Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Construction and preliminary characterization of an early generation multi-parent sorghum bicolor population for climate-adaptive breeding

Rachael Kisilu

rachael.kisilu@kalro.org

University of Nairobi

Kahiu Ngugi

University of Nairobi

Kimani Wilson

CornGro Limited

Wamalwa Lydia

University of Nairobi

Research Article

Keywords: Sorghum, MAGIC population, Multi-parent population, Drought adaptation, Stay-green, Genetic diversity

Posted Date: October 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11109960/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Multi-parent populations can broaden allelic diversity and recombination available for genetic analysis and crop improvement. In this study, we developed and preliminary characterized an early generation partially developed multi-parent advanced generation intercross (pMAGIC) population for sorghum (*Sorghum bicolor* (L.) Moench) derived from eleven genetically diverse founder lines selected for complementary agronomic traits, including drought tolerance, stay-green, early maturity, and high yield. The population, comprised 176 S_2-derived lines and the eleven founders, with Gadam included as a local check. Genotyping using DArTseq technology identified 5,512 high-quality SNP markers, with moderate polymorphic information content (PIC) across the population. Population structure analyses revealed residual genetic differentiation among nine sub-groups (F_{st} = 0.16) and admixture among progenies, consistent with contribution of multiple founders. Days to flowering (DTF), used as a proof-of-concept trait, was evaluated across multiple environments for and substantial phenotypic variation among the S_2-derived lines, including transgressive segregation beyond parental ranges, demonstrating the enhanced recombination and allelic diversity captured in the pMAGIC population. Genome-wide linkage disequilibrium declined to half of its maximum value at approximately 339.34 kb. Exploratory genome-wide association analysis (GWAS) identified nine significant SNPs associated with DTF across multiple chromosomes, with several associated intervals containing candidate genes with reported roles in flowering time regulation, including photoreceptor signaling and transcriptional regulation. These finding demonstrate that the population has retained substantial genetic and phenotypic diversity during early-generation development and provides a useful foundation for further population advancement and genetic analysis. Continued selfing, population expansion, and higher-density genotyping are expected to enhance its resolution and utility for dissecting complex agronomic traits in sorghum.

INTRODUCTION

Sorghum (*Sorghum bicolor* (L.) Moench) is a globally important cereal crop that serves as a staple food, feed, and industrial raw material, particularly in arid and semi-arid regions of sub-Saharan Africa and Asia. Its inherent tolerance to drought and high temperatures makes it a key crop for climate-resilient agriculture. The species harbors substantial genetic diversity distributed across major racial groups, providing a valuable resource for crop improvement (Kumar et al., 2023). However, despite this diversity, genetic gains in sorghum breeding have remained relatively modest due to the complex inheritance of key agronomic traits and limitations associated with conventional breeding approaches.

The dissection of complex traits in sorghum requires mapping populations that capture broad allelic diversity and allow high-resolution genetic analysis. Traditional biparental populations are limited by low allelic richness and restricted recombination, which reduces their power to detect quantitative trait loci (QTLs) and limits mapping resolution. In contrast, multi-parent populations provide a more powerful framework by incorporating multiple founder lines, thereby increasing genetic diversity and recombination (Mackay et al., 2014).

Multi-parent Advanced Generation Inter-Cross (MAGIC) populations have emerged as a valuable resource for genetic analysis and crop improvement. These populations are developed by intercrossing multiple founder lines over several generations, resulting in highly recombined genomes with reduced population structure and smaller linkage disequilibrium blocks. MAGIC populations have been successfully developed and utilized in several cereal crops, including wheat, rice, maize, and barley, where they have facilitated high-resolution mapping and the identification of loci controlling complex traits (Arrones et al., 2020; Bandillo et al., 2013). Their ability to capture multiple alleles at a locus and generate novel recombinants makes them particularly useful for genomics-assisted breeding.

In sorghum, multi-parent populations are increasingly being explored as resources for combining genetic diversity from contrasting germplasm groups. Initial efforts demonstrated the feasibility of generating multi-parent populations using diverse founder lines and structured crossing schemes, resulting in populations with high levels of recombination and genetic diversity (Ongom & Ejeta, 2018). More recent studies have further characterized sorghum MAGIC populations and demonstrated their utility in capturing useful variation within elite germplasm pools and enabling high-resolution genetic analysis (Kumar et al., 2023). These populations provide a powerful platform for both trait dissection and the identification of superior breeding lines. However, the utility of any particular population depends on its founder composition, crossing history, degree of advancement, retained genetic diversity, and phenotypic variation. In the present study, the population was designed to combine complementary attributes from drought-adapted and farmers-preferred varieties. Three stay-green and drought tolerant donor lines were crossed with eight high-yielding and locally adapted lines, followed by intercrossing and selfing to generate an early-generation multi-parent population. This strategy was intended to bring together alleles associated with drought adaptation, stay-green expression, maturity, and grain and biomass productivity within a common breeding population.

Days to flowering (DTF) was selected as a proof-of-concept trait for the initial characterization of the population. Flowering time is an important adaptive trait in sorghum that plays a critical role in determining environmental adaptation, yield potential, and cropping system suitability (Sannemann et al., 2015). DTF is a quantitatively inherited trait influenced by multiple genes and environmental factors including photoperiod and temperature (Mangal et al., 2025). Due to its relatively high heritability and ease of measurement, DTF is commonly used as a model trait for validating newly developed populations and assessing their utility for genetic studies. In multi-parent populations, flowering time traits often exhibit substantial phenotypic variation, making them suitable for proof-of-concept analyses (Mackay et al., 2014; Sannemann & Emma, 2015).

In this study, we developed and preliminarily characterized an early-generation partially-developed MAGIC (pMAGIC) population comprising 176 S_2-derived lines from eleven founder lines. We evaluated the population for genetic diversity, population structure, linkage disequilibrium, and phenotypic variation for DTF across three locations and two seasons. We further conducted an exploratory genome-wide association analysis of DTF to identify genomic regions warranting further investigation. Collectively, these analyses provide an initial assessment of the genetic and phenotypic composition of the

population and establish a foundation for its continued advancement, higher-density genotyping, and future genetic dissection of agronomically important traits in sorghum.

MATERIALS AND METHODS

Plant material, population development and phenotyping

A multi-parent, partially developed MAGIC (pMAGIC) population of sorghum (*Sorghum bicolor* (L) Moench) was developed from eleven genetically diverse founder lines selected to combine complementary agronomic attributes, including drought tolerance, stay-green expression, early maturity, and grain and biomass productivity (Table 1, Supplementary Table S1). The founder set comprised three stay-green and drought-adapted donor lines (E36-1, B35, IESB371) and locally adapted, high yielding lines (KARI Mtama 1, KM30-K, KM32-1, KM32-2, LEG17, IESV99032-2DL, IESV92038-1DL, and ICSV142011).

The population was initiated using a North Carolina Design II (NC-II) mating scheme, in which the three stay-green donor lines were crossed with the eight locally adapted lines to generate factorial F_1 families. The resulting F_1 progenies were subsequently intermated pairwise to increase the number of recombination events and combine genetic contributions from multiple founders. The resulting multi-parent progenies were advanced through two generations of selfing to produce partially inbred S_2 families. A total of 176 S_2-derived families were selected for further evaluation. The eleven founder lines and the local check (Gadam) were included as reference genotypes, giving a total of 188 entries in the evaluation and genotyping panel.

The population was therefore considered an early-generation or partially developed MAGIC population because it had not yet undergone sufficient generations of selfing to achieve advanced inbred-line status. This distinction was retained throughout the study when interpreting genetic diversity, linkage disequilibrium, and marker–trait associations.

Phenotypic evaluation

Days to flowering (DTF) was selected as the initial proof-of-concept trait for evaluating phenotypic segregation in the developing pMAGIC population. DTF was recorded at the reproductive stage as the number of days from planting to the time when approximately 50% of plants within a plot had flowered.

The 176 S_2-derived lines, eleven founders and Gadam check were evaluated across three locations and two growing seasons. Trials were established using an alpha-lattice design with two replicates. Each replicated consisted of ten blocks containing 19 plots, including two filler plots, thereby accommodating the 188 entries in the evaluation panel. The same experimental structure was used across the environments.

Table 1
Eleven founder lines of the sorghum pMAGIC population with local check

Founders	Agronomic characteristics
B35	Stay-green, Early maturity
LEG17	Yield, drought susceptible
IESB371	Yield, drought susceptible
KM34-2	Yield, drought susceptible
KARI MTAMA 1	Yield, drought susceptible
KM30-K	Yield, drought susceptible
KM32-2	Yield, drought susceptible
IESV99034-2DL	Yield, drought susceptible
ICSV142011	Yield, drought susceptible
E36-1	Stay-green, Early maturity
GADAM	Yield, Early maturity, drought tolerant
IESV92038-1DL	Yield, drought susceptible

For each environment, DTF observations were recorded for the experimental units according to the field protocol. Phenotypic means across the three locations and two seasons were used to describe the range of flowering time among founders, the check, and the S_2-derived population. The observed range of the progeny was compared with the parental range to assess evidence of transgressive segregation.

DArtseq genotyping and data processing

Young leaf tissue was collected from each of the 188 genotypes, lyophilized, and submitted to SEQART AFRICA (https://www.seqart.net/) at the International Livestock Research Institute (ILRI), Nairobi, Kenya, for DNA extraction and genotyping. Genomic DNA was extracted from the plant material and subjected to quality assessment before Diversity Arrays Technology sequencing low density (DArTseqLD) genotyping.

DArTseqLD method is a reduced-representation genotyping approach that uses restriction enzymes t to target informative regions of the genome (Kilian et al., 2012). The resulting SNP markers were aligned to the *S.bicolor* reference genome v3.1.1 (McCormick et al., 2018) obtained from Phytozome (Goodstein et al., 2012).

SNP quality control was performed using VCFtools (Danecek et al., 2011). Markers were retained when they had a call rate of at least 90% and a minor allele frequency (MAF) greater than 5%. Markers that did

not meet wither criterion were excluded from subsequent analyses. The distribution of the retained SNPs across the ten sorghum chromosomes were visualized using the CMplot function implemented in the rMVP R package (Yin et al., 2021).

The SNP markers characterization including call-rate, polymorphic information content (PIC) and mutation types (transitions *vs* tranversions) was implemented in dartR package in R (Gruber et al., 2018). Functional annotation of SNPs according to their genomic positions was performed using ANNOVAR (Wang et al., 2010).

Population structure, genetic diversity and phylogenetic analysis

Population structure was assessed using the ADMIXTURE (version 1.3.0) (Alexander & Lange, 2011). ADMIXTURE estimates ancestry proportions of individual genomes under a specified number of ancestral populations (K). Analyses were performed for K = 2 to K = 10 using 10-fold cross-validation. The K value corresponding to the minimum cross-validation error was used to describe the major ancestry components observed in the genotyped panel. ADMIXTURE results were visualized using a customized R script.

The Principal component analysis (PCA) was conducted using the adegenet package in R (Jombart 2008) to visualize genetic relationships among the founders, S_2-derived lines, and Gadam. The first two principal components (PCs) were plotted in R software (R Core Team, 2024) based on the number of sub-populations identified in ADMIXTURE, to visualize genetic clustering and admixture. To further analyze and understand the genetic variation of our population, a discriminant analysis of principal components (DAPC) was performed in the R package adegenet v2.1.3 (Jombart, 2008). Analyses were evaluated using one to ten clusters, and the Bayesian information criterion (BIC) was used to assess the number of clusters providing the best representation of the genetic data. Genetic diversity was assessed using observed heterozygosity (H_o), expected heterozygosity (H_e), unbiased expected heterozygosity (uH_e), and inbreeding coefficient (F_{is}). These statistics were calculated using the dartR package in R(Gruber et al., 2018). The F_{is} was calculated as: $F_{is} = (H_e\text{-}H_o)/H_e$. Analysis of molecular variance (AMOVA) was calculated based on the description by Nei (Masatoshi Nei, 1972) using the poppr package in R (Kamvar et al., 2014). Geneflow (number of migrants [Nm]) among the sub-populations was estimated using the formula: $Nm = (1\text{-}F_{st})/4F_{st}$, where F_{st} is the fixation index (Wright, 1968).

The genetic differentiation among the sub-populations was estimated using the Nei's pairwise fixation indices (F_{st}) (M Nei, 1987) in hierfstat package of R (Goudet, 2005). These F_{st} pairs were used to construct a neighbor-joining (NJ) phylogenetic tree using the ape package in R (Paradis et al., 2004). The clustering of individual within the sub-populations was illustrated by calculating the genetic distances among the accessions using the Euclidean distance method. This pairwise genetic distance matrix was then used to construct a hierarchical cluster dendogram with the UPGMA agglomeration method using the hclust function in R (R Core Team, 2024). The dendogram was exported in Newick format using the

ape package for refined annotation in the Interactive Tree of Life (iTOL) version 6.5.2 online tool (https://itol.embl.de/, accessed on 8th August 2026) (Letunic & Bork, 2024).

A network analysis of parental contributions to each of the sub-populations determined by ADMIXTURE was plotted to visualize and quantify the relationships between the founder lines and derived populations. The dataset of sub-populations, individual pMAGIC lines, and cross information was processed in R version 4.3.2 (R Core Team, 2024) using the packages tidyverse (Wickham et al., 2019) for data wrangling, igraph (Gábor et al., 2025), tidygraph (Pedersen, 2025), and ggraph (Pedersen, 2025) for network construction and visualization, and VisNetwork (Thieurmel, 2022) for interactive exploration.

Linkage disequilibrium

Genome-wide linkage disequilibrium (LD) was estimated in Trait Analysis by aSSociation Evolution, and Linkage (TASSEL) software version 5.2.88 (Bradbury et al., 2007) using the filtered SNP markers. The pairwise association among all the alleles were generated to obtain the correlation coefficient (r^2) using a sliding 50-SNP window. The averages of the r^2 across all 10 sorghum chromosomes were plotted against the chromosomal physical distance using the R software (Team, 2024) to simulate the LD decay curve which was then used to estimate the distance at which LD parameter (r^2) dropped to half of its maximum value.

Exploratory genome-wide association analysis

Genome-wide association analysis was conducted using days to flowering as a proof-of-concept trait to explore the potential of the developing pMAGIC population for preliminary genetic dissection. Association analysis was performed using the Fixed and Random Model Circulating Probability Unification (FarmCPU) model implemented in the rMVP R package (Yin et al., 2021). A kinship matrix was calculated using the VanRaden method (VanRaden, 2008) and the top three principal components (PCs) were included as fixed-effect covariates to account for population structure (Price et al., 2006). Association testing was carried out using 5,512 filtered DArTseq SNP markers, and a modified Bonferroni correction was applied to establish the genome-wide significance threshold (Kimani et al., 2020).

Identification of genes within associated genomic regions

For SNPs significantly associated with DTF, genomic regions surrounding each marker were examined using the S. bicolor reference genome v3.1.1. The interval used for gene interrogation was informed by the LD decay distance estimated for the pMAGIC population. Genes located within the corresponding genomic intervals were identified using the annotated reference genome available through Phytozome (Goodstein et al., 2012).

RESULTS

Phenotypic variation

Days to flowering (DTF) varied among the eleven founder lines, the Gadam check, and the 176 S_2-derived pMAGIC lines (Table 2). Among the founders, mean DTF ranged from 51.5 ± 0.96 days in IESB371 to 61.5 ± 0.29 days in IESV99034-2DL. Most founders had mean flowering times between approximately 54 and 56 days. Gadam was the earliest flowering entry, with a mean DTF of 49.5 ± 0.96 days.

The S_2-derived population had a mean DTF of 57.18 ± 0.17 days and a substantially wider phenotypic range of 45 to 75 days (Table 2). The progeny therefore included lines that flowered earlier than the earliest founder and lines that flowered later than the latest founder, indicating transgressive segregation. This broader distribution demonstrates substantial phenotypic segregation for DTF among the early-generation progeny.

Table 2
Mean, standard error, minimum (Min) and maximum (Max) values scored for flowering time in days as the mean across three locations for two seasons, of the pMAGIC founder lines (N = 11) one check (Gadam) and 176 S_2-derived lines

Group	Mean ± SE	Min	Max
E36-1	54.5 ± 0.65	53	56
B35	55.75 ± 0.48	55	57
IESB371	51.5 ± 0.96	49	53
KARI MTAMA 1	54.75 ± 0.63	53	56
KM30-K	55.25 ± 0.75	54	57
KM32-2	56.25 ± 0.48	55	57
LEG17	56 ± 0.41	55	57
IESV99034-2DL	61.5 ± 0.29	61	62
IESV92038-1DL	55 ± 0.91	53	57
ICSV142011	58.25 ± 1.65	55	62
KM34-2	55.75 ± 0.63	54	57
GADAM	49.5 ± 0.96	48	52
S_2 176 pMAGIC lines	57.18 ± 0.17	45	75

Genetic marker characterization

Genotyping of the 188 entries using DArTseq method generated 9,385 single nucleotide polymorphism (SNP) markers. Of these, 8,352 (~ 89%) were mapped to the Sorghum bicolor reference genome v3.1.1. Following filtering for marker call rate and minor allele frequency, 5,512 SNPs were retained for downstream analyses.

The retained markers had a mean call rate of 97.7% and a mean polymorphic information content (PIC) of 0.28, with PIC values ranging from 0.09 to 0.49 (Fig. 1a, b).The analysis of mutation types, including the frequency of transitions (Ts; A↔G and T↔C substitutions) and transversions (Tv; A↔T, A↔C, T↔G, or C↔G substitutions), revealed frequencies of 50.38% and 49.62%, respectively (Fig. 1c). The Ts/Tv ratio was 1.02.

Functional annotation showed that 39.15% of the retained SNPs were located in exonic regions, 20.74% in intergenic regions, and 11.91% in intronic regions (Fig. 1d). SNPs located upstream and within the 5' UTR accounted for 8.27% and 7.31%, respectively, while 3' UTR and downstream variants accounted for 5.60% and 5.35%. Splicing variants represented 1.67% of the markers. Among exonic SNPs, 63.17% were classified as nonsynonymous (missense), 36.24% as synonymous, and 0.69% as stop-gain variants (Fig. 1e).Population structure analysis and genetic relationships.

Population structure and genetic relationships

ADMIXTURE analysis based on the 5,512 filtered SNP markers identified K = 9 as the model with the lowest cross-validation error, resulting in nine genetic groups designated Pop1 to Pop9 (Fig. 2a; Supplementary Figure S1). Founder lines were distributed among these groups, with B35 associated with Pop1, LEG17 and IESB371 with Pop2, KM34-2 with Pop4, KARI MTAMA 1, KM30-K, KM32-2 and IESV99034-2DL with Pop5, ICSV142011 with Pop6, E36-1 with Pop8, and IESV92038-1DL and Gadam with Pop9.To further explore the contribution of the founder lines to each of the 9 sub-populations, we created a parent-population network structure (Fig. 2B, Supplementary Fig. 2b). B35 and E36-1 showed the highest average contributions across the groups, at 19.3% and 15.2%, respectively. B35 contributed 29.1%, 26.5%, 24.3%, and 16.7% to Pop1, Pop2, Pop7, and Pop8, respectively, while E36-1 contributed 27.3% to Pop8 and more than 17% to Pop4 and Pop6. IESB371 and LEG17 contributed 24.7% to Pop3 and 18.4% to Pop2, respectively. KARI MTAMA 1 contributed 16.2% to Pop8 and 11.2% to Pop5, while ICSV142011 contributed 23.3% to Pop6. KM34-2, KM30-K, and IESV92038-1DL each showed relatively lower contributions across the inferred groups.

PCA showed partial overlap among several genetic groups, with the first two principal components explaining 19.2% of the total genetic variation (Fig. 2c). Pop1 and Pop7 clustered closely along the negative PC1 axis, whereas Pop2 and Pop3 showed partial overlap. Pop4, Pop5, Pop6, Pop8, and Pop9 showed greater overlap, indicating shared genetic relationships among these groups.

DAPC using six retained principal components separated Pop1, Pop6, and Pop7 from the central cluster formed by Pop2, Pop3, Pop4, Pop5, Pop8, and Pop9 (Fig. 2d; Supplementary Figure S3). Pop1 and Pop7

were positioned closely in the DAPC space, consistent with their proximity in the PCA.3.4 Genetic diversity and molecular variance.

Genetic diversity estimates varied among the nine inferred genetic groups (Table 3). Observed heterozygosity (H_o) ranged from 0.13 in Pop9 to 0.33 in Pop4, while expected heterozygosity (H_e) ranged from 0.21 in Pop6 to 0.32 in Pop4. Unbiased expected heterozygosity (uH_e) ranged from 0.21 in Pop5 and Pop6 to 0.33 in Pop4. The highest diversity estimates were observed in Pop3 and Pop4, whereas Pop7 and Pop9 showed comparatively lower diversity. The inbreeding coefficient (Fis) ranged from 0.05 in Pop4 to 0.30 in Pop9, with an overall mean of 0.18.

Table 3
Genetic diversity indices for the nine sorghum MAGIC sub-populations

Sub-population	**n.Ind**	**H_o**	**H_e**	**uH_e**	**F_{is}**
pop1	30	0.24	0.3	0.3	0.16
pop2	16	0.19	0.26	0.27	0.24
pop3	27	0.25	0.3	0.31	0.16
pop4	28	0.33	0.32	0.33	0.05
pop5	36	0.17	0.22	0.22	0.2
pop6	10	0.16	0.21	0.22	0.22
pop7	11	0.19	0.25	0.26	0.22
pop8	20	0.18	0.21	0.21	0.11
pop9	10	0.13	0.21	0.22	0.3
Average	20.89	0.2	0.25	0.26	0.18

H_o, observed heterozygosity; H_e, expected heterozygosity; F_{is}, inbreeding coefficient; uH_e, unbiased expected heterozygosity.

AMOVA partitioned 14.92% of the total genetic variation among the nine groups, 14.61% among lines within groups, and 70.47% within individual lines (Table 4). Overall genetic differentiation among groups was significant (Fst = 0.16, $P < 0.001$). The corresponding effective number of migrants was estimated as Nm = 1.33.

Table 4
Analysis of molecular variance (AMOVA) among the 188 sorghum MAGIC lines based on the 5,512 SNPs.

Source	Df	Sum Sq	Mean Sq	Components of covariance (%)	P-value	F_{st}	Nm
Between sub-populations	8	85296.21	10662.03	14.92	< 0.001	0.16	1.33
Between lines within sub-populations	179	268797.7	1501.66	14.61	0.001		
Within lines	188	199574.9	1061.57	70.47	0.001		
Total	375	553668.7	1476.45	100			

df, degrees of freedom; Sum Sq, sum of squares; Mean Sq, mean square; Fst, fixation index; Nm, number of migrants polymorphism.

Pairwise F_{st} values ranged from 0.08 between Pop1 and Pop7 to 0.27 between Pop8 and Pop9 (Table 5). Most pairwise F_{st} estimates were below 0.20. The lowest differentiation was observed between Pop1 and Pop7 (0.08), Pop2 and Pop5 (0.09), and Pop4 and Pop5 (0.10), whereas the highest differentiation was observed between Pop8 and Pop9 (0.27), followed by Pop7 and Pop8 (0.24) and Pop1 and Pop9 (0.24).

Table 5
Pairwise genetic differentiation index (F_{st}) values of the MAGIC lines sub-populations

pop1	pop1	pop2	pop3	pop4	pop5	pop6	pop7	pop8	pop9
	NA								
pop2	0.13	NA							
pop3	0.17	0.12	NA						
pop4	0.14	0.1	0.11	NA					
pop5	0.16	0.09	0.13	0.1	NA				
pop6	0.19	0.15	0.2	0.16	0.13	NA			
pop7	0.08	0.16	0.18	0.13	0.15	0.21	NA		
pop8	0.22	0.14	0.22	0.12	0.1	0.14	0.24	NA	
pop9	0.24	0.12	0.2	0.18	0.19	0.2	0.27	0.2	NA

Genetic relationships and clustering

The neighbor-joining dendrogram based on pairwise F_{st} values showed patterns of genetic differentiation broadly consistent with the ADMIXTURE and PCA analyses (Fig. 3).

Clustering based on Euclidean genetic distances separated the 188 entries into three major clades (Fig. 4). Clade I was predominantly composed of individuals from Pop3 together with KM30-K and IESB371, whereas Clade II was largely composed of Pop1 and Pop7 individuals and the founder B35. Clade III comprised individuals from the remaining genetic groups and showed greater admixture. Within this clade, individuals from Pop5 and Pop6 formed relatively distinct sub-clusters.

Linkage disequilibrium

Genome-wide linkage disequilibrium (LD) was assessed using the 5,512 filtered SNP markers. Pairwise LD, measured as r^2, declined with increasing physical distance between markers across the sorghum genome (Fig. 5). The mean LD value declined to one-half of its maximum value ($r^2 = 0.23$) at an average physical distance of approximately 339.34 kb.

Exploratory GWAS of days to flowering

An exploratory genome-wide association analysis was conducted using DTF as a proof-of-concept trait. Using the FarmCPU model and 5,512 filtered SNP markers, nine SNPs were significantly associated with DTF at the genome-wide significance threshold of $P = 2.48 \times 10^{-5}$ (Fig. 6; Supplementary Table S1). The significant markers were distributed across chromosomes 1, 3, 4, 5, 6, and 10.

Two significant SNPs on chromosome 4, S4:9387261 and S4:9428776, were located within the same LD block and therefore represented a closely linked association signal. The remaining significant markers occurred on chromosomes 1, 3, 5, 6, and 10.

Figure 6. Manhattan plot showing the GWAS for days to flowering (DTF) in the sorghum pMAGIC population. The x-axis represents the genomic position of SNP markers across the ten sorghum chromosomes, while the y-axis shows the statistical significance of marker–trait associations expressed as $-\log_{10}$(p-values). Each point corresponds to a single SNP, and different colors distinguish chromosomes. The horizontal dashed line indicates the genome-wide significance threshold ($p = 2.48 \times 10^{-5}$).

Genes within DTF-associated genomic regions

Genomic regions surrounding the nine significant SNPs were examined using the estimated LD decay distance of 339.34 kb. A total of 498 genes were identified within these intervals, of which 397 (79.7%) had functional annotations in the S. bicolor reference genome v3.1.1 (Supplementary Table S2).

Several genes within the associated intervals had annotations related to processes potentially involved in flowering-time regulation (Table 6). These included Sobic.010G138000 on chromosome 10, annotated as similar to Cryptochrome 2 (CRY2); Sobic.001G433000 on chromosome 1, annotated as a FAR1 family

protein; and Sobic.004G102301 on chromosome 4, annotated as an auxin response factor. Additional genes included Sobic.004G101500, annotated as a putative bZIP protein, Sobic.003G110932, annotated as a NAM/NAC domain protein, and Sobic.006G276566, annotated as an RWP-RK domain protein.

Table 6
Genes with known or putative roles in flowering time regulation

SNP	Chr	SNP position (bp)	P-value	Candidate gene	Dist (kb)[£]	Annotation
S10:21874788	10	21874788	2.17E-05	Sobic.010G138000	282.07	similar to Cryptochrome 2
S1:71371498	1	71371498	1.13E-05	Sobic.001G433000	207	similar to FAR1 family protein
S4:9387261	4	9387261	9.38E-06	Sobic.004G102301	129.86	Auxin response factor (Auxin_resp)
S4:9387261	4	9387261	9.38E-06	Sobic.004G101500	7.75	similar to Putative bZIP protein
S3:10205658	3	10205658	1.70E-05	Sobic.003G110932	202.14	No apical meristem (NAM) protein (NAM)
S6:60773067	6	60773067	2.03E-05	Sobic.006G276566	16.88	RWP-RK domain (RWP-RK)

[£] Distance of the gene from the significant SNP

DISCUSSIONS

The development of a multi-parent population represents a powerful strategy for dissecting the genetic architecture of complex agronomic traits (Meng et al., 2016). Multi-parent populations that are created by combining genomes from multiple diverse founders and subjecting them to intercrossing and recombination capture substantially more allelic diversity than traditional biparental populations, while generating finely recombined mosaic genomes that facilitate high-resolution mapping (Kumar et al., 2023; Scott et al., 2020).

In this study, we developed an early-generation, partially developed MAGIC (pMAGIC) population of sorghum from eleven founders selected to represent contrasting but complementary agronomic attributes. The population included stay-green and drought-adapted donor lines together with locally adapted and high-yielding lines. Rather than considering the MAGIC design itself as novel, the value of the present population lies in the particular founder combination and its potential use as a breeding and genetic resource for sorghum improvement.

Genotyping of the population identified 5,512 SNPs (> 5% MAF and > 90% callrate). DArTseq is a GBS method that uses a reduced representation sequencing that targets gene-rich regions of the genome (Kilian et al., 2012) as was evident from the high percentage of genic SNPs (39.15% exonic and 11.91% intronic SNPs) in this study. The predominance of exonic SNPs suggests a high potential of this population for use in identifying the genetic basis of various traits which is fundamental for accelerated sorghum breeding. The PIC values of the SNPs in this study was between 0.09 and 0.49 with an average of 0.28 which was comparable with previous studies in sorghum (Enyew et al., 2022; Mudaki et al., 2023; Wondimu et al., 2021; Yahaya et al., 2023). The PIC is a measure of allelic diversity and ability of the markers to differentiate between genotypes based on their genetic relationship (Botstein et al., 1980). Markers with PIC values greater than 0.5 are considered highly informative, those between 0.5 and 0.25 are moderately informative, and those below 0.25 are slightly informative. Fifty three percent of these SNPs had PIC values above 0.25 which was moderately informative thus suitable for application in marker-aided sorghum breeding. The frequency of transitions was slightly more than transversions in this study (Ts/Tv = 1.02). This shows that transition mutations were tolerated better than transversion mutations during natural selection (C. Guo et al., 2017). This was similar to other studies (Kiranmayee et al., 2020; Mudaki et al., 2023; Wondimu et al., 2021).

The observed genetic structure provides insight into the current state of population development. ADMIXTURE stratified the population into 9 genetic groups. This provides critical insights into the complex interplay of founder lines contributions and the resulting population architecture. This stratification was consistent with PCA which showed clear yet partially overlapping clusters, indicating both differentiation and admixture among sub-populations, a pattern commonly observed in breeding populations undergoing recombination and selection (Price et al., 2006). The DAPC confirmed these distinctions by separating the most divergent groups (e.g., Pop1, Pop6, Pop7) while grouping the more admixed sub-populations into a central cluster, aligning with its strength in maximizing between-group variation (Jombart et al., 2010). This level of grouping was expected and desirable for multi-parental populations which are intentionally designed to capture broad allelic diversity while generating complex mosaics of founder haplotypes, thereby increasing mapping resolution (Yang et al., 2021).

The network analysis showed the relatedness and connectivity of the founder lines and the derived sub-populations. This visualization highlights how genomic segments from the different parents have been distributed and recombined across the breeding material following the crossing and objective selection steps. This characterization is essential for understanding the diversity and structure of newly developed breeding materials (Bandillo et al., 2013). Our results show a well-connected network among most founders and sub-populations, indicating successful incorporation of multiple parental genomes rather than a narrow diversity, while the presence of both tightly clustered and more distant nodes reflects natural recombination patterns and selection dynamics (Dell'Acqua et al., 2015). Such diversity provides a strong foundation for downstream applications including association mapping, marker development, and targeted improvement for traits such as drought tolerance (Bernardino et al., 2024; Min et al., 2025; Tsehaye et al., 2024). Identifying central versus more isolated founders further helps to pinpoint parents

likely to carry key or rare alleles, supporting informed decisions on line advancement and parental choice while maintaining broad genetic variation (Technow et al., 2014).

The average observed heterozygosity (H_o = 0.20) was slightly lower than the expected heterozygosity (H_e = 0.25), indicating a moderate deficit of heterozygotes within the population. Such a pattern is common to self-pollinating plant species including sorghum where successive inbreeding generations reduce heterozygosity relative to Hardy-Weinberg expectations (Mace et al., 2013). This was confirmed by a positive inbreeding coefficient (F_{is} = 0.18). Furthermore, the AMOVA showed that most of the variation (70.47%) was observed within the lines which was consistent with expectations for multi-parental populations. The extensive intercrossing and recombination in MAGIC populations promote high allelic mixing and diversity within individuals (Arrones et al., 2020). However, the 14.9% variation among the 9 sub-populations and the F_{st} value of 0.16, indicated existence of some genetic structure. The existence of genetic structure in our population differed from the observations by (Ongom & Ejeta, 2018). This could be due to incomplete fixation during selfing, unequal founder contributions or cryptic relatedness among lines sharing more genomic segments from the same founders.

The pairwise Euclidean genetic distance matrix estimated among individuals in this study provides an insight into the genetic variability of our multi-parental population. The broad distribution of genetic distances reflects substantial allelic diversity generated through the use of multiple founders and successive intermating, a hallmark of MAGIC populations (Cavanagh et al., 2008; Mackay et al., 2014), and is expected to enhance mapping resolution by reducing long-range linkage disequilibrium (Flint-Garcia et al., 2003). The distribution of founder lines across multiple clusters rather than discrete groupings indicates effective mixing of founder haplotypes and moderate population structure, consistent with reports in other cereal MAGIC populations (Bandillo et al., 2013; Ongom & Ejeta, 2018; Stadlmeier et al., 2018).

LD is the non-random association of alleles at different loci, shaped by recombination, mutation, selection or genetic drift. LD is a key measure of variation in breeding populations and is fundamental in determining the resolution and power of genome-wide association studies and genomic selection (Flint-Garcia et al., 2003; Gupta et al., 2005). In the present sorghum population, LD decayed to half of its maximum value (r^2=0.23) at approximately 339.34 kb, indicating a moderate extent of LD consistent with expectations for a structured, predominantly self-pollinating species. Comparable LD decay has been reported in MAGIC populations of sorghum (40kb and 220kb at r^2 = 0.1 and 0.2, respectively) (Ongom & Ejeta, 2018) and rice (200-400kb at r^2 = 0.24) (Zaw et al., 2019). However, the LD decay observed in this study was faster compared to that reported in several other MAGIC populations, reflecting enhanced historical recombination and increased haplotype resolution (Ayaad et al., 2021; Kumar et al., 2023; Meng, Guo, et al., 2016).

Days to flowering provided a useful initial trait for evaluating phenotypic segregation. The S_2-derived lines showed a broader flowering-time range (45–75 days) than the founder lines (51.5–61.5 days), indicating transgressive segregation. The occurrence of progeny flowering outside the parental range

suggests that recombination generated new combinations of alleles affecting flowering time. Importantly, this observation demonstrates phenotypic segregation in the developing population but does not by itself establish the genetic mechanisms responsible for the observed variation. Evaluation across three locations and two seasons further indicates that the population can be assessed under contrasting environmental conditions, although the current analysis was primarily intended to characterize the population and provide an initial genetic analysis of DTF.

The exploratory GWAS in this study demonstrated the effectiveness of the early-generation multi-parent population for detection of complex traits in sorghum using days to flowering (DTF) as a proof-of-concept trait. Flowering time is a key adaptive trait in plant species and strongly influence grain yield (Sannemann & Emma, 2015). The timing of floral transition is shaped by interactions between environmental cues, particularly photoperiod and temperature. These cues act through multiple genetic pathways, including photoperiod response, hormonal signaling, and endogenous developmental regulators. These pathways have been extensively characterized in the model plant *Arabidopsis thaliana* and other plant species including sorghum (Blümel et al., 2015; Jung et al., 2017; Mangal et al., 2025).

In sorghum, flowering time variation is classically controlled by the maturity loci *Ma1-Ma6*, which modulate photoperiod sensitivity and flowering behavior across environments (Mangal et al., 2025). Several of the associated genomic regions identified in this study were located on chromosomes that harbor known *Ma* loci. These associations did not consistently fall within the same LD decay intervals as the canonical Ma genes, may be due to the limited marker density used in this proof-of-concept analysis, thereby limiting definitive co-localization. Despite this, candidate gene mining revealed multiple genes with well-established roles in flowering time regulation across plant species, providing biological support for the detected associations. Notably, Sobic.010G138000, encoding a protein homologous to Cryptochrome 2 (CRY2), implicates photoperiod perception in flowering time variation. CRY2 plays a central role in photoperiodic flowering by linking light signals to downstream floral integrators in Arabidopsis (El-Din El-Assal et al., 2003; H. Guo et al., 1998; Zhao et al., 2022). The identification of Sobic.001G433000, a FAR1 family protein, further supports the involvement of phytochrome-mediated signaling pathways. FAR1-related proteins act downstream of phytochrome A to regulate light-responsive gene expression and flowering time, with conserved functions reported across monocots and dicots (Cheng et al., 2021; Tang et al., 2024). Additionally, Sobic.004G102301, encoding an auxin response factor (ARF), highlights the contribution of hormonal signaling to flowering time control, consistent with evidence that auxin integrates environmental and developmental signals during the floral transition (Okushima et al., 2005). Several transcription factor families previously implicated in flowering regulation were also identified, including bZIP, NAM/NAC, and RWP-RK domain-containing proteins. These transcription factors are known to integrate photoperiodic, hormonal, and endogenous cues controlling flowering time and developmental phase transitions (Lee et al., 2025; Liu et al., 2023; Sakuraba et al., 2022). This analysis provides a strong foundation for future studies employing increased marker density and functional validation to refine the genetic architecture of flowering and maturity traits in sorghum.

CONCLUSION

A partially developed MAGIC population comprising 176 S_2-derived lines was developed from eleven sorghum founders selected to combine drought-adapted and stay-green germplasm with locally adapted and high-yielding material. The population exhibited substantial genetic and phenotypic variation, detectable founder contributions, population structure, and transgressive segregation for days to flowering. Genome-wide LD declined to approximately half of its maximum mean r^2 at 339.34 kb, while exploratory GWAS identified nine SNPs associated with days to flowering across six chromosomes and several genes within the corresponding genomic intervals with functions potentially related to flowering-time regulation.

The findings demonstrate that the population has retained useful genetic and phenotypic diversity during its early stages of development and provides a promising resource for combining complementary sources of drought adaptation, stay-green expression, local adaptation, and productivity. The exploratory proof-of-concept GWAS with DTF identified multiple genomic regions associated with flowering time and highlighted biologically relevant candidate genes involved in photoperiod response and developmental regulation. Continued population advancement, phenotypic evaluation, and higher-density genotyping will be important for increasing the level of recombination and genetic resolution and for evaluating the population's broader value for genetic dissection and genomics-assisted breeding of agronomically important traits in sorghum.

Abbreviations

AMOVA
analysis of molecular variance
ARF
auxin response factor
BIC
Bayesian information criterion
DAPC
discriminant analysis of principal components
DArTseq
Diversity Arrays Technology sequencing
DArTseqLD
Diversity Arrays Technology sequencing low density
DNA
deoxyribonucleic acid
DTF
days to flowering
FarmCPU
Fixed and Random Model Circulating Probability Unification

F_{is}
inbreeding coefficient
F_{st}
fixation index
GBS
genotyping-by-sequencing
GWAS
genome-wide association study
He
expected heterozygosity
Ho
observed heterozygosity
LD
linkage disequilibrium
MAF
minor allele frequency
MAGIC
multi-parent advanced generation intercross
NC-II
North Carolina Design II
NJ
neighbor-joining
Nm
number of migrants
PCA
principal component analysis
PIC
polymorphic information content
pMAGIC
partially developed multi-parent advanced generation intercross
QTL
quantitative trait locus
SNP
single nucleotide polymorphism
TASSEL
Trait Analysis by aSSociation Evolution and Linkage
UHe
unbiased expected heterozygosity
UPGMA
unweighted pair group method with arithmetic mean

UTR
untranslated region.

Declarations

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Kisilu Rachael, 0000-0003-3932-9623

Kahiu Ngugi, 0000-0002-3845-2569

Lydia Wamalwa, 0000-0001-8787-8438

Kimani Wilson, 0000-0001-8653-8781

SUPPLEMENTAL MATERIAL

The supplemental materials are provide as a word document that contains Supplementary tables and figures.

Author Contribution

R.K. and K.N. conceptualized the research; R.K. , K.N. and L.W. approved the study design; K. W. and R.K. analyzed the data; R.K. and K.W. wrote the manuscript. All the authors reviewed the manuscript.

Acknowledgement

The authors would like to thank the former United States Agency for International Development (USAID) for their funding through Kenya Agricultural and Livestock Research Organization (KALRO).

Data Availability

All relevant data have been share in the supplementary materials.

References

1. Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics, 12(1), 246. https://doi.org/10.1186/1471-2105-12-246
2. Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., Prohens, J., & Gramazio, P. (2020). The Dawn of the Age of Multi-Parent MAGIC Populations in Plant Breeding: Novel Powerful Next-Generation Resources for Genetic Analysis and Selection of Recombinant Elite Material. Biology, 9(8). https://doi.org/10.3390/biology9080229
3. Ayaad, M., Han, Z., Zheng, K., Hu, G., Abo-Yousef, M., Sobeih, S. E. S., & Xing, Y. (2021). Bin-based genome-wide association studies reveal superior alleles for improvement of appearance quality using a 4-way MAGIC population in rice. Journal of Advanced Research, 28, 183–194. https://doi.org/https://doi.org/10.1016/j.jare.2020.08.001
4. Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., Tung, C.-W., McCouch, S., Thomson, M., Mauleon, R., Singh, R. K., Gregorio, G., Redoña, E., & Leung, H. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice, 6(1), 11. https://doi.org/10.1186/1939-8433-6-11
5. Bernardino, K. da C., Guilhen, J. H. S., de Menezes, C. B., Tardin, F. D., Schaffert, R. E., Bastos, E. A., Cardoso, M. J., Gazaffi, R., Rosa, J. R. B. F., Garcia, A. A. F., Guimarães, C. T., Kochian, L., Pastina, M. M., & Magalhaes, J. V. (2024). Genetic loci associated with sorghum drought tolerance in multiple environments and their sensitivity to environmental covariables. TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, 137(11), 259. https://doi.org/10.1007/s00122-024-04761-3
6. Blümel, M., Dally, N., & Jung, C. (2015). Flowering time regulation in crops—what did we learn from Arabidopsis? Current Opinion in Biotechnology, 32, 121–129. https://doi.org/https://doi.org/10.1016/j.copbio.2014.11.023
7. Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics, 32(3), 314–331.
8. Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23. https://doi.org/10.1093/bioinformatics/btm308
9. Cavanagh, C., Morell, M., Mackay, I., & Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Current Opinion in Plant Biology, 11(2), 215–221. https://doi.org/10.1016/j.pbi.2008.01.002
10. Cheng, M.-C., Kathare, P. K., Paik, I., & Huq, E. (2021). Phytochrome Signaling Networks. Annual Review of Plant Biology, 72(Volume 72, 2021), 217–244. https://doi.org/https://doi.org/10.1146/annurev-arplant-080620-024221
11. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Group, 1000 Genomes Project Analysis.

(2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

12. Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A. L., Aung, H. H., Nelissen, H., Baute, J., Frascaroli, E., Churchill, G. A., Inzé, D., Morgante, M., & Pè, M. E. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biology, 16(1), 167. https://doi.org/10.1186/s13059-015-0716-z
13. El-Din El-Assal, S., Alonso-Blanco, C., Peeters, A. J. M., Wagemaker, C., Weller, J. L., & Koornneef, M. (2003). The role of cryptochrome 2 in flowering in Arabidopsis. Plant Physiology, 133(4), 1504–1516. https://doi.org/10.1104/pp.103.029819
14. Flint-Garcia, S. a., Thornsberry, J., & Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. Ann Rev Plant Bio, 54, 357–374.
15. Gábor Csárdi and Tamás Nepusz and Vincent Traag and Szabolcs Horvát and Fabio Zanini and Daniel Noom and Kirill Müller and David Schoch and Maëlle Salmon. (2025). igraph: Network Analysis and Visualization in R. InterJournal, Complex Systems. https://doi.org/10.5281/zenodo.7682609
16. Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research, 40(Database issue), D1178–D1186. https://doi.org/10.1093/nar/gkr944
17. Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes, 5(1), 184–186. https://doi.org/https://doi.org/10.1111/j.1471-8286.2004.00828.x
18. Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources, 18(3), 691–699. https://doi.org/https://doi.org/10.1111/1755-0998.12745
19. Guo, C., McDowell, I. C., Nodzenski, M., Scholtens, D. M., Allen, A. S., Lowe, W. L., & Reddy, T. E. (2017). Transversions have larger regulatory effects than transitions. BMC Genomics, 18(1), 394. https://doi.org/10.1186/s12864-017-3785-4
20. Guo, H., Yang, H., Mockler, T. C., & Lin, C. (1998). Regulation of Flowering Time by Arabidopsis Photoreceptors. Science, 279(5355), 1360–1363. https://doi.org/10.1126/science.279.5355.1360
21. Gupta, P. K., Rustgi, S., & Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Molecular Biology, 57(4), 461–485. https://doi.org/10.1007/s11103-005-0257-z
22. Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129
23. Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11(1), 94. https://doi.org/10.1186/1471-2156-11-94

24. Jung, C., Pillen, K., Staiger, D., Coupland, G., & von Korff, M. (2017). Editorial: Recent Advances in Flowering Time Control. Frontiers in Plant Science, Volume 7–2016. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2016.02011
25. Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2, e281. https://doi.org/10.7717/peerj.281
26. Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms BT - Data Production and Analysis in Population Genomics: Methods and Protocols (F. Pompanon & A. Bonin (eds.); pp. 67–89). Humana Press. https://doi.org/10.1007/978-1-61779-870-2_5
27. Kimani, W., Zhang, L., Wu, X., Hao, H., & Jing, H. (2020). Genome-wide association study reveals that different pathways contribute to grain quality variation in sorghum (Sorghum bicolor). BMC Genomics, 21(112), 1–19.
28. Kiranmayee, K. N. S. U., Hash, C. T., Sivasubramani, S., Ramu, P., Amindala, B. P., Rathore, A., Kishor, P. B. K., Gupta, R., & Deshpande, S. P. (2020). Fine-Mapping of Sorghum Stay-Green QTL on Chromosome10 Revealed Genes Associated with Delayed Senescence. In Genes (Vol. 11, Issue 9, p. 1026). https://doi.org/10.3390/genes11091026
29. Kumar, N., Boatwright, J. L., Brenton, Z. W., Sapkota, S., Ballén-Taborda, C., Myers, M. T., Cox, W. A., Jordan, K. E., Kresovich, S., & Boyles, R. E. (2023). Development and characterization of a sorghum multi-parent advanced generation intercross (MAGIC) population for capturing diversity among seed parent gene pool. G3 Genes|Genomes|Genetics, 13(4), jkad037. https://doi.org/10.1093/g3journal/jkad037
30. Lee, W., Kim, S. H., Park, J., Yoon, S. H., Cho, S. W., Lee, N., Ito, S., Imaizumi, T., Hong, J. C., Chung, W. S., & Song, Y. H. (2025). The bZIP3-AS1 complex promotes CONSTANS-induced FLOWERING LOCUS T activation in a daylength-dependent manner. Journal of Integrative Plant Biology, 67(11), 2982–2998. https://doi.org/https://doi.org/10.1111/jipb.70014
31. Letunic, I., & Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Research, 52(W1), W78–W82. https://doi.org/10.1093/nar/gkae268
32. Liu, J., Qiao, Y., Li, C., & Hou, B. (2023). The NAC transcription factors play core roles in flowering and ripening fundamental to fruit yield and quality. Frontiers in Plant Science, 14, 1095967. https://doi.org/10.3389/fpls.2023.1095967
33. Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genetics, 12(2). https://doi.org/10.1371/journal.pgen.1005767
34. Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., Campbell, B. C., Hu, W., Innes, D. J., Han, X., Cruickshank, A., Dai, C., Frère, C., Zhang, H., Hunt, C. H., Wang, X., Shatte, T., Wang, M., Su, Z., …

Wang, J. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nature Communications, 4, 2320. https://doi.org/10.1038/ncomms3320

35. Mackay, I. J., Bansept-Basler, P., Barber, T., Bentley, A. R., Cockram, J., Gosman, N., Greenland, A. J., Horsnell, R., Howells, R., O'Sullivan, D. M., Rose, G. A., & Howell, P. J. (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3 (Bethesda, Md.), 4(9), 1603–1610. https://doi.org/10.1534/g3.114.012963
36. Mangal, H., Linders, K., Turkus, J., Shrestha, N., Long, B., Cebert, E., Kuang, X., Torres-Rodriguez, J. V., & Schnable, J. C. (2025). Genes and pathways determining flowering time variation in temperate-adapted sorghum. The Plant Journal: For Cell and Molecular Biology, 122(5), e70250. https://doi.org/10.1111/tpj.70250
37. McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., Kennedy, M., Amirebrahimi, M., Weers, B. D., McKinley, B., Mattison, A., Morishige, D. T., Grimwood, J., Schmutz, J., & Mullet, J. E. (2018). The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. The Plant Journal, 93(2), 338–354. https://doi.org/10.1111/tpj.13781
38. Meng, L., Guo, L., Ponce, K., Zhao, X., & Ye, G. (2016). Characterization of Three Indica Rice Multiparent Advanced Generation Intercross (MAGIC) Populations for Quantitative Trait Loci Identification. The Plant Genome, 9(2), plantgenome2015.10.0109. https://doi.org/https://doi.org/10.3835/plantgenome2015.10.0109
39. Meng, L., Zhao, X., Ponce, K., Ye, G., & Leung, H. (2016). QTL mapping for agronomic traits using multi-parent advanced generation inter-cross (MAGIC) populations derived from diverse elite indica rice lines. Field Crops Research, 189, 19–42. https://doi.org/https://doi.org/10.1016/j.fcr.2016.02.004
40. Min, H., Wang, K., Wang, T., Cheng, X., Habyarimana, E., Wang, Y., Hu, D., Wang, Y.-H., & Wang, L. (2025). Association mapping and candidate gene identification for drought tolerance in sorghum. Frontiers in Plant Science, Volume 16-2025. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2025.1629615
41. Mudaki, P., Wamalwa, L. N., Muui, C. W., Nzuve, F., Muasya, R. M., Nguluu, S., & Kimani, W. (2023). Genetic Diversity and Population Structure of Sorghum (Sorghum bicolor (L.) Moench) Landraces Using DArTseq-Derived Single-Nucleotide Polymorphism (SNP) Markers. Journal of Molecular Evolution. https://doi.org/10.1007/s00239-023-10108-1
42. Nei, M. (1987). Molecular evolutionary genetics. Columbia University Press.
43. Nei, Masatoshi. (1972). Genetic Distance between Populations. The American Naturalist, 106(949), 283–292. http://www.jstor.org/stable/2459777
44. Okushima, Y., Mitina, I., Quach, H. L., & Theologis, A. (2005). AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. The Plant

Journalfile:///C:/Users/Wkimani/Downloads/Citations-20260121T024247.Ris, 43(1), 29–46. https://doi.org/https://doi.org/10.1111/j.1365-313X.2005.02426.x

45. Ongom, P. O., & Ejeta, G. (2018). Mating Design and Genetic Structure of a Multi-Parent Advanced Generation Intercross (MAGIC) Population of Sorghum (Sorghum bicolor (L.) Moench). G3 (Bethesda, Md.), 8(1), 331–341. https://doi.org/10.1534/g3.117.300248
46. Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics, 20(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412
47. Pedersen, T. L. (2025). tidygraph: A Tidy API for Graph Manipulation (version 1.3.1.9000). https://github.com/thomasp85/tidygraph
48. Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. a, & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909. https://doi.org/10.1038/ng1847
49. R Core Team. (2024). R: A Language and Environment for Statistical Computing. https://www.r-project.org
50. Sakuraba, Y., Zhuo, M., & Yanagisawa, S. (2022). RWP-RK domain-containing transcription factors in the Viridiplantae: biology and phylogenetic relationships. Journal of Experimental Botany, 73(13), 4323–4337. https://doi.org/10.1093/jxb/erac229
51. Sannemann, W., & Emma, B. (2015). Multi-parent advanced generation inter-cross in barley : high-resolution quantitative trait locus mapping for flowering time as a proof of concept. https://doi.org/10.1007/s11032-015-0284-7
52. Sannemann, W., Huang, B. E., Mathew, B., & Léon, J. (2015). Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. Molecular Breeding, 35(3), 86. https://doi.org/10.1007/s11032-015-0284-7
53. Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., Clark, M., Dell'Acqua, M., Dixon, L. E., Filippi, C. V, Fradgley, N., Gardner, K. A., Mackay, I. J., O'Sullivan, D., Percival-Alwyn, L., Roorkiwal, M., Singh, R. K., Thudi, M., Varshney, R. K., … Mott, R. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity, 125(6), 396–416. https://doi.org/10.1038/s41437-020-0336-6
54. Stadlmeier, M., Hartl, L., & Mohler, V. (2018). Usefulness of a Multiparent Advanced Generation Intercross Population With a Greatly Reduced Mating Design for Genetic Studies in Winter Wheat. Frontiers in Plant Science, Volume 9–2018. https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2018.01825
55. Tang, H., Jing, D., Liu, C., Xie, X., Zhang, L., Chen, X., & Li, C. (2024). Genome-Wide Identification and Expression Analyses of the FAR1/FHY3 Gene Family Provide Insight into Inflorescence Development in Maize. In Current Issues in Molecular Biology (Vol. 46, Issue 1, pp. 430–449). https://doi.org/10.3390/cimb46010027
56. Team, R. C. (2024). R: A Language and Environment for Statistical Computing (4.4.2). R Foundation for Statistical Computing. https://www.r-project.org/

57. Technow, F., Schrag, T. A., Schipprack, W., Bauer, E., Simianer, H., & Melchinger, A. E. (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics, 197(4), 1343–1355. https://doi.org/10.1534/genetics.114.165860
58. Thieurmel, A. B. V. and C. and B. (2022). visNetwork: Network Visualization using “vis.js” Library (R package version 2.1.2). https://cran.r-project.org/package=visNetwork
59. Tsehaye, Y., Menamo, T. M., Abay, F., Tadesse, T., & Bantte, K. (2024). Multi-locus genome-wide association study for grain yield and drought tolerance indices in sorghum accessions. The Plant Genome, 17(4), e20505. https://doi.org/https://doi.org/10.1002/tpg2.20505
60. VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science, 91(11), 4414–4423. https://doi.org/https://doi.org/10.3168/jds.2007-0980
61. Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research, 38(16), e164. https://doi.org/10.1093/nar/gkq603
62. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686
63. Wondimu, Z., Dong, H., Paterson, A. H., Worku, W., & Bantte, K. (2021). Genetic diversity, population structure, and selection signature in Ethiopian sorghum [Sorghum bicolor L. (Moench)] germplasm. G3 Genes|Genomes|Genetics, 11(6), jkab087. https://doi.org/10.1093/g3journal/jkab087
64. Wright, S. (1968). Evolution and the Genetics of Populations, Volume 1: Genetic and Biometric Foundations. University of Chicago Press. https://books.google.de/books?id=Eq0DxAEACAAJ
65. Yang, C. J., Edmondson, R. N., Piepho, H.-P., Powell, W., & Mackay, I. (2021). Crafting for a better MAGIC: systematic design and test for Multiparental Advanced Generation Inter-Cross population. G3 (Bethesda, Md.), 11(11). https://doi.org/10.1093/g3journal/jkab295
66. Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., & Liu, X. (2021). rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. Genomics, Proteomics & Bioinformatics, 19(4), 619–628. https://doi.org/https://doi.org/10.1016/j.gpb.2020.10.007
67. Zaw, H., Raghavan, C., Pocsedio, A., Swamy, B. P. M., Jubay, M. L., Singh, R. K., Bonifacio, J., Mauleon, R., Hernandez, J. E., Mendioro, M. S., Gregorio, G. B., & Leung, H. (2019). Exploring genetic architecture of grain yield and quality traits in a 16-way indica by japonica rice MAGIC global population. Scientific Reports, 9(1), 19605. https://doi.org/10.1038/s41598-019-55357-7
68. Zhao, Z., Dent, C., Liang, H., Lv, J., Shang, G., Liu, Y., Feng, F., Wang, F., Pang, J., Li, X., Ma, L., Li, B., Sureshkumar, S., Wang, J.-W., Balasubramanian, S., & Liu, H. (2022). CRY2 interacts with CIS1 to regulate thermosensory flowering via FLM alternative splicing. Nature Communications, 13(1), 7045. https://doi.org/10.1038/s41467-022-34886-2

Figures

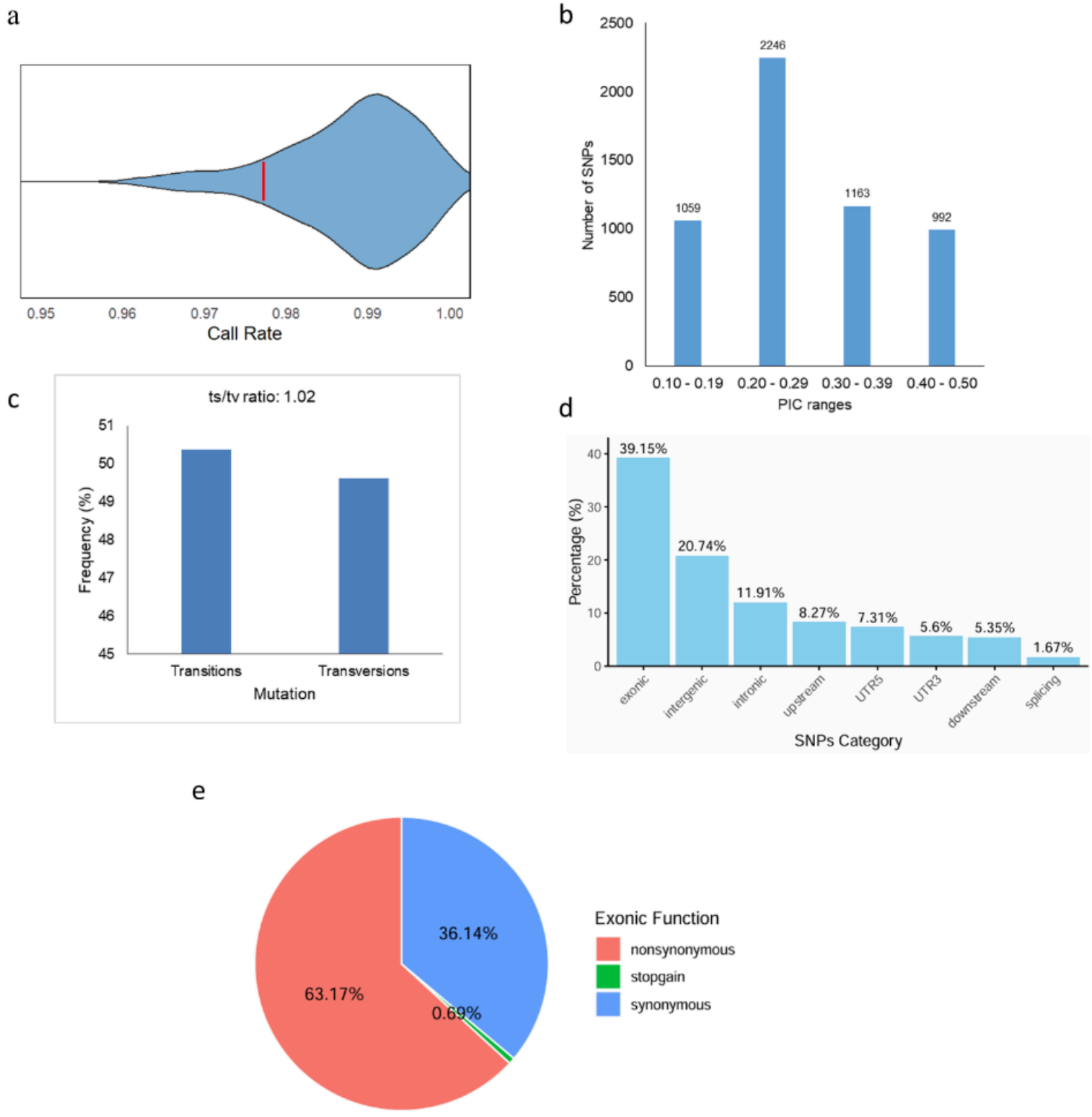


Figure 1

Genetic markers characterization for the pMAGIC population. **a**call rate of SNP markers per individual, **b**Polymorphic information content (PIC) range values of the 5,512 SNP markers, **c** Proportion of SNP transitions and transversions mutation types among the 5,512 SNPs, **d** Functional annotation of the

SNPs based on their genomic regions, and **e** annotation of exonic SNPs based on predicted functional impact.

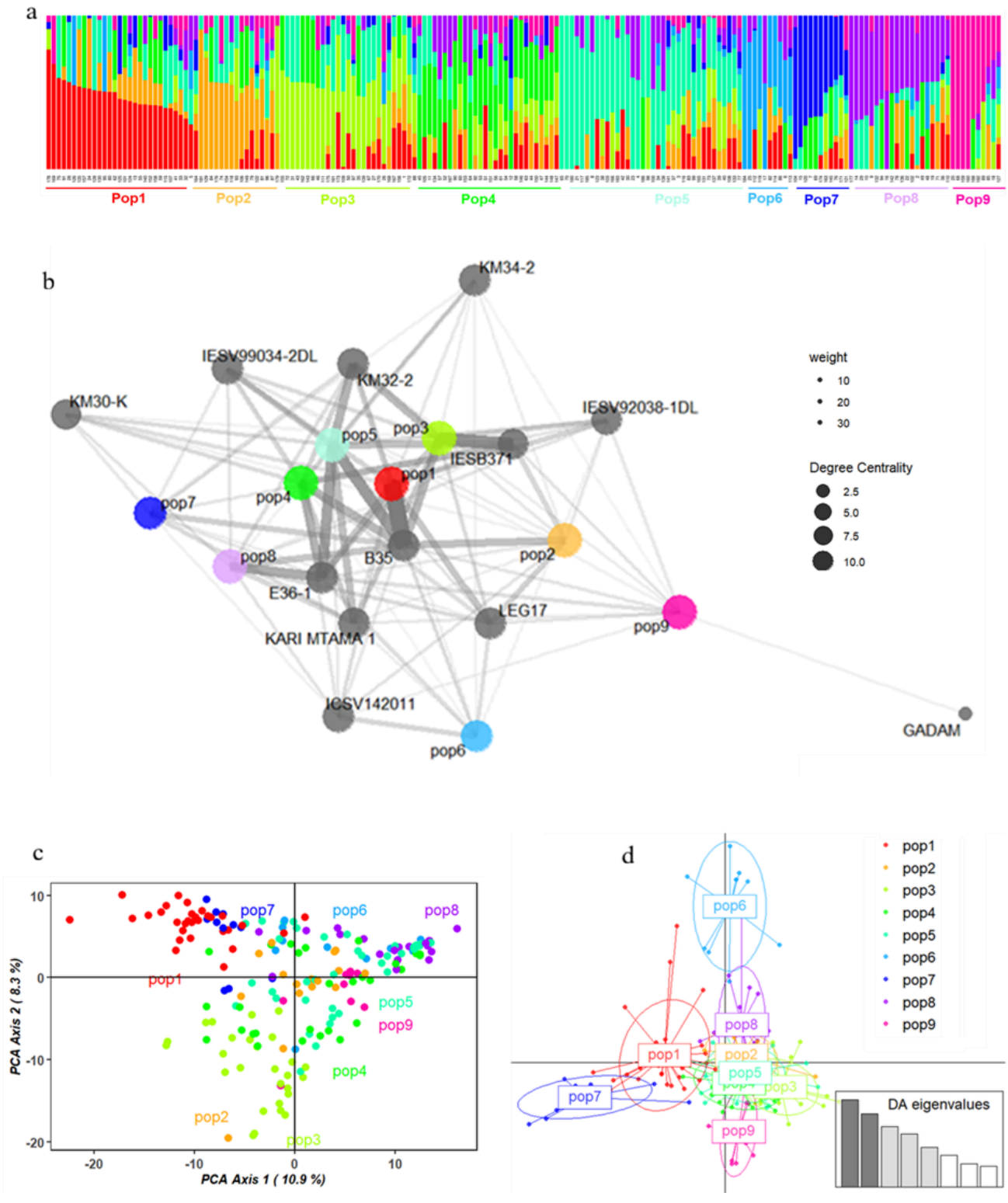


Figure 2

Population structure of the sorghum pMAGIC population and their founder lines. **a**Hierarchical organization of genetic relationship of 188 sorghum lines. Different colours represent different sub-

populations. Each bar represent an individual line. **b** Network Analysis of parental contributions in the sorghum pMAGIC population. The nodes represent parental lines (gray) and derived sub-populations (coloured according to the sub-populations in ADMIXTURE analysis). The size of the dot corresponds to degree centrality (number of connections) indicating the extent of connectivity within the network. Edge thickness represents the number of shared pMAGIC lines between parents and the sub-populations. **c**Principal Component Analysis (PCA) plot of the pMAGIC lines. A total of 5,512 SNP markers were used in the analysis. **d**Discriminant analysis of principal components (DAPC) of the pMAGIC lines and their founders. The sub-groups in PCA and DAPC were colour-coded based on the nine sub-populations of the ADMIXTURE results.

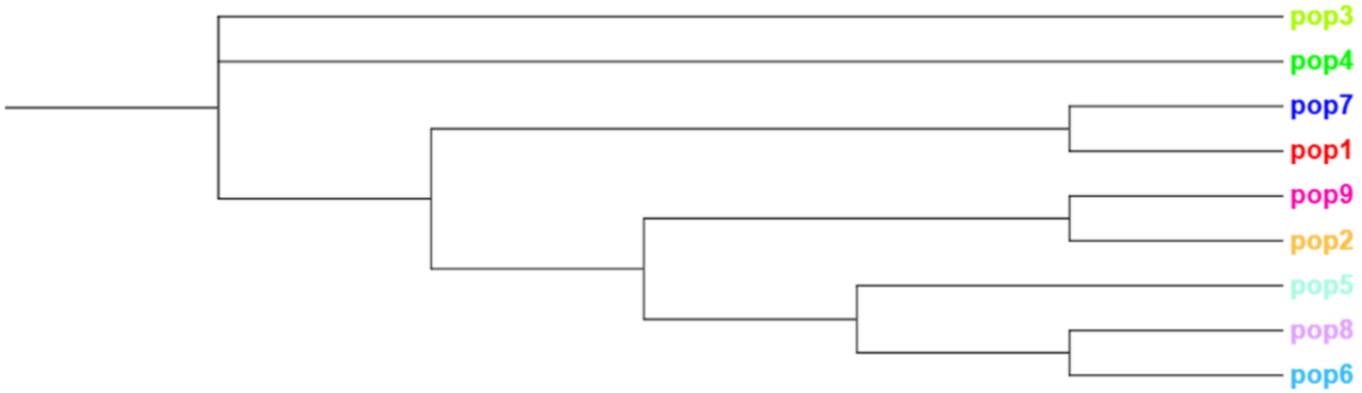


Figure 3

Neighbour-joining phylogenetic tree constructed using the fixation index (F_{st}) values illustrating the relationship among the nine sub-populations of the sorghum MAGIC population.

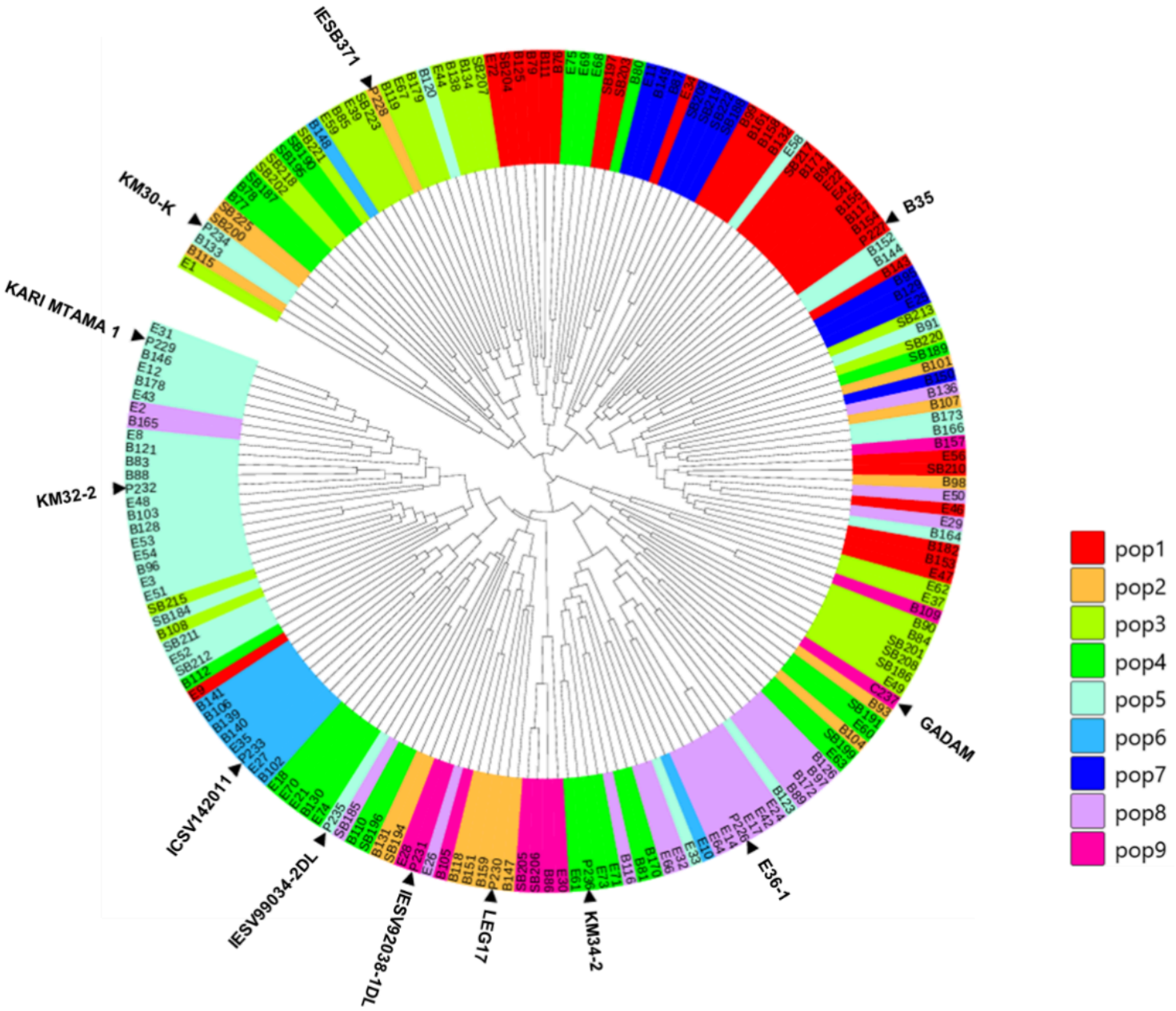


Figure 4

Clustering analysis based on Euclidean genetic distance to show relationship between the 188 sorghum MAGIC lines using 5,512 SNP markers. The arrow heads indicate the parental lines used to construct the MAGIC population. Colours were assigned based on their respective sub-populations as determined by ADMIXTURE.

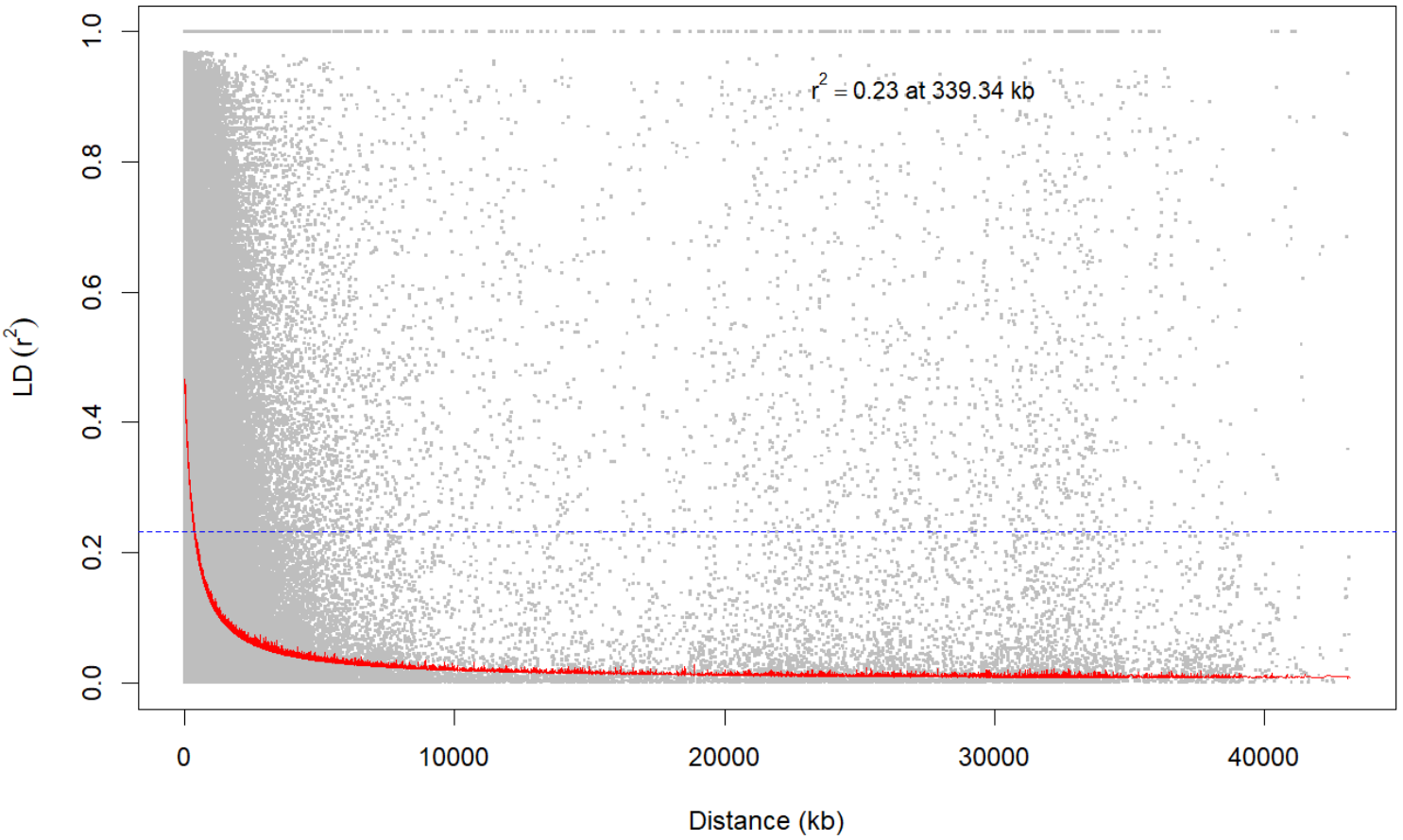


Figure 5

Linkage disequilibrium (LD) decay plot of the sorghum MAGIC population based on pairwise r^2 values among SNPs plotted against the distance across the *Sorghum bicolor* genome. The values on the y-axis are the squared correlation coefficients (r^2) between alleles at a loci while x-axis represents the physical distance in kilobases. Each grey point represents a pairwise r^2 value of two SNPs at a given distance.

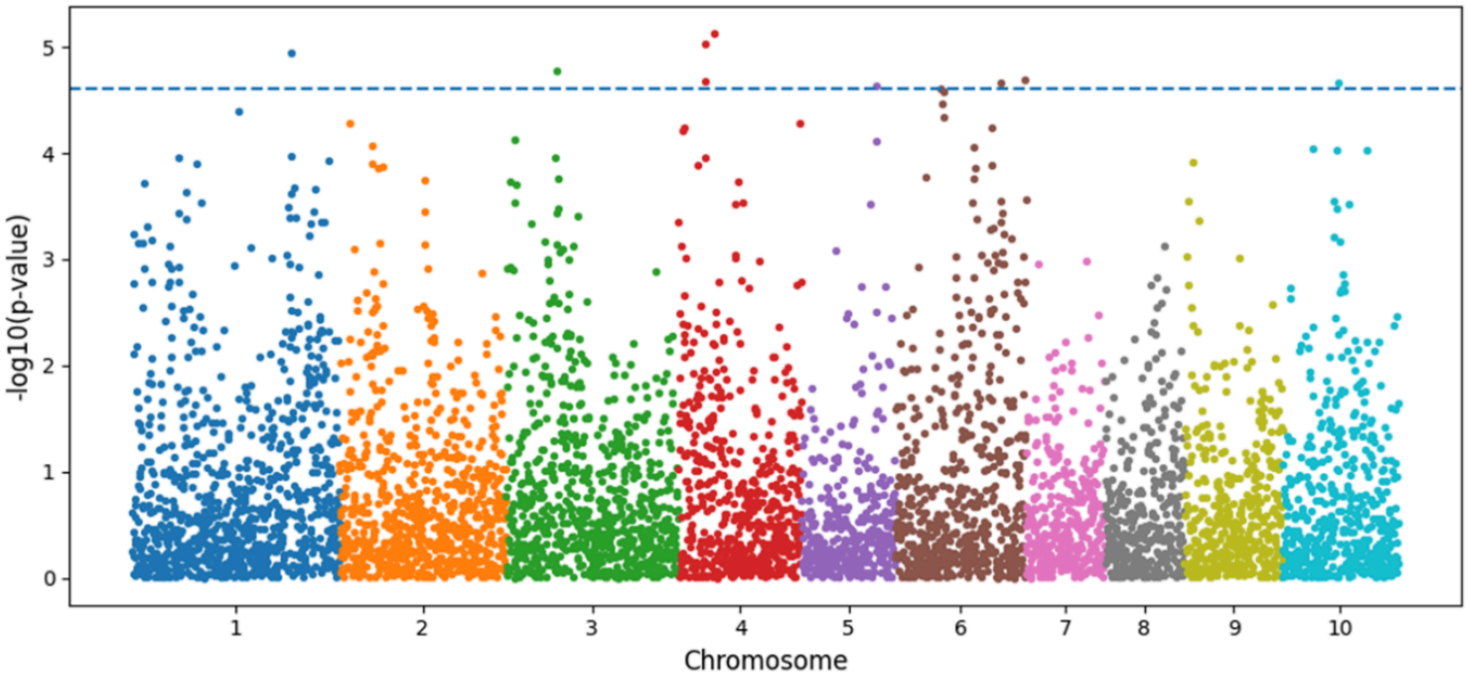


Figure 6

Manhattan plot showing the GWAS for days to flowering (DTF) in the sorghum pMAGIC population. The x-axis represents the genomic position of SNP markers across the ten sorghum chromosomes, while the y-axis shows the statistical significance of marker–trait associations expressed as $-\log_{10}$(p-values). Each point corresponds to a single SNP, and different colors distinguish chromosomes. The horizontal dashed line indicates the genome-wide significance threshold ($p = 2.48 \times 10^{-5}$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaldata.docx