



High-density DArTSeq SNP markers revealed wide genetic diversity and structured population in common bean (*Phaseolus vulgaris* L.) germplasm in Ethiopia

Yonas Moges Gelaw^{1,2} · John S. Y. Eleblu² · Kwadwo Ofori² · Berhanu Amsalu Fenta³ · Clare Mukankusi⁴ · Ermias Assefa Emam⁵ · Samuel Offei²

Received: 8 September 2022 / Accepted: 28 April 2023 / Published online: 30 June 2023

© The Author(s) 2023

Abstract

Introduction Common bean is one of the widely consumed food security crop in Africa, Asia, and South America. Understanding genetic diversity and population structure is crucial for designing breeding strategies.

Materials Two hundred and eighty-nine germplasm were recently collected from different regions of Ethiopia and introduced from CIAT to estimate genetic diversity and population structure using 11,480 DArTSeq SNP markers.

Results The overall mean genetic diversity and polymorphic information content (PIC) were 0.38 and 0.30, respectively, suggested the presence of adequate genetic diversity among the genotypes. Among the geographical regions, landraces collected from Oromia showed the highest diversity (0.39) and PIC (0.30). The highest genetic distance was observed between genotypes collected from SNNPR and CIAT (0.49). In addition, genotypes from CIAT were genetically more related to improved varieties than the landraces which could be due to sharing of parents in the improvement process. The analysis of molecular variance revealed that the largest proportion of variation was due to within the population both in geographical region (63.67%) and breeding status (61.3%) based classification. Model-based structure analysis delineated the 289 common bean genotypes into six hypothetical ancestral populations.

Conclusions The genotypes were not clustered based on geographical regions and they were not the main drivers for the differentiation. This indicated that selection of the parental lines should be based on systematic assessment of the diversity rather than geographical distance. This article provides new insights into the genetic diversity and population structure of common bean for association studies, designing effective collection and conservation for efficient utilization for the improvement of the crop.

Keywords DArTSeq · Genetic diversity · Heterozygosity · PIC · SNP markers

✉ Yonas Moges Gelaw
yonasmoges2006@gmail.com

¹ Haramaya University, Dire Dawa, Ethiopia

² West Africa Centre for Crop Improvement, University of Ghana, Legon, Accra, Ghana

³ Melkassa Agricultural Research Center, Ethiopian Institute of Agricultural Research (EIAR), Adama, Ethiopia

⁴ International Centre for Tropical Agriculture (CIAT), Kawanda, Kampala, Uganda

⁵ Bio and Emerging Technology Institute, Addis Ababa, Ethiopia

Introduction

Common bean (*Phaseolus vulgaris* L.) is a diploid ($2n = 2x = 20$) annual leguminous self-pollinated crop with estimated genome size ranged from 588 to 637 Mbp [1, 2]. It is an important component of agrobiodiversity to provide food, feed, cash income and improve soil fertility through its mutual association with rhizobium bacteria, thereby reducing the costs of chemical fertilizer and its negative impacts on the environment [3]. Thus, contribute to improved livelihood of smallholder farmers throughout the tropics [4, 5].

Africa is considered as the secondary center of diversity of common bean [6]. Common bean is one of the food security and commercial crops in Ethiopia widely cultivated in all regions. It largely contributes to food and nutrition

security and income generation for smallholder farmers [7]. It is one of the main components of nutrition sensitive agriculture which works to significantly reduce malnutrition from Ethiopia in 2030 [8]. In Ethiopia, it is grown by over 3.95 million households and is one of their main source of income. Between 2004 and 2017, income from bean exports increased by over seven-fold from USD 18 to 130 million. It is produced in an area of 355,550 ha and a production of 619,094.4 tonnes with a productivity of 1.74 ton ha⁻¹. The country stands 11th in area and 4th in production among the 36 countries that produce common bean in sub-Saharan Africa [9].

The productivity of common bean is low due to biotic and abiotic factors. The improved varieties are mainly introductions from other countries and do not adapted well to the diverse soil and climatic conditions of the country. On the other hand, landraces are well adapted to the local condition and sources for useful trait. Common bean cultivars found in the hands of the farmers are very diverse and the crop plays a diversified role within the Ethiopian farming system and food culture [10]. It has a wider importance in the lowlands of Ethiopia because of its better adaptability to the persistent warmer climates and prevalent disease pressures. Landraces are an important component of agrobiodiversity; however, most of them are at risk of extinction because of climate change and replacement with improved varieties. The genetic diversity present among the common bean genotypes conserved in the gene bank and landraces cultivated by the farmers can be a source of useful traits to be exploited by breeding programs. Collection and characterization of common bean germplasm from the different regions of Ethiopia should be considered as a high priority. Information on the genetic diversity and population structure of common bean landraces helps to preserve biodiversity and better management of the available genepool. It also helps, for better exploitation of genetic resources in breeding programs, to develop common bean varieties that increase productivity.

Phenotypic characterization and analysis of pedigree information are widely used in the improvement of common bean but are less efficient in the selection of elite materials due to environment and genotype x environment interaction. To overcome the challenge and accelerate the improvement efficiency, molecular markers could be used as a complement to accelerate the common bean breeding.

To use molecular markers in the diversity study, there needs to identify markers that are cheap, easy to work with, and have high reproducibility. Thus, various molecular markers have been used to study common bean genetic diversity and population structure: RFLP [11]; AFLP [12, 13]; RAPD [14, 15]; SSR [16–18]; ISSR [19]. However, these markers have limitations associated with poor genome coverage, low marker density, and cost ineffectiveness [20]. Recently, SNP markers are widely used in common bean

diversity study, population structure, and genome-wide association studies (GWAS) [21–24]. In Ethiopia, limited studies were conducted to characterize common bean germplasm using small number of germplasm and molecular markers such as ISSR [23, 25]; SSR [26]; SNP [24].

DArTSeq SNPs have been used in many cereal crops: sorghum [27], maize [28, 29], durum wheat [30], and legume crops: cowpea [31] and chickpea [32]. However, such studies are limited in common bean. Therefore, the objective of this study was to investigate the extent and pattern of genetic diversity and population structure of common bean genotypes using high-density DArTSeq SNP markers.

Materials and methods

Plant material

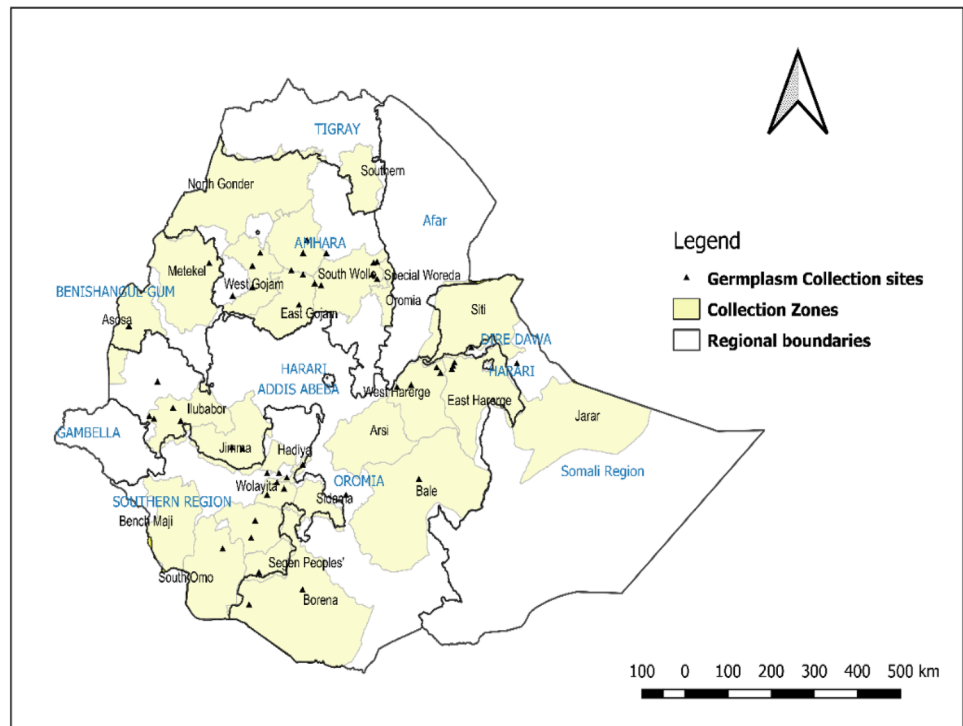
The present study consisted of 300 common bean genotypes including recent collections by Lowland Pulse Research Program of Melkassa Agricultural Research Center (MARC) from major common bean-growing areas of Ethiopia, Ethiopian Biodiversity Institute (EBI), improved varieties and genotypes from International Centre for Tropical Agriculture (CIAT). Information on the geographical description such as name of the locations and altitude was collected during the collection process. From the 164 common bean landraces, 85 were collected from Amhara, 38 from Oromia, 29 from Southern Nations and Nationalities and Peoples' Region (SNNPR), 10 from Benshangul Gumuz, and two from Somali region. The collection site of the genotypes is depicted in Table S1. In addition, 62 improved and pipeline varieties and 74 introductions from CIAT were included in the study. The improved and pipeline varieties were sourced from Haramaya University and Agricultural Research centers such as Melkassa, Hawassa, and Sirinka, while introductions were received from CIAT.

All the common bean materials used in this study, hereafter called genotypes, were regarded as a population and each grouping based on geographic regions and breeding status was regarded as subpopulation. The detailed description of the collection site is presented in Fig. 1.

DNA extraction and sequencing

Four representative seeds of each common bean genotypes were grown in a germination chamber. DNA extraction was done from a pooled sample of two-week-old seedling's leaf samples. The leaf samples were stored at -80 °C and genomic DNA was extracted from frozen leaf tissue using a modified CTAB protocol. The DNA quantity was checked using NanoDrop Spectrophotometer 2000c (ND-2000 v3.5, NanoDrop Technologies Inc). The DNA concentration was

Fig. 1 Collection sites of common bean landraces in Ethiopia.



quantified with a Qubit fluorometer (Life Technologies, Carlsbad, CA, USA) followed by running 0.8% agarose gel electrophoresis. The DNA concentration was adjusted to 50–100 ng μl^{-1} and sent for sequencing using GBS technology as described by [33] using DArTSeq™ technology platform in Biosciences eastern and central Africa (BeCA)-ILRI Hub, Nairobi, Kenya.

High-throughput DArTseq technology was applied to genotype the common bean genotypes. Complexity reduction was done using PstI to enrich genomic representations with single-copy sequences. Polymerase chain reaction (PCR) adapters were ligated to the PstI fragment ends followed by PCR amplification of adapter-ligated fragments. Genotype by sequencing was conducted using a combination of DArT complexity reduction methods and next-generation sequencing (NGS) following [34]. Next-generation sequencing was carried out using Illumina HiSeq2500. DArTSeq markers scoring was achieved using DArTsoft14 software plugin in KDCOMPUTE application (<http://www.kddart.org/kdcompute.html>). Two types of DArTSeq markers were scored, silicoDArT and SNP markers. Both markers were aligned to the Common bean_v9 reference genome of common bean to identify the chromosome positions.

Data filtering process and DArTSeq SNP calling

Initially, we received 24,191 single-row SNP markers from DArT Pty Ltd., which were polymorphic across common bean genotypes and distributed across the 11

chromosomes. Markers with unknown position were first removed from the analysis. The parameters such as call rate and MAF were determined in TASSEL platform v5.2.65. Accordingly, SNPs with $\geq 80\%$ call rate and MAF of ≥ 0.05 were retained for further analysis. In addition, genotypes with missing values of $\geq 30\%$ were removed from the analysis. Accordingly, 289 genotypes and 11,480 polymorphic SNPs were used for downstream analysis (Table S2).

Population structure and genetic diversity analysis

Population stratification into genetically similar groups was performed using STRUCTURE software v2.3.4 [35] based on Bayesian clustering model. The burn-in period of 50,000 and a Monte Carlo Markov Chain (MCMC) iteration of 100,000 were used by considering an admixture model with allele frequencies uncorrelated. The number of clusters was set using K values between 2 and 10 with 10 iterations. The optimal K value was detected by ΔK [36] using a web-based structure harvester software <http://taylor0.biology.ucla.edu/structureHarvester/>. Clusters were aligned from the independent runs using the in-file generated by structure harvester using CLUMPP server [37]. DISTRICT v1.1. was used to generate bar plots from average runs for the most probable K value [38].

Genetic relationship and summary statistics

Genetic distance and kinship matrix of 289 pairwise common bean genotypes based on 11,480 SNPs markers were calculated using Identity By State (IBS) method implemented in TASSEL v5.2.64 [39]. Diversity-relevant metrics such as expected heterozygosity (H_e), overall genetic diversity (GD), polymorphic information content (PIC), major allele frequency, and inbreeding coefficient (F_{is}) were computed using PowerMarker v3.25 software [40]. Arlequin v3.5.2.2. Software [41] was used to estimate the genetic variation between and within geographical regions and breeding status through the analysis of molecular variance.

For the estimation of genetic differentiation and relationships, distance between the genotypes and between pre-grouped populations were measured based on shared allele distances using the KD compute plugin system <https://kdcompute.seqart.net/kdcompute/login>. Clustering of the genotypes and PCA for the genetic relationships was computed using the allele-sharing matrix using KDCompute plugin system. Principal coordinate analysis was computed using GenAlEx software V6.5 [42].

Results

SNP marker variation and distribution

Whole-genome DArTSeq sequencing of common bean genotypes yielded 24,191 silicoDArT markers which were reproducible and aligned to the common bean reference genome v9 [43]. Among them, only the positions of 22,067 (91.2%) markers were provided. Scaffolds (170) and other markers (1954) were not located into any of the chromosomes.

In addition, genotypes were filtered for call rate and seven genotypes with $\leq 30\%$ call rate were removed and only 289 genotypes were used for the analysis. Based on these criteria, the remaining 11,480 polymorphic and high-quality SNPs distributed across the 11 common bean chromosomes were used for further analysis (Table S2).

The highest and lowest marker densities were observed on chr_2 (1280 SNPs) and on chr_4 (835 SNPs), respectively, with an average of 1044 markers per chromosome (Fig. 2). The average call rate and reproducibility of these markers were 0.94 and 1.00, respectively (Table 1). The average call rate, one ratio, heterozygosity, and PIC were consistent across the chromosomes.

Fig. 2 Distribution of SNP markers across the common bean chromosomes.

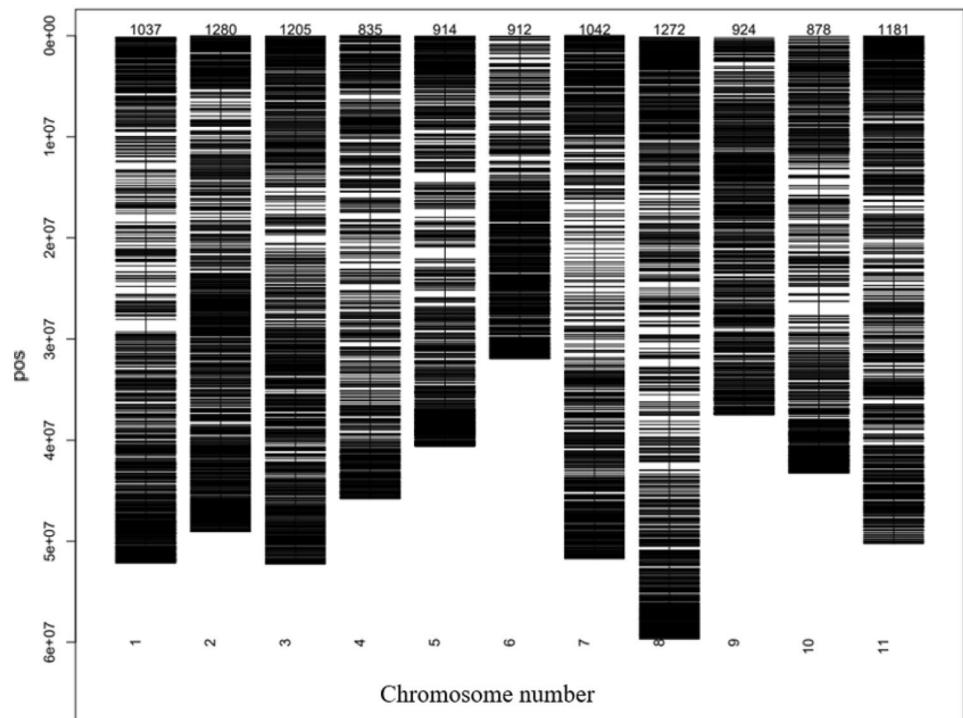


Table 1 Quality estimates of DArTSeq-derived SNP markers across the common bean chromosomes

Chromosome	Call rate	One ratio	Heterozygosity	Polymorphic information content	Average reproducibility
Chr_1	0.94	0.52	0.04	0.28	1.00
Chr_2	0.95	0.52	0.04	0.29	1.00
Chr_3	0.94	0.52	0.04	0.30	1.00
Chr_4	0.94	0.52	0.04	0.30	0.99
Chr_5	0.94	0.52	0.04	0.30	0.99
Chr_6	0.94	0.52	0.04	0.30	0.99
Chr_7	0.94	0.52	0.04	0.30	1.00
Chr_8	0.94	0.52	0.04	0.30	0.99
Chr_9	0.94	0.52	0.04	0.30	1.00
Chr_10	0.93	0.52	0.04	0.30	0.99
Chr_11	0.93	0.52	0.05	0.32	0.99
Mean	0.94	0.52	0.05	0.30	1.00

Table 2 Diversity estimates of 289 common bean genotypes using 11,480 DArTSeq SNP markers

Parameter	Minimum	Maximum	Mean
Genetic diversity	0.07	0.50	0.38
Polymorphic information content	0.06	0.38	0.30
Minor allele frequency	0.03	0.50	0.28
Heterozygosity	0.00	0.80	0.05
Reproducibility	0.9	1	1.00
Inbreeding coefficient	−0.65	1.00	0.85

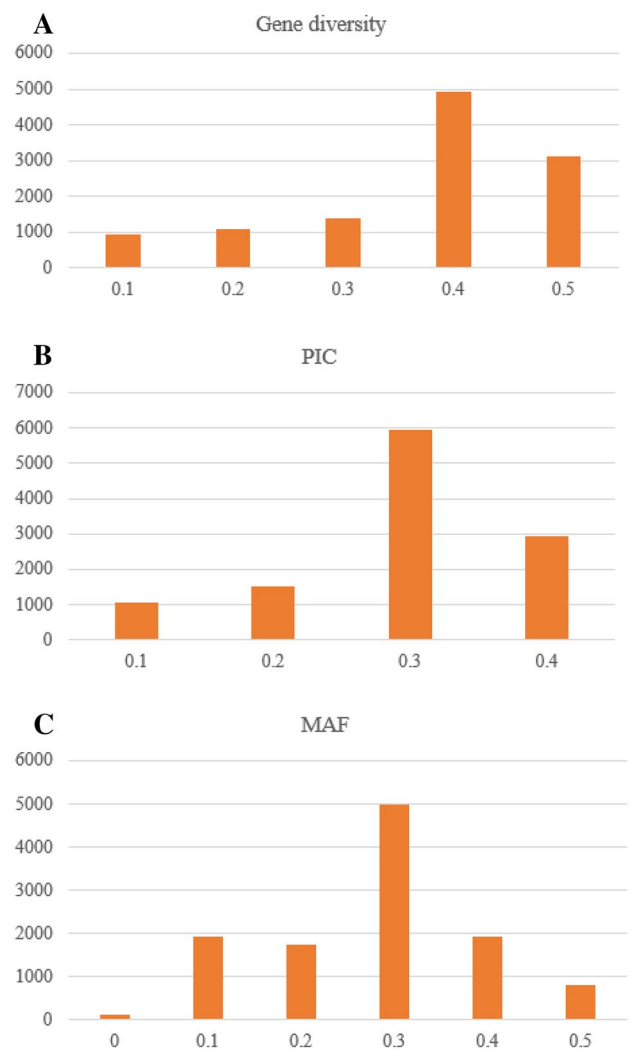
In this study, more transitions (58.23%) than transversions type SNPs (41.77%) were observed with a transition to transversion (Ts/Tv) ratio of 1.39:1 (6685/ 4795). It was also found that the C/T and A/G transitions were more than the T/C and A/G transitions. On the other hand, more transversions were observed in A/T, T/A, G/T, and C/T than G/C, T/G, A/C, and C/G.

The overall mean genetic diversity of the genotypes was 0.38 which ranged from 0.07 to 0.50 (Table 2). Polymorphic information content varied from 0.06 to 0.38 with an average of 0.30. Minor allele frequency (MAF) varied from 0.03 to 0.50 with a mean of 0.28. The mean inbreeding coefficient was 0.85. Single-nucleotide polymorphism markers exhibited a wide range of genetic diversity, PIC, and MAF across the common bean chromosomes (Fig. 3).

Genetic divergence based on geographic origin and breeding status of the genotypes

Genetic diversity based on geographic origin

Among the geographical regions, the highest genetic diversity (0.39) was found in the landraces collected from Oromia

**Fig. 3** Frequency distribution of genetic diversity, polymorphic information content (PIC), and Minor allele frequency (MAF) of 11,480 polymorphic SNPs for the 289 common bean germplasm.

region, followed by genotypes from CIAT (0.32), while the lowest diversity was obtained in the Benshangul Gumuz collections (0.11) (Table 3). The maximum heterozygosity (0.08) was found from the CIAT collections. Polymorphic information content across the geographical regions ranged from 0.11 (Benshangul Gumuz) to 0.30 (Oromia). The value of inbreeding coefficient (FIS) ranged from 0.72 for the landraces collected from Benshangul Gumuz to 0.88 for landraces collected from Oromia and improved varieties.

Genetic divergence between landraces and improved varieties

Genetic diversity of the landraces and improved varieties was 0.27 and 0.37, respectively, with a mean of 0.38. (Table 3). Heterozygosity was low both for the landraces (0.04) and for the improved varieties (0.06). Improved varieties showed higher (0.29) PIC as compared with the landraces (0.23).

Genetic relationship among genotypes

Pre-grouping of the genotypes into the different regions of collection and improved varieties revealed that the highest genetic distance was observed between SNNPR and CIAT (0.49), while genotypes from Oromia and Amhara showed close similarity (0.03) (Table 4). In general, genotypes from CIAT were closer to the improved varieties compared with landraces collected from Ethiopia. Grouping of genotypes revealed six distinct clusters. The result also showed that the genotypes were not clustered based on their geographical location, which could be due to the exchange of genetic materials across the geographic regions (Fig. 4).

Analysis of molecular variance

In the present study, the analysis of molecular variance of the common bean genotypes based on pre-defined geographic regions and breeding status is presented in Table 5. The result revealed that the largest proportion of variation was due to within populations based on both geographical region (63.8%) and breeding status (61.3%) classifications. Similarly, variation among populations accounted for 23.58% and 27.06% was due to geographical regions and breeding status, respectively. Fixation index, a measure of population differentiation, revealed high fixation index of 0.88 and 0.87 for within individual's variations for breeding status and geographic regions-based classification, respectively. Among populations, the fixation index of 0.83 and 0.84 was observed due to geographic regions and breeding status, respectively.

Table 3 Estimates of major allele frequency, gene diversity, heterozygosity, PIC, and inbreeding coefficient of 287 common bean germplasm classified based on geographic location and breeding status

Population	No. of genotype	genetic diversity	Heterozygosity	PIC
Amhara	81	0.27	0.04	0.23
Benshangul Gumuz	10	0.13	0.04	0.11
CIAT	74	0.32	0.08	0.26
Improved varieties	59	0.32	0.04	0.26
Oromia	38	0.39	0.05	0.30
SNNPR	25	0.14	0.04	0.12
Landrace	154	0.27	0.04	0.23
Improved varieties	133	0.37	0.06	0.29
Mean		0.38	0.05	0.30

CIAT: International Centre for Tropical Agriculture, PIC: Polymorphic Information Content, SNNPR: Southern Nations Nationalities and People's Region

Table 4 Pairwise Ne's genetic distance showing genetic differentiation among the different regions of common bean collection sites

Geographic area	Amhara	Benshangul Gumuz	CIAT	Improved	Oromia	SNNPR
Amhara	0					
Benshangul Gumuz	0.1	0				
CIAT	0.38	0.47	0			
Improved	0.16	0.24	0.10	0		
Oromia	0.04	0.13	0.30	0.10	0	
SNNPR	0.08	0.07	0.49	0.27	0.13	0

Genetic diversity and population structure

The optimal K value to group the common bean genotypes into subpopulations was found at $K=6$ (Fig. 5). In the first level of subpopulation grouping, $K=2$, the common bean genotypes were roughly classified into Andean and Mesoamerican genepools (Fig. 5). One hundred and eighty-eight (65.05%) genotypes were assigned in the Mesoamerican group, while the remaining 96 (33.22%) genotypes were classified in the Andean genepool. (Fig. 6).

The remaining five genotypes were admixed including three landraces, namely SWM012, 211,266, 211,305, and two breeding lines from CIAT, *i.e.*, NUS-1 and NUA-355. The threshold of membership coefficient (q_i) to assign genotypes with a group was considered as $q_i \geq 0.70$. The known high seed Fe concentration line MIB 465 was assigned in the Mesoamerican genepool, while RWR2154 and CAL-96 were included in the Andean genepool as reference lines in the respective genepool.

At $K=6$, the Mesoamerican genepool was separated into five distinct classes, each subpopulation consisting of 35, 21, 35, 1, and 50 genotypes (coded as MA1, MA2, MA3, MA4, and MA5, respectively (Fig. 6). Fifty-five genotypes were considered as admixture. The Andean genepool (A1) remains unseparated, except four genotypes. The first subpopulation (MA1) of the Mesoamerican genepool consisted of genotypes mainly collected from Amhara region and were small seeded. The second subpopulation (MA2) consisted of landraces mainly collected from Amhara and Oromia region. The third subpopulation (MA3) consisted of landraces mainly collected from Amhara and SNNPR. The fourth subpopulation (MA4) consisted of a single genotype, NUA-355 from CIAT. The fifth subpopulation (MA5) consisted of small- and medium-seeded improved varieties and breeding lines and landraces mainly from Amhara regions and few from Benshangul Gumuz and SNNPR regions.

To further confirm the genetic structure of the genotypes, principal component analysis was employed and the first two principal components explained 91% of the total variation (Figure S1). The genotypes were grouped into Andean and Mesoamerican genepools with some admixture.

The principal coordinate analysis (PCoA) revealed that the genotypes were clustered based on the population

structure result but the genotypes were not distinct due to the presence of admixture and gene flow among the different subpopulations particularly in the Mesoamerican genepool (Figure S2).

The net nucleotide distance among the six subpopulations generated based on population structure revealed that the maximum distance was recorded between population five and six, while the lowest distance was found between population three with population four and five (Table 6). The highest expected heterozygosity (0.28) was found in population two, while the lowest (0.01) was in population five. Similarly, mean fixation index among the populations ranged from 0.60 (Pop2) to 0.95 (Pop 5).

Discussion

SNP distribution and variation

Information on genetic diversity and relationship among common bean genotypes helps in the sustainable and efficient utilization of common bean genetic resources by breeding programs. Assessing the genetic diversity is very vital to get a better understanding of sources of novel alleles to be used in the crop improvement program. Various types of molecular markers have been widely used to estimate genetic diversity and population structure of common bean [43, 44]. Currently, DArTSeq-based SNP markers became a subject of choice because of high-throughput, high-genome coverage and are less costly [45].

Previously, studies were carried out to characterize the common bean genotypes in different countries and sources of germplasm from wild to domesticated using different molecular markers [46, 47]. However, few studies are available on the use of high-density DArTSeq SNP markers for diversity and population structure study in common bean [22, 48].

The distribution of the DArTSeq markers was almost uniform across the chromosomes with the largest number of SNPs recorded in chr_2 (1280 SNPs), followed by chr_8 (1272 SNPs), while the minimum SNPs were recorded

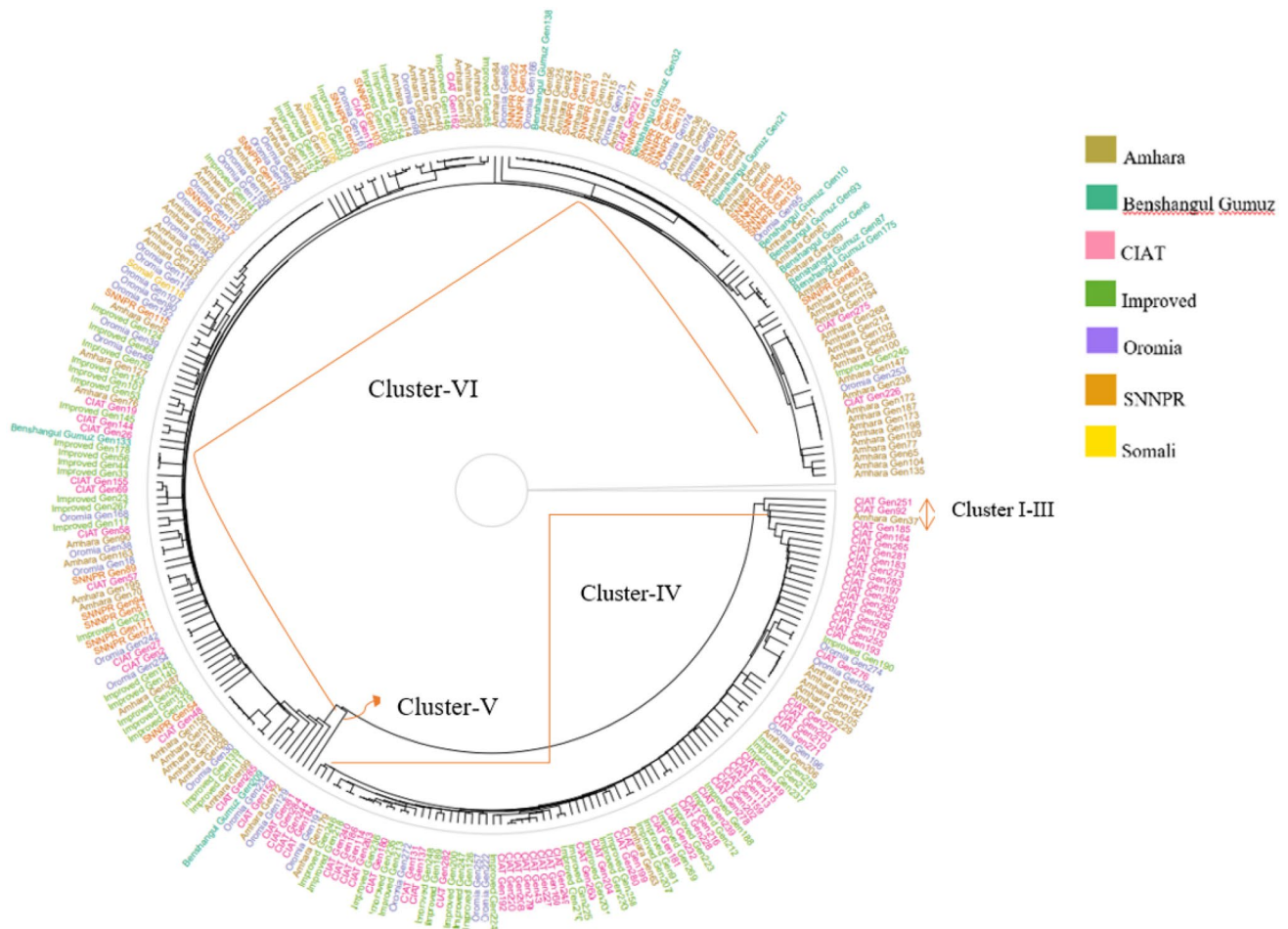


Fig. 4 Clustering of 289 common bean germplasm using Euclidean distance. The regions of collection and genotype are represented by different colors.

Table 5 Analysis of molecular variance for common bean genotypes based on geographic origin and breeding status

Source of variation	df	Sum of squares	Variance component	Percentage of variation	Fixation indices	<i>P</i> -value
Based on geographical regions						
Among populations	5	258,892.0	538.8 Va	23.53	FIS=0.83	FIS=0.001
Within populations	281	901,478.3	1457.5 Vb	63.67	FST=0.24	FST=0.001
Within individuals	287	84,100.5	293.0 Vc	12.80	FIT=0.87	FIT=0.001
Total	573	1,244,470.8	2289.4			
Based on breeding status						
Among populations	1	198,461	680.13 Va	27.06	FIS=0.84	FIS=0.001
Within populations	287	968,117	1540.49 Vb	61.3	FST=0.27	FST=0.001
Within individuals	289	84,463.5	292.26 Vc	11.63	FIT=0.88	FIT=0.001
Total	577	1,251,041	2512.88			

Df: Degree of freedom

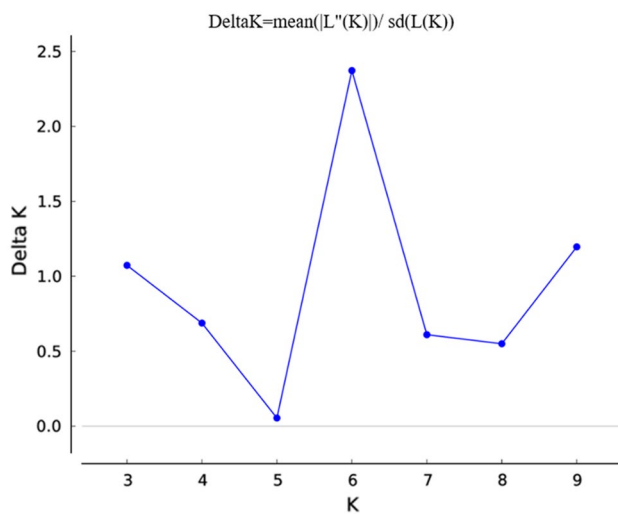


Fig. 5 Optimal K value to group the 289 common bean genotypes into subpopulations.

in chr_4 (835). Consistent to this study, [48–50] reported higher number of SNPs on chr_2.

Transition to transversion (Ts/Tv) ratio is one of the methods to check the quality of the SNP calls. The studied 289 common bean genome revealed that A/G and T/C were the most prevalent with a Ts/Tv ratio of 1.39:1 suggesting the transition mutations were more common than transversion mutation following methylation. The result is in agreement with [50] who reported a Ts/Tv ratio of 1.24 in 188 Brazilian common bean germplasm using 6286 SNP markers. Similarly, [22] reported a Ts/Tv ratio of 1.75 in 50 Italian common bean germplasm using 11,866 SNP markers.

In the present study, the overall mean genetic diversity, heterozygosity, MAF, and inbreeding coefficient were 0.38, 0.05, 0.28, and 0.85, respectively. This showed that the studied common bean genotypes have a wide diversity that can be exploited in the common bean improvement program. Contrary to this study [51] reported lower values for heterozygosity and genetic diversity after a study of 183 Turkish common bean accessions. In agreement with this study, [52] studied 60 common bean genotypes from eastern and southern Africa using 16,565 silicoDArT markers and reported overall genetic distance, observed heterozygosity and MAF of 0.28, 0.45, and 0.24, respectively.

Genetic diversity and relationship

The highest heterozygosity (6.0%) was observed in the improved varieties compared with landraces collected from different regions of Ethiopia. The result was further confirmed by the highest heterozygosity of materials sourced from CIAT and improved varieties with respective heterozygosity of 8 and 4.69% in the geographic origin-based classification. This indicated that the landraces were almost fixed to homozygosity compared with the improved varieties.

Polymorphic information content that measures the relevance of markers in a diversity study ranged from 0.06 to 0.38 with an average of 0.30. The maximum PIC in SNP markers is 0.50 due to its bi-allelic nature. Accordingly, the SNP markers used in this study were highly informative compared with other studies [49] and [48] who reported respective PIC values of 0.25 and 0.22 in common bean. The higher PIC value in this study could be due to the higher number of genotypes obtained from diverse sources and large number of SNP markers. For multi-allelic markers

Fig. 6 Population structure for 289 common bean genotypes in $K=6$. Each color represented a cluster. The arrow indicates the delineation of the genotypes based on regions of collection at $K=6$. BG: Benshangul Gumuz, CIAT: International Centre for Tropical Agriculture, SNNPR: Southern Nations Nationalities and People Region. The arrows demark the genotypes in each region.

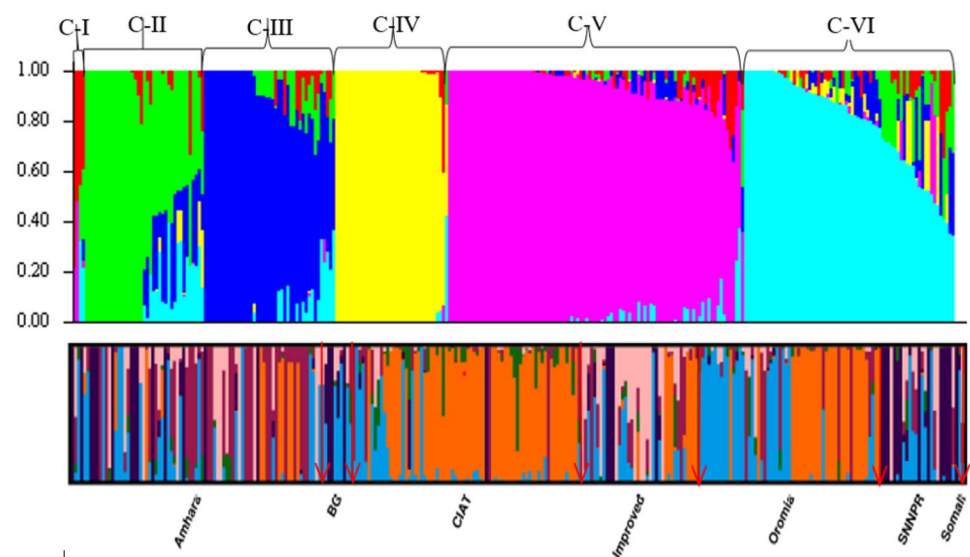


Table 6 Net nucleotide distance, expected heterozygosity and mean fixation index among the common bean populations

Population	Net nucleotide distance						Expected heterozygosity	% of membership
	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6		
Pop1							0.03	0.14
Pop2	0.37						0.28	0.19
Pop3	0.13	0.33					0.12	0.2
Pop4	0.14	0.34	0.1				0.12	0.28
Pop5	0.17	0.39	0.1	0.15			0.01	0.17
Pop6	0.48	0.48	0.4	0.43	0.5		0.08	0.03

Pop: population and each population numbering (1–5) showed MA1 to MA5, respectively, and Pop6 is A1

(e.g., microsatellites), PIC is regarded as slightly informative, moderately informative, and highly informative when the PIC value is <0.25 , $0.25–0.50$ and >0.50 , respectively [53]. The relatively lower PIC value of SNP markers could be compensated with their large number and wide-genome coverage.

Based on the breeding status, the highest genetic diversity and PIC were observed in the improved varieties compared with the landraces. From the landraces, those collected from Oromia and Amhara region showed higher genetic diversity and PIC. The higher genetic diversity of landraces collected from Oromia region indicated that landraces could have useful traits which pave the way for the collection of additional landraces in areas where there is an expected rich genetic diversity.

Pre-grouping of the genotypes showed that the highest genetic distance was observed between genotypes collected from CIAT and SNNPR (0.49). The CIAT genotypes were closer to improved varieties than all other regions of collection. This supports the fact that many of the improved varieties in Ethiopia and Africa are introduced from the CIAT collections and breeding lines and many of them have close pedigree [54, 55]. Clustering of the common bean genotypes revealed that they were not clustered based on their geographical origin and no clear explanation was given for the grouping and the result is in agreement with [6].

The AMOVA was conducted to display the variation among and within the $K=6$ subgroups and the result indicated that the greatest variability was observed within populations in both geographical region and breeding status-based classification followed by among populations.

Population structure

In the present study, the highest delta K value occurred at $K=6$ and grouped the 289 common bean genotypes into six subpopulations with a membership coefficient (q_i) ≥ 0.70 . At $K=2$, the genotypes were classified into Andean and Mesoamerican genepool. The classification was mostly based

on seed size in which Mesoamerican genotypes contained genotypes with small to medium seed size, while the Andean genotypes contained genotypes with medium to large seed size. At $K=6$, the Mesoamerican genepool further separated into five subclusters, while the Andean genepool remains the same as the K value was increasing, except four genotypes. A second peak was observed at $K=9$ which might be associated with the presence of different races within the Andean and Mesoamerican genepool. In agreement with this study, two peaks were reported in common bean at $K=3$ and $K=7$ [18, 28, 42, 43] have also reported two peaks. In addition, the membership coefficient revealed that there were admixtures that could be associated with the background of the genotypes used in this study as common bean genotypes might be hybridized for different traits.

In Ethiopia, the majority of the landraces and commercial cultivars were Mesoamerican type and are widely cultivated both for consumption and export, while the Andean genepool (large seeded) have relatively shorter period of cultivation history and the majority of them were introduced from CIAT as breeding lines and some are released after multi-location trial. In addition, the longtime cultivation and planting of mixed landraces in the same plot may result in the variation among the Mesoamerican genepool. The population structure detected by delta K varied with the number of populations, source of germplasm, and number and type of markers used. Accordingly, various studies reported different delta K values. In other studies, an optimal population structure at $K=6$ was reported [6, 56].

The eastern African region is considered as the secondary center of origin of common bean which could be the reason for such wide population structure [6]. In the present study, at $K=6$, the Mesoamerican genepool was separated into five subpopulations, while the Andean genepool remained unseparated. In agreement with this result, [44] reported the Mesoamerican genepool was separated into four subpopulations at $K=5$, while the Andeans remained unseparated. The results of population structure revealed that the Mesoamerican genepool was more diverse than the Andean genepool.

This could be due to large sample size of Mesoamerican genepool and long years of cultivation in Ethiopia under different cropping system.

Conclusions

The present study revealed the presence of wide genetic diversity and structured population in common bean. The average overall genetic diversity, PIC, MAF, and inbreeding coefficient were 0.38, 0.30, 0.28, and 0.85, respectively. Genetic diversity based on geographical locations showed that landraces collected from Oromia region showed the highest genetic diversity, PIC, and inbreeding coefficient.

Pre-grouping of the genotypes based on geographical origin showed that genotypes from SNNPR and CIAT were the most distant (0.492). Genotypes collected from CIAT were distant than genotypes collected from regions in Ethiopia. However, it is relatively close to the improved varieties due to sharing of parents in the improvement process. Clustering also showed that the genotypes were not clustered based on their origin, which could be due to transboundary movement of the germplasm across the regions. This showed that geographical regions were not the main drivers for differentiation and selection of the parents should be based on systematic assessment of the diversity. The population structure revealed that the optimal number of subpopulation was six ($K=6$). The results showed that common bean collections revealed a wide range of diversity. At $K=2$, the genotypes were clustered into Andean and Mesoamerican genepool. Further separation of the genotypes was due to the diversity, large sample size, and long years of cultivation of the Mesoamerican genepool, while the Andeans remained unseparated. The genotypes were not clustered based on geographical regions and they were not the main drivers for the differentiation and. This indicated that selection of the parental lines should be based on systematic assessment of the diversity rather than geographical distance.

The present investigation proved the potential of highly informative DArTSeq-derived SNP markers for population structure and genetic diversity studies. From this study, a core set of diagnostic SNP markers can be built and used in the genetic diversity and population structure of common bean for association studies, designing effective collection and conservation for efficient utilization for the improvement of the crop.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11033-023-08498-y>.

Acknowledgements This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation [OPP1198373] through the AVISA project. The authors are thankful to Integrated Genotyping

Service and Support(IGSS), BeCA, ILRI for their partial support for genotyping.

Author contributions YMG, JE, KO, BA, and SO contributed to conceptualization; YMG, BA, and JE contributed to methodology; YMG, JE, and EAE contributed to software; YMG and JE contributed to investigation; BA and CM contributed to resources; YMG contributed to data curation, writing—original draft preparation, visualization, and formal analysis; JE, KO, SO, BA, and CM contributed to writing—review and editing; JE, KO, BA, and SO contributed to supervision; and BA and CM contributed to funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding The authors have not disclosed any funding.

Declarations

Conflicts of interest The authors declare no conflict of interest.

Ethical approval This article does not contain any studies with human participants and animals and ethical approval is not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bennett MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. *Ann Bot* 76:113–176
- Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Ann Bot* 95(1):45–90
- Reckling M, Bergkvist G, Watson CA, Stoddard FL, Zander PM, Walker RL et al (2016) Trade-offs between economic and environmental impacts of introducing legumes into cropping systems. *Front Plant Sci* 7(669):1–15
- Graham PH, Vance CP (2003) Update on legume utilization legumes : importance and constraints to greater use. *Plant Physiol* 131:872–877
- Foyer CH, Lam HM, Nguyen HT, Siddique KHM, Varshney RK, Colmer TD et al (2016) Neglecting legumes has compromised human health and sustainable food production. *Nat Plants* 2(8):1–10
- Asfaw A, Blair MW, 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):1–12
- Fenta BA, Negash K, Shiferaw T, Tumsa K (2018) Progress of common bean breeding and genetics research in Ethiopia. *Ethiop J Crop Sci Spec Issue* 6(3):1–15
- Federal Democratic Republic of Ethiopia (2016) Federal Democratic Republic of Ethiopia Sekota Declaration Implementation Plan (2016–2030). Addis Ababa, Ethiopia, pp 1–19

9. FAO (2023) Food and Agriculture Organisation of the United Nations). 2021. Statistics division (FAOSTAT). Crops and livestock products, Rome
10. Ruelle ML, Asfaw Z, Dejen A, Tewolde-Berhan S, Nebiyu A, Tana T et al (2019) Inter- and intraspecific diversity of food legumes among households and communities in Ethiopia. *PLoS ONE* 14(12):1–19
11. Velasquez VLB, Gepts P (1994) RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. *Geome* 37:256–263
12. Boczkowska M, Bulin Z, Nowosielski J (2012) AFLP analysis of genetic diversity in five accessions of Polish runner bean (*Phaseolus coccineus* L.). *Genet Resour Crop Evol* 59:473–478
13. De AFA, Simões L, Gonçalves A, Miglioranza É (2016) AFLP analysis of genetic diversity in determinate and indeterminate snap bean accessions. *Acta Scientiarum* 38(1):29–34
14. Mavromatis AG, Arvanitoyannis IS, Korkovelos AE, Giakountis A, Chatzitheodorou VA, Goulas CK (2010) Genetic diversity among common bean (*Phaseolus vulgaris* L.) Greek landraces and commercial cultivars: nutritional components, RAPD and morphological markers. *Span J Agric Res* 8(4):986
15. Ligarreto GA, Orlando MW (2014) Identification of the variability of a common bean collection through morphological, physiological, biochemical, and molecular relationships. *Agronomía Colombiana* 32(2):159–169
16. Kumar A, Singh P, Rai N, Bhaskar G, Datta D (2014) Genetic diversity of French bean (*Phaseolus vulgaris* L.) genotypes on the basis of morphological traits and molecular markers. *Indian J Biotechnol (IJBT)* 13(2):207–213
17. Scarano D, Rubio F, Ruiz JJ, Rao R, Corrado G (2014) Morphological and genetic diversity among and within common bean (*Phaseolus vulgaris* L.) landraces from the Campania region (Southern Italy). *Sci Hortic* 180:72–78. <https://doi.org/10.1007/s11033-023-08498-y>
18. Valentini G, Gonçalves-Vidigal MC, Elias JCF, Moiana LD, Mindo NNA (2018) Population structure and genetic diversity of common bean accessions from Brazil. *Plant Mol Biol Rep* 36(5–6):897–906
19. Cabral PDS, de Souza LC, da Costa GF, Silva FHL, Soares TCB (2018) Investigation of the genetic diversity of common bean (*Phaseolus vulgaris*) cultivars using molecular markers. *Genet Mol Res* 17(4):1–11
20. Alam M, Neal J, O'Connor K, Kilian A, Topp B (2018) Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. *PLoS ONE* 13(8):1–20
21. Campa A, Murube E (2018) Disequilibrium in a Spanish common bean diversity panel revealed through genotyping-by-sequencing. *Genes* 9:518
22. Lioi L, Zuluaga DL, Pavan S, Sonnante G (2019) Genotyping-by-sequencing reveals molecular genetic diversity in Italian common bean landraces. *Diversity* 11(9):1–14
23. Raatz B, Mukankusi C, Lobaton JD, Male A, Chisale V, Amsalu B et al (2019) Analyses of African common bean (*Phaseolus vulgaris* L.) germplasm using a SNP fingerprinting platform: diversity, quality control and molecular breeding. *Genet Resour Crop Evol* 66(3):707–722
24. Tigist SG, Melis R, Sibiya J, Amelework AB, Keneni G, Tegene A (2019) Population structure and genome-wide association analysis of bruchid resistance in Ethiopian common bean genotypes. *Crop Sci* 59(4):1504–1515
25. Kefyalew D, Teklehaimanot H, Tileye F (2014) Genetic diversity study of common bean (*Phaseolus vulgaris* L.) germplasm from Ethiopia using inter simple sequence repeat (ISSR) markers. *Afr J Biotechnol* 13(36):3638–3649
26. Fisseha Z, Kyallo M, Tesfaye K, Harvey J, Dagne K, Opyio S et al (2018) Integrating phenotypic evaluation with a molecular diversity assessment of an Ethiopian collection of common bean landraces. *Afr Crop Sci J* 26(2):315–326
27. Girma G, Nida H, Tirfessa A, Lule D, Bejiga T, Seyoum A et al (2020) A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits. *Plant Genome* 13:e20055
28. Obeng-bio E, Badu-apraku B, Ifie BE, Danquah A, Blay ET, Dadzie MA et al (2020) Genetic diversity among early provitamin A quality protein maize inbred lines and the performance of derived hybrids under contrasting nitrogen environments. *BMC Genet* 21(78):1–13
29. Stanley A, Menkir A, Paterne A, Ifie B, Tongoona P, Unachukwu N, Meseke S, Mengesha W, Gedil M (2020) Genetic diversity and population structure of maize inbred lines with varying levels of resistance to striga hermonthica using agronomic trait-based and SNP markers. *Plants* 9(1223):2–18
30. Alemu A, Feyissa T, Letta T, Abeyo B (2020) Genetic diversity and population structure analysis based on the high density SNP markers in Ethiopian durum wheat (*Triticum turgidum* ssp. durum). *BMC Genet* 21(18):1–12
31. Seo E, Kim K, Jun T, Choi J, Kim S, Sun H et al (2020) Population structure and genetic diversity in Korean cowpea germplasm based on SNP markers. *Plants* 9:1190
32. Verma S, Gupta S, Bandhiwal N, Kumar T, Bharadwaj C, Bhatia S (2015) High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arjetinum* L.) using Genotyping-by-Sequencing (GBS). *Nat Publ Group* 5(17512):1–14
33. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5):1–10
34. Killian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. In: Bonin FP, A Bonin (eds) *Data production and analysis in population genomics: methods and protocols, methods in molecular biology*. Springer, New York, 67–89.
35. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959
36. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE : a simulation study. *Mol Ecol* 14:2611–2620
37. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801–1806
38. Rosenberg NA (2004) DISTRICT: a program for the graphical display of population structure. *Mol Breed* 4:137–138. <https://doi.org/10.1007/s11033-023-08498-y>
39. 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
40. Liu K, Muse SV (2005) PowerMaker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9):2128–2129
41. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567
42. Kuzay S, Hamilton-Conaty P, Palkovic A, Gepts P (2020) Is the USDA core collection of common bean representative of genetic diversity of the species, as assessed by SNP diversity? *Crop Sci* 60:1398–1414
43. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J et al (2014) A reference genome for common bean

- and genome-wide analysis of dual domestications. *Nat Genet* 46(7):707–713
44. Blair MW, Brondani RVP, Díaz LM, Del Peloso MJ (2013) Diversity and population structure of common bean from Brazil. *Crop Sci* 53(5):1983–1993
 45. Mercati F, Leone M, Lupini A, Sorgonà A, Bacchi M, Abenavoli MR et al (2013) Genetic diversity and population structure of a common bean (*Phaseolus vulgaris* L.) collection from Calabria (Italy). *Genet Resour Crop Evol* 60(3):839–852
 46. Jehan T, Lakhanpaul S (2006) Single nucleotide polymorphism (SNP) - methods and applications in plant genetics: a review. *Indian J Biotechnol* 5(4):435–459
 47. Gill-langarica HR, Muruaga-martínez JS, Vargas-vázquez MLP, Experimental C (2011) Genetic diversity analysis of common beans based on molecular markers. *Genet Mol Biol* 34(4):595–605
 48. Avila T, Blair MW, Reyes X, Bertin P (2012) Genetic diversity of bean (*Phaseolus*) landraces and wild relatives from the primary centre of origin of the Southern Andes. *Plant Genet Resour* 10(1):83–92
 49. Nkhata W, Shimelis H, Melis R, Chirwa R, Mzengeza T, Mathew I et al (2021) Genome-wide association analysis of bean fly resistance and agro-morphological traits in common bean. *PLoS ONE* 16(4):1–24
 50. Nemli S, Kaygisiz Aşçıoğul T, Ateş D, Eşiyok D, Tanyolaç MB (2017) Diversity and genetic analysis through DArTseq in common bean (*Phaseolus vulgaris* L.) germplasm from Turkey. *Turk J Agric For* 41(5):389–404
 51. Nadeem MA, Habyarimana E, Çiftçi V, Nawaz MA, Karaköy T, Comertpay G et al (2018) Characterization of genetic diversity in Turkish common bean gene pool using phenotypic and whole-genome DArTseq-generated silicoDArT marker information. *PLoS ONE* 13(10):1–28
 52. Valdisser PAMR, Pereira WJ, Filho JEA, Müller BSF, Coelho GRC, De MIPP et al (2017) In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics* 18(423):1–19
 53. Wilson Nkhata HS, 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):1–23
 54. Eltahir S, Sallam A, Belamkar V, Emara HA (2018) Genetic diversity and population structure of F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. *Front Plant Sci* 9(76):1–9
 55. Andriatsitohaina MR, Nagadya R, Okii D, Obilil I, Mukankusi CM, Chirwa R, Zulu RM, Lung'aho MG, Ruraduma C, Ugen MA, Kidane T, Karanja D, Mazuma E, Musoni A, Sefume L, Meshac M (2017) Common Bean variety releases in Africa, vol 2. Harvard Dataverse
 56. Luo Z, Brock J, Dyer JM, Kutchan T, Schachtman D, Augustin M et al (2019) Genetic diversity and population structure of a *Camelina sativa* spring panel. *Front Plant Sci* 10:1–12

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