



OPEN

Genetic diversity and population structure of maize inbred lines using phenotypic traits and single nucleotide polymorphism (SNP) markers

Sweetbird Phindile Dube^{1,2✉}, Julia Sibiya¹ & Funso Kutu²

Understanding germplasm's genetic diversity is essential for developing new and improved cultivars with stable yields under diverse environments. The objective of this study was to determine the genetic diversity and population structure of 128 maize inbred lines sourced from the International Institute of Tropical Agriculture (IITA), the International Maize and Wheat Improvement Centre (CIMMYT), and the University of KwaZulu-Natal (UKZN) using 11,450 informative single nucleotide polymorphism (SNP) markers. The inbred lines revealed highly significant ($p < 0.001$) levels of variability for the key phenotypic traits. The SNP markers had a mean gene diversity (GD) and polymorphic information content (PIC) of 0.40 and 0.31, respectively, indicating the existence of substantial genetic variation across the germplasm panel. The model-based population structure analysis identified three subpopulations ($K = 3$) among the inbred lines. This corroborated the phylogenetic analysis using phenotypic traits and molecular markers which classified the inbred lines into three groups. The findings of this study identified considerable genetic diversity for the selection of inbred lines with favourable alleles for multiple traits and could be useful to initiate marker-assisted selection (MAS) to identify significant loci associated with agronomic performance and multiple-stress tolerance.

In South Africa, the maize (*Zea mays* L.) crop is a significant employer and source of foreign currency due to its multiplier effects. The crop has strong linkages throughout the economy, upstream to the input industries and downstream into the milling, animal feed, and food processing industries^{1,2}. The hectareage planted with maize in South Africa varies depending on weather and market conditions, but on average, 2.5 to 2.75 million hectares of hybrid maize are planted each year, yielding 12 million tons of grain³. This translates to surplus maize exported to Zimbabwe, Botswana, Lesotho, Namibia, Eswatini, and Mozambique. South Africa is the only country in Africa that is in the top 10 of maize producers in the world⁴. However, climate change-induced biotic and abiotic stresses threaten maize productivity in tropical areas. Maize is severely affected by extreme weather events such as high temperatures, and unpredictable rainfall patterns, resulting in heat and drought stress or flooding⁵. Approximately 90% of maize in South Africa is produced on dry land under rain-fed conditions⁶.

Rainfall has decreased drastically over the past few years causing a dramatic decline in maize volumes due to the El Nino induced droughts. The variations in the frequency and amount of rainfall received have deleterious effects on maize grain yields mainly when water deficits occur during flowering (anthesis and silking)⁷. In addition, frequent droughts and the rise in atmospheric temperatures create conducive conditions for spawning of pests, diseases, and parasitic weeds causing severe maize grain yield losses⁸. Maize ear rots and stalk rots are pathogens showing an increasing impact in changing climates⁹. In Southern Africa, due to changes in environmental conditions, Refs.^{10,11} observed an increase in the incidence and severity of maize foliar diseases such as *Phaeosphaeria* leaf spot (PLS; causal agent *Phaeosphaeria maydis* (Henn.) Rane, Payak & Renfro) and northern leaf blight (NLB; causal agent *Exserohilum turcicum* Pass. Leonard & Snuggs) with previously resistant cultivars being affected. Thus, the selection of germplasm with improved resistance or tolerance to biotic and abiotic

¹School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa. ²School of Agricultural Sciences, University of Mpumalanga, Nelspruit 1200, South Africa. ✉email: phindile.dube@ump.ac.za

stresses is critical in developing resilient maize production systems adapted to climate change-induced stresses. The sub-Saharan Africa is a centre of diversity for tropical and subtropical maize which has evolved stress adaptation genes for wide adaptation in the harsh savannah climates. Hence, this germplasm can expand the genetic diversity required for the development of climate-resilient maize cultivars in South Africa.

The Consultative Group of International Agricultural Research (CGIAR) centers comprising the International Maize and Wheat Improvement Centre (CIMMYT) and International Institute of Tropical Agriculture (IITA) have contributed immensely to the genetic improvement of tropical maize in Africa. Several inbred lines with tolerance to a diverse range of biotic and abiotic stresses have been developed by CIMMYT and IITA¹². The CIMMYT program in Africa is divided into the East and Southern Africa regional hubs. Yield, drought, heat, low N, Low P, *Striga*, grey leafspot, maize streak virus, rusts, and ear rots have been the primary targets for selection at the two hubs¹³.

However, the recent fall armyworm outbreaks in Southern Africa and the maize lethal necrosis virus (MLN) in East Africa presented new challenges warranting immediate action. To date, several multiple-stress tolerant CIMMYT maize lines (CMLs), including CMLs 536–544 and 571–572, and 587–592, bred in Southern Africa are available for public and private maize breeding programs in Africa¹⁴. The MLN Screening facility in East Africa, Kenya identified notable resistance to the virus, and integrated the resistance genes into African-adapted germplasm. Currently, maize is a principal cereal in West Africa because of the Central Africa Maize Collaborative Research Network (WECAMAN) established by IITA and all the national programs in that region. The IITA's breeding strategy of combining *Striga* resistance, earliness, drought- and low-N tolerance enhanced maize yield production in multiple stress environments. Maize grown in West and Central Africa should mature early before the onset of terminal drought which triggers *Striga* parasitic infections. It is noteworthy that both programs integrate elite exotic germplasm to expand the genetic and phenotypic elasticity of tropical and subtropical maize in the region. For instance, some of CIMMYT's inbred lines developed in Africa are derived from Tuxpeño Sequía, and La Posta Sequía populations¹⁴, with the CIMMYT heterotic group A coming from populations and lines including the Tuxpeño, Kitale, BSSS (Iowa Stiff Stalk Synthetic), B73 and Salisbury white (N), in contrast the CIMMYT group B contains populations and lines from ETO, Ecuador 573, Lancaster, Mo17 and Southern cross (SC)^{10, 15}. Elite exotic germplasm introduces valuable new alleles that broaden the genetic divergence required to exploit heterosis for hybrid breeding¹⁶.

South Africa is the top maize producer in SSA, due to a vibrant private sector-owned maize seed industry, bolstered by biotechnology. About 85% of the maize grown in South Africa is genetically modified (GM) maize¹⁷, containing GM traits such as glyphosate and insect resistance, thus improving grain yield. While these traits have resulted in considerable yield gains compared to conventional maize production, introgression of these traits will potentially result in deleterious genetic 'bottlenecks'. In addition, most commercial breeding programs developed new inbred lines by recycling elite lines via pedigree which also narrows genetic divergence. Furthermore, most maize hybrid breeding programs, particularly in southern and eastern Africa, have utilized elite maize inbred lines from nine heterotic groups^{10, 15} and derived from the "P" heterotic group (derivatives from Natal Potchefstroom Pearl, the SC, N and K64r derivatives as well as the broader CIMMYT A and B groups^{15, 18} among others. There is a need, to broaden the genetic diversity of local germplasm by integrating tropical and subtropical germplasm adapted to SSA environments¹⁹. This is because native traits will be critical in improving the resilience of maize production systems under threat from the negative effects of climate change.

Knowledge of genetic relationships between local maize germplasm, and tropical and sub-tropical maize inbred lines will guide parental selection and the devising of appropriate breeding designs. Genetic diversity analysis differentiates breeding populations and assists in classifying inbred lines into definite heterotic groups^{20, 21}. Trait phenotyping can be used in conjunction with molecular makers in the designation of entries into heterotic groups, particularly for highly heritable traits²². While morphological and agronomic polymorphisms are highly amenable to the environment and developmental stage of the plant, molecular markers, on the other hand, are very stable. Single nucleotide polymorphisms (SNPs) are the predominant form of naturally occurring genetic variation and offer adequate variation to distinguish closely related individuals. As a result, they have become the preferred DNA marker of choice for diversity studies. Additionally, SNPs are useful for studying genetic variation due to their low cost per data point, lower genotyping error rates, locus-specificity, high genomic abundance, potential for high throughput analysis, and codominance^{23–25}.

Thus, SNPs provide an opportunity for the assessment of genetic diversity among local South African inbred lines and exotic tropical and subtropical maize germplasm developed for national maize breeding programs in SSA²⁶. Gaining insight into how locally bred lines are related to lines developed by the CGIAR, suffices in the exploitation of heterosis in local breeding programs. Therefore, the objectives of this study were to assess genetic diversity and interrelationships present among locally bred maize inbred lines and the CGIAR tropical and subtropical maize lines adapted to SSA conditions, to determine heterotic groups and select unique genotypes for biotic and abiotic stress tolerance breeding. The set of local South African germplasm used in the study represents important genetic resources for use in public and private maize breeding programs in the country.

Materials and methods

Source of plant material

One hundred and twenty-eight (128) diverse maize inbred lines obtained from the International Maize and Wheat Improvement Centre (42), the International Institute of Tropical Agriculture (50), and the University of KwaZulu-Natal (36) were used in this study. The CIMMYT and IITA inbred lines were selected for their nutritional content and resistance/tolerance to biotic and/or abiotic stresses. The UKZN inbred lines were chosen to represent a sample of locally developed inbred lines to be used in commercial hybrid maize breeding. The CIMMYT lines were developed in Kenya and Zimbabwe (East and Southern Africa regional hubs, respectively) and

were derived from the Tuxpeño Sequía, La Posta Sequía, and the Drought-Tolerant Population (DTP) yellow and white gene pools. The IITA germplasm was bred in Nigeria from several broad-based germplasm sources with resistance to *Striga* and MSV, as well as tolerance to drought, including TZE-W Pop DT STR C0, TZE-Y Pop DT STR C0, and TZE Comp 5-Y C6. The description of the germplasm used in the study is summarised in Table S1.

Description of experimental sites phenotyping, data collection, and data analysis

The test genotypes were evaluated under field conditions to complement molecular marker data. Two field trials were established at Ukulinga Research farm (29° 40' S, 30° 24' E, 800 m) of UKZN and Cedara Research Station (29° 32' S, 30° 17' E, 1076 m) during the 2020/2021 cropping seasons. The experiments were arranged in 8 × 16 alpha lattice design with 2 replicates. Each plot consisted of 1 row, 5 m long, with inter and intra-row spacing of 0.8 and 0.3 m, respectively. The plant population density was 41,666.6 plants ha⁻¹. Field management followed the recommended maize agricultural practices.

Plant count (PC) was recorded as the total number of plants per plot counted after thinning. Days-to-anthesis (DA) and days-to-silking (DS) were recorded as the number of days from emergence to the date when 50% of the plants in a plot had produced pollen and silks, respectively. Anthesis-silking interval (ASI) was calculated as days to silking minus days to anthesis. Plant height (PH) expressed in centimetres (cm) was measured from base to tassel branching to the soil surface to the top of the tassel of 10 representative plants selected from each plot. Scoring the severity of rust (RST) disease of maize was done using a scale of 1–9, where 1 = 0% of leaf surface diseased (no rust) and 9 = 81–100% of leaf surface diseased. Field weight (FW) was measured as the total weight of unshelled cobs in kilograms (kg). Cob count (CC) was counted as number of cobs per plot at harvest. The number of kernel rows per ear (KR) and kernels per row per ear (KRE) were determined as counts. Ear length (EL) was recorded (cm) as the length of a cob from the tip to the base using the ruler, while the diameter of the cob (ED) was measured using a vernier calliper (mm). Grain moisture (MOI) content was measured using a moisture tester during shelling. Shelling percentage (SHP) was calculated as field weight minus grain weight and the difference was expressed as a percentage. A sample of 100 kernels, grain weight per ear (GWE) from each plot was weighed. Grain yield (GY) expressed in t ha⁻¹ was obtained from grain weight per plot adjusted to 12.5% grain moisture following CIMMYT (1999).

$$GY = \left(\frac{GW}{NP} \right) \times 10 \times \left(\frac{100 - MO}{87.5} \right),$$

where GY—grain yield (kg/ha), GW—grain weight at harvest (kg/ha), MO—moisture content (%) of grains at harvest, 87.5—Standard dry matter of grain at 12.5% as required by maize grain market authorities in South Africa, NP—Net plot area (number of rows × intra-row spacing × number of stations × inter-row spacing).

Data generated were subjected to analysis of variance after testing for normality and homogeneity of variance using R software. Trait BLUPs were computed using DeltaGen (<https://www.deltagen.agr.nz>). Furthermore, the association among traits was deduced using Pearson correlation coefficients on the corplot package in R software²⁷.

Genotyping

DNA extraction, genotyping, and quality control

Fresh leaf samples were collected from 3 weeks old seedlings within each genotype and was shipped to SEQART AFRICA (<https://www.seqart.net>) in Nairobi, Kenya for DNA extraction and genotyping by sequencing. The DARtSeq protocol was used to genotype samples using a SNP chip covering the ten maize chromosomes. The initial 50,941 SNPs from the genotyping-by-sequencing (GBS) pipeline were filtered by imputation to remove SNPs with 20% missing data and 5% minor allele frequency (MAF) using the *snpReady* package on R²⁸. A total of 11,450 informative SNP markers were used after data imputation. Individual genotypes with > 20% missing data were removed.

Genetic diversity and population structure analysis

The polymorphic information content (PIC), major allele frequency, the number of alleles, heterozygosity, and gene diversity were estimated using the R package “adegenet”²⁹. Analysis of molecular variance (AMOVA) and genetic diversity was performed using GenAlex³⁰ after grouping the inbred lines based on the source of collection. Population structure was determined using the STRUCTURE v2.3.4 software³¹. The burn-in period and Markov Chain Monte Carlo (MCMC) length were set at 10,000 iterations, and the model was run by varying the number of clusters (K) from 1 to 10 with 10 alterations for each K. The appropriate K value was estimated by implementing the Evanno method using the STRUCTURE Harvester program³². A joint analysis of phenotypic and genotypic data was conducted. A phenotypic distance matrix was generated based on Gower’s distance, while the genotypic distance matrix was generated using Jaccard’s coefficient on the R software²⁷. The phenotypic, and genotypic matrices were used to generate hierarchical clusters using the package “cluster” in R software²⁷.

Results

Analysis of phenotypic traits and relationships among traits

The combined analysis of variance for the 15 quantitative traits studied is shown in Table 1. Analysis of variance revealed highly significant differences ($p < 0.001$) among genotypes for all the studied traits. The environmental factor exhibited significant differences for all the evaluated traits except for the days to anthesis (DA), ear length (EL), ear diameter (ED), and kernels per row per ear (KRE). The replicates revealed significant differences for days to days to 50% anthesis and silking (DA and DS), plant height (PH), cob count (CC), field weight (FW),

Source	DF	PC	DA	DS	ASI	PH	EL	ED	KR	KRE	CC	FW	MOI	GWE	SHP	GY
Genotypes (G)	127	26.8***	83.7***	82.8***	3.8***	4755.0***	28.9***	173.9***	19.6***	161.3***	82.6***	3.9***	9.4***	0.0***	73.1***	7.1***
Environment (E)	1	23.6*	4.3	356.1***	438.8***	74,165.0***	11.9	2.1	38.6**	5.0	109.7*	37.1***	962.5***	0.0***	1507.8***	57.4***
Rep (R)	1	12.5	143.4**	127.0**	0.5	18,296.0***	0.6	81.3	1.2	7.1	87.0*	6.2***	4.6	0.0**	0.4	11.4***
GxE	127	8.1*	19.9	21.3	2.1*	660.0***	7.1**	65.7*	6.8**	34.2	26.5***	0.6***	3.5	0.0*	15.7	1.5***
ExR	1	0.1	43.4	18.4	5.3	63,401.0***	3.4	196.1*	26.7*	88.6	32.5	1.2	38.9**	0.0**	4.9	2.0
REP:BLK	42	6.4	26.8*	21.7	2.2*	623.0**	7.3*	71.3*	5.5	39.7	29.4**	0.5*	3.8	0.0*	15.1	1.5**
Residuals	212	5.8	18.0	16.8	1.5	323.0	4.8	48.8	4.3	28.0	16.3	0.3	4.1	0.0	14.2	0.9

Table 1. Analysis of variance showing mean square values for the 15 morpho-agronomic traits for 128 maize genotypes. Significant codes: 0 "***", 0.001 "***" and 0.01 "*".

grain weight per ear (GWE), and grain yield (GY). The genotype \times environment (GxE) interaction displayed significant differences ($p < 0.01$) among all the traits except for days to 50% anthesis and silking (DA and DS), gain moisture (MOI), and shelling percentage (SHP).

The mean performance of the genotypes showed that days to 50% anthesis and silking (DA and DS) ranged from approximately 83 to 91 days. Anthesis and silking interval range from -0.4 to 1.8 with a mean of 0.3 . Plant height ranged from 99 to 274 cm with a mean of 179 cm. Cob count exhibited a mean of 15 and ranged between 3 and 27 per plot. Grain weight per cob ranged from 0 to 0.2 g with a mean of 0.1 g. Considerable variation was observed for grain yield ranging from 0.9 to 5.9 t ha $^{-1}$ with a mean performance of 2.8 t ha $^{-1}$. The top performing genotypes in terms of grain yield comprised TZISTR1190, TZISTR1261, CML540, CML571, and TZISTR1119, recording 5.9 , 5.8 , 5.6 , 5.6 , and 5.5 t ha $^{-1}$, respectively. The least performing genotypes were CML545, 18 UK1-54 and TZISTR1162 recording 1.3 , 0.9 and 0.9 t ha $^{-1}$ respectively. The mean performances for all the phenotypic traits of the genotypes are summarized in Table S2.

The association among the 16 measured phenotypic traits is depicted by the heat map of Pearson correlation coefficients. The blue and red squares are negative and positive correlations, respectively (Fig. 1). Strong correlations were observed between grain yield and cob count (0.76), plant height (0.72), ear length (0.77), ear diameter (0.67), and field weight (0.95). However, grain yield was negatively and significantly correlated with days to anthesis (-0.18), days to silking (-0.22), anthesis-silking interval (-0.21) and rust (-0.74).

Genetic diversity analysis

Population structure and genetic diversity analyses

The gene diversity (GD) ranged between 0.03 and 0.50 , with a mean of 0.40 . While the polymorphic information content (PIC) range was between 0.03 and 0.38 , with a mean of 0.31 . The minor allele frequency (MAF) ranged from 0.01 to 0.50 , with an average value of 0.34 . The observed heterozygosity had the mean average of 0.67 varied between 0.59 and 0.83 . Fixation index (F) ranged from -1.1 to -0.5 with a mean of -0.69 (Table 2). The results from AMOVA displayed significant genetic differences among individuals and populations (Table 3). Two percent of the variance was due to genetic differentiation among the populations, in contrast individuals within the panel accounted for 98% of the variance. The pairwise population matrix of Nei Genetic Distance was the highest between genotypes from CIMMYT and UKZN (0.05) followed by IITA and UKZN (0.04) and lastly between CIMMYT and IITA with 0.03 . The greatest genetic identity was observed between CIMMYT and IITA (0.97) followed by IITA and UKZN (0.96) and the least was observed between CIMMYT and UKZN (0.95) (Table 4).

The population structure analysis identified three subpopulations among the inbred lines based on the optimal $K = 3$ determined according to Evanno's method (Figs. 2, 3). The distribution of genotypes into clusters was based on 70% kinship. Sub-populations 1, 2, and 3 comprised 54% , 16% , and 30% of the total genotypes, respectively. The allele frequency divergence between subpopulations and the expected heterozygosity between genotypes within the same subpopulations is presented in (Table 5). Subpopulations 1 and 2 exhibited the highest allele frequency divergence of 0.08 , followed by subpopulations 1 and 3 with 0.07 . The least allele frequency was recorded for subpopulations 2 and 3 with 0.12 . Within the three subpopulations, the expected heterozygosity between genotypes varied from 0.24 , 0.12 , and 0.16 for subpopulations 1, 2 and 3, respectively (Table 5).

Cluster analysis of phenotypic and genotypic data for the 128 genotypes

The dendrogram based on phenotypic data showed three clusters: I, II, and III, each consisting of 81% , 9% , and 10% membership, respectively (Fig. 4, Table S3). Cluster I was an admixture group with 103 genotypes, having 35 IITA, 36 CIMMYT, and 32 UKZN material. Most of the genotypes (TZSTR1102, 17CED MAK1-61/62, and 18 UK1-3) in this cluster were associated with high anthesis-silking interval (ASI) coupled with reduced grain yield. Cluster II consisted of most high yielding genotypes (TZISTR1261, TZISTR1261, CML540, CML571, and TZISTR1119), with 8 from IITA and 4 from CIMMYT. The third cluster comprised 13 genotypes associated with the lowest yield, the least being CML545 and 18 UK1-54 with each recording 0.9 t ha $^{-1}$.

Based on the genetic data, the dendrogram showed that Clusters I, II, and III consisted of 23% , 26% , and 51% membership, respectively (Fig. 5, Table S4). The first cluster consisted of the 30 genotypes with 29 from UKZN and 1 from CIMMYT. The second cluster consisted of 14 and 19 genotypes from IITA and CIMMYT, respectively. The last cluster was the largest, with 65 genotypes, of which 35 were from IITA. The 25 and 5 genotypes were

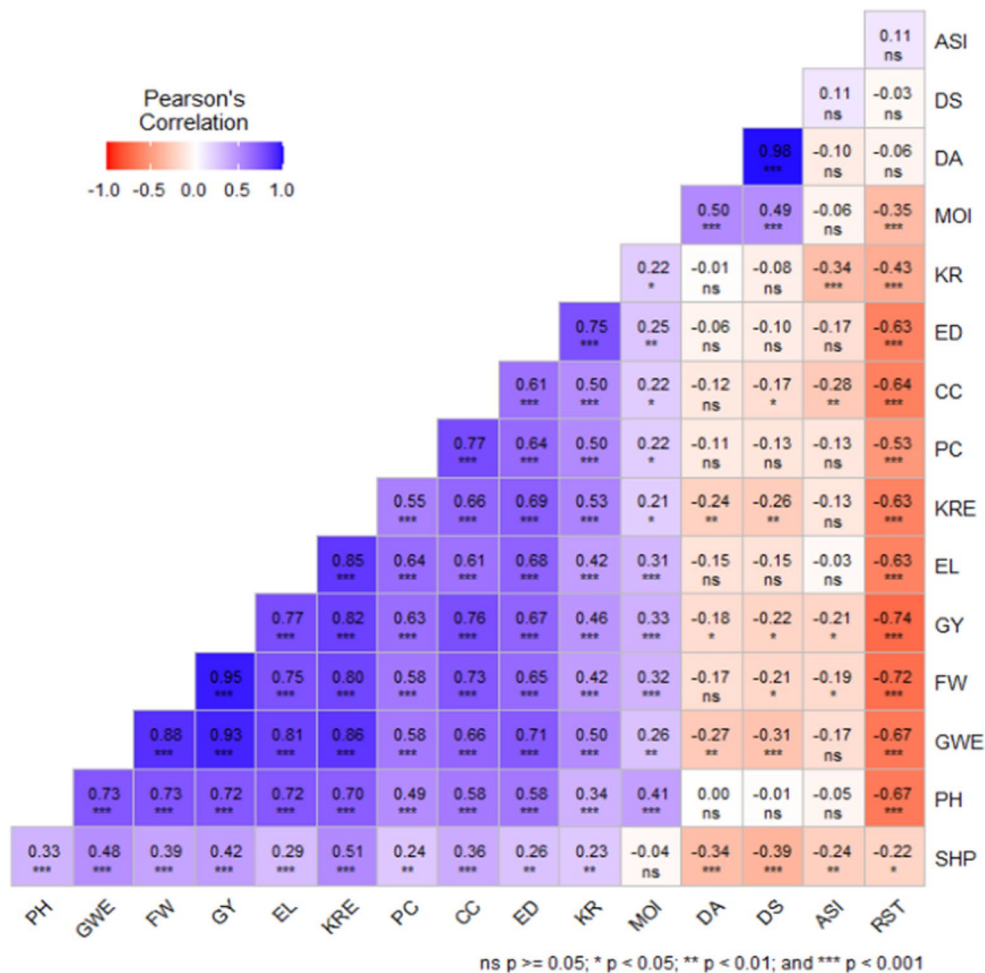


Figure 1. Heat map showing positive (blue squares) and negative (red squares) correlations.

Statistics	GD	PIC	MAF	Ho	F
Mean	0.40	0.31	0.34	0.67	-0.69
Lower	0.03	0.03	0.01	0.59	-1.1
Higher	0.50	0.38	0.50	0.83	-0.5

Table 2. Diversity indices statistics of the maize inbred lines based on SNP markers, *GD* gene diversity, *PIC* polymorphic information content, *MAF* minor allele frequency, *Ho* observed heterozygosity, *F* Fixation index.

Source	df	SS	MS	Est. var.	%
Among population	2	8460.69	4230.35	31.48	2
Within individuals	127	316,160.00	1638.13	1638.13	98
Total	253	421,428.05		1669.61	100

Table 3. Molecular analysis of variance of maize populations based on 11,450 SNP markers.

from CIMMYT and UKZN, respectively. The CIMMYT and IITA exchange a lot of planting material, which could be why they are grouped in the same cluster.

	UKZN	IITA	CIMMYT
UKZN	1	0.04	0.05
IITA	0.96	1	0.03
CIMMYT	0.95	0.97	1

Table 4. Pairwise population matrix of nei genetic distance (lower diagonal) and genetic identity (above diagonal).

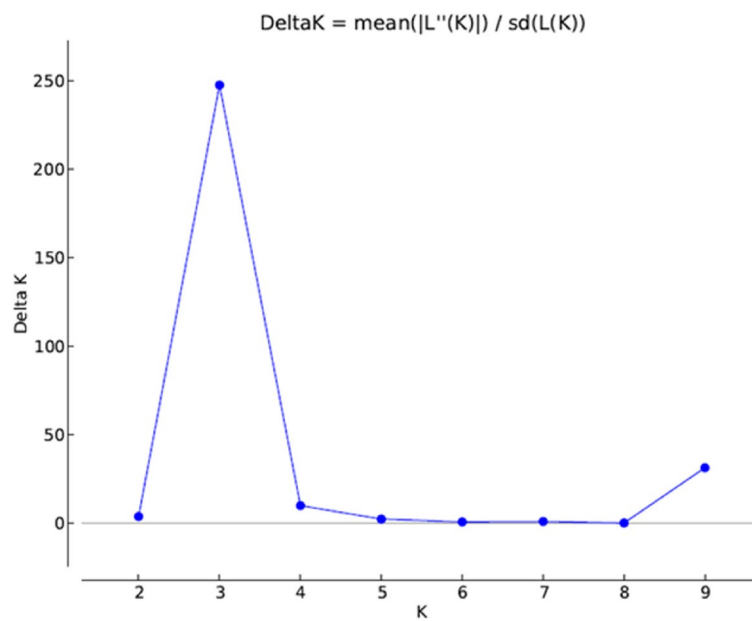


Figure 2. Delta K showing the number of populations.

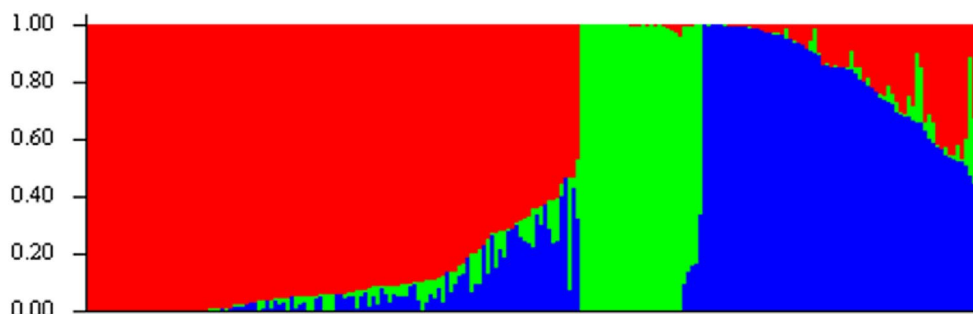


Figure 3. Bar plot of population sorted by kinship matrix.

Discussion

Genetic diversity studies are critical in identifying the individual genotypes among closely and distantly related groups for the start of new breeding initiatives. To breed novel hybrids with outstanding agronomic performance, multiple stress tolerance and high-quality yield, there is a need to understand the genetic interrelationships among the inbred lines. Hence, the rationale of this study was to gain genetic relationships among South African-bred maize inbred lines, tropical and subtropical inbred lines developed across the African continent.

The study revealed significant genetic variation among the inbred lines using phenotypic traits and SNP markers. Analysis of variance showed highly significant differences among genotypes for grain yield and its associated component traits, indicating substantial genetic variation among the genotypes. Grain yield and its associated yield components are quantitatively inherited and are of polygenic inheritance which can enable selection between genotypes within a population³³. The significant differences observed in days to 50% anthesis and silking (DA and DS) among genotypes are essential for crop development in drought-prone environments. Early

	FST allele frequency divergence among subpopulations			Membership%
	I	II	III	
I	–	0.0810	0.0737	0.54
II	0.0810	–	0.1159	0.16
III	0.0737	0.1159		0.30
Expected heterozygosity within subpopulations				
	I	II	III	
	0.244	0.115	0.158	

Table 5. Allele frequency divergence among sub-populations and expected heterozygosity (average distance) between genotypes within the same subpopulations.

flowering and maturity are desired characteristics because they contribute to drought escape. This is particularly important in the SSA context, adversely affected by climate change resulting in poor rainfall^{34,35}.

The phenotypic diversity observed in the test population was supported by the analysis of molecular variance. Genomic data is utilized in determining phylogenetic relationships in germplasm collections³⁶. The high number of SNP markers utilized in this study enabled a more precise estimate of the population structure. Gene diversity ranged between 0.03 and 0.5, with a mean of 0.4. Reference³⁷ reported similar findings with the genetic diversity varying from 0.01 to 0.5. The polymorphism information content (PIC) predicts the relevance of a genetic marker for linkage analysis^{38,39}. In this current study, the PIC average value was 0.31, indicating that the SNP markers employed could effectively discriminate between the genotypes. These results are in support of those reported by Refs.^{39,40}. Single nucleotide polymorphisms are bi-allelic; therefore, their PIC values are often lower than SSRs; however, they offer higher genetic resolution^{41,42}. The mean observed heterozygosity (Ho) of 0.67 reported in this study was considerably higher than those reported for inbred lines by Refs.^{43,44}. This suggest(s) that some of the inbred lines were not homozygous. Further selfing and substantial selection is strongly advised to fix these inbred lines⁴⁵. The inbreeding coefficient (F) quantifies the likelihood that two alleles at every locus within an organism are identical by descent from the two parents' shared ancestor(s)⁴⁶. In this study, the F value ranged from –1.1 to –0.5, with an average of –0.69. Reference⁴⁷ similarly observed a negative F mean value of –0.29. A negative F value in a population sample indicates the presence of excess heterozygotes⁴⁸.

The panel of selected SNPs used in this study effectively revealed the polymorphisms existing among and within the inbred line populations. The within inbred line variation accounted for the most significant proportion of the variation observed compared to among population differences. A constant germplasm exchange between IITA and CIMMYT which may explain the lack of high genetic identity and low distance genetic divergence observed between the two centres. For example, parental germplasm used to initiate the CIMMYT program for drought and low soil N stress tolerance in Kenya was sourced from the CIMMYT Southern Africa regional centre in Zimbabwe and early maturing populations from the International Institute of Tropical Agriculture (IITA)⁴⁹. Likewise, CIMMYT has been the major source of white maize germplasm for most SSA commercial seed companies because of the prominence of white maize in the region, particularly in Southern Africa^{10,15}. This presents genetic bottlenecks that limit breeding gains. Reference⁵⁰ reported a wide divergence between Southern Africa sub-tropical germplasm and temperate maize lines suggesting the potential utility of temperate germplasm in expanding the genetic base of African germplasm. The combination of stress tolerance and high-yield performance can produce broad adaptation in hybrids for South Africa. The Water Efficient Maize for Africa (WEMA) program is a good example of the efficacy of synergy between the private and public sector germplasm. Under the project, Monsanto, and CIMMYT developed germplasm combining drought tolerance and optimum yield performance. The resultant hybrids have a 20% yield advantage under drought conditions and outperform most commercial hybrids in optimum environments⁵¹.

The SNP analysis revealed the existence of three subpopulations (K = 3). These results are in agreement with those reported by Ref.³⁷ who reported the existence of 3 subpopulation among 94 early maturing tropical maize inbred lines using SNP markers. Similarly, the dendrogram grouped the genotypes into three primary clusters showing partial existence of origin as a source of diversity. In genetic diversity studies, reports of the grouping of genotypes according to their geographic origins are common⁴⁷. However, the clustering of most inbred lines was not based on ancestry suggesting that maize inbred lines derived from the same populations do not always have the same selection. This also indicates that the study panel consisted of unique inbred lines. The UKZN inbred lines were bred for performance in high-potential environments while the CGIAR material was developed for adaptation in stress environments.

Conclusions

The results of this study revealed that the maize inbred lines displayed phenotypic variation supported by genetic diversity, which will enable the selection and breeding of stress-tolerant maize hybrids. The genotypes exhibited highly significant variation in key agronomic traits such as DA, DS, PC, PH, and GY. High genetic distances between paired inbred lines demonstrated the distinctiveness of the evaluated genotypes and the availability of substantial genetic variation that could be utilized in the maize breeding program. The inbred lines were partially classified into three heterotic groups based on source background. Superior hybrids can be developed

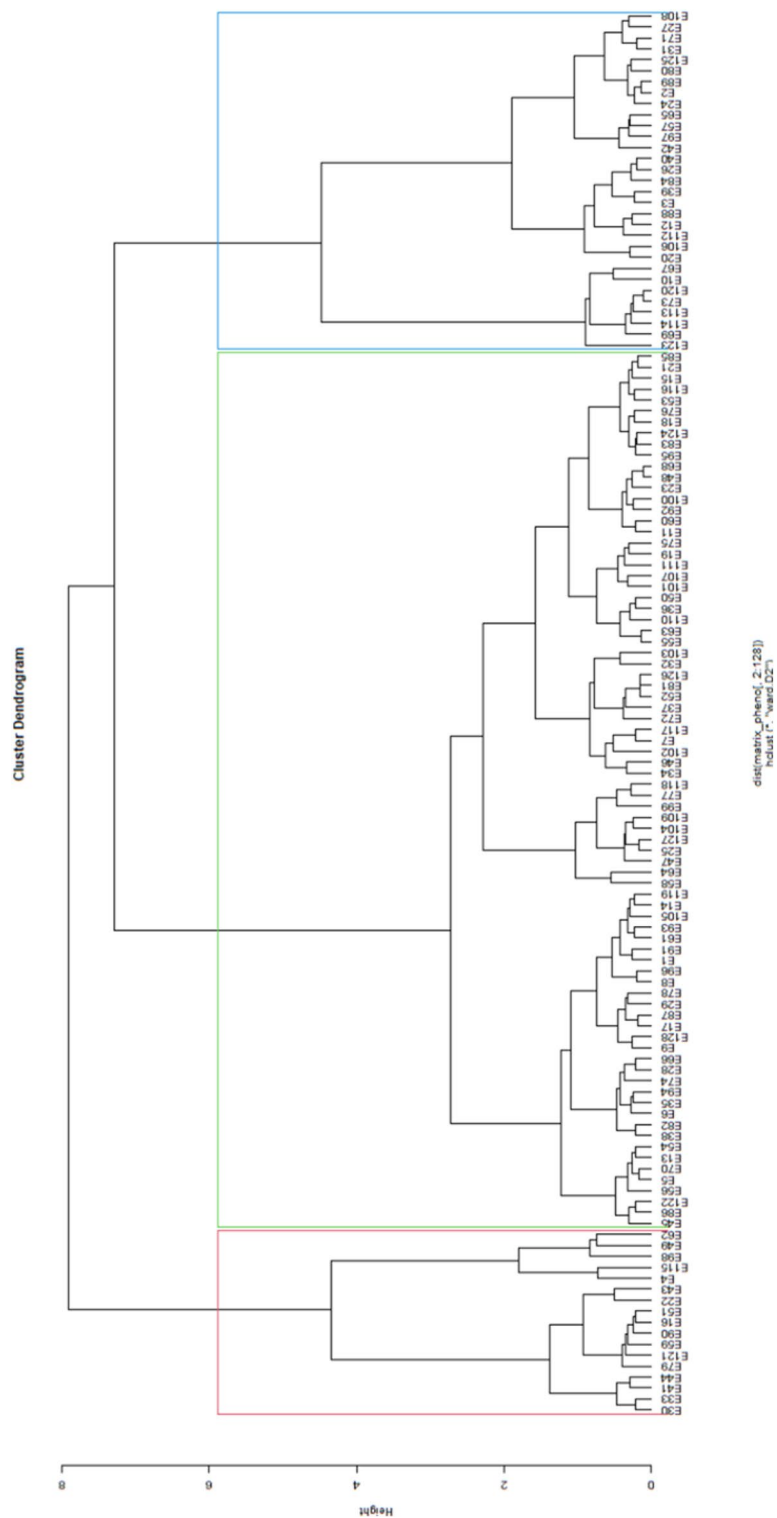


Figure 4. Dendrogram displaying the relationship between the 128 genotypes of maize based on the phenotypic matrix.

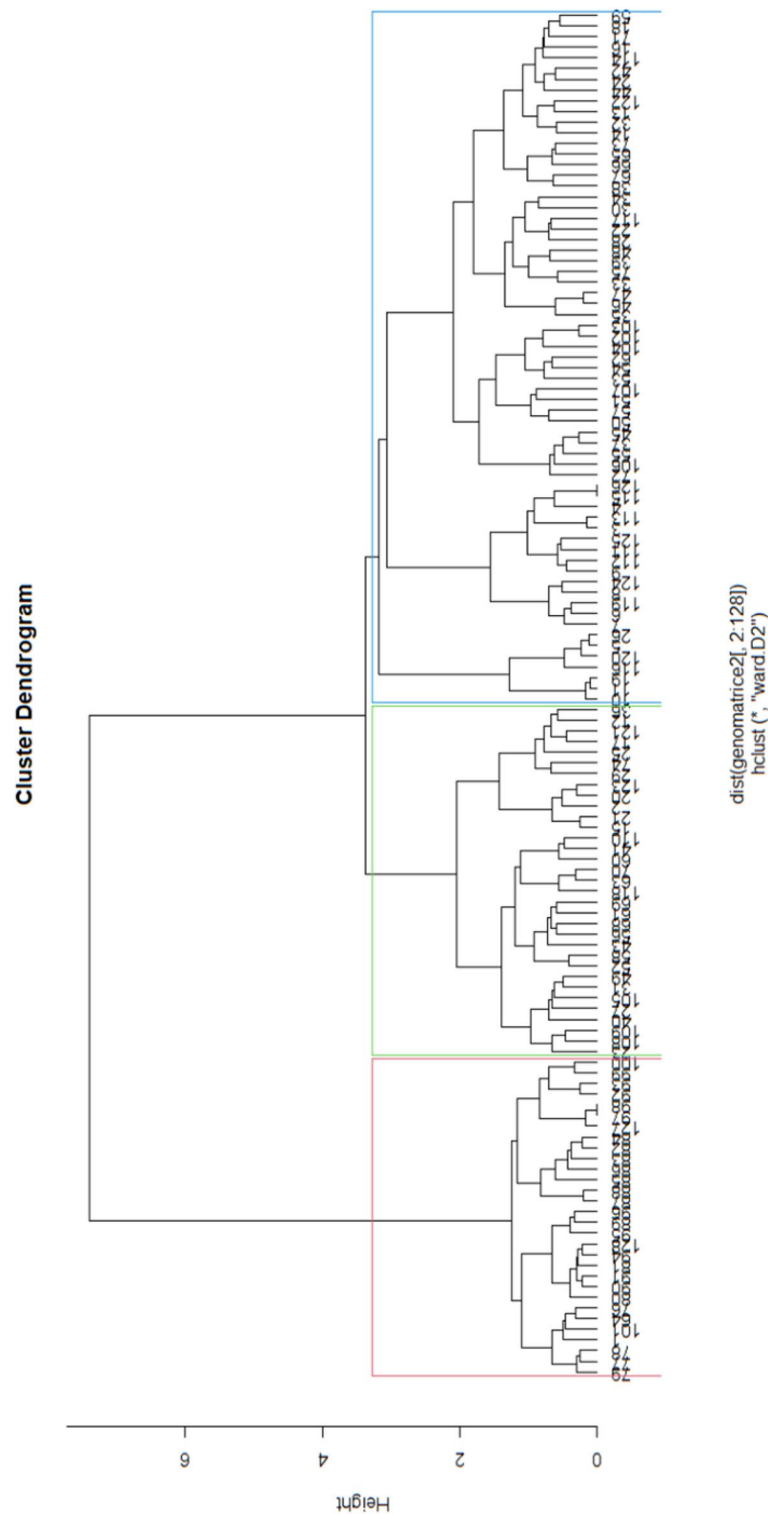


Figure 5. Dendrogram illustrating the relationship between the 128 genotypes of maize based on the genotypic matrix.

from inbred lines exhibiting the most comprehensive genetic distance within populations identified. Combining ability studies are recommended to confirm the agronomic performance and breeding values of the most divergent parental lines.

Data availability

All relevant data are within the paper and its Supporting Information files. The use and collection of plant material comply with relevant institutional, national, and international guidelines and legislation.

Received: 19 June 2023; Accepted: 13 October 2023

Published online: 19 October 2023

References

1. Wilkinson, J. The final foods industry and the changing face of the global agro-food system. *Sociol. Rural.* **42**, 329–346. <https://doi.org/10.1111/1467-9523.00220> (2002).
2. Shoko, R., Belete, A. & Chaminuka, P. Estimating the supply response of maize in South Africa: A Nerlovian partial adjustment model approach. *Agrekon* **55**, 237–253. <https://doi.org/10.1080/03031853.2016.1203802> (2016).
3. Diko, A. & Jun, W. Influencing factors of maize production in South Africa: The Case of Mpumalanga, free state and North West Provinces. *Asian J. Adv. Agric. Res.* **14**, 25–34. <https://doi.org/10.9734/AJAAR/2020/v14i130121> (2020).
4. FAO. World Food and Agriculture – Statistical Yearbook 2022. Rome. <https://doi.org/10.4060/cc2211en> (2022).
5. Bradshaw, C. D. *et al.* Unprecedented climate extremes in South Africa and implications for maize production. *Environ. Res. Lett.* **17**, 084028. <https://doi.org/10.1088/1748-9326/ac816d> (2022).
6. Mangani, R. *et al.* Potential impacts of extreme weather events in main maize (*Zea mays* L.) producing areas of South Africa under rainfed conditions. *Reg. Environ. Change* **19**, 1441–1452 (2019).
7. Bänziger, M., Edmeades, G. & Lafitte, H. Physiological mechanisms contributing to the increased N stress tolerance of tropical maize selected for drought tolerance. *Field Crop Res* **75**, 223–233. [https://doi.org/10.1016/S0378-4290\(02\)00028-X](https://doi.org/10.1016/S0378-4290(02)00028-X) (2002).
8. Tadele, Z. Raising crop productivity in Africa through intensification. *Agronomy* **7**, 22 (2017).
9. Pfordt, A., Ramos Romero, L., Schiwiek, S., Karlovsky, P. & von Tiedemann, A. Impact of environmental conditions and agronomic practices on the prevalence of Fusarium species associated with ear- and stalk rot in maize. *Pathogens* **9**, 236 (2020).
10. Sibiya, J., Tongoona, P., Derera, J., van Rij, N. & Makanda, I. Combining ability analysis for Phaeosphaeria leaf spot resistance and grain yield in tropical advanced maize inbred lines. *Field Crops Res.* **120**, 86–93. <https://doi.org/10.1016/j.fcr.2010.09.001> (2011).
11. Sibiya, J., Tongoona, P., Derera, J. & van Rij, N. Genetic analysis and genotype × environment (G × E) for grey leaf spot disease resistance in elite African maize (*Zea mays* L.) germplasm. *Euphytica* **185**, 349–362. <https://doi.org/10.1007/s10681-011-0466-2> (2012).
12. Krishna, V. V. *et al.* (International Maize and Wheat Improvement Center, 2021).
13. Kondwakwenda, A. *et al.* *Food Security for African Smallholder Farmers* 3–26 (Springer, 2022).
14. Prasanna, B. M. *et al.* Beat the stress: Breeding for climate resilience in maize for the tropical rainfed environments. *Theor. Appl. Genet.* **134**, 1729–1752. <https://doi.org/10.1007/s00122-021-03773-7> (2021).
15. Derera, J. *Genetic Effects and Associations Between Grain Yield Potential, Stress Tolerance and Yield Stability in Southern African Maize (Zea mays L.) Base Germplasm* (2005).
16. Silva, K. J. *et al.* High-density SNP-based genetic diversity and heterotic patterns of tropical maize breeding lines. *Crop Sci.* **60**, 779–787. <https://doi.org/10.1002/csc2.20018> (2020).
17. Senyolo, M. P., Long, T. B., Blok, V. & Omta, O. How the characteristics of innovations impact their adoption: An exploration of climate-smart agricultural innovations in South Africa. *J. Clean. Prod.* **172**, 3825–3840. <https://doi.org/10.1016/j.jclepro.2017.06.019> (2018).
18. Mickelson, H. R., Cordova, H., Pixley, K. V. & Bjarnason, M. S. Heterotic relationships among nine temperate and subtropical maize populations. *Crop Sci.* **41**, 1012–1020. <https://doi.org/10.2135/cropsci2001.4141012x> (2001).
19. Musundire, L., Derera, J., Dari, S., Tongoona, P. & Cairns, J. E. Molecular characterisation of maize introgressed inbred lines bred in different environments. *Euphytica* **215**, 1–14. <https://doi.org/10.1007/s10681-019-2367-8> (2019).
20. Begna, T. Role and economic importance of crop genetic diversity in food security. *Int. J. Agric. Sci. Food Technol.* **7**, 164–169. <https://doi.org/10.17352/2455-815X.000104> (2021).
21. Swarup, S. *et al.* Genetic diversity is indispensable for plant breeding to improve crops. *Crop Sci.* **61**, 839–852. <https://doi.org/10.1002/csc2.20377> (2021).
22. Kumar, A., Longmei, N., Kumar, P. & Kaushik, P. Molecular marker analysis of genetic diversity in maize: A review. *OBM Genet.* **6**, 1–1. <https://doi.org/10.21926/obm.genet.2201150> (2022).
23. Mir, R. R., Varshney, R. K. & Henry, R. J. Future prospects of molecular markers in plants. *Mol. Mark. Plants.* <https://doi.org/10.1002/9781118473023> (2013).
24. Tian, H. *et al.* New resources for genetic studies in maize (*Zea mays* L.): A genome-wide Maize6H-60K single nucleotide polymorphism array and its application. *Plant J.* **105**, 1113–1122. <https://doi.org/10.1111/tpj.15089> (2021).
25. Ananda Kumar, B. V. *et al.* Genomic prediction in doubled haploid maize (*Zea mays*) populations under water stress at flowering and well-watered conditions using high-density single-nucleotide polymorphisms. *Plant Breed.* **141**, 566–573. <https://doi.org/10.1111/pbr.13027> (2022).
26. Kondwakwenda, A., Sibiya, J., Amelework, A. B. & Zengeni, R. Diversity analysis of provitamin A maize inbred lines using single nucleotide polymorphism markers. *Acta Agric. Scand. B Soil Plant Sci.* **70**, 265–271. <https://doi.org/10.1080/09064710.2020.1718198> (2020).
27. Mitelpunkt, A., Galili, T., Shachar, N., Marcus-Kalish, M. & Benjamini, Y. *Proc. International Joint Conference on Biomedical Engineering Systems and Technologies* 566–573.
28. Granato, I. S. C. *et al.* snpReady: A tool to assist breeders in genomic analysis. *Mol. Breed.* **38**, 102. <https://doi.org/10.1007/s11032-018-0844-8> (2018).
29. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129> (2008).
30. Peakall, R. & Smouse, P. E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295. <https://doi.org/10.1111/j.1471-8286.2005.01155.x> (2006).
31. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959. <https://doi.org/10.1093/genetics/155.2.945> (2000).
32. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **14**, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> (2005).
33. Mackay, T. F. Q&A: Genetic analysis of quantitative traits. *J. Biol.* **8**, 1–5. <https://doi.org/10.1186/jbiol133> (2009).

34. Nelimor, C., Badu-Apraku, B., Tetteh, A. Y., Garcia-Oliveira, A. L. & Nguetta, A.S.-P. Assessing the potential of extra-early maturing landraces for improving tolerance to drought, heat, and both combined stresses in maize. *Agronomy* **10**, 318. <https://doi.org/10.3390/agronomy10030318> (2020).
35. Badu-Apraku, B., Obisesan, O., Abiodun, A. & Obeng-Bio, E. Genetic gains from selection for drought tolerance during three breeding periods in extra-early maturing maize hybrids under drought and rainfed environments. *Agronomy* **11**, 831. <https://doi.org/10.3390/agronomy11050831> (2021).
36. Osuman, A. S. *et al.* Genetic diversity, population structure and inter-trait relationships of combined heat and drought tolerant early-maturing maize inbred lines from West and Central Africa. *Agronomy* **10**, 1324. <https://doi.org/10.3390/agronomy10091324> (2020).
37. Boakyewaa Adu, G. *et al.* Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *PLoS ONE* **14**, e0214810. <https://doi.org/10.1371/journal.pone.0214810> (2019).
38. Netravati, M. Genetic diversity analysis in aromatic rice genotypes using microsatellite based simple sequence repeats (SSR) marker. *Afr. J. Biotechnol.* **12**, 4238–4250. <https://doi.org/10.5897/AJB2013.12137> (2013).
39. Zhang, X. *et al.* Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. *BMC Genom.* **17**, 1–16. <https://doi.org/10.1186/s12864-016-3041-3> (2016).
40. Wu, X. *et al.* Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. *Theor. Appl. Genet.* **127**, 621–631. <https://doi.org/10.1007/s00122-013-2246-y> (2014).
41. Helyar, S. J. *et al.* Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Mol. Ecol. Resour.* **11**, 123–136. <https://doi.org/10.1111/j.1755-0998.2010.02943.x> (2011).
42. Suvi, W. T., Shimelis, H., Laing, M., Mathew, I. & Shayanowako, A. I. T. Assessment of the genetic diversity and population structure of rice genotypes using SSR markers. *Acta Agric. Scand. Sect. B Soil Plant Sci.* **70**, 76–86. <https://doi.org/10.1080/09064710.2019.1670859> (2020).
43. Abu, P. *et al.* Genetic diversity and inter-trait relationship of tropical extra-early maturing quality protein maize inbred lines under low soil nitrogen stress. *PLoS ONE* **16**, e0252506. <https://doi.org/10.1371/journal.pone.0252506> (2021).
44. Josia, C. *et al.* SNP-based assessment of genetic purity and diversity in maize hybrid breeding. *PLoS ONE* **16**, e0249505. <https://doi.org/10.1371/journal.pone.0249505> (2021).
45. Semagn, K. *et al.* Quality control genotyping for assessment of genetic identity and purity in diverse tropical maize inbred lines. *Theor. Appl. Genet.* **125**, 1487–1501. <https://doi.org/10.1007/s00122-012-1928-1> (2012).
46. Wright, S. Coefficients of inbreeding and relationship. *Am. Nat.* **56**, 330–338. <https://doi.org/10.1086/279872> (1922).
47. Kasoma, C., Shimelis, H., Laing, M. D., Shayanowako, A. I. & Mathew, I. Revealing the genetic diversity of maize (*Zea mays* L.) populations by phenotypic traits and DArTseq markers for variable resistance to fall armyworm. *Genet. Resour. Crop Evol.* **68**, 243–259. <https://doi.org/10.1007/s10722-020-00982-9> (2021).
48. Abady, S. *et al.* Assessment of the genetic diversity and population structure of groundnut germplasm collections using phenotypic traits and SNP markers: Implications for drought tolerance breeding. *PLoS ONE* **16**, e0259883. <https://doi.org/10.1371/journal.pone.0259883> (2021).
49. Das, B. *et al.* Identification of donors for low-nitrogen stress with maize lethal necrosis (MLN) tolerance for maize breeding in sub-Saharan Africa. *Euphytica* **215**, 80. <https://doi.org/10.1007/s10681-019-2406-5> (2019).
50. Richard, C., Osiru, D., Lubberstedt, T. & Mwala, M. *Genetic Diversity and Heterotic Grouping of the Core Set of Southern African and Temperate Maize (Zea mays L.) Inbred Lines Using SNP Markers* (2016).
51. Lumpkin, T. A. & Armstrong, J. *Staple Crops, Smallholder Farmers and Multinationals: The 'Water-Efficient Maize for Africa' Project* (AgEcon Search, 2009).

Acknowledgements

Authors are grateful to the Alliance for a Green Revolution in Africa (AGRA), and the nGAP scholarship for funding the research project. The authors also acknowledge and appreciate Dr Admire Shayanowako for assisting with the data analysis. Thanks to Seqart Africa in Nairobi, Kenya, for genotyping by sequencing services.

Author contributions

S.P.D. conducted the experiment, collected phenotypic data, and wrote the manuscript, F.K., J.S. reviewed the manuscript.

Funding

The funding was provided by the Alliance for a Green Revolution in Africa (AGRA) and the New Generation of Academics Program (nGAP).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-44961-3>.

Correspondence and requests for materials should be addressed to S.P.D.

Reprints and permissions information is available at www.nature.com/reprints.

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



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/>.

© The Author(s) 2023