

Genetic diversity in sorghum (*Sorghum bicolor* L. Moench) accessions using SNP based Kompetitive allele-specific (KASP) markers

Thulo Sejake^{1,2}, Nemera Shargie^{1*}, Riann Christian², Assefa B. Amelework⁴ and Toi J. Tsilo^{2,3}

¹Agricultural Research Council - Grain Crops, Private Bag X1251, Potchefstroom 2520, South Africa

²University of South Africa, College of Agriculture and Environmental Sciences, Private Bag X6, Florida 1710, South Africa

³Agricultural Research Council - Small Grain, Private Bag X29, Bethlehem 9700, South Africa

⁴Agricultural Research Council - Vegetable and Ornamental Plants, Private Bag x 293, Pretoria, 0001, South Africa

*Corresponding author: ShargieN@arc.agric.za

Abstract

Genetic diversity analysis is an important component in conventional and marker-assisted breeding. The objective of this study was to assess the level of genetic diversity among 100 sorghum accessions, which were selected randomly from the Sorghum National Germplasm Bank maintained at Agricultural Research Council, South Africa. A total of 136 Kompetitive Allele-Specific PCR (KASP) markers were used in this study. The KASP markers were previously derived from single-nucleotide polymorphic (SNP) analysis of the world-wide sorghum accessions by other research groups. A total of 110 KASP markers were polymorphic and recorded an average polymorphic information content (PIC) value of 0.3, which indicated high level of discrimination of the markers. The markers had an average gene diversity and observed heterozygosity of 0.3 and 0.10, respectively. Analysis of molecular variance revealed a significantly high variation among accessions (83% and 89%) than within accessions (10% and 11%) based on breeding status and geographic origin, respectively. Genetic distance varied from 0.0 between SA0672 and SA0673, SA1282 and SA0670 to 0.57 between Hakika and SA1442 with an average mean of 0.30. The dendrogram and model-based population analysis identified three and four distinct groups in 95 sorghum accessions, respectively. These results imply the presence of genetic diversity and lack of genetic bottleneck within the Sorghum National Germplasm Bank, which could be highly relevant for sorghum breeding and germplasm maintenance.

Keywords: KASP markers; polymorphic information content; heterozygosity; analysis of molecular variance; genetic diversity.

Abbreviations: AMOVA_Analysis of Molecular Variance; CTAB_hexa-deCyl Tri-methyl Ammonium Bromide; F_{IS} _Inbreeding coefficient; GD_Genetic Distance; H_e _gene diversity; H_o _observed heterozygosity; I _Shannon's Information Index; KASP_Kompetitive Allele-Specific PCR; N_a _Number of alleles; N_e _ number of effective alleles; N_m _Gene flow; PIC_Polymorphic Information Content; SNP_Single Nucleotide Polymorphic; SSR_Simple Sequence Repeats.

Introduction

Sorghum (*Sorghum bicolor* L. Moench) is the fifth most important cereal crop providing food and feed throughout the world. In South Africa, sorghum meal is often consumed as a stiff porridge (brown porridge), also known as Mabele, and it is also used to ferment traditional beer (DAFF, 2010). It is gluten-free, which makes it an excellent substitute for wheat for consumers with gluten intolerance (Koenders, 2010). Despite its economic importance in South Africa, sorghum research is inadequately supported compared to other food crops like maize and wheat. Its productivity, especially under small-scale farming conditions, remains low due to various challenges including lack of income to buy production inputs, use of landraces with inadequate quality seed and susceptibility to pests and diseases (Rowhani et al., 2011). It is essential to discriminate and document the genetic variability in sorghum accessions because the information makes the conservation process easier and helps in the improvement of the crop. Genetic diversity provides plant breeders with relevant genetic variation to

practice effective selection and release of new and distinct cultivars. Assessment of genetic diversity can be accomplished using morphological traits, biochemical and molecular markers (Winter and Kahl, 1995; Narain, 2000; Mohammadi and Prasanna, 2003). Conventional breeding used morphological traits and markers to assess genetic diversity in crops. However, use of morphological features in genetic analysis has a limitation, including the influence by environmental conditions where plants are grown, thus reducing their value for efficient characterization, especially for traits with low heritability (Buchekeyi et al., 2009; Elangovan et al., 2013; Sami, 2013). Biochemical markers are more reliable than morphological markers but are also limited by environmental influences and changes in different developmental stages of the plant. The limitations of both morphological and biochemical traits have led to the evolution of molecular markers, which are unlimited in nature and are not affected by environmental conditions. There are various types of molecular markers including the

currently used Single Nucleotide Polymorphism (SNP) markers. These SNP markers are scattered throughout the genomes and occur more frequently than any other type of markers (Semagn et al., 2006, Govindaraj et al., 2015), which makes SNPs the markers of choice in the current study. Many genotyping platforms can be employed to generate SNP data, one of which include Kompetitive Allele-Specific PCR (KASP) genotyping, a homogeneous SNP genotyping technology initially developed by LGC genomics for in-house genotyping. Several studies have reported the effectiveness of KASP genotyping, including a study by Hiremath et al. (2012) who converted 2486 Chickpea SNPs to KASP markers in which 1341 (66.8%) markers were polymorphic. These markers were then validated on 58 genotypes and had the Polymorphic Information Content (PIC) value of 0.12. The same markers were used to assess the genetic diversity among 58 chickpea genotypes, and recorded genetic dissimilarity of 0.37. The other study was reported by Khera et al. (2013), who converted 96 set of SNPs to KASP markers in groundnuts, in which 73 markers were polymorphic. These polymorphic KASP markers were then employed on 280 groundnuts genotypes to estimate genetic diversity features and recorded an average PIC value of 0.32. The extent of genetic diversity of Southern African sorghum germplasm collections from different gene banks have been reported (Uptmoor et al., 2003; Ng'uni et al., 2012; Gerrano et al., 2014; Motlhaodi et al., 2014; Mofokeng et al., 2014), however there are limitations which include the sampling size, depth and extent of the analysis, and large number of accessions that are continuously maintained at these gene banks. The Agricultural Research Council (ARC) National Sorghum Germplasm Bank, for example, maintained over 4000 sorghum accessions. Therefore, the aim of the present study was to assess the level of genetic diversity among randomly selected sorghum accessions maintained at the National Sorghum Germplasm Bank in South Africa using SNP based KASP markers to inform subsequent breeding and conservation programmes.

Results

Characteristics of the KASP markers across sorghum accessions

Of the 136 KASP markers, only 110 KASP markers were highly polymorphic in a set of 95 sorghum accessions studied. The average PIC value was 0.3, in which only 4.55% of the markers had PIC values less than 0.10, 14.55% in the range of 0.11 to 0.20, 28.18% in the range of 0.21 to 0.30, 22.73% in the range of 0.31 to 0.40 and the other 30.0% had PIC values in the range of 0.41 to 0.50 (Figure 1). Most of the markers (52%) had PIC values of more than 0.30, which indicated high level of discrimination of the marker among the sorghum accessions. Similarly, the KASP markers recorded an average gene diversity (H_e) of 0.3, in which 5.45% of the markers had H_e values less than 0.10, 13.64% in the range of 0.11 to 0.20, 27.27% in the range of 0.21 to 0.30, 23.64% in the range of 0.31 to 0.40 and 30.0% in the range of 0.41 to 0.50 (Figure 2). Furthermore, the markers had an average observed Heterozygosity (H_o) of 0.10. About 70.0% of the markers had H_o values lower than 0.05, 22.73% in the range of 0.06 to 0.10, and 6.36% in the range 0.11 to 0.15, and only 0.91% in the range 0.16 to 0.20 (Figure 3).

Genetic diversity within and among sorghum accessions

Genetic diversity within and among sorghum populations was based on the breeding status of the accessions and the

geographic origin as reflected in Table 1 (a and b). The R-lines (a) recorded the highest Shannon's information index (0.54) while B-lines had the lowest value (0.38) with an average of 0.494. The observed Heterozygosity (H_o) ranged from 0.038 in imported genotypes (b) to 0.053 in R-lines, with an average of 0.043. On the other hand, the expected Heterozygosity (H_e) ranged from 0.237 in B-lines to 0.372 in R-lines compared to an average of 0.325. The highest inbreeding coefficient was observed in landraces (0.899) while the lowest was observed in B-lines (0.801), with an average of 0.862.

Analysis of molecular variance (AMOVA) is represented in Table 2 (a and b), in which 83% of the variation was observed among accessions (highly significant, $P < 0.001$) while 10% was observed within accessions (highly significant, $P < 0.001$) and only 7% among populations based on breeding status (a). Similarly, 89% variation was observed among accessions (highly significant, $P < 0.001$) and only 11% was observed within accessions based on geographic origin (b). The fixation index based on geographic origin was higher (0.082) than the one based on breeding status (0.001).

Genetic distances and clustering

Wide genetic distances were observed among accessions indicating a wide genetic diversity among sorghum accessions. The maximum genetic distance was observed between Hakika and SA1442 (0.57), while the minimum distance of 0.00 was observed between SA0672 and SA0673 and between SA1282 and SA0670. The average genetic distance between the accessions was 0.30. The summary data of genetic distance between the accessions is presented in Figure 4.

Cluster analysis was done based on neighbour-joining algorithm using the UPGMA and a dendrogram was generated on the dissimilarity matrix. The accessions were grouped into three distinct clusters as shown in Figure 5 with different colours indicating the clustering patterns among the genotypes. The clustering patterns of the genotypes did not match with the geographical origin probably due to gene flow. Cluster I had only four genotypes comprising of three from SADC region and one from United States. Cluster II had 27 genotypes comprising of 13 accessions from SADC region and 14 R-lines from United States. Cluster III contained the highest proportion of the genotypes 64 (67%) and dominated by the South African collections, 34 accessions, followed by 25 accessions from the United States and five accessions from the rest of SADC countries. The landraces from South Africa formed a close group, although they broadly clustered with other accessions from other countries. It is interesting to note that accessions SA1282 and SA0670, and SA0672 and SA0673, both pairs from the United States were genetically 100% similar.

Population structure

Population structure was constructed to reveal the genetic relationship and to aid accession selection. The structure analysis was performed by setting the range of K as 1 to 10. The results showed that K = 4 was the optimum, which indicates at least four distinct subgroups present in the 95 sorghum accessions. Figure 6 presents the population structure for K = 4 where each colour presents a different genetic cluster. In the structure, cluster 1 was the smallest consisting of only 11 accessions, in which five were from the USA followed by five accessions from the SADC regions and

one accession was from Australia. Cluster 2 consisted of 32 accessions, in which 18 were from the SADC regions and 14 were from the USA. Cluster 3 was the largest with 35 accessions, in which 23 were from South Africa and 12 were from the USA. The last cluster (4) consisted of nine accessions from the USA and eight accessions from the SADC regions. However, there was no congruence of clustering pattern of accessions between UPGMA method and model-based approach observed.

Discussion

Characteristics of the KASP markers across sorghum accessions

The KASP markers revealed a wide genetic diversity among the studied sorghum accessions. The average PIC mean value of the markers was 0.3, in which 52% of the markers had PIC values of more than 0.30, which indicated high levels of discrimination and informativeness of the markers among the sorghum accessions (Botstein et al., 1980). The PIC in the current study is higher than the values reported by Mengistu et al. (2020), who used SNP markers to assess genetic variation in sorghum accessions from Ethiopia and reported an average PIC value of 0.22. Although there are not many Southern African studies published in sorghum using KASP or SNP markers, there are however many papers that reported on the use of Simple Sequence Repeats (SSR) markers. Some of these papers include a study by Smith et al. (2000) who employed 15 SSRs in addition to other markers to assess genetic diversity among elite sorghum inbred lines and reported an average PIC value of 0.58. Mbeyagala et al. (2012) also reported a PIC value of 0.65 in landraces from Uganda using 21 SSR markers. Mofokeng et al. (2014) also employed 30 SSR markers to assess the extent of genetic diversity among South African sorghum genotypes and reported a PIC value of 0.50, which all implied that SSR markers were having higher PIC values. Amelework et al. (2015) reported a mean PIC value of 0.61 in their study of 200 Ethiopian sorghum landraces using 30 SSRs. Although SSR markers seemed to have higher PIC values than KASP/SNP markers in sorghum genotypes, studies in other crops had similar average PIC value to the current study using KASP markers. These include a study by Khera et al. (2013) who reported a PIC value of 0.32 in the validation set and 0.31 in the reference set of groundnuts using KASP markers. Roorkiwal et al. (2014) also reported PIC value of 0.23 in 94 genotypes of *Cicer species* using SNP markers.

The low levels of observed heterozygosity (0.10) in the current study could be associated with the inbreeding nature of sorghum as it is a self-pollinated crop (Motlhaodi et al., 2014). The average expected heterozygosity in this study was high (0.3) which could be due to the high level of genetic diversity in sorghum accessions studied. Low levels of gene diversity and observed heterozygosity have been reported by Ng'uni et al. (2011) in Zambian sorghum accessions using 10 SSR markers, which were on average 0.07 of H_e and 0.04 of H_o . However, higher levels of gene diversity and observed heterozygosity in sorghum accessions were also reported by Mbeyagala et al. (2012) who reported H_e of 0.68 and H_o of 0.18 in Ugandan sorghum landraces using 21 SSR markers. Motlhaodi et al. (2014) also reported gene diversity of 0.46 and observed heterozygosity of 0.09 in sorghum accessions from Botswana using 10 SSR markers. Amelework et al. (2015) reported a mean H_o of 0.10 and H_e of 0.59 in Ethiopian sorghum landraces using 30 SSRs. The

low H_o values observed in all the studies is expected from a crop like sorghum that is maintained by continued self-fertilization and low out crossing value of less than 5% (Amelework et al., 2015).

Genetic diversity within and among sorghum accessions

The average Shannon's information index (I) within the population was 0.494 with R-lines having the highest value (0.54) while B-lines had the lowest value (0.38). Lower values of Shannon's information index indicate that the population is diverse, while higher values indicate less diversity within populations (Shannon, 1948). Therefore the high value of I in the R-lines indicated the dominance of certain alleles in the population resulted in less diversity in the population as the accessions share common ancestry and most originated from the United States. The lower value of Shannon's information index in the B-lines indicate that the accessions have different genetic background and they originated from different countries which included Australia, Botswana, Lesotho, United States of America and South Africa. It was interesting to note that the low heterozygosity observed in landraces, while high heterozygosity was observed in R-lines. Similarly, the average gene diversity within genotypes per population was higher in R-lines and lower in B-lines. The average inbreeding coefficient within the genotypes was 0.86 with the highest observed in landraces and the lowest in B-lines. This phenomenon could simply be because the B-lines are more diverse as they originated from different countries other than the landraces, which originated from South Africa.

AMOVA revealed a highly significant variation among and within sorghum accessions based on both breeding status and geographic origin. The total variation was significantly high among accessions (83% and 89% based on breeding status and geographic origin, respectively) and significantly low variation within accessions (10% and 11% based on breeding status and geographic origin, respectively). This indicates that large variation existed among regions while low variation existed among accessions, indicating that accessions from the same origin are similar but different to those from different origins. This could be due to the fact that accessions from the same region may share common ancestry and environments, which makes them diverse from those in different genetic background. Similar results were also reported by Ng'uni et al. (2012) who employed 10 SSR markers to assess the genetic diversity among Southern African sorghum accessions from Malawi, Tanzania and Zambia and reported more variation among accessions (43.52%) and less variation within accessions (13.25%) based on agro-ecological regions. Using the same approach with SSR markers, Mbeyagala et al. (2012), however, reported a high variation within Ugandan sorghum landraces (96.42% and 94.72% based on breeding status and geographic origin, respectively) and low variation among sorghum landraces (3.58% and 5.28% based on breeding status and geographic origin, respectively), which are different from the findings in the current study. It was also interesting to note that the fixation index based on geographic origin was higher (0.082) than the one based on breeding status (0.001). Grouping of accessions based on geographic origin showed moderate genetic variation while grouping based on the breeding status of the accessions showed higher genetic variation. Similar results were also reported by Mbeyagala et al. (2012).

Table 1. Genetic diversity within and among sorghum populations formed based on breeding status and geographic origin.

Population	Genetic parameter				
	N	I	H _o	H _e	F _{IS}
(a) Breeding Status					
B-lines	48 (0.26)	0.38 (0.016)	0.039 (0.018)	0.237 (0.014)	0.801 (0.020)
R-lines	35 (0.14)	0.54 (0.009)	0.053 (0.016)	0.372 (0.013)	0.836 (0.013)
Landraces	12 (0.08)	0.53 (0.013)	0.038 (0.015)	0.367 (0.013)	0.899 (0.013)
(b) Geographic origin					
Local	42 (0.25)	0.479 (0.014)	0.045 (0.004)	0.312 (0.012)	0.849 (0.012)
Imported	53 (0.13)	0.505 (0.017)	0.039 (0.004)	0.342 (0.014)	0.866 (0.017)
Overall population					
Total	95 (0.33)	0.494 (0.014)	0.043 (0.004)	0.325 (0.012)	0.862 (0.011)

N = number of genotypes tested; I = Shannon information index; H_o = average observed gene diversity within genotypes per population; H_e = average gene diversity within genotypes per population; F_{IS} = inbreeding coefficient. The values in the bracket represent standard error.

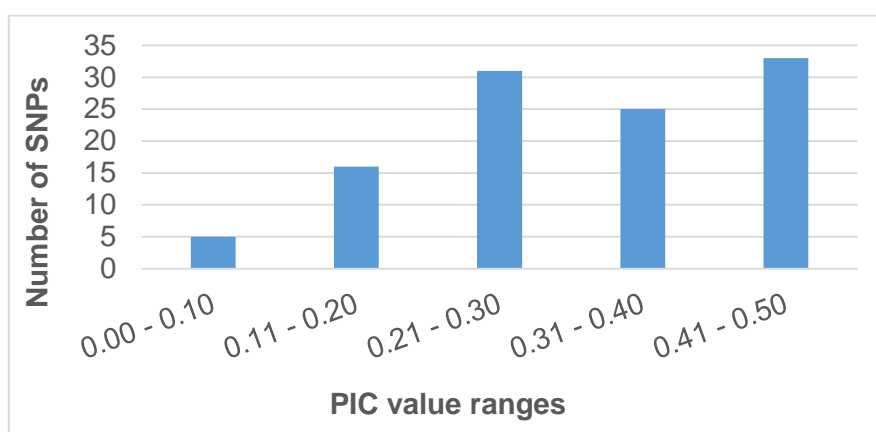


Figure 1. Polymorphic Information Content (PIC) values of KASP markers used for diversity analysis.

Table 2. AMOVA among 95 sorghum genotypes classified based on breeding status (a) and geographic origin (b) using 110 SNP markers.

Source	d.f.	SS	MS	Est. Var.	Per Var.	P(rand >= data)
(a) Breeding status						
Among Populations	2	246.229	123.115	1.438	7%	FST ≤ 0.001
Among Individuals	92	3466.276	37.677	17.717	83%	FIS ≤ 0.001
Within Individuals	95	213.000	2.242	2.242	10%	FIT ≤ 0.001
(b) Geographic origin						
Among Populations	1	47.876	47.876	0.101	0%	FST ≤ 0.082
Among Individuals	93	3664.629	39.405	18.581	89%	FIS ≤ 0.001
Within Individuals	95	213.000	2.242	2.242	11%	FIT ≤ 0.001

d.f. = degree of freedom, SS = sum of squares, MS = mean sum of squares, Est. Var = Estimated variance, Per Var.= Percent variation, FST = Fixation index.

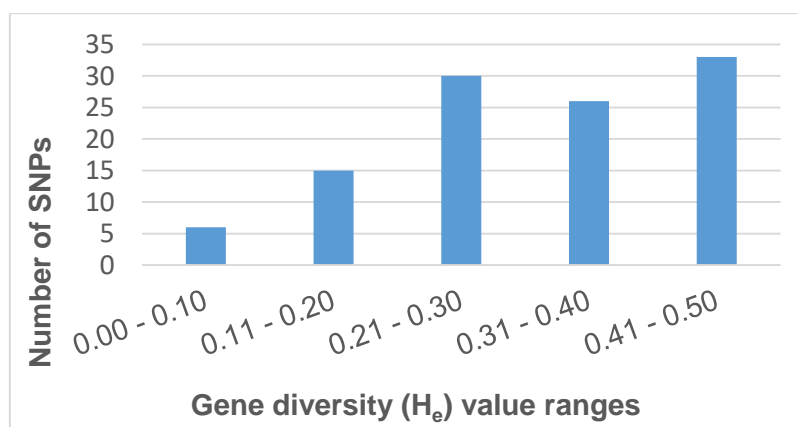


Figure 2. Gene Diversity (H_e) values of KASP markers used for diversity analysis.

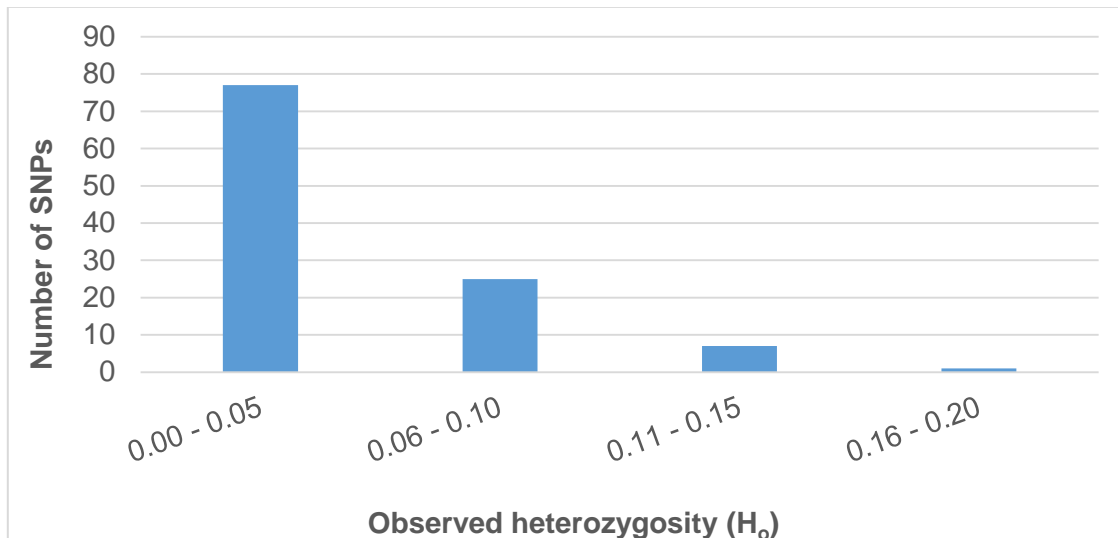


Figure 3. Observed Heterozygosity (H_o) values of KASP markers used for diversity analysis.

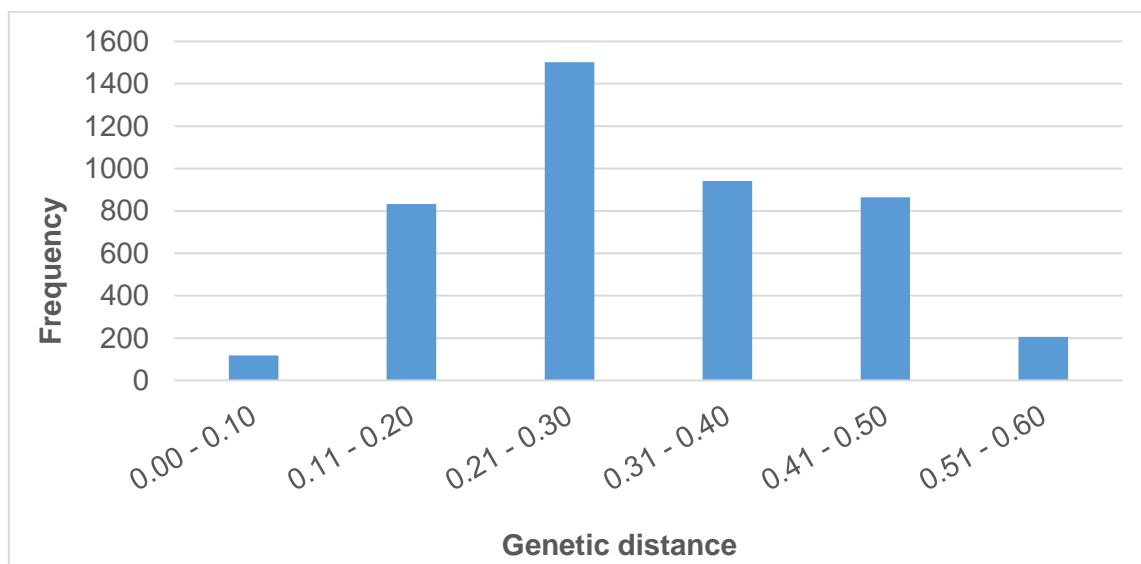


Figure 4. Summary of genetic distance values between 95 sorghum accessions using 110 KASP markers.

Genetic distance and clustering

The genetic distance between accessions in the current study ranged from 0.0 between SA0672 and SA0673, SA1282 and SA0670 to 0.57 between Hakika and SA1442, with an average mean of 0.30. Accessions SA0672 and SA0673 are very similar as they are both R-lines originating from the United States. Although SA1282 and SA0670 are very similar, they are, however, B- and R-line, respectively and their similarity could be due to common ancestry. In contrary, Hakika and SA1442 are very dissimilar owing to the fact that Hakika is an R-line introduced from ICRISAT-Kenya and SA1442 is a B-line from United States. Similar accessions clustered together whereas those not similar clustered differently. The average genetic distance between sorghum genotypes in the current study using KASP markers is lower than in other sorghum studies that used SSR markers. Ng'uni et al. (2011) reported highly variable genetic distance between Zambian sorghum accessions varying from 0.198 to 3.75 using 10 SSR markers. Using the same approach Mofokeng et al. (2014) reported genetic distances ranging from 0.0 to 8.4 with an average of 5.7 in South African sorghum germplasm using 30 SSR markers. The use of KASP

markers to estimate genetic distances has also been reported in other crops that include a study by Hiremath et al. (2012) who reported wide genetic distances in Chickpea genotypes ranging from 0.02 to 0.74 with an average of 0.37. Khera et al. (2013) also employed KASP markers in groundnuts and reported genetic dissimilarity between groundnuts genotypes ranging from 0.0 to 0.45 with an average of 0.13. Estimates of genetic distances between pairs of genotypes are important in parental selection for maintenance of genetic diversity in breeding programmes. Cluster analysis based on neighbour-joining algorithm using the UPGMA method revealed clear grouping of 95 sorghum accessions into three major groups. Cluster I was the smallest and consisted of only four genotypes SA2714, SA0117, SA2020 and SA1117. Cluster II had 27 genotypes comprising 13 genotype from SADC region and 14 R-lines from United States. This cluster was mostly dominated by R-lines from United States, with an exception of three B-lines (SA0221, M100 and SA1092) from Swaziland, South Africa and Botswana, respectively. Cluster III was the largest and most diverse cluster, which consisted of 64 genotypes. This cluster comprised mostly of 42 B-lines followed by 13 R-lines

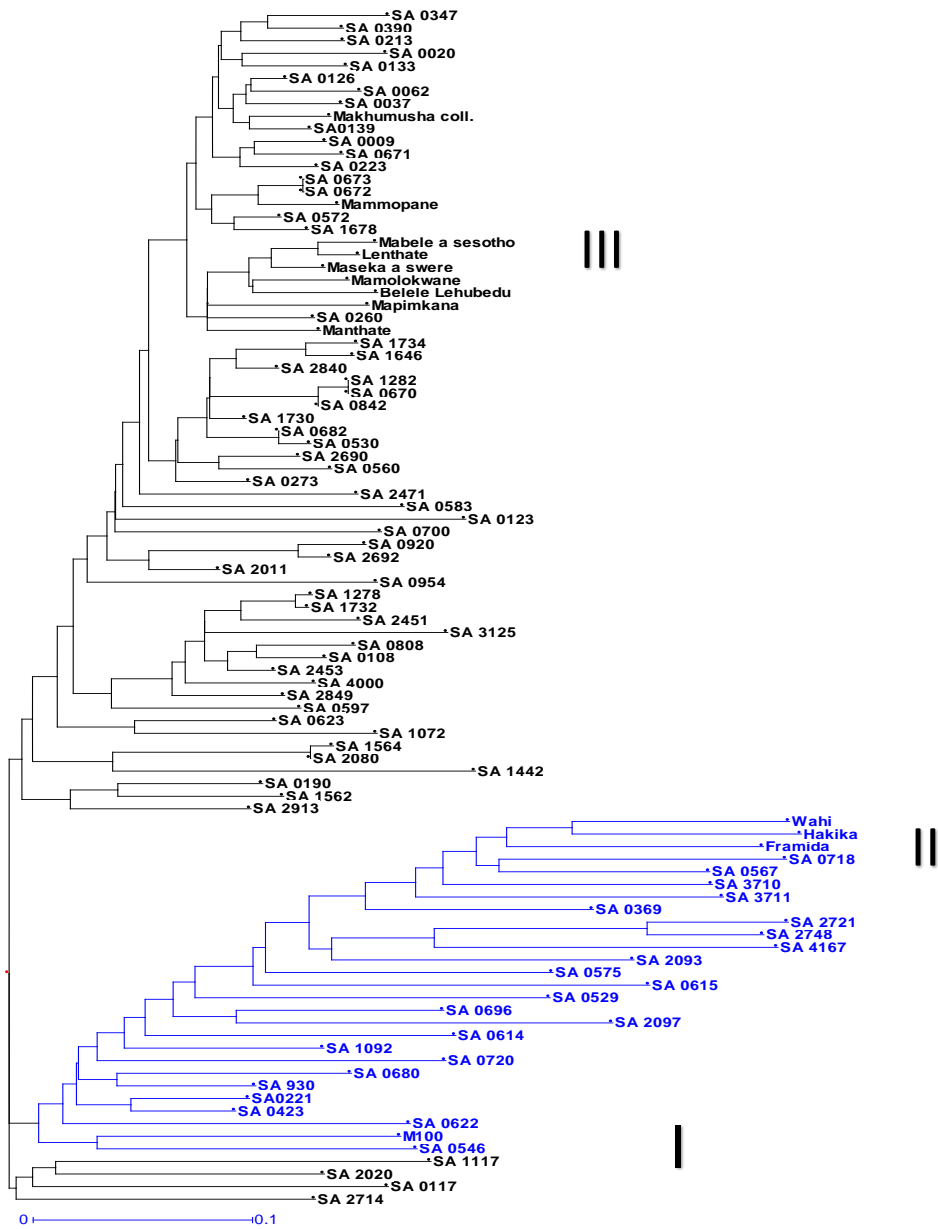


Figure 5. Dendrogram showing genetic relationship among 95 sorghum genotypes tested using 110 KASP markers.

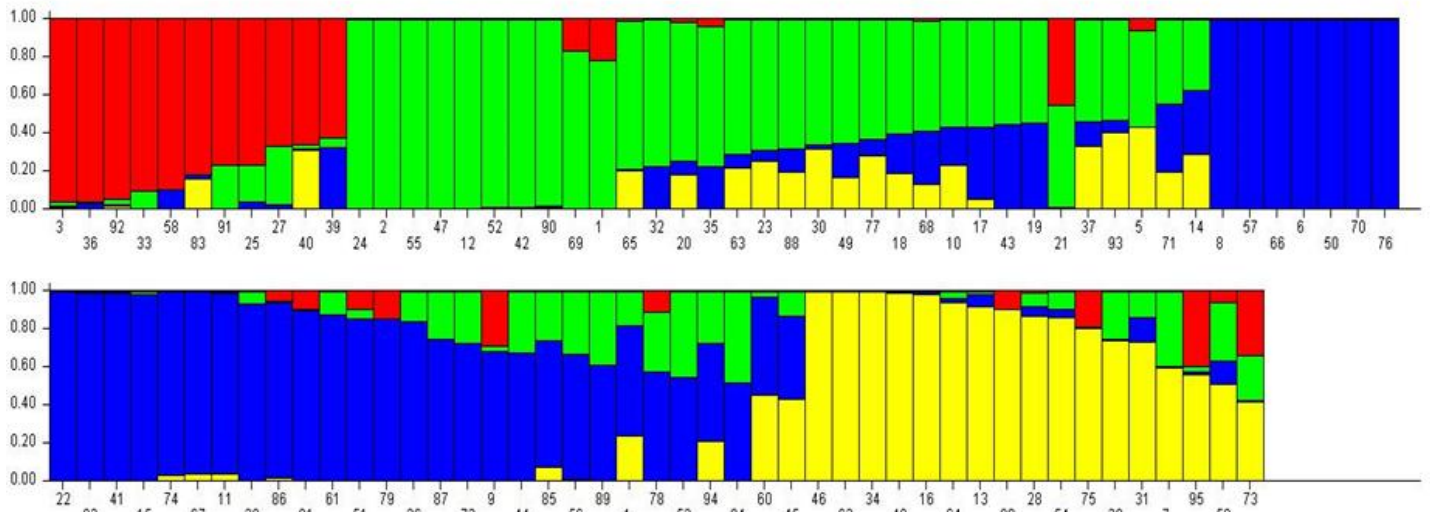


Figure 6. Model-based population structure analysis of sorghum accessions.

and nine landraces from different geographical regions. The clustering of genotypes was independent of geographical origin in that genotypes from different geographic origin were clustered in the same group. The clustering pattern in the current study is similar to the study by Sejake et al. (2020) who employed agro-morphological traits in the same sorghum accessions to assess genetic diversity, and grouped the accessions into four clusters.

Population structure

Understanding the population structure of the ARC National Sorghum Germplasm Bank is imperative to establish new strategies to preserve and utilize its genetic diversity, and is also a prerequisite for Genome Wide Association Studies (GWAS). The present study identified four major groups, in which cluster size ranged from 11 to 35 accessions. Several other studies have reported on the analysis of population structure of sorghum, some of which include Amelework et al. (2015, 2016); Weerasooriya et al. (2016); Cuevas et al. (2018); Pandian et al. (2019). In the current study, the smallest cluster (I) consisted of B- and R-lines from USA and the SADC region, while one B-line was the exception in the cluster. Accessions in this cluster were characterized by early flowering, low panicle weight, thousand grain weight and grain yield. Cluster II was dominated by 16 B-lines and two R-lines from the SADC region followed by 10 B-lines and four R-lines from USA. In this cluster, accessions had short leaf length and plant height, and lowest number of seeds per panicle. The largest cluster (III) was dominated by 14 B-lines and nine landraces from South Africa. This was followed by eight R-lines and four B-lines from USA. This cluster was characterized by late flowering, tall plants, high panicle weight, thousand seed weight and grain yield. The last cluster (IV) consisted of 17 R-lines from USA and SADC region, which were characterized by long leaf length, wider leaf and panicle width. Sejake et al. (2020) reported phenotypic characterization of these sorghum accessions used in the current study.

Materials and methods

Plant material

A total of 100 accessions was randomly selected from over 4000 sorghum accessions maintained at the ARC National Sorghum Germplasm Bank in Potchefstroom, South Africa. The accessions consisted of B-lines, R-lines and landraces collected over the years from different parts of Southern Africa and the world (Table S1).

DNA isolation and KASP genotyping

A total of 100 sorghum accessions were grown in a greenhouse for three to four weeks, and DNA was extracted from young fresh leaves using the modified Hexa-deCyl Trimethyl Ammonium Bromide (CTAB) method of Saghai-Moroof et al. (1984). Normalized DNA samples (50 ng/μl) were shipped for genotyping to LGC Genomics in United Kingdom (UK). The genotyping comprised of 136 KASP markers that had coverage of the 10 chromosomes of sorghum (Table S2). The 136 KASP markers were selected from the LGC Genomics database and each marker was selected within the interval of 10 centiMorgan (cM) along each sorghum linkage group to cover the whole sorghum genome (Table S3). KASP genotyping reaction consisted of DNA sample, KASP assay mix and KASP master mix. KASP assay comprised of two allele-specific forward primers, one

reverse primer and SNP assay provided in a single 2D barcoded tube. KASP master mix comprised of Fluorescence Resonance Energy Transfer (FRET) reporting system (universal FAM and HEX labelled cassettes), DNA polymerase, ROX passive reference dye, (Deoxyucleotide Triphosphates) dNTPs, Magnesium Chloride (MgCl₂) and an optimized buffer. A volume of 4μl of DNA sample was transferred into 96-well PCR plates. A master-mix was prepared by adding 40μl of 2x reaction mix and 1 μl of assay mix, and then 4.2μl of master mix was added to each DNA sample in the PCR plates. The plates were heat sealed and placed in the PCR thermal cycle programmed at 94°C for 15 minutes, 94°C for 10 seconds and 60°C for one minute for 35 cycles followed by 35°C for 30 seconds. After the KASP PCR run, the plates were fluorescently read and the data analyzed (Full protocol available at <http://www.lgcgenomics.com/genotyping/kasp-genotyping-reagents/how-does-kasp-work>).

Data analyses

Of the 136 markers, 26 markers fell within a category that was not used for further analyses. This category included markers with null alleles, monomorphic SNPs, SNPs with minor allele frequency of less than 2% and SNPs that had missing alleles for more than 30% of the genotypes. The 110 markers were further used for the study. Genotypes SA0109, SA0526, SA0550, SA0755, and SA1556 were also not used for further analyses because they either did not have any band amplified on them for all the SNPs or amplified bands for a very few markers. Genotypic data were subjected to analyses with various measures of genetic diversity within and among genotypes using GenAlex software version 6.5 (Peakall and Smouse, 2012). Genetic diversity parameters such as total number of alleles per locus (N_s), number of effective alleles per locus (N_e), observed heterozygosity (H_o), Shannon's Information Index (I), gene diversity (H_e) and polymorphic information content (PIC) were determined following protocol of Nei and Li (1979). To examine the degree of population differentiation, other genetic parameters such as differentiation (FST), gene flow (Nm), Nei's unbiased genetic distance (GD) and identity (GI) were estimated using GenAlex. The partitioning of total genetic variation into, within and among populations was done using the molecular analysis of variance (AMOVA) procedure of GenAlex.

Cluster analysis was done based on neighbor-joining algorithm using the un-weighted pair group method with arithmetic average (UPGMA) procedure in DARwin 5.0 software (Perrier and Jacquemoud-Collet, 2006). A dendrogram was generated on the dissimilarity matrix. To investigate the genetic relationships among accessions, genetic distances between all pairs of individual accessions were estimated to draw a dendrogram. Bootstrap analysis was performed for node construction using 10,000 bootstrap values.

To assess population structure of the sorghum accessions, model-based approach was used on STRUCTURE software version 2.3.4 (Pritchard et al., 2000). The burn-in periods were set at 10 000 with 10 000 Markov chain-Monte Carlo (MCMC) repetitions after each burn-in. The K-value was set between 1 and 10 and STRUCTURE was ran with 20 iterations for each sub-population (K). The most likely value for K was estimated using Evanno's ΔK method (Evanno et al., 2005) by STRUCTURE HARVESTER (Earl and vonHoldt, 2012).

Conclusions

This study assessed the extent of genetic diversity and population structure in sorghum accessions maintained at ARC National Sorghum Germplasm Bank in South Africa using SNP based KASP markers. The accessions were randomly selected from the gene bank and the KASP markers were selected within every 10 cM of the sorghum chromosomes. The results of this study highlight that the germplasm used were genetically diverse and they can be used as a good foundation to select potential genotypes for further genetic improvement and broadening of the genetic base of the crop. The accessions were clearly separated based on their breeding status and geographic origin and they formed three major clusters on UPGMA method. However, model-based population analysis identified four distinct groups. The most diverse genotypes were SA1442, Hakika, SA1117 and SA0123 based on the KASP markers. These genotypes, together with others on the clusters, can be used as parents in crossing programmes and germplasm conservation.

Conflict of interest

The authors have not declared any conflict of interests.

Acknowledgements

This work was supported by the Research and Technology Fund/National Research Foundation [RTF150422117506, 2016].

References

- Amelework B, Shimelis H, Tongoona P, Laing M, Mengistu F (2015) Genetic variation in lowland sorghum (*Sorghum bicolor* (L.) Moench) landraces assessed by simple sequence repeats. *Plant Genetic Resources: Characterization and Utilization* 13:131-141.
- Amelework B, Shimelis H, Tongoona P, Laing M, Mengistu F (2016) Genetic diversity of lowland sorghum landraces assessed with morphological and microsatellite markers. *Australian Journal of Crop Science* 10 (3):291-298.
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32 (3):314-31.
- Bucheyeki TL, Gwanama C, Mgonja M, Chisi M, Folkertsma R, Mutegi R (2009) Genetic variability characterisation of Tanzania sorghum landraces based on simple sequence repeats (SSRs) molecular and morphological markers. *African Crop Science Journal* 17:71-86.
- Cuevas HE, Prom LK, Rosa-Valentin G (2018) Population structure of the NPGS Senegalese sorghum collection and its evaluation to identify new disease resistant genes. *PLoS ONE* 13 (2):1-15.
- Department of agriculture, forestry and fisheries (DAFF), (2010) Annual report. South Africa 214p.
- Earl DA and vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359-361.
- Elangovan M, Kiran babu P, Seetharama N, Patil JV (2013) Genetic diversity and heritability characters associated in sweet sorghum (*Sorghum bicolor* L. Moench). *Sugar Tech* 16 (2):200-210.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611-2620.
- Gerrano AS, Labuschagne MT, van Biljon A, Shargie NG (2014) Genetic diversity assessment in sorghum accessions using qualitative morphological and amplified fragment length polymorphism markers. *Scientia Agricola* 71:394-401.
- Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genetics Research International* 2015:1-14.
- Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, Whaley AM, Carrasquilla-Garcia N, Gaur PM, Upadhyaya HD, Kishor PB, Shah TM, Cook DR, Varshney RV (2012) Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnology Journal* 10:716-732.
- Khera P, Upadhyaya HD, Pandey MK, Roorkiwal M, Sriswathi M, Janila P (2013) Single nucleotide polymorphism-based genetic diversity in the reference set of peanut (*Arachis spp.*) by developing and applying cost-effective Kompetitive allele specific polymerase chain reaction genotyping assays. *Plant Genome* 6:3-10.
- Koenders D (2010) Feasibility study to include sorghum and meat into the WFP basket in Kenya. Report for the WFP, Kenya.
- Mbeyagala EK, Kiambi DD, Okori P, Edema R (2012) Molecular diversity among sorghum (*Sorghum bicolor* (L.) Moench) landraces in Uganda. *International Journal of botany* 8 (3):85-95.
- Mengistu G, Shimelis H, Laing M, Lule D, Assefa E, Mathew I (2020) Genetic diversity assessment of sorghum (*Sorghum bicolor* (L.) Moench) landraces using SNP markers. *South African Journal of Plant and Soil* 37:220-226.
- Mofokeng A, Shimelis H, Tongoona P, Laing M (2014) A genetic diversity analysis of South African sorghum genotypes using SSR markers. *South African journal of plant and soil* 31 (3):145-152.
- Mohammadi SA and Prasanna BM (2003) Analysis of genetic diversity in crop plants - salient statistical tools and considerations. *Crop Science* 43:1235-1248.
- Mothaodi T, Geleta M, Bryngelsson T, Fatih M, Chite S, Ortiz R (2014) Genetic diversity in ex-situ conserved sorghum accessions of Botswana as estimated by microsatellite markers. *Australian Journal of Crop Science* 8 (1):35-43.
- Narain P (2000) Genetic diversity conservation and assessment. *Current Science* 79:170-175.
- Nei M and Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 76 (10):5269-73.
- Ng'uni D, Geleta M, Bryngelsson T (2011) Genetic diversity in sorghum (*Sorghum bicolor* (L.) Moench) accessions of Zambia as revealed by simple sequence repeats (SSR). *Hereditas* 148:52-62.
- Ng'uni D, Geleta M, Hofvander P, Fatih M, Bryngelsson T (2012) Comparative genetic diversity and nutritional quality variation among some important southern African sorghum accessions [*Sorghum bicolor* (L.) Moench]. *Australian Journal of Crop Science* 6 (1):56-64.
- Pandian S, Satish L, Shilpha J, Ramesh M (2019) Genetic Diversity Analysis reveals strong population structure in sorghum germplasm collection. *Proceedings of the National Academy of Sciences* 90 (1):179-190.
- Peakall R, Smouse P (2012) GenAIEx V5: Genetic Analysis in Excel. *Populations Genetic Software for Teaching and Research. Bioinformatics* 28 (19):2537-9.

- Perrier X, Jacquemoud-Collet JP (2006) DARwin software. Website available at: <http://darwin.cirad.fr/darwin> (Accessed on 29 August 2020).
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Roorkiwal M, Von Wettberg EJ, Upadhyaya HD, Warschefsky E, Rathore A (2014) Exploring germplasm diversity to understand the domestication process in *Cicer* spp. using SNP and DArT Markers. *PLoS ONE* 9 (7):1-10.
- Rowhani P, Lobell DB, Linderman M, Ramankutty N (2011) Climate variability and crop production in Tanzania. *Agricultural and Forest Meteorology* 151:449-460.
- Saghai-Marooif MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphism in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 81:8014-8019.
- Sami RA (2013) Heritability studies in some sweet sorghum (*Sorghum bicolor* L. Moench) genotypes. *Journal of biology, Agriculture and Healthcare* 3 (17):49-51.
- Sejake T, Shargie N, Christian R, Tsilo TJ (2020) Assessment of genetic diversity in sorghum germplasm using agromorphological traits. *South African Journal of Plant and Soil* 2020:01-13.
- Semagn K, Bjørnstad Å, Ndjondjop MN (2006) An overview of molecular marker methods for plants. *African Journal of Biotechnology* 5 (25):2540-2568.
- Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal* 27:379-423.
- Smith JS, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. *Crop Science* 40:226-232.
- Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K, Ordon F (2003) Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. *Theoretical and Applied Genetics* 106 (7):1316-1325.
- Weerasooriya DK, Maulana FR, Bandara AY, Tirfessa A, Ayana A, Mengistu G, Nouh K, Tesso TT (2016) Genetic diversity and population structure among sorghum (*Sorghum bicolor*, L.) germplasm collections from Western Ethiopia. *African Journal of Biotechnology* 15 (23):1147-1158.
- Winter P, Kahl G (1995) Molecular marker technologies for plant improvement. *World Journal of Microbiology and Biotechnology* 11:438-448.