

RESEARCH ARTICLE

Genetic differentiation of a southern Africa tepary bean (*Phaseolus acutifolius* A Gray) germplasm collection using high-density DArTseq SNP markersSaul Eric Mwale^{1,2*}, Hussein Shimelis¹, Wilfred Abincha³, Wilson Nkhata^{1,4}, Abel Sefasi⁵, Jacob Mashilo¹

1 School of Agricultural, Earth and Environmental Sciences, African Centre for Crop Improvement (ACCI), University of KwaZulu-Natal, Pietermaritzburg, South Africa, **2** Biological Sciences Department, The African Centre of Excellence in Neglected and Underutilized Biodiversity (ACENUB), Mzuzu University, Luwingu, Mzuzu, Malawi, **3** Kenya Agricultural and Livestock Research Organization (KALRO), Non-Ruminant Research Institute, Kakamega, Kenya, **4** Alliance of Bioversity International Institute of Tropical Agriculture (CIAT), Chitedze Agricultural Station, Lilongwe, Malawi, **5** Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi

* mwale.s@mzuni.ac.mw

OPEN ACCESS

Citation: Mwale SE, Shimelis H, Abincha W, Nkhata W, Sefasi A, Mashilo J (2023) Genetic differentiation of a southern Africa tepary bean (*Phaseolus acutifolius* A Gray) germplasm collection using high-density DArTseq SNP markers. PLoS ONE 18(12): e0295773. <https://doi.org/10.1371/journal.pone.0295773>

Editor: Aditya Pratap, ICAR-Indian Institute of Pulses Research, INDIA

Received: August 31, 2023

Accepted: November 29, 2023

Published: December 14, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0295773>

Copyright: © 2023 Mwale et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information](#) files.

Abstract

Genetic resources of tepary bean (*Phaseolus acutifolius* A. Gray) germplasm collections are not well characterized due to a lack of dedicated genomic resources. There is a need to assemble genomic resources specific to tepary bean for germplasm characterization, heterotic grouping, and breeding. Therefore, the objectives of this study were to deduce the genetic groups in tepary bean germplasm collection using high-density Diversity Array Technology (DArT) based single nucleotide polymorphism (SNP) markers and select contrasting genotypes for breeding. Seventy-eight tepary bean accessions were genotyped using 10527 SNPs markers, and genetic parameters were estimated. Population structure was delineated using principal component and admixture analyses. A mean polymorphic information content (PIC) of 0.27 was recorded, indicating a relatively low genetic resolution of the developed SNPs markers. Low genetic variation (with a genetic distance [GD] = 0.32) existed in the assessed tepary bean germplasm collection. Population structure analysis identified five sub-populations through sparse non-negative matrix factorization (snmf) with high admixtures. Analysis of molecular variance indicated high genetic differentiation within populations (61.88%) and low between populations (38.12%), indicating high gene exchange. The five sub-populations exhibited variable fixation index (F_{ST}). The following genetically distant accessions were selected: Cluster 1: Tars-Tep 112, Tars-Tep 10, Tars-Tep 23, Tars-Tep-86, Tars-Tep-83, and Tars-Tep 85; Cluster 3: G40022, Tars-Tep-93, and Tars-Tep-100; Cluster 5: Zimbabwe landrace, G40017, G40143, and G40150. The distantly related and contrasting accessions are useful to initiate crosses to enhance genetic variation and for the selection of economic traits in tepary bean.

Funding: Yes. Kirkhouse Trust SCIO provided the funding for the genotyping of tepary bean germplasm.

Competing interests: No The authors have declared that no competing interests exist.

Introduction

Tepary bean (*Phaseolus acutifolius* A. Gray) is a climate-smart legume crop that provides food and nutrition security in arid and semi-arid regions of the world [1]. It is a self-pollinating diploid ($2n = 2x = 22$) species with a genome size of approximately 647 million base pairs (Mbp) [2]. The grains are a source of proteins, lipids and essential mineral elements [3]. The crop is tolerant to drought and heat stress [1,4,5], allowing its cultivation in dry environments. It is highly resistant to diseases including bacterial blight [*Xanthomonas campestris* pv. *Phaseoli*] [6], *Fusarium* wilt (*Fusarium oxysporum*) [7], and bean golden mosaic virus [8]. As a result, the crop served as a useful gene donor for disease-resistance breeding in common bean [9,10].

In Southern Africa, tepary bean is mostly cultivated by smallholder farmers in Botswana, South Africa, Zimbabwe, Zambia, and Malawi using genetically unimproved landrace varieties [11,12]. This is attributed to limited breeding efforts to develop improved varieties with farmer, consumer and market-preferred traits [13]. The crop yield in the region is approximately 500 kilograms per hectare [14,15] compared to the potential yield of greater than 2000 kilograms per hectare [16]. To address the imminent challenge of low productivity in Southern Africa, the University of KwaZulu-Natal's African Center for Crop Improvement (ACCI) acquired a diverse germplasm panel of tepary bean from the International Centre for Tropical Agriculture (CIAT-Columbia) [17]. The collected genetic resources of tepary bean are yet to be explored using dedicated genomic resources. Reportedly, genetic analysis using various molecular markers developed for common bean (*Phaseolus vulgaris* L.) such as simple sequence repeat (SSR), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and sequenced characterized amplified region (SCAR) revealed a low to moderate genetic variation when applied in tepary bean [18–25]. There is a need to assemble genomic resources specific to tepary bean for germplasm characterization, selection, and breeding.

The reference genome to which the molecular marker covers the genome influences germplasm characterization accuracy and predictability [26]. Single nucleotide polymorphism (SNP) markers are widely abundant across the genome and quickly unravel diversity; hence, they are the most preferred for diversity and association mapping studies [27,28]. Gene-based SNP markers, which mainly uncover functional variation, were used to assess the population structure and genetic diversity of wild and cultivated tepary bean genotypes using the common bean reference genome [26]. However, the genotyping assay was limited in capturing several loci (< 60%) in tepary bean since it was aligned to the common bean reference genome [26]. A draft tepary bean reference genome was developed recently, providing enormous opportunities for developing tepary bean genomic resources [1]. Genotyping-by-sequencing (GBS) is one of the most cost-effective approaches for concurrent SNPs identification and genotyping [29]. Diversity Array Technology Sequencing (DArTSeq) is a GBS platform that combines the principles of genome complexity reduction methods with high throughput sequencing. It allows the simultaneous identification of numerous SNPs across a genome at an affordable cost [30]. DArTseq SNP markers have been used for quantitative trait loci (QTL) mapping, genome-wide association studies, the development of linkage maps, genetic diversity, and population structure analyses in several legume crops, including pigeonpea, chickpea, soybean, and common bean [30–34]. Genetic diversity and population structures were examined in Interspecific Mesoamerican X Wild Tepary (IMAWT) population using GBS with high-density SNP markers aligned to the common bean reference genome [5]. Further, GBS with high-density SNP markers mapped to the tepary bean reference genome in a tepary bean diversity panel has been used to determine the extent of genetic diversity and population structure [35]. Nevertheless, the southern Africa tepary bean diversity panel collection, comprising landraces,

released lines, and breeding lines, has not yet been characterized with high-density DArTSeq SNP markers matched to the tepary bean reference genome. Genetic analysis of the southern Africa tepary bean germplasm collection could potentially capture a broader range of genetic variation and provide useful information for heterotic grouping, genome-wide association mapping, marker-assisted selection for precision, and speed breeding. Therefore, the objectives of this study were to deduce the genetic groups in tepary bean germplasm collection using high-density DArT-based single nucleotide polymorphism markers and select contrasting genotypes for breeding.

Materials and methods

Plant genetic materials

A panel of 78 tepary bean germplasm collection comprising released varieties, breeding lines, and landraces were used for the study. The accessions were sourced from the International Center for Tropical Agriculture (CIAT)/Colombia, the United States Department of Agriculture (USDA), and farmers in South Africa and Zimbabwe. The names of the genotypes and their sources of origin are presented in “Table 1”.

DNA extraction and SNP genotyping

The above tepary bean germplasm were genotyped at SEQART Africa in the International Livestock Research Institute (1.2693° S, 36.7216° E) in Nairobi. Genomic DNA was extracted using the TANBEAD Plant Extraction Kit (Diagnocine, Hackensack, NJ, USA). The genomic DNA extracted was in the range of 50–100 ng/ul. DNA quality and quantity were checked on 0.8% Agarose Gel. The genomic DNA was subjected to restriction digestion using MstI and PstI as rare and frequent cutters, respectively. Ligation of the digested DNA fragments was accomplished by both common and barcode adapters. This was followed by selective amplification via polymerase chain reaction (PCR) of the adapter-ligated fragments. Pooling and purification of the PCR products were accomplished through a QIAquick PCR purification kit (QIAGEN GmbH, Hilden, Germany). An Illumina HiSeq 2500 (Macrogen, Seoul, Korea) that utilizes single reads was used to sequence the purified PCR products. DNA libraries were constructed according to Kilian et al. [36]. DArTseq marker scoring was achieved using DArTsoft 14, an in-house marker scoring pipeline based on algorithms [37–39]. Two types of DArTseq markers were scored: silica-DArT markers and SNP markers, which were both scored as binary for the presence (1) or absence (0) of the restriction fragment with the marker sequence in the genomic representation of the sample. 11,318 SNP markers were aligned to 11 chromosomes of the reference genome of *P. acutifolius_580_v1.0* to identify chromosome positions [1]. The markers used in the current study were highly reproducible, with polymorphic information content (PIC) that varied from 0.01 to 0.05 and a mean call rate of 0.93 ranging from 0.81 to 1.00.

SNP quality control. Quality control was implemented for the SNP data using the *raw.data* function of the *snpReady* package [40] in R statistical software version 4.3.1 [41]. SNP markers were filtered by eliminating markers with no chromosomal position and a minor allele frequency (MAF) of <5%. The percent similarity among identical markers was assessed, and markers with the highest missing data were eliminated. A total of 10527 markers with a SNP call rate greater than 95% were retained and used for genetic and population structure analysis. The *MVP.report* function of *rMVP* package in R was used to develop SNP density plot of the filtered SNPs [42]. Previously this analysis was done using *CMplot* package which has now been integrated in *rMVP* package.

Table 1. Names and origins of tepary bean germplasm collection used in the study.

Genotype Code	Genotype name or designation	Origin	Genotype Code	Genotype name or designation	Origin	Genotype Code	Genotype name or designation	Origin
G1	G40001	Mexico	G27	G40129	Mexico	G55	TARS-TEP-32	USDA
G2	G40005	El Salvador	G28	G40132	Mexico	G56	TARS-TEP-49A	USDA
G3	G40013	Nicaragua	G29	G40133	Mexico	G57	TARS-TEP-49B	USDA
G4	G40014	Nicaragua	G30	G40134	Mexico	G58	TARS-TEP-51	USDA
G5	G40017	El Salvador	G31	G40135	Mexico	G59	TARS-TEP-52	USDA
G6	G40019	Mexico	G32	G40136	Mexico	G60	TARS-TEP-54	USDA
G7	G40020	Mexico	G33	G40137	Mexico	G61	TARS-TEP-58A	USDA
G8	G40022	USA	G34	G40138	Mexico	G62	TARS-TEP-58B	USDA
G9	G40023	USA	G35	G40139	Mexico	G63	TARS-TEP-60	USDA
G10	G40031	Mexico	G36	G40140	Mexico	G64	TARS-TEP-64	USDA
G11	G40032	Guatemala	G37	G40143	Mexico	G65	TARS-TEP-73	USDA
G12	G40033	Mexico	G38	G40144A	Mexico	G66	TARS-TEP-77	USDA
G13	G40035	Mexico	G39	G40145	Mexico	G67	TARS-TEP-83	USDA
G14	G40036	Mexico	G40	G40147	Mexico	G68	TARS-TEP-85	USDA
G15	G40042	USA	G41	G40148	Mexico	G69	TARS-TEP-86	USDA
G16	G40059	El Salvador	G42	G40150	Mexico	G71	TARS-TEP 23	USDA
G17	G40062	Nicaragua	G43	G40157	Mexico	G72	PI-310801	Nicaragua
G18	G40063	USA	G44	G40158	Mexico	G73	G40119	Mexico
G19	G40065	USA	G45	Zimbabwe landrace	Zimbabwe	G74	PI-440786	USDA
G20	G40066A	USA	G46	G40173A	Mexico	G75	G40200	Costa Rica
G21	G40068	USA	G48	G40201	Costa Rica	G79	TARS-TEP97	USDA
G22	G40069	USA	G49	G40237	Mexico	G80	TARS-TEP112	USDA
G23	G40084	Mexico	G50	Uchokwane	South Africa	G81	TARS-TEP101	USDA
G24	G40111	Mexico	G51	SONORA	Sonora	G83	TARS-TEP93	USDA
G25	G40125	Mexico	G53	TARS-TEP-10	USDA	G84	TARS-TEP51	USDA
G26	G40127	Mexico	G54	TARS-TEP-22	USDA	G85	TARS-TEP100	USDA

USDA = United States Department of Agriculture.

<https://doi.org/10.1371/journal.pone.0295773.t001>

Genetic diversity and population structure analysis

Genetic diversity analyses were conducted to determine the polymorphic information content (PIC), minor allele frequency (MAF), observed heterozygosity (H_o), genetic diversity (GD), fixation index, additive variance and dominance variance. The genetic parameters were computed using the *snpReady* package [40] utilizing the *popgen* function in R statistical software [41]. A Landscape and Ecological Associations (LEA) R package was used to determine the population genetic structure [43–45]. Using *snmf* function of LEA, ancestry coefficients were estimated where Cross-entropy criteria was used to identify the optimum number of K populations where $K = 1$ to 10 with six repetitions [46,47]. Principal component analysis was implemented in LEA package using *pca* function to validate the identified subpopulation through the elbow method [48]. The individuals in the subpopulations were identified using the developed Q matrix in R statistical software [41]. Admixture plot of individuals were plotted using the

developed Q matrix using *plot* function of base R. The results obtained from population analysis were subjected to Principal Coordinate Analysis (PCoA) and analysis of molecular variance (AMOVA) using the poppr package [49] in R statistical software [41]. Neighbor joining hierarchical cluster analysis was performed in R statistical software using a dendextend package [41].

Results

Marker characterization

Genetic parameters derived by SNP markers are presented in “Table 2”. PIC ranged from 0.10 to 0.40, with a mean of 0.27. Most SNP markers were found within the following PIC categories; 0.2 to 0.03 followed by 0.3–0.4 and the least was 0–0.1 (“Fig 1”). The mean H_o and H_e were 0.30 and 0.32, respectively. The mean V_a and V_d were 3415.19 and 1227.16, respectively. Markers revealed a mean effective population size (N_e) of 488.28. The majority of the SNP markers had minor allele frequencies ranging from 0.1 to 0.2 and the least MAF ranging from 0.4 to 0.5. The SNP markers that explained the most genetic diversity were within the category 0.2 to 0.4 and the least informative were between 0.0 to 0.1 (“Fig 1”). The SNP markers were widely distributed across the 11 chromosomes with a SNP density ranging from 0 to > 65 (“Fig 2”).

Population structure of the assessed tepary bean germplasm

Population structure analyses based on cross-entropy criteria revealed that the population was optimally structured into 5 sub-populations ($K = 5$) (“Fig 3”). The result was further validated by observing the principal component screen plot, which showed five principal components (PCs). Results from admixture analysis revealed the presence of genetic admixture across the five subpopulations (“Fig 4”).

The mean coefficient of inbreeding (F) for the five subpopulations was 0.08 and variable among the populations (“Fig 5”). Further, the extent of genetic divergence was highest between subpopulations 1 and 5 (0.23), subpopulations 1 and 2 (0.21), subpopulations 3 and 5, and moderate for subpopulations 1 and 3 (0.08) (“Table 3”).

Analysis of molecular variance

Analysis of molecular variance revealed a 38.12% variation between the population and 61.87% within population (“Table 4”). This was further visualized using principal coordinate

Table 2. Genetic parameters computed from genetic diversity assessment of tepary bean germplasm using high-density SNP markers.

Genetic parameter	Overall Mean	Minimum	Maximum
GD	0.32	0.10	0.50
PIC	0.27	0.09	0.38
MAF	0.22	0.05	0.5
H_o	0.30	0.06	0.81
H_e	0.32	0.10	0.50
V_a	3415.19		
V_d	1227.16		
N_e	488.28		

GD = Genetic Diversity; PIC = Polymorphic information content; MAF = Minor allele frequency; H_o = observed heterozygosity; H_e = Expected heterozygosity; V_a = additive variance; V_d = dominance variance; N_e = effective population size.

<https://doi.org/10.1371/journal.pone.0295773.t002>

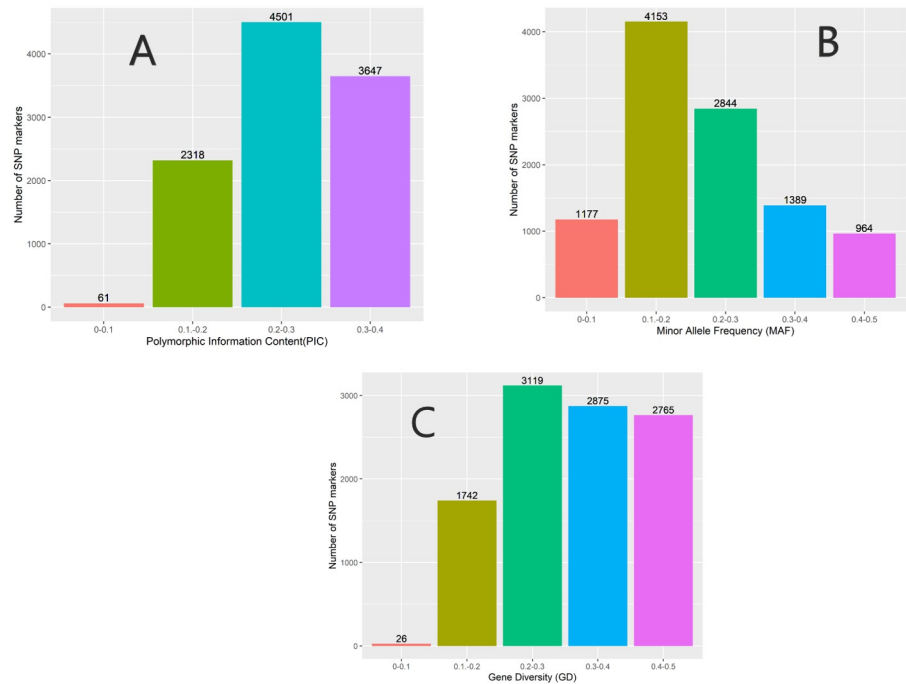


Fig 1. Single nucleotide polymorphism markers characteristics: A = Polymorphic Information Content (PIC), B = Minor allele frequency (MAF) and C = Gene diversity (GD).

<https://doi.org/10.1371/journal.pone.0295773.g001>

analysis (PCoA), which showed that clusters 1, 3, and 4 showed genetic lineage, thus making between-population variance less than within-population variance (“Fig 6”).

Genetic grouping of tepary bean

Neighbor-joining hierarchical clustering assorted the population into five clusters with different numbers of genotypes as follows: Cluster 1 (24 genotypes), Cluster 2 (8 genotypes), Cluster 3 (13 genotypes), Cluster 4 (11 genotypes), and Cluster 5 (22 genotypes).

The landraces G50 and G45 were clustered with released varieties and breeding lines in Clusters 1 and 5, respectively (“Fig 7”). Most genotypes sourced from Mexico were grouped in Clusters 2, 3, 4 and 5. Similarly, genotypes from USDA were primarily grouped in Cluster 1 (“Fig 7”).

Discussion

A well-characterized genetic resource is useful for breeding and genetic analysis. Genetic analysis using molecular markers complements phenotypic selection to develop heterotic groups and genotype selection for breeding. Genetic gains for yield, grain quality, and biotic and abiotic stress tolerance depend on the magnitude of the response to selection. However, limited genomic resources are developed for tepary bean for genetic differentiation and selection. Previously, germplasm characterization in this crop has been based on the use of molecular markers developed for common bean [19,26]. Therefore, the present study aimed at deducing the genetic groups in tepary bean germplasm collection using high-density DArT-based single nucleotide polymorphism (SNPs) markers and selecting contrasting genotypes for breeding.

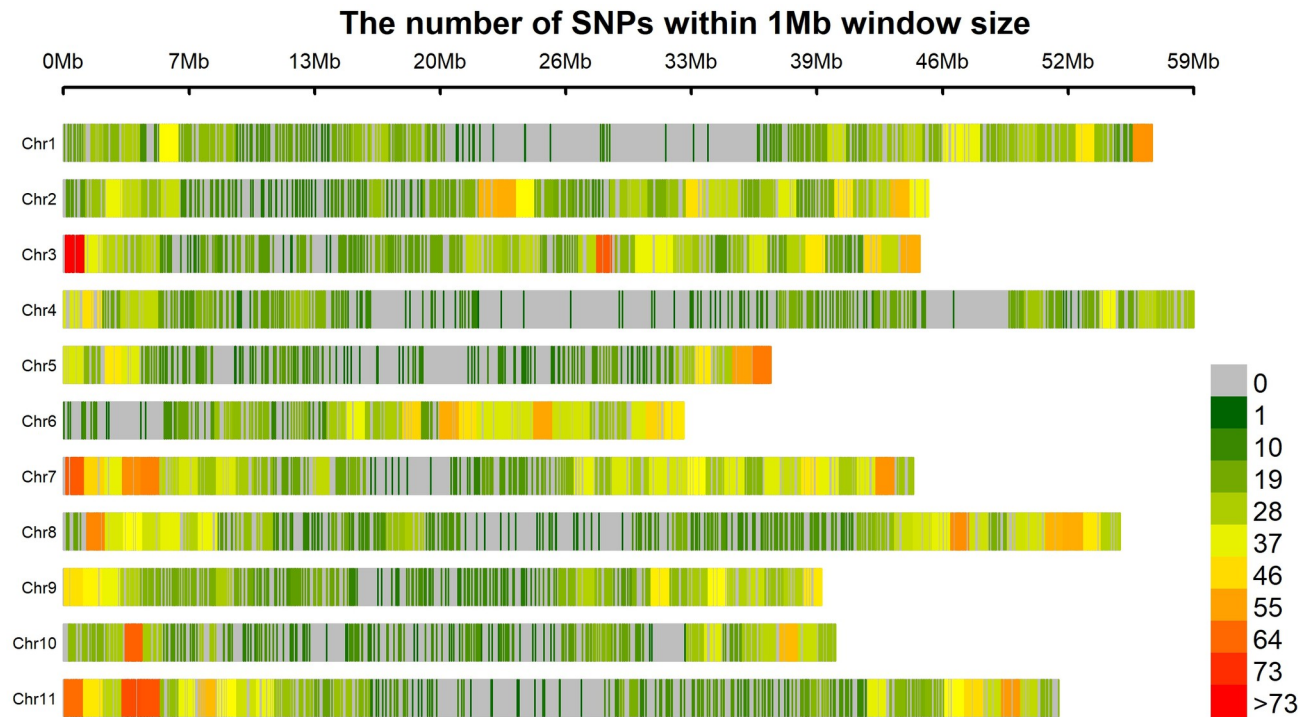


Fig 2. Single nucleotide polymorphism density plot mapped in 78 tepary bean genotypes.

<https://doi.org/10.1371/journal.pone.0295773.g002>

In the present study, the least number of the derived SNPs markers recorded PIC values ranging between 0 to 0.2 and were considered less informative (“Fig 1”). The low PIC value could be attributed to the bi-allelic nature of the SNP markers, which limits them to values of less than or equal to 0.5 [50,51]. The majority of the developed SNPs markers recorded PIC values ranging from 0.2 to 0.3, and 0.3 to 0.5 (“Fig 1”). These SNPs markers are considered informative. The SNPs markers revealed moderate genetic diversity in the studied tepary bean germplasm (“Table 2”). The AMOVA (“Table 4”) indicated fewer variations between populations than within the population and supplemented the moderate genetic diversity.

The moderate genetic divergence typical of cultivated tepary bean in our study corroborates with the study findings of Blair et al. [18], Gujaria-Verma et al. [26], and Mhlaba et al. [19]. This is attributed to localized domestication and distribution of the crop in the USA and

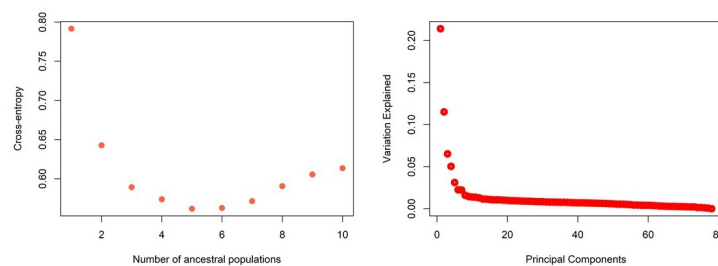


Fig 3. Tepary bean population structure based on cross-entropy (left) and principal component analysis (right) employing 10527 high-density SNP markers.

<https://doi.org/10.1371/journal.pone.0295773.g003>

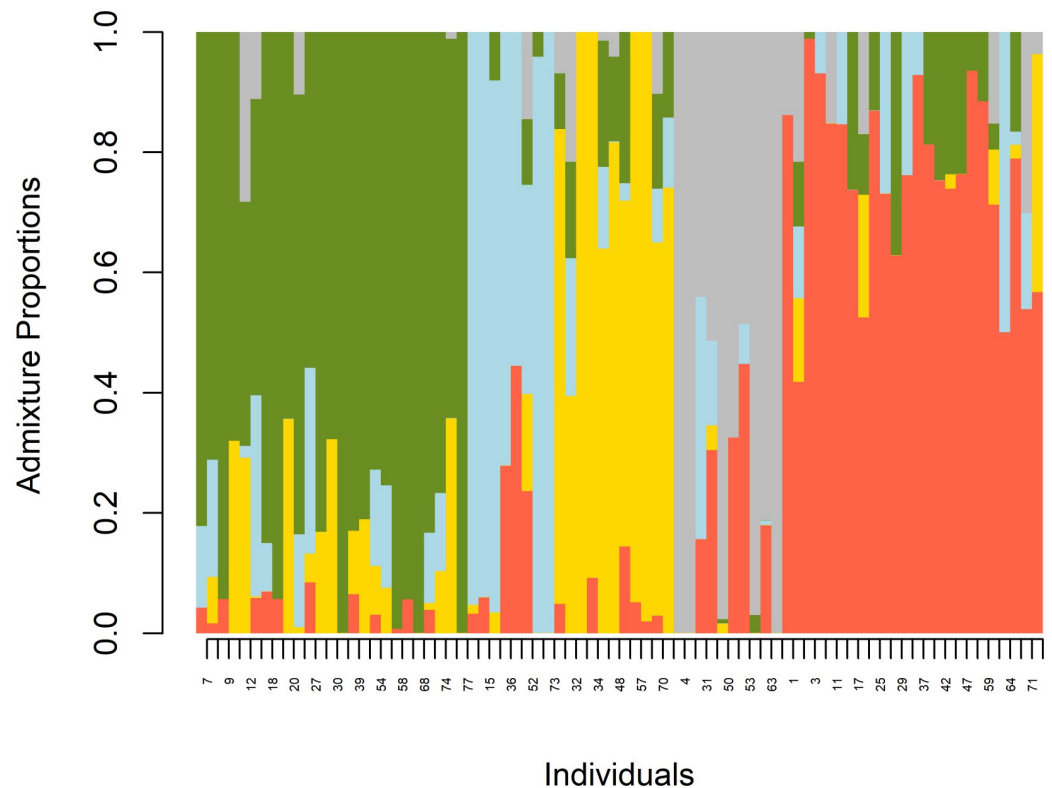


Fig 4. Population admixtures in 78 tepary bean genotypes using 10527 high-density SNP markers.

<https://doi.org/10.1371/journal.pone.0295773.g004>

Mexico, which culminated in a narrow genetic base. The low natural hybridization and outcrossing rates in tepary bean could also be attributed to its moderate genetic diversity [18,19]. However, the existing genetic diversity within the studied population (“Table 4”) implies that sufficient genetic variation is present for crop improvement. Moghaddam et al. [1] also reported high within-population variation in tepary bean, concurring with the present findings. Farmers and breeders may have selected tepary bean genotypes for certain agronomic traits, resulting in considerable genetic differentiation among populations [52].

The observed and expected heterozygosity were marginally different (“Table 2”), suggesting the preponderance of homozygous alleles in dominant and recessive forms at many loci in the assessed tepary bean collection. The high additive variance supports this compared to the dominance variance observed in the present study. The higher tendency favouring homozygosity is expected as the tepary bean is a self-pollinated crop with a limited outcrossing rate [13,18]. The occurrence of both dominant and recessive alleles has implications for selection and breeding. Specifically, the upregulation of dominant alleles may enhance the selection of well-adapted genotypes, while the same may hinder the selection of recessive alleles [32]. The moderate fixation index value or coefficient of inbreeding (“Table 3”) suggests high gene exchange emanating from artificial and natural hybridization, which culminates in moderate genetic differentiation. Fixation indices are classified as low (< 0.05), moderate (0.05–0.15), and high (> 0.15) and show the level of genetic divergence or similarity in the assessed population [53]. Therefore, strategic crossing for the development of new varieties should target genetically divergent genotypes selected in different clusters (“Fig 7”). These can enhance

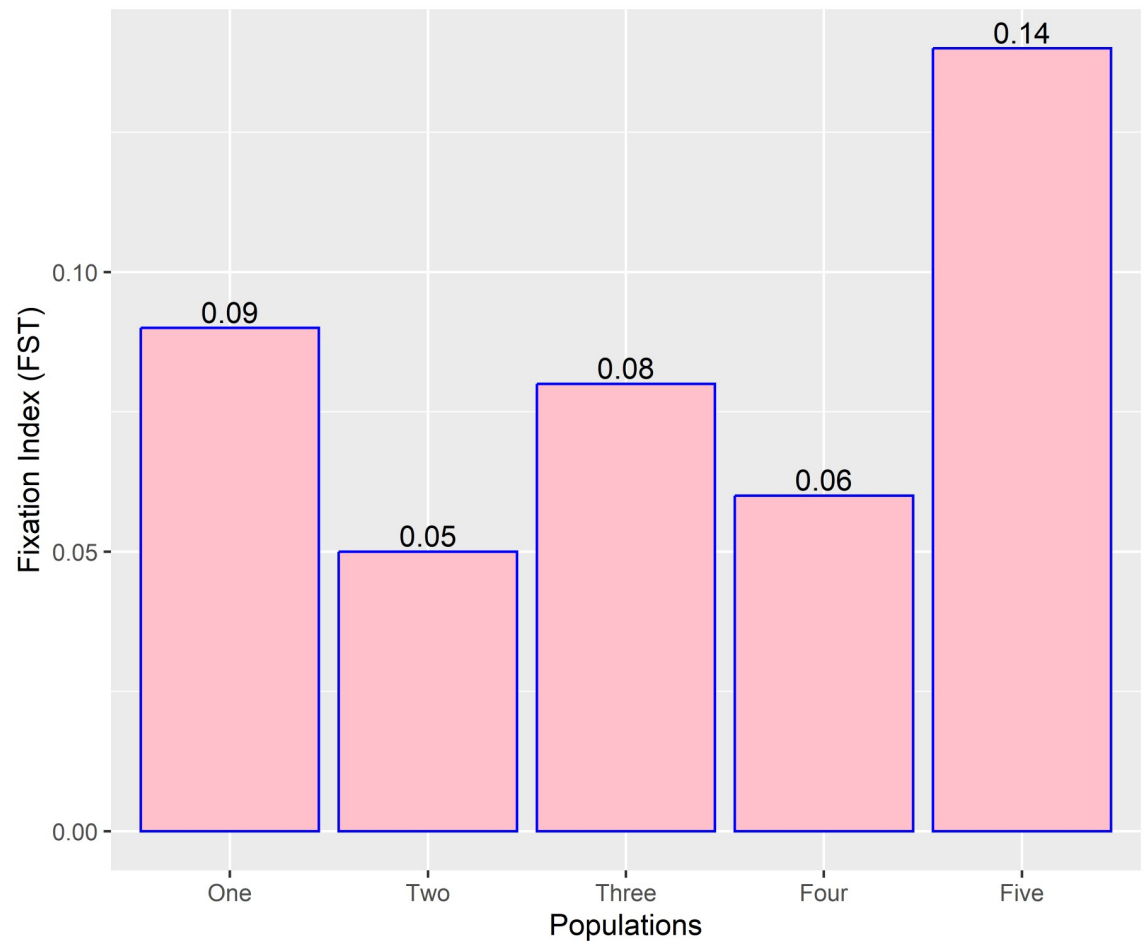


Fig 5. Fixation index (FST) of five tepary bean subpopulations determined from population structure analysis based on SNP markers.

<https://doi.org/10.1371/journal.pone.0295773.g005>

genetic gains for economic traits, including grain yield, nutrient compositions, heat and drought tolerance, and disease resistance.

The clustering or grouping of tepary bean genotypes into five sub-populations following their geographical origins in the USA and Mexico reiterated the single domestication theory, which apparently caused a genetic bottleneck and limited genetic variation in tepary bean [21]. The grouping of landraces like Uchokwane (G50) and Zimbabwe landrace (G45) with released and breeding varieties in Clusters 1 and 2 signifies the presence of admixtures that could have

Table 3. Population pair-wise fixation index of 78 tepary bean genotypes genotyped using SNPs markers.

Subpopulations	Subpopulation 1	Subpopulation 2	Subpopulation 3	Subpopulation 4	Subpopulation 5
Subpopulation 1	1	0.21	0.08	0.11	0.23
Subpopulation 2		1	0.18	0.14	0.20
Subpopulation 3			1	0.09	0.21
Subpopulation 4				1	0.06
Subpopulation 5					1

<https://doi.org/10.1371/journal.pone.0295773.t003>

Table 4. Analysis of molecular variance in tepary bean germplasm collection based on SNP markers.

Source of variation	DF	SS	MS	Variance Estimated	% variance
Between population	4	138098.70	34524.68	2097.93	38.12%
Within population	73	248584.00	3405.26	3405.26	61.87%
Total	77	386682.80	5021.85	5503.19	100%

DF, degrees of freedom; SS, sum of squares; MS, mean square.

<https://doi.org/10.1371/journal.pone.0295773.t004>

arisen due to historical exchanges of seeds through the informal seed system or the exchange of germplasm between breeding programs. The number of subpopulations in our study was slightly lower than the six subpopulations reported by Bornowski et al. [35]. This could be attributed to differences in the diversity panels and genotyping platform used.

The following genotypes are recommended for new variety design: Tars-Tep 112, Tars-Tep 10, Tars-Tep 23 selected from Cluster 1, G40022, Tars-Tep-93 and Tars-Tep -100 from Cluster 3, Zimbabwe landrace, G40017, G40143 and G40150 from Cluster 5. Genotypes, including Tars-Tep 23 and G40150, have high yield potential and tolerate drought and heat [54,55]. G40022 is tolerant to salt [56]. G40150 is resistant to common bacterial blight (CBB) while Tars-Tep 23 has broad resistance to common bacterial blight and rust [54,57].

Conclusion

The present study appraised SNP markers and determined their usefulness in assessing the extent of genetic diversity and population structure in tepary bean germplasm collections. The SNP markers revealed moderate genetic diversity in the germplasm collection. The study

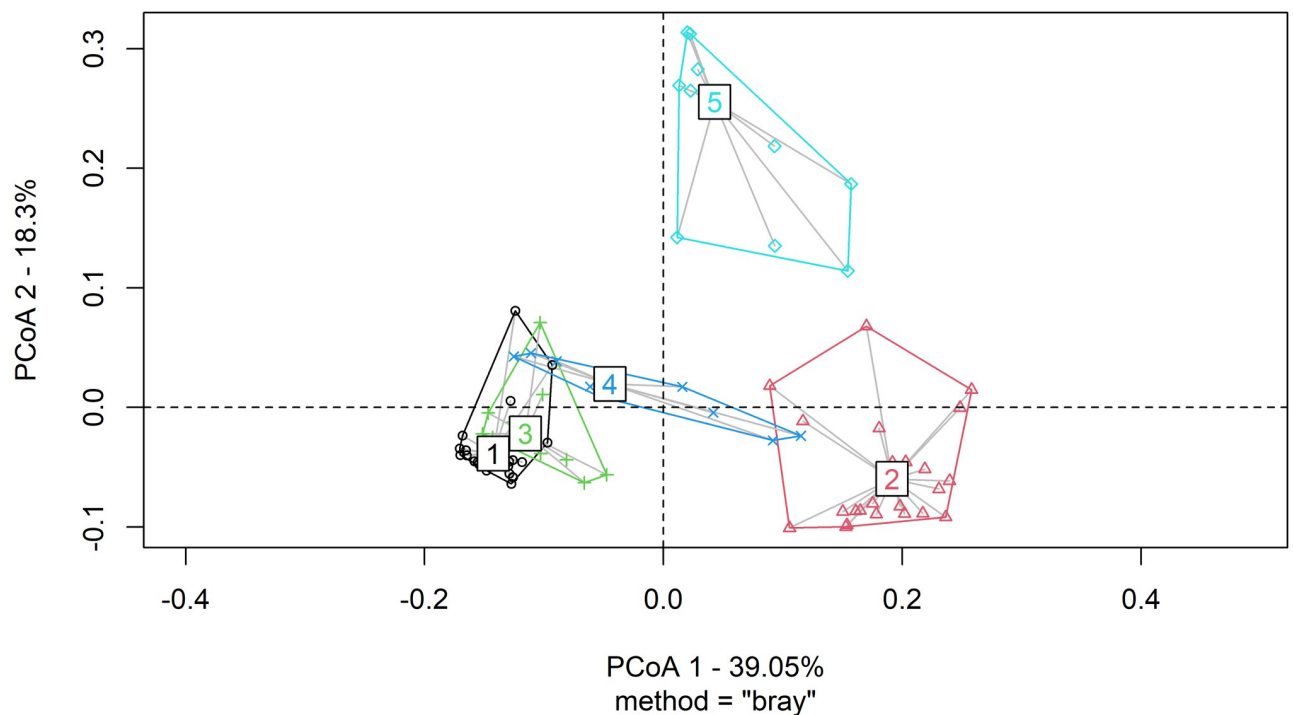


Fig 6. Principal coordinate analysis (PCoA) of tepary bean genotypes based on population structure using SNPs markers.

<https://doi.org/10.1371/journal.pone.0295773.g006>

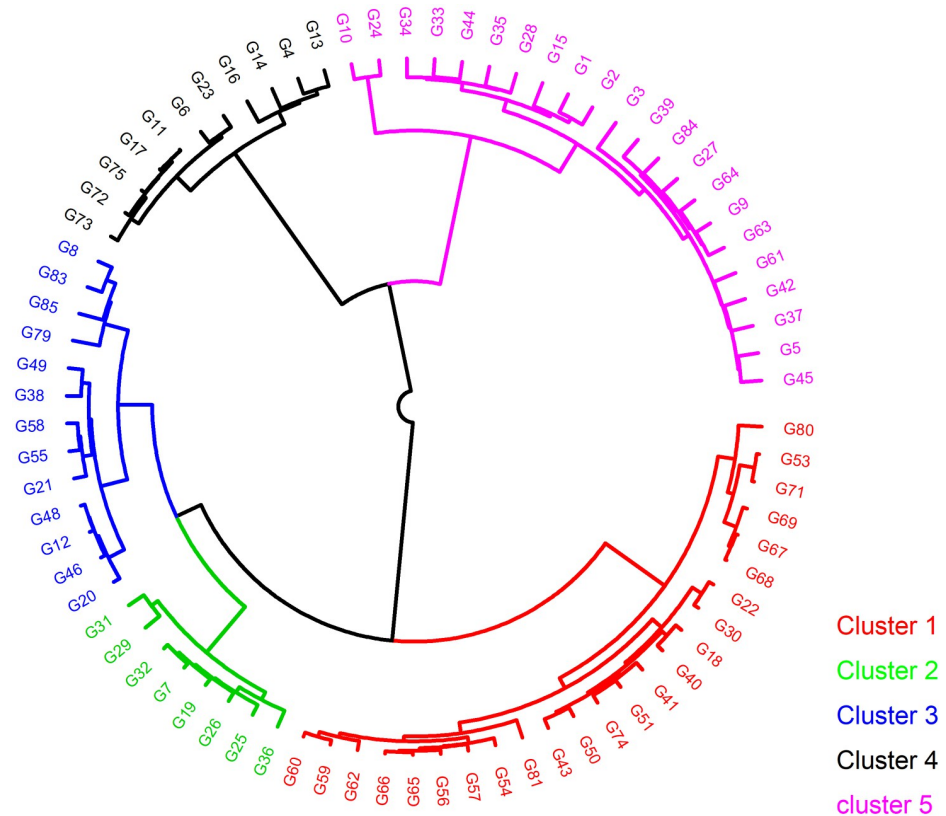


Fig 7. Dendrogram showing the five genetic clusters (Cluster 1 to 5) among 78 tepary bean genotypes based on SNPs markers. See codes of genotypes in “Table 1”.

<https://doi.org/10.1371/journal.pone.0295773.g007>

identified five distinctive sub-populations that were genetically differentiated. This will guide genotype selection and subsequent crosses to develop breeding populations and enhance genetic gains for economic traits. The following distantly related genotypes were selected, namely: Tars-Tep 112, Tars-Tep 10, Tars-Tep 23, Tars-Tep-86, Tars-Tep-83, and Tars-Tep 85 from Cluster 1, G40022, Tars-Tep-93, and Tars-Tep-100 from Cluster 3, Zimbabwe landrace, G40017, G40143, and G40150 from Cluster 5. The selected and contrasting accessions are valuable genetic resources to initiate crosses to enhance genetic variation and integrate traits in tepary bean and genetically related legume crops.

Supporting information

S1 Table. Hapman file for DArT- based SNP markers.
(CSV)

S2 Table. Divgenos file.
(CSV)

Acknowledgments

The authors would like to express their deepest gratitude to the staff of SEQART Africa for the Dartseq genotyping service and to the CIAT-Malawi staff for processing the seed samples.

Author Contributions

Conceptualization: Saul Eric Mwale, Hussein Shimelis, Jacob Mashilo.

Data curation: Saul Eric Mwale.

Formal analysis: Saul Eric Mwale, Wilfred Abincha.

Funding acquisition: Saul Eric Mwale.

Investigation: Saul Eric Mwale.

Methodology: Saul Eric Mwale, Wilfred Abincha.

Resources: Saul Eric Mwale.

Software: Saul Eric Mwale, Wilfred Abincha.

Supervision: Hussein Shimelis.

Validation: Saul Eric Mwale, Hussein Shimelis, Wilfred Abincha, Wilson Nkhata, Abel Sefasi, Jacob Mashilo.

Visualization: Saul Eric Mwale, Hussein Shimelis, Wilfred Abincha, Jacob Mashilo.

Writing – original draft: Saul Eric Mwale.

Writing – review & editing: Saul Eric Mwale, Hussein Shimelis, Wilfred Abincha, Wilson Nkhata, Abel Sefasi, Jacob Mashilo.

References

1. Moghaddam SM, Oladzad A, Koh C, Ramsay L, Hart J, Mamidi S, et al. The tepary bean genome provides insight into evolution and domestication under heat stress. *Nat. Commun.* 2021; 12:1–14. <https://doi.org/10.1038/s41467-021-22858-x> PMID: 33976152
2. Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. *Mol. Biol. Rep.* 1991; 9, 208–221.
3. Bhardwaj HL, Hamama AA. Oil and fatty acid composition of tepary bean seed. *Hort Sci.* 2005; 40, 1436–1438.
4. Muñoz LC, Debouck DG, Rivera M, Muñoz JE, Alpala D, Sarsu F, et al. Mutation breeding for heat and drought tolerance in tepary bean (*Phaseolus acutifolius* A. Gray). *Aust. J. Crop Sci.* 2021; 60–68.
5. Cruz S, Lobató n J, Urban MO, Ariza, Suarez D, Raatz B, et al. Interspecific common bean population derived from *Phaseolus acutifolius* using a bridging genotype demonstrate useful adaptation to heat tolerance. *Front. Plant Sci.* 2023; 14.
6. Miklas PN, Smith JR., Singh SP. Registration of common bacterial blight resistant dark red kidney bean germplasm line USDK-CBB-15. *Crop Sci.* 2006; 46, 1005–1007
7. Salgado MO, Schwartz HF, Brick MA, Pastor-Corrales MA. Resistance to *Fusarium oxysporum* f.sp. phaseoli in tepary bean (*Phaseolus acutifolius*). *Plant Dis.* 1994; 78, 357–360.
8. Miklas PN, Santiago J. Reaction of select tepary bean to bean golden mosaic virus. *Hort Sci.* 1996; 31, 430–432.
9. Thomas CV, Waines JG. Fertile backcross and allotetraploid plants from crosses between tepary beans and common beans. *J. Hered.* 1984; 75, 93–98.
10. Rao I, Beebe S, Polania J, Ricaurte J, Cajiao C, Garcia R, et al. Can tepary bean be a model for improvement of drought resistance in common bean? *Afr. Crop Sci. J.* 2013; 21, 265–281.
11. Molosiwa OO, Kgokong SB, Makwala B, Gwafila CM, & Ramokapane MG. Genetic diversity in tepary bean (*Phaseolus acutifolius*) landraces grown in Botswana. *J. Plant Breed. Crop Sci.* 2014; 6, 194–199.
12. Gwata ET, Shimelis H, & Matova PM. Potential of improving agronomic attributes in tropical Legumes using two mutation breeding techniques in Southern Africa. *Alternative crops and cropping systems*, Petr Konvalina, IntechOpen. 2016.
13. Mhlaba ZB, Mashilo J, Shimelis H, Assefa AB, and Modi AT. Progress in genetic analysis and breeding of tepary bean (*Phaseolus acutifolius* A. Gray): A review. *Sci. Hortic.* 2018a; 112–119.

14. Konvalina P. Alternative Crops and Cropping Systems. IntechOpen. 2016 <https://doi.org/10.5772/61408>.
15. Mwale SE, Shimelis H, Mafongoya P, Mashilo J. Breeding tepary bean (*Phaseolus acutifolius*) for drought adaptation: A review. *Plant Breed.* 2020; 1–13. <https://doi.org/10.1111/pbr.12806>
16. Pratt RC, Grant L, Velasco-Cruz C, & Lauriault L. Field performance of selected and landrace tepary bean varieties in diverse southwestern USA irrigated production environments. *Legum. Sci.* 2023; 5(1), e157. <https://doi.org/10.1002/leg3.157>
17. Mhlaba ZB, Amelework B, Shimelis HA, Modi AT, Mashilo J. Variance components and heritability of yield and yield-related traits in tepary bean (*Phaseolus acutifolius*). *S. Afr. J. Plant Soil.* 2018b; 1–12.
18. Blair MW, Pantoja W, Muñoz LC. First use of microsatellite markers in a large collection of cultivated and wild accessions of tepary bean (*Phaseolus acutifolius* A. Gray). *Theor Appl. Genet.* 2012; 125:1137–47.
19. Mhlaba ZB, Amelework B, Shimelis HA, Modi AT, Mashilo J. Genetic interrelationship among tepary bean (*Phaseolus acutifolius* A. Gray) genotypes revealed through SSR markers. *Aust. J. Crop Sci.* 2018c; 12(10): 1587–1595.
20. Gepts P, Llaca V, Nodari RO, Panella L. Analysis of Seed Proteins, Isozymes, and RFLPs for Genetic and Evolutionary Studies in *Phaseolus*. In: Linskens, H.F., Jackson, J.F. (eds) *Seed Analysis. Modern Methods of Plant Analysis.* 1992. vol 14. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-01639-8_4.
21. Schinkel C, Gepts P. Phaseolin diversity in the tepary bean, *Phaseolus acutifolius* A. Gray. *Plant Breed.* 1988; 101, 292–301.
22. Pedraza F, Gallego G, Beebe S, Tohme J. Marcadores SCAR y RAPD para la resistencia a la bacteriosis común (CBB). In: Singh, S.P., Voyses, O. (Eds.), *Taller de mejoramiento de frijol para el Siglo XXI: Bases para una estrategia para America Latina.* CIAT, Cali, Colombia. 1997; 130–134.
23. Park SO, Coyne DP, Dursun A, Jung G. Identifying randomly amplified polymorphic DNA (RAPD) markers linked to major genes for common bacterial blight resistance in tepary bean. *J. Am. Soc. Hortic. Sci.* 1998; 123, 278–282.
24. Muñoz LC, Duque MC, Debouck DG, Blair MW. Taxonomy of tepary bean (*Phaseolus acutifolius*) and wild relatives as determined by amplified fragment length polymorphism (AFLP) markers. *Crop Sci.* 2006; 46:1744–54.
25. Yu K, Park SJ, Poysa V. 2000. Marker-assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. *Plant Breed.* 2000; 119, 411–415.
26. Gujaria-Verma N, Ramsay L, Sharpe AG, Sanderson LA, Debouck DG, Tar'an B, et al. Gene-based SNP discovery in tepary bean (*Phaseolus acutifolius*) and common bean (*P. vulgaris*) for diversity analysis and comparative mapping. *BMC Genomics.* 2016; 17, 239.
27. Cichy KA, Wiesinger JA, Mendoza FA. Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 2015; 128 (8):1555–67. <https://doi.org/10.1007/s00122-015-2531>
28. Raatz B, Mukankusi C, Lobaton JD, Male A, Chisale V, Amsalu B, et al. Analyses of African common bean (*Phaseolus vulgaris* L.) germplasm using a SNP fingerprinting platform: diversity, quality control and molecular breeding. *Genet. Resour. Crop Evol.* 2019; 66(3):707–22. <https://doi.org/10.1007/s10722-019-00746>
29. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One.* 2011; 6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379> PMID: 21573248
30. Valdisser PAMR Pereira WJ, Almeida JE Muüller BSF, Coelho GRC, and de Menezes IPP. In-depth genome characterization of a Brazilian common bean core collection using dartseq high-density Snp genotyping. *BMC Genomics.* 2017; 18:423. <https://doi.org/10.1186/s12864-017-3805-4> PMID: 28558696
31. Yohane EN, Shimelis H, Laing M, Shayanowako A. Genetic diversity and grouping of pigeonpea [*Cajanus cajan* Millspaugh] Germplasm using SNP markers and agronomic traits. *PLoS ONE.* 2022; 17(11): e0275060. <https://doi.org/10.1371/journal.pone.0275060> PMID: 36327283
32. Ahmed SM, Alsamman AM, Jighly A, Mubarak MH, Al-Shamaa K, Istanbuli T, et al. Genome-wide association analysis of chickpea germplasms differing for salinity tolerance based on DArTseq markers. *PLoS ONE.* 2021; 16(12): e0260709. <https://doi.org/10.1371/journal.pone.0260709> PMID: 34852014
33. Lukanda MM, Dramadri IO, Adjei EA. et al. Genetic Diversity and Population Structure of Ugandan Soybean (*Glycine max* L.) Germplasm Based on DArTseq. *Plant Mol Biol Rep.* 2023; 41: 417–426. <https://doi.org/10.1007/s11105-023-01375-9>

34. Nkhata W, Shimelis H, Melis R, Chirwa R, Mzengeza T, Mathew I, et al. 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*. 2020; 15(12). <https://doi.org/10.1371/journal.pone.0243238> PMID: 33338076
35. Bornowski N, Hart JP, Vargas Palacios A, Ogg B, Brick MA, Hamilton JP, et al. Genetic variation in a tepary bean (*Phaseolus acutifolius* A. Gray) diversity panel reveals loci associated with biotic stress resistance. *The Plant Genome*. 2023; 16, e20363. <https://doi.org/10.1002/tpg2.20363> PMID: 37332263
36. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H. et al. Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol. Biol.* 2012; 888, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5 PMID: 22665276
37. Alam M, Neal J, O'Connor K, Kilian A, Topp B. Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. *PLOS ONE*. 2018; 13(8): e0203465. <https://doi.org/10.1371/journal.pone.0203465> PMID: 30169500
38. Adu BG, Akromah R, Amoah S, Nyadanu D, Yeboah A, et al. High-density DArT-based SilicoDArT and SNP markers for genetic diversity and population structure studies in cassava (*Manihot esculenta* Crantz). *PLOS ONE*. 2021; 16(7): e0255290. <https://doi.org/10.1371/journal.pone.0255290> PMID: 34314448
39. Gelaw YM, Eleblu JSY, Ofori K. et al. High-density DArTSeq SNP markers revealed wide genetic diversity and structured population in common bean (*Phaseolus vulgaris* L.) germplasm in Ethiopia. *Mol Biol Rep.* 2023; 50: 6739–6751. <https://doi.org/10.1007/s11033-023-08498-y> PMID: 37389701
40. Granato, Neto-Fritsche. 'snpReady package'. 2017. <https://cran.r-project.org/web/packages/snpReady/vignettes/snpReady-vignette.html>.
41. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2020. <http://www.R-project.org>.
42. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, et al. rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics, Proteomics and Bioinformatics*. 2021; 19(4): 619–628. <https://doi.org/10.1016/j.gpb.2020.10.007> PMID: 33662620
43. Frichot E, and Francois O. LEA: An R package for Landscape and Ecological Association studies. 2015.
44. Patterson N, Price AL, and Reich D. Population structure and eigen analysis. *PLoS Genet.* 2006; 2:20.
45. Pritchard JK, Stephens M and Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. <https://doi.org/10.1093/genetics/155.2.945> PMID: 10835412
46. Alexander DH, and Lange K. Enhancements to the Admixture algorithm for individual ancestry estimation. *BMC Bioinformatics*. 2011; 12:246. <https://doi.org/10.1186/1471-2105-12-246> PMID: 21682921
47. Frichot E, Mathieu F, Trouillon T, Bouchard G and Francois O. Fast and efficient estimation of individual ancestry coefficients. *Genetics*. 2014; 196:973–983.
48. Byun J, Han Y, Gorlov IP, Busam JA, Seldin MF, & Amos CI. Ancestry inference using principal component analysis and spatial analysis: A distance-based analysis to account for population substructure. *BMC Genomics*. 2017; 18(1):1–12. <https://doi.org/10.1186/s12864-017-4166-8> PMID: 29037167
49. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *Peer J*. 2014. <https://doi.org/10.7717/peerj.281> PMID: 24688859
50. Luo Z, Brock J, Dyer JM, Kutchan TM, Augustin M, Schachtman DP, et al. Genetic diversity and population structure of a *Camelina sativa* spring panel. *Front. Plant Sci.* 2019; 10:184. <https://doi.org/10.3389/fpls.2019.00184> PMID: 30842785
51. Singh N, Choudhury DR, Singh AK, Kumar S, Srinivasan K, Tyagi RK, et al. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS ONE*. 2013; 8(12):1–14. <https://doi.org/10.1371/journal.pone.0084136> PMID: 24367635
52. Kinhoègbè G, Djèdatin G, Saxena RK, Chitikineni A, Bajaj P, Molla J, et al. Genetic diversity and population structure of pigeonpea (*Cajanus cajan* [L.] Millspaugh) landraces grown in Benin revealed by Genotyping-By-Sequencing. *PLoS ONE*. 2022; 17(7): e0271565. <https://doi.org/10.1371/journal.pone.0271565>
53. Vitorino CA, Nogueira F, Souza IL, Araripe J, Venere PC. Low genetic diversity and structuring of the Arapaima (*Osteoglossiformes*, Arapaimidae) population of the Araguaia-Tocantins basin. *Front. genet.* 2017; 8:159. <https://doi.org/10.3389/fgene.2017.00159> PMID: 29114261
54. Porch T, Barrera S, Berny J, Díaz-Ramírez J, Pastor-Corrales M, Gepts P, et al. Release of tepary bean TARS-Tep 23 germplasm with broad abiotic stress tolerance and rust and common bacterial blight resistance. *J. Plant Regist.* 2021; 1–11.

55. Mwale SE, Shimelis H, Nkhata W, Sefasi A, Fandika I, Mashilo J. Genotype-by-Environment Interaction in Tepary Bean (*Phaseolus acutifolius* A. Gray) for Seed Yield. *Agronomy*. 2023; 13(1):12. <https://doi.org/10.3390/agronomy1301001>
56. Bayuelo-Jiménez JS, Debouck DG, Lynch JP. Salinity tolerance in Phaseolus species during early vegetative growth. *Crop Sci*. 2002; 42, 2184–2192.
57. Vargas A, Porch TG, Beaver J. Evaluation of the tepary bean (*Phaseolus acutifolius*) CIAT germplasm collection for response to common bacterial blight and bean common mosaic necrosis virus. *Ann. Rep. Bean. Improv. Coop*. 2014; 181–182.