

Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands

Asrat Asfaw · Matthew W. Blair · Conny Almekinders

Received: 20 June 2009 / Accepted: 30 August 2009 / Published online: 16 September 2009
© Springer-Verlag 2009

Abstract The East African highlands are a region of important common bean production and high varietal diversity for the crop. The objective of this study was to uncover the diversity and population structure of 192 landraces from Ethiopia and Kenya together with four genepool control genotypes using morphological phenotyping and microsatellite marker genotyping. The germplasm represented different common bean production ecologies and seed types common in these countries. The landraces showed considerable diversity that corresponded well to the two recognized genepools (Andean and Mesoamerican) with little introgression between these groups. Mesoamerican genotypes were predominant in Ethiopia while Andean genotypes were predominant in Kenya. Within each country, landraces from different collection sites were clustered together indicating potential gene flow between regions within Kenya or within Ethiopia. Across countries, landraces from the same country of origin tended to cluster

together indicating distinct germplasm at the national level and limited gene flow between the two countries highlighting divided social networks within the regions and a weak trans-national bean seed exchange especially for landrace varieties. One exception to this may be the case of small red-seeded beans where informal cross-border grain trade occurs. We also observed that genetic divergence was slightly higher for the Ethiopian landraces compared to Kenyan landraces and that Mesoamerican genotypes were more diverse than the Andean genotypes. Common beans in eastern Africa are often cultivated in marginal, risk-prone farming systems and the observed landrace diversity should provide valuable alleles for adaptation to stressful environments in future breeding programs in the region.

Introduction

Common bean (*Phaseolus vulgaris* L.) is one of the principal grain legumes of eastern and southern Africa, occupying more than 4 million ha annually and providing food for more than 100 million people (Wortmann et al. 1998). It is the second most important source of dietary protein and the third most important source of calories for lower income African households after cassava and maize (Broughton et al. 2003). Total production in sub-Saharan Africa is around 3.5 metric tons with 62% of production in East African countries of Burundi, DR Congo, Ethiopia, Kenya, Rwanda, Tanzania and Uganda making this the most important regions for the crop within the continent. Although the crop is basically cultivated for home consumption in much of East Africa, it is also rapidly evolving into a cash crop in certain countries with Ethiopia earning about US\$ 6.25 million (equivalent to 60 million ET Birr) from bean exports in 2005 (Teshale Assefa, personal

Communicated by F. Muehlbauer.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1154-7) contains supplementary material, which is available to authorized users.

A. Asfaw
Awassa Agricultural Research Center,
P.O. Box 6, Awassa, Ethiopia

M. W. Blair (✉)
Centro Internacional de Agricultura Tropical (CIAT),
AA 6713 Cali, Colombia
e-mail: m.blair@cgiar.org

C. Almekinders
Technology and Agrarian Development Group,
Wageningen University and Research Centre, Hollandseweg 1,
6706 KN Wageningen, The Netherlands

communication 2007). Regional trade is significant for some trading partners, for example from Ethiopia, Tanzania and Uganda into Kenya.

Cultivated common beans originated in Latin America from two recognized centers of domestications about 7,000–8,000 years ago (Gepts and Debouck 1991). The multiple regions of domestications endowed the crop with relatively high diversity that is broadly classified into two gene pools: Mesoamerican and Andean (Gepts et al. 1986; Singh et al. 1991a, b, c). The two gene pools further differentiate into different races, such as Mesoamerica, Durango, Jalisco and Guatemala in the Mesoamerican gene pool and Nueva Granada, Peru and Chile in the Andean gene pool (Singh et al. 1991b; Beebe et al. 2000). The gene pool and race differences have been validated using various marker systems including seed size, phaseolin (seed storage protein) patterns, plant morphology, isozymes, RFLP, RAPD, AFLP and microsatellite markers (Singh et al. 1991a, b, c; Becerra and Gepts 1994; Beebe et al. 2000, 2001; Islam et al. 2004; Blair et al. 2006, 2009).

Common beans are believed to have been introduced together with maize into the east coast of Africa by Portuguese and Spanish traders in the sixteenth and seventeenth century (Greenway 1945; Gentry 1969). Since then farmers have developed farming practices adapted to local conditions by preservation and exploitation of useful alleles which have resulted in a range of morphologically diverse landraces (Wortmann et al. 1998; Sperling 2001). Moreover, with recent efforts to improve on-farm level productivity by many national bean-breeding programs in Africa, new germplasm sources have been continually introduced to African farming systems from different parts of the world since the 1980s (CIAT 2005). The existence of both gene pools (Andean and Mesoamerican) in Africa has furthermore been documented (Martin and Adams 1987) and probably is a result of original introductions and subsequent imports of novel germplasm. Given the wide range of landraces on the continent, Africa can be considered to be a secondary center of diversity for common beans (Allen and Edje 1990; Wortmann et al. 1998; Sperling 2001).

Despite recognition of the genetic diversity of common beans in eastern and southern Africa, it is not clear if the observed variation in and between landraces is the result of the original differences between the various introductions or whether they result from a continuous process of natural hybridization and selection by farmers and by the environment. Gene flow within and between gene pools and races via spontaneous out-crossing in farmer's field or crossing programs in formal breeding could result in intermediate phenotypes that do not correspond well to any of the single race or gene pool divisions (Beebe et al. 2001; Islam et al. 2004; Díaz and Blair 2006; Blair et al. 2007). Understanding the pattern of population-genetic structure and diversity

of bean landraces and cultivars (hereafter accessions) and their relationships with the Andean and Mesoamerican gene pools can therefore provide information on gene flow and be of great importance for future common bean breeding in the region. However, to date, the diversity assessment exercises in the region have mainly been limited to agro-morphological traits and no comprehensive marker evaluation of bean landraces has been conducted.

This study, therefore, aims to examine the genetic diversity and relationships among and within accessions from two East African countries (Ethiopia and Kenya) in relation to the Andean and Mesoamerican gene pools using microsatellite marker analyses combined with morphological evaluation. Ethiopia and Kenya accessions were selected for analysis because these countries are among the ten largest common bean producers in sub-Saharan Africa (Hillocks et al. 2006) and bean production is very diverse in terms of agro-ecology, social settings and production systems (from monocrop to relay-cropping or intercropping) found in both countries. Throughout these countries, the crop is grown mainly by farmers with low external inputs and landraces remain the dominant source of seed for planting, although popular modern varieties have also been released in recent years, suggesting that landrace diversity may be under threat of being replaced by modern varieties in the near future, increasing the urgency of this germplasm characterization.

Materials and methods

Plant materials

A total of 192 accessions collected from a range of common bean production ecologies in Ethiopia and Kenya together with four control genotypes for the Andean and Mesoamerican gene pools were grown in a greenhouse at CIAT for DNA extractions and analysis. The East African accessions were selected on the basis of their origin from the CIAT genetic resource unit collection (<http://www.ciat.cgiar.org/urg/beans.htm>) with 99 genotypes from Ethiopia and 89 genotypes from Kenya and the majority being landraces and only a few being commercial cultivars (supplementary information). The control genotypes were selected on the basis of their use in previous studies with microsatellite markers (Blair et al. 2006; Díaz and Blair 2006; Blair et al. 2007). These were: 'Calima' (G4494) a variety from Colombia, and 'Chauca Chuga' (G19833), a landrace from Peru as the Andean control genotypes; as well as DOR364 (or 'Dorado'), a variety from CIAT/El Salvador, and ICA Pijao, a variety from Colombia as Mesoamerican controls. The seeds of these genotypes were also provided by the CIAT germplasm bank.

Morphological measurements

Morphological variables were measured on plants raised in both the greenhouse and the field at CIAT headquarters in Palmira (1,000 m altitude, mean growing temperature 24°C, Mollisol soil), Colombia in 2007. In the greenhouse, plants were grown from August to October 2007 in plastic pots carefully packed with 5 kg of field soil from Palmira mixed with river sand on a 2:1 w/w (weight-by-weight) ratio. A total of four plants were grown for each accession across two pots. For the field evaluation, 20 seeds of each accession were planted in November 2007 in a single 2-m long row with inter- and intra-row spacing of 60 and 10 cm, respectively. The plants were provided with optimum conditions for crop growth and development both under greenhouse and field conditions. The assessed morphological variables were bracteole size (small, medium or large), bracteole shape (cordate, ovate, lanceolate or triangular), outer base of the standard (banner petal), corolla type (smooth or striped), flower color (white, light purple or dark purple), growth habit (determinate bush, indeterminate bush, indeterminate prostrate or indeterminate climbing bean) and stem pigmentation (absent, light red or dark red). All variables were evaluated on four comparably aged plants per accession according to CIAT (1987) and Singh et al. (1991b). Additionally, primary and secondary seed colors and seed size were recorded after harvest for field grown seed.

Microsatellite marker evaluation

For molecular level diversity assessment, total genomic DNA for each accession was isolated from a bulked leaf tissue sample of 1-week old, paper-germinated plants from six randomly selected seeds using a CTAB extraction method as described by Afanador et al. (1993). Since these are CIAT genebank accessions that generally were purified from original landraces with each seed type within a landrace receiving a separate entry, we assumed that we were dealing with mainly single genotypes but that any heterozygosity would be captured within the six-plant sample. The DNA quality was evaluated on 1% agarose gels and quantified with QUANTITY ONE v. 4.0.3 software (Bio-Rad Lab., Hercules, CA). DNA was then diluted to 5 ng/μL for further use in the genotyping experiments. Microsatellite marker evaluation involved a total of 38 fluorescently labeled microsatellites selected to represent both cDNA-based and genomic markers and for high polymorphism information content (Blair et al. 2006, 2009). The genomic microsatellites included BM139, 140, 141, 143, 151, 156, 165, 172, 175, 183, 187, 188A, 188B, 2001, 205; BMd12, 36, AG1 and GATs54, 91 (Gaitán et al. 2002). The gene-based microsatellites included BMd1, 2, 8, 15, 16, 17, 18,

20, 45, 46, 47, 51 (Blair et al. 2003), and PV-ctt001a, PV-ctt001, PV-ag003, PV-ag001, PV-at001, PV-at003 (Yu et al. 2000). PCR amplifications were carried out on a MJ Research Inc. PTC-100 thermo-cycler using 96-wall plates as described by Blair et al. (2006) with a 13-μL final reaction volume that included 3 μL of genomic DNA, 1.5 μL of each primer at concentrations of 0.16 μM, 0.78 μL of Mg buffer at a concentration of 1.5 mM, 0.72 μL of 1× PCR buffer (10 mM pH 7.2 Tris-HCl, 50 mM of KCl), 0.13 μL of dNTP at concentration of 0.2 mM and 0.15 μL of 1.0 unit *Taq* polymerase and 5.22 μL ddH₂O. The PCR products of different size and contrasting fluorescent labels were pooled and diluted with sterile deionized water to equalize signal strength. The DNA fragments from pooled PCR amplifications were then separated by capillary electrophoresis using an ABI3730 DNA analyzer (Applied Biosystems, Foster City, CA). The fragment analysis data from ABI3730 system were analyzed and allele sizes scored with GENEMAPPER version 3.7 software (Applied Biosystems). The observed allele size was then adjusted for the discrete allele size using AlleloBin software (http://test1.icrisat.org/gt-bt/download_allelebin.htm) and allele sizes for the control genotypes (Calima, G19833, DOR364 and ICA Pijao) were confirmed to be of the same sizes as in Blair et al. (2006, 2007) and Díaz and Blair (2006).

Genetic diversity analysis

The pattern of genetic diversity within and among accessions and across the countries of collection (Ethiopia vs. Kenya) was assessed for both morphological and molecular data using several software programs. Morphological marker data from greenhouse and field trials were averaged and subjected to frequency distribution analyses in SAS statistical package version 9.1.3 (SAS Institute 2003). The morphological traits were scored for each genotype based on presence and absence. These data were used to generate a binary matrix of presence and absence which was used for principal coordinate analysis (PCoA) and for creating the matrix of average taxonomic distance (i.e., DIST coefficient in the procedure) between accessions, respectively, in the SIMQUAL and SIMINT subprograms of NTSYS-pc, version 2.10 (Rohlf 2002). Genetic relationships within and among accessions from the two East African countries based on genotypic data were assayed with a neighbor-joining method in DARWIN 5.0 software (Perrier et al. 2003; Perrier and Jacquemoud-Collet 2006). Genetic distance matrices were generated using the Peakall et al. (1995) method of calculating individual by individual genetic distances from co-dominant markers. Accordingly, for each SSR marker, with *i*-th, *j*-th, *k*-th and *l*-th different alleles, a set of squared distances was defined as $d^2(ii, ii) = 0$, $d^2(ij, ij) = 0$, $d^2(ii, ij) = 1$, $d^2(ij, ik) = 1$, $d^2(ij, kl) = 2$, $d^2(ii, jk) = 3$,

and $d^2(ii, jj) = 4$. Genetic distance matrices for each locus were summed across loci assuming statistical independence. The genetic distance values were then subjected to PCoA as implemented by GENALEX version 6.1 software (Peakall and Smouse 2007). Patterns revealed by the first three coordinates of each accession were plotted using the Graph module and the G3D procedure of the software program SAS. Genetic diversity parameters such as number of alleles (N_A), number of effective alleles (N_E), Number of private alleles (N_{PA}), observed heterozygosity (H_O), standardized allelic richness (A_R), gene diversity (G_D), Shannon's information richness (I), fixation index (F), percent polymorphic loci were estimated with FSTAT version 2.9.3 (Goudet 2001) for each pre-determined group based on origin of the accession of collection (by country) and genepool assignment as differentiated by neighbor-joining analysis and PCoA.

Molecular analysis of variance and population structure analysis

Partitioning of total genetic variation into within and among genepool diversity and country of origin was performed with a molecular analysis of variance (AMOVA) procedure in GENALEX. To infer pattern of population structure, both population level and individual-based clustering approaches were employed. Global F_{ST} and pairwise F_{ST} were estimated using Weir and Cockerham's Q (Weir and Cockerham 1984). F_{ST} values and significance of estimates were calculated with FSTAT. Other parameters such as gene flow, Nei's unbiased genetic distance and identity were computed to assess the degree of population differentiation using GENALEX. The Bayesian genotypic clustering method INSTRUCT (Gao et al. 2007) was used to validate population-based approaches and to infer population structure among the genotypes. INSTRUCT is an extended Bayesian clustering approach of STRUCTURE (Pritchard et al. 2000) that absorbs inbreeding or selfing rate for population inference. It quantifies the contribution of two forms of non-random mating: inbreeding (mating among relatives) and population substructure (limited dispersal of gametes) when determining the pattern of existing genetic variation (Gao et al. 2007). INSTRUCT was run for $K = 2$ to $K = 6$ in mode 2 for joint inference of population selfing rate and population sub-structure for five independent chains, each chain with 200,000 iteration steps, 100,000 burn-ins, and a thinning interval of ten steps, assuming different starting points. Graphical representations of population assignments from INSTRUCT were produced from the program DISTRUCT (Rosenberg 2002).

Results

Morphological diversity

Significant variation was observed for most morphological traits measured on the East African accessions with two to four character states found per trait as shown in Table 1. Significant differentiation of accessions from the two countries was observed for bracteole size, growth habit and seed size whereas for base of the standard, bracteole shape, flower color and stem anthocyanin pigmentation the difference was non-significant. In these cases, the majority of the accessions from both Ethiopia and Kenya had smooth outer base of the standard and no stem anthocyanin pigmentation.

A greater proportion of the accessions from Ethiopia had larger bracteole size, cordate or ovate bracteole shape, white flower color and smaller seed size, characteristics typical of the Mesoamerican genepool; and a larger proportion of accessions from Kenya had predominantly medium to large bracteole size, lanceolate or triangular bracteole shape and medium to larger seed size typical of the Andean genepool. The accessions from both countries were showing a range of growth habits and seed colors; however, type-III growth habit was prevalent in Ethiopian accessions and type I and II growth habit was prevalent in Kenyan accessions.

The dominant primary seed colors throughout the Ethiopian accessions were white, red and tan/brown whereas in Kenyan accessions purple, cream, yellow and red-seeded genotypes were common. A majority of the accessions in both countries were of a single primary color and had no secondary seed color; however, among those with secondary seed colors, red and cream mottled seed types were more prevalent in Ethiopia and Kenya, respectively.

Analysis of the morphological variables showed grouping of Andean and Mesoamerican genotypes combined with probable introgression between the genepools as shown by the PCoA in Fig. 1. In this graph, the first and second dimensions (Dim-1 and Dim-2) explained 21.0 and 10.6% of the total variation in the data set, respectively. Together, the first two-dimensions explained 31.69% of the total variation; and overall the PCoA analysis separated the Mesoamerican control genotypes from Andean control genotypes with concomitant clustering of some accessions into their respective gene pools. Many accessions, meanwhile, occupied intermediate positions between the two genepools and the control genotypes for the two genepools, probably due to introgression and/or shared morphological markers such as seed color and growth habit of the accessions in them.

Table 1 Frequency distribution for morphological traits evaluated for East African accessions in relation to genepool control genotypes

Trait ^a	Origin		χ^2 -test	Controls ($N = 4$) ^b
	Ethiopian ($N = 99$)	Kenyan ($N = 89$)		
Base of standard				
Smooth	72	67	0.1586ns	2
Stripped	27	22		2
Bracteole size				
Small	20	17	18.96**	2
Medium	28	51		0
Large	51	21		2
Bracteole shape				
Cordate	27	13	5.128ns	1
Ovate	23	24		1
Lanceolate	33	38		0
Triangular	15	11		2
Flower color				
White	53	38	5.696ns	3
Light purple	23	35		0
Dark purple	23	16		1
Growth habit ^a				
I	15	29	12.95**	0
II	23	23		3
III	41	18		1
IV	20	19		0
Stem anthocyanin pigmentation				
Absent	74	75	3.870ns	2
Light red	15	11		1
Dark red	10	3		1
Seed size				
Small	56	20	27.10**	2
Medium	33	39		0
Large	10	30		2

Number (N) of each set of accessions is given when listing country of origin

ns non-significant

** Highly significant ($P < 0.001$)

^a Morphological traits evaluated in greenhouse and verified in field plantings at CIAT in Palmira, Colombia during summer 2007. Conventions are: I determinate bush, II indeterminate bush, III indeterminate prostrate, IV indeterminate climbing beans; seed size only measured from field trial

^b Control genotypes were 'Calima' (G4494) a variety from Colombia and 'Chauca Chuga' (G19833) a landrace from Peru as Andean control genotypes and DOR364 (or 'Dorado') a variety from CIAT/El Salvador and ICA Pijao a variety from Colombia as Mesoamerican control

Genetic associations among accessions

Genotyping results with the fluorescent microsatellite markers were also used to cluster the accessions, and genetic associations among accessions from Ethiopia and Kenya with respect to Andean and Mesoamerican control genotypes as shown in Figs. 2 and 3. In these graphs, distinct clusters were apparent with the SSR markers unambiguously assigning accessions to the Andean and Mesoamerican genepools both with neighbor-joining dendrograms (Fig. 2) and with the 3D plot of the PCoA based on

pairwise genetic distances (Fig. 3). Within each country, accessions from the same collection site were often in different clusters and likewise accessions from different collection sites were clustered together (Fig. 2I, III) indicating the possibility of gene flow between sites and regions within Ethiopia and within Kenya. When comparing across countries in the overall analysis (Fig. 2II), accessions from the same country of origin tended to cluster together especially with the Andean genotypes indicating distinct germplasm at the national level and perhaps some cross-border gene flow between the countries.

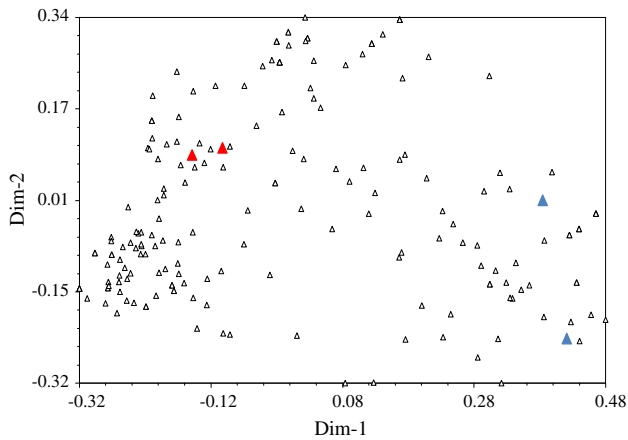


Fig. 1 Principal coordinates analysis of the 192 Ethiopian and Kenyan accessions based on nine morphological traits. *Filled triangles to the left* indicate placement of Andean control genotypes and *filled triangles to the right* indicate Mesoamerican control genotypes

Results of the PCoA were in agreement with those of the neighbor-joining dendrograms, with two major groups detected: one clearly representing the Andean gene pool and the other the Mesoamerican gene pool. The division of the accessions into two major groups showed that there was correspondence between the grouping of East African bean landraces and the respective gene pools in the primary centers of diversity. However, the further differentiation into

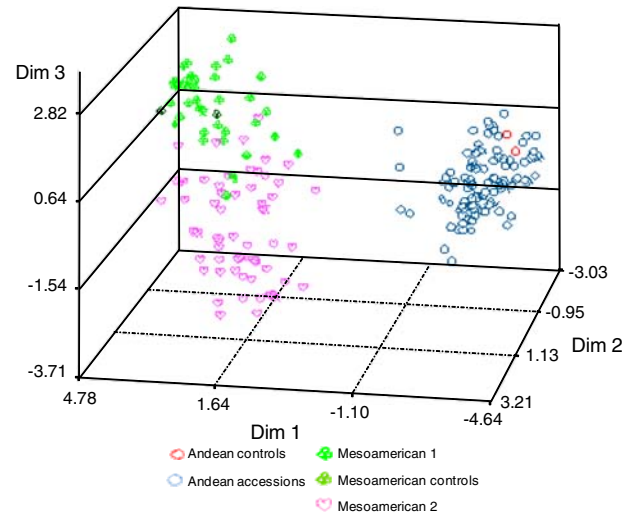


Fig. 3 Principal coordinate analysis based on microsatellite markers showing spatial distribution of Ethiopian and Kenyan accessions compared to Andean and Mesoamerican control genotypes. Each dimension explains 56.37% (Dim1), 12.01% (Dim2) and 11.30% (Dim3) of variation. The three dimensions together explained 79.68% of total variation present in the data set

recognized bean races belonging to these gene pools was not apparent, although some sub-grouping was observed in the analysis. The overall variation explained by the principle coordinate analysis was 79.7% with dimensions 1, 2 and 3 explaining 56.4, 12.0 and 11.3%, respectively.

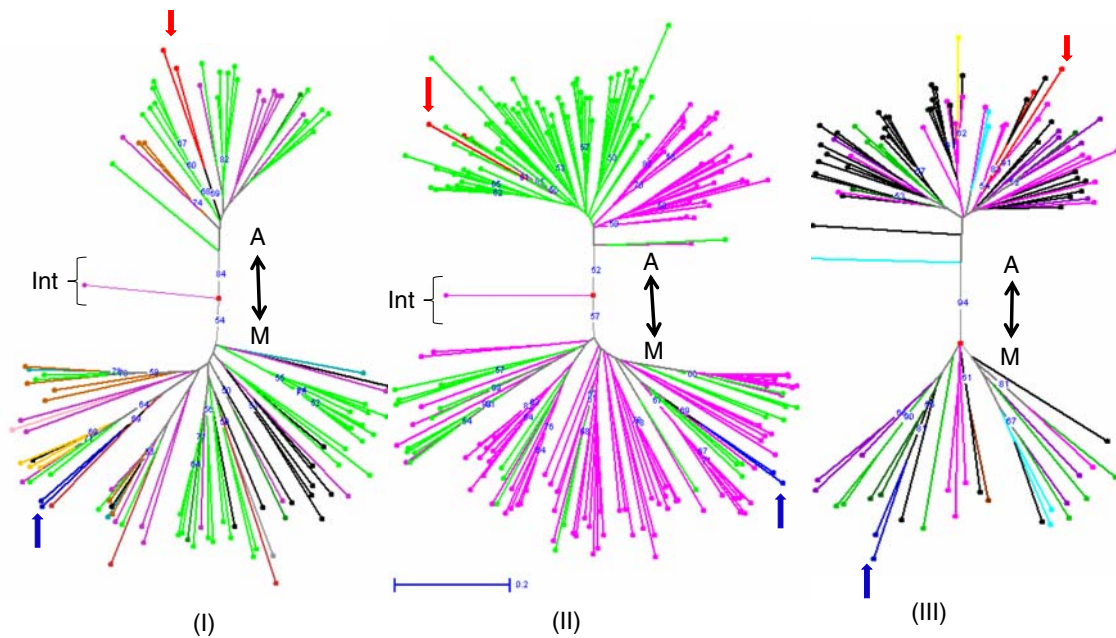


Fig. 2 Neighbor-joining dendrograms depicting genetic relationship between common bean accessions from Kenya and Ethiopia with respect to Andean and Mesoamerican control genotypes. **I** Ethiopian accessions, **II** global accessions (full set of the study materials) and **III** Kenyan accession. *Different line shading* represent different collection

sites within each of the countries (**I**, **III**) and country of origin (**II**). *Downward facing arrows* indicate Andean controls and *upward facing arrows* indicate Mesoamerican controls. *A* Andean, *M* Mesoamerican, *Int* introgression as explained in the text. *Numbers along branches* indicate bootstrap support (shown only for values greater than 50)

Table 2 Mean SSR diversity for 38 microsatellite loci in Ethiopian, Kenyan and genepool control genotypes

	No.	N_A	N_E	N_{PA}	A_R	G_D	H_O	I	F	t	P (%)
Total Andean	95	7.26	2.75	2.47	7.14	0.47	0.11	1.02	0.74	0.15	100.0
Total Mesoamerican	96	7.66	3.28	2.97	7.59	0.60	0.14	1.26	0.76	0.14	100.0
Ethiopian	99	7.68	3.72	1.95	7.51	0.64	0.14	1.36	0.79	0.12	100.0
<i>Andean Eth.</i>	28	3.95	2.23	0.26	3.79	0.44	0.11	0.82	0.76	0.14	86.8
<i>Mesoamerican Eth.</i>	70	6.45	3.09	1.05	5.15	0.59	0.15	1.18	0.73	0.16	100.0
Kenyan	89	8.16	3.39	2.42	8.09	0.59	0.12	1.28	0.81	0.11	100.0
<i>Andean Kya.</i>	65	5.97	2.56	1.29	4.70	0.45	0.12	0.95	0.71	0.17	94.7
<i>Mesoamerican Kya.</i>	24	5.40	3.28	0.58	5.28	0.59	0.13	1.17	0.77	0.13	94.7
Total Eth-Kenya	188	10.11	4.00	7.68	6.56	0.65	0.13	1.45	0.80	0.11	100.0
Andean-Meso checks	4	2.55	2.34	0.13	2.46	0.50	0.04	0.81	0.94	0.03	89.5
Grand total	192	10.24	4.03	–	10.20	0.65	0.13	1.45	0.81	0.11	100.0

N number of genotypes, N_A number of different alleles, N_E effective number of alleles, N_{PA} number of private alleles, A_R allele richness, G_D gene diversity according to Nei (1978), H_O observed heterozygosity, I Shannon's information index, F fixation index, $t = (1 - F)/(1 + F)$ out-crossing rate, P (%) percent polymorphic loci. *Eth* Ethiopian genotypes, *Kya* Kenyan genotypes

Genetic diversity within and among accessions and country of collection

All of the microsatellite markers used in this study were polymorphic. The proportions of polymorphic loci were 89.5% in the control genotypes and 100% in both Ethiopian and Kenyan accessions (Table 2). A total of 389 alleles were detected among the 192 bean accessions with an average of 10.24 alleles per marker. The number of alleles per markers ranged from 2 in BMd46 to 35 in Pv-at001, with the mean number of effective alleles per locus not significantly different among the two East African collections but slightly lower in Kenyan (3.39) than in Ethiopian (3.72) accessions. Meanwhile, the mean number of private alleles per population was slightly higher for Kenyan (2.42) versus Ethiopian (1.25) accessions although allele richness was not significantly different between countries of origin and genepools.

AMOVA results showed that 66% of allelic diversity was attributed to individuals within genepool ($P < 0.001$) while only 34% was distributed among genepools. No significant variation for molecular diversity was observed between countries of collection denoting shared alleles among them. However, Ethiopian accessions had slightly higher level of gene diversity compared to Kenyan accessions. Within the country of origin and between the genepools, accessions within the Mesoamerican group of East African accessions had slightly higher gene diversity than those within the Andean group. Similarly, Shannon's information index was slightly higher for Ethiopian than for Kenyan accessions and for Mesoamerican genepool accessions compared to Andean genepool representatives. The observed heterozygosity and probable out-crossing values

were low for all the study materials reflecting the inbreeding nature of the common bean crop. However, the heterozygosity and out-crossing values were slightly higher for the East African accessions (0.11–0.15) compared to the control genotypes (0.04) which might be explained by these genebank accessions resulting from the collection of varietal mixtures which are common in many farmer fields in the region. Higher observed heterozygosity was observed among Ethiopian accession than among Kenyan accessions overall and for the Mesoamerican genepool genotypes in Ethiopia and Andean genepool genotypes in Kenya.

Population differentiation and structure

Genetic differentiation in the East African bean landraces and cultivars was also analyzed with POPGENE (Yeh et al. 1997) and F_{ST} values among pairs of populations were found to range from 0.037 to 0.632 with an overall average of 0.273 (Table 3). Population differentiation was higher between genepools ($F_{ST} = 0.189$, $P < 0.001$) than between countries of origin ($F_{ST} = 0.06$, $P < 0.001$). However, for the comparison between the countries of origin, Andean genepool accessions were more highly differentiated ($F_{ST} = 0.331$, $P < 0.001$) than Mesoamerican genepool accessions ($F_{ST} = 0.04$, $P < 0.001$). Correspondingly, some level of gene flow ($Nm = 3.927$) existed between the two neighboring East African countries, which was higher for Mesoamerican representatives ($Nm = 6.421$) than for Andean representatives ($Nm = 3.940$). Average Nei's unbiased genetic distance was high between genepools (0.665) but low between countries of origin (0.195). Within genepool, the Mesoamerican representatives presented lower genetic distances than the Andean genepool representatives

Table 3 Pairwise genetic differentiation, gene flow, unbiased Nei's genetic distance and identity among and between gene pools and countries of origin in East African landraces and cultivars

Gene pool ^a	F_{ST}						GD					
	AC	MC	AE	ME	AK	MK	AC	MC	AE	ME	AK	MK
Andean control ($N = 2$)	–	0.632	0.194	0.341	0.168	0.368	–	1.407	0.259	1.034	0.233	1.099
Mesoamerican control ($N = 2$)	0.141	–	0.478	0.217	0.472	0.205	0.245	–	1.590	0.398	1.453	0.309
Andean Ethiopia ($N = 28$)	1.040	0.274	–	0.202	0.060	0.229	0.772	0.204	–	0.666	0.108	0.811
Mesoamerican Ethiopia ($N = 70$)	0.483	0.902	0.985	–	0.208	0.037	0.355	0.672	0.514	–	0.699	0.088
Andean Kenya ($N = 65$)	1.235	0.279	3.940	0.954	–	0.217	0.792	0.234	0.898	0.497	–	0.724
Mesoamerican Kenya ($N = 24$)	0.430	0.972	0.842	6.421	0.902	–	0.333	0.734	0.444	0.916	0.485	–
	Nm						GI					

AC Andean control, MC Mesoamerican control, AE Andean from Ethiopia, ME Mesoamerican from Ethiopia, AK Andean from Kenya, MK Mesoamerican from Kenya. Genetic differentiation (F_{ST}) and unbiased Nei's genetic distance (GD) in upper diagonals in right and left panels of the table. Gene flow (Nm) and Nei's unbiased genetic identity (GI) in lower diagonals in left and right panels, respectively

^a Introgression was excluded and number (N) of accessions in each germplasm group was indicated

in each country. Genetic identity was fairly high between the two countries (0.823); however, it was low between gene pools (0.204–0.507) and intermediate within gene pools (0.673–0.916).

Population structure analysis with INSTRUCT confirmed the existence of the two gene pools for the East African Highland common bean accessions (Fig. 4). The analysis for $K = 2$ populations showed individual genotypes from the two countries distributed between the two gene pools which was congruent with neighbor-joining and PCoA that clearly separated the Mesoamerican and Andean gene pools. At $K = 3$, the Mesoamerican gene pool genotypes further separated into two sub-groups with a low level of admixture, while the Andean gene pool genotypes did not show any separation. At $K = 4$, the Mesoamerican accessions further subdivided into three groups but no meaningful interpretation of population structure could be made. At $K = 5$, the Andean group separated according to country of origin with very little admixture between Ethiopian and Kenyan Andeans, supporting earlier analysis that depicted distinct germplasm at this national level separation. At $K = 6$, the Andean groups further differentiated into three groups, principally in Kenya where two subgroups were highly admixed; while the Mesoamerican gene pool maintained the same sub-grouping as observed at $K = 4$. INSTRUCT software predicated $K = 6$ as the optimum population structure in the study material, therefore no further population subdivisions were modeled. The morphological characteristics predominant in each sub-population at $K = 6$ are given in Table 4.

Discussion

The level of polymorphism in landraces and cultivars from Ethiopia and Kenya was found to be considerable, espe-

cially with microsatellite marker analysis. Our result identified common beans from this region as distinguishable into both Andean and Mesoamerican gene pools as described by various authors (Gepts et al. 1986; Singh et al. 1991a, b, c; Becerra and Gepts 1994; Islam et al. 2002; Blair et al. 2006, 2009). The conservation of the gene pool separation typical of the primary centers of diversity has been observed before for bean in southern Africa (Martin and Adams 1987) and is also a hallmark of bean diversity in other secondary centers of diversity outside of the Americas, such as Southwest Europe (Rodiño et al. 2006) and China (Zhang et al. 2008).

The separation of East African bean landraces into the two recognized gene pools was stronger with SSR markers than with morphological markers indicating the success of this marker type in detecting gene pools in common beans. Similar results were obtained in previous studies of accessions from primary and secondary centers of diversity analyzed with SSRs (Blair et al. 2006, 2009; Zhang et al. 2008). Despite the limitations of morphological analysis, the similarity distance matrices obtained using SSR markers was significantly correlated with that obtained with morphological markers ($r = 0.49731$, $P < 0.001$) based on the MXCOMP procedure of NTSYS-pc and testing using the normalized Mantel Z-statistics (Rohlf 2002). However, this positive and significant correlation might be misleading as distance matrices of morphological markers did not produce completely congruent patterns of population structure with that of SSR-based genetic distance matrices, which were of better resolution, indicating an under-estimate of genetic relationships with morphological markers. Discrepancy of clustering based on morphological and molecular markers has been attributed to hybridization or mutation that leads to divergent morphological or molecular profiles (Singh et al. 1991b). In addition, lower heritability or

Fig. 4 Population structure for 192 common bean accessions from the East African Highlands compared to Andean and Mesoamerican control genotypes at $K = 2$ to $K = 6$. Predetermined group names indicated below figure are *AC* Andean control genotypes, *AE* Andean genotypes from Ethiopia, *AK* Andean genotypes from Kenya, *MC* Mesoamerican control genotypes, *ME* Mesoamerican genotypes from Ethiopia and *MK* Mesoamerican genotypes from Kenya

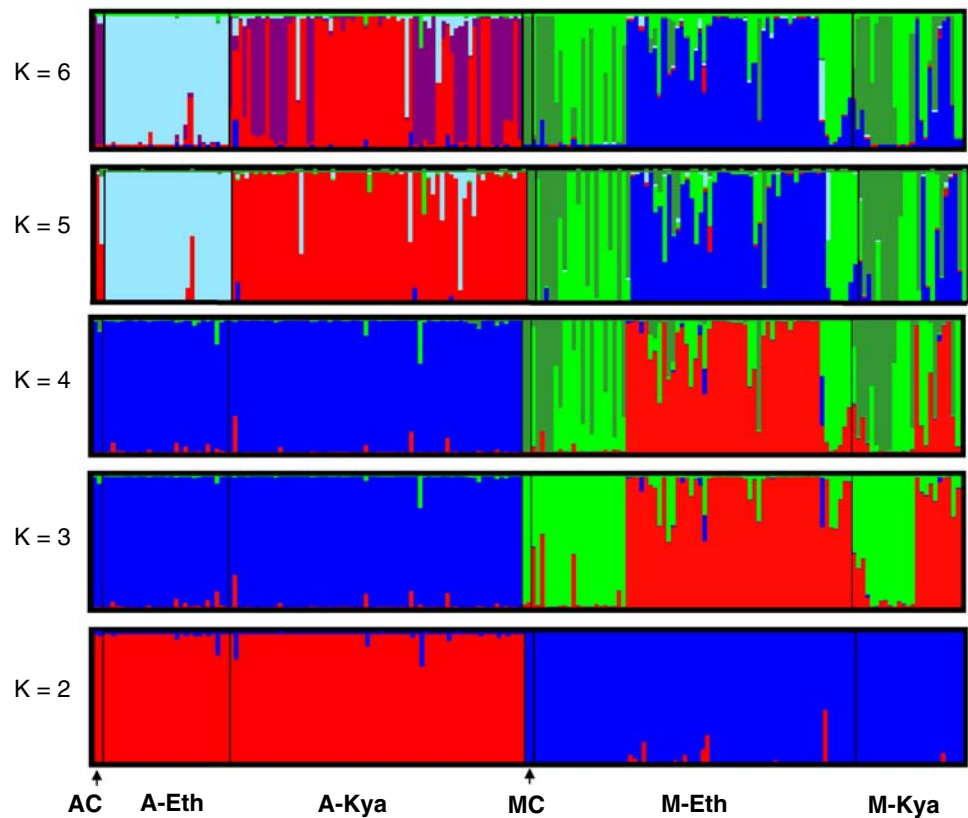


Table 4 Some characteristics of sub-populations identified at $K = 6$ population structure level for the East African common bean landraces

Population	Origin	N ^a	Growth habit ^b				Seed size			Bracteole size			Bracteole shape			
			I	II	III	IV	S	M	L	S	M	L	C	O	L	T
A1	Kenya	36	16	11	6	3	3	16	17	5	29	2	4	7	18	6
A2	Kenya	26	12	6	5	3		17	11	7	19			8	12	4
A3	Ethiopia	28	15	7	3	3	3	16	9	9	16	3	2	2	14	9
	Kenya	3	1	2				1	2	2	2				2	1
M1	Ethiopia	11		6	5		9	2				11	8	2	1	
	Kenya	10		3	5	2	9		1	2		8	3	5	2	
M2	Ethiopia	40		7	18	15	25	14	1	10	12	18	5	13	17	4
	Kenya	5			1	4	2	2	1	2	1	2	1	1	3	
M3	Ethiopia	19		3	14	2	17	2				19	12	5	1	1
	Kenya	9		1	1	7	6	3				9	5	3	1	

S small, M medium, L large, C cordate, O ovate, L lanceolate, T triangular

^a Number (N) of genotypes in each sub-population with control genotypes and introgressed individuals not shown

^b Growth habit as described in Table 1

similarity of character states can also lead to poor separation based on morphological characteristics. Therefore, the use of informative molecular markers as a prior clustering criterion to improve the resolution power of morphological markers in common bean germplasm characterization is valid as was suggested by Singh et al. (1991a).

Microsatellite analysis showed generally low levels of introgression between the gene pools compared to morpho-

logical analysis where character states were shared between the accessions belonging to each of the gene pools. PCoA of morphological traits showed many intermediate genotypes while the same analysis for the SSR markers showed very few genotypes that were intermediate between the Andean and Mesoamerican clusters. Furthermore, in the neighbor-joining dendrograms only one genotype from Ethiopia (G18863) was intermediate between the gene pools. In line

with this, low gene flow and high genetic differentiation between genepools were observed in the present analysis. However, within genepool, gene flow was higher both between countries and within each country especially for Mesoamerican representatives.

These results suggest that the genetic divergence in East African bean landraces could be due to the original differences in introduced germplasm from the primary centers of origin combined with spontaneous out-crossing in farmer field and further farmer selection for adaptation to production niches and uses. For example, early flowering was common in many of the Andean accessions compared to the Mesoamerican accession in both greenhouse and field evaluation (data not shown). This lack of flowering synchronization could make inter-genepool hybridization a less likely phenomenon for the East African highland germplasm, even if varietal mixtures of both Andean and Mesoamerican phenotypes are a common farming practice in many parts of the region.

Another conclusion from the molecular analysis was that there was low gene flow between the two countries compared to gene flow within each country. This led to recognizably distinct germplasm at the national level especially for Andean genepool accessions. The molecular diversity was also reflected in the diversity for seed color, size and shape (data not shown) and the fact that the countries share very few seed types in the Andean genepool and only a few seed types in the Mesoamerican genepool. This might highlight different informal or formal institutional introductions and weak trans-national bean seed exchange or social and commercial networks especially for landrace varieties. One exception to this may be the case of small red-seeded beans in cross-border grain trade from southern Ethiopia to certain parts of Kenya. Apart from gene flow, an additional reason for the divergence in germplasm between Ethiopia and Kenya may be the existence of different farmers' selection preferences in each country in accordance with ecological adaptation, cooking value and market orientation. In fact, in Ethiopia, the small white and small red color classes are the preferred bean seed classes for export and local consumption, respectively; whereas in Kenya, large-seeded, red mottled seed types have high market preference (Wortmann et al. 1998).

Despite the distinct germplasm at the national level and at the genepool level, further race structure was not apparent in the East African common bean landraces. While all the population-genetic analyses employed in this study showed good congruence in the division of the landraces into two genepools, the number of sub-groupings varied depending on the analysis conducted. For example, INSTRUCT analysis suggested six sub-populations within the landraces with three of these corresponding to Andean genepool groupings and the other three to Mesoamerican

genepool groupings; however, these six populations were not evident in the PCoA for the SSR markers where only three groupings, two in the Mesoamerican genepool and one in Andean genepool were found. Some of the genotypes clustered together in the neighbor-joining dendrograms were assigned in different groups with INSTRUCT.

Andean diversity was found to be relatively high but difficult to subdivide with the Andean genepool control genotypes, Calima and G19833, representing Nueva Granada and Peru races, respectively (Blair et al. 2007) clustered together in both principal coordinate and population structure analyses. The closer placement of the two Andean control genotypes and concomitant overlap with other accessions from Ethiopia and Kenya might indicate the representation of the East African Andean genotypes, especially those from Kenya as part of a race Nueva Granada/race Peru complex. This was evident in the distinction of the Andean groups in Table 4 where one group (A2) consisted of medium to large-seeded genotypes with small to medium sized bracteoles and a range of growth habits while another group (A1) was made up of only Kenyan genotypes with medium to large seeded and cylindrical or kidney seed shape genotypes having small to medium bracteole size, and ovate, lanceolate or triangular bracteole shape that corresponds to race Nueva Granada descriptors based on Singh et al. (1991b). The final Andean group (A3) consisted of Andean genotypes that were mostly from Ethiopia that had small to medium bracteole size, lanceolate or triangular bracteole shape, medium to large seed size, predominantly type I or II growth habit, and oval or rounded, cream spotted or tan seed shape.

Within the main Mesoamerican genepool grouping, most accessions clustered into three subgroups: M1 group included the Mesoamerican control genotypes DOR364 and ICA Pijao which are designated as race Mesoamerica in Díaz and Blair (2006) indicating the probable representation of East African Mesoamerican landraces by this race. This group possessed smaller seed size, larger cordate bracteole size and type II or III growth habit that corresponded to the race Mesoamerica description of Singh et al. (1991b). Another grouping under this genepool, M2, represented by 19 landraces from Ethiopia and nine landraces from Kenya had all large bracteole size and small to medium red, white or black seed with considerable admixture with the M1 populations suggesting that this represents another sub-grouping of race Mesoamerica. All small red-seeded genotypes including the dominant 'Red Wolayta' from Ethiopia were included under this group indicating that small red-seeded beans in Ethiopia have a narrow genetic background as compared to white and black beans that were distributed in the two other Mesoamerican sub-populations. Meanwhile, the majority of the small white-seeded genotypes from Ethiopia were represented in the third Mesoamerican

sub-population, M3, which included 40 accessions from Ethiopia and five from Kenya almost all with small to medium seed size, indeterminate prostrate (III) or indeterminate climbing (IV) growth habit characteristic of the Durango–Jalisco race complex (Singh et al. 1991b; Díaz and Blair 2006). Hence, we suspect the apparent representation of this race complex in East Africa but further analysis would be needed to confirm this. A comparison of the East African beans to Latin American germplasm from the Caribbean, Central America, Mexico or Brazil as likely sources of germplasm sent to East Africa would also be valuable as would a comparison to European germplasm from the ex-Colonial countries that probably served as transit points for this diversity.

In this regard, many of the small red beans preferred in Ethiopia are typical of Central America (Singh et al. 1991a) and could have arrived through trade via Spain. Meanwhile, red mottled beans preferred in Kenya are typical of the Caribbean and could have followed the same route. Durán et al. (2005) characterized a large set of landraces of this seed type and found separation of the gene pools based on morphological characteristics and RAPD markers with most of the large-seeded genotypes coming from the Eastern Caribbean. Rodiño et al. (2003, 2006) found small white beans in the Iberian Peninsula, and both studies observed inter-gene pool introgression that may have produced new seed types.

In conclusion, our study found that population structure for the East African common bean landraces was based mainly on gene pool origin and that introgression or gene flow was moderate. Given that beans in this region are often cultivated in marginal, risk-prone production ecologies (Wortmann et al. 1998), it will be interesting to correlate genetic diversity with drought tolerance and adaptation potential in future association mapping work. The results presented here also pave the way for rational use of East African germplasm and strategic crossing plans that could be used to identify transgressive segregation based on distinct germplasm at the national or regional level. In this regard, further phenotyping could identify the genotypes that would be the most valuable gene sources in future breeding programs in the region. Finally, the results also suggest that a considerable amount of common bean genetic diversity is present in East Africa motivating renewed conservation efforts for the region.

Acknowledgments We are grateful to the Genetic Resource Unit and Bean Project of CIAT as well as the Institute for Genomic Diversity at Cornell University for germplasm and assistance rendered during phenotyping and genotyping. The Generation challenge program (GCP) is gratefully acknowledged for funding associated with this study. We also thank the Cornell Computational Biology Service Unit for assistance in data analysis.

References

- Afanador L, Hadley S, Kelly JD (1993) Adoption of a mini-prep DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L.). *Bean Improv Coop* 36:10–11
- Allen DJ, Edje OT (1990) Common bean in Africa farming system. In: Smithson JB (ed) 1990 Progress in improvement of common bean in Eastern and Southern Africa. *Bean Research*, vol 5. CIAT Africa Workshop Series No 20, p 32
- Becerra VL, Gepts P (1994) RFLP diversity of common bean (*Phaseolus vulgaris*) in its centers of origin. *Genome* 37:256–263
- Beebe S, Skroch P, Tohme J, Duque MC, Pedraza F, Nienhuis J (2000) Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. *Crop Sci* 40:264–273
- Beebe S, Rengifo J, Gaitan-Solis E, Duque MC, Tohme J (2001) Diversity and origin of Andean landraces of common bean. *Crop Sci* 41:854–862
- Blair MW, Pedraza F, Buendía H, Gaitán E, Beebe S, 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 S (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 113:100–109
- Blair MW, Díaz JM, Hidalgo R, Díaz LM (2007) Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 116:29–43
- Blair MW, Díaz LM, Buendía 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*. doi:10.1007/s00122-009-1064-8
- Broughton WJ, Hernandez G, Blair MW, Beebe SE, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.)—model food legumes. *Plant Soil* 252:55–128
- CIAT (1987) Standard system for the evaluation of bean germplasm. Van Schoonhoven A, Pastor C (Compilers). CIAT, Cali
- CIAT (2005) Utilization of bean genetic diversity in Africa. Highlights of CIAT in Africa, no. 21. CIAT, Cali. <http://www.ciat.cgiar.org>
- 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
- Durán LA, Blair MW, Giraldo MC, Macchiavelli R, Prophete E, Nin JC, Beaver JS (2005) Morphological and molecular characterization of common bean landraces and cultivars from the Caribbean. *Crop Sci* 45:1320–1328
- Gaitán E, Duque MC, Edwards K, Tohme J (2002) Microsatellite repeats in common bean (*Phaseolus vulgaris* L.): isolation, characterization, and cross-species amplification in *Phaseolus* spp. *Crop Sci* 42:2128–2136
- Gao H, Williamson S, Bustamante CD (2007) An MCMC approach for joint inference of population structure and inbreeding rates from multi-locus genotype Data. *Genetics*. <http://cbsuapps.tc.cornell.edu/InStruct.aspx> (online)
- Gentry HS (1969) Origin of the common bean. *Phaseolus vulgaris*. *Econ Bot* 23:55–69
- Gepts P, Debouck D (1991) Origin, domestication and evolution of the common beans (*Phaseolus vulgaris* L.). In: Schoonhoven A, Voysest O (eds) Common beans: research for crop improvement. CAB International, Wallingford, pp 7–54
- Gepts P, Osborn T, Rashka K, Bliss F (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

- Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (Version 2.9.3). <http://www.unil.ch/pop-gen/softwares/fstat.html>
- Greenway P (1945) The origin of some East Africa food plants. *East Afr Agric For J* 10:177–180
- Hillocks RJ, Madata CS, Chirwa R, Minja EM, Msolla S (2006) *Phaseolus* bean improvement in Tanzania, 1959–2005. *Euphytica*. doi:10.1007/s10681-006-9112-9
- Islam FM, Basford KE, Reden RJ, Gonzalez AV, Kroonenberg PM, Beebe SE (2002) Genetic variability in cultivated common bean beyond the two major gene pools. *Genet Resour Crop Evol* 49:271–283
- Islam FM, Beebe S, Muñoz M, Tohme J, Redden RJ, Basford KE (2004) Using molecular markers to assess the effect of introgression on quantitative attributes of common bean in the Andean gene pool. *Theor Appl Genet* 108:243–252
- Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. *Theor Appl Genet* 78:809–817
- Martin GB, Adams MW (1987) Landraces of *Phaseolus vulgaris* (Fabaceae) in northern Malawi. II. Generation and maintenance of variability. *Econ Bot* 41:204–215
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590
- Peakall R, Smouse PE (2007) GENALEX 6.1: genetic analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra. Available via <http://www.anu.edu.au/BoZo/GenALEX>
- Peakall R, Smouse PE, Huff DR (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of buffalograss *Buchloe dactyloides*. *Mol Ecol* 4:135–147
- Perrier X, Jacquemoud-Collet JP (2006). DARwin software. Available from <http://www.darwin.cirad.fr/darwin>
- Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Enfield Science Publishers, Montpellier, pp 43–76
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rodiño AP, Santalla M, De Ron AM, Singh SP (2003) A core collection of common bean from the Iberian Peninsula. *Euphytica* 131:165–175
- Rodiño AP, Santalla M, González MA, de Ron AM, Singh SP (2006) Novel genetic variation in common bean from the Iberian Peninsula. *Crop Sci* 46:2540–2546
- Rohlf F (2002) NTSYS pc. Numerical Taxonomy System Exeter Publishing, Setauket
- Rosenberg NA (2002) Distruct: a program for the graphical display of structure results. Available from <http://www.cmb.usc.edu/~noahr/distruct>
- SAS Institute (2003) SAS/STAT Users guide 9.1.3. SAS Institute, Cary
- Singh SP, Gepts P, Debouck DG (1991a) Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ Bot* 45:379–396
- Singh SP, Molina A, Urrea C, Gepts P (1991b) Genetic diversity in cultivated *Phaseolus vulgaris*. II. Marker-based analysis of morphological and agronomic traits. *Crop Sci* 31:23–29
- Singh SP, Nodari R, Gepts P (1991c) Genetic diversity in cultivated common bean: I. Allozymes. *Crop Sci* 31:19–23
- Sperling L (2001) The effect of the civil war on Rwanda's bean seed systems and unusual bean diversity. *Biodivers Conserv* 10:989–1009
- Van Rheenen HA (1979) Diversity of food beans in Kenya. *Econ Bot* 33(4):448–454
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Wortmann CS, Kirkby RA, Eledu CA, Allen DJ (1998) Atlas of common bean (*Phaseolus vulgaris* L.) production in Africa. CIAT Pan-African Bean Research Alliance, vol 133
- Yeh FY, Boyle R, Ye T, Mao Z (1997) POPGENE, the user-friendly shareware for population genetic analysis, version 1.31. Molecular Biology and Biotechnology Centre, University of Alberta, Alberta
- Yu K, Park J, Poysa V, Gepts P (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *J Hered* 91:429–434
- Zhang X, Blair MW, Wang S (2008) Genetic diversity of Chinese common bean (*Phaseolus vulgaris* L.) landraces assessed with simple sequence repeat (SSR) markers. *Theor Appl Genet* 117:629–640