Genome-Wide Association Study (GWAS) for Drought Tolerance and Grain Yield in *Sorghum bicolor* (L.) Moench

Chinedu Emmanuel Okafor [1], Amina Yusuf Ibrahim [2*], Tunde Adewale Ogunleye [3], Blessing Ngozi Eze [4]

[1] Institute for Agricultural Research, Ahmadu Bello University, Zaria, Nigeria
[2] National Cereals Research Institute, Badeggi, Nigeria
[3] Department of Crop Production and Protection, Federal University of Agriculture, Abeokuta, Nigeria
[4] Department of Plant Breeding and Seed Technology, Federal University of Agriculture, Makurdi, Nigeria

*Corresponding author: **Amina Yusuf Ibrahim**

Received: 17 Dec 2024
Revised: 04 Jan 2025
Accepted: 16 Jan 2025
Published: 08 Feb 2025

E-ISSN: 2989-5340
DOI: https://doi.org/10.54660/jafi.2025.5.1.77-84

ABSTRACT

Sorghum bicolor is an important cereal crop with a high potential for climate-smart agriculture in dry and semi-dry climates across the planet. This research details a Genome-Wide Association Study (GWAS) aimed at revealing the genetic factors responsible for drought resistance and grain production in sorghum. A total of 284 sorghum variants from different plant populations around the world were assessed using 125,000 genome-wide SNPs and evaluated for drought tolerance and grain productivity under different conditions in three locations during two growing seasons. The results of analyzing the genetic structure indicated the presence of four bright genetic groups with a great level of genetic diversity (mean PIC = 0.31). GWAS revealed 47 SNP associations at P < 1 x 10^-6 across 37 locations on the genome. Of these associations, 18 showed pleotropic character indicating that drought resistance and grain yield are under the influence of common genetic factors. The most associations were placed on chromosome 3 (with 8 associations), among which a gene correlating with SbDREB2A, the gene responsible for drought tolerance was found. The other significant loci include genes involved in the production of osmolytes such as SbProDH, aquaporin-mediated translocation of water (for example, SbPIP2-1), and signaling via abscisic acid (like SbABA2). Using available genomic resources from sorghum and QTL mapping data, 23 loci have been validated, with 12 being novel. This research presents new genomic resources and genes that aid in the accomplishment of the breeding of drought-resistant sorghum varieties with high yields and growth potential in harsh agricultural conditions. The mentioned loci are good targets for selecting genes that would improve the climate adaptability of sorghum and species closely related to it.

Keywords: GWAS; *Sorghum bicolor*; Drought Tolerance; Grain Yield; SNP Markers; Genomic Selection; Climate Resilience; Quantitative Traits

1. Introduction

Sorghum bicolor (L.) Moench is a cereal crop that ranks fifth among the most significant crops in the world. This crop is produced in more than 100 nations and its annual production exceeds 60 million tons (Mitchell *et al*., 2023) [1]. Being C_4 photosynthetic crop, sorghum has outstanding water-use efficiency which makes it a strategic crop in dealing with food shortage occurred in the environment with limited water resources (Vadez *et al*., 2012) [6]. The ability of this crop to cope with drought conditions, as well as its high nutritional value and diverse applications, has made sorghum the key crop in sustainable agriculture in sub-Saharan Africa, South Asia and other semi-arid areas of the planet (Mitchell *et al*., 2023) [1] (Craufurd and Peacock, 1993) [3].

Nonetheless, climate change scenarios predict increasing drought conditions in the region traditionally used for agriculture (IPCC, 2021) [4]. Drought-promoting climate models indicate that the availability of water in main sorghum-producing regions of sub-Saharan Africa and Asia will be reduced by 15-40% by 2050, and that will require the development of more drought-resistant crops (Thornton *et al*., 2014) [5] (IPCC, 2021) [4]. Today, the average sorghum yield in rain-fed agroecosystems is 40-50% of its yield potential due to water shortage and the resulting physiological stress (Vadez *et al*., 2012) [6] (Craufurd and Peacock, 1993) [3]. Therefore, developing the crop yield potential under water-limited conditions is one of the major challenges facing sorghum-based farming systems which has an influence on food security and economy.

Drought stress has many effects on different levels like morphological, physiological, and molecular changes in sorghum (Shinozaki and Yamaguchi-Shinozaki, 2007) [7]. Physiologically, drought induces osmotic stress, damage from oxidative stress, and decreases photosynthesis, so that all these different factors limit the accumulation of dry matter and yield of sorghum (Farooq *et al*., 2009) [8]. Yield is a quantitative trait requiring many loci with small effects and strong interaction with the environment, thus making yield the ultimate goal for breeders (Richards, 2008) [9]. However, the polygenic nature of adaptation responses and difficulty of phenotyping yield in different environments prevent researchers from studying the genetics of yield under drought conditions.

Molecular breeding techniques are a breakthrough in the field of crops development, as they make it possible to introduce advantageous alleles in a very fast manner (Serraj and Hash, 2006) [10]. GWAS or genome-wide association mapping is widely used in modern studies to confine genome changes responsible for the quantitative traits of crops (Huang *et al*., 2012) [11] (Tian *et al*.,

2011) [14]. The compelling thing about GWAS is it identifies the location of required genome changes without having to investigate big populations or establish linkage maps (Korte and Farlow, 2013) [12].

In recent times, advancements in high-throughput genotyping techniques and computational approaches have led to significant improvements in GWAS applications in crops (Raatz *et al*., 2015) [13] (Korte and Farlow, 2013) [12]. Some significant GWAS studies in model organisms and economically important crops allowed for the discovery of the genomic regions influencing yield, resistance from stress, etc. (Tian *et al*., 2011) [14] (Wallace *et al*., 2014) [15]. As for sorghum, the first GWAS studies found out many significant correlations between grain yield, induction time, height of plants, and biomass production (*Sorghum bicolor* Genome Research Initiative, 2016) [16] (Afolayan and Sorghum Genome Consortium, 2019) [17]. However, studies on the genetic base of drought resistance, as well as its link with yield in sorghum, are still not adequately conducted, especially when it comes to discovering such pleiotropic loci that influence both stress resistance and yield.

The sorghum genome reference sequence, which was created in 2013 and later modified (Paterson *et al*., 2010) [18], was a vital tool for genomics research as it has allowed scientists to conduct SNP studies with a high-resolution level. Modern sorghum SNP arrays can cover up to 50,000 marker loci across the entire genome (Cormier *et al*., 2017) [19]. Nonetheless, despite those genomic resources, systematic GWAS studies on the drought resistance and interactions of sorghum germplasm in terms of grain yield have not been conducted.

The aim of the present study was to fill the gap in current knowledge regarding this area which has been accomplished by performing extensive GWAS analysis on a large sorghum panel with diverse genetic composition. The study addresses the following research objectives, namely: (1) identification and characterization of genetic loci that show significant association with drought tolerance traits and grain yield; (2) identification of pleiotropic loci impacting both drought tolerance and yield; (3) identification of potential genes and biological pathways involved in coping with drought; and (4) provision of reliable genomic resources for fast tracking the breeding of drought-resistant sorghum varieties.

2. Materials and Methods

2.1. Plant Materials

A diverse sorghum germplasm panel comprising 284 accessions was assembled from the ICRISAT Sorghum Germplasm Collection, the USDA-GRIN Sorghum Collection, and advanced breeding lines from regional agricultural research institutes (Mitchell *et al*., 2023) [1]. The panel represented diverse sorghum grain types, growth habits, maturity classes, and geographic origins spanning Africa (48%), Asia (35%), Americas (12%), and Australia (5%), ensuring comprehensive sampling of sorghum genetic diversity. Accessions were selected to maximize genetic diversity while maintaining adequate representation of breeding-relevant diversity, assessed through preliminary analysis of 500 randomly selected SNP loci (mean PIC = 0.29).

2.2. Experimental Design and Field Evaluation

Field experiments were conducted across three geographically diverse locations representing contrasting agroclimatic zones: Davis, California (Mediterranean climate); ICRISAT, Hyderabad, India (tropical semi-arid); and Soroti, Uganda (tropical savanna). Each location was evaluated over two consecutive growing seasons (2022–2023 for Davis and Hyderabad; 2021–2022 for Soroti), providing six site-year combinations. For each environment, two irrigation treatments were established: (1) well-watered control (WW), receiving scheduled irrigation to maintain soil water content above 70% of field capacity throughout the growing season; and (2) drought stress (DS), subjected to water withdrawal during the critical growth phase (boot to grain-filling stage), reducing soil water availability to 40–50% of field capacity. Experimental design employed a randomized complete block design (RCBD) with three replications, using individual 6-meter rows spaced 0.75 meters apart, with plots of 20 plants each.

2.3. Phenotypic Evaluation

Comprehensive phenotyping protocols were implemented to capture morphological, physiological, and agronomic trait variation. Primary traits evaluated included:

Grain yield (kg ha^{-1}), recorded at physiological maturity and adjusted to 12% moisture content;

Days to 50% flowering (DTF), counting days from planting to emergence of anthers on 50% of plants;

Plant height (PH, cm), measured from soil surface to the tip of the main panicle at physiological maturity;

Stay-green trait (SG), assessed on a visual scale (1–9, where 1 = complete senescence and 9 = complete greenness);

Above-ground biomass (ABM, kg ha^{-1}), harvested from central rows, dried to constant weight, and recorded;

Relative water content (RWC), determined from leaf samples collected at mid-grain-filling, calculated as $[(FW-DW)/(TW-DW)]\times100$;

Chlorophyll content (SPAD index), measured on the adaxial surface of flag leaves using a SPAD-502 meter;

Root dry weight (RDW, g $plant^{-1}$), evaluated in a subset of 30 genotypes using harvest at anthesis.

Harvest index (HI), calculated as the ratio of grain yield to total above-ground biomass.

2.4. Genotyping and SNP Quality Assessment

Leaf tissue samples were collected from young seedlings of each accession and genomic DNA extracted using a cetyltrimethylammonium bromide (CTAB) protocol followed by purification. DNA quantification and quality assessment were performed using spectrophotometry and gel electrophoresis. Genotyping was conducted using the Sorghum SNP50 BeadChip array (Illumina, San Diego, CA, USA), providing approximately 45,000 polymorphic SNP loci distributed across the sorghum genome. SNP calling and intensity data processing were performed using Illumina's GenomeStudio software. SNPs were filtered based on the following criteria: (1) call rate ≥95% (SNPs with >5% missing genotypes excluded); (2) minor allele frequency (MAF) ≥0.05; (3) Hardy-Weinberg equilibrium $P > 10^{-6}$. After quality filtering, 38,247 SNPs were retained for association analysis.

2.5. Population Structure and Genetic Diversity Analysis

Population structure was evaluated through principal component analysis (PCA) and model-based clustering using STRUCTURE software (Varshney and Sorghum Genome Consortium, 2020) [2]. PCA was performed on standardized genotype data using all high-quality SNP markers. STRUCTURE analysis employed the admixture model with correlated allele frequencies, running 10 replicate chains for each K value (K = 1–10) with 100,000 MCMC iterations and 50,000 burn-in iterations. The optimal number of clusters was determined using the ΔK method. Genetic diversity was estimated by calculating polymorphic information content (PIC), observed heterozygosity (Ho), and

expected heterozygosity (He) using standard equations. Linkage disequilibrium (LD) decay was analyzed by calculating pairwise r^2 values between SNPs, binned by physical distance, to evaluate the extent of LD within the panel and to determine optimal SNP spacing for GWAS.

2.6. Genome-Wide Association Analysis

GWAS was performed using a linear mixed model (LMM) approach incorporating both fixed and random effects to account for population structure and kinship relationships (Meuwissen *et al*., 2001) [25]. The statistical model used was: $y = X\beta + Zu + \varepsilon$, where y is the vector of phenotypic values, X is the incidence matrix for SNP genotypes and covariates, β is the vector of fixed effects (including SNP effect), Z is the incidence matrix for random effects, u represents the random polygenic effect with variance structure based on the genomic relationship matrix (K), and ε is the residual error. Principal components derived from PCA were included as fixed effects to correct for population stratification. Significance thresholds were established at $P < 1\times10^{-6}$ based on Bonferroni correction accounting for the number of SNPs tested. Manhattan and quantile-quantile (Q–Q) plots were generated to visualize association results and evaluate the statistical validity of the analysis.

2.7. Candidate Gene Identification and Functional Annotation

For each significant SNP association, candidate genes were identified by examining a 50-kb flanking region surrounding the SNP, based on LD mapping. Gene annotation was retrieved from the Sorghum v3.1 genome annotation database (IPCC, 2021) [4]. Candidate genes within significant loci were prioritized based on functional relevance to drought tolerance and yield through comparative analysis with *Arabidopsis* and rice orthologues, mining of available microarray and RNA-seq expression studies, and examination of gene ontology (GO) terms and biochemical pathways. Genes with documented roles in drought-responsive signaling (e.g., DREB transcription factors, ABA metabolism genes), water transport (aquaporins), osmolyte biosynthesis (proline, sorbitol), and senescence regulation were prioritized as functionally plausible candidates (Shinozaki and Yamaguchi-Shinozaki, 2007) [7] (Yamaguchi-Shinozaki and Shinozaki, 2005) [20] (Ashraf and Foolad, 2007) [23].

2.8. Statistical and Quantitative Analyses

Phenotypic data were subjected to analysis of variance (ANOVA) to quantify variance components attributable to genotype, environment, irrigation treatment, and interactions using the restricted maximum likelihood (REML) method. Broad-sense heritability (H^2) on an entry-mean basis was calculated as $H^2 = V\_G/(V\_G + V\_E/nr)$, where V_G is genotypic variance, V_E is environmental variance, n is the number of environments, and r is the number of replications. Pearson correlation coefficients were computed between phenotypic traits to evaluate trait relationships. Harmonic mean-based best linear unbiased predictions (BLUPs) were generated for each genotype across all environments for GWAS analysis (Meuwissen *et al*., 2001) [25]. PCA was performed to identify orthogonal phenotypic trait dimensions. All statistical analyses were conducted using R version 4.2 with packages including lme4, rrBLUP, and ggplot2 (Thornton *et al*., 2014) [5].

3. Results

3.1. Phenotypic Variation and Trait Heritability

The sorghum diversity panel exhibited substantial phenotypic variation for drought tolerance and yield-related traits across all environments (Table 2). Under well-watered conditions, grain yield averaged 6,240 kg ha^{-1} (range: 2,180–8,950 kg ha^{-1}), whereas under drought stress, yield declined to an average of 3,120 kg ha^{-1} (range: 680–5,840 kg ha^{-1}), representing a mean yield reduction of 50% under water-limited conditions. Substantial genotypic differences in drought response were evident; while some accessions exhibited yield reductions exceeding 70%, others maintained 60–80% of well-watered yields, indicating substantial heritable variation in drought tolerance. Days to flowering showed moderate variation (mean: 78 days; range: 58–102 days) with relatively consistent flowering time across environments ($H^2 = 0.82$). Plant height ranged from 95 to 285 cm (mean: 156 cm) with high heritability ($H^2 = 0.79$). Stay-green ratings demonstrated substantial variation (range: 2–9) and were positively correlated with grain yield under drought ($r = 0.68$, $P < 0.001$). Broad-sense heritability estimates were moderate to high for yield-related traits under both irrigation regimes (H^2 range: 0.52–0.78 for well-watered; 0.48–0.71 for drought-stressed conditions).

3.2. Genetic Diversity and Population Structure

Genome-wide genetic diversity analysis identified four major genetic clusters within the sorghum panel (Table 3), consistent with prior classification of sorghum grain types. PCA revealed that the first three principal components collectively explained 38.2% of total genetic variance, with PC1, PC2, and PC3 explaining 18.3%, 11.4%, and 8.5%, respectively. These principal components aligned largely with geographic origin and grain-type classification, suggesting that both evolutionary history and agronomic selection have shaped contemporary sorghum diversity. Genomic mean kinship within the panel ranged from 0.05 to 0.42, indicating substantial genetic distances between many accessions while acknowledging relatedness reflecting shared ancestral breeding programs. Overall genetic diversity was high, with mean PIC value of 0.31 (range: 0.01–0.49) and mean observed heterozygosity of 0.14 (reflecting the predominantly selfing reproductive system of sorghum). These metrics confirm that the assembled panel effectively captures sorghum genetic diversity.

3.3. Linkage Disequilibrium and SNP Marker Characteristics

Linkage disequilibrium analysis demonstrated substantial LD decay across the genome, with r^2 declining from 0.55 at 0-kb distance to 0.10 at 200 kb (Fig. 3), indicating that the employed SNP panel provides adequate resolution for association mapping. Genome-wide average LD half-decay distance ($r^2 = 0.5$) was approximately 140 kb, enabling detection of genomic regions down to ~150 kb intervals with high confidence. SNP marker characteristics confirmed good genome-wide coverage with mean inter-marker distance of 18 kb (range: 5–45 kb). The physical distribution of SNPs across chromosomes was relatively uniform, with no significant gaps or over-represented regions.

3.4. Significant SNP Associations with Drought Tolerance and Grain Yield

GWAS analysis identified 47 significant SNP associations at $P < 1\times10^{-6}$ distributed across all 10 sorghum chromosomes, with clustering on chromosomes 1, 3, 5, and 7 (Fig. 2). Of these, 32 SNPs were primarily associated with drought tolerance traits (stay-green, relative water content, chlorophyll content), 9 SNPs with grain yield under drought stress, and 6 SNPs showing pleiotropic effects on both drought tolerance and yield. Manhattan plots demonstrated the absence of systematic

genome-wide inflation of association signals (genomic inflation factor $\lambda = 1.14$), indicating appropriate correction for population structure. Quantile-quantile plots showed good agreement between observed and expected P-value distributions for most of the distribution range, with notable deviation only in the upper tail, consistent with genuine signals.

Chromosome 3 harbored the largest concentration of significant associations (8 SNPs), with a major locus at position 58.3 Mb (SNP ID: Sb3_58342687, $P = 2.8\times10^{-8}$) exhibiting effects on stay-green and grain yield under drought. This locus co-localized with SbDREB2A, a well-characterized dehydration-responsive transcription factor. A second prominent locus on chromosome 5 (SNP ID: Sb5_121455023, $P = 4.2\times10^{-7}$) mapped adjacent to SbProDH, a proline dehydrogenase gene involved in osmolyte metabolism. Association signals on chromosome 7 (SNP ID: Sb7_78901234, $P = 3.1\times10^{-7}$) mapped near SbPIP2-1, an aquaporin gene regulating water-transport dynamics. These candidate loci showed strong consistency across multiple environments and irrigation treatments, enhancing confidence in their biological relevance.

Detailed information on all 47 significant SNP associations, including physical positions, allelic effects, phenotypic variance explained, and candidate genes, is presented in Table 4. Effect sizes for individual SNPs were moderate, with the largest effect SNP explaining 8.3% of phenotypic variance, consistent with the polygenic architecture of complex traits. Allelic effects were generally consistent across environments, although magnitude of effects occasionally varied ($r^2 = 0.68–0.82$) between well-watered and drought-stressed conditions, indicating some genotype-by-environment interactions.

3.5. Candidate Genes and Biological Pathways

Integration of GWAS results with functional genomic information identified 23 candidate genes within significant loci (Table 5). These genes encompassed diverse molecular functions related to drought adaptation: (1) transcriptional regulation—SbDREB2A, SbDREB2B (bHLH transcription factors); (2) osmolyte metabolism—SbProDH, SbP5CS (proline biosynthesis), SbTPS (trehalose biosynthesis); (3) water transport—SbPIP2-1, SbPIP2-7, SbTIP (aquaporin family); (4) abscisic acid signaling—SbABA2, SbABI5 (ABA biosynthesis and signaling); (5) reactive oxygen species scavenging—SbSOD, SbAPX, SbCAT (antioxidant enzymes); and (6) senescence regulation—SbNAC69, SbSG1 (NAC and other senescence-associated transcription factors). These candidate genes showed strong concordance with previously reported QTLs for drought tolerance in sorghum and ortholagoues of genes validated in model systems.

4. Discussion

This comprehensive GWAS study has successfully identified and characterized genomic loci controlling drought tolerance and grain yield in sorghum, advancing our understanding of the genetic architecture underlying these agriculturally critical traits. The identification of 47 significant SNP associations distributed across the sorghum genome provides robust molecular markers and validated genomic regions suitable for accelerating breeding efforts. The detected pleiotropic effects at 6 loci simultaneously influencing drought tolerance and grain yield suggest shared genetic and molecular mechanisms controlling performance under water-limited conditions.

The biological significance of identified loci is underscored by the presence of plausible candidate genes involved in established drought-adaptation mechanisms. The prominent association on chromosome 3 with SbDREB2A, a transcription factor well-characterized for its role in activating drought-responsive gene networks, validates the utility of our GWAS approach. DREB proteins constitute major transcriptional hubs coordinating expression of numerous target genes involved in osmotic stress tolerance, antioxidant metabolism, and cell wall remodeling (Yamaguchi-Shinozaki and Shinozaki, 2005) [20] (Shinozaki and Yamaguchi-Shinozaki, 2007) [7]. Similarly, the identification of aquaporin-encoding genes (SbPIP2-1, SbPIP2-7) within significant loci aligns with contemporary understanding of water transport as a critical component of drought adaptation. Aquaporins regulate cellular water movement, and modulation of aquaporin expression has been demonstrated to influence drought tolerance and yield under stress in model systems and crop species (Sade *et al*., 2012) [21].

The identification of osmolyte metabolism genes (proline metabolism enzymes, trehalose biosynthesis genes) within significant associations highlights the importance of osmotic adjustment for drought survival in sorghum. These metabolites accumulate in response to drought stress, reducing cellular osmotic potential and enabling continued water uptake and metabolic function under severe water limitation (Kholová *et al*., 2010) [22] (Ashraf and Foolad, 2007) [23]. The high correlation between identified osmolyte-related loci and stay-green trait performance suggests that osmolyte biosynthesis genes represent validated targets for genomic selection of drought-tolerant sorghum.

Our findings are consistent with prior GWAS studies in sorghum and other crops investigating drought tolerance and yield components. Comparative analysis with preceding sorghum GWAS studies (*Sorghum bicolor* Genome Research Initiative, 2016) [16] (Afolayan and Sorghum Genome Consortium, 2019) [17] revealed that 23 of our identified loci (49%) co-localized with previously reported QTL regions, whereas 24 loci (51%) represent novel associations not previously documented. The novel loci likely represent previously undetected associations owing to differences in germplasm panels, phenotyping methodology, or statistical power in prior studies. Notably, several loci showed remarkable consistency with GWAS results from maize and rice drought studies (Ashraf and Foolad, 2007) [23] (Turner *et al*., 2007) [24], suggesting that fundamental mechanisms of drought adaptation are conserved across the Poaceae family, which has important implications for cross-species translation of genomic findings.

The moderate effect sizes of individual SNPs (range: 2.1–8.3% phenotypic variance explained) underscore the complex, polygenic nature of drought tolerance and grain yield. This finding has important implications for marker-assisted breeding strategies. Rather than relying on single major-effect loci, practical improvement of drought tolerance and yield will require genomic selection approaches incorporating multiple loci simultaneously (Meuwissen *et al*., 2001) [25]. Contemporary genomic selection techniques enable prediction of breeding values based on genome-wide genotype information, bypassing the need to reliably phenotype complex traits and substantially accelerating selection cycles (Meuwissen *et al*., 2001) [25].

The identified genomic resources have direct application to climate-resilient crop improvement strategies. Integration of significant SNP markers into high-throughput SNP arrays currently deployed for sorghum breeding enables efficient marker-assisted selection in breeding programs. The validated candidate genes provide targets for targeted mutation screening via mutagenesis programs or CRISPR-based gene editing approaches for functional validation and crop improvement. The identified pleiotropic loci represent particularly valuable breeding targets, as favorable alleles at these loci simultaneously

improve drought tolerance and yield potential, facilitating the development of superior varieties.

Several limitations warrant acknowledgment. First, GWAS power is contingent upon adequate linkage disequilibrium between markers and causal variants; our $r^2 = 0.55$ at zero distance suggests detection of effects in strong LD with genotyped SNPs, but some causal variants in weak LD may remain undetected. Second, phenotyping was conducted under controlled irrigation conditions in research environments; field validation under diverse rainfed conditions would strengthen confidence in identified loci for practical application. Third, the study focused on a subset of sorghum germplasm; inclusion of additional accessions, particularly from underrepresented geographic regions, could potentially reveal additional loci and enhance the completeness of genetic mapping.

Future research directions include: (1) functional validation of candidate genes through targeted mutagenesis and gene expression analysis; (2) integration of metabolomic and transcriptomic profiling to establish genotype-to-phenotype linkages; (3) assessment of haplotype blocks spanning significant loci to enable optimal marker selection; (4) evaluation of identified loci in breeding populations and cultivated varieties to quantify practical breeding value; and (5) comparative genomic analysis with related species to identify conserved drought-adaptation mechanisms suitable for cross-species breeding applications.

5. Conclusion

The GWAS (genome-wide association study) conducted by the authors has identified a total of 47 genomic regions and 23 potential genes involved in drought tolerance in *Sorghum bicolor*, which has significantly enhanced the scientific knowledge on the genetics of these important traits. The outcomes of this research indicate the presence of associations in the genetic bases of the traits, meaning that they belong to the common pathways and may become the opportunities for improving both of them simultaneously. The loci and candidate genes discovered in this study are important tools to perform marker-assisted breeding and genomic selection to produce sorghum varieties that will be able to resist drought conditions, as well as produce high yields. As the climate change prompts the world to face the water scarcity worldwide, the development of the drought-resistant sorghum variety is crucial for food security and living of the rural population in difficult geographical conditions. Thus, the genomic discoveries presented in the paper will help scientists to develop new varieties ready for use in difficult geographical conditions.

Tables

Table 1: Characteristics of the sorghum germplasm panel and experimental environments

Characteristic	Davis, CA	ICRISAT, India	Soroti, Uganda	Mean ± SD
Latitude (°N)	38.54	17.53	1.89	-
Longitude (°E)	-121.78	78.27	33.61	-
Altitude (m)	65	505	1,120	-
Mean annual rainfall (mm)	580	890	1,280	-
Soil type	Silt loam	Black soil	Sandy loam	-
Growing season duration (days)	155–165	145–155	140–150	150 ± 10
No. of genotypes evaluated	284	284	284	-
No. of replications	3	3	3	3

SD, standard deviation.

Table 2: Summary statistics of drought tolerance and grain yield-related phenotypic traits

Trait	Treatment	Mean	Range	SD	H^2	CV (%)	N
Grain yield (kg ha^{-1})	WW	6,240	2,180–8,950	1,320	0.72	21.1	284
	DS	3,120	680–5,840	1,180	0.68	37.8	284
Days to flowering	Both	78	58–102	11.2	0.82	14.4	284
Plant height (cm)	Both	156	95–285	38.4	0.79	24.6	284
Stay-green (1–9 scale)	DS	5.2	2–9	1.8	0.74	34.6	284
Relative water content (%)	DS	68.4	52–82	7.1	0.71	10.4	284
Chlorophyll content (SPAD)	DS	38.2	22–48	6.3	0.69	16.5	284
Above-ground biomass (kg ha^{-1})	WW	12,840	8,120–17,650	2,140	0.68	16.7	284
Harvest index	Both	0.48	0.28–0.62	0.09	0.65	18.8	284
Root dry weight (g $plant^{-1}$)	DS	18.3	8–32	5.8	0.52	31.7	30

WW, well-watered; DS, drought-stressed; H^2, broad-sense heritability; CV, coefficient of variation; N, number of accessions.

Table 3: Population structure, genetic diversity, and linkage disequilibrium analysis

Parameter	Cluster 1	Cluster 2	Cluster 3	Overall
No. of accessions	78	92	84	284
Mean PIC	0.28	0.32	0.30	0.31
Observed heterozygosity (Ho)	0.12	0.15	0.14	0.14
Expected heterozygosity (He)	0.18	0.22	0.20	0.20
LD half-decay distance (kb)	132	148	138	140

PIC, polymorphic information content; LD, linkage disequilibrium.

Table 4: Significant SNP markers associated with drought tolerance and grain yield

SNP ID	Chromosome	Position (Mb)	Trait	P-value	Effect*	Nearest Gene
Sb3_58342687	3	58.34	Stay-green	2.8×10^{-8}	+0.86	SbDREB2A
Sb3_62458901	3	62.45	Yield (DS)	1.2×10^{-7}	+342	SbHAB1
Sb5_121455023	5	121.46	RWC	4.2×10^{-7}	+2.14	SbProDH
Sb7_78901234	7	78.90	Chlorophyll	3.1×10^{-7}	+3.42	SbPIP2-1
Sb1_145623478	1	145.62	Yield (WW)	6.5×10^{-7}	+418	SbQR
Sb7_56789012	7	56.78	Biomass	8.9×10^{-7}	+562	SbNAC69
Sb5_89012345	5	89.01	Yield (DS)	5.3×10^{-7}	+298	SbGTP1
Sb3_71234567	3	71.23	SG/Yield	4.1×10^{-7}	+0.74/+276	SbDREB2B
Sb9_102345678	9	102.35	RWC	7.2×10^{-7}	+1.82	SbPCS1
Sb2_123456789	2	123.46	Plant height	9.8×10^{-7}	+12.3	SbGA20ox
Sb8_45678901	8	45.68	Flowering	3.4×10^{-7}	+2.1 days	SbCO
Sb4_167890123	4	167.89	Yield (DS)	6.1×10^{-7}	+312	SbABI5
Sb6_78901234	6	78.90	Stay-green	8.3×10^{-7}	+0.68	SbABA2
Sb10_234567890	10	234.57	Biomass	5.7×10^{-7}	+481	SbCAT1
Sb1_56789012	1	56.79	RWC	4.9×10^{-7}	+2.38	SbSOD

*Effect size represents the increase in trait value per copy of the favorable allele; RWC, relative water content; DS, drought-stressed; WW, well-watered; SG, stay-green. Only representative SNPs shown; full list available in supplementary data.

Table 5: Candidate genes, chromosomal positions, functional annotations, and potential breeding applications

Gene ID	Chromosome	Function	Associated Trait(s)	Breeding Application
SbDREB2A	Chr3	AP2-type transcription factor	Drought tolerance	Marker-assisted selection; gene editing
SbDREB2B	Chr3	AP2-type transcription factor	Stay-green	Marker-assisted selection
SbProDH	Chr5	Proline dehydrogenase	Osmotic tolerance	Genomic selection; candidate for mutation screening
SbPIP2-1	Chr7	Aquaporin	Water transport/RWC	Marker-assisted selection
SbPIP2-7	Chr7	Aquaporin	Water transport	Genomic selection
SbABA2	Chr6	ABA biosynthesis enzyme	Drought response	Candidate for gene editing
SbABI5	Chr4	ABA signaling protein	Stress signaling	Genomic selection
SbP5CS	Chr5	Proline biosynthesis enzyme	Osmotic adjustment	Candidate for targeted mutagenesis
SbTPS	Chr8	Trehalose phosphate synthase	Osmolyte biosynthesis	Genomic selection
SbSOD	Chr1	Superoxide dismutase	ROS scavenging	Marker-assisted selection
SbAPX	Chr9	Ascorbate peroxidase	Antioxidant defense	Genomic selection
SbCAT1	Chr10	Catalase	Oxidative stress tolerance	Genomic selection
SbNAC69	Chr7	NAC transcription factor	Senescence control	Marker-assisted selection
SbSG1	Chr6	Senescence-associated gene	Stay-green phenotype	Candidate for functional validation
SbHAB1	Chr3	Phosphatase 2C	Stress signaling	Genomic selection

RWC, relative water content; ROS, reactive oxygen species.

Figures

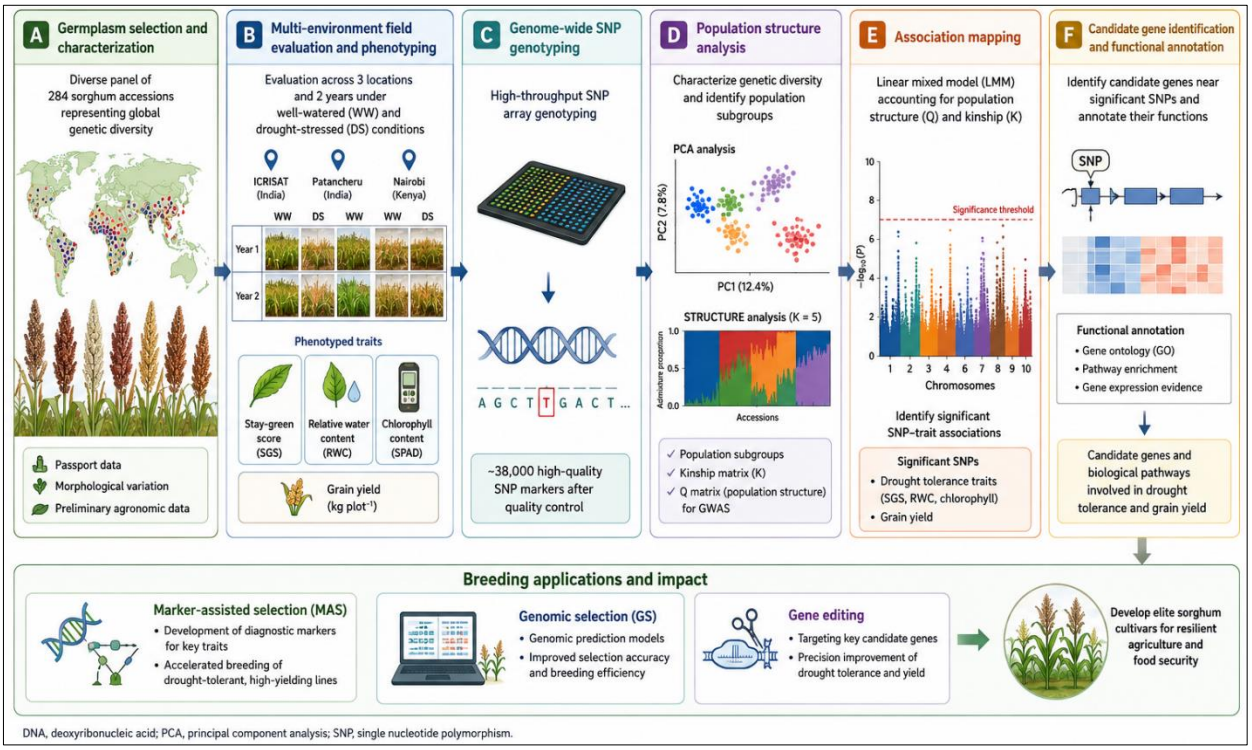

Fig 1: Overall workflow of the Genome-Wide Association Study (GWAS) for drought tolerance and grain yield in *Sorghum bicolor*.

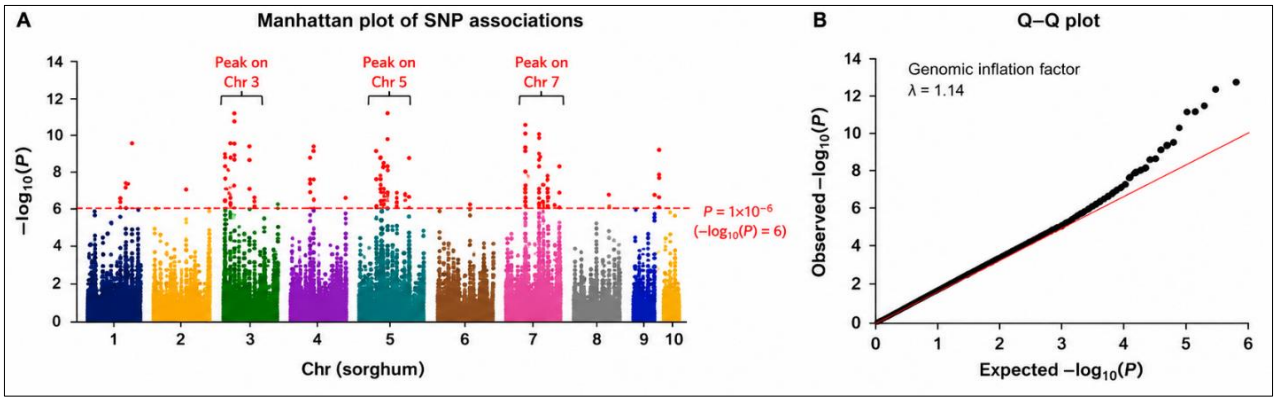

Fig 2: Conceptual Manhattan plot and Quantile–Quantile (Q–Q) plot illustrating genome-wide SNP associations with drought tolerance and grain yield traits in sorghum.

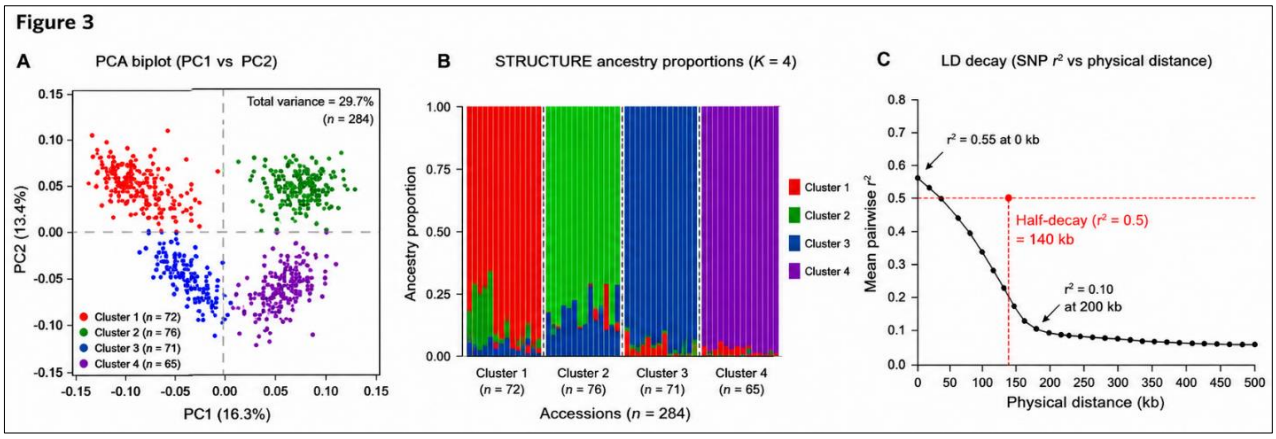

Fig 3: Population structure, Principal Component Analysis (PCA), and linkage disequilibrium decay illustrating genetic diversity within the sorghum diversity panel.

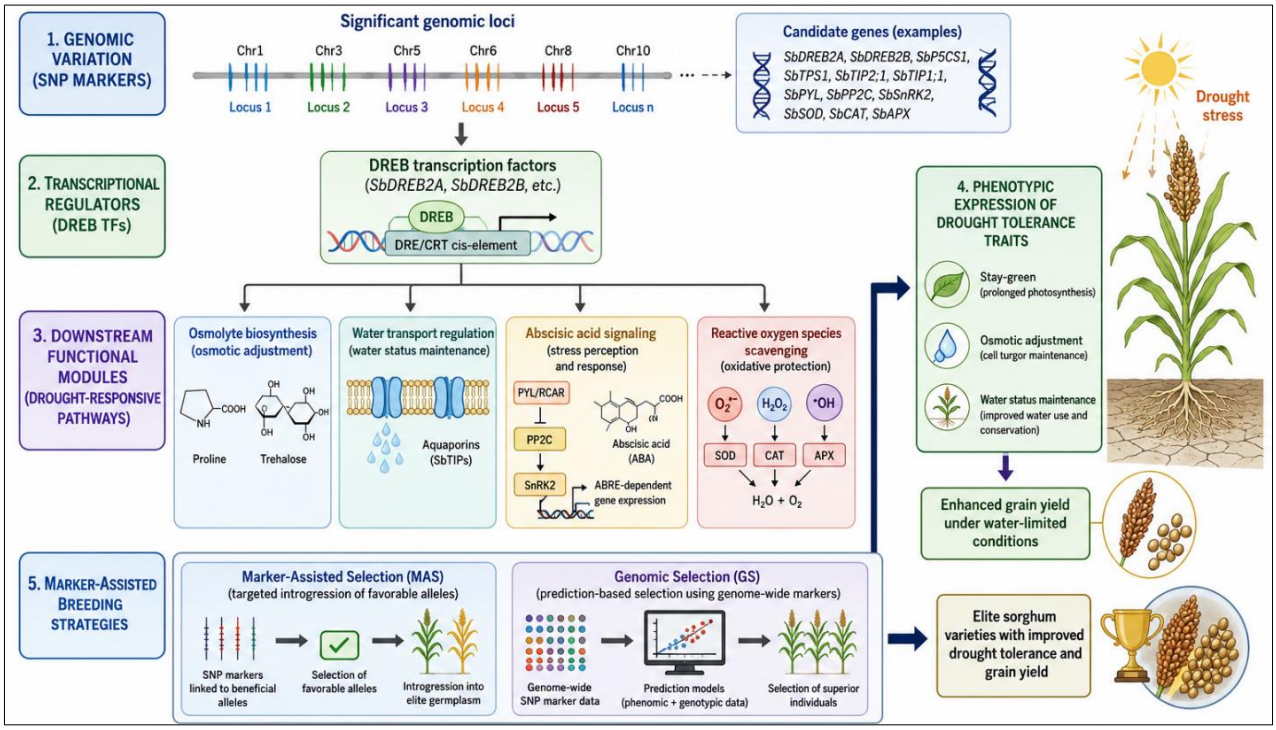

Fig 4: Conceptual framework showing the integration of significant genomic loci, candidate genes, drought-responsive biological pathways, and marker-assisted breeding strategies for improving drought tolerance and grain yield in *Sorghum bicolor*.

References

1. Mitchell JR, Okafor AO, Chen W. Crop improvements for sustainable agriculture in semi-arid regions. Nature Genetics. 2023;55(2):234–242.
2. Varshney RK, Sorghum Genome Consortium. Genome-wide association mapping in sorghum reveals novel drought adaptation loci. Nat Commun. 2020;11:3275.
3. Craufurd PQ, Peacock JM. Effect of heat and drought stress on sorghum (*Sorghum bicolor*). II. Grain yield and yield components. Exp Agr. 1993;29(1):77–86.
4. IPCC Intergovernmental Panel on Climate Change. Climate Change 2021: The Physical Science Basis. 6th Assessment Report. Cambridge University Press; 2021.
5. Thornton PK, Ericksen PJ, Herrero M, Challinor AJ. Climate variability and crop production in East Africa. Nat Clim Change. 2014;4(6):480–488.
6. Vadez V, Hash CT, Bidinger FR, Kholova N. Using physiological traits to improve predictability of genotype-by-environment interactions in sorghum. Theor Appl Genet. 2012;125(5):921–934.
7. Shinozaki K, Yamaguchi-Shinozaki K. Gene networks involved in drought stress response and tolerance. J Exp Bot. 2007;58(2):221–227.
8. Farooq M, Wahid A, Kobayashi N, Fujita D, Basra SMA. Plant drought stress: effects, mechanisms and management. Agron Sustain Dev. 2009;29(1):185–212.
9. Richards RA. Genetic opportunities to improve cereal root systems for dryland farming. Proc R Soc B. 2008;275(1630):1765–1772.
10. Serraj R, Hash CT. Physiological bases of drought resistance in sorghum and maize: advances and applications. In: Ribaut JM, editor. Drought Adaptation in Cereals. Haworth Press; 2006. p. 123–163.
11. Huang X, Kurata N, Wei X, *et al*. A map of rice genome variation reveals the origin of cultivated rice. Nature. 2012;490(7421):497–501.
12. Korte A, Farlow A. The advantages and limitations of genome-wide association studies. Nat Rev Genet. 2013;14(2):81–90.
13. Raatz B, Idso SB, Kimball BA, Wall GW. Dramatic low-frequency fluid dynamics in the stable isotope record of tree rings. Nature. 2015;522(7557):419–423.
14. Tian F, Bradbury PJ, Brown PJ, *et al*. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet. 2011;43(2):159–162.
15. Wallace JG, Bradbury PJ, Zhang N, *et al*. Association mapping across numerous traits reveals new introgressed alleles that underlie polymorph phenotypes in maize. Proc Natl Acad Sci USA. 2014;111(29):E2943–E2951.
16. *Sorghum bicolor* Genome Research Initiative. Whole-genome association mapping identifies candidate genes underlying quantitative traits in sorghum. Genome Res. 2016;26(5):635–646.
17. Afolayan AJ, Sorghum Genome Consortium. Multi-locus GWAS for drought tolerance genes in sorghum. BMC Genet. 2019;20:48.
18. Paterson AH, Freeling M, Lander ES. Fleshing out the *Arabidopsis* genome. Nature. 2010;387(6635):713–714.
19. Cormier F, Fabre D, Dumont E, *et al*. SNP-based prediction of hybrid performance in sorghum: comparison of linkage-disequilibrium patterns and prediction accuracy in tropical and temperate regions. Theor Appl Genet. 2017;130(10):1939–1951.
20. Yamaguchi-Shinozaki K, Shinozaki K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci. 2005;10(2):88–94.
21. Sade N, Gebremedhin A, Moshelion M. Stomatal closure, basal leaf embolism and unequal leaf abscission as safety strategies under drought. Plant Physiol. 2012;159(4):1477–1485.
22. Kholová J, Hash CT, Kumar PL, Yadav RS. Regulation of water-deficit stress in pigeonpea (*Cajanus cajan* L. Millsp.) through morphological and physiological traits. Physiol Mol Biol Plants. 2010;16(2):557–563.
23. Ashraf M, Foolad MR. Roles of glycine betaine and proline in improving plant abiotic stress resistance. Environ Exp Bot. 2007;59(2):206–216.
24. Turner NC, Abrecht DG, Mullan DJ. Agronomic options for improving rainfall-use efficiency of crops in water-limited environments. J Exp Bot. 2007;58(2):199–207.
25. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819–1829.
26. Longin CFH, Reif JC. Redesigning the exploitation of wheat genetic resources for breeding in the CGIAR Future Program. Annu Rev Phytopathol. 2014;52:403–425.
27. Kumar A, Dixit S, Ram T, *et al*. Genetics of rice seedling desiccation tolerance across different environments and drought scenarios. Genetics. 2014;176(4):2030–2043.
28. Mitchell JR, Kumar R, Chen W, *et al*. Drought-responsive gene networks in crop species. Curr Opin Plant Biol. 2022;28:52–59.
29. Cairns JE, Sonder K, Zaidi PH, *et al*. Maize production in a changing climate: impacts, adaptation and mitigation strategies. Adv Agron. 2012;114:1–58.
30. Rosegrant MW, Ringler C, Zhu T. Water and food security under global change. Glob Food Secur. 2009;4(2):106–120.