrylamide gel electrophoresis and silver stained (Creste *et al.*, 2001) (Figure S1). SSRs bands were manually scored.

Table 1 - Information for the 58 microsatellites that were used to assess the 500 common bean accessions, the core collection (180). The annealing temperatures (Ta), sizes fragments, numbers of alleles and polymorphism index values (PIC) and the discriminatory power (DP) are given for each marker. The first 15 are genomic-SSR *loci*, and the other 47 are EST-SSR *loci*.

SSRs used in genotyping analysis	Motif	Annealing temperature (Ta °C)	Fragment length/size (pb)	Number of alleles (500)	Number of alleles (180)	PIC (500)	DP (500)	PIC (180)	DP (180)
SSR-IAC20	(GA) ₇ AA (GA) ₂	56	190-194	2	2	0.37	0.49	0.35	0.81
SSR-IAC24	(AC) ₇ (AT) ₆	56	172-176	3	3	0.26	0.28	0.26	0.74
SSR-IAC52	(GA) ₁₁	56	160-216	7	6	0.64	0.69	0.61	0.88
SSR-IAC58	(TG) ₁₀	56	192-200	3	3	0.41	0.48	0.44	0.83
SSR-IAC62	(AG) ₁₄	45.3	202-216	8	8	0.81	0.84	0.82	0.95
SSR-IAC63	(AC) ₆	59.8	202-208	2	2	0.37	0.56	0.37	0.85
SSR-IAC66	(GA) ₁₀	56	260-300	10	10	0.86	0.87	0.87	0.96
SSR-IAC68	(CT) ₈	56	252-276	3	3	0.57	0.72	0.51	0.90
SSR-IAC127	(TA) ₃ T (TGA) ₃ G (TA) ₃	63.3	198-200	2	2	0.35	0.46	0.35	0.81
SSR-IAC136	(CA) ₇ (AT) ₅	56.7	232-270	6	6	0.72	0.64	0.73	0.88
SSR-IAC156	(TC) ₃ TG (GC) ₂	56.7	238-240	2	2	0.35	0.5	0.31	0.80
SSR-IAC160	(TG) ₂ (TA) ₂ (TG) ₅	56.7	180-184	2	2	0.36	0.49	0.37	0.83
SSR-IAC167	(TG) ₇ (CG) ₃	56.7	150-180	2	2	0.35	0.45	0.32	0.78
SSR-IAC179	(AC) ₆	50	100-104	2	2	0.36	0.50	0.34	0.81
SSR-IAC181	(AT) ₂ AC (AT) ₃ (AG) ₅	58.4	204-208	2	2	0.37	0.57	0.37	0.85
PvM01	(TTC) ₇	65	240-300	6	6	0.61	0.68	0.61	0.89
PvM02	(CTT) ₆	65	148-198	4	4	0.60	0.65	0.61	0.88
PvM03	(TTC) ₆	65	180-198	3	3	0.47	0.58	0.49	0.85
PvM04	(TTC) ₁₀	55	198-298	7	7	0.77	0.79	0.76	0.92
PvM07	(ATG) ₆	55	200-210	4	4	0.56	0.63	0.58	0.88
PvM11	(AGA) ₆	55	154-156	2	2	0.21	0.32	0.20	0.75
PvM13	(GAA) ₆	55	248-260	5	5	0.68	0.72	0.67	0.89
PvM14	(AATC) ₅	55	148-152	3	3	0.44	0.55	0.44	0.83
PvM15	(TGCA) ₅	55	182-310	3	3	0.29	0.34	0.26	0.75
PvM17	(ATGA) ₅	55	202-208	8	7	0.51	0.58	0.50	0.85
PvM21	(AT) ₁₄	55	226-320	12	12	0.86	0.90	0.86	0.97
PvM22	(TC) ₅	55	220-226	3	3	0.37	0.47	0.38	0.82
PvM28	(TCA) ₅	55	180-190	2	2	0.31	0.38	0.33	0.79
PvM36	(ATC) ₅	55	204-208	2	2	0.36	0.47	0.37	0.82
PvM40	(CTG) ₆	55	130-222	6	5	0.58	0.68	0.57	0.88
PvM45	(CT) ₂₄	55	174-200	5	5	0.56	0.62	0.55	0.86
PvM52	(CA) ₇	45	500-600	5	5	0.57	0.65	0.60	0.88
PvM53	(TC) ₆	55	200-204	2	2	0.36	0.50	0.37	0.83
PvM56	(TA) ₆	55	262-264	2	2	0.34	0.45	0.30	0.78

Table 1 - cont.

SSRs used in genotyping analysis	Motive	Annealing temperature (Ta °C)	Fragment length/size (pb)	Number of alleles (500)	Number of alleles (180)	PIC (500)	DP (500)	PIC (180)	DP (180)
PvM58	(TC) ₅	55	230-234	2	2	0.32	0.47	0.30	0.80
PvM61	(AC) ₆	55	196-200	2	2	0.37	0.5	0.37	0.82
PvM62	(TC) ₅	55	176-178	2	2	0.35	0.49	0.32	0.79
PvM66	(AT) ₅	65	182-200	5	5	0.46	0.57	0.48	0.86
PvM68	(AT) ₆	55	174-178	3	3	0.17	0.21	0.16	0.71
PvM73	(AAG) ₆	45	600-598	2	2	0.37	0.5	0.37	0.82
PvM75	(GAT) ₅	45	224-226	2	2	0.36	0.49	0.35	0.81
PvM79	(GAA) ₅	65	128-130	2	2	0.36	0.49	0.36	0.81
PvM93	(GA) ₅	65	206-216	2	2	0.18	0.24	0.17	0.72
PvM95	(AC) ₉	65	390-400	2	2	0.32	0.41	0.34	0.80
PvM97	(TA) ₅	65	168-170	2	2	0.33	0.46	0.33	0.81
PvM98	(TC) ₅	65	108-120	3	3	0.18	0.21	0.17	0.71
PvM100	(AT) ₇	65	198-304	4	3	0.25	0.33	0.22	0.75
PvM118	(TC) ₈	65	200-210	3	3	0.56	0.65	0.59	0.88
PvM120	(TC) ₆	65	200-204	3	2	0.37	0.49	0.36	0.81
PvM123	(CT) ₉	55	198-220	2	2	0.26	0.31	0.22	0.74
PvM124	(TA) ₅	55	500-598	2	2	0.37	0.5	0.37	0.82
PvM126	(TC) ₇	65	130-140	3	3	0.20	0.22	0.19	0.72
PvM127	(TC) ₅	55	260-270	3	3	0.43	0.50	0.43	0.82
PvM132	(AG) ₅	55	290-296	2	2	0.27	0.32	0.27	0.76
PvM148	(CAA) ₇	65	180-196	2	2	0.32	0.41	0.33	0.80
PvM150	(TCT) ₅	65	190-200	3	3	0.47	0.58	0.45	0.85
PvM151	(TTA) ₆	65	202-208	2	2	0.31	0.38	0.33	0.79
PvM153	(AGA) ₅	65	272-274	2	2	0.22	0.27	0.21	0.91
-	-	-	-	-	-	Mean 0.29	Mean 0.37	Mean 0.29	Mean 0.865

Data analysis

The size of alleles was scored in base pairs (bp) by visual comparison with a 100-bp DNA ladder and the value was converted to gene and genotypic frequencies. After the binary allele scoring (1 or 0, respectively), genotyping was performed using the allele number in decreasing order, that is to say, the alleles of largest size received the highest numbers, declining towards the lower size alleles. In the case of diploids, such as common beans, the scoring was considered twice when the band was homozygous and the genotype heterozygous, in which case both alleles were scored. The resulting matrix was used for obtaining genetic distances in Tools for Population Genetic Analyses (TFPGA) software, version 1.3 (Miller, 1997).

The percentage of polymorphisms obtained with each primer was calculated from this matrix. The genetic distances (GDs) were calculated from the SSR and EST-SSR data for all possible inbred pairs using modified Roger's genetic distance (MRD; Goodman and Stuber, 1983) implemented in the TFPGA program. Cluster analyses were performed using UPGMA with the incorporated NTSYS-pc computer package (Rohlf, 2000), version 2.1. Clustering stability was tested using a Bootstrap procedure based on 10,000 re-samplings with the BooD program (Coelho, 2002).

The polymorphism information content (PIC) values for SSRs were calculated using the following equation:

$$PIC = \sum_{i=1}^n f_i^2 - \sum_{j=i+1}^{n-1} 2f_i^2 f_j^2$$

where n is the number of alleles and f_i and f_j are the frequencies of the i^{th} and j^{th} allele, respectively (Lynch and Walsh, 1998).

The discrimination power (DP) values for the k^{th} primer were calculated using the formula:

$$DP_k = 1 - \sum_{j=1}^l p_j \frac{Np_{j-1}}{N-1}$$

where N is the number of individuals, and p_j is the frequency of the j^{th} pattern (Tessier *et al.*, 1999). The PIC was used to measure the information of a given marker locus for the pool of genotypes, while DP was used to measure the efficiency of the SSRs in identifying varieties by taking into account the probability that two randomly chosen individuals will have different patterns.

Wright's F statistics for SSRs were estimated using the GDA program (Lewis and Zaykin, 2000). This analysis was used to compare the structure of genetic diversity of the base collection with the core collection. Analysis of molecular variance (AMOVA) was used for estimating population differentiation directly from molecular data and testing hypotheses about such differentiation. The analyses were carried out using Arlequin 3.5 software (Excoffier and Lischer, 2010). The significance of the fixation indices was

tested by a permutation procedure with 10,000 iterations. The Arlequin 3.5 software was also used to estimate diversity fraction (F_{ST}) generated by SSRs analyses. AMOVA was performed with the base collection and core collection criteria. We used "among populations" to compare the base population and the core collection and "within population" to indicate the variability within each population.

Bootstrapping (Efron and Tibshirani, 1993) was used to determine whether the number of polymorphic SSRs used for genetic similarity estimation was adequate for a precise estimation of molecular markers among the 500 genotypes (Tivang *et al.*, 1994). The polymorphic markers were submitted to sampling with replacement to create new samples from the original data. The genetic similarities for each of these subsets were calculated from 1000 bootstrap estimates of the SSRs for each of these combinations. The coefficients of variation (CV) were used to construct box plots for each sample size. These analyses were carried out with R software (R Development Core Team, 2014). The exponential function was adjusted to estimate the number of loci needed to obtain a 10% CV. The median and maximum CV values were used to evaluate the accuracy of the genetic distance estimates. Although the mean CV is often used in the literature, caution is needed when dealing with molecular marker data for which there is no assurance that the CV values are distributed symmetrically.

The genetic structure of the sample was investigated using the Bayesian clustering algorithm implemented in STRUCTURE v.2.2 (Pritchard *et al.*, 2000a). The Admixture model was used for the base dataset with no previous population information and the "no-correlated allele frequencies between populations" option. Ten runs were applied using a burn-in period of 200,000 iterations, a run length of 500,000 Monte Carlo Markov Chain (MCMC) iterations, and a number of clusters varying from $K = 2$ to $K = 20$. The *ad hoc* statistic ΔK defined by Evanno *et al.* (2005) was used to determine the most probable number of clusters. The mean of the absolute values of $L' (K)$ was divided by the standard deviation, where $L' (K)$ stands for the mean likelihood plotted over 10 runs for each K . A hierarchical analysis of variance was carried out to test the significance of the differentiation among populations and clusters as defined by Structure software.

Construction of the core collection

In order to select the 180 accessions for a common bean core subset (Table S2), the following sampling criteria were applied: (i) the same percentage of each Structure group was selected to be integrated into the core collection; (ii) 105 accessions were selected equally from each structure group on the basis of the greatest genetic distance between accessions within each group and according to the genetic distance matrix and dendrogram (Figure S2); (iii) maintenance of 75 carioca tement cultivars, widely culti-

vated in the State of São Paulo (Brazil) under the leadership of the Agronomic Institute (IAC).

Results and Discussion

Molecular marker polymorphism of the base collection and genetic analyses

Genetic diversity among 500 common bean accessions was assessed from a total of 200 informative *loci*. The average number of alleles per locus of genomic-SSRs was 3.73, ranging from 2 to 10 alleles, and for EST-SSRs, it was 3.35. The highest numbers of observed alleles were found for SSR-IAC66 and PvM21 (Table 1). Our study showed an average of 2.8 alleles per locus, and found only three alleles for SSR-IAC66, corroborating the previous evaluation by Hanai *et al.* (2007) of 40 genomic-SSRs and 40 EST-SSRs in the Andean and Mesoamerican genotypes. Of the total number of markers in our study, 26 genomic-SSRs and 31 EST-SSRs exhibited a polymorphic pattern, with 2-7 alleles per locus and PvM21 showing 12 alleles. Hanai *et al.* (2010) evaluated the genetic diversity of an additional set of 100 EST-SSRs in 24 common bean genotypes, of which 54 were polymorphic, with an average of 2.7 alleles per locus.

The polymorphism information content (PIC) ranged from 0.26 to 0.86 for genomic markers and 0.17 to 0.86 for genic markers, and SSR-IAC66 and EST-SSR PvM21 were the *loci* with the highest PIC values (Table 1). The DP values ranged from 0.28 (SSR-IAC24) to 0.87 (SSR-IAC66) for genomic-SSRs and 0.21 (PvM68 and PvM98) to 0.90 (PvM21) for EST-SSRs (Table 1). The high PIC and DP values obtained for the SSR-IAC66 and PvM21 markers suggest their potential in accessing the genetic diversity in common beans. Benchimol *et al.* (2007) assessed the genetic diversity of 20 common bean genotypes belonging to the Andean and Mesoamerican gene pools with genomic-SSRs and found PIC values varying from 0.05 to 0.83. Perseguini *et al.* (2011) obtained lower PIC values (0.03 to 0.70) for a set of 60 carioca common beans, suggesting that this estimator is strongly influenced by the number and diversity of the genotypes under evaluation.

The boxplot chart (Figure 1) revealed that 10 CV% was obtained for approximately 33 markers, indicating that the number of microsatellites used in this study was sufficient to explain the genetic diversity content with good genome coverage. The number of markers is an important parameter to be considered in genetic diversity studies. Clustering analyses, which use a pairwise diversity matrix as input, require that the number of markers accurately estimates the diversity values. In the SSR diversity studies of cultivated genotypes, the number of markers varied considerably. In common beans, the number of SSRs that were used to evaluate the genetic diversity within core collections ranged from 36 (Blair *et al.*, 2009) to 58 markers (McClellan *et al.*, 2012).

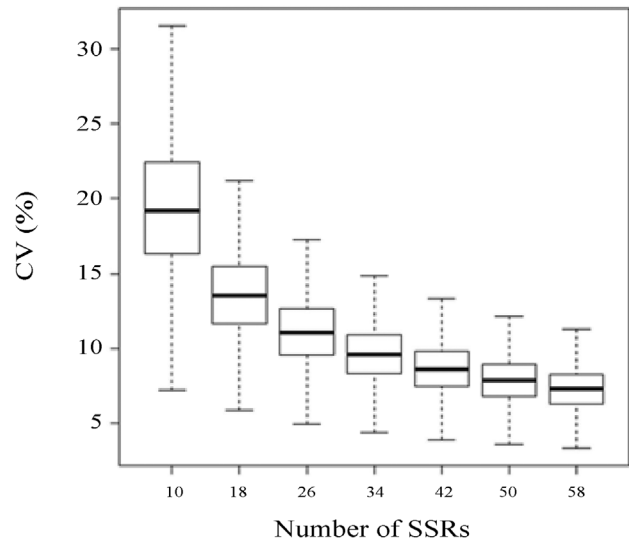


Figure 1 - Boxplot graph obtained by Bootstrap analysis of the data generated by genotyping 500 common bean accessions with 58 microsatellites.

The UPGMA dendrogram generated for the base collection revealed several groups, structured mostly in accordance to the grain morphology and genotype origin (Figure S2). To better understand the genetic organization of the 500 genotypes, Structure analyses were performed and found that the most appropriate number of groups (K) was 15 according to Evanno *et al.* (2005). Comparison of the clustering pattern determined by Structure with the UPGMA dendrogram indicated a strong correlation between the groups resolved in both analyses (Figure S3). The organization pattern of groups was inferred from the breeding institution (Groups 2, 3, 4, 6, 7, 9, 11, 12, 14, and 15). In fact, there are examples of crop species where breeding selection had resulted in domesticated populations displaying higher interpopulation differentiation than that by the wild populations (Doebley, 1989). This phenomenon and subsequent admixture (including crossing between cultivars) may maintain a high level of genetic diversity in breeding populations of domesticated species (Hernandez-Verdugo *et al.*, 2001). Perseguini *et al.* (2011) reported that carioca tegument genotypes clustered according to their breeding program. Such tendency may be attributed to a different artificial selection pressure in each breeding program that may render genetic differentiation. There is evidence that selection can be detected from patterns of polymorphism, and these signatures of artificial selection acting on alleles may be captured starting with $p < 0.2$ with reasonably high probability (Innan and Kim, 2004).

Analysis of genetic diversity of core collection

After evaluating the genetic structure of the base collection, we reduced the number of genotypes to form a core collection suitable for associative mapping purposes. The reduction was performed to remove possible redundant genotypes. Therefore, 36% reduction in the number of indi-

viduals in each group was performed in the base collection to obtain a representative core subset.

The choice of the most appropriate method for determining the core collections for association studies is an open issue requiring further investigation. To compare the performance of current state-of-the-art methods used to construct core subsets suitable for associative mapping of cultivated olive (*Olea europaea* L.), El Bakkali *et al.* (2013) found that a sample size of 94 entries captures the total diversity and is suitable for field assessments with many replicates for association mapping. Linkage disequilibrium observed in this study was mainly explained by a genetic structure effect estimated by Structure analyses.

In our study, the Bayesian method performed by Structure proved especially efficient for developing a core collection that can capture the allele diversity from a broad, diverse Brazilian germplasm collection, which comprises accessions with different agronomic features, such as disease resistance (anthracnose, angular leaf spot, and Fusarium wilt) and drought tolerance. Study of the genetic structure of 279 common bean genotypes, by using 67 microsatellite markers and four sequence characterized amplified regions (SCARs) by Burle *et al.* (2010), supported the efficiency of the Bayesian approach for germplasm analysis of genetic diversity and population structure.

The strategy used to establish the core collection (Table S2) in this study resembles the approach by Blair *et al.* (2009). Similar to a core collection formation that is generated by selecting a small percentage of the base collection to represent most of the total genetic variation with a minimum of redundancy (Oliveira *et al.*, 2010), the accessions chosen to integrate the diversity panel should also preserve as much of genetic variability as possible. Therefore, to ensure the adequacy and usefulness of the chosen accessions for associative mapping, it is necessary to assess whether the characteristics and variability of the base collection have been maintained.

Similarly to the base collection, the number of alleles present in the core collection varied between 2 and 10 alleles for the genomic-SSRs and from 2 to 12 alleles for the EST-SSRs. The average number of alleles per locus was slightly reduced (from 3.73 to 3.66 and from 3.35 to 3.26 for genomic-SSRs and EST-SSRs, respectively) suggesting that the allele richness was preserved in the reduced sample. The highest PIC and DP values were 0.87 and 0.96 for SSR-IAC66 and 0.86 and 0.97 for EST-SSR PvM21, respectively, indicating a high discriminatory power of these markers (Table 1). McClean *et al.* (2012) evaluated a common bean core collection using 58 SSRs, and showed that the number of alleles varied between 2 and 8 alleles per locus. Blair *et al.* (2009) evaluated 604 genotypes from the CIAT germplasm collection and reported PIC values ranging from 0.007 to 0.97. The number of alleles per locus and PIC in our core collection were in agreement with those in previous studies.

The core collection dendrogram divided the accessions into clusters similar to those observed in the base dendrogram. The genetic distances varied at a similar magnitude from 0.13 to 0.88 (Figure S4), suggesting that the genetic variability was maintained and was still quite extensive within the core subset.

The best K value obtained by the Bayesian analysis (Figure 2) divided the core accessions into four different groups (Figure 3), congruent with the Andean and Mesoamerican gene pools and the breeding program institution from which they were derived. Some accessions were grouped by grain size.

Group 1 of the Structure analysis (Table 2, Figure 3) was composed predominantly of Andean large-seeded genotypes directed for export driven by market demand, such as Feijão Suíço, Chileno/Branco, Branco Argentino, Amendoim, Bagajo, Jalo, and Jalo-110. Another feature observed in this group was the reddish color of the tegument that characterizes the Red Kidney and Vermelhinho cultivars and most of the lines derived from the CAL-143 x IAC-UNA (C x U) and IAC-UNA x CAL-143 (U x C) crosses used for the UC map (Campos *et al.*, 2011; Oblessuc *et al.*, 2012, 2014).

The accessions clustered in the remaining three groups (2, 3, and 4; Table 2, Figure 3) had smaller seeds, they were of the Mesoamerican type, and were distributed according to the breeding institution. The genotypes allocated to group 2 showed carioca grain tegument with economic importance in the Brazilian market and had been extensively exploited by the IAC and the IAPAR (Agronomic Institute of Paraná, Brazil) breeding programs until the late 1990s, when common bean improvement in Brazil moved toward the development of cultivars that were more resistant to biotic and abiotic stress.

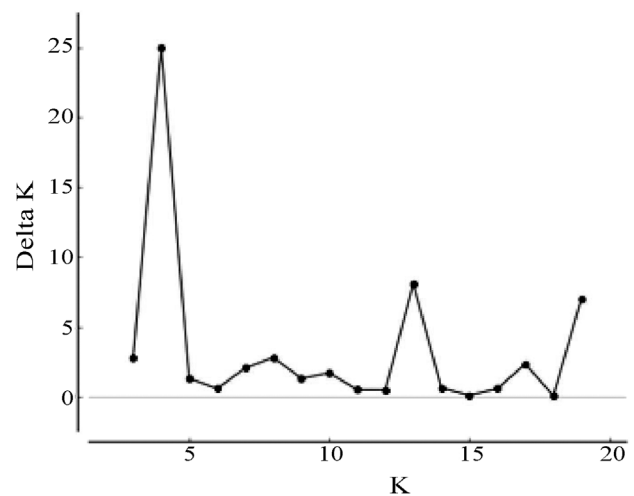


Figure 2 - Graphical representation of the optimal number of groups in the program Structure inferred using the criterion of Evanno *et al.* (2005). The analysis was based on data obtained from 58 microsatellite loci in core collection evaluated for genetic diversity.

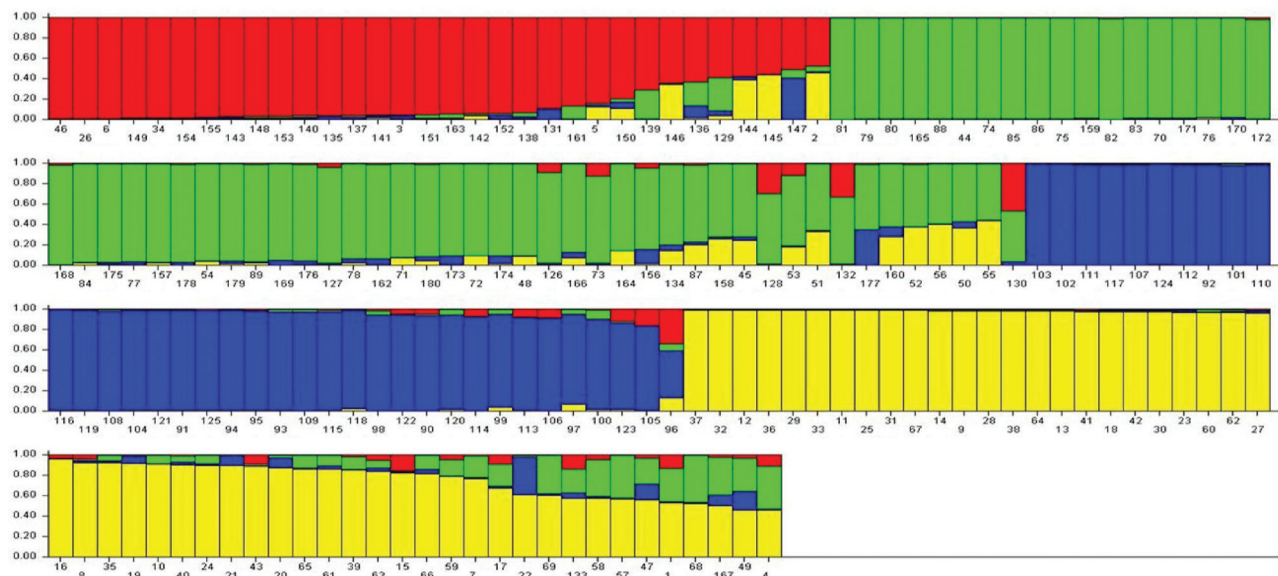


Figure 3 - Representation of the core collection according to the Bayesian analysis of the program Structure. The accessions evaluated were divided into four groups ($K = 4$). The names of the genotypes are given in Table S2 (The numbers correspond to the names of the genotypes). The red color corresponds to Groups 1, color Green corresponds to Group 2, color Blue corresponds to Group 3 and color Yellow corresponds to Group 4.

Table 2 - The 180 accessions clustered into the four groups generated by the Structure analysis and their respective traits.

Structure colors	Group	Accessions	Principals characteristics
Red	1	FeijãoSuiço, Chileno/Branco, Vermelhinho, Bagajo, Jalo-110, Jalo, Amendoim, Gen05C6-4-5-1-2, BrancoArgentino, UxC-1.1, UxC-2.20, UxC-1.2, UxC-1.19, UxC-1.5, UxC-3.9, UxC-4.17, UxC-9.2, UxC-9.16, CxU-1.3, CxU-1.5, CxU-1.7, CxU-1.19, CxU-2.11, CxU-2.16, CxU-7.8, UxC-1.8, UxC-1.10, UxC-6.13, UxC-2.18, UxC-3.3, CAL-143, Red Kidney	Grain size (typically Andean)
Green	2	Flor de Mayo, 2-Mar, Michelite, DOR-390, DOR-391, DOR-476, Turrialba-1, AND-279, RAZ-56, RAZ-49, Carioca Comum, Carioca Lustroso, Carioca MG, Carioca Precoce, H96A28-P4-1-1-1-1, H96A102-1-1-152, H96A31-P2-1-1-1-1, IAC-Alvorada, IAC-Apuã, IAC-AYSÓ, IAC-Carioca, IAC-Carioca Akytã, IAC-Carioca Aruã, IAC-Carioca Pyatã, IAC-Carioca Tybatã, IAC-Votuporanga, IAC-Ybaté, IAPAR-81, IAPAR-31, Pérola, Gen05C5-2-5-1-2, Gen05C5-2-10-1-1, Gen05C6-3-5-2-1, Gen05C6-5-2-2-1, Gen05C6-5-7-1-2, Gen05C7-4-1-1-1, VAX1, A0774, BAT447, SEA-5, IAC-UNA, Sanilac, FEB-176, FEB-177, J/39-2-3-1, J/61-5-3-1, J/43-5-1, J/43-1-1-1, J/39-1-3-2, M/100-4-3-1, F/19-6, F/19-3-1, E/20-2-1, D/15-3-1, C/11-2-2, (1108xBoreal/Brese), 29/24-6-1-1, 22/16-1-3-2	Most have 'carioca grain type' with great importance to IAC and IAPAR breeding programs
Blue	3	TO, Gen05P3-1-6-1, Gen05P4-2-6-2, Gen05P5-3-8-1, Gen05P5-3-8-2, Gen05P5-4-8-2, Gen05Pr11-1-2-2, Gen05Pr11-1-7-1, Gen05Pr11-2-3-1, Gen05Pr11-2-13-1, Gen05Pr11-2-14-2, Gen05Pr11-3-5-1, Gen05Pr11-6-5-1, Gen05Pr11-6-12-2, Gen05PR12-2-5-1-2, Gen05PR12-2-2-1-1, Gen05PR12-2-4-1-2, Gen05PR13-1-8-1-2, Gen05PR13-1-8-1-1, Gen05PR13-1-6-1-2, Gen05PR13-2-2-1-2, Gen05PR13-2-1-1-2, Gen05C1-3-2-1-1, Gen05C1-3-3-1-1, Gen05C2-1-1-2-1, Gen05C2-1-6-1-1, Gen05C2-1-1-1-3, Gen05C2-1-1-1-1, Gen05C3-2-4-1-1, Gen05C3-2-4-1-7, Gen05C4-3-1-1-2, Gen05C4-3-1-1-1, Gen05C4-4-3-1-2, Gen05C4-6-2-1-2, Gen05C5-1-2-2-2, Gen05C5-1-2-1-1	Recent crossings performed at the IAC breeding program
Yellow	4	Frijol Negro, ECU-311, México-115, Baetão (30273), Preto-208, Preto-184, Honduras-32, Guatemala-479, Jamapa (CNF-1671), Mulatinho (VP-102), Tupi, Rosinha G2, Preto do Pocrone, Porrillo-1, México-498, Small White 59 Preto, Perry Marron, Mortiño, Rosado-13, Porrillo Sintético, Puebla-152 (CIAT), ARA-1, Caeté (preta), IAC-Maravilha, FEB179, Jamapa (CIAT), Puebla-152 (CNF-1807), EMP-81, ARC-3, ARC-4, LP-90-91R.Bac., EMP-407, FEB180, Oito e Nove, Alemão, Bat-93, Pinto-114, G2333, PI-165426, RAZ-55, Batista Brilhante (CB), 82 PVBZ-1783, A-449, Aporé, Branquinho, BRS-Cometa, BRS-Horizonte, BRS-Pontal, BRS-Requinte, BRSMG-Talismã, CampeãoII, Caneludo, Gen05C7-3-2-2-2, J/54-5-1	Most are from CIAT and EMBRAPA

Group 3 (Table 2, Figure 3) included genotypes obtained from recent crosses conducted by the Agronomic Institute between 2000 and 2007, which were designed for the introgression of resistance genes to major diseases for carioca and black tegument cultivars. It was possible to observe changes in the genetic basis of these accessions compared to those clustered in groups 2 and 4, as the IAC breeding program has begun to focus on the maintenance of tegument and grain features in these cultivars as its main goal, in addition to high grain yield and nutritional quality.

The uppermost hierarchical level of the population structure that was identified using the ΔK (Evanno *et al.*, 2005) suggested that the 180 genotypes were divided into four groups; however, when $K = 2$ was considered (Figure S5), the samples were divided into two main genetic groups. A shared profile of alleles between the Andean and Mesoamerican genotypes was observed, most likely because some of the genotypes present in both parental crosses have both Andean and Mesoamerican origin (Figure S5). This mix is a result of the breeding process of common bean adopted by the institutions in Brazil. The two main clusters observed with the Structure analysis reflect our previous knowledge of the occurrence of two major wild gene pools of *P. vulgaris* (Blair *et al.*, 2009; Rossi *et al.*, 2009). Morphological and molecular markers showed that derived landraces are also generally organized into two gene pools and contain a subset of the wild-type genetic diversity (Gepts and Bliss 1986; Gepts *et al.*, 1986a,b; Beebe *et al.*, 2001; Debouck *et al.*, 1993; McClean *et al.*, 2004; McClean and Lee 2007).

AMOVA between the base and the core collection found only 2.75% change from the base collection to the core collection, but 97.2% of variation within each collection; in other words, most of the genetic variability of the base collection was retained in the core collection (Table 3).

According to the GDA analyses, the average expected heterozygosity (H_e) and observed heterozygosity (H_o) in the base collection were both 0.031, and in the core collection, they were equal to 0.034. The frequency of private alleles in the two collections indicated that there was no loss of genetic variability with the reduction of the base collection (500 genotypes) to the core collection (180 genotypes); however, it is worthy of note that in the base collec-

tion, three private alleles were found in loci PVM40, PVM73, and SSR-IAC181, whereas in the core collection, two private alleles were found in loci PVM04 and PVM40, and additionally, private allele PVM40 was preserved. Brown (1989a) proposed that a core collection should contain about 10% of the base collection. This sampling procedure should conserve about 0.80% of the alleles that occur in the base collection. Miklas *et al.* (1999) reported that a sample size of 10% is adequate to represent the genetic diversity of a base collection in common beans. The AMOVA and GDA results demonstrated that the methodology used to establish the core collection was appropriate because it maintained the genetic diversity present in the base collection.

The core collection for association mapping should include samples of mixed and/or admixed individuals from the most different genetic backgrounds. The presence of several genetic origins within the panels in different and unknown proportions induces linkage disequilibrium between unlinked loci and may increase the rate of false positives that are statistically associated with the analyzed trait without actually being causally involved in its phenotypic variation (Mezmouk *et al.*, 2011).

For proper use of genetic resources of a germplasm bank, it is essential to know the genetic diversity among the available accessions. The knowledge of genetic diversity also allows selection of the appropriate genotype and selection methods, depending on the available resources and genetic distance between recombinant genotypes and according to the objectives of the breeding program (Singh, 2001).

This study represents an efficient approach in developing a core collection suitable for association mapping studies by proper sampling of the core collection entries and assessment of the structure and relatedness within the samples. It is important to remark that the 180 selected genotypes are highly variable for important agronomic traits such as resistance to important common bean diseases (anthracnose, angular leaf spot, and bacterial blight) and drought tolerance. The proposed core collection should be periodically updated by including additional common bean germplasm in the base collection and adding novel molecular markers such as SNPs. At the current state, the developed core collection will be useful for conducting field assessments, and it is suitable for developing a long-term

Table 3 - Analysis of variance considering the base collection of 500 accessions and the core collection containing 180 accessions (Group 1 - Base collection and Group 2 -Core Collection).

Sources of variation	Sum of squares	Variance components	Percentage variation	p-value	Average F-Statistics over all <i>loci</i> Fixation Indices
					Fst
Among Groups	213.355	0.40356	2.75377	< 0.001	
Within Group 1 and Group 2	18416.941	14.25113	97.24623	< 0.001	
Total	18630.296	14.65469			0.02754

strategy for genome-wide association studies in common beans.

Acknowledgments

This research was supported by grants from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo - process numbers 2009/05284-1 and 2009/02502-8) and CNPq (process number 477239/2010-2).

References

- Abdurakhmonov IY and Abdulkarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int J Plant Genomics* 2008:e574927.
- Angioi SA, Rau D, Lanni L, Bellucci E, Papa R and Attene G (2010) The genetic make-up of the European landraces of the common bean. *Plant Genet Resour* 9:197-201.
- Asfaw A, Blair M and Almekinders C (2009) Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor Appl Genet* 120:1-12.
- Beebe SE, Rengifo J, Gaitan-Solis E, Duque MC and Tohme J (2001) Diversity and origin of Andean landraces of common bean. *Crop Sci* 41:854-862.
- Benchimol LL, Campos T, Carbonell SAM, Colombo CA, Chiorato AF, Formighieri EF and Souza AP (2007) Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. *Genet Resour Crop Evol* 54:1747-1762.
- Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, Logozzo G, Stougaard J, McClean P, Attene G, *et al.* (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc Natl Acad Sci USA* 109:E788-E796.
- Blair MW, Iriarte G and Beebe S (2006) QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean wild common bean (*Phaseolus vulgaris* L.) cross. *Theor Appl Genet* 112:1149-1163.
- Blair MW, Diaz JM, Hidalgo R, Diaz LM and Duque MC (2007) Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 116:29-43.
- Blair MW, Díaz LM, Buendía HF and Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theor Appl Genet* 119:955-972.
- Brown AHD (1989a) Core collection: a practical approach to genetic resources management. *Genome* 31:818-824.
- Brown AHD (1989b) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR and Williams JT (eds), *The Use of Plant Genetic Resources*. University Press Cambridge, Cambridge, pp 136-156.
- Brown AHD (1995) The core collection at the crossroads. In: Hodgkin T, Brown AHD, van Hintum TJL and Morales EAV (eds) *Core Collections of Plant Genetic Resources*. John Wiley & Sons, New York, pp 3-20.
- Brown AHD and Spillane C (1999) Implementing core collections: principles, procedures, progress, problems and promise. In: Johnson RC and Hodgkin T (eds) *Core Collections for Today and Tomorrow*. International Plant Genetic Resources Institute, Rome, pp 1-9.
- 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. *Theor Appl Genet* 121:801-813.
- Campos T, Oblessuc PR, Sforça DA, Cardoso JMK, Baroni RM, Sousa ACB, Carbonell SAM, Chioratto AF, Rubiano, LLB and Souza AP (2011) Inheritance of growth habit detected by genetic linkage analysis using microsatellites in the common bean (*Phaseolus vulgaris* L.). *Mol Breed* 27:549-560.
- Chioratto AF, Carbonell SAM, Dias LAS, Moura RR, Chiavegato MB and Colombo CA (2006) Identification of common bean (*Phaseolus vulgaris*) duplicates using agromorphological and molecular data. *Genet Mol Biol* 29:105-111.
- Coelho ASG (2002) Programa BootD: Avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de *bootstrap* com número variado de marcadores. Software. Laboratório de Genética Vegetal, Universidade Federal de Goiânia.
- Creste S, Tulmann A and Figueira A (2001) Detection of single sequence repeat polymorphism in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Reporter* 19:299-306.
- Debouck DG, Toro O, Paredes OM, Johnson WC and Gepts P (1993) Genetic diversity and ecological distribution of *Phaseolus vulgaris* in northwestern South America. *Econ Bot* 47:408-423.
- Doebley J (1989) Isozymic evidence and evolution of crop plants. In: Soltis ED and Soltis PM (eds) *Isozymes in Plant Biology*. Oregon Dioscordes, Portland, pp 165-191.
- Efron B and Tibshirani RJ (1993) *An Introduction to the Bootstrap*. v. 57 of Monographs on Statistics and Applied Probability. Chapman and Hall, New York, 436 pp.
- El Bakkali A, Haouane H, Moukhli, A, Costes, E, van Damme P and Khadari B (2013) Construction of core collections suitable for association mapping to optimize use of mediterranean olive (*Olea europaea* L.) genetic resources. *PLoS One* 8:e61265.
- Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 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. *Mol Ecol Resour* 10:564-567.
- Frankel OH and Brown AHD (1984) Plant genetic resources today: a critical appraisal. In: Holden JHW and Williams JT (eds) *Crop Genetic Resources: Conservation and Evaluation*. G. Allen and Unwin, London, pp 249-257.
- Gepts P and Bliss FA (1986) Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. *Econ Bot* 40:469-478.
- Gepts P, Kmiecik K, Pereira P and Bliss FA (1986a) Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. I. The Americas. *Econ Bot* 42:73-85.
- Gepts P, Osborn TC, Rashka K and Bliss FA (1986b) Phaseolin protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. *Econ Bot* 40:451-468.

- Gepts P (1998) Origin and evolution of common bean: past events and recent trends. *Hort Sci* 33:1124-1130.
- Goodman MM and Stuber CW (1983) Races of maize. VI. Isozyme variation among races of maize in Bolivia. *Maydica* 28:169-187.
- Hanai LL, Campos T, Camargo LEA, Benchimol LL, Souza AP, Melotto M, Carbonell SAM, Chioratto AF, Consoli L, Formighieri EF, *et al.* (2007) Development, characterization and comparative analysis of polymorphism at common bean-SSR loci isolated from genic and genomic sources. *Genome* 50:266-277.
- Hanai LL, Santini L, Camargo LEA, Fungaro MHP, Gepts P, Tsai SM and Vieira MLC (2010) Extension of the core map of common bean with EST-SSR, RGA, AFLP, and putative functional markers. *Mol Breed* 25:25-45.
- Hernandez-Verdugo S, Luna-Reyes R and Oyama K (2001) Genetic structure and differentiation of wild and domesticated populations of *Capsicum annum* (Solanaceae) from Mexico. *Plant Syst Evol* 226:129-142.
- Hoisington D, Khairallah M and Gonzalez-De-Leon D (1994) Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. CIMMYT, Mexico DF, 102 pp.
- Innan H and Kim Y (2004) Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci USA* 101:10667-10672.
- Logozzo G, Donnoli R, Macaluso L, Papa R, Knüpfner H and Zeuli PS (2007) Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. *Genet Resour Crop Evol* 54:1763-1779.
- Lynch M and Walsh JB (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, 980 pp.
- Mackay I and Powell W (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci* 12:57-63.
- Marita JM, Rodriguez JM and Nienhuis J (2000) Development of an algorithm identifying maximally diverse core collections. *Genet Resour Crop Evol* 47:515-526.
- McClellan PE, Lee RK and Miklas PN (2004) Sequence diversity analysis of dihydroflavonol 4-reductase intron 1 in common bean. *Genome* 47:266-280.
- McClellan PE and Lee RK (2007) Genetic architecture of chalcone isomerase non-coding regions in common bean (*Phaseolus vulgaris* L.). *Genome* 50:203-214.
- McClellan PE, Terpstra J, McConnell M, White C, Lee R and Mamidi S (2012) Population structure and genetic differentiation among the USDA common bean (*Phaseolus vulgaris* L.) core collection. *Genet Resour Crop Evol* 59:499-515.
- Mezmouk S, Dubreuil P, Bosio M, Decousset L, Charcosset A, Praud S and Mangin B (2011) Effect of population structure corrections on the results of association mapping tests in complex maize diversity panels. *Theor Appl Genet* 122:1149-1160.
- Miklas PN, Belorme R, Hannan R and Dickson M (1999) Using a subsample of the core collection to identify new sources of resistance to white mold in common bean. *Crop Sci* 39:569-573.
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE and Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194-2202.
- Oblessuc PR, Baroni RM, Garcia AAF, Chioratto AF, Carbonell SAM, Camargo LEA and Benchimol LL (2012) Mapping of angular leaf spot resistance QTL in common bean (*Phaseolus vulgaris* L.) under different environments. *BMC Genet* 13:e50.
- Oblessuc PR, Baroni RM, Pereira SG, Chioratto AF, Carbonell SAM, Briñez B, Silva LCS, Garcia AAF, Camargo LEA, Kelly JD, *et al.* (2014) Quantitative analysis of race-specific resistance to *Colletotrichum lindemuthianum* in common bean. *Mol Breeding* 34:1313-1329.
- Oliveira MF, Nelson RL, Geraldi IO, Cruz CD and de Toledo JFF (2010) Establishing a soybean germplasm core collection. *Field Crops Res* 119:277-289.
- Persegui JM, Chioratto AF, Zucchi MI, Colombo CA, Carbonell SAM, Mondego JMC, Gazaffi R, Garcia AAF, Campos T, Souza AP, *et al.* (2011) Genetic diversity in cultivated carioca common bean based on molecular marker analysis. *Genet Mol Biol* 34:88-102.
- Pritchard JK, Stephens M and Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Pritchard JK, Stephens MN, Rosenberg N and Donnelly P (2000b) Association mapping in structured populations. *Am J Hum Genet* 67:170-181.
- R Development Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- Rohlf FJ (2000) NTSYS-pc: Numerical taxonomy and multivariate analysis system, Version 2.1. New York, USA.
- Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G and Papa R (2009) Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol Appl* 2:504-522.
- Singh SP (2001) Broadening the genetic base of common bean cultivars: A review. *Crop Sci* 41:1659-1675.
- Spagnoletti-Zeuli PL and Qualset CO (1993) Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theor Appl Genet* 87:295-304.
- Tautz D (1989) Hypervariability of simple sequences as a general source of polymorphic markers. *Nucleic Acids Res* 17:6463-6471.
- Tessier C, David J, This P, Boursiquot JM and Charrier A (1999) Optimizations of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theor Appl Genet* 98:171-177.
- Tivang JG, Nienhuis J and Smith OS (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. *Theor Appl Genet* 89:259-264.
- van Hintum TJJ, Brown AHD, Spillane C and Hodgkin T (2000) Core Collection of Plant Genetic Resources. International Plant Genetic Resources Institute, Rome, 51 pp.
- Zhu C, Gore M, Buckler ES and Yu J (2008) Status and prospects of association mapping in plants. *Plant Genome* 1:5-20.

Internet Resources

- Lewis PO and Zaykin D (2000) Genetic data analysis: Computer program for the analysis of allelic data. ver. 1.0 (d15). <http://alleyn.eeb.uconn.edu/gda/2000> (June 30, 2011).

Miller M (1997) TFGPA - Tools for population genetic analyses. version 1.3. Northern Arizona University, <http://herb.bio.nau.edu/~miller/tfpga.htm> (June 30, 2011).
R Development Core Team (2011) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, <http://www.R-project.org/> (June 30, 2011).

Supplementary Material

The following online material is available for this article:

Figure S1 - Profile of the microsatellite PvM98.

Figure S2 - UPGMA cluster analysis of the modified Roger's genetic distances.

Figure S3 -Representation of the base collection according to the Bayesian analysis of the Structure program.

Figure S4 - UPGMA cluster analysis of the modified Roger's genetic distances.

Figure S5 -Representation of the core collection.

Table S1 -Names of the base collection.

Table S2 - Common bean genotypes selected to compose the core collection.

This material is available as part of the online article from <http://www.scielo.br/gmb>.

Associate Editor: Everaldo Gonçalves de Barros

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.