BMC Genomics

RESEARCH Open Access

Genetic dissection of hundred-kernel weight through combined genome-wide association study and linkage analysis in tropical maize

Mengfei Shi[1,2], Fuyan Jiang[1], Ranjan K. Shaw[1], Babar Ijaz[1], Jiachen Sun[1,2] and Xingming Fan[1*]

Abstract

Background Hundred-kernel weight (HKW) is a crucial determinant of maize yield. Understanding the genetic mechanisms underlying HKW is vital for maize breeding programs aimed at enhancing productivity. This study aimed to explore the genetic basis of HKW in maize using a multi-parent population (MPP), developed by crossing the common male parent Ye107 with five female parents representing a range of kernel sizes and weights. The MPP was evaluated under two distinct environmental conditions (19DH and 19BS).

Results Genotyping-by-sequencing (GBS) identified 591,483 high-quality single nucleotide polymorphisms (SNPs), which were used for a genome-wide association study (GWAS) and linkage analysis. The GWAS revealed 21 SNPs significantly associated with HKW, with *Zm00001d028188*, a gene involved in cell wall synthesis, emerging as a key candidate located on chromosome 1. This gene, encodes Galacturonosyltransferase 1 (GAUT1) and overlapped with two identified quantitative trait loci (QTLs): *qHKW1-2* and *qHKW1-3*, which were further validated through linkage analysis.

Conclusions This study identified critical genetic loci and candidate genes, such as *Zm00001d028188*, involved in regulating HKW in maize. The findings provide valuable genomic resources for maize breeding, potentially contributing to the development of high-yielding maize varieties through an enhanced understanding of the genetic control of HKW.

Keywords Maize, Hundred-kernel weight, Multi-parent population, Genome-wide association study

Introduction

Maize (*Zea mays* L.) ranks among the most important crops globally, playing a vital role in food, feed, and industrial production [1]. As natural disasters, and both biotic and abiotic stresses continue to rise, global food security challenges are becoming more pronounced. To address these challenges and ensure food security, it is necessary to explore new germplasm resources to improve maize yield. Hundred-kernel weight (HKW) is one of the key factors affecting maize yield and is a primary target for enhancing maize production [2]. As a quantitative trait, HKW is regulated by numerous genes and influenced by environmental factors [1, 2]. Therefore, understanding the genetic variation and molecular mechanisms underlying HKW in maize is crucial for enhancing maize production.

The genetic basis of quantitative traits has traditionally been analyzed through linkage analysis in biparental populations [1, 3]. Advancements in crop genomics and marker-assisted selection have made QTL-based approaches increasingly effective in elucidating the genetic basis of kernel-related traits in maize, thereby

*Correspondence:
Xingming Fan
xingmingfan@163.com
[1] Institute of Food Crops, Yunnan Academy of Agricultural Sciences, Kunming 650205, China
[2] College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201, China

improving breeding efficiency for higher yields [4–7]. Extensive research into genetic variation in kernel yield has led to the identification of several QTLs associated with maize kernel traits [8–12]. Previous studies have identified several QTLs affecting maize HKW at low resolution [2, 13–19]. The low frequency of recombination events in the biparental populations limits the precision of QTL mapping and the development of effective marker-assisted tools. Moreover, the large and complex maize genome further complicates the fine mapping of QTLs and the cloning of associated genes [1]. However, next-generation sequencing technologies and high-throughput SNP markers have made GWAS a powerful tool for fine-mapping of QTLs [1].

GWAS have been effectively applied to discover QTLs and potential genes associated with maize kernel traits [20–24], facilitating the identification of alleles linked to ancestral cultivars and strains carrying these alleles [25, 26]. Furthermore, GWAS enables fine mapping of QTLs, as the rapid decay of linkage disequilibrium (LD) due to historical recombination events enhances the resolution of genetic loci [27]. For example, Beló et al. while performing GWAS in maize, revealed a correlation between the SNP haplotypes and oleic acid content in 553 maize inbred lines [28]. Hao et al. [29] utilized a RIL population and an association mapping panel to identify four QTLs linked to maize kernel-associated traits, including kernel length (KL), kernel width (KW), kernel thickness and HKW. Similarly, Liu et al. [30] discovered 50 QTLs associated with kernel size features in the intermated B73 × Mo17 (IBM) Syn10 doubled haploid population, with eight QTLs consistently observed across three diverse environments. Since the availability of the B73 reference genome in maize [31], GWAS has become a standard tool for exploring multigenic traits and conducting large cohort analyses [32], providing rapid and precise insights into the relationship between genetic makeup and phenotypic variation. Despite its effectiveness, the accuracy of GWAS can be compromised by false associations resulting from population structure. However, these limitations can be mitigated by using multi-parent populations (MPP) [33]. Higher Genetic Diversity in MPP which is derived from crossing multiple diverse parental lines leading to much greater allelic and genetic variation which improves the chance of detecting rare alleles and provides a wider allelic series for trait dissection. Compared to traditional biparental mapping, MPPs introduce more recombination events and reduce genetic heterogeneity, thus improving mapping resolution and reducing false positives [34]. Though MAGIC also has high diversity, but usually involves more structured crossing, which can limit some allele combinations. MPP also increases Mapping Resolution due to the accumulated recombination events across generations and more founders, MPPs offer finer mapping resolution whereas in biparental population low recombination leading to larger linkage blocks. In MAGIC though good for mapping resolution, but it depends on the number of founders and design. Some MAGIC designs are more limited in recombination than others. Therefore, MPPs holds great potential for identifying superior genotypes with enhanced HKW performance.

Tropical and subtropical maize exhibit considerable genetic diversity, which remains largely underexplored [35]. Furthermore, maize maintains a relatively conserved gene structure and function across both temperate and tropical germplasms [36]. In this study, five inbred lines from subtropical and tropical regions, known for their extensive genetic diversity and high HKW, were selected as parental lines from the Reid, Non-Reid, and Suwan heterotic groups. The objectives of this study were to: (1) investigate the genetic mechanisms underlying HKW in maize; (2) identify QTLs and SNPs associated with HKW using combined GWAS and linkage analysis; and (3) identify candidate genes and conduct functional characterization to uncover the genetic factors influencing HKW in maize, thereby laying the foundation for increasing maize yield.

Materials and methods

Plant materials, multi-parent population construction and field experimental design

In this study, five inbred lines; CML312, YML32, CML373, CML395 and Q11 were used as female parents, while the temperate inbred line Ye107 served as the common male parent for crossing. The F_1s were self-pollinated for seven generations using the single-seed method to develop a multi-parent population consisting of five RIL subpopulations: pop1 (CML312 × Ye107), pop2 (YML32 × Ye107), pop3 (CML373 × Ye107), pop4 (CML395 × Ye107), and pop5 (Q11 × Ye107). A total of 813 RILs were obtained, exhibiting extensive genetic variation. Field experiments were conducted in two ecological environments in Yunnan Province, China: Dehong (19DH; Longitude: 98.6°E, Latitude: 24.4°N) and Baoshan (19BS; Longitude: 98.9°E, Latitude: 24.9°N). The pedigrees, ecological types, and HKW data of the six parental lines are presented in Table 1.

The experimental trials were conducted using a randomized complete block design (RCBD) in the 19DH and 19BS environments, with three replicates at each location. Each experimental plot was 4.0 m in length, with 0.70 m row spacing and 25 cm plant spacing, containing 14 plants per row. The trials were managed using standard agronomic practices. For each recombinant inbred line (RIL), ten plants were randomly selected, and

Table 1 Details of parental lines used to develop the multi-parent population

Parent	Pedigree	Heterotic Group	Ecological Type	Hundred kernel weight (g)	Subpopulations	Number of RILs
Ye107	Derived from US hybrid DeKalb XL80	Reid	Temperate	21	-	-
CML312	S89500-F2-2–2-1–1-B*5–2-16-1(DH)	Non-Reid	Subtropical	24	pop1 (CML312 ×Ye107)	125
YML32	Suwan 1(S)C9-S8-346–2 (Kei 8902)–3-4-4–6	Suwan	Tropical	22	pop2 (YML32 ×Ye107)	156
CML373	P43SR-4–1-1-2-1-B-8–1-B-B-B-1–1–3–6(DH)	Non-Reid	Subtropical	26	pop3 (CML373 ×Ye107)	156
CML395	90323B-1-B-1-B*4–1-1-2-1 (DH)	Non-Reid	Tropical	27	pop4 (CML395 ×Ye107)	196
Q11	Derived from US hybrid	Reid	Temperate	23	pop5 (Q11 ×Ye107)	180

All the plant materials used in this study are maize recombinant inbred lines, which are commonly employed in research worldwide. Therefore, registration of these materials is not required

hundred-kernel weight (HKW) was measured by sampling 100 seeds per plant. Each measurement were replicated three times, and the average HKW was calculated as the mean HKW of the RIL. To mitigate the impact of environmental factors on phenotypes, a mixed linear model of the lme4 package in R (v3.3.3) [37] was used to analyze the raw phenotypic data of each RIL subpopulation across two different environments. This analysis was conducted using Best Linear Unbiased Prediction (BLUP) values, with the following formula:

$$Y = \mu + Line + Loc + (Line \times Loc) + \text{Rep}\ (Loc) + \varepsilon \quad (1)$$

where, Y represents the phenotype, μ is the intercept, *Line* refers to the effects of different genotypes, and *Loc* represents environmental effects. Rep denotes replications, while ε captures the residual error. $Line \times Loc$ illustrates the interaction between the genotype and environment, and Rep (Loc) accounts for the nested effect of replications within each environment.

Genomic DNA isolation and Genotyping-by-sequencing (GBS)

Genomic DNA was extracted from seedlings leaves using the cetyltrimethylammonium bromide (CTAB) method [38]. The DNA was then digested and purified using the restriction enzymes PstI and MspI (New England BioLabs, Ipswich, MA, USA). Barcode adapters were ligated to the DNA fragments using T4 ligase (New England BioLabs). Genotyping-by-sequencing (GBS) libraries were constructed following the GBS protocol [39]. The ligated samples were pooled and purified using the QIAquick PCR Purification Kit (QIAGEN, Valencia, CA, USA). Polymerase chain reaction (PCR) amplification was performed using primers complementary to the adapters. PCR products were then purified and quantified using the Qubit dsDNA HS Assay Kit (Life Technologies). Fragments between 200 and 300 bp were size-selected using the E-gel system (Life Technologies). The library concentration was measured using a Qubit 2.0 fluorometer and the Qubit dsDNA HS Assay Kit (Life Technologies). Sequencing was performed on an Ion Proton sequencer (Life Technologies, software version 5.10.1) using P1v3 chips. Prior to analysis in TASSEL 5.0, 80 poly (A) bases were added to the 3' ends of all sequencing reads. Plink v1.9 was used to filter SNPs, with parameters set to -geno 0.2 and -maf 0.05 to exclude loci with missing rates higher than 20% and minimum allele frequencies lower than 5%. SNP calling was performed with the Genome Analysis Toolkit software [40] with the maize B73 reference genome [41] and annotation was completed with the ANNOVAR software [42].

Phylogenetic tree, principal component analysis, and LD analysis

The TreeBeST1.9.2 (https://github.com/Ensembl/treebest) software was used to calculate the distance matrix between RILs and construct a phylogenetic tree.

Population structure analysis was performed using Admixture v1.3.0 [43]. The K value was initially set for cross-validation, with the assumption that the K value showing the minimum cross-validation error corresponds to the optimal number of clusters.

Principal component analysis (PCA) was performed using GCTA v1.5.0, and the results were visualized using the scatterplot3 d package in R v4.3.2 [44].

Linkage disequilibrium (LD) decay was assessed using genome-wide SNPs with Pop LD decay v3.42 [45]. The R^2 value, which estimate the extent of LD between markers (spanning from 0 to 1, with values closer to 1 indicating stronger LD), was calculated using default parameters. The LD decay plot was generated using the'Plot _ OnePop.pl'script. LD decay analysis helped determine the minimum number of markers required

for GWAS, as well as the detection efficiency and accuracy of the study [46].

GWAS for HKW

GWAS analysis was conducted using the mixed linear model (MLM) implemented in the Genome-wide Efficient Mixed Model Association (GEMMA) statistical software package (v0.98.3) [47]. The MLM analysis was performed using the following formula:

$$y = X_a + Z_\beta + W_\mu + e \quad (2)$$

where y represents the phenotypic trait, X is the matrix of fixed effects, α is the estimated parameter of fixed effects, Z is the association matrix of SNP, β is the effect of SNPs, W is the matrix of random effects, μ is the vector of random individual effects, e is the random residual. The residuals (e) were assumed to have a mean of zero and variance $e \sim (0, \delta_e^2)$ [48]. A significance threshold of $-\log10(p) > 4.5$ was set to identify SNPs significantly associated with maize HKW. SNPs meeting or exceeding the threshold were extracted using bedtools 1.7.0 [49]. The GWAS results were visualized using the Manhattan and QQ plots generated with the CMplot 3.6.2 package in R v4.3.2 [44]. The significance levels of individual SNPs was adjusted using the Bonferroni correction, a multiple comparison test to control for type I error and reduce false positives [50].

Candidate genes associated with HKW were identified within a 50 kb region upstream and downstream of each SNP, using the B73 reference genome from MaizeGDB (http://www.maizegdb.org/, retrieved on August 15, 2024). Functional annotation was performed to gain insights into the biological roles of the candidate genes.

Construction of linkage maps and QTL mapping

Allelic data of the SNPs identified through GBS to construct the linkage maps for the five RIL subpopulations. Linkage maps were constructed using JoinMap v4.0, with a minimum logarithm of odds (LOD) threshold set at ≥ 5.0.

QTL analysis for HKW was performed using QTL Cartographer 2.0 using the composite interval mapping ()CIM method [51]. A permutation test with 1000 iterations was conducted to determine the LOD threshold, which was set at 2.5 ($p < 0.05$) [52]. The proportion of phenotypic variation explained (PVE) by each QTL was calculated using the correlation coefficient square (R^2). The QTLs were designated with a "q" prefix, followed by the trait abbreviation (HKW), the chromosome number, and the marker position.

Haplotype analysis

Haploview v4.2 software [53] was used to analyze HKW-related genes across two environments and to identify the superior haplotypes. A high-density genome-wide SNP dataset was employed to generate a haplotype map, highlighting haplotypes significantly associated with HKW based on SNP locations and LD patterns. Genes within these haplotypes were further characterized to assess evaluate functional roles and contributions to HKW regulation.

Results

Phenotypic analysis of HKW

This study developed a multi-parent population (MPP) comprising 813 $F_{2:7}$ RILs derived from five parental lines to investigate the genetic basis of HKW in maize (Fig. 1a). Kernel weight differences among the parents of the MPP are shown in Fig. 1b, with the male parent Ye107 exhibiting significantly lower HKW compared to all female parents (Fig. 1c). ANOVA revealed significant variations in HKW across the five subpopulations (Fig. 1d-e) in two environments: Dehong (19DH), Baoshan (19BS). Descriptive statistics and the coefficient of variation (CV) for HKW in the five RIL subpopulations across two environments (19DH and 19BS) are provided in Table 2. The CV for HKW ranged from 14.3% (pop2) to 26.8% (pop4) in the 19DH environment and from 15.6% (pop2) to 28.9% (pop4) in the 19BS environment. To assess the phenotypic robustness, the correlation coefficients for HKW between the 19DH and 19BS environments were calculated, revealing highly significant positive correlations of 0.767, 0.734, 0.987, 0.979, and 0.975 ($p < 0.001$) across the five subpopulations (Fig. 2). Additionally, the skewness of each subpopulation was below 1.0, indicating minimal bias in the distribution of HKW. ANOVA further showed significant effects of genotype and environment on HKW across the two regions ($P < 0.0001$ and $P < 0.01$, respectively), as well as a significant genotype × environment interaction ($P < 0.0001$) (Table 3). These results indicate that both genetic and environmental factors contributed substantially to HKW variation in maize. Overall, the MPP demonstrated extensive variation in HKW, with consistency across the environments, suggesting the reliability of the phenotypic data for subsequent GWAS and QTL mapping analyses.

Pearson correlation analysis ($p < 0.001$) revealed a strong and significant correlation in the performance of RILs across different environments, with correlation coefficients ranging from 0.73 to 0.99 within each population (Fig. 2). The correlation indicated that the RILs exhibited consistent hundred-kernel weight across environments, displaying stable phenotypes under

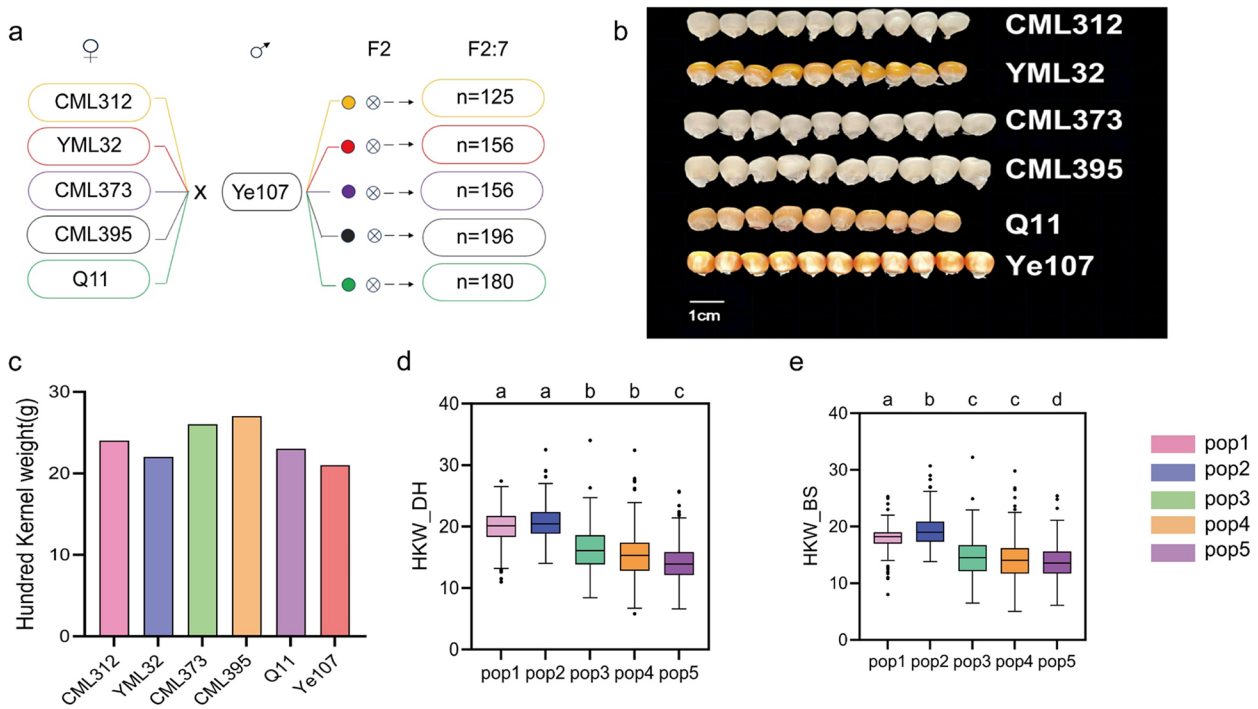

Fig. 1 Population development and comparison of the hundred-kernel weight. **a** The population construction scheme of the multi-parent population; (**b**) Kernel phenotypes of each parent; (**c**) Hundred-kernel weight of the parental lines; **(d)** Hundred-kernel weight of the RIL subpopulations in the 19DH environment; (**e**) Hundred-kernel weight of the RIL populations in the 19BS environments

Table 2 Phenotypic evaluation of the RIL subpopulations for HKW across two environments

Populations	Environment	Mean	Standard Deviation	Skewness	Kurtosis	Coefficient of Variation (CV)	Correlation Coefficient (r)
pop1	19DH	19.854	3.243	−0.432	0.272	0.163	0.767
	19BS	18.09	3.192	−0.267	0.846	0.176	
pop2	19DH	20.692	2.965	0.701	1.43	0.143	0.734
	19BS	19.38	3.03	0.968	1.47	0.156	
pop3	19DH	16.356	3.871	0.674	1.979	0.237	0.987
	19BS	14.605	3.769	0.716	2.376	0.258	
pop4	19DH	15.525	4.163	0.766	1.49	0.268	0.979
	19BS	14.219	4.108	0.711	1.441	0.289	
pop5	19DH	14.141	3.46	0.63	0.823	0.245	0.975
	19BS	13.649	3.295	0.581	1.058	0.241	

varying environmental conditions. This suggested that HKW is largely unaffected by environmental factors, underscoring the strong genetic influence that ensures trait consistency. Collectively, these findings provide a foundation for further QTL mapping and association studies to identify genetic loci linked to HKW in maize.

SNP characterization and LD decay analysis

The genotyping-by-sequencing (GBS) revealed 591,483 high-quality SNPs, distributed across the ten maize chromosomes. The number of SNPs on chromosomes 1–10 were as follows: 82,889; 67,249; 67,934; 75,637; 57,899; 47,683; 53,494; 50,217; 44,840; and 43,631, respectively (Fig. 3a). Chromosome 1 contained the

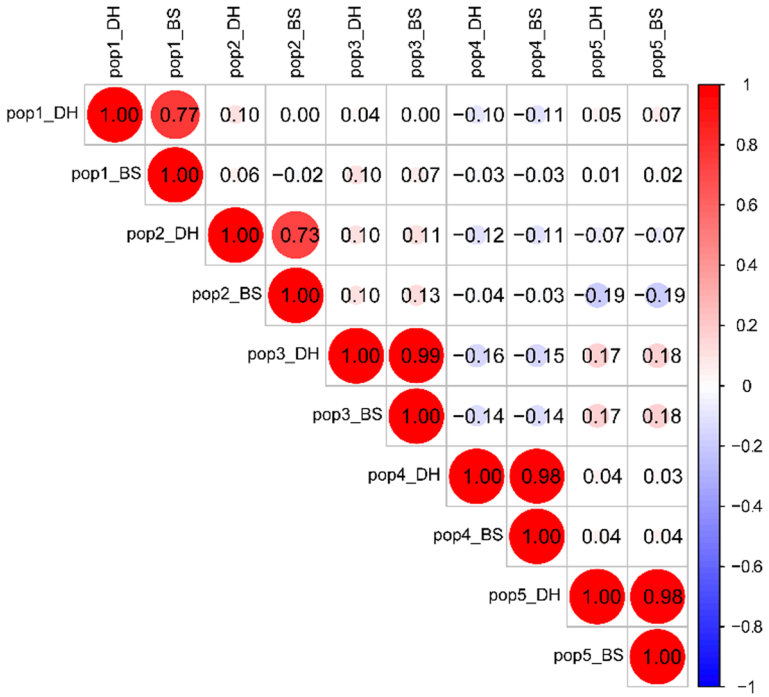

Fig. 2 Heat map illustrating the overall correlation of the RILs within each of the five subpopulations across two environments

Table 3 Analysis of variance (ANOVA) for hundred-kernel weight (HKW) of the multi-parent population (MPP) at two locations in Yunnan Province, China

Source of variation	df	Sum of squares	Mean squares	F value
Genotype	812	93,726	115.4****	11.75
Environment	1	2403	2403**	108.6
Genotype × Environment	812	10,219	12.59****	1.293
Error	1624	15,801	9.73	

** indicates $p < 0.01$

**** indicates $p < 0.0001$

highest number of SNPs (82,889), while chromosome 10 had the lowest (43,631). The SNP density per mega base (Mb) for chromosomes 1 to 10 was 269.99, 265.05, 275.11, 288.26, 306.30, 258.60, 273.99, 293.31, 277.25, and 280.65, respectively, indicating an even distribution of SNPs across the genome. In the filtered SNP dataset, the average missing rate was ≤0.19, and the average minor allele frequency (MAF) was 0.20, indicating that the dataset was suitable for subsequent GWAS analysis (Fig. 3b, c).

The 591,483 SNPs were also used to evaluate the LD decay in the multi-parent population. At an R^2 value of 0.38, LD was estimated to decay at a physical distance of approximately 50 kb (Fig. 3d). This rapid LD decay suggested that selection has influenced genetic variation across the genome, leading to distinct LD block that can aid in identifying loci responsible for traits of interest in future GWAS studies.

Phylogenetic tree, principal component analysis, and population structure of HKW

The phylogenetic tree demonstrated that the 813 RILs were grouped into five major subgroups (Fig. 4a). The principal component analysis (PCA) revealed that the RILs could be divided into five clusters, each representing one subpopulation. However, noticeable overlap was observed among populations, potentially indicating intra-population variation or outliers based on PC1, PC2, and PC3 (Fig. 4b).

Population structure analysis confirmed that 813 RILs were divided into five subgroups when K=5 (Fig. 4c). Overall, the findings from the population structure and phylogenetic analyses were in agreement with the principal component analysis. Nonetheless, the presence of heterogeneous clusters or hybrids with differing allele frequencies indicated gene introgression during the breeding process.

GWAS for HKW

Significant variation in the HKW was observed among the RILs of the MPP in both the 19DH and 19BS environments (Fig. 5). GWAS was performed using the

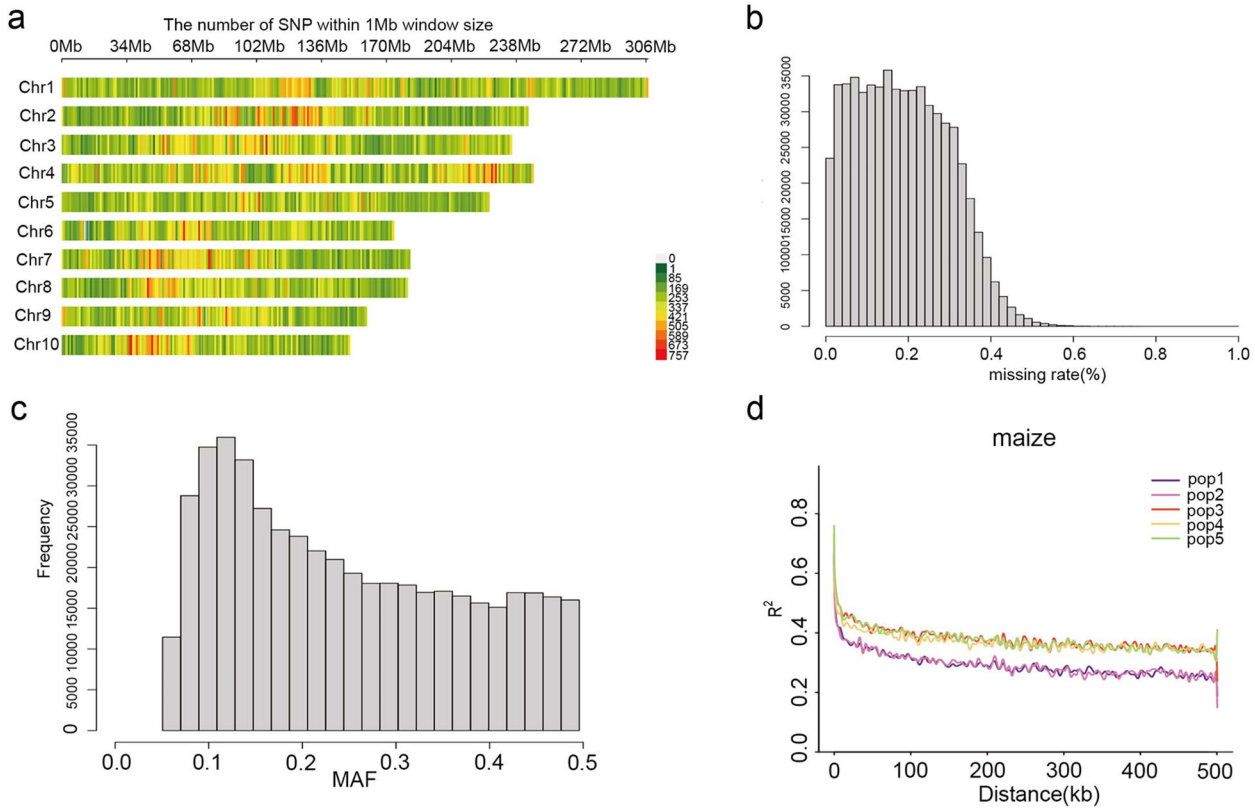

Fig. 3 Genetic diversity and LD decay map. **a** SNP distribution across chromosomes in 1-Mb intervals, with SNP count represented by a scale from green-to-red colour scale; (**b**) Distribution of SNP missing rate percentage; (**c**) Distribution of MAF in the RILs; (d) LD decay (R^2) plot across the five RIL subpopulations

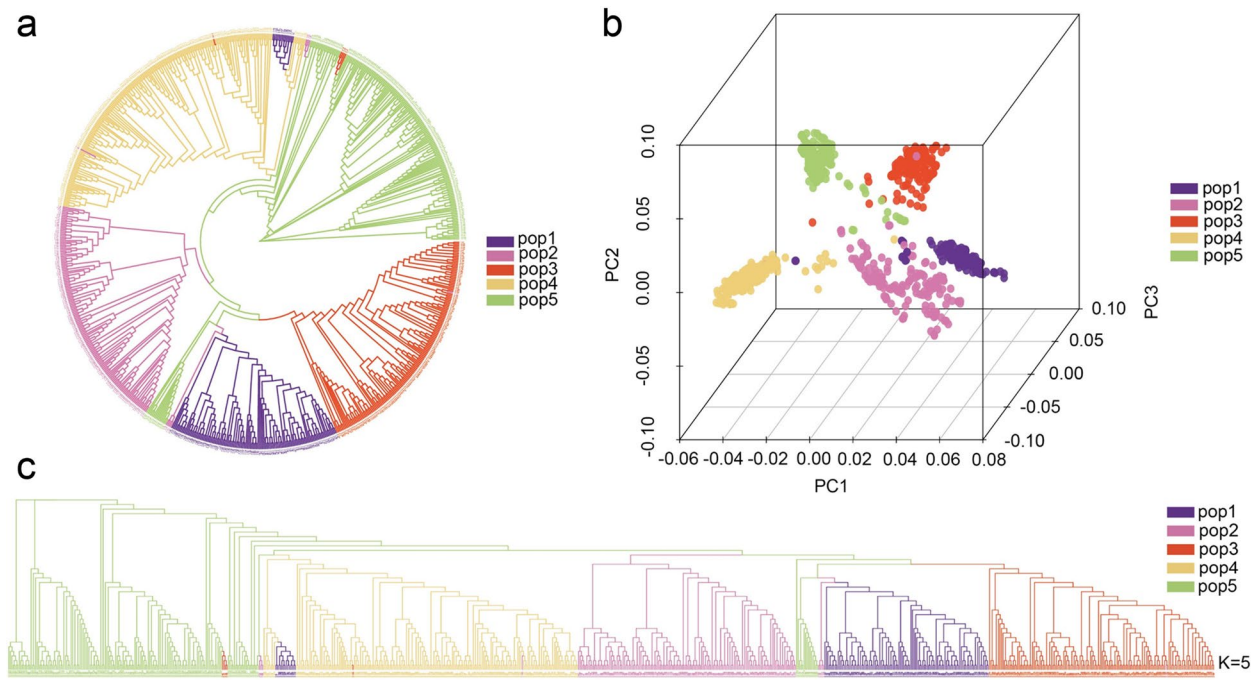

Fig. 4 Genetic diversity analysis of the 813 RILs of the multi-parent population. **a** Phylogenetic tree; (**b**) Principal component analysis; (**c**) Bayesian clustering diagram of the 813 RILs at K = 5

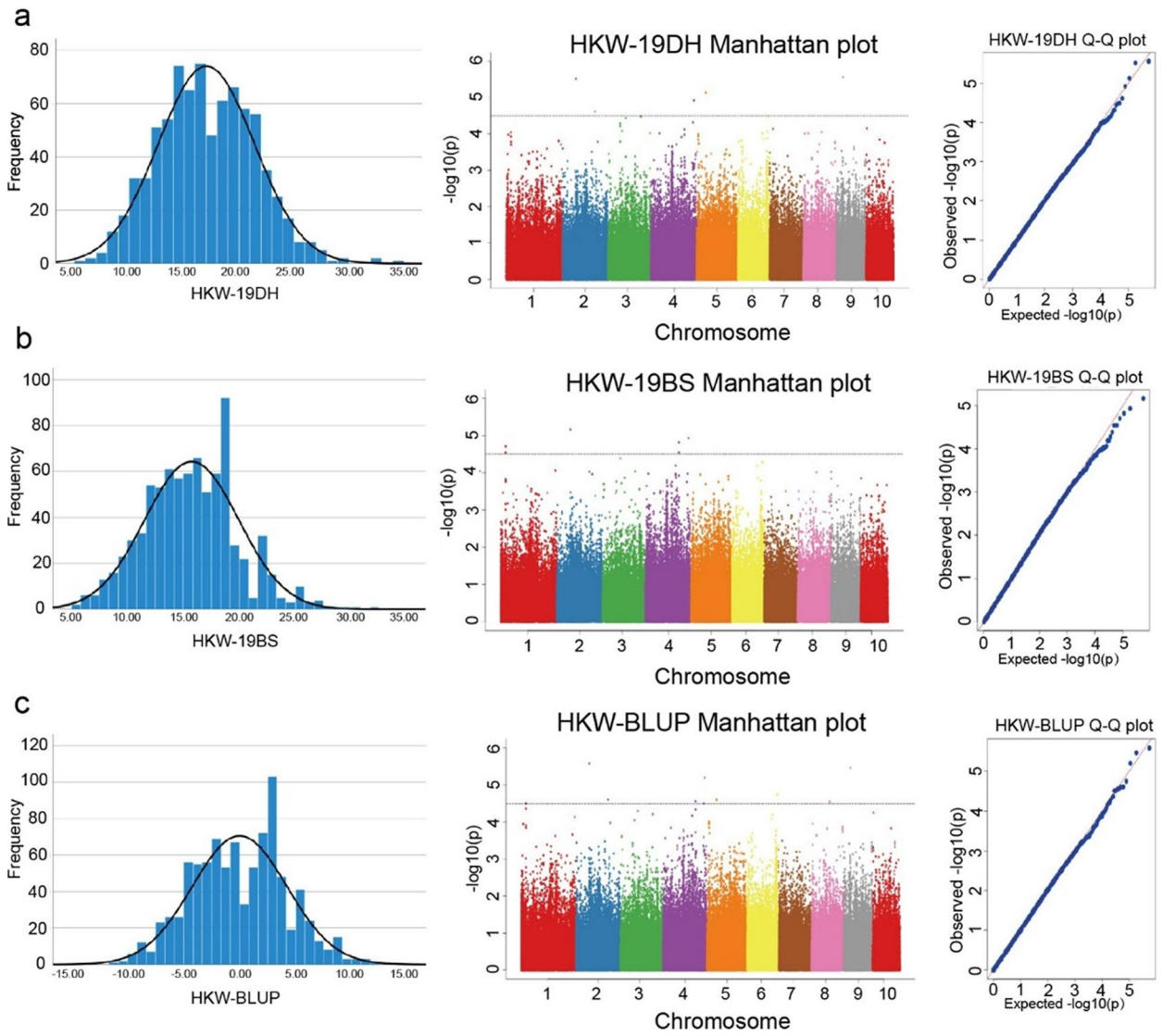

Fig. 5 GWAS results for HKW. **a** Frequency distribution, Manhattan plot, and Q-Q plot of HKW in the 19DH environment; (**b**) Frequency distribution, Manhattan plot, and Q-Q plot of HKW in the 19BS environment; (**c**) Frequency distribution, Manhattan plot, and Q-Q plot of HKW based on BLUP values

MLM, which incorporated population structure and kinship matrix as covariates to minimize the risk of false-positives. The Q-Q plots confirmed effective control of false-positive associations for HKW (Fig. 5a-c). At a significant threshold of -log10(p) $\geq$4.5, 21 SNPs significantly associated with HKW were identified (Table 4). Among these, five SNPs were identified in the 19DH environment, six in the 19BS environment, and 10 based on the BLUP values (Fig. 5; Table 4). The significant SNPs were distributed across chromosomes 1, 2, 4, 5, 6, 8, and 9). Notably, several SNPs were consistently detected across multiple environments. For instance, SNP-25750352 on chromosome 1 was identified in both the 19BS environment and BLUP values. SNP-73091888 on chromosome 2 was detected in the 19DH and 19BS environments, as well as for the BLUP values, while SNP-176207154 on chromosome 2 was consistently identified in both the 19DH environment and BLUP values. Similarly, SNP-233438083 on chromosome 4 was identified across the 19DH, 19BS environments, and BLUP values, and SNP-182679880 on chromosome 4 was identified in the 19BS environment and BLUP values. Additionally, SNP-51610984 on chromosome 5 and SNP-36178213 on chromosome 9 were both identified in the 19DH environment and BLUP values.

Table 4 SNPs significantly associated with HKW identified across three environments

Env	Chr	SNP	ref	alt	-log10(P)	Additive Effect	Dominance Effect
19DH	2	73,091,888	G	A	5.52	−3.84	0.22
	2	176,207,154	G	A	4.61	-	-
	4	233,438,083	C	T	4.92	−0.5	0.82
	5	51,610,984	T	G	5.13	1.2	0.53
	9	36,178,213	G	A	5.56	0.9	0.68
19BS	1	25,750,352	T	A	4.7	0.8	−0.59
	1	25,750,317	C	T	4.54	−1.7	−0.77
	2	73,091,888	G	A	5.17	−3.44	1.42
	4	182,679,880	T	C	4.82	1.37	0.49
	4	233,438,083	C	T	4.93	−0.18	0.76
	4	182,680,064	A	G	4.55	−1.14	0.47
BLUP	1	25,750,352	T	A	4.5	0.81	0.2
	2	73,091,888	G	A	5.58	−2.97	0.65
	2	176,207,154	G	A	4.6	-	-
	4	182,679,880	T	C	4.56	0.65	0.57
	4	228,845,293	C	T	4.51	−0.51	0.5
	4	233,438,083	C	T	5.19	−0.25	0.95
	5	51,610,984	T	G	4.59	1.1	0.72
	6	165,885,324	A	T	4.74	0.92	0.09
	8	101,736,982	T	A	4.54	−1.06	0.38
	9	36,178,213	G	A	5.46	0.76	0.68

QTL mapping of HKW

QTL mapping for HKW was conducted in five subpopulations across two distinct environments. The LOD threshold for identifying QTL was set at ≥ 2.5 [54]. As the SNPs identified through GWAS overlapped with the QTL intervals identified in pop3 (CML373 × Ye107), the linkage map construction results of pop3 are described here. A high-density genetic map of pop3 was constructed using 2,265 polymorphic SNPs, with a total genetic distance of 4926.27 cM, and an average inter-marker distance of 2.17 cM. Chromosome 1 contained the highest number of SNPs (327), while chromosome 6 had the lowest (163) (Table 5). The genetic map information for the remaining four subpopulations is also provided Table 5.

In pop1, three QTLs for HKW were identified: *qHKW4-1*, *qHKW1-1,* and *qHKW6-1*, explaining 7.3%, 7.8%, and 7.4% of the phenotypic variance, respectively (Fig. 6; Table 6). Among them, QTL *qHKW4-1* on chromosome 4 exhibited the highest LOD (3.26) and the largest additive effect (0.97) across both environments. Two additional QTLs (*qHKW1-1* on chromosome 1 and *qHKW6-1* on chromosome 6) were also identified in the 19BS environment, with additive effects of −1.010 and 0.923, respectively.

In pop2, two QTLs, *qHKW3-1* on chromosome 3 and *qHKW4-2* on chromosome 4 (LOD 4.3), were detected in the 19DH environment (Fig. 6), accounting for 6.9% and 6.5% of the phenotypic variation, respectively (Table 6).

Table 5 Summary statistics of genetic linkage maps for the five subpopulations

Populations	Total SNP_ number	SNP_number (max)	SNP_number (min)	Length (cM)	Inter-marker Distance (cM)
pop1	981	Chr1 (180)	Chr10 (38)	1045.83	1.07
pop2	638	Chr1 (115)	Chr8 (37)	581.28	0.91
pop3	2265	Chr1 (327)	Chr6 (163)	4926.27	2.17
pop4	2021	Chr1 (278)	Chr9 (128)	4593.45	2.27
pop5	1443	Chr1 (226)	Chr10 (71)	2875.24	1.99

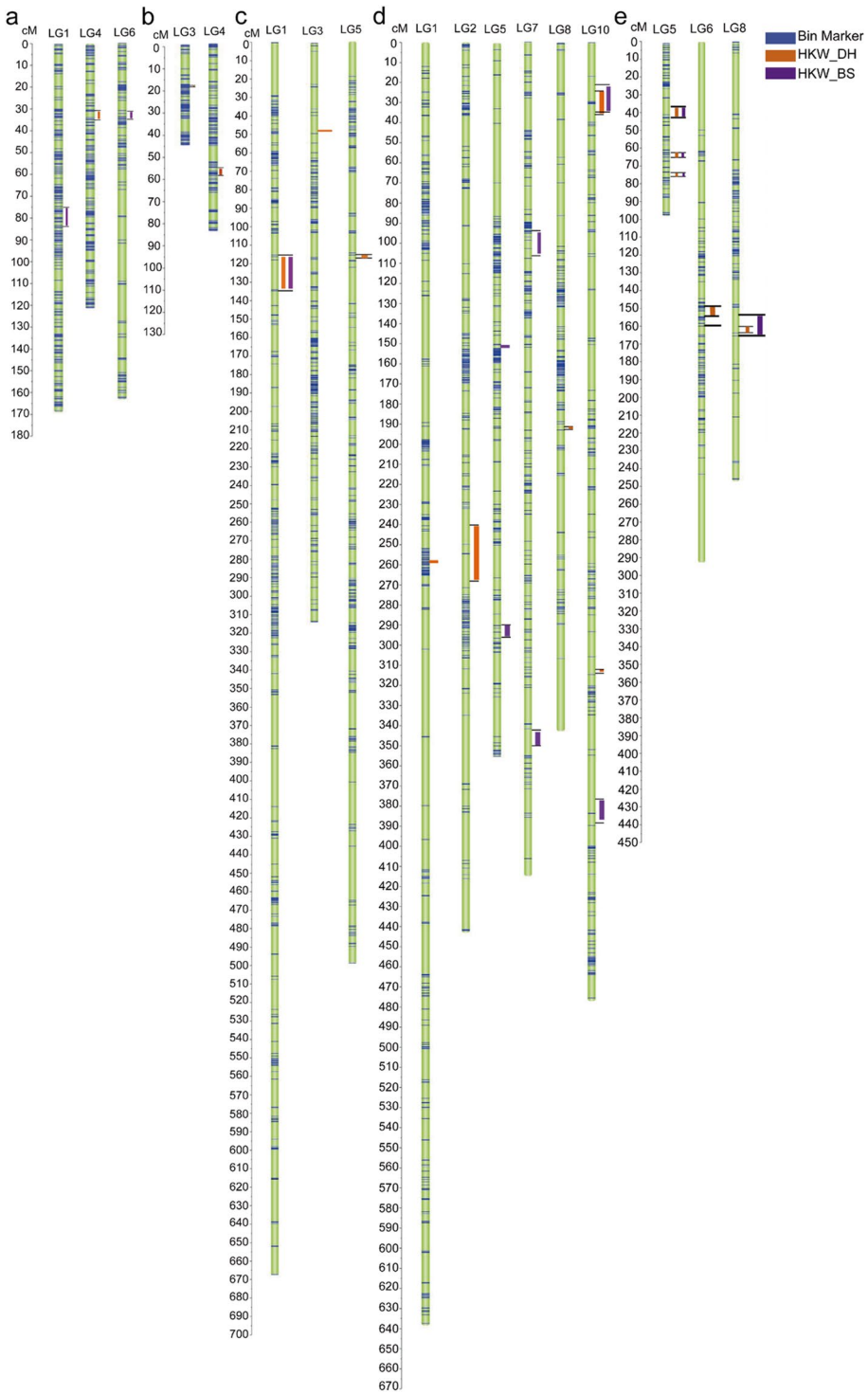

Fig. 6 QTL mapping for HKW in five subpopulations (**a**) Pop1; (**b**) Pop2; (**c**) Pop3; (**d**) Pop4; (**e**) Pop5. Blue represents bin markers, orange indicates the 19DH environment, and purple represents the 19BS environment

In pop3, four QTLs: *qHKW1-2*, *qHKW3-2*, *qHKW5-1* and *qHKW1-3* were identified, explaining 6.3%, 5.6%, 5.9% and 6.2% of phenotypic variation, respectively (Fig. 6; Table 6). Notably, *qHKW1-3* on chromosome 1 displayed the highest LOD score of 4.34.

In pop4, eleven QTLs were detected, explaining phenotypic variation ranging from 4.7%–8.0% across both

Table 6 List of QTLs linked to HKW identified in five RIL subpopulations

Populations	QTL	Chr	Environment	Position (cM)	Mapping Interval (Mb)	LOD	additive	R^2
pop1	qHKW4-1	4	19DH	32.43–35.25	174,665,020–179,452,217	3.26	0.972	0.073
	qHKW1-1	1	19BS	76.36–84.14	129,544,100–161,031,973	2.72	−1.01	0.078
	qHKW6-1	6	19BS	31.99–35.39	117,608,108–123,131,912	3.2	0.923	0.074
pop2	qHKW3-1	3	19DH	18.45–18.76	194,541,734–197,589,267	2.79	0.893	0.069
	qHKW4-2	4	19DH	58.6–65.52	21,832,586–56,256,452	4.3	−0.951	0.065
pop3	qHKW1-2	1	19DH	116.6–135.58	16,415,666–37,617,674	3.88	0.939	0.063
	qHKW3-2	3	19DH	48.86	172,317,351	2.5	−0.899	0.056
	qHKW5-1	5	19DH	118.64–119.21	13,061,656–16,193,567	2.55	−0.872	0.059
	qHKW1-3	1	19BS	116.6–135.58	16,415,666–37,617,674	4.34	0.92	0.062
pop4	qHKW1-4	1	19DH	258.5	110,271,115	4.18	−8.372	0.078
	qHKW2-1	2	19DH	241.93–268.55	121,041,846–143,902,975	4.82	−1.44	0.074
	qHKW8-1	8	19DH	191.54–192.67	70,331,236–70,065,277	2.79	1.062	0.051
	qHKW10-1	10	19DH	26.58–39.67	93,255,109–94,521,734	3.61	1.27	0.061
	qHKW10-2	10	19DH	323–325.06	99,629,009–115,803,388	2.74	−1.511	0.051
	qHKW5-2	5	19BS	153.62	94,062,594	2.58	3.75	0.047
	qHKW5-3	5	19BS	293.12–297.41	209,017,559–212,846,310	2.84	−1.03	0.05
	qHKW7-1	7	19BS	98.19–109.01	102,426,594–103,961,439	4.51	−1.809	0.055
	qHKW7-2	7	19BS	351.36–358.36	79,299,094	3.02	1.318	0.069
	qHKW10-3	10	19BS	23.58–37.67	93,255,109–94,521,734	4.1	1.402	0.08
	qHKW10-4	10	19BS	385.38–400.4	32,566,456–34,038,128	3.6	1.166	0.062
pop5	qHKW5-4	5	19DH	35.62–42.75	197,451,243–206,096,177	3.31	0.851	0.053
	qHKW5-5	5	19DH	63.02–64.65	18,288,409–135,237,637	7.12	−6.334	0.139
	qHKW5-6	5	19DH	73.71–74.71	122,872,553	3.1	−4.219	0.09
	qHKW6-2	6	19DH	149.38–157.96	42,219,374–62,328,938	3.47	−0.961	0.054
	qHKW8-2	8	19DH	160.2–163.53	76,588,355–95,585,833	2.61	−0.793	0.048
	qHKW5-7	5	19BS	35.62–42.75	197,451,243–206,096,177	3.2	0.785	0.049
	qHKW5-8	5	19BS	63.02–64.65	18,288,409–135,237,637	7.95	−6.371	0.152
	qHKW5-9	5	19BS	73.71–74.71	122,872,553	3.66	−4.28	0.102
	qHKW6-3	6	19BS	149.38–161.25	41,513,372–62,328,938	3.56	−0.845	0.046
	qHKW8-3	8	19BS	154.2–164.53	76,588,355–95,585,833	3.48	−0.83	0.057

environments. The QTL *qHKW2-1* showed the highest LOD value of 4.82 in both environments (Fig. 6). Among these, six QTLs (*qHKW8-1*, *qHKW10-1*, *qHKW5-2*, *qHKW7-2*, *qHKW10-3* and *qHKW10-4*) exhibited positive additive effects, suggesting they positive regulatory role in HKW.

In pop5, ten QTLs were detected, explaining phenotypic variation ranging from 4.6% to 15.2%. The QTL *qHKW5-8* had the highest LOD of 7.95 across both environments. Additionally, *qHKW5-4* and *qHKW5-7* exhibited positive additive effects, indicating positive effects on HKW (Fig. 6).

Identification of candidate genes regulating HKW in maize through combined GWAS and QTL mapping

Through co-localization analysis of GWAS and QTL mapping, candidate genes regulating HKW in maize were identified. A comparative analysis of the QTL mapping and GWAS revealed a SNP on chromosome 1 (SNP-25750352), identified through GWAS in both the 19DH and 19BS environments which overlapped with the QTL intervals of *qHKW1-2* (Fig. 7a) and *qHKW1-3* (Fig. 7b) identified in pop3 for the 19DH and 19BS environments, respectively. These QTLs explained 6.2%−6.3% of the phenotypic variation for HKW (Table 6). A comprehensive screening of candidate genes within a 50 kb region upstream and downstream of SNP-25750352 led to the identification of four potential candidate genes; *Zm00001d028185*, *Zm00001d028186*, *Zm00001d028187*, and *Zm00001d028188* (Table 7). Functional annotation was conducted using MaizeGDB, InterPro, UniProt, and NCBI databases, as well as published relevant literatures. Based on functional annotation, one gene, *Zm00001d028188* was identified as a likely candidate related to HKW (Table 8) (Fig. 7c). *Zm00001d028188*, which spans 6,693 bp, was located was located 9,968

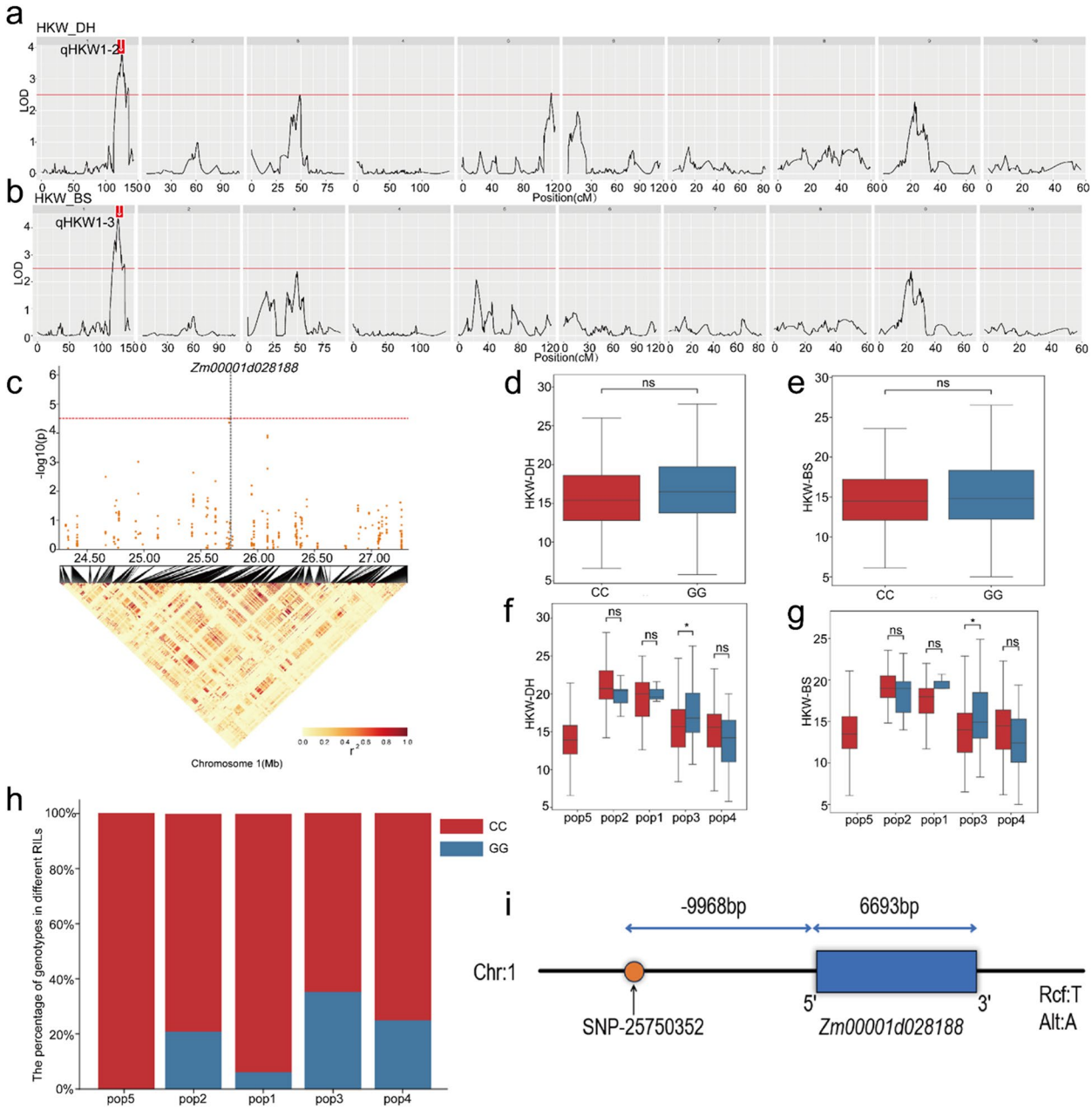

Fig.7 Identification of candidate genes associated with maize HKW. **a** QTLs identified in pop3 across different chromosomes in the 19DH environment; (**b**) QTLs identified in pop3 across different chromosomes in the 19BS environment; (**c**) LD block of candidate genes of *Zm00001d028188*; (**d**) Comparison of two haplotypes (CC and GG) for HKG in the 19DH environment and (**e**) 19BS environment; (**f**) Comparison of the two haplotypes for HKW in each subpopulation in the 19DH environment and (**g**) 19BS environment; (**h**) Proportion of the two haplotypes in five subpopulations; (**i**) Genomic position of *Zm00001d028188* and its associated SNP

bp downstream of SNP-25750352 (Fig. 7i). This gene encodes Galacturonosyltransferase 1 (GAUT1), a key enzyme responsible for transferring galacturonic acid into UDP-GalA (uridine 5'-diphosphate galacturonic acid), which is essential for the synthesis of pectin polysaccharides in plant cell wall components.

Haplotype analysis

Through QTL mapping and GWAS analyses, candidate genes associated with HKW in maize were identified. A comparative analysis of the QTL mapping and GWAS results revealed that a SNP on chromosome 1 (SNP-25750352) was located within the QTL intervals

Table 7 Candidate genes located within 50 kb of significant SNPs and Functional Annotation

Marker	Allele	Start	End	Candidate Gene	Functional Annotation
SNP-25750352	T/A	25,738,664	25,741,208	Zm00001 d028185	Remorins are plant-specific plasma membrane-associated proteins
		25,742,203	25,745,382	Zm00001 d028186	Pentatricopeptide repeat (PPR) proteins
		25,755,280	25,759,748	Zm00001 d028187	O-Glycosyl hydrolases (3.2.1.)
		25,760,320	25,767,012	Zm00001 d028188	Galacturonosyltransferase 1 (GAUT1)

Table 8 Candidate genes identified through combined GWAS and QTL mapping analyses

Marker	Chromosome	Position	Mapping Interval	Candidate Gene	Gene Annotation
SNP-25750352	1	25750352 bp	25,700,317–25,800,352	Zm00001 d028188	Galacturonosyltransferase 1 (GAUT1)
qHKW1-2、qHKW1-3	1	135.58 cM	16,415,666–37,617,674		

of *qHKW1-2* (Fig. 7a) and *qHKW1-3* (Fig. 7b). These QTLs explained 6.2%–6.3% of the phenotypic variation for HKW, indicating the strong reliability and consistency of these QTL intervals (Table 6).

Haplotype analysis was conducted for the significant SNP, SNP-25750352, which was located in close proximity to the candidate gene *Zm00001d028188* on chromosome 1 (Fig. 7c). *Zm00001d028188*, spanning 6.693 kb was located 9.968 kb downstream of SNP-25750352 (Fig. 7i). The analysis was conducted to identify the superior haplotype involved in regulating HKW in maize. Haplotype analysis revealed two haplotypes: Hap A (CC) and Hap B (GG) in 813 RILs from the MPP. Among these, haplotype B exhibited higher HGW, with a significant difference observed between Hap A and Hap B (Fig. 7 d, e) in both the 19DH and 19BS environments (Fig. 7 f, g). Therefore, Hap B was hypothesized to be the superior haplotype enhancing HKW in maize. Both haplotypes exhibited varying frequencies across different subpopulations (Fig. 7h). Hap A (CC) was observed across all subpopulations, while the superior haplotype, Hap B (GG) was present in pop1, pop2, pop3, and pop4, with pop3 exhibiting the highest frequency of Hap B (Fig. 7 h). This suggested that the high HKW observed in the pop3 subpopulation may be attributed to the Hap B (GG) haplotype. The parental line CML373 is likely to carry this superior haplotype, indicating that pop3 warrants further attention in breeding programs.

These findings provide valuable insights into the genetic and functional basis of HKW, highlighting *Zm00001d028188* and its associated SNP as potential targets for improving maize yield traits through marker-assisted selection.

Table 9 Additive and dominant effects of SNP-25750352 across different environments

Env	SNP	Chr	Position	additive effect	dominant effect
19DH	SNP-25750352	1	25,750,352	1.166	–0.453
19BS				0.815	–0.395
BLUP				1.251	–0.406

Additive and dominant effect analysis of SNPs associated with HKW

Analysis of the SNP, SNP-25750352, associated with HKW, revealed strong additive and dominant effects under different environmental conditions (Table 9). The results indicated that this SNP consistently demonstrated a positive effect on maize HKW, highlighting its potential role. This finding underscores the role of both additive and dominant effects in controlling maize HKW. Notably, the additive effects observed in the 19DH, 19BS environments, as well as in the BLUP values were positive, indicating that the allele carried by the parent Ye107 contributed to increased HKW.

Discussion

Comparison of significant loci linked to HKW with previously reported QTLs

A key objective of maize breeding programs is to enhance grain production. Factors such as plant density, the number of ears per plant, and ear weight determine grain yield per unit area. Ear weight, in turn, is influenced by the number of rows per ear, the number of kernels per row, and kernel weight. These traits are

inherited quantitatively and are affected by genotype, environments, and genotype-by- environment interactions. In this study, a multi-parent population (MPP) consisting of 813 RILs were developed using five tropical and subtropical inbred lines as female parents and one temperate inbred line as the male parent. Through QTL mapping and GWAS, multiple QTLs and one candidate gene related to HKW were identified. Previous studies have reported several QTLs linked to HKW across various populations and environments, located at different physical positions. In this study, a candidate gene on chromosome 1 was identified, and comparison with previous studies revealed overlapping QTLs mapped to chromosome 1. For instance, Hao et al. [29] reported *qHKW1-1*, located within the physical interval of 14,638,761–16,825,635 bp in an RIL population, with an LOD score of 4.60 and a PVE of 7.39%, which overlapped with the QTL intervals (*qHKW1-2* and *qHKW1-3*) identified in our study. Similarly, Liu et al. [15] identified *qHKW1-4*, spanning a physical interval of 26,697,162–43,754,904 bp, with an LOD value of 6.82 and a PVE of 12.23%. In the present study, two QTLs, *qHKW1-2* and q*HKW1-3* were identified on chromosome 1 and were tightly linked to HKW, with LOD values ranging from 3.88 to 4.34, and a PVE of 6.2–6.3%. One SNP (SNP-25750352) identified during GWAS was significantly associated with HKW and was located within the intervals of these QTLs, spanning a physical distance of 16,415,666–37,617,674 bp. Comparative analysis revealed that these QTLs are located within a critical region on chromosome 1, that overlaps with the intervals reported in earlier studies [15, 29]. The SNP co-located within these QTL intervals may represent a novel candidate locus associated with HKW, offering an important clue for future investigations into the genetic regulation of HKW in maize.

Genetic basis of HKW in maize

This study identified a significant QTL for HKW in pop3 (CML373 × Ye107), with haplotype analysis showing considerable variation in HKW-related haplotypes across the five subpopulations. HKW, a crucial determinant of kernel yield, is a quantitative trait regulated by multiple genes and influenced by environmental factors [1, 2]. Previous research has shown that HKW in maize is controlled by multiple genes, each typically having a small effect. Over the past three decades, approximately 1,920 QTLs associated with kernel size and kernel weight have been reported across diverse maize populations and environments, of which 528 were specifically associated with HKW [55]. Despite this large number of identified kernel-related QTLs, results are often been inconsistent, and beneficial alleles are rarely identified. For instance, previous studies have identified multiple QTLs associated with HKW, including five QTLs identified tightly linked to HKW in an $F_{2:3}$ population [56], nine QTLs and five significant SNPs in a RIL population of 204 lines [29], and 15 QTLs in an $F_{2:3}$ population of 270 individuals [15]. In general, yield components, such as (ear row number (ERN), kernel number per row (KNR), and kernel weight (KW) are positively correlated with kernel yield [57]. For instance, Huo et al. [14] found significant correlations between ear length (EL), KNR, ERN and ear weight (EW), while Xie et al. [58] observed strong correlations between ear diameter (ED) and kernel number per ear (KNE), along with moderate correlations with kernel yield.

Kernel development in maize begins with vegetative growth, when the shoot apical meristem (SAM) initiates the formation of leaves and axillary meristems (AMs), which give rise to ear primordia. As the plant transitions to reproductive growth, the SAM and AMs develop into inflorescence meristems (IMs), which eventually form the ear-like inflorescence. The peripheral regions of the IM generate indeterminate spikelet meristems (SPMs), each of which produces two spikelet meristems (SMs). Each SM gives rise to an upper flower meristem (FM) and a basal flower meristem (BM), though the BM usually aborts, and only the upper floret matures into a kernel [59]. Thus, the regulation of ear initiation, size, spikelet morphology, and developmental determinacy is crucial for shaping ear architecture and improving maize yield.

In this study, male parents with low HKW were selected for three key reasons: (1) Crossing low-HKW male parents with high-HKW female parents may produce hybrids with better yield or other agronomic advantages; (2) Male parents with low HKW may harbor favorable alleles that contribute to traits such as stress resistance, adaptability, or quality; (3) Such crosses may support the development of new varieties aligned with specific market demands or breeding objectives.

Screening and identification of candidate genes regulating HKW in maize

A comprehensive analysis of the genetic basis of hundred kernel weight (HKW) in maize was conducted using a MPP consisting of five subpopulations and comprising a total of 813 $F_{2:7}$ RILs. This study integrated QTL mapping, GWAS, and candidate gene identification to uncover the genetic basis underlying HKW in maize. The results highlighted a significant SNP (SNP-25750352), associated with HKW, consistently identified in the 19BS environment and in BLUP values. This SNP is located 9.968 Mb upstream of the candidate gene *Zm00001d028188*. Previous genetic studies have shown that mutations in *ZmBES1*/*BZR1-5* reduced kernel size

(length and width) and HKW in maize, while its overexpression enhances seed length, width, and weight in rice and Arabidopsis [60]. In addition, certain transcription factors, such as *ZmGRAS11*, positively regulate HKW by controlling the expansion of maize [61].

To further explore the genetic basis of HKW, candidate genes were screened within a 50 kb region upstream and downstream of the significant SNPs. Functional annotation of these genes was performed using databases such as MaizeGDB, InterPro, and NCBI, along with information from relevant literatures. The analysis identified four candidate genes, *Zm00001 d028185*, *Zm00001 d028186, Zm00001d028187,* and *Zm00001d028188*, with *Zm00001d028188*, located on chromosome 1, confirmed as the potential candidate gene regulating HKW.

Zm00001d028185 encodes Remorins, a group of plant plasma membrane proteins involved in membrane-cytoskeleton interactions, potentially affecting maize seed development [62]. *Zm00001d028186* encodes pentatricopeptide repeat (PPR) proteins, which are key RNA-binding proteins that regulate RNA stability and protein synthesis, thereby affecting seed size and quality [63]. *Zm00001d028187* encodes O-Glycosyl hydrolases, enzymes that hydrolyze the glycosidic bonds, which may play a significant role in maize seed development [64].

Zm00001d028188 encodes Galacturonosyltransferase 1 (GAUT1), a key enzyme responsible for transferring galacturonic acid into UDP-GalA (uridine 5'-diphosphate galacturonic acid), which is essential for the synthesis of pectin polysaccharides in plant cell walls. This process is fundamental to the synthesis of pectin, an essential component for growth, development, and environmental adaptation [65]. The GAUT gene family plays diverse roles in cotton fiber development, including fiber elongation and fiber thickening [66]. Persson et al. [67] showed that GAUT2 influences the accumulation of glucuronate xylan and hypergalacturonic acid, contributing to the linkage of the xylan polymer to the secondary cell wall and thereby maintaining the structural integrity of secondary cell wall. Caffall et al. [68] further confirmed that that expression of GAUT gene plays key role in cell wall biosynthesis. GAUT1 may affect water absorption, nutrient storage and seed quality by regulating cell wall flexibility and strength, thus affecting the hundred-kernel weight [69]. Further investigation into the functional and regulatory mechanisms of GAUT1 will provide valuable insights for maize improvement. Targeting GAUT1 through gene editing or transgenic approaches could enhance HKW, leading to improved yield and quality. Since this study was tested at two sites, 19DH and 19BS, further studies need to be evaluated in field experiments at more sites to determine the performance of HKW. Also, the findings needs to be further validated by independent studies, populations and genetic backgrounds to validate these genomic regions/candidate gene, before jumping into genetic manipulation for crop improvement efforts. In conclusion, *Zm00001d028188* could be a potential candidate gene involved in the genetic regulation of maize ear and kernel development, and could serve as a promising target in breeding programs aimed at producing higher-yielding and better-quality maize varieties.

Additive and dominant effects on HKW

The genetic control of HKW in maize is complex and polygenic, with individual genes contributing to HKW through cumulative additive effects. Genetic analyses have revealed that the additive effects vary among different loci [15, 17]. Some genes enhance HKW indirectly by improving kernel size, filling efficiency, or developmental processes, thus playing a crucial role in maize breeding. Genes with larger additive effects holds potential to significantly improve HKW [15], providing a theoretical framework for selection breeding. In addition to additive effects, HKW is influenced by dominant effects and environmental factors. In dominant effects, a dominant allele masks the effect of a recessive allele, while environmental effects modulate the trait expression in response to external factors. The interaction between additive, dominant, and environmental factors determines the overall expression of HKW and its adaptability across diverse environments [13]. Therefore, effective breeding strategies must consider these combined effects to achieve stable improvements in hundred-kernel weight. Advances in molecular breeding have enabled more precise analysis of additive effects, providing a theoretical basis for their application in practical breeding. These developments facilitate the improvement of HKW under varying environmental conditions, ultimately supporting higher yield and improved maize quality.

The influence of negative additive effects of QTLs and its biological significance

In this study, we found that multiple QTLs showed negative additive effects on maize HKW traits. For example, in pop1, *qHKW1-1* is located on chromosome 1 and has an additive effect of −1.01. In pop2, *qHKW4-2* is located on chromosome 4 and exhibited a negative additive effect of −0.951, indicating that alleles at these loci reduce HKW. The manifestation of these effects across the five populations may be attributed to several factors: (1) Certain QTLs may have adverse effects on grain weight in specific genetic backgrounds but may be favorable under different conditions. For example, yield-related QTLs may contribute to the tradeoff between grain number and grain weight [30].

(2) Some QTLs may affect grain weight by regulating the grain filling process, which can be constrained by overall plant growth and nutrient distribution [70]. (3) Additionally, these effects may involve gene-environment (G × E) interactions that influence trait stability and performance. For example, QTLs in some crops have shown opposite effects under varying cultivation conditions [71]. Overall, the presence of QTLs with negative additive effects highlights the complexity of HKW regulation in maize and underscores the importance of considering population-specific and environmental contexts in breeding and genetic improvement strategies.

Conclusion

In this study, five RIL subpopulations were developed by crossing tropical and subtropical inbred lines with a temperate inbred line to explore the genetic basis of hundred kernel weight (HKW) in maize. Through combined QTL mapping and GWAS analyses, co-localized loci linked to HKW were identified. QTL mapping using the tropical germplasms led to the discovery of two novel QTLs, *qHKW1-2* and *qHKW1-3*, on chromosome 1. Notably, one significant SNP, SNP-25750352, was co-located within these QTL intervals on chromosome 1 and found in close proximity to the candidate gene *Zm00001 d028188*. This gene encodes GAUT1, a galacturonic acid transferase involved in plant cell wall biosynthesis, especially by transferring galacturonic acid from UDP-GalA to pectin polysaccharides. The identification of SNP-25750352 and *Zm00001 d028188* contributes to a deeper insights into the genetic regulation underlying HKW in maize. The identification of this SNP and candidate gene in the tropical germplasms underscores their significant role in regulating maize HKW. Therefore, exploring tropical maize germplasms to uncover novel genes affecting HKW could be a key strategy for improving maize yield. Further research should focus on functionally validating these candidate genes using gene editing technologies such as CRISPR-Cas9 or RNA interference (RNAi).

Abbreviations

HKW	Hundred-kernel weight
MPP	Multi-parent population
RIL	Recombinant inbred lines
GBS	Genotyping-by-sequencing
SNP	Single nucleotide polymorphisms
GWAS	Genome-wide association study
QTL	Quantitative trait locis
GAUT1	Galacturonosyltransferase 1
LD	Linkage disequilibrium
KW	Kernel width
BLUP	Best Linear Unbiased Prediction
PCR	Polymerase chain reaction
PCA	Principal component analysis
MLM	Mixed linear model
PVE	Phenotypic variation explained
CV	Coefficient of Variation
MAF	Minor allele frequency
KW	Kernel number per row

Acknowledgements

We would like to extend our heartfelt thanks to Tianwang Lu, Mingtao Cao, Yangtao He and others who contributed to this research. Their invaluable support, expertise, and efforts were crucial to the success of the study. We also express our gratitude to the funding agencies for their generous support for this project.

Authors' contributions

M.F.S. was responsible for conceptualization, data collection and analysis, original draft writing, and revision. F.Y.J. was responsible for project administration, data collection and analysis. R.K.S. and B.I. were responsible for manuscript review and editing. J.C.S. was responsible for data collection. X.M.F. was responsible for conceptualization, funding acquisition, project administration, resource management, supervision, and manuscript review and editing. All authors read and approved the final manuscript.

Funding

This study was supported by Xingdian Talent Support Program of Yunnan Province, the High-level Scientific and Technological Talents and Innovation Team Program of Yunnan Province (202405 AS350030), the National key research and development plan.

Data availability

The raw data of genotyped individuals in this study are available through NCBI database (https://dataview.ncbi.nlm.nih.gov/object/PRJNA1196311?reviewer=ooaf52n2qvreh9l9bgkg8slgo5).

Declarations

Ethics approval and consent to participate

The seeds of the parents and offspring of MPP were collected from Prof. Xingming Fan's group at the Institute of Food Crops, Yunnan Academy of Agricultural Sciences. Appropriate permissions were obtained for all materials used in this study. We complied with all relevant institutional, national, and international guidelines and legislation.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 March 2025 Accepted: 7 May 2025
Published online: 16 May 2025

References

1. Xiao Y, Liu H, Wu L, Warburton M, Yan J. Genome-wide Association Studies in Maize: Praise and Stargaze. Mol Plant. 2017;10(3):359–74.
2. Zhang Z, Liu Z, Hu Y, Li W, Fu Z, Ding D, Li H, Qiao M, Tang J. QTL analysis of Kernel-related traits in maize using an immortalized F2 population. PLoS ONE. 2014;9(2): e89645.
3. Hu J, Guo C, Wang B, Ye J, Liu M, Wu Z, Xiao Y, Zhang Q, Li H, King GJ, et al. Genetic Properties of a Nested Association Mapping Population Constructed With Semi-Winter and Spring Oilseed Rapes. Front Plant Sci. 2018;9:1740.
4. Chen L, Li C, Li Y, Song Y, Zhang D, Wang T, Li Y, Shi Y. Quantitative trait loci mapping of yield and related traits using a high-density genetic map of maize. Mol Breed. 2016;36(9):134.
5. Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, et al. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat Genet. 2011;43(12):1266–9.

6. Tello-Ruiz M, Sen T, Dharmawardhana P, Ren L, Schaeffer M, Naithani S, Amarasinghe V, Thomason J, Harper L, Gardiner J, et al. Maize Metabolic Network Construction and Transcriptome Analysis. The Plant Genome. 2013;6:1–12.
7. Raihan MS, Liu J, Huang J, Guo H, Pan Q, Yan J. Multi-environment QTL analysis of grain morphology traits and fine mapping of a kernel-width QTL in Zheng58 × SK maize population. Theor Appl Genet. 2016;129(8):1465–77.
8. Li C, Li Y, Sun B, et al. Quantitative trait loci mapping for yield components and kernel-related traits in multiple connected RIL populations in maize. Euphytica. 2013;193:303–16. https://doi.org/10.1007/s10681-013-0901-7.
9. Ma X, Tang J, Teng W, Yan J, Meng YJ, Li J. Epistatic interaction is an important genetic basis of grain yield and its components in maize. Mol Breeding. 2007;20:41–51.
10. Qin, W., Li, Yx., Wu, X. et al. Fine mapping of qKL1.07, a major QTL for kernel length in maize. Mol Breeding. 2016;36:8. https://doi.org/10.1007/s11032-015-0419-x.
11. Su C, Wang W, Gong S, Zuo J, Li S, Xu S. High Density Linkage Map Construction and Mapping of Yield Trait QTLs in Maize (Zea mays) Using the Genotyping-by-Sequencing (GBS) Technology. Front Plant Sci. 2017;8:706.
12. Yang C, Zhang L, Jia A, Rong T. Identification of QTL for maize grain yield and kernel-related traits. J Genet. 2016;95(2):239–47.
13. Alvarez Prado S, López CG, Senior ML, Borrás L. The genetic architecture of maize (Zea mays L.) kernel weight determination. G3-Genes Genomes Genetics. 2014; 4(9):1611–1621.
14. Huo D, Ning Q, Shen X, Liu L, Zhang Z. QTL Mapping of Kernel Number-Related Traits and Validation of One Major QTL for Ear Length in Maize. PLoS ONE. 2016;11(5): e0155506.
15. Liu Y, Wang L, Sun C, Zhang Z, Zheng Y, Qiu F. Genetic analysis and major QTL detection for maize kernel size and weight in multi-environments. Theor Appl Genet. 2014;127(5):1019–37.
16. Peng B, Li Y, Wang Y, Liu C, Liu Z, Tan W, Zhang Y, Wang D, Shi Y, Sun B, et al. QTL analysis for yield components and kernel-related traits in maize across multi-environments. Theor Appl Genet. 2011;122(7):1305–20.
17. Wang G, Zhao Y, Mao W, Ma X, Su C. QTL Analysis and Fine Mapping of a Major QTL Conferring Kernel Size in Maize (Zea mays). Front Genet. 2020;11: 603920.
18. Zhang Z, Wu X, Shi C, Wang R, Li S, Wang Z, Liu Z, Xue Y, Tang G, Tang J. Genetic dissection of the maize kernel development process via conditional QTL mapping for three developing kernel-related traits in an immortalized F2 population. MOLECULAR GENETICS AND GENOMICS. 2016;291(1):437–54.
19. Liu C, Zhou Q, Dong L, Wang H, Liu F, Weng J, Li X, Xie C. Genetic architecture of the maize kernel row number revealed by combining QTL mapping using a high-density genetic map and bulked segregant RNA sequencing. BMC Genomics. 2016;17(1):915.
20. Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genet. 2014;10(9): e1004573.
21. Liu L, Du Y, Huo D, et al. Genetic architecture of maize kernel row number and whole genome prediction. Theor Appl Genet. 2015;128(11):2243–54. https://doi.org/10.1007/s00122-015-2581-2.
22. Xue S, Bradbury PJ, Casstevens T, Holland JB. Genetic Architecture of Domestication-Related Traits in Maize. Genetics. 2016;204(1):99–113.
23. Xiao Y, Tong H, Yang X, Xu S, Pan Q, Qiao F, Raihan MS, Luo Y, Liu H, Zhang X, et al. Genome-wide dissection of the maize ear genetic architecture using multiple populations. New Phytol. 2016;210(3):1095–106.
24. Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J, et al. Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biol. 2015;16(1):167.
25. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet. 2010;42(11):961–7.
26. Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, et al. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet. 2011;43(2):163–8.
27. Flint-Garcia SA, Thornsberry JM. Buckler ESt: Structure of linkage disequilibrium in plants. Annu Rev Plant Biol. 2003;54:357–74.
28. Beló A, Zheng P, Luck S, Shen B, Meyer DJ, Li B, Tingey S, Rafalski A. Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. MOLECULAR GENETICS AND GENOMICS. 2008;279(1):1–10.
29. Hao D, Xue L, Zhang Z, Cheng Y, Chen G, Zhou G, Li P, Yang Z, Xu C. Combined linkage and association mapping reveal candidate loci for kernel size and weight in maize. Breed Sci. 2019;69(3):420–8.
30. Liu M, Tan X, Yang Y, Liu P, Zhang X, Zhang Y, Wang L, Hu Y, Ma L, Li Z, et al. Analysis of the genetic architecture of maize kernel size traits by combined linkage and association mapping. Plant Biotechnol J. 2020;18(1):207–21.
31. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009;326(5956):1112–5.
32. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012;90(1):7–24.
33. Yu J, Buckler ES. Genetic association mapping and genome organization of maize. Curr Opin Biotechnol. 2006;17(2):155–60.
34. Yu J, Holland JB, McMullen MD, Buckler ES. Genetic design and statistical power of nested association mapping in maize. Genetics. 2008;178(1):539–51.
35. Wu Y, San Vicente F, Huang K, Dhliwayo T, Costich DE, Semagn K, Sudha N, Olsen M, Prasanna BM, Zhang X, et al. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theor Appl Genet. 2016;129(4):753–65.
36. Edmeades GO, Trevisan W, Prasanna BM, Campos H. Tropical Maize (Zea mays L.). Cham:Springer. In: Genetic Improvement of Tropical Crops; 2017. https://doi.org/10.1007/978-3-319-59819-2_3.
37. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. ArXiv e-prints .2014;arXiv:1406.
38. Stewart CN Jr, Via LE. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques. 1993;14(5):748–50.
39. Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE. 2012;7(2): e32253.
40. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
41. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, et al. Improved maize reference genome with single-molecule technologies. Nature. 2017;546(7659):524–7.
42. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16): e164.
43. Mussmann SM, Douglas MR, Chafin TK, Douglas ME. ADMIXPIPE: population analyses in ADMIXTURE for non-model organisms. BMC Bioinformatics. 2020;21(1):337.
44. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, et al. rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. Genomics Proteomics Bioinformatics. 2021;19(4):619–28.
45. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35(10):1786–8.
46. Lu Y, Shah T, Hao Z, Taba S, Zhang S, Gao S, Liu J, Cao M, Wang J, Prakash AB, et al. Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. PLoS ONE. 2011;6(9): e24861.
47. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44(7):821–4.
48. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, et al. Mixed linear model approach adapted for genome-wide association studies. Nat Genet. 2010;42(4):355–60.
49. Strable J, Wallace JG, Unger-Wallace E, Briggs S, Bradbury PJ, Buckler ES, Vollbrecht E. Maize YABBY Genes drooping leaf1 and drooping leaf2 Regulate Plant Architecture. Plant Cell. 2017;29(7):1622–41.

50. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scand J Stat. 1979;6(2):65–70.
51. Zeng ZB. Precision mapping of quantitative trait loci. Genetics. 1994;136(4):1457–68.
52. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994;138(3):963–71.
53. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263–5.
54. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989;121(1):185–99.
55. Baye W, Xie Q, Xie P. Genetic Architecture of Grain Yield-Related Traits in Sorghum and Maize. Int J Mol Sci. 2022;23(5):2405. https://doi.org/10.3390/ijms23052405. Published 2022 Feb 22.
56. Zhao Y, Su C. Mapping quantitative trait loci for yield-related traits and predicting candidate genes for grain weight in maize. Sci Rep. 2019;9(1):16112.
57. Li M, Guo X, Zhang M, Wang X, Zhang G, Tian Y, Wang Z. Mapping QTLs for grain yield and yield components under high and low phosphorus treatments in maize (Zea mays L.). Plant Science. 2010;178:454–62.
58. Lu M, Xie C, Li X-H, Hao Z-F, Li M-S, Weng J, Zhang D-G, Bai L, Zhang S-H. Mapping of quantitative trait loci for kernel row number in maize across seven environments. Mol Breeding. 2011;28:143–52.
59. Li M, Zhong W, Yang F, Zhang Z. Genetic and Molecular Mechanisms of Quantitative Trait Loci Controlling Maize Inflorescence Architecture. Plant Cell Physiol. 2018;59(3):448–57.
60. Sun F, Ding L, Feng W, Cao Y, Lu F, Yang Q, Li W, Lu Y, Shabek N, Fu F, et al. Maize transcription factor ZmBES1/BZR1-5 positively regulates kernel size. J Exp Bot. 2021;72(5):1714–26.
61. Li Y, Ma S, Zhao Q, Lv D, Wang B, Xiao K, Zhu J, Li S, Yang W, Liu X, et al. ZmGRAS11, transactivated by Opaque2, positively regulates kernel size in maize. J Integr Plant Biol. 2021;63(12):2031–7.
62. Reymond P, Kunz B, Paul-Pletzer K, Grimm R, Eckerskorn C, Farmer EE. Cloning of a cDNA encoding a plasma membrane-associated, uronide binding phosphoprotein with physical properties similar to viral movement proteins. Plant Cell. 1996;8(12):2265–76.
63. Nakamura T, Yagi Y, Kobayashi K. Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. Plant Cell Physiol. 2012;53(7):1171–9.
64. Vlodavsky I, Goldshmidt O, Zcharia E, Metzger S, Chajek-Shaul T, Atzmon R, Guatta-Rangini Z, Friedmann Y. Molecular properties and involvement of heparanase in cancer progression and normal development. Biochimie. 2001;83(8):831–9.
65. Sterling JD, Atmodjo MA, Inwood SE, Kumar Kolli VS, Quigley HF, Hahn MG, Mohnen D. Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase. Proc Natl Acad Sci USA. 2006;103(13):5236–41.
66. Senmiao FAN, Aiying LIU, Xianyan ZOU, Zhen Z, Qun G, Wankui G, Junwen LI, Gong J, Yuzhen SHI, Xiaoying D et al: Evolution of pectin-synthesis-relevant galacturonosyltransferase gene family and its expression during cotton fiber development; 2021.
67. Persson S, Caffall KH, Freshour G, Hilley MT, Bauer S, Poindexter P, Hahn MG, Mohnen D, Somerville C. The Arabidopsis irregular xylem8 mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. Plant Cell. 2007;19(1):237–55.
68. Caffall KH, Pattathil S, Phillips SE, Hahn MG, Mohnen D. Arabidopsis thaliana T-DNA mutants implicate GAUT genes in the biosynthesis of pectin and xylan in cell walls and seed testa. Mol Plant. 2009;2(5):1000–14.
69. Lund CH, Stenbæk A, Atmodjo MA, Rasmussen RE, Moller IE, Erstad SM, Biswal AK, Mohnen D, Mravec J, Sakuragi Y. Pectin Synthesis and Pollen Tube Growth in Arabidopsis Involves Three GAUT1 Golgi-Anchoring Proteins: GAUT5, GAUT6, and GAUT7. Front Plant Sci. 2020;11: 585774.
70. Huang J, Lu G, Liu L, Raihan MS, Xu J, Jian L, Zhao L, Tran TM, Zhang Q, Liu J, et al. The Kernel Size-Related Quantitative Trait Locus qKW9 Encodes a Pentatricopeptide Repeat Protein That Aaffects Photosynthesis and Grain Filling. Plant Physiol. 2020;183(4):1696–709.
71. El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MGM. Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends Plant Sci. 2014;19(6):390–8.

Publisher's Note