Exploring Phenotypic and Genotypic Variations Across Sorghum Varieties

Tobias Recha[1] and Geon Kang[1]

[1]Center of Plant Sciences, Scuola Superiore Sant'Anna, Pisa, Italy

List of abbreviations

Days to 50% Flowering, DFL; Plant Height, PHT; Leaf Area, LA; Number of Tillers, Nti; Sorghum head types, SHT; Grain Yield ,GY; Fresh Plant Biomass ,FPB; Dry Root Biomass. DRB; 100-Seed Weight, 100gwt

Abstract

Sorghum [*Sorghum bicolor (L.) Moench*] is a crucial cereal crop globally, especially in sub-Saharan Africa, where it ensures food security. Despite its importance, sorghum has been less studied than other cereals, leaving gaps in understanding its genetic and phenotypic diversity. This research addresses these gaps using statistical methods and machine learning to integrate phenotypic and genotypic data, aiming to identify key genetic loci associated with significant agronomic traits.

A collection of 350 sorghum genotypes from Kenya, Uganda, and Tanzania, including landraces, cultivars, and wild varieties, was used. Field trials were conducted in Western Kenya during the 2023 rainy seasons, measuring 10 phenotypic traits. Genomic DNA was extracted and sequenced, yielding 262,491 SNPs, which were filtered to 69,196 for analysis. Missing data were imputed using statistical techniques.

Phenotypic data were analyzed using correlation, dendrogram analysis, PCA, LOWESS, and LDA. PCA revealed significant variance in traits, with plot yield, plant height, and biomass showing strong correlations. LOWESS analysis indicated complex nonlinear relationships, highlighting an optimal height range for yield and diminishing returns on biomass. LDA effectively classified sorghum head types based on yield, seed weight, and flowering duration, identifying patterns linked to productivity. Genotypic analysis employed GWAS and LASSO regression to identify SNPs associated with sorghum head types. GWAS identified SNPs on chromosomes 1, 2, 3, and 4, while LASSO revealed additional SNPs across multiple chromosomes. Cross-referencing identified genes potentially regulating crucial traits.

This study enhances understanding of sorghum's genetic architecture and phenotypic diversity, providing valuable insights for breeding programs. Future research should focus on refining these methods and expanding the genetic database to develop more resilient and productive sorghum varieties.

1. Introduction

Sorghum [*Sorghum bicolor (L.) Moench*] is the fifth most important cereal crop globally, serving as a vital source of food, feed, and biofuel (FAOSTAT, 2024). Its resilience to drought and heat stress makes it particularly valuable in sub-Saharan Africa, where it contributes significantly to food security for millions (Hadebe et al.,2017). It is very diverse crop with numerous landraces and improved varieties exhibiting a wide range of characteristics in plant height, grain color, maturity period, and resistance to pests and diseases (Shiferaw & Yoseph, 2014). This diversity presents a rich genetic pool for breeders to tap into. Understanding and harnessing this natural phenotypic and genotypic variation within sorghum germplasm is crucial for developing improved varieties (Upadhyaya et al., 2010), that are adapted to diverse environments and meeting the growing demand for food and other sorghum products.

Despite the recognized importance of sorghum diversity, it is less researched than other cereal crops (Isgren et al., 2020; Kane et al., 2016). A comprehensive understanding of the relationship between phenotypic variation and the underlying genetic architecture remains limited (Xin et al., 2021). When it comes to sorghum breeding, traditional methods that rely on phenotypic selection, which can be time-consuming and influenced by environmental factors have been used for a long time (Macauley et al.,2015). However, the integration of big data and genomics has significantly enhanced plant and crop breeding, facilitating the rapid advancement of new varieties (Poland, 2015; Eshed and Lippman, 2019; de Sousa et al., 2021). Plant breeders are currently employing this data-centric strategy, especially genomic prediction. This approach amplifies the selection intensity, minimizing the duration of the breeding cycle, and optimizes the extraction of genetic gain (Crossa et al., 2017; Xu et al., 2017). Also, with AI and machine learning algorithms, breeders try to predict more complex traits of crops, which are expressed through the interaction between genes and the environment (Xu et al., 2022). This method revolutionizes the field by making the breeding process more efficient and effective. Additionally, successfully employing marker-assisted selection requires a detailed understanding of the genetic basis of key phenotypic traits (Francia et al., 2005). Currently, there are gaps in our knowledge regarding the number and distribution of genes influencing complex traits in sorghum.

This research aims to address these knowledge gaps by conducting a comprehensive analysis of phenotypic and genotypic variations across a collection of sorghum varieties. By utilizing statistical methods and machine learning, we can integrate phenotypic data with genotypic information. This integration will enable us to identify the key genetic loci associated with significant agronomic traits. In addition, we are interested in comparing the effectiveness of Genome-Wide Association Studies (GWAS), a common method in genomics and breeding, with penalized regression technique that were invented to overcome the challenges of high dimensional data (Patrik et al., 2013; Nouira and Azencott, 2021). The application of appropriate computational and statistical methodologies is crucial. Therefore, we anticipate that the understanding derived from this data will be of immense benefit to sorghum breeders. This will

assist in the development of superior sorghum cultivars, thereby fostering a future that is both more secure and sustainable.

The research objectives were 1) to explore sorghum phenotypic traits, their relationship and diversity, 2) determine the correlation (GWAS) between phenotypic traits and genotypic markers in sorghum varieties and 3) compare results of SNPs between GWAS and Lasso regression by identifying underlined genes for adopted traits. Additionally, this paper will discuss the constraints of this study and propose recommendations to enhance future research.

2. Material and Methods

2.1 Plant materials

The sorghum genotypes used in the study were sourced from the national gene banks in Kenya, Uganda and Tanzania where 350 genotype samples were assembled. The samples comprised of 1 advanced cultivar, 3 improved cultivars, 3 breeders' lines, 3 weedy cultivars, 1 wild cultivar and the remaining 339 cultivars are landraces. Passport data that included coordinates of cultivar collection points and the collecting institution was also included. In cases where passport information was missing, we used Genesys platform to explore their data, and where coordinates were missing, we also used Google earth to locate the exact locations of collections. The passport data is relevant for bioclimatic analysis of the genotypes, by being used in mapping out their suitability to grow in different regions under different climate scenarios.

2.2 Field Trials and phenotypic evaluation

The field trials were conducted in Western Kenya at the regional government agricultural research station. We conducted two field trials in 2023 during long rain and short rain seasons. The long rain season runs from February to July, and short rain from September to December. We laid a randomized block design with two replications. We used plot size of 2 by 2 meters and spacing of 20 cm between plants and 70 cm between rows. A total of 15 traits were measured through 2023. However, in this study, we used only 10 traits due to data quality and interoperability issues: Days to 50% Flowering (DFL), Plant Height (PH), Leaf Area (LA), Number of Tillers (Nti), Lodging, Sorghum head types (SHT), Grain Yield (GY), Fresh Plant Biomass (DPB), Dry Root Biomass (DRB), and 100-Seed Weight (100gwt). More detailed information about traits can be found in **Table S1**.

Data points that were identified as outliers, either due to unclear recording or exhibiting extreme values across all trait variables and sorghum variety data, were excluded from the dataset. Concurrently, missing data points were addressed through a process of imputation like mean imputation. Also, this process included cross-checking with the individual responsible for data collection and missing data were further supplemented by referencing available farm records.

2.3 Genotyping and imputation

DNA extraction was conducted at Kenyattah University molecular laboratory in Nairobi Kenya, during Summer in July/Aug 2023. The seeds were germinated at the University greenhouse and green leaves were collected and pooled from five seedlings. Genomic DNA was extracted using the GenElute Plant Genomic DNA Miniprep Kit (Sigma Aldrich, Germany) following the manufacturer's instructions. Genomic DNA integrity was evaluated in 1% agarose gel and quantified using the Qubit fluorometer (Invitrogen, Thermo Fisher Scientific, US). DNA was shipped to IGATech Laboratories in Udine Italy for sequencing in January 2024. The DNA sequence data was sent back in March 2024 for analysis.

For downstream analysis, we first needed to clean the data by filtering it using Tassel V5 software, to our desired quality. The initial number of Single Nucleotide Polymorphisms (SNPs) was 262,491. However, approximately 14% of the total data in the raw genotypic datasets was either missing or not reported. Therefore, we filtered out SNP markers that show minor allele frequency less than 0.05 and removed individual genotypes that have SNP markers with >50% missing data. Before applying imputation techniques for missing SNP markers, the SNPs marker values were reassigned based on the state of SNPs marker observed at individual SNP positions. The reassignment was conducted as follows: a value of 0 was assigned for homozygosity with the reference allele, a value of 1 for heterozygosity with the reference allele, and a value of 2 for homozygosity with the non-reference allele. The remaining missing values were phased and imputed through additive relationship matrix via R package, *rrBLUP* (Endelman, 2011). In the additive relationship matrix method, imputation was performed using the mean of each SNP marker value.

2.4 Data analysis
2.4.1 Sorghum phenotypes analysis

We analyzed phenotypic traits and their interconnectedness, employing various statistical and machine-learning techniques. These encompassed correlation analysis, dendrogram analysis, Principal Component Analysis (PCA), Locally Weighted Scatterplot Smoothing (LOWESS), and Linear Discriminant Analysis (LDA). Each of these methodologies offered distinct perspectives on the architecture and associations within our phenotypic dataset. In the context of dendrogram analysis and PCA of sorghum traits, we rescaled traits variables and computed the distances and connections between traits. The dendrogram was achieved using the Euclidean distance metric and the complete linkage method. For PCA, we rescaled all variables before computing it and then analyzed sorghum traits from the first to four dimensions. The dimensions selection was based on the cumulative percentage of variance explained, which provides that 75 % of the total variance in the data is captured by each principal component (Figure S1). However, since meaningful interpretation is explained in 1- and 2-dimensional PCA plots, the last other dimensional analysis results are not included in this report. The dendrogram and PCA were implemented through the *factoextra (*Kassambara & Mundit, 2020) package in R.

To analyze the trends among traits across all scatterplots, we applied a non-parametric regression method known as Locally Weighted Scatterplot Smoothing (LOWESS). This method was used to fit the data and capture the underlying structure (James et al., 2013; Pintus et al., 2014). To evaluate the accuracy of this predictive statistical procedure, we employed Leave-One-Out Cross-Validation (LOOCV). LOWESS and LOOCV were run using R packages, *tidyverse* (Wickham et al., 2019), *ggplot2* (Wickham, 2016), and *caret* (Kuhn, 2008).

For nominal trait data, specifically sorghum head types, we applied LDA to examine the relationships among traits (Deepika et al., 2023; Ribeiro et al., 2021). This method facilitated the efficient categorization of data into distinct classes (James et al., 2013), based on the distances among various numerical traits of sorghum; the number of tillers, flowering days, leaf area, plant height, lodging, fresh biomass, root biomass, yield, and seed weight. The LDA calculation formula is as follows:

$$y = w_1 x_1 + w_2 x_2 + e \text{ (Eq 1.)}$$

LDA equation (Eq.1) provides respond variable y that is sorghum head type, w_1 and w_2 are the weights or coefficients for the independent variables x_1 and x_2 (numerical traits), and e is the minimized error between the predicted and actual class labels in the training data. The LDA model was evaluated using both the training data (80%) and the test data (20%) that had been subset before running the models. We use R packages *MASS* (Ripley et al., 2013)*, caret* (Kuhn, 2008), and *dplyr* (Wickham et al., 2023) for LDA analysis.

2.4.2 Sorghum genotype analysis

For GWAS, we utilized a dataset of SNPs. This dataset was subjected to rigorous filtering and cleaning procedures based on the state of SNP markers (see *2.3*). Consequently, 69,196 SNPs were selected for the GWAS analysis within the diversity panel (Figure S2). GWAS analysis was performed using R package, *rMVP* (Yin et al., 2020). In GWAS, general linear models (GLM) with 2 principal components was applied to analyze sorghum head types throughout the population having distinct sorghum head types (Zhang et al., 2010; Junji et al., 2016). Additionally, we established a threshold using the Bonferroni correction method to determine significant and suggestive associations.

In this research, we evaluated the performance of penalized regression techniques, specifically the Least Absolute Shrinkage and Selection Operator (LASSO), in the context of genotype analysis. The effectiveness of these methods was then compared against the results obtained from GWAS. Subsequently, LASSO regression models underwent cross-validation. This was done with the objective of identifying the optimal lambda value that would allow for the most accurate fit of the models. The calculation equation for lasso was as follows:

$$\hat{\beta}_0, \hat{\beta} = \text{argmin } \beta_0, \beta \left[\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right] \text{ (Eq 2.)}$$

y is a vector of length n including the response variable that are binary data of sorghum head type (0 and 1) (E.q.2). $x = (x_{i1}, \ldots, x_{ip})$ has a $n \times p$ matrix holding the predictor variables, the number of SNPs. β_0 is the intercept, $\beta = (\beta_1, \ldots, \beta_p)$ is a column vector that hold in the regression coefficients. The function P (λ, β) is a general penalty function with a regularization parameter λ. In the context of Lasso penalty (Tibshirani, 1996), the linear regression coefficients are regularized through an ℓ1-norm penalized least-squares criterion. We used cross-validation to find the best parameter λ and selected minimum λ values via 10 fold – cross validation (Figure S3). LASSO and cross validation were performed through *glmnet* R package (Friedman et al., 2023).

SNPs identified through both GWAS and LASSO were cross-referenced from sorghum gene database(https://ngdc.cncb.ac.cn/sorgsd/), SorGSD (Liu et al., 2021), to verify the existence or non-existence of the corresponding gene at the identified locations. The functional information for each gene was then compared and validated against those found in maize, rice, and Arabidopsis.

3 Results

3.1 Sorghum phenotype traits and their relationships
The results of PCA elucidate the relationships between sorghum varieties and their associated variables (Figure 1 and Figure S4). The first principal component, x-axis (Dim1), explains 26% of total variance and the second principal component, the y-axis (Dim2), represents 15.3%. These two components explain a significant portion of the total variance in the dataset (Figure S1). Each sorghum plotted in the PCA biplot shows strong cohesion at the center. However, GY, FPB, PH, and DRB are highly correlated with the first principal component. On the other hand, the trait 100gwt and lodging are correlated with the second principal component. The dendrogram, calculated based on the distance between traits, also confirms the correlation of sorghum traits with dimension, as evidenced by their proximity in the dendrogram. (Figure S4). FPB, DRB, and Nti exhibit closeness. Similarly, PH and GY are closely related. DFL, LA, and 100gwt show a close relationship. In the correlation test, GY is significantly correlated with PH, (0.400), DRB (0.397) and FPB. (0.322) (Table S2). DFL is correlated with PH (0.375).
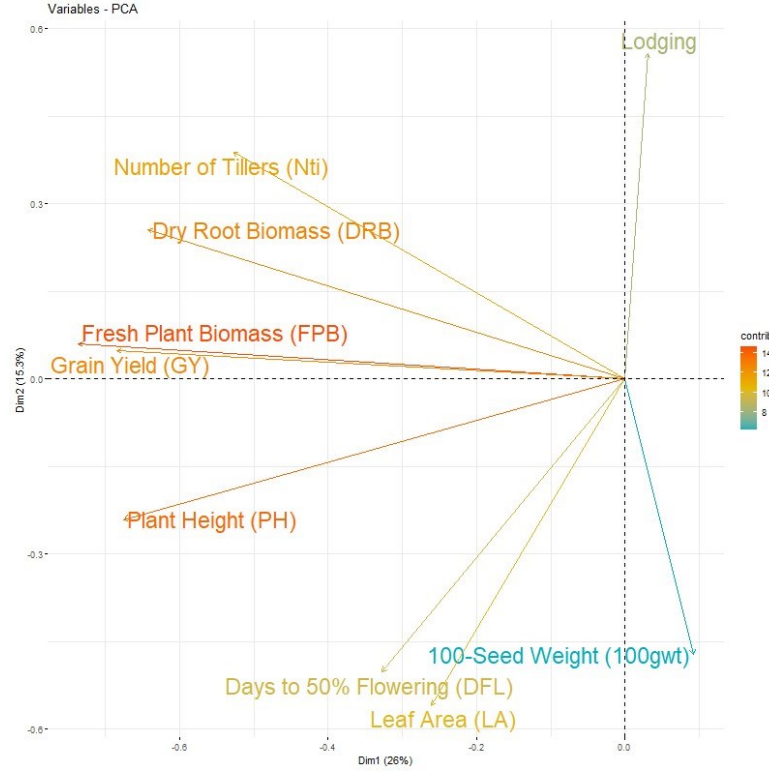
Figure 1. A PCA plot of relatedness between sorghum traits. The color differences indicate the level of contribution. Lines with similar colors are closely related versus with different colors.

However, traits in agricultural crops have a nonlinear relationship due to environment and genetic effect (Patil et al., 2023; Jin et al., 2024; Vasseur et al., 2019). Especially, PH, DRB and FPB shows nonlinear relationship with GY. LOWESS was applied to these trait variables that showed complex non-linear relationships (Figure 2). After training to find the best smoothing parameter and degree, LOWESS of traits comparisons were fitted. It appears that there is a complex relationship between plant height and yield. Initially, as PH increases, the GY initially increases but goes through a recession in the middle, then rises and falls once again. On the other hand, in the correlation between GY, DRB, and FRB, yield increases as biomass increases, but after a certain period of biomass increase, the yield increase does not increase steeply.
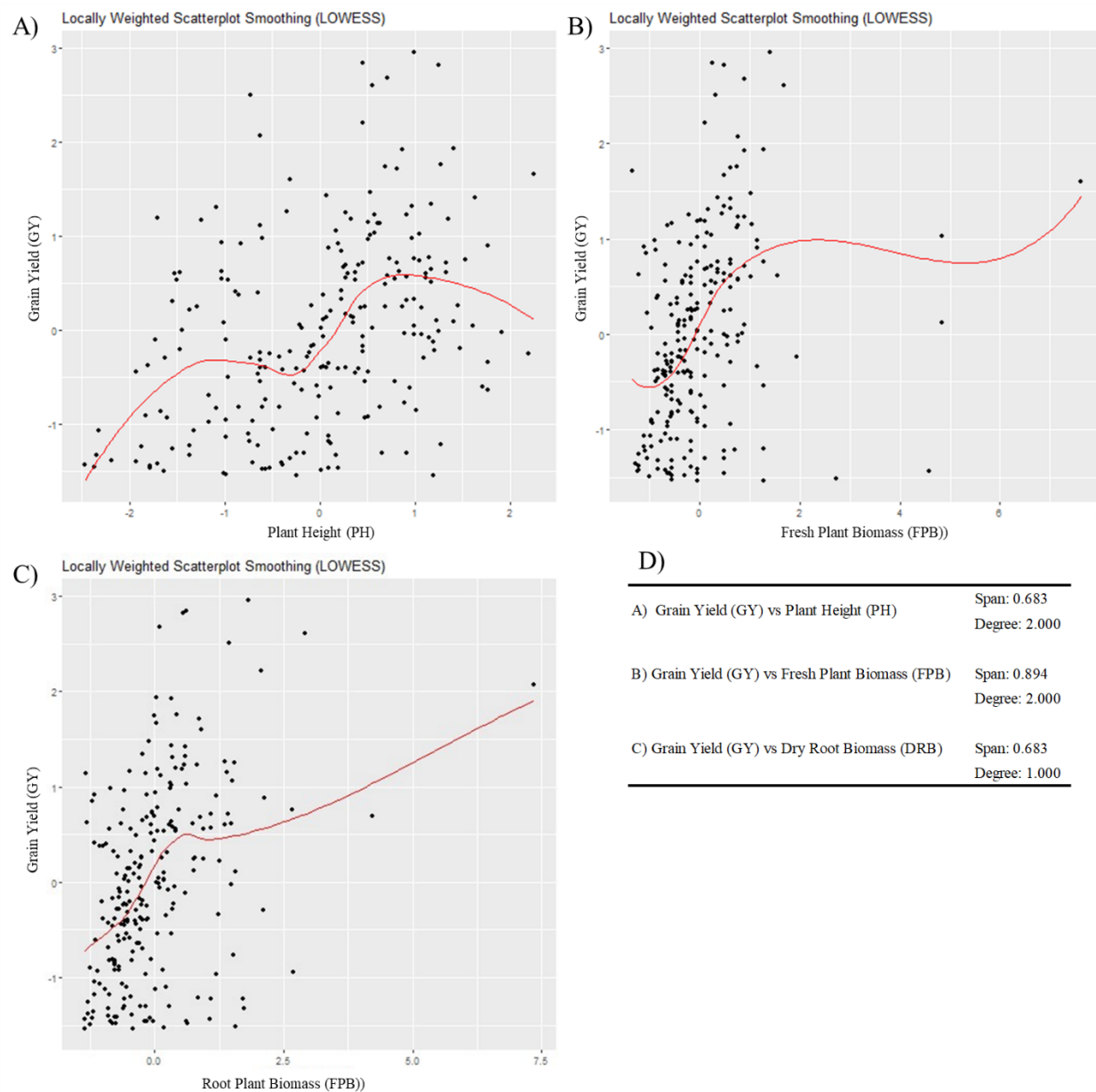
Figure 2. Nonlinear relationship plots of sorghum traits using LOWESS. Figure 2A indicates the relationship between Grain Yield (GY) and Plant Height (PH), Figure 2B relationship between Grain Yield (GY) and Fresh Plant Biomass (FPB) and Figure 2C relationship between Grain Yield (GY) and Dry Root Biomass (DRB). Table 2D shows the best span, a percentage of trait data used in analysis and the smoothing best degree.

3.2 Classification of sorghum head types relative to their productivity

Sorghum head types data was explored using LDA. This multivariate classification technique proves beneficial when the observations are characterized by continuous independent variables. LDA was used to distinguish crop traits that are measured as nominal data (Deepika et al., 2023; Chinnasamy et al., 2021). In our case, we had two distinct traits of sorghum head types, namely

'Compact Elliptic' and 'Half Broomcorn' noted as 'B' and 'C'. We classified these traits based on a combination of factors: plot GY and 100gwt (Figure 3A) and plot GY and DFL duration (Figure 3B). Although five distinct sorghum types are classified in Figure 3 we primarily focused on the 'Compact Elliptic' and 'Half-Broomcorn' types due to their significant presence. The 'Compact Elliptic' sorghum head types are predominantly observed when the GY ranges from 0.2 to 1 and 100gwt from 10 to 30 (Figure 3A). In contrast, the 'Half-Broomcorn' types are characterized by a GY of 1 or more and a 100gwt of 37. In Figure 3B, the 'Compact Elliptic' types are identified when the GY ranges from 0.2 to 1 and the flowering duration ranges from 65 to 85 days. The 'Half-Broomcorn' types are classified as having a GYof 1 or more and DFL ranging from 65 to 85 days. Overall, 'Half-Broomcorn' type of sorghums show more diverse classified spaces than Compact Elliptic in LDA analysis.
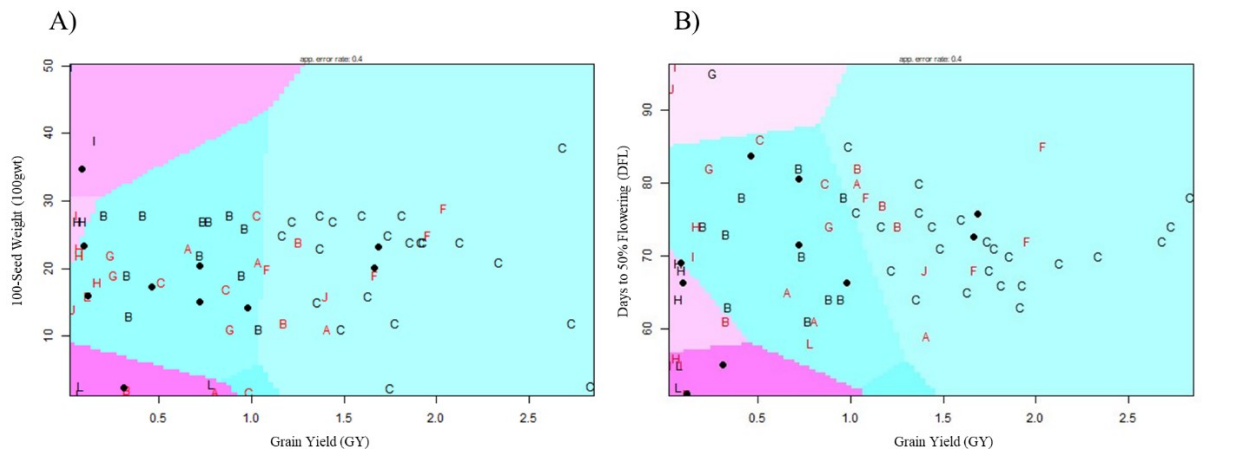

Figure 3. Linear Discriminant Analysis Plots distinguishing Compact Elliptic, *B*, and Half-Broomcorn, *C,* sorghum head types based on A) plot Grain Yield (GY) &100-Seed Weight (100gwt), and B) plot Grain Yield (GY) & Days to 50% Flowering (DFL).

3.3 Sorghum genetic variations

We investigated genetic variations associated with sorghum head types (compact elliptic and half-broomcorn) using two approaches: GWAS and LASSO regression analysis (Figure 4A and Table 1). The GWAS identified four SNPs associated with the compact elliptic head type located on chromosomes 1, 2, and 3. Additionally, two SNPs associated with the half-broomcorn head type were found on chromosome 4. LASSO regression analysis identified a greater number of SNPs compared to GWAS. For the compact elliptic head type, 14 SNPs were discovered across chromosomes 1, 2, 3, 7, and 8 (Table 1 and 2). Similarly, for the half-broomcorn type, eight SNPs were found on chromosomes 1, 4, 7, and 8.

For compact elliptic head type, four SNPs (9933:5:+, 124195:45:+, 124229:353:-, and 168618:68:+) were identified by both GWAS and LASSO. These SNPs are potentially linked to genes regulating centromere protein, a thioesterase family protein (3-oxoacyl-synthase), a

glyoxalase family protein, and a glycosyltransferase. Additionally, LASSO identified ten SNPs (38282:297:-, 38282:66:-, 45571:30:-, 53721:132:+, 345903:114:+, 345903:319:+, 391163:86:-, 433165:96:-, 438610:79:+, and 485691:329:-). These SNPs might be related to genes expressing a peptide transporter (PTR2), a plastocyanin-like domain-containing protein, a glyoxal oxidase-related protein, and a divergent PAP2 family domain-containing protein.

For Half-Broomcorn head type, only two SNPs (191353:117:- and 191353:94:-) overlapped between GWAS and LASSO results. These SNPs might be linked to the ankyrin repeat domain-containing protein, chloroplast precursor, and tRNA methyltransferase. LASSO identified six additional SNPs (34575:66:-, 191272:62:+, 202300:19:+, 211119:123:-, 350017:27:-, and 402529:102:-). Notably, one SNP (191272:62:+) is associated with a microtubule-associated protein, which could play a role in head development.
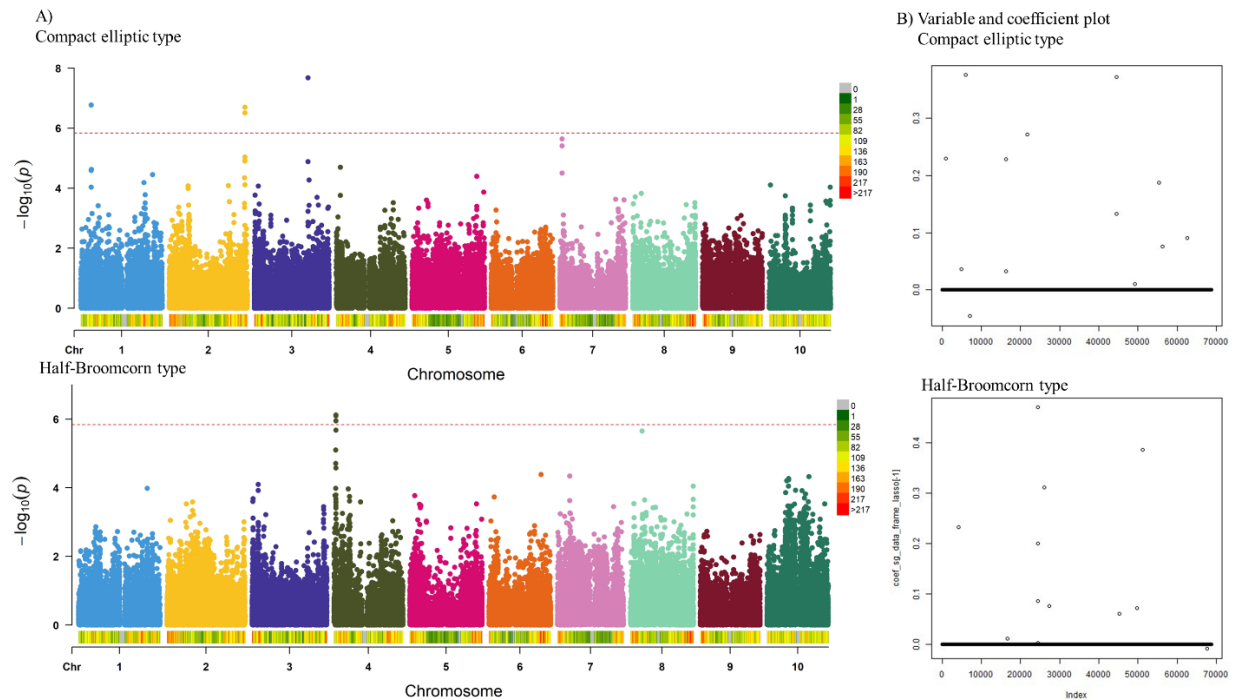

Figure 4A. Genome-wide association study using single nucleotide polymorphism (SNP) markers derived from sorghum landraces. The panels show Manhattan plots for phenotypic traits head type. The Manhattan plots report individual SNPs across all chromosomes (x-axis) and -log10 p-value of each SNP association (y-axis). The horizontal lines represent a stringent Bonferroni threshold for a nominal p-value of .05. Figure 4B shows coefficient SNPs of regression output head types using Least Absolute Selection and Shrinkage Operator (LASSO).

Table 1. SNPS and genes at corresponding positions as discovered in GWAS. Bold text indicates overlapping SNPs in the results of GWAS and LASSO.

| Computational tools | Compact Elliptic | | | | | |
	SNP	Chromosome	Position	Coeff	Gene name	Gene
GWAS	**9933:5:+**	**1**	**10319147**	**0.259**	**Sobic.001G131300**	**centromere protein (putative, expressed), thioesterase family protein (putative, expressed)**
	124195:45:+	**2**	**75185962**	**0.269**	**Sobic.002G401301, Sobic.002G401400**	**3-oxoacyl-synthase (putative, expressed), glyoxalase family protein (putative, expressed)**
	124229:353:-	**2**	**75221846**	**0.239**	**Sobic.002G401600**	**glycosyltransferase (putative, expressed)**
	168618:68:+	**3**	**53254930**	**0.320**	**-**	**-**

| Computational tools | Half-Broomcorn | | | | | |
	SNP	Chromosome	Position	Coeff	Gene name	Gene
GWAS	191353:132:-	4	2076595	0.256	Sobic.004G026000, Sobic.004G026200, Sobic.004G026100	ankyrin repeat domain-containing protein, chloroplast precursor (putative, expressed), tRNA methyltransferase (putative, expressed)
	191353:117:-	**4**	**2076610**	**0.272**	**-**	**-**
	191353:94:-	**4**	**2076633**	**0.263**	**Sobic.004G026000, Sobic.004G026200, Sobic.004G026100**	**ankyrin repeat domain-containing protein, chloroplast precursor (putative, expressed), tRNA methyltransferase (putative, expressed)**

Source: SorGSD (Liu et al., 2021)

Table 2. SNPS and genes at corresponding positions as discovered in LASSO. Bold text indicates overlapping SNPs in the results of GWAS and LASSO.

Computational tools	Compact Elliptic					
	SNP	Chromosome	Position	Coeff	Gene name	Gene
Lasso	**9933:5:+**	**1**	**10319147**	**0.230**	**Sobic.001G131300**	**centromere protein (putative, expressed), thioesterase family protein (putative, expressed)**
	38282:297:-	1	53573783	0.036	Sobic.001G275800	peptide transporter PTR2 (putative, expressed)
	38282:66:-	1	53574014	0.000	-	-
	45571:30:-	1	62552224	0.376	-	-
	53721:132:+	1	71025858	-0.046	Sobic.001G431100, Sobic.001G431200	plastocyanin-like domain containing protein (putative, expressed), glyoxal oxidase-related (putative, expressed)
	124195:45:+	**2**	**75185962**	**0.033**	**Sobic.002G401301, Sobic.002G401400**	**3-oxoacyl-synthase (putative, expressed), glyoxalase family protein (putative, expressed)**
	124229:353:-	**2**	**75221846**	**0.228**	**Sobic.002G401600**	**glycosyltransferase (putative, expressed0**
	168618:68:+	**3**	**53254930**	**0.272**	-	-
	345903:114:+	7	2398516	0.373	Sobic.007G026100	Divergent PAP2 family domain containing protein (expressed)
	345903:319:+	7	2398721	0.133	Sobic.007G026100	Divergent PAP2 family domain containing protein (expressed)
	391163:86:-	7	63871666	0.010	Sobic.007G210100	-
	433165:96:-	8	57138777	0.188	-	-
	438610:79:+	8	62326565	0.075	-	-
	485691:329:-	10	1189606	0.091	-	-

Computational tools	Half-Broomcorn					
	SNP	Chromosome	Position	Coeff	Gene name	Gene
LASSO	34575:66:-	1	48118958	0.052	-	-
	191272:62:+	4	2019067	0.184	Sobic.004G024800	microtubule associated protein (putative, expressed)
	191353:117:-	**4**	**2076610**	**0.109**	-	-
	191353:94:-	**4**	**2076633**	**0.414**	**Sobic.004G026000, Sobic.004G026200, Sobic.004G026100**	**ankyrin repeat domain-containing protein, chloroplast precursor (putative, expressed), tRNA methyltransferase (putative, expressed)**
	202300:19:+	4	14333705	0.282	-	-
	211119:123:-	4	26683113	0.024	-	-
	350017:27:-	7	7098186	0.027	-	-
	402529:102:-	8	11742333	0.355	-	-

Source: SorGSD (Liu et al., 2021)

4. Discussion

The results in Figure 1 show that root mass (root.biomass.kg), plant growth (fresh.biomass.kg), and number of shoots (tiller) are all closely related in sorghum. This reflects a shared focus on growth. More extensive roots allow for better nutrient and water uptake, which can fuel increased plant growth (fresh biomass) and potentially lead to more shoots (tillers) being produced (Wang et al., 2006). Taller sorghum plants (plant.height.cm) might have a double advantage for yield (plot.yield). First, they can capture more sunlight due to their height, which fuels growth and grain production (George-Jaeggl et al., 2021). Second, taller plants often have larger root systems, allowing them to access more water and nutrients from the soil, further boosting yield potential. This positive correlation suggests that selecting taller varieties within breeding programs could be a strategy to increase sorghum grain production. Earlier flowering (flowering day), potentially smaller leaves, and heavier grains (grain weight) might be interconnected. This could be a resource allocation strategy. Plants that flower earlier might prioritize grain production over extensive leaf growth, directing resources towards larger, heavier seeds(Ahmed et al., 2016). This trade-off could ensure successful reproduction even with slightly less leaf area for photosynthesis.

This study employed LOWESS, a technique suited for non-linear relationships, to explore connections between sorghum traits. Analyzing plant height and yield (Figure 2) revealed a fascinating complexity. Initially, yield increases with rising plant height, suggesting taller plants might be more productive. However, a dip in yield occurs at some intermediate height, followed by another rise and a final decline. This pattern suggests there might be an optimal height range for maximizing yield, potentially due to factors like resource allocation or limitations on light interception at extreme heights and lodging susceptibility (Hu et al., 2018).

Similarly, LOWESS uncovered interesting dynamics between yield, root biomass, and fresh biomass. While yield generally increases with increasing biomass, the rate of yield increase slows down after a certain point of biomass accumulation. This suggests diminishing returns - larger plants eventually reach a point where additional biomass doesn't translate proportionally into higher yield (Anfinrud et al, 2013). This could be due to factors like limitations in the plant's ability to convert biomass into grain or potential energy allocation trade-offs.

Linear Discriminant Analysis (LDA) in Figure 3 successfully classified the two sorghum head types (Compact Elliptic and Half-Broomcorn) based on plot yield, seed weight, and flowering duration. Compact Elliptic heads are dominant when plot yield is between 0.2 and 1 unit (potentially lower yielding) and seed weight falls between 10 and 30 units. Half-Broomcorn heads are more prevalent at higher plot yields (1 or more) and a distinct seed weight of 37 units. This suggests a potential trade-off between yield and seed size for these head types. Interestingly, both head types have similar flowering durations (65-85 days) but are differentiated by plot yield. Compact Elliptic is associated with lower yields (0.2-1 units). These findings highlight the effectiveness of LDA in separating head types based on a combination of yield, seed weight, and

flowering duration highlighting potential breeding strategies. Focusing on these traits could help develop sorghum varieties tailored for specific yield or seed size goals.

Our research scrutinized the performances of different correlation levels induced by LASSO and GWAS techniques. In evaluating the performance of GWAS and LASSO for classifying sorghum head types, we observed that LASSO identified more SNPs compared to GWAS. This finding aligns with previous studies that compare the statistical performance GWAS and Penalized regression approaches (Waldmann et al.,2013). According to Waldmann et al (2013) phenotypes (in our case, sorghum head types) contain almost no error variance leading to a significant difference between the number of SNPs identified by LASSO and GWAS. Nevertheless, both computational methodologies consistently detected SNPs believed to be genes in the Compact Elliptic and Half-Broomcorn types (Table 1).

The genes identified through GWAS and LASSO are crucial for plant metabolism. The different sorghum head types (compact and half-broomcorn) may result from these genetic differences. In the half-broomcorn case, we found genes that regulate Ankyrin repeat domain-containing protein. This protein, associated with the SNPs at coordinates 191353:94:- and 191353:132:-, plays a significant role in cell growth, development, and response to hormones and environmental stresses (Zhao et al., 2020; Ann et al., 2023). Additionally, in rice, ankyrin repeat domain-containing protein influences the number of secondary branches per panicle, a key yield-related trait. Considering the results of Figure 3, genes regulating ankyrin repeat domain-containing protein may significantly affect the formation of sorghum head phenotypes.

Interestingly, for specific SNPs (namely 168618:68:+ in Compact Elliptic and 191353:117:- in Half-Broomcorn type), no corresponding gene was identified. However, the SNP at 191353:117:- is closely situated, with a mere difference of 23 base pairs, to the 191353:94:- SNP, which is linked with the genes Sobic.004G026000, Sobic.004G026200, and Sobic.004G026100. Conversely, SNP 168618:68:+ is associated with the Sobic.003G202600 gene related to a single MYB histone in the nearby base pair (Liu et al., 2021). This phenomenon could be attributed to linkage disequilibrium, where genetic variants tend to be inherited together more frequently than anticipated by random chance (Hazelett et al., 2016). Notably, the SNPs we identified reside in non-coding regions, yet they are in close proximity to the actual gene locations. Alternatively, the none-coding region indicated by the SNP may be a short DNA region, an enhancer, that can bind to a protein to increase the likelihood of transcription of a specific gene (Tak and Farnham, 2015; Nicolae et al., 2010; González et al., 2019;).

While both GWAS and LASSO successfully identified SNPs associated with sorghum head type, they offer distinct advantages. GWAS excels in identifying a smaller set of high-confidence SNPs. This makes GWAS results easier to interpret and validate, providing a strong foundation for understanding the genetic basis of head type. On the other hand, LASSO has the potential to

uncover a wider range of SNPs, offering a more comprehensive view of the genetic landscape. This additional information might lead to the discovery of novel genes or regulatory elements influencing head development. However, these LASSO-identified SNPs might require further validation to confirm their true association with the trait. Together, GWAS and LASSO offer a complementary approach, with GWAS providing a strong starting point and LASSO potentially revealing a more intricate picture of the genetic factors at play.

5. Conclusion

This study provides a comprehensive analysis of the phenotypic and genotypic variations in sorghum varieties, elucidating key relationships between agronomic traits and their genetic underpinnings. The phenotypic data revealed significant correlations among traits such as plant height, root biomass, and yield, highlighting the interdependent nature of growth parameters in sorghum. Height plants and those with more extensive root systems were associated with higher yields, suggesting that selecting for these traits could enhance grain production. Additionally, the observed trade-offs between early flowering, smaller leaf area, and heavier grains emphasize the complexity of resource allocation in sorghum, where prioritizing certain traits might optimize reproductive success and yield potential.

The application of non-linear analysis techniques, such as LOWESS, provided deeper insights into the intricate relationships between sorghum traits. The study uncovered that while yield generally increases with plant height and biomass, there are optimal ranges beyond which further increases do not proportionally enhance yield. This finding suggests that breeders should aim for a balance in plant height and biomass to maximize yield, avoiding extremes that could lead to diminishing returns or increased susceptibility to lodging. Understanding of these relationships aids in formulating breeding strategies that can exploit these optimal ranges, thereby improving sorghum's productivity under varying environmental conditions.

Genotypic analysis through GWAS and LASSO regression identified several SNPs associated with important traits, providing a genetic basis for observed phenotypic variations. Identifying SNPs common to both methods, particularly those linked to key agronomic traits like head type, underscores the robustness of these findings. These genetic markers can be utilized in marker-assisted selection to accelerate the breeding of superior sorghum varieties. The study's integration of phenotypic and genotypic data presents a powerful approach to understanding and harnessing sorghum diversity, facilitating the development of varieties that are well-adapted to diverse environments and capable of meeting the growing demand for food and biofuel. This research contributes valuable knowledge to the field of sorghum breeding, promoting a future that is more secure and sustainable through improved crop varieties. Investigating identified SNPs and their genes, along with environmental influences on traits, are needed for future research. This enhances our understanding of sorghum adaptation. This knowledge, integrated with genomic data, is crucial for developing climate-resilient sorghum varieties.

7. References

Ahmed, A., Hassan, M., & Naim, A. (2016). Evaluation of Some Local Sorghum Genotypes in North Kordofan of Sudan Semi-Arid Agro-Ecological Environment. International Journal of Agriculture and Forestry, 6, 54-57. https://doi.org/10.5923/j.ijaf.20160601.06

Ahn, E., Botkin, J., Ellur, V., Lee, Y., Poudel, K., Prom, L. K., & Magill, C. (2023). Genome-wide association study of seed morphology traits in senegalese sorghum cultivars. Plants, 12(12), 2344. https://doi.org/10.3390/plants12122344

Anfinrud, R., Cihacek, L., Johnson, B., Ji, Y., & Berti, M. (2013). Sorghum and kenaf biomass yield and quality response to nitrogen fertilization in the Northern Great Plains of the USA. Industrial Crops and Products, 50, 159-165 https://doi.org/10.1016/J.INDCROP.2013.07.022.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., ... & Varshney, R. K. (2017). Genomic selection in plant breeding: methods, models, and perspectives. Trends in plant science, 22(11), 961-975. https://doi.org/10.1016/j.tplants.2017.08.011

de Sousa, K., van Etten, J., Poland, J., Fadda, C., Jannink, J. L., Kidane, Y. G., ... & Dell'Acqua, M. (2021). Data-driven decentralized breeding increases prediction accuracy in a challenging crop production environment. Communications biology, 4(1), 944. https://doi.org/10.1038/s42003-021-02463-w

Deepika, C., Gnanamalar, R. P., Thangaraj, K., Revathy, N., & Karthikeyan, A. (2023). Linear discriminant analysis of grain quality traits in rice (Oryza sativa L.) using the digital imaging technique. Journal of Cereal Science, 109, 103609. https://doi.org/10.1016/j.jcs.2022.103609

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. The plant genome, 4(3). https://doi.org/10.3835/plantgenome2011.08.0024

Eshed, Y., & Lippman, Z. B. (2019). Revolutions in agriculture chart a course for targeted breeding of old and new crops. Science, 366(6466), eaax0025. https://doi.org/10.1126/science.aax0025

Food and Agriculture Organization of the United Nations. (2024). FAOSTAT statistical database. Rome: FAO. Retrieved from https://www.fao.org/faostat/en/

Francia, E., Tacconi, G., Crosatti, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., & Valè, G. (2005). Marker assisted selection in crop plants. Plant Cell, Tissue and Organ Culture, 82, 317-342. https://doi.org/10.1007/s11240-005-2387-z.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., ... & Yang, J. (2023). glmnet: Lasso and elastic-net regularized generalized linear models. *Astrophysics Source Code Library*, ascl-2308.

George-Jaeggli, B., Lefèvre-Arbogast, S., Hunt, C., Cruickshank, A., & Jordan, D. (2021). Tall 3-dwarfs: oxymoron or opportunity to increase grain yield in sorghum?. Planta, 253. https://doi.org/10.1007/s00425-021-03629-w.

González, A., Artufel, M., & Rihet, P. (2019). TAGOOS: genome-wide supervised learning of non-coding loci associated to complex phenotypes. Nucleic Acids Research, 47(14), e79-e79. https://doi.org/10.1093/nar/gkz320

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. Retrieved from http://ggplot2.org, https://github.com/hadley/ggplot2

Hadebe, S., Modi, A., & Mabhaudhi, T. (2017). Drought Tolerance and Water Use of Cereal Crops: A Focus on Sorghum as a Food Security Crop in Sub-Saharan Africa. Journal of Agronomy and Crop Science, 203, 177-191. https://doi.org/10.1111/JAC.12191.

Hazelett, D. J., Conti, D. V., Han, Y., Al Olama, A. A., Easton, D., Eeles, R. A., ... & Coetzee, G. A. (2016). Reducing GWAS complexity. Cell Cycle, 15(1), 22-24. https://doi.org/10.1080/15384101.2015.1120928

Isgren, E., Andersson, E., & Carton, W. (2020). New perennial grains in African smallholder agriculture from a farming systems perspective. A review. *Agronomy for Sustainable Development, 40*, 1-14. https://doi.org/10.1007/s13593-020-0609-8

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York, NY: Springer

Jin, H., Tross, M. C., Tan, R., Newton, L., Mural, R. V., Yang, J., ... & Schnable, J. C. (2024). Imitating the "breeder's eye": Predicting grain yield from measurements of non-yield traits. The Plant Phenome Journal, 7(1), e20102. https://doi.org/10.1002/ppj2.20102

Kane, D. A., Roge, P., & Snapp, S. S. (2016). A systematic review of perennial staple crops literature using topic modeling and bibliometric analysis. PloS one, 11(5), e0155788. https://doi.org/10.1371/journal.pone.0155788

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. Retrieved from https://CRAN.R-project.org/package=factoextra

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Liu, Y., Wang, Z., Wu, X., Zhu, J., Luo, H., Tian, D., ... & Jing, H. C. (2021). SorGSD: updating and expanding the sorghum genome science database with new contents and tools. Biotechnology for Biofuels, 14(1), 165. https://doi.org/10.1186/s13068-021-02016-7

Macauley, H., and Ramadjita, T. (2015). *Cereal crops: Rice, maize, millet, sorghum, wheat.* Retrieved from https://www.afdb.org/fileadmin/uploads/afdb/Documents/Events/DakAgri2015/Cereal_Crops-_Rice__Maize__Millet__Sorghum__Wheat.pdf

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., & Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS genetics, 6(4), e1000888. https://doi.org/10.1371/journal.pgen.1000888

Nouira, A., & Azencott, C. A. (2021). Multitask group Lasso for Genome Wide association Studies in diverse populations. *In PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022* (pp. 163-174). https://doi.org/10.1142/9789811250477_0016

Patil, A. E., Deosarkar, D. B., Khatri, N., & Ubale, A. B. (2023). A comprehensive investigation of Genotype-Environment interaction effects on seed cotton yield contributing traits in Gossypium hirsutum L. Using multivariate analysis and artificial neural network. Computers and Electronics in Agriculture, 211, 107966. https://doi.org/10.1016/j.compag.2023.107966

Pengcheng Hu, Scott C. Chapman, Xuemin Wang, Andries Potgieter, Tao Duan, David Jordan, Yan Guo, Bangyou Zheng, Estimation of plant height using a high throughput phenotyping platform based on unmanned aerial vehicle and self-calibration: Example for sorghum breeding, European Journal of Agronomy, Volume 95, 2018, Pages 24-32, ISSN 1161-0301, https://doi.org/10.1016/j.eja.2018.02.004.

Pintus, E., Sorbolini, S., Albera, A., Gaspa, G., Dimauro, C., Steri, R., ... & Macciotta, N. P. (2014). Use of locally weighted scatterplot smoothing (lowess) regression to study selection signatures in p iedmontese and i talian b rown cattle breeds. Animal Genetics, 45(1), 1-11. https://doi.org/10.1111/age.12076

Poland, J. (2015). Breeding-assisted genomics. Current opinion in plant biology, 24, 119-124. https://doi.org/10.1016/j.pbi.2015.02.009

Ribeiro, J. P. O., de Medeiros, A. D., Caliari, I. P., Trancoso, A. C. R., de Miranda, R. M., de Freitas, F. C. L., ... & dos Santos Dias, D. C. F. (2021). FT-NIR and linear discriminant analysis to classify chickpea seeds produced with harvest aid chemicals. Food Chemistry, 342, 128324. https://doi.org/10.1016/j.foodchem.2020.128324

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. Cran r, 538, 113-120. Retrieved from https://cran-r.c3sl.ufpr.br/web/packages/MASS/MASS.pdf

Shiferaw, W., & Yoseph, T. (2014). Collection, characterization and evaluation of sorghum (Sorghum bicolor (L.) Moench) landraces from South omo and Segen peoples Zone of South Nation Nationality Peoples Region, Ethiopia.. International Research Journal of Agricultural Science and soil Science, 4, 76-84. http:/dx.doi.org/10.14303/irjas.2014.029

Su, J., Pang, C., Wei, H., Li, L., Liang, B., Wang, C., ... & Yu, S. (2016). Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. BMC genomics, 17, 1-16. https://doi.org/10.1186/s12864-016-2875-z

Tak, Y. G., & Farnham, P. J. (2015). Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics & chromatin, 8, 1-18. https://doi.org/10.1186/s13072-015-0050-4

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vasseur, F., Fouqueau, L., de Vienne, D., Nidelet, T., Violle, C., & Weigel, D. (2019). Nonlinear phenotypic variation uncovers the emergence of heterosis in Arabidopsis thaliana. PLoS biology, 17(4), e3000214. https://doi.org/10.1371/journal.pbio.3000214

Waldmann, P., Mészáros, G., Gredler, B., Fuerst, C., & Sölkner, J. (2013). Evaluation of the lasso and the elastic net in genome-wide association studies. Frontiers in genetics, 4, 58395. https://doi.org/10.3389/fgene.2013.00270

Wang, H., Inukai, Y., & Yamauchi, A. (2006). Root Development and Nutrient Uptake. Critical Reviews in Plant Sciences, 25, 279 - 301. https://doi.org/10.1080/07352680600709917.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R,Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL,Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." _Journal of Open Source Software_, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.4. Retrieved from dplyr.tidyverse.org

Xin, Z., Wang, M., Cuevas, H. E., Chen, J., Harrison, M., Pugh, N. A., & Morris, G. (2021). Sorghum genetic, genomic, and breeding resources. Planta, 254(6), 114. https://doi.org/10.1007/s00425-021-03742-w

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., ... & Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding. Journal of Experimental Botany, 68(11), 2641-2666. https://doi.org/10.1093/jxb/erx135

Xu, Y., Zhang, X., Li, H., Zheng, H., Zhang, J., Olsen, M. S., ... & Qian, Q. (2022). Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. Molecular Plant, 15(11), 1664-1695. https://doi.org/10.1016/j.molp.2022.09.001

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., ... & Liu, X. (2020). rMVP: Memory-Efficient, Visualize-Enhanced, Parallel-Accelerated GWAS Tool. R package version, 1(4).

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., ... & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. Nature genetics, 42(4), 355-360. https://doi.org/10.1038/ng.546

Zhao, J. Y., Lu, Z. W., Sun, Y., Fang, Z. W., Chen, J., Zhou, Y. B., ... & Min, D. H. (2020). The ankyrin-repeat gene GmANK114 confers drought and salt tolerance in Arabidopsis and soybean. Frontiers in Plant Science, 11, 584167. https://doi.org/10.3389/fpls.2020.584167

8. Supplement

Table S1. Details of sorghum traits collected during on station trials

| Trait	Description	Measurement Stage	Measurement Method
Days to 50% Flowering (DFL)	Number of days taken for 50% of plants in a plot to initiate flowering	Reproductive stage (flowering initiation)	Daily visual observation in the field to identify flowering panicles on at least half the plants in a plot.
Plant Height (PH)	Distance from the base of the plant to the tip of the main stem	Reproductive stage (flowering)	Measuring the height of 3 representative plants in a plot from ground level to the tip of the panicle (excluding awns) using a measuring stick.
Leaf Area (LA)	Area of the longest fully expanded leaf on the main stem	Vegetative or reproductive stage	Measuring the length of the main vein of 3 representative leaves from base to tip using a measuring stick or ruler.
Number of Tillers (Nti)	Total number of secondary stems arising from the base of the main stem	Vegetative stage (before flowering)	Counting the total number of tillers on 3 representative plants within a plot.
Lodging	Degree to which the main stems are leaning or lying on the ground	Reproductive stage (maturity)	Categorical scoring (e.g., erect, partially lodged, completely lodged) based on visual observation in the field.
Sorghum head types (SHT)	Phenotypic of sorghum head types	Reproductive stage (maturity	Recording sorghum head types as nominal data
Grain Yield (GY)	Total weight of threshed grain produced per plot	Reproductive stage (maturity)	Harvesting the grain from the plot, threshing it to remove the kernels from the panicle, and weighing the dry grain.
Fresh Plant Biomass (FPB)	Total weight of all above-ground plant material	Reproductive stage (maturity)	Harvesting the entire above-ground portion of representative plants, weighing them fresh in the field.
Dry Root Biomass (DRB)	Total weight of the root system	Reproductive stage (maturity)	Carefully excavating the root system of representative plants, washing away soil particles, drying the roots in an oven, and recording the dry weight.
100-Seed Weight (100gwt)	Average weight of 100 mature seeds	After threshing	Randomly sampling 100 seeds from the harvested grain, drying them thoroughly, and weighing them collectively. The weight is then divided by 100 to obtain the average seed weight.
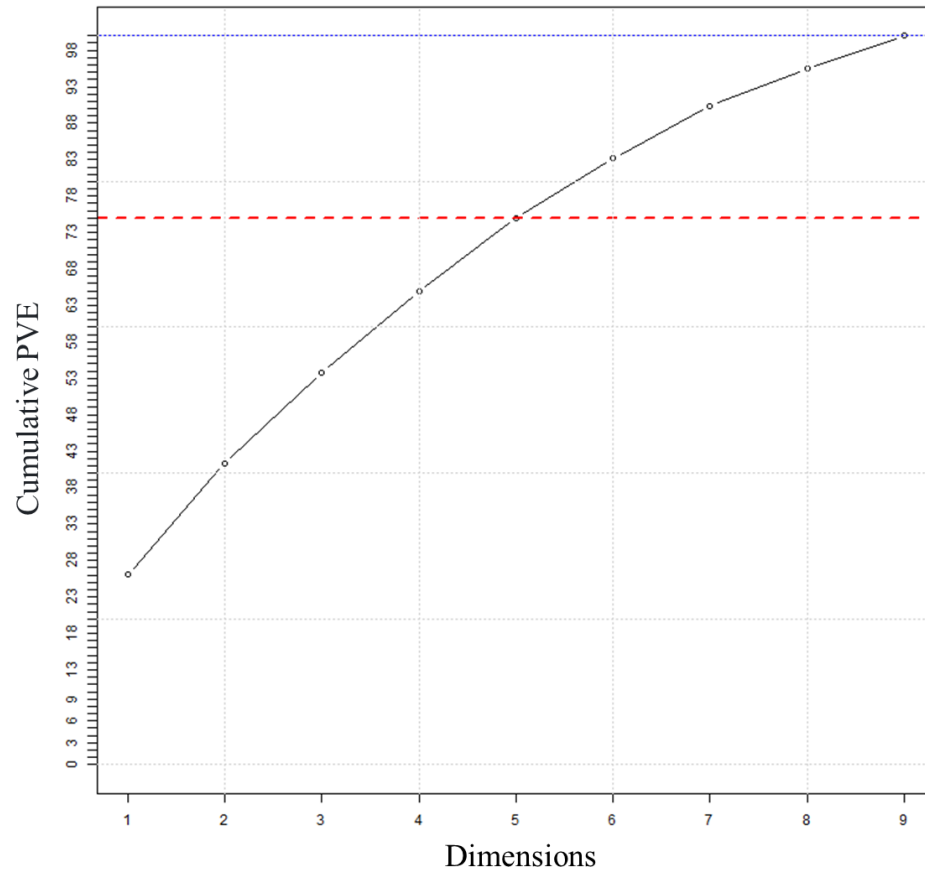
Figure S1. The cumulative percentage of variance explained (PVE) based each principal component. Red dot line present 75 % of the total variance in the data
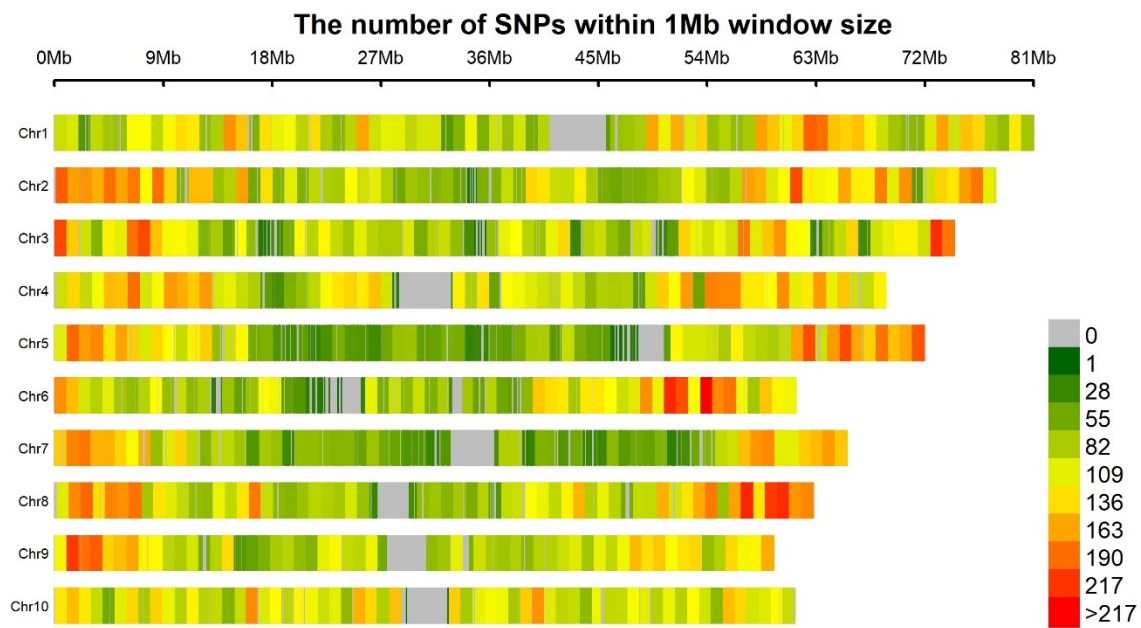
Figure S2. Number of SNPs within 1 Mb window for all the 10 chromosomes
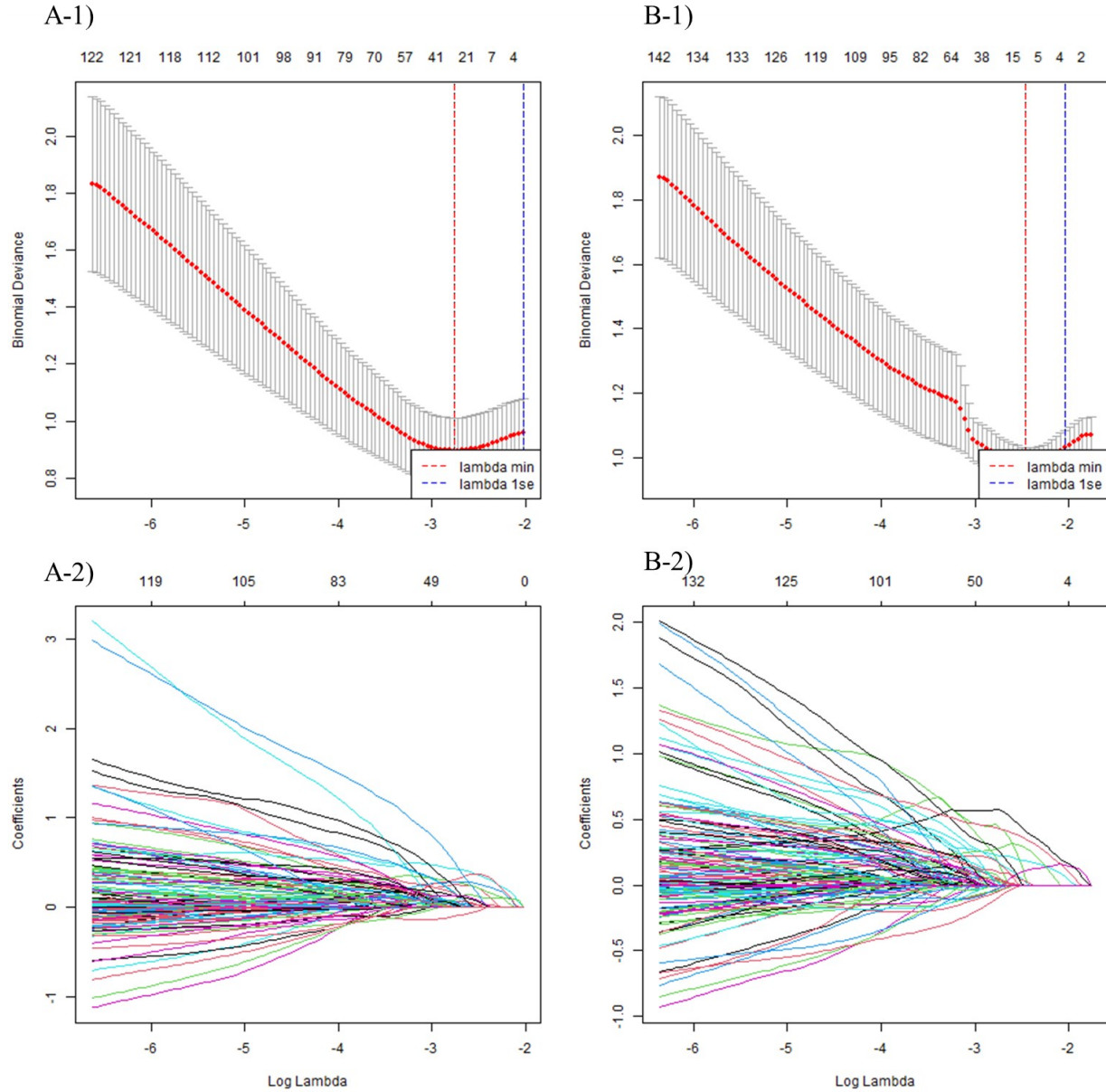
Figure S3. The mean-squared error (MSE) and the number of SNPs, A-1) & B-1, and coefficient path A-2) & B-2) in the Least Absolute Shrinkage and Selection Operator (LASSO) regression as functions of log(λ) for the 10-fold cross-validation analyses. A is ab Compact Elliptic' and B is 'Half Broomcorn'.

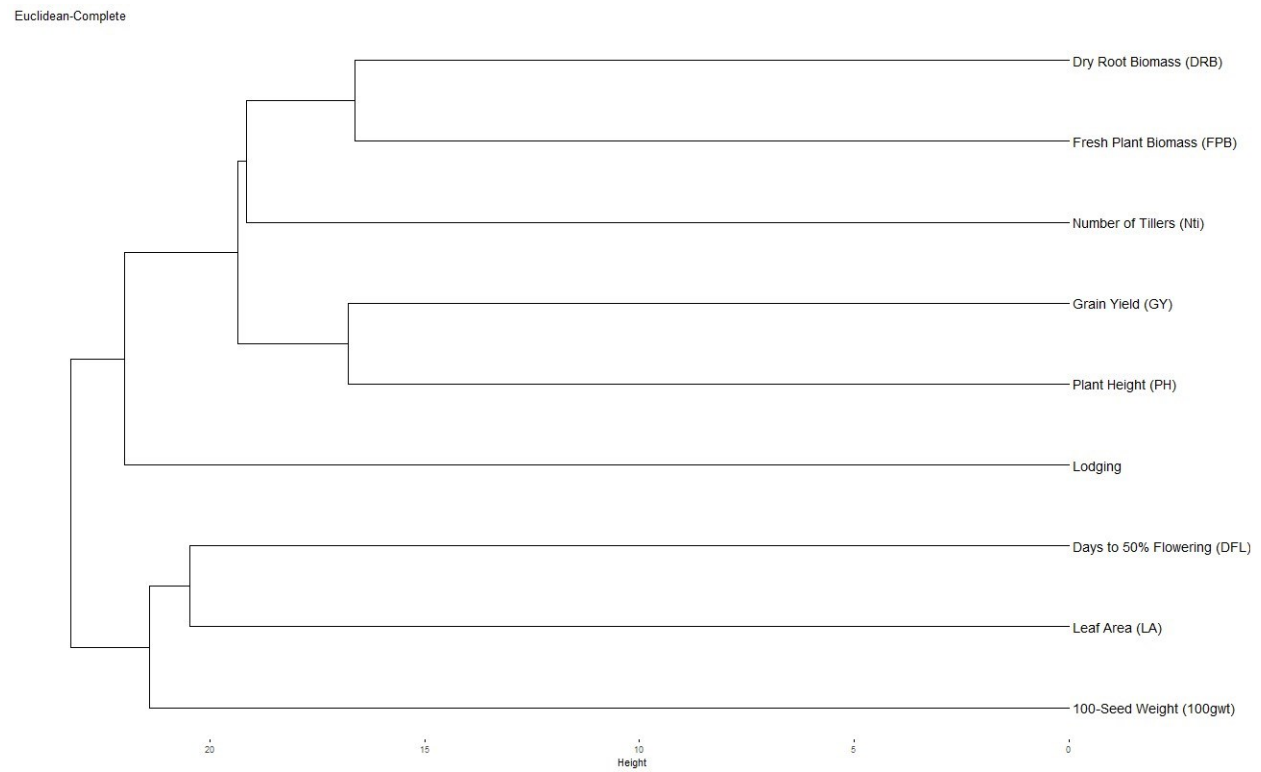**Figure S4**. Dendrogram of phenotypic characteristics of sorghum calculated using Euclidean distance.

Table S2. Pearson's correlation coefficient between sorghum traits. The p values show the significance of The hypothetical test for Pearson's correlation coefficient.

Traits	tillers	p_value		leaf_area	p_value		plant.height.cm.	p_value		days.toflowering	p_value	
Number of Tillers (Nti)	1.000	0.000	***	-0.145	0.025	*	0.204	0.002	**	0.097	0.137	
Leaf Area (LA)	-0.145	0.025	*	1.000	0.000	***	0.166	0.010	*	0.108	0.097	
Plant Height (PHT),	0.204	0.002	**	0.166	0.010	*	1.000	0.000	***	0.375	0.000	***
Days to 50% Flowering (DFL)	0.097	0.137		0.108	0.097		0.375	0.000	***	1.000	0.000	***
Fresh Plant Biomass (FPB)	0.395	0.000	***	0.147	0.024	*	0.312	0.000	***	0.150	0.021	*
Dry Root Biomass (DRB)	0.219	0.001	**	0.075	0.253		0.204	0.002	**	0.007	0.919	
Grain Yield (GY)	0.205	0.002	**	0.181	0.005	**	0.400	0.000	***	-0.015	0.820	
100-Seed Weight (100gwt)	-0.076	0.245		0.058	0.376		-0.008	0.897		0.025	0.706	

	fresh.biomass.kg.	p_value		root.biomass.kg.	p_value		plot.yield	p_value		X100.seedweight	p_value	
Number of Tillers (Nti)	0.395	0.000	***	0.219	0.001	**	0.205	0.002	**	-0.076	0.245	
Leaf Area (LA)	0.147	0.024	*	0.075	0.253		0.181	0.005	**	0.058	0.376	
Plant Height (PHT),	0.312	0.000	***	0.204	0.002	**	0.400	0.000	***	-0.008	0.897	
Days to 50% Flowering (DFL)	0.150	0.021	*	0.007	0.919		-0.015	0.820		0.025	0.706	
Fresh Plant Biomass (FPB)	1.000	0.000	***	0.412	0.000	***	0.322	0.000	***	0.037	0.573	
Dry Root Biomass (DRB).	0.412	0.000	***	1.000	0.000	***	0.397	0.000	***	-0.149	0.022	*
Grain Yield (GY)	0.322	0.000	***	0.397	0.000	***	1.000	0.000	***	-0.053	0.414	
100-Seed Weight (100gwt)	0.037	0.573		-0.149	0.022	*	-0.053	0.414		1.000	0.000	***

Note: Significance values: *** <0.001; ** <0.01; * <0.05; . <0.