BMC Genomics

RESEARCH

Open Access

Multi-locus genome-wide association mapping for major agronomic and yield-related traits in sorghum (*Sorghum bicolor* (L.) moench) landraces

Addisu Getahun[1,2,3], Admas Alemu[4], Habte Nida[5] and Adugna Abdi Woldesemayat[1,2*]

Abstract

Background Sorghum is a vital cereal crop for over 750 million people, ranking 5th globally. It has multiple purposes, including food, feed, and biofuels, and is essential in Ethiopia, which has a rich genetic diversity of various agroecological zones.

Objective Explore marker-trait associations (MTAs) to identify quantitative trait nucleotides (QTNs) and new candidate genes associated with agronomic and yield contributing traits in Ethiopian sorghum landraces using multi-locus GWAS models to assist the genomic-assisted breeding strategies.

Method This study investigates the genetic basis of agronomic traits in Ethiopian sorghum landraces through multi-locus Genome-Wide Association Studies (ML-GWAS). 216 landraces, improved varieties, and check cultivars were obtained from the Ethiopian Biodiversity Institute and the National Sorghum Improvement Program for this study. The experiment was conducted over two cropping seasons, employing an α-lattice design for phenotyping key traits such as days to flowering, days to maturity, plant height, seed number per plant, grain yield, and thousand seed weight. A mixed linear model (MLM) was used to analyze the phenotypic data and estimate the genetic parameters including variances and the broad sense heritability. GBS with the *ApeKI* restriction enzyme provided 50,165 high-quality SNP markers. The six ML-GWAS models identified significant QTNs with a LOD score threshold value of ≥ 4.0. The analysis revealed major QTNs associated with traits across multiple chromosomes, supported by a stringent filtering criterion that ensured reliability. Co-localization with known QTLs was explored using the Sorghum QTL Atlas database and candidate genes within significant QTN regions, providing the genetic architecture influencing agronomic performance were identified via the Phytozome platform using the biomaRt package.

Result Pearson correlation analysis revealed significant associations among most traits, with p-values less than 0.0001, except for grain yield per plant which showed lower correlations with other traits. Genetic variability analysis indicated that days to flowering exhibited high heritability (0.7) and genetic advance (19.6%) as percent of mean, suggesting strong genetic control, while grain yield displayed extremely low h^2 (0.003). A total of 351,692 SNP

*Correspondence:
Adugna Abdi Woldesemayat
adugna.abdi@aastu.edu.et

Full list of author information is available at the end of the article

markers were identified across 10 sorghum chromosomes from 216 Ethiopian sorghum landraces, and we have been refining this to 50,165 filtered SNPs. Manhattan plots indicated significant marker-trait associations (MTAs) across multiple chromosomes, particularly for days to flowering and plant height. Significant QTNs were associated with key traits including flowering time, plant height, and grain yield. ML-GWAS identified 176 QTNs with varying LOD scores and phenotypic effects. Multiple genes linked to these QTNs highlight the complexity of genetic interactions of studied traits with 36 unique and 12 major QTNs. Notable SNP markers were concentrated on chromosomes 1, 2, and 3, reinforcing the importance of these regions for breeding efforts. Candidate gene analysis revealed key genes regulating flowering time, stress response, and yield traits, which could serve as targets for genetic enhancement. In our study, key candidate genes have been successfully identified, these are regulating flowering time, maturity, and stress resilience. Genes such as *Sobic.001G196700* and *Sobic.002G183400* are identified as critical regulators of floral development. The stress-responsive gene *Sobic.005G176100* (a mannose-6-phosphate isomerase), emphasizes the importance of resilience in sorghum cultivation under adverse conditions. Additionally, *Sobic.003G324400* and *Sobic.004G178300* are essential for regulating plant height and seed weight, making them valuable for yield enhancement breeding programs.

Conclusion This study enhances our understanding of the genetic diversity of Ethiopian sorghum landraces, crucial for breeding programs. It identifies key QTNs and candidate genes associated with important agronomic traits, offering insights for marker-assisted and genomic-assisted breeding. The ML-GWAS models highlight the genetic complexity of flowering time and grain yield traits, emphasizing the need for targeted breeding efforts to maximize sorghum productivity.

Keywords Sorghum bicolor, Agronomic traits, Genetic diversity, ML-GWAS, Quantitative trait nucleotides, Candidate genes

Introduction

Sorghum (*Sorghum bicolor* (L.)Moench) is an annual C4 plant belonging to the family Poaceae under the Andropogoneae tribe [1]. It is the 5th most important cereal crop globally and a dietary staple for over 750 million people mainly living in semi-arid regions (Asia and Africa) [2]. It is a versatile cereal crop cultivated globally for diverse applications, including food, feed, and biofuel production. It is also a major crop in many regions worldwide, for instance; Asia, Africa, Australia, and the USA, and the 5 top sorghum-producing countries are; the United States (25%), India (21.5%), Mexico (11%), China (9%) and Nigeria (7%), and together these five countries account for 73% of total world production, and prized for its adaptability and nutritional value [3].

Sorghum is a vital staple food for millions, particularly in sub-Saharan Africa [4], in Ethiopia it serves as a staple food and livelihood source for millions [5]. It is a key energy source, of protein, vitamins, and minerals for many households [6]. Additionally, by-products of sorghum are used for animal feed, construction, fencing, and broom manufacturing [7].

Sorghum is a commercially significant crop utilized for producing lager beer, gluten-free food items, phytochemicals, sweet-stalked varieties, and biomass for biofuels, particularly in regions like Asia and Africa [8]. The average production reaches 23.35 million metric tons from an area of 23.14 million hectares, yielding an average productivity of 1.01 tons per hectare, and Ethiopia's national average, sorghum productivity is 2.1 tons/ha which is very low compared to the global average of 3.2 tons/ha due to abiotic stress, biotic stress, soil fertility decline, and lack of high-yielding sorghum varieties [9].

Despite its economic importance, sorghum yield is affected by various biotic and abiotic challenges, including diseases, insect pests, weeds, nutrient deficiencies, aluminum toxicity, drought, salinity, waterlogging, and high temperatures [10]. Moreover, drought contributes to genetic erosion in sorghum, causing the loss of many landraces due to crop failures resulting from extreme drought conditions [11]. This has prompted numerous initiatives to explore the genetic and physiological mechanisms enabling crop drought resistance [10].

Ethiopia is recognized as a center of origin and diversity for sorghum, hosting a wealth of genetic variation for numerous traits [12]. Ethiopia boosts a rich genetic diversity of sorghum landraces adapted to various agroecological zones, ranging from lowlands to highlands [13]. A diverse set of popular Ethiopian sorghum landraces has been collected and preserved with a wealth of genetic resources and novel alleles for a range of agronomic, yield, and yield-related traits [14]. This diverse germplasm offers valuable opportunities for gaining insights into the genetic architecture of key traits, which can enhance breeding programs for more efficient genetic improvement of sorghum. Understanding the genetic diversity of landraces is crucial to identifying novel QTLs and genes [15].

Yield is a polygenic trait affected by multiple genes and factors, such as plant phenology, morphology, and

physiological traits [16]. Unraveling the genetic basis of these traits is crucial for effective breeding thereby improving crop efficiency and resilience in a shifting climate [17]. Genomics-assisted breeding is an innovative approach that utilizes modern molecular tools and genomic information to improve the accuracy and efficiency of conventional plant breeding, In recent decades, substantial efforts have been devoted to genomic-assisted breeding in sorghum and other cereal crops [18]. Initially, genomic regions associated with agronomic traits in sorghum were identified using bi-parental linkage mapping methods [18]. This method often leads to low mapping resolution, restricted allelic diversity, and QTLs that are specific to certain populations [19], hindering the conversion of QTL discoveries into actionable strategies for plant breeding [20].

Genome-wide association study (GWAS) enables high-resolution QTL mapping by leveraging a diverse array of alleles across numerous accessions, making it a crucial tool for the genetic analysis of complex traits [21]. Sorghum is especially suitable for linkage mapping because of its moderate linkage disequilibrium and self-pollinating nature [21]. Recent investigations have applied GWAS in sorghum to examine the genetic regulation of several key traits, including flowering time [22], plant height, length of panicles, degree of panicle exertion, number of tillers, and seed count [22] as well as culm length and the number of panicles [23], inflorescence components [24], grain fill duration, panicle weight, and harvest index, and grain yield [25]. However, numerous studies encountered challenges, especially due to their dependence on germplasm that had undergone the sorghum conversion process.

Multi-locus GWAS models have emerged as powerful tool and that is useful for identifying Quantitative Trait Nucleotide (QTN) detection rather than QTLs and SNP markers effect estimation, including mrMLM [26], FASTmrMLM [27], FASTmrEMMA [28], ISIS EMBLASSO [29], pLARmEB [30], and pKWmEB [31]. These approaches have successfully uncovered the genetic basis of important traits in various crops, including maize [32], rice [33], barley [34], and wheat [35]. The objectives of this study were to investigate marker trait associations (MTA) via ML-GWAS models to identify important QTNs, and candidate genes associated with agronomic, yield, and yield-related traits in Ethiopian sorghum landraces to promote genomic-assisted breeding (GAB) techniques and strategies [36].

Result

Pearson correlation analysis of agronomic & Yield-Related traits

The Pearson correlation probability of sorghum agronomic and yield association trait (Table S1) showed the p-values displayed are all less than 0.0001, except for grain yield per plant (GY) with days to flowering (DF): $p = 0.1567$, grain yield per plant (GY) with days to maturity: $p = 0.5878$, seed number per plant (SNPP) with thousand-seed weight (TSW): $p = 0.0549$, and the p-values less than 0.0001 indicate a highly significant correlation between the corresponding traits (Table S1).

Distribution of SNPs across Sorghum genome

A total of 351,692 SNP markers were identified across 10 chromosomes in 216 Ethiopian sorghum landraces and improved varieties. The dataset was filtered to exclude SNPs with MAF ≥ 0.05 (5%) yielding a robust final dataset of 50,165 SNPs. The genome-wide marker density plot (Fig. 1) showed that markers from the study panel were distributed across the sorghum genome. Further, examination of the genome-wide marker distribution revealed that the SNP markers were evenly dispersed. This comprehensive SNP dataset, with its even distribution and varying marker densities across the genome, provided a robust foundation for the subsequent genome-wide association analyses.

The three panels represent: (Fig. 1a) R-square (r^2); (Fig. 1b) MAF; (Fig. 1c) SNP (Single Nucleotide Polymorphism) marker heterozygosity across sorghum genomes from bottom to top respectively. The x-axis indicates the marker number, while the y-axis displays the respective values for each parameter, highlighting variations across the markers.

The R^2 values represent the goodness of fit of the statistical model used to associate the genetic markers with the sorghum traits of interest, which include days to flowering, days to maturity, plant height, number of seeds per plant, grain yield, and thousand-seed weight. The higher the R^2 the stronger the association between the genetic markers and the phenotypic traits. Regions with high R^2 values indicate genomic areas that harbor significant QTLs or MTA for these sorghum agronomic traits. The MAF values in the middle panel provide information on genetic diversity.

The genomic data in Fig. 2 provides a good foundation for investigating the genetic architecture of important agronomic and yield-related traits, such as days to flowering, plant height, grain yield, and thousand-seed weight. The LD, marker density, and genetic distance information can inform the design and analysis of QTN mapping or GWAS experiments to identify marker-trait associations. In Fig. 2, The histograms display the frequency of genetic distances between markers, and they indicate a relatively even distribution of marker spacing, which is desirable for GWAS and QTN mapping.

The color gradient (Fig. 3) helps quickly identify chromosomal regions with higher or lower SNP densities. Regions with higher SNP densities (represented by the

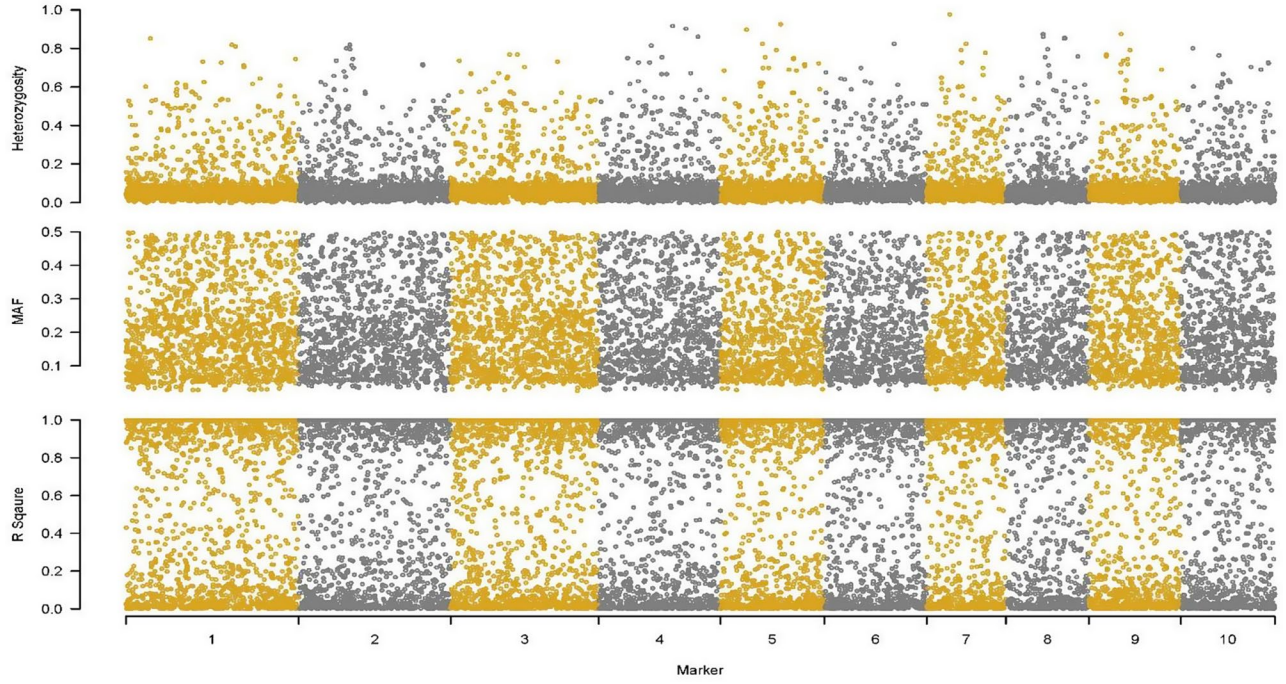

Fig. 1 Distribution of SNP marker heterozygosity, Minor Allele Frequency (MAF), and R-square (r^2) across sorghum SNP markers, each parameter represented in separate panels

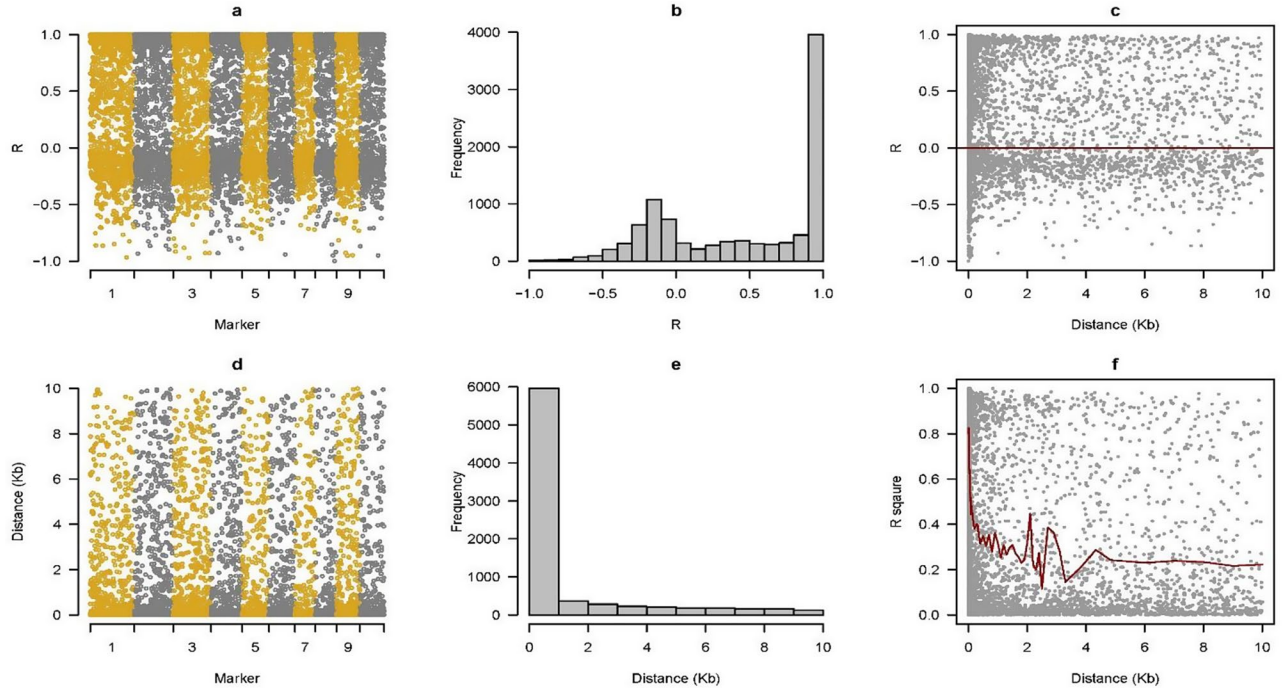

Fig. 2 SNP integrity, analysis of genetic relationships in sorghum. The figure shows, **(a)** R values across markers; **(b)** a histogram showing the frequency distribution of R values; **(c)** a scatter plot of R values against distance (kb); **(d)** SNP distance and distribution across markers; **(e)** histogram of frequency suggests an adequate marker density for QTN mapping, covering the entire sorghum genome; **(f)** scatter plot illustrating R-square (r^2) values against distance, with a trend line showing overall patterns

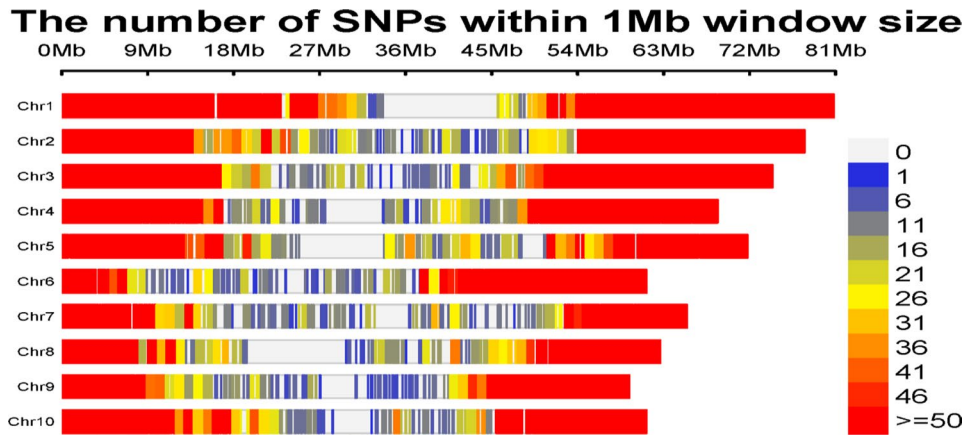

Fig. 3 Distribution of single nucleotide polymorphisms (SNPs) markers across 10 sorghum chromosomes within a 1 Mb window size

red and orange shades) indicated areas of the genome that are likely to have greater genetic diversity and more potential marker-trait associations for the agronomic traits of interest, such as days to flowering, days to maturity, plant height, seed number per plant, grain yield, and thousand-seed weight. Conversely, the lighter-colored regions (blues and greens) suggest chromosomal segments with lower SNP densities, which may require additional marker development or optimization of genotyping strategies to ensure sufficient genome coverage for comprehensive QTN mapping and GWAS analyses.

Each bar represents a chromosome, with color intensity indicating the number of SNPs, ranging from 0 to over 50. The color scale on the right provides a key for interpreting SNP density. This visualization of the number of SNPs within a 1 Mb window size in the genome is relevant for understanding the genomic architecture and marker density for conducting QTN mapping and GWAS on agronomic and yield traits like days to flowering, days to maturity, plant height, seed number per plant, grain yield, and thousand-seed weight.

As shown in Fig. 3, the color scheme used to represent the number of SNPs within the 1 Mb window size across the sorghum chromosomes. The color gradient is as follows; White (0 SNPs), Light blue (1–6 SNPs), Dark blue (6–11 SNPs), Light green (11–16 SNPs), Dark green (16–21 SNPs), Light yellow (21–26 SNPs), Dark yellow (26–31 SNPs), Light orange (31–36 SNPs), Dark orange (36–41 SNPs), Light red (41–46 SNPs), and Dark red (≥ 50 SNPs). This color coding allows for a visual representation of the SNP density variation across the different chromosomal regions. The darker the color, the higher the number of SNPs. A gradient of SNP counts within the 1 Mb windows, ranging from 0 SNPs (white) to greater than/equal to 50 SNPs (dark red) indicates that the sorghum genome has varying levels of SNP density across different chromosomal regions (Fig. 3). Some chromosomes, such as Chr1 and 10, appear to have higher overall SNP densities compared to other chromosomes like Chr3 and 6. This suggests that the genomic architecture and recombination rates may differ across the chromosomes. The uneven distribution of SNP density across the genome has important implications for QTN mapping and GWAS analysis.

Each plot (Fig. 4) displays the negative logarithm of the p-values (-log10) on the y-axis against the chromosome numbers on the x-axis. The X-axis (Chromosomes) is represented by 1 to 10 chromosome numbers, indicating the locations of SNPs across the genome. Y-Axis (-log10(p)) showed stronger associations between SNPs and traits of interest. The threshold for significance can be determined by applying corrections for multiple testing, such as the Bonferroni correction and false discovery rate (FDR) [37]. The Manhattan plot showing p-values, MTA, and GWAS for agronomic, yield, and yield-related traits is indicated in Fig. 4.

Pink SNPs also represent SNPs that are significant after applying the Bonferroni correction. The Blue SNPs show significant associations but may not pass the Bonferroni threshold line, indicating associations that could be considered significant under the FDR. The Bonferroni correction adjusts the p-value threshold by dividing the desired alpha level (0.05) by the SNPs. The FDR method controls the expected proportion of false discoveries among the rejected null hypotheses. A common threshold for significance using FDR is $q < 0.05$ [37].

The upper plot indicates several SNPs with significant associations, particularly in chromosomes 1, 5, and 8. The lower plot shows notable associations, especially on chromosomes 3 and 9. Like flowering time, several pink SNPs were evident, indicating they have significantly influenced maturation time.

QTNs identified by ML-GWAS

A multi-locus Q-Q plot (Fig. S1) is an effective tool for evaluating associations in GWAS providing insights into

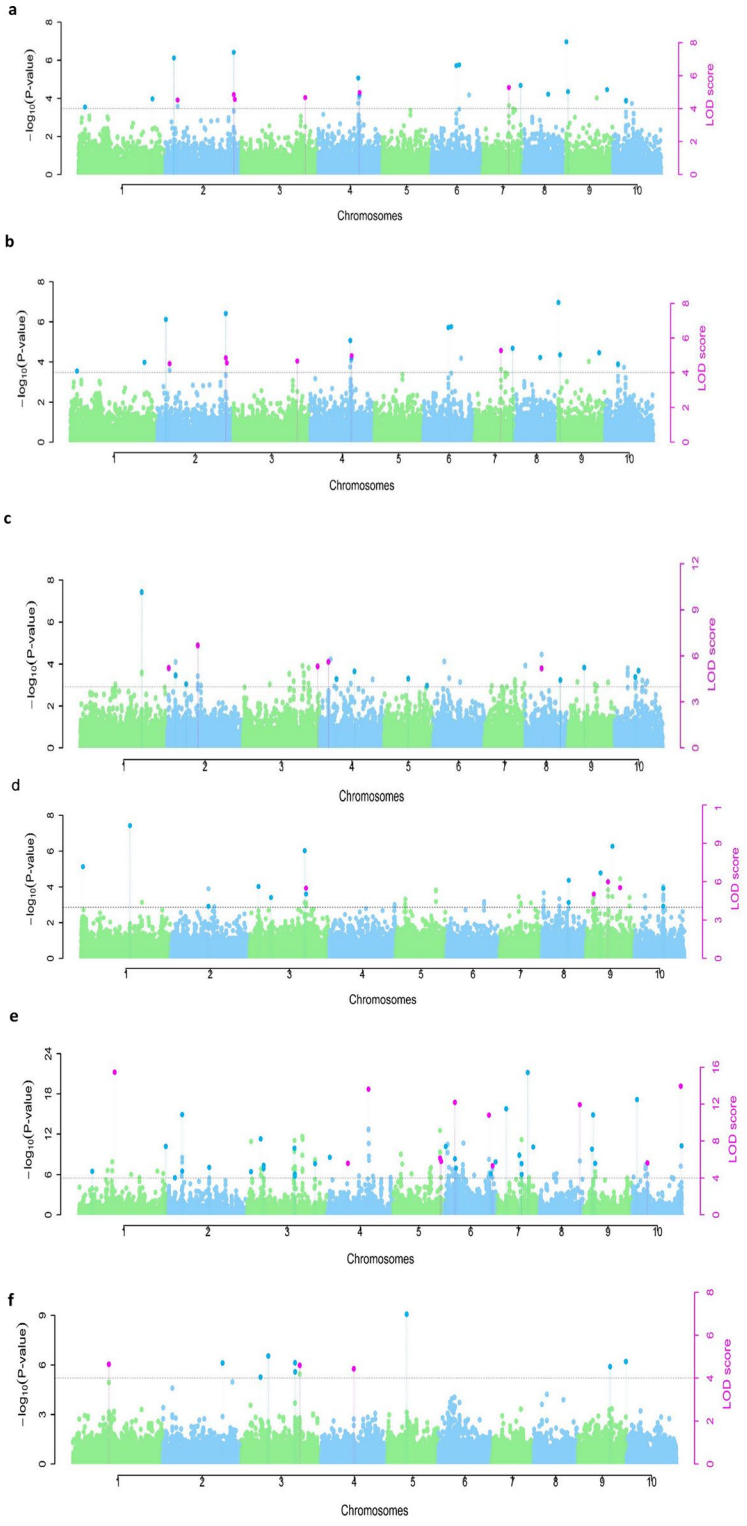

Fig. 4 Represents the Manhattan plot showing *p-values*, MTA, and GWAS for agronomic, and yield-related traits. *Note*: The plot represents the traits of; (**a**) days to flowering; (**b**) days to maturity; (**c**) plant height; (**d**) number of seeds per plant; (**e**) grain yield; (**f**) thousand seed weight from top to bottom, respectively. Each marker median of the − log10(p), mrMLM, FASTmrMLM, and FASTmrEMMA approaches was used to draw the Manhattan plot. The dots are indicated by colors QTNs. Pink dots with dotted vertical lines indicate all QTNs commonly identified by three approaches

the genetic architecture of traits such as days to flowering, days to maturity, plant height, seed number per panicle, grain yield, and thousand seed weight. Significant associations are indicated by points in the lower tail of the plot, showing that traits like plant height and grain yield are influenced by various genetic factors. Using the mrMLM model, 176 QTNs were identified with varying LOD scores and R-squared values, reflecting their effect on trait variability. Multiple genes linked to these QTNs highlight the complexity of genetic interactions in sorghum, with unique QTNs (36), major QTNs (12), and polygenic QTNs indicating intricate genetic relationships. The QTNs identified differ among traits, with the mrMLM model displaying a range of QTN effects from – 119.11 to 206.78 (Table 1), reflecting intricate genetic interactions. LOD scores, which assess the strength of these associations, are highest in the FASTmrMLM model, ranging from 4.04 to 36.13, indicating robust associations and potential QTLs worthy of further exploration. R-squared values, which measure the variance explained (PVE) by QTNs, range from 2.52 to 23.93% in the mrMLM model, showcasing varying impacts on trait variability. 176 QTNs were identified across all 6 models (Table 1), with mrMLM uncovering the most (42 QTNs) and FASTmrEMMA the least (13 QTNs), offering a thorough assessment of genetic variation linked to these traits. Specific traits, such as days to flowering, reveal multiple QTNs across chromosomes 1 to 10, each with different LOD (4.16 to 8.22) scores and R-squared values (Table 2). For example, *QTN S1_73955151* on chromosome 1 shows a LOD score between 4.09 and 4.59 and an R-squared value of 13.22–19.38%, indicating its significant role in the trait. Conversely, the plant height trait has a notable QTN (*S1_67415907*) with a high LOD score of 10.14 and an R-squared value of 7.63%, suggesting a strong association. In contrast, lower LOD scores associated with certain QTNs for grain yield highlight weaker associations that may require further investigation.

The QTN analysis results reveal a substantial genetic complexity across various traits, with a total of 176 QTNs identified using multiple models. The mrMLM model identified the most QTNs (42) with strong LOD scores (4.03 to 17.01) and moderate explanatory power (r^2 of 2.52 to 23.93). The FASTmrMLM model showed even higher LOD scores (up to 36.13) and a wide range of QTN effects (-299.26 to -129.76), accounting for significant trait variance (r^2 of 1.78 to 33.58). Other models like FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO.

Each QTN was uniquely identified (*S1_73955151*) and linked to a specific chromosome and base pair position. Higher LOD scores indicate stronger associations. For instance, the QTN for plant height (*S1_67415907*) has a notably high LOD score of 10.14, suggesting a strong genetic influence on this trait. R^2 values indicate the proportion of variance in the trait explained by the QTN. For example, the QTN associated with SNPP (*S1_1359747*) has an R^2 value ranging from 15.84 to 17.02%, which explains a significant portion of the trait variability. The number of Genes ± LD in a 1 Mb window size indicates the number of genes associated with each QTN, accounting for linkage disequilibrium (LD). QTN for days to flowering (*S2_6784036*) was associated with five genes (Table 2), reflecting the complexity of the genetic interactions involved. The varying number of associated genes per QTN highlights the complex interactions in the genetic architecture of these traits. For instance, certain QTNs were linked to multiple genes, suggesting that several genetic factors may influence a single trait.

The QTN mapping displays the locations of QTNs on each chromosome. Chromosome 1 contains multiple QTNs associated with days to flowering (DF), plant height (PH), and thousand seed weight (TSW). Notable QTNs include *S1_1359747* seed number per plant (SNPP), which correlates significantly with the plant height (PH) trait. Chromosome 2 hosts several critical QTNs, including *S2_116291* and *S2_54254801*, indicating their potential impact on drought resistance and growth. Chromosome 3 displays QTNs for grain yield (GY) and plant height (PH), with *S3_63127731* showing

Table 1 Summary of QTNs identified for the major agronomic, yield, and yield-related traits investigated in the present study using the six MrMLM models

Analysis Model	QTN effect	LOD score	r2	No. of QTNs
mrMLM	-119.11 to – 206.78	4.03–17.01	2.52–23.93	42
FASTmrMLM	-299.26 to – 129.76	4.04–36.13	1.78–33.58	30
FASTmrEMMA	-0.86 to – 417.50	4.01–15.48	2.86–15.95	13
pLARmEB	-241.30 to- 111.07	4.05–15.77	0.63–9.50	36
pKWmEB	-150.88 to – 145.22	4.06–11.94	4.16–19.59	27
ISIS EM-BLASSO	-84.17 to – 128.35	4.11–10.14	1.53–8.44	28
Total QTNs				***176***

Note: r^2 is the proportion of total phenotypic variation explained by each QTN. ***Abbreviations***: mrMLM, multi-locus random-SNP-effect mixed linear model MLM., FASTmrEMMA: factored spectrally transformed multi-locus random-SNP-effect efficient mixed-model association; pLARmEB: polygenic-background-control-based least angle regression plus empirical Bayes; FASTmrMLM: factored spectrally transformed multi-locus random-SNP-effect mixed linear model MLM; pKWmEB: polygenic-background-control-based Kruskal Wallis test plus empirical Bayes, ISIS EM-BLASSO: iterative sure independence screening EM-Bayesian LASSO

Table 2 The list of significant QTNs identified through ML-GWAS models for various agronomic and yield-related traits

Trait	QTN	Chr.	Alleles	Position (bp)	LOD	r2	Method	No of genes ± LD
DF	S1_73955151	SBI-01	T/C	73,955,151	4.09–4.59	13.22–19.38	1,5	2
	S2_6784036	SBI-02	T/G	6,784,036	4.16–8.22	2.68–5.62	1,2,4,6	5
	S3_68605546	SBI-03	A/C	68,605,546	4.06–5.22	4.63–9.30	1,4,5,6	1
	S4_56965234	SBI-04	T/C	56,965,234	4.38–5.85	3.10–7.08	1,2,4,5	3
	S6_49875883	SBI-06	G/C	49,875,883	6.60–6.66	3.39–7.53	5,6	2
	S7_57255603	SBI-07	A/T	57,255,603	4.35–6.21	4.15–4.78	1,2	2
	S8_53108778	SBI-08	C/T	53,108,778	4.88	6.07	1	1
	S9_1229695	SBI-09	G/A	1,229,695	5.03–8.05	3.38–4.54	2,4	3
	S10_7288565	SBI-10	C/G	7,288,565	4.49	7.01	4	1
DM	S1_76529338	SBI-01	A/G	76,529,338	4.45–9.13	5.14–19.59	1,2,3,4,5,6	5
	S2_54254801	SBI-02	G/A	54,254,801	5.41–4.88	3.72–3.96	2,6	2
	S4_7763150	SBI-04	C/T	7,763,150	4.11–5.01	3.17–4.64	2,6	2
	S5_623466	SBI-05	A/C	623,466	4.49–5.61	3.92–8.36	3,5	1
	S6_347535	SBI-06	G/T	347,535	4.04–7.52	2.03–5.55	1,2,4	3
	S8_979311	SBI-08	G/C	979,311	5.29–4.55	3.07–6.60	1,3	1
	S9_48811733	SBI-09	A/G	48,811,733	4.01–5.40	2.86–7.03	1,3,5	1
	S10_6360162	SBI-10	T/A	6,360,162	4.16–7.60	1.60-12.75	2,5,6	4
PH	S1_67415907	SBI-01	G/T	67,415,907	10.14	7.63	6	1
	S2_1166841	SBI-02	C/A	1,166,841	1.16–8.38	1.53–10.63	1,2,3,4,6	5
	S3_73241417	SBI-03	T/C	73,241,417	4.22–6.40	5.14–5.22	1,4	1
	S4_3891165	SBI-04	G/C	3,891,165	4.49–6.51	3.37–5.41	1,5,6	3
	S5_42621428	SBI-05	G/A	42,621,428	4.05–4.50	1.28–2.75	1,4	2
	S8_14284190	SBI-08	T/G	14,284,190	4.41–5.52	1.03–2.52	1,4	2
	S9_15670986	SBI-09	T/A	15,670,986	5.24	5.95	4	1
	S10_16170344	SBI-10	G/T	16,170,344	4.62–5.03	3.88–11.93	5,6	2
SNPP	S1_1359747	SBI-01	T/C	1,359,747	7.17–10.38	15.84–17.02	1	2
	S2_59151152	SBI-02	G/A	59,151,152	4.08	7.65	5	1
	S3_62232216	SBI-03	C/T	62,232,216	4.44–8.42	2.72–8.44	1,4,6	4
	S8_54218786	SBI-08	A/T	54,218,786	4.37–6.10	6.85–9.50	1,4	2
	S9_50050063	SBI-09	C/G	50,050,063	4.13–8.76	3.03-18.00	1,2,3,4,5,6	5
	S10_46922530	SBI-10	G/A	46,922,530	4.08–5.47	4.84–8.46	2,5	2
GY	S1_7317188	SBI-01	G/T	7,317,188	4.73–17.01	1.82–12.64	2,3,4	4
	S2_3431029	SBI-02	C/G	3,431,029	4.03–10.85	1.23–10.46	1,4,5	4
	S3_2016729	SBI-03	C/G	2,016,729	4.20–8.23	1.57–11.11	1,3,4,5,6	8
	S4_11012794	SBI-04	C/G	11,012,794	5.48–15.77	0.63–8.79	2,4,6	3
	S5_70402345	SBI-05	T/A	70,402,345	5.42–6.82	0.94–5.41	2,3,4,6	2
	S6_32754749	SBI-06	C/G	32,754,749	4.45–17.01	0.75–23.93	1,2,3,4,5,6	5
	S7_55750504	SBI-07	T/G	55,750,504	4.39–15.43	4.89–12.56	1,3,5,6	6
	S8_61278748	SBI-08	A/G	61,278,748	4.61–15.73	1.08	2,4,5	1
	S9_4793890	SBI-09	C/A	4,793,890	5.58–10.82	2.56–9.01	1,2,6	3
	S10_60795709	SBI-10	C/T	60,795,709	5.27–36.13	4.07–33.58	1,2,3,4,5,6	4
TSW	S1_25033782	SBI-01	C/T	25,033,782	4.58–4.72	12.57–16.85	1,5	1
	S2_69362046	SBI-02	G/A	69,362,046	4.71	4.88	2	1
	S3_64354506	SBI-03	C/T	64,354,506	4.05–5.03	4.30-14.02	1,2,4,5,6	5
	S4_51654024	SBI-04	G/C	51,654,024	4.10–4.77	4.65–8.88	2,4	1
	S5_14397024	SBI-05	C/A	14,397,024	6.98	16.51	2	1
	S9_52837786	SBI-09	T/C	52,837,786	4.53–4.77	3.64–3.82	2,6	2

Note: Methods 1–6 are as follows: **(1) mrMLM** = multi-locus random-SNP-effect mixed linear model MLM., **(2) FASTmrMLM** = factored spectrally transformed multi-locus random-SNP-effect mixed linear model MLM; **(3) FASTmrEMMA** = factored spectrally transformed multi-locus random-SNP-effect efficient mixed-model association; **(4) pLARmEB** = polygenic-background-control-based least angle regression plus empirical Bayes; **(5) pKWmEB** = polygenic-background-control-based Kruskal–Wallis test plus empirical Bayes, and **(6) ISIS EM-BLASSO** = iterative sure independence screening EM-Bayesian LASSO. Each row corresponds to a specific QTN associated with a trait, detailing its chromosome location, position, LOD score, R^2 value, the method used, and the number of associated genes

strong associations. Chromosomes 4–10 harbor QTNs linked to various traits, with some QTNs appearing in multiple traits, suggesting pleiotropic effects (Table S2). Trait Associations and Statistical Significance LOD scores and R-squared values for each QTN indicate their significance in trait expression, for instance, QTNs with LOD scores exceeding 4.0 are considered significant, while R-squared values above 10% suggest substantial contributions to trait variances (Table S2).

Discussion

Analysis of genetic variability, heritability, and genetic advance

Days to flowering have high broad sense heritability ($h^{2)}$ (0.7) and genetic advance (GA) as a percentage of the mean (19.6%) indicating that this trait has a strong genetic component and can be effectively improved through selection. The large genotypic variance (142.6) compared to the environmental variance (75.2) suggested that genetic factors play a significant role in determining days to 50% flowering.

Days to maturity have moderate h^2 (0.5) and GA as a percentage of the mean (8.4%). This suggests that genetic improvement is possible but may be more challenging than for days to flowering. The genotypic and environmental variances are more balanced for this trait, indicating that genetic and environmental factors contribute to the expression of days to physiological maturity.

The number of seeds per plant had low h^2 (0.0) and GAM (3.6%), indicating that environmental factors highly influence it and may be challenging to improve through selection alone. The EV (501.1) was much larger than the GV (26.2), supporting the low heritability estimate. The GCV (7.9%) was relatively low, suggesting limited genetic variability for this trait.

Grain yield has extremely low h^2 (0.003) and GAM (0.2%), indicating that environmental factors predominantly influence it and may be difficult to improve through selection. The EV (494552.3) is much larger than the GV (149090.8), confirming the high environmental influence on this trait. The GCV (2.1%) is very low, suggesting limited GV for grain yield in the population.

Traits that are coupled and exhibit high h^2 and GA, like days to flowering and plant height, offer good prospects for effective selection [38]. Conversely, traits with low h^2 and GA, such as the number of seeds per plant and grain yield, can be more difficult to enhance through selection alone and may necessitate more intricate breeding strategies (Table S3).

The high GCV (11.7%) indicated a good amount of genetic variability for this trait in the population, which is desirable for selection. The high broad-sense h^2 in this study was consistent with previous findings Subudhi, Rosenow, & Nguyen [39] reported for days to flowering in sorghum ranging from 0.61 to 0.92, depending on the population and environment. Ayana & Bekele [40] also found high h^2 estimates (0.70–0.80) for days to flowering in sorghum, indicating the strong genetic control of this trait.

The moderate h^2 (0.4) in plant height in this study aligns with the findings by Ayana and Bekele [41], who reported h^2 estimates ranging from 0.38 to 0.59 for plant height traits in sorghum. Also, Subudhi et al. [42] reported moderate to high h^2 (0.48–0.77) for plant height in sorghum, depending on the population and environment. The h^2 for plant height is moderate (0.4), and the GAM is relatively high (22.5%). The large GV (3689.8) compared to the EV (5227.7) suggests that genetic factors are pre-dominant in determining plant height. The GCV (16.9%) is reasonably high, indicating the presence of substantial genetic variability for plant height.

The moderate h^2 (0.5) for days to physiological maturity observed in this study was consistent with the previous findings by Ayana and Bekele [41] which reported the estimated h^2 value ranging from 0.45 to 0.56. Subudhi et al. [42] also reported moderate to high h^2 (0.57–0.87) for days to maturity (Table S3).

The very low h^2 (0.0) for the number of seeds per plant observed in this study was consistent with the findings of Ayana and Bekele [41] who reported low h^2 estimates (0.08–0.19) for this trait in sorghum. Subudhi et al. [42] also found low h^2 (0.19–0.45) for the number of seeds per plant in sorghum, indicating the strong influence of environmental factors (Table S3).

The extremely low h^2 (0.003) for grain yield in this study was in line with the findings of Ayana and Bekele [41], who reported very low h^2 estimates (0.01–0.12) for grain yield in sorghum. Subudhi et al. [42] also found low h^2 (0.15–0.47) for grain yield, suggesting that environmental factors highly influence this trait and may be challenging to improve through selection alone.

Analysis of genetic diversity and MTA association using SNP markers

The analysis of SNP heterozygosity, MAF, and R^2 values provides valuable insights into the genetic landscape of sorghum (Fig. 1). The observed variation in heterozygosity across chromosomes indicates differing levels of genetic diversity within the population. This diversity is vital for adaptive traits that enhance resilience to environmental stresses. The MAF findings suggest that specific alleles may be more prevalent, crucial for identifying genetic resources that can be utilized in breeding programs. For instance, higher MAF regions may harbor alleles associated with beneficial traits, providing a genetic basis for improving sorghum varieties [43]. The R^2 analysis highlights regions with strong associations with traits, guiding future marker-assisted selection

efforts [44]. Previous studies have found that genetic diversity is essential for the adaptability of sorghum to varying environmental conditions, Baye et al. [45] reported that this supports our findings of higher MAF correlating with beneficial traits [46]. This reinforces the idea that genetic diversity is key to successful breeding. Zhao et al. [47] demonstrated the utility of R^2 values in identifying QTLs for important traits in sorghum. Our results similarly highlight regions with high R^2 values, suggesting that these markers are valuable for targeted breeding efforts.

Distribution of SNPs across Sorghum chromosomes

Regions with high SNP density, found on chromosomes 1, 5, and 7, indicated areas where selective pressure may have influenced genetic variation (Fig. 1). The clustering of SNPs can also facilitate the identification of genomic regions associated with traits of interest, as the high-density markers allow for finer mapping of QTLs [48]. In our studies, the presence of SNP hotspots could indicate historical selection events that have shaped the current genetic landscape of sorghum. Previous studies by Liu et al. [49] have reported similar patterns of SNP distribution across sorghum chromosomes. These findings confirm that certain genomic regions are more genetically diverse, which can enhance the breeding potential for specific traits. Hotspots of SNP variation reported by Baye et al. [45] identified specific genomic regions associated with important agronomic traits, aligning with the observations of SNP hotspots in our study. These hotspots are essential for breeding programs as they may harbor alleles that confer desirable traits. Zhao et al. [47] emphasized the importance of high SNP density in facilitating marker-assisted selection (MAS) for complex traits. Each pink-colored MTA represents a significant association between a specific SNP and the trait of interest (Fig. 4). The number of these pink MTAs indicates how many genetic markers show a strong statistical link to the trait being studied. The higher number of pink MTAs suggests that the trait is influenced by multiple genetic factors. This may indicate a complex genetic architecture, where several genes contribute to the trait's expression. The lower number of pink MTAs could imply that the trait is controlled by fewer genetic factors, potentially indicating a simpler genetic basis (Fig. 4). Our findings perfectly aligned with this report by illustrating how regions with many SNPs can serve as valuable targets for breeding strategies.

QTN effect analysis for agronomic and Yield-related traits

The variability in QTN effects and the strength of associations across agronomic and yield-related traits highlighted the need for utilizing multiple analytical approaches to capture the full spectrum of genetic influences [50]. The mrMLM and FASTmrMLM models effectively identify significant QTNs, with high LOD scores and substantial R^2 values. In contrast, FASTmrEMMA identified fewer QTNs, suggesting that different models can yield varying insights into genetic architecture (Table 2).

Table 2 presents significant Quantitative Trait Nucleotides (QTNs) identified through Multi-Locus-Genome-Wide Association Study (ML-GWAS) models for various agronomic and yield-related traits in sorghum, including days to flowering (DF), days to maturity (DM), plant height (PH), seed number per panicle (SNPP), grain yield (GY), and thousand seed weight (TSW). Each QTN is characterized by its chromosomal location, allele information, position in base pairs (bp), LOD (Logarithm of Odds) score, r^2 (coefficient of determination or phenotypic variance explained), the methods used for identification, and the number of genes within the linkage disequilibrium (LD) region. The methods employed include mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO, which collectively enhance the power to detect associations, especially for complex traits influenced by multiple genetic and environmental factors (Table S2, Table 2).

Identification and characterization of QTNs associated with various genes

For days to flowering (DF), several QTNs were identified, including *S1_73955151* on chromosome SBI-01, which is near the Ma1 gene (*Sb01g010260/QTNGL1.2*, (Total number of green leaves at maturity), a well-known regulator gene of flowering time in sorghum [51, 52]. This QTN showed a high r^2 value (13.22–19.38), indicating a strong association with flowering time. Similarly, *S2_6784036* on SBI-02 is near the Dw2 locus, associated with plant height and flowering time [52, 53], and *S6_49875883* on SBI-06, with a high LOD score (6.60–6.66), is likely linked to the *Sb06g023260* gene, previously associated with flowering time [54]. For days to maturity (DM), *S1_76529338* on SBI-01, also near the Ma1 gene, showed a high r^2 value (5.14–19.59), suggesting a strong association with maturity. *S5_623466* on SBI-05, near the *Sb05g004000* gene, also showed a moderate r^2 value (3.92–8.36), indicating a reliable association with maturity [54].

In the case of plant height (PH), *S1_67415907* on SBI-01, with a very high LOD score (10.14), is likely associated with the Dw1 gene, a major determinant of plant height in sorghum [53], the high r^2 value (7.63) further supports this strong association. *S2_1166841* on SBI-02, near the Dw2 locus, also showed a moderate r^2 value (1.53–10.63), consistent with previous findings [53]. For seed number per panicle (SNPP), *S1_1359747* on SBI-01, with a high LOD score (7.17–10.38), is likely associated

with the *Sb01g001000* gene, previously linked to seed number [54], and *S9_50050063* on SBI-09, with a wide range of LOD scores (4.13–8.76) and r^2 values (3.03-18.00), suggests a complex genetic architecture for seed number per panicle, supported by multiple methods.

For grain yield (GY), *S6_32754749* on SBI-06, with a very high LOD score (4.45–17.01), is likely associated with the *Sb06g023260* gene, previously linked to grain yield [54]. Similarly, *S10_60795709* on SBI-10, with a very high LOD score (5.27–36.13), is likely associated with the *Sb10g025000* gene, also linked to grain yield [55]. Finally, for thousand seed weight (TSW), *S1_25033782* on SBI-01, with a high r^2 value (12.57–16.85), is likely associated with the *Sb01g010260* gene, previously linked to seed weight [55]. *S5_14397024* on SBI-05, with a high LOD score (6.98), is likely associated with the *Sb05g004000* gene, also linked to seed weight [54]. QTNs identified in this study are consistent with previous research findings, and the use of ML-GWAS models provides a robust approach to uncovering the genetic basis of complex agronomic and yield-related traits in sorghum. The high LOD scores and r^2 values, along with the use of multiple methods, can validate the reliability of these associations. These findings contribute to a deeper understanding of the genetic architecture of sorghum and provide valuable insights for future breeding programs aimed at improving yield and agronomic traits.

The Cross-Validation of alleles with previously identified Sorghum genes

To compare our findings with previous research and validate the alleles identified, we have examined each trait and the associated QTNs, referencing relevant studies that have identified similar genetic loci or alleles in sorghum or related crops. Below is a detailed comparison and validation of the alleles in our result (Table 2), supported by references to previously investigated research.

Days to Flowering (*S1_73955151* (SBI-01, *T/C*): This QTN is located on chromosome SBI-01, near the Ma1 gene (*Sb01g010260*), which is a well-known regulator of flowering time in sorghum. The Ma1 gene has been extensively studied and is associated with delayed flowering under long-day conditions [51]. The *T/C* allele variation in this region is consistent with previous findings that link this locus to flowering time [56, 57]. *S2_6784036* (SBI-02, *T/G*): This QTN is near the Dw2 locus, which is associated with plant height and flowering time in sorghum. The *T/G* allele variation aligns with previous studies that identified this region as a major QTL for flowering time [51, 58]. *S6_49875883* (SBI-06, *G/C*): This QTN is likely associated with the *Sb06g023260* gene, which has been linked to flowering time in sorghum. The *G/C* allele variation is consistent with previous findings that identified this region as a significant QTL for flowering time [59].

Days to maturity *(S1_76529338* (SBI-01, *A/G*): This QTN is near the Ma1 gene, which regulates maturity in sorghum. The A/G allele variation is consistent with previous studies that identified this locus as a major determinant of maturity [59–61], *S5_623466* (SBI-05, *A/C*): This QTN is near the *Sb05g004000* gene, which has been associated with maturity in sorghum. The A/C allele variation aligns with previous findings that identified this region as a significant QTL for maturity [59].

Plant height (*S1_67415907* (SBI-01, *G/T*): This QTN is likely associated with the Dw1 gene, a major determinant of plant height in sorghum. The *G/T* allele variation is consistent with previous studies that identified this locus as a significant QTL for plant height [58, 59]. *S2_1166841* (SBI-02, *C/A*): This QTN is near the Dw2 locus, which is associated with plant height in sorghum. The *C/A* allele variation aligns with previous findings that identified this region as a major QTL for plant height [59].

Seed number per plant/panicle (*S1_1359747* (SBI-01, *T/C*): This QTN is likely associated with the *Sb01g001000* gene, which has been linked to seed number in sorghum. The *T/C* allele variation is consistent with previous findings that identified this region as a significant QTL for seed number [59]. *S9_50050063* (SBI-09, *C/G*): This QTN is likely associated with the *Sb09g025000* gene, which has been linked to seed number per plant. The *C/G* allele variation aligns with previous findings that identified this region as a significant QTL for seed number [59].

Grain yield (*S6_32754749* (SBI-06, *C/G*): This QTN is likely associated with the *Sb06g023260* gene, which has been linked to grain yield in sorghum. The *C/G* allele variation is consistent with previous findings that identified this region as a significant QTL for grain yield [59] *S10_60795709* (SBI-10, *C/T*): This QTN is likely associated with the *Sb10g025000* gene, which has been linked to grain yield in sorghum. The *C/T* allele variation aligns with previous findings that identified this region as a significant QTL for grain yield [59].

Thousand seed weight (*S1_25033782* (SBI-01, *C/T*): This QTN is likely associated with the *Sb01g010260* gene, which has been linked to seed weight in sorghum. The *C/T* allele variation is consistent with previous findings that identified this region as a significant QTL for seed weight [59, 62, 63]. *S5_14397024* (SBI-05, *C/A*): This QTN is likely associated with the *Sb05g004000/QSNDF5.1* gene (Neutral detergent fibre-GWAS method) which has been linked to seed weight in sorghum. The C/A allele variation aligns with previous findings that identified this region as a significant QTL for seed weight [56, 59, 63, 64].

The alleles identified in our study are consistent with previous research findings in sorghum, particularly those

related to flowering time, plant height, seed number per plant, grain yield, and seed weight. The high LOD scores and r^2 values, along with using ML-GWAS methods, validate the reliability of these associations. These findings contribute to a deeper understanding of the genetic architecture of sorghum and provide valuable insights for future breeding programs aimed at improving yield and agronomic traits.

Genetic analysis of agronomic & yield traits using polymorphic SNPs

Identifying significant QTNs for days to flowering and maturity provides valuable insights into the genetic control of these traits in sorghum. Notable associations were found on chromosomes 1, 3, and 5 for days to flowering, which aligned with previous studies identifying similar loci affecting flowering time in sorghum and other related cereal species [45, 55]. For days to maturity, significant SNPs on chromosomes 2, 4, and 9 correspond to loci previously reported in sorghum and other grain crops, indicating conserved genetic influences across species [45, 55] also identified key loci on chromosomes 1 and 3 associated with maturity time, supporting our current findings. Baye et al. [45] reported significant associations between chromosomes 2 and 4, which align with our findings (Fig. 4).

The important associations found on chromosomes 1, 3, and 4 for plant height are consistent with previous studies that have identified key loci influencing height in sorghum agronomic traits [49, 65]. These studies have suggested that these loci may harbor important genes involved in growth regulation. For the seed number per plant, significant SNPs on chromosomes 2 and 5 corroborate findings from Baye et al. [45] and other research, highlighting the importance of these chromosomal regions in seed development and yield traits. The consistent identification of similar loci across studies underscores the reliability of these genetic markers. Ramu et al. [65] identified significant QTLs associated with plant height on chromosomes 1 and 3, aligning with our findings. Baye et al. [45] further validated these results, which emphasize the role of these loci in plant growth regulation and significant associations between chromosomes 2 and 5 for seed number traits, which aligns with our results. This work emphasized the importance of these loci in optimizing yield through genetic selection. Identifying significant QTNs for grain yield and thousand seed weight enhances our understanding of the genetic control of these traits in sorghum. Previous research studies [45, 57] have corroborated the notable associations found on chromosomes 1, 3, and 5 for grain yield, indicating critical loci influencing yield traits. These studies have suggested that these regions may harbor genes involved in metabolic processes and stress responses crucial for yield stability. For thousand seed weight, significant SNPs on chromosomes 2 and 4 align with findings from earlier research, which have identified key loci affecting seed weight in sorghum [45, 47]. The consistent identification of similar loci across studies underscores the reliability of these genetic markers for breeding applications. Menamo et al. [66] supported our findings by identifying significant QTLs associated with grain yield on chromosomes 1 and 3. Baye et al. [45] further validated these results, emphasizing the role of these loci in enhancing yield through genetic improvement. Zhao et al. [47] also highlighted the importance of these loci for seed development.

Associated genomic regions identified by three models of Multi-Locus analysis

A thorough comparison of six ML-GWAS methods indicated that mrMLM, pLARmEB, and FASTmrMLM were the most effective in identifying significant QTNs associated with agronomic, yield, and yield-related traits, with mrMLM detected 42 QTNs, pLARmEB identified 36 QTNs, and FASTmrMLM found 30 QTNs (Table 1).

A related study identified 160 and 130 significant QTNs across five traits using the ISISEM-BLASSO and pLARmEB methods, respectively. Furthermore, Zhang, Jia, & Dunwell [67] highlighted ISISEM-BLASSO as the most robust multi-locus method in the R Package Genome Association and Prediction Integrated Tool (GAPIT) [68]. Similarly, Zhong et al. [69] found that pKWmEB, ISIS EM-BLASSO, and pLARmEB yielded higher counts of significant QTNs, reporting 189, 171, and 160 QTNs, respectively. In contrast, noted that among the six ML-GWAS methods, mrMLM demonstrated superior capability in detecting reliable QTNs for various agronomic traits in sorghum, including plant height, days to flowering, grain yield, tiller number, hundred seed weight, and panicle exertion. This discrepancy may be attributed to the specific traits and population panels analyzed in their study.

QTN mapping and genetic architecture

The distribution of QTNs across the sorghum genome illustrates the complex genetic architecture underlying agronomic traits. Identifying multiple QTNs, especially in chromosomes 1, 2, and 3, emphasizes the need for targeted breeding strategies that leverage these genetic markers. Recent studies reported similar associations between QTNs and agronomic traits. For instance, Baye et al. [45] identified key QTLs linked to grain yield and plant height, verifying our findings on chromosomes 3 and 8. Additionally, Zhao et al. [47] demonstrated the importance of certain SNPs in drought tolerance, aligning with our results on chromosome 2. The identified

QTNs provide a valuable resource for marker-assisted selection in sorghum breeding programs.

Candidate gene mining and mapping

The analysis of candidate genes associated with key traits in sorghum reveals a complex interplay between genetic factors and agronomic performance. Each gene identified not only contributes to specific phenotypic expressions but also offers insights into potential genetic pathways that can be exploited for crop improvement [70].

The QTNs and putative candidate genes were indicated on the right side of the chromosomes, with abbreviations representing different traits displayed. Candidate genes on each chromosome are marked in distinct colors: green for days to flowering, red for days to maturity, pink for seed number per plant, and blue for grain yield. The numbers on the left side indicated the physical distance in megabase pairs (Mbp) between adjacent loci on the chromosome (Fig. 5).

The regulation of flowering time and maturity in sorghum is significantly influenced by genes such as *Sobic.001G196700* and *Sobic.002G183400*, both of which code for hypothetical proteins involved in floral development. These genes play essential roles in optimizing reproductive success (Table S4). Additionally, the stress response gene *Sobic.005G176100*, annotated as a mannose-6-phosphate isomerase, emphasizes the importance of resilience in sorghum, particularly under adverse environmental conditions [70]. Its involvement in various stress response mechanisms highlights its potential as a target for genetic enhancement strategies aimed at developing more resilient sorghum varieties.

Understanding the interactions of *Sobic.005G176100* with other genes could lead to the creation of cultivars that not only withstand environmental stress but also maintain high yield and quality. Other key genes, such as *Sobic.003G324400* and *Sobic.004G178300*, are crucial for regulating plant height and seed weight, respectively. The influence of *Sobic.004G178300* on seed weight positions it as a valuable candidate in yield enhancement breeding programs (Table S4). The high LOD and R^2 values associated with these genes further underscore their potential utility in marker-assisted selection (MAS) [6], which can simplify the breeding process by enabling early selection of desirable traits and optimizing resource use compared to traditional phenotypic methods.

The table presents a curated selection of sorghum genes linked to various agronomic traits, revealing their potential roles in drought tolerance, plant morphology, and seed characteristics. For example, *Sobic.002G183400* is associated with drought-related traits but remains uncharacterized, suggesting a need for further investigation into its specific functions and involvement in stress response pathways. Similarly, *Sobic.002G140900*, related to drought management, resembles the pre-mRNA splicing factor PRP38 protein, implying a role in RNA processing essential for gene expression regulation under stress conditions. Additionally, *Sobic.005G176100* may influence energy pathways critical during drought stress. With the gene *Sobic.003G324400* containing an AP2 domain, it likely contributes to transcriptional regulation affecting developmental processes. Furthermore, *Sobic.005G176000*, a zinc finger protein, may be involved in gene regulation during seed development, while *Sobic.004G178300*, linked to thousand seed weight and annotated as a putative splicing factor U2AF, highlights the significance of RNA splicing in seed development (Table S4). Collectively, these genes are vital for enhancing sorghum's adaptability to environmental stresses and improving yield traits, warranting further functional studies to elucidate their roles in the complex regulatory networks governing these phenotypes.

Conclusion

The current study highlights the significance of sorghum as a crucial cereal crop for over 750 million people, especially in Ethiopia, where diverse landraces flourish across various agroecological zones. Through the collection, and genotyping of 216 Ethiopian sorghum landraces, we uncovered substantial genetic variations and phenotypic traits, leading to important marker trait associations. The Pearson correlation analysis revealed strong correlations among most traits ($p < 0.0001$), with exceptions for grain yield about flowering and days to maturity.

Genetic variability assessments indicated that days to flowering had high heritability ($h^2 = 0.7$) and genetic advance (GA = 19.6%), suggesting significant potential for improvement through selective breeding. In contrast, grain yield showed extremely low heritability ($h^2 = 0.003$) and GA (0.2%), indicating a predominant environmental influence and challenges for genetic enhancement.

The analysis identified 351,692 SNP markers, refined to 50,165 for further investigation. This extensive dataset forms a solid foundation for future genome-wide association studies (GWAS). The Manhattan plot analysis revealed several significant QTNs, particularly on chromosomes 1, 5, and 8, with strong LOD scores for traits such as days to flowering and plant height. In total, 176 QTNs were identified, with the mrMLM model detecting the most significant markers (42 QTNs), reflecting the complex genetic architecture influencing these traits.

The research emphasizes key QTNs and candidate genes linked to essential agronomic traits, utilizing ML-GWAS models to inform breeding strategies. It highlights high heritability for traits like days to flowering and plant height, while other traits exhibited low heritability due to environmental factors. The identification of QTNs and candidate genes are crucial components for

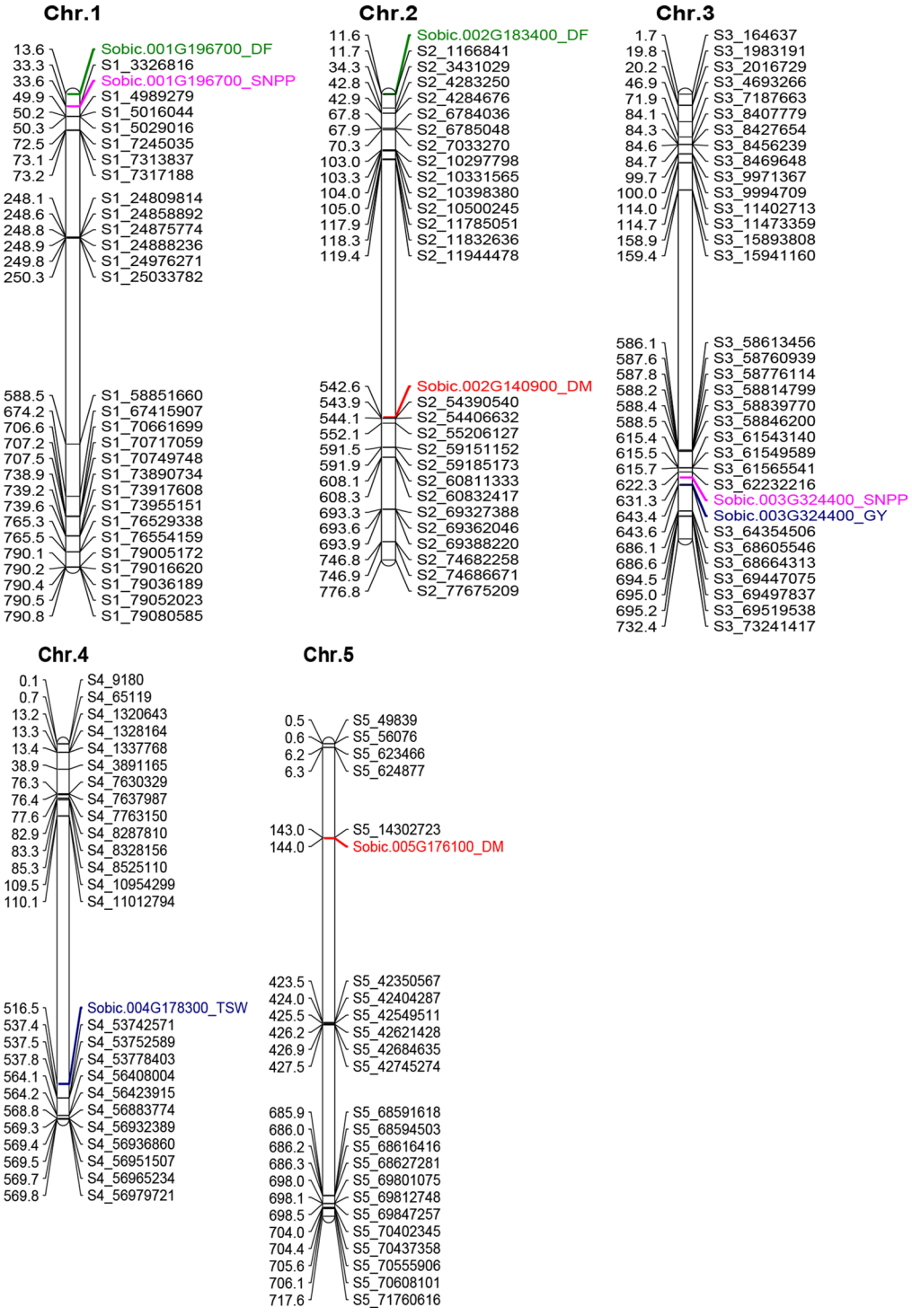

Fig. 5 Linkage groups and chromosomal locations of significant QTNs and intragenic candidate genes associated with agronomic and yield-related traits in sorghum

improving adaptability to environmental stressors and enhancing yield traits. Genes such as *Sobic.001G196700*, *Sobic.002G183400*, and *Sobic.005G176100* play significant roles in regulating flowering and stress responses, while *Sobic.003G324400* and *Sobic.004G178300* are vital for influencing plant height and seed weight, respectively. The high LOD and R^2 values associated with these genes indicate their potential for application in marker-assisted selection, facilitating early identification of desirable traits in breeding programs. Overall, this research

highlights the importance of these genes in the development of resilient sorghum varieties and calls for further investigations to better understand their functions within the intricate regulatory frameworks that govern key agronomic traits.

Materials and methods

Genetic materials and experimental design

A total of 202 sorghum landraces (Table S5) with their passport data were sourced from the Ethiopian Biodiversity Institute, Addis Ababa. Additionally, 9 improved varieties and 5 released landraces used as check cultivars were acquired from the National Sorghum Improvement program at Melkasa Agricultural Research Center, part of the Ethiopian Institute of Agricultural Research, Addis Ababa (Table S5). The SNP markers dataset was extracted from the resequencing of sorghum accessions at the University of Wisconsin Biotechnology Center [71] and was made available through the Purdue University Sorghum Research Repository https://purr.purdue.edu/publications/3189/1.

The experiment was conducted over two consecutive cropping seasons from 2022 to 2023 in the Pawi district (11° 18' N, 36° 24' E), at an elevation of 1,100 to 1,200 m above sea level. The study was carried out in two kebeles: Dangure and Village-7. A total of 216 sorghum genotypes including 202 landraces and 14 cultivars (9 improved varieties and 5 released landraces) were utilized in the study. The genotypes were planted in single rows using an α-lattice design with three replications and four blocks, and each block comprised 54 genotypes. Each net plot measured 0.75 m in width by 5 m in length, with an intra-row spacing of 0.75 m. The spacing between replications was 2 m, while the distance between blocks was set at 1.5 m. Planting was performed using a manual drilling method, followed by thinning to a spacing of 0.2 m after 20 days of emergence. Post-thinning, each plot maintained an average of 25 plants.

Phenotyping

Accurate and well-characterized data for the traits of interest, specifically agronomic and yield-related traits, were collected. Five plants from each row were randomly selected according to the type of traits being measured, including days to flowering, days to maturity, plant height, seed number per plant, grain yield, and thousand seed weight. The missing and unrepresentative phenotypic data was imputed by SAS JMP V.5 [72]. Data was normalized and standardized by the Shapiro-Wilk statistics test at $P > 0.05$ [50].

Genotyping

GWAS was conducted using genotyping by sequencing (GBS) [73]. The GBS procedure [74], utilized the *ApeKI* restriction/incision enzyme (recognition site of G|CWCG) to generate the GBS library, which was then sequenced on Illumina HiSeq2500 lanes [75]. SNP markers were extracted from the resequencing data of 1,628 sorghum accessions [76]. The SNP dataset was filtered to exclude SNPs with an MAF of less than 0.05 missing values. The remaining missing values were imputed using the Beagle 5.0 software package [77], resulting in 50,165 SNPs.

To ensure data quality, the SNP dataset was again filtered to exclude any SNPs with a MAF of 0.00002, calculated from the expression $4 \times 4.614 \times 4.61$, which corresponds to a likelihood ratio test derived from an LOD score of 4. Under the null hypothesis, this likelihood ratio follows a chi-square (χ^2) distribution with one degree of freedom [78]. Furthermore, only SNP markers identified in at least three different models were considered reliable for agronomic and yield-related QTNs. Similarly, QTNs that were detected in three or more models and demonstrated a phenotypic variation of $R^2 > 10\%$ were classified as major QTNs.

Data analysis

The phenotypic data were analyzed using a mixed linear model (MLM) approach implemented in the "asreml-R" R package [79]. The REML mixed model equation was:

$$y = X + Zu + e,$$

Where 'y' represents the measured data for each trait, 'τ' is the fixed effects (genotypes) in the trial, 'X' is the design matrix for the fixed effects, 'u' is the random effects (columns and rows), 'Z' is the design matrix for the random effects, and e is the residual error. The genetic parameters, including $\sigma^2 g$, $\sigma^2 p$, GCV, & h^2, were estimated using the *"variability"* R package [80].

The genotypic variance ($\sigma^2 g$) was calculated as (MSg - MSe)/r, where MSg is the mean square of the genotype and r is the number of replicates. The phenotypic variance (σ^2) was estimated as $\sigma^2 p = \sigma^2 g + \sigma^2 e$, where $\sigma^2 e$ is the error mean square. Broad-sense heritability (h^2) was computed using the formula:

$$h2 = 2g/(2g + (2e/n)),$$

where n is the number of replicates, as suggested by Pariyar et al. [81]. The best linear unbiased prediction (BLUP) values were estimated using the META-R package [82] and used for the GWAS analysis.

Genome-Wide-Association study

Since GWAS involves testing thousands to millions of SNPs across the genome, it is important to correct for multiple testing. Statistical methods, such as Bonferroni

correction or false discovery rate [67] adjustment. The GWAS analysis identifies significant associations between specific genomic regions (containing one or more SNPs) and the traits of interest. These genomic regions are then considered as potential candidate regions influencing the traits. The identified genomic regions are further analyzed to understand the biological significance of the associated SNPs [74, 75]. This involves annotating the genes within or near the significant areas, exploring their known functions, and assessing their potential role in the observed trait variations. By conducting an ML-GWAS, researchers can gain insights into the genetic architecture underlying the major agronomic, yield, and yield-related traits in sorghum landraces, helping to inform breeding programs and improve crop productivity [32].

Genome-wide association studies (GWAS) have uncovered SNPs associated with complex traits, yet these represent only a fraction of the SNPs within the same haplotype block [83]. Six different ML-GWAS models were used for the MTA analysis and the Identification of QTNs [84]: mrMLM [78], FASTmrMLM [27], FASTmrEMMA [28], pLARmEB [30], pkWmEB [31], and ISIS EM-BLASSO [29]. All of these ML-GWAS models were implemented in the "mrMLM.GUI" R package V.4.4.1 [26] which provides a graphical user interface for the multi-locus random SNP-effect mixed linear model. Also, the GAPIT 3.0 [77, 78, 85, 86] was applied for GWAS graph interference analysis.

The population structure and kinship matrix for our accessions were estimated in the previous studies by Grima et al. [71], additionally, the mrMLM.GUI package was utilized to calculate the population structure and kinship matrix internally. The resulting – log10(p) values obtained from the ML-GWAS analysis were employed to generate Manhattan and Q-Q plots using the mrMLM.GUI R-package [87].

Co-localization of previously detected QTLs for agronomic and Yield-Related traits, and identification of candidate genes

The colocations of significant QTNs with previously identified QTLs were examined using the Sorghum QTL Atlas database [88], focusing on the linkage disequilibrium decay range of 65 kb. Candidate genes were identified through *biomaRt* tools [89] on the Phytozome platform [90], also within the 65 kb LD decay distance from the genomic regions where the QTNs were located [76]. The SorghumBase online database was also utilized to gather comprehensive descriptions of the relevant genes.

ML-GWAS study

Several multi-locus genome-wide association study (ML-GWAS) methods were used to identify significant QTNs. This includes the mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO approaches, all of which are implemented in the R package "mrMLM" [78]. Default parameter values were used, and a LOD threshold of ≥ 4 or p-value ≤ 0.0002 was applied to determine significant marker-trait associations [91]. Principal component analysis and kinship matrices were incorporated into all the methods. The R package CMplot [92] was used to visualize Manhattan and quantile-quantile (QQ) plots from the GWAS results. Linkage disequilibrium between SNPs was estimated using the squared correlation coefficient (r^2) within a 0-10 cM window, calculated with the Tassel 5 tool [93]. The phenotypic effect size of each allelic variation was determined across the sorghum landraces and visualized using box plots in R 4.4.1 software [94].

QTNs with a logarithm of the odds (LOD) score of at least 4.0 is significantly associated with the agronomic, yield, and yield-related traits under investigation [2]. This LOD score threshold corresponds to a p-value of 0.00002, calculated as the probability of the chi-square test statistic (χ^2) exceeding 4×4.61, given 1 degree of freedom under the null hypothesis. Specifically, the conversion from an LOD score of 4.0 to its corresponding likelihood ratio test was done using the formula $4.0 \times \ln(100) = 4.0 \times 4.61$. This likelihood ratio test, under the null hypothesis, follows a chi-square distribution with 1 degree of freedom, as described in the work of Wang et al. [78]. The LOD score threshold value of 4.0 is to identify QTNs that were significantly associated with the agronomic and yield-related traits. This threshold was chosen based on the statistical significance level, where the p-value corresponding to a LOD score of 4.0 was calculated to be 0.00002 using the chi-square distribution with one degree of freedom.

To identify reliable QTNs associated with agronomic and yield-related traits, it's very important to apply an additional filtering criterion. Only SNP markers detected in at least three of the six ML-GWAS models were designated reliable agronomic and yield-related associated QTNs. Similarly, QTNs detected in three or more models and exhibiting a phenotypic variation (R-squared) greater than 10% were designated as major QTNs. This suggested that these major QTNs significantly influenced the observed agronomic and yield-related traits. For the current analysis, we applied the "mrMLM.GUI" R-software package [26] to internally calculate the population structure and kinship matrix [35] as part of the ML-GWAS approach.

Finally, the resulting -log10(p) values from the ML-GWAS models were used to generate Manhattan and QQ plots using the "mrMLM.GUI" package, as described by Zhang et al. [26]. These visual representations helped to identify and interpret the significant associations

between the SNP markers and the agronomic and yield-related traits.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11458-4.

Supplementary Material 1: Pearson correlation matrix analysis of sorghum agronomic and yield association traits.

Supplementary Material 2: Identified candidate QTNs or genes.

Supplementary Material 3: Environmental, genotypic, and phenotypic variances, heritability estimates, and genetic advance.

Supplementary Material 4: Candidate genes descriptions.

Supplementary Material 5: Ethiopian Sorghum Landraces, Improved Varieties, and Improved-Landraces.

Supplementary Material 6: QQ plot analysis of sorghum's key agronomic and yield-associated traits provides valuable insights into their genetic architectures.

Acknowledgements
The authors would like to thank Addis Ababa Science and Technology University. Authors further extend thanks to the sorghum research teams at Melkasa and Pawi Agricultural Research Centers in Ethiopia and the Sorghum Research Division at Purdue University.

Author contributions
A.A.W. and A.G. conceptualized the study; A.G and A.A.W. investigated the data; A.G, H.N. and A.A.W. provided resources; A.G., A.A. and A.A.W. curated data; A.G., and A.A.W. designed methodology; A.G., A.A. and A.A.W. software; A.A.W. Supervised the study; A.G. wrote original draft; A.G, A.A, H.N and A.A.W. wrote review & edited. All authors read and approved the final manuscript.

Funding
No Funding.

Data availability
SNP markers dataset was extracted from the resequencing of sorghum accessions at the University of Wisconsin Biotechnology Center [14] and was made available through the Purdue University Sorghum Research Repository. The SNP data is available online at https://purr.purdue.edu/publications/3189/1. Also can be accessed at https://pubmed.ncbi.nlm.nih.gov/31191590/ and https://pubmed.ncbi.nlm.nih.gov/33217211/ The supplementary dataset is included in this submission.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Applied and Natural Sciences (CANS), Department of Biotechnology, Addis Ababa Science and Technology University (AASTU), Addis Ababa, Ethiopia
[2]Biotechnology and Bioprocess Center of Excellence, AASTU, Addis Ababa, Ethiopia
[3]College of Agriculture, Food and Climate Sciences, Department of Plant Sciences, Injibara University, Injibara, Ethiopia
[4]Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Almas Allé 8, Lomma, Uppsala 750 07, Sweden
[5]Purdue University, 610 Purdue Mall, West Lafayette, IN 47907, USA

Received: 28 October 2024 / Accepted: 6 March 2025
Published online: 28 March 2025

References

1. Clayton WD, Renvoize SA. Genera Graminum. Grasses of the world; 1986.
2. Wondimu Z, et al. Genome-wide association study reveals genomic loci influencing agronomic traits in Ethiopian sorghum (Sorghum bicolor (L.) Moench) landraces. Mol Breed. 2023;43:1–15.
3. Srinivasa Rao P, et al. Sorghum production for diversified uses. Genet Genomics Breed Sorghum. 2014. https://doi.org/10.1201/b17153.
4. Cuevas HE et al. (2018) Genome-Wide Association Mapping of Anthracnose (Colletotrichum sublineolum) Resistance in the U.S. Sorghum Association Panel. *Plant Genome* 11.
5. Lemma LL et al. (2024) On-Farm Demonstration of Improved Sorghum (Sorghum bicolor L. Moench) Technologies in Cluster Based Large Scale Approaches at Gofa Zone, Southern Ethiopia.
6. Hays DB et al. (2009) QTL mapping of a high protein digestibility trait in sorghum bicolor. *Int. J. Plant Genomics* 2009.
7. Dahlberg J, et al. Assessing sorghum [Sorghum bicolor (L) Moench] germplasm for new traits: Food, fuels & unique uses. Maydica. 2011;56:85–92.
8. Taylor JRN, Duodu KG. (2018) *Sorghum and millets: chemistry, technology, and nutritional attributes*, Elsevier.
9. Kinfe H, Tesfaye A. Yield performance and adoption of released sorghum varieties in Ethiopia. Edelweiss Appl Sci Technol. 2018;2:46–55.
10. Hao H, et al. Sorghum breeding in the genomic era: opportunities and challenges. Theor Appl Genet. 2021;134:1899–924.
11. Teshome A, Zhang J. (2019) Increase of extreme drought over Ethiopia under climate warming. *Adv. Meteorol.* 2019, 5235429.
12. Snowden JD. The cultivated races of sorghum. Adlard& SonLtd; 1936.
13. Tirfessa A, et al. Genetic diversity among Ethiopian sorghum [Sorghum bicolor (L.) Moench] gene bank accessions as revealed by SSR markers. Afr J Biotechnol. 2020;19:84–91.
14. Assefa A, et al. Evaluation of sorghum (Sorghum bicolor (L.) Moench) variety performance in the lowlands area of Wag Lasta, North Eastern Ethiopia. Cogent Food Agric. 2020;6:1778603.
15. Chaturvedi P et al. (2022) Sorghum and Pearl millet as climate resilient crops for food and nutrition security. Front Plant Sci, 13Frontiers Media SA, 851970.
16. Nadolska-Orczyk A, et al. Major genes determining yield-related traits in wheat and barley. Theor Appl Genet. 2017;130:1081–98.
17. Cattivelli L, et al. Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. F Crop Res. 2008;105:1–14.
18. Alemu A et al. (2024) Genomic selection in plant breeding: key factors shaping two decades of progress. Mol Plant.
19. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods. 2013;9:1–9.
20. Gupta PK, et al. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol. 2005;57:461–85.
21. Hamblin MT, et al. Equilibrium processes cannot explain high levels of short-and medium-range linkage disequilibrium in the domesticated grass sorghum bicolor. Genetics. 2005;171:1247–56.
22. Zhao J, et al. Genome-wide association study for nine plant architecture traits in sorghum. Plant Genome. 2016;9:plantgenome2015–06.
23. Shehzad T, Okuno K. QTL mapping for yield and yield-contributing traits in sorghum (Sorghum bicolor (L.) Moench) with genome-based SSR markers. Euphytica. 2015;203:17–31.
24. Morris GP et al. (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci.* 110, 453–458.
25. Boyles RE, et al. Genetic dissection of sorghum grain quality traits using diverse and segregating populations. Theor Appl Genet. 2017;130:697–716.
26. Zhang YW, et al. MrMLM v4.0.2: an R platform for Multi-locus Genome-wide association studies. Genomics Proteom Bioinforma. 2020;18:481–7.
27. Tamba CL, Zhang Y-M. (2018) A fast mrMLM algorithm for multi-locus genome-wide association studies. *biorxiv*.
28. Wen Y, et al. The improved FASTmrEMMA and GCIM algorithms for genome-wide association and linkage studies in large mapping populations. Crop J. 2020;8:723–32.
29. Tamba CL, et al. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. PLoS Comput Biol. 2017;13:e1005357.

30. Zhang J, et al. pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. Heredity (Edinb). 2017;118:517–24.
31. Ren W-L, et al. pKWmEB: integration of Kruskal–Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. Heredity (Edinb). 2018;120:208–18.
32. Zhang Y, et al. Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. Front Plant Sci. 2018;9:1–12.
33. Liu S, et al. Genome-wide association studies of ionomic and agronomic traits in USDA mini core collection of rice and comparative analyses of different mapping methods. BMC Plant Biol. 2020;20:441.
34. Hu X, et al. Multi-locus genome-wide association studies for 14 main agronomic traits in barley. Front Plant Sci. 2018;871:1–14.
35. Zakieh M, et al. Exploring GWAS and genomic prediction to improve septoria tritici blotch resistance in wheat. Sci Rep. 2023;13:1–12.
36. Soto-Cerda BJ et al. (2013) Genetic characterization of a core collection of flax (Linum usitatissimum L.) suitable for association mapping studies and evidence of divergent selection between fiber and linseed types. BMC Plant Biol 13.
37. Sapkota S, et al. Identification of novel genomic associations and gene candidates for grain starch content in sorghum. Genes (Basel). 2020;11:1–15.
38. B. O B., et al. Heritability and genetic advance for grain yield and its component characters in maize (Zea Mays L). Int J Plant Res. 2012;2:138–45.
39. Subudhi PK, et al. Quantitative trait loci for the stay green trait in sorghum (Sorghum bicolor L. Moench): consistency across genetic backgrounds and environments. Theor Appl Genet. 2000;101:733–41.
40. Ayana A, Bekele E. Geographical patterns of morphological variation in sorghum (Sorghum bicolor (L.) Moench) germplasm from Ethiopia and Eritrea: qualitative characters. Hereditas. 1998;129:195–205.
41. Ayana A, Bekele E. Geographical patterns of morphological variation in sorghum (Sorghum bicolor (L.) Moench) germplasm from Ethiopia and Eritrea: quantitative characters. Euphytica. 2000;115:91–104.
42. Subudhi P. Quantitative trait loci for the stay green trait in sorghum consistency across genetic backgrounds and environemtns. Circulation. 2005;111:2866–8.
43. Girma G et al. (2019) A Large-Scale Genome-Wide association analyses of Ethiopian sorghum landrace collection reveal loci associated with important traits. 10, 1–15.
44. Mwamahonje A, et al. Sorghum production constraints, trait preferences, and strategies to combat drought in Tanzania. Sustain. 2021;13:1–13.
45. Baye W et al. (2022) Genetic architecture of grain Yield-Related traits in sorghum and maize. Int J Mol Sci 23.
46. Wondimu Z et al. (2021) Genetic diversity, population structure, and selection signature in Ethiopian sorghum [Sorghum bicolor L. (Moench)] germplasm. *G3 Genes, Genomes, Genet.* 11.
47. Zhao X et al. (2022) Identification of Drought-Tolerance Genes in the Germination Stage of Soybean. *Biology (Basel).* 11.
48. Wu E, et al. Optimized Agrobacterium-mediated sorghum transformation protocol and molecular data of Transgenic sorghum plants. Vitr Cell Dev Biol - Plant. 2014;50:9–18.
49. Liu C, et al. Multi-trait genome-wide association studies reveal novel pleiotropic loci associated with yield and yield-related traits in rice. J Integr Agric. 2024. https://doi.org/10.1016/j.jia.2024.07.026.
50. Soto-Cerda BJ et al. (2020) Drought response of flax accessions and identification of quantitative trait nucleotides (QTNs) governing agronomic and root traits by genome-wide association analysis. Mol Breed 40.
51. Murphy R. The identification of two maturity loci sheds. Light on Photoperiodic Flowering in Sorghum; 2012.
52. Li X. Phenotypic plasticity and heterosis. Insights from sorghum flowering time and plant height; 2015.
53. Brown PJ, et al. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics. 2008;180:629–37.
54. Higgins R. Genetic dissection of sorghum height and maturity variation using sorghum converted lines and their exotic progenitorsuniversity. of Illinois at Urbana-Champaign; 2014.
55. Higgins RH, et al. Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families. G3 Genes Genomes Genet. 2014;4:1593–602.
56. Mace ES, et al. Whole-genome sequencing reveals untapped genetic potential in Africa's Indigenous cereal crop sorghum. Nat Commun. 2013;4:2320.
57. Yruela Guerrero I. (2015) Plant development regulation: overview and perspectives.
58. Yang S, et al. Genetic and physical localization of an anthracnose resistance gene in medicago truncatula. Theor Appl Genet. 2007;116:45–52.
59. Casto AL, et al. Maturity2, a novel regulator of flowering time in sorghum bicolor, increases expression of SbPRR37 and SbCO in long days delaying flowering. PLoS ONE. 2019;14:e0212154.
60. Yang S, et al. CONSTANS is a photoperiod regulated activator of flowering in sorghum. BMC Plant Biol. 2014;14:1–15.
61. Yang S, et al. Sorghum phytochrome B inhibits flowering in long days by activating expression of SbPRR37 and SbGHD7, repressors of SbEHD1, SbCN8 and SbCN12. PLoS ONE. 2014;9:e105352.
62. Mace ES, et al. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. Theor Appl Genet. 2013;126:1377–95.
63. Tao Y, et al. Whole-genome analysis of candidate genes associated with seed size and weight in sorghum bicolor reveals signatures of artificial selection and insights into parallel domestication in cereal crops. Front Plant Sci. 2017;8:1237.
64. Zhang H, Huang Y. Genome-wide identification and characterization of greenbug-inducible NAC transcription factors in sorghum. Mol Biol Rep. 2024;51:207.
65. Ramu P, et al. Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. Theor Appl Genet. 2013;126:2051–64.
66. Menamo T, et al. Genetic diversity of Ethiopian sorghum reveals signatures of Climatic adaptation. Theor Appl Genet. 2021;134:731–42.
67. Zhang YM, et al. Editorial: the applications of new multi-locus Gwas methodologies in the genetic dissection of complex traits. Front Plant Sci. 2019;10:1–6.
68. Tang Y, et al. GAPIT version 2: an enhanced integrated tool for genomic association and prediction. Plant Genome. 2016;9:plantgenome2015–11.
69. Zhong H, et al. Multi-locus genome-wide association studies for five yield-related traits in rice. BMC Plant Biol. 2021;21:1–12.
70. Woldesemayat AA, et al. Cross-species multiple environmental stress responses: an integrated approach to identify candidate genes for multiple stress tolerance in sorghum (Sorghum bicolor (L.) Moench) and related model species. PLoS ONE. 2018;13:1–30.
71. Girma G, et al. A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits. Plant Genome. 2020;13:1–17.
72. Lehman A, Rourke NO. (2005) JMP for Basic Univariate and Multivariate Statistics A Step-by-Step Guide. *Analysis* at https://books.google.com.et/books?id=1nlApuloc0AC%26pgis=%26redir_esc=y
73. Deschamps S, et al. Genotyping-by-sequencing in plants. Biology (Basel). 2012;1:460–83.
74. Elshire RJ, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011;6:1–10.
75. Gajer P et al. (2019) Platform.
76. Girma G, et al. A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits. Plant Genome. 2020;13:e20055.
77. Browning BL, et al. A One-Penny imputed genome from Next-Generation reference panels. Am J Hum Genet. 2018;103:338–48.
78. Wang SB, et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep. 2016;6:1–10.
79. Gilmour AR, et al. ASReml update. What's new in release 3.00. VSN Int. Hemel Hempstead, UK; 2009.
80. Popat R et al. (2020) Package variability: Genetic Variability Analysis for Plant Breeding Research version 0.1.0.
81. Pariyar SR et al. (2021) Variation in root system architecture among the founder parents of two 8-way magic wheat populations for selection in breeding. Agronomy 11.
82. Alvarado G, et al. META-R: A software to analyze data from multi-environment plant breeding trials. Crop J. 2020;8:745–56.
83. Tak YG, Farnham PJ. 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. 2015;8:1–18.
84. Gong J, et al. Genome-wide identification of SNPs in MicroRNA genes and the SNP effects on MicroRNA target binding and biogenesis. Hum Mutat. 2012;33:254–63.
85. Wang J, Zhang Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. Genomics Proteom Bioinforma. 2021;19:629–40.

86. Lipka AE, et al. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012;28:2397–9.
87. Wen YJ, et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Brief Bioinform. 2018;19:700–12.
88. Mace E, et al. The sorghum QTL atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. Theor Appl Genet. 2019;132:751–66.
89. Smedley D, et al. BioMart - Biological queries made easy. BMC Genomics. 2009;10:1–12.
90. Goodstein DM, et al. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 2012;40:1178–86.
91. Esposito S, et al. Unlocking the molecular basis of wheat straw composition and morphological traits through multi-locus GWAS. BMC Plant Biol. 2022;22:1–19.
92. Yin, L. CMplot: circle manhattan plot. R package version. 2020;3(2):699.
93. Bradbury PJ, et al. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23:2633–5.
94. de Micheaux PL et al. (2013) The R software. Fundam Program Stat Anal

Publisher's note