Theor Appl Genet (2011) 123:827–845
DOI 10.1007/s00122-011-1630-8

ORIGINAL PAPER

SNP marker diversity in common bean (*Phaseolus vulgaris* L.)

Andrés J. Cortés · Martha C. Chavarro ·
Matthew W. Blair

Received: 9 December 2010 / Accepted: 31 May 2011 / Published online: 23 July 2011

Abstract Single nucleotide polymorphism (SNP) markers have become a genetic technology of choice because of their automation and high precision of allele calls. In this study, our goal was to develop 94 SNPs and test them across well-chosen common bean (*Phaseolus vulgaris* L.) germplasm. We validated and accessed SNP diversity at 84 gene-based and 10 non-genic loci using KASPar technology in a panel of 70 genotypes that have been used as parents of mapping populations and have been previously evaluated for SSRs. SNPs exhibited high levels of genetic diversity, an excess of middle frequency polymorphism, and a within-genepool mismatch distribution as expected for populations affected by sudden demographic expansions after domestication bottlenecks. This set of markers was useful for distinguishing Andean and Mesoamerican genotypes but less useful for distinguishing within each gene pool. In summary, slightly greater polymorphism and race structure was found within the Andean gene pool than within the Mesoamerican gene pool but polymorphism rate between genotypes was consistent with genepool and race identity. Our survey results represent a baseline for the choice of SNP markers for future applications because gene-associated SNPs could themselves be causative SNPs for traits. Finally, we discuss that the ideal genetic marker combination with which to carry out diversity, mapping and association studies in common bean should consider a mix of both SNP and SSR markers.

Introduction

Cultivated common beans (*Phaseolus vulgaris* L.) are a key source of nutrients and dietary protein for over 500 million people in Latin America and Africa (Broughton et al. 2003). Common bean originated as a crop in America where diversification occurred along a long arc from South to Central America from the original range in Ecuador and northern Peru. After that, domestication processes in each region gave rise to Andean and the Mesoamerican cultivars (Gepts 1998; Gepts and Debouck 1991; Gepts et al. 1986). The state of Jalisco was probably the Mesoamerican domestication center, although this does not preclude more than one domestication event in the same region from Guatemala to the central volcanic axis of Mexico (Chacón et al. 2005; Kwak et al. 2009). For the Andean genepool, southern Bolivia and possibly Northern Argentina and Southern Peru may have constituted the center of domestication (Chacón et al. 2005).

Additional structure within each of these genepools has been described. Races Nueva Granada, Peru and Chile are identifiable within the Andean genepool by microsatellite data (Blair et al. 2007), while races Mesoamerica, Durango-Jalisco and Guatemala are observable within the Mesoamerican genepool (Blair et al. 2009; Díaz and Blair

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1630-8) contains supplementary material, which is available to authorized users.

A. J. Cortés · M. C. Chavarro · M. W. Blair
Centro Internacional de Agricultura Tropical (CIAT),
Apartado Aéreo 6713, Cali, Colombia

A. J. Cortés
Universidad de los Andes, Carrera 1 N 18A-12,
J302, Bogotá, Colombia

M. W. Blair (✉)
International Center for Tropical Agriculture (CIAT),
1380 N.W. 78th Ave, Miami, FL 33126, USA
e-mail: mwblaircgiar@gmail.com

2006; Kwak and Gepts 2009). Both sets of races followed various pathways of dissemination through the world, generating new secondary centers of diversity in Africa and Asia (Blair et al. 2010a). Introgression between genepools and between cultivated and wild genotypes has occurred repeatedly (Blair et al. 2006a, 2007; Papa and Gepts 2003).

When planning to analyze common bean diversity, the ideal molecular marker should be highly polymorphic and evenly distributed across the genome, as well as provide co-dominant, accurate and reproducible data which can be generated in a high-throughput and cost-effective manner (Yan et al. 2010). Although SSR markers have most of these properties, they are not always low cost due to labor and time investment involved. Single nucleotide polymorphism (SNP) markers meet the above criteria, together with the potential for high throughput and low cost genotyping. SNPs can be used in the same manner as other genetic markers for linkage map construction, genetic diversity analysis, marker–trait association and marker-assisted selection. In common bean, SNP markers have been basically used to date to perform linkage map construction and synteny analysis (Galeano et al. 2009a; b; McConnell et al. 2010), but not diversity analysis.

A considerable resource of SNPs were identified in ESTs through data mining of the transcriptome of nitrogen-fixing root nodules, phosphorus-deficient roots, developing pods, and leaves of the Mesoamerican genotype Negro Jamapa 81, and leaves of the Andean genotype G19833 by Ramirez et al. (2005). Later, a limited set of 25 SNP were identified from sequence fragments obtained from the Mesoamerican genotype DOR364 and the Andean genotype G19833 (Gaitan et al. 2008). More recently, Galeano et al. (2009a) developed SSCP markers based on these ESTs and named these BSNP (bean SNP) markers, while McConnell et al. (2010) generated gene-based SNP markers from the Andean genotype JaloEEP558 and the Mesoamerican genotype BAT93 using CAPs and dCAPs approaches. Finally, Hyten et al. (2010b) used the same genotypes to design and validate non-genic SNPs using a reduced representation library from multiple rounds of nested digestions with sequencing carried out by 454 pyrosequencing and Solexa technologies.

In summary, five methods have been used to validate and exploit SNP resource in common bean. First, cleaved amplified PCR fragment techniques (CAPs and dCAPs) were used to convert EST-based polymorphisms into genetic markers (Hougaard et al. 2008; McConnell et al. 2010). Second, a medium-throughput system named Luminex-100 was used to confirm SNP calls in DNA from ten common bean genotypes (Gaitan et al. 2008). Third, CELI mismatch digestions were used to analyze and map SNP-based, EST-derived markers, finding that the method worked well with SNPs located in the middle of the amplified fragment (Galeano et al. 2009b). Fourth, 325 amplicons were tested for SSCP polymorphism based on the previous gene derived markers (Galeano et al. 2009a). Finally, 827 non-genic SNPs were validated using the GoldenGate technology of Illumina. The only mapped SNPs to date are those from the BSNP series developed first by Galeano et al. (2009b) for ecotilling and then by Galeano et al. (2009a) for testing of SSCP polymorphisms.

When comparing each one of these methodologies for SNP detection, there are advantages and disadvantages to consider. For example, the CELI technique was not useful for identifying polymorphism in contigs and amplicons with two or more SNPs (Galeano et al. 2009b). Meanwhile, SSCP markers depended on SNPs that changed DNA conformation, while CAPs and dCAPs were enzyme-specific and expensive to use (Galeano et al. 2009a; McConnell et al. 2010). Hence, a high throughput screening is not yet truly available or transferable for SNPs in common bean. This can be seen in the fact that Luminex-100 is an expensive and obsolete technology (Gaitan et al. 2008), and GoldenGate technology does not offer genotyping of flexible numbers of markers and therefore is expensive per sample (Hyten et al. 2010a, b). An alternative is therefore needed, and this must be based on allele-specific genotyping technique offering a wider spectrum of genotyping possibilities for medium to high scale projects without sacrificing economy, efficiency and quality.

KASPar technology for SNP detection offers many benefits in terms of the issues raised above, such as in cost efficiency and time to datapoint acquisition (Bauer et al. 2009; Borza et al. 2010; Cuppen 2007; Nijman et al. 2008). However, for KASPar to be implemented, it is mandatory to couple this technology with the establishment of a resource of well-validated SNPs from gene or non-gene sources. In this sense, our interest was to build on the validated EST-based SNPs from Galeano et al. (2009a, b) as well as candidate genes that are important in adaptation to drought by applying the KASPar technology to a diverse set of genotypes that had been previously studied with SSRs by Blair et al. (2006a) or that are important for drought tolerance breeding.

Therefore, the overall objective of this work was to consolidate a SNP marker resource that would be useful to access genetic diversity in wild and cultivated common bean and that might also serve a larger project dedicated to molecular breeding for drought tolerance. We addressed the following questions: (1) how useful are SNP markers for detecting genetic diversity within a diverse set of 70 common bean accessions spanning both the Andean and Mesoamerican genepools?; (2) did the genetic variation at the SNP markers capture the essentials of population structure in common bean?; and (3) were their patterns of nucleotide variation correlated with population structure and domestication events in the species as a whole?

Materials and methods

Plant material

A total of 70 common beans (*P. vulgaris*) were used in this study (Table 1). These included 39 genotypes from Blair et al. (2006a), 15 parental lines used by Makunde et al. (2007) for the study of drought tolerance and 16 other parental lines used in the breeding programs of Center for Tropical Agriculture (CIAT). The genotypes represented parents of genetic mapping populations being studied at the International CIAT for the inheritance of disease resistance (common bacterial blight caused by *Xanthomonas axonopodis* pv. *phaseoli*, angular leaf spot caused by *Phaeoisariopsis griseola*, anthracnose caused by *Colletotrichum lindemuthianum* and bean golden yellow mosaic virus), insect resistance (*Apion godmani* and *Thrips palmi*), abiotic stress tolerance (aluminum, heat, drought and low phosphorous), grain quality (micronutrient content), growth habit and yield components.

Among the 70 common bean genotypes were a total of 28 Andean (27 cultivated and 1 wild) and 42 Mesoamerican (40 cultivated and 2 wild) genotypes. The three wild accessions represented accessions from Argentina, Colombia and Mexico with the first and the last known to be closest to the domesticated accessions from the work of Blair et al. (2006a). All three have been used for advanced backcross population development (Blair et al. 2006b). Among the cultivated genotypes, 33 were advanced breeding lines from CIAT and the remainder were landraces or locally bred varieties. The advanced lines included three from the BAT series, one from the BRB series, five from the DOR series, two from the MAM series, one from the MAR series, three from the SEA series, one from the SEL series, three from the SEQ series and three from the VAX series. BAT, DOR, MAM, MAR, SEA, and VAX lines have Mesoamerican seed types while BRB and SEQ lines have Andean seed types. MAM lines have mixed Durango and Mesoamerica race pedigrees while the SEL and VAX lines have some tepary bean (*P. acutifolius*) ancestry. Germplasm accessions included representatives of the Nueva Granada and Peru races within the Andean genepool and representatives of the Guatemala, Durango-Jalisco, and Mesoamerica races within the Mesoamerican genepool according to previous classifications (Blair et al. 2006a, 2007; Díaz and Blair 2006). The growth habit of each genotype was classified from I (determinate bush) to IV (indeterminate climber) according to Singh (1982).

SNP markers analysis

Four seeds from single seed descent that were derived by self-pollination were used for the DNA preparation. Total DNA was extracted by the Germplasm Characterization Laboratory of CIAT by the protocol of Afanador and Hadley (1993). Quantification was made with a Hoefer DyNA Quant 2000 fluorometer for dilution to a standard concentration of 10 ng/µl. A total of 84 gene-based and 10 non-genic BSNP markers were carefully selected from previous studies (Galeano et al. 2009a, b; Hyten et al. 2010a; Ramirez et al. 2005) taking into account the polymorphism content, the quality of the flanking regions (at least 50 bp to each side) and the absence of introns in the case of gene-based BSNPs. A greater number of genic BSNPs were developed given our interest in identifying causative SNP polymorphisms that might be associated with drought tolerance or other traits. Presence of introns in the flanking regions was determined through a blast search against *Glycine max* genome (Schmutz et al. 2010) and BAC end sequences of common bean (Córdoba et al. 2010; David et al. 2008). Blast search was made with a gap open penalty of 5, a gap extension penalty of 2, a match score of 2, and a mismatch score of −3. Detailed information on newly designed BSNPk marker (Bean SNP detected by KASPar evaluation) can be found in Supplemental table 1.

Genotyping was done using KASPar technology, which uses a competitive allele-specific PCR combined with a FRET quenching reporter oligonucleotide probe (Cuppen 2007). Two allele-specific (AS) oligonucleotides of about 40 bp in length and one common oligonucleotide (CP) of about 20 bp in length were designed using Primer-Picker for each one of the 94 SNPs (KBioscience, UK). All of them are standard unmodified and unlabelled oligonucleotides. SNP genotyping was then carried out for the 70 cultivated and wild accessions of common bean with 94 designed primer triplets. The three oligonucleotides for each assay were dissolved in 10 mM Tris–HCl (pH 8) to a 100 µM concentration, mixed together as a SNP assay mix (12 µl AS1 + 12 µlAS2 + 30µlCP + 46*µ*; Tris–HCl pH8) and 2 µl aliquots were distributed into individual wells of 96 well plates by a Tecan Robot (Genesis RSP200 liquid handling workstation including an integrated 96-channel pipetting head TEMO96). Assay plates were frozen at −20°C until use. Each SNP was typed in a total volume of 4 µl in the following reaction mixture: 6 ng DNA, 22 mM $MgCl_2$, KTaq, 1 µl 4× reaction mix, and 2 µl pre-plated assay mix. Amplification was performed in Applied Biosystems GeneAmp 9700 thermocyclers running the following program: 94°C 15′ then 20 cycles of 94°C 10″, 57°C 5″ and 72°C 10″, followed by 18 cycles of 94°C 10″, 57°C 20″ and 72°C 40″. Finally, fluorescence scanning of the reactions was done with a BMG labtech Pherastar scanner (KBioscience).

Data analysis

Fluorescence signals were interpreted by the KlusterCaller 1.1 software (KBioscience) where a discriminant analysis

Table 1 Common bean genotypes used for assessment of SNP diversity and their accession number, phaseolin status, race and gene pool identity, origin and growth habit

Genotype	Ph	Genepool	Race	Status	Origin	GH
G4494	T	Andean	P	Cultiv	Colombia	I
G19833	H	Andean	P	Landrace	Peru	III
G19839	T	Andean	P	Landrace	Peru	III
G21078	T	Andean	P	Landrace	Argentina	IV
G21657	C	Andean	P	Cultiv	Bulgaria	III
Radical Cerinza	T	Andean	P	Cultiv	Colombia	I
BRB191	T	Andean*	NG	Line	CIAT	II
G5273	T	Andean	NG	Cultiv	Mexico	II
JaloEEP558	T	Andean	NG	Cultiv	Brazil	III
SEQ1027	T	Andean	NG	Line	CIAT	III
AFR298	na	Andean	na	Cultiv	CIAT	I
CAL143	na	Andean	na	Cultiv	CIAT	I
CAL96	na	Andean	na	Cultiv	CIAT	I
Canario70	na	Andean	na	Cultiv	Mexico	II–III
DOR303*	na	Andean	na	Line	CIAT	II
G122	na	Andean	na	Landrace	India	I
G19892	T	Andean	na	Wild	Argentina	IV
G21242	C	Andean*	na	Landrace	Colombia	IV
G24404	C	Andean*	na	Wild	Colombia	IV
G4523	na	Andean	na	Cultiv	Colombia	I
IJR	na	Andean	na	Cultiv	India	na
Montcalm	na	Andean	na	Cultiv	USA	I
Natal Sugar	na	Andean	na	Cultiv	Africa	II
PAN127	na	Andean	na	Line	Africa	II
R.C.WONDER	na	Andean	na	Cultiv	Africa	I
RAA21	na	Andean	na	Line	CIAT	I–II
SAB259*	na	Andean	na	Line	CIAT	I
SEQ1003	na	Andean	na	Line	CIAT	I
SUG131	na	Andean	na	Line	CIAT	I–II
A55	na	Mesoamerican	na	Line	CIAT	II
Arroyo Loro	na	Mesoamerican	na	Cultiv	Dom. Rep.	na
G24390	M	Mesoamerican	na	Wild	Mexico	IV
Morales	na	Mesoamerican	na	Cultiv	Pto Rico	na
Pinto Villa	na	Mesoamerican	na	Cultiv	Mexico	III
SEC16	na	Mesoamerican	na	Line	CIAT	II
SEQ11	na	Mesoamerican	na	Line	CIAT	II
SER16	na	Mesoamerican	na	Line	CIAT	II
SER22	na	Mesoamerican	na	Line	CIAT	II
SER8	na	Mesoamerican	na	Line	CIAT	II
VAX1	na	Mesoamerican	na	Line	CIAT	II
VAX3	na	Mesoamerican	na	Line	CIAT	II
MAR1	S	Mesoamerican	M	Line	CIAT	II
BAT477	S	Mesoamerican	M	Line	CIAT	II
BAT881	S	Mesoamerican	M	Line	CIAT	II
BAT93	S	Mesoamerican	M	Line	CIAT	II
DOR364	S	Mesoamerican	M	Cultiv	El Salvador	II
DOR390	S	Mesoamerican	M	Cultiv	Mexico	II
DOR476	S	Mesoamerican	M	Line	CIAT	II

Table 1 continued

Genotype	Ph	Genepool	Race	Status	Origin	GH
DOR714	S	Mesoamerican	M	Line	CIAT	II
G11350	S	Mesoamerican	M	Landrace	Mexico	III
G14519	S	Mesoamerican	M	Landrace	USA	IV
G21212	B	Mesoamerican	M	Landrace	Colombia	II
G3513	S	Mesoamerican	M	Landrace	Mexico	II
G4090	Sd	Mesoamerican	M	Landrace	El Salvador	II
G4825	B	Mesoamerican	M	Landrace	Brazil	III
ICA Pijao	B	Mesoamerican	M	Cultiv	Colombia	II
JAMAPA	S	Mesoamerican	M	Landrace	Mexico	II
MD23-24	S	Mesoamerican	M	Line	EAP	II
SEA21	S	Mesoamerican	M	Line	CIAT	II
SEL1309	S	Mesoamerican	M	Line	CIAT	II
Tio Canela	S	Mesoamerican	M	Cultiv	EAP	II
VAX6	S	Mesoamerican	M	Line	CIAT	II
G11360	S	Mesoamerican	J	Landrace	Mexico	IV
G855	Sb	Mesoamerican	J	Landrace	Mexico	IV
G2333	S	Mesoamerican	G	Landrace	Mexico	IV
G685	Sb	Mesoamerican	G	Landrace	Guatemala	IV
MAM38	S	Mesoamerican	D	Cultiv	CIAT	III
MAM49	S	Mesoamerican	D	Cultiv	CIAT	III
SEA15	S	Mesoamerican	D	Line	CIAT	II
SEA5	S	Mesoamerican	D	Line	CIAT	II

Ph Phaseolin type, Races: *D–J* Durango–Jalisco, *G* Guatemala, *NG* Nueva Granada, *P* Peru, *M* Mesoamerica, *GH* growth habitat as described in materials and methods, *na* not applicable, * an inter-genepool category would be more adequate

*An inter-genepool category would be more adequate

was performed using the transformed fluorescent signals to confirm allele calls. Allele assignments and frequencies for the 70 common bean accessions were then used to calculate the polymorphic information content (PIC) for each SNP marker according to Anderson et al. (1993) using the formula $\mathrm{PIC}_i = 1 - \Sigma p_{ij}^2$; were p_{ij} is the frequency of the allele j for each marker i. The minimum allele frequency (MAF) was also calculated and the full (global) data set was subjected to a principal component analysis (Hair et al. 1992). Values of total diversity (H_{t}), intra-population diversity (H_{s}) and population differentiation (F_{st}) were also calculated (Nei 1987; Wright 1969). Furthermore, population structure was evaluated without an *a priori* criterion of stratification using STRUCTURE 2.3.2 (Pritchard et al. 2000) as described in Blair et al. (2009). Analyses had a burn-in length of 50,000 iterations and a run length of 100,000 iterations after burning. Five replicates were carried out for each K value (sub-population numbers between 2 and 5). The consistency of results among replicated runs was evaluated. Neighbor-joining tree construction and nodal support evaluation using 1,000 bootstrap replicates were carried out with the program Mega4 (Tamura et al. 2007). In addition, the levels of genetic diversity were computed based on the average number of single nucleotide differences between any pair of accessions (π) (Nei 1987). These calculations were carried out with the program DnaSP 5.10 (Rozas et al. 2003) both for Andean and Mesoamerican genepools together (at the global level) and separately (at the intra-genepool level). Statistical evidence for the uniqueness of each genotype in the two genepools was calculated as the total number of SNP pairwise differences against the rest of the accessions using Shapiro–Wilk, Kruskal–Wallis and randomization tests. Moreover, the nucleotide diversity of Nei was contrasted against an estimated distribution (evolutionary background) based on SSR data from Blair et al. (2006a, 2009). Such comparison took into account the differences between SNP and SSR markers in terms of number of alleles and information content (Yan et al. 2010). Finally, mismatch distributions and folded site frequency spectra were made for all the accessions and for each genepool using also DnaSP 5.10 (Rozas et al. 2003). These observed distributions were compared with the expectations of the Wright–Fisher neutral model using coalescent simulations with 5,000 repetitions (Wakeley 2008). Hence, we could compare how much our data diverged from the neutral model of molecular evolution and from the evolutionary background in common bean.

Results

Characterization of SNP markers

KASPar genotyping was compared with Sanger sequencing previously carried out for BSNPk92, BSNPk93 and BSNPk94 (*Dreb2* genes) as an initial test of SNP calling quality. All the assignments were correct so we continued genotyping the remaining BSNPk markers using KASPar evaluations (Fig. 1). Across all the tested SNPs the average PIC content was of 0.437. However, non-genic SNPs had a higher average PIC (0.440) than gene-based SNPs (0.436) (Table 2). Only six SNPs presented a PIC <0.2 (BSNPk94, BSNPk77, BSNPk16, BSNPk17, BSNPk81, BSNPk57) all of which were gene-based (Supplemental figure 1). Eighty-one BSNPk markers had a PIC value higher than 0.400. The maximum PIC was up to 0.500 for BSNPk2, BSNPk18, BSNPk24, BSNPk25, BSNPk28, BSNPk35, BSNPk47, BSNPk48, BSNPk71, BSNPk78, BSNPk79, and BSNPk86, all of which except for the last were gene-based SNPs. A value of 0.5 corresponded to the theoretical maximum PIC for bi-allelic markers according to the definition of Anderson et al. (1993). Nucleotide diversity as measured by pairwise differences varied according to the BSNP marker used in a comparison of the gene-based and non-genic SNPs (Supplementary figure 2). MAF ranked between 0.043 (BNSPk94) and 0.5 (BSNPk71), which is the maximum possible theoretical value. Average MAF

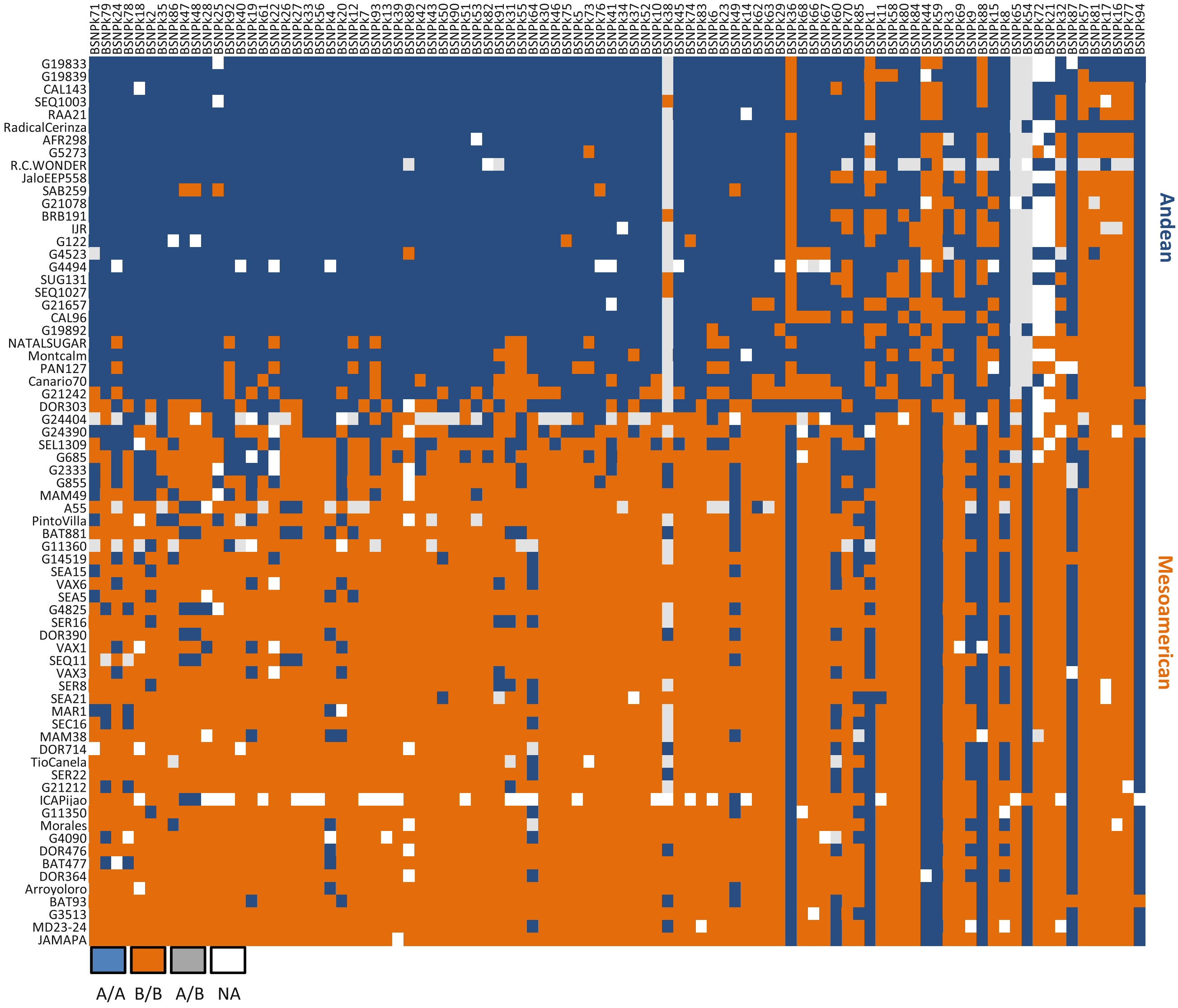

Fig. 1 Bitmap of polymorphism for 70 cultivated common bean accessions (*rows*) and 94 SNPs (*columns*). SNPs are organized according to their PIC values and accessions are presented according to the first principal component

Table 2 Frequency, polymorphism information content (PIC), minimum allele frequency (MAF), and nucleotide diversity (π) of 94 common bean SNP (84 gene-based and 10 non-genic) markers

BSNPk	Number				Frequency					PIC	π	BSNPk	Number				Frequency					PIC	π
	AA	BB	AB	NA	A	B	AA	BB	AB				AA	BB	AB	NA	A	B	AA	BB	AB		
BSNPk1	45	21	3	1	0.67	0.33	0.65	0.30	0.04	0.44	0.47	BSNPk48	31	37	0	2	0.46	0.54	0.46	0.54	0.00	0.50	0.49
BSNPk2	37	32	1	0	0.54	0.46	0.53	0.46	0.01	0.50	0.46	BSNPk49	43	26	0	1	0.62	0.38	0.62	0.38	0.00	0.47	0.48
BSNPk3	19	48	3	0	0.29	0.71	0.27	0.69	0.04	0.41	0.46	BSNPk50	29	40	1	0	0.42	0.58	0.41	0.57	0.01	0.49	0.49
BSNPk4	38	29	1	2	0.57	0.43	0.56	0.43	0.01	0.49	0.49	BSNPk51	29	40	0	1	0.42	0.58	0.42	0.58	0.00	0.49	0.48
BSNPk5	28	41	0	1	0.41	0.59	0.41	0.59	0.00	0.48	0.48	BSNPk52	27	42	1	0	0.39	0.61	0.39	0.60	0.01	0.48	0.49
BSNPk6	26	42	1	1	0.38	0.62	0.38	0.61	0.01	0.47	0.48	BSNPk53	28	39	2	1	0.42	0.58	0.41	0.57	0.03	0.49	0.39
BSNPk7	29	39	1	1	0.43	0.57	0.42	0.57	0.01	0.49	0.44	BSNPk54	46	1	22	1	0.83	0.17	0.67	0.01	0.32	0.29	0.39
BSNPk8	50	17	1	2	0.74	0.26	0.74	0.25	0.01	0.38	0.40	BSNPk55	28	40	1	1	0.41	0.59	0.41	0.58	0.01	0.48	0.49
BSNPk9	49	19	1	1	0.72	0.28	0.71	0.28	0.01	0.41	0.44	BSNPk56	30	39	0	1	0.43	0.57	0.43	0.57	0.00	0.49	0.34
BSNPk10	27	42	0	1	0.39	0.61	0.39	0.61	0.00	0.48	0.46	BSNPk57	6	62	2	0	0.10	0.90	0.09	0.89	0.03	0.18	0.31
BSNPk11	22	47	0	1	0.32	0.68	0.32	0.68	0.00	0.43	0.46	BSNPk58	22	48	0	0	0.31	0.69	0.31	0.69	0.00	0.43	0.43
BSNPk12	29	39	2	0	0.43	0.57	0.41	0.56	0.03	0.49	0.49	BSNPk59	49	21	0	0	0.70	0.30	0.70	0.30	0.00	0.42	0.44
BSNPk13	29	39	0	2	0.43	0.57	0.43	0.57	0.00	0.49	0.48	BSNPk60	45	23	2	0	0.66	0.34	0.64	0.33	0.03	0.45	0.47
BSNPk14	25	42	0	3	0.37	0.63	0.37	0.63	0.00	0.47	0.45	BSNPk61	30	38	1	1	0.44	0.56	0.43	0.55	0.01	0.49	0.48
BSNPk15	18	49	1	2	0.27	0.73	0.26	0.72	0.01	0.40	0.26	BSNPk62	26	44	0	0	0.37	0.63	0.37	0.63	0.00	0.47	0.47
BSNPk16	3	63	2	2	0.06	0.94	0.04	0.93	0.03	0.11	0.11	BSNPk63	25	44	1	0	0.36	0.64	0.36	0.63	0.01	0.46	0.48
BSNPk17	4	62	1	3	0.07	0.93	0.06	0.93	0.01	0.13	0.31	BSNPk64	39	27	3	1	0.59	0.41	0.57	0.39	0.04	0.48	0.41
BSNPk18	33	29	1	7	0.53	0.47	0.52	0.46	0.02	0.50	0.50	BSNPk65	0	40	28	2	0.21	0.79	0.00	0.59	0.41	0.33	0.39
BSNPk19	37	30	0	3	0.55	0.45	0.55	0.45	0.00	0.49	0.50	BSNPk66	24	44	1	1	0.36	0.64	0.35	0.64	0.01	0.46	0.45
BSNPk20	38	29	0	3	0.57	0.43	0.57	0.43	0.00	0.49	0.47	BSNPk67	23	44	0	3	0.34	0.66	0.34	0.66	0.00	0.45	0.45
BSNPk21	8	44	0	18	0.15	0.85	0.15	0.85	0.00	0.26	0.47	BSNPk68	23	42	1	4	0.36	0.64	0.35	0.64	0.02	0.46	0.43
BSNPk22	27	35	1	7	0.44	0.56	0.43	0.56	0.02	0.49	0.48	BSNPk69	19	48	2	1	0.29	0.71	0.28	0.70	0.03	0.41	0.43
BSNPk23	26	43	1	0	0.38	0.62	0.37	0.61	0.01	0.47	0.49	BSNPk70	23	45	2	0	0.34	0.66	0.33	0.64	0.03	0.45	0.48
BSNPk24	33	32	3	2	0.51	0.49	0.49	0.47	0.04	0.50	0.50	BSNPk71	33	33	3	1	0.50	0.50	0.48	0.48	0.04	0.50	0.48
BSNPk25	29	35	0	6	0.45	0.55	0.45	0.55	0.00	0.50	0.50	BSNPk72	8	41	1	20	0.17	0.83	0.16	0.82	0.02	0.28	0.47
BSNPk26	30	39	1	0	0.44	0.56	0.43	0.56	0.01	0.49	0.49	BSNPk73	28	41	0	1	0.41	0.59	0.41	0.59	0.00	0.48	0.48
BSNPk27	30	39	0	1	0.43	0.57	0.43	0.57	0.00	0.49	0.49	BSNPk74	27	42	0	1	0.39	0.61	0.39	0.61	0.00	0.48	0.48
BSNPk28	30	36	0	4	0.45	0.55	0.45	0.55	0.00	0.50	0.48	BSNPk75	28	41	1	0	0.41	0.59	0.40	0.59	0.01	0.48	0.49
BSNPk29	25	44	0	1	0.36	0.64	0.36	0.64	0.00	0.46	0.48	BSNPk76	28	41	0	1	0.41	0.59	0.41	0.59	0.00	0.48	0.29
BSNPk30	28	41	1	0	0.41	0.59	0.40	0.59	0.01	0.48	0.49	BSNPk77	3	65	1	1	0.05	0.95	0.04	0.94	0.01	0.10	0.30
BSNPk31	28	40	1	1	0.41	0.59	0.41	0.58	0.01	0.48	0.35	BSNPk78	35	33	1	1	0.51	0.49	0.51	0.48	0.01	0.50	0.50
BSNPk32	7	57	2	4	0.12	0.88	0.11	0.86	0.03	0.21	0.35	BSNPk79	35	34	1	0	0.51	0.49	0.50	0.49	0.01	0.50	0.47
BSNPk33	30	39	0	1	0.43	0.57	0.43	0.57	0.00	0.49	0.49	BSNPk80	21	47	1	1	0.31	0.69	0.30	0.68	0.01	0.43	0.29
BSNPk34	27	41	1	1	0.40	0.60	0.39	0.59	0.01	0.48	0.49	BSNPk81	5	63	2	0	0.09	0.91	0.07	0.90	0.03	0.16	0.33

Table 2 continued

BSNPk	Number				Frequency					PIC	π	BSNPk	Number				Frequency					PIC	π
	AA	BB	AB	NA	A	B	AA	BB	AB				AA	BB	AB	NA	A	B	AA	BB	AB		
BSNPk35	32	37	1	0	0.46	0.54	0.46	0.53	0.01	0.50	0.48	BSNPk82	29	40	0	1	0.42	0.58	0.42	0.58	0.00	0.49	0.49
BSNPk36	45	25	0	0	0.64	0.36	0.64	0.36	0.00	0.46	0.47	BSNPk83	27	42	0	1	0.39	0.61	0.39	0.61	0.00	0.48	0.45
BSNPk37	27	41	1	1	0.40	0.60	0.39	0.59	0.01	0.48	0.48	BSNPk84	21	48	1	0	0.31	0.69	0.30	0.69	0.01	0.43	0.44
BSNPk38	9	24	36	1	0.39	0.61	0.13	0.35	0.52	0.48	0.48	BSNPk85	46	23	1	0	0.66	0.34	0.66	0.33	0.01	0.45	0.48
BSNPk39	29	39	0	2	0.43	0.57	0.43	0.57	0.00	0.49	0.49	BSNPk86	31	36	2	1	0.46	0.54	0.45	0.52	0.03	0.50	0.36
BSNPk40	29	36	3	2	0.45	0.55	0.43	0.53	0.04	0.49	0.49	BSNPk87	58	7	2	3	0.88	0.12	0.87	0.10	0.03	0.21	0.31
BSNPk41	27	40	1	2	0.40	0.60	0.40	0.59	0.01	0.48	0.49	BSNPk88	48	18	1	3	0.72	0.28	0.72	0.27	0.01	0.40	0.43
BSNPk42	29	40	1	0	0.42	0.58	0.41	0.57	0.01	0.49	0.49	BSNPk89	25	34	1	10	0.43	0.58	0.42	0.57	0.02	0.49	0.48
BSNPk43	28	39	3	0	0.42	0.58	0.40	0.56	0.04	0.49	0.47	BSNPk90	29	40	1	0	0.42	0.58	0.41	0.57	0.01	0.49	0.49
BSNPk44	46	20	0	4	0.70	0.30	0.70	0.30	0.00	0.42	0.46	BSNPk91	28	40	2	0	0.41	0.59	0.40	0.57	0.03	0.49	0.49
BSNPk45	27	42	0	1	0.39	0.61	0.39	0.61	0.00	0.48	0.48	BSNPk92	31	38	0	1	0.45	0.55	0.45	0.55	0.00	0.49	0.49
BSNPk46	28	41	1	0	0.41	0.59	0.40	0.59	0.01	0.48	0.49	BSNPk93	29	39	1	1	0.43	0.57	0.42	0.57	0.01	0.49	0.29
BSNPk47	32	38	0	0	0.46	0.54	0.46	0.54	0.00	0.50	0.50	BSNPk94	66	3	0	1	0.96	0.04	0.96	0.04	0.00	0.08	0.49

was of 0.36 for all the SNPs, of 0.357 for non-genic SNPs and of 0.361 for gene-based SNPs. Eleven SNPs presented MAF values <0.2, all of which were gene-based. In total, 76 SNPs had values higher than 0.3. For example, the high polymorphic BSNPk markers included BSNPk18, BSNPk78, BSNPk24, BSNPk79, BSNPk71 all with MAF values higher than 0.46. All of them were gene-based SNPs.

Correspondence between PIC and MAF values was straightforward. Correlation coefficient between both measures was high with an r value of 0.964. Correlation coefficients of 0.983 and 0.969 were found for the non-genic and gene-based SNPs, respectively. In general, PIC values were biased toward higher magnitudes, while MAF values were inclined toward smaller values. Finally, there were no inconsistencies between the rankings according to PIC and MAF values so these could be used interchangeably for decisions about the markers. The SNPs with the highest level of heterozygosity were BSNPk38, BSNPk54, and BSNPk65 with 36, 28 and 22 heterozygotes, respectively. Heterozygosity signal was carefully checked using the software Snpviewer2 from KBioscience to confirm any values given to heterozygotes. In all cases, the heterozygotes and both homozygotes were observed and duplicated gene fragments were not observed to be a problem. Correlated with the PIC and MAF findings, the average number of single nucleotide differences between any pair of accessions (π) was globally extensive. The average values for this parameter were 0.445, 0.446 and 0.442, respectively, for total, gene-based and non-genic BSNPk markers.

Comparison of pairwise differences achieved using SNP markers from this study and SSR markers from Blair et al. (2006a) plotted against each other suggested a quadratic, non-linear relationship of the form $\mathrm{PD_{SSR}}/\mathrm{PD_{SNP}} = k - h\,(\mathrm{PD_{SNP}})$. This meant that the genetic distance measured by SSRs was "k" times greater than that measured by SNPs and that the proportion of SSR marker calls scored as similar due to homoplasy is directly proportional to the genetic distance measured by the SNPs. This pattern was significantly higher (p value = 0.012) when inter-genepool rather than intra-genepool comparisons were considered (Supplemental figure 3). The global relative differentiation measured by the SSR markers was 2.49 times greater than the differentiation measured by the SNP markers, and the rate of homoplasy was 1.03 times greater with SSRs than with SNPs.

Relationships between accessions

The Andean wild accession G24404 and the Mesoamerican wild accessions G24390 were distinguished from the domesticated common bean genotypes of each genepool both in the dendogram (Fig. 2) and in the first and the

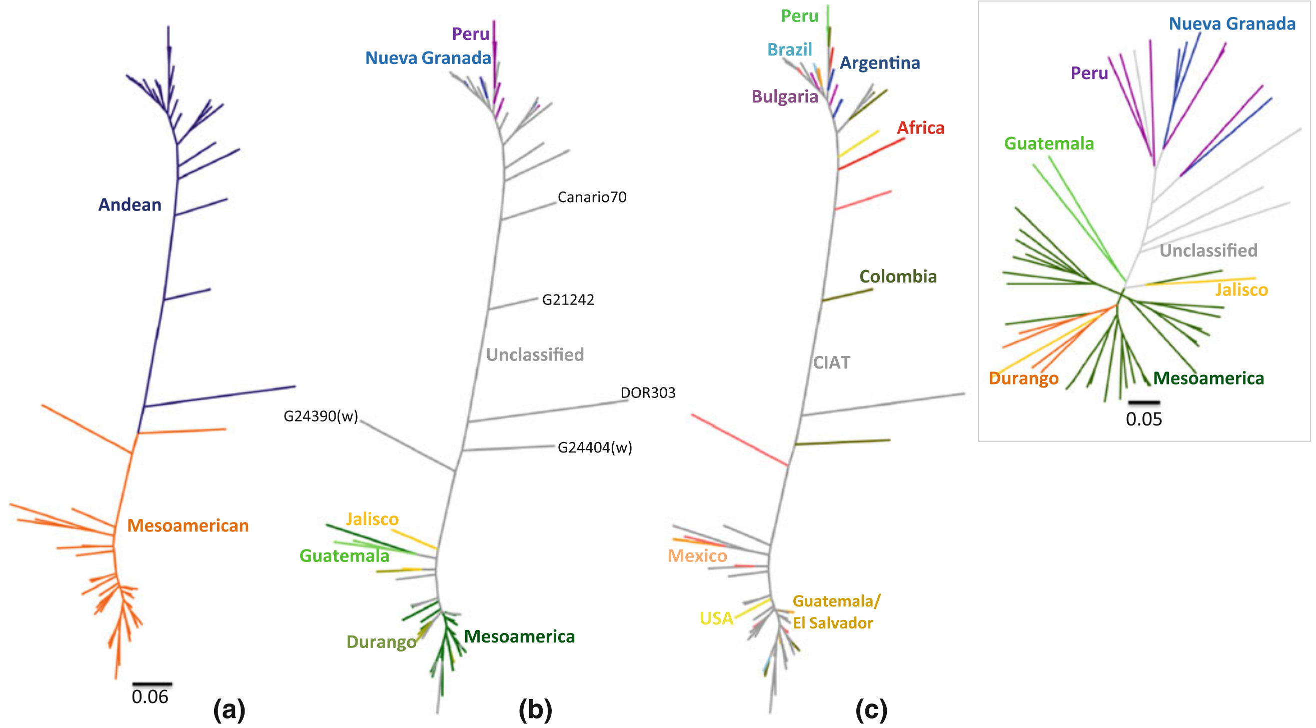

Fig. 2 Neighbor-joining trees for the results of 94 SNP markers evaluated on 70 cultivated and wild accessions according to their (**a**) gene pool, (**b**) race and (**c**) country of origin. Discontinuous red lines in first subfigure indicate reticulations. Accession names are shown for cases where introgression is recognizable. *Inset* shows SSR-based dendogram and Andean and Mesoamerican gene pool from Blair et al. (2006a)

fourth components of the principal component analysis (Fig. 3) and by the structure analysis (Fig. 4). Among the cultivated genotypes of common beans there were two principal clusters corresponding to the Andean and Mesoamerican genepools in the both analyses. These were predominantly separated by the first dimension of the principal component analysis. Within the Andean group there was evidence for two subgroups one clustering around the Peruvian accessions and the other around the Nueva Granada accessions but only separated by the fourth component and 1.47% of total variance. These same groups could be identified in the neighbor-joining trees of SSR markers from Blair et al. (2006a, b) as seen for 43 genotypes from Table 1 as analyzed by Darwin and shown in the small insert in Fig. 3.

Among the Andean genotypes, the Peru race contained type IV (G21078), type III (G19833, G19839, and G21657) and type I (Radical Cerinza) growth habit beans; while the Nueva Granada race (BRB191, G5273, JaloEEP558, and SEQ1027) did not include climbing beans. Within the Mesoamerican group there was less distinction of discrete groups or race structure although the Guatemala race genotypes (G685 and G2333) were associated and were separated from both Jalisco and Mesoamerica race genotypes in the neighbor-joining, the structure and the principal component analysis. Durango-Jalisco race genotypes G855, G11360, and MAM49 were intermediate between Guatemala and Mesoamerica race genotypes. However, MAM38, SEA5, and SEA15, of the same race, were not well separated from the Mesoamerica genotypes although they were clustered together.

Based on structure analysis some of the cultivated Andean genotypes such as Canario70 and DOR303 from lowland or hot environments as well as the mid-elevation climbing bean G21242 showed signs of introgression from the Mesoamerican genepool. In the first case, these individuals were identified because there was a high uncertainty of assignment to one of the gene pools ($K = 2$). In the second case, G21242 was recognized as introgressed because it did not cluster with the rest of the accessions and instead was placed in an intermediate position along the first component. Finally, one Mesoamerican genotype (SEL1309) showed signs of introgression from the Andean genepool although this is probably an artifact of its interspecific pedigree. No further sub-structure was revealed after considering genepools independently in the multivariate PCoA and Bayesian analyses (Supplementary figures 4 and 5, respectively).

Diversity was calculated for each of the groups and subgroups described above using the gene-based and non-genic SNP markers as well as the overall dataset (Table 3). Intra-population diversity (H_s) was highest when

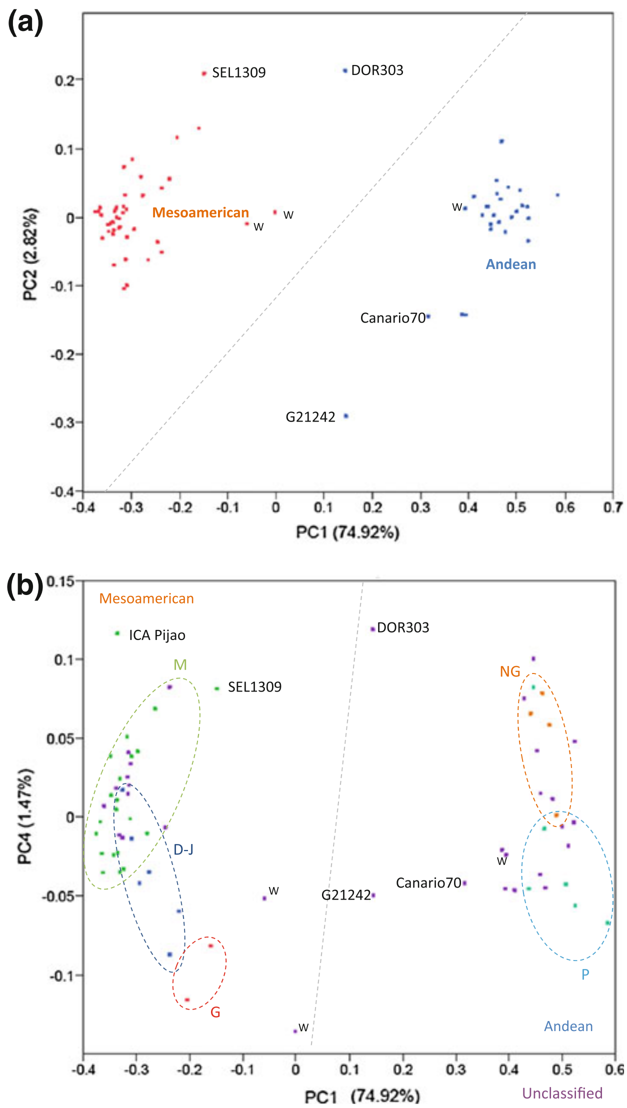

Fig. 3 Principal component analysis (PCoA) for 70 common bean accessions based on 94 SNP markers; **a** the first two components; **b** the first and the fourth components. The third component did not offer further discrimination. Accession names are shown for cases where introgression is recognizable. *W* wild accessions

comparing the wild and cultivated common beans, intermediate between the Andean and Mesoamerican genepools, and lower when comparing the races. Population differentiation (F_{ST}) did not distinguish the cultivated and wild common beans as they had overlapping patterns of diversity ($F_{ST} = 0.024$). On the other hand, population differentiation was evident when comparing the Andean and Mesoamerican genepools ($F_{ST} = 0.667$). Intra-population diversity within the Andean genepool was slightly higher than within the Mesoamerican genepool and this pattern was observed for both gene-based and non-genic BSNPk markers. Greater population differentiation was observed with the races ($F_{ST} = 0.812$) compared to the total variation.

Intra-population diversity within each of the races was lower than within the corresponding genepool as a whole. Within the Andean genepool, race Peru had higher diversity compared to race Nueva Granada, while within the Mesoamerican genepool, the races Durango-Jalisco, Guatemala and Mesoamerica had comparable levels of diversity. These patterns did not depend on whether the analysis was carried with gene-based or non-genic SNPs or both. Finally, observed and total heterogeneities were not higher for the non-genic SNPs compared to the gene-based SNPs in all the inter-population comparisons (status, genepools and races). Average heterozygosity (average number of heterozygous loci along all the surveyed accessions) was 2.42, but was more extensive in the Andean genepool (3.43) than in the Mesoamerican genepool (1.76). The Andean genotypes G24404 (wild) and Red Canadian Wonder; and the Mesoamerican parents G11360 and A55 presented the highest levels of heterozygosity with 22, 17, 12 and 12 heterozygous loci from the 94 which were surveyed. However, a total of 65 accessions presented less than 6 heterozygous loci. Average heterozygosity in wild accessions was 8.3, while it was 2.2 in cultivated genotypes.

Uniqueness for each one of the parents was calculated as the total number of pairwise differences against the rest of the accessions. Hence 69 different comparisons were considered for each parent. In this sense, the average Andean genotype uniqueness was larger than the average Mesoamerican genotype distinctiveness (3,080 vs. 2,282). Meanwhile, the uniqueness metric based on the Shapiro–Wilk test for each genepool was not normally distributed except for the Andean genepool (global p value < 0.001, Andean p value $=$ 0.8752, Mesoamerican p value < 0.001). The Kruskal–Wallis test showed that uniqueness was significantly lower for Mesoamerican accessions compared to Andean accessions ($F = 175$, p value < 0.001).

Furthermore, a randomization test based on 1,000 repetitions that used a t test statistic for unequal sample sizes and variances generated a distribution that was significantly lower (p value < 0.001) than the observed value of 17.81, confirming the observation of greater uniqueness for Andean genotypes but not for Mesoamerican genotypes than would be expected by chance (Supplemental figure 6). The parents with the highest uniqueness were G19839, Radical Cerinza and G19833 (3,346, 3,400 and 3,485, respectively). The parents with least distinctiveness were ICA Pijao and G11360 (1,792 and 1,960, respectively). Pairwise differences similarly were significantly larger between Andean accessions than between Mesoamerican accessions ($F = 4.25$, p value $= 0.043$).

Mismatch distributions, or weighted average number of polymorphic SNPs between an accession and the rest of the samples (Fig. 5) and site frequency spectra (Fig. 6) reinforced some of the previous patterns of relationships between

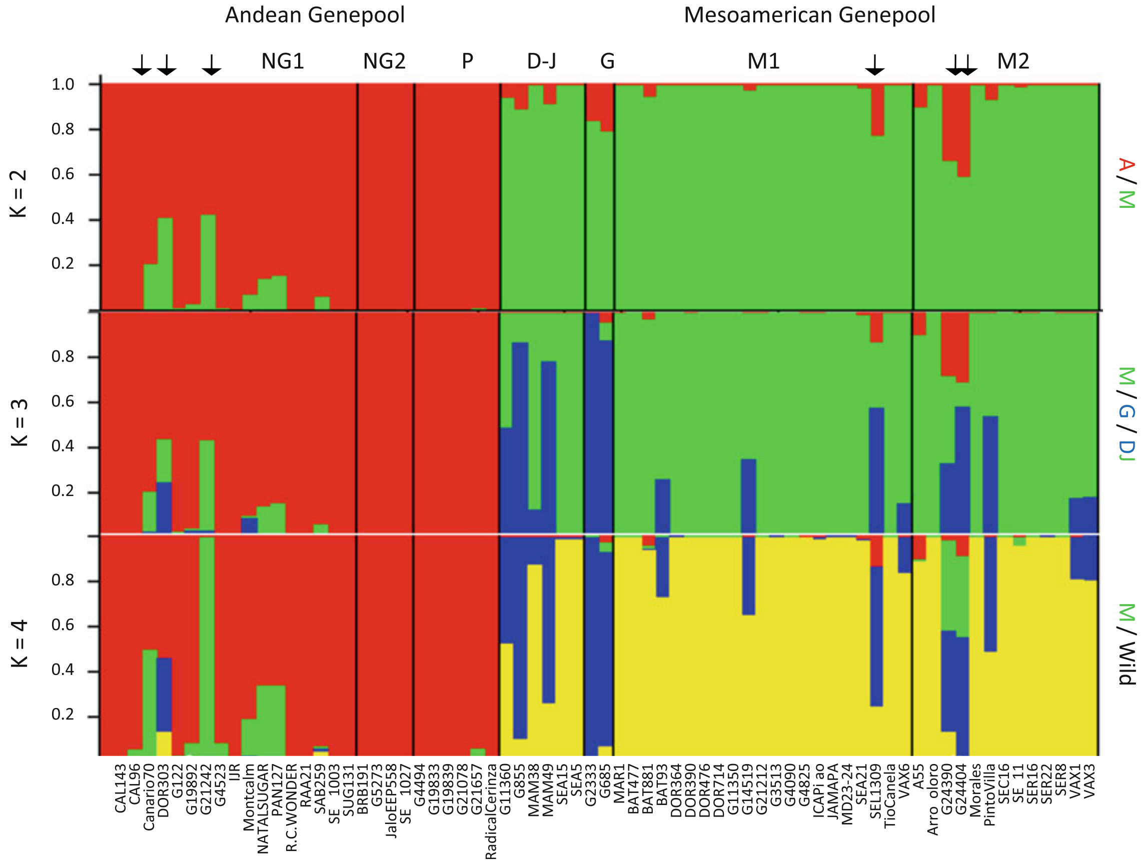

Fig. 4 Structure analysis from $K = 2$ to $K = 4$ presented for the Andean and Mesoamerican gene pools with sub-group abbreviations as D–J (Durango–Jalisco complex), G (race Guatemala) M1, M2 (race Mesoamerica subgroups); NG1, NG2 (race Nueva Granada subgroups) and P (race Peru). *Arrows* show cases of inter-genepool introgression detailed in the PCoA analysis

Table 3 Observed total (H_T) and intra population (H_s) diversity for genotypes belonging to wild and cultivated common bean, to Andean and Mesoamerican gene pools and to races within each gene pool

Category	N	Observed heterogeneity			Value
		Total BSNPk (94)	Gene-based BSNPk (84)	Non-genic BSNPk (10)	
Total	70	0.437	0.436	0.440	H_t
Status	70	0.017	0.016	0.024	F_{st}
Cultivated	67	0.435	0.435	0.438	H_s
Wild	3	0.302	0.306	0.256	H_s
Gene pools	70	0.644	0.645	0.677	F_{st}
Mesoamerican	42	0.145	0.145	0.134	H_s
Andean	28	0.171	0.169	0.154	H_s
Races	39	0.781	0.780	0.812	F_{st}
Nueva Granada	4	0.063	0.062	0.038	H_s
Peru	6	0.104	0.103	0.078	H_s
Durango–Jalisco	6	0.116	0.115	0.090	H_s
Guatemala	2	0.066	0.065	0.088	H_s
Mesoamerica	21	0.096	0.098	0.090	H_s

accessions. Particularly, a bimodal mismatch distribution was observed when all the samples were considered, as a consequence of the genepool structure. This pattern decayed for the Andean gene pool but was absent for the Mesoamerican gene pool. Finally, the goodness-of-fit with the predicted neutral model was inversely related with the bimodality (*p* values of the Kolmogorov–Smirnov test: 0.0002, 0.015 and 0.04 for the global, the Andean and the Mesoamerican analysis).

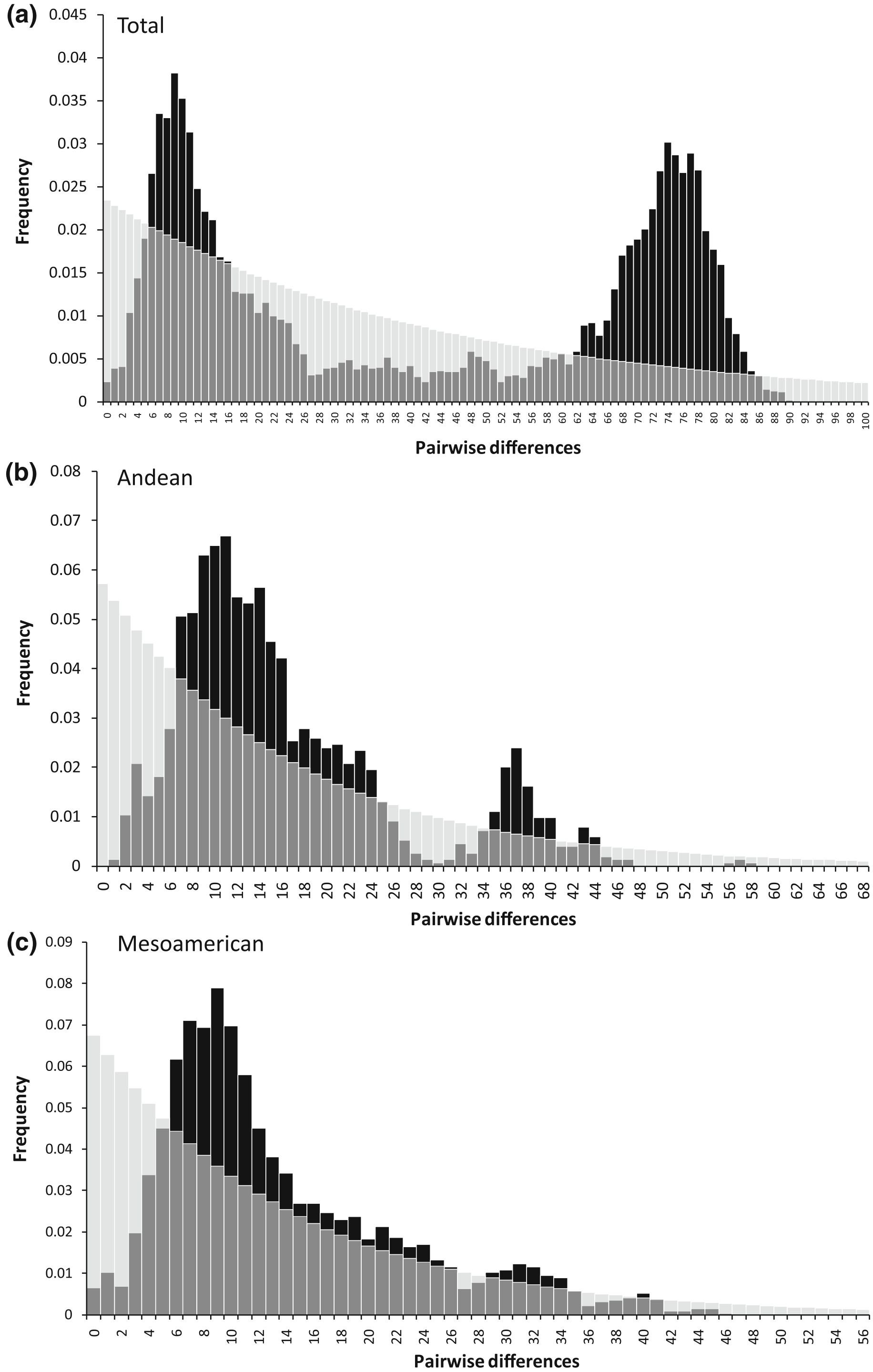

Fig. 5 Predicted (*gray bars*) and observed (*black bars*) mismatch distributions (weighted average number of polymorphic SNPs between an accession and the rest of the samples) **a** for all the cultivated common bean accessions, **b** for 28 accessions from the Andean gene pool, **c** and for the 42 accessions from the Mesoamerican gene pool. Predictions were based on the Wright–Fisher neutral model

In a similar vein, folded site frequency spectrum for all the samples showed an excess of high frequency SNPs especially for the Andean genotypes and an excess of low frequency SNPs especially for the Mesoamerican genepool. Again, the goodness of fit with the predicted neutral model was inversely related with the number of high frequency SNPs (*p*- values of the Kolmogorov–Smirnov test: 0.0001, 0.025 and 0.042 for the global, the Andean and the Mesoamerican analysis, respectively).

Comparisons of within and inter-genepool parental combinations

Among the parental combinations represented in this diversity survey were crosses between cultivars, between

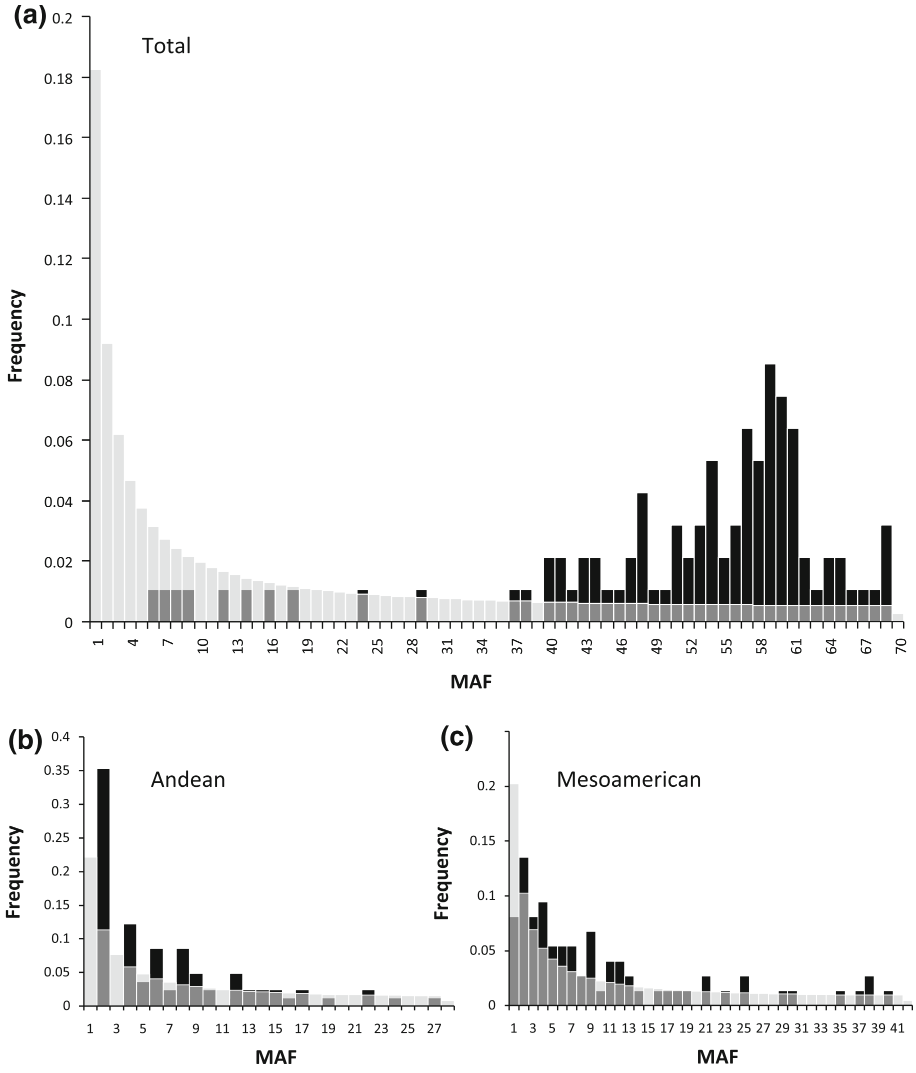

Fig. 6 Site frequency spectrum based on proportion of SNP markers with a specific minimum allele frequency (MAF); **a** for the entire group of cultivated and wild common bean, **b** for 28 accessions from the Andean gene pool, and **c** for the 42 accessions from the Mesoamerican gene pool

genepools and between wild accessions and cultivated genotypes (Table 4). The inter-genepool (Andean × Mesoamerican) combinations had higher polymorphism rates (71.7%) than the intra-genepool combinations (15.4%). The most polymorphic of the inter-genepool combinations was DOR364 × G19833 (Mesoamerican × Peru races) which had an average level of polymorphism of 86.2%, however, many of the other combinations between cultivated Mesoamerican and Andean beans had similar levels of polymorphism (from 53.2 to 86.2% on average). Among the inter-genepool combinations, DOR364 × G19833 was included because it was used to create the microsatellite map in Blair et al. (2003), while BAT93 × JaloEEP558 was included because it was the basis for the integrated genetic map of Freyre et al. (1998). Both of these inter-genepool comparisons had similar levels of polymorphism although the combination DOR364 × G19833 was slightly higher in overall polymorphism (81.9 and 86.2%, respectively) perhaps due to the distance of the G19833 Peru race accession from small red Mesoamericans.

Among the within-genepool crosses, comparisons between Andean genotypes had higher polymorphism (24.9%) on average than comparisons between Mesoamerican genotypes (10.2%). This was especially notable with the parental combination of the cultivated Radical Cerinza and the wild accession G24404 (38.3%) from the Andean genepool followed by the parental combination of the cultivated race Peru genotypes G21078 with G21242 (34%) compared to the highest polymorphism parental combination of the Mesoamerican parents G14519 and G4825 (18.1%).

Within the Mesoamerican genepool combinations, the intra-racial combinations between Durango or Jalisco and Mesoamerica race genotypes showed lower average polymorphism (4.8%) than the within-race combinations between Mesoamerica race genotypes (11.7%). Examples of intra-racial parental combinations with moderate polymorphism included DOR476 × SEL1309 (15.5%) and BAT881 × G21212 (14.9%). By comparison, polymorphism was extremely low in the inter-racial combination

Table 4 Level of polymorphism in parental combinations across or within Mesoamerican (M) and Andean (A) gene pools for gene-based and non-genic SNP markers

Source	Parental combination		Type of cross	Gene-based (84)		Non-genic (10)		Total (94)	
	Female parental	Male parental		No. Poly	%Poly	No.Poly	%Poly	No. Poly	%Poly
Blair et al. (2006a, b)	G11360	G11350	M(j) × M(m)	1	1.2	0	0.0	1	1.1
	G21657	G21078	A(p) × A(p)	5	6.0	1	10.0	6	6.4
	G21078	G21242	A(p) × A(na)	31	36.9	1	10.0	32	34.0
	G14519	G4825	M(m) × M(m)	14	16.7	3	30.0	17	18.1
	DOR364	G19833	M(m) × A(p)	73	86.9	8	80.0	81	86.2
	DOR364	BAT477	M(m) × M(m)	6	7.1	0	0.0	6	6.4
	DOR364	G3513	M(m) × M(m)	4	4.8	1	10.0	5	5.3
	BAT881	G21212	M(m) × M(m)	13	15.5	1	10.0	14	14.9
	Radical Cerinza	G24404	A(p) × A(w)	30	35.7	6	60.0	36	38.3
	Radical Cerinza	G24390	A(p) × M(w)	45	53.6	5	50.0	50	53.2
	DOR390	G19892	M(m) × A(w)	59	70.2	7	70.0	66	70.2
	DOR476	SEL1309	M(m) × M(m)	23	27.4	1	10.0	24	25.5
	BAT93	JaloEEP558	M(m) × A(ng)	67	79.8	10	10.0	77	81.9
	VAX6	MAR1	M(m) × M(m)	9	10.7	1	10.0	10	10.6
	G2333	G19839	M(g) × A(p)	56	66.7	7	70.0	63	67.0
	G855	BRB191	M(j) × A(ng)	56	66.7	7	70.0	63	67.0
	BRB191	MAM38	A(ng) × M(d)	62	73.8	7	70.0	69	73.4
	G5273	MAM38	A(ng) × M(d)	66	78.6	7	70.0	73	77.7
	BRB191	MAM49	A(ng) × M(d)	58	69.0	6	60.0	64	68.1
	MAM49	G5273	M(d) × A(ng)	58	69.0	6	60.0	64	68.1
	SEQ1027	G4090	A(ng) × M(m)	60	71.4	7	70.0	67	71.3
	TioCanela	DOR714	M(m) × M(m)	0	0.0	1	10.0	1	1.1
	SEA5	MD23-24	M(d) × M(m)	8	9.5	0	0.0	8	8.5
Others	A55	G122	M × A	49	58.3	7	70.0	56	59.6
	G122	Montcalm	A × A	10	11.9	2	20.0	12	12.8
	SEA5	CAL96	M × A	59	70.2	9	90.0	68	72.3
	SEA15	CAL96	M × A	59	70.2	8	80.0	67	71.3
	SEA5	CAL143	M × A	65	77.4	9	90.0	74	78.7
	SEA15	CAL143	M × A	67	79.8	8	80.0	75	79.8
	SEA5	BRB191	M × A	59	70.2	9	90.0	68	72.3
	SEA15	BRB191	M × A	61	72.6	8	80.0	69	73.4
	DOR303	IJR	A × A	26	31.0	5	50.0	31	33.0

Inter genepool and inter race combinations indicated by abbreviations where *A* Andean and *M* Mesoamerican. Genepools followed by an additional letter in parenthesis where *d* Durango, *g* Guatemala, *j* Jalisco, *m* Mesoamerica, *ng* Nueva Granada, *p* Peru race and *w* wild accession

G11360 × G11350 (1.1%). Polymorphism was equally low for the parental comparison between two parents of the same grain color class such as Tio Canela and DOR714, which both belong the race mesoamerica. This is a type of cross within small red seeded genotypes of the same subrace. The combinations between the wild and cultivated parents of different genepools, Radical Cerinza × G24390 and DOR390 × G19892 were similar to the averages of Andean × Mesoamerican combination within the cultivated genotypes (53.2 and 70.2%, respectively).

Although the differences within parental combinations were significant, there was no statistically significant difference when all the possible parental hypothetical crosses across and within both genepools were compared. This was shown by the non-normal distribution (Supplemental figure 7) of all the possible pairwise differences within each genepool (Shapiro–Wilk test: global p value < 0.001, Andean p value < 0.001 and Mesoamerican p value < 0.001). Despite this, a set of Kruskal–Wallis tests showed that pairwise differences were significantly larger between Andean accessions than between Mesoamerican accessions ($F = 4.25$, p value $= 0.043$). Finally, another randomization test using the t statistic (for unequal sample sizes and variances) generated a distribution that was slightly lower than the observed value of 1.34 (p value = 0.482).

Discussion

Single nucleotide polymorphisms are an example of molecular markers with a potential role for diversity and association analysis in any genome because they are the most abundant polymorphism which can be used to uncover diversity (Chagné et al. 2007). This paper is the first attempt to examine SNP variation in common bean. Furthermore, this research integrates different lines of evidence to propose useful applications of SNP markers for the study of common bean genetics. Additionally, the significance of our results is supported by the fact that we evaluated a broad set of tropically adapted wild and cultivated dry bean genotypes of various growth habits, as well as improved and unimproved germplasm from various seed classes ranging from carioca, small red, black, large red, red mottled, cream-mottled to yellow-mottled types, and that we used a flexible technological platform, namely KASPar, that can be easily validated for any additional combination of germplasm or gene and non-genic fragments.

BSNPk markers capture the essentials of population structure in common bean

In terms of the diversity assessment, four main observations were detected. First, SNP polymorphism in common beans are extensive in inter-genepool comparisons even in conserved gene sequences. The high within species diversity of SNPs in common bean reflected the dual domestication events of the Andean and Mesoamerican genepools and a greater level of inter-genepool hybridization, as was suggested previously (Blair et al. 2006a). Second, SNP categorization of germplasm agrees with previous analysis of the origins of cultivated common bean conducted with isozymes (Singh et al. 1991), RFLPs (Becerra-Velazquez and Gepts 1994; Sonnante et al. 1994), RAPDs (Beebe et al. 2000), AFLPs (Beebe et al. 2001; Tohme et al. 1996), and SSRs (Blair et al. 2009) in terms of identifying a wide chasm between the Andean and Mesoamerican genepools.

In terms of other sub-divisions, SNP markers allowed the identification of two Andean clusters corresponding to the Nueva Granada and Peru races and three Mesoamerican clusters corresponding to the Mesoamerica, Guatemala and Durango-Jalisco races. The Guatemala race was the most distant from the other Mesoamerican clades, and the Mesoamerica and Durango-Jalisco races formed the most undifferentiated sub-populations. However, SSR markers previously revealed further structure within the Nueva Granada, Peru, Mesoamerica and Durango-Jalisco races and a better differentiation between races of the same genepool (Blair et al. 2009). We must note that we did not include race Chile genotypes as this will be part of a separate study.

A third interesting aspect of this study would be that the genepool diversity with SNPs was higher in the Andean genepool than within the Mesoamerican genepool, at least in terms of real parental crosses if not statistically significant for all pairwise comparisons. As a result, SNPs may be more useful for genetic mapping in Andean x Andean parental combinations than for Mesoamerican x Mesoamerican parental combinations. The same feature was observed in the first systematic diversity survey of common bean carried out with SSR markers (Blair et al. 2006a). Three hypotheses have been proposed to explain the greater diversity within Andean beans. The first hypothesis proposes that this pattern may be a reflection of the multiple growth habits and agroecological origins of Andean accessions, while the similarity of some of the Mesoamerican genotypes may be a result of their ancestry from an inter-racial mix of parents (Blair et al. 2006a). Otherwise, the higher diversity of the Andean genotypes could be a consequence of the selection of genotypes from a greater range of agroecologies typical of the regions where Nueva Granada and Peru race cultivars are grown (Singh et al. 1991). Self-pollinating with 95–99% inbreeding, heterozygotes cannot be ruled out due to some level of outcrossing usually of <5% but higher in wild accessions. This was found specifically for G24404, the wild accession from Colombia, which was the individual with the highest heterogeneity along all the surveyed SNPs. The wild nature of G24404 and its location in a region of confluence between the Mesoamerican and Andean genepools make this observation not surprising. Similar observations for this accession were made previously with SSR markers (Blair et al. 2006a). By comparison, many Mesoamerican genotypes are from the CIAT breeding program or from Central America. Higher diversity within the Andean genepool may have also been due to introgression of Mesoamerican or wild accession alleles into this genepool (Beebe et al. 2001; Kwak and Gepts 2009), which is reinforced by our findings. Hence, we propose a scenario in which agroecological variation, agricultural practices and introgression have contributed to the higher heterogeneity within the Andean genepool. We also cannot rule out earlier and more widespread domestication in the Andes than in Mesoamerica.

This evolutionary scenario is also related with the unexpected patterns of pairwise differences, site frequencies and single nucleotide diversity. Demographic process, such as bottlenecks and population expansions, imprint the genomes by different mechanisms, causing the departure of genetic variation from the neutral expectations (Behar et al. 2010; Reagon et al. 2010). Recent bottlenecks are associated with high values of single nucleotide diversity (π) because only medium frequency polymorphisms can avoid being eliminated by the demographic filter where modern genetic variants coalesce at a faster rate than ancient genetic variants (Wakeley 2008). In our study, we found patterns of recent bottlenecks when inter-genepool comparisons were considered. Ancient bottlenecks may be evident in intra-genepool comparisons but these would be associated with low values of single nucleotide diversity because accumulation of unique genetic variants increases as a function of time and population expansion whereby modern genetic variants coalesce at a slower rate than ancient genetic variants (Fay and Wu 1999). Although balancing selection and selective sweeps tend to achieve, respectively, the same increase and reduction in the nucleotide variation (Xia et al. 2010), a genome-wide selective imprint is not frequent (Caicedo et al. 2007). On the other hand, independent domestication events, extensive population structure and recent bottlenecks tend to homogenize haplotype blocks within the same population, fix polymorphisms in different populations, and eliminate low frequency polymorphism. Consequently, few haplotypes with high frequency are generated, corresponding to high values of single nucleotide diversity (Xia et al. 2010).

In this study, we observed a global excess of middle frequency SNPs through the folded site frequency spectra and presume that genepool structure accounted for this tendency. We also saw a global bimodal mismatch distribution and extensive single nucleotide diversity especially within the Andean genepool. This aspect is congruent with the higher heterogeneity within the Andean common bean. Even more interesting is the fact that both Andean and Mesoamerican unimodal mismatch distributions are squeezed toward the right of the predicted Wright-Fisher neutral distribution, as was expected under the model of population expansion of Rogers (1995). Hence, within-genepool variation suggests that population explosions occurred after the two independent domestication bottlenecks, particularly within the Andean genepool.

Overall, polymorphism detected in the diversity panel agreed well with genepool and race structure in beans according to Blair et al. (2009). For example, SNP polymorphism was low in parental comparisons from the same race within the Mesoamerican genepool, slightly higher for parents from different races within the Mesoamerican genepool or for races within the Andean genepool, and still higher for crosses between genepools, as was also observed by Blair et al. (2006a). Observed heterozgosity was low for common bean SNPs; however, since DNA was pooled from four plants, it was not possible to distinguish between heterogeneous homozygous lines or truly heterozygous lines.

The utility of SNPs and SSRs is not redundant

As discussed above, difference in polymorphism rate were equally evident when using cDNA derived and non-genic SNPs over all the inter-genepool and intra-genepool comparisons. Parental comparisons made in this study were representative of the types of parental combinations used in common bean research and show the value of recently developed SNPs for efficient genetic analysis of *Phaseolus*. SNP markers are especially useful for inter-genepool comparisons, but not for intra-race combinations as with SSRs (Blair et al. 2006a, 2009). This could be a consequence of the effect of saturation (Felsenstein 2006). In this sense, the high mutation rate of SSRs, in comparison with the lower mutation rate of SNPs, inflates homoplasy of alleles at the level of inter-genepool comparisons.

This is especially true if we compare the *K* alleles model and the stepwise mutation model of SSR evolution against the infinite site model of SNP evolution. Hence, the probability that the same allele at the inter-genepool level comes from a common ancestor (identity by descent) is higher for SNP markers than for SSR markers. Instead, the probability that the same allele at the inter-genepool level comes from repetitive mutation (identity by state) is theoretically higher for SSR than for SNP markers. Our results reinforced this aspect and imply that homoplasy certainly has to be minimized for any application that relies on the shared alleles as a base for kinship. Therefore, in terms of saturation rate, it is more convenient to use SNP markers the inter-genepool level, and SSR markers at an inter-race, intra-genepool scale. Given the previous considerations, we propose that the ideal genetic diversity survey either for mapping or association studies in common bean should consider both SNP and SSR markers.

Although SNPs are more abundant and less susceptible to saturation at deeper evolutionary scales than the second ones, SSRs are more polymorphic within races and between populations that have diverged recently. SSRs markers have been recognized as a useful alternative for association mapping (Bahram and Inoko 2007), although they face some limitations (Jorgenson and Witte 2007). The conformation of core sets of molecular markers especially those associated with genes and causative mutations (Liu et al. 2011), is particularly useful in common bean because of its complex evolutionary history and its extensive stratification. For example, *P. vulgaris* emerged as a single species one million years ago, genepool structure dates to 20,000 years before present in the wild ancestors of modern common bean, domestication was 5,000–8,000 years ago, and secondary diversification centers were shaped less than 500–4,000 years before present (Schoonhoven and Voysest 1991). Other similar arsenals of molecular markers have been proposed as useful in rice (Zhao et al. 2010), apple (Koopman et al. 2007) and maize (Yan et al. 2010) for uncovering different levels of variation across complex genepools, races and even species.

In conclusion, our study represents a baseline for the choice of SNP markers for future applications because they were carefully chosen from transcriptome projects and candidate genes, and therefore will be useful to analyze the history of selection and diversification for specific loci. In addition, these markers will allow mapping in wide inter-genepool crosses, will assist the phylogenetic analysis of the genus, and will facilitate the analysis of specifica traits along the evolutionary history. Hence, any genome-wide scan of SNP polymorphism in common bean should include the set of 84 gene-based SNPs which we validated and surveyed in the present work. However, we emphasize that SSR markers are still essential to access the stratification, the parental polymorphisms and evolutionary processes that occurred within each genepool.

Overall, we can conclude that our experience with KASPar technology (Cuppen 2007) has been successful: out of 6,580 reactions, only 165 failed (2.5%) and there was an entire correspondence between Sanger sequencing and KASPar genotyping. In terms of costs, all the genotyping was carried out with less than US$1200 (US$0.18 per reaction); and the delivery time, including the design and genotyping phases, was of 5 weeks. We concluded from a cost comparison that the failure rate, *per* genotype price and delivery time were lower and competitive in comparison with other SNP genotyping technologies (Chagné et al. 2007). For example, while GoldenGate and Sequenom technologies are comparable in terms of price and quality, they are less flexible in terms of sample size and number of SNPs genotyped and less efficient in terms of design phase delivery time (Yan et al. 2010). In addition, the need for high quality DNA appears to be minimal with the KASPar technology, but perhaps limiting with Illumina genotyping. To date, KASPar genotyping kits are available for a number of wild and domesticated species of animals and plants (KBioscience) and are in common use for genotyping of humans (Bauer et al. 2009), rats (Nijman et al. 2008) and fish (Borza et al. 2010). We predict that this technology will spread rapidly in common bean because of its flexibility, quality, efficiency and competitive prices to generate fast and cost-effective SNP genotyping platforms. Finally, the establishment of a consolidated and validated resource of SNPs in a representative sample of common bean allowed the identification of evolutionary patterns that have modulated genetic diversity, such as bottlenecks and selective sweeps, which have been previously thought to be the dominant demographic and selective forces shaping levels of nucleotide variation in crop species (Caicedo et al. 2007; Camus et al. 2008). An integrative study of this kind will open the possibility for a genome-wide-scan aimed to detect QTLs as part of a mapping studies (Ioannidis et al. 2009), or further candidate gene studies (Blair et al. 2010b). Further development of BSNP markers is expected to assist in fine mapping and candidate gene analysis.

Acknowledgments The authors wish to thank the personnel of the Genetic Resource Unit and Bean Programs for seed multiplication. In particular, we are grateful with L. Díaz and A. Hoyos for providing us with the SSR data and passport information, respectively, and with V.M. Mayor for facilitating DNA of drought tolerance lines. We also acknowledge the KBioscience Center for support in SNP genotyping, are grateful to S. Madriñán for advice and analytical support and to the anonymous reviewers for their valuable comments. This research was supported by the Tropical Legume I project of the Generation Challenge Program.

References

Afanador LK, Hadley SD (1993) Adoption of a mini-prep DNA extraction method for RAPD marker analysis in common bean. Bean Improv Coop 35:10–11

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Bahram S, Inoko H (2007) Microsatellite markers for genome-wide association studies. Nat Rev Genet 8

Bauer F, Elbers CC, Adan RAH, Loos RJF, Onland-Moret NC, Grobbee DE, van Vliet-Ostaptchouk JV, Wijmenga C, van der Schouw YT (2009) Obesity genes identified in genome-wide association studies are associated with adiposity measures and potentially with nutrient-specific food preference. Am J Clin Nutr 90:951–959

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

Beebe S, Skroch PW, Tohme J, Duque MC, Pedraza F, Nienhuis J (2000) Structure of genetic diversity among common bean landraces of Mesoamerican origin based on Correspondence Analysis of RAPD. Crop Sci 40:264–273

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

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, Chaubey G, Kutuev I, Yudkovsky G, Khusnutdinova EK, Balanovsky O, Semino O, Pereira L, Comas D, Gurwitz D, Bonne-Tamir B, Parfitt T, Hammer MF, Skorecki K, Villems R (2010) The genome-wide structure of the Jewish people. Nature 466:238–242

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

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

Blair MW, Iriarte G, Beebe S (2006b) 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, Duque MC (2007) Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 116:29–43

Blair M, Diaz LM, Buendia HF, 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

Blair MW, Chaves A, Tofino A, Calderon JF, Palacio JD (2010a) Extensive diversity and inter-genepool introgression in a world-wide collection of indeterminate snap bean accessions. Theor Appl Genet 120:1381–1391

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

Borza T, Higgins B, Simpson G, Bowman S (2010) Integrating the markers Pan I and haemoglobin with the genetic linkage map of Atlantic cod (*Gadus morhua*). BMC Res Notes 3:261

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

Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, Bustamante CD, Purugganan MD (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet 3:1745–1756

Camus L, Chevin LM, Cordet CT, Charcosset A, Manicacci D, Tenaillon MI (2008) Patterns of molecular evolution associated with two selective sweeps in the *Tb1-Dwarf8* region in maize. Genetics 180:1107–1121

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

Chagné D, Batley J, Edwards D, Forster JW (2007) Single nucleotide polymorphism genotyping in plants. In: Oraguzie NC, Rikkerink EHA, Gardiner SE, Silva HNd (eds) Association mapping in plants. Springer, NY, pp 77–94

Córdoba JM, Chavarro C, Schlueter JA, Jackson SA, Blair MW (2010) Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. BMC Genomics 11:436

Cuppen E (2007) Genotyping by allele-specific amplification (KASPar). Cold Spring Harb Protocols, pp 172–173

David P, Sevignac M, Thareau V, Catillon Y, Kami J, Gepts P, Langin T, Geffroy V (2008) BAC end sequences corresponding to the B4 resistance gene cluster in common bean: a resource for markers and synteny analyses. Mol Genet Genomics 280:521–533

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

Fay JC, Wu CI (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol 16:1003–1005

Felsenstein J (2006) Inferring phylogenies. Sinauer Associates, New York

Freyre R, Skroch P, Geffroy V, Adam-Blondon AF, Shirmohamadali A, Johnson W, Llaca V, Nodari R, Pereira P, Tsai SM, Tohme J, Dron M, Nienhuis J, Vallejos CE, Gepts P (1998) Towards an integrated linkage map of common bean. 4. Development of a core map andalignment of RFLP maps. Theor Appl Genet 97:847–956

Gaitan E, Choi IY, Quigley C, Cregan P, Tohme J (2008) Single nucleotide polymorphisms in common bean: their discovery and genotyping using a multiplex detection system. Plant Genome 1:125–134

Galeano CH, Fernandez AC, Gomez M, Blair MW (2009a) Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and synteny analysis of common bean (*Phaseolus vulgaris* L.). BMC Genomics 10:629

Galeano CH, Gomez M, Rodriguez LM, Blair MW (2009b) CEL I nuclease digestion for SNP discovery and marker development in common bean (*Phaseolus vulgaris* L.). Crop Sci 49:381–394

Gepts P (1998) Origin and evolution of common bean: past events and recent trends. HortScience 33:1119–1135

Gepts P, Debouck DG (1991) Origin, domestication, and evolution of the common bean. In: van Schoonhaven A, Voysest O (eds) Common beans: research for crop improvement. Centro Internacional de Agricultura Tropical (CIAT), Cali

Gepts P, Osborn TC, Rashka K, Bliss FA (1986) 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

Hair JF, Anderson RE, Tatham RL, Black WC (1992) Multivariate data analysis with readings. Macmillan Publishing Company, New York

Hougaard BK, Madsen LH, Sandal N, Moretzsohn MC, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, Bertioli DJ, Stougaard JF (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. Genetics 119:2299–2312

Hyten DL, Cannon SB, Song QJ, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB (2010a) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38

Hyten DL, Song Q, Fickus EW, Quigley CV, Lim J-S, Choi I-Y, Hwang E-Y, Pastor-Corrales M, Cregan PB (2010b) High-throughput SNP discovery and assay development in common bean. BMC Genomics 11:475

Ioannidis J, Thomas G, Daly MJ (2009) Validating, augmenting and refining genome-wide association signals. Nat Rev Genet 10:319–329

Jorgenson E, Witte JS (2007) Microsatellite markers for genome-wide association studies. Nat Rev Genet 8

Koopman WJM, Li Y, Coart E, Weg WEVD, Vosman B, Roldán-Ruiz I, Smulders MJM (2007) Linked vs. unlinked markers: multilocus microsatellite haplotype-sharing as a tool to estimate gene flow and introgression. Mol Ecol 16:243–256

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

Kwak M, Kami JA, Gepts P (2009) The putative mesoamerican domestication center of *Phaseolus vulgaris* is located in the Lerma-Santiago Basin of Mexico. Crop Sci 49:554–563

Liu S, Zhou Z, Lu J, Sun F, Wang S, Liu H, Jiang Y, Kucuktas H, Kaltenboeck L, Peatman E, Liu Z (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. BMC Genomics 12:53

Makunde GS, Beebe S, Blair MW, Chirwa R, Lungu D (2007) Inheritance of drought tolerance traits in Andean × Andean and Andean × Mesoamerican F2 Populations. Bean Improv Coop 50:159–160

McConnell M, Mamidi S, Lee R, Chikara S, Rossi M, Papa R, McClean P (2010) Syntenic relationships among legumes revealed using a gene-based genetic linkage map of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 121:1103–1116

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nijman IJ, Kuipers S, Verheul M, Guryev V, Cuppen E (2008) A genome-wide SNP panel for mapping and association studies in the rat. BMC Genomics 9:95

Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. Theor Appl Genet 106:239–250

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

Ramirez M, Graham MA, Blanco-Lopez L, Silvente S, Medrano-Soto A, Blair MW, Hernandez G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol 137:1211–1227

Reagon M, Thurber CS, Gross BL, Olsen KM, Jia YL, Caicedo AL (2010) Genomic patterns of nucleotide diversity in divergent populations of US weedy rice. BMC Evol Biol 10:180

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. Evolution 49:608–615

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu SQ, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du JC, Tian ZX, Zhu LC, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schoonhoven Av, Voysest O (1991) Common beans: research for crop improvement. Centro Internacional de Agricultura Tropical, Cali

Singh SP (1982) A key for identification of different growth habits of *Phaseolus vulgaris* L. Bean Improv Coop 25:92–95

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

Sonnante G, Stockton T, Nodari R, Becerra Velasquez V, Gepts P (1994) Evolution of genetic diversity during the domestication of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 89:629–635

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

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

Wakeley J (2008) Coalescent theory: an introduction. Harvard University, Cambridge

Wright S (1969) Evolution and the genetics of populations, volume 2: the theory of gene frequencies. University of Chicago Press, Chicago

Xia H, Camus-Kulandaivelu LT, Stephan W, Tellier AL, Zhang Z (2010) Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. Mol Ecol (online version)

Yan JB, Yang XH, Shah T, Sanchez-Villeda H, Li JS, Warburton M, Zhou Y, Crouch JH, Xu YB (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451

Zhao KY, Wright M, Kimball J, Eizenga G, McClung A, Kovach M, Tyagi W, Ali ML, Tung CW, Reynolds A, Bustamante CD, McCouch SR (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. Plos One 5:e10780