

Article

Not peer-reviewed version

Genome-Wide Association Study of Seed Morphology-Related TRAITS in Sorghum Mini Core and Senegalese Lines

Ezekiel Ahn ^{*} , Sunchung Park , Zhenbin Hu , Vishnutej Ellur , Minhyeok Cha , Yoonjung Lee , Louis K. Prom , Clint W. Magill ^{*}

Posted Date: 21 February 2024

doi: 10.20944/preprints202402.1209.v1

Keywords: Sorghum; Seed; Mini core collection; Senegalese; Seed morphology; Seed color; GWAS



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Genome-Wide Association Study of Seed Morphology-Related TRAITS in Sorghum Mini Core and Senegalese Lines

Ezekiel Ahn ^{1,*†}, Sunchung Park ^{1,†}, Zhenbin Hu ², Vishnutej Ellur ³, Minhyeok Cha ⁴, Yoonjung Lee ⁵, Louis K. Prom ⁶ and Clint Magill ^{7,*}

¹ USDA-ARS Sustainable Perennial Crops Laboratory, Beltsville Agricultural Research Center, Beltsville, MD 20705, USA

² USDA-ARS Animal Genomics and Improvement Laboratory, Beltsville, MD 20705, USA

³ Molecular Plant Sciences, Washington State University, Pullman, WA 99164, USA

⁴ Department of Biotechnology, Korea University, Seoul, 02841, Republic of Korea

⁵ Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108, USA

⁶ USDA-ARS Southern Plains Agricultural Research Center, College Station, TX 77845, USA

⁷ Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX 77843, USA

* Correspondence: ezekiel.ahn@usda.gov (E.Z.); c-magill@tamu.edu (C.M.)

† These authors contributed equally to this work.

Abstract: Sorghum (*Sorghum bicolor* L.) ranks fifth as the most crucial cereal crop globally, yet its seed morphology remains relatively unexplored. This study investigated seed morphology in sorghum based on 115 mini core and 130 Senegalese germplasms. Eight seed morphology traits encompassing size, shape, and color parameters were assessed. Statistical analyses explored potential associations between these traits and resistance to three major sorghum diseases: anthracnose, head smut, and downy mildew. Furthermore, genome-wide association studies (GWAS) were conducted using phenotypic data from over 24,000 seeds and over 290,000 publicly available single nucleotide polymorphisms (SNPs) through the Genome Association and Prediction Integrated Tool (GAPIT) R package. Significant SNPs associated with various seed morphology traits were identified and mapped onto the reference sorghum genome to identify novel candidate defense genes.

Keywords: sorghum; seed; mini core collection; Senegalese; seed morphology; seed color; GWAS

1. Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] ranks as the fifth significant cereal grain in terms of global production and cultivated area [1]. It uses less water and endures climate change better than other cereals. In light of climate change and rising global temperatures, sorghum could be a feasible solution for growers [2]. Due to its high nutritional content, drought tolerance, minimal input requirements, and remarkable environmental adaptability, sorghum is a crucial crop for food security [3,4,5]. Sorghum is a widely cultivated crop grown in over 100 countries, particularly in dry, hot, and arid regions [6]. The largest Sorghum producers are the United States of America, Nigeria, Sudan, Mexico, Ethiopia, and India [7]. In the United States, it is mainly grown in rainfed conditions in dry regions on ultisol and mollisol soil types known as the "Sorghum Belt" in Kansas, Texas, Colorado, Oklahoma, and South Dakota [8,9].

Sorghum provides essential nutrients and phytochemicals. It contains protein, dietary fiber, and important minerals [10]. Polyunsaturated fatty acids such as linoleic, oleic, palmitic, linolenic, and stearic can be found in sorghum. Vitamins, particularly those from the B group and fat-soluble vitamins (A, D, E, and K), are abundant in sorghum. Furthermore, sorghum is a rich source of secondary plant metabolites and macro- and microelements such as phenolic acids, flavonoids,

sterols, policosanols, and antioxidants [11,12]. Compared to other primary cereal grains, sorghum is distinguished by its more significant amount of resistant starch and slowly digestible starch, which contributes to reducing spikes in blood sugar levels after meals [13]. Sorghum's diverse bioactive polyphenols can lower the risk of nutrition-linked chronic diseases. Additionally, its high molecular weight tannins are known to alter the functionality of proteins and starch, offering the potential for developing novel bioactive ingredients and enhancing food quality [14]. Sorghum's abovementioned factors make it a rare climate-resilient crop that can address food and nutrition security.

Sorghum is a multipurpose crop used in biofuel production, forage, ethanol production, and fodder preservation. Sorghum, particularly sweet sorghum, is a promising crop for biofuel production since it is a C4 crop that requires little input, has a high sugar content, and ease of extractability [15]. After human consumption, the remainder of sorghum is mainly utilized for animal feed [16]. The ideal mineral and fatty acid balance of sorghum and its protein source suitability for aquafeed production have recently increased its popularity as an aquafeed [17].

The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) GeneBank has almost 37,000 Sorghum accessions, 2,247 of which were selected to form a smaller group of germplasm known as the core collection. However, this core collection was also overwhelming. The core collection was evaluated for 11 qualitative and 10 quantitative traits, yielding 21 hierarchical groupings. From each cluster, about 10% or at least one accession was selected to create a mini core of 242 accessions [18]. The Sorghum mini core contains 10% of the core's accessions, or 1% of the entire collection, representing homogeneity for geographical origin, biological races, qualitative features, means, variances, phenotypic diversity indices, and phenotypic correlation. As a result, it is widely used for evaluating various agronomic traits and resistant traits, both biotic and abiotic, in the current genomic studies [18,19,20].

Sorghum germplasm lines from Senegal have emerged as a valuable source of essential agronomic traits, particularly resistance to biotic stresses such as fungal diseases [21]. Extensive Genome-wide association studies (GWAS) have dissected sorghum resistance against various fungal pathogens in the germplasms [21,22,23]. However, research on other agronomically important traits, such as seed morphology, has received limited attention.

Morphological variation in seed traits includes variations in seed size and shape. The morphology of seeds is a crucial agricultural characteristic as it reflects a combination of genetic, physiological, and environmental aspects, all of which significantly impact crop yield, quality, and market value [24]. Apart from market value, seed morphology has proved beneficial in determining taxonomic relationships in plant families. As a result, both seed shape and size are relevant parameters for assessing plant biodiversity [24]. In addition, investigating the biodiversity of seeds can help characterize intra- and inter-species variation, genotypic discrimination, and correlation—all of which are important for breeding to achieve the target levels of seed yield and quality [24,25].

Wang *et al.* [26] evaluated sorghum mini core panel in multiple locations with 6,094,317 single nucleotide polymorphism (SNP) markers and identified one locus for recurring peduncles and eight loci for panicle length, width, and compactness. Sakamoto *et al.* [27] used multi-trait GWAS to analyze 329 sorghum germplasms from different origins and found SNPs that may be related to seed morphology, such as SNP loci S01_50413644, S04_59021202, and S05_9112888. GWAS conducted on the 300 diverse accessions of the sorghum association panel (SAP) with 265,487 SNPs identified 30 SNPs that were strongly associated with traits measured at the seedling stage under cold stress, and 12 SNPs were significantly associated with seedling traits under heat stress [28]. Our previous study evaluated 162 Senegalese germplasms for seed area size, length, width, length-to-width ratio, perimeter, circularity, the distance between the intersection of length and width (IS) and center of gravity (CG), and seed darkness and brightness with 193,727 publicly available SNPs and identified multiple candidate genes potentially associated with seed morphology [29].

This study investigated seed morphology in a diverse panel of sorghum accessions, encompassing a subset of mini core collection (115 lines including IS19975 originated from Senegal) and germplasms from Senegal (130 lines excluding IS19975). Eight key quantitative traits related to seed size, shape, and color were evaluated in over 24,000 seeds. The selection of these accessions

prioritized the public availability of SNP data, facilitating GWAS to map genetic determinants of the observed phenotypic variation. By statistical analyses, the study explored potential connections between these traits and resistance to three major sorghum diseases: anthracnose, head smut, and downy mildew within the mini-core lines. Lastly, employing the Genome Association and Prediction Integrated Tool (GAPIT) R package, the study conducted GWAS using phenotypic data from the seeds and over 290,000 publicly available SNPs. This analysis identified and mapped significant SNPs associated with various seed morphology traits onto the reference sorghum genome.

2. Materials and methods

2.1. Phenotypic evaluations for seed morphology-related traits

Following the methodology of Ahn *et al.* [29], sorghum seed morphology was evaluated in 245 mini core and Senegalese germplasm lines (accession details in Supplementary Data S1) from the USDA-ARS Plant Genetic Resources Conservation Unit, Griffin, Georgia, to quantify and compare variation in key seed morphology traits. Seed area size (mm²), length (mm), width (mm), LWR, perimeter (mm), circularity (0-1 range, 0: not circular to 1: complete circle), distance between the intersection of length and width (IS) and center of gravity (CG), and seed color brightness (0-255 scale, 0: darkest, 255: brightest) were measured using digital image analysis with ImageJ software [18]. Data from BTx623 [29] was included as a control, representing a sorghum line with well-characterized seed morphology.

A comprehensive evaluation of seed morphology was conducted across 245 sorghum cultivars. Each cultivar had between 75 and 120 individual seeds analyzed for morphological traits. High-resolution images of the seeds were captured using a Canon imageRUNNER ADVANCE C7270 scanner (Canon Inc, Tokyo, Japan) and saved in JPEG format. SmartGrain (version 1.3) high-throughput phenotyping software quantified key traits: Area size, length, width, LWR, perimeter, circularity, and distance between IS and CG [18]. Any potential errors identified in the SmartGrain output were corrected by manually inspecting each image. ImageJ software (version 1.54d) was employed on 50 seeds per accession to assess seed color variation, applying a multi-point function to measure darkness and brightness levels across the accessions [19].

2.2. Statistical analysis

Following Ahn *et al.* [29], we performed Tukey's HSD test in JMP Pro 15 (SAS Institute, Cary, NC, USA) to analyze statistical differences among all tested accessions for each trait. Pearson's correlation coefficient analysis was used in JMP Pro 15 to identify potential pairwise correlations between seed morphology-related traits. Phenotypic data in mini core lines for anthracnose, head smut, and downy mildew resistance were obtained from Ahn *et al.* [19]. Student's t-tests were performed in JMP Pro 15 to compare seed morphology traits between resistant and susceptible cultivars for each of the three diseases. Additionally, a principal component analysis (PCA) and clustering variables analysis were conducted using phenotypic data to explore relationships between all traits, followed by logistic regression analysis between seed morphology traits and disease resistance traits in sorghum mini core lines.

2.3. GWAS

Genome-wide association study

The SNP data was extracted from an integrated sorghum SNPs dataset based on the sorghum reference genome version 3.1.1, which was genotyped initially using genotyping-by-sequencing (GBS) [30,31,32]. Missing data were imputed using Beagle 4.1 [33], and further filtering was performed, ensuring a minor allele frequency of at least 0.05. We conducted genome-wide association analyses using the R-package, GAPIT version 3 [34]. The analyses employed the Fixed and random model Circulating Probability Unification (FarmCPU) [35]. Population stratification was corrected by PCA, with the optimal number of principal components determined through a Bayesian information

criterion-based analysis within GAPIT. For the association analysis in GAPIT, the following parameters were used: PCA.total = 3 and model = FarmCPU, while all other parameters were kept at their default values. To identify significant SNP-trait associations, we applied a stringent Bonferroni correction, with a threshold of $-\log_{10}$ (*p*-value) of 6.77 or greater. Subsequently, we estimated pairwise LD (r^2) between the significant SNPs and nearby SNPs located within 100 kb both upstream and downstream of the significant SNPs. LD blocks were defined by merging SNPs that exhibited an r^2 value of at least 0.5. LD blocks represent genomic regions where SNPs are co-inherited due to strong LD. If an LD block was smaller than 20 kb, it was extended up to 20 kb. Within these regions, candidate genes were identified based on gene annotations from the sorghum reference genome available at the Phytozome 12 (<https://phytozome.jgi.doe.gov>) (version 3.1.1, GCF_000003195.3) [36].

3. Results

3.1. Seed morphologies

A two-tailed ANOVA was conducted on the 246 accessions, including the control accession BTx623, which showed significant differences with $p < 0.0001$ for all evaluated traits (Supplementary Data S1). Table 1 lists the top five cultivars for each trait, and phenotypic data is also available in Supplementary Data S1. For instance, the area size of IS11473 was $25.60 \pm 1.60 \text{ mm}^2$, while that of IS12697 was $4.92 \pm 0.83 \text{ mm}^2$ (Table 1 and Figure 1). Similarly, the seed colors of IS7987 and IS9108 showed the most significant contrast among the tested accessions (Table 1 and Figure 2). Significant morphological variations were identified across the population for other traits as well.

Table 1. Top seed morphology accessions across the accessions.

Largest area size (mm^2)		Smallest area size (mm^2)	
Accession	Mean \pm S.D.	Accession	Mean \pm S.D.
IS11473	25.60 ± 1.60	IS12697	4.92 ± 0.83
PI514404	21.61 ± 3.75	IS13264	7.76 ± 1.26
IS7987	21.24 ± 2.70	PI514394	7.93 ± 1.00
PI253986	20.04 ± 3.63	PI514308	8.07 ± 0.89
IS28141	19.32 ± 2.75	PI514474	8.18 ± 0.78
Longest perimeter (mm)		Shortest perimeter (mm)	
IS11473	20.57 ± 0.65	IS12697	8.54 ± 0.75
PI514404	18.16 ± 1.70	PI514394	11.04 ± 0.72
IS7987	17.96 ± 1.21	IS13264	11.08 ± 1.20

PI253986	17.28 ± 1.54	PI514308	11.13 ± 0.63
IS28141	17.15 ± 1.21	PI514474	11.25 ± 0.57
Longest length (mm)		Shortest length (mm)	
IS11473	6.32 ± 0.24	IS12697	3.03 ± 0.29
IS7987	5.73 ± 0.37	PI514434	3.74 ± 0.23
PI514404	5.59 ± 0.49	PI514394	3.89 ± 0.24
IS28141	5.55 ± 0.41	PI514308	3.90 ± 0.20
IS12804	5.46 ± 0.41	IS9108	3.91 ± 0.28
Longest width (mm)		Shortest width (mm)	
IS11473	5.59 ± 0.23	IS12697	2.17 ± 0.18
PI514404	5.19 ± 0.48	IS13264	2.61 ± 0.18
IS7987	5.01 ± 0.35	IS3121	2.76 ± 0.26
IS28141	5.00 ± 0.39	PI514394	2.79 ± 0.20
IS11026	4.99 ± 0.29	PI514474	2.82 ± 0.15
Highest LWR		Lowest LWR	
IS12804	1.73 ± 0.17	IS10302	1.06 ± 0.03
IS1233	1.59 ± 0.08	IS11026	1.07 ± 0.05
IS13264	1.58 ± 0.21	PI514323	1.07 ± 0.04
IS3121	1.47 ± 0.11	PI514283	1.08 ± 0.05
PI514471	1.46 ± 0.08	PI514288	1.08 ± 0.04

Highest circularity (0-1 scale)		Lowest circularity (0-1 scale)	
IS13294	0.87 ± 0.01	IS12804	0.73 ± 0.05
IS2872	0.87 ± 0.01	IS11473	0.76 ± 0.03
IS13893	0.86 ± 0.01	IS1233	0.78 ± 0.02
IS9108	0.86 ± 0.01	IS14090	0.79 ± 0.03
IS12937	0.85 ± 0.01	IS27034	0.79 ± 0.03
The longest distance between IS and CG (mm)		Shortest distance between IS and CG (mm)	
IS27034	0.40 ± 0.20	PI514394	0.18 ± 0.12
IS11026	0.39 ± 0.19	IS12697	0.18 ± 0.10
PI514288	0.37 ± 0.24	PI514434	0.18 ± 0.10
IS11473	0.37 ± 0.19	IS2872	0.19 ± 0.11
IS14090	0.36 ± 0.18	PI514468	0.19 ± 0.12
Brightest (0-255 scale)		Darkest (0-255 scale)	
IS7987	237.14 ± 6.64	IS9108	82.80 ± 10.00
IS32439	234.96 ± 11.57	IS11619	87.62 ± 13.74
IS32349	234.14 ± 12.62	IS9177	87.68 ± 9.00
IS7305	232.78 ± 17.78	IS13264	93.12 ± 22.34
PI514446	230.68 ± 12.47	PI11374	93.30 ± 12.03

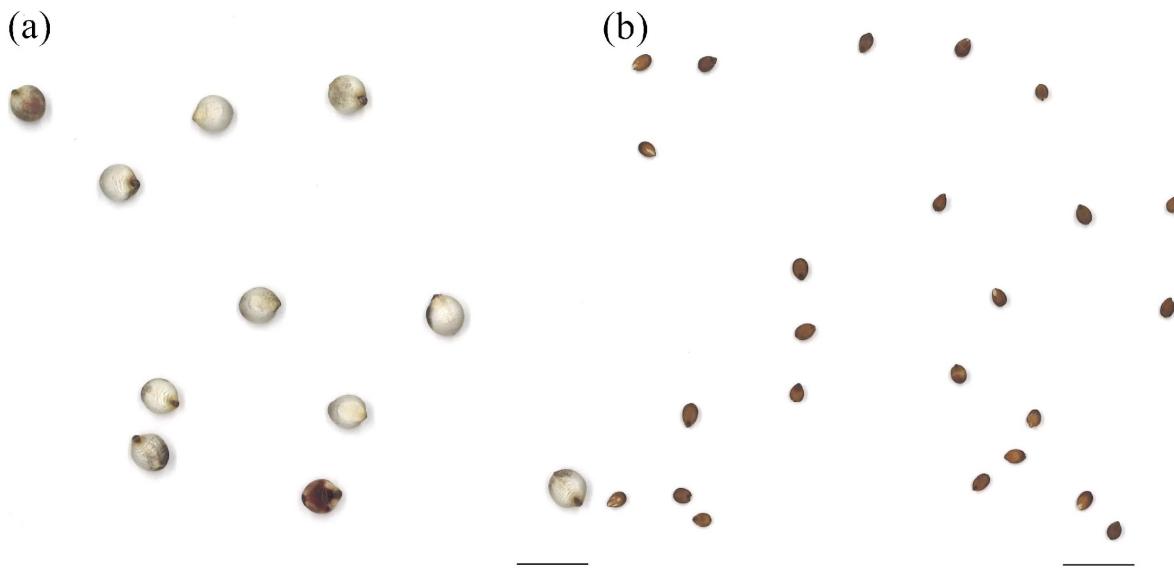


Figure 1. A comparison of the area sizes for IS11473 (PI329738) and IS12697 (PI302116). The seed of (a) IS11473 has one of the largest areas among the seeds compared, while the seed of (b) IS12697 has one of the smallest areas. The scale bars on the bottom right corner indicate 1 cm applied to both (a) and (b).

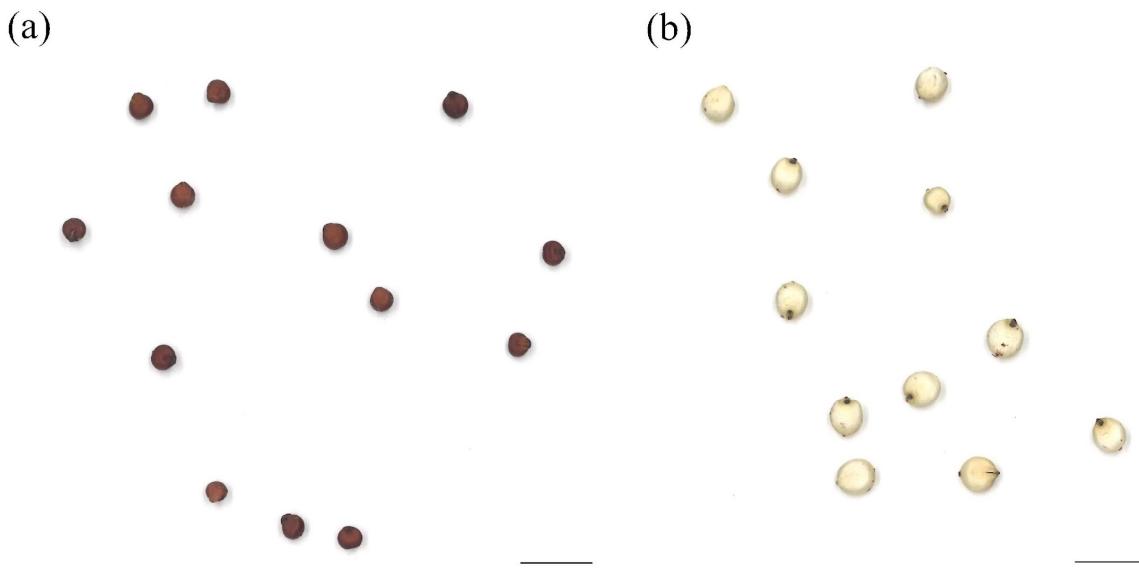


Figure 2. A comparison of the seed colors for IS9108 (PI682465) and IS7987 (PI685210). The seed of (a) IS9108 has one of the darkest colors among the mini core and Senegalese germplasms, while the seed of (b) IS7987 has one of the brightest colors. The scale bar indicates 1 cm applied to both (a) and (b).

3.2. Correlations in seed morphology-related traits

Pearson's correlation analysis revealed relationships between seed morphology traits in the mini core and Senegalese sorghum lines (Figure 3 and Table 2). Interestingly, seed brightness exhibited positive correlations with area size, perimeter, length, and width, contrasting with a previous study by Ahn *et al.* [29], which found no association in 162 Senegalese germplasms. Furthermore, PCA identified two principal components (PC1 and PC2) accounting for 77.6% of the total variance in seed morphology (Figure 4). The partial contribution of variables (Figure 5) revealed that PC1 is primarily driven by seed size-associated traits (area size, perimeter, length, and width). In contrast, PC2 reflects differences in seed shape, encompassing both circularity and length-to-width ratio. PC3, explaining less variance, is mainly associated with seed color. This result is consistent with the previous study

on the Senegalese line by Ahn *et al.* [29]. Cluster analysis of the phenotypic data in Table 3 formed three groups, with grain size, shape, and color being separated, similar to the results obtained from PCA analysis.

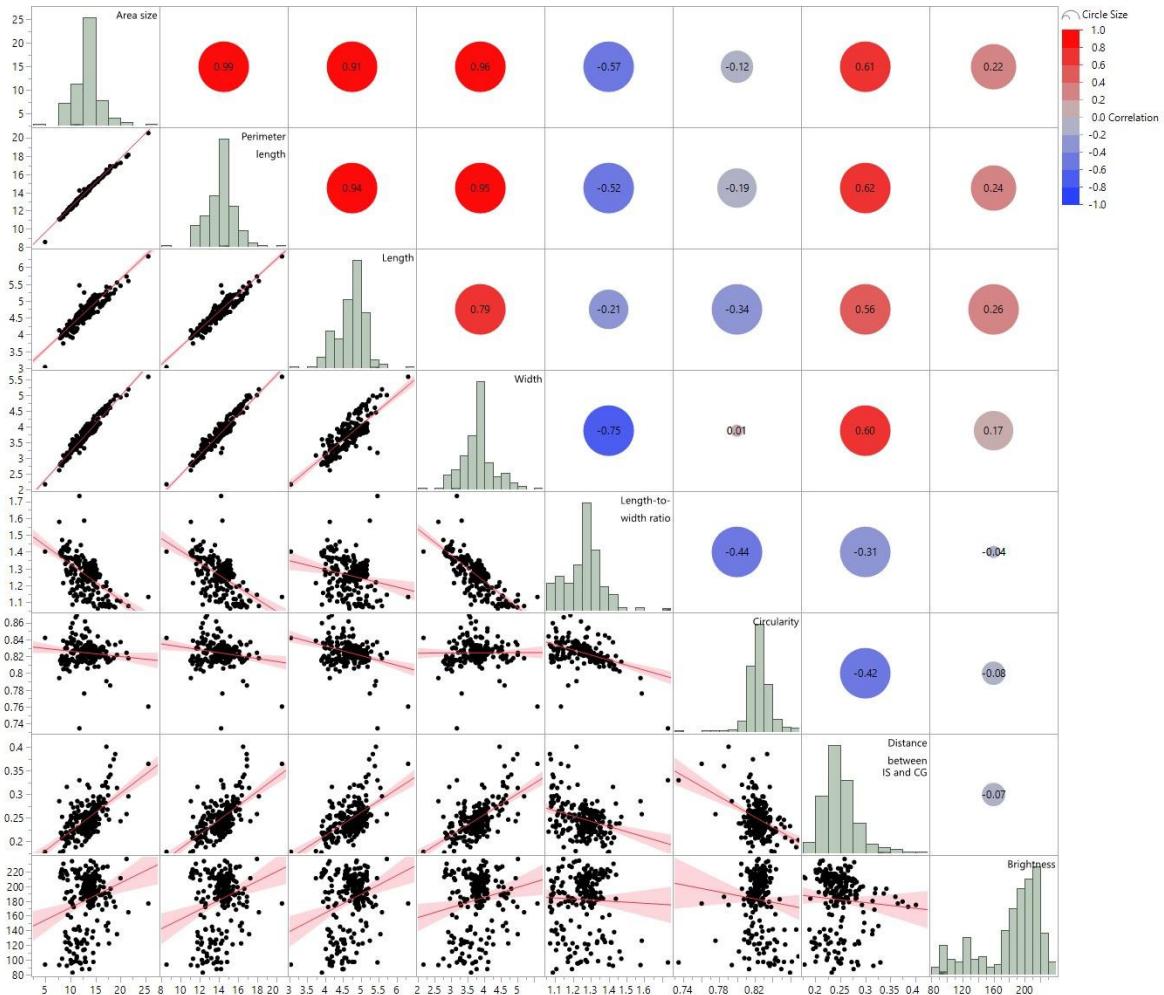


Figure 3. Scatter plots displaying correlations between the traits based on Pearson's r. The correlations are additionally shown with a heatmap and fit lines.

Table 2. Detailed correlations in eight seed morphology-related traits. ***= $P < 0.0001$, **= $P < 0.001$ and *= $P < 0.01$.

	Area size (mm ²)	Perimeter (mm)	Length (mm)	Width (mm)	LWR	Circularity	IS and CG	Brightness
Area size (mm ²)	1.00***	0.99***	0.91***	0.96***	-0.57***	-0.12	0.61***	0.22**
Perimeter (mm)	0.99***	1.00***	0.94***	0.95***	-0.52***	-0.19*	0.62***	0.24**
Length (mm)	0.91***	0.94***	1.00***	0.79***	-0.21**	-0.34***	0.56***	0.26
Width (mm)	0.96***	0.95***	0.79***	1.00***	-0.75***	0.01	0.60***	0.17**

LWR	-0.57***	-0.52***	-0.21**	-0.75***	1.00***	-0.44***	-0.31***	-0.04
Circularity	-0.12	-0.19*	-0.34***	0.01	-0.44***	1.00***	-0.42***	-0.08
IS and CG	0.61***	0.62***	0.56***	0.60***	-0.31***	-0.42***	1.00***	-0.07
Brightness	0.22**	0.24**	0.26***	0.17**	-0.04	-0.08	-0.07	1.00***

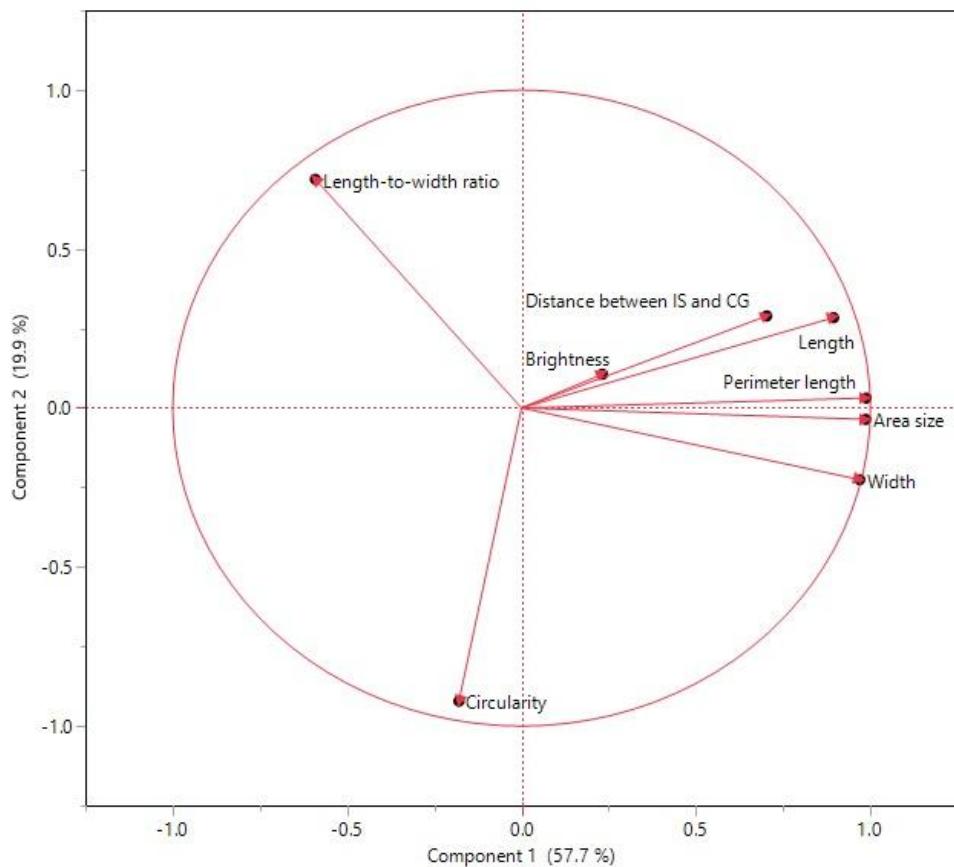


Figure 4. The principal component analysis of all seed morphology-related traits from tested sorghum germplasms. The plot displays PC1 vs PC2.

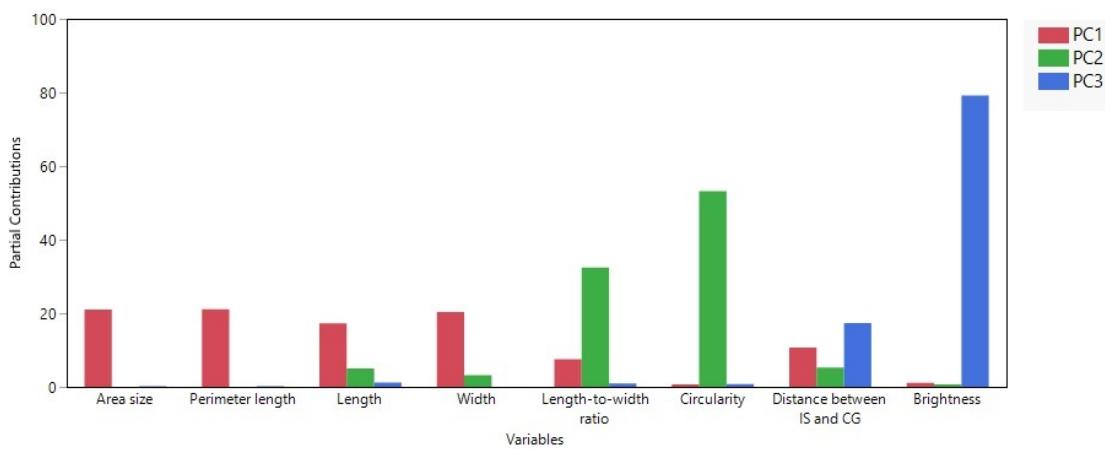


Figure 5. The partial contributions of variables to seed morphology traits in sorghum mini core and Senegalese lines are shown in the plot. The partial contributions toward PC1 (red), PC2 (green), and PC3 (blue) are displayed for each trait.

Table 3. Cluster variables analysis among the seed morphology-related traits. Three clusters were formed based on seed characteristics: Size, color, and shape.

Cluster	Members	R ² with its own cluster	R ² with the next closest	1-R ²
1	Perimeter length	0.98	0.06	0.02
	Area size	0.97	0.07	0.03
	Width	0.90	0.20	0.12
	Length	0.86	0.07	0.15
	Distance between IS and CG	0.52	0.01	0.49
2	Circularity	0.72	0.05	0.29
	Length-to-width ratio	0.72	0.27	0.38
3	Brightness	1.00	0.04	0.00

3.3. Seed morphology traits and anthracnose, head smut, and downy mildew resistance in mini core lines

We compared eight seed morphology traits between resistant and susceptible cultivars for anthracnose, head smut, and downy mildew in mini core lines using a two-tailed Student's t-test. No statistically significant differences were observed for anthracnose and downy mildew. However, for head smut, five of the eight traits exhibited associations with susceptibility (Table 4). Resistant lines had significantly larger area size, perimeter, and width than susceptible lines. Additionally, the resistant group displayed a slightly more circular shape, as indicated by both LWR and circularity. Notably, seed brightness did not show any significant difference between the groups.

Table 4. Comparison of seed morphology traits in sorghum mini core lines between head smut resistant and susceptible groups. * = P < 0.05, ND = no significant difference.

	Area size (mm ²)	Perimeter (mm)	Length (mm)	Width (mm)	LWR	Circularity (0-1 scale)	IS and CG (mm)	Brightness (0-255 scale)
Resistant	14.1	14.51	4.73	4	1.19	0.83	0.26	152.12
Susceptible	12.76	13.82	4.59	3.75	1.24	0.82	0.25	160
Significance based on p-value	*	*	ND	*	*	*	ND	ND

Similarly, logistic regression analyses revealed relationships between seed morphology traits and disease resistance in the mini-core sorghum lines (Table 5). The distance between IS and CG was associated with anthracnose and downy mildew. Identical to the t-test in Table 3, head smut was associated with the five traits but not with length, IS and CG, and brightness. These results suggest that larger, wider, and more circular seeds might be more resistant to head smut infection in this sorghum germplasms.

Table 5. Logistic regression analysis of seed morphology traits and disease resistance in sorghum mini core lines. *= $P < 0.05$ (Chi-square test), NS = No significant association.

	Area size (mm ²)	Perimeter (mm)	Length (mm)	Width (mm)	LWR	Circularity (0-1 scale)	IS and CG (mm)	Brightness (0-255 scale)
Anthracnose	NS	NS	NS	NS	NS	NS	*	NS
Head smut	*	*	NS	*	*	*	NS	NS
Downy mildew	NS	NS	NS	NS	NS	NS	*	NS

3.4. GWAS

Overall, 68 single-nucleotide polymorphisms (SNPs) surpassed the Bonferroni correction threshold for association with seed morphology traits. The number of identified SNPs varied across traits, ranging from 5 for IS and CG to 13 for area size. A detailed list of significant SNPs for each trait is provided in Table S1, while Manhattan plots visualizing their genomic distribution are presented in Figure 6. Furthermore, we identified over 100 candidate genes potentially linked to the significant SNPs based on their location within the LD block of the significant SNPs in the genome. These candidate genes are listed in Table S2, offering valuable resources for future investigations through functional validations in sorghum.

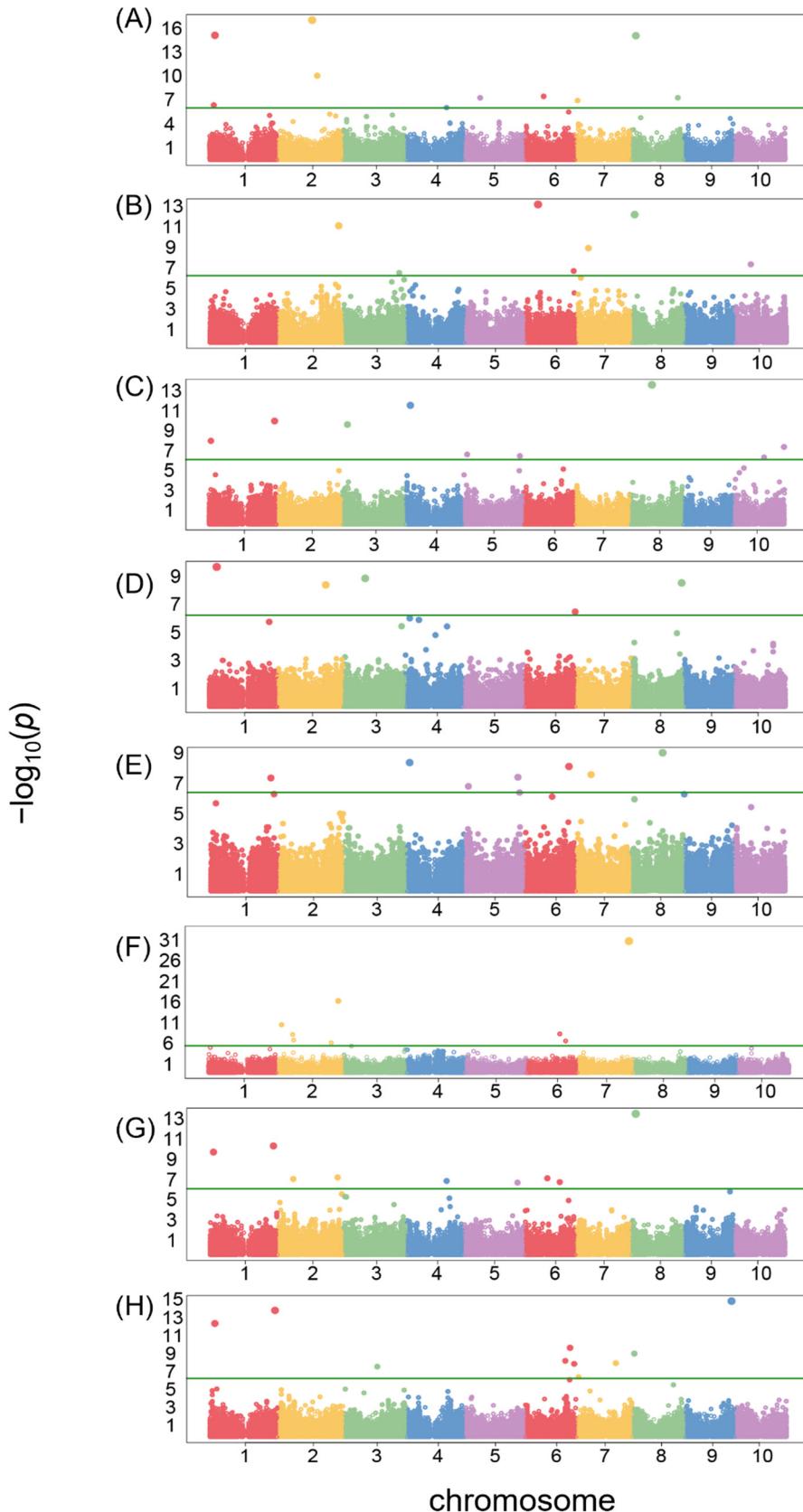


Figure 6. Manhattan plots of GWAS results showing the significant SNPs associated with eight phenotypic traits across the genome. The traits included: (A) Area size; (B) Brightness; (C) Circularity; (D) Distance between IS and CG; (E) Length; (F) Length to width ratio; (G) Perimeter length; (H) Width. The dots in the Manhattan plot represent SNP markers. The green line corresponds to a Bonferroni-corrected p -value threshold of 1.7E-7 ($-\log_{10}(p) = 6.8$).

4. Discussion

Seed morphology plays a key role in biological and ecological processes such as seed dormancy, germination, dispersal, persistence, evolution, and adaptation [37]. Despite its versatility, high-stress tolerance, and diverse applications as grain, forage, and biomass [38], sorghum seed morphology remains relatively unexplored. Correlation analysis of mini core and Senegalese accessions identified significance among the traits, identical to the patterns observed in previous studies with Senegalese germplasm [29]. Both PCA plots and partial contribution analyses yielded highly similar results, strengthening the consistency of these findings [29]. The observed consistency in correlation patterns across both studies could be attributed to the overlap of some Senegalese accessions. However, analyzing just the mini-core accessions in this study yielded nearly identical results, suggesting broader generalizability of these findings (data available in Supplementary Data S1).

Furthermore, recent studies identified potential linkages between sorghum seed morphology traits and host resistance against fungal pathogens. A recent study identified significant negative correlations between grain mold severity and seed weight in sorghum [39]. Similarly, Ahn *et al.* [29] identified correlations between seed morphology traits (circularity and the distance between IS and CG) and the formation of spots on seedling leaves. These spots appeared when seedlings were inoculated with *Sporisorium reilianum*, a causal pathogen causing head smut, and submerged under water [40]. Though spotted plants are considered susceptible, the cause of the spots is unclear. They might be a direct result of fungal infection or, alternatively, a defense mechanism triggered by the seedlings. Regardless of their origin, the association between spot appearance rate and seed morphology traits is notable. While no statistically significant links between seed morphology and anthracnose/downy mildew susceptibility were found except for IS and CG, five out of eight tested traits exhibited associations with head smut susceptibility. The head smut data applied in this study is from syringe needle inoculation (hypodermic injection), with resistance/susceptibility confirmed by the presence/absence of infected heads in mature plants [19]. The observed correlations between seed morphology and head smut resistance might be rooted in the distinct infection processes of *S. reilianum*. Unlike anthracnose caused by *Colletotrichum sublineola*, which involves direct contact infection by conidia, head smut relies on systemic fungal growth originating from soilborne spores infecting plants during seed germination and seedling emergence [41]. This suggests that certain seed morphological traits may influence plant structures or defenses that impact internal fungal spread, but the precise mechanism remains unknown.

Our GWAS analysis revealed over 100 candidate genes linked to seed morphology traits (Table S2). Intriguingly, several genes with similar functions appeared as top candidates for multiple traits, suggesting shared genetic influences as suggested in correlation analysis. For example, UDP-glycosyltransferases ranked among the top hits for area size, circularity, and distance between IS and CG, indicating their potential impact on seed size and shape. In rice, UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection [42]. Genes associated with zinc finger motifs emerged as candidates for length and LWR, indicating their potential influence on grain size and shape. This is further supported by the C2H2 zinc-finger protein LACKING RUDIMENTARY GLUME 1 (LRG1) in rice, which directly regulates spikelet formation and consequently impacts grain size and yield [43]. Likewise, F-box genes associated with LWR and brightness support findings in rice where the F-box protein FBX206 and OVATE family proteins network modulate brassinosteroid biosynthesis for grain size control [44]. Furthermore, leucine-rich repeat protein genes linked to length and brightness and the cytochrome P450 superfamily associated with area size and circularity support their roles in plant development, stress responses, and metabolism [29,45,46,47,48]. Notably, GW10, a P450 subfamily member, regulates grain size and number in rice [49]. These examples, alongside the entire candidate gene list in Table S2, offer valuable resources for future research and potential candidates for breeding programs aiming to improve sorghum seed morphology and grain yield. Multiple genes previously identified as top candidates in our earlier work [29] resurfaced as key genes in this study. This repeated association strongly suggests their genuine involvement in shaping seed morphology traits. These genes warrant

particular attention for further functional validation studies to explore their roles in determining seed morphology.

5. Conclusion

By analyzing 245 mini-core and Senegalese accessions, we identified significant phenotypic diversity and correlations among the seed morphology traits. Furthermore, we investigated the potential linkage between seed morphology and disease resistance in sorghum. Seed morphology traits associated with head smut resistance through both seedling leaf spot appearance rate [29] and systemic syringe inoculation response suggest a potential link between morphology and sorghum response. Furthermore, GWAS analysis identified SNPs potentially linked to seed morphology that can be targeted for functional validation using gene-editing tools like CRISPR-Cas9.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Supplementary data (Excel spreadsheets in a zip file) S1: <Phenotypic data> and Supplementary Tables 1 and 2.

Authors contributions E.A.: Conceptualization, Project administration, Funding acquisition, Supervision, Data Curation, Investigation, Writing - Original Draft. S.P.: Software, Supervision, Validation, Formal analysis, Writing - Review & Editing. Z.H.: Software, Formal analysis, Writing - Review & Editing. V.E.: Methodology, Writing - Review & Editing. M.C.: Writing - Review & Editing. Y.L.: Methodology, Writing - Review & Editing. L.K.P.: Methodology, Writing - Review & Editing. C.M.: Conceptualization, Supervision, Writing - Review & Editing.

Funding: This research was funded by AFRI, NIFA, USDA grant number 20156800423492 and by The Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet, a United States Agency for International Development Cooperative grant number AID-OAA-A-13-00047 with title name: Enabling marker assisted selection for sorghum disease resistance in Senegal and Niger. EA, SP, ZH, and LKP were supported by the U.S. Department of Agriculture, Agricultural Research Service. Mention of any trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U. S. Department of Agriculture. USDA is an equal opportunity provider and employer, and all agency services are available without discrimination.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We express our gratitude to the reviewers for their valuable feedback.

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

1. FAOSTAT. Crops and livestock products. Available online: <http://www.fao.org/faostat/en/#data/QCL> (accessed on 16 September 2023).
2. Chadalavada, K.; Kumari, B.D.R.; Kumar, T.S. Sorghum mitigates climate variability and change on crop yield and quality. *Planta* **2021**, *253*, 113.
3. Abreha, K.B.; Enyew, M.; Carlsson, A.S.; Vetukuri, R.R.; Feyissa, T.; Motlhaodi, T.; Ng'uni, D.; Geleta, M. Sorghum in Dryland: Morphological, Physiological, and Molecular Responses of Sorghum under Drought Stress. *Planta* **2022**, *255*, 1–23.
4. Enyew, M.; Feyissa, T.; Carlsson, A.S.; Tesfaye, K.; Hammehag, C.; Seyoum, A.; Geleta, M. Genome-wide analyses using multi-locus models revealed marker-trait associations for major agronomic traits in *Sorghum bicolor*. *Front. Plant Sci.* **2022**, *13*, 999692.
5. Jankowski, K.J.; Sokólski, M.M.; Dubis, B.; Załuski, D.; Szempliński, W. Sweet sorghum—Biomass production and energy balance at different levels of agricultural inputs. A six-year field experiment in north-eastern Poland. *Eur. J. Agron.* **2020**, *119*, 126119.
6. Khalifa, M.; Eltahir, E.A.B. Assessment of global sorghum production, tolerance, and climate risk. *Front. Sustain. Food Syst.* **2023**, *7*, 1184373.

7. USDA FAS. Sorghum explorer. Available online: <https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=0459200> (accessed on 1 January 2024).

8. Demarco, P.A.; Mayor, L.; Rotundo, J.L.; Prasad, P.V.V.; Morris, G.P.; Fernandez, J.A.; Tamagno, S.; Hammer, G.; Messina, C.D.; Ciampitti, I.A. Retrospective study in U.S. commercial sorghum breeding: II. Physiological changes associated to yield gain. *Crop Sci.* **2023**, *63*, 867–878.

9. Dille, J.A.; Stahlman, P.W.; Thompson, C.R.; Bean, B.W.; Soltani, N.; Sikkema, P.H. Potential Yield Loss in Grain Sorghum (*Sorghum bicolor*) with Weed Interference in the United States. *Weed Technol.* **2020**, *34*, 624–629.

10. Tanwar, R.; Panghal, A.; Chaudhary, G.; Kumari, A.; Chhikara, N. Nutritional, Phytochemical and Functional Potential of Sorghum: A Review. *Food Chem. Adv.* **2023**, *3*, 100501.

11. Frankowski, J.; Przybylska-Balcerek, A.; Stuper-Szablewska, K. Concentration of Pro-Health Compound of Sorghum Grain-Based Foods. *Foods* **2022**, *11*, 216.

12. Khalid, W.; Ali, A.; Arshad, M.S.; Afzal, F.; Akram, R.; Siddeeq, A.; Kousar, S.; Rahim, M.A.; Aziz, A.; Maqbool, Z. Nutrients and bioactive compounds of *Sorghum bicolor* L. used to prepare functional foods: A review on the efficacy against different chronic disorders. *Int. J. Food Prop.* **2022**, *25*, 1045–1062.

13. Taylor, J.R.N.; Duodu, K.G. Resistant-Type Starch in Sorghum Foods—Factors Involved and Health Implications. *Starch* **2022**, *75*, 2100296.

14. Teferra, T.F.; Awika, J.M. Sorghum as a healthy global food security crop: Opportunities and challenges. Available online: <https://www.cerealsgrains.org/publications/cfw/2019/September-October/Pages/CFW-64-5-0054.aspx> (accessed on 16 September 2023).

15. Mathur, S.; Umakanth, A.V.; Tonapi, V.A.; Sharma, R.; Sharma, M.K. Sweet sorghum as biofuel feedstock: Recent advances and available resources. *Biotechnol. Biofuels* **2017**, *10*, 1–19.

16. Osman, A.; Abd El-Wahab, A.; Ahmed, M.F.E.; Buschmann, M.; Visscher, C.; Hartung, C.B.; Lingens, J.B. Nutrient Composition and In Vitro Fermentation Characteristics of Sorghum Depending on Variety and Year of Cultivation in Northern Italy. *Foods* **2022**, *11*, 3255.

17. Zarei, M.; Amirkolaei, A.K.; Trushenski, J.T.; Sealey, W.M.; Schwarz, M.H.; Ovissipour, R. Sorghum as a Potential Valuable Aquafeed Ingredient: Nutritional Quality and Digestibility. *Agriculture* **2022**, *12*, 669.

18. Upadhyaya, H.D.; Pundir, R.P.S.; Dwivedi, S.L.; Gowda, C.L.L.; Reddy, V.G.; Singh, S. Developing a mini core collection of sorghum for diversified utilization of germplasm. *Crop Sci.* **2009**, *49*, 1769–1780.

19. Ahn, E.; Hu, Z.; Perumal, R.; Prom, L.K.; Odvody, G.; Upadhyaya, H.D.; Magill, C.W. Genome wide association analysis of sorghum mini core lines regarding anthracnose, downy mildew, and head smut. *PLoS ONE* **2019**, *14*, e0216671.

20. Upadhyaya, H.D.; Vetriventhan, M.; Azevedo, V.C.R. Variation for photoperiod and temperature sensitivity in the global mini core collection of sorghum. *Front. Plant Sci.* **2021**, *12*, 571243.

21. Cuevas, H.E.; Prom, L.K.; Rosa-Valentin, G. Population Structure of the NPGS Senegalese Sorghum Collection and Its Evaluation to Identify New Disease Resistant Genes. *PLoS ONE* **2018**, *13*, e0191877.

22. Ahn, E.; Fall, C.; Prom, L.K.; Magill, C. Genome-Wide Association Study of Senegalese Sorghum Seedlings Responding to a Texas Isolate of *Colletotrichum Sublineola*. *Sci. Rep.* **2022**, *12*, 13025.

23. Ahn, E.; Prom, L.K.; Hu, Z.; Odvody, G.; Magill, C. Genome-wide Association Analysis for Response of Senegalese Sorghum Accessions to Texas Isolates of Anthracnose. *Plant. Genome* **2021**, *14*, e20097.

24. Cervantes, E.; Martín, J.J.; Saadaoui, E. Updated Methods for Seed Shape Analysis. *Scientifica* **2016**, *2016*, 5691825.

25. Dong, R.; Guo, Q.; Li, H.; Li, J.; Zuo, W.; Long, C. Estimation of morphological variation in seed traits of *Sophora moorcroftiana* using digital image analysis. *Front. Plant Sci.* **2023**, *14*, 1185393.

26. Wang, L.; Upadhyaya, H.D.; Zheng, J.; Liu, Y.; Singh, S.K.; Gowda, C.L.L.; Kumar, R.; Zhu, Y.; Wang, Y-H.; Li, J. Genome-wide association mapping identifies novel panicle morphology loci and candidate genes in sorghum. *Front. Plant Sci.* **2021**, *12*, 743838.

27. Sakamoto, L.; Kajiyama-Kanegae, H.; Noshita, K.; Takanashi, H.; Kobayashi, M.; Kudo, T.; Yano, K.; Tokunaga, T.; Tsutsumi, N.; Iwata, H. Comparison of shape quantification methods for genomic prediction, and genome-wide association study of sorghum seed morphology. *PLoS ONE* **2019**, *14*, e0224695.

28. Chopra, R.; Burow, G.; Burke, J.J.; Gladman, N.; Xin, Z. Genome-wide association analysis of seedling traits in diverse Sorghum germplasm under thermal stress. *BMC Plant Biol.* **2017**, *17*, 12.

29. Ahn, E.; Botkin, J.; Ellur, V.; Lee, Y.; Poudel, K.; Prom, L.K.; Magill, C. Genome-Wide Association Study of Seed Morphology Traits in Senegalese Sorghum Cultivars. *Plants* **2023**, *12*, 2344.

30. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **2011**, *6*, e19379.

31. Hu, Z.; Olatoye, M.O.; Marla, S.; Morris, G.P. An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum Genotypes. *Plant Genome* **2019**, *12*, 180044.

32. Upadhyaya, H.D.; Wang, Y.H.; Gowda, C.L.L.; Sharma, S. Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection. *Theor. Appl. Genet.* **2013**, *126*, 2003–2015.

33. Browning, B.L.; Browning, S.R. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **2016**, *98*, 116–126.

34. Lipka, A.E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399.

35. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S.; Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient genomewide association studies. *PLoS Genet.* **2016**, *12*, e1005767.

36. McCormick, R.F.; Truong, S.K.; Sreedasyam, A.; Jenkins, J.; Shu, S.; Sims, D.; Kennedy, M.; Amirebrahimi, M.; Weers, B.D.; McKinley, B.; et al. The *Sorghum bicolor* reference genome: Improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* **2018**, *93*, 338–354.

37. Diantina, S.; McGill, C.; Millner, J.; Nadarajan, J.; W. Pritchard, H.; Clavijo McCormick, A. Comparative Seed Morphology of Tropical and Temperate Orchid Species with Different Growth Habits. *Plants* **2020**, *9*, 161.

38. Takanashi, H. Genetic Control of Morphological Traits Useful for Improving Sorghum. *Breed. Sci.* **2023**, *73*, 57–69.

39. Prom, L.K.; Ahn, E.; Isakeit, T.; Magill, C. Correlations among grain mold severity, seed weight, and germination rate of sorghum association panel lines inoculated with *Alternaria alternata*, *Fusarium thapsinum*, and *Curvularia lunata*. *J. Agric. Crops* **2022**, *8*, 7–11.

40. Ahn, E.; Prom, L.K.; Fall, C.; Magill, C. Response of Senegalese Sorghum Seedlings to Pathotype 5 of *Sporisorium reilianum*. *Crops* **2022**, *2*, 142–153.

41. Wright, P.J.; Fullerton, R.A.; Koolaard, J.P. Fungicide control of head smut (*Sporisorium reilianum*) of sweetcorn (*Zea mays*). *New Zeal J Crop Hortic Sci* **2006**, *34*, 23–26.

42. Dong, N.Q.; Sun, Y.; Guo, T.; Shi, C.L.; Zhang, Y.M.; Kan, Y.; Xiang, Y.H.; Zhang, H.; Yang, Y.B.; Li, Y.C. UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nat. Comm.* **2020**, *11*, 1–16.

43. Xu, Q.K.; Yu, H.P.; Xia, S.S.; Cui, Y.J.; Yu, X.Q.; Liu, H.; Zeng, D.L.; Hu, J.; Zhang, Q.; Gao, Z.Y. The C2H2 zinc-finger protein *LACKING RUDIMENTARY GLUME 1* regulates spikelet development in rice. *Sci. Bull.* **2020**, *65*, 753–764.

44. Sun, X.; Xie, Y.; Xu, K.; Li, J. Regulatory networks of the F-box protein FBX206 and OVATE family proteins modulate brassinosteroid biosynthesis to regulate grain size and yield in rice. *J. Exp. Bot.* **2023**, erad397.

45. Osakabe, Y.; Maruyama, K.; Seki, M.; Satou, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Leucine-Rich Repeat Receptor-Like Kinase1 Is a Key Membrane-Bound Regulator of Abscisic Acid Early Signaling in Arabidopsis. *Plant Cell* **2005**, *17*, 1105–1119.

46. Shahollari, B.; Vadassery, J.; Varma, A.; Oelmüller, R. A Leucine-Rich Repeat Protein Is Required for Growth Promotion and Enhanced Seed Production Mediated by the Endophytic Fungus *Piriformospora indica* in *Arabidopsis thaliana*: Growth Regulation by P. Indica. *Plant J.* **2007**, *50*, 1–13.

47. Xu, J.; Wang, X.; Guo, W. The Cytochrome P450 Superfamily: Key Players in Plant Development and Defense. *J. Integr. Agric.* **2015**, *14*, 1673–1686.

48. Hansen, C.C.; Nelson, D.R.; Møller, B.L.; Werck-Reichhart, D. Plant Cytochrome P450 Plasticity and Evolution. *Mol. Plant* **2021**, *14*, 1244–1265.

49. Zhan, P.; Wei, X.; Xiao, Z.; Wang, X.; Ma, S.; Lin, S.; Li, F.; Bu, S.; Liu, Z.; Zhu, H. GW10, a member of P450 subfamily regulates grain size and grain number in rice. *Theor. Appl. Genet.* **2021**, *134*, 3941–3950.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.