



Genetic clustering, and diversity of African panel of released common bean genotypes and breeding lines

Winyfred Amongi · Stanley Tamusange Nkalubo · Mildred Ochwo-Ssemakula · Arfang Badji · Isaac Onziga Dramadri · Thomas Lapaka Odongo · Ephraim Nuwamanya · Phineas Tukamuhabwe · Paulo Izquierdo · Karen Cichy · James Kelly · Clare Mukankusi

Received: 18 July 2022 / Accepted: 19 February 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract Common bean is a grain legume of global importance especially for proteins and micronutrients. The crop is a staple food in sub-Saharan Africa, where it has gained importance in iron biofortification for people prone to anemia. However, biotic and abiotic constraints, long cooking time, and high phytic acid and polyphenols both of which affect bioavailable iron, hinder the production and health benefits. To inform breeding decisions, the study determined genetic diversity and population structure within 725 breeding lines, varieties, or landraces mostly from Uganda and South America. Genotyping by sequencing and diversity array technology (DartSeq) were used to generate single nucleotide polymorphic markers on Set1 (427) and Set2 (298) germplasm, respectively. The germplasm were grouped into Andean and Mesoamerican gene pools, with the latter as the

larger subpopulation. Analysis of molecular variance revealed 46% (Set1) and 50% (Set2) of genetic variation among the subpopulations, with fixation indices (F_{ST}) of 0.54 (Set1) and 0.71 (Set2) among Andean and Mesoamerican beans, respectively. The overall germplasm's gene diversities were 0.206 (Set1) and 0.332 (Set2). Admixtures were the most diverse (0.193) in both sets of germplasm. The germplasm exhibited high genetic diversity and as a result they have a high potential for use in plant breeding. Inter-gene pool crosses within and across market classes are possible and considering both approaches is expected to increase diversity to realize genetic gain. The structure and polymorphic information generated provided useful perspectives for genomic breeding and genome-wide association study using the population.

W. Amongi · M. Ochwo-Ssemakula · I. O. Dramadri · T. L. Odongo · E. Nuwamanya · P. Tukamuhabwe
Department of Agricultural Production, College of Agricultural and Environmental Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda

W. Amongi (✉) · C. Mukankusi
Alliance of Bioversity and CIAT, National Agricultural Research Laboratories Kawanda, P.O. Box 6247, Kampala, Uganda
e-mail: w.amongi@cgiar.org

S. T. Nkalubo · A. Badji
National Crops Resources Research Institute (NaCRRI-NARO), P.O. Box 7084, Kampala, Uganda

A. Badji · I. O. Dramadri
Makerere University Regional Center for Crop Improvement (MaRCCI), College of Agriculture and Environmental Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda

P. Izquierdo · K. Cichy · J. Kelly
Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI, USA

K. Cichy
USDA-ARS, Sugarbeet and Bean Research Unit, East Lansing, MI, USA

Keywords Common bean · Genetic diversity · Structure · Subpopulations · Principal component analysis

Introduction

Common bean (*Phaseolus vulgaris* L. $2n = 2x = 22$) is an important grain legume crop of global importance, especially in sub-Saharan Africa. The crop is rich in calories, proteins, and micronutrients, especially iron, zinc, thiamin, and folic acid (Myers and Kmiecik 2017). Over 400 million people in Africa directly consume beans (CIAT 2018). However, due to several biotic and abiotic production challenges, on farm yields of 394–1589 kg/ha in most African countries remain below the potential yield of 1500–2000 kg/ha of bush beans (CIAT 2008; Mukankusi et al. 2019).

Cultivated beans are from two distinct gene pools: the Mesoamerican and the Andean pools. The Mesoamerican gene pool is divided into three races: the medium-size seeded prostrate bush, “Durango,” from dry highland Mexico; the climbing bean, “Jalisco,” from the moist highlands of central Mexico; and the small-seeded bush types, Mesoamerica, from the lowland Central America and Mexico (Singh, et al. 1991). There are also three races of highland climbing beans from Peru, mid-altitude bush from Nueva Granada and prostrate bush or weak climbers from Chile in the Andean gene pool (Singh, et al. 1991). The weight of 100 seeds for Andean beans ranges from 35.0–50.0 g, whereas those for Mesoamerican and Durango types are approximately 20.0 g and 30.0 g, respectively (Singh, et al. 1991). In Africa, both Andean and Mesoamerican beans are grown at varying percentages across countries or regions, and with narrow diversity of Andean beans (Bellucci et al. 2014; Cichy et al. 2015). In eastern Africa, Andean beans are the more preferred (Farrow and Muthoni-Andriatsitohaina 2020).

The market for beans in Africa is mainly driven the grain-type. The segments are characterized by size and color attributes (Buruchara et al. 2011). Popular market classes include large red mottled (calima), small and large white, yellow, sugar (speckled), dark red kidney, small red, carioca (small striped), and pinto, with varying preferences for more than one of these market classes within countries (Buruchara et al. 2011). Because of their high market demand, the mottled and plain red beans account for approximately

43% of the production in East Africa, followed by sugar, yellow, and white beans at 18, 13, and 9%, respectively (Farrow and Muthoni-Andriatsitohaina 2020). Large red mottled, sugar and yellow beans dominate the production in Uganda (Buruchara et al. 2011; Farrow and Muthoni-Andriatsitohaina 2020).

Information on diversity and population structure are essential for breeding to result in progressive genetic improvement and for the conservation of genetic diversity. Analyses of African germplasm using molecular markers revealed high diversity in East, Central, and South African bean germplasm for crop improvement, with more diversity in Mesoamerican than Andean genotypes (Asfaw et al. 2009; Nhkata et al. 2021). Such findings point to the possibility of increased molecular breeding of common bean in the region. The use of molecular markers decreases the challenges of conventional methods that are time-consuming, environment-dependent, and result in marginal genetic gains (Crossa et al. 2017). A diverse population is a source of desirable genetic variants for breeding. It is also important as a reference for predicting models in genomic selection (Tianhua and Chengdao 2020). However, such populations are often highly structured, which can influence prediction accuracy, heritability and trait association based on whole genome analysis (Azevedo et al. 2017).

Several PCR-based markers, such as SSR and AFLP (Beebe et al. 2001; Zargar et al. 2016), and microsatellite-based methods (Asfaw et al. 2009) have been deployed to differentiate gene pools and understand diversity in common beans. Many recent studies have utilized SNP marker-based technologies because of their high throughput, low cost, and abundance of SNPs in the genome (Mammadov et al. 2012; Bellucci et al. 2014; Cichy et al. 2015; Nhkata et al. 2021). Diversity arrays technology sequencing (DArTseq) and genotyping by sequencing (GBS) are both widely adopted SNP-based methods for genetic studies that involve a genome complexity reduction step, which selectively avoids highly repetitive genomic regions (Elshire et al. 2011; Kilian et al. 2012). Performance differences were reported in the use of DArT and GBS data but strong significant [$P \leq 0.01$] correlations of genetic diversity matrices [$r = 0.6$ to 0.72], and similar structures across the main principal components between the two approaches showed that both are efficient (Elbasyoni et al. 2018; Darrier et al. 2019).

This study aimed to evaluate the genetic diversity and structure of breeding lines and released varieties in East Africa to determine the relevance of the population for genomic selection, for use in the Pan African Bean Research Alliance (PABRA). Two populations that consisted of 725 common bean genotypes were analyzed by high-density single nucleotide polymorphic (SNP) markers generated using either DarTseq or GBS.

Materials and methods

Plant material

In total, 427 breeding lines, including their parents, and a panel of 298 breeding lines and varieties released in eastern Africa, hereafter referred to as Set1 and Set2, respectively, were studied. The germplasm originated from breeding programmes in six African countries, South America and North America, with the majority

(582 lines) sourced from Uganda (Table 1). Set1 consisted of 33 parental and 394 advanced lines selected from 71 populations developed for common bacterial blight (CBB) and angular leaf spot (ALS) resistance, drought tolerance, fast cooking time, and seed iron/zinc improvement at the Alliance of Bioversity and CIAT in Uganda. Set2 included 82% breeding lines, 17% varieties, and 1% landraces. The major common bean market classes in Africa, which include red and red mottled (Calima), sugar (speckled), white, and yellow beans, consisted of 40.4% of all the germplasm (Table 1).

DNA extraction, genotyping, and filtering

The DNA from Set1 germplasm was extracted from the leaves of 10–14 days old plants using the cetyltrimethyl ammonium bromide (CTAB) method adapted from Doyle and Doyle (1987). The DNA quality was assessed by 0.8% Agarose gel electrophoresis, and then quantified by PicoGreen (Moreno and Cox

Table 1 Origin and market class within gene pools of 472 (Set1) and 298 (Set2) common bean genotypes

	Set1			Set2		
	Andean	Mesoamerican	Admixture	Andean	Mesoamerican	Admixture
Total	87	305	35	64	140	94
Origin						
Uganda	72	293	34	31	65	88
South America	13	12	1	23	64	3
North America				4	7	2
Rwanda	1	0		4	2	1
Malawi	1	0		1	0	0
Tanzania				1	0	0
DRC				0	1	0
Ethiopia				0	1	0
Color						
Black	0	43	1	1	39	11
Brown	3	6	2	1	4	0
Cream (Pinto, Carioca, Mulatinho)	2	159	5	13	17	20
Others	9	16	18	2	6	2
Pink	1	0	0	1	3	2
Purple	5	0	1	0	4	28
Red	7	51	2	9	35	2
Red mottled	15	0	0	10	11	25
Sugar (speckled)	23	1	3	20	0	2
White	9	26	3	3	20	2
Yellow	13	3	0	4	1	0

2010). DNA libraries for each sample were constructed following the GBS protocol of Elshire et al. (2011). The DNA from the germplasm was digested using ApeKI, and a unique barcoded adapter and a common adapter were used to ligate restriction fragments. Libraries, each comprising 96 germplasm samples were sequenced in a lane of an Illumina HiSeq4000 at Michigan State University using the sequencing format Single End 150-pb. The quality check on raw data was performed in FASTQC, a Java-based quality control tool, which provided per-base and per-read quality profiles (Andrews 2010). Cutadapt was used to search and remove adapters, and filter short or untrimmed reads (Martin 2011). Bowtie2 (Langmead and Salzberg 2012) was used to align the sequences against the reference genome *P. Vulgaris* v.2.1 (<https://phytozome.jgi.doe.gov/>). Bowtie2 extracted 'seed' substrings from the read and its reverse complement, aligned the extracted substrings to the reference in an ungapped fashion assisted by the full-text minute index, prioritized seed alignments and calculated their positions in the reference genome from the index, and finally extended seeds into full alignments by performing single-instruction multiple-data (SIMD)-accelerated dynamic programming (Langmead and Salzberg 2012). SNP calling was done with the software NGSEP (Duitama et al. 2014).

For Set2 germplasm, leaf samples from 10 to 14 days old plants were punched into 96-well plates and oven-dried at 40 °C for 12 h. A 200 g silica gel enclosed in an envelope was attached to each sealed plate before shipping to the Integrated Genotyping Service and Support (IGSS) in Nairobi, Kenya. DNA was extracted using the Nucleomag plant genomic DNA extraction kit (Macherey–Nagel). The quality and quantity were assessed on 0.8% agarose gels and using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). Genotypes were sequenced using diversity array technology sequencing (DartSeq). Genomic complexity reduction technology was used to construct a genomic DNA library (Kilian et al. 2012), which was purified and quantified for cluster generation in an automated clonal amplification system (cBOT Illumina). Sequencing was performed on HiSeq2500-Illumina. Reads were aligned to the Common bean v9 and Soybean v2 blast databases using an E-value of $5e-5$ and 90% minimum sequence identity.

Through GBS and DArT, 210,450 and 22,890 SNP markers were generated in Set1 and Set2 respectively.

Data were filtered in TASSEL v5.2.75 (Bradbury et al. 2007) using a minimum call rate of 80%, and minimum minor allele frequency (MAF) of 0.01, resulting in 3061 markers for Set1 and 11,636 markers for Set2 that were considered for further analysis. Linkage disequilibrium k-nearest neighbors' imputation (LD KNNI) was performed on the resulting SNP data in the same software.

Population structure and diversity analysis

Several methods exist for population structure analysis as elucidated by Raj et al. (2014) to bring out both advantages and disadvantages. The software "Structure" is commonly used but it is computationally challenging for large datasets. The program "fast-Structure" was developed to achieve accuracies comparable to "structure" and "admixture" by using the variational Bayesian framework for the posterior inference that greatly speeds up the inference of population (Raj et al. 2014). Subpopulations were generated in fastSTRUCTURE v.1.0 (Raj et al. 2014). The program was run with the default convergence criterion of 10^{-6} , considering a logistic prior distribution. To impose a logistic prior, a population-specific allele frequency was generated at each given locus, with the normal distribution having a locus-specific mean and a population-specific variance (Raj et al. 2014). The analyses were run on several subpopulations ranging from $k=1$ to $k=15$ for Set1 and $k=1$ to $k=7$ for Set2, with ten cross-validation replicate runs per dataset. The algorithm for multiple choices of K (choosek) was used to determine the right number of subpopulations for each dataset. The k with the highest log marginal likelihood and low cross-validation error was considered. The structure plots were then visualized using the Distruct package. The Principal Component Analysis (PCA) was carried out in TASSEL v5.2.75 (Bradbury et al. 2007) with default parameters, and the outputs were plotted in R software v4.1.2 through ggplot2 and ggrepel packages (Wickham 2016; Slowikowski et al. 2021). Different hierarchical analyses of molecular variance (AMOVA) to evaluate the amount of population genetic structure were performed in Arlequin v3.5.2 (Excoffier and Lischer 2015). Global AMOVA results as a weighted average over loci were generated, and significance tests were obtained by 1023 permutations. The F_{ST} based genetic distances for short divergence time were

generated from the same software. The pairwise F_{ST} p-values (significance tests) were obtained by 110 permutations. To confirm subpopulations, phylogenetic relationships among genotypes were also determined by generating maximum likelihood trees in iqtree v2.1.2 (Nguyen et al. 2015). The Ultrafast Bootstrap Approximation (UFBoot) method (Minh et al. 2013) was used with 1000 bootstrap replications. The Model Finder Plus (MFP) option was used to infer the best nucleotide substitution model to use for the phylogenetic reconstruction (Kalyaanamoorthy et al. 2017). The resulting tree was visualized in FigTree v1.4.4 (Rambaut 2018). The summary statistics of the SNP markers, which included major allele frequencies, gene diversity, heterozygosity (H_e), and the polymorphism information content (PIC), were computed using PowerMarker v3.25 (Liu and Muse 2005). Heterozygosity is the proportion of heterozygous individuals in the population. At a single locus, it is estimated as $\hat{H}_l = 1 - \sum_{u=1}^k \hat{p}_{lu}^2$, where p_{lu} is the frequency of an allele (A_n) at the l th locus in a population (Liu and Muse 2005). Gene diversity, often referred to as expected heterozygosity, is the probability that any two alleles chosen at random from within the population are different. An unbiased estimator of gene diversity at the l th locus is $\hat{D}_l = \frac{(1 - \sum_{u=1}^k \hat{p}_{lu}^2)}{(1 - \frac{1+f}{n})}$, where the inbreeding coefficient, f , was estimated from the data using the method of moments, and P_{lu} is the frequency of an allele (A_n) at the l th locus in a population (Liu and Muse 2005). The PIC measures polymorphism, which indicates the usefulness of a molecular marker for linkage analysis (Botstein et al. 1980). It was estimated as $\hat{PIC}_l = 1 - \sum_{u=1}^k \hat{p}_{lu}^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^k 2\hat{p}_{lu}^2 \hat{p}_{lv}^2$, where p_{lu} or p_{lv} is the frequency of an allele (A_n) or (A_v) at the l th locus in a population.

Results

Population structure based on SNP markers

Based on the highest log marginal likelihood, the optimal number of subpopulations (k) from fastSTRUCTURE analysis were 11 and two for Set1 and Set2, respectively (Fig. 1). The subpopulations in $k=11$ for Set1 were majorly associated with the pedigree of the germplasm; hence a k of 2 was considered to

represent Andean and Mesoamerican gene pools of common bean for both sets.

A cut-off of 0.80 for the mean of Q (membership probability) was considered for each line to be grouped as Andean or Mesoamerican. This resulted in 87 Andean (20.4%), 305 Mesoamerican (71.4%), and 35 Admixtures (8.2%) for Set1, and 64 Andean (31.5%), 140 Mesoamerican (47.0%) and 94 Admixtures (31.5%) for Set2. Based on the pre-grouping of the germplasm into Mesoamerican, Andean, or Admixtures for set1, $k=11$ was considered in-depth, and it revealed that subpopulations one to four and six, eight, and 10 consisted of 138 Mesoamerican parents and their progenies generated from 37 crosses (Table 2). Subpopulations seven and 11 consisted of 82 Andean beans from many pedigrees, and parents. Subpopulation nine consisted of 12 Mesoamerican and three admixture germplasm of various pedigrees (Table 2). All germplasm in subpopulation five were below the cut-off ($Q=0.80$), and up to 179 others from different subpopulations did not meet this criterion and were considered admixtures.

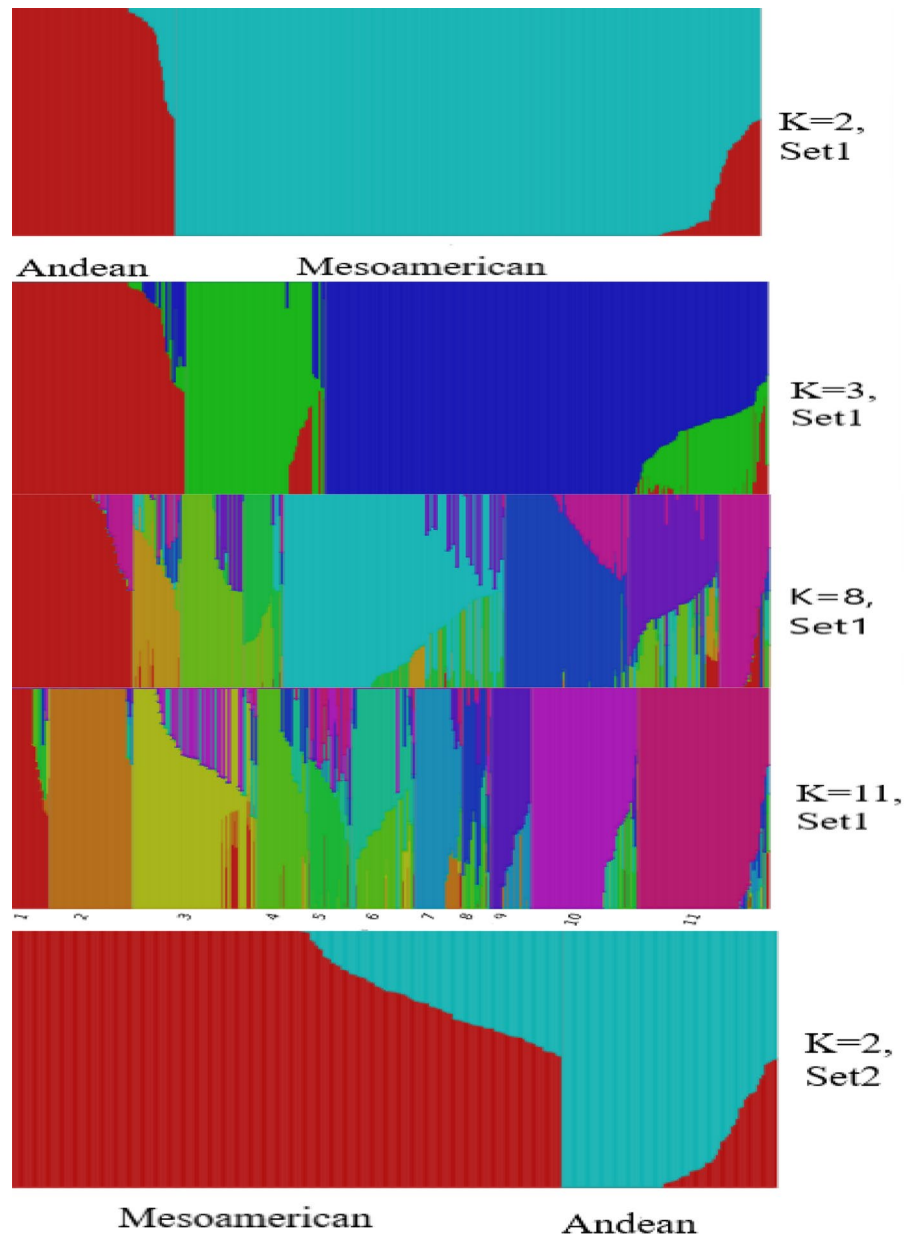
Analysis of molecular variance (AMOVA) explained significant ($P \leq 0.001$) differences among the subpopulations with 46% and 50% of the variation accounted for among populations in Set1 and Set2, respectively (Table 3). The fixation index (F_{ST}) showed that the three subpopulations in both sets were genetically distinct from each other, but stronger (0.54; 0.71) distinctiveness existed between Andean and Mesoamerican compared to these gene pools and admixtures (Table 3).

Further analysis to partition the variance obtained within populations based on market class, origin, or the release status showed that among populations accounted for <20 and <30% of the variation in Set1 and Set2, respectively (Table 4).

Based on principal component analysis (PCA), the first and second PCA accounted for 36% and 7% of the variation in Set1, and 42% and 10% in Set2, respectively (Fig. 2). In both sets, PCA_1 and PCA_2 spread the germplasm into three clear groups representing Andean, Mesoamerican, and Admixtures (Fig. 2), like in the fastSTRUCTURE result.

The maximum likelihood trees for both sets also showed more diversity within Mesoamerican (blue) and Admixtures (red), with Set2 showing Admixtures closer to each gene pool (Fig. 3).

Fig. 1 Population structure of 472 (Set1) and 298 (Set2) common bean genotypes as determined from the program fastSTRUCTURE for $k=11$ and $k=2$ subpopulations. In $k=11$, the groups were mainly based on the pedigree of the germplasm



Most of the Andean (green) and Mesoamerican germplasm were from Uganda, followed by South America, for both Set1 and Set2 (Table 1). Many of the Andean beans were of red mottled and sugar (speckled) market classes in both sets, whereas red, white, cream, and black dominated the Mesoamerican beans (Table 1).

Genetic diversity based on SNP markers

Before analysis, each of the two sets of bean germplasm was classified as Andean, Mesoamerican, and Admixtures. The proportion of the germplasm with the most common (major) allele for each SNP was 0.861 and 0.747 for Set1 and Set2, respectively (Table 5).

Table 2 The summary of subpopulations for set1 (k=11) indicating the composition, number of germplasm and some pedigree information

Sub population	Composition	Number of germplasm	Pedigree
1	Mesoamerican progenies	12	K131/NUA45, K132/JESCA, K132/MIB456, MASINDI YELLOW/RWV2070, MIB456/RWV2070, MIB456/JESCA, MIB456/MCM2001, MIB456/NUA45
2	Mesoamerican progenies	44	K131/MCM2001, K131/JESCA, K131/MIB456, K131/NUA45, K132/RWV2070, MASINDI YELLOW/RWV2070, MIB456/JESCA, MIB456/RWR2154, ACC714, SMC28
3	Mesoamerican progenies	20	K131/RWV2070, K131/MIB456, K131/MCM2001
4	Mesoamerican progenies	16	MIB456/NUA45, K131/RWV2070, K132/RWR2154, K132/RWV2070, MASINDI YELLOW/RWV2070, MCM2001/MIB456, NGWAKUNGWAKU/ACC714, MCM2001/NUA45, MIB456/MCM2001, K131/JESCA
8	Mesoamerican progeny	1	MIB456/MCM2001
6	Mesoamerican progenies + parent	3	MIB456/MCM2001, MCM2001/MIB456, K131/JESCA
10	Mesoamerican progenies + parent	42	K131/RWV2070, K131/MIB456, K132/MIB456, K131, KANYEBWA/(MEXICO54/AND277//G5856)
7	Andean	19	Various pedigrees and half sibs
11	Andean progenies and parents	63	Various pedigrees
9	Mesoamerican + Admixtures	12	Various pedigrees
5	Admixture	195	Various pedigrees

Table 3 AMOVA results as a weighted average over 3060 loci for Set1 and 11,636 loci for Set2, and fixation index (F_{ST}) among subpopulations

AMOVA					Pairwise F_{ST} ^a information		
Source of variation	Sum of squares	Variance components	Percentage variation	P-value	Sub populations	Andean	Mesoamerican
Set1					Set1		
Among populations	74,447	196	46.2	≤0.001	Andean		
Within populations	194,643	229	53.8		Mesoamerican	0.54***	
Total	269,090	425			Admixtures	0.27***	0.22***
Set2					Set2		
Among populations	452,207	1192	50.3	≤0.001	Andean		
Within populations	698,035	1178	49.7		Mesoamerican	0.71***	
Total	1,150,242	2369			Admixtures	0.42***	0.30***

^a F_{ST} , a measure of population differentiation due to genetic structure

AMOVA analysis of molecular variance

Gene diversity was highest in Admixtures and lowest in Andean beans for both sets. When all the germplasm were considered, the Set2 (0.332) germplasm had a higher gene diversity than Set1 (0.206). The proportion of heterozygotes for each SNP was generally low in both sets, and the PIC was higher in Set2 (0.263) than Set1 (0.172) (Table 5). When the germplasm were grouped

according to their market classes, gene diversity was generally similar within sets. The red and black (≈ 0.2), and red, cream, and pink (≈ 0.3) beans had the highest diversity in Set1 and Set2, respectively (Table 5). Based on the release status, slightly higher diversity was found in the varieties than in the breeding lines in both sets, although Set2 had higher diversity than Set1 (Table 5).

Table 4 Analysis of molecular variance (AMOVA) and results as a weighted average over different loci for Set1 and Set2 based on market classes, origin and the release status for Andean, Mesoamerican, and admixtures

Source of variation	Set1				Set2			
	Sum of squares	Variance components	Percentage variation	P-value	Sum of squares	Variance components	Percentage variation	P-value
	AMOVA design and results (average over 2508 loci) for Andean beans				AMOVA design and results (average over 9037 loci) for Andean beans			
<i>Market class</i>								
Among populations	2942.7	13.8	6.7	***	19,211.7	188.8	24.8	***
Within populations	32,057.0	192.0	93.3		68,596.7	571.7	75.2	
Total	34,999.8	205.8			87,808.4	760.5		
<i>Origin</i>								
Among populations	1229.2	12.4	5.9	***	14,294.0	140.4	19.0	***
Within populations	33,770.7	198.7	94.1		73,514.4	602.6	81.1	
Total	34,999.8	211.1			87,808.4	743.0		
<i>Release status</i>								
Among populations	1079.9	11.5	5.5	***	4963.6	62.5	8.6	***
Within populations	33,919.9	198.4	94.5		82,844.9	662.8	91.4	
Total	34,999.8	209.9			87,808.4	725.3		
	AMOVA design and results (average over 2790 loci) for Mesoamerican beans				AMOVA design and results (average over 11,172 loci) for Mesoamerican beans			
<i>Market class</i>								
Among populations	18,457.9	42.5	17.5	***	77,906.5	329.3	29.0	***
Within populations	120,570.3	200.0	82.5		218,791.0	807.8	71.0	
Total	139,028.3	242.4			296,697.5	1137.1		
<i>Origin</i>								
Among populations	1032.3	17.5	7.1	***	59,236.2	342.6	28.3	***
Within populations	137,995.9	227.0	92.9		237,461.4	867.1	71.7	
Total	139,028.3	244.4			296,697.5	1209.8		
<i>Release status</i>								
Among populations	621.9	12.7	5.3	***	10,352.7	101.4	9.0	***
Within populations	138,406.3	227.6	94.7		286,344.9	1030.6	91.0	
Total	139,028.3	240.3			296,697.5	1132.0		
	AMOVA design and results (average over 2378 loci) for Admixtures				AMOVA design and results (average over 10,828 loci) for Admixtures			
<i>Market class</i>								
Among populations	3501.3	40.0	12.8	***	44,369.9	250.0	14.4	***
Within populations	17,113.2	271.6	87.2		269,159.4	1487.1	85.6	
Total	20,614.5	311.7			313,529.3	1737.1		
<i>Origin</i>								
Among populations	514.7	56.4	16.0	***	14,683.6	427.8	20.8	***
Within populations	20,099.9	295.6	84.0		298,845.7	1624.2	79.2	
Total	20,614.5	352.0			313,529.3	2051.9		
<i>Release status</i>								
Among populations	514.663	56.4	16.0	***	6501.5	316.6	16.1	***
Within populations	20,099.8	295.6	84.0		307,027.8	1650.7	83.9	
Total	20,614.5	352.0			313,529.3	1967.3		

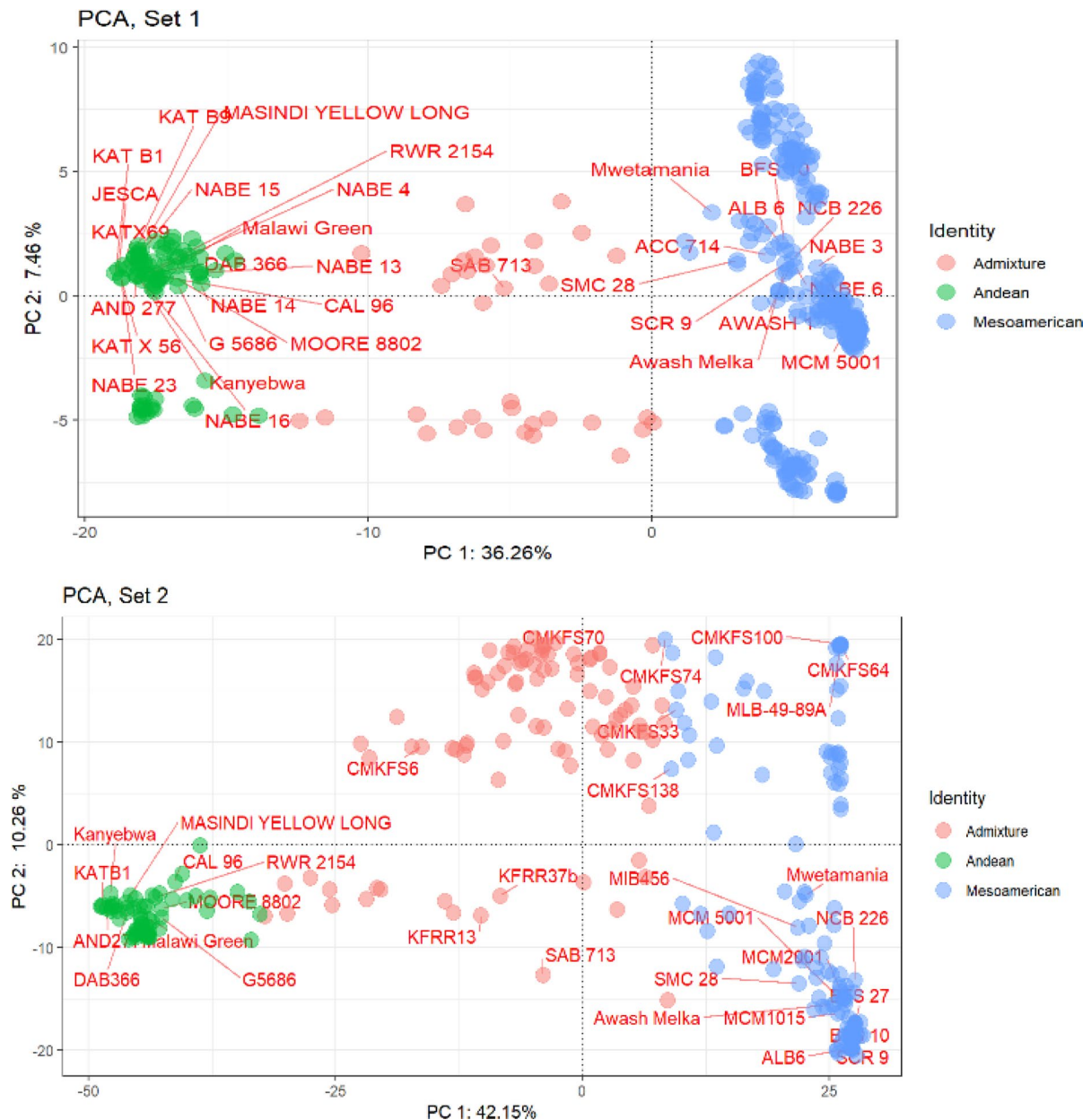


Fig. 2 Principle component analysis (PCA) of 472 (Set1) and 298 (Set2) common bean genotypes generated in Tasslev5 using 3061 and 11,636 single nucleotide polymorphic (SNP) markers and plotted in R software. (Color figure online)

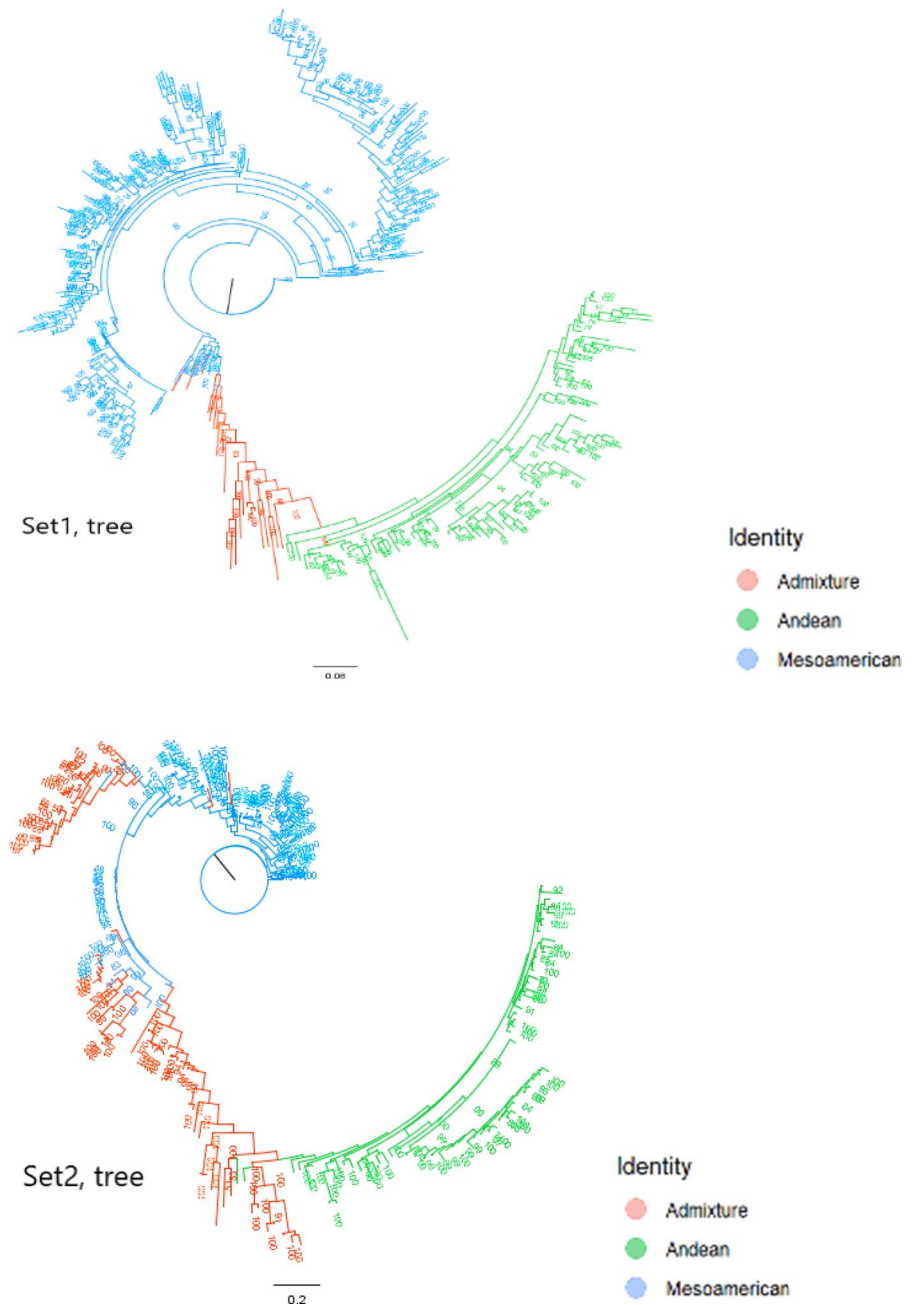
Discussion

Population structure based on SNP markers

The germplasm were classified into Andean and Mesoamerican subpopulations, with higher diversity in the latter. Other studies have also reported higher

diversity in the Mesoamerican gene pool (Bellucci et al. 2014; Cichy et al. 2015; Nkhata et al. 2021). These two gene pools are associated with geographical areas of origin and exist in both the wild and cultivated beans (Singh et al. 1991). Genetic diversity within the subpopulations was also high (>40%); however, only a small proportion could be explained

Fig. 3 Maximum likelihood tree of 472 (Set1) and 298 (Set2) common bean genotypes generated in iqtrees using 3061 and 11,636 single nucleotide polymorphic (SNP) markers and plotted in FigTree v1.4.4. Mesoamerican (blue), Andean (green) and Admixture (red). (Color figure online)



by market class (seed color), origin, or the release status. The Mesoamerican gene pool is widely known for its diversity in morphological and economically important phenotypes. The Andean gene pool was also reported to have high genetic diversity in seed color and size, agroecological adaptation, growth habit, disease resistance, and seed iron (Blair et al. 2007; Cichy et al. 2015). The observed variations

within the subpopulations could be attributed to the pedigrees. This could be seen in Set1, where the optimal number of subpopulations (k) was 11, two of which were Andean beans, and the others were Mesoamerican and Admixtures mostly grouped by their similar pedigrees. Some breeding lines were created by hybridizing parents within market classes, but others were created across market classes. While

Table 5 Genetic diversity in 472 (Set1) and 298 (Set2) common bean genotypes using 3061 and 11,636 single nucleotide polymorphic (SNP) markers

Groups	Major allele frequency	Gene Diversity	He	PIC	Major allele frequency	Gene Diversity	He	PIC
Gene pools	Set1				Set2			
Andean	0.913	0.131	0.091	0.112	0.922	0.119	0.027	0.101
Mesoamerican	0.900	0.149	0.099	0.126	0.869	0.183	0.049	0.151
Admixtures	0.860	0.193	0.101	0.158	0.860	0.193	0.101	0.157
All 427 or 298	0.861	0.206	0.097	0.172	0.747	0.332	0.064	0.263
Market class								
Red	0.850	0.214	0.099	0.175	0.745	0.330	0.064	0.261
Sugar (Speckled)	0.897	0.157	0.102	0.133	0.906	0.141	0.043	0.120
Cream (Pinto, Carioca, Mulatinho)	0.905	0.150	0.100	0.129	0.750	0.323	0.045	0.256
White	0.904	0.142	0.073	0.120	0.816	0.264	0.047	0.217
Yellow	0.887	0.166	0.091	0.137	0.849	0.218	0.012	0.178
Purple	0.891	0.154	0.120	0.127	0.810	0.245	0.080	0.194
Black	0.899	0.147	0.111	0.123	0.892	0.177	0.098	0.154
Brown	0.850	0.207	0.102	0.168	0.833	0.243	0.016	0.199
Pink					0.757	0.302	0.018	0.238
Others	0.855	0.202	0.084	0.165	0.803	0.271	0.083	0.218
Release status								
Breeding line	0.867	0.199	0.096	0.167	0.756	0.322	0.069	0.256
Variety	0.844	0.212	0.119	0.171	0.736	0.344	0.041	0.272
Origin								
Uganda	0.866	0.201	0.096	0.169	0.755	0.309	0.080	0.245
South America	0.825	0.223	0.124	0.178	0.789	0.299	0.034	0.242
North America					0.725	0.344	0.082	0.269
Rwanda					0.777	0.299	0.021	0.238

He heterozygosity, *PIC* polymorphic information content

hybridization among market classes increases the chance of using more superior parents, the recovery of the target seed color attributes, and seed size requires attention. Backcrossing (White et al. 1994), the use of molecular markers (Lei et al. 2020), and phenotype-based selection for seed size through the weight of 100 seeds are feasible options to recover the target seed size. These approaches are also possible for seed color and pattern, but the chance of recovery is increased by combining closely related types. This is perhaps because seed pattern and color are controlled by multiple genes, many of which exhibit epistatic interactions that define the observed phenotype (McClean et al. 2002).

The Andean, Mesoamerican, and Admixtures had a representative number of diverse market classes that are important for breeding specific product profiles

within and across gene pools and market classes. Elite x elite crosses within a market class and gene pool are recommended when possible because they generate adequate diversity for short-term breeding (Falk 2010). Recent simulations also show the feasibility of maintaining diversity for long-term genetic gain in elite line hybridization scheme when an optimized mating strategy is used (Allier et al. 2019). The Admixtures had no yellow colored bean, and only 16 others were present in the population. Increasing the diversity of this market class in the breeding population will enhance genetic gain within the yellow beans.

Information on population structure is relevant for the genomic selection-based breeding approach. Studies commonly report prediction accuracy based on random cross-validation, which can be strongly

inflated by the population or family structure, a characteristic many breeding populations share (Werner et al. 2020). Thus, the subpopulations could be used to train prediction models and select genetically diverse parents. Long-term maintenance of genetic variation, which is essential in the reference population for genomic selection (Tianhua and Chengdao 2020), necessitates selecting distantly related individuals or minimizing the inbreeding coefficient (Brisbane and Gibson 1995; Lindgren and Mullin 1997).

Genetic diversity based on SNP markers

The level of observed gene diversity showed that the admixtures were the most diverse in both Set1 and Set2, followed by Mesoamerican and Andean gene pools, which would encourage inter-gene pool crosses. The highest gene diversity was also previously reported in the admixtures (Cichy et al. 2015). Inter-gene pool crosses may be effective but attentiveness to key progress indicators of trait improvement is required to realize gains. Even in the admixtures (≈ 0.1), heterozygosity was low in Set1 and Set2. Common bean is a self-pollinating crop, and it is expected to be homozygous at most loci, especially when single seed descent was the breeding method (Snape and Riggs 1975) as observed here. Most of the genotypes were advanced breeding lines with mostly fixed genes, but inter-gene pool crosses were less homozygous. Low heterozygosity in common bean was reported (Cichy et al. 2015); however, higher heterozygosity (≈ 0.45) attributed to uncontrolled outcrossing was also reported (Nkhata et al. 2020). Regulating allelic diversity and heterozygosity in a breeding population is one way of sustaining long-term genetic gain (Akdemir and Sanchez 2016).

The diversity in all market classes showed the potential of the germplasm to be considered as a reference population for genomic selection as more than one product profile is possible. Based on this attribute, in addition to disease resistance or seed iron, seven product profiles, including large white, large red, large red mottled, small red, small white, sugar (speckled) bean, and yellow beans, were developed across three East African countries, including Uganda (Mukankusi et al. 2019). A product profile could be defined as the complete range of technical attributes

required for a new variety to be successfully released into a market segment (Ragot et al. 2018).

In consideration of the release status and origin of the germplasm, the observed genetic diversity indicated that breeding efforts in Uganda have generated some comparable diversity. Thus progenies (breeding lines) could be used as parents for further improvement rather than reusing the reference parents since this is especially important to realize a continuous genetic gain.

PIC reveals the value of a molecular marker for linkage analysis (Botstein et al. 1980). Similar PIC values of 1.5 to 2.2 like obtained in this study were reported in African germplasm using SNP-based markers (Cichy et al. 2015; Nkhata et al. 2021). Such levels of polymorphism have been used to determine whether common SNPs in the population are linked to traits of interest. While both Set1 and Set2 may be useful, Set2 recorded higher polymorphism probably because it consisted of 38% of germplasm from sources other than Uganda in comparison to Set1's 6%.

Conclusions

The structure and polymorphic information generated in the study provides useful perspectives for genomic breeding and genome wide association study using the population. In addition, the subpopulations exhibited a level of genetic diversity that can be used for breeding. Among gene pools, the market classes including the red, black, and cream had the highest genetic diversity. The observed variations within and among subpopulations were useful in guiding product profile development to realize genetic gain.

Acknowledgements The common bean breeding team at the Alliance of Bioversity International and CIAT at Kawanda, the National Crops Resources Research Institute (NaCRRI) at Namulonge, Uganda, the Tanzania Agricultural Research Institute (TARI) Maruku, Feed the Future Crop Improvement Innovation Laboratory (CIIL) Project, Cornell University and Michigan State University, United States of America.

Author contributions Designed the study: Prof. JK, Assoc. prof. KC, Dr. CM, Dr. SN and WA. Responsible for the project funding: Assoc. prof. KC; Dr. SN; Dr. CM. Supervised the Study: Dr. CM, Dr. NS, Dr. O-SM, Dr. AB, Dr. NE, Dr. OTL, Dr. TP. Conceived the structure of the manuscript and wrote the initial manuscript: WA. Conducted the experiments, and the molecular genotyping: WA, PI. Designed data analyses

and supervised interpretation of data: WA, Dr. AB. Analyzed the data: WA, Dr. AB. Reviewed the manuscript: Dr. AB, PI, Assoc. prof. KC, Dr. DOI, Dr. OTL, Dr. NE.

Funding USDA Funding Opportunity: 2017 Norman E. Borlaug International Agricultural Science and Technology Fellowship Program, USAID Feed the Future Innovation Lab for Crop Improvement (ILCI), and the Carnegie Corporation of New York through the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM), Grant Number RN/2020/RI/02.

Data availability All the data which is generated and analyzed during our study are included within the article.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

- Akdemir D, Sanchez JI (2016) Efficient breeding by genomic mating. *Front Genet* 7:210. <https://doi.org/10.3389/fgene.2016.00210>
- Allier A, Lehermeier C, Charcosset A, Moreau L, Teyssedre S (2019) Improving short- and long-term genetic gain by accounting for within-family variance in optimal cross-selection. *Front Genet* 10:1006. <https://doi.org/10.3389/fgene.2019.01006>
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Institute. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Accessed 1 June 2020
- Asfaw A, Blair WM, Almekinders C (2009) Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor Appl Genet* 120:1–12. <https://doi.org/10.1007/s00122-009-1154-7>
- Azevedo CF, Resende MDV, Fabyano FS, Nascimento M, Viana JMS, Valente MSF (2017) Population structure correction for genomic selection through eigenvector covariates. *CBAB* 17(4):350–358
- Beebe S, Rengifo J, Gaitan E, Duque MC, Tohme J (2001) Diversity and origin of Andean landraces of common bean. *Crop Sci* 41:854
- Bellucci E, Bitocchi E, Rau D, Rodriguez M, Biagetti E, Giardini A, Attene G, Nanni L, Papa R (2014) Genomics of origin, domestication and evolution of *Phaseolus vulgaris*. In: Tuberosa R, Graner A, Frison E (eds) *Genomics of plant genetic resources*. Springer, Netherlands, pp 483–507
- 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(1):29–43. <https://doi.org/10.1007/s00122-007-0644-8>
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19):2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Brisbane JR, Gibson JP (1995) Balancing selection response and inbreeding by including predicted stabilized genetic contributions in selection decisions. *Genet Sel Evol* 27:541–549. <https://doi.org/10.1186/1297-9686-27-6-541>
- Buruchara R, Chirwa R, Sperling L, Mukankusi C, Rubyogo JC, Muthoni R, Abang MM (2011) Development and delivery of bean varieties in Africa: the Pan-Africa Bean Research Alliance (PABRA) model. *Afr Crop Sci J* 19(4):227–245
- CIAT (Centro Internacional de Agricultura Tropical) (2008) The impact of improved bush bean varieties in Uganda. Highlights-CIAT in Africa, 43. CIAT, Kampala
- CIAT (Centro Internacional de Agricultura Tropical) (2018) Beans: what we do breeding better crops. (Web-exclusive story). <https://ciat.cgiar.org/what-we-do/breeding-better-crops/beans/>. Accessed 14 Dec 2019
- Cichy AK, Porch GT, Beaver SJ, Cregan P, Fourie D, Glahn PR, Grusak AM, Kamfwa K, Katuuramu ND, McClean P, Mndolwa E, Nchimbi-Msolla S, Pastor-Corrales AM, Miklas NP (2015) A *Phaseolus vulgaris* diversity panel for Andean Bean improvement. *Crop Sci* 55:2149–2160. <https://doi.org/10.2135/cropsci2014.09.0653>
- Darrier B, Russell J, Milner SG, Hedley PE, Shaw PD, Macaulay M, Ramsay LD, Halpin C, Mascher M, Fleury DL, Langridge P, Stein N, Waugh R (2019) A comparison of mainstream genotyping platforms for the evaluation and use of barley genetic resources. *Front Plant Sci* 10:544. <https://doi.org/10.3389/fpls.2019.00544>
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Duitama J, Quintero JC, Cruz DF, Quintero C, Hubmann G, Foulque-Moreno MR, Verstrepen KJ, Thevelein JM, Tohme J (2014) An integrated framework for discovery and genotyping of genomic variants from high-throughput sequencing experiments. *Nucleic Acids Res* 42(6):e44. <https://doi.org/10.1093/nar/gkt1381>
- Elbasyoni IS, Lorenz AJ, Guttieri M, Frels K, Baenziger PS, Poland J, Akhunov EA (2018) A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Sci* 270:123–130. <https://doi.org/10.1016/j.plantsci.2018.02.019>
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Excoffier L, Lischer H (2015) An integrated software package for population genetics data analysis. *Arlequin v3.5.2*. <http://cmpg.unibe.ch/software/arlequin35>
- Falk ED (2010) Generating and maintaining diversity at the elite level in crop breeding. *Genome* 53:982–991. <https://doi.org/10.1139/G10-081>

- Farrow A, Muthoni-Andriatsitohaina R (eds) (2020) Atlas of common bean production in Africa: second edition. Pan-Africa Bean Research Alliance (PABRA); International Center for Tropical Agriculture (CIAT). Nairobi, p242. <https://doi.org/10.7910/DVN/8W0QK7>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler AV, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14(6):587–589. <https://doi.org/10.1038/nmeth.4285>
- Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Uszynska KH, Jaccoud D, Hopper C, Kilian MA, Evers M, Peng K, Cayla C, Hok P, Uszynski G (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol* 888:67–89. https://doi.org/10.1007/978-1-61779-870-2_5
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359. <https://doi.org/10.1038/nmeth.1923>
- Lei L, Wang L, Wang S, Wu J (2020) Marker-trait association analysis of seed traits in accessions of common bean (*Phaseolus vulgaris* L.) in China. *Front Genet* 11:698. <https://doi.org/10.3389/fgene.2020.00698>
- Lindgren D, Mullin TJ (1997) Balancing gain and relatedness in selection. *Silvae Genet* 46:124–129
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9):2128–2129. <https://doi.org/10.1093/bioinformatics/bti282>
- Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. *Int J Plant Genomics* 728398:11. <https://doi.org/10.1155/2012/728398>
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17(1):10–12. <https://doi.org/10.14806/ej.17.1.200>
- McClearn PE, Lee RK, Otto Gepts CP, Bassett MJ (2002) Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). *J Hered* 93(2):148–152. <https://doi.org/10.1093/jhered/93.2.148>
- Minh BQ, Nguyen MAT, Haeseler AV (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30(5):1188–1195. <https://doi.org/10.1093/molbev/mst024>
- Moreno LA, Cox KL (2010) Quantification of dsDNA using the Hitachi F-7000 fluorescence spectrophotometer and Pico-Green dye. *J vis Exp* 5(45):2465. <https://doi.org/10.3791/2465>
- Mukankusi C, Raatz B, Nkalubo S, Berhanu F, Binagwa P, Kilango M, Williams M, Katungi E, Chirwa R, Beebe S (2019) Genomics, genetics, and breeding of common bean in Africa: a review of tropical legume project. *Plant Breed* 138:401–414. <https://doi.org/10.1111/pbr.12573>
- Myers JR, Kmiecik K (2017) Common bean: economic importance and relevance to biological science research. In: Perez de la Vega M, Santalla M and Marsolais F (eds) *The common bean genome. Compendium of plant genomes*. Springer, Cham, pp 2199–4781. https://doi.org/10.1007/978-3-319-63526-2_1
- Nguyen LT, Schmidt HA, Haeseler AV, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood. *Mol Biol Evol* 32(1):268–274. <https://doi.org/10.1093/molbev/msu300>
- Nkhata W, Shimelis H, Melis R, Chirwa R, Mzengeza T, Mathew I, Shayanowako A (2020) Population structure and genetic diversity analyses of common bean germplasm collections of East and Southern Africa using morphological traits and high-density SNP markers. *PLoS ONE* 15(12):e0243238. <https://doi.org/10.1371/journal.pone.0243238>
- Ragot M, Bonierbale M, Weltzien E (2018) From market demand to breeding decisions: a framework. lima (Peru). CGIAR Gender and Breeding Initiative. GBI Working Paper. No. 2. www.rtb.cgiar.org/gender-breeding-initiative. Accessed 15 Sept 2021
- Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197(2):573–589. <https://doi.org/10.1534/genetics.114.164350>
- Rambaut A (2018) FigTree: molecular evolution, phylogenetics and epidemiology. <http://tree.bio.ed.ac.uk/>. Accessed 20 Oct 2021
- Singh S, Gepts P, Debouck D (1991) Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ Bot* 45:379–396
- Slowikowski K, Schep A, Hughes S, Dang TK, Lukauskas S, Irlsson JO, Kamvar ZN, Ryan T, Christophe D, Hiroaki Y, Gramme P, Abdol AM, Barrett M, Cannoodt R, Krasowski M, Chirico M, Aphalo P (2021) ggplot2 v0.9.1: automatically position non-overlapping text labels with 'ggplot2'
- Snape JW, Riggs TJ (1975) Genetical consequences of single seed descent in the breeding of self-pollinating crops. *Heredity* 35:211–219. <https://doi.org/10.1038/hdy.1975.85>
- Werner CR, Gaynor RC, Gorjanc G, Hickey JM, Kox T, Abbadi A, Leckband G, Snowdon RJ, Stahl A (2020) How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. *Front Plant Sci* 11:592977. <https://doi.org/10.3389/fpls.2020.592977>
- White JW, Ochoa RM, Ibarra FP, Singh SP (1994) Inheritance of seed yield, maturity and seed weight of common bean (*Phaseolus vulgaris*) under semi-arid rainfed conditions. *J Agric Sci Camb* 122:265–273
- Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer, New York. ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>
- Zargar SM, Farhat F, Mahajan R, Bhakhri A, Sharma A (2016) Unraveling the efficiency of RAPD and SSR markers in diversity analysis and population structure estimation in common bean. *Saudi J Biol Sci* 23(1):139–149. <https://doi.org/10.1016/j.sjbs.2014.11.011>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”). Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com